POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Ingegneria Gestionale

Tesi di Laurea Magistrale

Sistema bancario e intelligenza artificiale: sviluppo di un processo di risk management, basato sui modelli di albero decisionale e rete neurale, per l'analisi del rischio di credito applicato al settore automotive italiano

Relatore
prof. Carlo Cambini

Candidato
Simone Bugetti

Anno Accademico 2021-2022

Alla mia famiglia

Sommario

Il settore finanziario rappresenta un settore di business estremamente complesso: una qualsiasi istituzione finanziaria che si trova a svolgere la propria operatività in questo ambiente deve necessariamente dimostrare di essere in grado di risolvere questa complessità, combinando la gestione dei rischi finanziari con il rispetto dei vincoli esterni.

A questo proposito, nel corso del tempo, le istituzioni finanziarie hanno effettuato ingenti investimenti, sia in risorse umane che tecnologiche, per rinnovare costantemente la propria capacità di gestione, nel tentativo di sviluppare modelli di misurazione dei rischi finanziari sempre più affidabili.

Le innovazioni adottate, qualificando e quantificando in misura migliore il grado di rischio associato alle diverse esposizioni creditizie, hanno consentito alle istituzioni finanziarie di estendere in maniera più efficace ed efficiente la possibilità complessiva di assumere rischio e il conseguente assorbimento di capitale.

In questo contesto, si inserisce il presente lavoro di tesi, che ha come obiettivo quello di progettare un vero e proprio processo di risk management per l'analisi del rischio di credito, basato sui principali modelli di intelligenza artificiale, quali alberi decisionali e reti neurali, esplorandone l'applicazione in uno dei più grandi settori industriali italiani, ovvero il settore automotive.

L'elaborato si suddivide in tre parti principali, che compongono un percorso, decisamente ambizioso e stimolante, con la finalità di sviluppare un modello di credit scoring per la stima della probabilità di insolvenza.

La prima parte, più descrittiva, contiene due capitoli e specifica il panorama di riferimento, illustrando, da un lato, la regolamentazione bancaria, da Basilea I a Basilea IV, con la spiegazione dei passaggi storici più importanti, unitamente agli emendamenti più significativi, e, dall'altro, il rischio di credito, con un focus sulle componenti fondamentali, quali perdita attesa e perdita inattesa.

La seconda parte comprende due capitoli e consente di avvicinarsi al problema oggetto della trattazione, descrivendo i modelli di credit scoring e introducendo il concetto di intelligenza artificiale, con la precisazione dei modelli, considerati allo stato dell'arte, di albero decisionale e rete neurale.

La terza parte, più sperimentale, contiene tre capitoli e descrive l'intera progettazione del modello di credit scoring basato su tecniche di intelligenza artificiale.

Per prima cosa, si comincia dalla costruzione del dataset di analisi, definito utilizzando i dati di bilancio delle imprese del settore automotive presenti nei database AIDA e ORBIS con il massimo orizzonte temporale disponibile.

In seguito, si prosegue con lo sviluppo, adottando Python come linguaggio di programmazione, del modello di albero decisionale per il processo di feature selection, oltre che del modello di rete neurale per la stima della probabilità di insolvenza.

In ultima istanza, si termina con la revisione dei risultati ottenuti, tramite il confronto con i modelli statistici di analisi discriminante e regressione logistica.

In questo modo, si dimostra come l'intelligenza artificiale possa essere applicata, anche in ambito economico-finanziario, per costruire modelli di credit scoring per la gestione del rischio di credito, in grado potenzialmente di diventare in un futuro, anche prossimo, lo standard de facto di tutto il settore finanziario, sostituendo i modelli passati e presenti impiegati nel corso della storia.

Ringraziamenti

Prima di procedere con la trattazione, mi è doveroso riservare uno spazio dell'elaborato per dedicare qualche parola a tutte le persone che, con il loro aiuto, mi sono state vicine in questo intenso percorso di crescita personale e professionale.

Ringrazio innanzitutto il mio relatore, il prof. Cambini, per aver sempre reagito con estrema disponibilità e tempestività nei confronti di ogni mia richiesta, accompagnandomi, con opportuni suggerimenti, lungo tutto il processo di stesura del lavoro di tesi.

Ringrazio anche tutti i professori che ho avuto l'onore e il piacere di incontrare durante il corso di Laurea Magistrale in Ingegneria Gestionale presso il Politecnico di Torino, per la passione, le conoscenze e le competenze trasmesse durante tutto il percorso accademico.

Un ringraziamento speciale ai miei genitori, Emanuela e Maurizio, per avermi sempre sostenuto e incoraggiato, sia nei momenti buoni che, soprattutto, nei momenti meno buoni, consentendomi di inseguire i miei obiettivi e permettendomi di arrivare fino a questo punto.

Un pensiero affettuoso anche a mio fratello Alessio, per avermi supportato e sopportato in modo costante nello svolgimento del percorso di studi, aiutandomi, con preziosi consigli, a celebrare i momenti di felicità e a superare i momenti di difficoltà.

Un grazie di cuore ai miei colleghi, alcuni dei quali diventati amici nel corso del tempo, per aver condiviso insieme un percorso così importante, con la possibilità di imparare sempre qualcosa di nuovo, giorno dopo giorno, in tutte le occasioni passate insieme.

Un grazie anche ai miei amici storici, ormai conoscenti da una vita, per tutti gli istanti di svago, caratterizzati da divertimento e spensieratezza, trascorsi insieme in questi anni.

Grazie infinite a tutti, senza di voi sarebbe stato sicuramente tutto più complicato.

Come ultima cosa, ringrazio me stesso per questo piccolo grande traguardo raggiunto, ricordando tutti i vari sacrifici portati avanti con grande determinazione e perseveranza per il conseguimento di questo tanto atteso quanto desiderato risultato.

Indice

Elenco delle tabelle 10

Elenco delle figure 11

I Prima Parte **13**

1 Regolamentazione bancaria 15
1.1 Introduzione . . . 15
1.2 Motivi della regolamentazione bancaria . . . 16
1.3 Comitato di Basilea . . . 18
1.4 Principi di regolamentazione . . . 19
1.5 Regolamentazione prima di Basilea I . . . 20
1.6 Basilea I (1988) . . . 22
1.6.1 Coefficiente patrimoniale . . . 22
1.6.2 Ponderazioni per il rischio . . . 23
1.6.3 Capitale regolamentare . . . 23
1.6.4 Limiti di Basilea I . . . 26
1.7 Basilea II (2004) . . . 28
1.7.1 Pillar I: Capital Requirement . . . 30
1.7.2 Pillar II: Supervisory Review . . . 34
1.7.3 Pillar III: Market Discipline . . . 35
1.8 Basilea III (2010) . . . 37
1.8.1 Definition of Capital . . . 39
1.8.2 Misure contro la prociclicità . . . 41
1.8.3 Leva finanziaria . . . 42
1.8.4 Misure per la liquidità . . . 44
1.8.5 Rischio di mercato e di controparte . . . 47
1.8.6 Banche sistemiche (G-SIBs, S-SIIs, SIFIs, D-SIBs, O-SIIs) . . . 48
1.9 Basilea IV (2017) . . . 49
1.9.1 Principali modifiche . . . 49
1.10 Normativa sulla risoluzione . . . 51
1.11 Considerazioni finali . . . 53

2 Rischio di credito 55
2.1 Introduzione . . . 55
2.2 Definizione di rischio di credito . . . 55
2.3 Tipologie di rischio di credito . . . 57
2.4 Componenti del rischio di credito . . . 58
2.5 Perdita attesa . . . 59
2.5.1 Exposure at Default (EAD) . . . 60
2.5.2 Probability of Default (PD) . . . 62
2.5.3 Loss Given Default (LGD) . . . 62
2.5.4 Relazione Probability of Default e Loss Given Default . . . 70
2.6 Perdita inattesa . . . 71
2.6.1 Value at Risk (VaR) . . . 71

II Seconda Parte **75**

3 Credit scoring 77
3.1 Introduzione . . . 77
3.2 Analisi univariata . . . 78
3.3 Analisi multivariata . . . 80
3.3.1 Analisi discriminante . . . 80
3.3.2 Modelli di regressione . . . 83
3.4 Modelli di intelligenza artificiale . . . 88
3.5 Limiti dei modelli di credit scoring . . . 88

4 Intelligenza artificiale 91
4.1 Introduzione . . . 91
4.2 Storia dell'intelligenza artificiale . . . 92
4.3 Concetti di machine learning e deep learning . . . 95
4.4 Alberi decisionali . . . 96
4.4.1 Architettura . . . 96
4.4.2 Apprendimento . . . 97
4.4.3 Criteri decisionali . . . 98
4.4.4 Ensemble . . . 101
4.4.5 Caratteristiche principali . . . 103
4.5 Reti neurali . . . 104
4.5.1 Neurone biologico . . . 104
4.5.2 Neurone artificiale . . . 106
4.5.3 Funzioni di attivazione . . . 109
4.5.4 Progettazione architettura . . . 119
4.5.5 Addestramento rete neurale . . . 121
4.5.6 Algoritmo error back-propagation . . . 122
4.5.7 Metodo gradient descent . . . 127
4.5.8 Caratteristiche principali . . . 128

III Terza Parte **131**

5 Analisi industria automotive 133
5.1 Introduzione . . . 133
5.2 Settore automotive . . . 133
5.2.1 Scenario internazionale . . . 134
5.2.2 Scenario italiano . . . 137
5.2.3 Database AIDA e ORBIS (Bureau Van Dijk) . . . 137
5.2.4 Download dei dati di bilancio . . . 138
5.2.5 Analisi dei dati di bilancio . . . 139
5.2.6 Inizializzazione database . . . 148
5.2.7 Pulizia e correzione dei dati . . . 152
5.2.8 Pre-processing . . . 153

6 Progettazione modello 163
6.1 Introduzione . . . 163
6.2 Train-Valid-Test split . . . 163
6.3 Metriche di performance . . . 165
6.4 Scelta delle variabili . . . 167
6.4.1 Implementazione . . . 168
6.4.2 Albero decisionale . . . 169
6.4.3 Ensemble . . . 172
6.4.4 Indicatori . . . 179
6.5 Modello di credit scoring . . . 179
6.5.1 Implementazione . . . 182
6.5.2 Rete neurale . . . 183
6.5.3 Commento . . . 189
6.5.4 Confronto con modelli statistici . . . 194

7 Conclusioni 201
7.1 Introduzione . . . 201
7.2 Riflessioni conclusive . . . 201

Elenco delle tabelle

1.1 Coefficienti di ponderazione previsti da Basilea I (approccio standard) . . . 31
2.1 Principali determinanti della Loss Given Default 65
4.1 Caratteristiche principali alberi decisionali 104
4.2 Caratteristiche principali reti neurali 129
5.1 Descrizione settore automotive (codice ATECO 29) 134
5.2 Elenco indicatori di bilancio impiegati per la costruzione dei modelli 151
5.3 Analisi di correlazione tra indicatori di bilancio. 157
6.1 Indicatori di bilancio in ordine di importanza relativa (Random Forest) . . 180
6.2 Indicatori di bilancio in ordine di importanza relativa (XGBoost) 181
6.3 Report tecnico: flag S/A - impresa (analisi di correlazione) 190
6.4 Report tecnico: flag S/A+L - impresa (analisi di correlazione) 191
6.5 Report tecnico: flag S/A - impresa (Random Forest e XGBoost) 192
6.6 Report tecnico: flag S/A+L - impresa (Random Forest e XGBoost) 193
6.7 Confronto dei risultati con analisi discriminante e regressione logistica . . . 197

Elenco delle figure

1.1 Sillogismo alla base della regolamentazione bancaria. 17
1.2 Evoluzione Equity to Asset Ratio nel sistema bancario americano. 21
1.3 Gerarchia delle principali componenti del capitale regolamentare. 26
2.1 Distinzione tra perdita attesa e perdita inattesa. 59
2.2 Distribuzione di probabilità Beta con $\alpha = \beta = 0.5$. 69
2.3 Distribuzione di probabilità delle perdite sui crediti. 73
3.1 Sintesi grafica dell'analisi univariata. 79
3.2 Sintesi grafica dell'analisi discriminante. 81
3.3 Distribuzione di probabilità logistica standard e normale standard. 86
4.1 Storia dell'intelligenza artificiale. 94
4.2 Schema dei concetti di machine learning e deep learning. 95
4.3 Struttura di un albero decisionale. 97
4.4 Entropia di una sorgente binaria. 99
4.5 Impurità di Gini di una sorgente binaria. 100
4.6 Caratteristiche Bagging vs Boosting. 102
4.7 Rappresentazione di un neurone biologico. 105
4.8 Modello di neurone artificiale di W. S. McCulloch e W. H. Pitts. 107
4.9 Architettura di una rete neurale. 108
4.10 Modello matematico di una rete neurale. 109
4.11 Binary Step Function. 110
4.12 Linear Activation Function. 111
4.13 Sigmoid Function. 112
4.14 Tahn Function. 113
4.15 Rectified Linear Unit Function. 114
4.16 Leaky Rectified Linear Unit Function. 115
4.17 Parametric Rectified Linear Unit Function. 115
4.18 Exponential Linear Unit Function. 116
4.19 Gaussian Error Linear Unit Function. 117
4.20 Swish Function. 118
4.21 Softmax Function. 118
4.22 Feed-Forward Neural Network (ANN). 119
4.23 Recurrent Neural Network (RNN). 120
4.24 Convolutional Neural Network (CNN). 121
4.25 Esempio applicazione gradient descent $y = x^2 + y^2 + x - y + 2xy$. 128

5.1 Produzione mondiale ed europea veicoli a motore. 134
5.2 Andamento produzione mondiale veicoli a motore. 135
5.3 Database AIDA. 138
5.4 Margine Primario di Struttura. 140
5.5 Margine Secondario di Struttura. 141
5.6 Leva Finanziaria Book Value. 142
5.7 Posizione Finanziaria Netta. 142
5.8 Current Ratio. 143
5.9 Acid Test. 144
5.10 Interest Coverage Ratio. 145
5.11 Return on Sales. 146
5.12 Return on Equity. 146
5.13 Return on Investment. 147
5.14 Return on Assets. 148
5.15 Distribuzione imprese sane e imprese anomale. 155
5.16 Matrice di correlazione tra indicatori di bilancio. 158
6.1 Rappresentazione grafica del concetto di Train-Valid-Test split. 164
6.2 ROC curve. 166
6.3 Confusion matrix. 166
6.4 Albero decisionale (flag di status S/A - impresa). 170
6.5 Albero decisionale (flag di status S/A+L - impresa). 171
6.6 Albero decisionale (testing set). 172
6.7 Albero decisionale (database). 172
6.8 Random Forest (testing set). 174
6.9 Random Forest (database). 174
6.10 Processo di feature selection utilizzando Random Forest. 175
6.11 XGBoost (testing set). 177
6.12 XGBoost (database). 177
6.13 Processo di feature selection utilizzando XGBoost. 178
6.14 Training-Validation-Testing split. 183
6.15 Sintesi grafica del processo di addestramento di una rete neurale. 187
6.16 Modelli di intelligenza artificiale ($1A^{RF}$ e $1F^{RF}$). 198
6.17 Modello di analisi discriminante. 199
6.18 Modello di regressione logistica. 199

Parte I

Prima Parte

Capitolo 1

Regolamentazione bancaria

1.1 Introduzione

Il settore finanziario e, nello specifico, il settore bancario, rappresenta un settore di business estremamente complesso. Considerando l'importanza per il corretto funzionamento del sistema economico globale, una banca che si trova a operare in questo ambiente, deve necessariamente dimostrare di essere in grado di risolvere questa complessità, combinando la gestione di una molteplicità di rischi finanziari e il rispetto di una serie di vincoli esterni.

In generale, una banca che progetta un sistema di risk management con l'obiettivo di valutare le attività a rischio e misurare l'assorbimento di capitale, deve obbligatoriamente tenere in considerazione i vincoli imposti dalla regolamentazione bancaria.

In questo momento, il principale vincolo regolamentare è costituito dalla presenza di requisiti patrimoniali obbligatori, inizialmente introdotti e successivamente revisionati dal Comitato di Basilea, che definiscono l'impianto di base della regolamentazione bancaria, armonizzando il processo di misurazione dei rischi e di quantificazione del capitale. Nel corso del tempo, la regolamentazione bancaria è stata recepita in modo progressivo dalle Autorità di Vigilanza di oltre 150 Paesi, tra cui l'Italia e gli altri Paesi dell'Unione Europea.

Il lavoro realizzato dal Comitato di Basilea risulta essere particolarmente importante, dal momento che la formulazione di requisiti patrimoniali obbligatori ha segnato il primo passo verso la ricerca di standard internazionali in materia di regolamentazione bancaria. In realtà, la normativa regolamentare, pur mantenendo nelle diverse revisioni uno schema metodologico generalmente invariato, ha seguito un processo evolutivo e migliorativo, attraverso la ricerca e lo sviluppo di regole sempre più efficaci ed efficienti.

A questo proposito, il Comitato di Basilea ha progressivamente effettuato numerosi interventi di modifica, scegliendo di conservare e, se necessario, migliorare le idee proposte originariamente, con la finalità di allineare le nuove regole alle best practice mondiali in termini di analisi e controllo dei rischi. Per valutare l'impatto di queste correzioni sul sistema bancario e, pertanto, comprendere al meglio la struttura regolamentare attuale, è necessario approfondire l'evoluzione che ha caratterizzato la regolamentazione bancaria nel corso della storia, dagli inizi fino ad oggi.

Di conseguenza, in questo capitolo, si riassume la regolamentazione bancaria, cercando, da una parte, di evidenziarne le motivazioni sottostanti, e, dall'altra, di illustrarne, con le dovute spiegazioni, i passaggi storici più importanti, in modo da cominciare con una breve descrizione dell'ambiente economico-finanziario prima che la regolamentazione bancaria moderna facesse prepotentemente il suo ingresso all'interno del panorama internazionale, e, in seguito, focalizzare l'attenzione sui relativi emendamenti più significativi, iniziando con Basilea I (1988) e passando da Basilea II (2004) a Basilea III (2010), per terminare il ragionamento con Basilea IV (2017), ultima versione della regolamentazione bancaria disponibile e consultabile al momento del presente lavoro di tesi.

In ultima analisi, durante questo percorso, si dedica uno spazio anche alla spiegazione delle normative più rilevanti a livello internazionale sulla risoluzione delle banche, oltre che uno sguardo prospettico sul futuro della regolamentazione bancaria stessa.

1.2 Motivi della regolamentazione bancaria

La regolamentazione bancaria si pone come principale obiettivo quello di assicurare che le banche possano sempre disporre di una quantità di capitale sufficientemente adeguata a fronteggiare i rischi consapevolmente assunti. In questo modo, le Autorità di Vigilanza sperano di rendere minima la probabilità che una qualsiasi banca possa fallire, nel tentativo di creare un sistema bancario stabile, con piena fiducia da parte di famiglie e imprese.

Per lavorare alla stabilità dell'ambiente economico-finanziario, le Autorità di Vigilanza hanno introdotto il concetto di capitale regolamentare, applicando una strategia facilmente riassumibile in un sillogismo fondamentale, che rappresenta il pilastro dei ragionamenti effettuati per stabilire le principali disposizioni in materia di regolamentazione bancaria.

Il sillogismo è composto da alcuni semplici passaggi logici, come visibile in figura 1.1. Prima di tutto, il traguardo da raggiungere risulta essere la stabilità macroeconomica dell'intero sistema economico, poiché un sistema economico stabile è in grado di funzionare al meglio, in modo da ottimizzare le principali variabili economiche: esprimere la massima crescita del PIL, sostenere la massima occupazione e garantire la minima inflazione.

Per ottenere la stabilità del sistema economico si deve necessariamente assicurare la stabilità, sia dal punto di vista macroeconomico che dal punto di vista microeconomico, del sistema finanziario. La stabilità del sistema finanziario si dimostra un requisito fondamentale per arrivare alla stabilità del sistema economico, in quanto un sistema finanziario instabile conduce inevitabilmente nel tempo verso una instabilità del sistema economico.

A questo punto, per ottenere la stabilità macroeconomica del sistema finanziario, si deve garantire la stabilità microeconomica del sistema finanziario, e, in modo analogo, per ottenere la stabilità microeconomica del sistema finanziario, si deve assicurare la solvibilità di ogni singola banca. In questo ambito, per raggiungere la solvibilità di ogni singola banca e, quindi, la stabilità dell'intero sistema bancario, le Autorità di Vigilanza hanno deciso di imporre obbligatoriamente il possesso di un requisito minimo di capitale.

In poche parole, il sillogismo può essere riassunto con semplicità nel modo seguente: se ogni banca ha capitale a sufficienza, ogni banca è solvibile, se ogni banca è solvibile, ogni banca è stabile e, infine, se ogni banca è stabile, allora è stabile il sistema bancario nel suo complesso, arrivando a dimostrare il completamento dell'obiettivo iniziale.

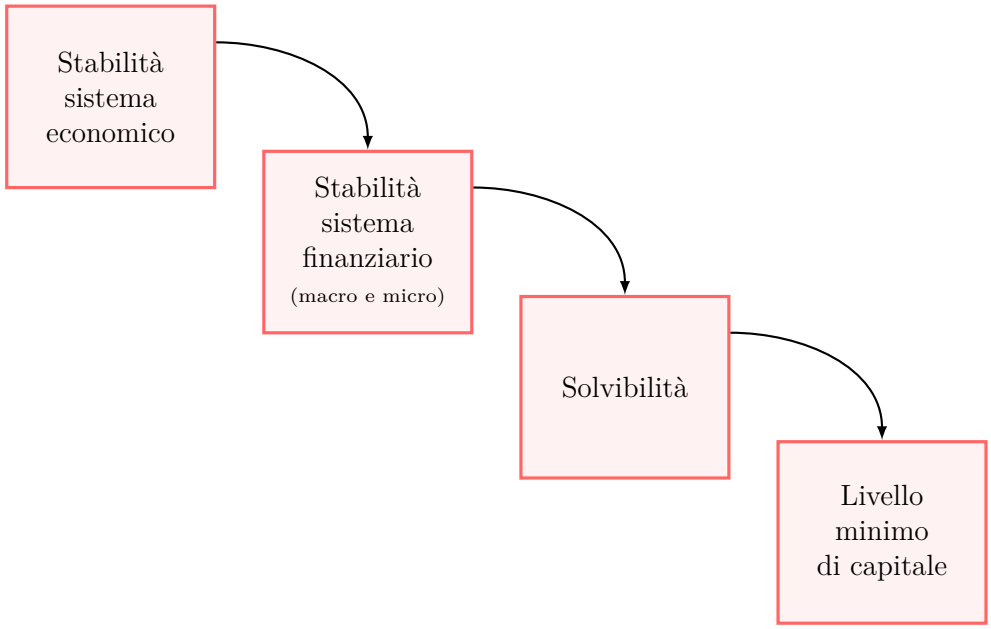


Figura 1.1. Sillogismo alla base della regolamentazione bancaria.

Questa impostazione è alla base del pensiero delle Autorità di Vigilanza in termini di regolamentazione bancaria, dal momento che attraverso una regola amministrativa microeconomica è possibile ottenere un grande risultato macroeconomico. Infatti, con il raggiungimento della stabilità del sistema finanziario, è ragionevole attendersi anche la stabilità dell'economia reale e, come risultato, la stabilità dell'intero sistema economico.

A questo riguardo, è doveroso effettuare una precisazione: il sillogismo rappresenta una condizione necessaria ma non sufficiente per la stabilità dell'intero sistema economico. I passaggi logici presentati alla base del ragionamento sono imprescindibili e indispensabili, ma presentano un importante problema strutturale, poiché non permettono di catturare completamente il rischio sistemico, essendo in grado di spiegare soltanto parzialmente il comportamento reale del sistema economico.

In questo modello, pertanto, è presente una discrasia, considerando che non necessariamente la stabilità di un sistema complesso e integrato può essere ottenuta dalla semplice stabilità dei singoli elementi che compongono il sistema. In particolare, in presenza di rischio sistemico, si potrebbe verificare una situazione in cui la soluzione ottimale per il singolo elemento potrebbe non essere la soluzione ottimale per il sistema nel suo complesso, compromettendo potenzialmente la validità del modello.

Tuttavia, nonostante questa possibile distorsione, i requisiti patrimoniali obbligatori hanno avuto un impatto molto significativo sul sistema bancario internazionale, svolgendo un ruolo essenziale nel determinare le politiche strategiche delle singole banche e nel rafforzare la stabilità del sistema bancario. In altre parole, l'introduzione del concetto di capitale regolamentare ha generato effetti benefici sul sistema economico, con grandi vantaggi sia dal punto di vista microeconomico che dal punto di vista macroeconomico.

In primo luogo, sul piano microeconomico, i requisiti patrimoniali obbligatori hanno fortemente condizionato le linee di sviluppo, le decisioni di investimento e la cultura aziendale delle banche, producendo nuovi modelli gestionali e organizzativi.

In secondo luogo, sul piano macroeconomico, i requisiti patrimoniali obbligatori hanno notevolmente influenzato il paradigma di patrimonializzazione dei diversi sistemi bancari, ottenendo una maggiore omogeneità e uniformità di funzionamento.

Riassumendo, il Comitato di Basilea, con l'implementazione di un vincolo obbligatorio sul capitale regolamentare, ha cercato di fornire gli incentivi corretti per spingere le banche verso una maggiore capitalizzazione e, in questo modo, rafforzare la stabilità del sistema bancario, in modo da allineare quest'ultimo su livelli sostenibili non solo nel breve termine, ma anche nel medio-lungo termine.

1.3 Comitato di Basilea

Il Comitato di Basilea per la Vigilanza Bancaria[1], è un'organizzazione internazionale di consultazione e cooperazione, costituita nel 1974 dai governatori delle Banche Centrali dei Paesi del Gruppo dei Dieci, o G-10, in seguito ad alcuni gravi episodi di crisi verificatisi sul mercato finanziario.

Il Gruppo dei Dieci, al momento della sua costituzione, nel 1962, era composto dai dieci paesi economicamente più sviluppati dell'epoca, ovvero Belgio, Canada, Francia, Germania, Giappone, Italia, Paesi Bassi, Regno Unito, Stati Uniti e Svezia, con l'aggiunta posteriore della Svizzera, unitasi successivamente, nel 1964, come undicesimo membro, mantenendo il nome del gruppo invariato.

Inizialmente, quindi, il Comitato di Basilea riuniva la presenza dei principali Paesi internazionali, caratterizzati da un elevato tasso di sviluppo e industrializzazione. In ogni caso, nel corso del tempo, il Comitato di Basilea, ha progressivamente incrementato il numero dei propri partecipanti, raggiungendo oggi la rappresentanza di 45 istituzioni appartenenti a 28 giurisdizioni diverse.

Attualmente, il Comitato di Basilea è composto da membri provenienti dalla maggioranza dei Paesi internazionali: Arabia Saudita, Argentina, Australia, Belgio, Brasile, Canada, Cina, Corea del Sud, Francia, Germania, Giappone, Hong Kong, India, Indonesia, Italia, Lussemburgo, Messico, Paesi Bassi, Regno Unito, Russia, Singapore, Spagna, Stati Uniti, Sudafrica, Svezia, Svizzera e Turchia. Ogni Paese è rappresentato da esponenti delle Banche Centrali e da eventuali altre Autorità di Vigilanza sul sistema bancario.

Il Comitato di Basilea, che si riunisce regolarmente a Basilea, in Svizzera, sotto il patronato della Banca per i Regolamenti Internazionali[2], è caratterizzato da grandissima credibilità e reputazione internazionale, anche se non dispone di alcun potere legislativo o autorità sovranazionale.

[1] In inglese, *Basel Committee on Banking Supervision* (BCBS).

[2] In inglese, *Bank for International Settlements* (BIS).

Le decisioni in materia di regolamentazione bancaria emanate dal Comitato di Basilea, sotto il profilo formale, non hanno un valore giuridico, ma influenzano in modo determinante le legislazioni dei vari Paesi internazionali, i quali possono decidere autonomamente di recepire nel proprio ordinamento nazionale eventuali modifiche o correzioni.

Queste proposte non rappresentano un vincolo dal punto di vista legale, ma hanno una prestigiosa valenza internazionale, poiché vengono accettate e adottate su base volontaria non solo dai Paesi membri, ma anche dalla maggior parte dei Paesi internazionali che non partecipano direttamente al Comitato di Basilea.

In ogni caso, il compito è quello di definire linee guida e raccomandazioni che risultino universalmente valide e applicabili dalle singole Autorità Politiche e Autorità di Vigilanza nazionali e internazionali, per la costituzione di un sistema bancario sostenibile.

1.4 Principi di regolamentazione

In generale, è possibile evidenziare due approcci principali in materia di regolamentazione.

Un primo approccio considera l'intervento regolamentare dannoso e superfluo per il sistema economico: il mercato è sempre in grado di autoregolarsi e non ha bisogno di alcun tipo di intervento da parte delle Autorità di Governo.

Un secondo approccio, al contrario, considera l'intervento regolamentare non soltanto opportuno per il sistema economico, ma anzi, necessario: il mercato da solo non è in grado di autoregolarsi e conduce inevitabilmente a condizioni non concorrenziali, asimmetrie informative, instabilità e, potenzialmente, crisi sistemiche.

Per il problema generale della regolamentazione, quindi, si possono distinguere due scuole di pensiero, che si differenziano in base alla possibilità e alla tipologia di intervento sul sistema economico, con conseguenze estremamente diversificate su quest'ultimo.

In questo ambito, non è possibile discutere la correttezza del primo o del secondo approccio senza analizzare attentamente il trade-off implicito tra concorrenza e stabilità.

Infatti, al variare della concorrenza e della stabilità, il sistema economico potrebbe essere condotto verso due scenari estremi e completamente opposti: massimizzando la concorrenza a discapito della stabilità, si potrebbe verificare sul mercato una situazione di incertezza e instabilità; viceversa, massimizzando la stabilità a discapito della concorrenza, si potrebbe determinare una situazione di blocco e stallo, senza garanzia di rinnovamento.

Considerando questo trade-off, la regolamentazione ideale deve contemporaneamente risultare efficace nel raggiungere gli obiettivi prefissati ed efficiente nel combinare insieme i benefici e i costi derivanti dalla presenza di specifiche regole sul mercato. L'esistenza della regolamentazione finanziaria e, nel dettaglio, della regolamentazione bancaria, implica la presenza di Autorità di Vigilanza destinate alla verifica del rispetto delle regole imposte, oltre che al loro aggiornamento in base allo stato del sistema economico.

A questo punto, è possibile approfondire il problema della regolamentazione bancaria, in modo da analizzare in modo sintetico e, allo stesso tempo, esaustivo i passaggi storici realmente verificatisi. A tale proposito, è possibile distinguere due impostazioni diverse:

- regolamentazione strutturale;
- regolamentazione prudenziale.

La regolamentazione strutturale impone limiti all'operatività delle banche, in modo da contenere direttamente i rischi che possono essere assunti. In poche parole, con questo tipo di regolamentazione, l'arbitrarietà delle banche viene condizionata dall'imposizione di una serie di restrizioni, come divieti all'entrata, all'espansione geografica, all'offerta di prodotti e servizi specifici e, soprattutto, alla concentrazione dei rischi.

La regolamentazione prudenziale, invece, non impone nessun tipo di limite connesso all'operatività delle banche, che viene lasciata completamente libera e priva di qualsiasi restrizione, ma prescrive delle forme di copertura per i rischi assunti.

Nel corso della storia, il settore bancario è stato inizialmente caratterizzato dalla presenza di una regolamentazione strutturale, decisamente pesante e severa, introdotta come risposta alla crisi finanziaria degli anni '30, con l'obiettivo di contrastare duramente il possibile ripetersi di eventi in grado di sconvolgere il sistema economico mondiale.

Successivamente, a partire dagli anni '80, la regolamentazione strutturale è stata progressivamente abbandonata in favore della regolamentazione prudenziale, che costituisce il modello di riferimento attuale per la formulazione della normativa regolamentare.

In sintesi, la deregolamentazione del settore bancario ha sostanzialmente significato il passaggio dalla prima alla seconda impostazione, con una trasformazione motivata dalla convinzione che la regolamentazione strutturale fosse troppo rigida, con gravi limitazioni alla concorrenza e, dunque, con effetti negativi sull'intero sistema economico.

Per superare queste problematiche, è stato introdotto il modello di regolamentazione prudenziale, che si fonda sulla definizione di capitale regolamentare e, in particolare, sulla disposizione di specifici requisiti patrimoniali, opportunamente valutati in relazione ai rischi assunti. In questo contesto, le banche non presentano nessun limite alla loro operatività e possono scegliere di assumere i rischi che desiderano, ma, in conformità con i rischi a cui sono esposte, devono possedere una quantità minima di capitale sufficiente per assorbire le perdite inattese fino a un certo livello di confidenza.

1.5 Regolamentazione prima di Basilea I

L'Accordo di Basilea del 1988, comunemente chiamato Basilea I, è stato il primo tentativo per fissare uno standard internazionale con cui poter misurare l'adeguatezza del capitale a disposizione delle banche e, per questo motivo, si tratta di una tappa fondamentale per la storia della regolamentazione bancaria. Negli anni immediatamente precedenti, per incrementare ulteriormente il grado di sviluppo del settore bancario e, di conseguenza, del sistema economico nel suo complesso, i problemi da risolvere erano molteplici.

Per prima cosa, nel corso del tempo, si era notata una continua diminuzione del grado di patrimonializzazione dei principali sistemi bancari. Il rapporto tra patrimonio netto e attività di bilancio si era, infatti, progressivamente ridotto per i sistemi bancari più avanzati, passando da un valore particolarmente elevato a fine '800 e inizio '900 fino a un punto di minimo[3] negli anni '40 e '70 del secolo scorso, come visibile in figura 1.2.

[3]Dati ricavati, con alcune rielaborazioni, da Federal Reserve [145], Bureau of the Census [138] e Federal Deposit Insurance Corporation [144].

Il modello di business delle banche, specialmente a seguito della diffusione sul mercato di innovazioni di ingegneria finanziaria, aveva subito numerose trasformazioni, spingendo l'intero sistema bancario verso una ricerca spasmodica di aumentare la redditività. Pertanto, considerando questo andamento, iniziava a essere evidente che una maggiore capitalizzazione delle banche non fosse solo auspicabile, bensì necessaria per garantirne la sopravvivenza e la sostenibilità nel medio-lungo periodo.

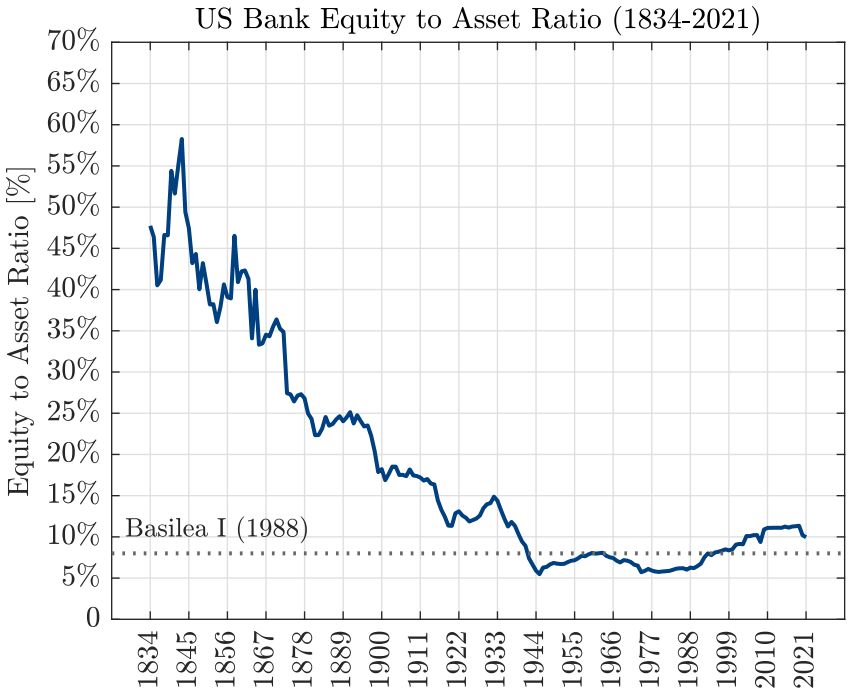


Figura 1.2. Evoluzione Equity to Asset Ratio nel sistema bancario americano.

In aggiunta, prima del 1988, le Autorità di Vigilanza regolavano autonomamente il patrimonio delle banche e decidevano i criteri per determinare l'adeguatezza patrimoniale fissando livelli minimi del rapporto tra patrimonio netto e attività di bilancio. Tuttavia, le definizioni del contenuto del patrimonio netto e le soglie minime del rapporto considerate accettabili, variavano, con profonde differenze, da Paese a Paese, determinando l'applicazione di requisiti patrimoniali in modo più o meno rigoroso.

Come risultato, per le istituzioni finanziarie appartenenti a sistemi bancari diversi, questa disparità determinava, a livello internazionale, profonde differenze concorrenziali. La competizione internazionale tra banche, infatti, non si realizzava in modo uniforme: le banche che operavano in Paesi con requisiti patrimoniali più elastici godevano di grandi vantaggi concorrenziali rispetto alle banche che, contrariamente, operavano in Paesi con requisiti patrimoniali più stringenti (in particolare, le banche giapponesi [105] [124] [132]).

A questo proposito, il desiderio di imporre requisiti patrimoniali omogenei e uniformi su scala internazionale appariva chiaramente come uno dei punti fondamentali per cercare di risolvere, almeno parzialmente, questa problematica.

Basilea I ha cercato, quindi, di rispondere parallelamente a queste esigenze, ponendosi come obiettivo sia quello di rafforzare la solvibilità e la solidità dei diversi sistemi bancari, per scoraggiare l'assunzione di rischi giudicati eccessivi, che quello di superare le distorsioni competitive connesse a diverse normative nazionali, per favorire la creazione di condizioni concorrenziali uniformi e ridurre le disparità competitive.

Nonostante abbia ricevuto numerose critiche a causa di un'eccessiva arbitrarietà e semplicità, questa normativa ha rappresentato un importante passo in avanti verso la regolamentazione bancaria moderna, dal momento che, per la prima volta, ha influenzato in modo decisivo la cultura aziendale delle banche, introducendo un significativo aumento delle risorse dedicate alla comprensione, alla misurazione e alla gestione dei rischi.

1.6 Basilea I (1988)

L'attenzione del Comitato di Basilea, con la formulazione dell'Accordo di Basilea del 1988, si è concentrata sulla necessità di introdurre un sistema di requisiti patrimoniali obbligatori per assicurare la copertura delle perdite inattese fino a un certo livello di confidenza.

1.6.1 Coefficiente patrimoniale

Basilea I richiede alle banche l'obbligo di rispettare in modo dinamico un requisito minimo di capitalizzazione: il capitale regolamentare[4], o *Regulatory Capital* (RC), deve essere almeno pari all'8% delle attività ponderate per il rischio, o *Risk Weighted Assets* (RWA).

In formule:

$$\frac{RC}{RWA} = \frac{RC}{\sum_{i=1}^{n} A_i \cdot w_i} \geq 8\% \ , \tag{1.1}$$

dove:

- RC = capitale regolamentare;
- RWA = attività ponderate per il rischio;
- A_i = attività i-esima;
- w_i = ponderazione per il rischio i-esima.

In sostanza, la normativa introdotta con Basilea I impone il rispetto di un vincolo pari all'8%, definito coefficiente patrimoniale, nel rapporto tra capitale regolamentare e attività ponderate per il rischio.

In realtà, le banche erano obbligate a rispettare un valore del coefficiente patrimoniale leggermente variabile, in base all'adesione o meno a un gruppo di imprese. Più precisamente, le banche che non appartenevano a un gruppo dovevano semplicemente presentare

[4] In alternativa, il capitale regolamentare è chiamato anche patrimonio di vigilanza.

un coefficiente patrimoniale almeno pari all'8%, mentre le banche che appartenevano a un gruppo erano caratterizzate da una maggiore complessità, dovendo soddisfare contemporaneamente due requisiti diversi per il soddisfacimento del coefficiente patrimoniale: un valore pari al 7% a livello individuale e un valore pari all'8% a livello consolidato.

Il rapporto tra patrimonio di vigilanza e attività ponderate per il rischio, chiamato anche con il nome di rapporto di Cooke, o *Cooke ratio*, prendendo spunto da P. Cooke, governatore della Bank of England e primo presidente del Comitato di Basilea, rappresenta la principale novità introdotta con la riforma di Basilea I. In questo calcolo entrano in gioco sia le voci di bilancio, *on-balance sheet*, sia le voci fuori bilancio, *off-balance sheet*, in modo da contribuire alla determinazione delle attività ponderate per il rischio, misurando l'esposizione creditizia complessiva della banca.

1.6.2 Ponderazioni per il rischio

A questo punto, è possibile discutere l'assegnazione dei pesi per le varie poste dell'attivo, che costituisce il punto centrale attorno al quale è stato creato il meccanismo di Basilea I. In particolare, l'assegnazione dei pesi risulta proporzionale al livello di rischiosità. Infatti, per le attività giudicate più rischiose viene assegnato un peso più elevato, richiedendo un requisito patrimoniale più consistente; viceversa, per le attività giudicate meno rischiose viene assegnato un peso più modesto, richiedendo un requisito patrimoniale più limitato. Le diverse voci dell'attivo vengono caratterizzate seguendo tre parametri principali: il grado di liquidità, la natura dei debitori e l'area geografica di residenza.

Le ponderazioni per il rischio risultanti sono le seguenti:

- 0%: per la cassa e i crediti verso Governi e Banche Centrali dei paesi OCSE;
- 10%: per i crediti verso pubblica amministrazione;
- 20%: per i crediti verso banche dei paesi OCSE, banche dei paesi non OCSE con durata residua inferiore a un anno e istituti bancari internazionali;
- 50%: per crediti ipotecari concessi per l'acquisto di immobili residenziali;
- 100%: per tutti gli altri crediti.

1.6.3 Capitale regolamentare

Definita la scala di ponderazione per il rischio, è necessario analizzare in dettaglio un ulteriore aspetto importante, istituito per la prima volta da Basilea I e mantenuto, seppur con qualche modifica, anche nelle versioni successive della regolamentazione bancaria, concentrando l'attenzione sulla definizione di capitale regolamentare. In questo ambito, il Comitato di Basilea distingue il capitale regolamentare in due componenti fondamentali: il patrimonio di base (Tier 1 Capital) e il patrimonio supplementare (Tier 2 Capital). Questa differenziazione è sostanzialmente basata sul concetto di qualità del capitale, che rappresenta la capacità di proteggere i terzi dalle conseguenze, più o meno pericolose, di eventuali perdite impreviste subite dalla banca, come mostrato in figura 1.3.

Il patrimonio di base è composto dalle poste patrimoniali più pregiate e comprende principalmente il capitale azionario versato, che contiene le azioni ordinarie emesse e interamente liberate e le azioni privilegiate irredimibili senza diritto di priorità cumulativo, le riserve palesi, come per esempio la riserva sovrapprezzo azioni, la riserva legale e gli utili accantonati a riserva, e alcuni fondi generali, non collegati direttamente alla copertura di specifiche perdite. Questi elementi, complessivamente, costituiscono l'Upper Tier 1.

A questi fattori, è possibile aggiungere determinati strumenti innovativi di capitale, come le azioni privilegiate, o *preference share*, a condizione che vengano rispettati alcuni requisiti minimi: devono avere natura permanente, non devono presentare condizioni restrittive che permettano agli investitori di conseguire un rimborso anticipato, l'eventuale rimborso anticipato su iniziativa dell'emittente deve essere preventivamente autorizzato dalle Autorità di Vigilanza e non deve avvenire prima di cinque anni dall'emissione, non devono possedere un diritto alla remunerazione differibile, devono essere in grado di assorbire le perdite anche senza l'intervento di una procedura di liquidazione o fallimento e devono essere subordinati a tutti i restanti creditori ordinari e subordinati.

Questi strumenti, inoltre, non possono rappresentare più del 15% del patrimonio di base, anche se l'eventuale eccesso può essere computato nel patrimonio supplementare e, per questo motivo, sono chiamati Lower Tier 1.

Per ottenere il valore corretto del patrimonio di base è necessario dedurre, dalla somma delle voci patrimoniali discusse precedentemente, l'ammontare dell'avviamento e di eventuali altre poste adottate a discrezione delle Autorità di Vigilanza nazionali. In Italia, queste ultime sono rappresentate dalle azioni proprie, dalle immobilizzazioni immateriali, dalle perdite registrate in esercizi precedenti e nell'esercizio corrente, dalle rettifiche di valore calcolate sul portafoglio di negoziazione per motivi di vigilanza, oltre che dai filtri prudenziali negativi del patrimonio di base.

Il patrimonio supplementare, al contrario, è formato dalle poste patrimoniali meno preziose e comprende principalmente le riserve occulte, le riserve da rivalutazione, gli accantonamenti o fondi generali per rischi, gli strumenti ibridi di patrimonializzazione e i prestiti subordinati ordinari.

Le riserve occulte, consentite dalla normativa di alcuni Paesi, sono speciali riserve non evidenziate in bilancio e alimentate con profitti non distribuiti, che devono presentare la stessa solidità di una riserva palese: non devono essere gravate da oneri e devono essere immediatamente libere e disponibili per fronteggiare perdite inattese.

Le riserve da rivalutazione sono riserve legate alla rivalutazione di attività specifiche originariamente registrate al costo storico e possono essere evidenziate in bilancio oppure anche solo latenti, ossia connesse a potenziali plusvalenze non ancora realizzate. In questo caso, le riserve da rivalutazione latenti, tenendo in considerazione eventuali fluttuazioni sfavorevoli dei prezzi di mercato delle attività e il prelievo fiscale connesso a eventuali plusvalenze generabili, possono essere ammesse nel calcolo del patrimonio supplementare per un massimo del 45% del loro valore totale.

Gli accantonamenti generali e i fondi generali per rischi su crediti sono voci del passivo sviluppate per rispondere a perdite future non ancora identificate o identificabili e possono essere utilizzati nel calcolo del patrimonio supplementare fino a un massimo dell'1,25% delle attività ponderate per il rischio.

Gli strumenti ibridi di patrimonializzazione sono strumenti ibridi di capitale-debito, che uniscono insieme elementi peculiari sia delle azioni che del debito e le cui caratteristiche, pur variando in maniera specifica da Paese a Paese, devono in ogni caso rispettare alcuni requisiti minimi: devono essere interamente versati, non garantiti e subordinati a tutti gli altri debiti, non possono essere rimborsati su iniziativa del creditore, non possono essere rimborsati senza la preventiva autorizzazione delle Autorità di Vigilanza, devono poter essere abbattuti a fronte di perdite senza la necessità di una procedura di liquidazione o fallimento e, in caso di diritto a remunerazione periodica non sopprimibile e non riducibile, devono consentire di differire la remunerazione se i profitti della banca non la permettono.

Nel complesso, le riserve occulte, le riserve da rivalutazione, gli accantonamenti o fondi generali per rischi e gli strumenti ibridi di patrimonializzazione contribuiscono a rappresentare, congiuntamente, l'Upper Tier 2.

I prestiti subordinati ordinari costituiscono titoli di debito non garantiti, di durata uguale ad almeno cinque anni e il cui rimborso risulta subordinato, in caso di problemi relativi alla banca in esame, al totale e completo soddisfacimento di tutti i creditori. La principale differenza rispetto agli strumenti ibridi di patrimonializzazione è rappresentata dal fatto che i prestiti subordinati ordinari non possono essere abbattuti in presenza di perdite, senza che questo implichi obbligatoriamente la messa in liquidazione della banca.

A questo riguardo, i prestiti subordinati ordinari non possono superare il 50% del patrimonio di base e sono chiamati Lower Tier 2.

Analogamente a quanto fatto per il patrimonio di base, per calcolare il valore corretto del patrimonio supplementare occorre effettuare alcune deduzioni dal computo totale: le previsioni di esiti dubbi sui crediti e le minusvalenze sui titoli.

In questi calcoli, la normativa prevede che il valore del patrimonio supplementare non possa superare il valore del patrimonio di base e, pertanto, non possa eccedere il 50% del valore del patrimonio complessivo, somma di Tier 1 Capital e Tier 2 Capital.

A questo punto, la somma del patrimonio di base e del patrimonio supplementare costituisce il patrimonio complessivo. Tuttavia, per arrivare alla formulazione definitiva del patrimonio regolamentare complessivo, è fondamentale sottrarre il valore associato alle partecipazioni in altre istituzioni finanziarie, in modo da evitare effetti distorsivi causati da fenomeni di *double gearing*. Questi elementi devono essere dedotti in modo equivalente, in misura pari al 50% dal patrimonio di base e al 50% dal patrimonio supplementare.

In qualche caso, al capitale regolamentare viene aggiunto anche un'ulteriore potenziale componente di capitale, ovvero il Tier 3 Capital, il quale, introdotto per la prima volta nel 1996 a seguito dell'emendamento effettuato per l'introduzione del rischio di mercato, risulta costituito da debito subordinato a breve scadenza non utilizzabile nel calcolo del patrimonio supplementare. In questo ambito, i prestiti subordinati che possono essere utilizzati devono rispettare alcune condizioni: devono essere interamente versati, non garantiti e subordinati a tutti gli altri debiti, devono avere una durata originaria minima di almeno due anni, non possono essere rimborsati anticipatamente senza approvazione delle Autorità di Vigilanza e devono prevedere una clausola di *lock-in*, attraverso la quale la banca può sospendere il rimborso del capitale e il pagamento degli interessi, qualora il patrimonio scenda al di sotto dei requisiti patrimoniali minimi, eventualmente maggiorati per una specifica percentuale.

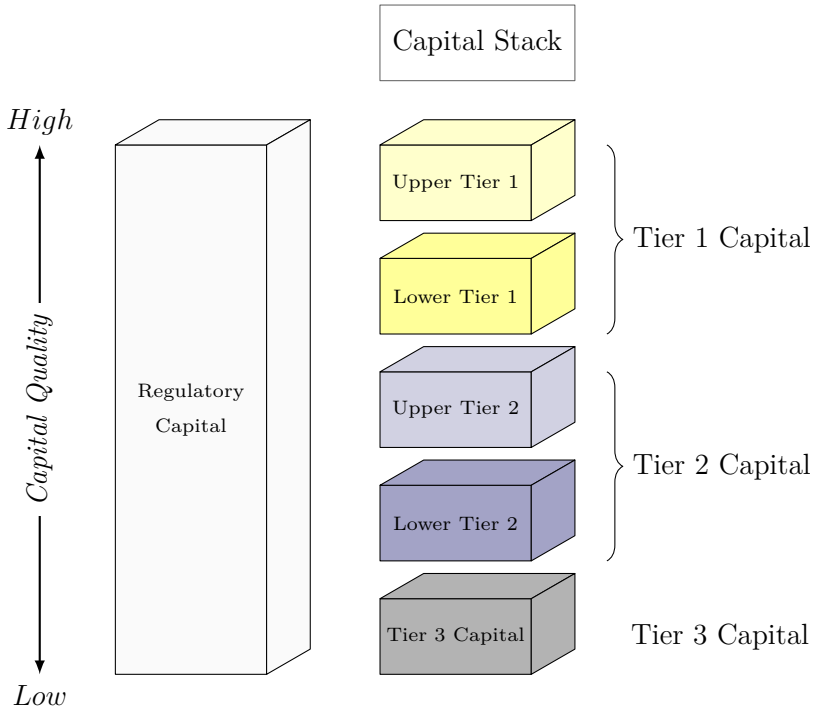


Figura 1.3. Gerarchia delle principali componenti del capitale regolamentare.

Il Tier 3 Capital è utilizzabile solo ai fini del rispetto del requisito patrimoniale per il rischio di mercato e non può eccedere oltre il 250% del patrimonio di base utilizzabile per la copertura di quest'ultimo. Al contrario, il requisito patrimoniale relativo al rischio di credito deve essere sempre rispettato attraverso il capitale regolamentare vero e proprio, composto da Tier 1 Capital e Tier 2 Capital.

1.6.4 Limiti di Basilea I

L'approccio proposto da Basilea I nel 1988 presentava numerosi limiti. Per prima cosa, il focus era quasi esclusivamente sul rischio di credito, nonostante le ponderazioni per il rischio provassero in qualche misura a tenere conto della liquidità delle diverse classi di attività finanziarie. Pertanto, lo schema di Basilea I, almeno nella sua versione originale, si concentrava solamente sul rischio di credito e ignorava completamente altre tipologie di rischio, come il rischio di mercato e il rischio operativo.

Per provare a correggere questo limite, il Comitato di Basilea, nel 1996, decise di integrare la regolamentazione bancaria con un emendamento che estendeva i requisiti patrimoniali obbligatori anche al rischio di mercato.

Basilea I presentava, inoltre, una scarsa differenziazione del rischio, dal momento che le ponderazioni per il rischio risultavano troppo approssimative e consideravano come un'unica categoria di rischio tutti i crediti verso imprese private. Così facendo, non si

riusciva a ottenere una buona differenziazione tra imprese con rating diverso, in quanto tutti i crediti verso imprese private sottintendevano erroneamente la stessa esposizione al rischio di credito e richiedevano, quindi, lo stesso requisito patrimoniale.

Un ragionamento sostanzialmente identico era valido anche nel confronto tra crediti verso Paesi OCSE e crediti verso Paesi non-OCSE. In questo caso, infatti, le ponderazioni per il rischio, non essendo in grado di riflettere la reale qualità creditizia dei vari Paesi, valutavano i crediti verso Paesi non-OCSE più rischiosi rispetto a quelli verso paesi OCSE, generando, in modo analogo, distorsioni tra Paesi con rating diverso.

Un'ulteriore mancanza era legata al mancato riconoscimento della diversificazione di portafoglio, che portava a trascurare completamente i benefici derivanti da una corretta gestione della diversificazione del rischio. Basilea I, per la misurazione del rischio su un generico portafoglio crediti, teneva conto solamente del livello di rischio delle singole esposizioni creditizie, ma non ne misurava la correlazione.

In questo caso, si distruggeva completamente l'incentivo per le banche di seguire una corretta diversificazione del rischio di credito, dato che non si riusciva a distinguere tra portafogli ben diversificati e portafogli fortemente concentrati, richiedendo per entrambi lo stesso ammontare di capitale.

In aggiunta, lo schema di Basilea I proponeva un limitato riconoscimento del legame tra scadenza dell'esposizione creditizia e rischio di credito, determinando una valutazione non completamente corretta della relativa esposizione creditizia, visto che non si riusciva a catturare la diversità del grado di rischio in funzione della vita residua e, dunque, ad assegnare un requisito patrimoniale maggiore alle esposizioni creditizie caratterizzate da un elevato orizzonte temporale.

Infine, il modello di Basilea I presentava un limitato riconoscimento degli strumenti di *risk mitigation*, con un'identificazione modesta e parziale dei benefici connessi all'utilizzo di garanzie o derivati creditizi. La presenza di attivi assistiti da garanzie reali o personali consentiva, infatti, di ottenere solamente una riduzione nelle ponderazioni, cancellando completamente l'incentivo per le banche a utilizzare questa tipologia di strumenti per ottenere un vantaggio in termini di riduzione dei rischi.

In conclusione, a causa dei limiti precedentemente discussi, è possibile affermare come le ponderazioni per il rischio formulate e presentate da Basilea I fossero troppo semplicistiche. I coefficienti introdotti dalla normativa regolamentare, infatti, rischiavano di generare pesanti distorsioni nel sistema bancario e contribuivano a creare incentivi a sviluppare operazioni di arbitraggio regolamentare per alterare i portafogli bancari con l'obiettivo di massimizzare il valore per gli azionisti, sotto il vincolo della regolamentazione bancaria.

In questo contesto, a fini regolamentari, non era possibile differenziare adeguatamente la concessione di un credito di alta qualità e la concessione di un credito di bassa qualità. Per questo motivo, le banche, cercando di aumentare la redditività, erano indirettamente spinte a privilegiare investimenti in crediti sempre più rischiosi, dal momento che i crediti con un elevato livello di rischio erano in grado di offrire, rispetto ai crediti con un livello di rischio inferiore, rendimenti più importanti, pur essendo entrambi soggetti allo stesso accantonamento di capitale.

L'incentivo conseguente era quello di sostituire esposizioni creditizie di alta qualità, caratterizzate da un requisito patrimoniale relativamente elevato e, dunque, costose in termini di capitale, con esposizioni creditizie di bassa qualità, caratterizzate, invece, da

un requisito patrimoniale comparativamente modesto, con la finalità di guadagnare spread più consistenti e incrementare la redditività.

Pertanto, il risultato finale era quello di peggiorare la qualità del portafoglio creditizio delle banche, con un incremento generale della rischiosità. A tale proposito, considerando la volontà del Comitato di Basilea di rendere più stabile il sistema bancario, è facile mettere in luce come questo risultato ex-post fosse in palese contrasto con l'obiettivo ex-ante.

Nonostante questi limiti, Basilea I è comunque riuscito a interrompere la progressiva riduzione della capitalizzazione per le grandi banche internazionali e a favorire la creazione di un *level playing field*, ovvero un sistema contraddistinto da principi concorrenziali omogenei e uniformi. A testimonianza di questo, negli anni successivi alla riforma, il coefficiente patrimoniale medio delle principali banche internazionali è aumentato nella maggior parte dei Paesi economicamente più sviluppati, superando molto spesso anche il valore minimo dell'8%.

Tuttavia, il Comitato di Basilea ha progressivamente continuato a lavorare per migliorare l'impianto originale di Basilea I, caratterizzato da un meccanismo semplice e uniforme ma troppo elementare per consentire una corretta valutazione dei rischi, inaugurando così il passaggio della regolamentazione bancaria verso Basilea II.

1.7 Basilea II (2004)

Il nuovo accordo sul capitale, noto come Basilea II, è il risultato di un processo lungo e laborioso, che inizia a giugno 1999, quando il Comitato di Basilea decide di avviare una modifica alla regolamentazione bancaria, proponendo una prima proposta di riforma. Quest'ultima, dopo un intenso processo di consultazioni e di simulazioni quantitative, subisce ulteriori modifiche, prima nel gennaio 2001 e dopo nell'aprile 2003, arrivando a una versione definitiva nel giugno 2004.

Basilea II rappresenta un passaggio fondamentale per la regolamentazione bancaria, introducendo un progetto di riforma basato su tre pilastri fondamentali, chiamati *pillar*, destinati a rinforzarsi mutuamente:

- requisiti patrimoniali minimi, o *Capital Requirement*;
- processo di controllo prudenziale, o *Supervisory Review*;
- disciplina di mercato, o *Market Discipline*.

Il primo pilastro introduce nuove regole, più strutturate, per il calcolo dei requisiti patrimoniali obbligatori, anche se il coefficiente patrimoniale previsto da Basilea I, secondo cui il patrimonio delle banche deve essere almeno pari all'8% delle attività ponderate per il rischio, non subisce variazioni e rimane immutato anche con Basilea II.

Tuttavia, adesso, con questa riforma si osservano numerose modifiche: da una parte, i requisiti patrimoniali per il rischio di credito vengono finalmente calcolati utilizzando una metodologia più sofisticata, in modo da riflettere coerentemente il rating delle controparti; dall'altra, i requisiti patrimoniali per il rischio di mercato non vengono modificati e restano invariati rispetto a quanto previsto dall'emendamento del 1996, ma, allo stesso tempo,

viene introdotto un nuovo requisito patrimoniale per la copertura dei rischi operativi, mancante precedentemente. Riassumendo, è possibile rappresentare questi concetti in modo estremamente sintetico con la seguente terminologia:

$$\frac{Capitale\ Regolamentare}{CR + MR + OR} \geq 8\% \ , \tag{1.2}$$

dove:

- CR = rischio di credito;
- MR = rischio di mercato;
- OR = rischio operativo.

Il secondo pilastro riguarda la supervisione sull'intero sistema bancario da parte delle Autorità di Vigilanza, in modo da assicurare l'adeguatezza patrimoniale e garantire il possesso di opportuni sistemi di misura e controllo dei rischi. Fondamentalmente, le Autorità di Vigilanza hanno il compito di verificare che la gestione dei rischi rispetti determinati standard sia dal punto di vista qualitativo che dal punto di vista quantitativo.

In questo contesto, si richiede alle Autorità di Vigilanza qualcosa in più rispetto al semplice riscontro dell'osservanza da parte delle banche dei requisiti patrimoniali stabiliti da Basilea II. In particolare, le Autorità di Vigilanza hanno il compito di individuare situazioni potenzialmente critiche, valutare eventuali problematiche e, infine, effettuare, in caso di necessità, interventi correttivi in tempo utile.

L'obiettivo è quello di esaminare approfonditamente le tecniche di risk management utilizzate dalle banche e, nel caso in cui dovessero essere riscontrate carenze, entrare in un dialogo attivo per incoraggiare lo sviluppo di tecniche sempre migliori.

Questo pilastro cerca di assicurare uniformità e coerenza a livello generale, anche se consente ai diversi Paesi una certa discrezionalità nell'applicazione delle regole, in modo da tenere in considerazione le diverse condizioni nazionali. Infatti, nell'ambito dei controlli bancari effettuati dalle Autorità di Vigilanza, è sempre possibile imporre requisiti patrimoniali più elevati rispetto a quelli previsti da Basilea II.

Il terzo pilastro riguarda la disciplina di mercato, ovvero la disciplina esercitata dal mercato dei capitali, che valuta l'operato delle banche e sceglie di premiare o punire rispettivamente quelle più o meno meritevoli. Nello specifico, si ritiene necessario un rafforzamento della disciplina esercitata dal mercato dei capitali, con la motivazione di fornire maggiori informazioni agli azionisti e, quindi, indurre, in modo diretto o indiretto, il management a perfezionare i processi alla base della gestione dei rischi.

Il mercato dei capitali rappresenta da sempre il miglior giudice per valutare e prezzare il contenuto di rischio di una banca, poiché consente di ottenere una valutazione oggettiva, con una penalizzazione nei confronti delle banche giudicate troppo rischiose, attraverso un aumento del costo del finanziamento o, ancora, un rifiuto alla concessione dello stesso.

Il mercato dei capitali, però, deve risultare efficace ed efficiente per svolgere una corretta operazione di monitoraggio e controllo sull'operato delle banche. Di conseguenza, quest'ultimo deve essere costantemente e adeguatamente informato sulle condizioni di capitalizzazione e sul contenuto di rischio delle banche.

Per ottenere questo, il terzo pilastro prescrive alle banche di soddisfare specifici criteri di disclosure, con la richiesta di adempiere a obblighi di trasparenza e comunicazione verso il mercato. In poche parole, le banche sono sollecitate a divulgare maggiori informazioni sulle decisioni in materia di esposizione ai rischi e di allocazione del capitale.

1.7.1 Pillar I: Capital Requirement

Uno dei principali obiettivi di Basilea II è quello di promuovere un sistema di requisiti patrimoniali maggiormente sensibili al grado di rischio effettivo dei portafogli bancari, riducendo il divario tra capitale economico, misurato dai modelli interni delle banche, e capitale regolamentare, imposto dalla regolamentazione bancaria.

A tale proposito, si è deciso di ridisegnare completamente le regole per il calcolo dei requisiti minimi di capitale, con la finalità di ottenere una struttura in grado di riflettere in modo più strutturato il rischio effettivo dei portafogli delle banche. In aggiunta, il patrimonio di vigilanza deve adesso coprire una maggior quantità e qualità di rischi, ovvero rischio di credito, rischio di mercato e, per la prima volta, rischio operativo.

A questo punto, è possibile approfondire le regole numeriche per il calcolo puntuale del patrimonio di vigilanza. Per quanto riguarda il rischio di credito, Basilea II riesce finalmente a differenziare tra prestiti emessi a una medesima categoria di controparti, in modo da richiedere una diversa copertura patrimoniale al variare della rischiosità.

In questo caso, la rischiosità di una generica attività finanziaria può essere valutata secondo due approcci: approccio standard e approccio basato sui sistemi di rating interni, o *Internal Ratings Based* (IRB). Questi metodi differiscono per il processo di assegnazione del rating, osservando che l'approccio standard utilizza una valutazione della rischiosità effettuata da istituzioni esterne alla banca, mentre l'approccio basato sui sistemi di rating interni utilizza una valutazione della rischiosità effettuata dalla banca in prima persona.

Primo pilastro: approccio standard

L'approccio standard riprende i concetti istituiti da Basilea I, introducendo una novità. Con questo metodo le ponderazioni per il rischio sono sempre determinate in base alla categoria a cui appartengono i diversi debitori, ma, adesso, la ponderazione è attribuita in funzione del rating assegnato alla controparte da una o più agenzie per la valutazione esterna del merito di credito, chiamate *External Credit Assessment Institution* (ECAI). Queste organizzazioni possono essere agenzie di rating oppure altre istituzioni finanziarie, a condizione di essere riconosciute dalle Autorità di Vigilanza nazionali.

In ogni caso, per accedere alle autorità abilitate, è necessario soddisfare una serie di requisiti minimi in termini di indipendenza, trasparenza e coerenza dei criteri di rating. Nello specifico, sono previsti sei requisiti: oggettività (metodologia di assegnazione dei giudizi rigorosa, sistematica e soggetta a validazione sulla base di esperienza storica), indipendenza, disponibilità di dati pubblici a supporto della validazione, trasparenza della metodologia, adeguatezza delle risorse e credibilità delle valutazioni.

Ricapitolando, l'approccio standard autorizza le banche a utilizzare le valutazioni di agenzie di rating esterne per la stima del merito creditizio. In questo modo, il merito creditizio viene valutato secondo due dimensioni diverse: categoria di controparte e rating.

Insomma, categorie di controparti diverse comportano scale di ponderazione diverse, ma, all'interno della stessa categoria di controparte, è finalmente possibile differenziare i risultati in base al rating. Pertanto, nel calcolo delle attività ponderate per il rischio, un rating migliore consente di utilizzare un peso minore, mentre un rating peggiore impone di utilizzare un peso maggiore.

Questo calcolo si ripercuote direttamente nella valutazione del patrimonio regolamentare minimo, che risulterà essere, quindi, un valore più basso nel caso di rating migliore e un valore più alto nel caso di rating peggiore, come visibile nella tabella 1.1.

Tabella 1.1. Coefficienti di ponderazione previsti da Basilea I (approccio standard).

	AAA $AA-$	$A+$ $A-$	$BBB+$ $BBB-$	$BB+$ $BB-$	$B+$ $B-$	$B-$ D	NR (NOT RATED)
Soggetti sovrani (Paesi e Banche Centrali)	0%	20%	50%	100%	100%	150%	100%
Banche opzione 1 opzione 2	 20% 20%	 50% 50%	 100% 50%	 100% 100%	 100% 100%	 150% 150%	 100% 50%
Imprese	20%	50%	100%	100%	150%	150%	100%
Retail	75%	75%	75%	75%	75%	75%	75%
Mutui ipotecari su immobili residenziali	35%	35%	35%	35%	35%	35%	35%
Mutui ipotecari su immobili non residenziali	da 50% a 100%, a discrezione delle Autorità di Vigilanza						

In questa tabella, è possibile osservare un meccanismo semplice e intuitivo: le righe descrivono le diverse classi di prenditori individuate da Basilea II, quali soggetti sovrani (Paesi e Banche Centrali), banche[5], imprese e retail, che costituisce una categoria in grado di raggruppare insieme imprese di piccola dimensione e soggetti privati, oltre che alcune tipologie specifiche di prestiti, mentre le colonne indicano i diversi rating che potrebbero essere assegnati a una controparte.

La scala di ponderazione dei crediti mette in evidenza due incongruenze fondamentali. La prima incongruenza coinvolge la categoria di rating che va da BBB+ fino a BB+. In questo sottoinsieme, infatti, si mescolano crediti di diversa qualità, dal momento che,

[5] I crediti verso banche possono essere ponderati, con possibilità di scelta per le Autorità di Vigilanza, seguendo due opzioni diverse, che si differenziano nel processo di assegnazione della ponderazione: l'opzione 1 utilizza il rating del Paese in cui ha sede la banca e l'opzione 2 utilizza il rating della banca.

al suo interno, esistono contemporaneamente due classi di rating molto diverse tra loro: *investment grade* e *speculative grade*.

In particolare, la classe di rating *investment grade* comprende crediti caratterizzati da ottima qualità e da un ridotto livello di rischio, con rating compresi tra AAA e BBB−; viceversa, la classe di rating *speculative grade* comprende invece crediti caratterizzati da qualità inferiore e da un livello di rischio superiore, con rating compresi tra BB+ e D.

La seconda incongruenza riguarda il confronto tra due tipologie di imprese: da un lato, le imprese appartenenti alla classe di rating compresa tra BB− e D; dall'altro, le imprese sprovviste di rating, ovvero appartenenti alla classe di rating NR. A questo proposito, Basilea II effettua una valutazione diversa per queste due categorie di imprese, assegnando un coefficiente di ponderazione pari al 100% nel primo caso e al 150% nel secondo caso.

Un'assegnazione di questo tipo risulta in evidente conflitto con la logica finanziaria relativa alla disponibilità di informazioni sulla qualità creditizia della controparte e, quindi, rischia di eliminare l'incentivo per le imprese a richiedere un rating, dato che quest'ultimo potrebbe imporre alla banca un requisito patrimoniale non proporzionale. In questo caso, infatti, un'impresa con un potenziale di reddito contenuto e una *debt capacity* limitata, visto che, se domandasse un rating potrebbe ottenere solamente un rating modesto, non avrebbe nessun incentivo a richiedere un rating.

Primo pilastro: approccio IRB

Come alternativa all'approccio standard, Basilea II prevede che le banche possano utilizzare l'approccio dei sistemi di rating interni e, pertanto, applicare nell'ambito della regolamentazione bancaria un modello per la misura del rischio sviluppato internamente. Ovviamente, per essere concretamente utilizzabile in ambito risk management per il calcolo del capitale regolamentare, il sistema di rating interno deve essere prima approvato dalle Autorità di Vigilanza nazionali.

In questo contesto, le banche sono ritenute, secondo una modalità parziale o totale, direttamente e personalmente responsabili per la valutazione del rischio di credito, stimando, in completa autonomia, il livello di rischio associato a ogni singolo credito, oltre che al portafoglio crediti nel suo complesso.

A questo riguardo, Basilea II, implicitamente o esplicitamente, individua sei driver di rischio principali, responsabili di determinare l'entità delle possibili perdite future in relazione alla concessione di uno o più crediti:

- Exposure at Default;
- Probability of Default;
- Loss Given Default;
- Maturity;
- Granularity;
- Correlation.

Più precisamente, i primi quattro fattori di rischio, EAD, PD, LGD e Maturity, rappresentano i parametri essenziali che un sistema di rating interno deve dimostrare di essere in grado di misurare in modo appropriato e adeguato per ogni singolo credito, mentre gli ultimi due fattori di rischio, Granularity e Correlation, costituiscono specifiche dimensioni ulteriori che intervengono nella misurazione del rischio di credito, passando dall'analisi di un singolo credito a quella di un portafoglio crediti.

A questo punto, le banche, al variare del grado di complessità ottenuto internamente dai loro modelli e dai loro database, possono essere ammesse, previa dimostrazione di aver raggiunto un livello sufficientemente avanzato, a utilizzare due approcci diversi, che si distinguono in base al processo seguito per la stima dei diversi fattori di rischio:

- approccio di base, o *Foundation*: dove le banche, impiegando metodologie interne, possono stimare soltanto la PD dei debitori e sono obbligate a fare riferimento a valori prefissati dalle Autorità di Vigilanza per la stima di EAD, LGD e Maturity.
- approccio avanzato, o *Advanced*: dove le banche, impiegando metodologie interne, possono stimare tutti i relativi profili di rischio, quali EAD, PD, LGD e Maturity, pur dovendone dimostrare prestazioni e solidità.

In sostanza, Basilea II segue un approccio evolutivo: inizialmente, in via cautelativa, si invitano le banche ad applicare l'approccio di base, ma successivamente, nel momento in cui queste ultime dimostrano di essere diventate sufficientemente sicure della qualità delle loro stime, con la possibilità di mostrare alle Autorità di Vigilanza le basi di dati e le metodologie utilizzate, si autorizzano a passare verso l'approccio avanzato.

Tuttavia, alle banche non si concede in nessun caso la possibilità di misurare in modo autonomo i fattori di rischio Granularity e Correlation del proprio portafoglio crediti, i quali risultano stabiliti in anticipo dalle Autorità di Vigilanza e, dunque, fissati su livelli standard, identici per qualsiasi banca.

In aggiunta, le banche che decidono di utilizzare il proprio sistema di rating interno per la valutazione della rischiosità e per il calcolo del capitale minimo, devono essere in grado di presentare una serie di requisiti minimi, consentendo una corretta misurazione del rischio di credito. Nello specifico, i requisiti minimi per la validazione di un sistema di rating interno sono numerosi, tra questi i più importanti sono i seguenti:

- Il sistema di rating interno deve valutare separatamente la PD e la LGD.
- I crediti devono essere distribuiti in modo omogeneo e uniforme tra le varie classi di rating, senza eccessive concentrazioni in una o più categorie specifiche.
- Il rating deve essere assegnato ai debitori prima della concessione del credito.
- Il rating deve essere rivisto e rivalutato periodicamente, valutando la possibilità di onorare gli impegni nonostante l'insorgere di condizioni avverse o eventi inattesi.
- Il rating deve essere utilizzato sia nella gestione che nel pricing dei crediti.
- Il sistema di rating interno deve presentare requisiti di documentazione formale in materia di costruzione, funzionamento e validazione di accuratezza e coerenza.

Per spiegare il principio generale con cui sono stati costruiti questi requisiti, è possibile evidenziare come i procedimenti di stima del rischio debbano essere in grado di differenziare adeguatamente livelli di rischio diversi, esprimendo una valutazione corretta, accurata e coerente con l'esperienza pregressa sperimentata della banca in esame.

Come ultima cosa, Basilea II ha stabilito una definizione univoca di insolvenza, affinché venga presa come riferimento senza ambiguità per la stima della probabilità di insolvenza. Più precisamente, un debitore viene considerato insolvente se e solo se ricorre almeno una tra le seguenti condizioni:

- condizione soggettiva: dove la banca considera improbabile che il debitore possa adempiere in modo completo alle proprie obbligazioni, con una valutazione che deriva da un pensiero soggettivo della banca, eventualmente in considerazione del fatto che, dal momento della concessione del credito, si potrebbero essere verificate condizioni particolari, come la svalutazione delle esposizioni creditizie originarie, la creazione di accantonamenti specifici, la concessione della ristrutturazione del debito, oltre che la presentazione da parte del debitore di una domanda di fallimento o di ammissione a una procedura di protezione nei confronti dei creditori.

- condizione oggettiva: dove la controparte risulta in ritardo sul pagamento di almeno una delle proprie obbligazioni per un intervallo temporale totale di oltre 90 giorni (termine che può essere prorogato a 180 giorni per crediti verso speciali controparti), con una valutazione che non deriva dal semplice pensiero soggettivo della banca, quanto piuttosto dal verificarsi di un determinato evento oggettivo.

1.7.2 Pillar II: Supervisory Review

Il primo pilastro, anche nella sua versione più avanzata, non risulta in grado di modellare con sufficiente flessibilità e duttilità alcuni aspetti sostanziali connessi al rischio di credito, come ad esempio la concentrazione e la correlazione tra crediti in un portafoglio crediti, o ancora la copertura di alcuni rischi specifici non adatti a essere misurati utilizzando schemi universali, anche se potenzialmente estremamente rilevanti.

Per di più, il corretto funzionamento dei modelli introdotti dal primo pilastro non è assoluto, ma dipende dall'ambiente di applicazione e, per questo motivo, discende in maniera cruciale dalle soluzioni organizzative previste dalle singole banche, oltre che dal livello di coinvolgimento dei vertici aziendali nelle politiche di misura e controllo del rischio.

A questo proposito, è possibile mettere in evidenza il ruolo fondamentale ricoperto dalle Autorità di Vigilanza, le quali, attraverso lo svolgimento di un'opera di supervisione sull'intero processo di misurazione dei rischi adottato dalle singole banche, possono valutare la struttura organizzativa adottata, verificare la validità e la solidità dei modelli quantitativi applicati e, in caso di necessità, imporre un cuscinetto di capitale addizionale oltre a quello previsto dalle regole di calcolo introdotte precedentemente.

Il secondo pilastro è fondato, quindi, sull'interazione tra banche e Autorità di Vigilanza, con l'obiettivo di mettere in parallelo oltre a determinate regole numeriche e oggettive, anche una proficua relazione di feedback tra attori diversi, per consentire di esaminare i soggettivi profili di rischio a cui possono andare incontro le diverse organizzazioni.

Questo processo di supervisione o controllo prudenziale, definito in modo sintetico *Supervisory Review Process* (SRP), dipende da quattro principi essenziali, specificati brevemente di seguito:

- le banche devono possedere un sistema di processi interni per la determinazione dell'adeguatezza patrimoniale, definito *Internal Capital Adequacy Assessment Process* (ICAAP), in modo tale che sia possibile non solo valutare con tecniche specifiche la propria dotazione complessiva di capitale in rapporto al profilo di rischio scelto, ma anche definire una strategia per il mantenimento del relativo livello di capitale.

- le Autorità di Vigilanza hanno il compito di esaminare con attenzione i processi, le tecniche e le strategie adottate dalle banche, nonché la capacità di assicurare il rispetto dei coefficienti patrimoniali obbligatori, formulando un giudizio complessivo contenuto in una valutazione, chiamata *Supervisory Review and Evaluation Process* (SREP), e, in caso di esito insufficiente, attivare specifiche misure correttive tramite alcune azioni prudenziali.

- le Autorità di Vigilanza si aspettano che le banche possano operare con una dotazione di capitale eccedente rispetto al capitale minimo regolamentare, ma, in ogni caso, possono sempre richiedere alle banche di mantenere un valore superiore a quanto obbligatorio, con l'eventuale eccesso di capitale che può essere spiegato in base alle caratteristiche della banca in esame, come l'esperienza e i risultati ottenuti nella gestione dei rischi, la natura dei mercati in cui opera e la volatilità dei ricavi.

- le Autorità di Vigilanza hanno il compito di intervenire, quanto più tempestivamente possibile, per evitare che il capitale in possesso di una banca possa scendere al di sotto del capitale minimo regolamentare, con la possibilità di richiedere soluzioni immediate in caso quest'ultimo non venga prontamente riportato e mantenuto al di sopra del relativo valore limite.

In Italia, la Banca d'Italia ha specificato provvedimenti aggressivi e penalizzanti, quali l'obbligo di possedere una quantità di patrimonio di vigilanza superiore rispetto al limite imposto dal primo pilastro, il divieto di svolgere speciali categorie di operazioni, il vincolo legato a specifiche attività o strutture territoriali e, infine, il divieto di distribuire utili.

Nel complesso, il secondo pilatro gioca un ruolo fondamentale nella regolamentazione bancaria moderna, poiché riconosce che il rischio a cui è esposta una banca non dipende soltanto da parametri quantitativi, i quali risultano estremamente necessari e importanti ma non sufficienti per descrivere pienamente il problema, richiedendo anche aspetti qualitativi in grado di completare il ragionamento, tra i quali sicuramente l'assetto organizzativo, la qualità dei processi aziendali e la capacità del management della banca.

1.7.3 Pillar III: Market Discipline

Il settore finanziario e, nel dettaglio, il settore bancario risulta estremamente peculiare, governato da regole singolari, in quanto i suoi principali attori, le banche, non possono essere considerate imprese "normali", bensì imprese quantomeno "speciali".

Prima di tutto, le banche giocano un ruolo essenziale all'interno del sistema economico, in quanto allocatori ottimali delle risorse in eccesso, con il compito di prestare capitale, trasformando complessivamente ammontare, scadenza e rischio, a tutte quelle imprese meritevoli che possono consentire di ottenere il ritorno marginale più elevato e, al contempo, massimizzare la generazione di nuovo PIL, ma anche canali preferenziali di trasmissione della politica monetaria e gestori di una porzione importante del sistema dei pagamenti.

Purtroppo, però, queste ultime costituiscono imprese relativamente poco trasparenti e presentano un elevato livello di opacità, rendendo particolarmente complesso valutare in modo corretto il rischio connesso ai loro investimenti, se non adottando particolari strategie di analisi, che risultano comunque di difficile applicazione.

Inoltre, sono prevalentemente finanziate da specifici soggetti, i depositanti, incapaci, nella maggior parte dei casi, di valutare correttamente il rischio e, conseguentemente, di fissare un prezzo adeguato, domandando tassi di interesse più elevati.

Infine, considerata la loro importanza, hanno diritto a una serie di strumenti specifici, come le anticipazioni della Banca Centrale o l'assicurazione dei depositi, che, congiuntamente, contribuiscono ad aggravare un fenomeno di azzardo morale, o *moral hazard*, con una rete di salvataggio che mitiga notevolmente gli incentivi a valutarne la stabilità.

Per queste ragioni, la regolamentazione bancaria si dimostra imprescindibile in funzione di un corretto funzionamento del settore bancario, in quanto interviene per cercare di alleviare questi problemi, imponendo la necessità di requisiti patrimoniali obbligatori.

In generale, un'impresa industriale o commerciale che, svolgendo il proprio business, decide di aumentare il relativo livello di rischio, aumentando la leva finanziaria e, quindi, aumentando il proprio indebitamento, si trova costretta a scontare, in qualche modo, queste stesse variazioni direttamente sul processo di finanziamento, almeno tramite un incremento del costo del debito, se non, addirittura, un rifiuto alla sua concessione.

Questo andamento è riassumibile nella disciplina di mercato, che costituisce un meccanismo fondato sull'operato del mercato dei capitali, il quale sceglie, contemporaneamente, di premiare le imprese più virtuose, concedendo di ampliare i finanziamenti concessi, oltre che ridurre i tassi di interesse richiesti, e di punire le imprese più rischiose, stabilendo, viceversa, di ridurre i finanziamenti concessi, oltre che ampliare i tassi di interesse richiesti.

Tuttavia, la disciplina di mercato produce effetti notevolmente meno tangibili quando il processo di finanziamento coinvolge una banca, dal momento che i creditori delle banche, per i motivi menzionati in precedenza, risultano poco efficaci ed efficienti nell'esercitare quella funzione di disciplina, piuttosto importante, che le stesse banche hanno il compito di assolvere verso le imprese non finanziarie, esaminandone la solidità e la solvibilità, svolgendone il monitoraggio e il controllo del rischio e, infine, imponendo, in merito a un eventuale finanziamento, condizioni coerenti e consistenti sul tasso di interesse applicato, arrivando persino a rifiutare il finanziamento in presenza di rischi giudicati eccessivi.

In questo contesto, si innesta il terzo pilastro, che intende rimuovere i fattori di disturbo che possono ostacolare un'adeguata disciplina di mercato nei confronti delle banche, in modo da imporre a queste ultime criteri di disclosure estremamente severi e rigidi, per trasmettere agli investitori, con un'informativa chiara e precisa, le informazioni principali circa lo stato di salute economico-finanziaria della banca in esame.

Basilea II fornisce un elenco delle informazioni ritenute rilevanti, ovvero essenziali per la corretta formazione del giudizio da parte di un generico investitore, che devono essere

obbligatoriamente diffuse verso il pubblico:

- la dimensione e la composizione del patrimonio e degli attivi a rischio;
- la ripartizione delle esposizioni creditizie tra le classi di probabilità di insolvenza calcolate, oltre che il relativo tasso di insolvenza registrato su ogni classe;
- i sistemi e i processi di gestione del rischio utilizzati;
- le pratiche contabili adottate e i criteri di allocazione del capitale.

Così facendo, è possibile mettere nelle migliori condizioni i creditori delle banche, in modo da consentire loro di valutare, in maniera più corretta e tempestiva, l'effettivo grado di rischio complessivo assunto dalla banca.

In generale, il terzo pilastro considera solamente uno dei problemi che impediscono un corretto svolgimento della disciplina di mercato nei confronti delle banche, ovvero una limitata trasparenza delle condizioni di esposizione al rischio, anche se, allo stesso tempo, trascura alcuni aspetti che, in qualche caso, possono risultare anche molto importanti.

A tale riguardo, è possibile menzionare ulteriori condizioni necessarie per il corretto svolgimento della disciplina di mercato, aggiungendo, accanto alla trasparenza, i seguenti elementi: l'esistenza di creditori non tutelati da garanzie pubbliche, l'assenza di particolari forme di garanzia implicita, come una politica di supervisione di tipo *too big to fail*, e il perseguimento da parte del management degli interessi dei propri azionisti, rispondendo ai segnali inviati dal mercato dei capitali.

1.8 Basilea III (2010)

La data di inizio di Basilea II è coincisa, più o meno, con l'inizio della crisi finanziaria del 2007-2009, la peggiore crisi sui mercati finanziari dopo quella del 1929. Per questo motivo, inizialmente, molti studiosi e analisti hanno ritenuto la normativa di Basilea II come la principale responsabile, commettendo un errore di valutazione.

La principale accusa che, per molto tempo, è stata rivolta nei confronti di Basilea II riguarda l'auto-regolamentazione del sistema bancario. Nello specifico, questa normativa è stata ritenuta colpevole di aver progressivamente favorito l'auto-regolamentazione del sistema bancario, considerando che, con questa riforma, nel calcolo del capitale regolamentare, è stato autorizzato l'utilizzo di modelli basati sui sistemi di rating interni, i quali stimano in modo autonomo variabili fondamentali per la valutazione del rischio di credito, come EAD, PD e LGD.

In realtà, le cause della crisi finanziaria del 2007-2009 sono da ricercare nell'evoluzione strutturale che ha sperimentato il sistema economico negli anni precedenti alla normativa, molto prima dell'introduzione e, soprattutto, dell'applicazione di Basilea II.

In questi anni, infatti, l'ambiente economico-finanziario mondiale era caratterizzato dalla presenza di tassi di interesse relativamente bassi, con effetti estremamente positivi su investimenti e crescita economica. Pertanto, nel corso del tempo, il rischio di default si era sistematicamente ridotto, causando una diminuzione direttamente proporzionale del premio per il rischio richiesto dagli investitori.

Con queste premesse, l'offerta di fondi da parte degli investitori superava ampiamente la relativa domanda, determinando un progressivo e significativo incremento del grado di liquidità nella maggior parte dei mercati finanziari.

Queste condizioni macroeconomiche sono risultate essere particolarmente favorevoli per lo sviluppo economico, ma hanno consentito l'insorgenza di tre punti deboli fondamentali: un elevato livello di leva finanziaria di individui e imprese, un'inadeguata comprensione del rischio di credito connesso a prestiti successivamente ceduti sul mercato finanziario e, infine, un'insufficiente attenzione al rischio di credito collegato a titoli risultanti da operazioni di cartolarizzazione, o *securitization*, quali ABS e CDO.

Dalla combinazione di questi elementi di vulnerabilità si è innestata la crisi finanziaria del 2007-2009, causando ripercussioni catastrofiche su tutto il sistema economico, con effetti trasmessi, in modo più o meno marcato, verso tutti i settori industriali.

Come risposta, Autorità di Vigilanza, Banche Centrali e Autorità di Governo dei principali Paesi internazionali sono stati costretti a intervenire tempestivamente a sostegno di tutte le istituzioni finanziarie in difficoltà, arrivando fino al salvataggio delle istituzioni tecnicamente insolventi. In questo modo, i costi della crisi finanziaria sono stati condivisi e sopportati non soltanto dagli azionisti delle banche, come correttamente auspicabile, ma anche dai contribuenti del settore pubblico.

In sintesi, la crisi finanziaria del 2007-2009 ha messo in evidenza alcune problematiche dell'accordo di Basilea II. Infatti, pur non rappresentandone la causa principale, Basilea II, con la propria struttura regolamentare, ha dimostrato di non essere sufficientemente avanzata per gestire la complessità associata al rischio di credito. Per questa ragione, è apparso evidente la necessità di effettuare delle modifiche per provare a risolvere definitivamente gli elementi di fragilità discussi in precedenza.

La volontà di non ripetere nuovamente un episodio di questo genere, particolarmente oneroso e gravoso per l'intero sistema economico, ha spinto il Comitato di Basilea a completare una profonda revisione della regolamentazione bancaria, che prevede l'aggiunta di nuove importanti disposizioni, superando Basilea II e arrivando a Basilea III.

Il primo documento su Basilea III è stato pubblicato nel dicembre 2009, mentre la versione definitiva è stata pubblicata solo successivamente, nel dicembre 2010, dopo una lunga serie di riunioni internazionali e considerazioni di risposta delle banche, oltre che diversi studi sull'impatto quantitativo della riforma.

La normativa definita da Basilea III comprende sette sezioni:

- definizione di patrimonio, o *Definition of Capital*;
- riserve patrimoniali, o *Capital Conservation Buffer*;
- riserve anticicliche, o *Counter-Cyclical Buffer*;
- indice di leva finanziaria, o *Leverage Ratio*;
- rischio di liquidità, o *Liquidity Risk*;
- rischio di mercato e di controparte, o *Market and Counterparty Risk*;
- rischio sistemico, o *Systemic Risk*.

Queste nuove regole sono state introdotte gradualmente tra il 2013 e il 2019.

1.8.1 Definition of Capital

Basilea III assegna alla definizione di patrimonio di vigilanza una valenza più restrittiva, in modo da aumentare la quantità e la qualità del capitale. Questo obiettivo viene perseguito con tre strumenti: un innalzamento dei requisiti minimi di capitale, l'applicazione di una serie di deduzioni prudenziali e la graduale esclusione di strumenti innovativi giudicati non sufficientemente solidi per assorbire le perdite.

In particolare, Basilea III attribuisce un ruolo di primaria importanza al capitale di maggiore qualità, ovvero il Common Equity Tier 1, composto solamente dal capitale azionario versato e dalle riserve di utili, che subisce un innalzamento dei requisiti minimi, passando dal 2% al 4,5% degli attivi ponderati per il rischio.

Inoltre, per incrementare ulteriormente il livello di sicurezza, Basilea III dispone la presenza di due principali aggregati di capitale: il patrimonio di base, o Tier 1 Capital, che deve ora rappresentare un valore superiore, passando dal 4% al 6% degli attivi ponderati per il rischio, e il patrimonio supplementare, o Tier 2 Capital.

Infine, un'ulteriore modifica della normativa regolamentare sulla definizione di capitale riguarda l'eliminazione del Tier 3 Capital, formato prevalentemente da prestiti subordinati a breve termine e originariamente ammesso solo per la copertura del rischio di mercato.

Tuttavia, come nelle precedenti versioni della regolamentazione bancaria, Basilea III impone per il patrimonio di vigilanza complessivo, ottenuto combinando insieme, con determinate modalità, Tier 1 Capital e Tier 2 Capital, un requisito patrimoniale invariato, almeno pari all'8% degli attivi ponderati per il rischio.

Il Tier 1 Capital è chiamato *going concern capital* (in inglese, letteralmente: "capitale di un'impresa in esercizio"), dato che questo aggregato di capitale è in grado di assorbire le perdite quando la banca si trova in uno stato di continuità aziendale, con capitale proprio positivo, in modo da conservare la propria operatività e scongiurare l'intervento di un processo di liquidazione.

In questo contesto, al variare della qualità del capitale, ossia la capacità di assorbimento di eventuali perdite future, è possibile classificare il Tier 1 Capital in due componenti fondamentali:

- Common Equity Tier 1 (CET1);
- Additional Tier 1 (AT1).

Il Common Equity Tier 1 rappresenta una componente di capitale di maggiore valore e comprende strumenti finanziari contrassegnati da specifiche caratteristiche, come la durata perpetua, l'impossibilità di rimborso se non in caso di liquidazione della banca e la mancanza di un obbligo di distribuire dividendi. In poche parole, questo sottoinsieme di capitale coinvolge le poste dell'attivo con elevato potenziale di assorbimento delle perdite, includendo principalmente le azioni, il fondo sovrapprezzo azioni e le riserve da utili non distribuiti.

L'Additional Tier 1 rappresenta, al contrario, una componente di capitale di minore valore e comprende strumenti finanziari in grado di rispettare alcuni requisiti, come la durata perpetua, la subordinazione rispetto ai restanti creditori della banca, inclusi i depositanti e i creditori subordinati, la mancanza di un obbligo di distribuire dividendi

e, infine, l'assenza di garanzie da parte della banca e di qualsiasi meccanismo di *step-up*. Questo sottoinsieme di capitale prevede, inoltre, la possibilità di un'opzione di rimborso anticipato esercitabile dall'emittente, con l'autorizzazione delle Autorità di Vigilanza, a distanza di almeno cinque anni dall'emissione.

Il Tier 2 Capital è detto invece *gone concern capital* (in inglese, letteralmente: "capitale di un'impresa in difficoltà"), visto che questo aggregato di capitale è in grado di assorbire le perdite quando la banca si trova in uno stato di crisi economico-finanziaria, con capitale proprio negativo, e, quindi, soltanto nel caso in cui la banca venga posta preventivamente in liquidazione o assoggettata a procedure essenzialmente simili.

A questo riguardo, il Tier 2 Capital cancella la distinzione, presente nella precedente normativa, tra Upper Tier 2 e Lower Tier 2, comprendendo adesso un'unica categoria di strumenti finanziari, ovvero prestiti subordinati, seppur con specifiche caratteristiche: una scadenza originale di almeno cinque anni, la presenza di un'opzione di rimborso anticipato da esercitare su iniziativa dell'emittente, previa autorizzazione delle Autorità di Vigilanza, dopo almeno cinque anni e l'assenza di clausole che colleghino il rendimento con il merito creditizio della banca.

In aggiunta, Basilea III applica una serie di rettifiche prudenziali alla definizione di capitale, in modo da evitare sorprese negative nelle fasi di difficoltà economico-finanziaria e, conseguentemente, garantire una corretta capacità di assorbimento delle perdite.

Per questo motivo, si corregge in via prudenziale il capitale delle banche, principalmente il Common Equity Tier 1, deducendo una serie di aggregati patrimoniali, quali investimenti in azioni proprie, *deferred tax asset*, avviamento e altre attività immateriali, investimenti non consolidati in altre istituzioni finanziarie, patrimonio di pertinenza di terzi nelle banche controllate e, infine, profitti e perdite derivanti dall'applicazione del principio del valore corrente, o *fair value*.

Riassumendo, Basilea III impone il rispetto dei seguenti requisiti patrimoniali:

$$CET1 \geq 4{,}5\% \cdot RWA \ . \tag{1.3}$$

$$Tier\ 1 = CET1 + AT1 \geq 6\% \cdot RWA \ . \tag{1.4}$$

$$RC = Tier\ 1 + Tier\ 2 \geq 8\% \cdot RWA \ . \tag{1.5}$$

Osservando questi valori, è facile evidenziare come le regole introdotte con Basilea III siano molto più stringenti rispetto alla precedente normativa in materia di regolamentazione bancaria. A tale proposito, infatti, è possibile osservare un potenziamento dei requisiti patrimoniali, sia analizzando l'aumento dei valori soglia percentuali che esaminando la circoscrizione del campo di definizione del capitale.

In sintesi, le Autorità di Vigilanza hanno scelto di modellare la struttura regolamentare sull'idea fondamentale che, nel caso in cui una banca dovesse diventare insolvente, prima di utilizzare nuovamente fondi pubblici per un salvataggio, i possessori di strumenti finanziari riconosciuti come capitale regolamentare sarebbero chiamati a sopportare, attraverso la conversione, parziale o totale, in azioni ordinarie e l'abbattimento per perdite, i costi di ristrutturazione associati.

1.8.2 Misure contro la prociclicità

La prociclicità del sistema finanziario rappresenta l'insieme dei meccanismi attraverso i quali, quest'ultimo, contribuisce ad accentuare e amplificare le fluttuazioni naturali del ciclo economico, aggravando in modo particolare le fasi di recessione, dal momento che, quando il sistema economico si trova immerso in una fase recessiva, le condizioni di salute delle imprese tendono inevitabilmente a deteriorarsi.

Basilea III introduce due strumenti contro la prociclicità, con la finalità di spingere le banche a monitorare con attenzione l'andamento del ciclo economico, accantonando una specifica porzione di capitale, in eccesso rispetto al requisito minimo regolamentare, durante le fasi di espansione e di redditività positiva, da utilizzare per contrastare le possibili perdite future durante le fasi di recessione e di redditività negativa.

Questi strumenti, come già anticipato in precedenza, sono i seguenti:

- riserve patrimoniali, o *Capital Conservation Buffer*;
- riserve anticicliche, o *Counter-Cyclical Buffer*.

In altre parole, Basilea III richiede alle banche la costituzione di cuscinetti aggiuntivi di capitale da utilizzare per l'assorbimento delle perdite in periodi di tempo caratterizzati da elevata turbolenza del sistema economico.

L'argomentazione alla base di questa scelta da parte delle Autorità di Vigilanza consiste nel fatto che, per una generica banca, risulta considerevolmente più semplice raccogliere capitale in condizioni di mercato positive piuttosto che in condizioni di mercato negative. Pertanto, queste disposizioni prudenziali si impegnano a garantire che le banche possano sempre approfittare di condizioni favorevoli del mercato per accumulare una quantità di capitale sufficientemente maggiore e, conseguentemente, essere in grado di assorbire eventuali perdite future, fronteggiando, invece, condizioni sfavorevoli del mercato.

Le regole sul *Capital Conservation Buffer* e sul *Counter-Cyclical Buffer* sono state istituite dai vari Paesi internazionali gradualmente tra il gennaio 2016 e il gennaio 2019.

Capital Conservation Buffer

Il primo meccanismo è il *Capital Conservation Buffer*, il quale prevede la costituzione di un cuscinetto addizionale di capitale, formato da Common Equity Tier 1, pari al 2,5% degli attivi ponderati per il rischio. Considerando il nuovo requisito patrimoniale minimo del 4,5% per il Common Equity Tier 1, Basilea III, con l'introduzione di questa regola supplementare, spinge il requisito patrimoniale totale per il Common Equity Tier 1 al 7% degli attivi ponderati per il rischio.

Il mancato rispetto di questo requisito di capitale prevede vincoli e limitazioni nella distribuzione degli utili, con una severità che dipende dall'ampiezza del divario rispetto al requisito di capitale totale. In poche parole, una banca, in difetto di non aver ancora accantonato un cuscinetto adeguato di capitale, è soggetta a restrizioni nella distribuzione degli utili fino alla corretta verifica dei requisiti patrimoniali.

Una disposizione di questo genere si sforza di assicurare che le banche, in condizioni di mercato normali, accumulino capitale supplementare con cui far fronte alle perdite nei

periodi di difficoltà, dal momento che, per una banca è molto più facile raccogliere capitale in condizioni di mercato normali piuttosto che in condizioni di mercato critiche.

Pertanto, per effetto del *Capital Conservation Buffer*, le banche, nelle fasi positive del ciclo economico, saranno incentivate ad accumulare in modo graduale capitale di elevata qualità fino al 7% dell'attivo, che diventa 8,5%, aggiungendo la parte di Tier 1 Capital non inclusa nel Common Equity Tier 1, e 10,5%, aggiungendo anche il Tier 2 Capital.

Al contrario, nelle fasi negative del ciclo economico, le banche saranno autorizzate a utilizzare questo cuscinetto di capitale per assorbire le perdite e, conseguentemente, a ridurre il Common Equity Tier 1 fino al 4,5% dell'attivo, senza necessariamente provocare gravose ripercussioni sulla propria normale operatività.

Counter-Cyclical Buffer

Il secondo meccanismo è il *Counter-Cyclical Buffer*, il quale comporta l'introduzione di un requisito di capitale addizionale fino a un valore massimo pari al 2,5% degli attivi ponderati per il rischio. La discrezionalità di imporre questo requisito di capitale dipende dalle Autorità di Vigilanza, che possono scegliere un valore compreso tra 0% e 2,5%, in base al livello di surriscaldamento del ciclo creditizio.

Per valutare la presenza di un surriscaldamento, è possibile utilizzare, come parametro di riferimento, la differenza tra il valore corrente e il trend di lungo periodo del rapporto tra prestiti bancari e PIL; tuttavia, l'utilizzo di questo parametro non è obbligatorio.

In ogni caso, l'applicazione del *Counter-Cyclical Buffer* non è mai automatica, ma dipende sempre dal giudizio espresso in materia dalle Autorità di Vigilanza, che, in caso di verifica positiva della fase di surriscaldamento del ciclo creditizio, devono annunciare l'implementazione di questo cuscinetto di capitale con un anno di anticipo per evitare shock imprevisti sul mercato finanziario, applicandolo nei confronti di tutte le banche in possesso di almeno un'esposizione creditizia nel Paese in esame.

Insomma, il *Counter-Cyclical Buffer* rappresenta un cuscinetto di capitale abbastanza simile ma, allo stesso tempo, diverso e ulteriore rispetto al *Capital Conservation Buffer*.

1.8.3 Leva finanziaria

In aggiunta ai requisiti patrimoniali basati sul calcolo degli attivi ponderati per il rischio, Basilea III introduce l'applicazione di un valore massimo alla leva finanziaria delle banche, denominato *Plain Leverage Ratio*. Nello specifico, si richiede il rispetto di un valore minimo, pari al 3%, sul calcolo di un indice di leva finanziaria non ponderato, definito come rapporto tra patrimonio di base (Tier 1 Capital) e attivo contabile.

$$Plain\ Leverage\ Ratio = \frac{Tier\ 1\ Capital}{Attivo} \geq 3\% \ . \tag{1.6}$$

Calcolare un indice di leva finanziaria significa sempre, con alcune differenze, calcolare il rapporto tra mezzi propri ed esposizione complessiva. In questo contesto, i mezzi propri sono costituiti dal patrimonio di base (Tier 1 Capital), mentre l'esposizione complessiva è determinata dalla somma delle esposizioni in bilancio e fuori bilancio, senza effettuare nessuna ponderazione per il rischio, a differenza di quanto previsto fino a questo momento.

In poche parole, Basilea III prescrive l'utilizzo di un indicatore di leva finanziaria basato su dati puramente contabili e non su dati corretti per il rischio, in grado di rifletterne la relativa rischiosità. Pertanto, il vincolo connesso alla leva finanziaria risulta essere come minimo controverso e contribuisce a generare più di qualche perplessità.

A questo proposito, la teoria economica ha da sempre messo in luce come l'imposizione di un requisito patrimoniale basato sul semplice rapporto tra capitale e attivo contabile risulti praticamente insufficiente per ridurre la probabilità di insolvenza di una banca.

La valutazione del rischio di credito, infatti, non dovrebbe assolutamente prescindere dall'operazione di ponderazione dell'attivo contabile per il rischio delle diverse classi di attività, in modo da non incorrere in giudizi completamente sbagliati [89] [62].

In questo contesto, l'assegnazione del requisito di capitale dovrebbe essere sempre eseguita in modo proporzionale alla rischiosità, poiché una assegnazione non proporzionale risulterebbe inevitabilmente sbagliata, considerando che, per motivi regolamentari, scegliere tra un credito con alto rischio e un credito con basso rischio non dovrebbe mai richiedere la stessa quantità di capitale.

Le argomentazioni utilizzate dal Comitato di Basilea per giustificare questa scelta rispondono principalmente a due obiettivi, ossia contenere il livello di leva finanziaria e, soprattutto, limitare gli eventuali errori di misurazione nel calcolo degli attivi ponderati per il rischio, nel tentativo di generare un sistema di sicurezza che possa intervenire nei casi di fallimento dello schema generale.

Tuttavia, è possibile mettere in evidenza almeno due importanti problemi in risposta a queste motivazioni, che non aiutano a comprenderne fino in fondo la logica sottostante.

Un primo problema riguarda la possibile eccessiva discrezionalità delle banche, legata principalmente all'utilizzo opportunistico dei sistemi di rating interni per la valutazione della rischiosità, nel calcolo degli attivi ponderati per il rischio.

A questo riguardo, un requisito patrimoniale basato sulla leva finanziaria contabile non può essere considerato uno strumento semplice e trasparente per compensare possibili errori di misurazione del sistema di ponderazione dei rischi. Così, un requisito patrimoniale calcolato seguendo questo metodo, non elimina il problema della valutazione del rischio, dal momento che utilizza una scala di pesi approssimativa, limitandosi solo ad attribuire una ponderazione pari al 100% per tutte le classi di attività.

Un secondo problema riguarda, invece, il procedimento di calcolo della leva finanziaria, poiché, calcolare l'esposizione complessiva e, dunque, esplicitare tutte le esposizioni che compongono il totale dell'attivo, non risulta un compito di facile risoluzione.

Per tutte queste ragioni, le giustificazioni presentate dal Comitato di Basilea per motivare l'introduzione di un requisito di capitale basato sulla leva finanziaria contabile possono sembrare prive di fondamento, oltre che prive di giustificabilità economica, anche nell'intento di aggiungere un ulteriore livello di sicurezza.

Ad ogni modo, è possibile immaginare che il Comitato di Basilea non abbia voluto mettere in discussione la validità delle metodologie basate sul calcolo degli attivi ponderati per il rischio, quanto piuttosto imporre alle banche il rispetto di due differenti criteri: da una parte, il rapporto tra patrimonio e attività ponderate per il rischio, con specifiche percentuali al variare della tipologia di capitale regolamentare considerato, per esempio Common Equity Tier 1, Tier 1 Capital e Tier 2 Capital; dall'altra, il *Plain Leverage Ratio*, ovvero il rapporto tra patrimonio ed esposizione complessiva non ponderata per il rischio.

Al variare delle caratteristiche idiosincratiche della banca in questione, è possibile che i due vincoli possano risultare più o meno stringenti; in ogni caso, la banca, per soddisfare le richieste di capitale, è comunque tenuta all'applicazione dinamica di entrambi.

Il vincolo basato sul calcolo del *Plain Leverage Ratio* è stato introdotto per la prima volta nella regolamentazione bancaria nel gennaio 2018, dopo un periodo transitorio di valutazione, iniziato nel 2013 e terminato nel 2017.

1.8.4 Misure per la liquidità

Prima della crisi finanziaria del 2007-2009, l'obiettivo principale della regolamentazione bancaria era quello di assicurare che le banche, svolgendo la propria attività, avessero sempre a disposizione una quantità di capitale sufficiente per contrastare i rischi assunti.

In realtà, nel corso della crisi, il Comitato di Basilea si è reso conto che la maggior parte dei problemi sperimentati dalle istituzioni finanziarie non erano provocati semplicemente dalla possibile inadeguatezza patrimoniale, ma erano causati, in modo decisamente più complesso e interconnesso, dalla mancanza di liquidità[6].

In poche parole, la crisi finanziaria del 2007-2009 ha messo in evidenza come la gestione del rischio di liquidità, stabilita dalle precedenti versioni della regolamentazione bancaria, non fosse adeguata a risolvere i problemi che affliggono, per sua natura, il settore bancario.

Le banche costituiscono, infatti, un business fortemente illiquido, considerando che possiedono il compito di raccogliere depositi a breve termine e, parallelamente, finanziare investimenti a medio-lungo termine, causando un disallineamento tra attivo e passivo, che non rappresenta un problema in condizioni normali, quando il mercato ritiene che la banca sia strutturalmente solida. Durante una crisi, però, il ragionamento precedente non risulta più valido, osservando che il mercato, in condizioni di stress, provoca inevitabilmente un aggravamento della relativa situazione economico-finanziaria e la diversa struttura per scadenza tra attivo e passivo diventa potenzialmente in grado di generare conseguenze disastrose, arrivando fino a spingere la banca in una crisi di liquidità.

A questo proposito, considerando i problemi di liquidità che hanno coinvolto la maggior parte delle banche internazionali, il Comitato di Basilea ha deciso di introdurre due nuovi requisiti di vigilanza prudenziale, in modo da garantire che le banche possano sopravvivere nei periodi di scarsa liquidità:

- indice di copertura della liquidità, o *Liquidity Coverage Ratio*;
- indice di provvista stabile netta, o *Net Stable Funding Ratio*.

Il calcolo di queste variabili, *Liquidity Coverage Ratio* e *Net Stable Funding Ratio*, non riguarda l'obbligo di superare requisiti patrimoniali aggiuntivi, ma la necessità per le banche di mantenere un livello minimo di attività liquide, o comunque a breve termine, con la finalità ultima di valutare in modo migliore il rischio di liquidità e rendere le banche autosufficienti in situazioni di mancanza di liquidità.

[6] In questo caso, è possibile citare come riferimento il fallimento di Lehman Brothers [127].

Il primo coefficiente di liquidità è stato introdotto nelle varie legislazioni internazionali tra il gennaio 2015 e il gennaio 2019, mentre il secondo coefficiente di liquidità è entrato in vigore da gennaio 2018.

Liquidity Coverage Ratio

L'indice di copertura della liquidità, o *Liquidity Coverage Ratio* (LCR), rappresenta un vincolo applicato alla liquidità di breve periodo, per assicurare che una banca possa mantenere, su base continuativa, attività liquide di elevata qualità in misura sufficiente a generare cassa e superare una situazione di difficoltà economico-finanziaria.

In sintesi, questo requisito prescrive che le attività liquide di elevata qualità siano regolarmente in grado di far fronte ai deflussi di cassa netti attesi nei successivi 30 giorni, stimati in base a uno scenario di stress, caratterizzato da gravi tensioni per la liquidità.

In formule:

$$LCR = \frac{HQLA}{TCNO_{30}^{s}} \geq 100\% \ , \tag{1.7}$$

dove:

- $HQLA$ = attività liquide di elevata qualità, o *High-Quality Liquid Assets*;
- $TNCO$ = deflussi di cassa netti attesi, o *Total Net Cash Outflows*.

Il numeratore rappresenta le attività liquide di elevata qualità e include una serie di attività facilmente e velocemente trasformabili in liquidità con una perdita di valore nulla o, comunque, trascurabile, le quali devono sempre presentare alcune caratteristiche peculiari, come basso rischio di credito e basso rischio di mercato, facilità e certezza di valutazione, bassa correlazione con attività contraddistinte da alto rischio, quotazione in un mercato sviluppato, mancanza di vincoli e, infine, ammissibilità in operazioni di sconto presso la Banca Centrale.

A questo proposito, per identificare correttamente le attività liquide di elevata qualità, il Comitato di Basilea ha predisposto un elenco di massima di attività ammissibili, distinguendo, in relazione alla qualità, tra attività di livello 1 e attività di livello 2.

Nello specifico, le attività di livello 1 rappresentano le attività di qualità superiore e possono essere utilizzate senza alcun tipo di limitazione, mentre le attività di livello 2, ovvero le attività di qualità inferiore, possono essere incluse nel computo delle attività liquide di elevata qualità soltanto fino al raggiungimento del 40% del valore totale.

Il denominatore rappresenta, invece, i deflussi di cassa netti attesi e prevede la stima della differenza tra i flussi di cassa cumulati in entrata e i flussi di cassa cumulati in uscita, misurati valutando una situazione di stress in un periodo di 30 giorni. Come risultato, il calcolo dei deflussi di cassa netti attesi consente di analizzare lo squilibrio nella posizione netta cumulata di liquidità di una banca in condizioni di considerevole criticità.

Anche in questo caso, il Comitato di Basilea, basandosi su effettivi risconti ottenuti durante la crisi finanziaria del 2007-2009, ha definito le condizioni generali da applicare per la simulazione dello scenario di stress e la stima dei deflussi di cassa netti attesi.

In sintesi, Basilea III richiede che il *Liquidity Coverage Ratio* sia costantemente almeno pari al 100%. Questo indicatore, misurando la capacità di contrastare i deflussi di cassa attesi nel breve periodo con attività potenzialmente liquide, garantisce che una banca possa essere in grado di sopravvivere durante una sollecitazione del mercato finanziario.

Net Stable Funding Ratio

L'indice di provvista stabile netta, o *Net Stable Funding Ratio* (NSFR), rappresenta, al contrario, un vincolo connesso al mantenimento di un rapporto equilibrato e sostenibile tra le fonti di finanziamento stabili a medio-lungo termine e, in modo del tutto parallelo, il fabbisogno di fondi a medio-lungo termine determinato dalla scadenza degli attivi.

Il *Net Stable Funding Ratio* misura la capacità generale di una banca di contrastare le esigenze di liquidità in una prospettiva di medio-lungo periodo, con la possibilità di imporre che il rapporto tra le risorse finanziarie stabili e il fabbisogno di risorse stabili legato alla struttura dell'attivo sia sempre maggiore del 100%. In altre parole, questo requisito richiede che l'ammontare disponibile di provvista stabile sia sempre superiore all'ammontare disponibile di provvista obbligatoria.

In formule:

$$NSFR = \frac{ASF}{RSF} \geq 100\% \ , \tag{1.8}$$

dove:

- ASF = risorse finanziarie stabili disponibili, o *Available Stable Funding*;
- RSF = risorse finanziarie stabili necessarie, o *Required Stable Funding*.

Il numeratore rappresenta le risorse finanziarie stabili disponibili ed è determinato come somma di quattro importanti elementi del passivo: il patrimonio netto, somma di Tier 1 Capital e Tier 2 Capital, le azioni privilegiate con scadenza pari o maggiore di un anno, le passività con scadenza pari o superiore a un anno e, infine, la porzione stabile delle passività senza scadenza o con scadenza inferiore a un anno, caratterizzata dall'aspettativa di rinnovo per anni successivi anche nel caso in cui si verifichi uno scenario di stress.

Per ottenere il risultato finale, si attribuisce a queste fonti di finanziamento un determinato coefficiente di stabilità, chiamato *Available Stable Funding Factor* (ASF Factor), che può assumere valori diversi al variare della qualità della fonte considerata, passando dal 50% per le fonti meno stabili al 100% per le fonti più stabili.

In questo contesto, la scelta di effettuare una ponderazione emerge dalla volontà di riflettere il possibile impatto della stabilità delle diverse fonti di finanziamento impiegate sul calcolo complessivo. A dimostrazione di questo ragionamento, infatti, è possibile immaginare di assegnare implicitamente un valore pari allo 0% per le fonti di finanziamento non direttamente utilizzabili per il computo delle risorse finanziarie stabili disponibili.

Il denominatore costituisce, invece, le risorse finanziarie stabili necessarie ed è calcolato, in modo simile, come somma di diversi elementi dell'attivo, ponderati tramite un determinato coefficiente di liquidità, chiamato *Required Stable Funding Factor* (RSF Factor), che consente di valutarne il relativo grado di liquidità.

A questo proposito, le attività caratterizzate da un elevato livello di liquidità sono attività che risultano essere facilmente e velocemente liquidabili sul mercato e, quindi, richiedono una esigenza relativamente limitata di finanziamento con fondi stabili. Al contrario, le attività con un modesto livello di liquidità sono attività che, specialmente in condizioni di stress, possono essere liquidate solo difficilmente e lentamente sul mercato e, dunque, necessitano di utilizzare, come finanziamento, fondi stabili.

Per questo motivo, il Comitato di Basilea ha deciso di includere una valutazione più approfondita, assegnando coefficienti di ponderazione diversi a seconda della tipologia di attività considerata: un coefficiente di ponderazione elevato per le attività meno liquide e, viceversa, un coefficiente di ponderazione moderato per le attività più liquide.

1.8.5 Rischio di mercato e di controparte

Nel corso della crisi finanziaria del 2007-2009, le principali banche internazionali hanno subito grandi perdite sul portafoglio trading, influenzate prevalentemente da una gestione non completamente corretta del rischio di mercato e di controparte. Considerando, altresì, che queste perdite sono risultate molto spesso superiori rispetto al capitale regolamentare detenuto a fronte delle corrispondenti attività, il Comitato di Basilea ha immediatamente stabilito di proporre alcune importanti modifiche alla normativa esistente, associata sia al rischio di mercato, originariamente introdotta nel 1996, che al rischio di controparte.

In questo periodo, infatti, è apparso evidente come fosse più che necessario apportare delle correzioni rilevanti alla regolamentazione bancaria, in relazione al calcolo puntuale del capitale regolamentare da possedere obbligatoriamente per contrastare in modo ottimale il rischio di mercato e di controparte, dal momento che l'eventuale fallimento di una banca avrebbe determinato pesanti ripercussioni a catena per tutte le altre.

Un primo fondamentale pacchetto di riforme, che ha preso il nome di Basilea II.5, è stato introdotto, con il duplice obiettivo di contenere e, soprattutto, non ripetere le pesanti conseguenze sul sistema economico. Queste sostanziali modifiche, concordate inizialmente nel luglio 2009 e applicate successivamente nel dicembre 2011, hanno portato le banche verso una maggior consapevolezza nella valutazione del rischio di mercato, con una crescita significativa del relativo capitale regolamentare.

Un secondo pacchetto di riforme, forse ancora più importante e ambizioso, chiamato *Fundamental Review of the Banking Book*, è stato sviluppato negli anni più recenti, ma, attualmente, non è ancora operativo per il calcolo di nessun requisito patrimoniale vincolante in materia di regolamentazione bancaria.

Riassumere le correzioni effettuate dal Comitato di Basilea nel corso del tempo non è un compito semplice, anche se è possibile suddividere queste ultime in due parti principali. In primo luogo, per quanto riguarda il rischio di mercato, si è deciso di inasprire i requisiti patrimoniali per i portafogli di trading, oltre che rafforzare il trattamento prudenziale dei rischi, spesso nascosti, sulle operazioni di cartolarizzazione e sulle attività fuori bilancio. In secondo luogo, per quanto riguarda il rischio di controparte, si è deciso di rendere ancora più stringente il calcolo dei requisiti patrimoniali sui rischi connessi a eventuali operazioni sugli strumenti finanziari derivati negoziati sul mercato over the counter.

Nel complesso, quindi, la regolamentazione bancaria ha progressivamente cercato di eliminare la discrezionalità dell'attività di prestito in relazione all'attività di trading.

1.8.6 Banche sistemiche (G-SIBs, S-SIIs, SIFIs, D-SIBs, O-SIIs)

La crisi finanziaria del 2007-2009 ha sottolineato il pericolo che il fallimento di istituzioni finanziarie di grandi dimensioni e di rilevanza sistemica possa trasmettersi a catena verso tutte le altre istituzioni, determinando pesanti esternalità negative. A questo proposito, durante la crisi, si sono verificati numerosi salvataggi, connessi a istituzioni finanziarie sistematicamente importanti, che hanno contribuito ad aumentare ulteriormente il livello di attenzione delle diverse Autorità di Vigilanza.

Per questo motivo, il Comitato di Basilea ha deciso di dedicare una particolare attenzione nei confronti delle banche ritenute di importanza sistemica, attribuendo a queste ultime un requisito patrimoniale aggiuntivo, in modo da incrementare il livello di sicurezza contenuto all'interno dello schema regolamentare.

In questo contesto, per semplicità, si usano frequentemente numerosi acronimi diversi, tra i quali è senza dubbio possibile nominare come più rilevanti, G-SIBs, S-SIIs e SIFIs, in modo da descrivere, impiegando concetti estremamente simili, anche se caratterizzati da sfumature leggermente diverse, rispettivamente, le banche globali sistematicamente importanti, o *Global Systemically Important Banks*, le istituzioni globali sistematicamente importanti, o *Global Systemically Important Institutions*, e, infine, le istituzioni finanziarie sistematicamente importanti, o *Systemically Important Financial Institutions*.

Nel complesso, si ritiene che le banche o, più in generale, le istituzioni finanziarie di importanza sistemica siano sempre da considerare *too big to fail*, quindi, letteralmente, "troppo grandi per essere lasciate fallire", a causa degli effetti conseguenti di un potenziale fallimento sul sistema finanziario globale e, di conseguenza, sul sistema economico globale. In poche parole, si suppone che le istituzioni finanziarie, così individuate, siano da salvare, anche nell'eventualità in cui manifestino significative difficoltà economico-finanziarie.

Per individuare le banche da includere nella categoria degli organismi sistematicamente importanti e valutare il relativo grado di rilevanza sistemica, il Comitato di Basilea ha prescritto il calcolo di un indicatore basato su cinque profili diversi: il grado di attività transfrontaliera, la dimensione, l'interconnessione, il grado di insostituibilità dei servizi infrastrutturali e la tipologia di attività. A questo punto, assegnando un punteggio per ciascun profilo e calcolando la media semplice dei cinque punteggi, si ottiene la misura definitiva del livello di rilevanza sistemica, utile per confrontare tra loro banche diverse.

Il risultato è il seguente: a punteggi più elevati si identificano istituzioni finanziarie ritenute di maggiore importanza sistemica e, pertanto, si assegnano requisiti patrimoniali addizionali più consistenti al Common Equity Tier 1, partendo da un minimo dell'1% fino ad arrivare a un massimo del 3,5% degli attivi ponderati per il rischio. Nello specifico, in relazione a quanto discusso precedentemente, è possibile ottenere una classifica, pubblicata annualmente dal *Financial Stability Board*, che consente di distinguere le istituzioni finanziarie di maggior rilevanza sistemica sulla base del capitale aggiuntivo in rapporto agli attivi ponderati per il rischio, individuando le classi 1%, 1,5%, 2%, 2,5% e 3,5%.

A volte si usano anche gli acronimi D-SIBs per menzionare le banche nazionali sistematicamente importanti, o *Domestic Systemically Important Banks*, e O-SIIs per menzionare le ulteriori istituzioni di importanza sistemica, o *Other Systemically Important Istitutions*. Queste istituzioni finanziarie possono essere assoggettate a maggiori vincoli regolamentari, principalmente in termini di requisiti patrimoniali, obblighi informativi e stress test.

1.9 Basilea IV (2017)

La crisi finanziaria del 2007-2009 ha rappresentato la peggior crisi dopo quella del 1929, con effetti devastanti e permanenti sul sistema economico globale per la maggior parte dei Paesi internazionali. Questo episodio ha profondamente sconvolto il pensiero di tutte le istituzioni finanziarie, portando il fallimento di alcune e il salvataggio di altre.

In questo contesto, il Comitato di Basilea e i diversi governi nazionali hanno lavorato per intervenire urgentemente, sviluppando una profonda revisione della regolamentazione bancaria, in modo da superare con successo i limiti presentati dalla precedente normativa. Basilea III, infatti, ha cercato di incrementare significativamente il patrimonio di vigilanza che le banche devono costantemente detenere per far fronte ai rischi verso cui sono esposte, con alcune modifiche importanti nelle principali dimensioni della struttura regolamentare, risultate particolarmente problematiche durante la crisi.

Nel dicembre 2017, il Comitato di Basilea ha inaugurato un passaggio ulteriore della regolamentazione bancaria, istituendo un nuovo pacchetto di riforme, ovvero Basilea IV. Le Autorità di Vigilanza, nel corso del tempo, hanno riconosciuto le molteplici accuse rivolte nei confronti dei sistemi di rating interni delle banche, responsabili di non possedere una struttura sufficientemente confrontabile, oltre che generare attivi ponderati per il rischio molto diversi, anche a parità di esposizione e controparte creditizia.

La normativa, prendendo in considerazione tutti questi aspetti, ha deciso di dedicare la maggior parte delle attenzioni nei confronti dei sistemi di rating interni delle banche e concentrare in questo punto le principali innovazioni dello schema regolamentare. Per questo motivo, è facile intuire il motivo per cui questa direttiva, chiamata principalmente con il nome di Basilea IV, viene definita, in alternativa, anche con il nome di Basilea III.5, osservando che non si tratta di un'autentica rivoluzione, quanto piuttosto di un giusto completamento della regolamentazione bancaria di Basilea III.

In ogni caso, l'applicazione della normativa è prevista, con un inserimento graduale a seconda dei vari Paesi internazionali, nel corso di tutto il 2022.

1.9.1 Principali modifiche

A questo punto, è possibile mettere in evidenza i principali aggiustamenti disposti da Basilea IV, focalizzando, ovviamente, l'attenzione sulle proposte legate al rischio di credito.

Rischio di credito: approccio standard

Prima di tutto, in relazione all'approccio standard, si introduce una notevole modifica alla struttura sottostante, in modo da aumentarne la sensibilità e riflettere in modo migliore il rischio effettivo connesso alle singole esposizioni creditizie. Per ottenere questo risultato, si utilizza una maggiore segmentazione per il calcolo del coefficiente di ponderazione da applicare per determinare il valore complessivo delle attività ponderate per il rischio delle diverse classi di attività finanziarie.

In aggiunta, si impone di adottare ulteriori miglioramenti alla normativa: da una parte, una riduzione del coefficiente di ponderazione all'85% per i crediti alle medie imprese;

dall'altra, un incremento della scala di ponderazioni, inserendo una nuova ponderazione, pari al 75%, assente precedentemente.

Tuttavia, in alcuni Paesi internazionali, si rimuove la possibilità di impiegare, per analizzare un'esposizione creditizia, i rating esterni prodotti da agenzie specializzate, imponendo alla banca di generare in completa autonomia delle valutazioni indipendenti, organizzate in categorie di rischio crescente in funzione della natura della controparte.

Così facendo, si dispone di utilizzare contemporaneamente tre categorie di rischio, denominate "A", "B" e "C", per esposizioni creditizie verso istituzioni finanziarie, nonché due categorie di rischio, denominate, "investment grade" e "altri", per esposizioni creditizie verso imprese non finanziarie. Questa scelta dipende dal volere delle Autorità di Vigilanza di eliminare dalla regolamentazione bancaria i riferimenti ai rating esterni assegnati da agenzie specializzate, considerati non sufficientemente adeguati.

Rischio di credito: approccio IRB

In seguito, per quanto riguarda l'approccio dei sistemi di rating interni, si elimina la possibilità di utilizzare la variante *Advanced* dei sistemi di rating interni, che autorizza le banche a stimare internamente non solo la PD ma anche la EAD, la LGD e la Maturity, per esposizioni creditizie verso imprese non finanziarie di grandi dimensioni e banche.

Per valutare questa tipologia di crediti, quindi, si impone l'obbligo di impiegare misure di EAD e LGD stabilite in modo completamente indipendente dal Comitato di Basilea e valide nella variante *Foundation* dei sistemi di rating interni. Spiegare la motivazione di questa scelta è relativamente facile, in quanto, analizzando il portafoglio complessivo di una banca, è facile osservare come, nel caso di esposizioni creditizie nei confronti di imprese non finanziarie di grandi dimensioni e banche, notoriamente di grande qualità, i default rappresentino un evento estremamente raro.

Con questi dati a disposizione, è possibile dimostrare la difficoltà nel produrre inferenze statistiche affidabili e procedure di backtesting robuste, rendendo problematico per le Autorità di Vigilanza il compito di verificare la validità delle stime prodotte dalle singole banche. In questo caso, per aumentare il livello di sicurezza, si è preferito modificare la normativa e rimandare all'utilizzo dell'approccio *Foundation*.

Allo stesso modo, si è stabilito di abolire la possibilità di usare stime interne per gli investimenti in azioni, oltre che modificare lo schema di calcolo del capitale regolamentare, aggiungendo una serie di floor, ovvero un insieme di valori minimi inderogabili, in relazione ad alcuni parametri di rischio prodotti dai sistemi di rating interni delle banche.

In ultima analisi, Basilea IV prevede un aggiornamento nel computo del requisito patrimoniale obbligatorio per il rischio di credito, in quanto si richiede che il capitale regolamentare calcolato con l'approccio standard diventi vincolante anche per le banche che scelgono di utilizzare l'approccio dei sistemi di rating interni.

Questo meccanismo, definito output floor, si ispira alla volontà di imporre un limite minimo inderogabile al risultato finale del calcolo del capitale regolamentare, senza richiedere nessun limite intermedio per i singoli parametri di rischio. In sostanza, la nuova normativa impone che il patrimonio di vigilanza, misurato con i sistemi di rating interni, non possa mai scendere al di sotto di una specifica percentuale del patrimonio di vigilanza, misurato, viceversa, con l'approccio standard.

Pertanto, i requisiti patrimoniali complessivi di una banca saranno sempre pari al massimo tra quelli determinati con i sistemi di rating interni e una percentuale, fissata dalle Autorità di Vigilanza, di quelli determinati con l'approccio standard. In questo contesto, il valore target della percentuale è impostato inizialmente al 50% nel 2022, ma è destinato a salire progressivamente, arrivando a un valore del 72,5% nel 2027.

Oltre che per il rischio di credito, si programma di applicare questa metodologia anche per la valutazione di altre tipologie di rischio, tra cui rischio di mercato e rischio operativo.

Banche sistemiche

Per le banche di rilevanza sistemica, si stabilisce di incrementare il vincolo relativo alla leva finanziaria, ovvero il *Plain Leverage Ratio*, che prescrive un valore minimo del 3% nel rapporto tra Tier 1 Capital ed esposizione complessiva. A questo proposito, si prevede di considerare il grado di importanza sistemica, aumentando il valore minimo obbligatorio del *Plain Leverage Ratio* in misura pari alla metà del cuscinetto di capitale aggiuntivo richiesto in termini di requisito patrimoniale sugli attivi ponderati per il rischio.

Così facendo, una banca, che subisce un requisito patrimoniale ulteriore rispetto al livello minimo di capitale previsto dai propri attivi ponderati per il rischio, è obbligata a rispettare per il *Plain Leverage Ratio* un vincolo superiore rispetto al valore originale, pari al 3%, con l'aggiunta della metà del requisito patrimoniale aggiuntivo.

1.10 Normativa sulla risoluzione

In ultima istanza, l'obiettivo è quello di affrontare, seppur in modo sintetico, la normativa sulla risoluzione delle banche e sui relativi salvataggi governativi, in modo da completare il quadro generale sulla struttura della regolamentazione bancaria attuale. Questi concetti non sono direttamente collegati ai modelli di risk management o ai requisiti patrimoniali obbligatori, ma rappresentano, comunque, un argomento di grande importanza, che, nel corso del tempo, ha influenzato significativamente l'intero schema regolamentare.

Durante la crisi finanziaria del 2007-2009, i governi dei principali Paesi internazionali, giustificati dalla necessità di salvare le istituzioni finanziarie in condizioni di insolvenza, sono stati costretti a intervenire a sostegno di queste ultime, indebolendo fortemente gli incentivi alla disciplina di mercato. In questo contesto, infatti, la maggior parte degli investitori, che, in precedenza, avevano sottoscritto obbligazioni bancarie subordinate, oltre che strumenti innovativi di capitale e strumenti ibridi di capitale, hanno riportato perdite limitate, grazie ai salvataggi bancari finanziati con soldi pubblici.

Gli interventi a supporto delle istituzioni finanziarie in difficoltà si contrappongono con decisione al principio di "no bail-out", ovvero nessun salvataggio, che costituisce un concetto indispensabile per un corretto funzionamento della disciplina di mercato.

Per risolvere questa incongruenza, si è deciso di introdurre una serie di normative per non replicare in nessun caso la situazione sperimentata nel corso della crisi finanziaria, dove i contribuenti hanno dovuto sopportare la maggior parte dei costi associati, mentre gli investitori hanno potuto conservare in larga misura il valore originario dell'investimento.

La scelta di definire normative sulla risoluzione delle banche risponde, quindi, all'esigenza di costruire uno schema di regole, in modo da assicurare, per le banche insolventi, un'uscita diligente e ordinata dal mercato, consentendo congiuntamente di raggiungere due obiettivi specifici: assoggettare a perdite gli investitori e proteggere il sistema economico da pesanti conseguenze sistemiche.

A questo proposito, è possibile menzionare due normative:

- *Total Loss Absorbtion Capacity* (TLAC);
- *Minimum Requirement of Eligible Liabilities* (MREL).

Queste discipline, a differenza dei requisiti patrimoniali tradizionali, non si propongono di ottimizzare la gestione dei rischi e, conseguentemente, minimizzare la probabilità di insolvenza di una banca, quanto piuttosto di ridurre l'impatto che una possibile uscita dal mercato possa esercitare sul funzionamento del resto del sistema economico.

Il *Total Loss Absorbtion Capacity*, o TLAC, è una normativa sulla risoluzione introdotta per la prima volta dal Comitato di Basilea nel 2015, la quale prevede l'obbligo, per le banche internazionali di importanza sistemica G-SIBs, di predisporre un cuscinetto minimo, composto da capitale o altre passività, per assorbire le perdite in caso di bisogno.

Nello specifico, questa normativa prescrive di soddisfare simultaneamente due requisiti di capitale regolamentare: un requisito espresso in funzione degli attivi ponderati per il rischio, fissato inizialmente al 16% ma destinato ad aumentare gradualmente fino al 18%, e un requisito espresso in funzione del totale dell'attivo contabile, stabilito inizialmente al 6% ma programmato per incrementare gradualmente fino al 6,75%.

In sintesi, una generica banca di rilevanza sistemica, per svolgere la propria attività, deve essere sempre in grado di presentare il soddisfacimento di entrambi i requisiti patrimoniali, in aggiunta ai cuscinetti di capitale integrativi, previsti dalla regolamentazione bancaria, come il *Capital Conservation Buffer*, il *Counter-Cyclical Buffer* e il cuscinetto di capitale riservato alle banche sistemiche.

Per soddisfare questi vincoli patrimoniali, si ammette di utilizzare dotazioni di capitale, quali Common Equity Tier 1, Additional Tier 1, Tier 2 Capital e altri specifici titoli di debito unsecured, ossia titoli di debito subordinati e con vita residua maggiore di un anno. In poche parole, è possibile avvalersi di strumenti finanziari che, in caso di risoluzione, possono essere convertiti in azioni e, conseguentemente, abbattuti per perdite.

Il *Minimum Requirement of Eligible Liabilities*, o MREL, è una normativa istituita dall'Unione Europea nel 2014, nell'ambito della *Bank Recovery and Resolution Directive*, la quale stabilisce una regolamentazione sulla risoluzione delle banche rilevanti, imponendo di detenere un ammontare di patrimonio e altre passività, adeguato a essere eliminato in caso di necessità per contrastare le perdite.

In particolare, questa normativa, applicabile a tutte le banche e non soltanto alle banche sistemiche, prescrive di rispettare tre diverse componenti di capitale: un primo aggregato, composto da patrimonio e passività, disponibile per assorbire le perdite, un secondo aggregato per assicurare una corretta capitalizzazione in caso di perdite e, infine, un terzo aggregato per mantenere in via precauzionale una protezione ulteriore. In questo caso, per adempiere a questi vincoli patrimoniali, si consente di utilizzare, in aggiunta agli strumenti finanziari descritti precedentemente, anche passività non subordinate.

In tema di aiuti di Stato, si applica il principio della condivisione degli oneri, definito burden sharing, ai sensi del quale, per la copertura delle perdite, è necessario che l'intervento del governo nel capitale di un'istituzione finanziaria in difficoltà sia sempre subordinato al completo utilizzo degli strumenti finanziari validi come capitale regolamentare.

Questo meccanismo stabilisce automaticamente che, prima di prevedere un'operazione di "bail-out", ovvero utilizzare soldi pubblici per il salvataggio di un'istituzione finanziaria in stato di insolvenza, è sempre necessario applicare un'operazione di "bail-in", in modo da convertire in azioni e, eventualmente, abbattere per perdite una parte delle passività, senza dover ricorrere alla messa in liquidazione. In questo contesto, i requisiti patrimoniali obbligatori si impegnano a garantire che le istituzioni finanziarie possano disporre di un volume di capitale sufficientemente adeguato a raggiungere questo obiettivo.

1.11 Considerazioni finali

In generale, la regolamentazione bancaria moderna, nel corso del tempo, dalla sua versione originale del 1988, Basilea I, alla sua versione attualmente definitiva del 2017, Basilea IV, è riuscita con successo a rimediare alle principali mancanze della normativa prudenziale del settore bancario. A questo punto, illustrati i principali aspetti positivi e negativi delle diverse riforme nelle rispettive sezioni, è possibile effettuare alcune riflessioni conclusive.

In questo percorso, la maggior parte delle disposizioni emanate dal Comitato di Basilea si è dimostrata essere un corpus di interventi auspicabili e desiderabili, ciascuno dei quali necessario per provare a rispondere ai problemi del suo tempo, con particolare riferimento alla crisi finanziaria del 2007-2009, colpevole di aver sconvolto in modo irreversibile tutto il sistema economico.

In questo contesto, però, alcune prescrizioni non hanno rappresentato un autentico salto di qualità nella struttura regolamentare, quanto piuttosto un maldestro tentativo, non giustificato dal punto di vista teorico, di introdurre regole il più possibile semplici e uniformi, in modo da contenere l'autonomia che, la regolamentazione bancaria, dopo la riforma di Basilea II, aveva progressivamente concesso alle banche, con la possibilità di progettare sistemi di misura e controllo del rischio in linea con le proprie caratteristiche.

In realtà, la crisi finanziaria del 2007-2009 non è scaturita dall'applicazione di modelli di risk management sempre più efficaci ed efficienti, che, sebbene costituiscano strumenti sicuramente ancora acerbi e recenti, hanno permesso di ottenere notevoli risultati, ma a causa di strategie gestionali inadeguate presenti nel modello di business delle principali banche internazionali, indirizzate verso l'ottenimento di profittabilità di breve periodo, con strumenti di ingegneria finanziaria di difficile valutazione a causa della loro complessità, e non verso la costruzione di una profittabilità sostenibile nel medio-lungo periodo.

Pertanto, è possibile precisare come, per fare in modo che la prossima crisi finanziaria sia scongiurata, o quantomeno, superata con modeste conseguenze, non si debba sostituire, in modo semplicistico, i modelli di risk management con banali regole di buon senso, ma spingere per continuare il rapporto tra banche e Autorità di Vigilanza, lavorando per migliorare gli strumenti di misura e controllo del rischio e, soprattutto, per integrare questi ultimi nel processo decisionale delle banche, il quale dovrebbe rappresentare, dopotutto, il vero e proprio obiettivo finale della regolamentazione bancaria.

Capitolo 2

Rischio di credito

2.1 Introduzione

In questo capitolo, si concentra l'attenzione sul rischio di credito, specificando brevemente la definizione generale, illustrando le diverse tipologie di rischio contenute al suo interno e, infine, spiegando le componenti fondamentali, ovvero perdita attesa e perdita inattesa.

Nello specifico, si approfondisce l'argomento del rischio di credito, analizzando nel dettaglio la distribuzione di probabilità associata alle perdite, in modo da introdurre le principali metodologie per la stima dei parametri alla base della perdita attesa, quali *Exposure at Default* (EAD), *Probability of Default* (PD) e *Loss Given Default* (LGD), unitamente a spiegare la logica alla base della perdita inattesa.

Come risultato: da una parte, si analizza, in ordine di complessità relativa, l'insieme degli ingredienti della perdita attesa, cominciando dalla *Exposure at Default*, tramite la possibile distinzione tra esposizioni creditizie a valore certo e a valore incerto, proseguendo con la *Loss Given Default*, tramite la descrizione delle principali determinanti, oltre che delle tecniche Market LGD e Workout LGD, e concludendo con la *Probability of Default*, autentico obiettivo del presente lavoro di tesi; dall'altra, si esamina la perdita inattesa, esplorandone il collegamento sia con il concetto di Value at Risk, abbreviato VaR, che con il concetto, alla base della regolamentazione bancaria, di capitale regolamentare.

2.2 Definizione di rischio di credito

Il rischio di credito rappresenta, senza dubbio, la principale tipologia di rischio per gli intermediari finanziari e costituisce il rischio che una controparte non sia in grado di adempiere ai suoi obblighi in merito al rimborso del capitale e al pagamento degli interessi. Nello specifico, il rischio di credito è formalmente definito come "*la possibilità che una variazione inattesa del merito creditizio di una controparte generi una corrispondente variazione inattesa del valore corrente della relativa esposizione creditizia*"[1].

[1] A. Resti e A. Sironi. *Rischio e valore nelle banche.* 2ª ed. Milano: EGEA, 2021 [108].

Questa definizione risulta estremamente semplice e puntuale, ma comprende alcuni concetti impliciti di fondamentale importanza, per i quali è necessario effettuare una serie di approfondimenti e precisazioni generali al fine di illustrarne il dominio di interesse.

In primo luogo, è possibile evidenziare come il rischio di credito comprenda due scenari di rischio diversi: il rischio di insolvenza e il rischio di migrazione. La definizione di rischio di credito, infatti, non può essere limitata solamente alla possibilità che, in un determinato intervallo di tempo, si verifichi un'insolvenza della controparte, ma deve necessariamente includere anche il deterioramento del relativo merito creditizio.

A questo proposito, non si può affrontare il rischio di credito considerando una semplice distribuzione di probabilità binaria dei possibili eventi, ovvero analizzando esclusivamente l'evento insolvenza e l'evento non insolvenza. Al contrario, per studiare correttamente il rischio di credito, si deve esaminare una distribuzione di probabilità, discreta o continua, nella quale l'insolvenza costituisce esclusivamente l'evento estremo.

In poche parole, si tratta di una gestione molto più complessa: la distribuzione di probabilità deve inevitabilmente considerare l'evento insolvenza, ma, allo stesso tempo, deve anche considerare la variazione del merito creditizio, prevedendo una serie di eventi in cui la controparte rimane solvibile, ma la relativa probabilità di insolvenza diventa, progressivamente, sempre più elevata. Così facendo, si può ottenere una rappresentazione decisamente migliore della realtà associata a una generica esposizione creditizia, considerando che una diminuzione della qualità creditizia della controparte, pur non comportando necessariamente l'insolvenza, contribuisce direttamente a ridurne il valore economico.

In secondo luogo, è possibile sottolineare come il rischio di credito, analizzando la definizione generale di rischio, debba essere valutato e misurato studiando una variazione obbligatoriamente inattesa del merito creditizio della controparte. Per discutere in modo corretto di rischio in ambito risk management, è necessario tenere presente che la reale componente di rischio associata a un evento futuro è rappresentata dalla possibilità che i risultati ex-post possano essere profondamente diversi rispetto a quelli immaginati ex-ante.

In questo contesto, il rischio di credito è connesso al fatto che, nel corso del tempo, si possa verificare un eventuale deterioramento della qualità creditizia della controparte assolutamente non prevedibile dall'istituzione finanziaria al momento della concessione del credito. In sostanza, la variazione della qualità creditizia della controparte deve essere inattesa, dal momento che l'evoluzione attesa delle condizioni economico-finanziarie viene sempre opportunamente valutata in anticipo, sia nella decisione di affidamento che nella determinazione della probabilità di insolvenza e, conseguentemente, del pricing del credito.

In ultima analisi, è possibile approfondire il concetto di esposizione creditizia, con la possibilità di chiarire fin da subito, definitivamente e adeguatamente, due ulteriori aspetti importanti per lo studio dello scenario associato al rischio di credito.

Prima, è possibile osservare come il concetto di esposizione creditizia debba essere interpretato con ampiezza e flessibilità di significato, poiché il rischio di credito non interessa solamente gli impieghi tradizionali di una banca, come i titoli o i prestiti in bilancio, ma coinvolge anche numerose applicazioni diverse, quali le posizioni fuori bilancio, gli strumenti finanziari derivati e le transazioni in titoli o in valute.

Dopo, è possibile constatare come il concetto di esposizione creditizia non si riferisca al valore storico, bensì al valore corrente, o *fair value*, in modo da utilizzare valutazioni basate sul valore economico per ottenere una corretta valutazione del rischio di credito.

2.3 Tipologie di rischio di credito

Il rischio di credito, come anticipato in precedenza, racchiude al suo interno una definizione sensibilmente estesa, che non consente un'immediata comprensione del problema in esame. Tuttavia, è possibile definire in modo molto più approfondito il dominio di interesse, specificando la possibile composizione dell'esposizione totale verso il rischio di credito alla quale possono andare incontro gli intermediari finanziari nel portare avanti la normale operatività sul mercato finanziario.

A tale proposito, il rischio di credito, in relazione al proprio campo di esistenza, incorpora le seguenti tipologie di rischio:

- rischio di insolvenza;
- rischio di migrazione;
- rischio di recupero;
- rischio di esposizione;
- rischio di spread;
- rischio di sostituzione;
- rischio di regolamento;
- rischio Paese.

Il rischio di insolvenza è il rischio associato all'insolvenza della controparte e, quindi, rappresenta il rischio che, nell'ambito di un'operazione creditizia, il debitore non sia in grado di adempiere, in totale o anche solo in parte, ai suoi obblighi di rimborso del capitale e di pagamento degli interessi nei confronti del creditore.

Il rischio di migrazione, chiamato anche rischio di *downgrading*, è il rischio associato al deterioramento del merito creditizio della controparte, con conseguente diminuzione del valore economico della relativa esposizione creditizia.

Il rischio di recupero è il rischio che il valore economico dell'ammontare concretamente recuperato da una controparte rivelatasi insolvente risulti, in qualche misura, inferiore rispetto a quanto previsto originariamente, generando una differenza che può essere spiegata analizzando la volatilità delle variabili che caratterizzano il processo di recupero, ovvero valore di realizzo, durata e tasso di interesse.

Il rischio di esposizione è il rischio che l'ammontare relativo a un'esposizione creditizia di una controparte subisca un aumento significativo in prossimità dell'insolvenza, portando a un incremento dell'esposizione creditizia complessiva.

Il rischio di spread è il rischio connesso a un eventuale aumento dell'avversione al rischio degli investitori, i quali, anche a parità di merito creditizio della controparte, potrebbero richiedere un maggior premio per il rischio e causare un conseguente incremento degli spread domandati dal mercato finanziario, costringendo, potenzialmente, il valore economico dell'esposizione creditizia a subire una diminuzione, anche senza riscontrare un peggioramento nella qualità creditizia della controparte.

Il rischio di sostituzione, chiamato anche rischio di pre-regolamento, è il rischio che la controparte di una specifica transazione in strumenti finanziari derivati, negoziati sul mercato over the counter, diventi insolvente prima della scadenza del contratto, obbligando a sostituire la posizione sul mercato finanziario a condizioni contrattuali differenti.

Il rischio di regolamento è il rischio che la controparte di una specifica transazione in strumenti finanziari derivati, negoziati sul mercato over the counter, risulti insolvente alla scadenza del contratto, non adempiendo ai propri obblighi di consegna o di pagamento.

Il rischio Paese è il rischio che una controparte non sia in grado di adempiere ai propri obblighi a causa di imposizioni esterne di natura politica o legislativa.

2.4 Componenti del rischio di credito

In termini generali, un rischio finanziario nasce a causa della situazione di incertezza nella quale gli operatori del mercato finanziario si trovano costretti a prendere le proprie decisioni economico-finanziarie, i cui risultati ex-post possono essere notevolmente diversi rispetto a quelli attesi ex-ante. Questa incertezza dipende dalla pluralità di risultati attesi che si possono manifestare in conseguenza delle decisioni assunte, generando un concetto di rischio comprendente il fatto che, sul mercato finanziario, gli eventi che si possono verificare ex-post potrebbero essere diversi da quelli che si erano immaginati ex-ante.

Globalmente, quindi, un rischio finanziario considera una generica distribuzione di probabilità e si concentra sia sulla parte positiva della distribuzione, che comprende gli eventi favorevoli, sia sulla parte negativa, che comprende invece gli eventi sfavorevoli. A questo proposito, la metrica più utilizzata per la valutazione di un rischio è la varianza o, in alternativa, lo scarto quadratico medio, in modo da considerare simultaneamente la volatilità connessa a entrambe le code della distribuzione.

Il rischio di credito riprende la definizione generale di rischio, anche se richiede una precisazione ulteriore, poiché, per studiare e valutare correttamente il rischio di credito, è necessario effettuare un passaggio concettuale fondamentale, ovvero introdurre una nuova metrica per il rischio. In questo contesto, infatti, non è possibile impiegare metriche per il rischio che tengano conto di entrambe le code della distribuzione, bensì è importante impiegare metriche per il rischio che tengano conto soltanto di una coda della distribuzione, come la semivarianza o il *Lower Partial Moment* (LPM).

Per questo motivo, nell'ambito del rischio di credito, è indispensabile lavorare con un concetto di rischio asimmetrico e, più specificatamente, un concetto di rischio sfavorevole, in modo da trascurare completamente gli eventi positivi e prendere in considerazione soltanto gli eventi negativi. In particolare, si è interessati ai casi in cui il sistema finanziario faccia evolvere le variabili di mercato in modo sfavorevole, focalizzando i ragionamenti esclusivamente sulla parte negativa della distribuzione e, dunque, sulle perdite.

Per esplorare il rischio di credito è necessario distinguere due componenti fondamentali:

- perdita attesa (*Expected Loss*, EL);
- perdita inattesa (*Unexpected Loss*, UL).

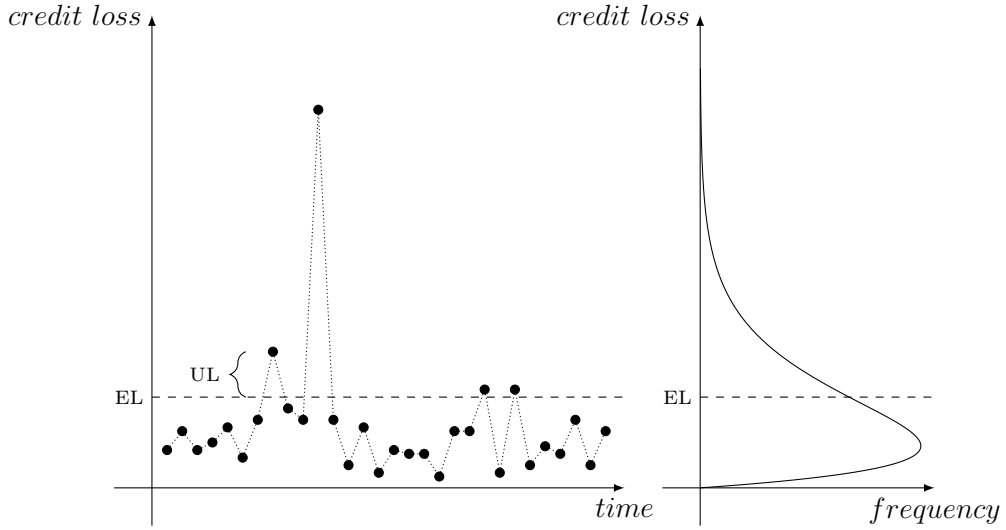


Figura 2.1. Distinzione tra perdita attesa e perdita inattesa.

2.5 Perdita attesa

La perdita attesa è definita come il valore medio della distribuzione delle perdite. Questa, per definizione, non rappresenta un rischio in senso stretto, dal momento che può essere stimata ex-ante dall'intermediario finanziario in questione e, conseguentemente, coperta aggiungendo un adeguato spread sul tasso di interesse attivo, in modo tale che, nel caso in cui si dovesse verificare esattamente quanto previsto, si possa ottenere ex-post proprio il rendimento netto prefissato al momento della concessione del credito.

In poche parole, la perdita attesa costituisce una perdita caratterizzata generalmente da un'elevata frequenza e da un importo relativamente contenuto, che, a causa di queste peculiarità, attraverso specifici modelli statistici, risulta in qualche modo misurabile e calcolabile. Per questo motivo, la perdita attesa è sostanzialmente considerabile come un costo inevitabile da sostenere per stare nel business e, come risultato, deve essere assorbita dal Conto Economico, tramite un accantonamento per rischi.

A questo punto, è possibile introdurre il problema della stima della perdita attesa, in quanto, per valutare in modo corretto la perdita attesa di un'esposizione creditizia, si richiede contemporaneamente il calcolo di tre parametri fondamentali:

- esposizione in caso di insolvenza (*Exposure at Default*, EAD);
- probabilità di insolvenza (*Probability of Default*, PD);
- tasso di perdita in caso di insolvenza (*Loss Given Default*, LGD).

In formule:

$$EL = \overline{EAD} \cdot PD \cdot \overline{LGD} \ . \tag{2.1}$$

In questo contesto, si ipotizza che i diversi fattori di rischio siano tra loro indipendenti, in modo da semplificare il ragionamento, in quanto, se così non fosse, per il calcolo della perdita attesa non sarebbe più sufficiente la stima della PD e il calcolo del valore atteso di EAD e LGD, ma, piuttosto, sarebbe necessario conoscere anche le covarianze tra i vari fattori di rischio, rischiando di complicare sensibilmente il problema in questione.

2.5.1 Exposure at Default (EAD)

L'esposizione in caso di insolvenza, abbreviata EAD dall'inglese *Exposure at Default*, è una variabile casuale che quantifica l'ammontare che può essere perso in caso di default della controparte. A questo proposito, l'esposizione in caso di insolvenza rappresenta una variabile che può assumere un certo livello di aleatorietà, dal momento che il suo valore dipende direttamente dalle caratteristiche dell'esposizione creditizia e, più nello specifico, dalla forma tecnica del finanziamento concesso.

In generale, un'esposizione creditizia può essere caratterizzata secondo due modalità: da un valore certo, quando è possibile conoscere ex-ante il valore esatto dell'esposizione complessiva, oppure da un valore incerto, quando invece è possibile conoscere il valore esatto dell'esposizione complessiva solamente ex-post. Con questa distinzione, a seconda che si tratti di un'esposizione creditizia a valore certo o a valore incerto, l'esposizione in caso di insolvenza può essere più o meno facilmente calcolabile.

Infatti, nel caso di un'esposizione a valore certo, l'esposizione in caso di insolvenza risulta deterministica e, conseguentemente, quantificabile in modo relativamente semplice e veloce, non essendo prevista alcuna rettifica ulteriore. Così, per fare un esempio, è possibile menzionare il caso di un mutuo, dove l'esposizione creditizia è sempre definita al momento della concessione dello stesso, senza nessuna condizione di incertezza.

In questo caso, il cliente non ha alcuna discrezionalità circa il piano di rimborso del finanziamento, in quanto l'esposizione creditizia è formata dal debito residuo previsto, periodo di tempo per periodo di tempo, secondo il piano di ammortamento del prestito.

Un ragionamento molto simile è applicabile a un titolo obbligazionario a tasso fisso, poiché, anche in questo caso, non si verifica nessuna condizione di incertezza, essendo i relativi flussi di cassa interamente fissati al momento dell'emissione.

Al contrario, nel caso di un'esposizione a valore incerto, a causa della presenza di una componente aleatoria, l'esposizione in caso di insolvenza non può essere calcolata immediatamente, ma necessita di una modifica ulteriore. Così, per fare un esempio, è possibile segnalare il caso di un'apertura di credito in conto corrente, dove una banca accetta di mettere a disposizione una determinata quantità di fondi a una controparte.

In questo caso, con l'apertura di un fido in conto corrente, la controparte beneficia di un'opzione a proprio favore per poter variare, entro specifici limiti, la dimensione del finanziamento, generando una situazione di incertezza. Pertanto, proprio a causa della discrezionalità concessa alla controparte di scegliere quale porzione di fido in conto corrente utilizzare, la dimensione effettiva dell'esposizione creditizia per cassa potrebbe oscillare nel tempo per effetto di decisioni esterne alla banca.

Questa potenziale aleatorietà si ripercuote inevitabilmente sul calcolo dell'esposizione in caso di insolvenza, dal momento che è possibile verificare come le controparti che entrano in difficoltà economico-finanziarie abbiano la tendenza a utilizzare la massima

disponibilità del fido in conto corrente, generando il rischio che l'esposizione aumenti, anche sensibilmente, in prossimità del default e superi le previsioni originarie.

Una medesima argomentazione è applicabile a uno strumento finanziario derivato, negoziato sul mercato over the counter, essendo il valore del relativo contratto dipendente dall'evoluzione nel tempo di determinati fattori di mercato.

In sostanza, per il calcolo dell'esposizione in caso di insolvenza a proposito di esposizioni creditizie a valore incerto, è necessario conoscere tre componenti essenziali, che specificano un finanziamento di questo tipo: la quota utilizzata, o DP dall'inglese *Drawn Portion*, la quota inutilizzata, o UP dall'inglese *Undrawn Portion*, e un fattore di conversione del credito, chiamato UGD *Usage Given Default* o, altrimenti, CCF *Credit Conversion Factor*.

In formule:

$$\overline{EAD} = DP + UP \cdot UGD \ . \tag{2.2}$$

Nello specifico, DP rappresenta la porzione di finanziamento già utilizzata al momento della valutazione, UP la porzione di finanziamento non ancora utilizzata e, infine, UGD il fattore di conversione del credito, ovvero la percentuale della quota di finanziamento ancora disponibile che si prevede venga utilizzata verso l'insolvenza.

A questo riguardo, occorre aggiungere un'osservazione relativa al pricing del credito. La scelta di includere nell'esposizione in caso di insolvenza non solo la quota utilizzata, ma anche una percentuale della quota inutilizzata potrebbe portare a stimare una maggiore perdita attesa e, pertanto, a richiedere alla controparte un prezzo più elevato, in modo da compensare anche il rischio implicito nella quota inutilizzata. Così facendo, la banca sarebbe obbligata a scaricare la maggiorazione di prezzo sullo spread applicato alla quota utilizzata, generando tassi di interesse molto elevati e riducendone la competitività.

A questo si aggiunge il fatto che questa situazione potrebbe causare problemi non solo alla singola banca, ma anche al sistema economico nel suo complesso. L'applicazione di spread superiori, infatti, potrebbe coinvolgere in modo particolare le imprese migliori, che generalmente utilizzano una percentuale relativamente modesta della quota totale disponibile e presentano un elevato rapporto tra quota inutilizzata e quota utilizzata.

A questo proposito, per evitare che la banca possa essere progressivamente spinta "fuori mercato", causando conseguenze dannose per l'intero sistema economico, la maggior perdita attesa, calcolata considerando la quota inutilizzata, può essere risolta attraverso due soluzioni diverse, che si differenziano in accordo con il processo seguito per la copertura della maggior perdita attesa.

Come prima soluzione, è possibile utilizzare per la copertura una commissione proporzionale alla quota inutilizzata, chiamata *back-end fee* o *commitment fee*, in modo tale da garantire un corretto valore della perdita attesa attraverso la ripartizione dei costi tra quota utilizzata e quota inutilizzata, utilizzando rispettivamente lo spread contenuto nel tasso di interesse e la commissione legata all'impegno a prestare.

Come seconda soluzione, è possibile emettere per la copertura affidamenti revocabili, che, in caso di necessità, possono essere bloccati e annullati senza preavviso, in modo tale da applicare un tasso di interesse più consono alla qualità creditizia della controparte, unitamente alla possibilità di proteggersi da possibili problemi futuri.

2.5.2 Probability of Default (PD)

La probabilità di insolvenza, abbreviata PD dall'inglese *Probability of Default*, rappresenta la probabilità che si verifichi un'insolvenza nella controparte, ovvero la probabilità che la controparte diventi inadempiente all'obbligazione di restituire il capitale prestato e gli interessi maturati. Questa variabile costituisce un argomento fondamentale, poiché rappresenta storicamente uno dei passaggi più critici da superare, nel tentativo di ottenere una valutazione, quanto più possibile corretta, del rischio di credito.

Per questo motivo, nel corso del tempo, sono state sviluppate diverse metodologie, con l'obiettivo di ottenere una stima sempre più realistica della probabilità di insolvenza. A tale proposito, è possibile distinguere due principali categorie di modelli: i modelli di credit scoring e i modelli basati sull'utilizzo di dati ricavati dal mercato dei capitali.

I modelli di credit scoring sono modelli di natura statistica che analizzano i principali indicatori economico-finanziari di una controparte e, assegnando a questi ultimi una specifica ponderazione proporzionale all'importanza relativa nel prevedere il fenomeno dell'insolvenza, realizzano una valutazione della qualità creditizia sintetizzata in un valore numerico, chiamato credit score, interpretabile come probabilità di insolvenza.

Al contrario, i modelli basati sull'utilizzo di dati ricavati dal mercato dei capitali impiegano le informazioni più importanti pubblicamente disponibili, come prezzo, rendimento e volatilità di azioni e obbligazioni, per effettuare una stima della probabilità di insolvenza.

In questa categoria, è possibile discriminare tra due modelli principali, a seconda del tipo di informazioni utilizzate: il primo utilizza gli spread tra i tassi di rendimento delle obbligazioni e i tassi risk-free, mentre il secondo utilizza i prezzi delle azioni e applica il modello di pricing delle opzioni sviluppato da F. S. Black, M. S. Scholes e R. C. Merton.

In questo lavoro di tesi, si è deciso di dedicare particolare attenzione ai modelli di credit scoring, considerati generalmente, valutandone la possibilità di fornire indicazioni più attendibili e più significative circa lo stato di salute economico-finanziaria di una controparte, la soluzione migliore per risolvere il problema oggetto della trattazione.

2.5.3 Loss Given Default (LGD)

Il tasso di perdita in caso di insolvenza, o LGD dall'inglese *Loss Given Default*, rappresenta la percentuale dell'esposizione creditizia che una banca prevede di non riuscire a recuperare in caso di default della controparte. A questo proposito, il tasso di perdita in caso di insolvenza può essere definito come il complemento a uno del tasso di recupero in caso di insolvenza, o RR dall'inglese *Recovery Rate*.

In formule:

$$LGD = 1 - RR \ . \tag{2.3}$$

La perdita in caso di insolvenza costituisce un elemento di difficile valutazione, considerando che, purtroppo, non è possibile conoscere il reale valore assunto da questa variabile, né al momento della concessione del credito né al momento dell'insolvenza del credito, e, per questo motivo, può essere calcolata con esattezza solamente al termine dell'intero processo di recupero.

Determinanti Loss Given Default

In generale, la perdita in caso di insolvenza dipende dall'interazione di una pluralità di fattori diversi; tuttavia, in base alla loro natura, è possibile mettere in evidenza quattro principali categorie (tabella 2.1):

- caratteristiche dell'esposizione;
- caratteristiche del debitore;
- caratteristiche della banca;
- fattori esterni macroeconomici.

Le caratteristiche dell'esposizione, essendo responsabili di spiegare in modo privilegiato il comportamento della perdita in caso di insolvenza, rappresentano la categoria più importante e comprendono la presenza di garanzie[2], reali o personali, il grado di subordinazione e la tipologia di contenzioso previsto per il recupero.

Prima di tutto, l'esistenza di eventuali garanzie influenza notevolmente il valore finale della perdita in caso di insolvenza. Le garanzie reali possono essere rappresentate sia da attività finanziarie che da attività non finanziarie e vengono classificate in base al loro grado di liquidità, ossia alla facilità e alla velocità di escussione e liquidazione sul mercato. Al contrario, le garanzie personali possono essere costituite da garanzie fornite da terzi, come ad esempio banche, imprese capogruppo e pubblica amministrazione.

In seguito, il grado di priorità di un'esposizione creditizia, chiamato sinteticamente *seniority*, costituisce un ulteriore elemento responsabile di modificare fortemente il valore della perdita in caso di insolvenza, disaggregando e differenziando in modo sensibile tra esposizione creditizia *senior*, specificata da una massima priorità di soddisfacimento, ed esposizione creditizia *junior*, subordinata al preventivo soddisfacimento dei vari creditori.

In ultima analisi, è possibile menzionare anche l'impatto derivante dalla tipologia di contenzioso previsto per il recupero, distinguendo tra contenzioso giudiziale in procedura, contenzioso giudiziale non in procedura, contenzioso stragiudiziale e cessione del credito.

Le caratteristiche del debitore raccolgono alcune delle caratteristiche più esplicative del soggetto finanziato e, nello specifico, il settore produttivo, il Paese di origine e i principali indicatori economico-finanziari.

Il settore industriale caratterizza in larga misura la specificità delle attività dell'impresa e, conseguentemente, influenza direttamente il processo di liquidazione, ovvero il processo attraverso il quale le attività dell'impresa possono essere vendute sul mercato e trasformate in risorse finanziarie per i creditori. Per questo motivo, al variare del settore industriale, investimenti più o meno idiosincratici rischiano di compromettere il grado di recuperabilità delle rispettive attività aziendali.

Inoltre, il Paese o la regione geografica dell'impresa determina la legislazione alla quale sottostare per continuare a operare, condizionando fortemente l'efficacia e l'efficienza del

[2] In inglese, sinteticamente *collateral*.

processo di recupero. In questo caso, infatti, Paesi diversi presentano legislazioni diverse e, corrispondentemente, procedure fallimentari diverse, con profonde differenze per quanto riguarda la facilità e la velocità del processo di recupero.

Infine, è possibile sottolineare anche l'impatto risultante da alcuni indici di bilancio, come il rapporto tra attività e passività di terzi oppure il rapporto tra EBITDA e fatturato, che consentono di valutare se una impresa divenuta insolvente sia ancora in grado di generare flussi di cassa in misura consistente.

Le caratteristiche della banca includono fondamentalmente due dimensioni differenti: le politiche aziendali interne e la competenza dell'ufficio legale, che, insieme, contribuiscono a specificare la strategia di gestione adottata.

In primo luogo, le politiche aziendali interne in tema di recupero crediti, come la frequenza al ricorso di accordi stragiudiziali e la tendenza alla cessione di crediti in contenzioso, influenzano notevolmente la perdita in caso di insolvenza. Queste procedure possono essere impiegate e applicate in specifiche situazioni per ottenere una riduzione, anche significativa, della perdita in caso di insolvenza.

L'utilizzo di accordi stragiudiziali e la vendita di crediti in contenzioso, rispetto al processo di liquidazione canonico, riducono il valore nominale dell'ammontare potenzialmente recuperabile, ma, allo stesso tempo, diminuiscono sensibilmente la durata del processo di recupero. In molti casi, il minor tempo di recupero è in grado di compensare o, addirittura, superare il minor valore nominale, generando un effetto economico-finanziario complessivo benefico nel calcolo della perdita in caso di insolvenza.

In secondo luogo, anche la competenza dell'ufficio legale, interno all'impresa, che si occupa del processo di recupero, costituisce un argomento importante. Questa capacità, quantificabile attraverso l'efficacia e l'efficienza dei servizi legali passati, condiziona in modo significativo la perdita in caso di insolvenza.

I fattori esterni macroeconomici contengono, a loro volta, due elementi essenziali: lo stato del ciclo economico e il livello dei tassi di interesse, che caratterizzano in misura equivalente tutte le esposizioni creditizie, essendo in grado di modificare la perdita in caso di insolvenza nell'intero sistema economico.

Il ciclo economico influenza pesantemente il valore finale al quale le diverse attività di un'impresa insolvente potrebbero essere liquidate, considerando che, proprio al variare del ciclo economico, il valore di liquidazione degli attivi aziendali potrebbe essere, anche sensibilmente, diverso. Come regola generale, è possibile dimostrare come, in presenza di una fase del ciclo economico di recessione, il valore di liquidazione degli attivi aziendali potrebbe essere inferiore, mentre, in presenza di una fase del ciclo economico di espansione, il valore di liquidazione degli attivi potrebbe essere superiore.

Il livello dei tassi di interesse influenza il valore attuale dei possibili recuperi futuri. La perdita in caso di insolvenza diverge al variare dei tassi di interesse, dal momento che, per un calcolo preciso, è necessario considerare gli importi recuperati nei diversi periodi di tempo. In questo modo, immaginando costanti tutte le altre condizioni, tassi di interesse elevati riducono il valore attuale e tassi di interesse ridotti aumentano il valore attuale.

In definitiva, è facile osservare come la perdita in caso di insolvenza rappresenti un aspetto di difficile valutazione, dipendente dall'interazione di numerosi fattori diversi, i quali, incidendo autonomamente e rinforzandosi reciprocamente, sono responsabili di determinare diversi valori di LGD, a seconda della tipologia di credito considerato.

Tabella 2.1. Principali determinanti della Loss Given Default.

Loss Given Default	
Categoria	Fattore
Caratteristiche dell'esposizione	Garanzie reali e personali Grado di subordinazione Tipologia di contenzioso
Caratteristiche del debitore	Settore produttivo Paese di origine Indicatori economico-finanziari
Caratteristiche della banca	Politiche aziendali interne Competenza ufficio legale
Fattori esterni macroeconomici	Stato del ciclo economico Livello dei tassi di interesse

Stima Loss Given Default

Per la stima del tasso di perdita in caso di insolvenza è possibile utilizzare due approcci:

- Market LGD;
- Workout LGD.

Market LGD prevede l'utilizzo di dati provenienti dal mercato dei capitali, impiegando, più precisamente, i prezzi delle esposizioni creditizie in default. In questo caso, il calcolo del tasso di perdita in caso di insolvenza viene effettuato sfruttando direttamente il prezzo delle obbligazioni emesse da imprese divenute insolventi, ossia i cosiddetti *distressed bond*, in modo da studiare il prezzo successivo all'insolvenza, il quale riflette le aspettative del mercato finanziario sull'entità e sulle tempistiche del recupero da parte dei creditori, spiegando implicitamente il tasso di perdita in caso di insolvenza.

Questo metodo può essere utilizzato solo in presenza di specifiche esposizioni creditizie dotate di un mercato secondario e, per questa ragione, può essere applicato con difficoltà, unicamente nel caso di imprese in possesso di debiti quotati sul mercato finanziario.

I dati del mercato dei capitali costituiscono, infatti, una fonte d'informazione privilegiata, dal momento che forniscono informazioni obiettive e aggiornate, ma, a causa della loro natura, risultano disponibili solo per obbligazioni emesse da imprese di grandi dimensioni.

Sfortunatamente, questo tipo di esposizioni creditizie sono molto rare nel portafoglio prestiti di una banca tradizionale, considerando che la maggior parte dei prestiti bancari tradizionali non risulta quotata, rendendo impossibile osservare un prezzo di mercato.

Workout LGD, al contrario, prevede l'utilizzo di dati interni della banca, attraverso l'applicazione di un approccio basato sulla misura degli effettivi flussi di cassa di recupero conseguiti nei diversi periodi temporali successivi al default.

Per utilizzare questo metodo, è necessario costruire un database interno, strutturato in maniera adeguata per conservare tutte le informazioni più rilevanti provenienti dalle insolvenze passate, come gli importi recuperati, i tempi di recupero e le procedure seguite.

Producendo questi dati, è possibile raggruppare e segmentare questi ultimi in base al tipo di esposizione, di controparte e di procedura seguita, in modo da identificare un numero finito di cluster con caratteristiche simili e, attraverso questi cluster, stimare il tasso medio di recupero in caso di insolvenza su eventuali default futuri.

In questo contesto, è possibile definire il tasso di recupero in caso di insolvenza nel modo seguente:

$$RR = \frac{\sum_{t=1}^{T} \frac{ER_t - AC_t}{(1+i)^t}}{EAD} \ , \tag{2.4}$$

dove:

- ER = importo recuperato nel periodo t (*expected recovery*);
- AC = costi amministrativi sostenuti nel periodo t (*administrative cost*);
- EAD = esposizione creditizia al momento del default
- i = tasso di attualizzazione;
- T = periodo di tempo stimato per realizzare il recupero.

Analizzando più nel dettaglio, il tasso di recupero in caso di insolvenza deve essere studiato non solo dal punto di vista economico, ma anche dal punto di vista finanziario, dal momento che è necessario tenere in considerazione tutti i fattori che possono contribuire a ridurre il valore economico finale di una specifica esposizione creditizia.

Per questo motivo, è fondamentale calcolare l'effetto finanziario del tempo che trascorre tra il momento dell'insolvenza e il momento dei recuperi, al netto delle diverse voci di costo, dirette e indirette, che una banca deve sostenere per recuperare un'esposizione creditizia in contenzioso. Considerando che i recuperi connessi alle esposizioni creditizie in default possono avere luogo in maniera graduale, con successivi incassi collocati in differenti istanti temporali, la durata del processo di recupero comprende tutti gli eventuali flussi di cassa intermedi tra il periodo di tempo t_1, che segnala l'evento default e l'inizio del processo di recupero, e il periodo di tempo T, che rappresenta, invece, la fine del processo di recupero.

Per applicare correttamente questa formula, occorre studiare in modo approfondito il tasso di attualizzazione, che rappresenta una componente fondamentale per il calcolo del valore attuale dei recuperi al momento del default. A tale riguardo, è possibile scegliere tra quattro tassi di attualizzazione diversi:

- tasso di interesse di trasferimento dei fondi;
- tasso contrattuale del finanziamento;
- tasso di interesse congruo per il rischio;
- tasso di interesse risk-free.

A questo punto, è doveroso effettuare alcune precisazioni in merito a questa distinzione. Innanzitutto, è possibile affermare come il tasso di interesse congruo per il rischio rappresenti senza dubbio l'unico tasso di attualizzazione corretto dal punto di vista teorico. Questo rappresenta, infatti, il tasso di attualizzazione economicamente più opportuno, dal momento che consente di tenere conto del rischio connesso agli importi recuperati, ossia il rischio che si verifica tra il tempo t_1, in cui si è verificato l'evento default, e il tempo T, in cui si è esaurito il processo di recupero, anche se, purtroppo, non risulta direttamente osservabile, oltre che difficilmente stimabile.

In modo alternativo, seppur non completamente adeguato, è possibile ricorrere all'uso del tasso di interesse risk-free, che comporta l'esclusione dalla valutazione del premio per il rischio, con l'ipotesi forte che il rischio sistematico sia implicitamente incluso in altri parametri utilizzati dalla regolamentazione bancaria.

Considerando questi problemi, è necessario valutare alcune scelte alternative, in modo da trovare soluzioni magari meno corrette dal punto di vista teorico, ma, allo stesso tempo, facilmente e concretamente applicabili nella realtà.

Una prima possibilità prevede l'utilizzo del tasso di interesse di trasferimento dei fondi, che rappresenta il tasso applicato dal controllo di gestione delle banche per il calcolo esatto della redditività delle singole filiali e, pertanto, costituisce il costo marginale del funding della banca, ovvero il tasso ombra impiegato per trasferire fondi tra filiali diverse.

Una seconda possibilità consiste nell'utilizzo del tasso contrattuale del finanziamento, che rappresenta il tasso originale del finanziamento stabilito al momento della concessione e, tenuta presente la sua disponibilità, costituisce il tasso più largamente e diffusamente impiegato dalle banche, essendo in grado di assicurare la soluzione di miglior compromesso tra giustificabilità economica e semplicità di utilizzo.

Calcolo Loss Given Default

Nel complesso, il calcolo della perdita in caso di insolvenza rappresenta un compito di difficile risoluzione e, come conseguenza, per ottenerne una stima affidabile è necessario affrontare e superare alcuni problemi statistici, in modo tale da passare dai dati relativi alle insolvenze passate alla stima della perdita in caso di insolvenza futura.

Per effettuare questo passaggio, è fondamentale costruire un database per raccogliere le principali informazioni ottenute dallo studio delle insolvenze al termine del processo di recupero. In questo modo, è possibile descrivere la distribuzione di frequenza empirica dei valori della perdita in caso di insolvenza attraverso l'utilizzo di alcuni indicatori sintetici e, con questi ultimi, effettuare una stima sulle possibili insolvenze future.

A questo punto, un primo tentativo per ottenere una stima del tasso di perdita in caso di insolvenza potrebbe essere effettuato calcolando la media di questa distribuzione. Tuttavia, un procedimento di questo tipo risulta troppo semplicistico e genera un risultato distorto, che può essere al massimo utile per ottenere una stima grossolana e veloce dei possibili tassi di perdita in caso di insolvenza futuri.

Prima di tutto, dal punto di vista concettuale, il risk management non si occupa di studiare il valore atteso di una specifica variabile, quanto piuttosto di analizzarne le deviazioni dal valore atteso, poiché in teoria soltanto queste ultime rappresentano un rischio in senso stretto, come ampiamente discusso in precedenza.

In aggiunta, dal punto di vista statistico, la media non emerge come un indicatore adatto a descrivere le distribuzioni di probabilità del tasso di perdita in caso di insolvenza o, equivalentemente, del tasso di recupero in caso di insolvenza.

In questo contesto, infatti, è possibile osservare due comportamenti completamente diversi: da una parte, alcune esposizioni creditizie, come il leasing o i mutui residenziali, tendono ad avere tassi di recupero molto elevati e vicini al 100%, con un tasso di perdita conseguentemente limitato; dall'altra, invece, altre esposizioni creditizie, come gli scoperti in conto corrente non assistiti da adeguate garanzie, tendono ad avere tassi di recupero molto modesti e prossimi allo 0%, con un tasso di perdita praticamente totale.

Per questa ragione, l'analisi delle distribuzioni statistiche consente di evidenziare come la maggior parte dei valori provenienti dalle osservazioni passate tenda ad addensarsi intorno al valore zero o al valore uno, con una probabilità molto elevata di sperimentare recuperi rispettivamente pari allo 0% o al 100% e una probabilità decisamente ridotta di sperimentare recuperi intermedi.

In poche parole, la distribuzione di probabilità del tasso di perdita in caso di insolvenza e del tasso di recupero in caso di insolvenza segue una distribuzione bimodale, caratterizzata da una forma tipica a "U", con una conseguente probabilità di osservare valori prossimi alla media relativamente modesta. Con distribuzioni di probabilità bimodali, i cui valori sono prevalentemente concentrati ai due estremi, infatti, il valore medio non produce un grande significato e non rappresenta una statistica descrittiva affidabile per specificare la distribuzione in esame, visto che rappresenta il valore meno probabile.

Il modello statistico più frequentemente utilizzato per descrivere il tasso di perdita in caso di insolvenza o, parallelamente, il tasso di recupero in caso di insolvenza è la distribuzione Beta[3], caratterizzata da due parametri che possono essere messi in relazione con la media e la varianza dei valori osservati, come visibile in figura 2.2.

In poche parole, i valori medi presentano scarso significato e grande volatilità. Per provare a superare questi problemi, è necessario abbandonare il tentativo di utilizzare un'unica media generale e provare a utilizzare una numerosità di medie condizionate, selezionate combinando insieme determinati valori delle principali variabili rilevanti, quali la forma tecnica del finanziamento, la categoria di debitore, la presenza di garanzie, il periodo temporale di concessione e molte altre variabili ritenute significative nello spiegare le differenze tra i tassi di perdita in caso di insolvenza osservati in passato.

Così facendo, è possibile stabilire una relazione tra il livello del tasso di perdita in caso di insolvenza e le principali caratteristiche delle esposizioni creditizie in esame, in modo da assegnare a insiemi di esposizioni creditizie simili un valore del tasso di perdita in caso di insolvenza appropriato.

Con dati sufficientemente numerosi, infatti, è possibile raggruppare le osservazioni in sottoinsiemi ragionevolmente omogenei, contraddistinti da volatilità limitata, arrivando a ottenere una strategia che consente di collegare le caratteristiche delle esposizioni creditizie considerate con il valore del tasso di perdita in caso di insolvenza osservato.

[3]La distribuzione Beta è definita come $B \sim B(\alpha, \beta)$, con $\alpha, \beta > 0$.

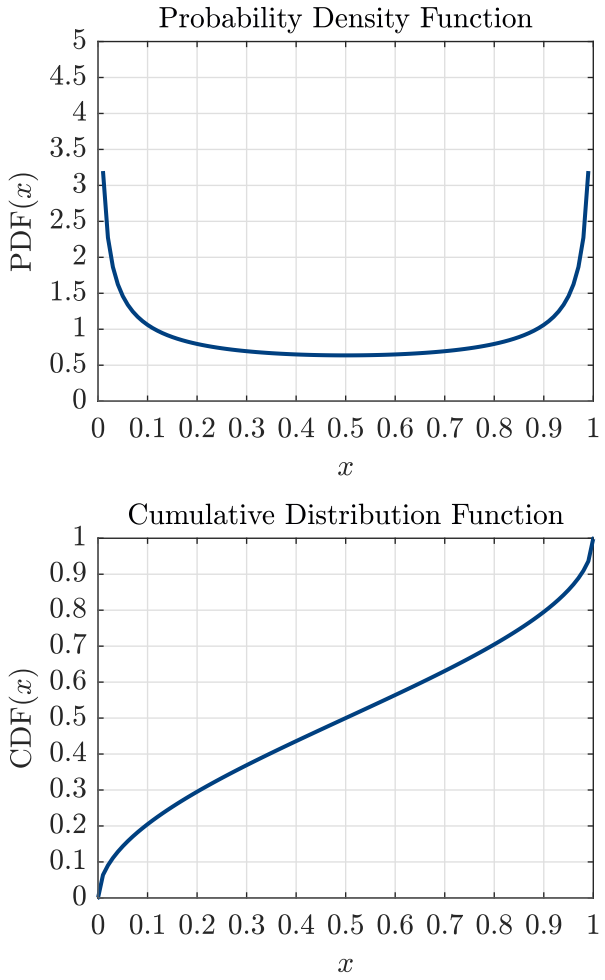


Figura 2.2. Distribuzione di probabilità Beta con $\alpha = \beta = 0.5$.

Per analizzare le caratteristiche di ogni singola esposizione creditizia finita in contenzioso, ottenendo una stima della perdita in caso di insolvenza media condizionata, è possibile utilizzare due metodologie diverse, che si differenziano in relazione al processo di stima impiegato:

- look-up table;
- modelli multivariati.

La metodologia delle look-up table prescrive la scomposizione delle informazioni presenti nel database costruito con i dati relativi ai tassi di perdita in caso di insolvenza passati, in modo da ottenere un insieme di categorie con caratteristiche simili. Dopo aver suddiviso il database in cluster omogenei, questo approccio prevede di calcolare il tasso di perdita in caso di insolvenza medio per ogni cluster e utilizzare quest'ultimo come stima

per tutte le esposizioni creditizie che presentano caratteristiche simili rispetto al cluster in esame.

La metodologia dei modelli multivariati, lineari o non lineari, richiede l'utilizzo di un modello statistico, con la finalità di verificare e analizzare la relazione tra il tasso di perdita in caso di insolvenza e le caratteristiche delle esposizioni creditizie andate in insolvenza. Nello specifico, questo approccio prevede di determinare un algoritmo di stima che sia in grado di spiegare, dopo aver esaminato le caratteristiche delle esposizioni creditizie, i valori osservati dei tassi di perdita in caso di insolvenza e, conseguentemente, esplicitare, dopo averne verificato la significatività, i legami tra le variabili.

Purtroppo, anche in questo caso, è possibile mettere in evidenza alcuni problemi specifici, derivanti implicitamente dalla natura del fenomeno in esame, ovvero l'insolvenza, che concorrono a complicare notevolmente il calcolo di una stima quanto più affidabile possibile.

In primo luogo, la presenza di una limitata disponibilità di dati, dal momento che, all'interno del sistema economico, l'insolvenza rappresenta, fortunatamente, un fenomeno relativamente raro e, quindi, con manifestazioni limitate.

In secondo luogo, la presenza di una considerevole quantità di variabili rilevanti, che contribuiscono ad aumentare la complessità e, contestualmente, ridurre la possibilità di realizzare insiemi sufficientemente omogenei di insolvenze.

Questi problemi risultano particolarmente accentuati per le banche di minor dimensione, le quali, per superare queste difficoltà, ricorrono alla pratica del *data-pooling*, ovvero alla condivisione, in forma completamente anonima, delle informazioni che riguardano le insolvenze delle proprie controparti. In questo modo, le banche possono contare su un database di dimensioni molto più rilevanti, con la possibilità di elaborare questa maggiore massa di informazioni e ricavare le statistiche di interesse.

2.5.4 Relazione Probability of Default e Loss Given Default

Nell'analisi del rischio di credito, è importante specificare come la probabilità di insolvenza e il tasso di perdita in caso di insolvenza rappresentino due variabili fondamentali, che dipendono direttamente dalla definizione che è possibile assegnare al concetto di default. Infatti, utilizzare una definizione di default diversa, più ampia o più ristretta, influenza pesantemente non solo il processo di stima della probabilità di insolvenza, ma anche il processo di stima del tasso di perdita in caso di insolvenza, complicando ulteriormente il ragionamento. In poche parole, è possibile evidenziare la presenza di un trade-off tra la severità della definizione di default e il valore assunto dalle due variabili in questione: probabilità di insolvenza e tasso di perdita in caso di insolvenza.

In questo contesto, è possibile definire l'insolvenza secondo due modalità differenti, illustrando integralmente la relazione che esiste tra PD e LGD.

Una prima modalità utilizza una definizione del default più stringente, in modo da includere soltanto il fallimento e le altre procedure concorsuali. In questo caso, il default viene definito come un evento assorbente, ovvero come una manifestazione definitiva e permanente, che riguarda controparti in stato di grande difficoltà economico-finanziaria, per le quali è sicuramente difficile ipotizzare un eventuale risanamento.

Una seconda modalità utilizza una definizione del default più estesa, che comprende qualsiasi semplice ritardo nei pagamenti previsti da contratto. In questo caso, il default viene definito come un evento non assorbente, ossia come una manifestazione modificabile e temporanea, che riguarda controparti in uno stato di difficoltà economico-finanziaria molto più modesta, per le quali è facile immaginare un ritorno in bonis.

Una definizione di insolvenza più ristretta considera una probabilità di default limitata e una perdita in caso di insolvenza elevata, poiché analizza eventi poco frequenti e classifica i debitori come insolventi soltanto in presenza di grandi difficoltà, prevedendo un rimborso contenuto dell'esposizione terminata in default.

Al contrario, una definizione di insolvenza più ampia considera una situazione opposta, con una probabilità di default elevata e una perdita in caso di insolvenza limitata, poiché analizza eventi molto più frequenti e classifica i debitori come insolventi anche in presenza di piccole difficoltà, prevedendo, a differenza della definizione precedente, un rimborso sostanziale dell'esposizione terminata in default.

In sostanza, la definizione di default emerge come un argomento molto importante. Una corretta gestione del rischio di credito, infatti, conduce inevitabilmente verso l'adozione di una definizione omogenea di insolvenza, considerando che, per evitare valutazioni incoerenti e distorte e, conseguentemente, stimare in modo appropriato il rischio di credito complessivo, è fondamentale utilizzare una definizione di default identica, sia nella stima della probabilità di insolvenza che nella stima del tasso di perdita in caso di insolvenza.

2.6 Perdita inattesa

La perdita inattesa è definita come la variabilità della perdita attorno al suo valore medio. Questa costituisce, ricorrendo alla definizione generale di rischio presentata in precedenza, il vero e proprio rischio di credito, ovvero il rischio che la perdita effettivamente osservata si dimostri superiore rispetto a quanto originariamente stimato.

In poche parole, la perdita inattesa costituisce una perdita caratterizzata generalmente da una bassa frequenza e da un importo estremamente elevato, che, a causa di queste peculiarità, non risulta in alcun modo stimabile, oltre che potenzialmente distruttiva. A questo riguardo, la copertura della perdita inattesa è sempre a carico del capitale, il quale deve risultare sufficientemente consistente per assorbire le perdite inattese fino a un certo livello di confidenza e rimanere nel business, garantendo la continuità aziendale.

2.6.1 Value at Risk (VaR)

Per discutere la perdita inattesa, è necessario introdurre un concetto fondamentale, quale il concetto di "valore a rischio", o Value at Risk (VaR), che rappresenta un approccio in grado di esprimere una misura del rischio complessivo a cui si è esposti, declinabile in base alla tipologia di rischio considerato, in questo caso, il rischio di credito.

Il Value at Risk identifica una misura di rischio ottenibile da specifici modelli in grado di misurare, aggregare e confrontare il rischio connesso a posizioni o portafogli di posizioni differenti, i quali, introdotti gradualmente dalle principali banche degli Stati Uniti durante gli anni '80, rappresentano oggi il punto di riferimento in ambito risk management per

tutte le istituzioni finanziarie. Questo concetto è stato sviluppato per la prima volta dalla banca J. P. Morgan, intorno al 1990, con il modello *RiskMetrics*, in risposta all'esigenza di provare a creare un'informativa sintetica, racchiusa in un singolo valore monetario, che consentisse di conoscere in maniera più chiara e precisa i rischi assunti dalla banca, migliorandone il processo di allocazione del capitale.

Il Value at Risk costituisce una misura di tipo probabilistico, la quale specifica la perdita massima che può essere subita su una posizione o su un portafoglio di posizioni, definito uno specifico livello di confidenza e uno specifico orizzonte temporale di riferimento, in modo tale che la probabilità che la perdita effettiva superi questo importo risulti sufficientemente bassa. In altri termini, il VaR identifica la massima perdita potenziale L che potrebbe essere subita in un orizzonte temporale T con un livello di confidenza c:

$$P(L > VaR_c^T) = 1 - c \ . \tag{2.5}$$

Il VaR può essere utilizzato per studiare attività finanziarie diverse, con la possibilità di esprimere una valutazione combinando insieme rischi diversi, e, per questo motivo, viene applicato per confrontare le alternative di impiego del capitale, esaminare la redditività del capitale allocato e assegnare il pricing alle singole operazioni in base al livello di rischio.

Il VaR applicato al rischio di credito, chiamato VaR creditizio, considera i seguenti parametri imposti, all'interno della regolamentazione bancaria, dal Comitato di Basilea: $T = 1$ anno e $c = 99{,}9\%$, con riferimento al modello *CreditMetrics*, definito sempre dalla banca J. P. Morgan nel 1997, ispirandosi al lavoro di R. C. Merton per la stima della PD.

In ultima analisi, è possibile mettere in evidenza alcuni limiti del VaR, considerando che non specifica le perdite potenziali oltre il livello di confidenza e costituisce una misura di rischio non coerente[4], violando l'assioma di subadditività. Per risolvere questi eventuali problemi, è possibile menzionare una misura di rischio alternativa e integrativa, come la "perdita media condizionata", o Expected Shortfall (ES), la quale risulta una misura di rischio coerente, definita come il valore atteso di tutte le possibili perdite superiori al VaR:

$$ES = E[L \mid L > VaR_c^T)] \ . \tag{2.6}$$

A questo punto, è interessante analizzare le principali differenze che intercorrono tra la definizione di perdita attesa e di perdita inattesa, sotto due dimensioni fondamentali.

Dal punto di vista matematico-statistico, si può spiegare la distinzione tra perdita attesa e perdita inattesa considerando un generico portafoglio impieghi. In questo caso, la perdita attesa di un portafoglio impieghi risulta uguale alla somma delle perdite attese dei singoli crediti che lo compongono, mentre la variabilità della perdita su un portafoglio impieghi risulta diversa, e generalmente minore, dalla somma delle variabilità delle perdite sui singoli crediti, diminuendo progressivamente al diminuire del grado di correlazione tra i diversi crediti.

A questo proposito, infatti, è possibile dimostrare due conseguenze molto importanti: da una parte, si può constatare come la perdita attesa non possa essere ridotta tramite

[4]Una misura di rischio coerente deve sistematicamente rispettare i seguenti assiomi fondamentali: invarianza alle traslazioni, subadditività, omogeneità positiva e monotonicità [10].

un'opportuna diversificazione del portafoglio crediti, per esempio combinando insieme crediti diversi appartenenti a più settori industriali o più aree geografiche; dall'altra, si può constatare come la perdita inattesa, ovvero la variabilità intorno alla perdita attesa, possa essere invece ridotta mediante una conveniente politica di ripartizione del rischio, con la possibilità che un'adeguata diversificazione del portafoglio crediti possa consentire di attenuare, anche in misura significativa, a parità di rendimento atteso, l'ammontare complessivo dell'esposizione al rischio di credito.

Dal punto di vista economico-finanziario, si può spiegare la distinzione tra perdita attesa e perdita inattesa esaminandone la responsabilità di copertura. In questo modo, la perdita attesa dovrebbe essere coperta con un accantonamento a riserva, registrato tra le voci di costo del Conto Economico, mentre la perdita inattesa dovrebbe essere coperta con il patrimonio della banca, ritenendo gli azionisti responsabili non solo di beneficiare di risultati superiori alle aspettative, prodotti da perdite minori rispetto a quelle previste, ma anche di sopportare i risultati inferiori alle aspettative, prodotti da perdite maggiori rispetto a quelle previste.

In sintesi, è possibile osservare una panoramica di quanto appena spiegato in figura 2.3.

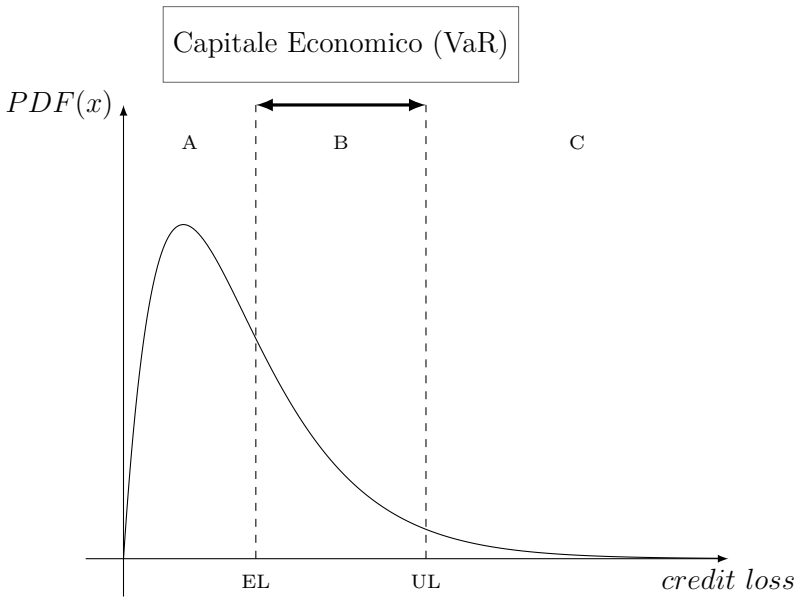


Figura 2.3. Distribuzione di probabilità delle perdite sui crediti.

Nello specifico, esaminando una generica distribuzione di probabilità associata alle potenziali perdite sui crediti, è facile mettere in evidenza tre zone diverse:

- A: *Expected Loss*, perdita coperta con pricing e accantonamenti;
- B: *Unexpected Loss*, perdita coperta con capitale;
- C: *Stress Loss*, perdita monitorata con analisi di scenario o di stress.

Parte II

Seconda Parte

Capitolo 3

Credit scoring

3.1 Introduzione

Lo studio della probabilità di insolvenza di una controparte rappresenta uno degli aspetti più importanti e più rilevanti che caratterizzano l'operatività degli intermediari finanziari e, nello specifico, degli intermediari bancari. Per questo motivo, le banche hanno da sempre investito una grande quantità di risorse, nel tentativo di modellare nel miglior modo possibile questa problematica e ottenere risultati in linea con le proprie esigenze.

I modelli di credit scoring rappresentano i modelli tradizionalmente più diffusi per la stima della probabilità di insolvenza. In particolare, si tratta di modelli avanzati di natura statistica che analizzano i principali indicatori economico-finanziari di una controparte, assegnano una ponderazione che riflette l'importanza relativa nel prevedere l'insolvenza e arrivano a una valutazione del merito creditizio sintetizzata in un valore numerico, chiamato credit score, rappresentativo della probabilità di insolvenza.

Le tecniche alla base dei modelli di credit scoring sono state elaborate e formulate a partire dagli anni '30, con il lavoro di R. A. Fisher (1936) e D. Durand (1941), ma la spinta decisiva allo sviluppo e alla diffusione di questi modelli è stata impressa solo successivamente, intorno agli anni '60, grazie al contributo di autori fondamentali, quali G. A. Donaldson (1962), W. H. Beaver (1966), E. I. Altman (1968) e J. A. Ohlson (1980).

Nel corso del tempo, sono stati sviluppati numerosi modelli di credit scoring, sempre più efficaci ed efficienti, con l'obiettivo di ottenere una stima sempre più affidabile della probabilità di insolvenza. Per comprendere in modo ottimale il dominio di interesse e discutere la possibilità di applicare algoritmi di intelligenza artificiale nell'ambito della valutazione del rischio di credito, è necessario ripercorrere le principali metodologie che sono state utilizzate nel corso della storia per la stima della probabilità di insolvenza.

A questo proposito, il capitolo si apre con una ricostruzione dello studio della probabilità di insolvenza, spiegando l'evoluzione da analisi univariata ad analisi multivariata, e si chiude con una presentazione delle principali categorie di modelli di credit scoring: l'analisi discriminante, i modelli di regressione (lineare multipla, Logit e Probit) e, infine, i modelli di intelligenza artificiale.

3.2 Analisi univariata

Lo studio sistematico della probabilità di insolvenza ha avuto inizio negli anni '60 del secolo scorso, con il lavoro di W. H. Beaver, che ha sviluppato il primo approccio scientificamente corretto per la stima del rischio di credito utilizzando i dati di bilancio delle imprese. Questo contributo ha rappresentato un passaggio fondamentale, dal momento che ha introdotto per la prima volta una serie di analisi economico-finanziarie che, in seguito, si sono sviluppate e hanno dato origine ai moderni modelli di credit scoring [60].

Per analizzare il rischio di credito, W. H. Beaver ha messo a confronto due campioni di imprese in un medesimo intervallo temporale di riferimento: un campione di imprese sane e un campione di imprese anomale. Nello specifico, il primo campione era composto da imprese che non avevano avuto eventi di default nel corso del periodo di tempo in esame, mentre il secondo campione era composto da imprese che, a differenza delle precedenti, avevano avuto eventi di default nello stesso periodo di tempo.

Nella costruzione del campione W. H. Beaver ha introdotto per la prima volta l'utilizzo del concetto di campione pareggiato, cercando di scegliere per ciascuna impresa anomala un'equivalente impresa sana, ovvero un'impresa che risiede nella stessa area geografica, che opera nello stesso settore industriale e che presenta la stessa classe dimensionale. In altri termini, si è selezionato un campione di imprese composto da un numero identico di imprese sane e imprese anomale, iniziando con la ricerca di imprese anomale e cercando, successivamente, di pareggiare il campione con l'aggiunta di imprese sane simili, in modo da effettuare un confronto il più possibile omogeneo.

Dopo aver completato la composizione del campione, W. H. Beaver ha lavorato per ricostruire la serie storica dei bilanci delle imprese in uno specifico intervallo di tempo, con l'obiettivo di calcolare una serie di indicatori di bilancio e osservarne l'evoluzione temporale per studiare il comportamento delle imprese sane e delle imprese anomale.

A questo punto, W. H. Beaver ha calcolato la media di ogni indicatore di bilancio sia per il campione di società sane che per il campione di società anomale, nel tentativo di confrontare i valori ottenuti e ricavarne informazioni sulla probabilità di insolvenza. In altre parole, si è esaminato l'evoluzione nel tempo della media di questi indicatori, partendo da un numero specifico di anni prima dell'insolvenza fino ad arrivare al momento dell'effettiva insolvenza.

In questo contesto, analizzando l'andamento di questi indicatori verso l'insolvenza, è facile osservare come la media delle società anomale si allontani progressivamente dalla media delle società sane, dimostrando l'applicazione del ragionamento proposto, come illustrato in figura 3.1.

Per lo studio della probabilità di insolvenza, W. H. Beaver ha quindi deciso di condurre un'analisi univariata, esaminando singolarmente i migliori indicatori di bilancio a disposizione e mettendone a confronto l'evoluzione storica tra il campione di imprese sane e il campione di imprese anomale, con l'obiettivo di individuare gli indicatori più diagnostici, in grado di discriminare al meglio tra i due campioni in esame. Osservando l'andamento degli indicatori di bilancio in periodi di tempo precedenti rispetto al momento di analisi, si è cercato di identificare le variabili economico-finanziarie maggiormente significative per anticiparne l'insolvenza.

In questo lavoro, W. H. Beaver ha cercato di analizzare e verificare numerosi indicatori di bilancio: indicatori di liquidità, indicatori di indebitamento e solvibilità, indicatori di interest coverage con margini di profitto, indicatori di interest coverage con flussi di cassa, indicatori di capacità potenziale di rimborso del debito e indicatori di flussi di cassa per il rimborso del debito. A questo riguardo, si è individuato come indicatore più diagnostico il rapporto tra cash flow e debito totale.

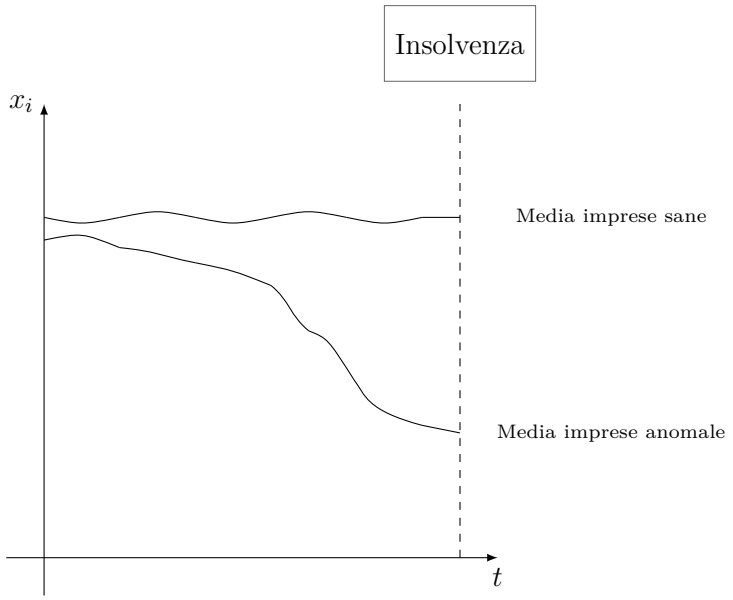


Figura 3.1. Sintesi grafica dell'analisi univariata.

Riassumendo, W. H. Beaver ha effettuato un'analisi univariata, prendendo in considerazione un solo indicatore per volta e scegliendo, tra tutti gli indicatori in esame, quelli che avevano la maggior capacità di segnalare un problema di solvibilità. In questo caso, infatti, osservare un rapido degrado di questi indicatori rappresentava un ottimo segnale sul fatto che la società in esame stava procedendo con passi veloci verso l'insolvenza.

Il ragionamento di W. H. Beaver, la base della scelta degli indicatori da utilizzare, è riferita ad alcuni lavori precedenti di G. A. Donaldson, che aveva sviluppato un approccio basato sui flussi di cassa per valutare la probabilità di insolvenza delle imprese [76].

G. A. Donaldson, nel corso dei suoi studi, ha eseguito numerose analisi empiriche, dalle quali è emersa la possibilità di interpretare la politica economico-finanziaria dell'impresa in base alla dinamica dei flussi di cassa. A questo proposito, secondo G. A. Donaldson, l'impresa può essere vista come un colossale serbatoio di liquidità, che nel tempo riceve dei flussi di cassa in entrata, o *cash inflows*, e dei flussi di cassa in uscita, o *cash outflows*.

In questa visione, l'impresa riesce a mantenersi solvibile nel caso in cui questo serbatoio di liquidità non si esaurisca mai, mantenendo sempre un margine di sicurezza in modo tale che i flussi di cassa in ingresso siano in grado di far fronte ai flussi di cassa in uscita. La riserva di liquidità serve come buffer contro variazioni inattese dei flussi di cassa: il

rischio di insolvenza corrisponde alla probabilità che questa riserva sia esaurita, spingendo l'impresa verso l'impossibilità di far fronte ai propri impegni alla scadenza.

In questo modo, W. H. Beaver ha ripreso l'impostazione concettuale e lo schema di ragionamento di G. A. Donaldson, scegliendo gli indicatori di bilancio che fossero il più possibile vicini a quella che, in quel periodo, era considerata la base teorica più importante delle osservazioni empiriche sul mondo delle imprese.

3.3 Analisi multivariata

G. A. Donaldson prima e W. H. Beaver poi hanno svolto il proprio lavoro intorno agli anni '60 del secolo scorso e, per questo motivo, hanno studiato e affrontato la finanza aziendale molto prima che la finanza quantitativa moderna facesse prepotentemente il suo ingresso nello sviluppo di modelli statistici e nella ricerca econometrica.

Come spiegato in precedenza, W. H. Beaver ha utilizzato gli strumenti a disposizione della sua epoca in modo estremamente intelligente, senza utilizzare nessun modello statistico, ma sviluppando un'analisi univariata, la quale costituisce il progenitore storico di una serie di importanti modelli successivi che hanno poi migliorato e reso più sofisticato il ragionamento sviluppato sull'analisi dell'insolvenza di un'impresa.

L'analisi univariata proposta da W. H. Beaver rappresenta quindi un primo tentativo di analizzare il rischio di credito, anche se presenta alcuni limiti che devono essere superati per migliorarne il ragionamento. Osservare un indicatore per volta costituisce un metodo semplice e veloce per analizzare la situazione economico-finanziaria di una impresa, ma non risulta sufficientemente avanzato e rischia di generare una valutazione approssimativa.

A questo proposito, è necessario effettuare un passaggio decisivo ulteriore, introducendo l'analisi multivariata, che consiste nello sviluppare un modello in grado di mettere insieme tutte le informazioni possibilmente ricavabili da un insieme di indicatori di bilancio e nel condensare queste ultime in un indicatore di sintesi, chiamato credit score.

3.3.1 Analisi discriminante

L'analisi discriminante, introdotta da R. A. Fisher nel 1936, rappresenta un modello statistico basato sull'identificazione di variabili in grado di analizzare e discriminare in modo ottimale tra imprese sane e imprese anomale. Questa tecnica applica l'analisi multivariata, raggiungendo un grande vantaggio dal punto di vista delle analisi economico-finanziarie, in quanto consente di valutare contemporaneamente più variabili.

In poche parole, l'analisi discriminante utilizza tutte le informazioni disponibili estratte da un campione di imprese in esame, con l'obiettivo di tracciare un confine e distinguere, nel modo più corretto possibile, tra imprese sane e imprese anomale.

In questo modo, si riesce a rendere più completo il ragionamento sviluppato sui dati contabili delle imprese: si cerca di verificare se una generica impresa è più vicina a un insieme di imprese sane o a un insieme di imprese anomale e, se è più vicina all'insieme di imprese sane, si dice che l'impresa ha basse probabilità di default, mentre, se è più vicina all'insieme di imprese anomale, si dice che l'impresa ha alte probabilità di default.

In figura 3.2 è possibile osservare l'idea alla base dell'analisi discriminante in un caso semplificato, in cui le imprese sane e le imprese anomale sono descritte da due sole variabili, x_1 e x_2.

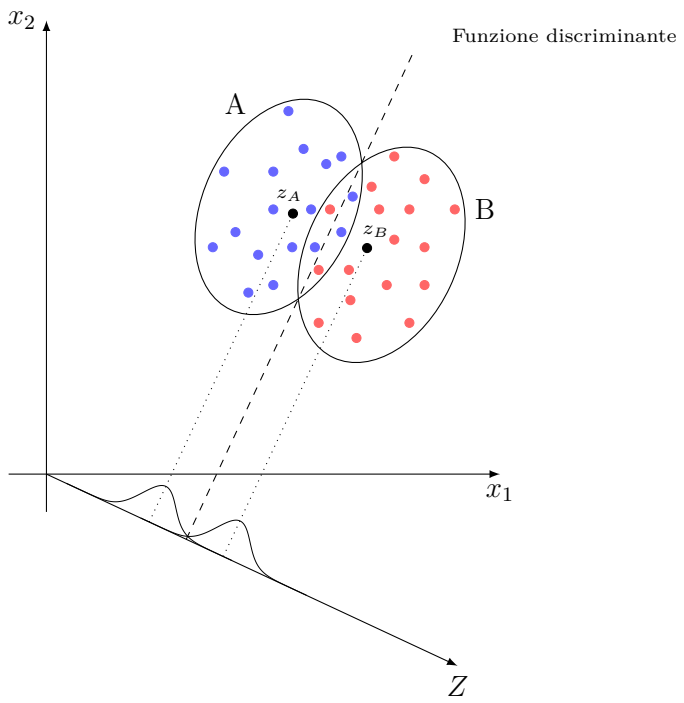


Figura 3.2. Sintesi grafica dell'analisi discriminante.

Sull'asse Z è rappresentato il credit score, che prende il nome di funzione discriminante, generato combinando insieme con determinati coefficienti le variabili di interesse.

Funzione discriminante

L'analisi discriminante, considerando la sua versione più semplice e più utilizzata, ovvero l'analisi discriminante lineare, costruisce il credit score z_i come combinazione lineare di una serie di variabili indipendenti $x_{i,j}$. In generale, considerando n variabili indipendenti, è possibile calcolare il credit score per la generica impresa i-esima nel modo seguente:

$$z_i = \sum_{j=1}^{n} \gamma_j \cdot x_{i,j} \ . \tag{3.1}$$

I coefficienti γ_j di questa combinazione lineare sono scelti, tra tutte le infinite soluzioni possibili, in modo da ottenere un credit score che discrimini in modo quanto più possibile corretto tra imprese sane e imprese anomale. In questo modo, il credit score deve essere in grado di massimizzare la distanza tra le medie dei due gruppi di imprese, z_A e z_B, che prendono il nome di centroidi.

A questo proposito, è possibile dimostrare che, per soddisfare la precedente condizione, il vettore dei coefficienti $\boldsymbol{\gamma}$ deve verificare la seguente condizione:

$$\boldsymbol{\gamma} = \boldsymbol{\Sigma}^{-1} \cdot (\mathbf{x}_A - \mathbf{x}_B) \ , \tag{3.2}$$

dove:

- $\boldsymbol{\Sigma}^{-1}$ = inversa della matrice di varianza/covarianza tra le n variabili indipendenti;
- $\mathbf{x}_A$ = vettore della media delle n variabili indipendenti per le imprese sane;
- $\mathbf{x}_B$ = vettore della media delle n variabili indipendenti per le imprese anomale.

A questo punto, è possibile assegnare, per ogni impresa, un punteggio del credit score, al variare dei valori assunti dalle variabili indipendenti del modello e, una volta fissata una specifica soglia, o cut-off point, differenziare tra imprese sane e imprese anomale.

Stima dei coefficienti

Per spiegare approfonditamente l'analisi discriminante è possibile seguire due step diversi.

Il primo passo dell'analisi discriminante consiste nell'approccio non parametrico, nel quale non si effettua nessuna ipotesi sulla forma della distribuzione delle diverse variabili di interesse. In questo caso, lavorando con due gruppi di imprese diverse, sane e anomale, l'obiettivo è costruire una funzione decisionale di tipo lineare, individuando i parametri che servono per calcolare la media ponderata delle variabili, in modo tale che gli errori di classificazione siano i minori possibili e, conseguentemente, la classificazione sia ottimale.

In pratica, si modella la scelta dei pesi con cui ponderare le variabili, con la finalità di distanziare il più possibile il gruppo delle imprese sane dal gruppo delle imprese anomale, ovvero massimizzando la variabilità tra i gruppi e minimizzando la variabilità nei gruppi.

Per ottenere questo risultato, è sufficiente massimizzare un rapporto specifico, ovvero il rapporto tra la varianza tra i gruppi e la varianza nei gruppi, facendo emergere una soluzione per il vettore dei pesi pari al prodotto tra il vettore della differenza tra le medie e l'inversa della matrice di varianza/covarianza. Il vettore dei pesi, così calcolato, costituisce il vettore ottimale, in grado di ridurre il più possibile gli errori di classificazione.

Il secondo passo dell'analisi discriminante riguarda l'approccio parametrico, che prevede di fare un'assunzione forte sulla forma della distribuzione delle variabili di interesse, ipotizzando una distribuzione normale multivariata. In questo caso, è possibile impostare la decisione di assegnare una generica impresa al gruppo delle imprese sane oppure al gruppo delle imprese anomale, analizzando la probabilità di appartenere ai due gruppi. Pertanto, si sceglie di assegnare una nuova osservazione al gruppo delle imprese sane se la probabilità di appartenere alle sane è maggiore rispetto alla probabilità di appartenere alle anomale e, viceversa, al gruppo delle imprese anomale se la probabilità di appartenere alle sane è minore rispetto alla probabilità di appartenere alle anomale.

A questo punto, è possibile osservare come si ottenga la stessa forma funzionale, sia nel caso di approccio parametrico che nel caso di approccio non parametrico, mettendo in evidenza il fatto che l'analisi discriminante conduce sempre alla stessa funzione decisionale. L'uguaglianza della forma funzionale costituisce, infatti, un aspetto molto importante, poiché significa che questo modello risulta robusto alla violazione delle ipotesi di partenza.

Z-score di E. I. Altman

E. I. Altman ha definito il primo modello di credit scoring basato sull'analisi discriminante, realizzando un significativo salto di qualità sulla valutazione dell'insolvenza di un'impresa. Questo modello, applicato inizialmente nel 1968 per le imprese quotate degli Stati Uniti e successivamente modificato e aggiornato, è funzione di cinque variabili indipendenti [3]:

$$z_i = 1{,}20 \cdot x_{i,1} + 1{,}40 \cdot x_{i,2} + 3{,}30 \cdot x_{i,3} + 0{,}60 \cdot x_{i,4} + 1{,}00 \cdot x_{i,5} \ , \tag{3.3}$$

dove:

- $x_{i,1}$ = capitale circolante/totale attivo;
- $x_{i,2}$ = utili non distribuiti/totale attivo;
- $x_{i,3}$ = EBIT/totale attivo;
- $x_{i,4}$ = valore di mercato del patrimonio netto/valore contabile dei debiti;
- $x_{i,5}$ = fatturato netto/totale attivo.

3.3.2 Modelli di regressione

Regressione lineare multipla

Il primo tentativo di utilizzare un modello di regressione per risolvere il problema associato alla stima della probabilità di default consiste nell'individuare le variabili che determinano l'insolvenza attraverso l'applicazione di un modello di regressione lineare multipla.

Prima di tutto, per applicare correttamente questo metodo, si seleziona un campione di imprese con dimensione sufficientemente elevata, in modo da ottenere un modello in grado di offrire risultati il più possibile attendibili e affidabili.

A questo punto, si suddivide il campione in due gruppi, caratterizzati da una variabile dipendente binaria y che descrive lo stato di salute di una impresa, assumendo soltanto due valori, 0 e 1. In relazione a questa variabile di stato, si modella una generica impresa i-esima, assegnando un valore $y_i = 0$ se l'impresa è sana e $y_i = 1$ se l'impresa è anomala.

In seguito, si procede con la selezione delle variabili indipendenti, calcolando m variabili economico-finanziarie $[x_{i,1}, x_{i,2}, \ldots, x_{i,m}]$ per ogni impresa del campione in esame.

In ultima analisi, si esegue la stima dei coefficienti, solitamente con il classico approccio dei Minimi Quadrati Ordinari[1], calcolando il modello di regressione lineare multipla:

$$y_i = \alpha + \sum_{j=1}^{m} \beta_j \cdot x_{i,j} + \varepsilon_i \ . \tag{3.4}$$

Una volta ottenuto il modello completo, si utilizza per stimare la probabilità di default.

[1] In inglese, *Ordinary Least Squares* (OLS).

Il modello di regressione lineare multipla rappresenta certamente un approccio di semplice applicazione, ma, pur sembrando apparentemente corretto, questo modello presenta diverse problematiche che occorre approfondire per una sua corretta comprensione.

Il primo problema riguarda il calcolo della probabilità di insolvenza, dal momento che, utilizzando questo modello, è possibile che la variabile dipendente y assuma valori esterni all'intervallo 0%-100%. Interpretando questa variabile come probabilità si ottiene quindi un ragionamento poco sensato, in quanto una situazione del genere risulta in palese contrasto con la definizione di probabilità, che implica valori compresi tra 0% e 100%. Per questo motivo, nel caso in cui si dovessero ottenere valori incongruenti della variabile y, che implicano valori della probabilità di default inferiori allo 0% o superiori al 100%, è necessario effettuare, per ripristinare la coerenza, un'operazione di troncamento del risultato in corrispondenza dei valori estremi.

Il secondo problema, possibilmente ancora più complesso, riguarda la varianza dei residui, che, nel modello di regressione lineare multipla, non risulta costante, ma risente di un problema di eteroschedasticità, violando le assunzioni alla base dell'approccio dei Minimi Quadrati Ordinari, in modo da condurre verso stime imprecise e distorte. Dal punto di vista econometrico è dunque possibile mettere in evidenza problemi di disturbi nella stima corretta dei regressori.

La regressione lineare multipla funziona bene se tutto il mondo è distribuito secondo una distribuzione normale o, se non è normale, quantomeno secondo una distribuzione simmetrica che tende alla distribuzione normale. In questo contesto, però, i residui non sono distribuiti normalmente e, conseguentemente, la stima con i minimi quadrati rischia di non essere una stima efficiente.

Per tutte queste considerazioni, il modello di regressione lineare risulta sostanzialmente inutilizzabile e si preferisce ricorrere a funzioni non lineari, come i modelli Logit e Probit, in grado di superare questi problemi e ottenere risultati sensibilmente migliori.

Logit e Probit

Nello studio dell'analisi discriminante si applica un'ipotesi importante dal punto di vista metodologico, immaginando che l'universo sia composto da due popolazioni distinte, la popolazione delle imprese sane e la popolazione delle imprese anomale, caratterizzate da valori completamente diversi delle variabili descrittive. A questo proposito, il modello di analisi discriminante si prefigge come obiettivo quello di osservare una generica impresa e di attribuire quest'ultima, con una certa probabilità di commettere degli errori, alla popolazione delle imprese sane o alla popolazione delle imprese anomale.

Al contrario, nei modelli di regressione Logit e Probit si applica un ragionamento diverso, dal momento che si ipotizza che l'universo sia composto soltanto da un'unica grande popolazione di imprese, alla quale appartengono tutte le imprese in questione.

Come risultato, non si distingue più tra imprese sane e imprese anomale: adesso le imprese che appartengono alla popolazione sono tutte caratterizzate da una variabile latente y^*, non osservabile, che ne rappresenta lo stato di salute economico-finanziaria.

Sfortunatamente, non è possibile distinguere e misurare questa variabile nella realtà; tuttavia, è possibile osservarne alcune realizzazioni concrete, considerando la variabile binaria y e valutando se nell'orizzonte temporale di interesse l'impresa è stata o meno

caratterizzata dall'evento default. Anche in questo caso si utilizza la terminologia introdotta precedentemente, ovvero $y_i = 0$ per imprese sane e $y_i = 1$ per imprese anomale, dal momento che, se la variabile latente stato di salute dell'impresa degenera in un default, si è soliti contrassegnarla per evidenziare l'evento creditizio negativo.

Riassumendo, il passaggio concettuale fondamentale è il seguente: si passa dall'approccio dell'analisi discriminante, dove l'universo è composto da due popolazioni di imprese diverse, all'approccio dei modelli Logit e Probit, dove, invece, l'universo è composto da una sola popolazione, complessiva di tutte le imprese, in cui ciascuna di queste imprese è descrivibile con una variabile latente che non è direttamente osservabile, ma di cui è possibile registrare delle realizzazioni concrete a specifici istanti di tempo.

Di conseguenza, utilizzando variabili economico-finanziarie che descrivono la situazione dell'impresa, è possibile modellare la probabilità che una generica impresa sia assimilabile alle imprese sane o alle imprese anomale, ovvero la probabilità che si realizzi davvero l'evento insolvenza. Per ottenere questo risultato, è necessario effettuare alcune ipotesi statisticamente forti sulla forma della distribuzione di probabilità, a seconda che si tratti di modello Logit o Probit.

A questo riguardo, nei modelli Logit e Probit si utilizzano le funzioni di ripartizione, poiché producono probabilità comprese tra 0 e 1, in modo da superare uno dei principali ostacoli alla corretta applicazione della regressione lineare multipla nell'ambito dei modelli di credit scoring: la funzione di ripartizione logistica standard per la regressione Logit e la funzione di ripartizione normale standard per la regressione Probit.

In questo modo, il modello Logit può essere definito come segue:

$$y_i = F(w_i) = \frac{1}{1 + e^{-(\alpha + \sum_{j=1}^{m} \beta_j \cdot x_{i,j})}} + \varepsilon_i \ , \tag{3.5}$$

dove F è la funzione di ripartizione logistica standard, w_i la variabile indipendente, ottenuta a partire dalla trasformazione lineare degli indicatori di bilancio x_{ij}, e ε_i la componente di disturbo casuale.

Analogamente, il modello Probit può essere definito come segue:

$$y_i = \Phi(w_i) = \Phi(\alpha + \sum_{j=1}^{m} \beta_j \cdot x_{i,j}) + \varepsilon_i \ , \tag{3.6}$$

dove Φ è la funzione di ripartizione normale standard, w_i la variabile indipendente, ottenuta a partire dalla trasformazione lineare degli indicatori di bilancio x_{ij}, e ε_i la componente di disturbo casuale.

I modelli Logit e Probit sono modelli di regressione non lineari specificatamente progettati per variabili dipendenti binarie, con un codominio limitato nell'intervallo (0,1), in modo da interpretare correttamente la probabilità di insolvenza, garantendo che la variabile dipendente y sia sempre compresa tra 0% e 100%, come visibile in figura 3.3.

Questi modelli risultano sostanzialmente simili, con risultati comparabili nella maggior parte delle applicazioni. L'unica differenza risiede nel fatto che la funzione Logit è caratterizzata da code più spesse rispetto alla funzione Probit, anche se, in pratica, non si osservano divergenze rilevanti tra i due modelli, producendo risultati diversi solo nel caso in cui il campione in esame sia caratterizzato da numerose osservazioni con valori estremi.

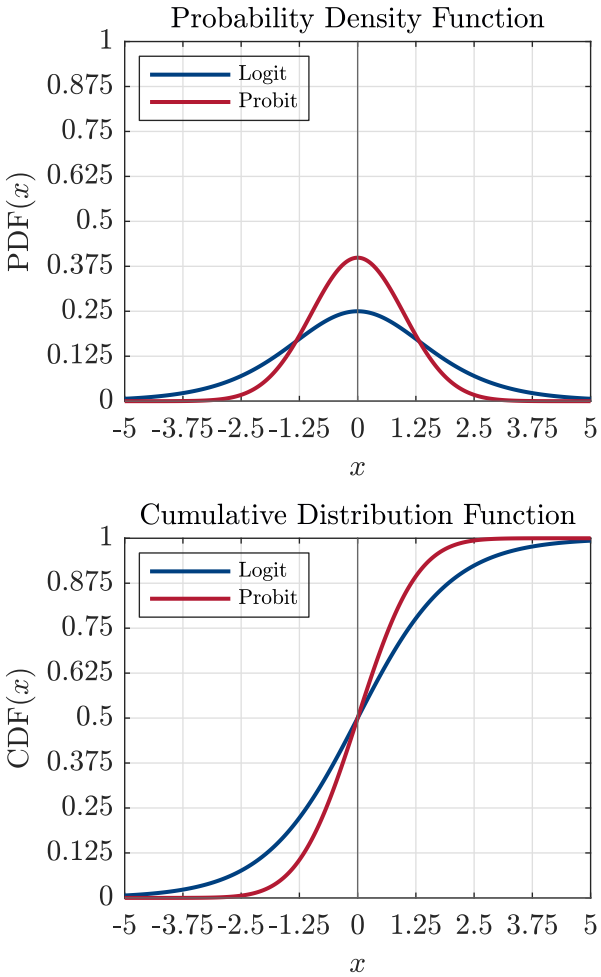


Figura 3.3. Distribuzione di probabilità logistica standard e normale standard.

Storicamente, si preferisce impiegare il modello Logit, il quale risulta più facilmente manipolabile dal punto di vista matematico, considerando che, soprattutto in caso di limitata potenza di calcolo, la funzione di ripartizione logistica può essere calcolata in modo molto più semplice e veloce rispetto alla funzione di ripartizione normale.

Per la stima dei coefficienti si utilizza il metodo della Massima Verosimiglianza[2], con il quale, osservando che la funzione di verosimiglianza rappresenta la distribuzione di probabilità congiunta dei dati in esame, analizzata come funzione dei coefficienti incogniti, si genera lo stimatore di Massima Verosimiglianza dei coefficienti incogniti, ottenuto a partire dai valori dei coefficienti che massimizzano la funzione di verosimiglianza.

[2]In inglese, *Maximum Likelihood* (ML).

In poche parole, lo stimatore di Massima Verosimiglianza sceglie i coefficienti incogniti in modo da massimizzare la funzione di verosimiglianza, che, a sua volta, costituisce la funzione di probabilità congiunta. Pertanto, questo approccio consente di scegliere i valori dei parametri al fine di massimizzare la probabilità che i dati osservati siano effettivamente estratti, generando delle stime che, in questo senso, rappresentano i valori dei parametri che "più verosimilmente" hanno prodotto i dati in questione.

In ogni caso, non esiste una formula esplicita in grado di definire una soluzione in forma chiusa per calcolare i coefficienti con cui devono entrare le variabili di interesse sia nel modello Logit che nel modello Probit e, per questo motivo, la funzione di verosimiglianza deve essere massimizzata per via numerica, ricorrendo a determinati algoritmi numerici applicati su un calcolatore, come ad esempio il metodo Newton-Raphson, che prende il nome dagli autori I. Newton e J. Raphson, il quale rappresenta appunto un metodo per il calcolo approssimato della soluzione di un'equazione nella forma $f(x) = 0$.

Logit di J. A. Ohlson

J. A. Ohlson ha definito nel 1980 il primo modello di credit scoring basato sul modello Logit, utilizzando un campione non bilanciato di imprese industriali degli Stati Uniti, in modo da sviluppare diversi modelli per la stima della probabilità di default al variare dell'orizzonte temporale esaminato. Il modello per la stima della probabilità di insolvenza, prendendo un anno come riferimento, è funzione di nove variabili indipendenti[3] [103]:

$$\begin{aligned} z_i = &- 1{,}32 - 0{,}41 \cdot x_{i,1} + 6{,}03 \cdot x_{i,2} - 1{,}43 \cdot x_{i,3} \\ &+ 0{,}08 \cdot x_{i,4} - 2.37 \cdot x_{i,5} - 1{,}83 \cdot x_{i,6} \\ &+ 0{,}29 \cdot x_{i,7} - 1{,}72 \cdot x_{i,8} - 0{,}52 \cdot x_{i,9} \ , \end{aligned} \tag{3.7}$$

dove:

- $x_{i,1}$ = ln(totale attivo/livello dei prezzi impliciti del PIL);
- $x_{i,2}$ = debiti totali/totale attivo;
- $x_{i,3}$ = capitale circolante/totale attivo;
- $x_{i,4}$ = passività correnti/attività correnti;
- $x_{i,5}$ = dummy (1 se debiti totali > totale attivo, 0 altrimenti);
- $x_{i,6}$ = risultato netto/totale attivo;
- $x_{i,7}$ = flusso di cassa della gestione/debiti totali;
- $x_{i,8}$ = dummy (1 se risultato netto < 0, 0 altrimenti);
- $x_{i,9}$ = variazione risultato netto.

[3] Per semplicità, si riporta il valore dei coefficienti arrotondato alla seconda cifra decimale.

3.4 Modelli di intelligenza artificiale

I modelli di credit scoring analizzati fino a questo momento, ovvero analisi discriminante e regressione Logit/Probit, si fondano su un approccio deduttivo, con l'obiettivo di spiegare le cause economico-finanziarie che possono condurre all'insolvenza di una controparte. Questi modelli cercano di identificare le variabili e le relazioni fondamentali tra variabili, in modo da spiegare l'equilibrio economico-finanziario, individuare le origini di eventuali criticità e, conseguentemente, prevedere l'insolvenza di una controparte.

In poche parole, l'analisi discriminante e i modelli di regressione seguono sempre un procedimento strutturale: analizzano le caratteristiche in grado di descrivere al meglio le condizioni di salute della controparte e, sulla base dei valori assunti da queste ultime, determinano la probabilità di insolvenza. In questo caso, la scelta delle variabili rilevanti, anche se calibrata con tecniche statistiche più o meno sofisticate, riflette sempre una scelta basata su un ragionamento economico.

Al contrario, i modelli di credit scoring che utilizzano tecniche di intelligenza artificiale si basano su un approccio completamente diverso e opposto, ovvero un approccio induttivo. I modelli di intelligenza artificiale non seguono un procedimento strutturale, partendo dalle idee di un analista e cercando di trovare una qualche conferma a queste idee in un campione di dati, bensì seguono un procedimento puramente empirico.

In questo modo, si parte da un campione di dati e, nel caso in cui venga riscontrata, studiando le osservazioni in esame, una certa regolarità, si utilizza questa regolarità, in modo completamente agnostico, nel tentativo di applicare la conoscenza appena acquisita per cercare di prevedere in tempo utile l'insolvenza di altre controparti.

Nel complesso, i modelli di credit scoring deduttivi sono modelli di tipo white box, comprensibili, trasparenti e costruiti poggiandosi su algoritmi robusti, che prevedono la presenza di test inferenziali per accertare la significatività dei diversi coefficienti stimati; viceversa, i modelli di credit scoring induttivi sono in genere modelli di tipo black box, con i quali è possibile generare, più o meno rapidamente, un risultato, senza chiaramente e agevolmente comprendere fino in fondo la logica sottostante.

Pertanto, sebbene un approccio deduttivo possa sembrare più corretto, almeno da un punto di vista razionale, un approccio induttivo può in realtà risultare molto utile quando è complesso o, addirittura, impossibile esplicitare le regole strutturali che governano un certo fenomeno, ovvero, nel caso specifico, l'insolvenza di una controparte.

I modelli di credit scoring che utilizzano tecniche di intelligenza artificiale sono quindi modelli induttivi, in grado di analizzare e studiare un insieme, anche complesso, di dati, imparare a riconoscere dei pattern al suo interno e, successivamente, effettuare previsioni.

A questa categoria di modelli sono riconducibili gli alberi decisionali e le reti neurali, che rappresentano i modelli di intelligenza artificiale designati per la progettazione di un modello di credit scoring per la gestione del rischio di credito, oggetto della trattazione.

3.5 Limiti dei modelli di credit scoring

In definitiva, i modelli di credit scoring costituiscono i principali modelli per analizzare il fenomeno dell'insolvenza, consentendo di raggiungere ottimi risultati, sia nel tentativo

di utilizzarli per la previsione dell'insolvenza, in modo da fissare una o più soglie minime di credit score, eventualmente separate da un'area "grigia" di difficile valutazione, per accettare o respingere le possibili richieste di credito, che nel tentativo di utilizzarli per la stima della probabilità di insolvenza, in modo da assegnare uno specifico livello di rischio, possibilmente aggregando il calcolo in un numero finito di classi, per ciascuna controparte.

Tuttavia, è doveroso precisare anche l'esistenza di alcuni limiti:

- definizione del concetto di default: l'evento insolvenza potrebbe essere valutato con sfumature diverse, le quali influenzano direttamente la definizione che è possibile assegnare al significato di impresa sana e di impresa anomala, condizionando, anche in maniera significativa, i risultati prodotti dai modelli di credit scoring.
- composizione del campione di analisi: le imprese analizzate, in linea teorica, anche se difficilmente applicabile nella realtà, dovrebbero appartenere al medesimo settore industriale, per rendere confrontabili i diversi indicatori economico-finanziari utilizzati, oltre che comporre un insieme non estremamente sbilanciato, includendo una composizione di imprese sane superiore rispetto alle imprese anomale.
- importanza delle variabili indipendenti: il potere discriminante relativo assunto dai molteplici indicatori economico-finanziari impiegati potrebbe variare ripetutamente nel corso del tempo, portando verso una modifica del conseguente peso all'interno dei modelli di credit scoring, in relazione alle fasi del ciclo economico, alla struttura del mercato di interesse o a molti altri fattori, sia interni che esterni.
- omissione di specifici fattori qualitativi: nei modelli di credit scoring potrebbero essere trascurati numerosi elementi qualitativi, alcuni dei quali molto importanti nel determinare l'insolvenza di un'impresa, seppur difficilmente inseribili e valutabili in un sistema quantitativo, come lo stato del ciclo economico, le prospettive future del mercato in questione, la reputazione dell'impresa, la competenza del management e la formazione del board.

Insomma, i modelli di credit scoring rappresentano una categoria di modelli in grado di assicurare una corretta gestione del rischio di credito, ma, nel loro utilizzo, è fondamentale prestare attenzione a tutti questi elementi di criticità, i quali, nel caso in cui non dovessero essere valutati con attenzione e tempestività, potrebbero condurre verso risultati non affidabili, con conseguenze estremamente dannose per l'operatività della banca in esame.

Pertanto, con l'obiettivo di ottenere un'adeguata progettazione e implementazione di un modello di credit scoring, è necessario esaminare a fondo le condizioni precedentemente elencate, prima di procedere alla sua effettiva costruzione, la quale inizia dalla scelta del campione di analisi, selezionando le imprese opportune, prosegue con la definizione del concetto di insolvenza, dividendo tra imprese sane e imprese anomale, oltre che con il calcolo dei principali indicatori di bilancio, e termina con la riduzione della dimensionalità, individuando le variabili più significative come risultato dell'analisi di correlazione.

Capitolo 4

Intelligenza artificiale

4.1 Introduzione

L'intelligenza artificiale, abbreviata AI dall'inglese *artificial intelligence*, è un insieme di tecnologie che interagiscono tra loro per consentire di progettare una macchina in grado di percepire, comprendere, imparare e agire con livelli di conoscenza, intelligenza e coscienza simili a quelli umani. In questo ambito, è possibile distinguere due concetti differenti: intelligenza artificiale generale e intelligenza artificiale ristretta[1].

L'intelligenza artificiale generale, o forte, si prefigge come obiettivo la creazione di macchine senzienti, capaci di emulare e replicare completamente l'intelligenza umana, con la possibilità, più o meno concreta, non solo di pensare in modo strategico, astratto e creativo, ma anche di gestire simultaneamente una serie di compiti complessi.

L'intelligenza artificiale ristretta, o debole, al contrario, non si impone l'obiettivo di possedere abilità cognitive generali, ma piuttosto si impegna a risolvere esattamente un singolo problema, non pienamente comprensibile nei limiti delle capacità cognitive umane, eseguendo un compito specifico o una serie di attività strettamente correlate.

In poche parole, l'intelligenza artificiale generale rappresenta il traguardo originale e finale della ricerca nel campo dell'intelligenza artificiale, anche se quest'ultimo è stato sostanzialmente e progressivamente abbandonato nel corso del tempo, verificando che l'intelligenza umana è un sistema troppo complesso per essere replicato completamente con metodi artificiali, a favore dell'intelligenza artificiale ristretta, che, invece, costituisce una finalità certamente più concreta, nel tentativo di sviluppare macchine in grado di risolvere problemi specifici simulando soltanto una o più funzioni umane complesse.

In questo capitolo, si discutono le nozioni fondamentali alla base della struttura dell'intelligenza artificiale, introducendo brevemente la storia dello sviluppo e della diffusione dagli inizi fino ad oggi, specificando la differenza tra i concetti di intelligenza artificiale, apprendimento automatico (machine learning) e apprendimento profondo (deep learning), e, infine, analizzando i principali modelli, oltre che i relativi ambiti di applicazione.

---

[1] In inglese, *artificial general intelligence* (AGI) e *artificial narrow intelligence* (ANI).

In ultima analisi, si approfondiscono i modelli di intelligenza artificiale designati per raggiungere l'obiettivo del presente lavoro di tesi, ovvero progettare e sviluppare un modello di credit scoring in grado di stimare la probabilità di default di una controparte e gestire il rischio di credito: alberi decisionali e reti neurali.

4.2 Storia dell'intelligenza artificiale

L'intelligenza artificiale nasce ufficialmente nel 1956, con l'organizzazione del famoso seminario estivo di Dartmouth, presso il Dartmouth College di Hanover, nel New Hampshire, da parte di J. McCarthy, M. L. Minsky, C. E. Shannon e N. Rochester, durante il quale per la prima volta venne fondata programmaticamente la nuova disciplina, iniziando dalla raccolta dei principali contributi sviluppati in passato fino a quel momento, proseguendo con la presentazione di alcuni studi in corso di svolgimento e terminando con l'analisi delle potenzialità più promettenti individuabili per il futuro. Questa storica conferenza, chiamata *Dartmouth Summer Research Project on Artificial Intelligence* (DSRPAI), aveva come obiettivo ambizioso, immaginando un grande sforzo collaborativo, quello di riunire insieme i principali esperti dell'epoca, nell'ambito dell'informatica dedicata allo sviluppo di sistemi intelligenti, per una discussione aperta e un confronto di idee e progetti. I partecipanti includevano autori molto importanti, come R. Solomonoff, O. G. Selfridge, T. More, A. L. Samuel, A. Newell e H. A. Simon, ciascuno dei quali negli anni successivi avrebbe contribuito in modo determinante alla ricerca e allo sviluppo di questa materia, svolgendo il proprio lavoro in direzione di nuove conoscenze e applicazioni.

La proposta del convegno immaginava la possibilità di creare un team di lavoro che avrebbe dovuto creare una macchina in grado di simulare ogni aspetto dell'apprendimento e dell'intelligenza umana. A questo proposito, è interessante sottolineare come la proposta includesse la seguente affermazione: «*every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it*», in grado di rappresentarne in modo sintetico e, allo stesso tempo, esplicativo la finalità [94].

A differenza di quanto si possa pensare, la questione delle origini non è di semplice risoluzione, per quanto la data di nascita convenzionale, come spiegato in precedenza, rappresenta senza dubbio un punto sul quale vi è accordo unanime da parte dell'intera comunità scientifica. Infatti, sebbene per descrivere la storia dell'intelligenza artificiale si decida comunemente di cominciare con gli sviluppi successivi al 1956, è doveroso precisare come il suo percorso reale inizi decisamente prima, facendo riferimento, più correttamente, alla cibernetica e alla nascita dei primi calcolatori elettronici.

Tuttavia, l'importanza di questo evento è indescrivibile: la conferenza di Darmouth rappresenta il momento storico in cui l'intelligenza artificiale ottenne il suo nome, la sua mission, il suo primo successo e, soprattutto, i suoi principali attori. Per questo motivo, anche se esistono lavori antecedenti fondamentali, condotti da autori molto importanti, tra i quali è possibile citare W. S. McCulloch, W. H. Pitts, D. O. Hebb e A. M. Turing, che hanno lavorato sotto diverse dimensioni su problemi direttamente correlati, quest'ultima viene tradizionalmente considerata come l'origine dell'intelligenza artificiale.

L'evoluzione dell'intelligenza artificiale è estremamente peculiare, essendo contraddistinta fin da subito da un'alternanza spasmodica di periodi di tempo positivi, di grande

ottimismo ed esaltazione, e periodi di tempo negativi, di grande pessimismo e depressione. Il suo progresso, dagli inizi fino ad oggi, può essere spiegato in modo molto semplice ricorrendo a un'analogia ed evocando come paragone una "giostra di successi e fallimenti"[2], all'interno della quale è possibile individuare sequenzialmente intervalli di tempo caratterizzati da progressi e miglioramenti, che hanno sperimentato la nascita di progetti molto ambiziosi, e intervalli di tempo caratterizzati da imprevisti e rallentamenti, che hanno rischiato di ostacolarne la diffusione. In questo contesto, i primi sono passati alla storia con il nome di "AI spring", mentre i secondi con il nome di "AI winter", per simboleggiare intuitivamente i relativi intervalli temporali.

Negli anni successivi al seminario, gli studi in materia di intelligenza artificiale iniziarono a moltiplicarsi, concentrandosi, almeno inizialmente, sull'intelligenza artificiale simbolica, un approccio di tipo knowledge-based per la creazione di modelli basati su rappresentazioni di problemi simboliche di alto livello, leggibili dall'uomo, logica e ricerca, la quale ha rappresentato il paradigma dominante delle analisi sull'intelligenza artificiale dagli anni '50 fino agli anni '90, con la convinzione di poter creare una macchina con intelligenza artificiale generale come obiettivo finale della ricerca.

Pertanto, gli intervalli di tempo compresi tra il 1957 e il 1974 e tra il 1980 e il 1987 sono definiti rispettivamente "1st AI spring" e "2nd AI spring", comprendendo periodi di tempo molto favorevoli per lo sviluppo di progetti di intelligenza artificiale. In particolare, nel primo caso, la ricerca si è concentrata sullo sviluppo di macchine in grado di eseguire una quantità sempre maggiore di operazioni, come risolvere equazioni algebriche, dimostrare teoremi geometrici e svolgere dialoghi in lingua inglese, mentre, nel secondo caso, la ricerca si è dedicata al progetto di sistemi esperti, che, servendosi di regole specifiche, sono in grado di automatizzare decisioni e compiti del mondo reale, riproducendo le prestazioni di una persona esperta in un determinato dominio di conoscenza o campo di attività.

In aggiunta, è interessante evidenziare due intervalli di tempo particolarmente problematici: il primo, dal 1974 al 1980, e il secondo, dal 1987 al 1993, chiamati rispettivamente "1st AI winter" e "2nd AI winter", per simboleggiare i momenti di difficoltà ed emergenza che l'intelligenza artificiale ha dovuto superare per completare il proprio sviluppo, durante i quali gli studi non progredivano, la ricerca restituiva risultati insufficienti e i finanziatori, principalmente governi, università e grandi imprese, iniziavano a ridurre o, addirittura, annullare il loro supporto.

Durante questi anni di alti e bassi, la maggior parte delle idee che inizialmente potevano sembrare valide proposte commerciali, attirando ingenti investimenti nel settore, finivano inevitabilmente per deludere le aspettative. Le motivazioni alla base di questi fallimenti erano molteplici, ma il problema principale, accomunando praticamente tutti i progetti di intelligenza artificiale, riguardava l'ambiente economico-sociale in cui operare, in quanto, al suo interno, non si erano ancora verificate le condizioni indispensabili per una loro corretta implementazione. A questo proposito, è possibile specificare come gli ingredienti fondamentali per lo sviluppo di modelli di intelligenza artificiale efficaci ed efficienti siano sostanzialmente tre: metodi e algoritmi, dati e potenza di calcolo.

[2] Letteralmente, "*roller coaster of success and setbacks*" [9].

Il percorso dell'intelligenza artificiale si è rivelato quindi essere particolarmente complesso, rendendo necessario non solo esplorare diverse possibilità ed effettuare numerosi tentativi, ma, soprattutto, attendere con estrema fiducia che lo scenario economico-sociale acquisisse le caratteristiche più opportune per favorirne lo sviluppo e la diffusione. Così, per esempio, si è dimostrato fondamentale l'arrivo di metodi e algoritmi, in grado di risolvere alcune questioni aperte di difficile risoluzione, la crescente presenza di dati e la relativa capacità di storage, con la possibilità di immagazzinare un numero sempre maggiore di dati a un costo sempre minore, e, infine, il continuo progresso della potenza di calcolo, sia lato CPU che lato GPU, per consentire di svolgere operazioni progressivamente più complesse riducendone contemporaneamente il tempo e il costo di esecuzione.

In questo modo, la svolta dell'intelligenza artificiale avvenne intorno agli '90, con l'introduzione di un vero e proprio cambio di approccio, ovvero l'apprendimento automatico, o machine learning, che, pur rappresentando una tecnica già conosciuta e utilizzata fin dall'inizio, in quanto alcuni ricercatori si erano sempre mostrati interessati alla possibilità che le macchine imparassero dai dati, anche se, purtroppo, si erano scontrati con il fatto che all'epoca i sistemi di tipo probabilistico erano invasi da problemi sia teoretici che pratici in termini di acquisizione e rappresentazione dei dati, affrontò una nuova rinascita.

In questo periodo, il machine learning iniziò a svilupparsi come un campo separato dall'intelligenza artificiale classica, iniziando a concentrare la propria attenzione non più sulla ricerca dell'intelligenza artificiale generale, bensì sulla possibilità di risolvere problemi di natura pratica con metodi e modelli presi in prestito dalla statistica e dalla teoria della probabilità. Questo approccio non si focalizzava sul tentativo di istruire la macchina alla determinazione di regole, ma sulla determinazione della probabilità dei risultati basati sui dati, seguendo un metodo molto più vicino al processo cognitivo umano.

Da questo momento in poi, l'ascesa dell'intelligenza artificiale si è rivelata inarrestabile, sviluppandosi radicalmente a partire dagli anni '90 e rappresentando oggi una tecnologia sostanzialmente di uso quotidiano. In definitiva, come è possibile osservare nella figura 4.1, la storia dell'intelligenza artificiale risulta costellata di successi e fallimenti, definita da interessi e finanziamenti altalenanti, e, sebbene non abbia certamente seguito un viaggio facile e veloce per arrivare al punto in cui si trova oggi, l'intelligenza artificiale, data la sua importanza, è destinata a crescere sistematicamente nel prossimo futuro.

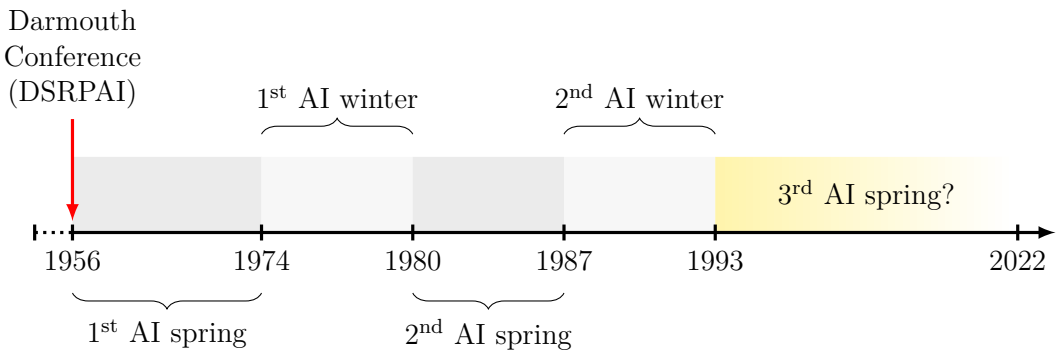


Figura 4.1. Storia dell'intelligenza artificiale.

4.3 Concetti di machine learning e deep learning

A questo punto, è necessario chiarire definitivamente la differenza fondamentale che esiste tra i concetti di intelligenza artificiale, apprendimento automatico (machine learning) e apprendimento profondo (deep learning), osservando lo schema proposto nella figura 4.2, costruito mutuando la teoria degli insiemi. L'intelligenza artificiale è la disciplina che studia lo sviluppo di sistemi intelligenti, dotati di un'intelligenza simile a quella umana. L'apprendimento automatico, o machine learning, è una branca dell'intelligenza artificiale che raccoglie un insieme di tecniche in grado di studiare un insieme di dati, processare le informazioni a disposizione, imparare la conoscenza ed effettuare analisi e predizioni. L'apprendimento profondo, o deep learning, è una branca dell'apprendimento automatico che si basa sull'utilizzo di reti neurali profonde, composte da numerosi strati.

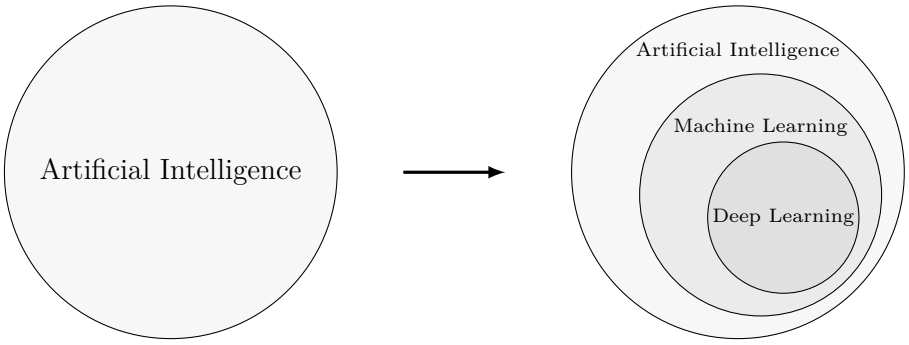


Figura 4.2. Schema dei concetti di machine learning e deep learning.

In questo lavoro di tesi, si lavorerà con i concetti di machine learning e deep learning. Come spiegato più nel dettaglio in precedenza, il machine learning ha rappresentato un momento cruciale per l'implementazione con successo dei concetti di intelligenza artificiale. La prima definizione di machine learning risale al 1959, quando A. L. Samuel inventò personalmente il termine, descrivendolo come «*field of study that gives computers the ability to learn without being explicitly programmed*». Tuttavia, qualche anno più tardi, nel 1966, T. M. Mitchell assegnò una definizione più completa e rigorosa, ancora oggi assolutamente valida: «*a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E*».

In particolare, il machine learning è composto da tre categorie di modelli, caratterizzati in base al contesto in cui si deve muovere la macchina, con l'obiettivo di imparare le regole generali e particolari per raggiungere la conoscenza:

- apprendimento supervisionato (*supervised learning*);
- apprendimento non supervisionato (*unsupervised learning*);
- apprendimento con rinforzo (*reinforcement learning*).

L'apprendimento supervisionato è una tipologia di modello in cui si fornisce alla macchina una numerosità sufficiente di esempi di obiettivi da raggiungere, illustrando le relazioni che esistono tra input, output e risultato. In questo modo, la macchina, studiando i dati a disposizione, deve essere in grado di imparare ed estrapolare una regola generale, che possa permettere, ogni volta che venga stimolata con un nuovo input, di scegliere l'output corretto per il raggiungimento dell'obiettivo.

I principali modelli che appartengono a questa categoria sono i modelli di regressione, dove l'obiettivo consiste nel determinare la relazione tra un insieme di variabili di input e una variabile di output, e i modelli di classificazione, dove l'obiettivo consiste invece nell'assegnare la classe più opportuna a uno specifico elemento.

L'apprendimento non supervisionato è una tipologia di modello in cui non si fornisce alla macchina nessun tipo di allenamento iniziale, spingendo quest'ultima a imparare esclusivamente dagli errori commessi. In questo caso, la macchina deve essere in grado di prendere delle decisioni ed effettuare delle scelte, senza essere stata prima educata sulle differenti possibilità di output al variare degli input selezionati.

I principali modelli che appartengono a questa categoria sono i modelli di clustering, in grado di analizzare le caratteristiche dell'insieme di dati, riconoscere eventuali similarità e, quindi, ottenere una suddivisione in classi distinte, non conosciute in anticipo, ma identificate analizzando le informazioni a disposizione.

L'apprendimento con rinforzo è una tipologia di modello in cui la macchina si trova a interagire con un ambiente dinamico, all'interno del quale le caratteristiche sono mutevoli, con l'obbligo di muoversi per portare a termine un obiettivo non avendo nessun tipo di indicazione se non, alla conclusione della prova, la possibilità di capire se sia riuscita o meno a raggiungere lo scopo iniziale.

Con il seguente lavoro di tesi, si concentra principalmente l'attenzione sui modelli di machine learning e deep learning più importanti, in grado di ottenere performance all'avanguardia e allo stato dell'arte, con la finalità di studiare in modo appropriato il dominio di interesse, ovvero gli alberi decisionali e le reti neurali.

4.4 Alberi decisionali

Gli alberi decisionali sono un algoritmo di apprendimento supervisionato, utilizzabile per risolvere un'ampia gamma di problemi, sia di regressione che di classificazione. L'obiettivo è quello di creare un modello che preveda il valore di una variabile target, imparando semplici regole decisionali ricavabili dalle informazioni delle caratteristiche dei dati.

4.4.1 Architettura

Gli alberi decisionali sono composti da una struttura ad albero gerarchica (figura 4.3), formata da un insieme di elementi diversi: nodo radice, nodi interni, nodi foglia e rami[3]. La struttura inizia con un nodo radice o nodo iniziale, che non possiede rami in entrata, ma

[3]In inglese, rispettivamente *root node*, *internal node*, *leaf node* e *branch*.

che con una serie di rami in uscita alimenta i nodi interni, chiamati anche nodi decisionali, fino al raggiungimento dei nodi foglia o nodi terminali.

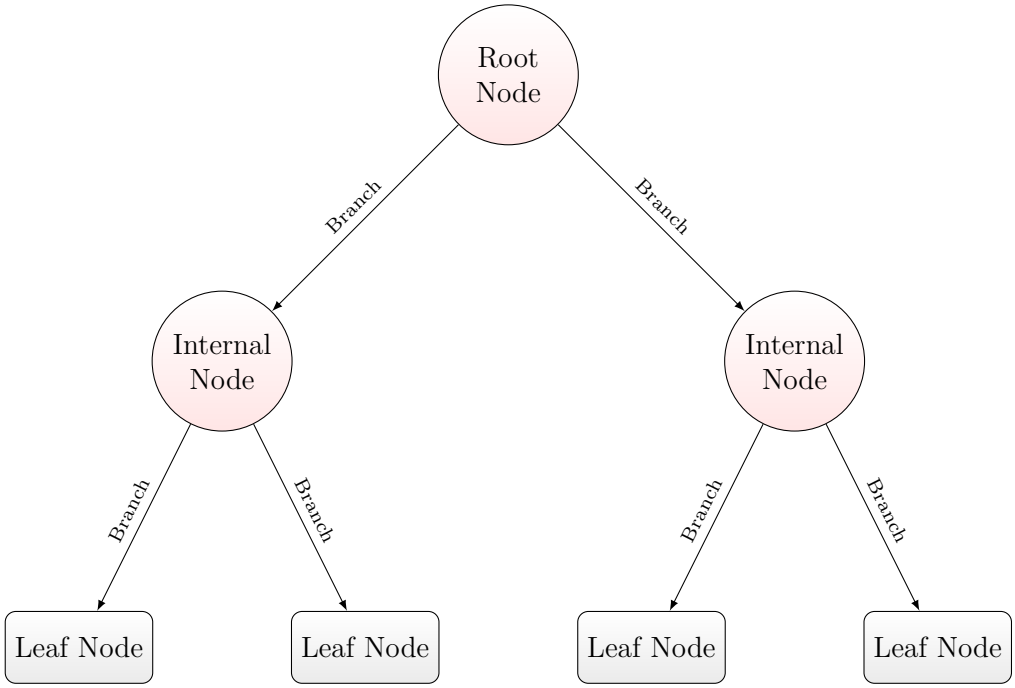


Figura 4.3. Struttura di un albero decisionale.

Analizzando le funzionalità disponibili al variare dell'ambito di applicazione, il nodo radice e i nodi interni effettuano valutazioni di specifiche caratteristiche per dividere gli elementi in esame e formare sottoinsiemi omogenei, che sono rappresentati dai nodi foglia, segnalando tutti i possibili risultati all'interno dei dati a disposizione.

Gli alberi decisionali coinvolgono decisioni multiple su una o più variabili di input in modo da separare in modo ottimale gli elementi e produrre il minor numero di errori possibile, pur consapevoli che, in quanto l'insieme di dati non è perfettamente separabile, non sarà possibile evitare errori. Pertanto, in un albero di decisione ogni nodo interno effettua un test su un attributo, ogni ramo corrisponde a uno dei possibili valori che l'attributo può assumere e ogni nodo foglia assegna una classificazione.

4.4.2 Apprendimento

L'apprendimento di qualsiasi albero decisionale utilizza una strategia "divide et impera", con la quale si prevede di studiare le variabili di input del problema, in grado di descrivere le caratteristiche del dataset in esame, e condurre una ricerca per identificare i punti di divisione ottimali all'interno dell'albero decisionale stesso. Questa tipologia di apprendimento costruisce l'albero decisionale top-down, nel quale il processo di suddivisione viene

ripetuto in modo ricorsivo dall'alto verso il basso, fino a quando tutti i record sono stati classificati nelle diverse variabili di output.

A questo riguardo, è interessante analizzare come scegliere il miglior attributo in ciascun nodo, in modo da ottimizzare il processo di suddivisione, generando sottoinsiemi di elementi più puri e omogenei possibili, ovvero selezionando tra le variabili di interesse il punto di decisione caratterizzato da un maggior potere discriminante, che costituisce l'abilità di separare chiaramente tra classi diverse.

Per esprimere questa idea matematicamente, è necessario ricorrere all'utilizzo del concetto di criterio decisionale, o *splitting criteria*, che rappresenta l'elemento essenziale degli alberi decisionali e la cui scelta risulta fondamentale per spiegare il meccanismo alla base del funzionamento di questa tipologia di modelli.

4.4.3 Criteri decisionali

Data una variabile aleatoria x, che può assumere k valori discreti, distribuita secondo $P = (p_1, p_2, \ldots, p_k)$, si definisce misura di impurità una funzione $\phi\colon [0,1]^k \to \mathbb{R}$ che soddisfa le seguenti condizioni:

- $\phi(P) \geq 0$;
- $\phi(P)$ è minima se $\exists i \mid p_i = 1$;
- $\phi(P)$ è massima se $\forall i$, $1 \leq i \leq k$, $p_i = \frac{1}{k}$;
- $\phi(P)$ è simmetrica;
- $\phi(P)$ è differenziabile.

Dato un training set S il vettore di probabilità della variabile target y è definito nel modo seguente:

$$P_y(S) = \left(\frac{|\sigma_{y=c_1} S|}{|S|}, \frac{|\sigma_{y=c_2} S|}{|S|}, \ldots, \frac{|\sigma_{y=c_{|dom(y)|}} S|}{|S|} \right) . \tag{4.1}$$

La bontà della divisione dovuta all'attributo discreto a_i è definita come la riduzione dell'impurità della variabile target y dopo aver partizionato S in accordo con i valori $v_{i,j} \in dom(a_i)$:

$$\Delta\Phi(a_i, S) = \phi(P_y(S)) - \sum_{j=1}^{|dom(a_i)|} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot \phi(P_y(\sigma_{a_i=v_{i,j}} S)) . \tag{4.2}$$

I principali criteri decisionali per selezionare l'attributo migliore in ciascun nodo, definiti in letteratura *impurity-based criteria*, sono i seguenti:

- Information Gain;
- Gini Impurity.

Information Gain

Information Gain, tradotto guadagno di informazioni, è un criterio decisionale che utilizza l'entropia come misura di impurità per selezionare l'attributo ottimale.

L'entropia è un concetto fondamentale, applicato prima nella termodinamica classica e dopo nella meccanica quantistica, che, nel corso del tempo, è stato esteso a numerosi ambiti, tra cui la teoria dell'informazione, la teoria dei segnali e l'economia. A tale proposito, l'entropia è stata introdotta per la prima volta nel campo della teoria dell'informazione da C. E. Shannon nel 1948, il quale, riprendendo e traslando la definizione originale data da L. E. Boltzmann e J. W. Gibbs, utilizza la seguente formula per definire il livello medio di informazione di una variabile aleatoria discreta (figura 4.4):

$$H(y, S) = \sum_{c_j \in dom(y)} -\frac{|\sigma_{y=c_j} S|}{|S|} \cdot \log_b \frac{|\sigma_{y=c_j} S|}{|S|} \ . \tag{4.3}$$

In questa formula, b rappresenta la base del logaritmo utilizzato, i cui valori comuni sono 2, 10 e il numero di Eulero e, in modo da stabilire anche la rispettiva unità di misura: bit per $b = 2$, ban per $b = 10$ e nat per $b = e$.

In poche parole, l'entropia può essere interpretata come il grado di disordine di un sistema: un aumento di entropia è associato verso un aumento del disordine del sistema, mentre una riduzione di entropia è associata verso una riduzione del disordine del sistema. Per questo motivo, l'entropia, rappresentando una misura di disordine relativo, può essere utilizzata per descrivere quanto il dataset in esame sia impuro e disomogeneo.

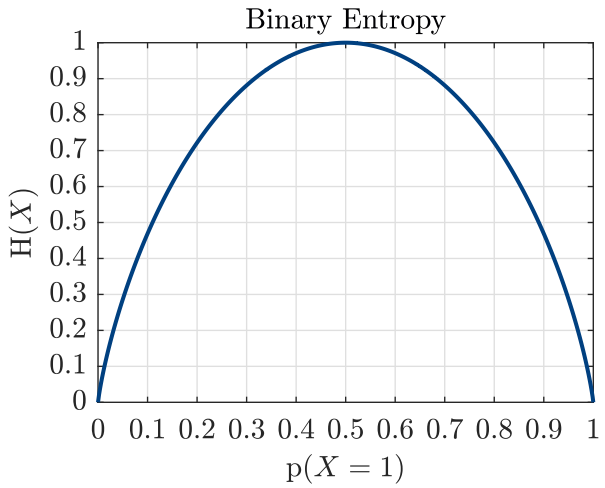


Figura 4.4. Entropia di una sorgente binaria.

A questo punto, è possibile definire il guadagno di informazioni come la differenza di entropia precedente e successiva a una divisione per un determinato attributo, calcolando la conseguente riduzione a seguito della trasformazione subita dal dataset.

Pertanto, il guadagno di informazione di S per l'attributo a_i è definito come la differenza tra l'entropia del dataset $H(y, S)$ e l'entropia condizionata $H(y, S \mid a_i)$:

$$IG(a_i, S) = H(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i = v_{i,j}} S|}{|S|} \cdot H(y, \sigma_{a_i = v_{i,j}} S) \ . \tag{4.4}$$

Utilizzando Information Gain come criterio decisionale per selezionare la caratteristica migliore in base alla quale effettuare la divisione e ottenere l'albero decisionale ottimale, è quindi necessario scegliere l'attributo con il maggiore guadagno di informazioni.

Gini Impurity

Gini Impurity, tradotto impurità di Gini, è un criterio decisionale che calcola la probabilità di classificare in modo errato un generico elemento del dataset, assegnandolo casualmente a una classe specifica secondo la relativa distribuzione di probabilità (figura 4.5):

$$GI(y, S) = 1 - \sum_{c_j \in dom(y)} \left(\frac{|\sigma_{y = c_j} S|}{|S|} \right)^2 \ . \tag{4.5}$$

Utilizzando Gini Impurity come criterio decisionale per selezionare la caratteristica migliore in base alla quale effettuare la divisione e ottenere l'albero decisionale ottimale, è quindi necessario scegliere l'attributo con la minore impurità di Gini.

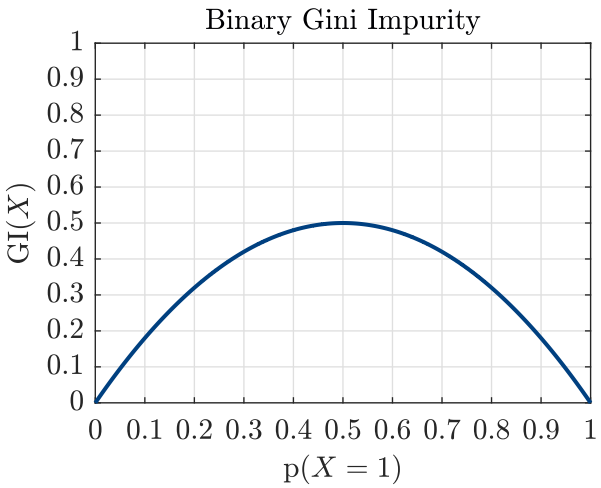


Figura 4.5. Impurità di Gini di una sorgente binaria.

In alternativa, parallelamente al caso precedente, è possibile utilizzare una formulazione diversa, definendo il guadagno di Gini di S per l'attributo a_i:

$$GG(a_i, S) = GI(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i = v_{i,j}} S|}{|S|} \cdot GI(y, \sigma_{a_i = v_{i,j}} S) \ . \tag{4.6}$$

In questo caso, per produrre la miglior suddivisione e ottenere l'albero decisionale ottimale, è necessario selezionare l'attributo con il maggiore guadagno di Gini.

Il guadagno di informazioni e l'impurità di Gini sono i criteri decisionali più utilizzati, in grado di ottenere i migliori risultati. Questi metodi sono praticamente equivalenti e non mostrano particolari differenze in termini di performance nella maggior parte delle applicazioni, ma, poiché il guadagno di informazioni richiede il calcolo del logaritmo, l'impurità di Gini viene molto spesso preferita per ottenere un'esecuzione più veloce, richiedendo un minore sforzo computazionale.

4.4.4 Ensemble

Gli alberi decisionali rappresentano un modello di intelligenza artificiale semplicemente comprensibile e interpretabile, ma con un potere predittivo limitato, soprattutto in caso di problemi caratterizzati da elevata complessità. Per superare questo problema, i moderni algoritmi basati sull'applicazione di alberi decisionali ottengono performance più elevate, superando l'utilizzo di un singolo albero decisionale e combinando insieme una molteplicità di alberi decisionali, con la finalità di formare i cosiddetti *ensemble*.

Queste tecniche rappresentano un concetto di machine learning molto importante, in cui l'idea di base consiste nell'addestrare numerosi modelli diversi impiegando lo stesso algoritmo di apprendimento, in modo tale che, con la combinazione congiunta di più classificatori, si ottenga un risultato più affidabile rispetto a un singolo classificatore.

A questo proposito, è possibile mettere in evidenza due tecniche *ensemble*, le quali, come illustrato graficamente in figura 4.6, utilizzano, combinandoli insieme, una serie di classificatori "deboli" per creare un classificatore "forte", in grado di ottenere prestazioni decisamente migliori:

- Bagging;
- Boosting.

Queste due tecniche *ensemble* risultano simili, ma presentano anche alcune differenze. In primo luogo, sia Bagging che Boosting producono N classificatori, generando, durante il processo di apprendimento, N dati aggiuntivi, i quali vengono costruiti utilizzando un campionamento casuale con ripetizione a partire dagli elementi del dataset originale. Ad ogni modo, in Bagging le osservazioni non sono pesate, comportando per ogni elemento la stessa probabilità di apparire nel nuovo training dataset; viceversa, in Boosting le osservazioni sono pesate, determinando per alcuni elementi la possibilità di apparire con maggior frequenza nel nuovo training dataset. Il risultato è quello di generare insiemi di dati diversi che vengono utilizzati per addestrare lo stesso modello di intelligenza artificiale, producendo, conseguentemente, classificatori diversi.

In secondo luogo, sia Bagging che Boosting costituiscono metodi capaci di determinare N classificatori partendo da un singolo classificatore. In ogni caso, Bagging sfrutta una fase di apprendimento parallela, dove ogni modello viene costruito in modo indipendente; al contrario, Boosting sfrutta una fase di apprendimento sequenziale, dove ogni classificatore viene allenato valutando l'esperienza del precedente classificatore, dal momento che, al termine di ogni step di allenamento, i pesi vengono ridistribuiti in base al risultato della classificazione, assegnando un incremento relativo ai dati classificati erroneamente, in modo tale che i successivi classificatori possano favorire la loro corretta identificazione.

In terzo luogo, Bagging e Boosting, per predire la classe di un nuovo dato, applicano N classificatori alla nuova osservazione, applicando una media, semplice o pesata, oppure selezionando la maggior parte di essi secondo uno specifico criterio. In formule:

$$E_{Boosting} = \frac{1}{N} \sum_{i=1}^{N} e_i \ . \tag{4.7}$$

$$E_{Bagging} = \frac{1}{N} \sum_{i=1}^{N} w_i e_i \ . \tag{4.8}$$

Infatti, in Bagging, il risultato è ottenuto facendo la media semplice della risposta degli N classificatori, mentre in Boosting il risultato è ottenuto da una media ponderata degli N classificatori, assegnando un peso maggiore ai classificatori con le migliori performance.

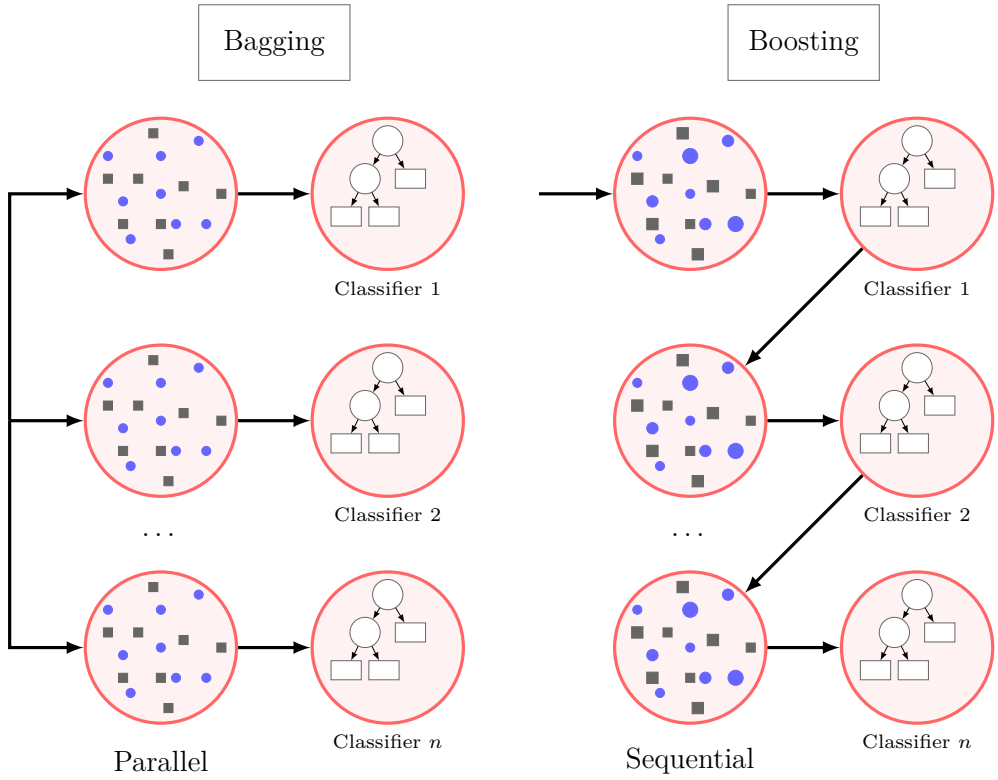


Figura 4.6. Caratteristiche Bagging vs Boosting.

Le principali difficoltà che si possono verificare durante il processo di apprendimento sono rumore, bias e varianza. Le diverse tecniche *ensemble*, la cui scelta non risulta unica e dipende dalle peculiarità del problema in esame, aiutano a minimizzare questi fattori, essendo progettate precisamente per migliorare la stabilità e l'accuratezza degli algoritmi di machine learning. In generale, se il modello presenta il fenomeno di underfitting, con performance relativamente modeste sul training set, conviene utilizzare Boosting, in

modo da diminuire la varianza, mentre se il modello presenta il fenomeno di overfitting, con performance eccessivamente elevate sul training set, conviene utilizzare Bagging, in modo da diminuire il bias.

In ultima analisi, è possibile nominare le principali applicazioni delle tecniche *ensemble*, Boosting e Bagging, appena discusse, ovvero rispettivamente Random Forest e XGBoost.

4.4.5 Caratteristiche principali

Gli alberi decisionali forniscono un valido strumento per risolvere una grande varietà di problemi in diversi ambiti di applicazione e, come strumento di supporto alle decisioni, presentano numerosi vantaggi, ma anche qualche svantaggio (tabella 4.1).

I principali vantaggi sono:

- Facili da utilizzare: gli alberi decisionali possono essere visualizzati nella propria struttura gerarchica, in modo da rendere facilmente comprensibile e interpretabile quali attributi siano i più importanti.
- Minima preparazione dei dati: gli alberi decisionali non richiedono praticamente nessuna preparazione dei dati, dal momento che non necessitano di operazioni di normalizzazione o standardizzazione e possono gestire non solo diversi tipi di dati (valori discreti e continui), ma anche dati con valori mancanti.
- Grande flessibilità: gli alberi decisionali possono essere applicati per risolvere problemi sia di regressione che di classificazione, con un funzionamento non influenzato dalle relazioni presenti tra agli attributi, gestendo automaticamente il problema della correlazione tra due o più variabili.
- Modello white-box: gli alberi decisionali, utilizzando la logica booleana, rendono facilmente interpretabili i risultati del modello.
- Grande robustezza: gli alberi decisionali mantengono buone prestazioni anche se le ipotesi alla base del modello sono in qualche modo violate dal modello reale con cui sono stati generati i dati a disposizione.

I principali svantaggi sono:

- Rischio overfitting: gli alberi decisionali possono generare modelli eccessivamente complessi e profondi, che non consentono di generalizzare correttamente nuovi dati (richiedendo, per superare il problema, l'utilizzo di tecniche di potatura, o *pruning*, che si distinguono in pre-potatura o post-potatura).
- Instabilità: gli alberi decisionali producono stimatori di varianza elevata, con piccole variazioni all'interno dei dati che possono generare una struttura dell'albero ottimale anche completamente diversa.
- Addestramento complesso: gli alberi decisionali richiedono per l'addestramento un maggior costo e tempo computazionale.

Tabella 4.1. Caratteristiche principali alberi decisionali.

Alberi decisionali	
Vantaggi	Svantaggi
Facili da utilizzare	Rischio overfitting
Minima preparazione dei dati	Instabilità
Grande flessibilità	Addestramento complesso
Modello white-box	
Grande robustezza	

4.5 Reti neurali

Le reti neurali, più esattamente reti neurali artificiali, sono un algoritmo di apprendimento supervisionato, in grado di risolvere numerosi problemi di regressione e di classificazione, il cui nome e la cui struttura è ispirata al cervello umano, imitando e simulando il modo in cui i neuroni biologici si inviano segnali.

4.5.1 Neurone biologico

Il cervello rappresenta l'organo fondamentale del sistema nervoso, essendo il principale responsabile delle funzioni cognitive umane che sovrintendono il lavoro dell'organismo: gioca un ruolo essenziale nel controllo del respiro e del movimento, genera continuamente pensieri ed emozioni, custodisce la memoria e gestisce numerose attività involontarie.

Il cervello è composto da circa 100 miliardi di neuroni, che descrivono le "cellule-base", specifiche del sistema nervoso, in grado di ricevere, elaborare e trasmettere informazioni da un punto a un altro sotto forma di impulsi elettrochimici ad alta velocità.

I neuroni sono le unità cellulari più complesse dell'organismo, costituiscono il tessuto nervoso e rappresentano il componente indispensabile del cervello umano. Questi ultimi, nonostante presentino sistematicamente una struttura morfologica differente al variare della particolare funzione che sono chiamati a svolgere o del luogo in cui sono collocati, sono sempre formati da tre parti principali, come visibile in figura 4.7:

- soma (corpo cellulare): corpo del neurone, all'interno del quale si trova il nucleo;
- dendriti: canali di input del neurone, responsabili della ricezione dei messaggi;
- assone: canale di output del neurone, incaricato della trasmissione dei messaggi.

I neuroni adempiono simultaneamente a due funzioni indispensabili e interconnesse: una funzione di input, con la ricezione delle diverse informazioni in entrata, sia dall'interno, attraverso tutti gli altri neuroni, che dall'esterno, attraverso i canali sensoriali, oltre che una funzione di output, con la trasmissione delle informazioni in uscita dove necessario.

Il trasferimento delle informazioni tra cellule avviene attraverso le sinapsi, presenti nel cervello con una numerosità stimata compresa tra 10^{13} e 10^{15}, che specificano i punti di

contatto tra neuroni e, più precisamente, l'unione dell'assone di un neurone precedente con i dendriti di neuroni successivi, diventando il ponte che consente di trasmettere il messaggio da una cellula all'altra.

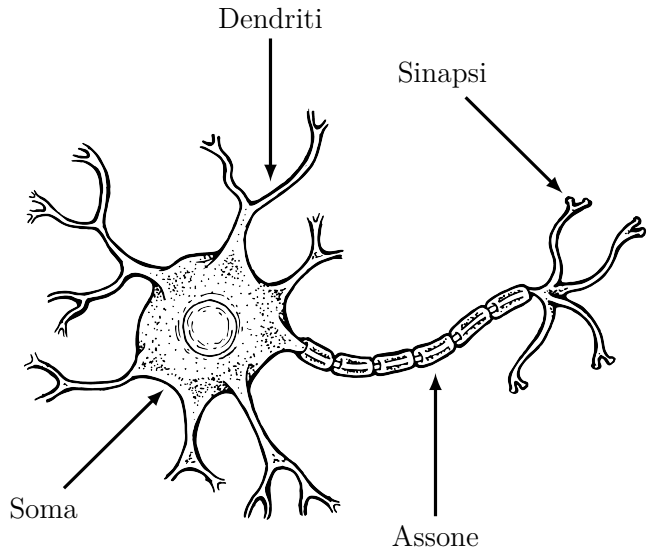


Figura 4.7. Rappresentazione di un neurone biologico.

I neuroni funzionano trasmettendo un impulso nervoso, ovvero un impulso di natura elettrochimica, il cui trasferimento è possibile in considerazione del fatto che la membrana plasmatica di un neurone risulta polarizzata, presentando una differenza di carica elettrica tra l'esterno e l'interno della cellula dovuta alla concentrazione di ioni positivi superiore all'esterno che all'interno. La distribuzione asimmetrica di carica elettrica genera una differenza di potenziale elettrico, definita potenziale a riposo, che viene mantenuto costante dall'azione della pompa sodio-potassio, una proteina di membrana che in modo attivo trasporta ioni sodio Na^+ con una direzione interno-esterno cellula e ioni potassio K^+ con una direzione esterno-interno cellula.

In questo contesto, gli ioni K^+ possono scorrere liberamente attraverso determinate proteine di membrana, dette proteine canale per il potassio, in modo da equilibrarne la relativa concentrazione, spostandosi dall'interno all'esterno della cellula; al contrario, gli ioni Na^+ tenderebbero a distribuirsi in modo uniforme attraverso altre determinate proteine di membrana, dette proteine canale per il sodio, ma questo passaggio non è consentito, dal momento che, quando il neurone si trova a riposo, i canali risultano chiusi. Di conseguenza, il potenziale a riposo rimane costante, al valore di circa $-70\,\mathrm{mV}$.

L'impulso nervoso è generato dall'improvvisa variazione della differenza di potenziale, prima di ripristinare le condizioni normali. Quando il neurone viene stimolato, infatti, con l'apertura di alcune proteine canale per il sodio, il potenziale della membrana aumenta, passando da un valore di $-70\,\mathrm{mV}$ a un valore di $-50\,\mathrm{mV}$, definito potenziale di soglia.

A questo punto, si verifica una sequenza di azioni, in grado di generare due processi

distinti che avvengono in ordine cronologico, definiti rispettivamente depolarizzazione e ripolarizzazione della membrana. Come prima cosa, una volta raggiunto il valore di soglia, si aprono molte proteine canale per il sodio, spingendo un gran numero di ioni sodio Na^+ dall'esterno all'interno della cellula e aumentando la concentrazione di cariche positive, in modo da invertire bruscamente il potenziale elettrico fino a un valore di circa +35 mV, chiamato potenziale d'azione. Immediatamente dopo, le proteine canale per il sodio si richiudono, mentre le proteine canale per il potassio, fino a questo momento chiuse, si riaprono e, grazie alla pompa sodio-potassio, vengono ristabilite le condizioni a riposo.

In sintesi, il potenziale d'azione è un fenomeno elettrochimico localizzato e, affinché si verifichi con successo la trasmissione dell'impulso nervoso, il processo di depolarizzazione deve trasmettersi dalla zona in cui è avvenuto all'area immediatamente adiacente, con una propagazione dello stimolo che avviene unicamente in una direzione, a causa del fatto che, nel tratto interessato dal potenziale d'azione, la pompa sodio-potassio opera per riportare le condizioni a riposo. Come risultato, l'impulso nervoso viene generato e trasmesso lungo la membrana plasmatica degli assoni delle cellule nervose.

4.5.2 Neurone artificiale

Le reti neurali si fondano sul funzionamento del cervello umano, prendendo ispirazione dalla propagazione delle informazioni tramite impulso nervoso tra neuroni e sinapsi. In questo caso, è immediato ottenere un parallelo tra neurone biologico e neurone artificiale: i neuroni rappresentano le unità elementari di elaborazione, mentre le sinapsi rappresentano le connessioni orientate e ponderate tra questi ultimi.

Modello di W. S. McCulloch - W. H. Pitts

Il primo storico modello computazionale di neurone artificiale, è stato presentato nel 1943 da W. S. McCulloch e W. H. Pitts con un lavoro destinato a diventare un pilastro della materia, intitolato "*A logical calculus of the ideas immanent in nervous activity*" [95].

Il neurone McCulloch-Pitts, chiamato anche neurone MCP, rappresenta il primo tentativo di utilizzare un semplice modello matematico per riprodurre un neurone biologico e costituisce un contributo storico fondamentale per lo sviluppo delle reti neurali moderne. L'idea di base è quella di astrarre il comportamento di un neurone biologico in un semplice modello matematico, in grado di generare schemi più o meno complessi tramite neuroni tra loro interconnessi, gestendo input e output binari, 0 e 1.

Come visibile in figura 4.8, il neurone artificiale può essere frazionato in due parti distinte: la prima parte g riceve un input e svolge una determinata operazione di calcolo, mentre la seconda parte f, a seguito del risultato ottenuto, prende una decisione in output.

Il funzionamento del modello è il seguente: il neurone artificiale riceve i segnali in ingresso sotto forma di input $x_i \in [0,1]$, effettua una somma ponderata con i rispettivi pesi w_i ed emette un segnale in uscita sotto forma di output $y \in [0,1]$.

Nello specifico, se la somma ponderata risulta maggiore o uguale a un determinato valore soglia, o *threshold*, T, il neurone si attiva, emettendo un segnale in uscita con output $y = 1$; viceversa, se la somma ponderata risulta minore il neurone si disattiva, emettendo un segnale in uscita con output $y = 0$.

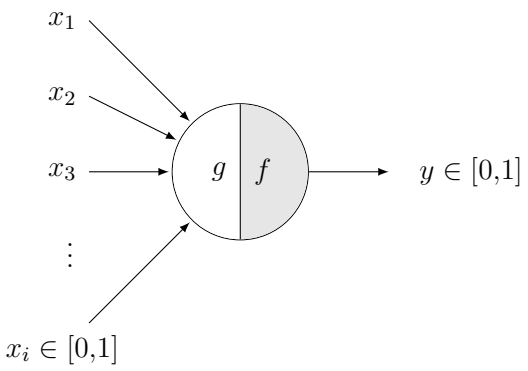


Figura 4.8. Modello di neurone artificiale di W. S. McCulloch e W. H. Pitts.

Formalmente, il modello di neurone artificiale può essere spiegato come segue:

$$y = \begin{cases} 1 & \text{se } \sum_{i=1}^{n} x_i w_i \geq T \\ 0 & \text{altrimenti} \end{cases} \tag{4.9}$$

dove:

- $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ = vettore dei segnali in input;
- $\mathbf{w} = [w_1, w_2, \ldots, w_n]$ = vettore dei pesi;
- T = valore soglia previsto dal modello.

In questo caso, è possibile notare un'evidente analogia con il neurone biologico: i dendriti ricevono i segnali in input, la soma processa l'informazione, l'assone emette il segnale in output e, in aggiunta, le sinapsi formano i punti di connessione tra neuroni.

Modello attuale

Nel corso del tempo, molti studiosi hanno svolto il proprio lavoro, riprendendo il progetto di W. S. McCulloch e W. H. Pitts, tra i quali è possibile citare D. O. Hebb, B. G. Farley, W. A. Clark, N. Rochester, J. H. Holland, L. H. Haibt, W. L. Duda, F. Rosenblatt, M. L. Minsky, S. A. Papert e molti altri, che hanno permesso di effettuare importanti passi in avanti per arrivare al modello attuale di rete neurale, sviluppando il neurone artificiale nella sua versione moderna.

In generale, come descritto in precedenza, una rete neurale è composta da una numerosità di elementi, chiamati neuroni, connessi tra loro tramite relazioni, chiamate sinapsi. I neuroni sono disposti a strati, definiti layers, dove i neuroni dello strato più esterno della rete neurale ricevono n variabili in inuput, le elaborano attraverso una funzione, lineare o non lineare, definita funzione di attivazione, e trasmettono in output il relativo risultato ai neuroni dello strato seguente. A loro volta, questi ultimi processano, con un'ulteriore funzione, gli input ricevuti e formulano in output un nuovo risultato allo strato seguente,

replicando questo processo fino al raggiungimento dell'ultimo strato della rete neurale. Infatti, dopo una serie di strati nascosti, chiamati hidden layer, il cui numero può variare in base alle scelte effettuate in fase di progettazione, la rete neurale produce in output il risultato finale del modello.

Così, una rappresentazione grafica di quanto appena discusso è illustrata in figura 4.9.

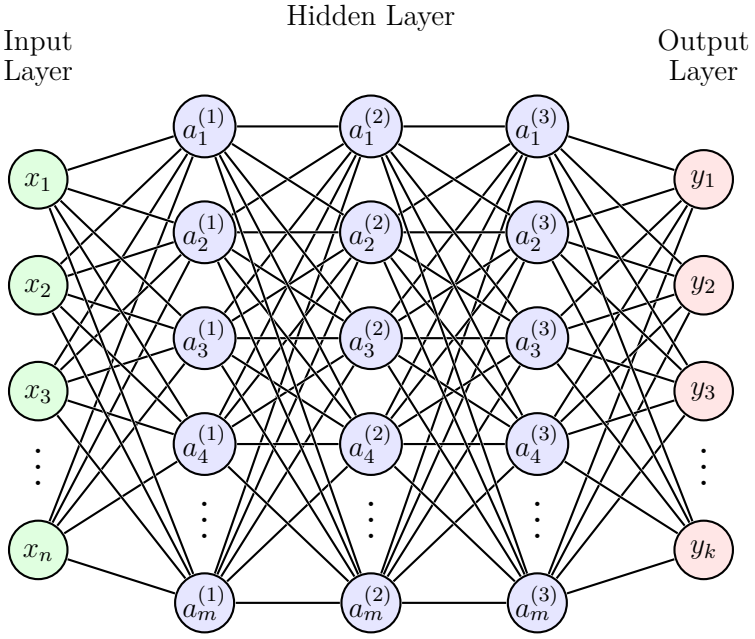


Figura 4.9. Architettura di una rete neurale.

Formalmente, questo modello matematico può essere spiegato come segue in figura 4.10, definendo il livello di eccitazione del neurone j, ovvero net_j:

$$y_j = \varphi(net_j) = \varphi(\sum_{i=1}^{n} x_{ij} \cdot w_{ij} + b_j) \ , \tag{4.10}$$

dove:

- $\mathbf{x} = [x_{1j}, x_{2j}, \ldots, x_{nj}]$ = vettore dei segnali in input;
- $\mathbf{w} = [w_{1j}, w_{2j}, \ldots, w_{nj}]$ = vettore dei pesi;
- b_j = bias;
- φ = funzione di attivazione.

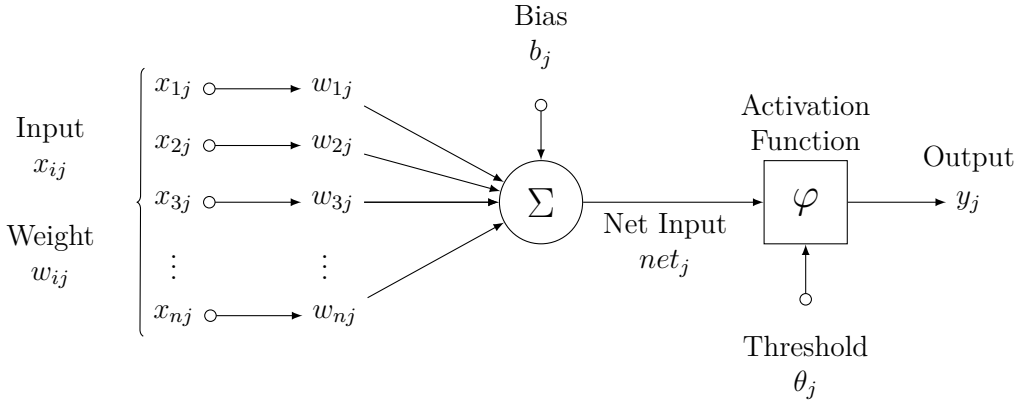


Figura 4.10. Modello matematico di una rete neurale.

4.5.3 Funzioni di attivazione

In una rete neurale, ogni singolo neurone artificiale elabora una numerosità di input e pesi per calcolare una somma ponderata, che, con l'intervento di una funzione di attivazione, viene convertita in un risultato in output. In poche parole, una funzione di attivazione, che può essere definita con la seguente forma funzionale, viene impiegata per mappare l'ingresso e l'uscita, ovvero trasformare l'input in output:

$$output = \varphi(\sum input \cdot weight + bias) \ . \tag{4.11}$$

Le funzioni di attivazione più impiegate per elaborare i calcoli di una rete neurale sono molteplici, con caratteristiche specifiche che oscillano al variare del campo di applicazione:

- Binary Step Function;
- Linear Activation Function;
- Sigmoid Function;
- Tahn Function;
- Rectified Linear Unit Function;
- Leaky Rectified Linear Unit Function;
- Parametric Rectified Linear Unit Function;
- Exponential Linear Unit Function;
- Gaussian Error Unit Function;
- Swish Function;
- Softmax Function.

Le funzioni di attivazione costituiscono una componente fondamentale del meccanismo di costruzione di qualsiasi rete neurale e possono essere lineari o non lineari. In realtà, le funzioni di attivazione lineari non vengono praticamente mai utilizzate, dal momento che non risultano sufficienti per formare un'approssimazione di funzione universale, impedendo alla rete neurale di imparare relazioni complesse nei dati. Pertanto, essendo la maggior parte dei problemi reali caratterizzati da estrema complessità, si ricorre all'uso di funzioni di attivazione non lineari per riuscire a modellare correttamente questa complessità.

Una funzione di attivazione non lineare è una funzione che presenta la caratteristica di non linearità, implicando che l'output non possa essere generato da una combinazione lineare di input. In aggiunta, questa funzione deve essere continua e differenziabile, in modo da poter rendere facilmente applicabile l'algoritmo di retro-propagazione dell'errore, o *error back-propagation*, che rappresenta il processo con il quale si addestra la rete neurale.

Binary Step Function

Binary Step Function è una funzione di attivazione lineare basata su un valore soglia, responsabile della decisione di attivazione per uno specifico neurone. In questo caso, la funzione confronta il valore in input con il valore soglia, restituendo il valore 1 e attivando il neurone se il valore in input risulta maggiore del valore soglia o, viceversa, restituendo il valore 0 e disattivando il neurone se il valore in input risulta minore del valore soglia.

Matematicamente, Binary Step Function può essere rappresentata come (figura 4.11):

$$BS(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} . \tag{4.12}$$

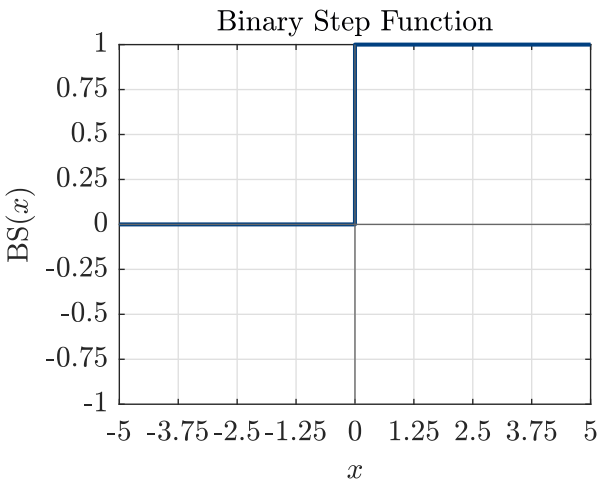


Figura 4.11. Binary Step Function.

Binary Step Function costituisce la funzione di attivazione più semplice, ma presenta alcune limitazioni: da una parte, non può gestire output con valori multipli e, quindi, non

può essere usata per attività di classificazione in problemi caratterizzati da classi multiple; dall'altra, il gradiente della funzione è nullo, causando potenziali problemi nel processo di *error back-propagation*. Calcolare il gradiente e ottenere un valore uguale a zero, significa che non è possibile aggiornare i pesi e i bias durante il processo di *error back-propagation*, dal momento che il gradiente viene proprio utilizzato per l'aggiornamento di questi ultimi.

Linear Activation Function

Linear Activation Function è una funzione di attivazione lineare, che risulta direttamente proporzionale al valore in input tramite una costante a, selezionabile liberamente.

Matematicamente, Linear Function può essere rappresentata come (figura 4.12):

$$LA(x) = ax \ . \tag{4.13}$$

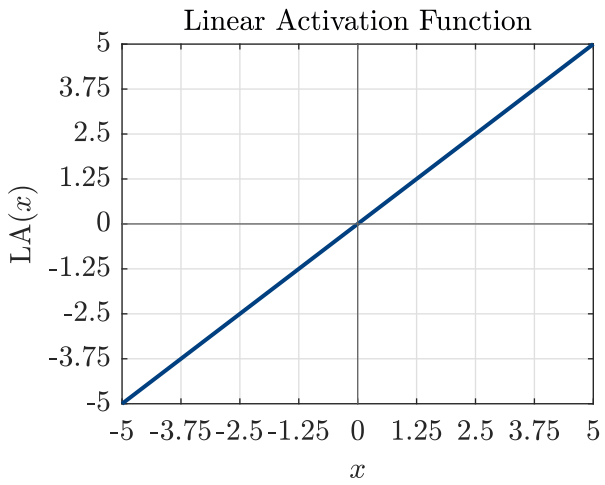


Figura 4.12. Linear Activation Function.

Linear Activation Function può essere utilizzata per risolvere il principale problema della Binary Step Function che produce un gradiente uguale a zero, in quanto non presenta nessuna componente di x nella funzione. In questo caso, il gradiente non risulta nullo, bensì uguale alla costante a utilizzata, indipendente dal valore in input, implicando in automatico l'aggiornamento dei pesi e dei bias nel processo di *error back-propagation* con un medesimo fattore di aggiornamento.

Tuttavia, non si osserva un particolare vantaggio complessivo, poiché la rete neurale, a causa dell'utilizzo dello stesso valore del gradiente per ogni iterazione, non migliorerà sensibilmente l'errore commesso e non sarà capace di osservare pattern complessi nei dati.

In aggiunta, impiegando questa funzione di attivazione, tutti gli strati di una rete neurale finiscono per collassare in un unico strato, dal momento che, indipendentemente dal numero di strati presenti, l'ultimo risulterà comunque una funzione lineare del primo, trasformando essenzialmente qualsiasi struttura, dalla più semplice alla più complessa, in una rete neurale di un solo livello.

Sigmoid Function

Sigmoid Function è una funzione di attivazione non lineare, continua e differenziabile in tutti i suoi punti, che produce una curva a forma di "S", ricevendo in input un qualsiasi valore e restituendo in output valori compresi nell'intervallo (0,1). La funzione non risulta simmetrica rispetto all'origine e, conseguentemente, prevede lo stesso segno per tutti i valori in output, rendendo il processo di addestramento più difficile e instabile.

Matematicamente, Sigmoid Function può essere rappresentata come (figura 4.13):

$$S(x) = \frac{1}{1 + e^{-x}} \ . \tag{4.14}$$

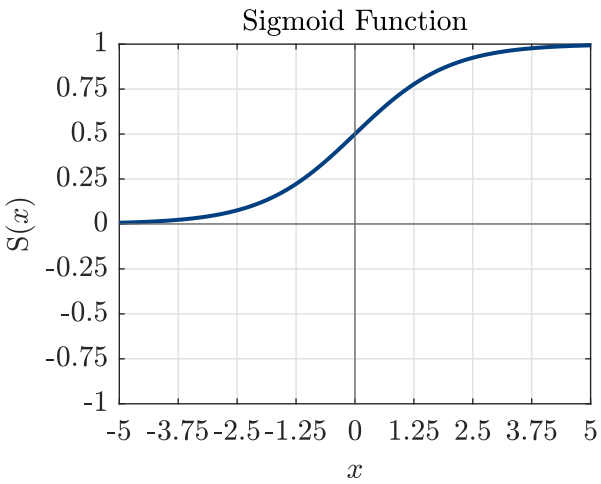


Figura 4.13. Sigmoid Function.

Sigmoid Function rappresenta la stessa funzione applicata per la regressione logistica e, grazie alle sue caratteristiche, costituisce una delle funzioni di attivazione più utilizzate. A tale proposito, questa funzione di attivazione viene comunemente impiegata per costruire modelli in cui è necessario prevedere in output una probabilità, essendo quest'ultima definita con un numero compreso tra 0 e 1.

Sigmoid Function presenta un unico limite: il fenomeno di *vanishing gradient problem*, ovvero il problema della scomparsa del gradiente, connesso direttamente alla derivata della funzione, la quale, presentando contemporaneamente gradienti elevati nella parte centrale e gradienti modesti nelle parti più esterne, può condurre a eventuali problemi quando i valori del gradiente si avvicinano a zero.

Tahn Function

Tahn Function, o Hyperbolic Tangent Function, è una funzione di attivazione non lineare molto simile alla Sigmoid Function, essendo continua e differenziabile in tutti i suoi punti e presentando la stessa forma a "S", anche se adesso, a differenza del caso precedente,

i valori in output sono compresi nell'intervallo $(-1,1)$. La funzione risulta simmetrica rispetto all'origine, in modo da poter facilmente mappare i valori in output non tutti con lo stesso segno, bensì con un segno sia positivo che negativo.

Matematicamente, Tahn Function può essere rappresentata come (figura 4.14):

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1 - e^{-x}}{1 + e^{-x}} \ . \tag{4.15}$$

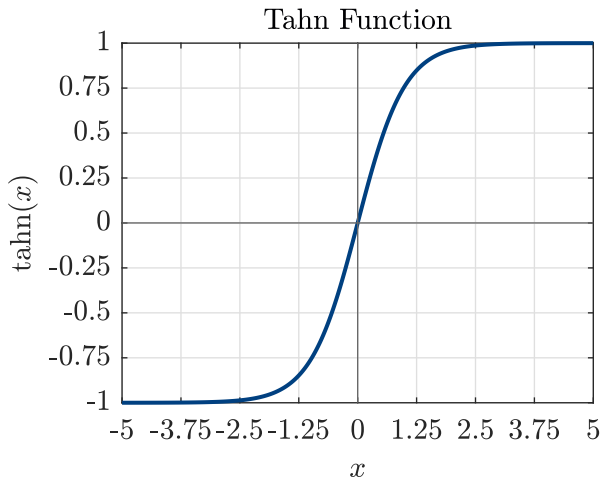


Figura 4.14. Tahn Function.

Anche Tahn Function, così come Sigmoid Function, è caratterizzata dalla presenza del medesimo fenomeno di *vanishing gradient problem*, con una caratteristica ulteriore: il gradiente della funzione risulta ancora più ripido.

Ad ogni modo, sebbene sia Sigmoid Function che Tahn Function presentino lo stesso problema, quest'ultima viene preferita nell'utilizzo generale, in considerazione del fatto che è centrata sullo zero e il gradiente non è limitato a muoversi in una specifica direzione.

Rectified Linear Unit Function

Rectified Linear Unit Function, abbreviata ReLU, è una funzione di attivazione non lineare, ampiamente utilizzata in numerose applicazioni diverse, essendo caratterizzata dalla possibilità di non attivare contemporaneamente tutti i neuroni della rete neurale. Infatti, con questa funzione di attivazione è possibile ottenere due risultati diversi: da una parte, attivare i neuroni, quando il risultato della trasformazione lineare risulta maggiore di zero, e, dall'altra, disattivare i neuroni, quando il risultato della trasformazione lineare risulta minore di zero.

Matematicamente, ReLU può essere rappresentata come (figura 4.15):

$$ReLU(x) = x^+ = \max(0, x) = \begin{cases} x & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} \ . \tag{4.16}$$

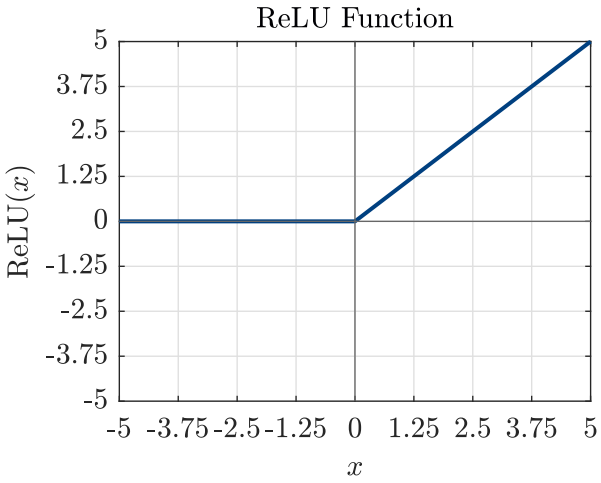


Figura 4.15. Rectified Linear Unit Function.

ReLU, considerando la sua natura, risulta computazionalmente più funzionale rispetto ad altre funzioni di attivazioni, come Sigmoid Function e Tahn Function, con la capacità di accelerare la convergenza del *gradient descent*, ovvero la tecnica di ottimizzazione utilizzata per identificare la direzione di modifica dei parametri, verso il minimo globale della funzione di perdita. Questa funzione rappresenta forse la funzione di attivazione più impiegata, ma presenta un limite: il *dying ReLU problem*.

Dal momento che il gradiente può assumere valore 0 per specifici valori, durante il processo di *error back-propagation*, i pesi e i bias di alcuni neuroni possono non venire mai aggiornati, generando neuroni "morti" che non vengono mai attivati.

In aggiunta, tutti i valori in input negativi vengono trasformati in 0, diminuendo l'abilità del modello di imparare correttamente i dati.

Leaky Rectified Linear Unit Function

Leaky Rectified Linear Unit Function, abbreviata LReLU, è una funzione di attivazione non lineare, che costituisce una versione leggermente migliorata di ReLU, nella quale, per valori negativi di x, invece che assegnare il valore 0, come nel caso precedente, si definisce una componente lineare di x estremamente piccola.

Matematicamente, LReLU può essere rappresentata come (figura 4.16):

$$LReLU(x) = \begin{cases} x & \text{se } x \geq 0 \\ 0{,}01x & \text{se } x < 0 \end{cases} . \tag{4.17}$$

Leaky ReLU presenta le stesse considerazioni effettuate per ReLU, con l'aggiunta di semplificare il processo di *error back-propagation* anche per valori di input negativi, superando il fenomeno del *dying ReLU problem*.

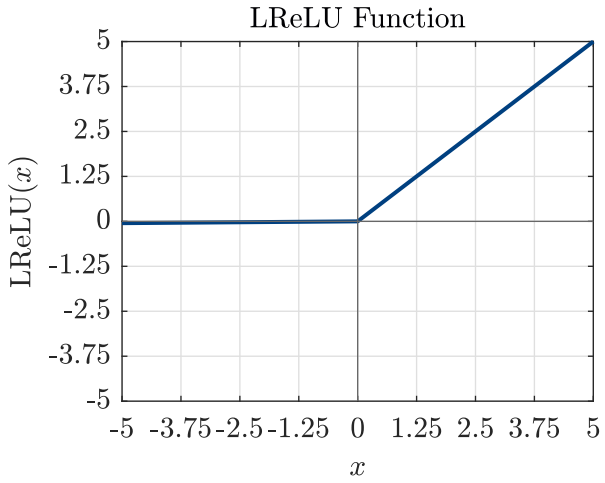


Figura 4.16. Leaky Rectified Linear Unit Function.

Parametric Rectified Linear Unit Function

Parametric Rectified Linear Unit Function, abbreviata PReLU, è una funzione di attivazione non lineare, che si impegna a risolvere il problema del gradiente nullo di ReLU per valori negativi di x, introducendo un nuovo parametro nella parte negativa della funzione.

Matematicamente, PReLU può essere rappresentata come (figura 4.17):

$$PReLU(x) = \begin{cases} x & \text{se } x \geq 0 \\ ax & \text{se } x < 0 \end{cases} . \tag{4.18}$$

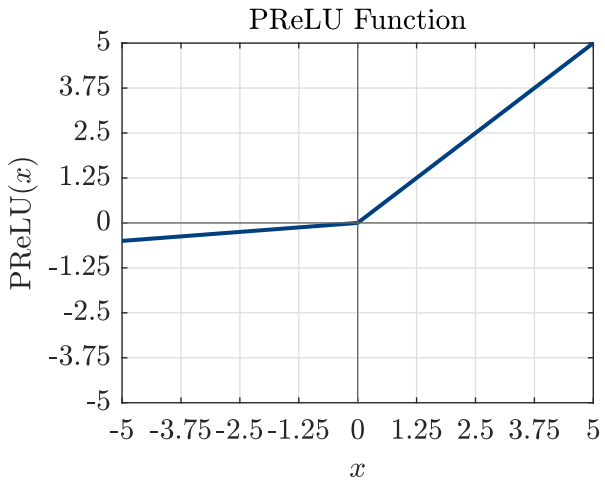


Figura 4.17. Parametric Rectified Linear Unit Function.

PReLU può essere considerata una generalizzazione di LReLU impostando $a = 0{,}01$ e viene generalmente impiegata quando LReLU continua a fallire per risolvere il fenomeno del *dying ReLU problem*, impedendo il passaggio agli strati successivi delle informazioni rilevanti. In questo caso, a risulta un parametro liberamente modificabile, anche se, per ottenere una convergenza ottimale, si è soliti assegnare il valore più appropriato tramite addestramento.

Exponential Linear Unit Function

Exponential Linear Unit Function, abbreviata ELU, è una funzione di attivazione non lineare, variante di ReLU, che modifica la pendenza della parte negativa della funzione, introducendo una curva logaritmica per definire i valori negativi di x, a differenza di LReLU e PReLU, che utilizzano rette, più o meno inclinate.

Matematicamente, ELU può essere rappresentata come (figura 4.18):

$$ELU(x) = \begin{cases} x & \text{se } x \geq 0 \\ \alpha(e^x - 1) & \text{se } x < 0 \end{cases} . \tag{4.19}$$

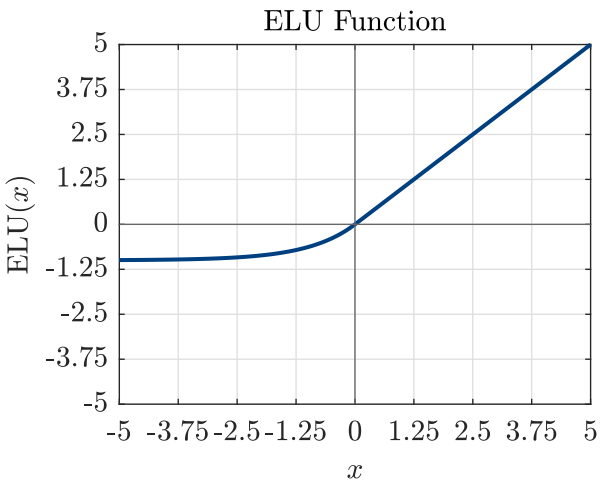


Figura 4.18. Exponential Linear Unit Function.

ELU rappresenta un'alternativa molto credibile a ReLU, poiché consente di evitare il fenomeno del *dying ReLU problem*, introducendo una curva logaritmica per valori negativi degli input e aiutando la rete neurale a spingere pesi e bias nella giusta direzione. Infatti, questa funzione di attivazione, a differenza di ReLU, che si appiattisce bruscamente, tende ad avere un andamento molto più morbido, fino al raggiungimento del valore $-\alpha$.

ELU presenta anche alcuni limiti: aumento del tempo computazionale, a causa delle operazioni matematiche esponenziali coinvolte, nessun addestramento per il valore di α e il fenomeno di *exploding gradient problem*, ovvero il problema dell'esplosione del gradiente, che può condurre ad aggiornamenti molto grandi per i valori dei pesi e dei bias del modello durante l'addestramento, rendendo instabile la rete neurale.

Gaussian Error Linear Unit

Gaussian Error Linear Unit Function, abbreviata GELU, è una funzione di attivazione non lineare in grado di combinare i benefici di diverse funzioni di attivazione e ottenere migliori prestazioni in specifici campi di applicazione.

Matematicamente, GELU può essere rappresentata come (figura 4.19):

$$GELU(x) = x \cdot \Phi(x) \ , \tag{4.20}$$

dove $\Phi(x)$ rappresenta la funzione di distribuzione cumulativa normale standard.

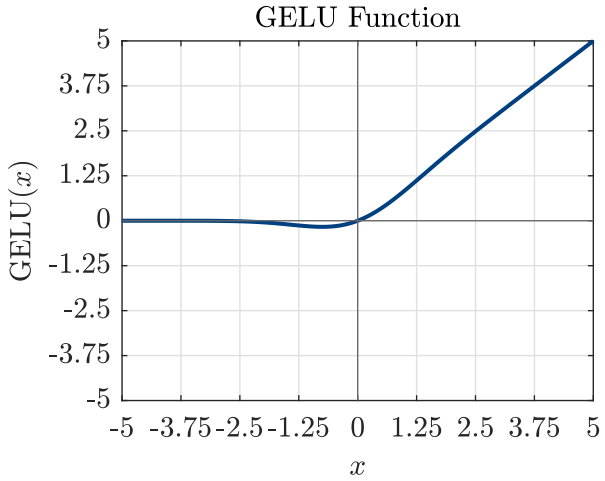


Figura 4.19. Gaussian Error Linear Unit Function.

Swish Function

Swish Function è una funzione di attivazione non lineare, che risulta non monotonica, implicando automaticamente il fatto che il valore della funzione possa diminuire anche all'aumentare dei valori degli input.

Matematicamente, Swish Function può essere rappresentata come (figura 4.20):

$$Swish(x) = \frac{x}{1 + e^{-x}} \ . \tag{4.21}$$

Swish Function è potenzialmente in grado di ottenere performance migliori di ReLU, soprattutto nel caso di reti neurali profonde addestrate su dataset con schemi interconnessi e articolati, dal momento che non cambia bruscamente direzione vicino allo zero, ma diminuisce lentamente da zero a valori minori di zero per poi aumentare nuovamente, in modo da catturare con più precisione le informazioni sottostanti i dati.

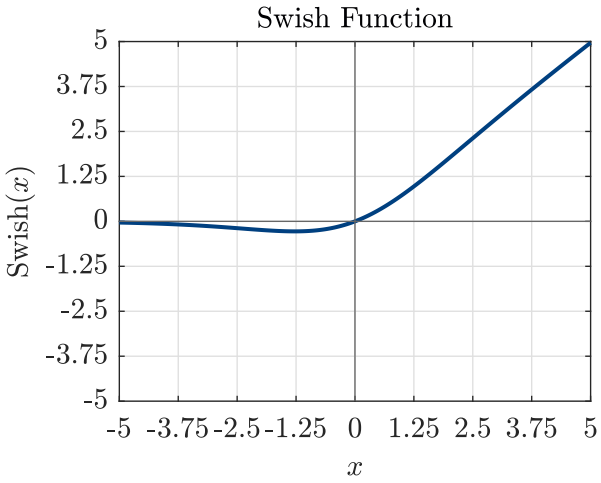


Figura 4.20. Swish Function.

Softmax Function

Softmax Function è una funzione di attivazione non lineare, costruita idealmente come combinazione di più Sigmoid Function, in grado di calcolare le probabilità relative di appartenere a una specifica classe.

Matematicamente, Softmax Function può essere rappresentata come (figura 4.21):

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \ . \tag{4.22}$$

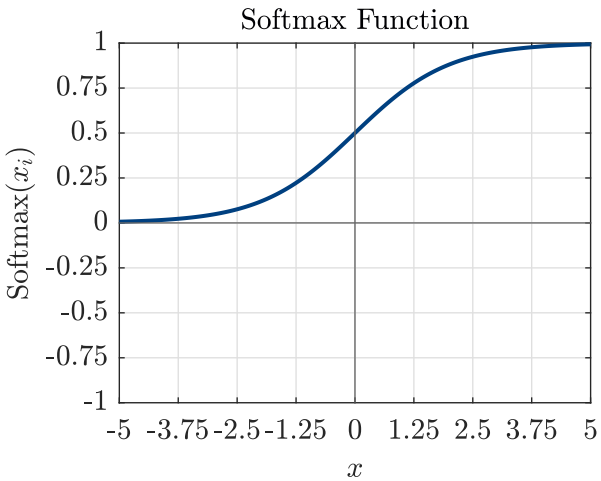


Figura 4.21. Softmax Function.

Softmax Function, a differenza della Sigmoid Function che viene impiegata per la classificazione binaria, viene comunemente utilizzata per problemi di classificazione multiclasse, ovvero per problemi con il compito di gestire più classi, ritornando un valore compreso tra 0 e 1, il quale può essere interpretato come la probabilità di appartenere a ogni classe.

4.5.4 Progettazione architettura

Con il termine architettura di una rete neurale artificiale si intende l'insieme delle scelte effettuate in fase di progettazione, le quali comprendono, in particolare, la costruzione dei livelli, la disposizione dei neuroni e l'organizzazione delle relazioni tra questi ultimi. Al variare di queste caratteristiche, è possibile distinguere diverse tipologie di reti neurali, ciascuna utilizzabile per raggiungere specifici obiettivi in diversi ambiti di applicazione:

- reti neurali feed-forward, o *Feed-Forward Neural Network*;
- reti neurali ricorrenti, o *Recurrent Neural Network*;
- reti neurali convoluzionali, o *Convolutional Neural Network*.

Feed-Forward Neural Network (ANN)

Le reti neurali feed-forward, definite *Feed-Forward Neural Network* (ANN), sono reti neurali composte da uno strato di input, uno o più strati nascosti e uno strato di output. In questa architettura il flusso informativo viaggia in una sola direzione, dal momento che non si osservano connessioni verso strati precedenti o all'interno dello stesso strato. Infatti, i neuroni di uno specifico strato possono solo ricevere l'input dai neuroni dello strato precedente e, parallelamente, possono solo trasmettere l'output ai neuroni dello strato successivo, come visibile in figura 4.22.

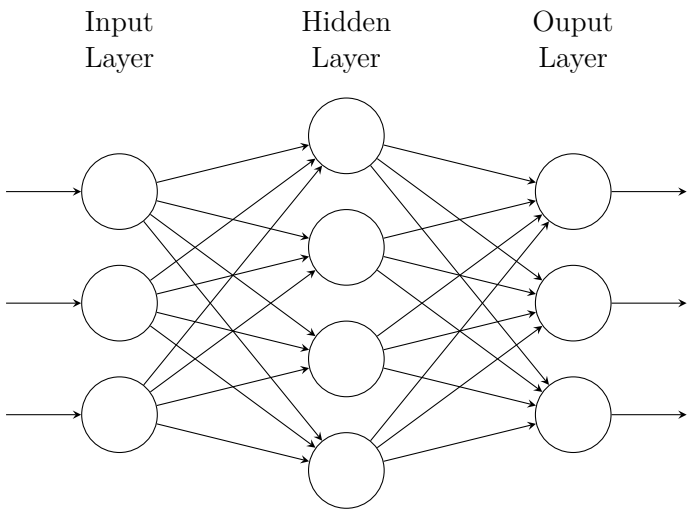


Figura 4.22. Feed-Forward Neural Network (ANN).

Recurrent Neural Network (RNN)

Le reti neurali ricorrenti, definite *Recurrent Neural Network* (RNN), sono reti neurali che si distinguono per la loro memoria, in quanto ricevono informazioni da input precedenti in grado di influenzare sia l'input che l'output correnti. A differenza delle reti neurali tradizionali, che presuppongono l'indipendenza tra input e output, le reti neurali ricorrenti producono un output che dipende dagli elementi precedenti all'interno della sequenza.

In poche parole, le reti neurali ricorrenti prevedono la presenza di connessioni di loop o feedback tra neuroni dello stesso livello o neuroni di livelli precedenti, che influenzano il modo in cui i neuroni prendono le decisioni, valutando non solo gli input correnti, ma anche quanto appreso dagli input ricevuti in precedenza, dal momento che la presenza di una sequenza rende questi ultimi in grado di ricordarsi lo stato del neurone precedente e passare le relative informazioni al loro corrispettivo nel futuro, con l'obiettivo di analizzare più correttamente i dati, come visibile in figura 4.23.

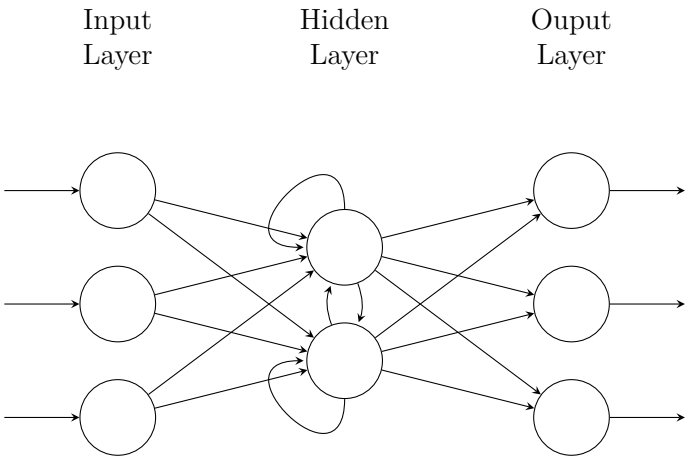


Figura 4.23. Recurrent Neural Network (RNN).

Questa tipologia di reti neurali processa dati sequenziali o in serie storica per cercare di individuare un qualche pattern al loro interno e, conseguentemente, risolvere problemi ordinali o temporali, come l'elaborazione del linguaggio naturale, lo svolgimento di analisi predittive, l'individuazione di anomalie, il riconoscimento vocale, il riconoscimento della grafia, la generazione di descrizioni da immagini e molto altro ancora.

Convolutional Neural Network (CNN)

Le reti neurali convoluzionali, o *Convolutional Neural Network* (CNN), sono reti neurali che sfruttano la disposizione e la localizzazione spaziale dei neuroni per ottenere i propri risultati. A questo proposito, le reti neurali convoluzionali sono composte da tre tipologie di strati: convolutional layer, che costituisce il primo strato, pooling layer, che rappresenta uno o più strati intermedi, e fully-connected layer, che costituisce l'ultimo strato.

In ogni strato, la rete neurale convoluzionale aumenta la propria complessità, identificando porzioni progressivamente maggiori di un'immagine per individuare e riconoscere uno o più oggetti. Nello specifico, i livelli iniziali, definiti low-level features extractor, si concentrano su semplici caratteristiche, come colore, bordi e forme elementari, ma, con la graduale e progressiva elaborazione dei dati estratti dalle immagini verso i livelli finali, definiti high-level features extractor, la rete neurale inizia a riconoscere elementi sempre più complessi, fino all'identificazione completa degli oggetti previsti, come visibile in figura 4.24.

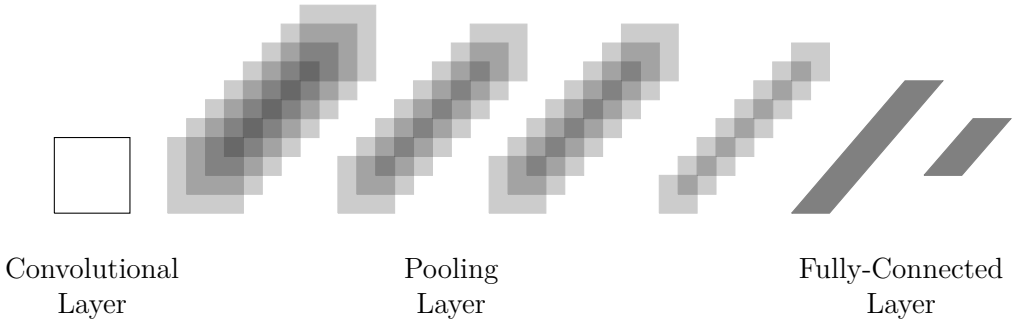


Figura 4.24. Convolutional Neural Network (CNN).

Questa tipologia di reti neurali viene tipicamente impiegata per il riconoscimento di immagini, con la possibilità di fornire un approccio scalabile per attività di classificazione e riconoscimento di oggetti, sfruttando i principi dell'algebra lineare, tra cui soprattutto la moltiplicazione di matrici, per identificare un qualche pattern all'interno di un'immagine, e, più in generale, per attività di *computer vision*, un campo dell'intelligenza artificiale che permette di sviluppare sistemi capaci di derivare informazioni significative da immagini digitali, video e altri tipi di input visivi e, sulla base di questi input, prendere decisioni.

4.5.5 Addestramento rete neurale

Le reti neurali mirano a riprodurre il meccanismo di apprendimento che risulta alla base della conoscenza e della memoria umana. Tuttavia, per raggiungere questo risultato, è necessario addestrare, o allenare, la rete neurale, in modo da renderla auspicabilmente in grado di riconoscere la funzione ignota che collega input e output del problema in esame. L'obiettivo dell'addestramento di una rete neurale risulta quindi abbastanza complesso, cercando di calibrare i parametri del modello, nel tentativo di ricercare i parametri più corretti per simulare le relazioni che esistono tra input e output.

Il processo di apprendimento di una rete neurale avviene attraverso un meccanismo di tipo iterativo, ovvero modificando iterativamente i valori dei coefficienti per cercare di ottenere risultati il più possibile vicini a quelli desiderati. Pertanto, addestrare una rete neurale consiste in un progressivo tentativo di identificare le corrette ponderazioni da assegnare alle variabili di input e alle sinapsi degli strati nascosti, per ottenere in output un risultato vicino a quello della funzione che si intende approssimare.

Per svolgere questo compito, impegnandosi a trovare una soluzione a quanto appena discusso, si utilizza l'algoritmo di retro-propagazione dell'errore, o *error back-propagation*.

Single-Layer Perceptron

Il percettrone, in inglese *perceptron*, è un modello di neurone artificiale sviluppato nel 1958 da F. Rosenblatt, ispirato sempre al lavoro precedente di W. S. McCulloch e W. H. Pitts, che si basa su una struttura a due soli livelli, uno di ingresso e uno di uscita, connessi tramite l'utilizzo di una funzione di attivazione lineare Step Function [111].

Una rete neurale *Single-Layer Perceptron* (SLP) rappresenta il modello più semplice di rete neurale e può essere addestrata tramite una semplice regola, chiamata *delta rule*, che prevede di modificare, nel caso in cui l'output desiderato y^d e l'output effettivo y^e non coincidano, il peso delle connessioni sinaptiche con la seguente formula:

$$w_i(t+1) = w_i(t) + \delta(y^d - y^e)x_i \ , \tag{4.23}$$

dove:

- $w_i(t+1)$ = peso della i-esima sinapsi al tempo $t+1$, dopo la modifica;
- $w_i(t)$ = peso della i-esima sinapsi al tempo t, prima della modifica;
- δ = coefficiente di apprendimento, o *learning rate*;
- x_i = valore risultante trasmesso dalla i-esima sinapsi.

Questo modello di rete neurale non risulta sufficientemente sofisticato per risolvere problemi complessi, in quanto non è in grado di gestire input non linearmente separabili. Per risolvere questo problema, è necessario introdurre un modello di rete neurale più completo, in grado di rimuovere e superare i limiti impostati dal *Single-Layer Perceptron*: il *Multi-Layer Perceptron*.

Multi-Layer Perceptron

Una rete neurale *Multi-Layer Perceptron* (MLP) è un modello di rete neurale complesso e generale, caratterizzato dalla presenza di uno o più strati nascosti e funzioni di attivazione non lineari. In questo caso, il processo di addestramento risulta decisamente più articolato.

Infatti, per l'addestramento di una rete neurale si è dovuto attendere fino al 1986 quando D. E. Rumelhart, G. E. Hinton e R. J. Williams proposero un algoritmo destinato a diventare di fondamentale importanza per l'allenamento di qualsiasi rete neurale, ovvero l'algoritmo di retro-propagazione dell'errore, o *error back-propagation* [114].

4.5.6 Algoritmo error back-propagation

Come spiegato precedentemente, una rete neurale moderna viene addestrata attraverso l'utilizzo dell'algoritmo di *error back-propagation*, il quale consente di approssimare la forma funzionale che collega tra loro le variabili del problema.

A questo punto, è possibile spiegare brevemente i principi alla base del funzionamento.

Per prima cosa, sfruttando un reticolo di funzioni elementari e adottando un set di coefficienti assolutamente casuali, si procede con il calcolo della rete neurale, ottenendo dei valori in output logicamente diversi se confrontati con i valori corretti.

In seguito, però, tramite ripetuti tentativi, chiamati sinteticamente epoche, si arriva a ridurre progressivamente l'errore commesso, calcolando dopo ogni iterazione una specifica funzione di perdita, definita *loss function*, e, conseguentemente, modificando ogni volta i coefficienti delle funzioni elementari sulla base dei risultati ottenuti.

Così facendo, dopo un numero sufficientemente elevato di iterazioni, si osserva per i valori in output un avvicinamento graduale verso i valori desiderati, attestando di fatto il corretto esito dell'addestramento della rete neurale in questione.

In definitiva, l'algoritmo di *error back-propagation* si compone di due fasi distinte, che avvengono temporalmente in sequenza, chiamate rispettivamente *forward phase*, in cui si procede con l'attivazione delle unità di input, la quale viene propagata in avanti attraverso le funzioni di attivazione, e *backward phase*, in cui si modifica il peso delle connessioni sinaptiche sfruttando alcune tecniche di ottimizzazione, con le quali l'errore delle unità di output viene propagato all'indietro.

Equazioni fondamentali

Formalmente, è possibile descrivere l'algoritmo di *error back-propagation* usando quattro equazioni fondamentali, in grado di descriverne completamente il relativo meccanismo di funzionamento. Per ottenere questo risultato, è necessario introdurre una notazione non ambigua per riferirsi agli elementi di base di una rete neurale:

- w^l_{jk} = peso della connessione che collega il k-esimo neurone nello strato $(l-1)$-esimo con il j-esimo neurone nello strato l-esimo.
- b^l_j = bias del j-esimo neurone nello strato l-esimo.
- a^l_j = attivazione del j-esimo neurone nello strato l-esimo.

Con questa notazione, è possibile mettere in relazione l'attivazione a^l_j del j-esimo neurone nello strato l-esimo con le attivazioni dello strato $(l-1)$-esimo, valutando una sommatoria calcolata su tutti i k neuroni dello strato $(l-1)$-esimo:

$$a^l_j = \sigma(\sum_k w^l_{jk} a^{l-1}_k + b^l_j) \ . \tag{4.24}$$

Inoltre, è possibile definire gli input pesati per la funzione di attivazione per il j-esimo neurone nello strato l-esimo:

$$z^l_j = \sum_k w^l_{jk} a^{l-1}_k + b^l_j \ . \tag{4.25}$$

Infine, è possibile mettere insieme le due equazioni precedenti:

$$a^l_j = \sigma(z^l_j) \ . \tag{4.26}$$

In questo contesto, riscrivendo le formule precedenti in modo più semplice e compatto utilizzando una formulazione matriciale, si ottiene:

$$a^l = \sigma(w^l a^{l-1} + b^l) \ . \tag{4.27}$$

$$z^l = w^l a^{l-1} + b^l \ . \tag{4.28}$$

$$a^l = \sigma(z^l) \ . \tag{4.29}$$

A questo punto, è quindi possibile approfondire l'algoritmo di *error back-propagation*. L'obiettivo ultimo di questo algoritmo è quello di capire come cambia la funzione di errore, modificando i pesi e i bias. Questo significa calcolare le derivate della funzione di errore: $\partial C/\partial w^l_{jk}$ e $\partial C/\partial b^l_j$.

Per ottenere questo risultato, è necessario introdurre una quantità intermedia δ^l_j, definita come l'errore del j-esimo neurone nello strato l-esimo:

$$\delta^l_j = \frac{\partial C}{\partial z^l_j} \ . \tag{4.30}$$

Di conseguenza, seguendo la stessa formulazione matriciale introdotta in precedenza, è possibile definire δ^l come il vettore degli errori associati allo strato l-esimo.

L'algoritmo di *error back-propagation* consente di calcolare δ^l per ogni strato e, in seguito, collegare questi errori con le quantità di interesse: $\partial C/\partial w^l_{jk}$ e $\partial C/\partial b^l_j$.

L'algoritmo di *error back-propagation* si basa su quattro equazioni fondamentali, le quali, combinate insieme, consentono di calcolare sia l'errore δ^l che il gradiente della funzione di errore.

La prima equazione rappresenta un'equazione per l'errore nell'output layer δ^L:

$$\delta^L_j = \frac{\partial C}{\partial a^L_j} \sigma'(z^L_j) \ . \tag{4.31}$$

Questa equazione può essere riscritta in una formulazione matriciale come segue[4]:

$$\delta^L = \nabla_a C \odot \sigma'(z^L) \ . \tag{4.32}$$

La seconda equazione costituisce un'equazione per l'errore δ^l in un generico strato, espresso in termini dell'errore nello strato successivo δ^{l+1}:

$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^L) \ . \tag{4.33}$$

Combinando le prime due equazioni è possibile determinare l'errore δ^l per qualsiasi strato della rete neurale: si inizia calcolando δ^L con la prima equazione e poi con la

[4] Il simbolo $\odot$ indica il prodotto di Hadamard, tale che $s \odot t = (s \odot t)_j = s_j t_j$.

seconda equazione si continua, calcolando δ^{L-1}, δ^{L-2}, δ^{L-3} e così via a ritroso per tutta la struttura della rete neurale, fino ad arrivare a δ^1.

La terza equazione rappresenta un'equazione per il tasso di variazione del costo rispetto a qualsiasi bias della rete neurale:

$$\frac{\partial C}{\partial b_j^l} = \delta_j^l \ . \tag{4.34}$$

La quarta equazione costituisce un'equazione per il tasso di variazione del costo rispetto a qualsiasi peso della rete neurale:

$$\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l \ . \tag{4.35}$$

Funzione di perdita

In questo contesto, per calibrare i valori dei diversi parametri, allenando la rete neurale, è fondamentale introdurre una metrica di controllo, con la quale valutare la correttezza dei risultati ottenuti e permettere, eventualmente, alla rete neurale di "auto-correggersi", con la finalità di migliorare le proprie performance.

In generale, una funzione di perdita[5], o *loss function*, è una funzione che consente di ottimizzare la scelta dei parametri di modello, confrontando, sistematicamente nel tempo, i valori target con i valori previsti, per misurare la capacità complessiva della rete neurale di modellare i dati in questione.

Durante la fase di addestramento, infatti, si cerca di ridurre al minimo questa differenza, aggiustando i possibili iperparametri in modo da trovare i pesi w_i e i bias b_i che minimizzano una perdita media J, definita attraverso una particolare funzione L, che dipende dai valori target y_i e dai valori previsti $\hat{y}_i$:

$$J(w, b) = \frac{1}{m} \sum_i^m L(y_i, \hat{y}_i) \ . \tag{4.36}$$

A questo riguardo, esistono molteplici funzioni di perdita, appositamente disegnate con lo scopo di rispondere a specifiche esigenze al variare del campo di applicazione:

- Mean Squared Error;
- Mean Absolute Error;
- Huber Loss;
- Binary Cross-Entropy;
- Categorical Cross-Entropy.

[5] In alternativa, la funzione di perdita viene anche chiamata funzione di errore o funzione di costo.

Mean Squared Error (MSE) è una funzione di perdita che calcola la media del quadrato delle differenze tra i valori target e i valori previsti:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \ . \tag{4.37}$$

Questa funzione di perdita presenta numerose proprietà che la rendono particolarmente adatta per il calcolo delle perdite. Analizzando più nel dettaglio le caratteristiche: da una parte, la funzione prevede il calcolo di una differenza al quadrato, non comportando alcuna differenza ai fini del computo della perdita se il valore previsto è maggiore o minore del valore target; dall'altra, la funzione risulta convessa, con un minimo globale chiaramente definito, consentendo una più semplice applicazione delle tecniche di ottimizzazione per il calcolo dei parametri di modello.

Tuttavia, un problema collegato a questo tipo di funzione di errore è rappresentato dal fatto che quest'ultima risulta estremamente sensibile agli outlier, penalizzando in modo eccessivo i valori con errori elevati, in quanto un valore previsto significativamente maggiore o minore del valore target, considerando che la distanza è al quadrato, produce una perdita decisamente elevata.

Mean Absolute Error (MAE) è una funzione di perdita che calcola la media del valore assoluto della differenza tra i valori target e i valori previsti:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \ . \tag{4.38}$$

Questa funzione di perdita viene utilizzata in specifiche applicazioni come alternativa alla funzione di perdita precedente, soprattutto nel caso in cui i dati in esame presentino un elevato numero di outlier, in modo da mitigarne il conseguente problema di sensibilità. Purtroppo, con questo tipo di funzione di errore, quando la distanza si avvicina a zero, le tecniche di ottimizzazione non lavorano correttamente, poiché la derivata della funzione non è definita in questo punto, restituendo un errore.

Huber Loss è una funzione di perdita introdotta da P. J. Huber nel 1964 e sviluppata per incorporare contemporaneamente i vantaggi di MSE e MAE:

$$HL = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 & \text{se } (y_i - \hat{y}_i) \leq \delta \\ \frac{1}{n}\sum_{i=1}^{n}\delta(y_i - \hat{y}_i - \frac{1}{2}\delta) & \text{se } (y_i - \hat{y}_i) > \delta \end{cases} \ . \tag{4.39}$$

Questa funzione di perdita consente di valutare l'entità dell'errore e, conseguentemente, calcolare equazioni diverse al variare del risultato, utilizzando un parametro δ selezionabile liberamente, in modo da superare i punti deboli e, allo stesso tempo, sfruttare i punti forti presentati da MSE e MAE.

In sostanza, quando l'errore è relativamente piccolo, ovvero se il valore assoluto della differenza tra il valore target e il valore previsto è minore o uguale a un valore soglia δ, si applica MSE, mentre, quando l'errore è relativamente grande, ovvero se il valore assoluto della differenza tra il valore target e il valore previsto è maggiore di un valore soglia δ, si applica MAE.

Binary Cross-Entropy Loss, chiamata anche Log Loss, viene usata nei problemi di classificazione binaria, in cui il modello riceve un input con il compito di classificarlo in una delle due categorie di interesse:

$$BCEL = -\frac{1}{n}\sum_{i=1}^{N}[y_i \cdot log(p_i) + (1 - y_i) \cdot log(1 - p_i)] \ . \tag{4.40}$$

Le reti neurali di classificazione funzionano emettendo un vettore di probabilità, che contiene la probabilità che l'input in esame rientri nelle categorie presenti, in modo da selezionare come output la categoria con la probabilità più elevata.

Nei problemi di classificazione binaria, l'output può assumere due soli valori, 0 e 1. Per determinare la perdita tra i valori target e i valori previsti, è necessario confrontare il valore effettivo, 0 e 1, con la probabilità che l'input sia allineato con la categoria scelta, ossia p_i e $(1 - p_i)$, che indicano rispettivamente la probabilità che la categoria sia 1 e 0.

Categorical Cross-Entropy Loss viene usata nei problemi di classificazione multiclasse, nel caso in cui il numero di classi sia maggiore di due:

$$CCEL = -\frac{1}{n}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} \cdot log(p_{ij}) \ . \tag{4.41}$$

Ricapitolando, l'algoritmo di *error back-propagation* consente di calcolare l'errore commesso in ogni iterazione, correggendo i parametri per avvicinarsi progressivamente alla soluzione del problema. In questo ambito, per individuare la direzione di modifica dei parametri, si utilizza una tecnica di ottimizzazione, chiamata *gradient descent*.

4.5.7 Metodo gradient descent

Il *gradient descent*, tradotto discesa del gradiente, è una tecnica di ottimizzazione che consente di calcolare la direzione ottimale per modificare i parametri del problema e convergere nel modo più velocemente possibile verso la soluzione cercata. L'obiettivo è quello di modificare i parametri w_i e b_i, in modo da minimizzare la funzione di perdita.

A questo proposito, la tecnica del *gradient descent* indica la direzione ottimale: per arrivare alla soluzione, si sceglie di modificare i parametri in direzione opposta al gradiente. Infatti, dal momento che, in analisi matematica, il gradiente di una funzione di più variabili rappresenta la direzione di maggior crescita della funzione, scegliendo la direzione opposta al gradiente, si ottiene automaticamente la direzione di maggior decrescita della funzione.

In poche parole, per minimizzare la funzione di perdita, si sceglie di effettuare passi iterativi in direzione opposta al gradiente, come mostrato graficamente in figura 4.25.

In ultima istanza, per completare il ragionamento, è possibile aggiungere una precisazione ulteriore, distinguendo tre tipologie di *gradient descent*, che si differenziano in base alla tipologia di processo seguito, pur mantenendo la stessa logica sottostante:

- Batch Gradient Descent (BGD);
- Stochastic Gradient Descent (SGD);
- Mini-Batch Gradient Descent (MBGD).

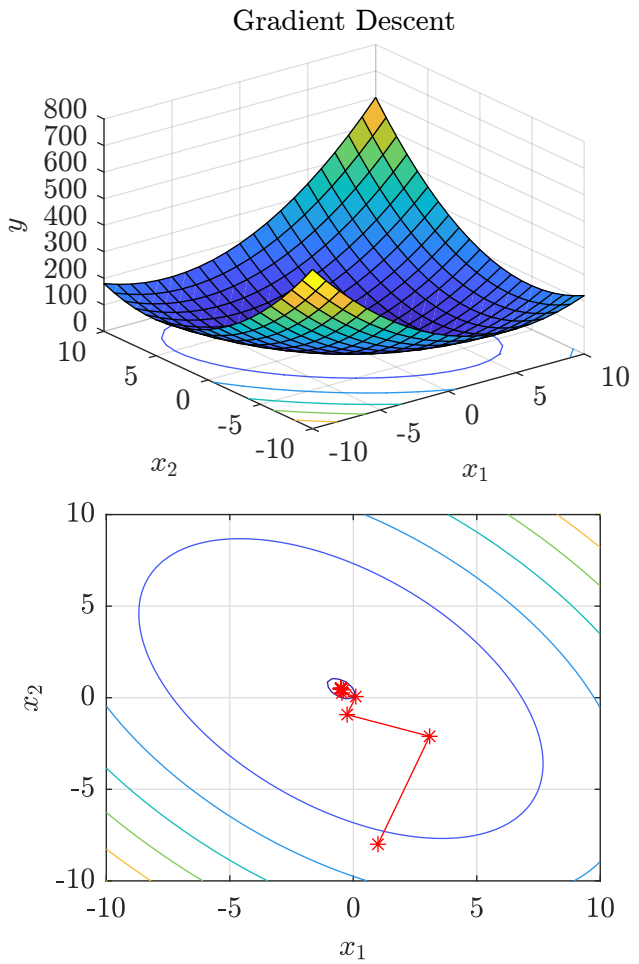


Figura 4.25. Esempio applicazione gradient descent $y = x^2 + y^2 + x - y + 2xy$.

4.5.8 Caratteristiche principali

Le reti neurali cercano di riprodurre il meccanismo di apprendimento presente all'interno del cervello umano, riuscendo a gestire problemi non immediatamente riconducibili a un normale algoritmo di calcolo, per quanto semplice o complesso possa essere, in modo da ottenere numerosi vantaggi, ma anche qualche svantaggio (tabella 4.2).

I principali vantaggi sono:

- Performance elevate: le reti neurali costituiscono un modello di intelligenza artificiale in grado di ottenere performance all'avanguardia e allo stato dell'arte, arrivando a rappresentare il punto di riferimento nella maggior parte delle applicazioni.
- Pattern complessi: le reti neurali possono imparare anche pattern caratterizzati da grande complessità, lineari e non lineari, gestendo un numero elevato di input diversi come immagini, video, suoni e testo.

- Grande flessibilità: le reti neurali possono essere impiegate per risolvere problemi sia di regressione che di classificazione, essendo possibile modificare l'architettura in base alle esigenze per adattarsi alle diverse applicazioni.
- Generalizzazione: le reti neurali riescono a generalizzare molto bene nuovi dati, evitando il fenomeno dell'overfitting.
- Aggiornabilità: le reti neurali risultano facilmente aggiornabili con nuovi dati.

I principali svantaggi sono:

- Modello black-box: le reti neurali rappresentano un modello di intelligenza artificiale estremamente complesso, la cui soluzione finale non risulta facilmente osservabile e analizzabile, in quanto non è possibile rispondere a domande sul come e sul perché si sia raggiunta una specifica soluzione.
- Determinazione architettura: le reti neurali non presentano una regola universalmente valida per determinare la struttura ideale, implicando l'esigenza di procedere con l'esperienza acquisita e con operazioni euristiche di *trial and error* e *thumb rule*.
- Traduzione problema: le reti neurali possono lavorare con informazioni numeriche, obbligando sostanzialmente a tradurre il problema di interesse in valori numerici prima di processarlo.
- Lavorazione dati: le reti neurali necessitano di un processo di preparazione dei dati e, nello specifico, un processo di *scaling* delle variabili, che comprende operazioni di normalizzazione o standardizzazione, per evitare possibili distorsioni causate da eventuali differenze di scala.
- Addestramento complesso: le reti neurali risultano computazionalmente onerose, sia per quanto riguarda il tempo che il costo.

Tabella 4.2. Caratteristiche principali reti neurali.

Reti neurali	
Vantaggi	Svantaggi
Performance elevate	Modello black-box
Pattern complessi	Determinazione architettura
Grande flessibilità	Traduzione problema
Generalizzazione	Lavorazione dati
Aggiornabilità	Addestramento complesso

Parte III

Terza Parte

Capitolo 5

Analisi industria automotive

5.1 Introduzione

Per applicare i modelli di credit scoring descritti precedentemente è necessario disporre di un database quanto più strutturato possibile, in modo da avere a disposizione informazioni in misura sufficiente, sia dal punto di vista quantitativo che dal punto di vista qualitativo.

In questo contesto, si è scelto di sviluppare modelli di intelligenza artificiale per l'analisi del rischio di credito, prendendo come ambiente di riferimento l'Italia e, nello specifico, un settore industriale italiano: il settore automotive, individuato dal codice ATECO 29.

Il capitolo si apre con un'analisi del settore industriale prescelto, in modo da descriverne brevemente le principali caratteristiche, unitamente agli ultimi andamenti, prosegue con un focus approfondito sul contesto italiano, autentico obiettivo del presente lavoro di tesi, e si chiude con la descrizione delle operazioni necessarie per ottenere il database in esame.

5.2 Settore automotive

Il settore automotive è un settore di business appartenente all'industria manifatturiera che si occupa di progettazione, sviluppo, costruzione, marketing e vendita di veicoli a motore. Nello specifico, il settore automotive è contrassegnato dal codice ATECO 29, che in generale descrive la "fabbricazione di autoveicoli, rimorchi e semirimorchi", ma che più precisamente comprende la produzione di autoveicoli (codice ATECO 29.1), la produzione di carrozzerie per autoveicoli, rimorchi e semirimorchi (codice ATECO 29.2) e, infine, la produzione di parti e accessori per autoveicoli e loro motori (codice ATECO 29.3)[1], come riassunto nella tabella 5.1.

Questo settore industriale ricopre un'importanza strategica fondamentale per l'Italia, la quale, presentando da sempre una filiera automotive di assoluta eccellenza internazionale, rappresenta uno dei maggiori Paesi produttori, con la presenza di alcuni tra i brand più storici e prestigiosi a livello mondiale.

[1] Il codice ATECO 29 e, più precisamente, il codice ATECO 29.3 non include pneumatici.

Tabella 5.1. Descrizione settore automotive (codice ATECO 29).

Codice ATECO	Descrizione attività
29	Fabbricazione di autoveicoli, rimorchi e semirimorchi
29.1	Fabbricazione di autoveicoli
29.2	Fabbricazione di carrozzerie per autoveicoli, rimorchi e semirimorchi
29.3	Fabbricazione di parti e accessori per autoveicoli e loro motori

5.2.1 Scenario internazionale

Il settore automotive rappresenta uno dei principali settori di business a livello globale. Secondo gli ultimi dati a disposizione[2], nel 2021, la produzione mondiale di veicoli a motore ha superato il valore di 80 milioni di unità. Analizzando più nel dettaglio questo numero: il 32,54% viene prodotto in Cina, l'11,44% negli Stati Uniti, l'1,16% nel Regno Unito, il 9,79% in Giappone e il 15,17% nell'Unione Europea, dove Germania, Francia, Italia e Spagna rappresentano tradizionalmente i principali Paesi produttori, sia per numero di impianti di produzione che per numero di vetture prodotte, come visibile in figura 5.1.

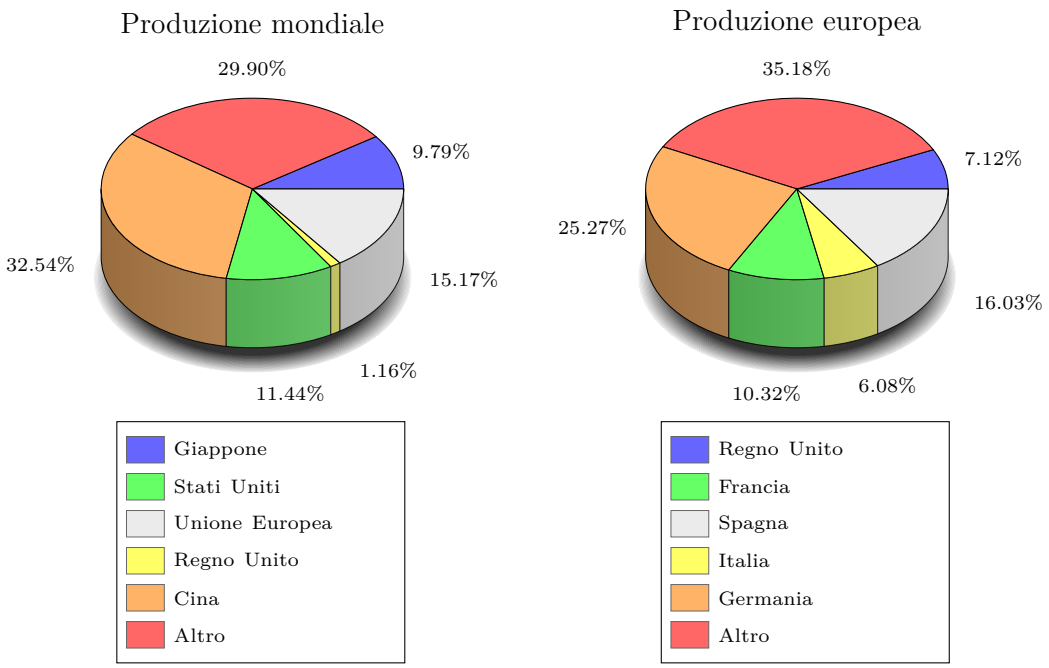


Figura 5.1. Produzione mondiale ed europea veicoli a motore.

[2]Dati ricavati da OICA (International Organization of Motor Vehicle Manufacturers) [154].

Insomma, è possibile sottolineare come un numero ridotto di Paesi internazionali contribuiscano, complessivamente, a generare oltre il 70% della produzione mondiale, mentre il restante 30% è suddiviso, in modo più o meno equilibrato, nel resto del mondo.

A questo proposito, nei mercati internazionali, l'industria automobilistica si caratterizza per una significativa impronta regionale, soprattutto in relazione alle catene di fornitura che collegano saldamente materie prime, semilavorati e prodotti finiti. In questo contesto, Stati Uniti, Germania, Giappone e, più recentemente, Cina costituiscono il nucleo attorno al quale gravitano rispettivamente i mercati dell'America, dell'Europa e dell'Asia.

Pertanto, è possibile osservare come il settore automotive sia fortemente concentrato, con una tendenza che sembra addirittura destinata a rinforzarsi nel prossimo futuro, in considerazione del fatto che, la filiera automotive costituisce una delle filiere industriali più complesse, con un veicolo composto mediamente da oltre 20000 componenti, prodotte da diversi fornitori internazionali in diversi stabilimenti produttivi.

Per ricostruire l'andamento del settore automobilistico negli ultimi anni, con l'obiettivo di disegnare una panoramica quanto più approfondita possibile del comparto di interesse, è necessario iniziare parlando della crisi finanziaria del 2007-2009, che ha causato, anche in questo ambito, pesanti ripercussioni lungo tutta la catena del valore, determinando un crollo del mercato con una perdita totale di circa il 15% e, conseguentemente, un faticoso ritorno in positivo soltanto dopo due anni, nel 2010, come mostrato in figura 5.2.

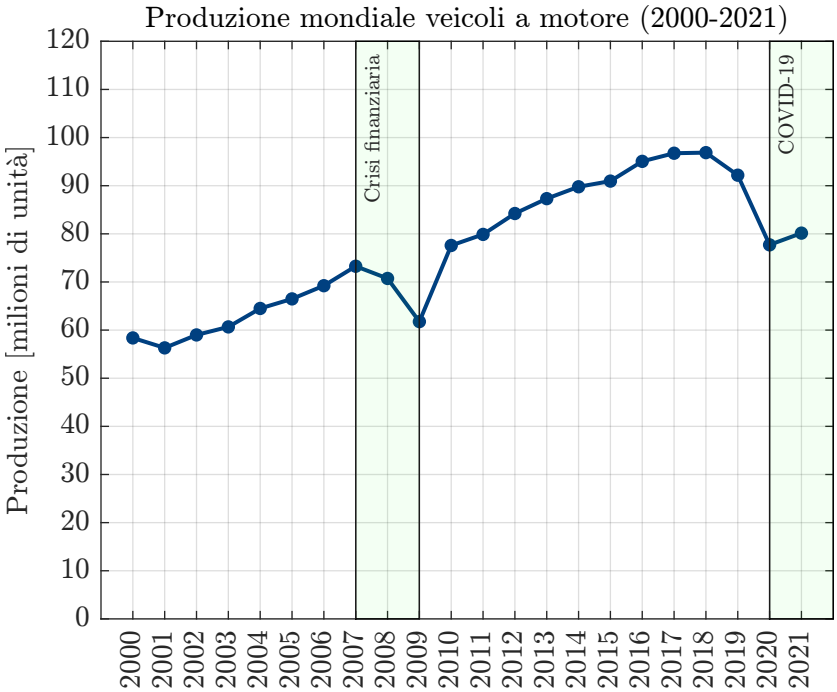


Figura 5.2. Andamento produzione mondiale veicoli a motore.

In seguito, però, il settore automobilistico ha attraversato un periodo molto positivo. Dopo la crisi finanziaria del 2007-2009, infatti, il settore automotive ha sperimentato

a livello globale una fase di grande espansione, facendo registrare un incremento della produzione con un tasso di crescita medio annuo relativamente elevato, pari al 6%.

Nel biennio 2018-2019, al contrario, il settore automotive ha inaugurato un nuovo periodo negativo, caratterizzato da una congiuntura sfavorevole che ha provocato una contrazione della produzione superiore al 4%, giustificabile con alcuni eventi che hanno progressivamente invaso il mercato, come il declino delle motorizzazioni diesel, l'emersione del segmento di veicoli ibridi/elettrici e, infine, la diffusione di nuovi modelli di consumo incentrati sul concetto di auto non più come prodotto, bensì come servizio.

In questa prospettiva, non esattamente radiosa, si è aggiunta la crisi causata dalla diffusione della pandemia da COVID-19, che ha contribuito ad aggravare ulteriormente il momento di difficoltà, determinando una chiusura prolungata degli stabilimenti produttivi dei principali Paesi produttori, iniziata nel 2020 e continuata con effetti differenti, come logica conseguenza di tempi disomogenei nella diffusione dei contagi, nei singoli territori nazionali.

In poche parole, è fondamentale evidenziare come l'emergenza sanitaria si sia innestata in una fase di stagnazione del settore automotive, rappresentata da una serie di profondi cambiamenti, in grado di generare a livello globale significative trasformazioni nella filiera industriale, concentrata su enormi investimenti in ricerca e sviluppo, per la creazione, da una parte, di motorizzazioni tradizionali a minor impatto ambientale grazie alla riduzione di emissioni inquinanti, e, dall'altra, di batterie elettriche sempre più efficaci ed efficienti.

In conclusione, è doveroso dedicare un commento a quanto accaduto negli ultimi anni. La pandemia da COVID-19, evento straordinario e imprevedibile, ha determinato un pesante rallentamento dell'economia globale, con effetti rilevanti nella maggior parte dei settori di business. L'industria automotive non ha certamente rappresentato un'eccezione.

Sebbene le stime iniziali evidenziassero, per il 2020, un crollo della produzione superiore al 20% e un orizzonte temporale di circa tre anni per recuperare i livelli di produzione pre-crisi, le conseguenze osservabili sono state solo leggermente inferiori rispetto alle attese.

In realtà, non è ancora possibile produrre delle conclusioni significative, essendo ancora completamente immersi nella pandemia da COVID-19, seppur nel progressivo tentativo di ritorno alla normalità e di ripartenza per l'economia globale; tuttavia, sfruttando gli ultimi dati a disposizione, è possibile ottenere alcune informazioni interessanti, confrontando il valore della produzione del 2021 con il valore della produzione nei due anni precedenti: il 2019, ultimo anno pre-crisi, e il 2020, caratterizzato, appunto, dall'impatto della crisi sanitaria.

A questo proposito, è facile osservare come la produzione globale del settore automotive sia drasticamente crollata negli ultimi anni, cominciando da un valore di partenza di oltre 92 milioni di unità nel 2019, transitando per un valore di 78 milioni di unità, nel 2020, con una riduzione di quasi il 16%, e arrivando a un valore di 80 milioni di unità, nel 2021, con un incremento di circa il 3%.

In ogni caso, è possibile affermare come l'industria automotive sia ancora molto lontana da recuperare il valore della produzione pre-crisi, dal momento che, considerando i valori della produzione 2019 e 2021, esiste comunque una differenza complessiva di oltre il 13%.

Una volta completata una panoramica completa, anche se estremamente sintetica, del settore automotive internazionale, è possibile concentrare l'attenzione in modo particolare sul comparto italiano, che rappresenta il dominio di interesse, oggetto della trattazione.

5.2.2 Scenario italiano

In Italia, il settore automotive costituisce un settore di business notevolmente importante per tutta l'economia nazionale, considerando che l'intera filiera produttiva automotive comprende, complessivamente, tra attività dirette e attività indirette, oltre 5500 imprese e circa 278000 addetti, generando un fatturato di 106 miliardi di euro, che rappresenta l'11% del fatturato dell'industria manifatturiera e il 6,2% del PIL[3].

La filiera automotive italiana, nei mercati internazionali, riveste un ruolo di assoluta protagonista, posizionandosi nei segmenti a più elevato valore aggiunto, sia tramite il contributo della produzione di autoveicoli di alto livello e autoveicoli commerciali che tramite il contributo delle specializzazioni produttive dei distretti della componentistica.

La competitività del settore automotive italiano è stata per molto tempo superiore rispetto a quella dell'industria manifatturiera in generale, anche se, purtroppo, negli ultimi anni è facile osservare un andamento negativo, ulteriormente aggravato in considerazione del fatto che questo settore risulta uno dei settori più colpiti dalla pandemia da COVID-19.

L'emergenza sanitaria ha comportato la chiusura di numerosi stabilimenti produttivi, generando pesanti ripercussioni sul piano operativo, economico-finanziario e logistico per la maggior parte dei player internazionali e, parallelamente, anche per le imprese italiane. Considerando la forte dipendenza sia in termini di capacità produttiva che di importazioni di componentistica, il risultato è stato un vero e proprio shock nella supply chain del settore, con effetti, ancora una volta, sia a livello nazionale che internazionale.

A questo si è aggiunto il progressivo esaurimento delle scorte di materie prime, semilavorati e altri elementi o componenti essenziali per la continuità operativa e produttiva degli *Original Equipment Manufacturer* (OEM), che ha contribuito in modo sempre più oneroso ad alimentare ulteriormente il periodo di difficoltà.

Ma oltre al blocco della produzione, che si è tradotto in un problema sul lato dell'offerta, il settore è stato colpito anche dal lato della domanda: moltissimi consumatori, infatti, sia famiglie che imprese, di fronte all'incertezza del periodo in questione, hanno preferito scegliere di rimandare o annullare l'acquisto di un nuovo veicolo a motore.

In questo contesto, tra i Paesi europei, l'Italia rappresenta uno dei Paesi che ha subito con maggiore intensità l'impatto della crisi causata dalla pandemia da COVID-19, ma, nonostante questa flessione senza precedenti, le statistiche illustrate in precedenza dimostrano la rinnovata importanza di questa industria come "motore" che continua a spingere l'economia dell'Italia, indipendentemente da quanto possa succedere.

5.2.3 Database AIDA e ORBIS (Bureau Van Dijk)

Per la raccolta delle informazioni per l'analisi del rischio di credito nel settore automotive, si è scelto di utilizzare i database AIDA e ORBIS, grazie a una licenza messa a disposizione dal Politecnico di Torino. In questo modo, è stato possibile ricavare i dati di bilancio delle imprese appartenenti all'industria automotive italiana e, conseguentemente, costruire un database di dimensioni opportune per le analisi successive.

[3]Dati ricavati da ANFIA (Associazione Nazionale Filiera Industria Automobilistica) [134].

I database AIDA e ORBIS appartengono all'impresa Bureau Van Dijk, recentemente acquisita da Moody's e integrata in Moody's Analitics, che consente la ricerca, la consultazione, l'analisi e l'elaborazione delle principali informazioni anagrafiche, commerciali e, soprattutto, economico-finanziarie, rispettivamente di tutte le società di capitali che operano in Italia e di tutte le società quotate o non quotate che operano nel mondo [139].

In questo ambito, AIDA e ORBIS consentono di svolgere in modo efficace ed efficiente una serie di attività di ricerca, poiché, mettendo a disposizione informazioni difficilmente reperibili, rappresentano uno strumento molto potente per svolgere analisi avanzate.

In questo lavoro, si è deciso di utilizzare il database AIDA, sebbene ORBIS rappresenti un database più ampio, contenendo osservazioni di imprese provenienti da tutto il mondo, dal momento che è stato ritenuto, dopo accurate e profonde riflessioni, maggiormente idoneo per svolgere le analisi di interesse in riferimento al mercato italiano.

Quest'ultimo, consentendo l'analisi informatizzata delle aziende italiane, contiene al suo interno, per ciascuna società, numerose informazioni utili, come la scheda anagrafica e commerciale completa, la serie storica fino a 10 anni del bilancio d'esercizio redatto secondo lo schema proposto dalla IV direttiva CEE, oltre che il relativo bilancio ottico.

5.2.4 Download dei dati di bilancio

Con l'accesso al database AIDA, si è potuto effettuare il download dei dati di bilancio di tutte le imprese italiane appartenenti al settore automotive, scegliendo la massima dimensione temporale disponibile, ovvero un orizzonte temporale che comprende la serie storica degli ultimi 10 anni, dal 2012 al 2021.

Per poter scaricare i dati di bilancio in questione dal database AIDA, si è eseguito una ricerca per "Classificazione merceologica" nella categoria "Attività" (figura 5.3).

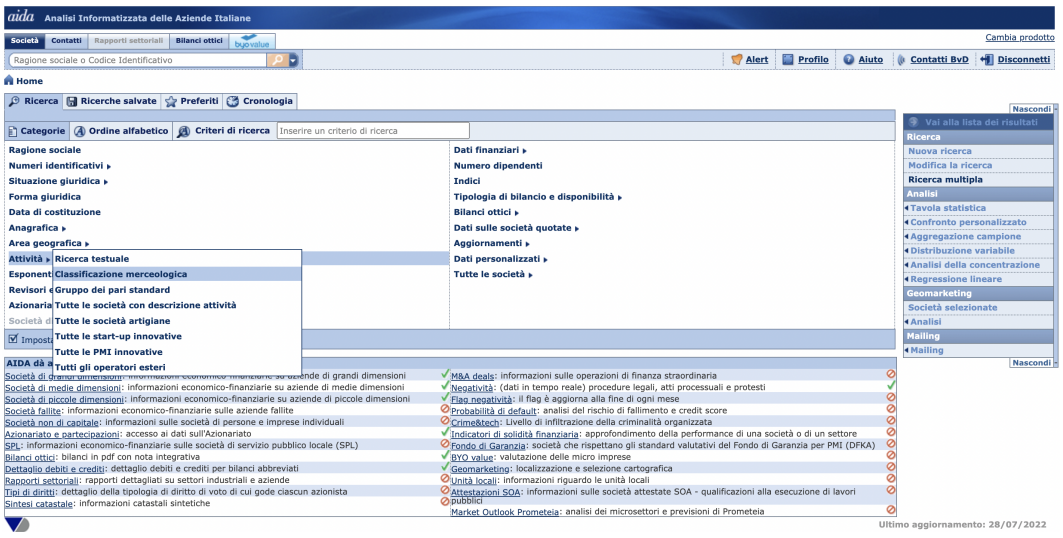


Figura 5.3. Database AIDA.

A questo punto, si è selezionato il settore di attività, con il codice ATECO 29, che individua le imprese appartenenti al settore automotive, ottenendo un campione di analisi composto da 3100 imprese.

In ultima analisi, si è impostato il set di informazioni di interesse, che comprende le informazioni anagrafiche, le informazioni commerciali e le informazioni di bilancio, quali Stato Patrimoniale, Conto Economico e Rendiconto Finanziario, oltre che i principali indicatori di bilancio associati, effettuandone il download, in relazione a tutte le imprese del campione, per ciascun anno dell'orizzonte temporale in esame, 2012-2021.

La scelta di lavorare con il maggior numero di dati di bilancio reperibili è stata eseguita con l'obiettivo di rilevare quanti più possibili eventi anomali, ovvero eventi di default, in modo da ottenere un campione di imprese sufficientemente adeguato alla costruzione di modelli significativi.

5.2.5 Analisi dei dati di bilancio

Prima di procedere alla definizione del modello di credit scoring, è interessante effettuare una breve analisi preliminare dei dati di bilancio delle imprese del settore automotive italiano, con l'obiettivo di approfondire i principali andamenti del campione estratto dal database AIDA, oggetto dei successivi studi. Così facendo, è possibile ottenere non solo una prima osservazione sul campione di imprese, utile per ricavare informazioni di base del problema, ma anche una descrizione più dettagliata della situazione dell'industria automotive italiana, in modo da completare il ragionamento iniziato precedentemente.

Per raggiungere questo risultato, si è deciso di aggregare i dati di bilancio delle imprese costituenti il campione, in modo da calcolare le statistiche descrittive più importanti e, conseguentemente, analizzare l'industria automotive italiana, nell'orizzonte temporale di riferimento, 2012-2021. Nello specifico, si è scelto di utilizzare come statistica descrittiva la mediana, ritenuta, in questa applicazione, la soluzione migliore possibile per descrivere il centro di un insieme di dati, dal momento che la media risulta influenzata in maniera sensibile da valori eccezionalmente alti o bassi, definiti outlier, e, quindi, non affidabile.

Infatti, analizzando le caratteristiche del database in esame, la mediana ha permesso di superare questi valori estremi e studiare la distribuzione dei vari indicatori di bilancio, trascurandone la parte esterna e concentrandosi sulla parte centrale, in modo da restituire risultati molto più compatibili con l'intento originale.

A tale proposito, si è calcolato una serie di indici di bilancio utili per rappresentare la condizione economico-finanziaria di un'impresa, eseguendo un'analisi di bilancio completa, strutturata secondo quattro dimensioni principali:

- analisi della solidità patrimoniale;
- analisi della struttura finanziaria;
- analisi della liquidità;
- analisi della redditività.

Analisi della solidità patrimoniale

L'analisi della solidità patrimoniale ha come obiettivo quello di valutare la capacità di un'impresa di mantenere nel tempo una situazione di equilibrio strutturale, analizzando la struttura di finanziamenti e investimenti, unitamente al livello di indipendenza dai terzi.

Per valutare la solidità patrimoniale delle imprese del settore automotive, si è scelto di calcolare due indici di bilancio principali:

- Margine Primario di Struttura (MS);
- Margine Secondario di Struttura (KCN).

Il Margine Primario di Struttura è calcolato come la differenza tra il patrimonio netto e le attività immobilizzate nette, e consente di valutare quanto l'impresa sia in grado di coprire i propri asset immobilizzati, ovvero gli asset di medio-lungo periodo, con il proprio patrimonio netto. In poche parole, questo indicatore permette di valutare la capacità di un'impresa di coprire gli investimenti in strutture fisse e immobilizzate con i mezzi propri.

$$MS = KN - AIN \ . \tag{5.1}$$

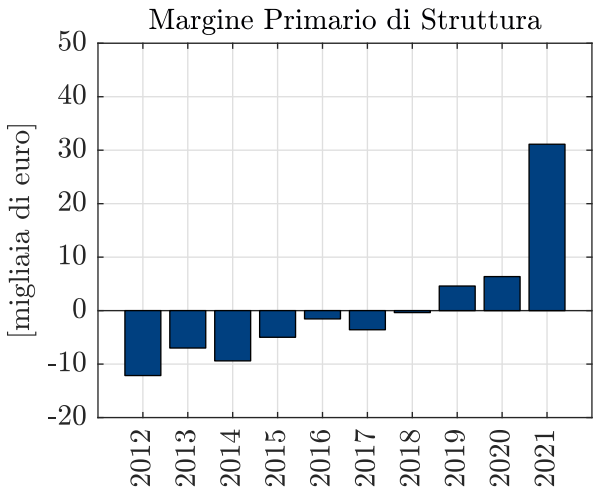


Figura 5.4. Margine Primario di Struttura.

Il Margine Secondario di Struttura, detto anche Capitale Circolante Netto, è ricavato come la differenza tra le passività permanenti nette e le attività immobilizzate nette. Con questo indicatore, è possibile valutare la capacità di un'impresa di finanziare le proprie attività immobilizzate attraverso le fonti durevoli a medio-lungo termine, ottenendo una valutazione della solidità patrimoniale dell'impresa meno stringente e restrittiva rispetto all'indicatore precedente, poiché adesso si considerano anche altre fonti di finanziamento, oltre al patrimonio netto.

$$KCN = PPN - AIN \ . \tag{5.2}$$

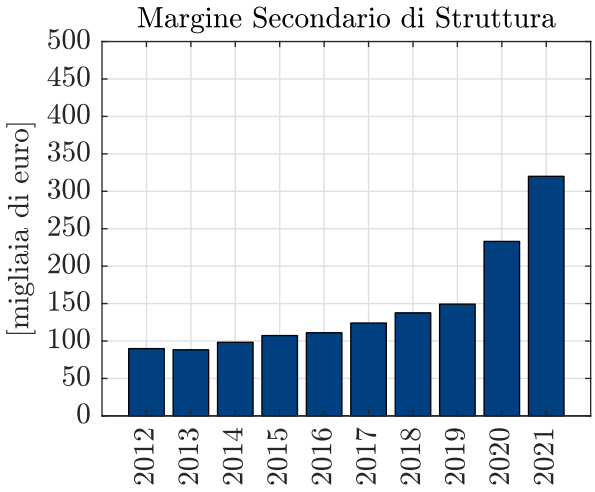


Figura 5.5. Margine Secondario di Struttura.

Osservando i grafici in figura 5.4 e 5.5, è possibile affermare come l'industria automotive italiana sia caratterizzata da una buona solidità patrimoniale generale. In questo caso, infatti, sia MS che KCN assumono valori relativamente buoni durante tutto l'orizzonte temporale in esame, con una crescita praticamente continua anno su anno, specialmente nell'ultimo periodo.

Analisi della struttura finanziaria

L'analisi della struttura finanziaria ha come obiettivo quello di esaminare la struttura del capitale di un'impresa, ovvero l'insieme delle fonti di finanziamento utilizzate per finanziare la propria attività operativa, in modo da mettere in evidenza le diverse possibili scelte gestionali portate avanti dal management.

Per valutare la struttura finanziaria delle imprese del settore automotive, si è scelto di calcolare due indici di bilancio principali:

- Leva Finanziaria Book Value (LDF);
- Posizione Finanziaria Netta (DN).

La Leva Finanziaria Book Value, ovvero calcolata a valori contabili, è calcolata come il rapporto tra i debiti finanziari e il patrimonio netto, e descrive la tendenza dell'impresa a finanziarsi attraverso l'emissione di debito. In altre parole, questo indicatore consente di analizzare le scelte adottate dall'impresa per finanziare la propria attività operativa, scegliendo tra capitale proprio e capitale di terzi.

$$LDF = \frac{DF}{KN} \ . \tag{5.3}$$

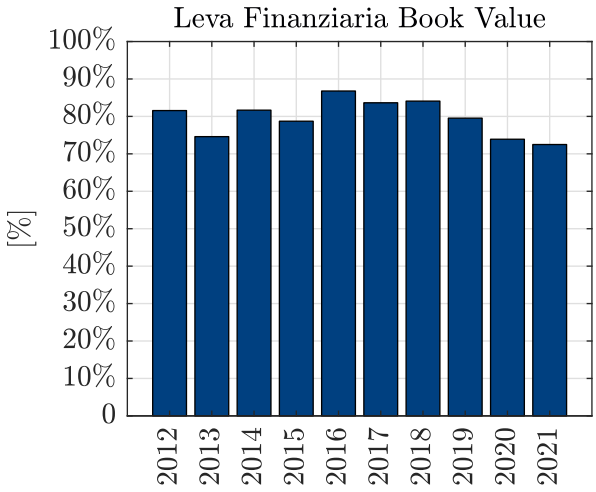


Figura 5.6. Leva Finanziaria Book Value.

La Posizione Finanziaria Netta, detta anche Indebitamento Finanziario Netto, è calcolata come la differenza tra i debiti finanziari e le attività finanziarie, e permette di osservare le condizioni di liquidità dell'impresa, valutando sia il livello complessivo di indebitamento che la potenziale capacità di restituzione del debito.

$$DN = DF - AF \ . \tag{5.4}$$

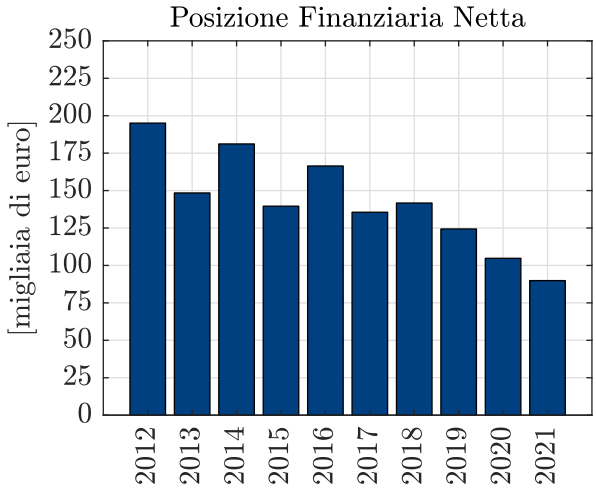


Figura 5.7. Posizione Finanziaria Netta.

Esaminando i grafici in figura 5.6 e 5.7, è possibile precisare come l'industria automotive italiana sia caratterizzata da un significativo ricorso al debito come fonte di finanziamento

principale, in quanto LDF assume valori piuttosto elevati e praticamente costanti lungo tutto il periodo di tempo di interesse, mentre DN, in modo parallelo, presenta valori abbastanza elevati, anche se, a differenza dell'indicatore precedente, risulta in leggero miglioramento con il passare del tempo.

Analisi della liquidità

L'analisi della liquidità ha come obiettivo quello di evidenziare la capacità di un'impresa di far fronte ai propri impegni di breve periodo, mettendo a confronto, congiuntamente, il grado di liquidità delle attività rispetto al grado di esigibilità delle passività.

Per valutare la liquidità delle imprese del settore automotive, si è scelto di calcolare tre indici di bilancio principali:

- Current Ratio (CR);
- Acid Test (AT);
- Interest Coverage Ratio (ICR).

Il Current Ratio è calcolato come il rapporto tra le attività correnti e le passività correnti, e rappresenta la capacità di un'impresa di ripagare i propri debiti di breve periodo attraverso la disponibilità generabile dai propri asset. In sostanza, questo indicatore fornisce informazioni sulla capacità di generare liquidità da parte dell'impresa con scadenza entro 12 mesi.

$$CR = \frac{AC}{PC} \ . \tag{5.5}$$

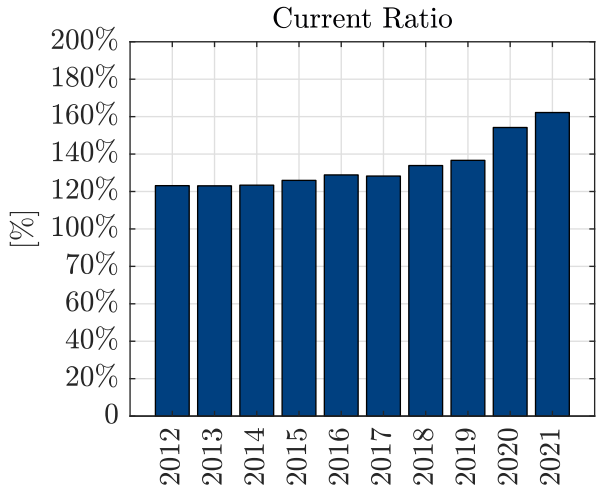


Figura 5.8. Current Ratio.

L'Acid Test, esattamente come l'indicatore precedente, permette di misurare la capacità dell'impresa di generare liquidità sufficiente nel breve periodo per soddisfare i propri debiti, anche se adesso la misura risulta più cautelativa, poiché si considerano solamente le attività con la maggiore probabilità di trasformarsi in liquidità nel prossimo periodo, sottraendo dalle attività correnti il valore delle scorte di magazzino, che rappresentano la voce più a rischio in seguito a una variazione della domanda.

$$AT = \frac{AC - MG}{PC} \ . \tag{5.6}$$

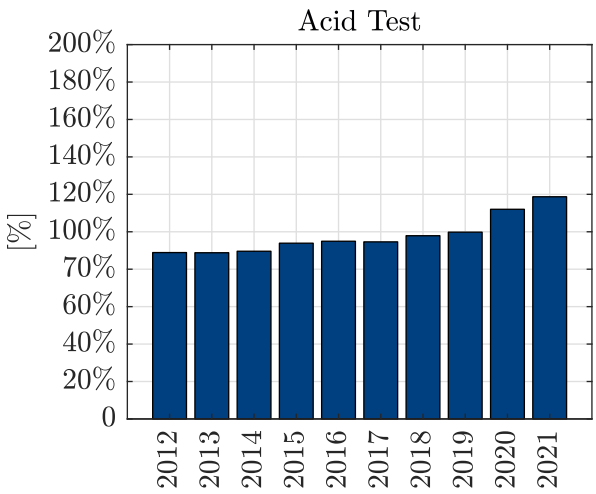


Figura 5.9. Acid Test.

Guardando i grafici in figura 5.8 e 5.9, è possibile affermare come l'industria automotive italiana sia caratterizzata dalla presenza di una buona liquidità generale, considerando che sia CR che AT assumono valori ragionevolmente elevati, dimostrando la capacità di possedere asset facilmente e velocemente trasformabili in liquidità in misura sufficiente per ripagare i propri debiti.

In aggiunta, è possibile sottolineare come la tendenza per entrambi gli indicatori presi in esame sia decisamente positiva durante tutto l'orizzonte temporale in questione, con una crescita anche piuttosto importante, specialmente negli ultimi anni.

L'Interest Coverage Ratio rappresenta la capacità di un'impresa di far fronte al pagamento degli oneri finanziari attraverso il proprio EBITDA, ovvero la possibilità di sostenere il costo del debito attraverso il risultato dell'attività operativa. Questo indicatore fornisce il numero di volte in cui il risultato dell'attività operativa, rappresentato in questo caso dall'EBITDA, riesce a coprire gli oneri finanziari e, pertanto, risulta fondamentale per valutare la solvibilità di un'impresa e la relativa capacità potenziale di indebitamento.

$$ICR = \frac{EBITDA}{OF} \ . \tag{5.7}$$

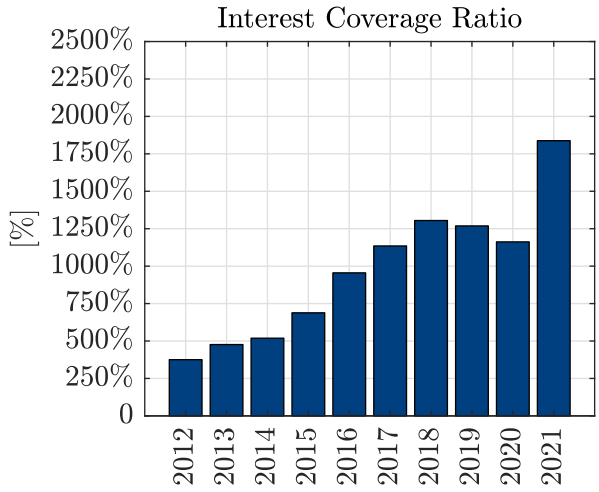


Figura 5.10. Interest Coverage Ratio.

In questo caso, osservando il grafico in figura 5.10, è possibile sottolineare come ICR presenti valori estremamente elevati, con una tendenza positiva praticamente lungo tutto l'orizzonte temporale di interesse, salvo una lieve flessione verificatasi negli ultimi anni, confermando non solo un'ottima solvibilità e capacità di rimborso del debito, ma anche la potenziale possibilità di espandere ulteriormente il proprio indebitamento in futuro.

Analisi della redditività

L'analisi della redditività ha come obiettivo quello di misurare la profittabilità generale di un'impresa sotto alcune dimensioni diverse, ovvero l'attitudine a produrre con continuità nel tempo risorse sufficienti a coprire i costi e, possibilmente, a generare profitti.

Per valutare la redditività delle imprese del settore automotive, si è scelto di calcolare quattro indici di bilancio principali:

- Return on Sales (ROS);
- Return on Equity (ROE);
- Return on Investment (ROI);
- Return on Assets (ROA).

Il Return on Sales è un indicatore di redditività delle vendite, ottenuto attraverso il rapporto tra l'EBITDA e il fatturato netto di una specifica impresa. Questo indicatore misura il margine di reddito generato per ogni unità di valore prodotto, determinando quanta parte di fatturato viene assorbita dalla gestione operativa e, quindi, descrivendo il risultato operativo dell'impresa in esame dal punto di vista commerciale.

$$ROS = \frac{EBITDA}{FN} \ . \tag{5.8}$$

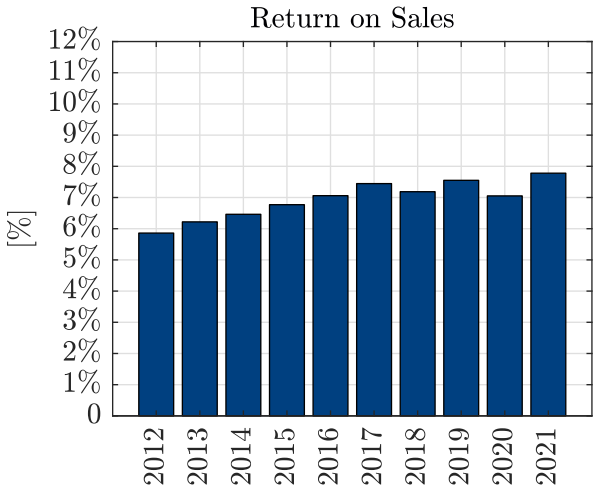


Figura 5.11. Return on Sales.

Il Return on Equity è un indicatore di redditività contabile di un'impresa, ottenuto attraverso il rapporto tra il reddito d'esercizio e il patrimonio netto. Questo indicatore misura il rendimento del capitale investito nell'impresa dagli azionisti, determinando la possibilità di generare profitti e la capacità di remunerare il capitale di rischio.

$$ROE = \frac{RE}{KN} \ . \tag{5.9}$$

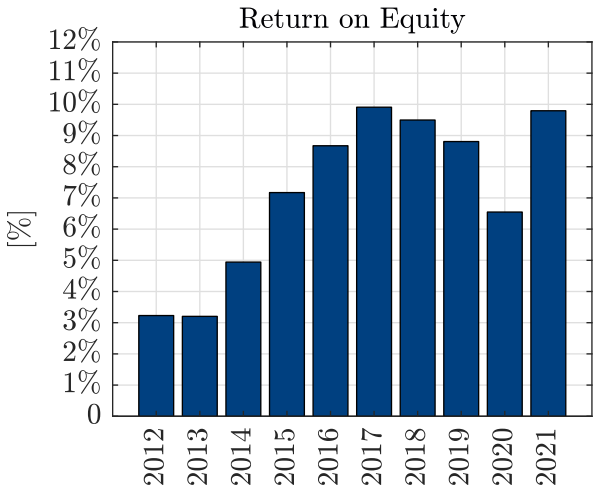


Figura 5.12. Return on Equity.

Il Return on Investment è un indicatore di redditività operativa di un'impresa, calcolato come il rapporto tra l'EBIT e il capitale operativo netto. Questo indicatore valuta la

redditività di un'impresa dal punto di vista operativo, misurando la capacità di remunerare il capitale investito, ovvero sia il capitale di rischio che il capitale di debito.

$$ROI = \frac{EBIT}{KON} \ . \tag{5.10}$$

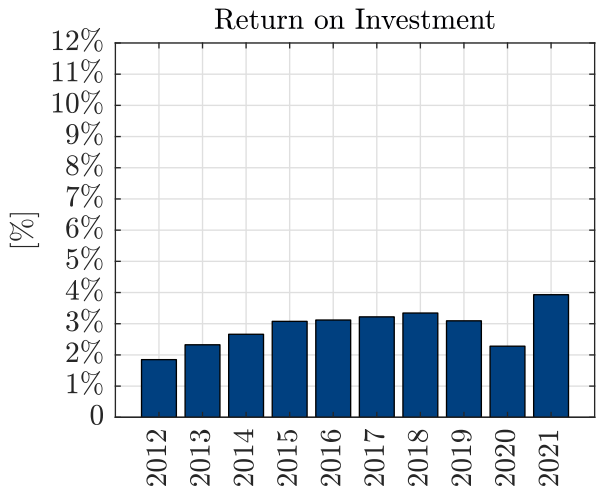


Figura 5.13. Return on Investment.

Il Return on Assets è un indicatore di redditività delle attività, calcolato come il rapporto tra il reddito d'esercizio e il totale delle attività. Questo indicatore misura la redditività di un'impresa in relazione alle risorse utilizzate per svolgere la propria attività operativa, valutando il rendimento con il quale gli asset dell'impresa vengono impiegati.

$$ROA = \frac{RE}{Attivo} \ . \tag{5.11}$$

Gli indici di redditività consentono di studiare i risultati economici di un'azienda, in quanto non forniscono informazioni tra loro alternative o sostitutive, bensì integrative, differenziandosi per la prospettiva adottata: il ROS valuta la gestione di un'impresa, indagando la capacità di trasformare i ricavi in profitti tramite la valutazione, da una parte, delle condizioni operative interne, e, dall'altra, delle situazioni di mercato esterne, il ROE presenta una prospettiva *equity side*, con l'obiettivo di calcolare la remunerazione del capitale di rischio, e, infine, il ROI e il ROA offrono una prospettiva *asset side*, con la finalità di determinare la remunerazione degli investimenti effettuati.

In questo contesto, è possibile affermare come l'industria automotive italiana sia caratterizzata da valori di redditività relativamente modesti, con una conseguente limitata appetibilità sul mercato. Infatti, il ROS (figura 5.11) assume valori positivi e di entità moderata, dimostrando una performance operativa non esaltante, anche se è possibile evidenziare una leggera tendenza positiva nell'orizzonte temporale in esame. Anche il ROE (figura 5.12) presenta valori non particolarmente elevati, con un andamento prima crescente e dopo decrescente, sottolineando una ridotta capacità di remunerare il capitale

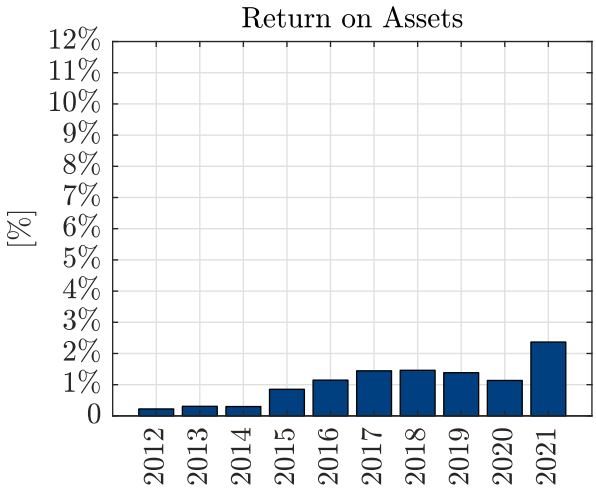


Figura 5.14. Return on Assets.

investito nell'azienda dagli azionisti. Infine, il ROI e il ROA (figura 5.13 e 5.14) mostrano valori addirittura inferiori, segnalando una ridotta capacità di remunerare il capitale investito, senza considerare l'origine delle fonti di finanziamento utilizzate.

5.2.6 Inizializzazione database

Una volta effettuato il download dei dati di bilancio delle imprese italiane del settore automotive, si è deciso di effettuare alcune operazioni fondamentali per i successivi studi, creando un file Microsoft Excel che contenesse tutte le informazioni necessarie organizzate anno per anno per il campione di imprese oggetto della trattazione.

Una delle informazioni di maggiore importanza per il raggiungimento dell'obiettivo della tesi, presente nel database AIDA, riguarda l'informazione anagrafica associata allo stato giuridico dell'impresa, unitamente all'eventuale procedura seguita.

A questo proposito, si è deciso di eseguire un'operazione di assegnazione di cinque flag, numerati da 0 a 4, in modo da classificare sulla base di specifiche procedure le imprese del campione in esame e, conseguentemente, individuare le imprese sane e le imprese anomale, descrivendone la situazione di salute economico-finanziaria:

- Flag 0: impresa classificata sana, senza particolari segnalazioni.
- Flag 1: impresa classificata anomala, caratterizzata da una delle seguenti procedure:
  - amministrazione straordinaria
  - accordo di ristrutturazione dei debiti
  - concordato fallimentare
  - concordato preventivo
  - fallimento

  - chiusura del fallimento
  - stato di insolvenza
  - chiusura per fallimento o liquidazione
  - liquidazione coatta amministrativa
  - scioglimento per atto dell'autorità
  - motivo non precisato
  - altre cause

- Flag 2: impresa classificata sana, caratterizzata da una delle seguenti procedure:
  - liquidazione
  - liquidazione volontaria
  - scioglimento e liquidazione
  - scioglimento
  - chiusura della liquidazione
  - chiusura dell'unità locale
  - cessazione di ogni attività
  - cancellata d'ufficio ai sensi art 2490 c.c.
  - scioglimento e messa in liquidazione
  - chiusura per liquidazione
  - scioglimento senza liquidazione finalizzato alla cancellazione ai sensi dell'art. 40 C. 2 DL N. 76/2020
  - cessazione delle attività nella provincia
  - cessazione d'ufficio

- Flag 3: impresa classificata sana, caratterizzata da una delle seguenti procedure:
  - scissione
  - trasferimento sede all'estero
  - fusione mediante costituzione di nuova società
  - fusione mediante incorporazione in altra società

- Flag 4: impresa classificata sana, caratterizzata da una delle seguenti procedure:
  - cancellata dal registro delle imprese
  - trasferimento in altra provincia
  - cancellata d'ufficio a seguito istituzione cciaa di fermo
  - cessata d'ufficio perché già iscritta nel registro ditte e non transitata nel registro imprese
  - provvedimento di cancellazione dal registro imprese
  - cancellazione d'ufficio dell'iscrizione al r.i.

— cancellata d'ufficio ai sensi dell'art. 40 C.5 DL 76/2020 CONV. L. 120/2020

Classificando le imprese con l'utilizzo dei vari flag assegnati nel modo appena discusso, si è scelto di definire quattro colonne flag di sintesi, che consentono di riassumere il lavoro svolto in precedenza per la successiva implementazione dei modelli di credit scoring:

- Flag di status sana/anomala - impresa: individua se l'impresa in esame è sana o anomala, assegnando uno 0 se l'impresa è sana, ovvero presenta una procedura che rientra nelle fattispecie "Flag 0", "Flag 2", "Flag 3" e "Flag 4", oppure un 1 se l'impresa è anomala, ovvero presenta una procedura che rientra nella fattispecie "Flag 1".

- Flag di status sana/anomala - anno: individua l'anno in esame in cui si è verificata la procedura che rientra nella fattispecie "Flag 1", assegnando uno 0 se l'impresa non ha subito nessuna procedura nell'anno in questione, oppure un 1 se, al contrario, l'impresa ha subito la procedura nell'anno in questione.

- Flag di status sana/anomala o in liquidazione - impresa: individua se l'impresa in esame è sana o anomala, assegnando uno 0 se l'impresa è sana, ovvero presenta una procedura che rientra nelle fattispecie "Flag 0", "Flag 3" e "Flag 4", oppure un 1 se l'impresa è anomala, ovvero presenta una procedura che rientra nelle fattispecie "Flag 1" e "Flag 2".

- Flag di status sana/anomala o in liquidazione - anno: individua l'anno in esame in cui si è verificata la procedura che rientra nelle fattispecie "Flag 1" e "Flag 2", assegnando uno 0 se l'impresa non ha subito nessuna procedura nell'anno in questione, oppure un 1 se, al contrario, l'impresa ha subito la procedura nell'anno in questione.

In poche parole, è stato necessario definire il concetto di insolvenza, effettuando una classificazione dei risultati emersi dal campione di imprese estratto dal database AIDA, per differenziare tra imprese sane e imprese anomale.

A questo punto, è stato possibile aggregare le informazioni più importanti estratte dai dati di bilancio, analizzando con attenzione le singole voci di Stato Patrimoniale, Conto Economico e Rendiconto Finanziario in modo da calcolare 57 indicatori di bilancio riguardo la redditività, la solidità patrimoniale, la produttività e la struttura operativa e, infine, la liquidità e la struttura finanziaria, con l'obiettivo di caratterizzare nel modo più approfondito possibile la condizione di salute economico-finanziaria delle imprese del campione in esame, come visibile nella tabella 5.2.

In questo lavoro, si è deciso di studiare il fenomeno dell'insolvenza di una controparte, valutandone i due principali flag di interesse[4]: il "flag di status sana/anomala - impresa" e il "flag di status sana/anomala o in liquidazione - impresa".

[4]Sinteticamente: "flag di status S/A - impresa" e "flag di status S/A+L - impresa".

Tabella 5.2. Elenco indicatori di bilancio impiegati per la costruzione dei modelli.

Indicatori						Nomenclatura
1	$\ln(FN)$	20	$\frac{AU-(PS-OS)}{AT}$	39	$\frac{KNT}{DT+KN}$	FN = Fatturato Netto
2	$\ln(AT)$	21	$\frac{EBITDA}{AT}$	40	$\frac{KNT}{DT-LIQ+KN}$	AT = Totale Attivo
3	$\frac{DT}{FN}$	22	$\frac{EBITDA-OF}{AT}$	41	ROS	AC = Attività Correnti
4	$\frac{DF}{FN}$	23	$\frac{OFN}{AT}$	42	ROE	PC = Passività Correnti
5	$\frac{PC}{FN}$	24	$\frac{RER}{AT}$	43	$ROE\ AI$	DF = Debiti Finanziari
6	$\frac{VA}{FN}$	25	$\frac{DT}{EBITDA}$	44	ROI	DT = Debiti Totali
7	$\frac{EBITDA}{FN}$	26	$\frac{DF}{EBITDA}$	45	ROA	DFL = Debiti Finanziari a lungo
8	$\frac{EBIT}{FN}$	27	$\frac{OFN}{EBITDA}$	46	$\frac{AS}{CP}$	DTL = Debiti Totali a lungo
9	$\frac{OFN}{FN}$	28	$\frac{OFN}{EBIT}$	47	$\frac{WL}{CP}$	KN = Patrimonio Netto
10	$\frac{OF}{FN}$	29	$\frac{DT}{VA}$	48	$\frac{AMM}{CP}$	KNT = Patrimonio Netto Tangibile
11	$\frac{RER}{FN}$	30	$\frac{DF}{VA}$	49	$\frac{CB}{CP}$	VA = Valore Aggiunto
12	$\frac{RE}{FN}$	31	MS	50	$\frac{AS}{CO}$	$EBITDA$ = Margine Operativo Lordo
13	$\frac{AC}{AT}$	32	KCN	51	$\frac{WL}{CO}$	$EBIT$ = Margine Operativo Netto
14	$\frac{KNT}{AT}$	33	$\frac{AC}{PC}$	52	$\frac{AMM}{CO}$	OF = Oneri Finanziari
15	$\frac{KN}{AT}$	34	$\frac{AC-MG}{PC}$	53	$\frac{CB}{CO}$	OFN = Oneri Finanziari Netti
16	$\frac{LIQ}{AT}$	35	$\frac{LIQ}{PC}$	54	$\frac{DFL}{DF}$	LIQ = Liquidità
17	$\frac{FN}{AT}$	36	$\frac{OFN}{AU}$	55	$\frac{DTL}{DT}$	AU = Autofinanziamento Lordo
18	$\frac{RS+RE}{AT}$	37	LBV	56	$\frac{KN}{DT}$	RS = Riserve di Patrimonio Netto
19	$\frac{AU}{AT}$	38	DN	57	ICR	MG = Magazzino
						RE = Risultato Netto
						RER = Risultato Netto Rettificato
						PS = Proventi Straordinari
						OS = Oneri Straordinari
						$ROE\ AI$ = ROE ante imposte
						AS = Acquisti di Servizi
						WL = Costo del Lavoro
						AMM = Ammortamento
						CB = Consumi di Beni Intermedi
						CP = Costi di Produzione
						CO = Costi Operativi

5.2.7 Pulizia e correzione dei dati

Una volta terminato il processo di download delle informazioni dal database AIDA, di assegnazione dei flag e di calcolo degli indicatori di bilancio di interesse, si è provveduto a effettuare una serie di operazioni fondamentali per la corretta implementazione di qualsiasi modello di credit scoring.

In primo luogo, si è deciso di svolgere una verifica di correttezza dei dati di bilancio, analizzando approfonditamente lo Stato Patrimoniale, sia lato Attivo che lato Passivo, il Conto Economico e il Rendiconto Finanziario. Per ottenere questo risultato, si è scelto di utilizzare Microsoft Excel, generando una numerosità di colonne di controllo che potessero eventualmente verificare la presenza di errori:

- colonna di controllo dell'Attivo dello Stato Patrimoniale;
- colonna di controllo del Passivo dello Stato Patrimoniale;
- colonna di controllo del Conto Economico;
- colonna di controllo del Rendiconto Finanziario.

In questo modo, conoscendo le regole di costruzione del bilancio, è stato possibile ricostruire correttamente i valori subtotali e totali, rispettivamente di Stato Patrimoniale, Conto Economico e Rendiconto Finanziario, aggregando le singole voci di bilancio attraverso semplici operazioni di addizione e sottrazione, allo scopo di riscontrare possibili errori. In particolare, si è verificato che tutte le colonne di controllo introdotte per la revisione riportassero valori adeguati e, soprattutto, coerenti con le informazioni estratte dal database AIDA.

Nel caso in cui sia stata individuata qualche disuguaglianza, si è provveduto a studiare nel dettaglio il problema, individuando la causa dell'incongruenza e, infine, correggendola. A questo proposito, è interessante segnalare come la quasi totalità dei problemi individuati era causata semplicemente da errori nelle singole voci di bilancio, le quali, combinandosi insieme, costituiscono le varie categorie dello Stato Patrimoniale, i saldi intermedi del Conto Economico e i flussi parziali del Rendiconto Finanziario.

Insomma, per ottenere questo risultato, si è dovuto applicare un'analisi individuale del bilancio di ogni specifica impresa per ogni specifico anno appartenente all'orizzonte temporale di interesse, in modo da accertare che tutte le colonne di controllo restituissero valori nulli, dimostrando, quindi, la correttezza e la veridicità delle informazioni esportate dal database AIDA.

In secondo luogo, si è deciso di svolgere un'operazione di controllo e correzione dei dati di bilancio, effettuando il procedimento, anche in questo caso, su Microsoft Excel. In questo caso, si è scelto di esplorare il database alla ricerca di celle contenenti alcuni valori singolari o distorti. Così, per esempio, i valori "n.d." e "n.s." sono stati sostituiti con il valore zero, mentre gli errori "#NUM!" e "#DIV/0!" sono stati sostituiti con valori compatibili con il ragionamento economico.

Questo lavoro si è rivelato essere molto oneroso, dal momento che è stato necessario studiare gli errori segnalati in modo approfondito per individuare le cause e risolvere

definitivamente il problema. Tuttavia, si tratta di un'operazione preliminare fondamentale, con l'obiettivo di verificare la correttezza dei dati in possesso e rendere questi ultimi *machine readable*, per avere una solida base con la quale costruire modelli affidabili.

Al termine di questo processo, si è provveduto a eliminare dal campione di imprese i record che non presentavano informazioni, caratterizzati da valori completamente nulli di Stato Patrimoniale, Conto Economico e Rendiconto Finanziario, ottenendo dal campione di partenza un database composto da un numero totale di record pari a 15879.

5.2.8 Pre-processing

Una corretta organizzazione e strutturazione di qualsiasi database costituisce un passaggio fondamentale per cercare di ottenere risultati soddisfacenti attraverso l'applicazione di modelli numerici più o meno complessi. In questo contesto, la fase di pre-processing emerge come una fase cruciale per la buona riuscita di tutti gli studi successivi.

Dopo aver concluso il processo di pulizia e correzione, quello che si ottiene è un database composto da 15879 record, dove ogni record comprende 57 indicatori di bilancio calcolati a partire dai dati di bilancio delle imprese in esame, in modo da descriverne la condizione di salute economico-finanziaria.

Prima di applicare modelli di credit scoring basati su tecniche di intelligenza artificiale, è necessario introdurre i principali strumenti utilizzati, indispensabili per svolgere alcune operazioni di preparazione del database, oltre che per implementare i modelli designati di albero decisionale e rete neurale.

Analisi database

Per svolgere il lavoro di tesi, sviluppando modelli di intelligenza artificiale in grado di calcolare la probabilità di insolvenza di una controparte, si è deciso di utilizzare come tecnologia di riferimento Python, il quale rappresenta uno dei linguaggi di programmazione più famosi e usati al mondo.

Python rappresenta un linguaggio di programmazione dinamico orientato agli oggetti, utilizzabile per molte tipologie di sviluppo software. Quest'ultimo, infatti, presenta numerosi punti di forza, che hanno giocato un ruolo fondamentale nella sua costante diffusione degli ultimi anni. In questo ambito, si ritiene utile citarne alcuni tra i più importanti.

Prima, Python è completamente gratuito, con licenza open-source, e, sebbene possa essere usato e distribuito senza restrizioni di copyright, possiede da sempre una community molto attiva, ricevendo miglioramenti e aggiornamenti con grande continuità.

Dopo, Python è un linguaggio di programmazione versatile, in grado di garantire un supporto multi-paradigma e multi-piattaforma, trattandosi di un linguaggio interpretato con le funzionalità più disparate, il cui codice può essere eseguito su diverse piattaforme, come Windows, Linux, MacOS, Android, iOS e molte altre, garantendo parallelamente performance molto elevate.

Infine, Python è decisamente facile da usare, essendo un linguaggio di programmazione di alto livello che risulta allo stesso tempo semplice e potente, in quanto la sua sintassi e i diversi moduli e funzioni presenti sono consistenti, intuitivi e agevoli da imparare.

Per questo motivo, Python consente di esplorare numerosi ambiti di applicazione: sviluppo di siti web e applicazioni, realizzazione di interfacce grafiche, amministrazione di sistema, automazione delle attività, calcolo scientifico e numerico, analisi di database, elaborazione di videogiochi, lavori di grafica e modellazione 3D e, non ultimo, sistemi di intelligenza artificiale, con particolare riferimento al machine learning e al deep learning.

Conclusa questa premessa, il principale vantaggio di Python riguarda senza dubbio la presenza di librerie: ogni installazione di Python include la standard library, la quale costituisce una collezione di oltre 200 moduli per svolgere numerosi compiti e attività. Inoltre, è sempre possibile scaricare e installare migliaia di moduli aggiuntivi di terze parti, progettati e mantenuti dalla community, in caso di operazioni e applicazioni particolari.

Per svolgere il presente lavoro di tesi si è reso necessario utilizzare i seguenti strumenti: Macbook Pro M1 come calcolatore, Python 3.10.4 come linguaggio di programmazione, Anaconda come package manager, oltre che PyCharm e Jupiter Notebook come ambienti di programmazione. In aggiunta, per la creazione di un ambiente specificatamente adatto allo sviluppo di modelli di intelligenza artificiale, si è reso fondamentale importare le seguenti librerie open-source, ognuna caratterizzata da specifiche funzioni peculiari:

- Pandas;
- NumPy;
- SciPy;
- Statsmodel;
- Matplotlib;
- Seaborn;
- Scikit-learn;
- TensorFlow;
- Keras.

Come detto precedentemente, il database è formato da 15879 record, ciascuno composto da 57 indicatori di bilancio. Al fine di migliorare la comprensione, si è deciso di strutturare il database anche in una versione alternativa, suddividendo il file principale Microsoft Excel in tre file CSV, "indicatori.csv", dedicato agli indicatori di bilancio, che rappresentano i dati di input del problema, oltre che "flagS/A.csv" e "flagS/A+L.csv", dedicati ai flag di status per le imprese sane e per le imprese anomale, che rappresentano invece i dati di output del problema.

A questo punto, è possibile osservare come il database in esame non risulti bilanciato, essendo i record di valore 0 molto maggiore in numero rispetto ai record di valore 1, sia considerando il problema di "flag di status S/A - impresa" che, in misura minore, considerando il problema di "flag di status S/A+L - impresa".

Infatti, come illustrato graficamente in figura 5.15, si osserva nel primo caso un tasso di default pari all'1,852%, con l'identificazione di 294 eventi di default, mentre nel secondo caso un tasso di default pari al 13,685%, con l'identificazione di 2173 eventi di default.

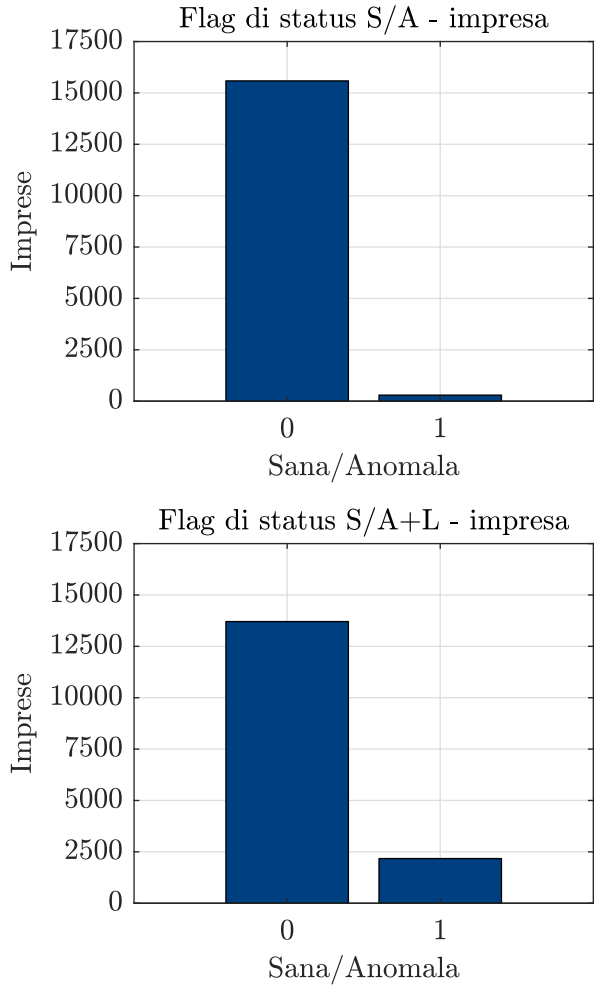


Figura 5.15. Distribuzione imprese sane e imprese anomale.

Questo sbilanciamento, nel caso in cui portasse verso una distorsione dei risultati finali, dovrebbe essere tenuto presente in fase di progettazione del modello di credit scoring, adottando particolari strategie, per risolvere il bias a favorire predizioni negative, presente internamente a causa della specifica composizione del campione in esame.

In ogni caso, i valori relativi al "flag di status S/A - impresa" potrebbero rischiare di risultare troppo esigui per sviluppare modelli di credit scoring completamente attendibili e affidabili, poiché, fortunatamente, il fenomeno dell'insolvenza rappresenta un evento sufficientemente raro, del quale si dispone di poche osservazioni nel corso del tempo.

Per questa ragione, si ritiene utile portare avanti il progetto scegliendo di analizzare parallelamente anche il "flag di status S/A+L - impresa", con una metodologia che, seppur non completamente corretta dal punto di vista teorico, consente di studiare il problema da una prospettiva ulteriore.

Gestione outliers

Il primo passo per la pulizia del dataset consiste nella gestione degli outliers, ossia le osservazioni che assumono valori estremi rispetto a tutte le altre manifestazioni disponibili. Questi ultimi sono valori anomali rispetto alla distribuzione, a causa di errori derivanti da inserimenti o misurazioni sbagliate, che, se non correttamente e tempestivamente gestiti, possono condurre verso un'alterazione della valutazione.

Per risolvere questo problema, si è effettuato la saturazione dei valori assunti da ogni indicatore di bilancio al 5° e al 95° percentile, in modo da limitare automaticamente i valori di ogni feature all'interno di questo intervallo.

In questo modo, si è provveduto a calcolare il 5° e il 95° percentile di ogni indicatore di bilancio, con una procedura implementata utilizzando sia Python che Microsoft Excel, sostituendo ogni valore inferiore al 5° percentile e ogni valore superiore al 95° percentile, rispettivamente con il 5° e con il 95° percentile.

Questa tecnica permette di gestire adeguatamente gli outlier, allineando ogni indicatore di bilancio con alcuni valori soglia definiti dai percentili presi come riferimento, senza la necessità di eliminare completamente il record in questione, rischiando così di perdere le preziose informazioni contenute al suo interno.

La rimozione degli outlier costituisce, infatti, un aspetto fondamentale per consentire al modello di intelligenza artificiale di modellare in maniera ottimale la distribuzione dei dati, senza che un valore troppo alto o troppo basso possa portare verso acquisizioni non corrette delle informazioni potenzialmente rilevabili.

Indicatori di bilancio

Dopo aver completato il processo di pulizia dei dati, è necessario analizzare e individuare le variabili di input, per ridurre la dimensionalità e selezionare le variabili ottimali, in grado di discriminare al meglio tra i flag di output. Per ottenere le variabili con cui alimentare i modelli di intelligenza artificiale, si è deciso di effettuare due operazioni preliminari:

- analisi della correlazione tra gli indicatori di bilancio e i flag di output;
- analisi della correlazione tra gli indicatori di bilancio.

Per scegliere le variabili di input è importante valutare la correlazione tra gli indicatori di bilancio e i flag di output, in modo da capire quanto ogni indicatore apporti informazioni in relazione al risultato che si vuole ottenere. Il calcolo della correlazione tra gli indicatori di bilancio e i flag di output costituisce, infatti, un passaggio essenziale per individuare le variabili più significative, in grado di differenziare al meglio tra imprese sane e anomale.

A questo punto, osservando la tabella 5.3, è possibile evidenziare le variabili di input più importanti, caratterizzate da maggiore capacità diagnostica e, conseguentemente, con una logica di inserimento prioritaria all'interno dei modelli di intelligenza artificiale.

La scelta delle variabili di input è stata effettuata prendendo in considerazione due aspetti differenti e risolvendo un risultante trade-off: da un lato, si è selezionato le variabili con maggiore capacità diagnostica, valutandone la relativa correlazione con i flag di output; dall'altro, si è rifiutato le variabili che, sebbene fossero molto rilevanti, producevano una correlazione troppo elevata con una variabile ancora più rilevante.

Tabella 5.3. Analisi di correlazione tra indicatori di bilancio.

Top	Indicatori di bilancio			
	Flag di status S/A - impresa	Correlazione	Flag di status S/A+L - impresa	Correlazione
1°	$\frac{KNT}{AT}$	0,1557	$\frac{DT}{VA}$	0,3326
2°	$\frac{KN}{AT}$	0,1553	ROE	0,3315
3°	$\frac{KNT}{DT+KN}$	0,1547	$\frac{DT}{EBITDA}$	0,3186
4°	MS	0,1482	$\ln(FN)$	0,3152
5°	LBV	0,1447	$\frac{DF}{VA}$	0,3110
6°	$\frac{KNT}{DT-LIQ+KN}$	0,1410	$\frac{KNT}{AT}$	0,2994
7°	$\frac{DT}{FN}$	0,1311	LBV	0,2991
8°	$\frac{DT}{VA}$	0,1289	$ROE\ AI$	0,2082
9°	$\frac{DF}{VA}$	0,1241	$\frac{AU-(PS-OS)}{AT}$	0,2955
10°	$\frac{PC}{FN}$	0,1232	$\frac{DF}{EBITDA}$	0,2947
11°	$\frac{DT}{EBITDA}$	0,1183	$\frac{KNT}{DT+KN}$	0,2913
12°	$\frac{DF}{EBITDA}$	0,1177	$\frac{KN}{AT}$	0,2885
13°	$\frac{DF}{FN}$	0,1147	$\frac{AS}{CO}$	0,2885
14°	ROE	0,1120	$\frac{AU}{AT}$	0,2876
15°	DN	0,1044	$\frac{EBITDA}{AT}$	0,2856
16°	$\frac{RER}{FN}$	0,1021	$\frac{EBITDA-OF}{AT}$	0,2851
17°	$ROE\ AI$	0,1000	$\frac{RER}{AT}$	0,2841
18°	$\frac{OFN}{EBITDA}$	0,0982	$\frac{AS}{CP}$	0,2839
19°	$\frac{OFN}{EBIT}$	0,0930	ROA	0,2733
20°	$\frac{FN}{AT}$	0,0916	$\frac{CB}{CP}$	0,2667

Pertanto, scegliendo due o più variabili con elevata correlazione, anche se con elevata capacità diagnostica, è possibile dimostrare come si verifichi sempre una perdita di segno per la variabile più debole, con la possibilità di osservare una conseguente riduzione di significatività del modello in questione.

I risultati dell'analisi di correlazione tra gli indicatori di bilancio calcolati nel presente lavoro di tesi è riassumibile in una matrice di correlazione, come illustrato graficamente in figura 5.16.

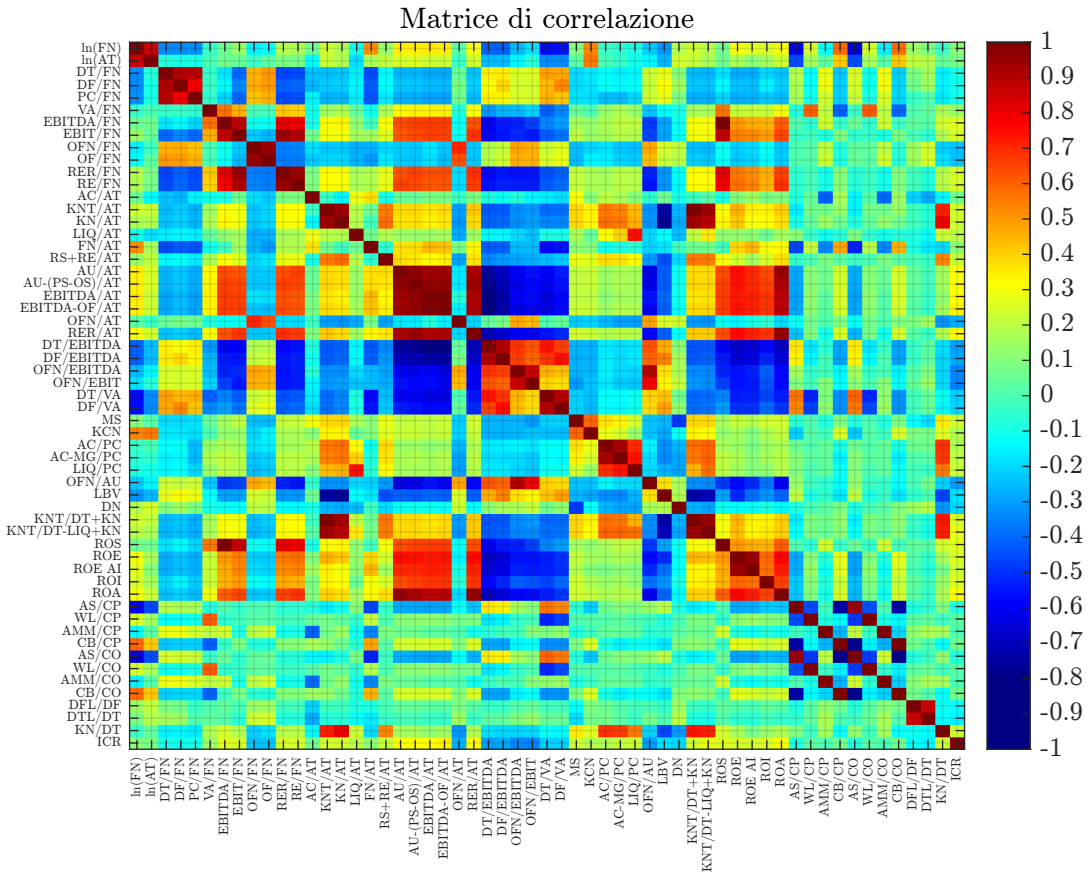


Figura 5.16. Matrice di correlazione tra indicatori di bilancio.

Normalizzazione/Standardizzazione

I modelli di intelligenza artificiale cercano di identificare dei pattern all'interno dei dati, osservando e confrontando i valori assunti da specifiche variabili di interesse, con l'obiettivo di imparare a prevederne il relativo comportamento futuro. Tuttavia, durante il processo di addestramento, si possono verificare svariati problemi, anche molto importanti, qualora i valori assunti dalle variabili di interesse risultino su scale drasticamente diverse.

Per questo motivo, è fondamentale prevedere alcune operazioni di trasformazione delle variabili di input, chiamate normalizzazione o standardizzazione, prima di procedere verso la vera e propria implementazione degli algoritmi di intelligenza artificiale, nel tentativo di riportare ogni variabile in esame sulla stessa scala e rendere ogni caratteristica equamente importante all'interno del problema.

La trasformazione consente di eliminare le differenze di scala presenti tipicamente in origine nelle diverse variabili di input, in modo tale da riportare i valori di queste ultime in uno specifico range. In questo modo, è possibile eliminare il rischio che, quando il modello compara i valori assunti dalle variabili, la feature con la scala più grande domini completamente la feature con la scala più piccola, distorcendo i risultati finali.

Le più diffuse tecniche di trasformazione delle variabili di input sono le seguenti:

- normalizzazione Min-Max;
- standardizzazione Z-score.

La normalizzazione Min-Max costituisce uno dei metodi più comuni per normalizzare i dati e calcola i valori assunti da ogni feature utilizzando la seguente trasformazione:

$$x^*_{Min-Max} = \frac{x - min}{max - min} \ , \tag{5.12}$$

dove:

- max = valore massimo della variabile;
- min = valore minimo della variabile.

Questa tecnica riporta i valori di qualsiasi feature nell'intervallo [0,1]: per ogni feature il valore minimo viene trasformato in 0, il valore massimo viene trasformato in 1 e, come risultato, qualsiasi altro valore intermedio viene corretto con un numero decimale compreso tra 0 e 1. In poche parole, si tratta di una tecnica molto semplice e potente, che, purtroppo, presenta anche un problema, non essendo in grado di gestire molto bene gli outlier.

La standardizzazione Z-score rappresenta, viceversa, una strategia per standardizzare i dati che calcola i valori assunti da ogni feature con la seguente trasformazione:

$$x^*_{Z-score} = \frac{x - \mu}{\sigma} \ , \tag{5.13}$$

dove:

- μ = media della variabile;
- σ = deviazione standard della variabile.

Anche in questo caso, la logica alla base è molto intuitiva: se un valore è esattamente uguale alla media della feature in questione, viene modificato in 0, mentre se un valore è superiore o inferiore viene trasformato rispettivamente in un numero positivo o negativo. Questa tecnica consente di trasformare i dati, gestendo in modo migliore gli outlier, anche se non riporta i valori esattamente sulla stessa scala, bensì su scale ragionevolmente simili.

Trasformare i dati è un passaggio fondamentale della fase di pre-processing, osservando che, anche avendo a disposizione un database approfondito, con feature dettagliate, se si dimenticasse di operare una trasformazione dei valori, una di queste feature potrebbe dominare completamente tutte le altre, comportando un'enorme perdita di informazione.

In questo lavoro, si è scelto di utilizzare entrambe le tecniche di trasformazione, sia normalizzazione Min-Max che standardizzazione Z-score, in modo da comparare i risultati.

Distanza di Mahalanobis

La distanza di Mahalanobis è una misura di distanza introdotta da P. C. Mahalanobis nel 1936, che può essere utilizzata per il rilevamento di outlier in dati multidimensionali, dal momento che consente di capire quanto un particolare valore osservato possa appartenere o meno a una determinata distribuzione di probabilità. Analiticamente:

$$d(x)_{Mahalanobis} = \sqrt{(x-\mu)^T \cdot \mathbf{\Sigma}^{-1} \cdot (x-\mu)} \ , \tag{5.14}$$

dove:

- μ = media della variabile;
- $\mathbf{\Sigma}^{-1}$ = matrice di varianza/covarianza.

La distanza di Mahalanobis è la variante multivariata della distanza euclidea, con la potenzialità di tenere presente la correlazione all'interno dell'insieme di dati. Infatti, con questo metodo è possibile ottenere una stima approssimativa in merito alla possibilità che un generico valore risulti un inlier, ovvero un valore che credibilmente appartiene alla distribuzione di probabilità considerata, oppure un outlier, ovvero un valore anomalo.

In poche parole, questo metodo analizza la distanza di un'osservazione dalla relativa distribuzione di probabilità, esaminandone la correlazione all'interno dell'insieme di dati e, conseguentemente, trasformandone i valori in relazione all'intensità di quest'ultima: maggiore è la correlazione e maggiore è il ridimensionamento.

Come risultato, questo metodo determina una tecnica in grado di individuare gli outlier e, in caso di necessità, rimuoverli, in modo da rendere ancora più stabile il modello.

Bilanciamento dei dati

Osservando i dati mostrati in precedenza, il campione di analisi risulta sbilanciato, comportando possibili problemi in un compito di classificazione binaria, quale il presente lavoro di tesi, poiché il modello, effettuando le previsioni, potrebbe mostrare una tendenza verso la classe più numerosa, che domina la funzione di perdita nella fase di addestramento.

Per la gestione di dataset sbilanciati, è possibile usare tre tecniche diverse:

- Downsampling;
- Oversampling;
- Loss weighting.

In questo progetto, si è scelto di utilizzare la tecnica di oversampling e, nello specifico, il metodo SMOTE, o *Synthetic Minority Oversampling Technique*, che rappresenta una tecnica statistica in grado di incrementare, anche sensibilmente, il numero di osservazioni nel set di dati in modo bilanciato, generando nuove istanze a partire dall'analisi dei casi di minoranza esistenti, forniti come input del problema.

Capitolo 6

Progettazione modello

6.1 Introduzione

In questo capitolo, si affronta la progettazione di un modello di credit scoring basato su tecniche di intelligenza artificiale. Nello specifico, l'obiettivo del presente lavoro di tesi consiste nella ricerca di trovare una soluzione al complesso problema della gestione del rischio di credito, in modo da stimare la probabilità di insolvenza di una controparte, impiegando sia modelli statistici che modelli di intelligenza artificiale.

Per raggiungere questo obiettivo, si progetta un processo di risk management completo, strutturato in due parti essenziali: una prima parte dedicata alla selezione delle variabili più importanti da utilizzare per alimentare le analisi e una seconda parte dedicata allo sviluppo di modelli in grado di prevedere il fenomeno dell'insolvenza.

A questo proposito, il capitolo si apre con una panoramica sul concetto di divisione del database in training set, validation set e testing set, oltre che una spiegazione sulle principali metriche di performance impiegate per la valutazione dei risultati, prima di passare alla presentazione delle scelte effettuate in fase di progettazione dei modelli sviluppati.

In questo contesto, il capitolo prosegue descrivendo i diversi sottoinsiemi di variabili utilizzate per la costruzione dei modelli, impiegando, per il processo di feature selection, il modello di albero decisionale e mostrando sia i risultati prodotti dal modello elementare che i risultati prodotti dai modelli più sofisticati, quali Random Forest e XGBoost.

In ultima analisi, il capitolo si chiude con la presentazione del modello di rete neurale, esponendone i numerosi sistemi costruiti, in modo da specificare, per ciascuno di essi, le scelte progettuali adottate e i risultati ottenuti, con la finalità ultima di sperimentarne le performance in relazione ai modelli statistici di analisi discriminante e regressione logistica.

6.2 Train-Valid-Test split

Il problema da risolvere consiste in un problema di classificazione binaria, in modo da modellare il rischio di credito all'interno del settore automotive italiano, valutando la probabilità di insolvenza e classificando le imprese sane e le imprese anomale.

Per raggiungere questo obiettivo, è necessario effettuare un passaggio fondamentale, ovvero dividere il database originario in tre distinti dataset, ciascuno dei quali responsabile di una precisa operazione nell'applicazione di un modello di intelligenza artificiale:

- training set;
- validation set;
- testing set.

Il training set è un set di dati impiegato durante la fase di addestramento del modello di intelligenza artificiale, in modo da consentire a quest'ultimo di acquisire la conoscenza nascosta presente nei dati.

Il validation set è un set di dati, diverso dal training set, utilizzato sempre durante la fase di addestramento del modello di intelligenza artificiale, ma impiegato questa volta per validarne le performance.

Il testing set è un set di dati impiegato durante la fase di sperimentazione del modello di intelligenza artificiale, in modo da consentire di misurare e valutare le performance ottenute dal modello finale.

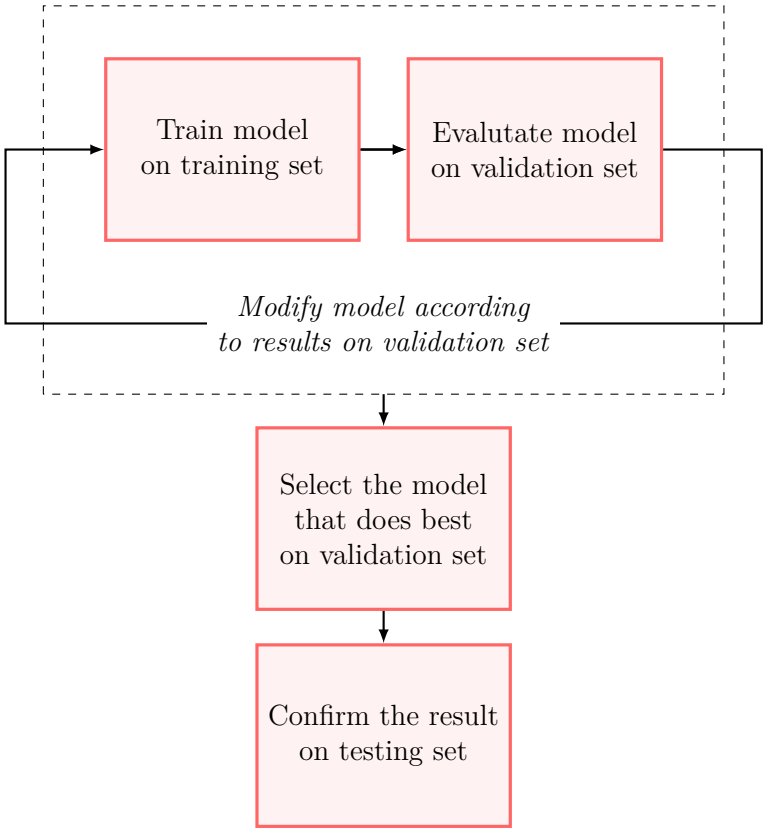


Figura 6.1. Rappresentazione grafica del concetto di Train-Valid-Test split.

In poche parole, è possibile riassumere il processo di applicazione di un modello di intelligenza artificiale in due fasi: training phase e testing phase.

La fase di training consiste nell'addestramento del modello di intelligenza artificiale sulle informazioni presenti nel training set, aggiornandone continuamente i parametri per ridurre al minimo l'errore di previsione, il quale viene calcolato con una funzione di perdita che permette di valutare la distanza tra i valori predetti e i valori osservati.

Durante la fase di training, si calcola la performance del modello sul validation set, che consente di valutare i fenomeni di overfitting e underfitting, oltre che perfezionare la scelta dei parametri, selezionando l'insieme di parametri ottimale per il modello finale.

In aggiunta, una volta definito l'insieme dei parametri, si replica la fase di training sul dataset formato unendo il training set e il validation set per ottenere il modello finale.

A questo punto, si procede con la fase di testing, durante la quale si utilizza il modello di intelligenza artificiale prodotto dopo la fase di training per valutare le performance sul testing set e, conseguentemente, osservarne i risultati complessivi.

In questo ambito, si è scelta una divisione per il Train-Valid-Test split rispettivamente pari a $\frac{2}{3}$ del database per la fase di training, suddiviso a sua volta in $\frac{2}{3}$ per il training set e $\frac{1}{3}$ per il validation set, e $\frac{1}{3}$ del database per la fase di testing, con il testing set.

6.3 Metriche di performance

Per valutare i risultati di un modello è necessario definire alcune metriche di performance, in modo tale da poter analizzare e confrontare le prestazioni ottenute dai diversi modelli.

In questo ambito, le metriche di performance più diffuse sono le seguenti:

- ROC/AUC;
- Confusion matrix;
- Accuracy;
- Precision;
- Recall.

Receiving Operating Characteristics, abbreviata ROC, rappresenta una metrica di performance molto diffusa e utilizzata nell'ambito dei modelli di credit scoring per problemi di classificazione binaria. Studiando a fondo a figura 6.2, è possibile osservare come nei due assi del grafico, x e y, siano rappresentati rispettivamente il totale delle imprese sane e il totale delle imprese anomale, oltre che due curve principali: la prima coincide con la diagonale del grafico e specifica un sistema di credit scoring completamente casuale, mentre la seconda combacia con il segmento superiore che delimita il grafico e descrive un sistema di credit scoring perfetto. Tra queste curve, è possibile inserire la curva del modello di credit scoring di cui si intende calcolare le performance, definita ROC curve.

Parallelamente, per valutare le performance di un modello, è necessario introdurre il concetto di Area Under the ROC Curve, abbreviata AUC, che rappresenta l'area sottesa dalla curva definita in precedenza e consente di confrontare l'area sottesa dal grafico del

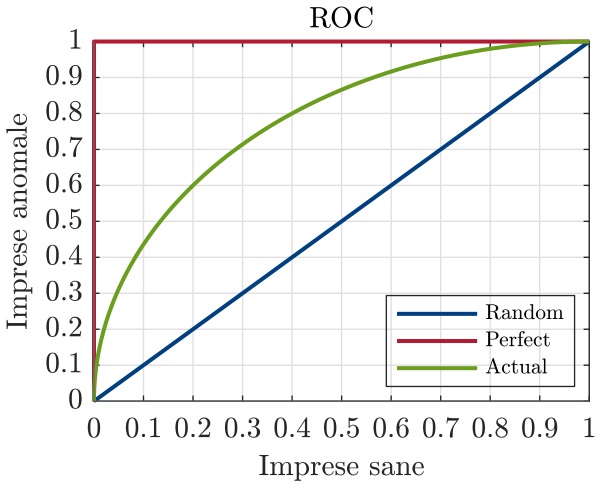


Figura 6.2. ROC curve.

modello di credit scoring in oggetto, con due benchmark, ovvero il modello casuale e il modello perfetto. Infatti, considerando che l'area sottesa dalle curve che specificano il modello casuale e il modello perfetto vale rispettivamente 0,5 e 1, i sistemi reali presentano un'area compresa tra questi due valori, risultando tanto migliori quanto si allontano dal modello casuale e si avvicinano al modello perfetto.

Confusion Matrix costituisce una metrica di performance estremamente intuitiva, che consente di valutare un modello di classificazione binaria, evidenziandone le osservazioni correttamente ed erroneamente classificate, con uno schema grafico suddiviso in quattro quadranti (figura 6.3): True Positive, True Negative, False Positive e False Negative.

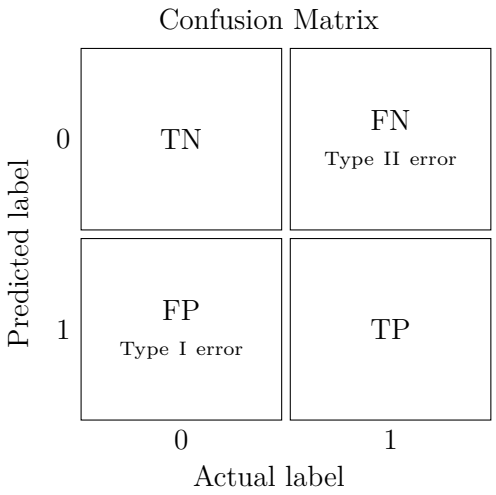


Figura 6.3. Confusion matrix.

In questo grafico, è doveroso sottolineare come sia possibile individuare due tipologie di errore, diversi dal punto di vista del costo, definiti errore del primo o del secondo tipo.

Accuracy, Precision e Recall sono metriche di performance relativamente simili, che, però, valutano i risultati di un modello sotto dimensioni diverse. A tale riguardo, Accuracy consente di esprimere la percentuale di osservazioni correttamente classificate sul totale delle osservazioni analizzate, riconoscendo, quindi, le imprese sane e le imprese anomale.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \ . \tag{6.1}$$

In aggiunta, Precision permette di valutare la percentuale di osservazioni anomale reali sul totale delle osservazioni classificate come anomale dal modello in questione, mentre, in modo simile, Recall permette di valutare la percentuale di osservazioni anomale reali che vengono classificate come anomale dal modello in questione.

$$Precision = \frac{TP}{TP + FP} \ . \tag{6.2}$$

$$Recall = \frac{TP}{TP + FN} \ . \tag{6.3}$$

6.4 Scelta delle variabili

Il punto di partenza consiste nella selezione delle variabili da utilizzare per la costruzione dei modelli di intelligenza artificiale: l'idea del presente lavoro di tesi prevede di combinare diversi algoritmi per ottenere un risultato finale quanto più possibile efficace ed efficiente.

In questo contesto, si è scelto di utilizzare il modello di albero decisionale e, più precisamente, le tecniche Random Forest e XGBoost, per il processo di feature selection.

Queste tecniche, come spiegato in precedenza, rappresentano una famiglia di algoritmi di machine learning estremamente diffusa, il cui successo può essere spiegato considerando che, generalmente, sono in grado di unire sia un buon potere predittivo che, soprattutto, una facile interpretabilità dei risultati, la quale consiste nel fatto che è relativamente semplice derivare l'importanza di ciascuna variabile dalla struttura dell'albero decisionale, ovvero calcolare quanto ogni variabile contribuisca alla decisione finale del modello.

Pertanto, Random Forest e XGBoost, considerando queste caratteristiche, risultano particolarmente adatti, non solo per risolvere problemi di regressione e di classificazione, ma anche per gestire il processo di feature selection.

Nel modello di albero decisionale è possibile definire diverse metriche per classificare l'importanza relativa delle variabili all'interno del modello, tra queste, le più importanti:

- Gain: che descrive il guadagno medio relativo, calcolato considerando tutti gli split in cui la feature in esame viene utilizzata.
- Weight: che descrive la frequenza relativa con cui la feature in esame viene utilizzata per dividere i dati attraverso tutti gli alberi decisionali previsti dal modello.
- Cover: che descrive la copertura media relativa, calcolata considerando tutti gli split in cui la feature in esame viene utilizzata.

In questo ambito, si è scelto di fare riferimento al Gain, in quanto rappresenta la metrica più rilevante per descrivere l'importanza di una variabile all'interno di un modello, essendo in grado di calcolare il contributo relativo di ogni variabile per la definizione dello stesso.

Comparando tra loro i valori assunti da questo attributo, infatti, è possibile classificare le variabili del modello in ordine di importanza relativa, ovvero in ordine di incidenza per la costruzione di un modello e, conseguentemente, per la generazione di una previsione.

6.4.1 Implementazione

A questo punto, è possibile descrivere il codice Python utilizzato per la progettazione di un processo di feature selection basato sull'applicazione dei modelli di albero decisionale, quali Random Forest e XGBoost, riportando i passaggi usando un flag come riferimento.

Per migliorare la comprensione, si è deciso di mostrare fin da subito, in modo da evitare inutili ripetizioni, alcune operazioni preliminari, come l'importazione delle librerie, il caricamento del database e la suddivisione di quest'ultimo in training set e testing set.

In primo luogo, si importa l'insieme di librerie usate per il processo di feature selection:

```
#library for data handling
import pandas as pd
#library for calculations
import numpy as np
#library for plotting
import seaborn as sns
import matplotlib
from matplotlib import rc
import matplotlib.pyplot as plt
import matplotlib.cm as cm
#library for pre-processing operations
from sklearn import datasets
from sklearn import preprocessing
#library for the Decision Tree classifier
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
#library for the Random Forest classifier
from sklearn.ensemble import RandomForestClassifier
#external package for XGBoost Classifier
import xgboost as xgb
#library for feature selection
from sklearn.feature_selection import SelectFromModel
#library for evaluation tools
from sklearn import metrics
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
#library for operations
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
```

In secondo luogo, si carica il dataset, dividendo le variabili di input dai flag di output:

```
#loading the dataset
df = pd.read_excel('AIDA_Export_Automotive_Industry.xlsx',"Risultati")
print('dimensionality',df.shape)
#indicatori di bilancio
df_indicatori = df.iloc[:,1:58]
#flag di status S/A - impresa
df_flag1 = df.iloc[:,58:59]
#flag di status S/A+L - impresa
df_flag2 = df.iloc[:,59:60]
```

6.4.2 Albero decisionale

Random Forest e XGBoost rappresentano due modelli di intelligenza artificiale basati sul concetto di albero decisionale e, per questo motivo, si ritiene opportuno e interessante iniziare, per lo sviluppo di un processo di feature selection, dalla costruzione di un modello di albero decisionale singolo, utile come punto di partenza per svolgere le analisi successive.

Prima di procedere, è doveroso effettuare una precisazione sul database impiegato, in quanto il processo in questione è stato eseguito utilizzando contemporaneamente il dataset trasformato sia con i valori normalizzati MinMax che con i valori standardizzati Z-score, anche se con risultati praticamente identici in entrambe le applicazioni.

Implementazione

Per l'implementazione del modello di albero decisionale semplice si è scelto di utilizzare, in merito alla fase di training del modello, sia l'intero database a disposizione che una divisione del database in training set e testing set, rispettivamente pari a $\frac{2}{3}$ e $\frac{1}{3}$, ricorrendo alla funzione *train_test_split*:

```
#splitting dataset into training set and testing set
X_train1,X_test1,y_train1,y_test1 = train_test_split(df_indicatori,df_flag1,
    test_size=1/3,random_state=42)
```

Dal momento che non si è interessati a impiegare il modello di albero decisionale per risolvere il problema di classificazione tra imprese sane e imprese anomale, si è ritenuto opportuno utilizzare non solo una divisione del database in training set e testing set, più corretta dal punto di vista teorico per la validazione dei risultati, ma anche l'intero database a disposizione, senza nessuna divisione, in modo da sfruttare tutta l'informazione contenuta al suo interno per individuare le variabili più significative.

A questo proposito, si inizializza il classificatore *DecisionTreeClassifier*, scegliendo adeguatamente gli iperparametri del modello, si addestra un albero decisionale singolo, utilizzando la funzione *fit*, e, infine, si predice il risultato, utilizzando la funzione *predict*:

```
#implementing a single Decision Tree model
clf = DecisionTreeClassifier(random_state=42,criterion='gini',max_depth=3)
#fitting the model
clf = clf.fit(X_train1,y_train1)
#predicting the result
y_pred1 = clf.predict(X_test1)
```

Per la definizione del modello di albero decisionale singolo si è deciso di utilizzare come criterio decisionale l'impurità di Gini, oltre che una limitata dimensione dell'albero, in modo da ottenere il giusto compromesso tra complessità e interpretabilità dei risultati.

Così facendo, è possibile osservare i risultati conseguiti dal modello appena prodotto, visualizzando graficamente la struttura dell'albero decisionale, con la funzione *plot_tree*:

```
#describing the model
plt.figure(figsize=(20,10))
tree.plot_tree(clf,
               filled=True,rounded=True,
               feature_names=X_train1.columns,
               class_names=['0','1'],
               fontsize=12)
plt.show()
```

In definitiva, è possibile controllare la predizione del modello sui dati, utilizzando, per comparare i risultati previsti e osservati, la matrice di confusione e l'Accuracy, rispettivamente tramite le funzioni *confusion_matrix* e *accuracy_score*, in modo da mostrare le performance del modello sia graficamente che analiticamente:

```
#computing the accuracy and visualizing the confusion matrix
print('Accuracy:',metrics.accuracy_score(y_test1,y_pred1))
mat = confusion_matrix(y_test1,y_pred1)
sns.heatmap(mat.T,square=True,annot=True,cbar=False,fmt=".0f")
plt.xlabel('Actual label')
plt.ylabel('Predicted label');
plt.show()
```

Analisi risultati

Il modello di albero decisionale singolo risulta estremamente semplice, con l'obiettivo di studiare il database a disposizione e svolgere alcune interessanti osservazioni preliminari.

Con la visualizzazione grafica del modello, infatti, è possibile analizzare la struttura dell'albero decisionale e interpretarne i risultati: in ogni nodo decisionale si può osservare la variabile di decisione selezionata per dividere i dati e il valore assunto da quest'ultima, il valore del criterio di decisione adottato, il numero di osservazioni presenti in ogni ramo e, infine, l'esito finale della classificazione (figura 6.4 e 6.5).

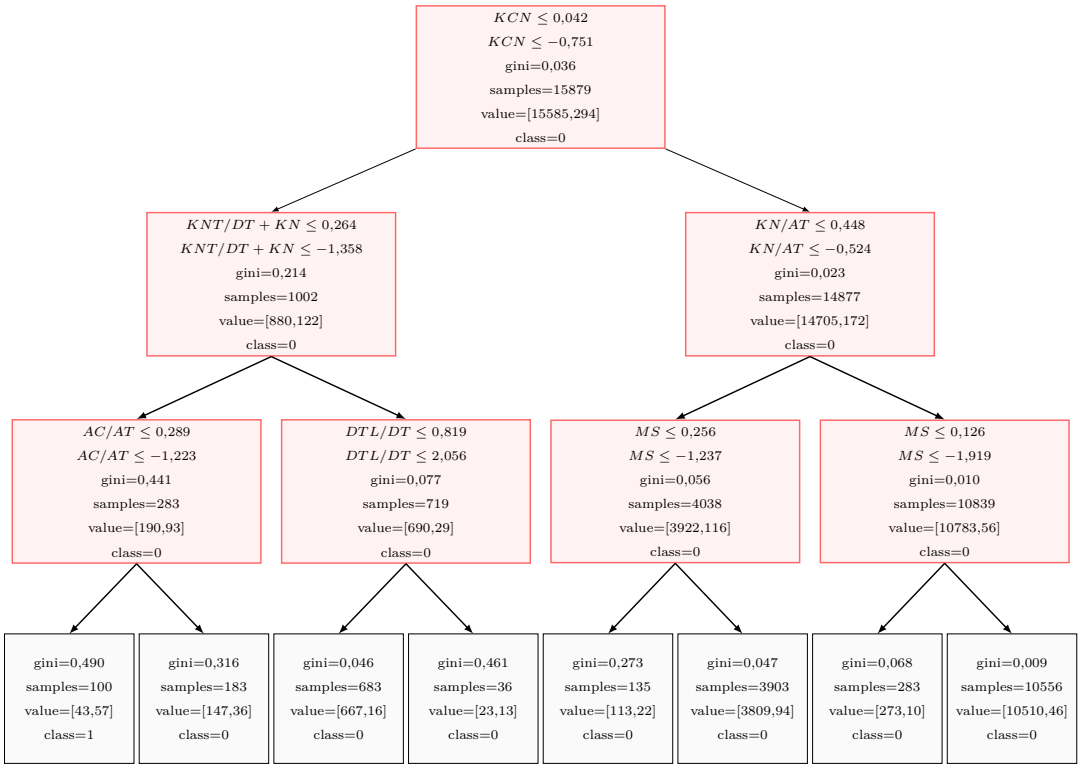


Figura 6.4. Albero decisionale (flag di status S/A - impresa).

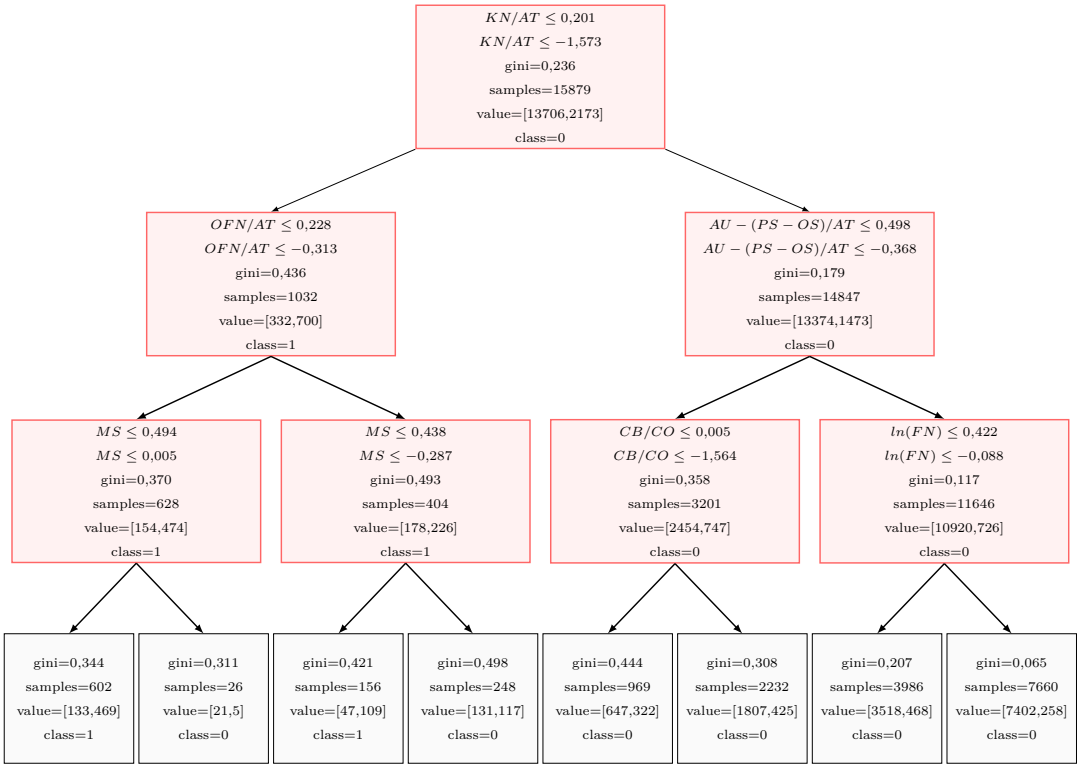


Figura 6.5. Albero decisionale (flag di status S/A+L - impresa).

Con la visualizzazione analitica del modello, invece, è possibile valutare le performance dell'albero decisionale, applicato ai due determinati problemi che si intende analizzare, ovvero "flag di status S/A - impresa" e "flag di status S/A+L - impresa", utilizzando sia il database completo che il database diviso in training set e testing set (figura 6.6 e 6.7).

Nel primo problema, si è ottenuto rispettivamente una Accuracy del 98,05%, a fronte di un numero di imprese correttamente classificate pari a 5190 su un totale di imprese analizzate nel testing set pari a 5293, e una Accuracy del 98,24%, a fronte di un numero di imprese correttamente classificate pari a 15599 su un totale di imprese analizzate nel database pari a 15879.

Nel secondo problema, si è ottenuto rispettivamente una Accuracy dell'88,23%, a fronte di un numero di imprese correttamente classificate pari a 4670 su un totale di imprese analizzate nel testing set pari a 5293, e una Accuracy del 88,82%, a fronte di un numero di imprese correttamente classificate pari a 14104 su un totale di imprese analizzate nel database pari a 15879.

In realtà, non si è particolarmente interessati alle performance del modello, in quanto il modello di albero decisionale semplice non risulta sufficientemente adeguato per modellare la complessità che governa il fenomeno dell'insolvenza di una controparte. Tuttavia, è importante studiare questo modello poiché, rappresentando il componente fondamentale dei modelli Random Forest e XGBoost, costituisce il punto di partenza ideale.

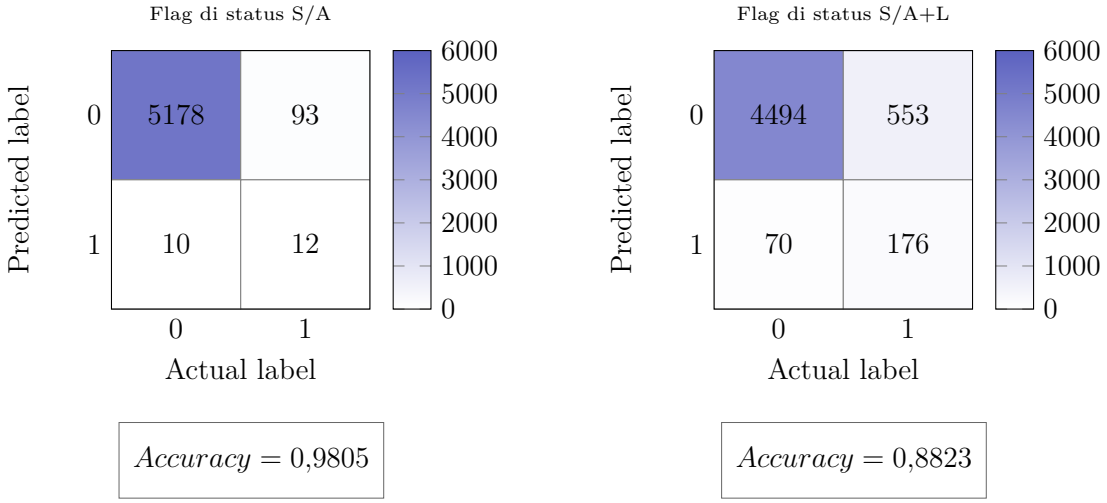


Figura 6.6. Albero decisionale (testing set).

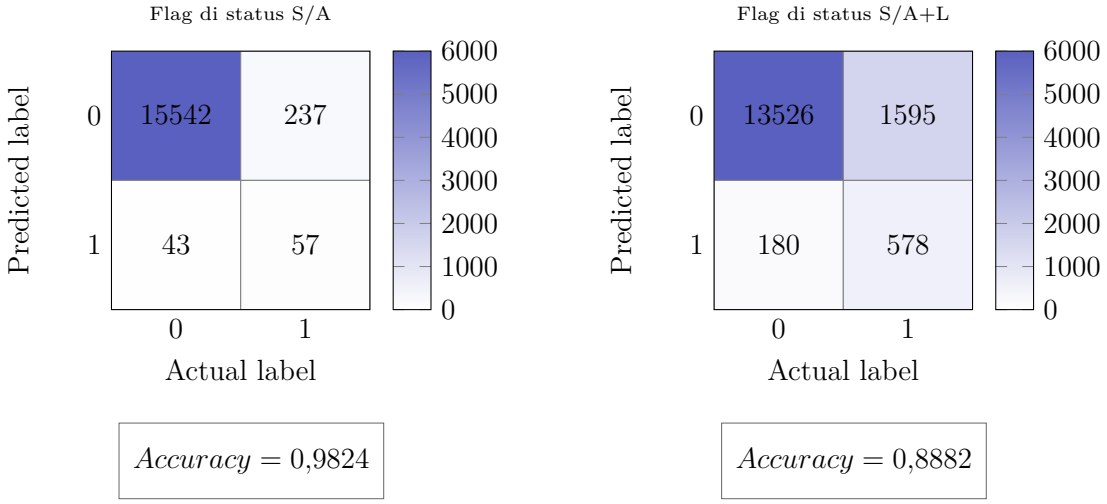


Figura 6.7. Albero decisionale (database).

6.4.3 Ensemble

Il semplice modello precedente, basato sull'uso di un singolo albero decisionale, non risulta sufficientemente sofisticato per gestire la complessità contenuta all'interno del presente problema di analisi, essendo affetto dai disturbi che accomunano tutti gli alberi decisionali.

Per questo motivo, è necessario utilizzare, sfruttando le diverse tecniche *ensemble*, un modello più complesso e articolato, generato applicando le strategie Bagging e Boosting, ovvero, rispettivamente, Random Forest e XGBoost, in grado di combinare una numerosità di alberi decisionali e ottenere, conseguentemente, risultati migliori.

Random Forest

Come in precedenza, per l'implementazione del modello Random Forest si è deciso di utilizzare, in merito alla fase di training del sistema, sia l'intero database a disposizione che una divisione del database in training set e testing set, rispettivamente pari a $\frac{2}{3}$ e $\frac{1}{3}$:

```
#splitting dataset into training set and testing set
X_train1,X_test1,y_train1,y_test1 = train_test_split(df_indicatori,df_flag1,
    test_size=1/3,random_state=42)
```

Pertanto, si procede a generare il classificatore *RandomForestClassifier*, scegliendo in maniera opportuna gli iperparametri del modello, si continua addestrando il modello Random Forest e si conclude calcolando la predizione sui dati in esame:

```
#implementing a Random Forest Classifier
rf_class = RandomForestClassifier(n_estimators=100,criterion='gini',oob_score=True)
#fitting the Random Forest model
rf_object = rf_class.fit(X_train1,np.array(y_train1).ravel())
#getting the prediction for the data
rf_prediction = rf_object.predict(X_test1)
```

In ultima analisi, si osserva la performance del modello, comparando i risultati ottenuti:

```
#computing the accuracy
print('Accuracy score:',accuracy_score(y_test1,rf_prediction))
#printing the oob_score
print('OOB-Score:',rf_object.oob_score_)
#visualizing the confusion matrix
mat = confusion_matrix(y_test1,rf_prediction)
sns.heatmap(mat.T,square=True,annot=True,cbar=False,fmt=".0f")
plt.xlabel('Actual label')
plt.ylabel('Predicted label');
plt.show()
```

Come è facile osservare analizzando le figure 6.8 e 6.9, le prestazioni sono nettamente migliori in tutte le applicazioni: nel caso di "flag di status S/A - impresa" si ottiene un incremento di Accuracy pari allo 0,44% sul testing set e all'1,78% sul database completo; nel caso di "flag di status S/A+L - impresa" si ottiene un risultato ancora superiore, con un incremento di Accuracy pari al 2,53% sul testing set e al 12,57% sul database completo.

Tuttavia, l'obiettivo di questa applicazione non consiste nel valutare le performance del modello in merito alla capacità di classificare le imprese sane e le imprese anomale presenti nel database, quanto piuttosto utilizzare questo algoritmo per mettere in evidenza le variabili più discriminanti, in modo da generare un processo di feature selection.

Per ottenere questo risultato, si utilizza la funzione *feature_importances_*, che consente di determinare l'importanza relativa di ciascuna delle variabili analizzate dal modello, utilizzando come metrica il Gain per definire il concetto di importanza:

```
#plotting the variables in order of importance
gain = rf_class.feature_importances_

plt.rcParams['figure.figsize']=(15,15)
sorted_idx = rf_class.feature_importances_.argsort()
plt.barh(df_indicatori.columns[sorted_idx],rf_class.feature_importances_[sorted_idx
    ])
plt.xlabel('Random Forest Feature Importance')
plt.xlabel('Balance Sheet Ratios')
plt.show()
```

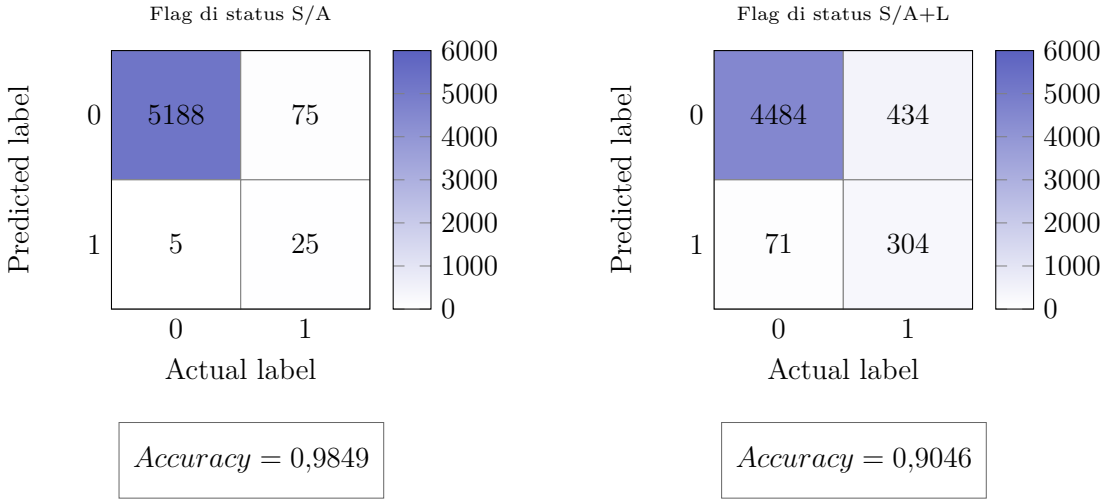


Figura 6.8. Random Forest (testing set).

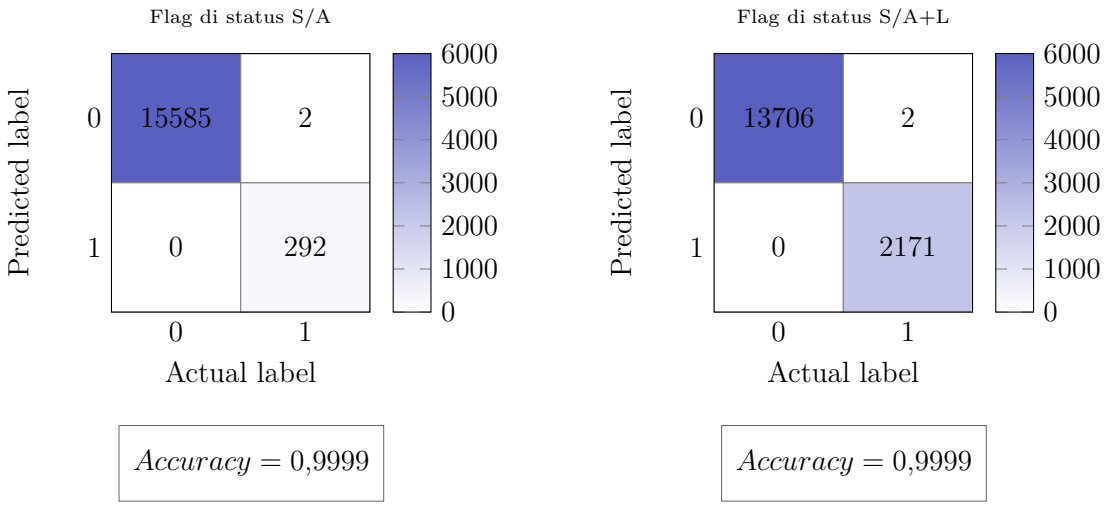
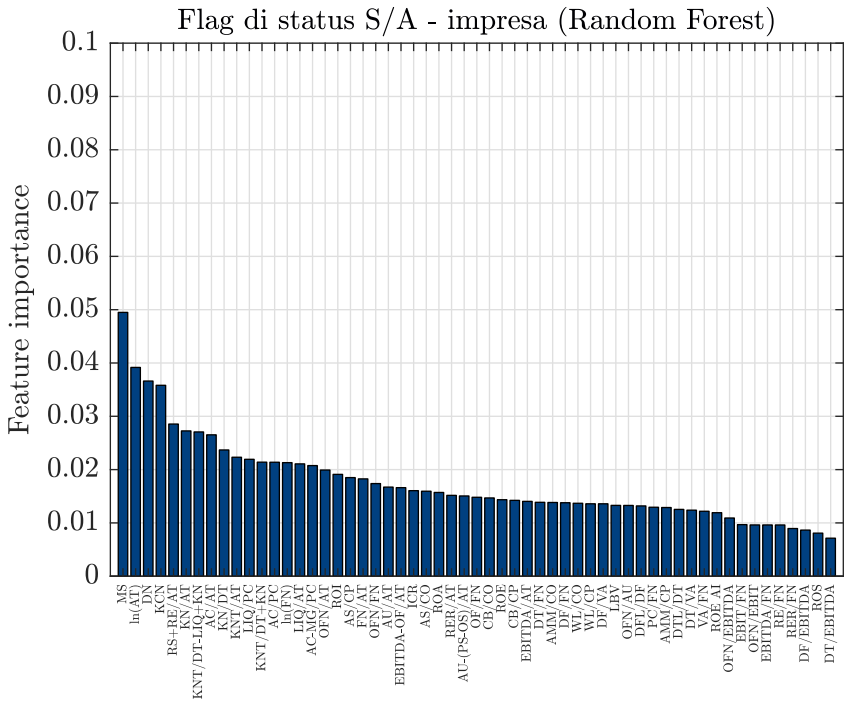


Figura 6.9. Random Forest (database).

In questo modo, è possibile rappresentare graficamente le variabili del problema in ordine di importanza relativa, come visibile in figura 6.10, suddividendo i risultati nei due flag di output in esame: "flag di status S/A - impresa" e "flag di status S/A+L - impresa".

Come facilmente intuibile, i due modelli producono in output risultati differenti, anche se, è doveroso sottolineare come la maggior parte degli indicatori più diagnostici risultino in comune a entrambi i modelli, dimostrando in qualche modo il loro significato economico.



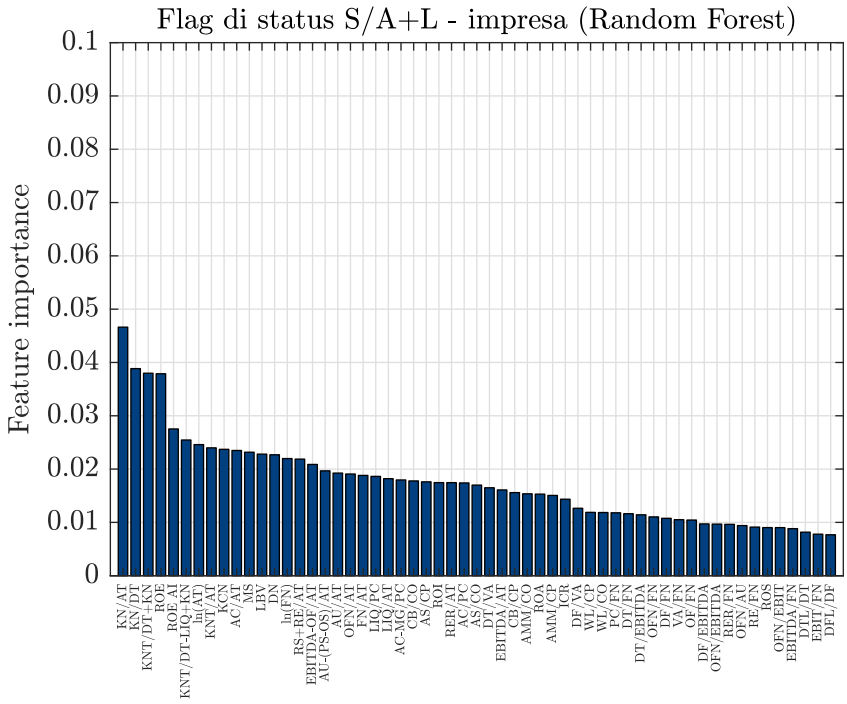


Figura 6.10. Processo di feature selection utilizzando Random Forest.

XGBoost

In modo analogo, è possibile utilizzare un modello alternativo, ovvero XGBoost, seguendo gli stessi passi seguiti in precedenza per il modello Random Forest e sintetizzati di seguito.

Allora, prima di tutto, per l'implementazione del modello XGBoost si è deciso di utilizzare, in merito alla fase di training del sistema, sia l'intero database a disposizione che una divisione del database in training set e testing set, rispettivamente pari a $\frac{2}{3}$ e $\frac{1}{3}$:

```
#splitting dataset into training set and test set
X_train1,X_test1,y_train1,y_test1 = train_test_split(df_indicatori,df_flag1,
    test_size=1/3,random_state=42)
```

Successivamente, si procede a costruire il classificatore *XGBClassifier*, scegliendo in maniera opportuna gli iperparametri del modello, si prosegue addestrando il modello XGBoost e si termina calcolando la predizione sui dati in esame:

```
#implementing a XGBoost Classifier
xg_class = (xgb.XGBClassifier(n_estimators=100,max_depth=5,learning_rate=0.1,
    subsample=0.5))
#fitting the XGBoost model
xg_object = xg_class.fit(X_train1,np.array(y_train1).ravel())
#getting the prediction for the data
xg_prediction = xg_object.predict(X_test1)
```

Per ultima cosa, si osserva la performance del modello, comparando i risultati ottenuti:

```
#computing the accuracy
print('Accuracy score:',accuracy_score(y_test1,xg_prediction))
#visualizing the confusion matrix
mat = confusion_matrix(y_test1,xg_prediction)
sns.heatmap(mat.T,square=True,annot=True,cbar=False,fmt=".0f")
plt.xlabel('Actual label')
plt.ylabel('Predicted label');
plt.show()
```

Anche in questo caso, analizzando le figure 6.11 e 6.12, le prestazioni sono leggermente migliori nelle varie applicazioni: nel caso di "flag di status S/A - impresa" si ottiene un incremento di Accuracy pari allo 0,39% sul testing set e all'1,24% sul database completo; nel caso di "flag di status S/A+L - impresa" si ottiene un risultato di nuovo superiore, con un incremento di Accuracy pari al 1,76% sul testing set e al 4,74% sul database completo.

A questo riguardo, sebbene entrambi i modelli, Random Forest e XGBoost, consentano di raggiungere ottimi livelli di Accuracy, è comunque possibile mettere in evidenza come Random Forest permetta di ottenere risultati ancora migliori rispetto a XGBoost, soprattutto nel caso di "flag di status S/A+L - impresa".

A questo punto, si utilizza la funzione *feature_importances_* indicata in precedenza, per determinare l'importanza relativa di ciascuna delle variabili analizzate dal modello, utilizzando come metrica il Gain per definire il concetto di importanza:

```
#plotting the variables in order of importance
gain = xg_class.feature_importances_

plt.rcParams["figure.figsize"]=(15,15)
sorted_idx = xg_class.feature_importances_.argsort()
plt.barh(df_indicatori.columns[sorted_idx],xg_class.feature_importances_[sorted_idx
    ])
plt.xlabel("XGboost Feature Importance")
plt.xlabel('Balance Sheet Ratios')
plt.show()
```

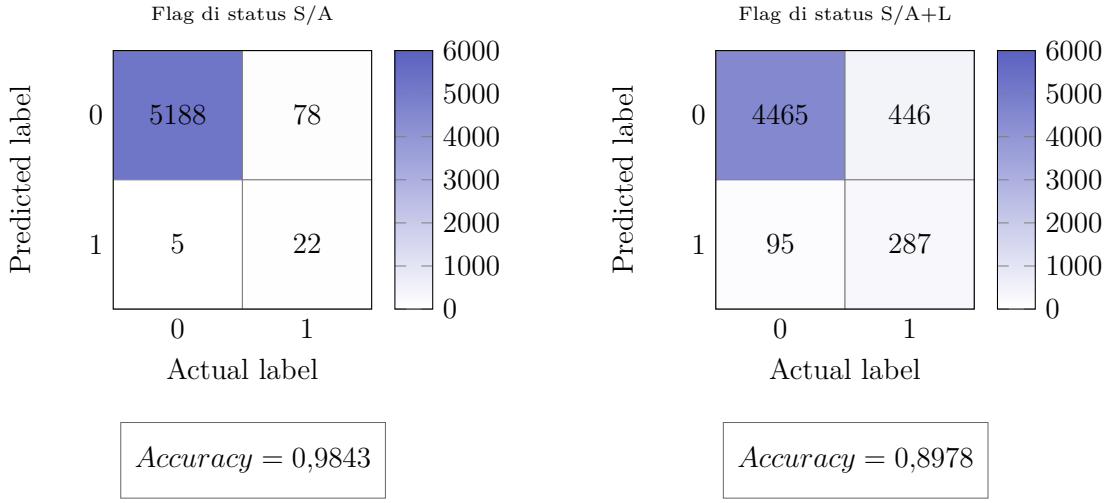


Figura 6.11. XGBoost (testing set).

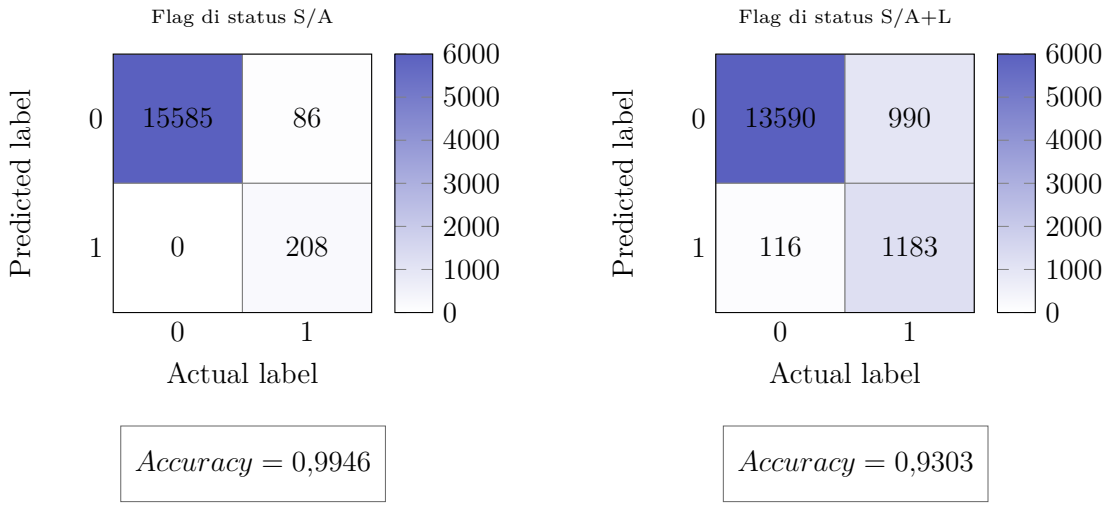


Figura 6.12. XGBoost (database).

Come risultato, è possibile effettuare le medesime considerazioni indicate in precedenza, rappresentando graficamente le variabili del problema in ordine di importanza relativa, come visibile in figura 6.13, suddividendo sempre i risultati nei due flag di output in esame: "flag di status S/A - impresa" e "flag di status S/A+L - impresa".

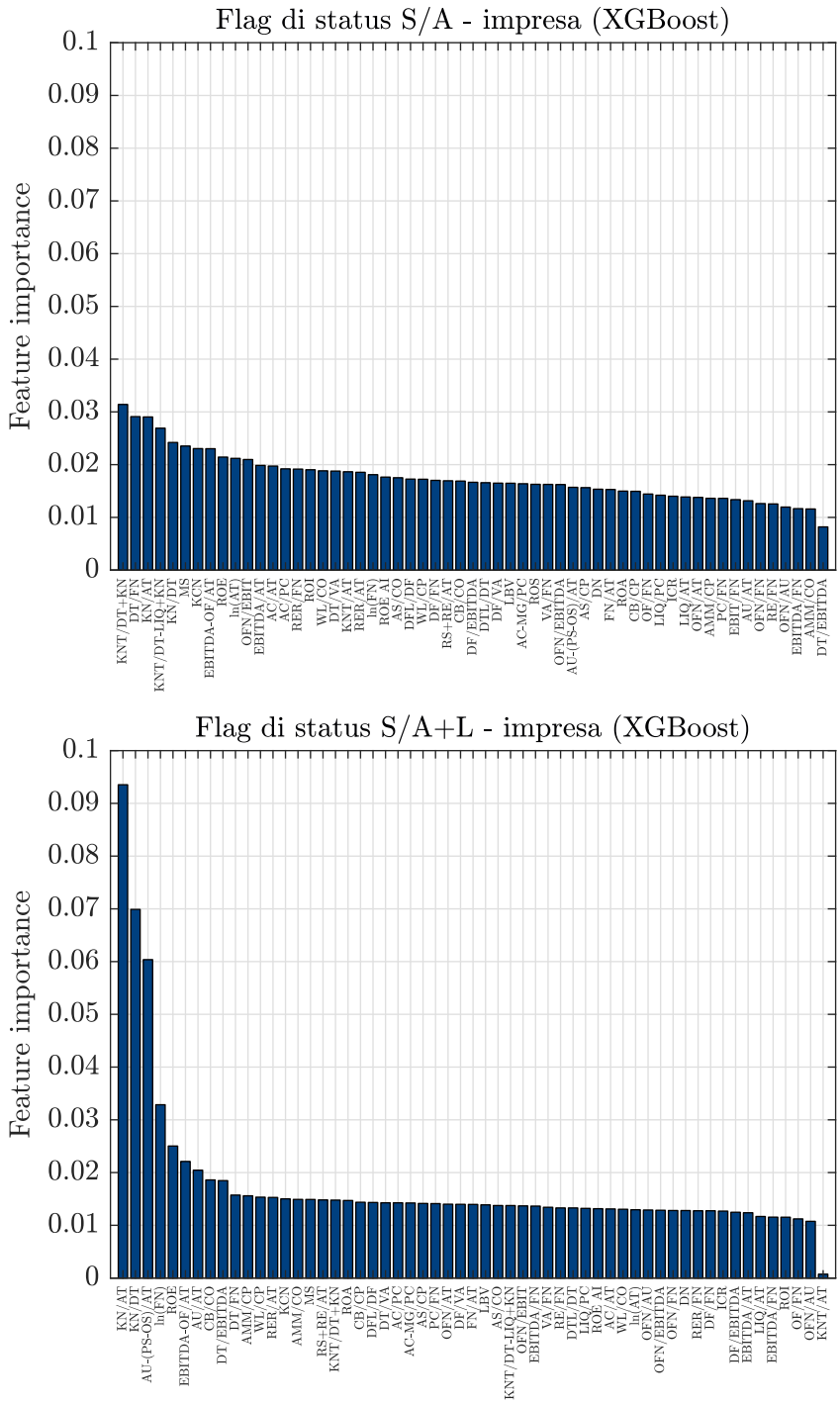


Figura 6.13. Processo di feature selection utilizzando XGBoost.

6.4.4 Indicatori

In generale, è possibile anche decidere di automatizzare il processo di feature selection e modificarlo, in base alle necessità, impiegando Random Forest e XGBoost, attraverso l'imposizione di un livello di cut-off per selezionare o meno una variabile in modo arbitrario.

In questo modo, *SelectFromModel* seleziona, utilizzando determinati livelli di cut-off, le variabili più rilevanti, ovvero le variabili la cui importanza risulta superiore rispetto all'importanza media calcolata su tutte le variabili del problema di analisi:

```
#automatic feature selection process
sel = SelectFromModel(RandomForestClassifier(n_estimators = 100))
sel.fit(df_indicatori,np.array(df_flag1).ravel())
sel.get_support()
sel_feat = df_indicatori.columns[(sel.get_support())]
print(sel_feat)
```

In questo caso, si è scelto di mettere in evidenza i primi 20 indicatori, differenziando i risultati sia nel caso di Random Forest (tabella 6.1) che nel caso di XGBoost (tabella 6.2).

Una volta ottenuta la classifica degli indicatori in ordine di importanza relativa, per risolvere il problema di classificazione in oggetto, si è dovuto analizzare i risultati acquisiti e filtrare questi ultimi, studiando non solo il relativo significato economico, ma anche l'esame di correlazione svolto in precedenza, con l'obiettivo di selezionare un insieme di variabili in grado di rappresentare correttamente il fenomeno dell'insolvenza di una controparte.

Queste variabili, ritenute in possesso di maggior potere diagnostico, si utilizzano per alimentare il modello di rete neurale, scelto come principale responsabile per risolvere il problema di stima della probabilità di insolvenza.

6.5 Modello di credit scoring

A questo punto, si introduce il cuore pulsante del presente lavoro di tesi, che comprende la progettazione di un modello di rete neurale in grado di gestire il rischio di credito, calcolando la probabilità di insolvenza e classificando le imprese sane e le imprese anomale.

Come spiegato in precedenza, la costruzione di una rete neurale efficace ed efficiente richiede sempre numerose scelte progettuali, alcune delle quali anche molto importanti per la corretta implementazione del modello, oltre che la corretta validazione dello stesso. In questo contesto, a complicare ulteriormente il ragionamento, contribuisce anche l'assenza di regole universalmente valide per l'adeguata selezione di queste ultime, affidando il compito all'esperienza e a regole euristiche, con processi di *trial and error* e *thumb rule*.

A questo proposito, si è deciso di seguire un percorso iterativo, più o meno articolato, di modifica degli iperparametri, nel tentativo di selezionarne i valori più opportuni, che si è concluso, dopo aver testato numerose alternative, arrivando a scegliere un'architettura non estremamente sofisticata, in modo da risolvere al meglio il trade-off tra bias e varianza, insito nella costruzione di qualsiasi modello di intelligenza artificiale.

L'obiettivo è sempre quello di ottenere la definizione di un modello, che risulti, da una parte, sufficientemente complesso, in grado di imparare relazioni difficili e intricate dallo studio dei dati in esame, e, dall'altra, non eccessivamente complesso, in grado di assicurare la capacità di generalizzare prendendo in considerazione nuovi dati.

Tabella 6.1. Indicatori di bilancio in ordine di importanza relativa (Random Forest).

Top	Indicatori (Random Forest)	
	Flag di status S/A - impresa	Flag di status S/A+L - impresa
1°	MS	$\frac{KN}{AT}$
2°	$\ln(AT)$	$\frac{KN}{DT}$
3°	DN	ROE
4°	KCN	$\frac{KN\ tangibile}{DT - LIQ + KN}$
5°	$\frac{RS + RE}{AT}$	$\frac{KN\ tangibile}{DT + KN}$
6°	$\frac{KN}{DT}$	$\frac{KN\ tangibile}{AT}$
7°	$\frac{KN\ tangibile}{AT}$	$ROE\ ante\ imposte$
8°	$\frac{LIQ}{PC}$	LBV
9°	$\frac{AC}{AT}$	$\ln(AT)$
10°	$\frac{LIQ}{AT}$	MS
11°	$\frac{KN\ tangibile}{DT - LIQ + KN}$	$\frac{AC}{AT}$
12°	$\frac{AC}{PC}$	KCN
13°	$\frac{AC - MG}{PC}$	DN
14°	$\frac{KN}{AT}$	$\ln(FN)$
15°	$\frac{OFN}{AT}$	$\frac{AU - (PS - OS)}{AT}$
16°	ROI	$\frac{RS + RE}{AT}$
17°	$\frac{KN\ tangibile}{DT + KN}$	$\frac{EBITDA}{AT}$
18°	$\ln(FN)$	$\frac{OFN}{AT}$
19°	$\frac{FN}{AT}$	$\frac{AC}{PC}$
20°	ROA	$\frac{LIQ}{AT}$

Tabella 6.2. Indicatori di bilancio in ordine di importanza relativa (XGBoost).

Top	Indicatori (XGBoost)	
	Flag di status S/A - impresa	Flag di status S/A+L - impresa
1°	MS	$\frac{KN}{AT}$
2°	$\ln(AT)$	$\frac{KN}{DT}$
3°	DN	ROE
4°	KCN	$\frac{KN\ tangibile}{DT - LIQ + KN}$
5°	$\frac{RS + RE}{AT}$	$\frac{KN\ tangibile}{DT + KN}$
6°	$\frac{KN}{DT}$	$\frac{KN\ tangibile}{AT}$
7°	$\frac{KN\ tangibile}{AT}$	$ROE\ ante\ imposte$
8°	$\frac{LIQ}{PC}$	LBV
9°	$\frac{AC}{AT}$	$\ln(AT)$
10°	$\frac{LIQ}{AT}$	MS
11°	$\frac{KN\ tangibile}{DT - LIQ + KN}$	$\frac{AC}{AT}$
12°	$\frac{AC}{PC}$	KCN
13°	$\frac{AC - MG}{PC}$	DN
14°	$\frac{KN}{AT}$	$\ln(FN)$
15°	$\frac{OFN}{AT}$	$\frac{AU - (PS - OS)}{AT}$
16°	ROI	$\frac{RS + RE}{AT}$
17°	$\frac{KN\ tangibile}{DT + KN}$	$\frac{EBITDA}{AT}$
18°	$\ln(FN)$	$\frac{OFN}{AT}$
19°	$\frac{FN}{AT}$	$\frac{AC}{PC}$
20°	ROA	$\frac{LIQ}{AT}$

6.5.1 Implementazione

In questo senso, è possibile descrivere il codice Python utilizzato per la progettazione di un modello di rete neurale per la stima della probabilità di insolvenza di una controparte.

Come spiegato in precedenza, si riporta ancora una volta l'intero procedimento seguito usando un flag come riferimento, considerando che i passaggi previsti dai due problemi in esame risultano esattamente speculari.

Per migliorare la comprensione, si è deciso di mostrare fin da subito, in modo da evitare inutili ripetizioni, alcune operazioni preliminari, come l'importazione delle librerie, il caricamento del database e la suddivisione di quest'ultimo in training set e testing set.

In primo luogo, si importa l'insieme di librerie usate per lo sviluppo della rete neurale:

```
#library for data handling
import pandas as pd
#library for calculations
import numpy as np
#library for plotting
import seaborn as sns
import matplotlib
from matplotlib import rc
import matplotlib.pyplot as plt
import matplotlib.cm as cm
#library for pre-processing operations
from sklearn import datasets
from sklearn import preprocessing
#library for the Decision Tree classifier
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
#library for the Random Forest classifier
from sklearn.ensemble import RandomForestClassifier
#external package for XGBoost Classifier
import xgboost as xgb
#library for feature selection
from sklearn.feature_selection import SelectFromModel
#library for evaluation tools
from sklearn import metrics
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
#library for operations
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
#library for Artificial Neural Network
import tensorflow as tf
import keras
from keras.models import Sequential
from keras.layers import Dense
```

In secondo luogo, si carica il dataset, dividendo le variabili di input dai flag di output:

```
#loading the dataset
df = pd.read_excel('AIDA_Export_Automotive_Industry.xlsx',"Risultati")
print('dimensionality',df.shape)
#indicatori di bilancio
df_indicatori = df.iloc[:,1:58]
#flag di status S/A - impresa
df_flag1 = df.iloc[:,58:59]
#flag di status S/A+L - impresa
df_flag2 = df.iloc[:,59:60]
```

6.5.2 Rete neurale

Per prima cosa, è necessario suddividere il database in oggetto, prevedendo un testing set, pari a $\frac{2}{3}$ delle osservazioni, che viene ulteriormente scomposto definendo un validation set, sempre con la proporzione $\frac{2}{3}$ e $\frac{1}{3}$, oltre che un training set, pari a $\frac{1}{3}$ delle osservazioni.

In poche parole, si procede a frazionare il database a disposizione in tre parti principali, seguendo quanto precedentemente anticipato, come visibile in figura 6.14:

- training set: $\frac{4}{9}$ del database, pari a 7057 osservazioni;
- validation set: $\frac{2}{9}$ del database, pari a 3529 osservazioni;
- testing set: $\frac{1}{3}$ del database, pari a 5293 osservazioni.

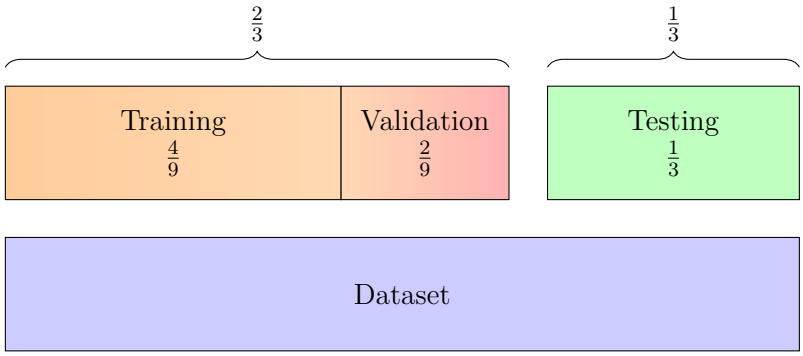


Figura 6.14. Training-Validation-Testing split.

Architettura

Così facendo, si inizia dalla divisione del database in esame come appena spiegato:

```
#splitting dataset into training set and testing set
X_train1,X_test1,y_train1,y_test1 = train_test_split(df_indicatori,df_flag1,
    test_size=1/3,random_state=42)
```

Una volta preparati i dati, si può passare a definire l'architettura della rete neurale, per la quale è necessario effettuare alcune scelte progettuali, riguardanti i seguenti elementi:

- input layer;
- output layer;
- hidden layer.

In altri termini, si deve stabilire il numero di strati che compongono la rete neurale, il numero di neuroni per ogni strato e la funzione di attivazione utilizzata da ogni strato.

Per quanto riguarda l'input layer e l'output layer, la scelta è relativamente facile, osservando che il numero di neuroni risulta fissato dal contesto di applicazione del modello.

Nello specifico, infatti, il numero di neuroni contenuti nell'input layer dipende sempre direttamente dal numero di variabili di input utilizzate dal modello in questione, mentre il numero di neuroni compresi nell'output layer risulta obbligatoriamente pari a uno, considerando che si tratta di un problema di classificazione binaria, in cui si deve distinguere tra imprese sane e imprese anomale.

Per quanto concerne invece la scelta del numero di hidden layer e del relativo numero di neuroni, non esistono regole specifiche e, per questo motivo, si è deciso di sperimentare numerose alternative diverse, nel tentativo di ricercare la soluzione ottimale.

Un ragionamento analogo è valido anche nella scelta delle funzioni di attivazione, con la consapevolezza che l'output layer sia obbligato, per costruzione, a utilizzare sempre Sigmoid Function come funzione di attivazione, la quale consente di modellare gli input ricevuti nel range (0,1), rendendo possibile interpretare il risultato come una probabilità.

```
#defining neural network architecture
model = Sequential()

#the input layer is fixed to have 10 neurons
#the hidden layer 1 is fixed to have n1 neurons with ReLu activation function
model.add(Dense(n1,
                activation='relu',
                input_shape=(10,)))
#the hidden layer 2 is fixed to have n2 neurons with Sigmoid activation function
model.add(Dense(n2,
                activation='sigmoid'))
#the hidden layer 3 is fixed to have n3 neurons with Sigmoid activation function
model.add(Dense(n3,
                activation='sigmoid'))
#the output layer is fixed to have 1 neuron
model.add(Dense(1,
                activation='sigmoid'))
model.summary()
```

In questo caso, si è scelto di utilizzare soltanto una specifica categoria di strati, chiamati *Dense Layer*, contraddistinti dalla capacità di generare una struttura in cui tutti i neuroni di uno strato precedente sono sempre connessi a tutti i neuroni di uno strato successivo.

Addestramento

Dopo aver completato la costruzione del modello di rete neurale, si può passare alla sua compilazione, prima di procedere ulteriormente verso le fasi cruciali di training e testing. Per fare questo, esattamente come in precedenza, è necessario definire alcuni parametri:

- loss function;
- metric;
- optimizer.

In altri termini, si deve definire la funzione di perdita adottata, le diverse metriche di performance impiegate e un optimizer, utilizzabile nel processo di addestramento.

Prima di tutto, dal momento che si deve risolvere un problema di classificazione, il quale consiste nel classificare una generica impresa come sana o anomala, è necessario utilizzare la Binary Cross Entropy come loss function, che si può dimostrare essere la soluzione ottimale per i problemi di classificazione binaria come quello in esame.

In aggiunta, è importante tenere traccia delle diverse metriche di performance rilevanti, tra queste, sicuramente, l'Accuracy, la quale, a differenza della Binary Cross Entropy, che viene usata principalmente durante la fase di training della rete neurale, consente di valutare le performance del modello durante la fase di testing, indicando la frazione di imprese classificate correttamente sul totale, ma anche AUC, Precision, Recall, oltre che True Positive, True Negative, False Positive e False Negative.

In ultima analisi, è interessante definire un optimizer, come *Adam method*, il quale permette di apportare alcuni miglioramenti specifici alla versione originale dell'algoritmo del *gradient descent*, in modo da potenziare e velocizzare l'apprendimento, oltre che evitare l'eventuale convergenza anticipata in punti di massimo o minimo locali.

```
#compiling the neural network
metrics=[tf.keras.metrics.TruePositives(name='TP',thresholds=T),
             tf.keras.metrics.TrueNegatives(name='TN',thresholds=T),
             tf.keras.metrics.FalsePositives(name='FP',thresholds=T),
             tf.keras.metrics.FalseNegatives(name='FN',thresholds=T),
             tf.keras.metrics.AUC(name='AUC',num_thresholds=1000),
             tf.keras.metrics.BinaryAccuracy(name='accuracy',threshold=T),
             tf.keras.metrics.Precision(name='precision',thresholds=T),
             tf.keras.metrics.Recall(name='recall',thresholds=T)]

model.compile(loss='binary_crossentropy',
              optimizer='Adam',
              metrics=metrics)
```

Per passare alla fase di addestramento del modello di rete neurale, è necessario prendere alcune decisioni in merito ad alcuni parametri, essenziali per il buon esito di quest'ultima:

- learning rate;
- epoch;
- batch size;

In particolare, il learning rate è un parametro che governa la modifica delle ponderazioni nei coefficienti della rete neurale dopo il calcolo della loss function per ogni iterazione, epoch è un parametro che indica il numero di iterazioni che il modello svolge durante la fase di apprendimento sul training set e il batch size è un parametro che individua il numero di campioni analizzati per ogni epoca. A questo riguardo, dopo attente valutazioni, si è deciso di impostare i seguenti valori: learning rate = 0,001, epoch = 250, batch size = 32.

A questo punto, si può procedere con l'addestramento del modello di rete neurale, sfruttando l'algoritmo di *error back-propagation*, attraverso la funzione *fit*:

```
#training the neural network
history = model.fit(X, #(X=X_train.to_numpy())
                    Y, #(Y=y_train.to_numpy())
                    epochs=250,
                    validation_split=0.2,
                    verbose=1,
                    shuffle=True)
```

La fase di apprendimento del modello di rete neurale risulta fondamentale, costituendo la parte principale di qualsiasi algoritmo di intelligenza artificiale e, pertanto, richiede alcune sottolineature addizionali in relazioni agli argomenti trasmessi in input al modello.

Validation split indica la frazione del dataset che viene riservata al validation set, il quale viene impiegato per il processo di validazione dei risultati durante la fase di training.

Shuffle consente di mescolare in modo casuale il training set prima della fase di training, non solo garantendo che i dati in input, appartenenti al training set e al validation set, vengano disegnati dalla stessa distribuzione, ma anche evitando che la disposizione delle osservazioni possa influenzare in qualche modo il modello verso un particolare output.

Verbose permette di visualizzare una serie di informazioni dettagliate durante tutta la fase di training, tracciando alcune metriche rilevanti, che risultano fondamentali per valutare l'architettura della rete neurale e, in caso di esito insoddisfacente, modificarla:

- Loss: errore Binary Cross Entropy totale sui training data;
- Val_loss: errore Binary Cross Entropy valutato sui validation data;
- Acc: Accuracy della previsione sui training data;
- Val_acc: Accuracy della previsione valutata sui validation data.

A questo proposito, è possibile eseguire il disegno di queste metriche per ogni epoca:

```
#plotting the BCE loss
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('Learning Curve')
plt.ylabel('Loss cross entropy')
plt.xlabel('Epoch')
plt.legend(['training BCE loss','validation BCE loss'])
plt.show()
#plotting the accuracy
plt.plot(history.history['accuracy'])
plt.plot(history.history['val_accuracy'])
plt.title('Learning Curve')
plt.ylabel('Accuracy')
plt.xlabel('Epoch')
plt.legend(['training accuracy','validation accuracy'])
plt.show()
```

Durante la fase di training, per un modello di rete neurale progettato adeguatamente, è sempre lecito attendersi che la Binary Cross Entropy possa diminuire e, allo stesso tempo, l'Accuracy possa aumentare, livellandosi dopo un certo numero di epoche per convergere verso uno specifico valore di Accuracy finale del modello, come visibile in figura 6.15.

Tuttavia, osservando i grafici e confrontando l'andamento delle due metriche appena considerate, ovvero Binary Cross Entropy e Accuracy, rispettivamente per il training set e il validation set, è possibile effettuare alcune considerazioni generali. Infatti, valutando le prestazioni sul training set e sul validation set, si ottengono alcune indicazioni importanti in merito alla struttura del modello di rete neurale e alla sua corretta implementazione.

Nello specifico, da una parte, nel caso in cui la rete neurale ottenga prestazioni migliori sui validation data rispetto che sui training data, è possibile che si possa verificare un problema di underfitting, dimostrando che il modello non risulta sufficientemente complesso

per modellare correttamente i dati e implicando l'aggiunta di maggiore complessità per risolvere il problema; dall'altra, nel caso in cui la rete neurale ottenga prestazioni migliori sui training data rispetto che sui validation data, è possibile che si possa verificare un problema di overfitting, dimostrando invece che il modello risulta eccessivamente complesso per modellare correttamente i dati e implicando la rimozione di ulteriore complessità per risolvere il problema.

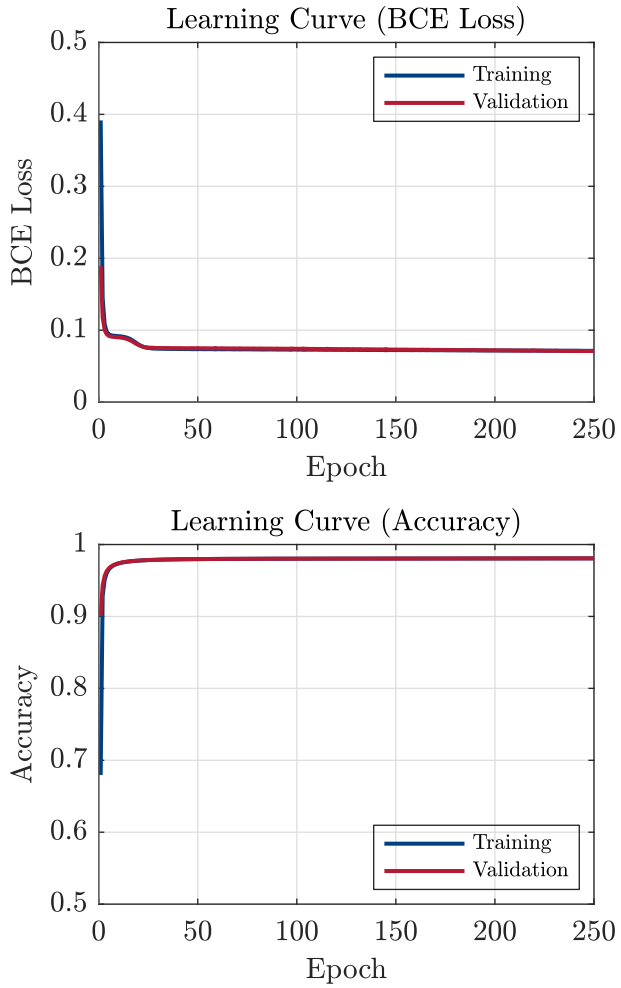


Figura 6.15. Sintesi grafica del processo di addestramento di una rete neurale.

Analisi risultati

Per ultima cosa, è possibile svolgere la fase di testing, applicando il modello di rete neurale per effettuare le previsioni sul testing set, oltre che, eventualmente, per valutare una nuova generica impresa, passando i valori delle variabili assunti da quest'ultima, in modo da assegnarla al gruppo delle imprese sane o delle imprese anomale, con la funzione *predict*:

```
#evaluating the neural network
model_evaluate = model.evaluate(X_test1,y_test1)
print(model_evaluate)
#getting the prediction for the data
model_prediction = model.predict(X_test1)
#prediction using a T=threshold
model_prediction = (model_prediction>T)
print(model_prediction)
```

Così facendo, il valore predetto costituisce un numero compreso tra 0 e 1 e rappresenta la stima della probabilità di insolvenza generata dal modello di rete neurale in esame.

```
#computing the accuracy
print('Accuracy score:',accuracy_score(y_test1,model_prediction))
#visualizing the confusion matrix
mat = confusion_matrix(y_test1,model_prediction)
sns.heatmap(mat.T,square=True,annot=True,cbar=False,fmt=".0f")
plt.xlabel('Actual label')
plt.ylabel('Predicted label');
plt.show()
```

A questo proposito, è necessario discutere la scelta del valore di cut-off da assegnare al modello per classificare tra imprese sane e imprese anomale. In questo caso, infatti, il modello di rete neurale fornisce in output un valore compreso tra 0 e 1, interpretabile come probabilità di insolvenza, ma, per risolvere il problema di classificazione, si deve stabilire un valore soglia, al di sotto del quale le imprese vengono classificate come sane e al di sopra del quale le imprese vengono classificate come anomale.

Anche in questo caso, si è deciso di svolgere numerosi tentativi, partendo dal valore canonico di 0,5 e muovendosi verso valori inferiori, in modo da rendere il ragionamento ancora più selettivo e critico, fino ad arrivare al valore del tasso di default osservato nelle due popolazioni in questione. Per ottenere il valore ideale, si è scelto di seguire il ragionamento economico, piuttosto che affidarsi a un semplice algoritmo di ottimizzazione.

In questo caso, però, è necessario dividere il ragionamento per quanto riguarda il problema di "flag di status S/A - impresa" e di "flag di status S/A+L - impresa". Pertanto, in aggiunta al valore standard di 0,5, nel primo caso, si è scelto di adottare come punto di lavoro ottimale, il valore del tasso di default osservato sul campione di imprese in esame, pari a 13,685%; al contrario, nel secondo caso, si è notato come questa scelta penalizzasse eccessivamente i risultati del modello, in quanto si tratta di un campione decisamente sbilanciato, con un tasso di default pari a 1,852%.

Così, per l'individuazione del punto di lavoro ottimale, si è deciso di testare il modello arrivando a valutare come threshold non solo il tasso di default della prima popolazione, ma anche il tasso di default della seconda popolazione e una media tra questi due valori.

In ultima istanza, si è deciso di calcolare una metrica ulteriore, quale una funzione di costo, che consenta di tenere in considerazione il diverso costo a cui può andare in contro una banca che sviluppa un sistema di risk mangement per la gestione del rischio di credito: FP (errore del I tipo), ovvero classificare un'impresa anomala tra quelle sane e, viceversa, FN (errore del II tipo), ovvero classificare un'impresa sana tra quelle anomale.

Nel primo caso, l'errore commesso produce un costo nascosto, non direttamente osservabile, il quale può essere considerato come un "costo-opportunità", corrispondente al possibile mancato guadagno, a seguito del rifiuto di concedere credito a un'impresa sana, rinunciando al margine di interesse, nonché, eventualmente, alle relative commissioni.

Nel secondo caso, l'errore commesso produce un costo visibile, facilmente osservabile, il quale corrisponde al capitale e agli interessi perduti attraverso l'insolvenza di un'impresa erroneamente classificata come sana, alla quale è stato concesso credito.

Sulla base di queste considerazioni, la banca sceglierà di concedere credito valutando il diverso costo associato alle due tipologie di errore, in modo da misurare il costo atteso dei due errori, pesati per la rispettiva probabilità di verificarsi. In pratica, la banca accetterà di concedere credito a tutte quelle imprese in cui il costo atteso dell'errore di II tipo risulti inferiore al costo atteso dell'errore di I tipo:

$$C(type\ II\ error) \cdot PD < C(type\ I\ error) \cdot (1 - PD)\ . \tag{6.4}$$

In poche parole, si è ritenuto opportuno calcolare e valutare una speciale funzione di costo, estremamente basilare, che tenesse conto del diverso costo associato ai due errori, assegnando due ponderazioni differenti all'errore di I o di II tipo:

$$Cost\ Function = w_{type\ I\ error} \cdot error^{FP} + w_{type\ II\ error} \cdot error^{FN}\ . \tag{6.5}$$

In questo caso, dopo attente riflessioni e valutazioni, calcolando una media dei valori reali più frequenti e aggiungendo uno specifico coefficiente di sicurezza, si è scelto di assegnare una ponderazione pari a $w_{type\ I\ error} = 1$ e $w_{type\ II\ error} = 100$, finendo per distanziare gli errori in modo estremamente cautelativo con due ordini di grandezza, in modo da effettuare un calcolo, sicuramente grossolano e approssimativo, ma utile per avere alcune intuizioni interessanti sul problema in questione.

6.5.3 Commento

Il modello di rete neurale appena presentato consente di gestire il rischio di credito.

Per commentare i risultati ottenuti, è possibile osservare i report tecnici presentati nella tabella 6.3 e 6.4 per il "flag di status S/A - impresa" e, in modo del tutto parallelo, nella tabella 6.5 e 6.6 per il "flag di status S/A+L - impresa", che riassumono una selezione dei principali modelli costruiti, ottenuti rispettivamente analizzando diversi sottoinsiemi di variabili, alcuni derivanti dall'analisi di correlazione e altri derivanti dal processo di feature selection, svolto applicando il modello di albero decisionale Random Forest e XGBoost.

Per ciascun modello sviluppato, è possibile verificare come sia stato dedicato uno spazio specifico all'interno della tabella: nella parte iniziale, sono state indicate le diverse scelte progettuali effettuate per la definizione dell'architettura, mentre, nella parte finale, sono state precisate le varie metriche di performance ottenute nella fase di training e validation, riservando uno spazio conclusivo alla fase di testing, in cui è stata calcolata l'Accuracy.

In generale, è possibile constatare come tutti i modelli sviluppati consentano di ottenere ottimi risultati, con valori estremamente positivi di Loss, AUC e Accuracy, conseguiti sia nella fase di training che nella fase di validation, permettendo, complessivamente, di raggiungere ottime prestazioni nella fase di testing, con valori di Accuracy molto elevati.

Tuttavia, è possibile mettere in evidenza come i migliori modelli, in grado di ottenere sistematicamente le migliori performance al variare dei diversi contesti di applicazione, risultino essere il modello $1A^{RF}$ per quanto riguarda il "flag di status S/A - impresa" e, allo stesso modo, il modello $1F^{RF}$ per quanto riguarda il "flag di status S/A+L - impresa".

Tabella 6.3. Report tecnico: flag di status S/A - impresa (analisi di correlazione).

Modello	Input Layer		Hidden Layer			Output Layer		Learning rate	Epoch	Batch size	Threshold	Training			Validation			Testing	
	Neuron	Activation Function	Layer	Neuron	Activation Function	Neuron	Activation Function					Loss	AUC	Accuracy	Loss	AUC	Accuracy	Accuracy	Cost
$1A^C$	10	ReLU	2	8|8	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0699	0,8461	0,9807	0,0666	0,8461	0,9807	0,9818	8016
$2A^C$	10	Sigmoid	2	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0667	0,8245	0,9818	0,0750	0,8246	0,9818	0,9785	10707
$3A^C$	10	ReLU	2	8|8	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,0671	0,8559	0,9730	0,0709	0,8559	0,9730	0,9724	7274
$4A^C$	10	ReLU	2	8|8	ReLU	1	Sigmoid	0,001	250	32	0,07768	0,0635	0,8553	0,9557	0,0771	0,8554	0,9557	0,9569	5970
$5A^C$	10	ReLU	2	8|8	ReLU	1	Sigmoid	0,001	250	32	0,01852	0,0665	0,8635	0,7871	0,0651	0,8635	0,7872	0,7876	3500
$1B^C$	10	ReLU	2	6|6	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0645	0,8388	0,9827	0,0834	0,8388	0,9827	0,9790	11100
$2B^C$	10	Sigmoid	2	6|6	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0692	0,8430	0,9822	0,0521	0,8430	0,9822	0,9796	10701
$3B^C$	10	ReLU	2	6|6	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,0658	0,8690	0,9714	0,0811	0,8690	0,9714	0,9766	6361
$4B^C$	10	ReLU	2	6|6	ReLU	1	Sigmoid	0,001	250	32	0,07768	0,0651	0,8721	0,9557	0,0730	0,8722	0,9557	0,9445	5541
$5B^C$	10	ReLU	2	6|6	ReLU	1	Sigmoid	0,001	250	32	0,01852	0,0692	0,8558	0,7725	0,0630	0,8559	0,7725	0,7663	4207
$1C^C$	10	ReLU	3	12|12|12	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0519	0,9014	0,9820	0,0736	0,9015	0,9820	0,9824	8409
$2C^C$	10	Sigmoid	3	12|12|12	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0675	0,8226	0,9819	0,0789	0,8227	0,9819	0,9805	9805
$3C^C$	10	ReLU	3	12|12|12	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,0539	0,9081	0,9707	0,0753	0,9081	0,9707	0,9703	5899
$4C^C$	10	ReLU	3	12|12|12	ReLU	1	Sigmoid	0,001	250	32	0,07768	0,0574	0,8864	0,9548	0,0950	0,8864	0,9548	0,9479	5028
$5C^C$	10	ReLU	3	12|12|12	ReLU	1	Sigmoid	0,001	250	32	0,01852	0,0540	0,9069	0,8130	0,0712	0,9070	0,8131	0,7768	3662
$1D^C$	10	ReLU	3	7|7|7	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0610	0,8730	0,9830	0,0694	0,8730	0,9830	0,9794	9811
$2D^C$	10	Sigmoid	3	7|7|7	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0698	0,8209	0,9811	0,0724	0,8209	0,9811	0,9805	10300
$3D^C$	10	ReLU	3	7|7|7	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,0710	0,8422	0,9726	0,0666	0,8422	0,9726	0,9760	6562
$4D^C$	10	ReLU	3	7|7|7	ReLU	1	Sigmoid	0,001	250	32	0,07768	0,0611	0,8847	0,9529	0,0816	0,8848	0,9529	0,9590	5860
$5D^C$	10	ReLU	3	7|7|7	ReLU	1	Sigmoid	0,001	250	32	0,01852	0,0635	0,8803	0,7854	0,0756	0,8803	0,7855	0,7995	3536
$1E^C$	10	ReLU	3	8|3|12	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0625	0,8763	0,9825	0,0658	0,8763	0,9825	0,9796	8721
$2E^C$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0654	0,8159	0,9817	0,0732	0,8160	0,9817	0,9775	10910
$3E^C$	10	ReLU	3	8|3|12	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,0628	0,8619	0,9755	0,0561	0,8619	0,9755	0,9724	7670
$4E^C$	10	ReLU	3	8|3|12	ReLU	1	Sigmoid	0,001	250	32	0,07768	0,0612	0,8752	0,9534	0,0861	0,8752	0,9534	0,9344	5396
$5E^C$	10	ReLU	3	8|3|12	ReLU	1	Sigmoid	0,001	250	32	0,01852	0,0564	0,8914	0,8014	0,0818	0,8914	0,8014	0,7869	3999
$1F^C$	10	ReLU	3	6|9|9|3	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0624	0,8656	0,9817	0,0933	0,8656	0,9817	0,9813	9504
$2F^C$	10	Sigmoid	3	6|9|9|3	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0718	0,7986	0,9800	0,0803	0,7987	0,9800	0,9822	9400
$3F^C$	10	ReLU	3	6|9|9|3	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,0589	0,8748	0,9728	0,0869	0,8748	0,9728	0,9758	7157
$4F^C$	10	ReLU	3	6|9|9|3	ReLU	1	Sigmoid	0,001	250	32	0,07768	0,0644	0,8980	0,9469	0,0636	0,8980	0,9469	0,9556	5185
$5F^C$	10	ReLU	3	6|9|9|3	ReLU	1	Sigmoid	0,001	250	32	0,01852	0,0592	0,8617	0,7823	0,0795	0,8617	0,7823	0,7859	4004

Top 10 variabili correlazione: KNT/AT, MS, LBV, DT/FN, DT/VA, DT/EBITDA, ROE, DN, OFN/EBITDA, FN/AT

Tabella 6.4. Report tecnico: flag di status S/A+L - impresa (analisi di correlazione).

Modello	Input Layer		Hidden Layer			Output Layer		Learning rate	Epoch	Batch size	Threshold	Training			Validation			Testing	
	Neuron	Activation Function	Layer	Neuron	Activation Function	Neuron	Activation Function					Loss	AUC	Accuracy	Loss	AUC	Accuracy	Accuracy	Cost
$1A^C$	10	ReLU	2	8|8	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,3013	0,8078	0,8823	0,3380	0,8079	0,8823	0,8866	48615
$2A^C$	10	Sigmoid	2	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3062	0,8062	0,8848	0,3042	0,8062	0,8848	0,8795	53900
$3A^C$	10	ReLU	2	8|8	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,3029	0,8147	0,7412	0,3180	0,8147	0,7412	0,7425	10183
$1B^C$	10	ReLU	2	6|6	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2995	0,8186	0,8852	0,3142	0,8186	0,8852	0,8787	53805
$2B^C$	10	Sigmoid	2	6|6	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3099	0,7999	0,8813	0,3083	0,7999	0,8813	0,8821	54183
$3B^C$	10	ReLU	2	6|6	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,2994	0,8164	0,7534	0,3132	0,8165	0,7534	0,7533	22789
$1C^C$	10	ReLU	3	12|12|12	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2891	0,8353	0,8862	0,3025	0,8353	0,8862	0,8795	50039
$2C^C$	10	Sigmoid	3	12|12|12	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,2980	0,8103	0,8836	0,3337	0,8103	0,8837	0,8827	50714
$3C^C$	10	ReLU	3	12|12|12	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,2889	0,8238	0,7695	0,3382	0,8238	0,7696	0,7682	24393
$1D^C$	10	ReLU	3	7|7|7	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2979	0,8212	0,8829	0,3115	0,8212	0,8829	0,8796	50137
$2D^C$	10	Sigmoid	3	7|7|7	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3073	0,8066	0,8823	0,3221	0,8067	0,8823	0,8823	53687
$3D^C$	10	ReLU	3	7|7|7	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,2932	0,8207	0,7735	0,3082	0,8207	0,7735	0,7527	24079
$1E^C$	10	ReLU	3	8|3|12	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,3104	0,8116	0,8815	0,3066	0,8116	0,8815	0,8876	48412
$2E^C$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3050	0,7991	0,8813	0,3370	0,7991	0,8813	0,8832	53583
$3E^C$	10	ReLU	3	8|3|12	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,3076	0,8197	0,7665	0,3156	0,8197	0,7665	0,7578	22171
$1F^C$	10	ReLU	3	6|9|9|3	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2997	0,8143	0,8832	0,3339	0,8143	0,8832	0,8795	54593
$2F^C$	10	Sigmoid	3	6|9|9|3	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3143	0,8007	0,8778	0,2997	0,8007	0,8778	0,8851	50504
$3F^C$	10	ReLU	3	6|9|9|3	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,3008	0,8170	0,7464	0,3090	0,8170	0,7464	0,7439	19310
$1G^C$	10	ReLU	3	8|10|8	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2921	0,8256	0,8849	0,3144	0,8256	0,8849	0,8793	50733
$2G^C$	10	Sigmoid	3	8|10|8	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3134	0,7942	0,8826	0,3004	0,7942	0,8826	0,8810	53793
$3G^C$	10	ReLU	3	8|10|8	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,2963	0,8161	0,7611	0,3215	0,8161	0,7611	0,7561	21784
$1H^C$	10	ReLU	3	15|7|15|5	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2826	0,8300	0,8859	0,3317	0,8301	0,8859	0,8821	48738
$2H^C$	10	Sigmoid	3	15|7|15|5	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3114	0,7993	0,8823	0,3191	0,7993	0,8823	0,8832	53583
$3H^C$	10	ReLU	3	15|7|15|5	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,2943	0,8243	0,7577	0,3324	0,8244	0,7576	0,7279	17379
$1I^C$	10	ReLU	3	10|3|10	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,3012	0,8276	0,8807	0,3111	0,8276	0,8807	0,8795	48356
$2I^C$	10	Sigmoid	3	10|3|10	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3054	0,8030	0,8837	0,3231	0,8031	0,8837	0,8823	53390
$3I^C$	10	ReLU	3	10|3|10	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,3040	0,8177	0,7525	0,3195	0,8177	0,7525	0,7476	20740
$1L^C$	10	ReLU	3	5|7|2	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,3064	0,8097	0,8858	0,2880	0,8098	0,8858	0,8802	54391
$2L^C$	10	Sigmoid	3	5|7|2	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3115	0,7864	0,8781	0,3087	0,7865	0,8781	0,8842	51004
$3L^C$	10	ReLU	3	5|7|2	ReLU	1	Sigmoid	0,001	250	32	0,13685	0,3094	0,8042	0,7503	0,3374	0,8042	0,7503	0,7712	21803
Top 10 variabili correlazione: DT/VA, ROE, DT/EBITDA, KNT/AT, LBV, AS/CO, AU/AT, ROI, FN/AT, OFN/EBITDA																			

Tabella 6.5. Report tecnico: flag di status S/A - impresa (Random Forest e XGBoost).

Modello	Input Layer		Hidden Layer			Output Layer		Learning rate	Epoch	Batch size	Threshold	Training			Validation			Testing	
	Neuron	Activation Function	Layer	Neuron	Activation Function	Neuron	Activation Function					Loss	AUC	Accuracy	Loss	AUC	Accuracy	Accuracy	Cost
$1A^{RF}$	10	ReLU	2	8|8	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0540	0,8774	0,9840	0,0702	0,8774	0,9840	0,9809	8615
$2A^{RF}$	10	Sigmoid	2	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0704	0,8090	0,9814	0,0801	0,8090	0,9814	0,9822	8311
$3A^{RF}$	10	Sigmoid	2	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,0705	0,8019	0,9722	0,0719	0,8019	0,9722	0,9730	6380
$4A^{RF}$	10	Sigmoid	2	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,07768	0,0697	0,8184	0,9569	0,0589	0,8184	0,9569	0,9611	6344
$5A^{RF}$	10	Sigmoid	2	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,01852	0,0711	0,8134	0,7999	0,0740	0,8135	0,8000	0,8479	3874
$1D^{RF}$	10	ReLU	2	7|7|7	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0659	0,8630	0,9836	0,0477	0,8630	0,9836	0,9800	9412
$2D^{RF}$	10	Sigmoid	2	7|7|7	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0679	0,7991	0,9807	0,0796	0,7991	0,9807	0,9804	9509
$3D^{RF}$	10	Sigmoid	2	7|7|7	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,0701	0,8118	0,9751	0,0621	0,8118	0,9751	0,9713	7280
$4D^{RF}$	10	Sigmoid	2	7|7|7	Sigmoid	1	Sigmoid	0,001	250	32	0,07768	0,0686	0,8164	0,9527	0,0631	0,8165	0,9528	0,9605	6050
$5D^{RF}$	10	Sigmoid	2	7|7|7	Sigmoid	1	Sigmoid	0,001	250	32	0,01852	0,0737	0,7933	0,7722	0,0791	0,7933	0,7723	0,8029	3320
$1E^{RF}$	10	ReLU	3	8|3|12	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0502	0,8585	0,9855	0,0700	0,8586	0,9855	0,9813	8118
$2E^{RF}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0729	0,8181	0,9812	0,0746	0,8181	0,9812	0,9821	9509
$3E^{RF}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,0689	0,8350	0,9725	0,0602	0,8351	0,9726	0,9662	7901
$4E^{RF}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,07768	0,0687	0,8107	0,9540	0,0815	0,8107	0,9540	0,9633	6035
$5E^{RF}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,01852	0,0726	0,8072	0,8037	0,0649	0,8072	0,8038	0,8256	4487
$1A^{XG}$	10	ReLU	3	8|8	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0677	0,8715	0,9814	0,0649	0,8715	0,9814	0,9807	8814
$2A^{XG}$	10	Sigmoid	3	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0716	0,8233	0,9816	0,0701	0,8234	0,9816	0,9824	8409
$3A^{XG}$	10	Sigmoid	3	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,0686	0,8115	0,9727	0,0769	0,8116	0,9727	0,9777	7444
$4A^{XG}$	10	Sigmoid	3	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,07768	0,0745	0,8333	0,9637	0,0703	0,8333	0,9637	0,9637	7617
$5A^{XG}$	10	Sigmoid	3	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,01852	0,0735	0,8197	0,7366	0,0551	0,8197	0,7366	0,7568	3762
$1D^{XG}$	10	ReLU	3	7|7|7	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0676	0,8493	0,9817	0,0741	0,8493	0,9817	0,9834	7018
$2D^{XG}$	10	Sigmoid	3	7|7|7	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0754	0,8157	0,9803	0,0687	0,8158	0,9803	0,9832	7217
$3D^{XG}$	10	Sigmoid	3	7|7|7	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,0718	0,8326	0,9719	0,0681	0,8326	0,9719	0,9756	6960
$4D^{XG}$	10	Sigmoid	3	7|7|7	Sigmoid	1	Sigmoid	0,001	250	32	0,07768	0,0655	0,8085	0,9580	0,0671	0,8086	0,9580	0,9673	8093
$5D^{XG}$	10	Sigmoid	3	7|7|7	Sigmoid	1	Sigmoid	0,001	250	32	0,01852	0,0722	0,8053	0,7726	0,0770	0,8054	0,7726	0,7672	2915
$1E^{XG}$	10	ReLU	3	8|3|12	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,0646	0,8582	0,9818	0,0702	0,8582	0,9818	0,9830	8406
$2E^{XG}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,0732	0,8305	0,9805	0,0705	0,8305	0,9805	0,9834	7414
$3E^{XG}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,0718	0,8400	0,9693	0,0519	0,8400	0,9693	0,9739	7365
$4E^{XG}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,07768	0,0713	0,8160	0,9545	0,0592	0,8160	0,9545	0,9635	6727
$5E^{XG}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,01852	0,0720	0,8286	0,7533	0,0785	0,8287	0,7534	0,7718	3683

Top 10 variabili Random Forest: MS, ln(FN), KCN, DN, RS+RE/AT, KN/DT, AC/AT, KN/AT, LIQ/PC, LIQ/AT
Top 10 variabili XGBoost: KNT/DT+KN, DT/FN, MS, KCN, EBITDA-OF/AT, ROE, ln(FN), OFN/EBIT, AC/AT, AC/PC

Tabella 6.6. Report tecnico: flag di status S/A+L - impresa (Random Forest e XGBoost).

Modello	Input Layer		Hidden Layer			Output Layer		Learning rate	Epoch	Batch size	Threshold	Training			Validation			Testing	
	Neuron	Activation Function	Layer	Neuron	Activation Function	Neuron	Activation Function					Loss	AUC	Accuracy	Loss	AUC	Accuracy	Accuracy	Cost
$1A^{RF}$	10	ReLU	2	8|8	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2877	0,8354	0,8851	0,3187	0,8355	0,8851	0,8914	47204
$2A^{RF}$	10	Sigmoid	2	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3008	0,8249	0,8823	0,2962	0,8249	0,8824	0,8900	47805
$3A^{RF}$	10	Sigmoid	2	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,3092	0,8136	0,7474	0,2959	0,8137	0,7474	0,7614	21360
$1C^{RF}$	10	ReLU	2	12|12|12	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2764	0,8449	0,8869	0,3209	0,8449	0,8869	0,8927	47197
$2C^{RF}$	10	Sigmoid	2	12|12|12	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3008	0,8214	0,8822	0,3157	0,8214	0,8822	0,8919	47300
$3C^{RF}$	10	Sigmoid	2	12|12|12	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,2979	0,8226	0,7543	0,3251	0,8226	0,7543	0,7536	18233
$1E^{RF}$	10	ReLU	3	8|3|12	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2950	0,8377	0,8847	0,3013	0,8377	0,8847	0,8874	45938
$2E^{RF}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,2980	0,8103	0,8836	0,3337	0,8103	0,8837	0,8827	49521
$3E^{RF}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,3023	0,8241	0,7391	0,2859	0,8241	0,7391	0,7644	22928
$1F^{RF}$	10	ReLU	3	6|9|9|3	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2913	0,8353	0,8886	0,3011	0,8353	0,8886	0,8840	48332
$2F^{RF}$	10	Sigmoid	3	6|9|9|3	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3013	0,8115	0,8811	0,3156	0,8115	0,8811	0,8927	47989
$3F^{RF}$	10	Sigmoid	3	6|9|9|3	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,2992	0,8191	0,7471	0,2983	0,8192	0,7471	0,8164	29187
$1L^{RF}$	10	ReLU	3	5|7|2	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2968	0,8317	0,8865	0,3048	0,8317	0,8865	0,8876	49105
$2L^{RF}$	10	Sigmoid	3	5|7|2	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,2997	0,8160	0,8829	0,2967	0,8160	0,8829	0,8870	49306
$3L^{RF}$	10	Sigmoid	3	5|7|2	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,2952	0,8140	0,7564	0,3021	0,8140	0,7564	0,7708	22498
$1A^{XG}$	10	ReLU	3	8|8	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2869	0,8412	0,8864	0,3064	0,8413	0,8864	0,8849	50406
$2A^{XG}$	10	Sigmoid	3	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,2959	0,8275	0,8814	0,3147	0,8275	0,8815	0,8899	49390
$3A^{XG}$	10	Sigmoid	3	8|8	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,2938	0,8269	0,7546	0,3188	0,8269	0,7546	0,7570	19205
$1C^{XG}$	10	ReLU	3	12|12|12	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2762	0,8507	0,8878	0,3016	0,8507	0,8878	0,8855	46245
$2C^{XG}$	10	Sigmoid	3	12|12|12	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,2924	0,8310	0,8846	0,2972	0,8310	0,8846	0,8857	52184
$3C^{XG}$	10	Sigmoid	3	12|12|12	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,2964	0,8300	0,7489	0,2935	0,8300	0,7489	0,7565	19802
$1E^{XG}$	10	ReLU	3	8|3|12	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2990	0,8292	0,8862	0,3166	0,8292	0,8862	0,8866	53070
$2E^{XG}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3009	0,8282	0,8818	0,2743	0,8282	0,8818	0,8865	49903
$3E^{XG}$	10	Sigmoid	3	8|3|12	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,2967	0,8220	0,7450	0,3079	0,8221	0,7450	0,7716	19920
$1F^{XG}$	10	ReLU	3	6|9|9|3	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2821	0,8437	0,8881	0,3150	0,8437	0,8882	0,8834	51404
$2F^{XG}$	10	Sigmoid	3	6|9|9|3	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,2986	0,8235	0,8812	0,3024	0,8235	0,8812	0,8887	48307
$3F^{XG}$	10	Sigmoid	3	6|9|9|3	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,2950	0,8195	0,7429	0,2976	0,8195	0,7429	0,7557	20796
$1L^{XG}$	10	ReLU	3	5|7|2	ReLU	1	Sigmoid	0,001	250	32	0,50000	0,2842	0,8385	0,8876	0,3016	0,8385	0,8876	0,8831	53089
$2L^{XG}$	10	Sigmoid	3	5|7|2	Sigmoid	1	Sigmoid	0,001	250	32	0,50000	0,3033	0,8212	0,8788	0,3098	0,8212	0,8788	0,8846	46989
$3L^{XG}$	10	Sigmoid	3	5|7|2	Sigmoid	1	Sigmoid	0,001	250	32	0,13685	0,2916	0,8273	0,7533	0,2910	0,8273	0,7533	0,7739	25155

Top 10 variabili Random Forest: KN/DT, KN/AT, ROE, LBV, ln(FN), MS, KCN, AU-(PS-OS)/AT, DN, RS+RE/AT
Top 10 variabili XGBoost: KN/AT, KN/DT, AU-(PS-OS)/AT, ln(FN), ROE, CB/CO, DT/EBITDA, DT/FN, AMM/CP, WL/CP

6.5.4 Confronto con modelli statistici

In ultima istanza, si è deciso di mettere alla prova il processo di risk management appena sviluppato, che prevede una prima fase di feature selection, eseguita tramite il modello di albero decisionale e, più precisamente, il modello Random Forest e XGBoost, oltre che una seconda fase di analisi del rischio di credito, impostata con il modello di rete neurale, confrontando i risultati con due modelli statistici essenziali, utilizzati con ottimi risultati nel corso della storia dal settore bancario: l'analisi discriminante e la regressione logistica.

Prima di mostrare i risultati ottenuti, è necessario descrivere il codice Python utilizzato per l'implementazione dei due modelli statistici appena elencati, limitandone i commenti al minimo indispensabile in modo da evitare di effettuare inutili ripetizioni rispetto a quanto già discusso in precedenza, oltre che appesantire ulteriormente la trattazione.

In questo senso, lo sviluppo dei modelli di analisi discriminante e regressione logistica segue i medesimi passaggi logici, facilitandone enormemente il compito di descrizione, prevedendo comunque l'esigenza di precisare alcuni elementi ulteriori.

Il primo passo, come sempre, è l'importazione delle librerie necessarie, il caricamento del database di interesse e la suddivisione di quest'ultimo in training set e testing set:

```
#library for data handling
import pandas as pd
#library for calculations
import numpy as np
#library for plotting
import seaborn as sns
import matplotlib
from matplotlib import rc
import matplotlib.pyplot as plt
import matplotlib.cm as cm
#library for pre-processing operations
from sklearn import datasets
from sklearn import preprocessing
#library for evaluation tools
from sklearn import metrics
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import roc_auc_score
#library for operations
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
#library for Linear Discriminant Analysis
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
#library for Logistic Regression (using statsmodel)
import statsmodels.api as sm
from scipy import stats
from statsmodels.formula.api import logit
#library for Logistic Regression (using scikit-learn)
from sklearn.linear_model import LogisticRegression
```

```
#loading the dataset
df = pd.read_excel('AIDA_Export_Automotive_Industry.xlsx',"Risultati")
print('dimensionality',df.shape)
#indicatori di bilancio
df_indicatori = df.iloc[:,1:58]
#flag di status S/A - impresa
df_flag1 = df.iloc[:,58:59]
#flag di status S/A+L - impresa
df_flag2 = df.iloc[:,59:60]
```

```
#split dataset into training set and testing set
X_train1,X_test1,y_train1,y_test1 = train_test_split(df_indicatori,df_flag1,
    test_size=0.3,random_state=42)
```

In seguito, è possibile costruire l'analisi discriminante lineare e la regressione logistica, rispettivamente grazie alle funzioni *LinearDiscrimantAnalysis* e *Logit*, che consentono di stimare i coefficienti del modello, mostrando in output i risultati ottenuti:

```
#fitting the LDA model (scikit-learn)
model = LinearDiscriminantAnalysis()
model.fit(X_train1,np.array(y_train1).ravel())
```

```
#fitting the Logit model (statsmodel)
logit = sm.Logit(y_train1,X_train1)
result = logit.fit(cov_type="hc1")
print(result.summary())
```

A questo punto, è fondamentale analizzare i risultati prodotti dai due modelli, in modo da valutare, per i rispettivi coefficienti, la significatività e la coerenza del segno, prima di procedere alla revisione delle prestazioni sul problema oggetto della discussione.

In questo caso, è necessario costruire un ciclo per elaborare il credit score, calcolare la probabilità di insolvenza e determinare la classificazione in base alle diverse threshold:

```
#initializing the variables
nA_test = 0
logit_test = [0]*n_test
PD_test = [0]*n_test
PS_test = [0]*n_test
y_pred_test = [0]*n_test

#setting manually the threshold
#threshold=T

for i in range(n_test):
    for j in range (n_params):
        logit_test[i] = logit_test[i]+result.params[j]*X_test1.iloc[i,j]
        PD_test[i] = 1/(1+np.exp(-logit_test[i]))
        PS_test[i] = 1-PD_test[i]
        if PD_test[i]>=threshold:
            y_pred_test[i] = 1
        else:
            y_pred_test[i] = 0
```

```
#initializing the variables
nA_test = 0
logit_test = [0]*n_test
PD_test = [0]*n_test
PS_test = [0]*n_test
y_pred_test = [0]*n_test

#setting manually the threshold
#threshold=T

for i in range(n_test):
    for j in range (n_params):
        LDA_test[i] = LDA_test[i]+params[j]*X_test1.iloc[i,j]
        PD_test[i] = 1/(1+np.exp(-LDA_test[i]))
        PS_test[i] = 1-PD_test[i]
        if PD_test[i]>=threshold:
            y_pred_test[i] = 1
        else:
            y_pred_test[i] = 0
```

Inoltre, è possibile osservare le performance generate dai due modelli, principalmente attraverso la valutazione dell'Accuracy e della matrice di confusione:

```
#computing the accuracy
print('Accuracy score:',accuracy_score(y_test1,y_pred_test))
#visualizing the confusion matrix
mat = confusion_matrix(y_test1,y_pred_test)
sns.heatmap(mat.T,square=True,annot=True,cbar=False,fmt=".0f")
plt.xlabel('Actual label')
plt.ylabel('Predicted label');
plt.show()
```

Infine, è possibile calcolare anche tutte le altre metriche di performance, seguendo un approccio esattamente identico ai modelli sviluppati in precedenza:

```
#initializing the metrics
TP = 0
TN = 0
FP = 0
FN = 0

#computing TP,FP,TN,FN
for i in range(n_test):
    if y_test.iloc[i,0]==1 and y_pred_test.iloc[i,0]==1:
        TP += 1
    if y_test.iloc[i,0]==0 and y_pred_test.iloc[i,0]==0:
        TN += 1
    if y_test.iloc[i,0]==0 and y_pred_test.iloc[i,0]==1:
        FP += 1
    if y_test.iloc[i,0]==1 and y_pred_test.iloc[i,0]==0:
        FN += 1

print('TP:',TP,'TN:',TN,'FP:',FP,'FN:',FN)

#printing the performance metrics
Accuracy = (TP+TN)/(TP+TN+FP+FN)
print('Accuracy:',float("{:.10f}".format(Accuracy)))

if TP+FP!=0:
    Precision = TP/(TP+FP)
    print('Precision:',float("{:.10f}".format(Precision)))
if TP+TN!=0:
    Recall = TP/(TP+FN)
    print('Recall:',float("{:.10f}".format(Recall)))
```

Per la costruzione dei diversi modelli di analisi discriminante e regressione logistica, si è deciso di utilizzare come punto di partenza le variabili più significative, emerse in seguito all'analisi di correlazione, come precedentemente spiegato per la costruzione dei modelli di intelligenza artificiale, ponendo l'attenzione su due aspetti fondamentali per la costruzione di qualsiasi modello statistico accettabile, ovvero la significatività dei coefficienti, oltre che la coerenza del loro segno in relazione al significato economico della variabile associata.

In questo senso, esplorando numerose possibilità per la definizione del modello finale, si è reso necessario eliminare tutte quelle variabili che, seppur importanti, non presentassero un coefficiente significativo[1] oppure un segno coerente con il proprio significato economico, in modo da definire un modello completamente accettabile dal punto di vista teorico.

[1]In questo lavoro di tesi, si è ipotizzato di lavorare con un livello di significatività del 5%.

Come risultato, si è potuto costruire i seguenti modelli:

- analisi discriminante:
  - Flag di status S/A: $\frac{KNT}{AT}$, MS, LBV, $\frac{DT}{FN}$, $\frac{DT}{VA}$, DN, $\frac{OFN}{EBITDA}$, $\frac{FN}{AT}$.
  - Flag di status S/A+L: $\frac{DT}{VA}$, ROE, $\frac{DT}{EBITDA}$, $\frac{KNT}{AT}$, LBV, $\frac{AS}{CO}$, ROI, $\frac{FN}{AT}$, $\frac{OFN}{EBITDA}$.
- regressione logistica:
  - Flag di status S/A: $\frac{KNT}{AT}$, MS, LBV, $\frac{DT}{FN}$, $\frac{FN}{AT}$.
  - Flag di status S/A+L: ROE, $\frac{DT}{EBITDA}$, $\frac{KNT}{AT}$, LBV, $\frac{AS}{CO}$, $\frac{FN}{AT}$, $\frac{OFN}{EBITDA}$.

Tabella 6.7. Confronto dei risultati con analisi discriminante e regressione logistica.

	Flag	Threshold	Performance Metrics					
			Accuracy	TP	TN	FP	FN	Cost
Modello AI $1A^{RF}$ e $1F^{RF}$	Flag di status S/A - impresa	$T = 0{,}50000$	0,9822	10	5189	11	83	8311
		$T = 0{,}13685$	0,9730	32	5118	80	63	6380
		$T = 0{,}07768$	0,9611	48	5039	144	62	6344
		$T = 0{,}01852$	0,8479	63	4425	774	31	3874
	Flag di status S/A+L - impresa	$T = 0{,}50000$	0,8927	242	4483	89	479	47989
		$T = 0{,}13685$	0,8164	446	3875	687	285	29187
Analisi Discriminante	Flag di status S/A - impresa	$T = 0{,}50000$	0,9788	9	5172	25	87	8725
		$T = 0{,}13685$	0,9618	39	5052	147	55	5647
		$T = 0{,}07768$	0,9363	49	4907	285	52	5485
		$T = 0{,}01852$	0,7954	56	4154	1054	29	3954
	Flag di status S/A+L - impresa	$T = 0{,}50000$	0,8734	275	4348	191	479	48091
		$T = 0{,}13685$	0,7863	451	3711	855	276	28455
Regressione Logistica	Flag di status S/A - impresa	$T = 0{,}50000$	0,9804	0	5189	0	104	10400
		$T = 0{,}13685$	0,9715	23	5119	76	75	7576
		$T = 0{,}07768$	0,9526	40	5002	209	42	4409
		$T = 0{,}01852$	0,7944	79	4126	1056	32	4256
	Flag di status S/A+L - impresa	$T = 0{,}50000$	0,8740	169	4457	122	545	54622
		$T = 0{,}13685$	0,7780	484	3634	967	208	21767

Come è possibile osservare nella tabella 6.7, che mette a confronto i principali modelli di intelligenza artificiale con i modelli statistici di analisi discriminante e regressione logistica, i risultati ottenuti sono, ancora una volta, molto interessanti, non solo perché permettono di ottenere livelli di Accuracy molto elevati in senso assoluto, sia analizzando nel dettaglio il caso di "flag di status S/A - impresa" che il caso di "flag di status S/A+L - impresa",

ma perché permettono di arrivare alle medesime considerazioni anche in senso relativo, confrontandone le performance con i principali modelli statistici.

In particolare, si conclude come il modello di intelligenza artificiale di riferimento, designato selezionando, tra i diversi modelli testati, il migliore in base ai risultati raggiunti, consenta di raggiungere, rispetto al modello di analisi discriminante e regressione logistica, livelli di Accuracy superiori praticamente in ogni tipologia di applicazione, modificando sia il flag di interesse che le possibili threshold (figura 6.16, 6.17 e 6.18).

Allo stesso tempo, è comunque doveroso sottolineare come le performance raggiunte dai modelli statistici risultino ugualmente ottime e non si discostino molto da quelle raggiunte dai modelli di intelligenza artificiale, dimostrando di essere strumenti sempre validi e applicabili, con la regressione logistica che, rispetto all'analisi discriminante, riesce tendenzialmente a ottenere risultati migliori, anche se in misura limitata.

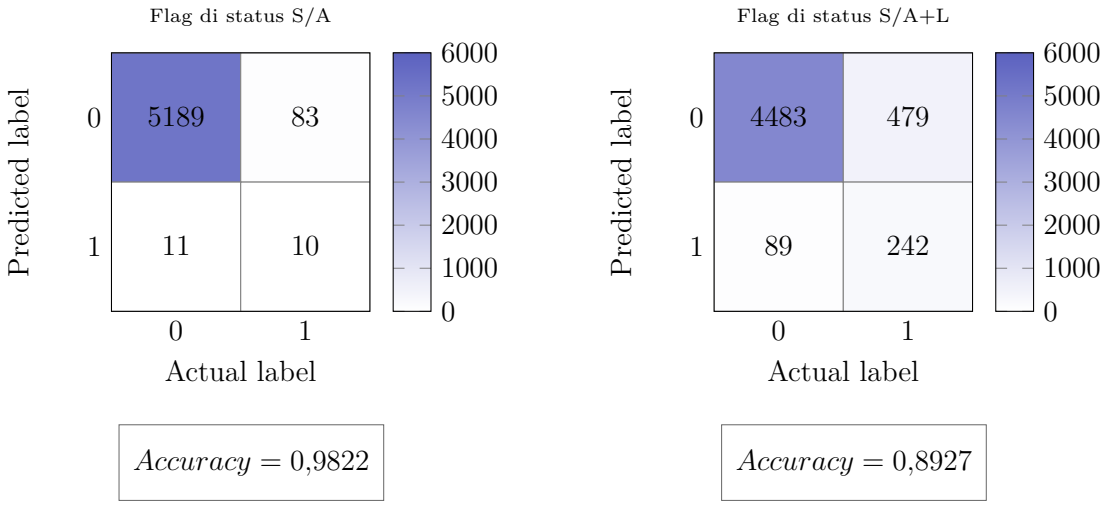


Figura 6.16. Modelli di intelligenza artificiale ($1A^{RF}$ e $1F^{RF}$).

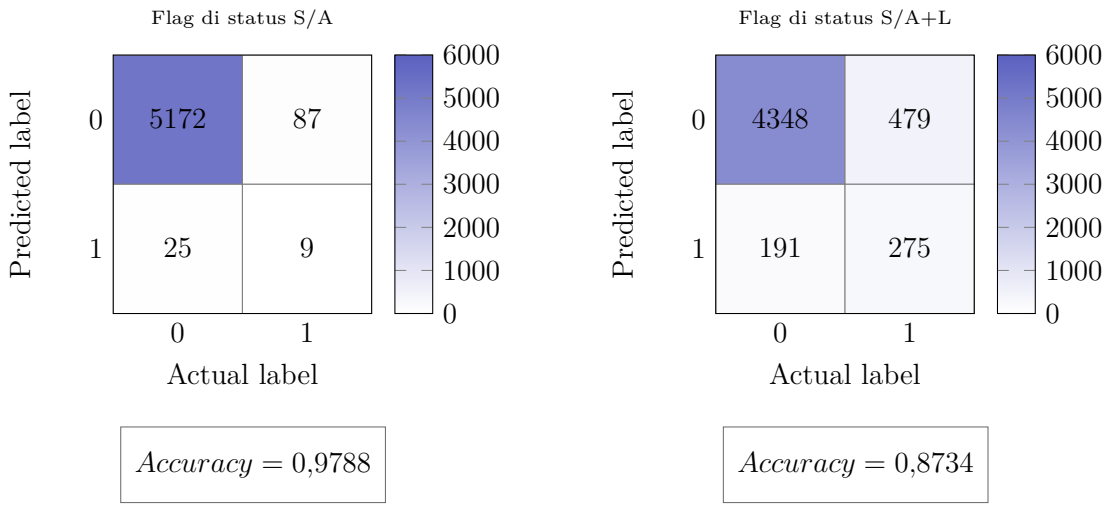


Figura 6.17. Modello di analisi discriminante.

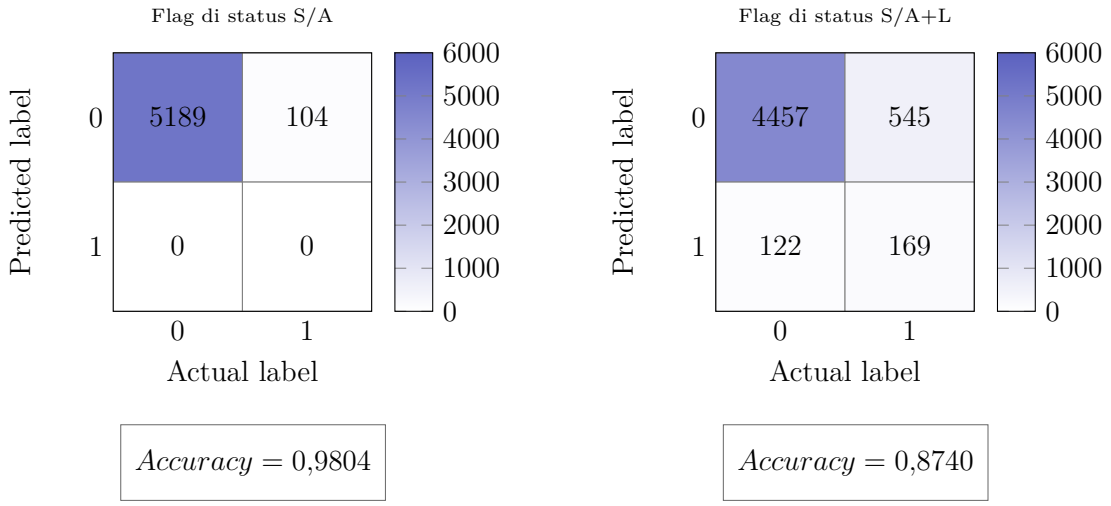


Figura 6.18. Modello di regressione logistica.

Capitolo 7

Conclusioni

7.1 Introduzione

L'obiettivo, estremamente ambizioso e stimolante, del presente lavoro di tesi è stato quello di immergersi completamente nell'universo del risk management, concentrando in modo particolare l'attenzione sulla probabilità di insolvenza di una controparte, nel tentativo di studiare e valutare una serie di approcci basati su tecniche di intelligenza artificiale. In poche parole, quindi, si è cercato di sperimentare l'applicazione di alcuni algoritmi di intelligenza artificiale e, nello specifico, machine learning e deep learning, nell'ambito del risk management applicato al settore bancario, con focus principale sul rischio di credito.

L'aspirazione di questa ricerca, che ha rappresentato il filo conduttore di tutto il lavoro, è stata quella di applicare i principali algoritmi di intelligenza artificiale, considerati allo stato dell'arte sulla base della loro efficacia ed efficienza, per sviluppare un modello di credit scoring per l'analisi del rischio di credito, e, possibilmente, mostrare le differenze con i modelli passati e presenti impiegati nel settore bancario nel corso della sua storia, come l'analisi discriminante, che ha rappresentato lo standard internazionale per molto tempo, dagli inizi fino agli anni '80 del secolo scorso, e, soprattutto, la regressione logistica, che, nel corso del tempo, ha progressivamente sostituito l'analisi discriminante, arrivando a rappresentare, grazie ai migliori risultati ottenuti, lo standard internazionale attuale.

Insomma, si è ritenuto interessante mostrare come l'intelligenza artificiale possa essere impiegata e applicata, anche in ambito economico-finanziario, per costruire modelli di credit scoring per la gestione del rischio di credito, in grado, potenzialmente, di diventare in un futuro, anche prossimo, lo standard de facto di tutto il settore industriale in esame.

7.2 Riflessioni conclusive

In conclusione, è possibile raccogliere alcune considerazioni finali, sviluppate nel corso del lavoro di tesi, studiando e lavorando per lo sviluppo di un modello di risk management per la gestione del rischio di credito. A questo proposito, è possibile suddividere il lavoro eseguito in tre parti: una prima parte dedicata all'analisi dell'ambiente di riferimento, con l'esame della regolamentazione bancaria e del rischio di credito, una seconda parte dedicata

all'analisi dei modelli di credit scoring più diffusi, oltre che allo studio dei principali modelli di intelligenza artificiale, e, infine, una terza parte dedicata allo sviluppo concreto di un vero e proprio processo di risk management, basato su tecniche di intelligenza artificiale, quali alberi decisionali e reti neurali, esplorandone l'applicazione in uno dei più grandi settori industriali italiani, ovvero il settore automotive.

Durante questo motivante percorso, si è arrivati ad acquisire una notevole conoscenza del fenomeno in esame, raccogliendo una serie di riflessioni di cui si ritiene utile discutere brevemente a completamento del lavoro svolto, in modo da sottolineare alcuni elementi essenziali e, magari, trasmettere un'ispirazione per possibili progetti futuri.

Prima di tutto, si è verificato l'importanza dei dati a disposizione per svolgere le analisi, i quali devono dimostrare di avere una sufficiente quantità e qualità per portare risultati consistenti e significativi. Questo aspetto, infatti, si è dimostrato essere molto importante, considerando che le informazioni utilizzate in questo progetto derivano direttamente dai database AIDA e ORBIS, che, sebbene risultino uno strumento di principale importanza per svolgere numerose analisi economico-finanziarie, presentano comunque un problema di incompletezza dei dati a disposizione, a causa del quale si è dovuto prevedere un'intensa fase di trasformazione, che implica la correzione dei valori errati o mancanti e, specialmente, la sostituzione di specifici valori singolari, per ripristinare il significato economico, con l'obiettivo di passare alla macchina informazioni tra loro coerenti.

In seguito, si è sperimentato il progetto di numerosi modelli di credit scoring, basati sia su tecniche statistiche che su tecniche di intelligenza artificiale, avendo cura di meditare in maniera critica e approfondita sulla loro progettazione. Nello specifico, i modelli statistici, quali analisi discriminante e regressione logistica, sono caratterizzati da una complessità relativamente modesta nella costruzione del modello, in quanto, per ricavare quest'ultimo, si rende unicamente necessario specificare in input le variabili del problema e applicare la metodologia prevista, per ottenere in output i coefficienti del modello, attraverso i quali è possibile calcolare il credit score per ogni impresa, modellare la probabilità di insolvenza e classificare, secondo diverse threshold, le imprese sane e le imprese anomale.

Le regole teoriche alla base dell'analisi discriminante o della regressione logistica sono state sviluppate ormai da molto tempo, rendendo la loro realizzazione e applicazione decisamente favorevole, considerando che possono poggiare su una solida base strutturale di riferimento, anche se, di contro, presentano un ridotto numero di gradi di libertà in fase di progettazione dell'architettura, osservando che i suddetti modelli statistici, per essere considerati accettabili, devono sempre sottostare al rispetto di specifiche regole teoriche. In questo caso, prima di analizzare i risultati di un modello statistico, si rende regolarmente necessario controllare che i coefficienti risultino statisticamente significativi e presentino un segno coerente con il significato economico della variabile, obbligando, in caso di esito non accettabile, al rifiuto del modello stesso, anche in caso di ottimi risultati.

Un ragionamento completamente opposto è valido nel caso dei modelli di intelligenza artificiale, i quali sono caratterizzati, per la fase di definizione dell'architettura, da una grande complessità e da infiniti gradi di libertà, portando vantaggi e svantaggi per un analista che si approccia a questo tipo di tecniche. In questo senso, da una parte, aumenta sensibilmente la difficoltà di gestione di un modello, poiché non si dispone di specifiche funzioni in grado di ottimizzare la struttura della rete neurale, come la scelta del numero di strati e del numero di neuroni, della tipologia di connessioni e di funzioni di attivazione,

o ancora di modellare gli iperparametri; dall'altra, aumenta parallelamente la capacità di costruzione di un modello dall'enorme potenziale computazionale, con la possibilità di imparare, elaborando in profondità, pattern anche molto complessi nei dati, apprendendo relazioni fra variabili, lineari e, soprattutto, non lineari, che i modelli statistici non riescono a catturare.

In aggiunta, si è osservata la differenza tra modelli statistici e modelli di intelligenza artificiale in relazione alla disponibilità di dati e al fabbisogno di potenza di calcolo, con i modelli di intelligenza artificiale che richiedono, rispetto ai modelli statistici, una base di dati più numerosa, in modo da poter estrarre tutta la conoscenza disponibile, oltre che una potenza di calcolo più consistente, in modo da svolgere con successo il processo di addestramento, che, prevedendo numerosi calcoli iterativi, potrebbe rivelarsi lungo e dispendioso.

In ultima istanza, si è dimostrato la possibilità di riconoscere eventuali problematiche durante tutto il processo di costruzione del sistema di analisi, applicandosi in modo differenziato a seconda che si tratti di modelli statistici o di modelli di intelligenza artificiale. I primi, essendo modelli deduttivi di tipo white-box, consentono di identificare facilmente e velocemente un eventuale problema nella costruzione di un modello e, una volta individuato, è possibile utilizzare la teoria economica per comprendere il problema e risolverlo. I secondi, al contrario, essendo modelli induttivi di tipo black-box, non permettono di comprendere il processo di elaborazione condotto dal modello, costringendo ad affidarsi, per risolvere i problemi, a metodi euristici, come processi di trial and error e thumb rule. Così, i modelli statistici si basano su fondamenta matematico-statistiche universalmente riconosciute, con regole di costruzione e validazione rigidamente definite in letteratura, mentre i modelli di intelligenza artificiale non presentano fondamenta teoriche per la loro costruzione e validazione, sia per la natura dello strumento che per la limitata vita utile.

Mettendo insieme tutte queste riflessioni, è possibile concludere come i modelli di intelligenza artificiale rappresentino un insieme di tecniche caratterizzate sicuramente da un'elevata complessità, sia dal punto di vista teorico che dal punto di vista pratico, ma, allo stesso tempo, in grado di sfruttare proprio questa complessità per ottenere risultati migliori, analizzando un fenomeno impegnativo, quale l'insolvenza di una controparte.

A questo proposito, considerando che l'applicazione dell'intelligenza artificiale a diversi casi di studio risulta un fenomeno abbastanza recente, osservando che la disponibilità di dati e di potenza di calcolo risulta costantemente in crescita ormai da diverso tempo e, soprattutto, sottolineando che la comunità scientifica in ambito di intelligenza artificiale risulta estremamente attiva e aperta, con la volontà di offrire, nel campo della conoscenza, numerose soluzioni open-source, oltre che condividere i risultati della ricerca, è facile immaginare un incremento esponenziale di progetti nel prossimo futuro.

Se questo è vero, allora, il prossimo processo di risk management per la gestione del rischio di credito non si costruisce continuando a lavorare unicamente con le metodologie utilizzate come standard de facto in tutto il settore bancario nel corso della storia, come l'analisi discriminante prima e la regressione logistica dopo, strumenti tutt'ora preziosi, ma continuando a lavorare per migliorare la conoscenza di intelligenza artificiale alla base di questi nuovi potenti strumenti e integrarli sempre più all'interno dei processi decisionali.

Bibliografia

[1] Accenture. *The ROI of AI*. Report. Accenture, 2019.

[2] C. Alexander. *The handbook of risk management and analysis*. 1ª ed. Chichester: John Wiley & Sons, 1996.

[3] E. I. Altman. «Financial ratios, discriminant analysis and the prediction of corporate bankruptcy». In: *Journal of Finance* 23 (1968), pp. 589–609.

[4] E. I. Altman, R. B. Avery, R. A. Eisenbeis e J. F. Sinkey. *Application of classification techniques in business, banking and finance*. 1ª ed. Greenwich, Connecticut: JAI Press, 1981.

[5] E. I. Altman, B. Brady, A. Resti e A. Sironi. «The link between default and recovery rates: theory, empirical evidence and implications». In: *Journal of Business* 78 (2005), pp. 2203–2228.

[6] E. I. Altman, R. G. Haldeman e P. Narayanan. «ZETA analysis: a new model to identify bankruptcy risk of corporations». In: *Journal of Banking and Finance* 1 (1977), pp. 29–54.

[7] E. I. Altman, A. Resti e A. Sironi. «Default recovery rates in credit risk modeling: a review of the literature and empirical evidence». In: *Journal of Finance Literature* 1 (2005), pp. 21–45.

[8] E. I. Altman e A. Saunders. «An analysis and critique of the BIS proposal on capital adequacy and ratings». In: *Journal of Banking and Finance* 25 (2001), pp. 25–46.

[9] R. Anyoha. «The history of artificial intelligence». In: *Science in the News, Harvard University* (2017).

[10] P. Artzner, F. Delbaen, J. M. Eber e D. Heath. «Coherent measures of risk». In: *Mathematical Finance* 9 (1999), pp. 203–228.

[11] A. F. Atiya. «Bankruptcy prediction for credit risk using neural networks: a survey and new results». In: *IEEE Transactions on Neural Networks* (2001).

[12] Banca d'Italia (2005). *Il bilancio bancario: schemi e regole di compilazione*. Circolare n. 262 del 22 dicembre 2005. Banca d'Italia. Roma, 2005.

[13] Banca d'Italia (2006). *Nuove disposizioni di vigilanza prudenziale per le banche*. Circolare n. 263 del 27 dicembre 2006. Banca d'Italia. Roma, 2006.

[14] Banca d'Italia (2008). *Guida per l'attività di vigilanza*. Circolare n. 269 del 7 maggio 2008. Banca d'Italia. Roma, 2008.

[15] Banca d'Italia (2010). *Linee guida per l'analisi di impatto della regolamentazione.* Circolare n. 277 del 20 luglio 2010. Banca d'Italia. Roma, 2010.

[16] Banca d'Italia (2013). *Disposizioni di vigilanza per le banche.* Circolare n. 285 del 17 dicembre 2013. Banca d'Italia. Roma, 2013.

[17] Banca d'Italia (2015). *Disposizioni di vigilanza per gli intermediari finanziari.* Circolare n. 288 del 3 aprile 2015. Banca d'Italia. Roma, 2015.

[18] Bank of England e Federal Reserve Board. *Convergence of capital adequacy in the United Kingdom and the United States.* Report. London-Washington: Bank of England e Federal Reserve Board, 2009.

[19] Basel Commitee on Banking Supervision (1987). *Proposal for the international convergence of capital measurement and capital standards.* Consultative. Basel: Bank for International Settlements, 1987.

[20] Basel Commitee on Banking Supervision (1988). *International convergence of capital measurement and capital standards.* Standard. Basel: Bank for International Settlements, 1988.

[21] Basel Commitee on Banking Supervision (1993a). *The prudential supervision of netting, market risks and interest rate risk.* Other. Basel: Bank for International Settlements, 1993.

[22] Basel Commitee on Banking Supervision (1993b). *The supervisory treatment of market risk.* Consultative. Basel: Bank for International Settlements, 1993.

[23] Basel Commitee on Banking Supervision (1994). *Amendment to the 1988 capital accord recognition of collateral.* Standard. Basel: Bank for International Settlements, 1994.

[24] Basel Commitee on Banking Supervision (1995). *Basel capital accord: treatment of potential exposure for off-balance-sheet items.* Standard. Basel: Bank for International Settlements, 1995.

[25] Basel Commitee on Banking Supervision (1996). *Amendment to the capital accord to incorporate market risks.* Standard. Basel: Bank for International Settlements, 1996.

[26] Basel Commitee on Banking Supervision (1997). *Core principles for effective banking supervision.* Consultative. Basel: Bank for International Settlements, 1997.

[27] Basel Commitee on Banking Supervision (1998). *Enhancing bank transparency.* Guideline. Basel: Bank for International Settlements, 1998.

[28] Basel Commitee on Banking Supervision (1999a). *Promoting financial stability.* Other. Basel: Bank for International Settlements, 1999.

[29] Basel Commitee on Banking Supervision (1999b). *Capital requirements and bank behaviour: the impact of the Basel accord.* Working paper. Basel: Bank for International Settlements, 1999.

[30] Basel Commitee on Banking Supervision (1999c). *A new capital adequacy framework.* Consultative. Basel: Bank for International Settlements, 1999.

[31] Basel Commitee on Banking Supervision (2000a). *Range of practice in banks' internal ratings systems*. Sound practice. Basel: Bank for International Settlements, 2000.

[32] Basel Commitee on Banking Supervision (2000b). *Credit ratings and complementary sources of credit quality information*. Other. Basel: Bank for International Settlements, 2000.

[33] Basel Commitee on Banking Supervision (2000c). *Principles for the management of credit risk*. Guideline. Basel: Bank for International Settlements, 2000.

[34] Basel Commitee on Banking Supervision (20001). *Basel II: the new Basel capital accord*. Consultative. Basel: Bank for International Settlements, 2001.

[35] Basel Commitee on Banking Supervision (2003). *Trends in risk integration and aggregation*. Other. Basel: Bank for International Settlements, 2003.

[36] Basel Commitee on Banking Supervision (2004a). *Modifications to the capital treatment for expected and unexpected credit losses*. Guideline. Basel: Bank for International Settlements, 2004.

[37] Basel Commitee on Banking Supervision (2004b). *Bank failures in mature economies*. Working paper. Basel: Bank for International Settlements, 2004.

[38] Basel Commitee on Banking Supervision (2004c). *Basel II: international convergence of capital measurement and capital standards (a revised framework)*. Stanrdard. Basel: Bank for International Settlements, 2004.

[39] Basel Commitee on Banking Supervision (2005a). *Studies on the validation of internal rating systems*. Working paper. Basel: Bank for International Settlements, 2005.

[40] Basel Commitee on Banking Supervision (2005b). *Enhancing corporate governance for banking organisations*. Consultative. Basel: Bank for International Settlements, 2005.

[41] Basel Commitee on Banking Supervision (2006a). *Regulatory and market differences: issues and observations*. Other. Basel: Bank for International Settlements, 2006.

[42] Basel Commitee on Banking Supervision (2006b). *The IRB use test: background and implementation*. Guideline. Basel: Bank for International Settlements, 2006.

[43] Basel Commitee on Banking Supervision (2006c). *Studies on credit risk concentration*. Working paper. Basel: Bank for International Settlements, 2006.

[44] Basel Commitee on Banking Supervision (2008a). *Liquidity risk: management and supervisory challenges*. Sound practice. Basel: Bank for International Settlements, 2008.

[45] Basel Commitee on Banking Supervision (2008b). *Principles for sound liquidity risk management and supervision*. Consultative. Basel: Bank for International Settlements, 2008.

[46] Basel Commitee on Banking Supervision (2009a). *Findings on the interaction of market and credit risk*. Working paper. Basel: Bank for International Settlements, 2009.

[47] Basel Commitee on Banking Supervision (2009b). *Core principles for effective deposit insurance systems*. Standard. Basel: Bank for International Settlements, 2009.

[48] Basel Commitee on Banking Supervision (2009c). *International framework for liquidity risk measurement, standards and monitoring*. Consultative. Basel: Bank for International Settlements, 2009.

[49] Basel Commitee on Banking Supervision (2010a). *Basel III: international framework for liquidity risk measurement, standards and monitoring*. Standard. Basel: Bank for International Settlements, 2010.

[50] Basel Commitee on Banking Supervision (2010b). *Basel III: a global regulatory framework for more resilient banks and banking systems*. Standard. Basel: Bank for International Settlements, 2010.

[51] Basel Commitee on Banking Supervision (2011). *Global systemically important banks: assessment methodology and the additional loss absorbency requirement*. Consultative. Basel: Bank for International Settlements, 2011.

[52] Basel Commitee on Banking Supervision (2012a). *Results of the Basel III monitoring exercise*. QIS (Quantitative Impact Study). Basel: Bank for International Settlements, 2012.

[53] Basel Commitee on Banking Supervision (2012b). *The policy implications of transmission channels between the financial system and the real economy*. Working paper. Basel: Bank for International Settlements, 2012.

[54] Basel Commitee on Banking Supervision (2012c). *A framework for dealing with domestic systemically important banks*. Consultative. Basel: Bank for International Settlements, 2012.

[55] Basel Commitee on Banking Supervision (2013a). *Liquidity stress testing: a survey of theory, empirics and current industry and supervisory practices*. Working paper. Basel: Bank for International Settlements, 2013.

[56] Basel Commitee on Banking Supervision (2013b). *Supervisory framework for measuring and controlling large exposures*. Consultative. Basel: Bank for International Settlements, 2013.

[57] Basel Commitee on Banking Supervision (2015a). *The interplay of accounting and regulation and its impact on bank behaviour*. Working paper. Basel: Bank for International Settlements, 2015.

[58] Basel Commitee on Banking Supervision (2015b). *TLAC holdings*. Consultative. Basel: Bank for International Settlements, 2015.

[59] Basel Commitee on Banking Supervision (2017). *Basel III: finalising post-crisis reforms*. Standard. Basel: Bank for International Settlements, 2017.

[60] W. H. Beaver. «Financial ratios as predictors of failures». In: *Journal of Accounting Research* 4 (1966), pp. 71–111.

[61] T. S. Beder. «VAR: seductive but dangerous». In: *Financial Analyst Journal* 51 (1995), pp. 12–24.

[62] A. Beltratti e R. M. Stulz. «Why did some banks perform better during the credit crisis? A cross-country study of the impact of governance and regulation». In: *National Bureau of Economic Research* (2009).

[63] Boston Consulting Group. *Five rules for fixing AI in business*. Report. Boston Consulting Group, 2022.

[64] M. K. Brunnermeier, A. Crocket, C. Goodhart, A. D. Persaud e H. S. Shin. *The fundamental principles of financial regulation*. Geneva Reports on the World Economy, 2009.

[65] R. Calcagno e E. Ughetto. *Metodi di finanziamento delle imprese*. Materiale didattico corso di Laurea Magistrale in Ingegneria Gestionale. Torino: Politecnico di Torino, 2021.

[66] C. Cambini e L. Buzzacchi. *Economia dei sistemi industriali*. Materiale didattico corso di Laurea Magistrale in Ingegneria Gestionale. Torino: Politecnico di Torino, 2020.

[67] C. Cambini, E. Grinza e L. Abrardi. *Analisi dei sistemi economici*. Materiale didattico corso di Laurea Magistrale in Ingegneria Gestionale. Torino: Politecnico di Torino, 2020.

[68] J. B. Caouette, E. I. Altman, P. A. Narayanan e R. W. J. Nimmo. *Managing credit risk: the great challenge for global financial markets*. 2ª ed. John Wiley & Sons, 2008.

[69] Cassa Depositi e Prestiti (CDP), EY e Luiss Business School. *Settore automotive e Covid-19: scenario, impatti e prospettive*. Report. Italia: Cassa Depositi e Prestiti (CDP), 2020.

[70] P. F. Christoffersen. *Elements of financial risk management*. 2ª ed. Academic Press, 2011.

[71] Comitato esecutivo ABI. *Riunione del 21 ottobre 2020*. Report. ABI. Italia, 2020.

[72] Commissione Europea. *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts*. Proposal. Commissione Europea. Bruxelles, 2021.

[73] Commissione Europea. *Annexes to the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts*. Proposal. Commissione Europea. Bruxelles, 2021.

[74] M. Crouhy, D. Galai e R. Mark. *The essentials of risk management*. 2ª ed. McGraw-Hill, 2006.

[75] Deloitte. *Artificial intelligence*. Report. Deloitte, 2018.

[76] G. A. Donaldson. «New framework for corporate debt policy». In: *Harvard Business Review* (1962).

[77] G. A. Donaldson. *Strategy for financial mobility*. 1ª ed. Boston: Harvard Business School Press, 1969.

[78] P. Embrechts, C. Klüppelberg e T. Mikosch. *Modelling extremal events for insurance and finance*. 1ª ed. Berlin: Springer-Verlag, 1997.

[79] B. G. Farley e W. A. Clark. «Simulation of self-organizing systems by digital computer». In: *Transactions of the IRE Professional Group on Information Theory* 4 (1954), pp. 76–84.

[80] Financial Stability Board. *Report of the financial stability forum on addressing procyclicality in the financial system*. Report. Financial Stability Board, 2009.

[81] Financial Stability Board. *2022 list of global systemically important banks (G-SIBs)*. Report. Financial Stability Board, 2022.

[82] G. Gabbi. «L'utilizzo delle reti neurali per la misurazione del rischio di credito». In: *La misurazione e la gestione del rischio di credito*. Roma: Bancaria, 1998.

[83] G. Gabbi. *La previsione nei mercati finanziari*. 1ª ed. Roma: Bancaria, 1999.

[84] O. S. D. Hart e L. Zingales. «A new capital regulation for large financial institutions». In: *American Law and Economics Review* 13 (2011), pp. 453–490.

[85] D. O. Hebb. *The organization of behavior: a neuropsychological theory*. 1ª ed. New York: John Wiley & Sons, 1949.

[86] J. C. Hull. *Risk management and financial istitutions*. 5ª ed. Hoboken: John Wiley & Sons, 2018.

[87] A. A. Jobst e D. F. Gray. *Systemic contingent claims analysis: estimating market-implied systemic risk*. Working paper. Washington: International Monetary Fund, 2013.

[88] A. Khashman. «A neural network model for credit risk evaluation». In: *International Journal of Neural Systems* 19 (2009), pp. 285–294.

[89] D. Kim e A. M. Santomero. «Risk in banking and capital regulation». In: *The Journal of Finance* 43 (1998), pp. 1219–1233.

[90] KPMG. *Artificial intelligence and machine learning*. Report. KPMG, 2022.

[91] M. Lam. «Neural network techniques for financial performance prediction: integrating fundamental and technical analysis». In: *Decision Support System* 37 (2004), pp. 567–581.

[92] J. R. Lay. *Banking on digital growth: the strategic marketing manifesto to transform financial brands*. 1ª ed. Lioncrest, 2020.

[93] O. Maimon e L. Rokach. *Data mining and knowledge discovery handbook*. 2ª ed. New York: Springer, 2010.

[94] J. McCarthy, M. L. Minsky, C. E. Shannon e N. Rochester. «A proposal for the Dartmouth summer research project on artificial intelligence». In: (1955).

[95] W. S. McCulloch e W. H. Pitts. «A logical calculus of the ideas immanent in nervous activity». In: *Bullettin of Mathematical Biophysics* 5 (1943), pp. 115–133.

[96] McKinsey. *Artificial intelligence: the time to act is now.* Report. McKinsey, 2018.

[97] L. J. Mester. «What's the point of credit scoring?» In: *Business Review* 3 (1997), pp. 3–16.

[98] M. L. Minsky e S. A. Papert. *Perceptrons: an introduction to computational geometry.* 1ª ed. Cambridge, Massachusetts: MIT Press, 1969.

[99] T. M. Mitchell. *Machine learning.* 1ª ed. McGraw-Hill, 1997.

[100] L. Nadotti, C. Porzio e D. Previati. *Economia degli intermediari finanziari.* 4ª ed. Milano: McGraw-Hill, 2022.

[101] A. Newell e H. A. Simon. «The logic theory machine: a complex information processing system». In: *IRE Transactions on Information Theory* 2 (1956), pp. 61–79.

[102] M. Nielsen. *Neural networks and deep learning.* 2019.

[103] E. I. Ohlson. «Financial ratios and the probabilistic prediction of bankruptcy». In: *Journal of Accounting Research* 18 (1980), pp. 109–131.

[104] E. Paolucci e P. Neirotti. *Strategia e organizzazione aziendale.* Materiale didattico corso di Laurea Magistrale in Ingegneria Gestionale. Torino: Politecnico di Torino, 2020.

[105] R. H. Pettway, T. Kanedo e M. T. Young. «International bank capital standards and the costs of issuing capital securities by Japanese banks». In: *Journal of Banking and Finance* 15 (1991), pp. 559–580.

[106] S. Piramuthu. «Financial credit-risk evaluation with neural and neurofuzzy systems». In: *European Journal of Operational Research* 112 (1999), pp. 310–321.

[107] PwC. *Solving AI's ROI problem.* Report. PwC, 2021.

[108] A. Resti e A. Sironi. *Rischio e valore nelle banche.* 2ª ed. Milano: EGEA, 2021.

[109] L. Rondi e G. Falavigna. *Economia e finanza d'impresa.* Materiale didattico corso di Laurea Magistrale in Ingegneria Gestionale. Torino: Politecnico di Torino, 2021.

[110] N. Rochester, J. H. Holland, L. H. Haibt e W. L. Duda. «Tests on a cell assembly theory of the action of the brain, using a large digital computer». In: *IRE Transactions on Information Theory* 2 (1956), pp. 80–93.

[111] F. Rosenblatt. «The perceptron: a probabilistic model for information storage and organization in the brain». In: *Psychological Review* 65 (1958), pp. 386–408.

[112] F. Rosenblatt. *Principles of neurodynamics.* 1ª ed. Washington: Spartan Books, 1962.

[113] G. Rotundo. «Neural networks for large financial crashes forecast». In: *Physica* 344 (2004), pp. 77–80.

[114] D. E. Rumelhart, G. E. Hinton e R. J. Williams. «Learning internal representations by error propagation». In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, Massachusetts: MIT Press, 1986, pp. 318–362.

[115] A. L. Samuel. «Some studies in machine learning using the game of checkers». In: *IBM Journal of Research and Development* 3 (1959), pp. 210–229.

[116] G. Scellato e F. Caviggioli. *Economia aziendale*. Materiale didattico corso di Laurea Magistrale in Ingegneria Gestionale. Torino: Politecnico di Torino, 2020.

[117] J. R. Sobehart e S. C. Keenan. «Measuring default accurately». In: *Risk* 14 (2001), pp. 31–33.

[118] J. R. Sobehart e S. C. Keenan. «The score for credit». In: *Risk* 17 (2004), pp. 54–58.

[119] K. Y. Tam e M. Y. Kiang. «Managerial applications of neural networks: the case of bank failure predictions». In: *Management Science* 38 (1992), pp. 926–947.

[120] A. M. Turing. «Computing machinery and intelligence». In: *Mind (Oxford University Press)* 59 (1950), pp. 433–460.

[121] Unione Europea. *Direttiva 2014/59/UE del Parlamento europeo e del Consiglio, del 15 maggio 2014, che istituisce un quadro di risanamento e risoluzione degli enti creditizi e delle imprese di investimento*. Bank Recovery and Resolution Directive (BRRD). Gazzetta Ufficiale dell'Unione Europea. 2014.

[122] Unione Europea. *Direttiva (UE) 2019/879 del Parlamento europeo e del Consiglio, del 20 maggio 2019, che modifica la direttiva 2014/59/UE per quanto riguarda la capacità di assorbimento di perdite e di ricapitalizzazione degli enti creditizi e delle imprese di investimento*. Bank Recovery and Resolution Directive (BRRD). Gazzetta Ufficiale dell'Unione Europea. 2019.

[123] F. Varetto. *Mercati, rischi e strumenti finanziari*. Materiale didattico corso di Laurea Magistrale in Ingegneria Gestionale. Torino: Politecnico di Torino, 2021.

[124] J. D. Wagster. «Impact of the 1988 Basle Accord on international banks». In: *Journal of Finance* 51 (1996), pp. 1321–1346.

[125] J. R. Walter. «US bank capital regulation: history and changes since the financial crisis». In: *Economic Quarterly* 105 (2019), pp. 1–40.

[126] D. West. «Neural network credit scoring models». In: *Computers and Operations Research* 27 (2000), pp. 1131–1152.

[127] R. Z. Wiggins, T. Piontek e A. Metrick. «The Lehman Brothers bankruptcy». In: *Yale Program on Financial Stability Case Study* (2014).

[128] J. W. Wilcox. «A simple theory of financial ratios as predictors of failures». In: *Journal of Accounting Research* 9 (1971), pp. 389–395.

[129] C. Wu e X. M. Wang. «A neural network approach for analyzing small business lending decisions». In: *Review of Quantitative Finance and Accounting* 15 (2000), pp. 259–276.

[130] G. Zhang, M. Y. Hu, B. E. Patuwo e D. C. Indro. «Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis». In: *European Journal of Operational Research* 116 (1998), pp. 16–32.

[131] G. Zhang, B. E. Patuwo e M. Y. Hu. «Forecasting with artificial neural networks: the state of the art». In: *International Journal of Forecasting* 14 (1998), pp. 35–62.

[132] S. A. Zimmer e McCauley R. N. «Bank cost of capital and international competition». In: *Federal Reserve Bank of New York Quarterly Review* 15 (1991), pp. 33–59.

Sitografia

[133] ABI (Associazione Bancaria Italiana). URL: https://www.abi.it/.

[134] ANFIA (Associazione Nazionale Filiera Industria Automobilistica). URL: https://www.anfia.it/it/.

[135] Banca d'Italia. URL: https://www.bancaditalia.it/.

[136] Bank of England. URL: https://www.bankofengland.co.uk/.

[137] BCBS (Basel Commitee on Banking Supervision). URL: https://www.bis.org/.

[138] Bureau of the Census. URL: https://www.census.gov/.

[139] Bureau Van Dijk (Moody's Analytics). URL: https://www.bvdinfo.com/it-it/.

[140] Deutsche Bundesbank. URL: https://www.bundesbank.de/en.

[141] EBA (European Banking Authority). URL: https://www.eba.europa.eu/.

[142] ECB (European Central Bank). URL: https://www.ecb.europa.eu/.

[143] European Commission. URL: https://ec.europa.eu/info/index_en.

[144] FDIC (Federal Deposit Insurance Corporation). URL: https://www.fdic.gov/.

[145] Federal Reserve Board. URL: https://www.federalreserve.gov/.

[146] Financial Stability Board. URL: https://www.fsb.org/.

[147] IIF (Institute of International Finance). URL: https://www.iif.com/.

[148] IMF (International Monetary Fund). URL: https://www.imf.org/en/Home.

[149] ISTAT (Istituto Nazionale di Statistica). URL: https://www.istat.it/.

[150] Keras. URL: https://keras.io/.

[151] Matplotlib. URL: https://matplotlib.org/.

[152] NumPy. URL: https://numpy.org/.

[153] OECD (Organisation for Economic Co-Operation e Development). URL: https://www.oecd.org/.

[154] OICA (International Organization for Motor Vehicle Manufacturers). URL: https://www.oica.net/.

[155] OICV-IOSCO (International Organization of Securities Commissions). URL: https://www.iosco.org/.

[156] Pandas. URL: https://pandas.pydata.org/.

[157] Scikit-learn. URL: https://scikit-learn.org/.

[158] SciPy. URL: https://scipy.org/.

[159] Seaborn. URL: https://seaborn.pydata.org/.

[160] Statsmodel. URL: https://www.statsmodels.org/.

[161] TensorFlow. URL: https://www.tensorflow.org/.

[162] The Economist. URL: https://www.economist.com/.