

POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale

**Corso di Laurea Magistrale
in Ingegneria Gestionale
percorso Finanza**

Tesi di Laurea Magistrale

**Valutazione del rischio di credito del settore
Telecomunicazioni: il contributo dell'intelligenza
artificiale mediante le reti neurali**



Relatore:
Prof. Franco Varetto

Candidato:
Giuseppe Milazzo

Anno accademico 2020/2021

“La scienza di oggi è la tecnologia di domani”
(Edward Teller)

INDICE

INTRODUZIONE	6
CAPITOLO I - Gli accordi di Basilea: la regolamentazione della Vigilanza Bancaria ...	8
1.1 Basilea I	8
1.1.1 Limiti e conseguenze di Basilea I	10
1.2 Basilea II	11
1.2.1 Primo pilastro	11
1.2.2 Secondo pilastro	17
1.2.3 Terzo pilastro	19
1.2.4 Limiti e conseguenze di Basilea II	19
1.3 Basilea III: la regolamentazione attuale.....	21
1.3.1 Considerazioni su Basilea III	26
1.4 Basilea IV: uno sguardo al futuro	26
CAPITOLO II – Il Rischio di Credito	28
2.1 Il rischio di credito	28
2.2 Le componenti del rischio di credito	30
2.2.1 La perdita attesa	30
2.2.2 La perdita inattesa	34
2.3 Credit scoring.....	36
2.3.1 Modelli statistici: la regressione logistica	37
2.3.2 Modelli di machine learning	40
CAPITOLO III – Le Reti neurali	42
3.1 Il neurone biologico	42
3.2 Il neurone artificiale.....	43
3.3 Tipi di funzione di attivazione	46
3.3.1 Step function	47
3.3.2 Linear Activation Function	48
3.3.3 Sigmoid Function	48
3.3.4 Tanh Function	49
3.3.5 ReLU Function.....	50
3.3.6 Leaky ReLU Function.....	51
3.3.7 Parameterised ReLU Function	51
3.4 Architettura delle reti neurali	52
3.4.1 Single-Layer Feedforward Architecture.....	54

3.4.2 Multiple-Layer Feed-forward Architectures	54
3.4.3 Recurrent or Feedback Architecture	55
3.4.4 Mesh Architectures	56
3.5 L'algoritmo back propagation.....	56
3.5.1 Loss Function	57
CAPITOLO IV - Il Settore scelto e l'operazione di preparazione del dataset	58
4.1 settore telecomunicazioni in Italia	58
4.2 Lo scarico dei dati da Aida	60
4.3 La pulizia del dataset e la correzione degli errori	62
4.4: Preprocessing	62
4.5.1 Dataset.....	63
4.5.2 Saturazione al 5° e 95° percentile	65
4.5.3 Standardizzazione e Normalizzazione	66
4.5.4 Rimozione degli outliers tramite distanza di Mahalanobis	69
4.5.5 Analisi di correlazione	71
5.1 Training-Validation-Test	75
5.2. Ribilanciamento dei dati: SMOTE	77
5.3. Architettura della rete	78
5.4. Configurazione iniziale.....	79
5.4.1 Inizializzazione dei parametri della rete.....	80
5.4.2 Strategia di ottimizzazione - SGD Optimizer	81
5.4.3 Funzione di loss.....	81
CAPITOLO VI – Analisi delle performance del modello.....	84
6.1 Fase di training.....	84
6.2 Fase di test e Metriche di valutazione.....	87
6.2.1 Accuratezza	88
6.2.2 Precision & Recall.....	88
6.3 Considerazioni su un ulteriore giro di rete neurale: il caso delle società in liquidazione.....	90
CONCLUSIONI.....	93
BIBLIOGRAFIA E SITOGRAFIA.....	94
RINGRAZIAMENTI	95

INTRODUZIONE

Uno degli aspetti principali di un soggetto che definiamo intelligente è certamente la sua capacità di apprendere ed uno dei motivi che ha spinto l'uomo ad interessarsi dell'intelligenza artificiale, ed in particolare delle reti neurali, è comprendere la sua stessa mente.

Tutti gli algoritmi di intelligenza artificiale si basano sul *machine learning* ovvero sull'apprendimento automatico. La capacità di un software di apprendere ed eseguire compiti grazie ad algoritmi, ispirati alla struttura del cervello, che imparano dai dati in modo iterativo, ha da sempre affascinato gli studiosi in svariati campi, non solo in quello informatico, ma anche in quello della medicina, dei trasporti, delle telecomunicazioni, della meteorologia, della psicologia ma in particolare della finanza. Il presente lavoro di tesi consiste nell'applicazione del concetto dell'intelligenza artificiale in ambito finanziario. In particolare, è stato valutato il rischio di credito delle imprese italiane che appartengono al settore telecomunicazioni attraverso la costruzione di un modello basato sulle reti neurali, capace di apprendere dai dati, elaborarli ed effettuare previsioni verosimili anche laddove gli output non sono noti a priori. L'elaborato consta di sei capitoli. Il primo capitolo tratta degli Accordi di Basilea, che rappresentano un insieme di regole elaborate dal Comitato di Basilea per la Vigilanza Bancaria per far fronte al rischio di credito. Dopo aver indicato le componenti del rischio di credito nel secondo capitolo, e quindi dopo aver spiegato in che campo viene applicato il modello, il terzo capitolo è stato dedicato alle reti neurali: è stato spiegato a livello teorico cosa sono, su cosa si basano e il meccanismo che sta alla base del loro funzionamento. Il quarto capitolo fa riferimento al settore scelto, quello delle telecomunicazioni, allo scarico dei dati dal database AIDA e alla non indifferente attività di pulizia del dataset. Nei due capitoli successivi vengono presentati i dettagli implementativi della rete neurale e delle relative fasi di training, validation e test, e l'analisi delle performance ottenute sul dataset analizzato: in particolare, prima vengono discusse le scelte architetturali e configurative inerenti il modello neurale e l'approccio di training; successivamente, si passa ad analizzare l'andamento della loss function sia sul training che sul validation set per valutare eventuali circostanze di overfitting/underfitting; infine, i risultati finali raggiunti sul dataset di test tramite il modello addestrato vengono presentati attraverso una

confusion matrix e altre metriche comunemente usate in classificazione come accuracy, precision e recall.

CAPITOLO I - Gli accordi di Basilea: la regolamentazione della Vigilanza Bancaria

I cosiddetti Accordi di Basilea, ampiamente trattati in questo primo capitolo, rappresentano un insieme di regole elaborate dal Comitato di Basilea per la Vigilanza Bancaria sotto la supervisione del Financial Stability Board. Tali regole vengono poi applicate dai Paesi aderenti su base volontaria. Al Comitato possono partecipare non solo i dieci paesi più industrializzati al mondo ma anche quelli che, su base volontaria, decidono di adottare le sue regole. I dieci paesi a cui ci si riferisce sono quelli del G-10 che nasce nel 1962 e comprende le dieci economie occidentali più avanzate: Belgio, Canada, Francia, Germania, Italia, Giappone, Olanda, Gran Bretagna, Stati Uniti, Svezia e Svizzera (quest'ultima ne fa parte dal 1984). Il Comitato, fin dalle sue origini, si propone di stabilire l'adeguatezza patrimoniale degli intermediari e di rafforzare la sicurezza e l'affidabilità del sistema finanziario. Pertanto, nel presente capitolo verranno illustrate le quattro regolamentazioni emanate dal Comitato, ovvero: Basilea I, Basilea II, Basilea III e Basilea IV. L'analisi descrittiva di tale capitolo ha come scopo principale quello di rappresentare i punti chiave e le motivazioni che hanno portato all'evoluzione – dal 1988 ad oggi - degli Accordi.

1.1 Basilea I

Alla fine degli anni Ottanta il contesto normativo sul quale operavano le banche era poco regolamentato. Per tali motivi, molti istituti bancari adottarono una condotta “aggressiva” che pose le basi per la celebre crisi esplosa solo pochi anni dopo. A ciò ne conseguì una forte regolamentazione che fu introdotta con il primo Accordo di Basilea raggiunto nel 1988. Tale regolamentazione, che inizialmente riguardava solo le banche internazionali, è stata adottata a seguito della pubblicazione di alcune proposte elaborate nel corso dei diversi anni e alle quali è seguito un processo consultivo nei Paesi del G-10. L'obiettivo principale era quello di non compromettere in nessun modo l'efficienza sulla gestione dei rischi e delle perdite, data la loro influenza sulla stabilità economica mondiale.

Prima di giungere all'Accordo del 1988, nel 1974, è stato istituito il Comitato di Basilea, ovvero un organismo internazionale di cooperazione. “Il Comitato di Basilea per la Vigilanza Bancaria – CBVB - è il principale organismo di definizione degli standard

internazionali per la regolamentazione prudenziale del settore bancario. Esso fornisce una sede per la cooperazione in materia di vigilanza bancaria. Il suo mandato consiste nel rafforzare la regolamentazione, la vigilanza e le prassi bancarie a livello mondiale allo scopo di accrescere la stabilità finanziaria.” (Comitato di Basilea per la Vigilanza Bancaria 2013). I membri del Comitato di Basilea sono le banche centrali e gli organismi direttamente competenti per la vigilanza bancaria dei paesi aderenti e cioè: Arabia Saudita, Argentina, Australia, Belgio, Brasile, Canada, Cina, Corea, Francia, Germania, Giappone, Hong Kong, India, Indonesia, Italia, Lussemburgo, Messico, Olanda, Regno Unito, Russia, Singapore, Spagna, Stati Uniti, Sudafrica, Svezia, Svizzera e Turchia. Il CBVB ha sede a Basilea, presso la Banca dei regolamenti Internazionali. Il CBVB non possiede alcuna autorità sovranazionale formale, infatti le sue decisioni non sono giuridicamente vincolanti. Per adempiere il proprio mandato il CBVB fa affidamento sugli impegni presi da ciascuno dei suoi membri.

La prima regolamentazione era fondata su un principio guida secondo il quale era definito un criterio uniforme per stabilire l’adeguatezza patrimoniale dei sistemi bancari, in modo da ridurre l’aggressività di alcuni istituti operanti in contesti normativi allora poco regolamentati. Giova precisare, infatti, che nel 1989 ben nove delle prime dieci banche mondiali erano giapponesi e la loro supremazia era dovuta principalmente all’utilizzo di un’elevata leva finanziaria che consentiva di ottenere alti tassi di redditività a fronte, però, di elevati rischi a cui erano sottoposte. Alla base dei requisiti minimi di capitalizzazione vi era da parte della banca l’assunzione di un impiego corrispondente alla presa in carico di un rischio il quale, oltre ad essere quantificato, doveva essere coperto da un proporzionato ammontare di capitale proprio. Originariamente l’accordo faceva riferimento solamente al rischio di credito. Tale rischio, riguarda la possibilità che la controparte diventi inadempiente. In seguito è stato introdotto anche il rischio di mercato. Questo prende in considerazione le possibili perdite inattese causate dalla variazione dei prezzi delle attività finanziarie. Il primo accordo di Basilea, inoltre, divide il capitale di vigilanza in due componenti, sulla base della loro qualità. La prima componente è chiamata *Tier 1*, rappresenta il patrimonio di base ed è composta dal capitale sociale e da riserve costituite da utili non distribuiti; la seconda componente *Tier 2*, con un limite del 50% del patrimonio complessivo, costituisce il patrimonio supplementare formato dalle riserve occulte, dal debito subordinato, dai fondi rischi e dagli ibridi di capitale e di debito. In base a quanto detto sopra, in formula abbiamo:

$$\frac{\text{Capitale di vigilanza}}{\text{Risk Weighted Asset} + 12.5 * \text{Rischi di Mercato}} \geq 8\%$$

Quindi, nel primo Accordo, il patrimonio di vigilanza doveva almeno essere pari all'8% delle attività ponderate per il rischio. Erano previsti cinque coefficienti di ponderazione fissi in funzione del tipo di prestatore, che erano così suddivisi:

- 0%: cassa, crediti verso governi centrali e banche centrali dei paesi OCSE;
- 10%: crediti verso enti pubblici;
- 20%: crediti verso banche OCSE, enti bancari internazionali e banche internazionali e banche non OCSE con durata residua inferiore ad 1 anno;
- 50%: crediti ipotecari su immobili residenziali;
- 100%: crediti e partecipazioni in imprese private ed i restanti crediti.

1.1.1 Limiti e conseguenze di Basilea I

All'accordo del 1988 va sicuramente riconosciuto l'avvio del processo di regolamentazione della vigilanza bancaria volto al raggiungimento della stabilità del sistema finanziario a livello globale. Tuttavia, essendo tale accordo il precursore di questo cambiamento, ha mostrato fin da subito i suoi limiti. Infatti, è emerso subito che le cinque classi di controparte erano troppo generiche e comprendevano attività con diversi livelli di rischio. Tale suddivisione, poco specifica, ha spinto le banche a finanziare le controparti di credito più rischiose in modo da poter garantire agli azionisti rendimenti più elevati. Pertanto, l'effetto di tale regolamentazione è stato diverso da quello sperato in quanto, a parità di capitale di vigilanza allocato, le banche costruivano portafogli più rischiosi rispetto ai precedenti, provocando una difficoltà nella ricerca di finanziamento da parte dei debitori più affidabili. Inoltre, nella classificazione in cinque classi mancano i riferimenti alla durata e alle garanzie associate al credito. Tali elementi, nel definire la rischiosezza del credito forniscono un contributo importante che, almeno in quella sede, furono trascurati. Inoltre, sono stati trascurati anche i possibili benefici da diversificare ottenibili attraverso l'adozione di strumenti derivati oppure tramite un'apposita costruzione del portafoglio impieghi.

1.2 Basilea II

I limiti dell'Accordo di Basilea I, legati principalmente all'incapacità di ridurre il rischio del sistema bancario, furono il punto di partenza per il raggiungimento dell'accordo successivo che vide quasi un decennio di elaborazioni e miglioramenti. I criteri di calcolo introdotti dal primo Accordo furono riformati così da rendere la regolamentazione quanto più sensibile al rischio di ogni prestito, anche attraverso la differenziazione di diverse tipologie di rischio per tipologie di clienti identiche. Per questo vennero introdotti i "rating", ovvero i giudizi assegnati dalla banca da cui far dipendere il requisito patrimoniale delle imprese. La versione definitiva del secondo accordo risale a giugno 2004 e si basa su tre pilastri importanti:

- *Minimum Capital Requirement*

Il primo pilastro fissa il requisito minimo di capitale di vigilanza fornendo informazioni dettagliate sul rischio di credito e introducendo il tema del rischio operativo.

- *Supervisory Review*

Il secondo pilastro definisce il ruolo delle autorità di vigilanza nazionali le quali, oltre a dover monitorare il rispetto dei requisiti minimi di capitale nel tempo, hanno la facoltà di aumentare i requisiti patrimoniali regolamentari.

- *Market Discipline*

Il terzo pilastro attiene alla divulgazione al pubblico delle informazioni riguardanti l'esposizione ai rischi e il patrimonio di vigilanza allocato. Tali obblighi di trasparenza permettevano al mercato di recepire le informazioni, elaborare e valutare le banche fissando i prezzi.

1.2.1 Primo pilastro

Il *primo pilastro* stabilisce i requisiti patrimoniali minimi che devono essere rispettati aggiungendo il rischio operativo ai rischi precedenti, ovvero il rischio di credito e di mercato. La banca è percepita come attività di impresa e, in quanto tale, è soggetta al rischio operativo. Il rischio operativo ricomprende gli errori delle risorse umane nel trattamento dei dati, nella gestione delle pratiche burocratiche legate al compimento delle operazioni finanziarie ma ricomprende anche l'utilizzo errato dei sistemi informativi. In base a quanto detto sopra, in formula abbiamo:

$$\frac{\text{Capitale di vigilanza}}{\text{Rischio di Credito} + \text{Rischi di mercato} + \text{Rischi operativi}} \geq 8\%$$

L'innovazione più importante introdotta da Basilea II riguarda le ponderazioni del rischio di credito determinate attraverso l'assegnazione delle controparti a precise classi di rating. Il nuovo accordo, per l'attribuzione delle classi di *rating*, prevede che gli istituti di credito possono fare riferimento alla valutazione effettuata dalle agenzie esterne, oppure possono utilizzare la propria classificazione se sono provvisti di un SRI (Sistema di Rating Interno) che soddisfa i requisiti minimi.

Giova precisare che vi sono diverse metodologie per il calcolo dei requisiti patrimoniali, in seguito verranno descritti due approcci:

- *Approccio standard;*
- *Sistema basato sui rating interni.*

1.2.1.1 Approccio Standard

La prima metodologia sopra citata per il calcolo dei requisiti patrimoniali consiste nella misurazione del rischio di credito tramite il *metodo standard* integrato da valutazioni esterne del merito creditizio. Il metodo Standard rappresenta un elemento chiave del Nuovo Accordo ed è assimilabile a quello previsto da Basilea I ma, in questo caso, i coefficienti di ponderazione dipendono non solo dalla tipologia del prestito ma anche dalla classe di merito della controparte, determinata sulla base dei *rating* elaborati da agenzie esterne alle banche e autorizzate dalle Autorità di Vigilanza. Nel caso in cui il *rating* di un debitore non risulti disponibile, si applicano i coefficienti di ponderazione presi da Basilea I. Nel definire tale metodo, il comitato ha attribuito un ruolo centrale alle ECAI (*Eligible External Credit Assessment Institution*) le quali assegnando i rating diventano un punto di riferimento nella determinazione del merito creditizio. Alle Autorità di vigilanza nazionali spetta stabilire quali ECAI siano affidabili. La Banca d'Italia, per esempio, ne ha riconosciute quattro: Fitch, Standard&Poor's, Moody's e Cevered Group. Inoltre, bisogna sempre accertarsi che essi rispettino diversi criteri, quali:

- **Obiettività:**
Il primo criterio prevede che la metodologia del processo di assegnazione dei rating deve essere esatta e sottoposta ad un controllo continuo. Ogni metodologia, per essere riconosciuta dalla autorità di vigilanza, deve essere applicata almeno per dodici mesi e preferibilmente per tre anni.
- **Indipendenza:**
Tale criterio prevede che una ECAI non deve essere influenzata da pressioni politiche, economiche o conflitti di interesse causati dalla composizione della struttura societaria dell'istituto di valutazione.
- **Trasparenza:**
Tale criterio richiede che tutte le valutazioni devono essere disponibili a parità di condizioni a qualsiasi istituto nazionale o portatore di interesse.
- **Pubblicità delle informazioni:**
In questo caso si fa riferimento ad informazioni che attengono alle metodologie di valutazione, alla definizione di default, all'orizzonte temporale, al significato di ciascun rating, ai tassi effettivi di inadempienza per ogni categoria di valutazioni e alle matrici di migrazione
- **Risorse:**
Per garantire un'alta qualità delle valutazioni, questo criterio prevede che la disponibilità delle risorse deve essere notevole.
- **Credibilità:**
Tale criterio è legato al rispetto dei precedenti punti e alle performance delle ECAI che sono monitorate nel tempo tramite il confronto tra i rating assegnati e la situazione reale della società valutata.

Ciò che differenziano principalmente Basilea I da Basilea II sono le notevoli modifiche effettuate alle ponderazioni delle varie esposizioni, le principali riguardano:

1. Crediti verso governi;
2. Crediti verso banche;
3. Crediti verso le imprese;
4. Crediti garantiti da ipoteche su immobili residenziali;
5. Categorie a più alto rischio.

Analizzando nello specifico le modifiche precedentemente elencate, abbiamo la seguente situazione.

Con riferimento al *punto 1*, crediti verso governi, le ponderazioni di rischio sono attribuite sulla base del rating assegnato dalle agenzie ai governi e alle loro banche centrali. La tabella che segue mostra le percentuali corrispondenti alle varie classi di *rating*.

Valutazione	Da AAA ad AA-	Da A+ ad A-	Da BBB+ a BBB-	Da BB+ a B-	Inferiore a B-	Senza Rating
Ponderazione	0%	20%	50%	100%	150%	100%

Tabella 1.1: Ponderazione per crediti verso governi

Con riferimento al *punto 2*, crediti verso le banche, sono previste due alternative. La prima assegna la ponderazione immediatamente successiva rispetto ai crediti concessi al governo del paese di riferimento. L'eccezione riguarda i paesi non quotati o con rating che vanno da BB+ a B- per i quali è prevista una soglia massima del 100%. La seconda utilizza la ponderazione assegnata dalle ECAI alle banche, attribuendo 50% a quelle sprovviste di valutazione e inoltre favorisce i crediti a breve termine con durata uguale o inferiore a tre mesi, per i quali assegna la ponderazione immediatamente precedente fino ad un minimo del 20%. Le tabelle che seguono mostrano le percentuali corrispondenti alle varie classi di *rating* nelle due opzioni.

Valutazione	Da AAA ad AA-	Da A+ ad A-	Da BBB+ a BBB-	Da BB+ a B-	Inferiore a B-	Senza Rating
Ponderazione	20%	50%	100%	100%	150%	100%

Tabella 1.2: Ponderazioni per crediti verso banche: OPZIONE 1

Valutazione	Da AAA ad AA-	Da A+ ad A-	Da BBB+ a BBB-	Da BB+ a B-	Inferiore a B-	Senza Rating
Ponderazione	20%	50%	50%	100%	150%	50%
Ponderazione per i crediti a breve termine	20%	20%	20%	50%	150%	20%

Tabella 1.3: Ponderazioni per crediti verso banche: OPAZIONE 2

Con riferimento al *punto 3*, crediti verso le imprese, anche in questo caso si utilizzano le ponderazioni sulla base del *rating* stabilito dagli istituti di valutazione del merito creditizio. Per le aziende non quotate si adotta una misura standard del 100% con il vincolo aggiuntivo di non poter ottenere una ponderazione migliore rispetto a quello dello Stato in cui operano. La tabella che segue mostra le percentuali corrispondenti alle varie classi di *rating*.

Valutazione	Da AAA ad AA-	Da A+ ad A-	Da BBB+ a BBB-	Da BB+ a B-	Inferiore a B-	Senza Rating
Ponderazione	20%	50%	100%	100%	150%	100%

Tabella 1.4: Ponderazioni per crediti verso imprese

Con riferimento al *punto 4*, crediti garantiti da ipoteche su immobili residenziali, la copertura di un tale tipo di credito consente di ridurre la ponderazione al 35%. Le autorità di vigilanza hanno il compito di verificare che questo valore di favore venga applicato solamente agli immobili residenziali.

Con riferimento al *punto 5*, categorie a più alto rischio, si applica una ponderazione maggiore o uguale al 150%. In questa categoria rientrano: i crediti verso i governi, ESP, banche, società di intermediazione mobiliare con *rating* inferiore a B-, le imprese con *rating* inferiore a BB-. Alle autorità di vigilanza nazionale spetta sempre decidere se aumentare la ponderazione nei casi di investimenti *venture capital* e *private equity*.

1.2.1.2 Sistema basato sui rating interni

Mentre il metodo standard, precedentemente analizzato, si rivolge alle banche di minore dimensione, il metodo basato sui rating interni è stato progettato per le banche più grandi che dispongono competenze e risorse sufficienti a progettare e implementare una metodologia interna di *rating*. Pertanto, tale metodo per il calcolo dei requisiti patrimoniali minimi si basa sull'utilizzo dei sistemi di rating interni (SRI). Per sistema di rating si intende l'insieme dei metodi, procedimenti, controlli, dati e sistemi informativi che fungono da supporto alla valutazione del rischio di credito, all'attribuzione dei gradi interni di merito e alla stima quantitativa delle inadempienze e delle perdite delle singole operazioni creditizie. Tale metodologia può essere utilizzata come alternativa alla precedente solamente qualora rispetti i seguenti criteri:

- Valuta separatamente la PD e la LGD;
- La distribuzione dei crediti sulle varie classi avviene senza particolari concentrazioni;
- Il rating è controllato periodicamente;
- Il rating è utilizzato dalla banca nella gestione dei crediti e nel pricing dei prestiti;
- L'intermediario deve possedere un sistema di validazione dell'accuratezza e coerenza del SRI e deve rispettare i requisiti di documentazione formale sul suo funzionamento.

Vista la loro importanza, la banca d'Italia ha voluto specificare e integrare i precedenti requisiti nel documento “nuove disposizioni di vigilanza prudenziale per le banche” (circolare n.263 del 27 dicembre 2006), definendo altri sei requisiti di seguito analizzati.

- Documentazione del sistema di rating
Gli intermediari devono documentare tutti i dettagli che sono legati al funzionamento del proprio sistema di *rating* interno, spiegandone le caratteristiche principali e i fondamenti teorici del modello. Nello specifico, vanno documentate le definizioni di default e di perdita adottate. Inoltre, devono essere registrate e conservate tutte le informazioni che

consentono di ricostruire il percorso di assegnazione del *rating* e le attività di controllo effettuate su di esso.

- **Completezza delle informazioni**

Le banche devono adottare delle metodologie interne in grado di valutare la completezza e la rilevanza dei dati e delle informazioni impiegate nel loro sistema interno di rating.

- **Replicabilità**

Agli intermediari è altresì richiesto di conservare le decisioni prese durante il percorso di assegnazione della valutazione, in modo tale da poter ripercorrere il processo di assegnazione del rating ed eventualmente ricalcolarlo per le singole posizioni.

- **Integrità**

Nel caso in cui sorga un conflitto di interessi, dovuto a situazioni in cui i soggetti responsabili dell'attribuzione del rating svolgano un'attività connessa al raggiungimento di obiettivi in termini di volumi o ricavi oppure abbiano il potere di decidere sull'erogazione del credito, devono essere previste cautele adeguate sul piano organizzativo e procedurale per evitare proprio che l'assegnazione del *rating* possa essere condizionata da tali soggetti.

- **Omogeneità**

I sistemi utilizzati dalle banche devono garantire che debitori ed attività che comportano rischi simili siano valutati allo stesso modo e, quindi, assegnate alla stessa classe di rating.

- **Univocità**

Tale carattere deve riguardare sia i debitori che le operazioni. Una volta riconosciuto che si tratta di una stessa controparte piuttosto che di una stessa attività, ogni esposizione deve essere valutata uniformemente.

1.2.2 Secondo pilastro

Il secondo pilastro dà particolare importanza al processo di controllo prudenziale svolto dalle Autorità di Vigilanza nazionali. Basilea II fissa quattro principi chiave che guidano

il processo di controllo prudenziale: controllo prudenziale dell'adeguatezza patrimoniale, vigilanza, prudenzialità e vigilanza preventiva.

- Le banche devono adottare delle metodologie di verifica dell'adeguatezza patrimoniale in riferimento al profilo di rischio adottato. Infatti, a tal proposito, dovrebbero essere effettuati *stress test* in modo da poter individuare aprioristicamente scenari negativi per la banca.
- Le autorità di vigilanza dovrebbero riesaminare regolarmente i processi interni di valutazione dell'adeguatezza patrimoniale, dei rischi assunti e la qualità del capitale posseduto.
- Le autorità di vigilanza devono evitare che il patrimonio di una banca scenda sotto la soglia corrispondente al suo profilo di rischio. Ciò deve essere previsto tramite l'imposizione di azioni correttive quali l'intensificazione dell'attività di vigilanza, le restrizioni al pagamento dei dividendi o di un piano di reintegro del patrimonio,
- Le autorità di vigilanza devono poter richiedere alle banche di detenere un patrimonio superiore rispetto allo standard minimo obbligatorio. Tale margine aggiuntivo serve principalmente per tener conto delle incertezze che colpiscono i vari mercati e il settore bancario.

L'applicazione di questi principi richiede una continua interazione tra banca e Autorità di Vigilanza denominata da Basilea II "processo di controllo prudenziale" (*Supervisory Review Process – SRP*) e che si articola in due momenti successivi e complementari:

- Processo interno di determinazione dell'adeguatezza patrimoniale (*Internal Capital Adequacy Assessment Process – ICAAP*): in questo caso la banca si dota di una struttura interna, con responsabilità ben individuate, che gestisce le procedure di determinazione dei requisiti minimi e, più in generale, del rischio complessivo dell'attività bancaria;
- Processo di revisione e valutazione prudenziale (*Supervisory Review and Evaluation Process – SREP*): una volta che la banca si è dotata delle strutture di cui alla prima fase, le Autorità di Vigilanza esprimono un giudizio sull'adeguatezza delle stesse e, se è il caso, richiedono anche l'adozione di misure correttive.

1.2.3 Terzo pilastro

Tale pilastro, invece, stabilisce i requisiti informativi che devono essere resi noti affinché gli operatori del mercato finanziario possano valutare l'operatività, le esposizioni ai rischi e l'adeguatezza del patrimonio. La pubblicazione delle informazioni rende più sicure e affidabili le banche. Infatti, tale pilastro si fonda sull'idea che mercati efficienti e correttamente informati circa la situazione economico-finanziaria della banca possano agire da disciplina sui suoi comportamenti ed evitando, così, l'assunzione di posizioni eccessivamente rischiose da parte della stessa. Ciò si traduce nell'obbligo da parte delle banche di fornire tutta una serie di dati riguardanti la gestione, la concentrazione dei rischi, la composizione del patrimonio di vigilanza, i metodi di *rating* e le misure adottate per contenere il rischio. È fatta salva la possibilità di omettere informazioni riservate che potrebbero favorire la concorrenza o non comunicabili a causa di vincoli giuridici. Qualora non vengano rispettati i requisiti di trasparenza può portare le autorità di vigilanza ad adottare dei provvedimenti che vanno da un semplice dialogo con le direzioni bancarie o semplici richiami fino a sanzioni pecuniarie. Le informazioni fornite devono rispettare la frequenza semestrale e il criterio di rilevanza poiché la mancanza o errata indicazione di un'informazione rilevante è in grado di modificare il giudizio o le decisioni degli utenti che ne sono a conoscenza.

1.2.4 Limiti e conseguenze di Basilea II

L'accordo di Basilea II, entrato in vigore nello stesso periodo in cui è iniziata la crisi del 2008, è considerato uno dei colpevoli della severità della crisi finanziaria. A dichiararne i suoi limiti è stato il suo stesso Comitato che li ha riassunti in sei punti chiave.

- Qualità e livello del capitale
È emerso che gli intermediari che sono stati salvati grazie agli enti governativi possedevano un patrimonio superiore al minimo imposto dalla regolamentazione. Ciò ha messo in dubbio che la qualità e la quantità del capitale necessario per prevenire lo stato di insolvenza bancaria fosse in realtà di molto superiore a quanto fissato da Basilea II. Giova precisare che tale affermazione risulta solo parzialmente vera, poiché il patrimonio delle banche divenute insolventi aveva una quota considerevole di strumenti finanziari ibridi. Quest'ultimi sono stati introdotti a seguito di

una consistente domanda da parte degli investitori desiderosi di ottenere rendimenti sicuri investendo in strumenti di debito per le banche.

- Procciclicità

Tale termine si riferisce alle dinamiche del sistema finanziario che tendono ad ampliare le fluttuazioni cicliche e quindi a migliorare le fasi espansive e, di contro, aggravare quelle recessive. La regolamentazione, infatti, in corrispondenza delle fasi recessive richiede un aumento dei requisiti provocando una contrazione dell'offerta di credito ed accentuando la fase ciclica negativa. Se si adotta una visione sulla singola banca, il ragionamento appare sensato ma se si guarda all'interno sistema finanziario emerge che ad una contrazione del credito da parte di tutte le banche corrisponde un aumento della fase recessiva. Questo principalmente avviene perché accresce il rischio di default dei debitori causando una situazione di difficoltà per le banche.

- Incremento incontrollato della leva finanziaria

Nel momento antecedente la crisi del 2000 la situazione finanziaria vedeva le banche adottare livelli elevati di leva finanziaria, sempre nel rispetto dei limiti regolamentari. L'arrivo della crisi e la conseguente contrazione della concessione del credito hanno spinto le banche ad accrescere il patrimonio riducendo i propri attivi. La velocità di tale processo, definito *processo di deleveraging*, ha sicuramente contribuito ad accrescere l'instabilità dei mercati finanziari.

- Liquidità

Il rischio di liquidità per le banche è emerso già prima della crisi del 2008. Questo allora era dovuto principalmente alla globalizzazione dei gruppi finanziari, alle cartolarizzazioni, alla tecnologia e alla concentrazione dei grandi gruppi finanziari. L'aumento del rischio di controparte e la conseguente perdita di fiducia provocata dalla crisi ha causato considerevoli agitazioni ai singoli intermediari.

- Banche sistemiche

Gli enti governativi, durante la crisi, sono intervenuti per salvare diverse banche e assicurazioni. Ciò è stato fatto principalmente per evitare una crisi sistemica che avrebbero provocato, poiché la loro interconnessione con le

altre istituzioni finanziarie era tale per cui un loro fallimento avrebbe un impatto ben più ampio della crisi dell'intermediario stesso.

- Rischi di mercato sul *trading book*

Durante la crisi del 2008 si sono registrate importanti svalutazioni delle attività che componevano i *trading book* delle banche, uno dei motivi del crollo del valore dei portafogli di negoziazione è il fatto che tali attività erano ricondotte al *fair value*. È emersa la fragilità dei nuovi requisiti minimi fissati dalla regolamentazione soprattutto a seguito della caduta dei mercati, i quali non sono più stati in grado di ricoprire le perdite. Una delle cause della loro leggerezza è da ricercare nei modelli utilizzati per il calcolo dei rischi, questi sono risultati talmente reattivi da rimuovere troppo velocemente gli episodi di crisi e considerare solo quelli riferiti agli scenari ottimistici precrisi.

1.3 Basilea III: la regolamentazione attuale

Basilea III è arrivata in un momento in cui bisognava sopperire al periodo buio dell'economia mondiale causato dalla crisi finanziaria del 2008. Infatti, come è stato detto nel precedente paragrafo, la crisi del 2008 ha fatto emergere tutti i limiti di una regolamentazione - Basilea II - che non è stata in grado di garantire la sopravvivenza degli intermediari rendendo necessario il salvataggio da parte degli enti governativi.

Affrontando brevemente la questione della crisi del 2008, va sottolineato che il punto d'origine di tale crisi è rappresentato dal collasso del mercato bancario statunitense dei mutui *subprime* che, attraverso il settore bancario ombra si è trasmesso ad altri intermediari finanziari non bancari e quindi all'intero sistema finanziario. Molti dei mutui *subprime* erano stati cartolarizzati e trasferiti dal sistema bancario a tutta una serie di intermediari. La difficoltà di alcune banche nel riscuotere i mutui rimasti nel loro portafoglio ha fatto sorgere dubbi anche sulla consistenza dei titoli aventi come sottostanti tali mutui nelle operazioni di cartolarizzazione. Ciò ha provocato un aumento dei tassi di interesse e una drastica riduzione della liquidità, spingendo molti governi ad intervenire con salvataggi e in alcuni casi con il passaggio di proprietà delle banche a favore dello Stato. Come ulteriore conseguenza di ciò si è verificata la crisi dei debiti sovrani, sebbene tale crisi - come nel caso dell'Italia - abbia ragioni più profonde che risiedono

nell'elevato rapporto debito/PIL e nell'instabilità del sistema politico. La crisi dei mutui *subprime* si estende all'intero mercato finanziario a causa di un noto fenomeno: *shadow banking*. Con questo termine si intende una forma di intermediazione finanziaria che si svolge al di fuori del sistema bancario regolamentato. Secondo il Financial Stability Board (2011) lo *shadow banking* può essere definito come: “un sistema di intermediazione creditizia che coinvolge entità ed attività esterne al sistema bancario regolare, e comporta un rischio sistemico e un rischio di arbitraggio regolamentare”. Lo *shadow banking*, sostanzialmente, trasforma attività opache, rischiose e a lungo termine in passività a breve termine e altamente liquide. Tuttavia, proprio per il fatto di finanziare attività a lungo termine con passività a breve termine, rende questo sistema molto fragile. Infatti, con la crisi del 2008 molte “banche ombra” sono fallite. Inoltre, successivamente alla crisi finanziaria del 2008 e la conseguente crisi dell'economia reale, i tassi di crescita del PIL si sono ridotti diventando addirittura negativi. Ciò è avvenuto insieme alle spese necessarie al salvataggio delle banche e l'attuazione degli stabilizzatori automatici previsti per fronteggiare la crisi economica, aggravando la situazione della finanza pubblica in gran parte dei Paesi. La Grecia, il Portogallo, l'Irlanda e l'Italia, che già presentavano un rapporto debito/Pil elevato, hanno sperimentato una crisi da parte dei mercati finanziari. Ciò, invece, non è accaduto in Francia o Germania, in quanto il rapporto debito/PIL era più accettabile e sostenibile nel periodo pre-crisi. Sebbene sia azzardato instaurare una relazione causa-effetto diretta tra crisi finanziaria e crisi dei debiti sovrani, nondimeno la prima può essere considerata uno dei fattori rilevanti.

Da ciò emerge che in quegli anni era forte l'esigenza dei presidenti delle Autorità di Vigilanza degli Stati membri di rendere ancora più stringenti gli accordi precedenti. Per questi motivi, il Comitato si è riunito ancora una volta per emanare i provvedimenti di Basilea III. Anche questo accordo si basa sulla struttura a tre pilastri ma ne rinforza le basi. In particolare, interviene su ogni pilastro, fissa le direttive per le istituzioni finanziarie di rilevanza sistemica e i requisiti globali di liquidità. I cambiamenti al precedente accordo di Basilea II riguardano principalmente i requisiti patrimoniali minimi i quali diventano più rigidi sia in termini di qualità che di quantità e questo avviene per assicurare un maggiore assorbimento delle perdite. Le modifiche sono notevoli e, per una trattazione specifica e agevole, vengo riassunti nei seguenti sei punti.

- Capitale

Basilea III aumenta la qualità e il livello del patrimonio di vigilanza ridefinendo la composizione del *Tier 1*, *Tier 2* ed eliminando il *Tier 3*. In particolare, il patrimonio di base deve non deve essere composto da strumenti ibridi innovativi ma deve essere principalmente composto da azioni ordinarie e riserve di utili non distribuiti (*common equity*). Inoltre, il livello massimo del *Tier 1* passa dal 4% al 6% delle attività ponderate per il rischio e il requisito minimo del *core Tier 1* è innalzato a 4.5% delle attività pesate per il rischio al netto degli aggiustamenti. Il *Tier 2* copre il 2% che, sommato al *Tier 1*, permette di raggiungere l'8% regolamentare.

Dopodiché, il nuovo accordo prevede che le banche debbano possedere un *buffer* di conservazione del capitale del 2.5% formato da *common equity*, portando la quota totale di quest'ultimo al 7%. La costruzione di tale *buffer* è garantita da limitazioni nella distribuzione degli utili da parte delle banche che non lo hanno ancora adottato. Il ruolo principale è quello di incoraggiare l'accumulo di riserve patrimoniali nei periodi economici particolarmente favorevoli e di fornire una possibile copertura aggiuntiva nei periodi di crisi. È stato altresì introdotto il *buffer* anticiclico, anch'esso composto da *common equity* in misura compresa tra lo 0 e il 2.5%. Il suo ruolo principale è quello di proteggere gli intermediari dalle possibili perdite causate da una politica di espansione del credito eccessiva. Sono le autorità nazionali a decidere sull'applicazione del *buffer*, i quali sono incaricati a valutare il surriscaldamento del ciclo creditizio sulla base della differenza tra prestiti bancari e il trend di lungo periodo del PIL.

- Copertura dei rischi

Basilea III stabilisce che le banche debbano adottare un trattamento patrimoniale più cautelativo e che siano effettuate analisi più rigorose del merito creditizio. Questo deve avvenire soprattutto nei confronti di quegli strumenti finanziari innovativi responsabili della crisi del 2008 come le cartolarizzazioni provviste di *rating* esterno e i derivati per i quali il portafoglio di negoziazione deve essere valutato tenendo conto di un

rischio incrementale derivato dai rischi di insolvenza, di migrazione di rating e di liquidità.

- Contenimento della leva finanziaria

Obiettivo fondamentale di tale accordo è quello di controllare la crescita dei livelli della leva finanziaria, infatti tale regolamentazione ha introdotto *leverage ratio*. Tale indice è calcolato come rapporto tra patrimonio di base *Tier 1* e il totale dell'attivo comprendente le esposizioni fuori bilancio. Tale indice, per evitare forti *deleveraging*, non potrà scendere al di sotto del 3%.

$$\text{Leverage ratio} = \frac{\text{Tier 1}}{\text{Attivo comprendente fuori bilancio}} \geq 3\%$$

- Liquidità

La crisi del 2008 ha fatto emergere importanti problematiche relative alla liquidità dei mercati finanziari e ciò ha spinto ad introdurre nell'accordo di Basilea III due nuovi indicatori di liquidità minima.

Il primo coefficiente è chiamato *Liquidity coverage ratio* e ha lo scopo di garantire la presenza di attività liquide di alta qualità (ALAQ) in grado di generare fondi per superare una fase di eventuali deflussi negativi nel breve periodo. Le attività che rientrano nel calcolo del numeratore sono attività con basso rischio di credito, di mercato, aventi valutazione certa e con una bassa correlazione con attività rischiose. Il denominatore, invece, fa riferimento alla differenza tra i flussi di cassa cumulati in uscita e in entrata in un periodo di stress di 30 giorni. L'indice deve assumere un valore superiore ad uno ed è così calcolato:

$$\text{Liquidity coverage ratio} = \frac{\text{ALAQ}}{\text{Deflussi di cassa (30 giorni)}} > 1$$

Il secondo coefficiente è il *Net stable funding ratio* e ha il compito di assicurare l'equilibrio tra le fonti di finanziamento e il fabbisogno a medio-lungo termine determinato dalla scadenza. La stabilità è garantita da una maggiore quantità di risorse finanziarie stabili (AFS) rispetto al

fabbisogno di risorse stabili richiesto dalla formazione dell'attivo (RSF). Il numeratore è formato da: *Tier 1*, *Tier 2*, azioni privilegiate e passività con scadenza maggiore o uguale ad un anno e passività con scadenza inferiore ad un anno per le quali la banca si aspetta un rinnovo oltre i dodici mesi. Tutte le componenti sono ponderate con un coefficiente che rappresenta la stabilità della fonte. Tale coefficiente va dallo 0% per le passività con scadenza inferiore all'anno fino al 100% per le più stabili. In generale, le ASF rappresentano tutte le fonti di finanziamento delle quali si potrà fare uso in condizioni di stress. Il denominatore è una stima del fabbisogno di risorse stabili connesso alla struttura dell'attivo e anche in questo caso, agli elementi che lo compongono è applicata una ponderazione che riflette il livello di liquidità dell'attività in esame dando maggior peso alle attività più difficilmente liquidabili.

L'indice, pertanto, è il seguente:

$$\text{Net stable funding ratio} = \frac{\text{ASF}}{\text{RSF}} > 1$$

- Istituzioni finanziarie di importanza sistemica

Gli intermediari che rientrano in questa categoria devono avere una maggiore capacità di assorbimento delle perdite poiché un possibile loro default risulterebbe assai più grave di quello di altri intermediari non rientranti in tale categoria. Proprio per la loro importanza, il Comitato ha fissato i criteri quantitativi e qualitativi per poterle classificare ed assegnare un coefficiente patrimoniale progressivo, che varia dall'1% al 2.5% da soddisfare con *Tier 1*.

- Gradualità

Nell'accordo di Basilea III è interessante anche osservare le modalità e le tempistiche con le quali il Comitato ha stabilito l'entrata in vigore per la stabilità del sistema economico. I requisiti sono entrati con tempistiche e gradualità diverse, con diverse fasi di sperimentazione e introduzione graduale. La gradualità è dovuta a diversi fattori, in primis riguarda la diversità con la quale gli intermediari dei vari paesi hanno ricevuto gli aiuti da parte dello Stato. In alcuni paesi, le autorità sono intervenute

immettendo massicce dosi di liquidità mentre in altri il governo non è stato in grado di contribuire al loro salvataggio. Basilea III, in particolare, è stata introdotta in modo graduale e diversificato in modo da evitare impatti negativi sulla crescita economica dei Paesi.

1.3.1 Considerazioni su Basilea III

L'accordo di Basilea III ha inciso in maniera positiva sulla quantità di credito disponibile. Infatti, gli incrementi dei requisiti minimi di capitale e l'introduzione dei vari *buffer* hanno agito in tal senso, aumentando la capacità di assorbimento di capitale e diminuendo il ricorso a forme di debito meno onerose.

Altro aspetto positivo è sicuramente rappresentato dalla maggiore qualità del capitale di *Tier 1* che aumenta dal 2% al 7% evidenziando l'incapacità nell'assorbimento delle perdite da parte degli strumenti ibridi e innovativa che non rientrano nel patrimonio di base.

Altro punto di forza di tale regolamentazione sono le considerazioni sulle fasi del ciclo economico. Infatti, si è deciso di sfruttare meglio i periodi positivi accumulando risorse destinati a sopperire alle perdite dei periodi più negativi del ciclo economico.

1.4 Basilea IV: uno sguardo al futuro

Nel 2017 il Comitato si è riunito per la quarta volta per dare vita alle riforme regolamentari che entreranno in vigore dal 2021 al 2027. Pertanto, Basilea IV non è ancora in vigore ma in questa sede verranno delineate solo gli scopi e le proiezioni e non i risultati e le conclusioni, come si è fatto per gli altri accordi. Innanzitutto, lo scopo principale di tale accordo è quello di aumentare l'attendibilità ai calcoli delle attività ponderate per il rischio (*Risk Weighted Asset* o RWA) e rendere più facilmente confrontabili gli indicatori patrimoniali delle banche. Queste nuove regole nascono da risultato di numerose analisi che hanno evidenziato una variabilità significativa tra gli RWA di diverse banche aventi una simile rischiosità di portafoglio. È emerso che gli intermediari utilizzano i loro modelli interni non per quantificare i rischi. Ma per minimizzare i requisiti patrimoniali. Infatti, sono stati introdotti dei limiti all'uso dei modelli interni, nello specifico:

- La sensitività dei rischi dell'Approccio Standard è stata accresciuta mediante l'adozione di una modifica dei fattori di ponderazione degli attivi più articolata e specifica;
- Le banche devono aggiornarsi maggiormente tramite specifiche ricerche circa l'attendibilità delle valutazioni espressa dai rating esterni per sviluppare un approccio *non-rating-based*. Per gli intermediari la cui attività è in paesi che non consentono di utilizzare la valutazione dei rating esterni possono applicare una ponderazione del 65% alle imprese *investment grade* e una al 100% alle restanti.
- Per evitare l'accumulo dei rischi eccessivi nell'uso dei sistemi di rating interno sono fissati dei livelli minimi di PD, LGD ed EAD.
- Inoltre, saranno anche introdotti degli *output floor* per stabilire che il valore complessivo degli RWA sia uguale al maggiore tra gli RWA ottenuti con gli approcci consentiti dalla regolamentazione e il 72,5% del totale degli RWA calcolati con il metodo standard.
- Le SIFI devono adottare un ulteriore leverage ratio buffer per evitare un eccessivo ricorso alle fonti di debito.

I possibili impatti di tale accordo hanno mostrato che molto probabilmente le banche dovranno aumentare il proprio capitale. Ciò vale soprattutto per gli istituti di importanza sistemica e per le banche che utilizzano un sistema di rating interno poiché sono proprio loro l'oggetto della regolamentazione. Tale accordo interesserà direttamente anche gli enti che si occupano dell'erogazione di mutui e dei leasing immobiliari in quanto dovranno applicare le nuove ponderazioni per il calcolo degli RWA. In conclusione, emerge che le nuove riforme renderanno sempre meno elastica l'operatività delle banche, incentivando l'utilizzo dei mezzi proprio o di strumenti alternativi di accesso al credito.

CAPITOLO II – Il Rischio di Credito

Tale capitolo, che volutamente segue la regolamentazione di Basilea, è fortemente ancorato a quello che ancora oggi rappresenta il principale fattore delle crisi bancarie: il rischio di credito. Al rischio di credito, come vedremo, possono essere associate diverse tipologie di rischio, quali: il rischio di insolvenza, il rischio di migrazione, il rischio di recupero, il rischio di esposizione, il rischio di spread e il rischio paese. Dopodiché verranno analizzate, sia descrittivamente che analiticamente le componenti del rischio di credito, ossia la perdita attesa e la perdita inattesa. Il supporto grafico, oltre a quello formulare, fornisce un supporto indispensabile per analizzare meglio tali componenti. Dopodiché, si prosegue con la presentazione del processo di valutazione di solvibilità effettuato dalle banche, noto come *credit scoring*, esaminando il modello di regressione logistica, per poi fare una panoramica sugli attuali modelli di *machine learning*.

2.1 Il rischio di credito

Il rischio di credito è definito come la possibilità che una variazione inattesa del rischio di merito di credito di una controparte, nei confronti della quale esiste un'esposizione, generi una corrispondente variazione inattesa del valore di mercato del credito. Da tale definizione si intuisce la necessità che la variazione del merito creditizio sia inattesa poiché, se così non fosse, sarebbe già considerata al momento della valutazione ed inclusa nel *pricing* del credito. La definizione evidenzia inoltre che lo stato di insolvenza rappresenta solamente l'evento estremo del rischio di credito, il quale va considerato in modo più ampio facendo riferimento ad una completa distribuzione di probabilità di eventi intermedi che individuano le diverse classi di rischio. Il rischio di credito, come già accennato nel precedente capitolo, è valutato mediante l'attribuzione di un rating, ossia un giudizio sulla capacità di un emittente o di un'emissione di far fronte agli impieghi finanziari entro certe scadenze. Tale giudizio è espresso da agenzie di rating sulla base di informazioni recapitate dal mercato o attraverso sistemi di rating interni.

Il rischio di credito è complesso da analizzare perché ad esso possono essere associate diverse tipologie di rischio:

- **Rischio di insolvenza**

Attiene al rischio che il debitore non sia in grado di adempiere ai suoi obblighi di pagamento di interessi e di rimborso del capitale. In tal caso, il creditore o l'investitore può subire una perdita anche totale.

- **Rischio di migrazione**

Rappresenta il rischio di un deterioramento del merito creditizio di una controparte (downgrade). Tale deterioramento può trovare riscontro concreto in un declassamento del rating del debitore ad opera di un'agenzia o ad opera degli analisti fidi della banca creditrice.

- **Rischio di recupero**

Si riferisce alla possibilità che il tasso di recupero connesso alle esposizioni nei confronti di controparti divenute insolventi si riveli inferiore a quanto originariamente stimato dalla banca. Tale diminuzione può trovare origine in diversi fattori, quali ad esempio un allungamento dei tempi connessi alle procedure giudiziali, un aumento dei tassi di interesse o una diminuzione del valore dei beni a garanzia.

- **Rischio di esposizione**

È relativo all'effettivo ammontare del prestito al momento dell'insolvenza. Si tratta di un rischio tipico delle linee di credito a valore aleatorio. Dove l'Exposure At Default (EAD) può differire, anche in modo sensibile, da quella corrente. Un esempio è lo scoperto di conto corrente dove il cliente è libero di utilizzare in misura variabile il fido accordato.

- **Rischio di spread**

Definito come la probabilità che, a parità di merito di creditizio, aumenti il premio al rischio, ossia lo spread, richiesto dal mercato di capitali.

- **Rischio paese**

Rappresenta il rischio di insolvenza di operatori, pubblici e privati, legato all'area geografica di provenienza e indipendente dalla loro volontà. Si tratta di paesi caratterizzati da diverse forme di instabilità, quali politica, legislativa, economica, sociale.

2.2 Le componenti del rischio di credito

Il rischio di credito, sulla base della definizione fornita nel paragrafo precedente, è distinto da due componenti: la perdita attesa (Expected Loss o EL) e la perdita inattesa (Unexpected Loss o UL):

- La perdita inattesa rappresenta la parte vera e propria di rischio in quanto non prevedibile dal modello impiegato e dunque rischiosa;
- La perdita attesa, invece, è già inclusa nelle attività di *pricing* della banca in quanto risulta nota a priori.

Graficamente:

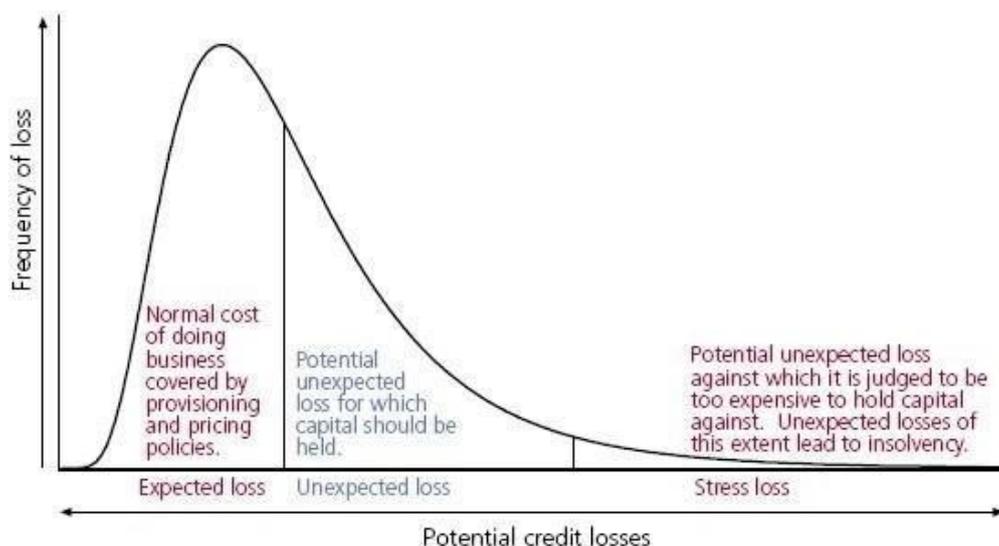


Figura 2.1: Componenti del rischio di credito in funzione della quantità e della frequenza

2.2.1 La perdita attesa

La perdita attesa è definita come il valore medio della distribuzione delle perdite che una banca si attende di conseguire su un singolo credito o su un portafoglio di crediti. Essendo attesa la banca si aspetta *ex-ante* di dover sostenere tale costo, poiché rappresenta una componente stabile.

La perdita attesa è calcolata come prodotto di tre elementi:

- l'esposizione in caso di insolvenza (*Exposure at Default* o EAD);
- La probabilità di insolvenza (*Probability of Default* o PD);
- La percentuale di perdita in caso di insolvenza (*Loss Given Default* o LGD).

In formula:

$$EL = EAD * PD * LGD$$

2.2.1.1 Exposure At Default

L'EAD rappresenta l'ammontare di perdita che grava sulla banca al momento del default. Dal momento che l'insolvenza in un periodo futuro non è sempre facile conoscere a priori questo valore. Sulla base di tale premessa, si possono distinguere due tipologie di esposizioni:

- A valore certo: la banca conosce il valore dell'esposizione *ex-ante* il valore esatto del finanziamento, un esempio è il mutuo.
- A valore incerto: la banca conosce il valore dell'esposizione con esattezza solo *ex-post*, ossia dopo il default. Un esempio è il fido di conto corrente.

Quindi, quando le esposizioni sono a valore certo, il valore dell'EAD è determinabile risalendo alla quota di debito non ancora restituita dal cliente, mentre nel secondo caso la presenza di una componente aleatoria non rende immediato il calcolo. La stima dell'EAD per le esposizioni a valore incerto deve necessariamente passare dalla determinazione di due quote diverse, ovvero quella di fido utilizzata (*Draw Portion* o DP) e quella ancora non utilizzata (*Undrawn Portion* o UP). A quest'ultima è associata un'altra variabile che rappresenta la percentuale di quota di fido non ancora utilizzata che si ritiene verrà utilizzata da debitore (*Usage Given Default* o UGD).

In formula:

$$EAD = DP + UP * UGD$$

2.2.1.2 Probability of Default

La PD è definita come la possibilità che la controparte diventi inadempiente agli obblighi che derivano dalla sua esposizione, in tal senso la PD è strettamente connessa al merito creditizio del debitore.

Per il calcolo della PD si possono distinguere tre differenti metodologie: la prima prevede che informazioni rilevanti siano estrapolate dai dati di mercato dei capitali; la seconda utilizza modelli analitico/soggettivi in grado di valutare entrambi gli aspetti quantitativi e qualitativi; la terza metodologia si avvale dei rating creditizi emanati da agenzie specializzate oppure dai sistemi di rating interni.

2.2.1.3 Loss Given Default

La LGD rappresenta la perdita subita dall'impresa finanziatrice nel momento in cui la controparte diventa insolvente. Analogamente, si può dire che la LGD è considerata come il complemento del *Recovery Rate* (RR), ossia il tasso che l'istituto di credito riesce a recuperare in seguito al default del debitore. Quindi, in altre parole, la perdita in caso di insolvenza è il tasso di perdita subito dai creditori in caso di insolvenza dell'impresa.

In formula:

$$LGD = 1 - RR$$

Dove:

- RR = Tasso di recupero (Recovery Rate)

Il *tasso di recupero* (RR) è influenzato da diversi fattori, per esempio dalle caratteristiche del finanziamento quindi dalla durata, dal grado di subordinazione rispetto ad altri creditori, dalla tipologia di contenzioso prevista per il recupero, dall'esistenza di garanzie (collaterals) reali o finanziarie, nonché della loro liquidità ed efficacia. Può essere altresì influenzato dalle caratteristiche dell'impresa finanziata, quindi in base all'area geografica e al settore produttivo di provenienza del debitore. La prima definisce la tipologia, l'efficacia e la rapidità della procedura di recupero utilizzata, mentre il settore produttivo influisce sul grado di liquidità dell'impresa. Inoltre, può essere influenzato dalle caratteristiche dell'impresa finanziatrice, quali i servizi legali interni e le politiche di recupero crediti che influenzano la rapidità e l'importo recuperato. Infine, i fattori esterni

vanno anch'essi ad influenzare il tasso di recupero. Con fattori esterni si intende la fase corrente del ciclo economico e il livello dei tassi di interesse che gravano sul valore attuale dell'importo recuperato a seguito dell'insolvenza del debitore.

Per considerare e soddisfare tutte le caratteristiche sopracitate i due approcci più utilizzati sono il *Market LGD* e il *Workout LGD*.

Il primo è un modello *market based* che basa la stima della LGD sul delta di prezzo di mercato dello strumento rappresentativo dell'esposizione prima e dopo la notizia di default del suo emittente. Tale teoria si basa sul presupposto che il mercato una volta recepite le nuove informazioni le rielabora e assegna un nuovo prezzo comprensivo delle aspettative sull'entità e sui tempi di recupero dei creditori e sulla possibilità da parte dell'intermediario di poter vendere immediatamente lo strumento finanziario relativo all'esposizione.

Il secondo, l'approccio *Workout LGD*, calcola il valore della LGD misurando i reali flussi di cassa recuperati in seguito all'evento di insolvenza considerando tutti gli elementi che determinano il Recovery Rate.

In formula:

$$RR = \frac{\sum_{t=1}^T \frac{ERt - ACt}{(1+i)^t}}{EAD}$$

Dove:

- i = tasso di attualizzazione:
 - Tasso di interesse di trasferimento di fondi (=costo marginale del "funding" della banca);
 - Tasso contrattuale del finanziamento andato in default (tale tasso non può riflettere il rischio del prestito post – default);
 - Tasso congruo per rischio, tenuto conto dei rischi su ER;
 - Tasso risk-free: nell'ipotesi in cui si assuma che il rischio (sistemico) e il premio per l'avversione ad esso siano già inclusi

in altri parametri (down LGD, asset correlations implicite nella regulation, etc);

- T = durata del processo di recupero;
- ER_t = *Expected Recovery*, importo recuperato nel periodo t ;
- AC_t = *Administrative Cost*, l'ammontare dei costi amministrativi sostenuti nel periodo t ;
- EAD = *Exposure at Default* – esposizione al momento del default;
- n = periodo di tempo stimato per realizzare il recupero.

2.2.2 La perdita inattesa

La perdita inattesa rappresenta una misura del grado di dispersione del tasso di perdita attorno al proprio valore atteso, quindi attorno all'*Expected Loss*. Tale componente rappresenta la vera fonte del rischio di credito in quanto la rischiosità è legata al verificarsi di un deterioramento inatteso della qualità creditizia.

Graficamente:

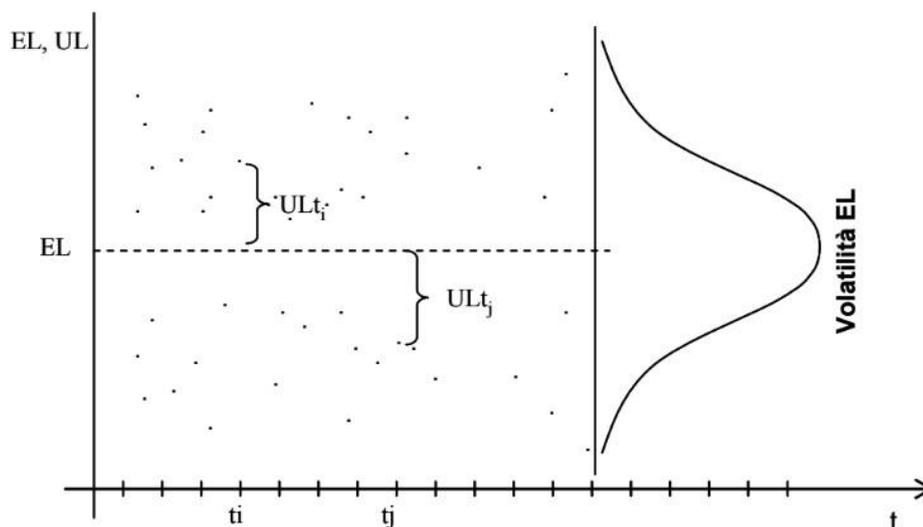


Figura 2.2: *Expected ed Unexpected Loss in funzione del tempo (t)*

Data la natura variabile della perdita inattesa, questa non può essere contabilizzata in conto economico ma deve trovare un'adeguata copertura nel patrimonio. Viceversa, la perdita attesa è contabilizzata in conto economico considerata come semplice voce di costo già incluso nelle aspettative dell'intermediario. Un'altra differenza rilevante è

rappresentata dalla possibilità di attenuazione, in particolare mentre la perdita inattesa può essere ridotta tramite un'opportuna diversificazione del portafoglio, la perdita attesa non può essere mitigata ma solo stabilizzata mediante un aumento del volume delle operazioni prese in considerazione per il suo calcolo. Quindi, la perdita attesa complessiva di un portafoglio è data dalla somma delle perdite attese sulle singole esposizioni, mentre la quantità da detenere in seguito all'aggregazione di più perdite inattese su più posizioni può risultare inferiore alla somma delle perdite inattese sulle singole posizioni grazie ai benefici di diversificazione.

Per valutare la perdita inattesa basta considerare un modello binomiale che prevede due soli eventi: default e non default.

	EVENTI	
	Default	Non default
Probabilità	PD	1-PD
Perdita	LGD	0

Tabella 2.4: Eventi con relative probabilità e perdite

In questo modo, le due perdite possono essere calcolate:

$$EL = PD * LGD + (1 - PD) * 0$$

$$UL = LGD * \sqrt{PD(1 - PD)}$$

L'espressione della volatilità (UL) ipotizza che il tasso di perdita (LGD) sia deterministico

La seconda formula sopracitata è quella maggiormente usata nonostante vi siano evidenze empiriche di una correlazione negativa tra i RR e le PD. Le basi su cui si fonda tale affermazione vanno ricercate nell'andamento del ciclo economico e dei tassi di interesse. È comunque valida sotto le ipotesi di tasso di perdita LGD deterministico e di indipendenza tra la PD e LGD. Ai fini di un calcolo più preciso, andrebbe considerata la

LGD stocastica e andrebbe introdotta la covarianza tra la PD e la LGD in modo da poter esprimere il grado di correlazione tra di esse.

Nel caso di LGD stocastica si ha:

$$UL = \sqrt{PD(1 - PD) * LGD^2 + PD * \sigma LGD^2}$$

L'espressione ipotizza indipendenza tra PD e LGD; se si ritiene che vi sia connessione tra le due variabili occorre introdurre anche il termine riguardante la covarianza. Quindi, considerando entrambe le ipotesi la formula diventa:

$$UL^2 = cov(PD^2, LGD^2) + [PD(1 - PD) + PD^2] * [\sigma^2_{LGD} + LGD^2] \\ - [cov(PD, LGD) + PD * LGD]^2$$

Si può notare che la volatilità del tasso di insolvenza (σ^2_{LGD}) ha un'influenza maggiore sulla UL quanto più elevata è la PD, ciò vale a dire che pesa di più nelle classi di rating peggiori.

2.3 Credit scoring

Il *credit scoring* è una tecnica di valutazione statistica impiegata dagli intermediari per valutare la solvibilità dei clienti. Tale tecnica è calcolata mediante l'uso di modelli previsionali che elaborano le informazioni disponibili dei clienti e le riducono ad un unico valore in grado di sintetizzare il rischio creditizio, noto come *credit score*. I principali vantaggi del *credit scoring* sono dovuti all'omogeneità di trattamento dei singoli clienti e dalla capacità di assistere in maniera rapida ed efficiente al processo decisionale di concessione del prestito.

Grazie ad Edward I. Altman, professore di Finanza della Stern School of Business presso la New York University, negli anni Sessanta nascono i primi credit scoring. Egli, in

particolare, diede vita al modello Z-score. Egli analizzò i dati di bilancio di 33 aziende in stato di default e 33 aziende solide sviluppando così un modello in grado di prevedere, con un grado di accuratezza del 95%, la probabilità di fallimento di un'impresa negli anni successivi. Giova precisare che, nonostante i limiti, tutt'oggi lo *Z-score* possiede tutti gli elementi che caratterizzano un modello di *credit scoring*.

Nel corso degli anni, grazie all'evoluzione delle tecniche statistiche e delle capacità computazionali, si è assistito ad uno sviluppo di nuovi modelli di diversa natura, partendo dai *logit* e *probit* di martin ed Ohlson fino ad arrivare alle più recenti tecniche di machine learning. Nonostante le numerose differenze, tutti i modelli perseguono la minimizzazione della funzione di errore, calcolata sui valori delle probabilità di default assegnate ad ogni società del campione in esame.

2.3.1 Modelli statistici: la regressione logistica

I primi modelli previsionali sviluppati sono stati di tipo statistico. Pertanto, il caso più semplice è rappresentato dalla regressione lineare a variabile singola che consente di studiare la relazione tra due variabili attraverso l'equazione di una retta generica.

In formula:

$$Y = \alpha + \beta X + \varepsilon$$

La variabile dipendente Y deve essere prevista attraverso l'analisi dei valori della variabile indipendente X .

Gli altri tre parametri presenti nella formula sopracitata sono:

- l'intercetta α che rappresenta il valore di Y quando la variabile indipendente X assume valore nullo;
- il coefficiente di regressione β che descrive la pendenza della retta, ovvero il cambiamento della variabile dipendente Y corrispondente ad una variazione unitaria della variabile X ;
- la componente di errore ε di natura casuale in quanto non può essere spiegata dalla variabile indipendente X .

La logica della regressione lineare semplice può essere modificata a piacere, ad esempio

è possibile includere un maggior numero di variabili indipendenti in modo da spiegare meglio il comportamento della variabile di risposta Y . Questi modelli statistici rientrano nella famiglia degli algoritmi di regressione lineare multipla che, date k variabili indipendenti, possono essere generalizzati con la seguente formula:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

La formula sopracitata presenta la stessa struttura della precedente, l'unica differenza riguarda la presenza di molteplici coefficienti di regressione $\beta_1, \beta_2, \dots, \beta_k$ definiti parziali in quanto esprimono l'entità della variazione della variabile Y rispetto al cambiamento della singola variabile indipendente X_1, X_2, \dots, X_k ad essi associato. Tali modelli risultano particolarmente versatili in quanto, apportando opportune trasformazioni, è possibile crearne delle versioni generalizzate che linearizzano le relazioni tra le variabili esplicative anche laddove non lo fossero in origine, adattando in questo modo il modello al caso di studio.

Un particolare modello lineare generalizzato è la regressione logistica. Essa è utilizzata quando la variabile Y , della quale si vuole spiegare il comportamento, è dicotomica, definita in 0 e 1, e avente distribuzione binomiale. Le variabili di tipo dicotomico sono utilizzate principalmente per problemi di classificazione, ovvero la determinazione della probabilità di appartenenza ad un gruppo piuttosto che ad un altro. Il modello di regressione logistica non è di tipo lineare in quanto la variabile Y non può ammettere valori tra $-\infty$ e $+\infty$ ma deve avere un comportamento asintotico in corrispondenza dei valori 0 e 1. La figura che segue ne mostra l'andamento al variare della X :

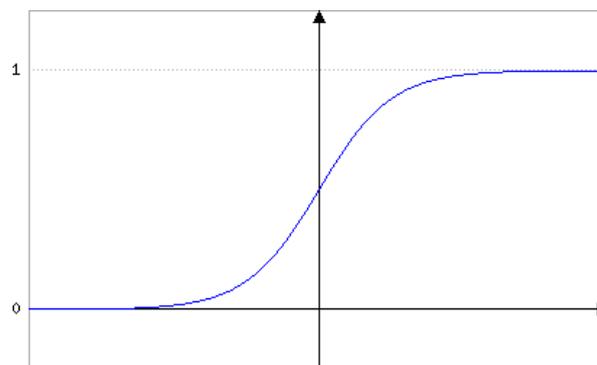


Figura 2.3: Funzione logistica

Nelle regressioni lineari la variabile di risposta è calcolata come il valore medio della variabile Y condizionata all'assunzione di un dato valore di X .

In termini matematici si ha:

$$E[Y|X] = \mu_Y$$

Diversamente, nella regressione logistica il valor medio è uguale alla probabilità che la variabile Y sia pari ad 1 condizionata all'assunzione del valore x della variabile indipendente X , ossia:

$$P(Y = 1|X = x) = \pi(x) = \alpha + \beta X$$

La probabilità così espressa possiede ancora un campo di esistenza tra $-\infty$ e $+\infty$, tuttavia è possibile restringerlo al range (0;1) applicando prima la trasformazione esponenziale e poi quella logistica:

$$P(Y = 1|X=x) = \pi(x) = \frac{e^{\alpha + \beta X}}{1 + e^{\alpha + \beta X}}$$

A questo punto, è possibile esprimere le probabilità mediante il rapporto tra le due categorie 0 e 1, considerando la complementarità degli eventi ossia $P(Y = 0) = P(Y = 1|X = x)$. Tale rapporto prende il nome di *odd* ed in genere si calcola dividendo le frequenze osservate in una classe con quelle osservate nell'altra:

$$\begin{aligned} odds(Y=1|X=x) &= \frac{\pi(x)}{1 - \pi(x)} \\ odds(Y=1|X=x) &= \frac{\frac{e^{\alpha + \beta X}}{1 + e^{\alpha + \beta X}}}{1 - \frac{e^{\alpha + \beta X}}{1 + e^{\alpha + \beta X}}} = \frac{\frac{e^{\alpha + \beta X}}{1 + e^{\alpha + \beta X}}}{\frac{1}{1 + e^{\alpha + \beta X}}} = e^{\alpha + \beta X} \end{aligned}$$

Calcolando il *logit*, ossia il logaritmo dell'*odd*, e applicando le proprietà dei logaritmi si ottiene:

$$\ln(odds(Y=1|X=x)) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta X$$

Nel caso di regressione logistica multipla avente k variabili indipendenti è possibile

ripercorrere lo stesso ragionamento arrivando alla seguente conclusione:

$$\pi(\mathbf{x}) = \frac{e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{1 + e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}$$

Riassumendo, è possibile affermare che i concetti di probabilità *odd* e *logit* trattati, forniscono sfumature diverse di uno stesso concetto. La probabilità è riferita all'appartenenza ad un gruppo, l'*odd* definisce quanto è più o meno probabile un evento rispetto ad un altro e il *logit*, calcolato come il logaritmo dell'*odd*, rende possibili le trasformazioni algebriche.

2.3.2 Modelli di machine learning

Negli ultimi anni, ricercatori delle istituzioni finanziarie e accademici hanno concentrato la loro attenzione sugli algoritmi di machine learning anche in ambito economico-finanziario. Un modello di machine learning è un file che è stato addestrato per riconoscere determinati tipi di pattern, a differenza dei modelli tradizionali, per i quali è necessario stabilire i passi che l'algoritmo dovrà effettuare. Si addestra, dunque, un modello su un insieme di dati, fornendogli un algoritmo che può utilizzare per ragionare e imparare da quei dati. Dopo aver addestrato il modello, è possibile utilizzarlo per ragionare su dati che non ha visto prima e fare previsioni.

Nel campo dell'apprendimento automatico, ci sono due tipi principali di algoritmi sulla base della tipologia di apprendimento: supervisionato e non supervisionato. La principale differenza tra i due tipi è che l'apprendimento supervisionato è effettuato utilizzando una verità fondamentale, o in altre parole, si ha una conoscenza preliminare di quali dovrebbero essere i valori di output per i nostri campioni. Pertanto, l'obiettivo dell'apprendimento supervisionato è apprendere una funzione che, dato un campione di dati e di output desiderati, approssimi al meglio la relazione tra input e output osservabile nei dati. Le tipologie di algoritmi adottati in questo campo sono le tecniche di regressione

e di classificazione. Fondamentalmente, la classificazione riguarda la previsione di un'etichetta e la regressione riguarda la previsione di una quantità.

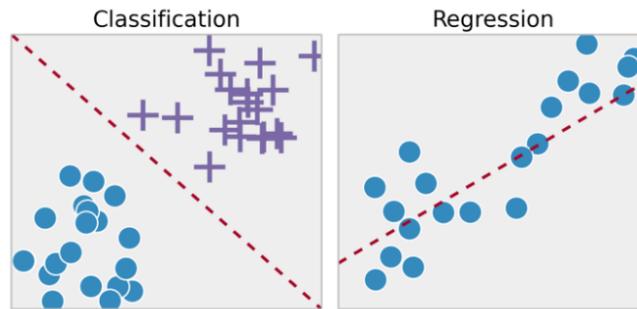


Figura 2.4: Algoritmi di Classificazione e Regressione

L'apprendimento senza supervisione, d'altra parte, non ha output etichettati, quindi il suo obiettivo è inferire la struttura naturale presente all'interno di un insieme di punti dati.

Nel caso del credit scoring, risulta più efficiente utilizzare i modelli di machine learning grazie alla loro capacità di acquisire in modo efficiente le informazioni da grandi set di dati, riconoscendo modelli non lineari e dipendenze temporali tra i dati. I modelli di classificazione del credito sono ampiamente applicati perché possono aiutare i responsabili delle decisioni finanziarie a gestire i problemi di classificazione del credito. Tra questi, le reti neurali artificiali sono state ampiamente accettate come metodi convincenti nel settore del credito.

Il modello si adatta inizialmente ad un set di dati di training, che è un insieme di dati utilizzati per adattare i parametri. Successivamente, il modello adattato viene utilizzato per prevedere le risposte delle osservazioni in un secondo set di dati denominato set di dati di validation. La valutazione diventa più parziale man mano che le competenze sul set di dati di validation sono incorporate nella configurazione del modello. Infine, con un set di dati di test è quindi possibile valutare le prestazioni del modello.

Nel prossimo capitolo saranno discusse le principali proprietà delle reti neurali e le teorie che sono alla base del loro funzionamento.

CAPITOLO III – Le Reti neurali

Il Deep Learning consiste in una tecnica di Machine Learning che impiega algoritmi capaci di simulare il cervello umano. In particolare, questi algoritmi si fondano sullo sviluppo di reti neurali: una rete neurale è un meccanismo di apprendimento automatico molto potente che fondamentalemente imita il modo in cui un cervello umano apprende.

Il cervello riceve lo stimolo dal mondo esterno, elabora l'input e quindi genera l'output.

Relativamente alle reti neurali, per prima cosa, si crea una raccolta di "neuroni" e successivamente è consentito loro di interagire. Man mano che il compito si complica, più neuroni formano una rete complessa, passando le informazioni tra di loro. In seguito, alla rete è chiesto di risolvere un problema, cosa che cerca di fare più e più volte, rafforzando ogni volta le connessioni che portano al successo e diminuendo quelle che portano al fallimento.

Le reti neurali consentono ai programmi per computer di riconoscere schemi e risolvere problemi comuni nei campi dell'IA, essendo capaci di modificare il loro comportamento in modo da adattarsi agli stimoli che provengono dall'esterno.

È possibile dunque affermare che il prototipo delle reti neurali sono le reti neurali biologiche e che il cervello dell'uomo è come un computer su cui sono immagazzinate le informazioni.

3.1 Il neurone biologico

I neuroni (chiamati anche cellule nervose) sono le unità fondamentali del cervello e del sistema nervoso. Esse sono le cellule responsabili di ricevere input sensoriali dal mondo esterno, di inviare comandi motori ai nostri muscoli e di trasformare e ritrasmettere i segnali elettrici ad ogni step. I nostri circa 100 miliardi di neuroni interagiscono strettamente con altri tipi di cellule, normalmente classificati come cellule della glia.

Un'analogia utile è pensare a un neurone come a un albero. Un neurone ha tre parti principali: dendriti, un assone e un corpo cellulare o soma, che possono essere rappresentati rispettivamente come rami, radici e tronco di un albero. Un dendrite (ramo

di un albero) è il punto in cui un neurone riceve input da altre cellule. I dendriti si ramificano mentre si muovono verso le loro punte, proprio come fanno i rami degli alberi, e hanno persino strutture simili a foglie chiamate spine.

L'assone (radici degli alberi) è la struttura di output del neurone, la parte trasmittente del neurone; quando un neurone vuole interagire con un altro neurone, invia un messaggio elettrico chiamato potenziale d'azione attraverso l'intero assone, ed è breve evento elettrico tipicamente generato nell'assone che segnala il neurone come "attivo". Un potenziale d'azione percorre la lunghezza dell'assone e provoca il rilascio del neurotrasmettitore nella sinapsi. Il potenziale d'azione e il conseguente rilascio del trasmettitore consentono al neurone di comunicare con altri neuroni. Il soma (tronco d'albero) è dove si trova il nucleo e quindi dove si trova il DNA del neurone e dove le proteine sono trasportate tra l'assone e i dendriti.

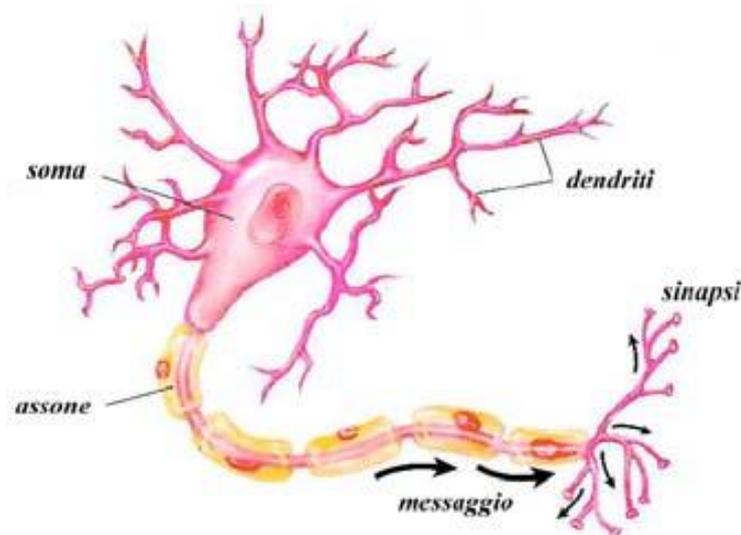


Figura 3.1: Il neurone biologico

La somiglianza tra le reti neurali artificiali e il cervello umano è che entrambi acquisiscono abilità nell'elaborazione dei dati e nella ricerca di soluzioni attraverso la formazione.

3.2 Il neurone artificiale

Il primo modello computazionale di un neurone è stato proposto da Warren McCulloch (neuroscienziato) e Walter Pitts (logico) nel 1943 nel loro trattato "A logical calculus of the ideas immanent in nervous activity". In questo articolo McCulloch e Pitts hanno cercato di capire come il cervello potesse produrre schemi altamente complessi

utilizzando molte cellule di base collegate tra loro. Il modello di McCulloch e Pitts di un neurone, chiamato brevemente neurone MCP, ha dato un importante contributo allo sviluppo di reti neurali artificiali, che modellano le caratteristiche chiave dei neuroni biologici. Come mostrato in figura 3.2, il modello può essere diviso in due parti. La prima parte g riceve un input, esegue un'aggregazione e, in base al valore aggregato, la seconda

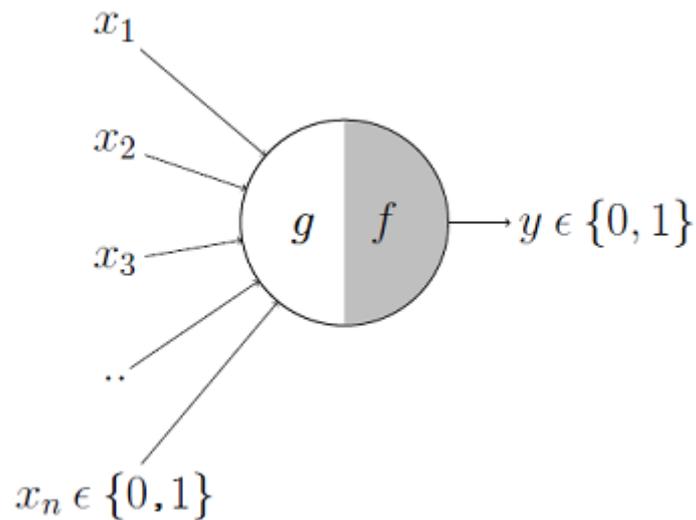


Figura 3.2: Il neurone MCP

parte f prende una decisione.

Con gli anni, si è sviluppato il neurone artificiale nella sua versione più moderna.

La rete neurale artificiale è costituita da semplici unità di calcolo "neuroni artificiali" e ciascuna unità è collegata alle altre unità tramite connettori di peso; quindi, queste unità

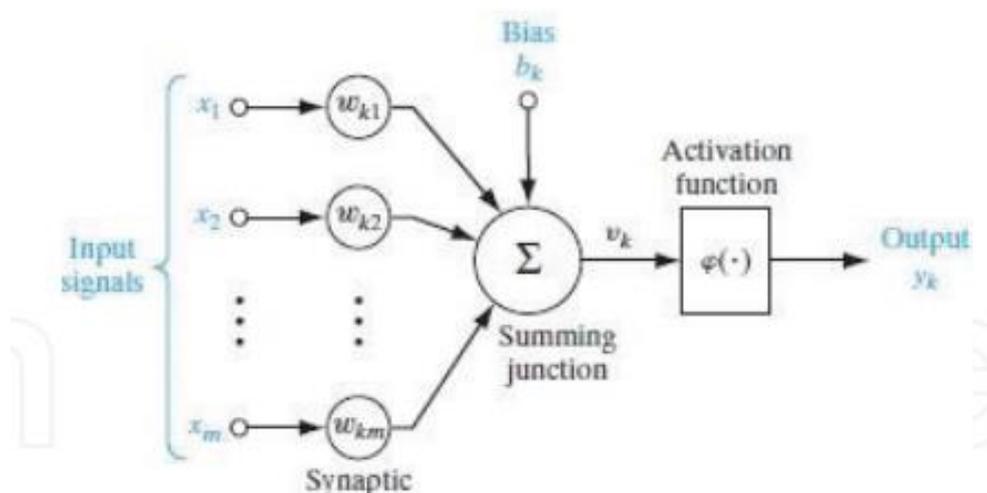


Figura 3.3: Il neurone artificiale moderno

calcolano la somma ponderata degli ingressi in arrivo e rilevano l'uscita utilizzando la funzione di attivazione.

Sulla base del diagramma a blocchi e della funzione della rete neurale, è possibile identificare tre elementi fondamentali del modello neurale:

- Le sinapsi che hanno un peso o una forza in cui il segnale di ingresso x_i connesso al neurone k viene moltiplicato per il peso sinaptico w_{ki} ;
- Un sommatore per sommare gli input ponderati;
- Una funzione di attivazione per produrre l'output di un neurone. Viene anche definita funzione di schiacciamento, in quanto schiaccia (limita) l'intervallo di ampiezza del segnale di uscita a un valore finito.

Il bias b_k ha l'effetto di aumentare o diminuire l'input netto della funzione di attivazione, a seconda che sia rispettivamente positivo o negativo.

Matematicamente, l'output sul neurone k può essere descritto come:

$$y_k = \varphi\left(\sum_{i=1}^m x_i \cdot w_{ki} + b_k\right)$$

dove

- $x_1, x_2, x_3, \dots, x_m$ sono i segnali di ingresso;
- $w_{k1}, w_{k2}, w_{k3}, \dots, w_{km}$ sono i rispettivi pesi del neurone;
- b_k è il bias;
- φ è la funzione di attivazione.

Per chiarire l'effetto del bias sulle prestazioni del neurone, l'output fornito nell'equazione precedente è elaborato in due fasi: la prima fase include gli input ponderati, e si prende in considerazione la somma denotata come S_k :

$$S_k = \sum_{i=1}^m x_i \cdot w_{ki}$$

Quindi, l'output del sommatore sarà dato da:

$$v_k = S_k + b_k$$

A seconda del valore del bias, la relazione tra l'input ponderato e l'output del sommatore verrà modificata come mostrato nella Figura 3.3.

Inoltre, il bias potrebbe essere anche considerato come un segnale di ingresso x_0 fissato a +1 con peso sinaptico uguale al bias b_k .

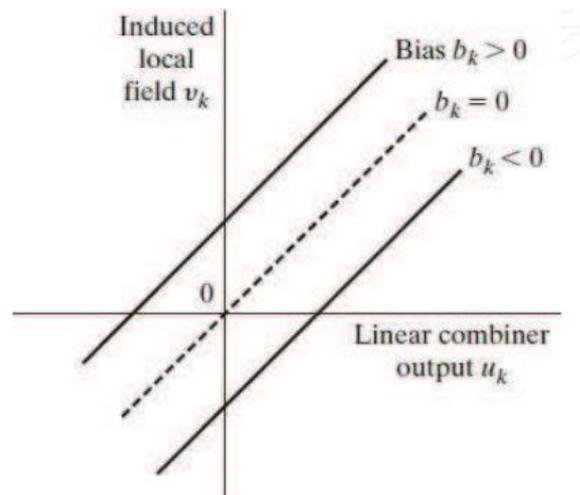


Figura 3.3: Effetto del bias

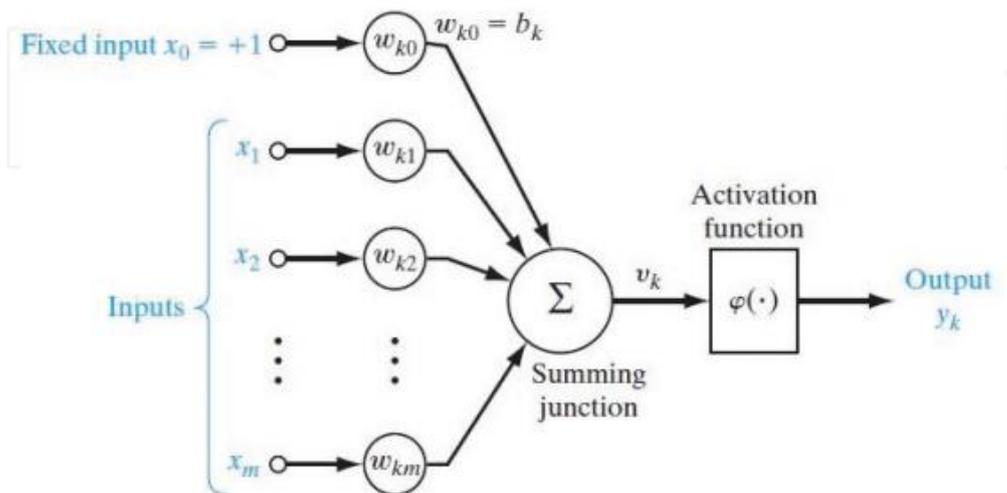


Figura 3.4: La struttura del neurone artificiale considerando il bias come input

3.3 Tipi di funzione di attivazione

Le funzioni di attivazione della rete neurale sono una componente cruciale del deep learning. Esse determinano l'output di un modello di deep learning, la sua accuratezza e anche l'efficienza computazionale dell'addestramento di un modello, che può creare o distruggere una rete neurale su larga scala. Le funzioni di attivazione hanno anche un effetto importante sulla capacità della rete neurale di convergere e sulla velocità di convergenza, o in alcuni casi, potrebbero impedire la convergenza delle reti neurali. Si tratta di equazioni matematiche che definiscono l'output di una rete. La funzione è

collegata a ciascun neurone e determina se deve essere attivato o meno, in base al fatto che l'input di ciascun neurone sia rilevante per la previsione del modello. Un ulteriore aspetto delle funzioni di attivazione è che devono essere efficienti dal punto di vista computazionale perché vengono calcolate su migliaia o addirittura milioni di neuroni per ogni campione di dati.

Le funzioni di attivazione più usate sono: Step Function, Linear Activation Function, Sigmoide, Tanh, ReLU, Leaky ReLU e Parameterised ReLU.

3.3.1 Step function

Una step function è una funzione di attivazione basata su una soglia di attivazione.

Essa è caratterizzata dalla seguente formula:

$$f(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases}$$

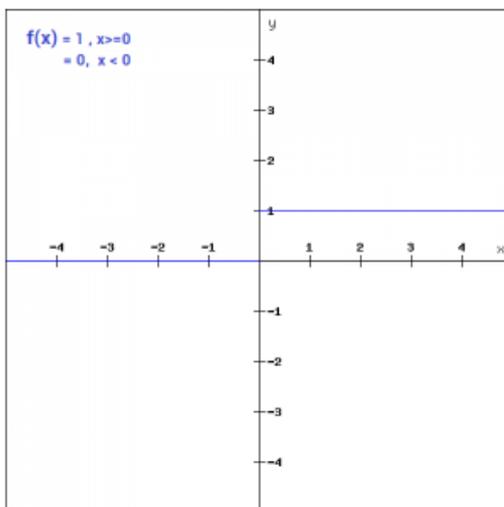


Figura 3.5: Step function

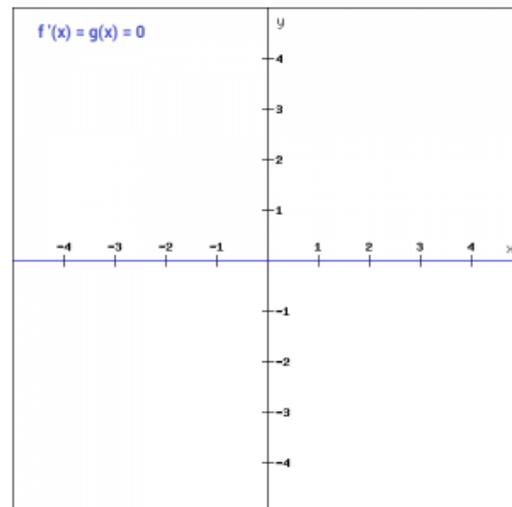


Figura 3.6: Derivata della step function

Se il valore di input è superiore o inferiore ad una determinata soglia, il neurone viene attivato e invia esattamente lo stesso segnale al livello successivo. L'output del neurone assume il valore di 1 se il campo locale indotto di quel neurone non è negativo e 0 altrimenti. Purtroppo, il gradiente della funzione gradino è zero, il che causa un ostacolo nel processo di back propagation, infatti, calcolando la derivata di $f(x)$ rispetto a x , risulta essere 0.

I gradienti vengono calcolati per aggiornare i pesi e le deviazioni durante il processo di back propagation. Poiché il gradiente della funzione è zero, i pesi e i pregiudizi non si aggiornano.

3.3.2 Linear Activation Function

Nella step function il gradiente è diventato zero. Questo perché non vi è alcun componente di x nella funzione. Per questo motivo, si introduce una funzione lineare. Possiamo definire la funzione come: $f(x)=ax$

Sebbene qui non diventi zero, il gradiente è una costante che non dipende dal valore di input x . Ciò implica che i pesi e i pregiudizi verranno aggiornati durante il processo di backpropagation, ma il fattore di aggiornamento sarebbe lo stesso.

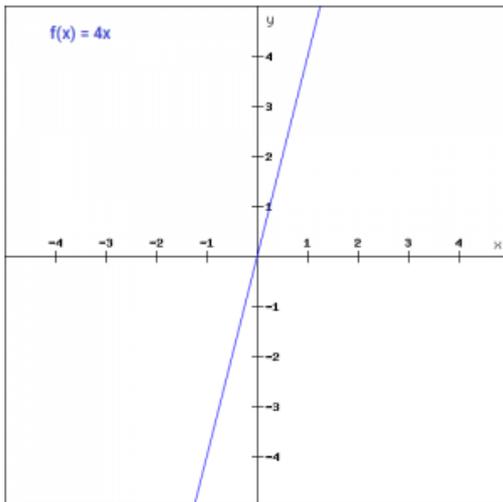


Figura 3.7: Linear Activation Function

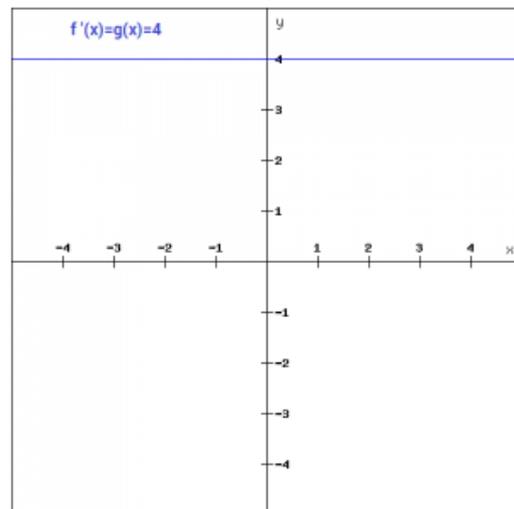


Figura 3.8: Derivata della Linear Activation Function

In questo scenario, la rete neurale non migliorerà realmente l'errore poiché il gradiente è lo stesso per ogni iterazione. La rete non sarà in grado di addestrarsi bene e acquisire i modelli complessi dai dati. Quindi, la funzione lineare potrebbe essere ideale per compiti semplici in cui l'interpretabilità è altamente desiderata.

3.3.3 Sigmoide Function

La funzione Sigmoide è una delle funzioni di attivazione non lineare più utilizzate. Questa funzione trasforma i valori tra 0 e 1.

Di seguito l'espressione matematica per la funzione sigmoide: $f(x) = \frac{1}{(1+e^{-x})}$

Un punto degno di nota qui è che, a differenza della step function e delle funzioni lineari, questa funzione è una funzione non lineare. Ciò significa essenzialmente che quando ho più neuroni che hanno la funzione sigmoide come funzione di attivazione, anche l'output è non lineare.

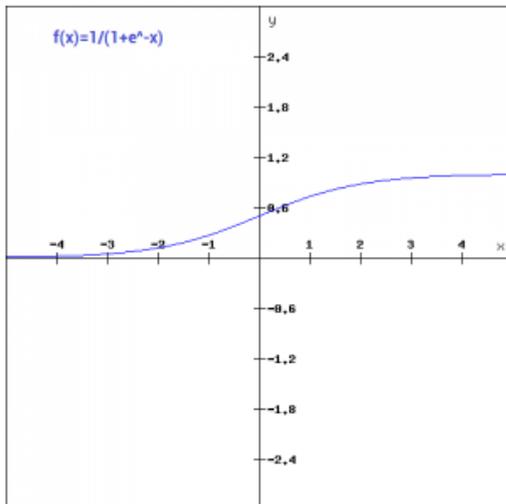


Figura 3.9: Funzione Sigmoide

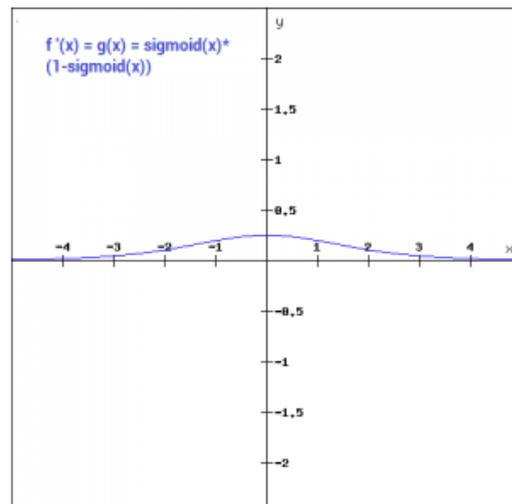


Figura 3.10: Derivata della Funzione Sigmoide

La funzione sigmoide causa un problema principalmente definito come problema del gradiente di fuga che si verifica perché convertiamo un input di grandi dimensioni nell'intervallo da 0 a 1 e quindi le loro derivate diventano molto più piccole che non danno un output soddisfacente. Per risolvere questo problema viene utilizzata un'altra funzione di attivazione come ReLU dove non abbiamo un piccolo problema derivativo.

3.3.4 Tanh Function

La funzione tanh è molto simile alla funzione sigmoide. L'unica differenza è che è simmetrica attorno all'origine. L'intervallo di valori in questo caso è compreso tra -1 e 1. Pertanto, gli input per i livelli successivi non saranno sempre dello stesso segno. La funzione tanh è definita come: $\tanh(x) = 2\text{sigmoid}(2x) - 1$.

Tutte le altre proprietà della funzione tanh sono le stesse della funzione sigmoide. Simile al sigmoide, la funzione tanh è continua e differenziabile in tutti i punti.

Il gradiente della funzione tanh è più ripido rispetto alla funzione sigmoide. Di solito il tanh è preferito alla funzione sigmoide poiché è centrato sullo zero e i gradienti non sono limitati a muoversi in una certa direzione.

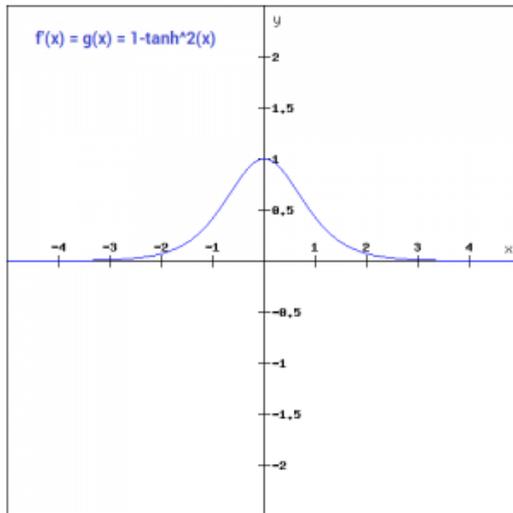


Figura 3.11: Tanh Function

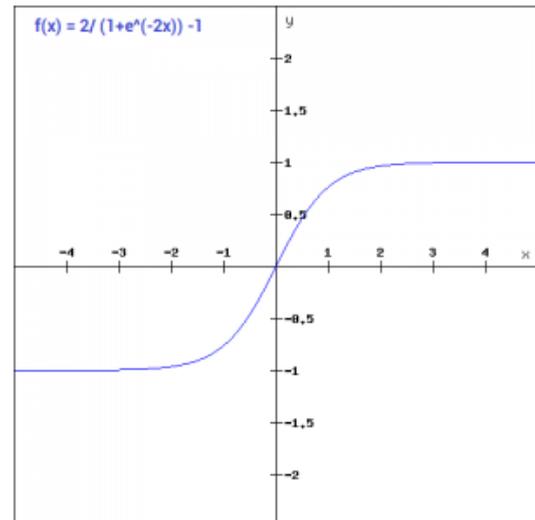


Figura 3.12: Derivata della Tanh Function

3.3.5 ReLU Function

La funzione ReLU è un'altra funzione di attivazione non lineare che ha guadagnato popolarità nel dominio del deep learning. ReLU sta per unità lineare rettificata. Il vantaggio principale dell'utilizzo della funzione ReLU rispetto ad altre funzioni di attivazione è che non attiva tutti i neuroni contemporaneamente. Ciò significa che i neuroni verranno disattivati solo se l'output della trasformazione lineare è inferiore a 0.

La funzione ReLU è: $f(x) = \begin{cases} x & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$

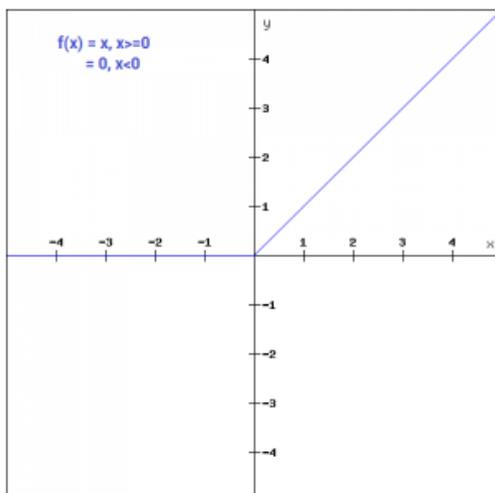


Figura 3.13: ReLU Function

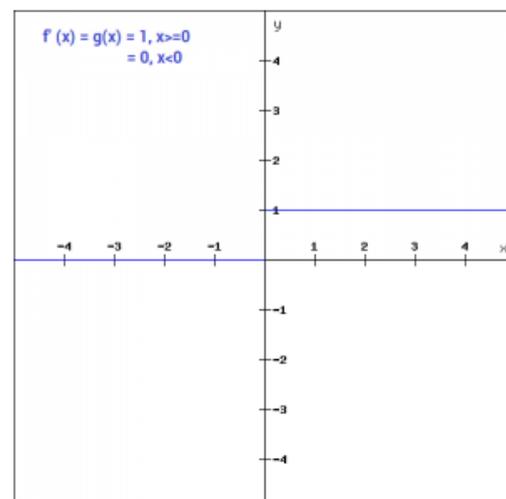


Figura 3.14: Derivata della ReLU Function

Osservando il grafico negativo del gradiente della funzione ReLU, il gradiente è zero. Per questo motivo, durante il processo di back propagation, i pesi e i bias di alcuni neuroni non vengono aggiornati. Questo può creare neuroni morti che non si attivano mai. Questo limite è superato dalla funzione ReLU "Leaky"

3.3.6 Leaky ReLU Function

La funzione Leaky ReLU non è altro che una versione migliorata della funzione ReLU. Come abbiamo visto che per la funzione ReLU, il gradiente è 0 per $x < 0$, che disattiverrebbe i neuroni in quella regione. La Leaky ReLU è nata per affrontare questo problema. Invece di definire la funzione Relu come 0 per valori negativi di x , la definiamo come una componente lineare estremamente piccola di x .

$$\text{La funzione è } f(x) = \begin{cases} x & \text{se } x \geq 0 \\ 0,01x & \text{se } x < 0 \end{cases}$$

Effettuando questa piccola modifica, il gradiente del lato sinistro del grafico risulta essere un valore diverso da zero. Quindi non incontreremmo più neuroni morti in quella regione

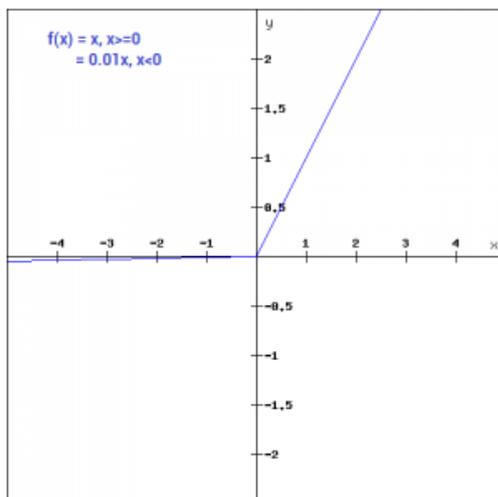


Figura 3.15: Leaky ReLU Function

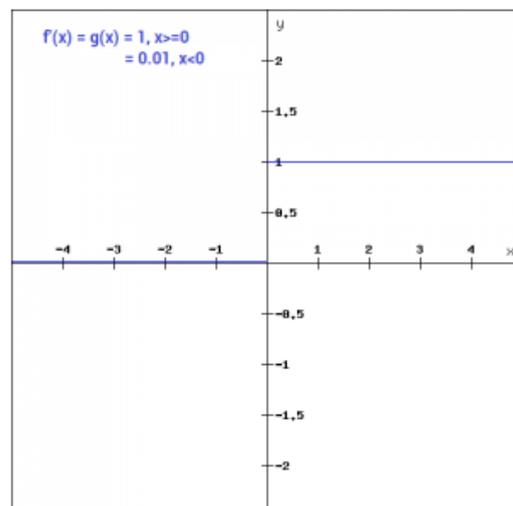


Figura 3.16: Derivata della Leaky ReLU Function

3.3.7 Parameterised ReLU Function

Questa funzione è un'altra variante di ReLU che mira a risolvere il problema del gradiente che diventa zero per la metà sinistra dell'asse. Il ReLU parametrizzato, come suggerisce

il nome, introduce un nuovo parametro come pendenza della parte negativa della funzione:

$$f(x) = \begin{cases} x & \text{se } x \geq 0 \\ ax & \text{se } x < 0 \end{cases}$$

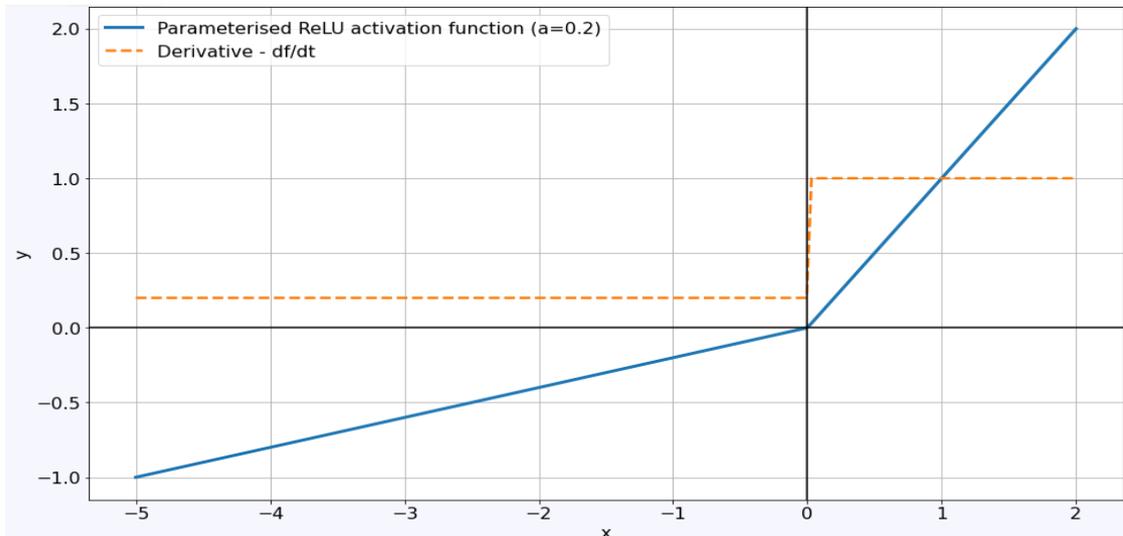


Figura 3.17: *Parameterised ReLU Function e la sua derivata*

Quando il valore di a è fissato a 0,01, la funzione agisce come una funzione Leaky ReLU. Tuttavia, nel caso di una funzione ReLU parametrizzata, anche " a " è un parametro addestrabile. La rete apprende anche il valore di " a " per una convergenza più rapida e ottimale. La funzione ReLU parametrizzata viene utilizzata quando la funzione ReLU che perde non riesce ancora a risolvere il problema dei neuroni morti e le informazioni rilevanti non vengono passate con successo al livello successivo.

3.4 Architettura delle reti neurali

L'architettura di una rete neurale artificiale definisce il modo in cui i suoi numerosi neuroni sono disposti, o posizionati, in relazione l'uno con l'altro. Queste disposizioni sono strutturate essenzialmente dirigendo le connessioni sinaptiche dei neuroni. La topologia di una data rete neurale, all'interno di una particolare architettura, può essere definita come le diverse composizioni strutturali che può assumere. Il successo della rete neurale risiede nell'attenta progettazione dell'architettura.

In generale, una rete neurale artificiale può essere divisa in tre parti, denominate layer, noti come:

- Layer di input: questo livello è responsabile della ricezione di informazioni (dati), segnali, caratteristiche o misurazioni dall'ambiente esterno. Questi input (campioni o modelli) sono normalmente normalizzati entro i valori limite prodotti dalle funzioni di attivazione. Questa normalizzazione si traduce in una migliore precisione numerica per le operazioni matematiche eseguite dalla rete.
- Layer nascosti, intermedi o invisibili: questi strati sono composti da neuroni responsabili dell'estrazione dei modelli associati al processo o al sistema analizzato. Questi livelli eseguono la maggior parte dell'elaborazione interna da una rete.
- Layer di output: questo livello è composto anche da neuroni, e quindi è responsabile della produzione e della presentazione degli output finali della rete, che risultano dall'elaborazione eseguita dai neuroni negli strati precedenti.

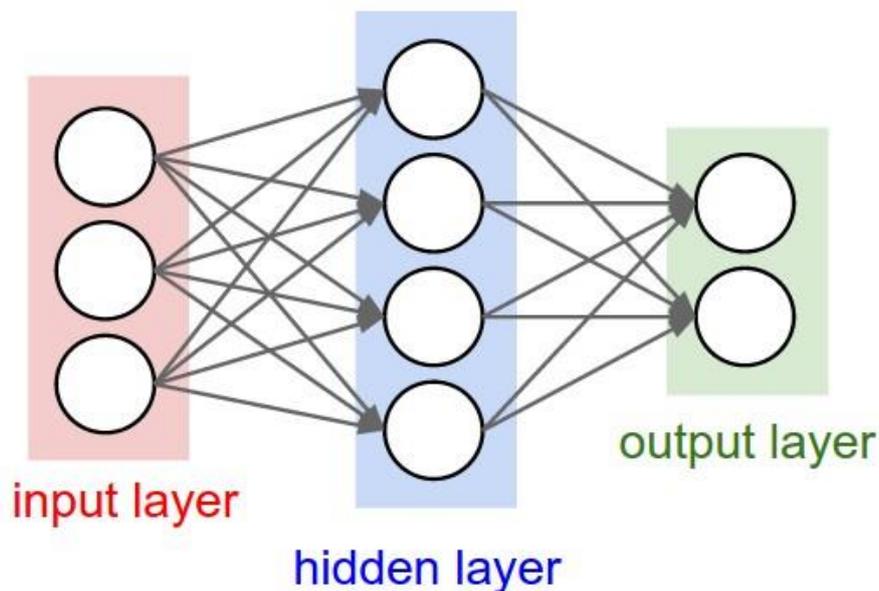


Figura 3.18: I Layer di una rete neurale

Le principali architetture delle reti neurali artificiali, considerando la disposizione dei neuroni, nonché come sono interconnesse e come sono composti i suoi strati, possono essere suddivise come segue:

- single-layer feedforward network;

- multilayer feedforward networks;
- recurrent networks;
- mesh networks.

3.4.1 Single-Layer Feedforward Architecture

Questa rete neurale artificiale ha un solo livello di input e un unico livello neurale, che è anche il livello di output. Le informazioni fluiscono sempre in un'unica direzione (quindi unidirezionale), che va dallo strato di input allo strato di output. Dalla Fig. 2.1, è possibile vedere che nelle reti appartenenti a questa architettura, il numero di uscite di rete coinciderà sempre con la sua quantità di neuroni. Queste reti vengono solitamente impiegate nella classificazione dei modelli e nei problemi di filtraggio lineare.

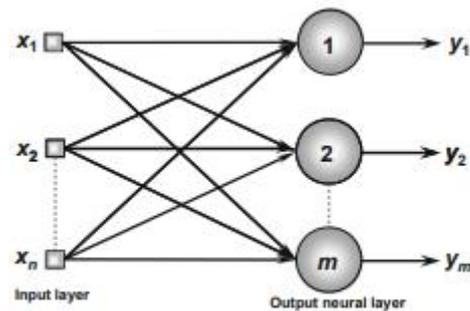


Figura 3.19: Esempio di una rete single-layer feedforward

3.4.2 Multiple-Layer Feed-forward Architectures

A differenza delle reti appartenenti all'architettura precedente, le reti feed-forward con più strati sono composte da uno o più strati neurali nascosti (Fig. 2.2). Sono impiegate nella soluzione di diversi problemi, come quelli relativi all'approssimazione delle funzioni, alla classificazione dei modelli, all'identificazione del sistema, al controllo dei processi, all'ottimizzazione, alla robotica e così via. La Figura 2.2 mostra una rete feedforward con più livelli composti da uno strato di input con n segnali campione, due strati neurali nascosti costituiti rispettivamente da n_1 e n_2 neuroni e, infine, uno strato neurale di output composto da m neuroni che rappresentano i rispettivi valori di output del problema analizzato.

Tra le principali reti che utilizzano architetture feedforward a più livelli ci sono le Multilayer Perceptron (MLP) e Radial Basis Function (RBF).

Dalla Figura 3.02, è possibile capire che la quantità di neuroni che compongono il primo strato nascosto è solitamente diversa dal numero di segnali che compongono lo strato di ingresso della rete. In effetti, il numero di strati nascosti e la rispettiva quantità di neuroni dipendono dalla natura e dalla complessità del problema mappato dalla rete, nonché dalla quantità e qualità dei dati disponibili sul problema. Tuttavia, allo stesso modo per le reti feedforward a strato semplice, la quantità di segnali di uscita coinciderà sempre con il numero di neuroni da quel rispettivo strato.

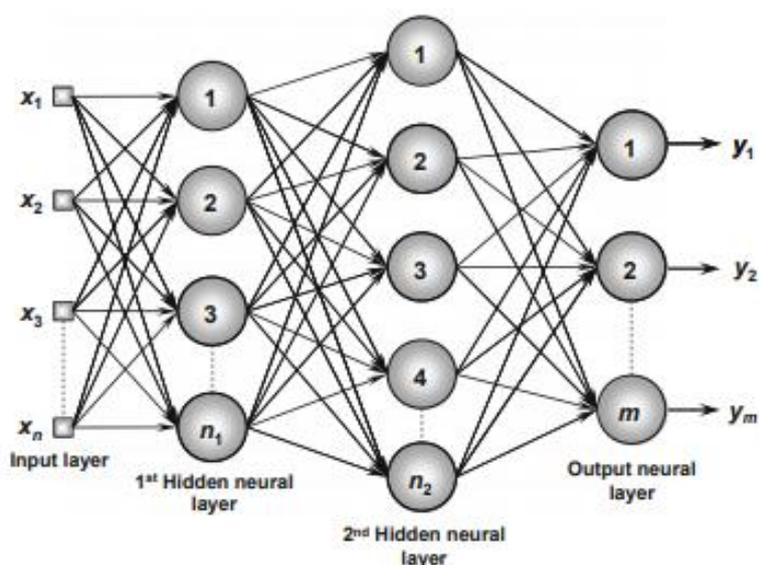


Figura 3.20: Esempio di una rete con architettura Multiple-Layer Feed-forward

3.4.3 Recurrent or Feedback Architecture

In queste reti, gli output dei neuroni vengono utilizzati come input di feedback per altri neuroni. La funzione di feedback qualifica queste reti per l'elaborazione dinamica delle informazioni, il che significa che possono essere impiegate su sistemi con varianti temporali, come la previsione di serie temporali, l'identificazione e l'ottimizzazione del sistema, il controllo del processo e così via.

La Figura 3.21 illustra un esempio di una rete con feedback, in cui uno dei suoi segnali di uscita viene rimandato allo strato intermedio. Pertanto, utilizzando il processo di feedback, le reti con questa architettura producono uscite di corrente tenendo anche in considerazione i valori di uscita precedenti

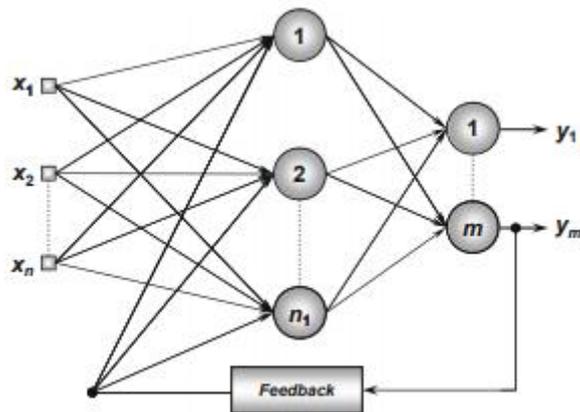


Figura 3.21: Esempio di una rete con feedback

3.4.4 Mesh Architectures

Le caratteristiche principali delle reti con strutture a maglia risiedono nel considerare la disposizione spaziale dei neuroni ai fini dell'estrazione di pattern, ovvero la localizzazione spaziale dei neuroni è direttamente correlata al processo di regolazione dei loro pesi e soglie sinaptiche. Queste reti servono un'ampia gamma di applicazioni e vengono utilizzate in problemi che coinvolgono il raggruppamento dei dati, il riconoscimento di modelli, l'ottimizzazione del sistema, i grafici e così via.

3.5 L'algoritmo back propagation

Le moderne reti neurali utilizzano una tecnica chiamata back propagation per addestrare il modello, che pone un maggiore sforzo computazionale sulla funzione di attivazione e sulla sua funzione derivata. Tale tecnica confronta il risultato ottenuto da una rete con l'output che si vuole ricavare: usando la differenza tra i due risultati prevede di cambiare i pesi delle connessioni tra i livelli della rete partendo dal livello output. La parte "all'indietro" del nome deriva dal fatto che il calcolo del gradiente procede a ritroso attraverso la rete, calcolando per primo il gradiente dell'ultimo strato di pesi e per ultimo il gradiente del primo strato di pesi. I calcoli parziali del gradiente da uno strato vengono riutilizzati nel calcolo del gradiente per lo strato precedente. Questo flusso all'indietro

delle informazioni sull'errore consente un calcolo efficiente del gradiente a ogni strato rispetto all'approccio ingenuo di calcolare il gradiente di ogni strato separatamente.

3.5.1 Loss Function

Una funzione di perdita è utilizzata per ottimizzare i valori dei parametri in un modello di rete neurale. Le funzioni di perdita associano un insieme di valori di parametri per la rete su un valore scalare che indica quanto bene quei parametri eseguono il compito che la rete è destinata a svolgere. Esistono diverse funzioni di perdita: alcune misurano l'errore quadratico medio tra l'output di una rete e un output target o desiderato. In questo caso si calcola semplicemente il quadrato della differenza tra Y e Y_{pred} ed successivamente la media su tutti i dati:

$$Loss = \frac{1}{n} * \sum_{i=0}^n (Y_i - Y_{pred_i})^2$$

Altre funzioni di perdita sono progettate specificamente per i modelli di classificazione; la cross-entropy, ad esempio è una funzione di perdita progettata per minimizzare la distanza tra la distribuzione della rete su etichette di classe e la distribuzione che le definisce set di dati.

CAPITOLO IV - Il Settore scelto e l'operazione di preparazione del dataset

Il settore di riferimento scelto per la costruzione del modello di machine learning per la valutazione del rischio di credito attraverso le reti neurali è quello delle telecomunicazioni, in particolare, delle telecomunicazioni fisse, mobili, satellitari e altre attività. In questo capitolo, si effettuerà un breve excursus del settore riguardante i principali andamenti che lo caratterizzano ed infine, si illustreranno nel dettaglio le operazioni di pulizia dei dati effettuate per ottenere il campione di analisi e le operazioni di preprocessing sulla base del quale si è costruito il modello.

4.1 settore telecomunicazioni in Italia

Il settore delle telecomunicazioni è uno dei campi più influenzati dalla trasformazione digitale. Nonostante questo, la digitalizzazione sembra aver messo a dura prova questo mercato più degli altri.

Con il termine telecomunicazioni si intende: *“l’insieme delle tecniche e dei procedimenti per la comunicazione a distanza di suoni (voce, musica), immagini, testi ecc., ai quali oggi ci si riferisce come segnali, quantità variabili nel tempo e di natura elettromagnetica, acustica o altra ancora, ormai quasi sempre rappresentati in forma numerica (o, più diffusamente, digitale)”*.

Le performance del settore delle telecomunicazioni in Italia sono risultate nel 2020 ancora significativamente negative in termini di variazioni rispetto all'anno precedente. Infatti, il volume d'affari complessivo si è affermato a 31.185 milioni di euro con una variazione del -3,4% rispetto al 2019.

Tale risultato ha un impatto ancora più negativo considerando che il 2020 è stato il tredicesimo anno consecutivo in cui le performance del settore dei servizi di telecomunicazioni che, in questo periodo, hanno visto diminuire in maniera non

indifferente il proprio valore totale. Infatti, nel corso dell'ultimo anno il mercato dei servizi fissi e mobili ha avuto una contrazione di più di un miliardo di euro.

I servizi di rete fissa hanno registrato una variazione del -4,2% rispetto al 2019 aggiudicandosi un valore di mercato di 9.750 milioni di euro.

Nel 2019 anche i servizi Internet hanno evidenziato un calo rispetto all'anno precedente. Tale calo è dovuto al fatto che nel corso del 2019 le offerte di connessione a prestazioni sempre più elevate (fibra e ultra-broadband) hanno evidenziato una maggiore competizione anche in termini di profitti medi per cliente.

Le linee fisse sono diminuite di quasi 740 mila unità nel corso del 2019 mentre le linee broadband sono aumentate di circa 430 mila unità nello stesso periodo superando la soglia dei 17,5 milioni di linee comprensive.

Tale tendenza è sempre più dovuta al solido e ininterrotto processo di abbandono delle connessioni tradizionali a vantaggio delle linee che possiedono prestazioni più elevate, solitamente per la sola connessione Internet, dato che la fonia viene sempre più fornita mediante servizi mobili.

Il mercato relativo ai servizi di rete mobile prosegue registrando anche nel corso del 2020 una tendenza particolarmente negativa.

I ricavi legati ai servizi mobili sono risultati, infatti, pari a 10.807 milioni di euro, con una perdita del -4,1% rispetto al 2019.

Nonostante un non indifferente aumento dei volumi di traffico dati, sia fisso che mobile, nel 2020 si è potuto constatare una tendenza negativa del comparto degli smartphone (-10,9%) dovuto alla chiusura improvvisa delle unità produttive e della capacità di spesa inferiore delle famiglie.

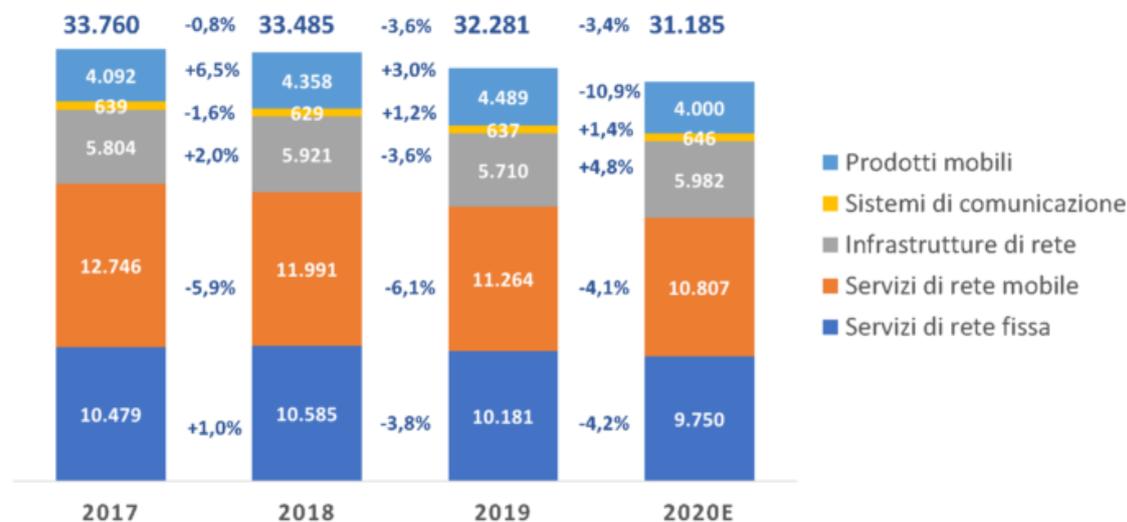


Figura 4.1: Andamento del valore di mercato dal 2017 al 2020 del settore Telecomunicazioni

In conclusione, le nuove tecnologie legate al 5G faranno strada a nuove occasioni di business in ambito non solo consumer ma anche industriale. Nel corso del prossimo decennio il 5G potrà abilitare molti miliardi di euro di investimenti nel processo di trasformazione digitale. È dunque possibile affermare che quello delle telecomunicazioni è un settore in continua evoluzione e altamente competitivo, in cui è necessario operare su due fronti. Se da un lato, infatti, risulta necessario migliorare e aumentare le prestazioni delle infrastrutture, in modo da aumentare i margini di produttività presumendo anche di ottimizzare con delle piattaforme ad implementazione e gestione agile, dall'altro lato, si deve riesaminare l'organizzazione dell'intero mercato, in modo da indirizzarsi verso servizi che permettano di generare un vantaggio competitivo realmente considerevole.

4.2 Lo scarico dei dati da Aida

Per costruire il dataset, è stato utilizzato il database AIDA, messo a disposizione del Politecnico di Torino. La banca dati Aida, che è l'acronimo di "Analisi Informatizzata Delle Aziende", contiene molte informazioni delle aziende italiane. Lo scarico dei dati è avvenuto su un file Excel, organizzato in due fogli:

- Foglio1: In questo foglio sono stati inseriti i dati generali sulle aziende considerate, i bilanci degli anni esaminati e le informazioni circa il loro stato giuridico e l'ultima procedura/cessazione subita: a tal proposito ad ogni azienda è stato associato un

flag, legato alla loro situazione finanziaria. I flag possono avere valore da 0 a 4 e sono così suddivisi:

- Flag 0: Sana senza particolari segnalazioni;
- Flag 1: Concordato preventivo, Fallimento, Amministrazione giudiziaria, Accordo di ristrutturazione dei debiti, Chiusura del fallimento, Altre cause, Liquidazione giudiziaria, Motivo non precisato, Stato di insolvenza, Sequestro giudiziario, Concordato fallimentare, Amministrazione controllata, Cancellazione per comunicazione piano di riparto, Amministrazione straordinaria, Chiusura per fallimento o liquidazione, Decreto cancellazione tribunale, Liquidazione coatta amministrativa, Scioglimento per atto dell'autorità, Sequestro conservativo di quote, Bancarotta;
- Flag 2: 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. (bilancio di liquidazione), Liquidazione, Scioglimento e messa in liquidazione, Chiusura per liquidazione, Scioglimento senza messa in liquidazione, Cessazione delle attività nella provincia, Cessazione d'ufficio;
- Flag 3: Fusione mediante incorporazione in altra società, Scissione, Trasferimento sede all'estero, Fusione mediante costituzione di nuova società, Cessione azienda;
- Flag 4: cessata, cancellata dal registro impresa, trasferimento in altra provincia, cancellata d'ufficio a seguito istituzione cciaa di fermo, di monza.
- Analisi dei bilanci: in questo secondo foglio sono stati sintetizzati i flag del Foglio 1 nel secondo modo:
 - FLAG DI STATUS S/A – SOCIETA: indica se la società è sana o anomala e vale 0 nei casi in cui le società rientrano nei casi “0”, “3” e “4”, “1” se le società rientrano nei casi “1” e “2”;
 - FLAG DI STATUS S/A – ANNO: vale “0” se le società non hanno subito nessuna procedura nell'anno di riferimento, “1” se hanno subito la procedura nell'anno in considerazione;

- FLAG SOC SANA IN LIQUIDAZIONE – SOC: vale "1" qualora le società avessero subito un processo di liquidazione in corrispondenza dell'anno, "0" in tutti gli altri casi.

Inoltre, sono stati effettuati dei controlli sugli errori, trattati nel capitolo successivo, dovuti alla non correttezza dei dati dei bilanci. Infine, sono stati calcolati 42 indicatori riguardo la redditività, la Produttività e la struttura operativa e la liquidità e la struttura finanziaria.

4.3 La pulizia del dataset e la correzione degli errori

La pulizia del dataset è il primo e cruciale passaggio per effettuare l'analisi dei dati. Il compito di classificazione sarebbe fuorviante e ridondante se i dati non fossero compresi e considerati completamente in anticipo. Questa attività comprende processi di manipolazione dei dati come il trattamento dei valori mancanti, la gestione dei valori infiniti, il rilevamento dei valori anomali e la prevenzione della perdita di dati.

Per prima cosa, sono stati sostituiti i valori “n.d.” ed “n.s” con 0 e si sono corretti gli eventuali errori #DIV/0! e #NUM!. Inoltre, per verificare la correttezza dello stato patrimoniale e del conto economico sono state inserite tre colonne che, attraverso operazioni di addizione e sottrazione, stabiliscono la presenza degli errori. In particolare, è stato notato che, a causa della mancanza di alcune voci parziali, l'attivo non coincidesse con la somma delle sue macro componenti, che l'attivo e il passivo non fossero uguali e che il conto economico non fosse completo. Per ogni riga, corrispondente ad un bilancio, si è controllato che tutti i tre controlli restituissero valori nulli e in caso contrario si sono indagate le cause e si sono corretti errori.

4.4: Preprocessing

Questa sezione è dedicata alla descrizione del campione di dati in analisi e ai dettagli implementativi relativi alla fase di preprocessing condotta prima del training della rete neurale. Nella prima parte del capitolo sono discussi la struttura del dataset, il suo

caricamento e le statistiche riguardanti la distribuzione degli indicatori; successivamente si prosegue alla discussione della prima tecnica di preprocessing applicata, ovvero la saturazione degli indicatori al 5° e al 95° percentile per la gestione dei valori anomali, o outliers; la terza parte è relativa alle tecniche di standardizzazione e normalizzazione del dataset, introdotte per rendere più stabile il training della rete neurale. Il capitolo si conclude con alla rimozione degli outliers attraverso la distanza di Mahalanobis e l'analisi di correlazione del dataset, quindi con la descrizione della fase di dimensionality reduction, in cui la dimensionalità del dataset viene ridotta in modo da eliminare indicatori ridondanti e/o poco significativi.

4.5.1 Dataset

Il dataset in analisi è composto da 20655 record, ciascuno definito da 42 indicatori. Dapprima, ci si è concentrati sul flag “FLAG DI STATUS S/A – SOCIETA”.

Il dataset organizzato in due file .csv, indicatori.csv, flag1.csv, contenenti rispettivamente per ogni record i valori degli indicatori e il flag di output. Per la lettura e il caricamento dei dati si è fatto uso della funzione `read_csv()` della libreria Pandas di Python.

Come è possibile notare nell'istogramma in figura..., il dataset non è bilanciato ma presenta una quantità di dati negativi (0) molto maggiore di quella dei dati positivi (1).

In entrambi i casi infatti, la percentuale di 0 nel dataset è del 78,9% mentre quella degli 1 è del 21,1%. Questo sbilanciamento si tradurrà nell'adozione di particolari strategie per risolvere il bias della rete a favore delle predizioni negative, come discusso nel capitolo 6.

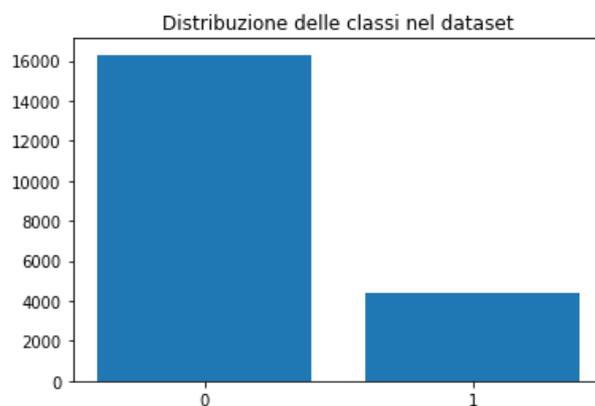


Figura 4.2: Distribuzioni delle classi

Nel grafico in Figura 4.3 sono riportate le distribuzioni dei valori per le varie features. E' possibile da subito osservare come la maggior parte dei valori assuma valori piccoli, molto vicino allo 0, mentre i valori di due indicatori si concentrino nel range [10,20].

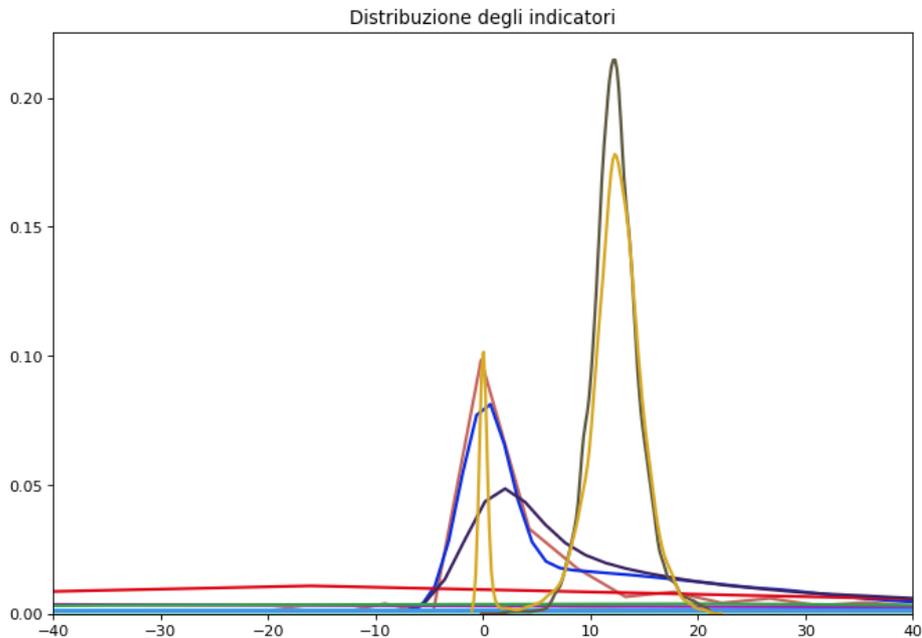


Figura 4.3: Distribuzioni delle features

Un'altra osservazione immediata è che le distribuzioni degli indicatori assumono forme distribuite (la scala di concentrazione di questo grafico è [0,20]%) e non definite.

Il grafico a boxplot, riportato in Figura 4.4, ci aiuta ad avere una visione più chiara della situazione.

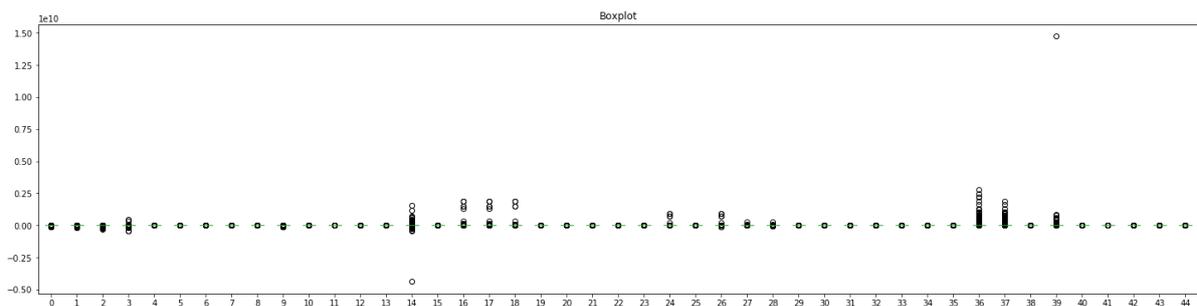


Figura 4.4: Grafico a boxplot delle features

Un grafico a boxplot è un grafico quantitativo che riassume in un modo univoco la distribuzione dei dati evidenziandone:

- Minimo
- 1° quartile o 25° percentile (Q1)
- Mediana

- 3° quartile o 75° percentile
- Massimo

Fornisce una rappresentazione intuitiva di come i dati sono dispersi e dell'eventuale presenza di anomalie: I punti fuori dal range [minimo, massimo], identificati con un cerchio nero, sono considerati outlier perchè sono troppo lontani dai valori centrali e quindi fuori dal range di valori prevalente,.

Nel grafico riportato è possibile notare più facilmente quanto esposto prima: gli indicatori 14,36 e 37 e 39 presentano il numero maggiore di outliers; in particolare gli indicatori 14 e 39 presentano due valori anomali estremamente lontani dal range [minimo,massimo].

4.5.2 Saturazione al 5° e 95° percentile

Il primo passo di pulizia dei dati effettuato è la saturazione dei valori assunti di ogni indicatore al 5° e al 95° percentile.

Si è deciso di svolgere il calcolo dei percentili direttamente su un foglio di calcolo Excel per semplicità. Il passo svolto in Python si è limitato a leggere il valore dei percentuali per ogni indicatore e sostituire nel dataset il valore di ogni indicatore minore del 5° percentile o maggiore del 95° percentile con il percentile stesso.

```
X_cleaned=X
for c in X.columns:
    X_cleaned[c]=X[c].where(X[c]>=percentili[c].iloc[0],percentili[c].iloc[0])
    X_cleaned[c]=X[c].where(X[c]<=percentili[c].iloc[1],percentili[c].iloc[1])
```

Questa tecnica consente di gestire i valori outlier per ogni indicatore allineandoli alle soglie definite dai percentili, senza così perdere informazioni come accadrebbe nel caso di rimozione totale del record con indicatore da allineare.

Un outlier infatti è un valore anomalo rispetto a una distribuzione e potrebbe perciò condurre ad errori di valutazione: generalmente si hanno valori anomali a causa di inserimenti sbagliati, misurazioni sbagliate o semplicemente per incertezze dovute ad eventi esterni.

È importante rimuovere gli outlier per permettere alla rete neurale di modellare ed astrarre al meglio la distribuzione dei dati: un valore anomalo potrebbe portare a modellazioni non corrette della conoscenza rappresentata nel dataset.

Le distribuzioni degli indicatori e i boxplot relativi dopo questa prima fase sono riportati nelle Figure 4.5 e 4.6.

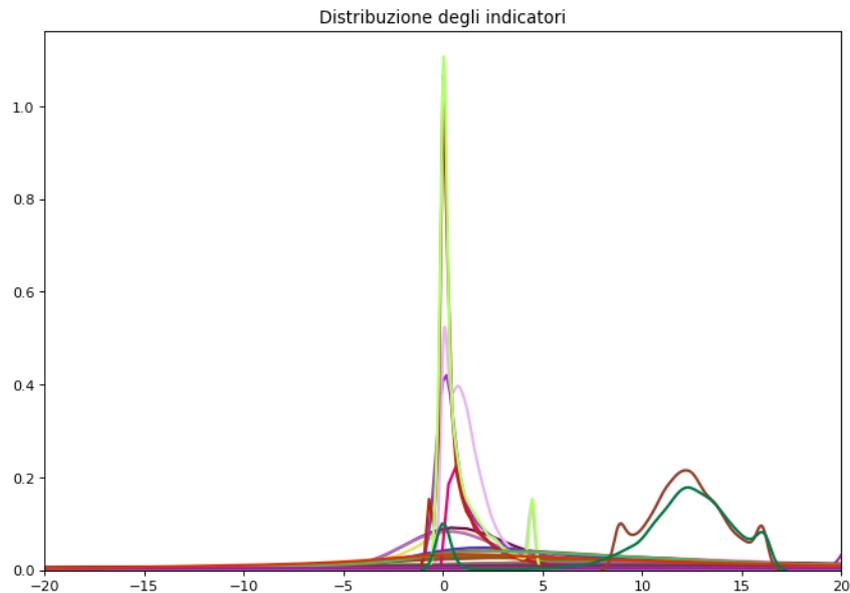


Figura 4.5: Distribuzioni delle features a seguito della saturazione

Guardando l'asse delle ascisse del primo grafico, è possibile notare immediatamente come il range di valori dei vari indicatori si è ristretto. La stessa cosa è visibile nel boxplot dove gli outliers presenti prima per gli indicatori 14 e 39 sono stati rimossi. Le distribuzioni per gli indicatori 14, 36 e 37 rimangono comunque troppo distribuite.

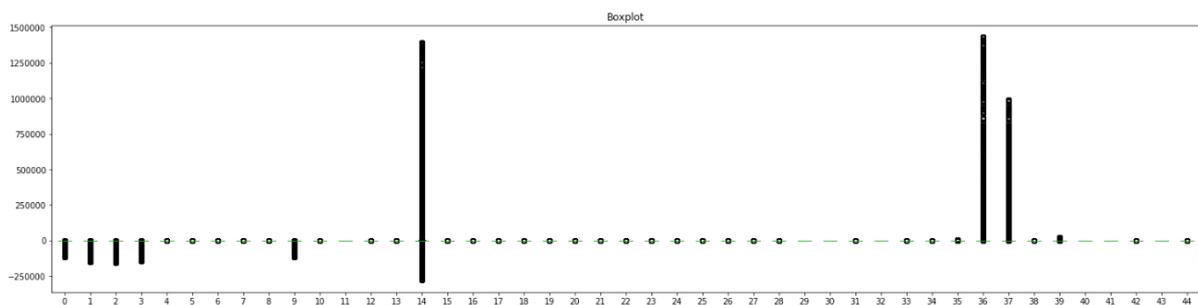


Figura 4.6: Grafico a boxplot delle features a seguito della saturazione

4.5.3 Standardizzazione e Normalizzazione

Questa fase è tipicamente introdotta nelle pipeline di preprocessing di qualsiasi algoritmo di machine learning o deep learning. Essa permette di eliminare le differenze di scala tra i vari indicatori, riportando i valori di tutti gli indicatori all'interno di un certo range (tipicamente $[0,1]$ o $[-1,1]$). Così facendo si evita che indicatori con una scala più ampia

dominano le metriche di distanza rispetto a indicatori con valori più piccoli; allo stesso tempo, restringere il range di valori a range standards intorno allo zero rende la fase di training più stabile con variazioni dei pesi del modello limitate, e ciò si traduce in una maggiore velocità di convergenza della rete nella fase di training.

In questo lavoro è stata analizzata l'applicazione al dataset sia della standardizzazione che della attraverso l'uso della libreria scikit-learn.

4.5.3.1 Normalizzazione con MinMax

Questa tecnica prevede di calcolare i valori assunti da ogni indicatore seguendo la formula:

$$y = (x - \min) / (\max - \min)$$

Questo processo riporta i valori in un range compreso tra [0,1]. Le figure 4.7 e 4.8 rappresentano rispettivamente le distribuzioni dei valori di ogni indicatore e il boxplot relativo a seguito dell'applicazione della normalizzazione min-max.

Questa operazione non modifica la forma delle distribuzioni.

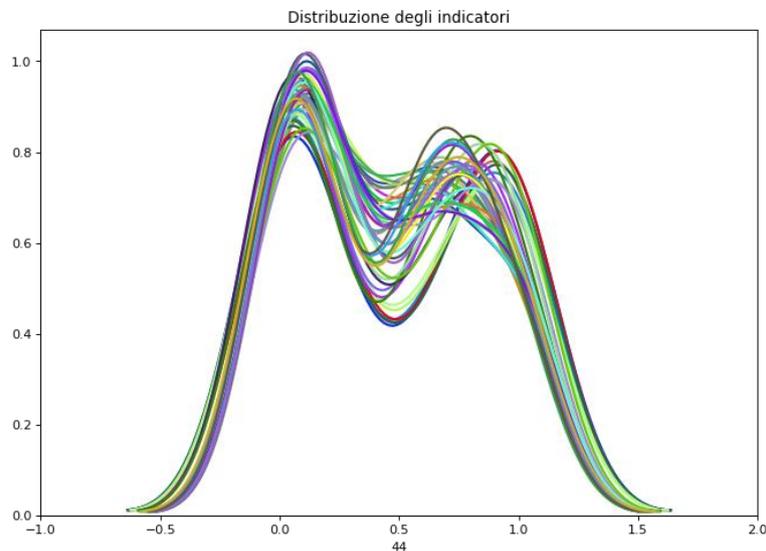


Figura 4.7: Distribuzioni delle features a seguito della normalizzazione

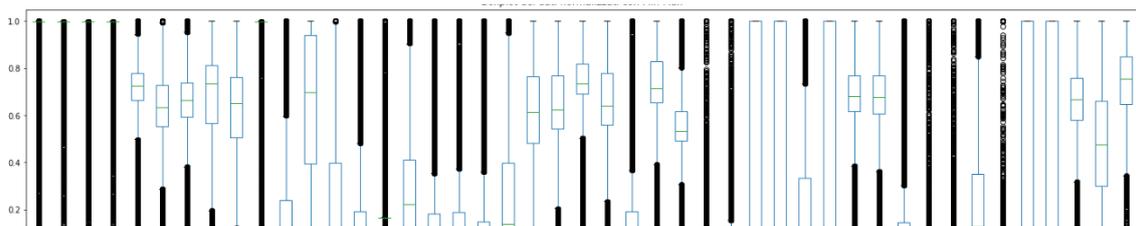


Figura 4.8: Grafico a boxplot delle features a seguito della normalizzazione

4.5.3.2 Standardizzazione dei dati con StandardScaler

La standardizzazione assume che i dati appartengono a una distribuzione Gaussiana per cui prevede che essi vengano ricalcolati in modo da avere media = 0 e standard deviation =1:

$$y = (x - \text{mean}) / \text{standard_deviation}$$

Ciò significa che il 68% dei valori per ogni indicatore risiederà nel range [-1,1].

Questa tecnica modifica la forma delle distribuzioni portandole ad assumere una distribuzione Gaussiana normale.

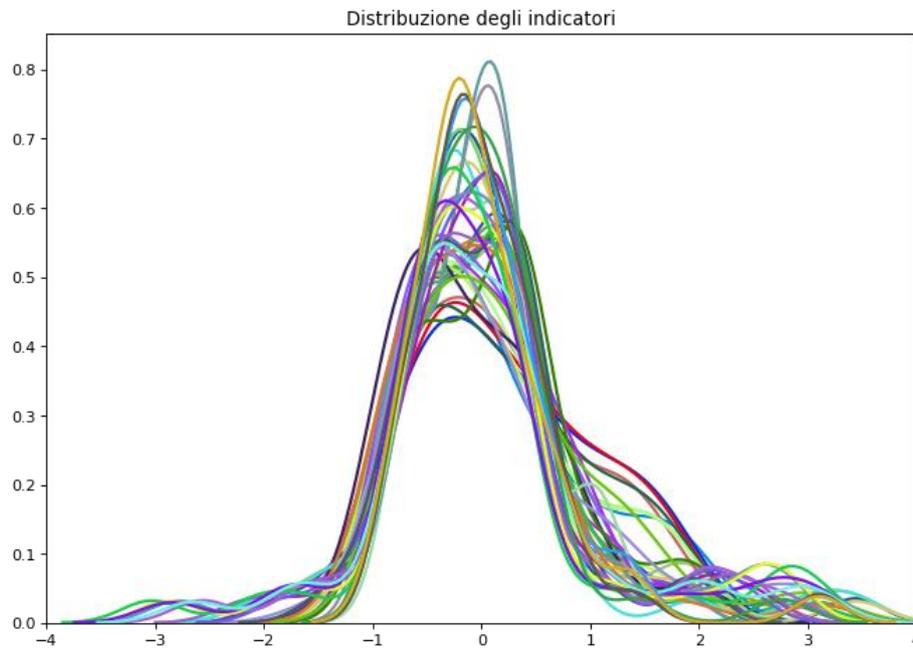


Figura 4.9: Distribuzioni delle features a seguito della standardizzazione

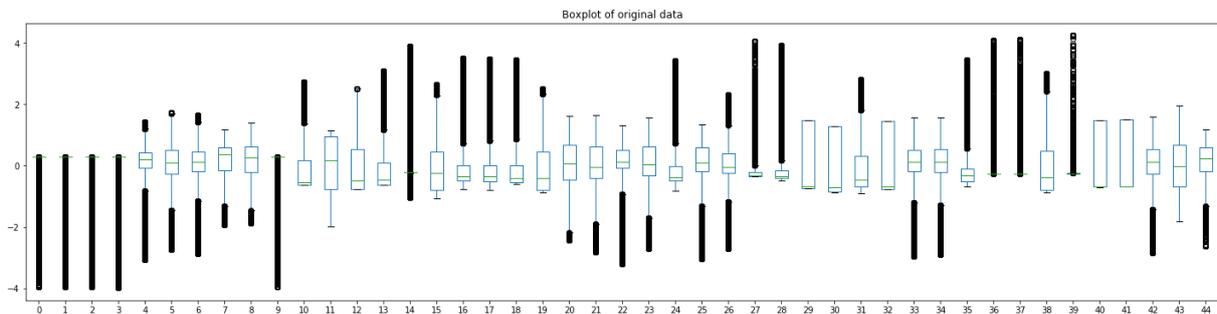


Figura 4.10: Grafico a boxplot delle features a seguito della standardizzazione

La standardizzazione ha portato a maggiore stabilità nella fase di apprendimento della rete per cui da questo momento in poi il resto dell'analisi farà riferimento al dataset standardizzato con StandardScaler.

4.5.4 Rimozione degli outliers tramite distanza di Mahalanobis

La distanza di Mahalanobis fornisce un metodo per il rilevamento dei valori anomali nei dati multidimensionali. È una metrica della distanza tra due punti che nella nostra analisi viene utilizzata per valutare la distanza di un'osservazione da una distribuzione (identificata dal valore medio di essa). Può essere vista come una variante multivariata della distanza euclidea ed è calcolata secondo la seguente formula:

$$D = \sqrt{(x - m)^T C^{-1} (x - m)}$$

Dove:

- D è la distanza di Mahalanobis;
- X è la matrice dei samples;
- m è la media dei valori;
- C è la matrice di covarianza.

La divisione per matrice di covarianza affronta il problema della correlazione dei dati, scalando la distanza in base alla forza della correlazione: maggiore è la correlazione, maggiore è il ridimensionamento.

Il grafico riportato in Figura 4.11, mostra la distribuzione della distanza di Mahalanobis per il dataset originale: sull'asse delle x è indicata la distanza di Mahalanobis e sull'asse delle y un valore intero progressivo che identifica il sample.

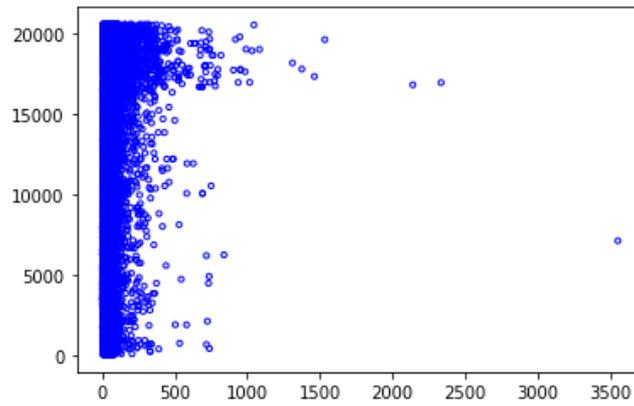


Figura 4.11: Distribuzione della distanza di Mahalanobis sul dataset iniziale

Successivamente i punti con distanza maggiore di 1500 sono stati considerati outlier e quindi rimossi dal dataset; la figura 4.12 mostra la distribuzione dopo la prima rimozione degli outlier.

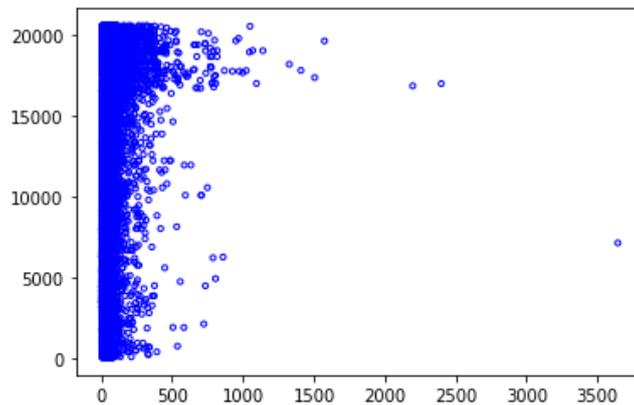


Figura 4.12: Distribuzione della distanza di Mahalanobis a seguito di una prima rimozione di outlier

Anche qui i punti con distanza maggiore di 600 sono stati rimossi ma con un effetto ininfluenza sulla distribuzione finale come si può osservare nella Figura 4.13.

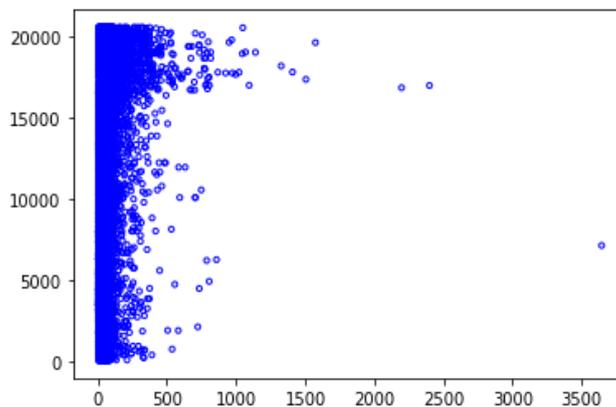


Figura 4.13: Distribuzione della distanza di Mahalanobis a seguito di una seconda rimozione di outlier

4.5.5 Analisi di correlazione

In questo capitolo si analizzerà la correlazione di ogni indicatore con il flag di output e la correlazione tra gli indicatori stressi.

Valutare la correlazione degli indicatori con i flag di output ci consente di avere una visione chiara di quanto ogni indicatore porti informazioni significative relativamente ai flag di output.

Riportiamo la lista ordinata in senso decrescente degli indicatori e della loro correlazione con l'output, ottenuta attraverso la seguente parte di codice:

```
Numeric_df = pd.DataFrame(X_normalized_cleaned_with_names)
Numeric_df['Y'] = Y_cleaned
corr= Numeric_df.corr()
corr_y = abs(corr["Y"])
#highest_corr = corr_y[corr_y > 0.1]
corr_y=corr_y.sort_values(ascending=False)
```

Ln(RIC)	0.224695
Ln(AN)	0.177298
OFN/EBITDA	0.176447
Debiti totali/EBITDA	0.175547
Debiti finanziari/EBITDA	0.174002
OFN/Autof lordo	0.173725
OFN/EBIT	0.157098
Deb totali/Ric	0.155606
PC/ric	0.153311
EBITDA/Ricavi	0.151373
Val Agg Oper/Ricavi	0.151200
EBIT/ricavi	0.147645
Utile corrente/ricavi	0.146645
Risultato netto rettif/ricavi	0.144911
Riserve+utile/AN	0.142499
Autof Lordo/AN (=cash flow/attivo)	0.140719
ROE ante imposte	0.139964
ROE	0.139840
Auto Lordo-comp straord/AN	0.137718
Ebitda-servizio debito/AN	0.136269
EBITDA/AN	0.134331
Risultato netto rettif/AN	0.129761
ROA	0.125798
Deb finanziari (stimati)/Ric	0.118070
OF/RIC	0.109092
OFN/RIC	0.102870
Deb finanziari (stimati)/VA	0.099072
costo lavoro/costi operativi	0.092644
Patr netto tang/AN	0.088898
Ricavi/AN	0.087398
Patr netto/AN	0.085521
Patr netto tan/Debiti tot+PN	0.081903
servizi esterni/costi operativi	0.070710
Patr netto tan/Debiti tot-Liq+PN	0.061243
Cap Circ/AN	0.056436
Val Agg/ITN	0.039584
ammortam materiali/costi operativi	0.030479
OFN/AN	0.027711
AC-mag/PC	0.027513
AC/PC	0.027220
Consumi/costi operativi	0.025126
Liq/PC	0.015458
Deb totali/VA	0.013744
Patr netto/debiti totali	0.006101
Liq/AN	0.005596

Risulta possibile osservare che gli indicatori risultano in generale debolmente correlati ai flag di output con valori di correlazione che raggiungono al massimo il 22,4%.

Per la visualizzazione della correlazione tra gli indicatori è stata utilizzata matrice di correlazione.

Nel codice sottostante, il metodo `corr()` della libreria Pandas è usato per valutare la matrice di correlazione attraverso il metodo di Pearson, conosciuto anche come PMCC:

```
f = plt.figure(figsize=(19, 15))
plt.matshow(X_normalized_cleaned_with_names.corr(), fignum=f.number)
plt.xticks(range(X_normalized_cleaned_with_names.shape[1]), X_normalized_cleaned_with_names.columns, fontsize=14, rotation=45)
plt.yticks(range(X_normalized_cleaned_with_names.shape[1]), X_normalized_cleaned_with_names.columns, fontsize=14)
bottom, top = plt.gca().get_ylim()
plt.ylim(bottom, top)
cb = plt.colorbar()
cb.ax.tick_params(labelsize=14)
```

Il coefficiente espresso per ogni coppia di indicatori è un numero compreso nel range [-1,1] e indica quanto la coppia di variabili è correlata: valori positivi indicano una correlazione positiva, valori negativi correlazione negativa; quanto più si è vicini allo 0, meno forte è la correlazione.

Nella figura 4.14 è riportata la matrice di correlazione per il dataset dopo gli step di preprocessing descritti nelle sezioni precedenti.

La leggenda dei colori aiuta a capire il grado di correlazione di ogni coppia di indicatori. Logicamente, sulla diagonale tutti i valori essere uguale a 1 poiché ogni indicatore è fortemente correlato a se stesso. Come è possibile notare ci sono molte regioni (maggiormente in alto a sinistra) fortemente correlate, il che ci suggerisce l'utilizzo di tecniche di dimensionality reduction introdotte nella sezione successiva. Questa forte correlazione non ci sorprende poiché il calcolo di un indicatore molto spesso dipende dai valori assunti da altri indicatori presenti nel dataset.

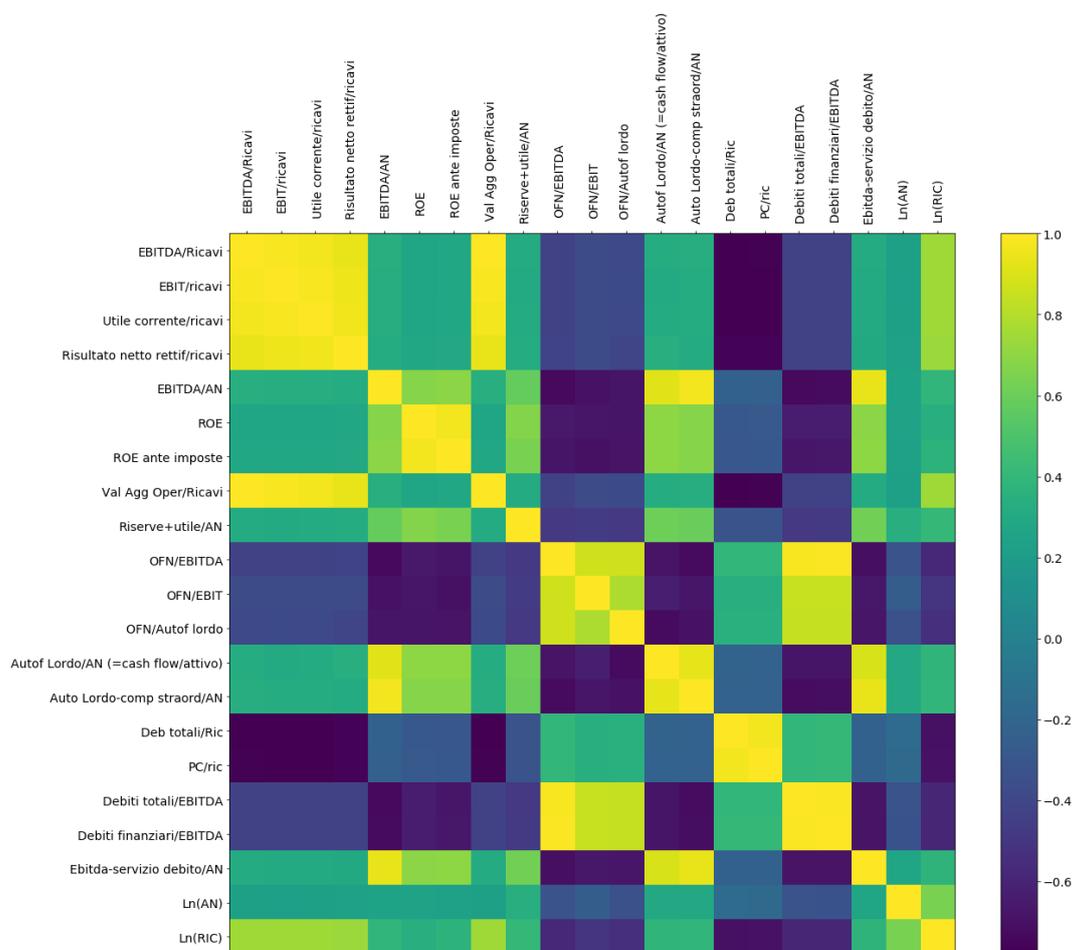


Figura 4.14: Matrice di correlazione tra gli indicatori

4.5.5.1 Riduzione della dimensionalità

A fronte dell'analisi di correlazione presentata prima, si è ritenuto necessario ridurre la dimensionalità del dataset con i seguenti obiettivi:

- Evitare il fenomeno della curse of dimensionality, tipico in dataset ad alta dimensionalità;
- Mantenere solo gli indicatori che forniscono informazioni rilevanti sui dati (alta correlazione con il flag di output e bassa correlazione con gli altri indicatori).

Il processo seguito si è diviso nei seguenti passi:

- Filtraggio degli indicatori con correlazione con il flag di output maggiore o uguale a 0.13;

- Identificazione delle coppie di indicatori con alta correlazione (valore assoluto maggiore di 0.7);
- Rimozione per ognuna delle coppie identificate, dell'indicatore con correlazione con il flag di output più basso.

Per definire le soglie di selezione ottimali e quindi per ottenere il sottoinsieme di indicatori migliore, il suddetto processo è stato applicato iterativamente.

Questa fase ci ha portato a ridimensionare il dataset considerando solo gli indicatori ad alta correlazione con il flag di output e a bassa correlazione tra di loro.

Il sottoinsieme di indicatori scelto è stato:

- ROE ante imposte
- Riserve+utile/AN
- OFN/EBITDA
- Deb totali/Ric
- Ln(AN)
- Ln(RIC)

È stato quindi ridotto il numero di indicatori da 45 a 6. Nella Figura 4.15 è riportata la matrice di correlazione che si ottiene dopo la riduzione della dimensionalità.

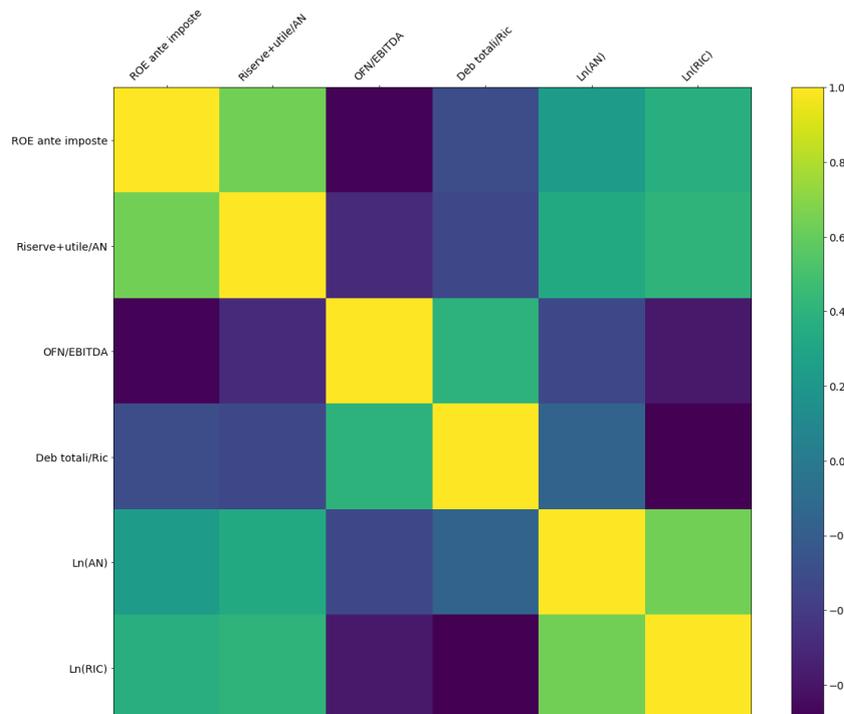


Figura 4.15: Matrice di correlazione a seguito della riduzione di dimensionalità

CAPITOLO V- La costruzione del modello di rete neurale

Lo scopo di questo capitolo è quello di discutere i dettagli progettuali e implementativi della soluzione neurale proposta.

Il problema affrontato è un task di classificazione binaria completamente supervisionato (fully supervised) su un dataset sbilanciato. La rete neurale proposta è stata quindi trainata sul dataset discusso nel precedente capitolo affinché sia in grado di predire una valutazione del rischio di credito delle società appartenenti al settore telecomunicazioni.

Il training di una rete neurale prevede solitamente tre fasi distinte condotte su split diversi del dataset originario: training, validation e test. La divisione dei dati e l'implementazione delle tre fasi vengono affrontati nella sezione 5.1. Nella sezione 5.2 sono riportate delle tecniche necessarie al fine di ribilanciare il dataset. Nella sezione 5.3 viene presentata l'architettura neurale proposta, la cui configurazione iniziale è discussa nella sezione 5.4 insieme all'inizializzazione dei parametri per la fase di training. L'ultima sezione di questo capitolo, la 5.5, si concentra sulla discussione della fase di validazione e sul tuning degli iperparametri.

Si noti che il codice è stato scritto in Python e che si è fatto uso del framework Pytorch per l'implementazione della rete neurale e delle funzioni ad essa relative.

5.1 Training-Validation-Test

Un problema si dice “fully supervised” quando il dataset è composto da dati etichettati con una delle classi che si vuole predire ed ogni classe è rappresentata nel dataset. In un problema di questo tipo, lo scopo della fase di training di una rete neurale è quello di modellare la conoscenza di un problema astraendo una funzione capace di associare un dato all'etichetta di classe corretta: una rete neurale infatti può essere vista come una serie di operazioni semplici e funzioni di attivazione che insieme modellano una funzione complessa non lineare; tuttavia, la “realtà” nota durante la fase di training si limita alla conoscenza rappresentata dal dataset: l'assunzione di fondo è infatti che il dataset

rappresenti un sampling della distribuzione dei dati della realtà che si vuole modellare; la fase di training quindi consiste nel minimizzare l'errore empirico della rete su tale sampling.

Dividere il dataset in training, validation e test split è una pratica comunemente usata in problemi di questo tipo.

Le etichette di classe si assumono visibili solo per il training e il validation set mentre quelle relative al test set vengono ignorate, poiché quest'ultimo si assume rappresenti un'approssimazione della realtà, in cui perciò non esiste nessuna conoscenza a priori sulle classi di appartenenza dei dati. Il test set viene quindi utilizzato solo per esplicitare le performance del modello finale.

La fase di training iniziale prevede che il modello venga trainato esclusivamente sul training set per consentire la valutazione delle performance del set di iperparametri adottato: durante questa fase, i parametri del modello vengono continuamente aggiornati affinché l'errore di predizione del modello sui dati di training sia minimo; l'errore è valutato attraverso una funzione di perdita (loss function) che può essere interpretata ad alto livello come la distanza media delle predizioni della rete sui dati dalla loro ground truth (cioè dalle etichette di classe previste).

Durante il training, le performance del modello vengono valutate sul validation set. Quest'ultimo viene utilizzato per valutare i fenomeni di overfitting/underfitting descritti nel capitolo 3 e per la fase di hyperparameter tuning, cioè la fase in cui, in base alle performance del modello trainato con diversi iperparametri, viene scelto l'insieme degli iperparametri da utilizzare per il modello finale: questa fase è molto importante e può essere meno o più strutturata.

Scelto il set di iperparametri definitivo, la fase di training viene ripetuta sull'unione del set di training e validation per generare il modello finale da valutare sul test set.

Il dataset originario è stato diviso in training-validation-test splits in base al rapporto 2/3-1/6-1/6; in ognuno degli split il rapporto tra la rappresentanza delle classi è stato mantenuto uguale a quello nel dataset di partenza. In particolare, i data split utilizzati sono:

- training: 13587 samples, tra cui 2883 positivi e 10704 negativi;
- validation: 3500 samples, tra cui 728 positivi e 2772 negativi;
- test: 3500 samples, tra cui 721 positivi e 2779 negativi.

5.2. Ribilanciamento dei dati: SMOTE

Come già discusso nel precedente capitolo, il dataset in analisi risulta sbilanciamento verso la classe negativa.

Operare su un dataset sbilanciato per ottenere una classificazione binaria è di solito molto difficile in quanto il modello potrebbe tendere a predire maggiormente la classe più rappresentata in quanto domina la funzione di perdita durante la fase di training; tuttavia, come anche nel nostro caso, la classe di minoranza è spesso quella a cui si è maggiormente interessati.

Ci sono molte tecniche per gestire dataset sbilanciati, tra cui le più comuni sono downsampling dei dati, loss weighting oppure oversampling.

Durante il corso dell'analisi, si è escluso la tecnica di downsampling in quanto avrebbe comportato una riduzione drastica della mole di dati considerata, e quindi ad una perdita di informazione significativa. Il loss weighting e l'oversampling invece sono entrambe tecniche che sono state testate, con l'ultima delle due che è stata scelta come tecnica definitiva in quanto ha portato a una fase di training più stabile e a performance più elevate.

L'approccio più semplice di oversampling è quello di duplicare i dati di minoranza (relativi al flag 1) affinché i dati di maggioranza (flag 0) vengano bilanciati: questo processo tuttavia, seppur porti ad ottenere un dataset bilanciato, non fornisce alcuna informazione aggiuntiva.

Una tecnica che abbiamo invece rilevato più significativo è lo SMOTE (Synthetic Minority Oversampling Technique), che permette di ottenere l'oversampling aggiungendo informazioni tramite la sintetizzazione di dati effettivamente nuovi: senza

scendere nei dettagli, se immaginiamo i dati in uno spazio multidimensionale identificato dagli indicatori, i nuovi samples vengono presi lungo le linee congiungenti due dati spazialmente vicini.

Per applicare l'agoritmo, ci siamo serviti della classe SMOTETomek della libreria imbalanced-learn.

```
from imblearn.combine import SMOTETomek
smt = SMOTETomek(random_state=42)
```

```
X_oversampled_train, Y_oversampled_train = smt.fit_resample(x_train, y_train)
```

5.3. Architettura della rete

La scelta dei dettagli architetturali della rete neurale ha seguito un processo iterativo di adattamento nel corso degli esperimenti.

Si è alla fine optato per un'architettura semplice e non troppo profonda con l'obiettivo di limitare la capacità di modellizzazione e quindi l'effetto di overfitting sul training set, molto probabile su un dataset sbilanciato.

```
class Network(nn.Module):

    def __init__(self, in_ch, out_ch):
        super().__init__()
        # Inputs to hidden layer linear transformation
        self.fc1 = nn.Linear(in_ch, 12)
        self.fc2 = nn.Linear(8, 12)
        self.fc3 = nn.Linear(12, out_ch)

    def forward(self, x):
        x = F.leaky_relu(self.fc1(x))
        x = F.leaky_relu(self.fc2(x))
        return self.fc3(x)
```

Come mostrato nel codice riportato, la rete è costituita da tre layers lineari, rispettivamente con 6, 8 3 12 nodi; l'output è binario; all'output di ogni strato lineare è applicata una funzione di attivazione leaky relu, descritta nel dettaglio nel capitolo 3.

5.4. Configurazione iniziale

Come già ampiamente discusso, trainare una rete in un approccio completamente supervisionato significa trovare la migliore configurazione dei parametri per minimizzare l'errore empirico di predizione della rete sui dati di training; nel prossimo capitolo la fase di training verrà dettagliata, ma è importante anticipare che l'aggiornamento dei parametri della rete durante il ciclo di training dipende dalla strategia di ottimizzazione adottata.

Durante la fase di training, mini-batch di dati, ovvero sottogruppi del train set, vengono iterativamente sottoposti alla rete. Per ognuno la rete effettua le predizioni sulle quali poi viene calcolata la funzione di perdita. La derivata della loss function rispetto ai parametri della rete (gradiente) rappresenta l'impatto di ogni parametro (weight) dei diversi nodi sul risultato finale e quindi l'aggiornamento necessario del parametro per ridurre il suddetto errore.

Lo scopo dell'aggiornamento iterativo dei parametri ad ogni step di forward sui mini-batch quindi è quello di minimizzare la loss-function media sul dataset. Trovare la configurazione dei parametri ottimale non è banale in quanto l'andamento della funzione di loss potrebbe convergere lentamente o incontrare zone di stallo o di minimi locali. Le strategie con cui i parametri vengono aggiornati dopo ogni step di forward nella rete di un mini-batch vengono dette strategie di ottimizzazione e sono fondamentali per assicurare una convergenza ottimale della funzione di perdita.

La più comune è la SGD, Stochastic Gradient Descent, espressa come di seguito:

$$W_{t+1} = W_t - \eta \nabla_W L_t(W_t)$$

La strategia di ottimizzazione adottata negli esperimenti è la SGD, Stochastic Gradient Descent.

Ad ogni step t i parametri della rete W vengono aggiornati di una quantità pari al gradiente della loss function rispetto ai parametri, scalato per una learning rate η .

La learning rate è un iperparametro che controlla la dimensione degli aggiornamenti dei parametri.

La dimensione dei mini-batch e la quantità di iterazioni di training da effettuare sono altri iperparametri fondamentali da definire.

La velocità di convergenza della rete a una configurazione tale da garantire performance accettabili dipende dalla scelta degli iperparametri e dalla configurazione iniziale dei parametri della rete.

5.4.1 Inizializzazione dei parametri della rete

La velocità di convergenza della fase di training dipende molto anche dal punto di partenza, definito ai valori iniziali dei pesi della rete. Negli esperimenti, i pesi dei neuroni sono stati inizializzati con la tecnica di Xavier mentre i termini di bias sono stati azzerati.

```
self.fc1.bias.data.fill_(0)
self.fc2.bias.data.fill_(0)
self.fc3.bias.data.fill_(0)

torch.nn.init.xavier_normal_(self.fc1.weight.data)
torch.nn.init.xavier_normal_(self.fc2.weight.data)
torch.nn.init.xavier_normal_(self.fc3.weight.data)
```

Inizializzare con valori troppo piccoli i pesi della rete infatti potrebbe portare al fenomeno del vanishing gradient, ovvero a una situazione in cui la rete non viene trainata a causa dei valori quasi nulli assunti dai gradienti.

Valori troppo grandi porterebbero a instabilità della rete o a fenomeni di loss explosion, in cui i valori dei pesi diventano troppo grandi finendo per generare dei NaN dopo le prime iterazioni.

L'inizializzazione di Xavier garantisce performance efficaci ed è comunemente utilizzata nel deep learning.

5.4.2 Strategia di ottimizzazione - SGD Optimizer

Si è usata la classe SGD di Pytorch definita nel modulo optim per l'ottimizzazione:

```
optimizer = optim.SGD(model.parameters(), lr=LR, momentum=MOMENTUM, weight_decay=5e-4)
```

La definizione dell'oggetto optimizer richiede l'insieme dei parametri da aggiornare (model.parameters()), il fattore di aggiornamento (learning rate (LR)), il momento e il Weight decay.

Il momento può essere visto come il peso di una media mobile dei gradienti ed è comunemente settato a 0.9 (anche nel nostro caso). Il Weight decay invece è un termine addizionale alla loss usato per la regolarizzazione. Questo termine consente di penalizzare i pesi troppo grandi, limitando il fenomeno dell'overfitting. Il SGD con weight-decay può essere formalizzato in questo modo:

$$w_i \leftarrow w_i - \eta \frac{\partial E}{\partial w_i} - \eta \lambda w_i.$$

dove:

- w_i sono i pesi della rete allo step i ;
- η è la learning rate
- $\frac{\partial E}{\partial w_i}$, è la derivata della funzione di loss rispetto ai pesi
- λ , è il weight decay.

5.4.3 Funzione di loss

Essendo un problema di classificazione binaria, la funzione di loss scelta è la cross-entropy binaria (BCEWithLogitsLoss), definita, in accordo a quanto riportato sulla documentazione ufficiale di Pytorch, come segue:

$L(x,y) = \text{mean}(L)$, dove $L = \{l_1, l_2, \dots, l_N\}$

Dove N rappresenta la dimensione dei mini-batch

5.5. Validation

Di seguito è riportato l'elenco degli iperparametri che è stato necessario definire per la fase di training e quindi oggetto della fase di ricerca dei parametri:

- Learning rate: definisce il fattore di aggiornamento dei parametri ad ogni iterazione di training così come definito nella SGD;
- Numero di epoche: il numero di volte che il dataset è visto dalla rete durante la fase di training;
- Weight decay: il parametro di regolarizzazione presente nell' optimizer;
- Dimensione del mini-batch

Come già anticipato, non fa parte di questa lista il momento, in quanto settato al valore comune di 0.9.

Si noti bene la distinzione tra “epoca” e “iterazione” di training: con “iterazione” si intende il singolo passaggio nella rete di un mini-batch che porta all'aggiornamento dei parametri con SGD; con “epoca” si fa riferimento all'insieme di iterazioni che definiscono un ciclo completo sul dataset. Configurare il numero di epoche piuttosto che quello di iterazioni ci consente di avere una visione più chiara sul numero di volte che tutti i dati sono effettivamente visti dalla rete.

Si noti anche che nel visualizzare l'andamento delle loss, si preferisce considerare le loss medie delle iterazioni di ogni epoca piuttosto che quelle delle singole iterazioni.

La fase di esplorazione dei parametri condotta può essere divisa nei seguenti passi:

- scelta della batch size, impostata a 30 empiricamente;
- Scelta empirica della Weight decay partendo da valori comunemente accettati;
- Scelta della learning rate;
- Scelta del numero di iterazioni.

Il processo non è stato condotto in maniera ordinata ma iterativa in quanto le performance di un parametro sono strettamente dipendenti dai valori degli altri parametri.

Per valutare le diverse performance della learning rate si sono considerati progressivamente range di esplorazioni più piccoli partendo da $[1e-6, 1e-3]$ e testando l'andamento della loss nelle prime epoche sia sul training che sul validation set:

se nelle prime iterazioni si verificavano fenomeni di loss explosion, la LR testata è stata considerata troppo alta, restringendo l'upper bound del range; al contrario, fenomeni di decrescita lenta o assente della loss, hanno portato a un restringimento del lower bound del range.

Arrivati a un range di valori che consentissero una decrescita continua della loss di training, si è passati alla valutazione del fenomeno di overfitting.

In questa fase, si è tenuta in considerazione l'andamento della loss di validazione in relazione a quella di training e il valore della weight decay.

Si è cercato quindi il giusto trade-off tra diminuzione della learning rate e aumento del fattore di Weight decay:

- Valori troppo alti di loss function portavano infatti a fenomeni di overfitting, con un'andamento crescente o piatto della loss di validazione in confronto a quello decrescente della loss di training;
- Valori troppo alti di weight decay, portavano ad azzeramento dei gradienti;

La scelta del numero di epoche si è basata sull'osservazione empirica degli andamenti delle loss function, terminando la fase di training laddove si verificasse un loro appiattimento.

Al termine di questa fase i valori scelti per la fase di training finale sono stati:

- Learning rate: $5e-4$
- Numero di epoche: 150
- Weight decay: $5e-4$
- Dimensione del mini-batch: 50

CAPITOLO VI – Analisi delle performance del modello

Questo capitolo è dedicato all'analisi quantitativa delle performance raggiunte nei nostri esperimenti con le scelte implementative descritte in precedenza.

La sezione 6.1 presenta nel dettaglio la fase di training del nostro modello, discutendo il codice e le performance sul training e validation set, con i grafici dei relativi trend.

La sezione 6.2 introduce la fase di test, commentata nel capitolo precedente e le metriche comunemente usate nel deep learning per la valutazione delle performance dei modello neurali nei problemi di classificazione binaria; dunque è stato presentato il risultato finale degli esperimenti e delle analisi effettuati attraverso l'applicazione delle metriche di valutazione precedentemente descritte. Infine, nella sezione 6.3 sono stati riportati i risultati ottenuti effettuando un'ulteriore valutazione su un altro flag.

6.1 Fase di training

Di seguito è riportata l'implementazione della fase di training:

```
critterion = nn.BCEWithLogitsLoss()
optimizer = optim.SGD(model.parameters(), lr=LR, momentum=MOMENTUM,
weight_decay=5e-4)
scheduler = torch.optim.lr_scheduler.StepLR(optimizer, step_size=1000, gamma=0.1)
train_loss = []
val_loss = []
valFlag=True

for epoch in range(1,MAX_ITER+1):
    model.train(True)

    running_loss=0.0
    total_train=0
    correct_train=0
    train = iter(train_loader)
    i=0

    for data, target, t in train:
        optimizer.zero_grad()
        out = model(data)
        loss = critterion(out,target.float())
        loss.backward()
        optimizer.step()
        running_loss+=loss.item()
        i+=1

    train_loss.append(running_loss/i)
```

```

print("[TRAIN] - Lr: "+str(scheduler.get_lr())+
      " Epoch: "+str(epoch)+"/"+str(MAX_ITER)+
      " Train loss: %.3f" % (running_loss/i))

if (valFlag==True):
    ## Validation
    model.eval()
    running_loss=0.0
    total_train=0
    correct_train=0
    val_iter = iter(val_loader)
    i=0
    for data, target, t in val_iter:
        out = model(data)
        loss = criterion(out,target.float())
        running_loss+=loss.item()
        i+=1

    val_loss.append(running_loss/i)
    print("[VAL] - Epoch: "+str(epoch)+"/"+str(MAX_ITER)+
          " Val loss: %.3f" % (running_loss/i))

```

Per ogni epoca (fino a `MAX_ITER`), i dati del training set vengono passati all'interno della rete sotto-forma di mini-batch; prima del forward di ogni mini-batch, i gradienti sono azzerati (`optimizer.zero_grad()`) per consentire di accumulare solo quelli relativi all'iterazione in corso; successivamente, il modello predice l'output relativo ai dati passati (`out=model(data)`) su cui viene calcolata la funzione di perdita rispetto alla ground-truth (`criterion(...)`).

A questo punto, la fase di backward, e quindi di calcolo dei gradienti dei parametri, viene avviata attraverso il metodo `backward()` della loss: in questo modo, ad ogni parametro della rete verrà associato un gradiente calcolato come la derivata della loss rispetto al parametro stesso, seguendo la cosiddetta *chain rule*.

L'aggiornamento effettivo dei parametri secondo la strategia di ottimizzazione SGD viene effettuata attraverso la chiamata a `step()` sull'optimizer.

Si noti che la fase di validazione è speculare a quella di training ma senza la parte relativa al processo di back propagation e quindi all'optimizer; in più prima del ciclo di validazione il modello viene posto in modalità valutazione attraverso la funzione `eval()` in modo da non tenere traccia dei gradienti; la modalità di training viene ripristinata prima di ogni ciclo di training attraverso il metodo `train(True)`.

Per osservare le performance del modello durante il training, si è scelto di tenere traccia del valore della funzione di perdita media sul train set e sul validation set ad ogni epoca.

Nelle Figura 6.1 sono riportati i grafici dell'andamento delle loss di training e di validation lungo le epoche.

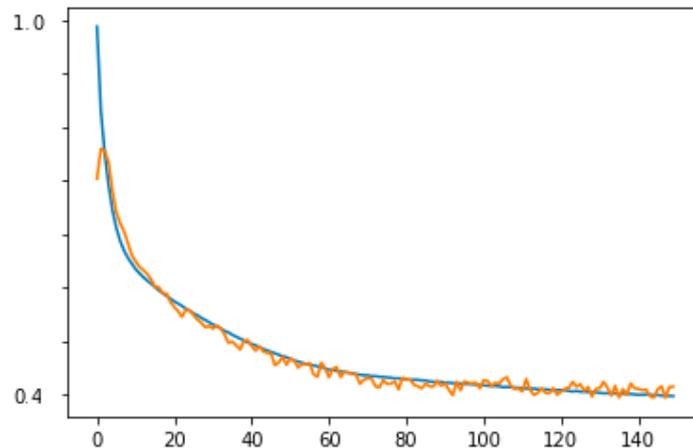


Figura 6.1: Loss function della fase di test e di validation

Si può facilmente notare che come ci aspettavamo, l'andamento delle loss function è decrescente esponenziale: nelle prime epoche la fase di aggiornamento dei parametri è stata maggiormente significativa il che giustifica l'andamento più ripido; all'aumentare delle epoche l'andamento delle loss si è appiattito il che identifica la fase di convergenza ad un minimo.

Vale la pena notare come l'andamento della loss sul validation set sia “frastagliata” rispetto a quello della loss sul training set che al contrario è molto smooth: la motivazione di quest'instabilità risiede nell'aver eseguito il ribilanciamento dei dati tramite SMOTE solo sul training set; la loss sul validation set è stata invece calcolata nella condizione di dataset sbilanciato.

Si noti anche che le scelte implementative hanno con successo evitato l'overfitting: infatti l'andamento delle loss è decrescente sia per il training che per il validation set.

Come anticipato nella sezione precedente, dopo aver analizzato le performance del modello per vari set di iperparametri e dopo aver scelto il set ottimale, il modello neurale finale è stato ottenuto ripetendo la fase di training sull'unione del training e del validation set.

Il grafico sottostante rappresenta l'andamento della loss durante la fase di training finale.

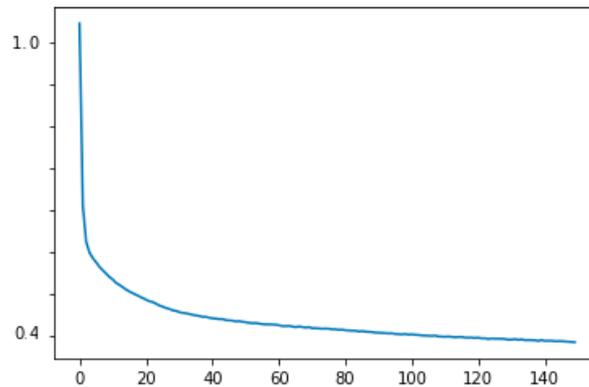


Figura 6.2: Loss function della fase di training finale

6.2 Fase di test e Metriche di valutazione

Per testare il modello ci si è serviti del seguente codice:

```
from sklearn.metrics import confusion_matrix

model.eval()

test_iter = iter(test_loader)

predicted = []
labels = []

for data, target, t in test_iter:

    out = model(data)

    #res = out.ge(.5).view(-1)

    _, res = torch.max(out, 1)

    for r, t in zip(res.numpy(), t.numpy()):
        predicted.append(r)
        labels.append(t)

predicted = np.array(predicted)
labels = np.array(labels)
```

Per analizzare i risultati del nostro modello, ci si è serviti delle comuni metriche di valutazione utilizzate in problemi di classificazione binaria: accuratezza, precisione e recall.

In un problema di classificazione binaria possiamo assegnare a ogni predizione uno dei seguenti tag:

- True Positive (TP): identifica un dato appartenente alla classe positiva, classificato correttamente;
- True Negative (TN): identifica un dato appartenente alla classe negativa, classificato correttamente;
- False Positive (FP): identifica un dato appartenente alla classe positiva, classificato erroneamente;
- False Negative (FN): identifica un dato appartenente alla classe negativa, classificato erroneamente;

La matrice in Figura 6.3 è detta confusion matrix e riassume quanto descritto:

		Ground truth	
		1	0
Output	1	TP	FP
	0	FN	TN

Figura 6.3: Confusion matrix

Di seguito vengono riportate le definizioni delle metriche di valutazione utilizzate, in funzione dei tag definiti sopra.

6.2.1 Accuratezza

L'accuratezza è una metrica globale calcolata sull'insieme delle classi:

$$Acc = (TP + TN) / (TP + TN + FN + FP)$$

6.2.2 Precision & Recall

La precisione e la recall sono metriche di valutazione definite per classe.

$$Precision_0 = \frac{TN}{TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall_0 = \frac{TN}{TN + FP}$$

$$Recall_1 = \frac{TP}{TP + FN}$$

Sebbene l'accuratezza sembri una metrica affidabile e riassuntiva delle performance globali del modello, essa non è un indice affidabile nei problemi di classificazione binaria, soprattutto in quelli caratterizzati da un forte sbilanciamento tra le classi come nel nostro caso.

Questo proprio perché essa è sensibile alla percentuale di rappresentanza tra le classi: infatti, le performance sulla classe di maggioranza dominano questa metrica.

Il nostro dataset è caratterizzato da una presenza di dati negativi pari all'80% e di dati positivi pari al 20%: se il modello predicesse la classe negativa per tutti i dati, l'accuratezza sarebbe pari all'80%.

Per questo motivo nel corso della nostra analisi ci siamo concentrati maggiormente sulla precisione e la recall.

In particolare, si noti che le performance su un dataset sbilanciato sono di solito molto più basse rispetto a quelle di dataset bilanciati a causa della difficoltà a risolvere completamente il bias verso la classe di maggioranza, che dipende molto non solo dall'approccio seguito ma anche dalla qualità dei dati.

Di seguito è riportata la Confusion Matrix ottenuta con il modello finale sul test set e le relative metriche misurate ottenute con il seguente codice:

```
import numpy as np

def _fast_hist(label_true, label_pred, n_class):
    mask = (label_true >= 0) & (label_true < n_class)
    #print(np.unique(label_pred))
    #print(np.unique(label_true))
    hist = np.bincount(
        n_class * label_true[mask].astype(int) + label_pred[mask],
```

```

        minlength=n_class ** 2,
    ).reshape(n_class, n_class)
    return hist

def scores(label_trues, label_preds, n_class):
    hist = np.zeros((n_class, n_class))
    for lt, lp in zip(label_trues.numpy(), label_preds.numpy()):
        if(lt.size > 0):
            hist += _fast_hist(lt.flatten(), lp.flatten(), n_class)
    acc = np.diag(hist).sum() / hist.sum()

    acc_cls = np.diag(hist) / (hist.sum(axis=1)+ 1e-3)

    cls_acc = dict(zip(range(n_class), acc_cls))
    recall_cls = dict(zip(range(n_class), acc_cls))
    precision_cls = np.diag(hist) / (hist.sum(axis=0) + 1e-3)
    precision_cls = dict(zip(range(n_class), precision_cls))

    return acc, cls_acc, precision_cls, recall_cls, hist

```

Si è deciso di riportare la Confusion Matrix sia sul test set originario (senza aver applicato la tecnica dello SMOTE) che su quello bilanciato per sottolineare quanto la dominanza di una classe influisca anche sugli errori di predizione.

Non applicando la tecnica dello SMOTE si hanno i seguenti risultati:

```

ACCURACY: 0.6394285714285715
CLS ACCURACY: {0: 0.6509533461844742, 1: 0.5950061095615679}
CLS PRECISION: {0: 0.8610181527757482, 1: 0.30664738624203985}
CLS RECALL: {0: 0.6509533461844742, 1: 0.5950061095615679}
array([[1809., 970.],
       [ 292., 429.]])

```

Applicando la tecnica dello SMOTE si hanno i seguenti risultati:

```

ACCURACY: 0.6393380970289583
CLS ACCURACY: {0: 0.6596462355598963, 1: 0.6190294776120806}
CLS PRECISION: {0: 0.6338993010844592, 1: 0.6452369089624034}
CLS RECALL: {0: 0.6596462355598963, 1: 0.6190294776120806}
array([[1754., 905.],
       [1013., 1646.]])

```

6.3 Considerazioni su un ulteriore giro di rete neurale: il caso delle società in liquidazione

È stato effettuato un ulteriore giro di rete neurale, considerando però come flag binario l'unione tra i seguenti due flag:

- FLAG DI STATUS S/A – SOCIETA;
- FLAG SOC SANA IN LIQUIDAZIONE – SOC.

Infatti, è stato constatato che tra le società che hanno subito dei processi di liquidazione sono presenti società liquidate per motivi di default: in questo modo si considerano queste società come anomale. Ovviamente, sarà commesso un errore dovuto alla presenza di società sane, liquidate per volontà dei soci a causa di incongruenze tra di loro oppure per raggiungimento dell'oggetto sociale (caso molto raro): Aida infatti non fa una distinzione tra questi casi.

A prova del fatto che sono state elaborate le reti neurali in maniera corretta, si riportano in figura 6.3 i grafici di training e di validation e il grafico della loss durante la fase di training finale in figura 6.4, che ovviamente sono decrescenti, come nel caso precedentemente studiato.

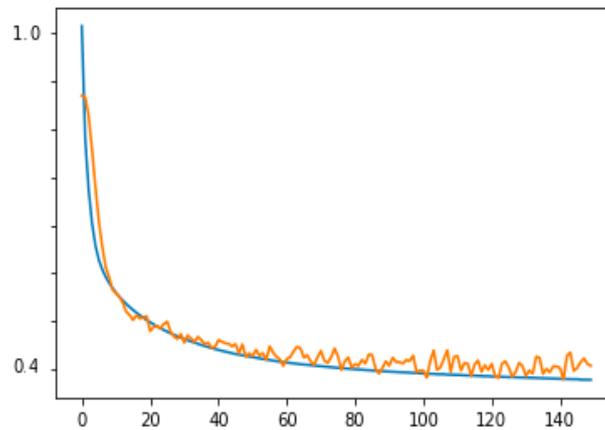


Figura 6.3: Loss function della fase di test e di validation per il Flag 2

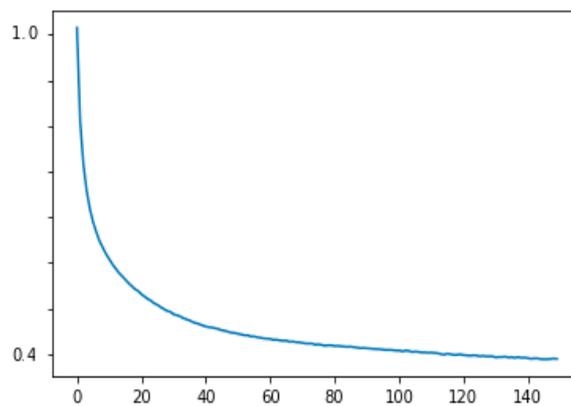


Figura 6.4: Loss function della fase di training finale del Flag 2

Dalle confusion matrix riportate di seguito è possibile notare che si hanno performance minori rispetto al primo giro di elaborazioni di reti neurali effettuato.

Ciò, molto probabilmente, è dovuto all'errore citato all'inizio della sezione, causato dal fatto che sono state considerate anomale società in liquidazione, che in realtà sono sane.

Non applicando la tecnica dello SMOTE si hanno i seguenti risultati:

```
ACCURACY: 0.6265714285714286  
CLS ACCURACY: {0: 0.6284170401377553, 1: 0.6194435841061332}  
CLS PRECISION: {0: 0.8644231249761876, 1: 0.30155490090946524}  
CLS RECALL: {0: 0.6284170401377553, 1: 0.6194435841061332}
```

array([[1747., 1033.],
[274., 446.]])

Applicando la tecnica dello SMOTE si hanno i seguenti risultati:

```
ACCURACY: 0.6427492447129909  
CLS ACCURACY: {0: 0.6382172816399994, 1: 0.6472807223260112}  
CLS PRECISION: {0: 0.6440546326011308, 1: 0.6414668257983436}  
CLS RECALL: {0: 0.6382172816399994, 1: 0.6472807223260112}
```

array([[1690., 958.],
[934., 1714.]])

CONCLUSIONI

Lo scopo di questa tesi è stato quello di introdurre una serie di approcci al rischio di credito basati sull'intelligenza artificiale.

Il processo seguito in questo lavoro di tesi ha prodotto i risultati attesi con performance sufficienti alla risoluzione del task prefissato.

Le performance potrebbero tuttavia essere incrementate in lavori futuri attraverso un'analisi più approfondita sui dati presenti nel dataset. Relativamente agli indicatori di bilancio, è possibile stabilire informazioni più corrette laddove i dati non siano disponibili oppure chiaramente imprecisi, attraverso un'analisi di regressione, positiva o negativa in base al flag assegnato alla società. Inoltre, si sarebbero raggiunte performance più elevate se si fosse tenuto conto in maniera strutturata della dipendenza temporale dei parametri aziendali nel corso degli anni, associando il valore FLAG DI STATUS S/A – SOCIETA, non solo rispetto alla situazione finale della società, ma anche tenendo conto degli indicatori che esprimono l'andamento delle società nel corso degli anni.

In conclusione, nel settore finanziario, l'applicazione di tecniche di intelligenza artificiale e machine learning ha visto un aumento di interesse nel corso degli anni. Un numero crescente di studi è stato condotto sulla gestione del rischio, sulla previsione di fallimenti e sui sistemi di valutazione del credito supportati dall'intelligenza artificiale. Analogamente alla natura di questi concetti, il machine learning ha anche un'importante applicazione nella previsione del default del prestito. Una migliore comprensione delle insolvenze dei prestiti, di ciò che contribuisce al rischio di insolvenze dei prestiti e di come prevederle, avrà enormi vantaggi per l'industria del mercato del debito.

BIBLIOGRAFIA E SITOGRAFIA

- 1) Dispense e slide dei corsi di Economia degli intermediari finanziari e Analisi finanziaria e creditizia per l'impresa tenuti dal professore Franco Varetto.
- 2) Federica Miglietta, "APPUNTI: DA BASILEA 1 A BASILEA 3", Università di Bari
https://www.uniba.it/docenti/miglietta-federica/lezionebasilea_2015.pdf
- 3) Andrea Sironi, M. Marsella, "La misurazione e la gestione del rischio di credito. Modelli, strumenti e politiche" 1998
- 4) Andrea Resti Andrea Sironi, "Rischio e valore nelle banche. Risk management e capital allocation", EGEA, Milano, II edizione, 13 agosto 2008.
- 5) Crescenzo Gallo, Reti Neurali Artificiali: Teoria ed Applicazioni Finanziarie
https://www.researchgate.net/publication/23693039_Reti_Neurali_Artificiali_Teoria_ed_Applicazioni_Finanziarie
- 6) A Beginner's Guide to Neural Networks and Deep Learning
<https://wiki.pathmind.com/neural-network>
- 7) AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?
<https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>
- 8) Michael Nielsen, "Neural Networks and Deep Learning"
- 9) Loss and Loss Functions for Training Deep Learning Neural Networks
<https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/>
- 10) Python AI: How to Build a Neural Network & Make Predictions
<https://realpython.com/python-ai-neural-network/>
- 11) Keras Tutorial: Deep Learning in Python
<https://www.datacamp.com/community/tutorials/deep-learning-python>
- 12) Machine Learning:Reti Neurali
<https://docplayer.it/108615174-Machine-learning-reti-neurali-sommario.html>

RINGRAZIAMENTI

Al termine di questo percorso, mi è doveroso dedicare qualche riga del mio elaborato alle persone che hanno contribuito con il loro supporto alla realizzazione dello stesso e che mi hanno sostenuto in questi anni di università.

In primis, vorrei tanto ringraziare il mio relatore, il Professore Franco Varetto, per la sua disponibilità tempestiva ad ogni mia richiesta e per il suo sostegno.

Ringrazio i miei genitori e i miei fratelli, per il loro supporto morale ed economico che mi hanno permesso di percorrere e concludere questo cammino.

Un grazie ai miei due migliori amici, Giovanni ed Attilio, che hanno avuto un peso determinante nel conseguimento di questo risultato.

Un grazie particolare alla mia seconda famiglia, i miei coinquilini, ufficiali e non, passati e presenti, che mi hanno sempre supportato e (forse più) sopportato ogni singolo giorno. Grazie a Marco, Riccardo, Giordana, Silvia, Enrico, Alessio, Davide e Mattia.

Ringrazio tutti i miei colleghi, diventati amici e confidenti, in particolare Chiara, Federica, Maria, Tony e Mich per aver condiviso con me in questi anni un percorso così bello e importante, ma non privo di momenti di difficoltà.

Ed infine, ringrazio me stesso per la determinazione e l'impegno concretizzato al fine di raggiungere questo traguardo, la voglia di sfidarmi, per averci provato ed esserci riuscito.