

POLITECNICO DI TORINO

**Corso di Laurea Magistrale
in Ingegneria Gestionale**

Tesi di Laurea Magistrale

**Credit Risk Scoring Model con metodologie di
data science**



Relatore

prof. Franco Varetto

Candidato

Niccolò Mangione

Anno Accademico 2018/2019

*Alla mia famiglia,
che mi ha permesso di intraprendere questo percorso
e di raggiungere questo obiettivo.
Ai miei amici, con cui ho condiviso tutto.
A Giada, la mia forza più grande.*

Sommario	8
1 Sella Personal Credit	10
1.1 Introduzione	10
1.2 Presentazione	10
1.3 Attività	10
1.4 Struttura della società	12
1.5 Business della società	12
1.5.1 Credito al consumo	13
1.5.2 Credito a PMI o Credito alle Imprese	14
1.5.3 Credito a Condominio	14
1.5.4 Cessione del Quinto dello Stipendio, della Pensione e Deleghe di Pagamento	14
1.5.5 Carte di Credito	15
1.5.6 Leasing	15
1.6 Andamento economico	16
1.7 Canali di distribuzione	18
1.8 Patrimonio	19
1.8.1 Fondi propri	20
1.8.2 Adeguatezza patrimoniale	21
1.9 Gestione dei rischi	21
1.9.1 Rischio di Credito	21
1.9.2 Rischi di Mercato	29
1.9.3 Rischio di prezzo	31
1.9.4 Rischi operativi	31
1.9.5 Rischio di liquidità	36
2 Rischio di credito – Regulation	38
2.1 Introduzione	38
2.1 Definizione	38
2.2 Componenti del Rischio di credito	38
2.2.1 Perdita attesa	39
2.2.2 Perdita inattesa	39
2.3 Tipologie di rischio di credito	40
2.4 Disposizioni normative sul rischio di credito	40
2.5 La crisi finanziaria del 2007-2009	42
2.6 La risposta del comitato di Basilea	43
2.6.1 Provvedimenti Microprudenziali	44
2.6.2 Provvedimenti Macroprudenziali	48

2.6.3	Applicazione degli standard.....	50
2.7	Il recepimento della normativa post Basilea 3 nel quadro normativo europeo	51
2.7.1	Il framework normativo	52
2.8	Approcci metodologici	53
2.8.1	Standard Approach – SA	54
2.8.2	Metodo basato sui rating interni	60
2.9	Verso Basilea 4	64
3	Rischio di credito: Modelli di scoring	66
3.1	Introduzione	66
3.2	Requisiti di un modello.....	66
3.3	Approcci metodologici	67
3.3.1	Modello univariato.....	67
3.3.2	Modelli multivariati	68
3.3.3	Analisi discriminante lineare	68
3.3.4	Z-Score di Altman.....	70
3.3.5	Affinamento del modello di Altman	71
3.3.6	Approfondimento al calcolo del cut-off.....	72
3.4	Regressione logistica	75
3.5	Confronto tra analisi discriminante e regressione logistica.....	77
3.6	Fasi del processo di stima di un modello.....	77
3.6.1	Selezione del campione.....	78
3.6.2	Selezione delle variabili di analisi	78
3.6.3	Stima del modello	79
3.6.4	Test del modello.....	88
3.6.5	Calibrazione del modello	89
3.6.6	Master Scale.....	90
3.6.7	Backtesting del modello.....	91
4	Descrizione del progetto	94
4.1	Introduzione	94
4.2	Scopo del progetto	94
4.3	Definizione dei clienti.....	95
4.4	Metodologie di misurazione del rischio di credito	96
4.4.1	Reti Neurali (Artificial Neural Network - ANN).....	96
4.4.2	Alberi decisionali (CART).....	99
4.4.3	Tecniche Random Forest	101
4.5	Osservazioni	102

5	La metodologia adottata	103
5.1	Introduzione	103
5.2	Portafoglio su cui si applica il modello	103
5.3	Modello predittivo	104
5.3.1	Analisi delle variabili	105
5.3.2	Modello di regressione e definizione della soglia	106
5.3.3	Definizione del modello predittivo	109
5.3.4	Fit del modello	110
5.3.5	Training e validazione.....	111
5.3.6	Definizione delle classi e assegnazione delle PD	114
5.3.7	Analisi di performance del modello.....	114
5.4	Osservazioni	116
6	Esplorazione dei dati – Analisi ed opportunità	118
6.1	Introduzione	118
6.2	Analisi delle variabili.....	118
6.3	Profilo di un buon pagatore	125
6.4	Dati innovativi	126
6.5	Osservazioni	127
7	Applicazione del modello.....	128
7.1	Introduzione	128
7.2	Pattern di variabili selezionate.....	129
7.3	Modello di regressione, fit e definizione della soglia.....	130
7.4	Training e test	131
7.4.1	Reject Inference	134
7.5	Definizione delle classi e assegnazione delle PD	135
7.6	Validazione del modello	141
7.7	Osservazioni	142
8	Conclusioni.....	144
9	Bibliografia.....	146

Sommario

Il presente lavoro di tesi è stato basato sul progetto Credit Risk Assessment di Sella Personal Credit con la consulenza di un team dell'azienda Moxoff , al fine di sviluppare un modello di valutazione del rischio di credito mediante l'utilizzo di modelli e metodi di data science e introdurre un set di variabili che attualmente SPC non raccoglie e non archivia, ma che potrebbero rivelarsi utili per il processo di valutazione della clientela. Il modello sviluppato dovrà essere in grado di percepire modifiche significative nella popolazione richiedente e quindi proporre delle modifiche ai regressori utilizzati nel modello di stima. Questo meccanismo di auto adattamento avviene mediante un riaddestramento effettuato automaticamente. Il progetto si pone inoltre lo scopo di ridurre al minimo l'interazione manuale da parte del valutatore, al fine di ridurre valutazioni soggettive e possibili errori. Il modello si basa sullo sviluppo di una regressione per la valutazione del rischio associato ad un cliente senza effettuare richieste al credit bureau. La tipologia di modello prevede l'utilizzo di un algoritmo di machine learning del tipo Random Forest e Neural Network che offre la possibilità di simulare in maniera alquanto semplice ed automatica una moltitudine di scenari e di casistiche. Tale metodologia ha il vantaggio di poter integrare in maniera altrettanto agevole eventuali nuove informazioni o variabili.

1 Sella Personal Credit

1.1 Introduzione

Nel seguente capitolo viene presentata l'azienda in cui è stato svolto il progetto di tesi. Si fornisce un'idea generale delle attività di business, in termini di prodotti e di servizi offerti e della distribuzione sul territorio, di seguito viene riportato l'andamento economico facendo riferimento ai dati dei bilanci più recenti, la struttura patrimoniale e la gestione dei rischi nel rispetto delle normative vigenti, particolare risalto viene dato al rischio di credito, per il quale vengono descritti i sistemi di controllo, misurazione e gestione di tali rischi e le tecniche di mitigazione adottate.

1.2 Presentazione

Sella Personal Credit, nata come Consel nel 1999 e così ridenominata il 1° ottobre 2018, appartiene al gruppo Sella. Le regole di seguito esposte sono espressione della "Carta di identità" che la Società ha adottato come punto di riferimento per il proprio sviluppo, riassunta nelle indicazioni relative a visione, missione e valori:

- Visione: "Credere che investire sulla persona, metterla al centro dell'attività, sia un acceleratore per la creazione di ricchezza".
- Missione: "Nello svolgimento dell'attività del credito al consumo, sia presso i negozi sia direttamente con i Clienti, è necessario poter disporre di un gruppo di lavoro motivato, dove la persona sia rispettata e protagonista. Per questo l'Azienda deve disporre di un gruppo di quadri e dirigenti che sappiano mettersi in gioco e in discussione costantemente, lavorando non solo sull'apprendimento tecnico e professionale, ma anche su quello che porta ad un irrobustimento della personalità mediante un cammino personale di crescita interiore".
- Valori: "La fiducia della Clientela, mantenendo un'immagine e una reputazione fatta di: correttezza, serietà, onorabilità, professionalità".

L'obiettivo è quello di entrare subito nel mercato del credito al consumo classico e delle più innovative forme di pagamento: un mercato che, sempre in crescita negli ultimi anni, vede ancora tuttavia il nostro Paese in posizione marginale rispetto al contesto europeo e ai livelli raggiunti dai Paesi occidentali più avanzati. Tramite il credito al consumo e le carte di credito ad utilizzo rateale, si assicura continuità e maggiore impulso ad una linea di forte attenzione alle nuove tecnologie, ai più moderni e innovativi ambiti di attività, agli strumenti finanziari e ai servizi alla clientela che da sempre ne hanno contraddistinto le scelte e le strategie.

1.3 Attività

La Società ha per oggetto l'esercizio nei confronti del pubblico dell'attività di concessione di finanziamenti prevista dall'articolo 106, comma 1, del D.Lgs. 1° settembre 1993, n. 385 e specificata dalle disposizioni attuative pro tempore vigenti nonché la prestazione dei servizi di pagamento di cui all'articolo 1, comma 1, lettera b), punti 4 e 5, del D.Lgs 27 gennaio 2010, n. 11 – per lo svolgimento dei quali è stato costituito con delibera del Consiglio di

Amministrazione del 27 gennaio 2011 un patrimonio destinato -, come previsto dagli articoli 106, 114-sexies e 114-octies del predetto D.Lgs 1° settembre 1993, n. 385 e dai dmt 6 luglio 1994 e 28 giugno 1996 o dalla normativa pro tempore vigente, ed in particolare:

- l'erogazione del credito al consumo in ogni sua forma, sia direttamente con finanziamenti ai consumatori, sia indirettamente con la concessione di credito ai venditori in stretta relazione all'operazione di vendita con pagamento differito, effettuata nei confronti dei consumatori persone fisiche, con scadenza e con i limiti di importo stabiliti dalla normativa pro tempore vigente;
- l'erogazione di finanziamenti a professionisti, artigiani e imprenditori, Società in genere, in ogni loro forma, che richiedano i finanziamenti per esigenze derivanti dalla propria attività;
- l'emissione o la gestione di carte di credito, in ogni loro forma prevista, anche con modalità rotativa ed utilizzabili eventualmente per l'erogazione del credito al consumo di cui all'articolo 121 del D.lgs. 1° settembre 1993, n. 385.

La Società può prestare e commercializzare alla propria clientela prodotti assicurativi a copertura del credito e altresì assumere, per la propria rete di vendita, l'incarico di agenzia finalizzato al collocamento di prodotti assicurativi e bancari per conto di imprese di assicurazione o banche.

La Società avendo costituito per l'espletamento della propria attività istituzionale apposite strutture ed in particolare per la valutazione del merito creditizio, per l'attività di recupero crediti e per l'attività di marketing e direct marketing, può prestare tali servizi a favore di terzi.

La Società può altresì compiere nel rispetto della normativa vigente tutte le attività previste dal presente articolo anche mediante la vendita a distanza, ove compatibile.

Con l'osservanza delle disposizioni vigenti e previo ottenimento delle prescritte autorizzazioni, ove occorrono, la Società può compiere ogni altra attività finanziaria nonché ogni altra attività connessa o strumentale alla realizzazione dell'oggetto sociale, ivi compresa l'assunzione di partecipazioni, anche di maggioranza, in Società italiane od estere aventi per oggetto attività correlate allo scopo sociale, con la tradizionale esclusione tuttavia di ogni operazione meramente aleatoria.

I canali commerciali attraverso cui avviene l'erogazione sono quelli tipici del credito al consumo: negozi, grande distribuzione e distribuzione organizzata. Le merceologie che maggiormente si prestano vanno dal settore mezzi di trasporto (auto nuove e usate, moto e motocicli, caravan e roulotte, biciclette), al credito finalizzato per acquisto di beni durevoli (elettrodomestici, telefonia, computer, hi-fi e foto-cine-ottica, mobili e arredamento, articoli sportivi e hobbistica), al settore dei servizi (viaggi e vacanze, palestre e piscine, corsi scolastici e di lingue, spese mediche e dentistiche...).

I vantaggi che il credito al consumo può assicurare: il punto vendita che realizza un finanziamento rateale, tramite un rapporto di convenzione con Sella Personal Credit, ha in mano uno strumento efficace per lo sviluppo della propria attività, che gli consente incassi rapidi e sicuri, la possibilità di erogare al proprio cliente un credito immediato e di aumentare il livello di fidelizzazione. Dal punto di vista del cliente-consumatore, credito al consumo significa acquisti più facili e rapidi, immediata disponibilità finanziaria, possibilità di dilazionare i pagamenti e non dover ricorrere a disinvestimenti di capitale.

1.4 Struttura della società

Costituiscono la struttura della Società:

a) Organi statutari

- Assemblea
- Organo con Funzione di Supervisione Strategica (Consiglio di Amministrazione)
- Presidente del Consiglio di Amministrazione
- Vicepresidente del Consiglio di Amministrazione
- Organo con Funzione di Gestione (Amministratore Delegato)
- Direttore Generale
- Organo con Funzione di Controllo (Collegio Sindacale)
- Comitato Fidi

b) Meccanismi:

- Comitato di Direzione Operativa
- Comitato Controlli
- Comitato Fidi
- Comitato del Personale
- Comitato Accettazione Clientela
- Comitato Controllo Rischi
- Comitato Andamento Spese Amministrative
- Comitato Progetti
- Comitato Commerciale
- Comitato Innovazione Tecnologica
- Comitato Agenzie di Recupero

c) Unità organizzative:

- Internal Audit
- Risk Management
- Compliance
- Risorse Umane
- IT e Organizzazione
- Credito
- Commerciale
- Affari Societari, Amministrazione e Bilancio

1.5 Business della società

La società fonda il suo business sulla cessione di crediti distinti in impieghi con clientela, con banche e verso enti finanziari, sia erogati direttamente sia acquistati da terzi, che prevedono pagamenti fissi o comunque determinabili, che non sono quotati in un mercato attivo e che non sono stati classificati all'origine tra le attività finanziarie disponibili per la vendita. La tabella 1.1 seguente riporta la composizione dei crediti divisi per le 3 macro-aree di business presi dai bilanci consolidati del 2016 e 2017:

Tabella 1.1: Composizione dei crediti Sella Personal Credit

	31/12/2015	31/12/2016	% var 15-16	31/12/2017	% var 16- 17
Crediti verso le banche	3.672.851	427.921	-88,3%	1.852.918	+326,7%
Crediti verso Enti Finanziari	12.784	1.724.989	Ns	20.298	-98,8%
Crediti verso Clienti	866.347.295	850.082.581	-1,9%	897.508.691	+5,6%
Totale	870.032.930	852.253.491	-2,0%	899.354.907	+5,5%

Si può osservare come il credito verso clienti sia la principale fonte di business della società, costituendo il 93% de crediti nell'anno 2015 e il 94% nei successivi 2 anni.

I crediti verso banche sono costituiti dai saldi attivi e competenze sui conti correnti.

I crediti verso Enti Finanziari sono incrementati nel 2016 a seguito dell'operazione di cartolarizzazione avvenuta nel mese di maggio 2017, a cui è seguita una forte riduzione nel 2017 per effetto dell'esercizio dell'operazione di estinzione anticipata.

1.5.1 Credito al consumo

Con credito al consumo si intende il credito per riacquisto di beni e servizi (prestito finalizzato) ovvero per soddisfare esigenze di natura personale (ad esempio: prestito personale). Non costituisce credito al consumo il prestito concesso per esigenze di carattere professionale del consumatore (ad esempio, acquisto di un'autovettura da utilizzare per il trasporto dei dipendenti della propria impresa). Il credito al consumo, tra i vari requisiti richiesti, per essere tale deve soddisfare i seguenti punti:

- essere compreso tra i 200 e i 75.000 euro
- non essere finalizzato all'acquisto o conservazione di immobili
- avere una garanzia di valore inferiore a quello del finanziamento, se riguarda un prestito garantito da beni mobili (per esempio da titoli)
- avere una durata inferiore o uguale a 5 anni, se riguarda un prestito garantito da
- ipoteca

Rientrano nel credito al consumo:

- Credito finalizzato

Tale prodotto prevede, a fronte dell'acquisto di un bene o servizio, l'erogazione di una somma pari al costo del bene o servizio acquistato e il suo progressivo ammortamento con un piano con rate predeterminate. La somma finale non è versata al cliente ma, su disposizione di quest'ultimo, al venditore del bene o servizio oggetto del finanziamento. L'offerta del finanziamento è operata dalla distribuzione commerciale attraverso il canale tradizionale e quello virtuale rappresentato da Internet. La rete commerciale stipula con gli esercenti apposite convenzioni che regolano l'attività distributiva.

- Credito personale

Credito che prevede l'erogazione di una somma di denaro da restituire secondo un piano di ammortamento costante a rate fisse. Viene considerato un prestito "non finalizzato" ovvero un'erogazione non evidentemente connessa all'acquisto di uno specifico bene. Il finanziamento è concluso solamente tra finanziatore e richiedente senza dover dichiarare il

motivo dell'esigenza monetaria. Per questo motivo, dal momento in cui la richiesta viene accettata il denaro è stanziato direttamente al consumatore richiedente e non al soggetto convenzionato (detto dealer ovvero, letteralmente, il commerciante). Proprio perché non finalizzato, questo tipo di prestito non prevede la sottoscrizione a contratto di un bene che possa garantire un'eventuale insolvenza, ciò rende il prodotto finanziario del prestito personale molto rischioso per la Società. Al fine di ridurre il livello di rischio associato a questa tipologia di prestiti, la Società può richiedere delle garanzie personali come la fidejussione. Il collocamento di crediti non finalizzati avviene utilizzando la rete di succursali Banca Sella, gli agenti in attività finanziaria e/o assicurativi ed attraverso il sito web della società. Il portafoglio già acquisito è altresì interessato da iniziative repeat business per i clienti rivenienti dal credito al consumo che hanno un profilo adeguato per la società.

1.5.2 Credito a PMI o Credito alle Imprese

Credito finalizzato all'acquisto di beni e servizi strumentali all'attività professionale (es. apparecchiature, prodotti e servizi per efficientamento energetico, corsi di formazione, prodotti targati); l'acquisto deve avvenire presso fornitori convenzionati con SPC; l'importo è erogato direttamente al Fornitore. Le Forme Giuridiche ammesse comprendono le Ditte individuali, le Società di persone (SNC - SAS), le Società di capitali (SRL - SPA), Cooperative, Associazioni. Non è prevista la forma tecnica di finanziamento tramite prestito personale diretto al Cliente.

1.5.3 Credito a Condominio

La Società è entrata nel mercato del credito alle famiglie su spese effettuate dai condomini in termini di credito finalizzato all'acquisto di beni specifici, ascensori, impianti di riscaldamento generali e relativa manutenzione; l'acquisto deve avvenire presso fornitori convenzionati con SPC; l'importo è erogato direttamente al Fornitore. Viene richiesta la firma dell'Amministratore di condominio, non è prevista la coobbligazione. Non è prevista la forma tecnica di finanziamento tramite prestito personale diretto.

1.5.4 Cessione del Quinto dello Stipendio, della Pensione e Deleghe di Pagamento

Particolare tipologia di prestito personale da estinguersi con cessione di quote dello stipendio o salario o pensione fino al quinto dell'ammontare dell'emolumento valutato al netto di ritenute. Questa tipologia di prestito è destinata a tutte le categorie di lavoratori dipendenti, sia dello Stato e del comparto para-statale (come specificamente previsto dal testo originale del provvedimento legislativo) sia delle aziende private (come definitivamente sancito dagli aggiornamenti previsti dalla Legge 80/05). Nell'ambito della medesima legge è stata estesa la possibilità di cedere parte della propria retribuzione anche ai pensionati di tutti gli Enti Pensionistici. Possono contrarre la richiesta anche i dipendenti delle aziende private, ma la Società si riserva la possibilità di valutare le garanzie. Le aziende vengono valutate per il capitale sociale, il numero di dipendenti e soprattutto si

valuta se, in passato, sono stati autorizzati altri contratti di cessioni ai propri dipendenti. Il prodotto è sviluppato direttamente dalla società; è collocato attraverso la propria rete, agenti in attività finanziaria e promosso dalle succursali del gruppo bancario di appartenenza e attraverso l'acquisizione dei crediti da società specializzati nel prodotto. Per quanto concerne il prestito con delega di pagamento è una particolare tipologia di prestito personale rivolta a dipendenti pubblici e privati. La rata viene addebitata direttamente in busta paga. A differenza della cessione del quinto dello stipendio il datore di lavoro può rifiutare di concedere la trattenuta e quindi impedire il perfezionamento dell'operazione.

1.5.5 Carte di Credito

Si tratta di un'operazione basata sulla presenza di una linea di credito rotativo utilizzabile attraverso una carta appartenente al circuito VISA.

Tale prodotto è distribuito direttamente dalla società ai propri clienti (mediante, canale web e canale telefonico), valutati attraverso un modello statistico andamentale e opportunamente segmentati attraverso iniziative di marketing diretto.

Dal 2016 è attivo lo strumento della carta su circuito privato: esso consiste in un affidamento utilizzabile mediante carta di credito privata che consente, nei limiti dell'importo assegnato, di effettuare acquisti con rimborso rateale di polizze assicurative e prodotti o servizi esclusivamente presso gli aderenti al circuito privato convenzionati con Sella Personal Credit S.p.A.

Tali aderenti possono essere agenti assicurativi o esercenti e-commerce.

Nel corso del 2017 inoltre è stato introdotto il prodotto Go! Credit, una linea di credito rotativa con un credito assegnato utilizzabile per mezzo di un'applicazione tramite smartphone e tablet.

1.5.6 Leasing

Detto anche locazione finanziaria, è l'operazione di finanziamento posta in essere da una banca o da un intermediario finanziario (locatore) consistente nella concessione in utilizzo per un determinato periodo di tempo e dietro il pagamento di un corrispettivo periodico (canone), di un bene acquistato o fatto costruire dal locatore da un terzo fornitore, su scelta e indicazione del Cliente (utilizzatore), che ne assume così tutti i rischi e conserva una facoltà al termine della predetta durata contrattuale di acquistare il bene ad un prezzo prestabilito ed eventualmente di prorogare il suo utilizzo a condizioni economiche predeterminate o predeterminabili. In sostanza, con l'operazione di locazione finanziaria (leasing finanziario):

- Il locatore concede all'utilizzatore l'uso del bene per un determinato periodo di tempo a fronte del pagamento di un corrispettivo periodico (canone);
- Il bene è acquistato dal locatore su scelta e indicazione dell'utilizzatore;
- L'utilizzatore ancorché non proprietario assume tutti i rischi relativi al perimento del bene acquisito in leasing;
- In favore dell'utilizzatore è prevista una opzione finale di acquisto del bene locato ad un prezzo contrattualmente predeterminato (c.d. opzione di riscatto).

Il prodotto, sviluppato direttamente dalla società, è distribuito attraverso un accordo di distribuzione con Biella Leasing S.p.A., società del Gruppo Banca Sella, per cui non si ha

più l'erogazione diretta di tale prodotto. La distribuzione del prodotto leasing è circoscritta al solo settore veicoli.

Per ciascun prodotto commercializzato la Società crea ed aggiorna le “Schede prodotto di Credito”. Le Schede Prodotto sono predisposte ed aggiornate tempo per tempo dall'Area Credito e/o dall'Ufficio Marketing e Sviluppo Prodotti (per quanto riguarda le schede prodotto della CQS). Le schede prodotto delineano le caratteristiche e non i requisiti che i crediti abbinati alla o generati dalla vendita di un determinato prodotto debbono avere per essere deliberati secondo i criteri standard definiti nel presente manuale. Ne deriva che, in percentuali non eccessive, parte della produzione può essere anche accettata ed erogata pur avendo uno o più variabili non in linea con quanto definito nella scheda prodotto, seguendo un ciclo di credito ed autorizzativo fuori standard. Particolari esigenze legate a determinati intermediari devono essere approvate dal Comitato Controllo Rischi. Inoltre, il Servizio Risk Management effettua annualmente un monitoraggio per individuare la parte di produzione fuori scheda prodotto e verificare quantitativamente e qualitativamente l'andamento.

1.6 Andamento economico

Il 2017 ha visto un ulteriore miglioramento dei risultati anche grazie al continuo e progressivo miglioramento del contesto economico, con una ripresa del mercato del lavoro e un clima di fiducia favorevole. A livello di scelte aziendali, la costante e continua attenzione rivolta verso la selezione dei canali distributivi, la progressiva riduzione del flusso dei crediti deteriorati e la razionalizzazione dei processi, hanno favorito un risultato di bilancio molto positivo sebbene la quota di mercato sia risultata sostanzialmente in linea con quella già evidenziata nel corso del 2016 e nel 2015, pari allo 0,88%.

I flussi relativi alle erogazioni hanno registrato un ulteriore incremento, pari all'11,99% in termini di volumi, mentre il numero delle operazioni ha subito nel corso del 2017 una contrazione del 19,45%, anche se meno marcata rispetto allo scostamento dell'anno precedente con il 2015. Tale contrasto conferma la scelta di revisione del processo produttivo con un mix più orientato verso importi medi più alti: l'importo medio dei prestiti finalizzati è passato da 1.807 € del 2016 a 2.155 € del 2017, così come le cessioni del quinto che hanno registrato un “ticket” medio cresciuto da 17.580 € del 2016 ai 19.290 € del 2017. Più nel dettaglio, la Società ha registrato nel periodo un leggero incremento in termini di numero di operazioni avendone erogati 107.069, rispetto ai 106.419 del 2016.

Per quanto riguarda i volumi erogati, SPC ha complessivamente concesso finanziamenti per un totale di 518,2 milioni di euro (462,7 nel corso del 2016), con un aumento dell'11,99% a fronte di un trend del mercato di + 9,6%.

Esaminando nel dettaglio i diversi settori, si evidenzia che l'attività sui prestiti personali è stata leggermente inferiore all'anno precedente (-3,75%), con un importo totale erogato di 105,5 milioni di euro e 11.722 operazioni; il settore auto (quindi auto, moto e veicoli business) ha registrato un significativo incremento, con un +15,34% sui volumi erogati (154,2 milioni di euro e 13.390 operazioni liquidate), così come il settore degli altri beni finalizzati che ha evidenziato una crescita del 16,62% (171,2 milioni di euro con 79.474 operazioni, in leggera contrazione del 2,23%).

Per quanto riguarda le carte di credito, il transato complessivo è stato pari a 40,6 milioni di euro.

Con riferimento alla cessione del quinto, SPC ha inoltre erogato 36 milioni di euro (+47,70% rispetto al 2016): tale incremento è in parte legato alle attività di acquisto dei crediti nella formula del pro-soluto e si è altresì consolidata l'attività di sviluppo commerciale anche grazie alla strutturazione di una rete agenziale costituita sia da figure specialistiche che da strutture più organizzate.

Il saldo per anno della distribuzione dei crediti netti per singolo prodotto viene riportata nella tabella 1.2:

Tabella 1.2: Distribuzione dei crediti netti verso la clientela

	31/12/2015	31/12/2016	% var 2015-16	31/12/2017	% var 2016-17
Prestiti auto	376.718.954	366.145.149	-2,8%	381.144.292	+4,1%
Prestiti finalizzati	147.486.089	137.553.938	-6,7%	164.617.369	+19,7%
Prestiti personali	219.773.679	226.549.100	+3,1%	230.212.772	+1,6%
CQS	90.060.451	98.120.487	+8,9%	104.590.245	+6,6%
Carte di credito	26.384.044	20.063.983	-24,0%	16.154.411	-19,5%
Leasing auto	4.866.166	1.303.755	-73,2%	354.595	-72,8%
Altri crediti	1.057.912	346.169	-67,3%	435.007	+25,7%
Totale	866.347.295	850.082.581	-1,9%	897.508.691	+5,6%

La voce altri crediti rappresenta i saldi attivi dei conti correnti postali.

Si rileva, tra il 2015 e il 2016, una diminuzione dei prestiti auto (-2,8%) e dei prestiti finalizzati (-6,7%), e un incremento dei prestiti personali (+3,1%). In aumento la cessione del quinto (8,9%), in calo le carte di credito (-24%), e, in maniera più evidente, il comparto leasing auto a seguito della decisione di non effettuare più erogazioni dirette del prodotto (-73,2%).

Tra il 2016 e il 2017 si rileva un incremento dei prestiti auto (+4,1%), dei prestiti finalizzati (+19,7%) e dei prestiti personali (+1,6%), in aumento anche rispetto al 2015. In aumento la cessione del quinto (+6,6%), che mantiene il trend in crescita dal 2015, in calo le carte di credito (-19,5%), anche rispetto al 2015, e, in maniera più evidente, il comparto leasing auto.

Il portafoglio dei crediti netti verso clienti è composto per la maggior parte dal prodotto prestito auto, che in media rappresenta il 43% del totale, seguito dal prestito personale (26%), prestito finalizzato (17%) e cessione del quinto (11%), mentre le carte di credito mostrano un trend in calo, passando dal 3% del 2015 all'1,80% del 2017. In figura 1.3 viene riportato l'andamento per anno della distribuzione percentuale per prodotto dei crediti netti verso clienti.

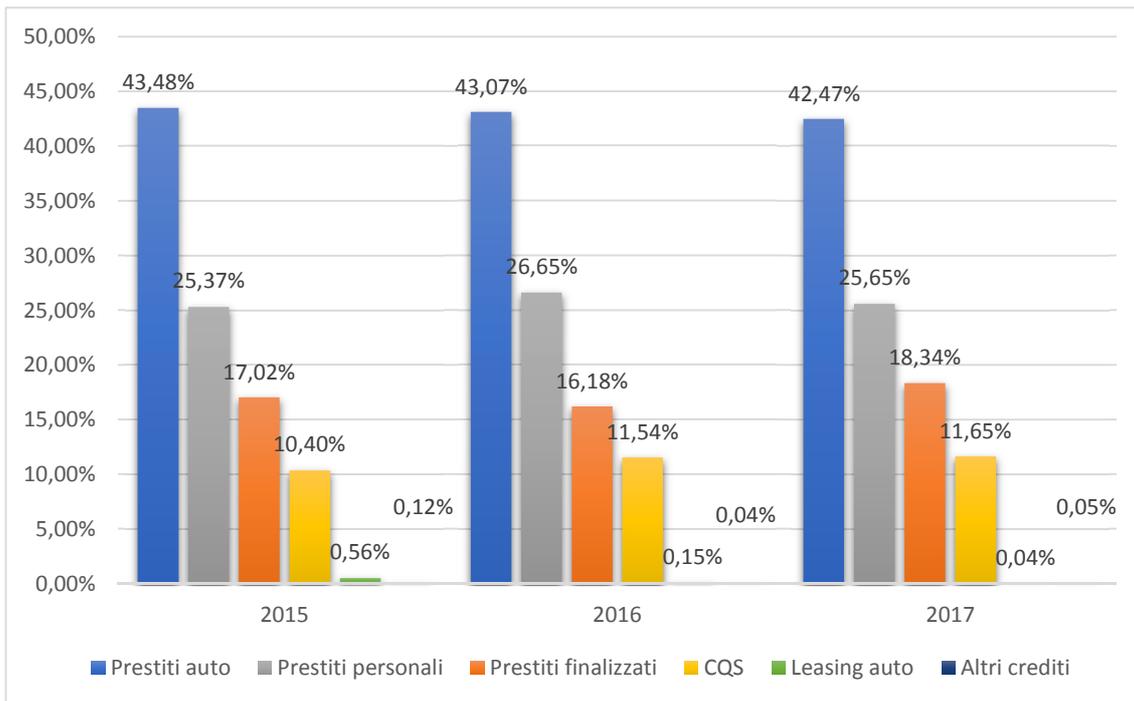


Figura 1.3: Distribuzione % dei crediti netti verso clienti

1.7 Canali di distribuzione

La rete commerciale di SPC è collocata sotto la responsabilità dell' Area Commerciale, la quale indirizza e presidia le attività di business, al fine di conseguire il raggiungimento degli obiettivi previsti in termini strategici e di budget definiti dalla società. In particolare, la rete distributiva si configura in:

- Rete Diretta: composta da risorse interne (commerciali) che seguono l'attività Captive rivolta alla Rete Agenziale dedicata ai prodotti assicurativi, con il compito di proporre i prodotti SPC alle singole agenzie e svolgono attività di sviluppo sul territorio di competenza divisi per zone (convenzionano esercizi commerciali, propongono i finanziamenti alla clientela dei convenzionati per intermediazione di questi ultimi, monitorano il regolare comportamento della rete convenzionati e presiedono alle attività formative di competenza commerciale). La rete diretta monitora, inoltre, il regolare comportamento della rete convenzionati e presidia le attività formative di competenza del commerciale. Ne fa parte altresì l'Ufficio Grandi accordi e Web. La Rete Diretta si articola in tre Aree geografiche (Nord-Est, Nord-Ovest, Centro-Sud), che gestiscono e coordinano le risorse commerciali dedicate alla promozione e vendita dei prodotti e servizi commercializzati dalla società.
- Rete Indiretta (Servizio Network Agenti): composta da agenti e mediatori creditizi iscritti agli albi competenti (Ivass, OAM e albo agenti italiani degli istituti di pagamento), la cui organizzazione, direzione e coordinamento è affidata al responsabile dell'ufficio rete network agenti. La rete indiretta si compone di:
 - Agenti in attività finanziaria: svolgono attività di intermediazione di prestiti personali, prestiti finalizzati, carte di credito e cessione del quinto;
 - Agenti assicurativi: svolgono attività rivolta alla rateizzazione dei premi delle polizze assicurative, oltre che al collocamento di prestiti personali attraverso agenzie convenzionate;

- Mediatori Creditizi, i quali svolgono attività di promozione di prodotti finanziari nei limiti di quanto previsti dalla normativa vigente.
- Commerciali dedicati alle succursali Banca Sella” (Specialist) – che seguono le attività commerciali rivenienti dall’accordo di distribuzione stipulato con Banca Sella.
- Reti Terze¹: oltre alla Rete Indiretta, fanno parte del Servizio Network Agenti anche le Reti Terze. Il responsabile della divisione Reti Terze è incaricato di curare le relazioni con i partner delle Banche del Gruppo Banca Sella (di seguito anche GBS) nello svolgimento di attività commerciali in cross selling;
 - gestire i rapporti con i soggetti delle altre Banche del GBS che promuovono e collocano i prodotti SPC dando attuazione a piani di penetrazione nel mercato;
 - collaborare, in sinergia con la Direzione Generale, all’elaborazione delle politiche di prezzo;
 - assicurarsi che i soggetti delle altre Banche del GBS abbiano conseguito un’adeguata formazione e conoscenza per poter sviluppare i prodotti commerciali SPC;
 - assicurare un costante controllo degli indici di redditività, produttività, concentrazione, relativi alla rete Banche individuando e ponendo in essere le eventuali azioni correttive e aggiornando il Responsabile Area Commerciale di SPC circa i risultati conseguiti.

La società, anche sulla base delle nuove “Disposizioni di vigilanza per gli Intermediari Finanziari”, opera controlli nel continuo sui soggetti esterni di cui si avvale per la distribuzione dei propri prodotti, curando il processo di selezione e di conferimento dei mandati, i processi organizzativi adottati per governare e gestire i relativi rischi e la relazione con visite periodiche e con un presidio costante. La società pone in essere una verifica costante della documentazione necessaria per avviare e mantenere il rapporto di collaborazione, ponendo particolare attenzione ad acquisire notizie precise e dettagliate, compresa l’eventuale attestazione circa l’iscrizione di tali soggetti ai rispettivi albi, controllando l’appartenenza ad un settore merceologico, verificando l’anzianità nel settore, la professionalità e la qualità sul mercato.

La struttura operativa, al 31/12/2017, consisteva in 7 succursali dirette, 48 agenti in attività finanziaria e circa 4.538 punti vendita convenzionati distribuiti su tutto il territorio italiano, con un andamento in continua crescita; nel 2015 si contavano 28 agenti e 3.500 punti vendita.

La distribuzione geografica delle filiali è la seguente: Lombardia (1), Veneto (2), Toscana (1), Lazio (1), Sicilia (1), Sardegna (1).

1.8 Patrimonio

Il patrimonio rappresenta una risorsa finanziaria e determina la capacità dell’impresa di operare in condizioni di solvibilità.

È costituito da risorse che sono stabilmente a disposizione della Società e che possono essere utilizzate per far fronte alla copertura di perdite generate dalla manifestazione di rischi derivanti dallo svolgimento della propria attività.

¹ Si definiscono “Reti Terze” le Banche, anche appartenenti al Gruppo Banca Sella, con cui SPC ha stipulato un accordo di collaborazione/distribuzione volto allo svolgimento di attività commerciali in cross selling.

La nozione di patrimonio è intesa nell'accezione attribuita dall'Autorità di Vigilanza distinguendo il patrimonio contabile (o patrimonio in senso stretto) dal patrimonio di vigilanza come stabilito dalla circolare 216 di Banca d'Italia. A tal fine viene rilevato trimestralmente il patrimonio di riferimento, che corrisponde al patrimonio di vigilanza (Tier 1); mensilmente sono monitorati, in collaborazione con il servizio Risk Management della Capogruppo, i dati quantitativi mensili di impiego e trimestralmente sono analizzate le informazioni desunte dalla segnalazione a Banca d'Italia relativa al rischio di credito.

1.8.1 Fondi propri

1. Capitale primario di classe 1 (Common Equity Tier 1 - CET1)
 Ai sensi della parte 2 "Fondi Propri" della Direttiva Europea 575/2013 (cd. CRR) il Patrimonio di Vigilanza di SPC consta nel solo capitale primario di classe 1 (CET1), composto dal capitale sociale, dalle riserve di utili, nonché dall'utile d'esercizio di competenza dell'anno quali elementi positivi e dalle attività immateriali quale unico elemento negativo.
2. Capitale aggiuntivo di classe 1 (Additional Tier 1 - AT1).
 Gli strumenti di capitale aggiuntivo di classe 1 e i relativi eventuali sovrapprezzi costituiscono gli elementi patrimoniali del capitale aggiuntivo di classe 1. Da tali elementi devono essere portati in deduzione gli eventuali strumenti di AT 1 propri detenuti anche indirettamente e/o sinteticamente e gli impegni al riacquisto degli stessi, nonché gli strumenti di capitale aggiuntivo, detenuti anche indirettamente e/o sinteticamente, emessi da altri soggetti del settore finanziario nei confronti dei quali si detengono o meno partecipazioni significative. Nella quantificazione degli anzidetti elementi deve tenersi conto anche degli effetti del "regime transitorio".
3. Capitale di classe 2 (Tier 2 – T2)
 Le passività subordinate le cui caratteristiche contrattuali ne consentono l'inquadramento nel T2, inclusi i relativi eventuali sovrapprezzi di emissione, costituiscono gli elementi patrimoniali del capitale di classe 2. Da tali elementi devono essere portati in deduzione le eventuali passività subordinate proprie detenute anche indirettamente e/o sinteticamente e gli impegni al riacquisto delle stesse, nonché gli strumenti di T2, detenuti anche indirettamente e/o sinteticamente, emessi da altri soggetti del settore finanziario nei confronti dei quali si detengono o meno partecipazioni significative. Nella quantificazione degli anzidetti elementi deve tenersi conto anche degli effetti del "regime transitorio".

Nella tabella 1.4 è riportato il dato quantitativo della suddivisione del capitale per classe, facendo riferimento ai bilanci consolidati degli anni 2016 e 2017.

Tabella 1.4: Ripartizione del capitale per classi

	31/12/2015	31/12/2016	31/12/2017
Common Equity Tier1-CET1	67.453.331	68.897.772	77.213.405
Additional Tier1- AT	-	-	-
Tier2-T2	-	-	-
Totale fondi propri	67.453.331	68.897.772	77.213.405

1.8.2 Adeguatezza patrimoniale

Nell'ambito dei requisiti patrimoniali per fronteggiare i rischi tipici dell'attività finanziaria (di credito, di controparte, di mercato, di cambio e operativi) la società adotta le metodologie di calcolo definite dalla Capogruppo.

A tale fine le attività di rischio e di controparte sono misurate secondo la metodologia standardizzata² avendo esposizioni per la maggior parte rientranti nel portafoglio retail. Inoltre, non avendo un portafoglio di negoziazione di vigilanza o un'operatività in cambi, non viene richiesta la misurazione dei rischi di mercato e del rischio di cambio.

Il requisito patrimoniale viene determinato come somma dei requisiti relativi alle singole tipologie di rischio.

La Società, per mezzo della Capogruppo, è dotata di processi e strumenti per determinare il livello di capitale interno adeguato (ICAAP - Internal Capital Adequacy Assessment Process) a fronteggiare ogni tipologia di rischio, anche diversi da quelli presidiati dal requisito patrimoniale complessivo, nell'ambito di una valutazione dell'esposizione, attuale e prospettica, che tenga conto delle strategie e dell'evoluzione del contesto di riferimento.

1.9 Gestione dei rischi

SPC attribuisce una forte rilevanza alla gestione e al controllo dei rischi, quali condizioni per garantire un'affidabile e sostenibile generazione di valore in un contesto di rischio controllato, proteggere la solidità finanziaria e la reputazione della società stessa e del Gruppo Bancario di cui fa parte e consentire una trasparente rappresentazione della rischiosità del proprio attivo.

Il presidio e il controllo dei rischi sono svolti dalle funzioni aziendali di controllo di secondo livello (Risk Management e Compliance) e terzo livello (Revisione Interna). In particolare, la funzione di Risk Management, in sinergia con la funzione di Risk Management della Capogruppo, ha la mission di contribuire attivamente al conseguimento di un'efficace gestione dei rischi attraverso l'identificazione, la misurazione e il controllo dei rischi di Primo e Secondo Pilastro di Basilea 2, operando in maniera strettamente legata alla gestione delle variabili economiche e patrimoniali e nel costante rispetto ed adeguamento ai cambiamenti imposti dalla normativa e allineati con le best practice di sistema.

Uno dei principi fondamentali cui la Società si ispira è l'efficiente monitoraggio/controllo del rischio credito, attività nell'ambito della quale sono stati sviluppati strumenti di controllo e gestione in linea sia con i principi contabili IAS/IFRS attualmente vigenti sia con riferimento agli standard di sistema.

1.9.1 Rischio di Credito

Il rischio di credito è definito come la possibilità che una variazione inattesa del merito creditizio di una controparte nei confronti della quale esiste un'esposizione, generi una corrispondente variazione inattesa del valore di mercato della posizione creditoria; ne deriva, pertanto, che:

² Standardised Measurement Approach (SMA) definito dal comitato di Basilea e pubblicato nel Consultative Paper (D355) del 3 giugno 2016.

- rischio di credito non significa solo possibilità di insolvenza di una controparte, in quanto anche il semplice deterioramento del merito creditizio deve considerarsi una manifestazione del rischio predetto;
- un'attenta valutazione delle controparti da affidare, seguendo principi e regole che determinano i criteri di concessione degli affidamenti, è attività imprescindibile per la corretta allocazione degli impieghi.

Il processo di gestione del rischio di credito costituisce un elemento fondamentale per garantire l'equilibrio economico e la stabilità degli intermediari finanziari. Tale processo in SPC è regolato da specifici principi che ne disciplinano le fasi di:

- istruttoria, ossia di acquisizione delle informazioni e della documentazione del cliente, nonché di valutazione del merito creditizio;
- delibera ed erogazione del credito, con verifiche di conformità e liquidazione delle quote;
- monitoraggio tramite le fasi di controllo andamentale, gestione dei crediti problematici, valutazione delle esposizioni deteriorate, criteri di classificazione delle esposizioni e misurazione del rischio, secondo le disposizioni impartite a tutto il Gruppo Banca Sella.

Nella fase di istruttoria, un primo presidio adottato per una generale mitigazione del rischio attiene la selezione dei clienti e l'istruttoria delle operazioni di credito. L'attività si focalizza sull'acquisizione della documentazione necessaria per effettuare un'adeguata valutazione del merito creditizio del cliente ed è la base per un data entry estremamente curato ed attento, finalizzato al calcolo automatico del punteggio di scoring socio-demografico per ogni operazione inserita, nonché all'interrogazione in forma automatizzata dei SIC (Sistemi di Informazioni Creditizie) e all'esame dei dati andamentali relativi agli eventuali rapporti già in essere. L'erogazione del credito è abbinata a giudizi sintetici (rating) che consentono la gestione di ciascuna pratica secondo tecniche oggettive ed uniformi. Il monitoraggio del rischio di credito post accettazione avviene attraverso tecniche e prassi consolidate ed una reportistica di dettaglio in grado di garantire l'intercettazione di eventuali andamenti anomali. Vengono inoltre prodotti, con periodicità mensile, report relativi all'andamento della delinquency (percentuale di prestiti con pagamenti insoluti all'interno del portafoglio), default static pool e recovery rate. In caso di necessità, sono intraprese specifiche azioni di mitigazione del rischio assunto sia per quanto riguarda la rischiosità della clientela in accettazione sia sui canali di provenienza piuttosto che attraverso l'applicazione di vincoli operativi (variazione cut-off delle PD di accettazione, blocco o chiusura dealer, introduzione vincoli operativi quali la tipologia di pagamento tramite addebiti SEPA SDD o restrizioni sulla tipologia di clientela finanziabile o sugli importi, ecc.). Inoltre, in linea con la normativa in vigore, i principali indicatori di rischio (Key Risk Indicator) sono monitorati attraverso il R.A.F. (Risk appetite Framework), condiviso mensilmente con il Risk Management di Capogruppo. In particolare, attraverso il R.A.F. vengono definite le soglie di Risk Appetite, Risk Capacity e Risk Tolerance per tutti i rischi principali (credito, operativo, riciclaggio, non conformità), oltre che per i rischi reputazionali, strategici e di liquidità.

1.9.1.1 Selezione dei canali di acquisizione della clientela

Nel credito al consumo la prima fase dell'attività prevede la selezione dei canali da convenzionare. L'attività di convenzionamento e selezione dei fornitori riveste importanza strategica per il corretto sviluppo della società ed è decisivo per il contenimento del rischio. Infatti, la qualità delle richieste di finanziamento è strettamente correlata alla provenienza delle operazioni: intrattenere rapporti commerciali con controparti marginali del mercato, occasionali per il flusso di lavoro, superficiali nella conduzione aziendale o addirittura in tensione di liquidità e/o con dubbio standing creditizio richiede più approfondite attività di valutazione stante il rischio sensibilmente più elevato rispetto allo standard.

Alla luce di quanto sopra, si è ritenuto opportuno regolamentare l'acquisizione dei rapporti attraverso la costituzione di un'unità organizzativa, l'ufficio convenzionati e mandati, deputata allo svolgimento delle attività di istruttoria e controllo operativo, nonché alla redazione di una dettagliata procedura volta a identificare la documentazione e l'iter idonei ad analizzare l'acquisizione dei dealer. Le attività inerenti al convenzionamento degli esercenti e la valutazione dei fornitori da cui pervengono le richieste di finanziamento sono di competenza dell'ufficio convenzioni e mandati che riporta all'Area Credito. Il Comitato Controllo Rischi, attraverso la Commissione Operativa gestisce le attività relative al rischio esercenti convenzionati/fornitori, sia per la valutazione dei nuovi rapporti sia per il monitoraggio di quelli in essere, tramite la produzione di indici di qualità, il coordinamento degli interventi di controllo, la proposta di chiusura in base alle performances e la valutazione di alert che possono essere emessi in relazione a singole posizioni canalizzate dagli esercenti. L'ufficio convenzioni e mandati ha quindi, nel processo di selezione dei canali di acquisizione della clientela, funzione operativa mentre il Comitato Controllo Rischi, nella sua funzione di monitoraggio, esercita un controllo di secondo livello. Ad ulteriore presidio è presente il Comitato Fidi, i cui componenti sono nominati dal Consiglio di Amministrazione ed a cui riporta le proprie delibere. Il Comitato monitora i rapporti attivi il cui trend di sviluppo indica la possibilità di effettuare presentazioni di richieste di affidamento complessivamente superiori a 1.000.000,00 (un milione) di euro l'anno, definisce gli eventuali interventi in funzione degli indicatori di rendimento e rischiosità di ciascun rapporto oggetto di osservazione, analizza le segnalazioni anomale sulle performance qualitative evidenziate dalla funzione Risk Management o dal Comitato Controllo Rischi e valuta i provvedimenti correttivi da questo proposti, definendo, se del caso, propri interventi. Dal 2016 SPC si avvale inoltre di una scorecard empirica di valutazione dello stato degli esercenti convenzionati e delle strategie da adottare in relazione a ciascuno dei canali attivi e produttivi, alimentata da dati interni ed esterni che valutano diverse variabili fra le quali il flusso complessivo di finanziamenti convogliato verso il sistema e il relativo andamento della rischiosità. Relativamente al credito al consumo finalizzato ed al leasing, le politiche di credito innanzi tutto partono da considerazioni relative ai beni o servizi da finanziare e dalle tipologie di distribuzione messe in atto dal dealer. Quindi, definendo a priori le tipologie di prodotti o servizi graditi e non graditi, è eseguita anche una prima selezione del rischio riveniente. Altra scelta strategica per quanto riguarda le politiche di credito è la valutazione degli affidamenti per mezzo di data entry operato dalla rete commerciale, ma verificato centralmente ed autonomamente in virtù di un sistema di acquisizione ottica di tutti i documenti relativi al contratto ed ai soggetti contraenti e di verifiche di conformità effettuate sulla base di strumenti ad elevato standard di affidabilità. Le informazioni relative alla clientela così acquisite sono trattate da sistemi

di scoring che permettono di implementare logiche statisticamente robuste, esaustive dell'universo delle informazioni disponibili, oggettive e controllabili. Lo scoring di PD andamentale, in abbinamento a caratteristiche sociodemografiche e parametri di credito prestabiliti, è utilizzato per la profilazione dei clienti già in portafoglio cui indirizzare offerte commerciali o di aumento fido (con specifico riferimento ai prodotti prestati personali e carte di credito). Tali proposte di repeat business sono ulteriormente affinate dal ricorso a modelli di scoring, di propensione e di sostenibilità appositamente sviluppati. Circa la selezione degli affidati per operazioni ex novo di prestito personale e carta di credito, siano essi rivenienti da attività dirette sia intermedie da agenti e mediatori creditizi, la società adotta le stesse politiche sopra esposte in merito agli affidamenti per operazioni di credito al consumo. In tale ambito, a supporto degli elementi di base per la valutazione del merito creditizio, la società ha predisposto uno specifico questionario dematerializzato (Colloquio di benvenuto) da sottoporre al richiedente³. Scopo del questionario è l'approfondimento della conoscenza preventiva in fase di raccolta delle informazioni propedeutiche alla valutazione, al fine di gestire la richiesta di prestito non solo in un'ottica di semplice solvibilità ma anche di sostenibilità in ragione del profilo derivato dalle informazioni assunte nella fase preliminare. Riguardo il prodotto cessione del quinto dello stipendio e delega di pagamento, stante l'obbligatorietà dell'assicurazione per il perfezionamento delle operazioni, ruolo centrale nella selezione ricoprono i criteri assuntivi ed il giudizio espresso dalle compagnie assicuratrici sull'azienda terza ceduta. Anche per tale prodotto finanziario tuttavia viene attuato un processo di valutazione della controparte con ricorso alle interrogazioni nei SIC e ad altre banche dati pubbliche coerentemente con l'approccio etico alla concessione di credito che la società adotta e persegue. Elementi caratterizzanti il prodotto cessione del quinto, per quanto attiene le policy aziendali, sono l'obbligatorietà di valutazione delle operazioni tenendo in considerazione la complessiva esposizione del cliente, sulla base delle informazioni a lui richieste e la definizione di una soglia minima di netto erogato, espressa in percentuale sul montante dell'operazione e definita in base al principio che le operazioni di cessione o delega di pagamento devono recare un concreto beneficio al cliente. Sempre in tema di cessione del quinto dello stipendio si segnala che la società ha interamente recepito nelle proprie procedure le indicazioni contenute nella comunicazione di Banca d'Italia con le "cautele e indirizzi per gli operatori" del 10 novembre 2009. In tema di fissazione dei tassi, nel caso di credito finalizzato, in considerazione della centralità dell' esercente convenzionato in tale mercato, il pricing è determinato in base ad un'offerta commisurata alle dimensioni del dealer ed alla necessità di acquisire e ritenere rapporti che assicurino elevati volumi d'attività e buona qualità del credito. Nel caso di prodotti rivolti direttamente al cliente finale, senza intermediazione dell' esercente, sono stati approntati profili di pricing basati sulla segmentazione rischio/propensione all'indebitamento. Al fine di assicurare il rispetto delle politiche di tasso aziendali è stato predisposto un controllo automatico delle condizioni concedibili, articolato per livelli di autonomia. L'autorizzazione delle condizioni finanziarie è gestita e controllata da uno specifico applicativo e dalla definizione di livelli di delibera. Ogni utente abilitato alla proposta di condizioni finanziarie può inserire a sistema le condizioni che intende attribuire all' esercente convenzionato. L'assegnazione del livello di delibera è associata alla funzione ricoperta dall'utente nella gerarchia della struttura commerciale. Nella circostanza in cui le condizioni proposte non rientrino nelle autonomie del proponente, è possibile proporre le stesse al livello gerarchico superiore per

³ Utilizzato anche in fase di controlli di [Riciclaggio](#)

approvazione. Il tasso applicato esprime il costo del finanziamento per il cliente finale; le provvigioni e le commissioni riconosciute definiscono il livello di remunerazione per l'esercente convenzionato. Sono previste soglie massime per famiglie di prodotto ai tassi proposti, superate le quali, vengono assegnati alla condizione i livelli di delibera più elevati tra quelli previsti, a seconda dell'entità dello scostamento. Per il prodotto carte di credito il tasso applicato alle carte attive può generalmente subire modifiche in conseguenza della variazione del costo della provvista. Ad ulteriore presidio per tutti i prodotti, volto al controllo dei tassi, il sistema effettua automaticamente il calcolo del TEG e il raffronto con i tassi soglia trimestralmente valorizzati in apposita tabella di sistema, aggiornata volta per volta ad ogni nuova comunicazione. L'inserimento delle nuove soglie ha efficacia retroattiva, vale a dire che tutte le operazioni non perfezionate, che eventualmente superino la nuova indicazione limite, non possono essere concluse, se non previo adeguamento del tasso applicato ai nuovi valori.

1.9.1.2 Sistemi di gestione, misurazione e controllo

Su tutte le tipologie di finanziamenti erogati da SCP (rateale classico finalizzato, prestiti personali, carte di credito, leasing, cessione del quinto e deleghe di pagamento), il primo presidio riguarda il processo di selezione dei canali e di acquisizione della clientela.

La valutazione del merito creditizio in fase di accettazione avviene con il calcolo in automatico del punteggio di scoring per ogni operazione, l'interrogazione delle principali banche dati disponibili sul mercato, l'esame dei dati andamentali relativi agli eventuali rapporti già in essere con SPC.

Le informazioni acquisite dai SIC, dalle banche dati pubbliche e dall'archivio anagrafiche interne sono parte integrante della valutazione delle richieste di finanziamento.

Lo scoring di accettazione è un algoritmo che utilizza una combinazione di informazioni disponibili al momento della richiesta di finanziamento, finalizzato al calcolo della probabilità di insolvenza del cliente (PD), ossia la probabilità che il richiedente diventi "cattivo pagatore" e cioè che raggiunga almeno 3 rate insolute nell'arco di 12 mesi.

Per ciascuna griglia di scoring sono definite opportune strategie di cut-off. Tali strategie vengono definite dalla Direzione Generale sulla base di simulazioni di scenario proposte dal Servizio Risk Management (in base al tasso di rifiuto e alla PD media della popolazione in accettazione) eseguite sempre in ottemperanza al Risk Appetite della società. Le griglie di scoring sono sotto la responsabilità della capogruppo e, prima del passaggio in produzione, sono validate sia dal punto di vista sia qualitativo (model design e metodologie utilizzate) sia quantitativo (potere discriminante, calibrazione e stabilità). Sono inoltre sottoposte a monitoraggio da parte del Risk Management della Controllante, volto a misurare l'efficacia nel tempo del sistema di scoring. In caso di funzionamento impreciso del sistema, le procedure di monitoraggio possono condurre al fine tuning oppure allo sviluppo ex novo delle griglie non più performanti a seconda dei risultati ottenuti. I valori di PD risultanti dall'elaborazione di un modello statistico vengono aggregati all'interno di fasce di rischio omogenee, determinando il rating di accettazione. In funzione del rating associato ad ogni singola operazione, la possibilità di rettificare l'esito dello scoring di accettazione (cd. override) può essere inibita o sottoposta a specifici livelli di autonomia. Le griglie di scoring tengono anche in considerazione gli indicatori di rischio previsti dal modello di credit bureau score (EURISC TOTAL VALUE) fornito da CRIF, che esprime un indicatore sintetico di rischio. Relativamente alle informazioni acquisite dai Credit Bureau, si

evidenzia che, oltre alle esposizioni rateali, è possibile rilevare anche quelle non rateali (fidi di conto e finanziamenti per anticipi) e diversi indicatori sintetici di rischio al fine di migliorare il set di informazioni.

Sono previsti inoltre i seguenti collegamenti:

- Consorzio per la tutela del credito (CTC);
- Archivio anagrafico relativo a soggetti presenti in specifiche liste, di cui: UE(Unione Europea); UST (Ministero del Tesoro degli Stati Uniti); FBI (Federal Bureau of Investigation) e archivio anagrafico relativo a Persone politicamente esposte (PEP), finalizzati alle attività relative alla normativa antiriciclaggio⁴;
- Sistema pubblico di prevenzione SCIPAFI⁵, gestito da Consap e sotto la tutela del Ministero dell'Economia e delle Finanze (MEF).

Le informazioni presenti nelle banche dati di cui sopra sono acquisite in modo automatico dal sistema informativo e si possono integrare automaticamente con i modelli di scoring, ovvero possono essere acquisite tramite interrogazione manuale. Il Servizio Risk Management ha il compito di selezionare i SIC da utilizzare nel processo di valutazione o monitoraggio delle richieste di finanziamento. Tutte le regole di credito che non rientrano nei modelli di score, sia per motivi tecnico statistici sia per motivi esperienziali (judgemental), sono gestite a livello di policy rules ed agiscono in concomitanza con il rating di accettazione. Il servizio Risk Management è la funzione aziendale deputata alla custodia e manutenzione delle matrici di esito scoring e rating di accettazione.

Il monitoraggio del portafoglio avviene su base mensile attraverso la produzione di specifici report che hanno come orizzonte temporale sia gli ultimi 12 mesi sia il cosiddetto “periodo maturo” (osservazione della performance da 9 a 24 mesi) – ma anche attraverso vintage analysis che verificano la performance attraverso un'organizzazione per coorti; si monitorano altresì gli indicatori di performance attraverso un approccio basato sugli ultimi 13 mesi per verificare i dati di produzione e il funzionamento della delibera in generale.

L'analisi è suddivisa per:

- Tipologia di canale di acquisizione della clientela: esercenti convenzionati, agenti, mediatori creditizi, succursali della società;
- Tipologia di operazione: credito al consumo, carte di credito, ecc;
- Tipologia di finalizzazione del prestito.

Viene inoltre misurato il bad rate⁶ del portafoglio di osservazione (operazioni con ritardo di pagamento maggiore o uguale a tre rate – c.d. cattivi – in rapporto al totale delle operazioni in bonis – c.d. buoni). L'indicatore così ottenuto consente di osservare l'andamento del portafoglio al fine di stimare il trend di miglioramento/peggioramento della qualità.

Unitamente al bad rate sono individuate:

- Le percentuali di operazioni respinte;
- Gli override, ossia le operazioni ove valutazioni ulteriori rispetto al giudizio del modello statistico hanno portato ad esito diverso da quello proposto dal modello stesso;
- Le operazioni no start, ovvero quelle posizioni ove l'insolvenza si è manifestata sulle prime tre scadenze di pagamento.

⁴ Si veda paragrafo [Riciclaggio](#) per maggiori dettagli

⁵ Si veda paragrafo [Frodi interne o esterne](#) per maggiori dettagli

⁶ Indicatore di rischio che considera le operazioni con almeno 3 rate insolte negli ultimi 12 mesi oppure su periodo di performance maturo.

È altresì disponibile un indicatore qualitativo determinato internamente per la valutazione degli esercenti convenzionati, il quale valuta la qualità dell'esercente, definisce l'assunzione di rischio nei confronti dello stesso in percentuale sul monte operazioni intermediato, influisce sul livello di delibera per l'attribuzione delle condizioni finanziarie e determina, in combinazione con altri elementi qualitativi, l'eventuale interruzione del rapporto.

Per quanto riguarda il presidio dei canali, è stato introdotto un modello di "Dealer Ranking" che ha come sottostante l'incrocio tra performance consuntivata degli ultimi 12 mesi e i modelli andamentali di PD e LGD⁷ sul cliente: attraverso un'aggregazione per canale, permettono di calcolare la perdita attesa del dealer stesso, che viene posta in comparazione con i dati di accantonamento previsti a budget per determinarne una classificazione.

Mensilmente si verifica altresì la redditività corretta per il rischio veicolata a ciascun dealer attraverso il calcolo del RORAC⁸ medio ponderato della produzione intermediata.

È altresì disponibile un indicatore qualitativo interno degli esercenti convenzionati che ha la funzione di segmentare il portafoglio dei dealers in classi che definiscono l'adeguatezza e la solidità del rapporto con l'intermediario. Tale indicatore ha come sottostante l'incrocio tra performance consuntivata degli ultimi 12 mesi e i modelli andamentali di PD e LGD sul cliente: attraverso un'aggregazione per canale, essi permettono di calcolare la perdita attesa del dealer stesso, che viene posta in comparazione con i dati di accantonamento previsti a budget per determinarne una classificazione. La definizione della classe di appartenenza genera riverberi sia in fase di delibera delle posizioni che in sede di possibilità di concessione di override. Inoltre, la classificazione peggiore determina l'inserimento dei dealer connotati da tale evidenza in una lista di rapporti da verificare all'interno della Commissione Operativa e del Comitato Controllo Rischi.

La classe dealer identifica quindi la qualità dell'esercente, definisce l'assunzione di rischio in percentuale sul monte operazioni intermediato, influisce sul livello di delibera per l'attribuzione delle condizioni finanziarie, determina, in combinazione con altri elementi qualitativi, l'interruzione del rapporto.

Per il monitoraggio dei canali di acquisizione della clientela è in essere la collaborazione con società specializzata per un monitoraggio costante delle strutture convenzionate. Sono stati definiti alcuni elementi di allerta quali indicatori di possibili situazioni anomale ed è stata elaborata apposita procedura per la segnalazione, registrazione e quantificazione di comportamenti anomali da parte degli stessi convenzionati.

Più in dettaglio:

- Acquisizione servizio di monitoraggio su anagrafiche degli esercenti
La società ha attivato un servizio che consente di ottenere una segnalazione qualora si verificano fatti attinenti la vita della società che siano depositati in CCIAA (Camera di Commercio, Industria e Artigianato) o emergano segnalazioni/pregiudizievoli a carico delle società convenzionate o dei loro esponenti/soci, di cui sia stata richiesta evidenza al fornitore del servizio.
Sono monitorati i rapporti attivi appartenenti a settori merceologici che per importo medio o modalità di vendita sono considerati a maggior rischio.
- Controlli di credito automatici

⁷ Loss given default o percentuale di perdita in caso di insolvenza misura il tasso di perdita su una singola linea di credito in caso di insolvenza della controparte.

⁸ Return on risk-adjusted capital, indicatore che mette in relazione, la redditività di ogni singolo impiego creditizio corretto per il rischio, con il capitale assorbito, ovvero costo del credito/rendimento al di sotto del quale le operazioni non dovrebbero essere accolte. L'entità delle perdite attese riflette tutti i fattori di rischio (credito, interesse, prezzo, operativi).

Sono presenti dei controlli di credito automatici che innalzano il livello firma qualora una richiesta di finanziamento proveniente da un esercente o da un soggetto fisico ad esso collegato sia inserita a sistema o sia in gestione al contenzioso.

- Note di irregolarità

Le comunicazioni di irregolarità consistono nella segnalazione, a cura dell'ufficio che ne viene a conoscenza, di comportamenti scorretti e di situazioni di potenziale frode da parte di convenzionati, clienti, venditori, agenti, mediatori creditizi, esattori esterni. Responsabile della raccolta delle note di irregolarità (ricezione, archiviazione, reportistica) è la funzione Risk Management. L'evidenza di una nota di irregolarità impone la gestione della stessa con risoluzione del problema evidenziato in un arco temporale ristretto pena la sospensione del rapporto. Qualora l'anomalia riscontrata sia di particolare gravità oppure immediatamente riconducibile all'esercente si procede con la sospensione immediata del rapporto. Parimenti, comportano la sospensione immediata del rapporto concentrazioni significative di segnalazioni sullo stesso canale. Al fine di monitorare il processo delle note di irregolarità è eseguito un controllo di linea accentrato volto a verificare la corretta gestione delle note, il rispetto delle tempistiche previste dalla norma tecnica che disciplina l'attività, l'escalation delle azioni definite per la mitigazione dei rischi connessi a tali segnalazioni.

Quale controllo di secondo livello, mensilmente è elaborata dalla funzione Risk Management un'analisi delle concentrazioni di rischio, al fine di consentire il monitoraggio dell'operatività dei convenzionati, agenti e mediatori creditizi.

Sono infine previsti presidi organizzativi dedicati alla valutazione preventiva ed al monitoraggio delle controparti con cui la società collabora per la collocazione dei propri prodotti: Comitato Fidi e Ufficio convenzionati e mandati.

In considerazione del core business di SPC, il rischio di credito è evidentemente il principale rischio cui la Società risulta esposta. In generale, l'operatività di SPC è quasi del tutto focalizzata su clientela di tipo retail/mass, dove il rischio in oggetto risulta differenziato e "polverizzato". Infatti, l'attivo nel suo complesso si caratterizza per un numero medio di clienti molto elevato, con esposizione media bassa e duration⁹ residua contenuta. In generale, quindi, il rischio presenta manifestazioni caratterizzate da elevata parcellizzazione.

Con periodicità mensile, il Servizio Risk Management della Società provvede ad aggiornare i valori di perdita attesa per i prodotti relativi al credito al consumo (Prestito Personale, Prestito Finalizzato e Prestito Auto) e alle carte di credito, sia per le posizioni in bonis sia per quelle deteriorate, valori che sono funzionali al calcolo delle svalutazioni forfettarie e analitiche. La perdita attesa deriva da modelli matematici per il calcolo del Tasso di Decadimento (aggiornati mensilmente) e della Loss Given Default (aggiornata annualmente).

Per quanto riguarda, invece, la stima della perdita attesa relativa ai prodotti, leasing e cessioni del quinto, essa viene desunta attraverso logiche expert based e di benchmarking, ossia sono ottenute in maniera esperienziale e sono riconducibili alle medesime percentuali di svalutazione per stato regolamentare tratte da un benchmark effettuato su competitor con le medesime tipologie di prodotti. Tali stime vengono riviste con periodicità almeno annuale

⁹ È un indicatore della durata finanziaria di un titolo, cioè della sua vita residua.

o, in alternativa, ogni volta che in applicazione vengano ravvisati scostamenti significativi tali da pregiudicare la robustezza e la coerenza delle stime stesse.

Relativamente al prodotto leasing, stante la residualità del portafoglio in essere (in base all'accordo commerciale con Sella Leasing, la Società intermedia esclusivamente le richieste di credito per questo prodotto), esso viene gestito principalmente attraverso il monitoraggio mensile dello stock dei deteriorati.

Per quanto riguarda il prodotto Cessione del quinto dello stipendio, vengono prodotte mensilmente le analisi sul portafoglio in essere contenenti:

- lo stato delle pratiche (pratiche regolari, sinistri in gestione, reclami, messe in mora, a recupero);
- percentuale di insoluto su outstanding complessivo;
- percentuale di accodato su outstanding complessivo;
- percentuale di impagato su outstanding complessivo.

1.9.1.3 Tecniche di mitigazione del rischio di credito

Alla luce di quanto sopra esposto, l'attività di erogazione del credito avviene a fronte di una valutazione iniziale dei possibili clienti. La valutazione del merito creditizio è fondata sull'effettiva capacità del debitore di far fronte agli impegni assunti sulla base della sua capacità di generare flussi finanziari adeguati al sostenimento dell'investimento in corso; oggetto di analisi è anche la verifica di un'equilibrata struttura patrimoniale.

Nel processo di erogazione e monitoraggio del credito non vengono tuttavia sottovalutate, soprattutto con riferimento alla clientela cui è associata una probabilità di insolvenza più elevata, le forme di protezione dal rischio di credito concesse dalla tipologia di forma tecnica e dalla presenza di garanzie. Le garanzie tipicamente acquisite dalle controparti sono garanzie personali e, in alcuni casi residuali, garanzie reali.

1.9.2 Rischi di Mercato

Con il termine rischio di mercato si intende la possibilità che variazioni inattese dei fattori di mercato (tassi di interesse, tassi di cambio, variazioni di prezzo dipendenti dalle fluttuazioni delle variabili di mercato e da fattori specifici degli emittenti o delle controparti) determinino una variazione al rialzo o al ribasso dei valori di mercato delle posizioni in strumenti finanziari e/o delle valute detenute da un intermediario finanziario. Con riferimento al portafoglio di negoziazione ai fini di vigilanza, il rischio di mercato comprende:

- il rischio di posizione in strumenti di debito e di capitale: rischio derivante dall'oscillazione del prezzo dei valori mobiliari per fattori attinenti all'andamento dei mercati e alla situazione della società emittente. Si scompone in rischio generico (rischio di perdite causate da un andamento sfavorevole dei prezzi/tassi della generalità degli strumenti finanziari negoziati) e in rischio specifico (rischio di perdite causate da una sfavorevole variazione del prezzo degli strumenti finanziari negoziati dovuta a fattori connessi con la situazione dell'emittente). Nel rischio di posizione è incluso il rischio base, inteso come "rischio di perdite causate da variazioni non allineate dei valori di posizioni di segno opposto, simili ma non identiche".

- il rischio di concentrazione: rischio legato all'eccessiva esposizione verso clienti o gruppi di clienti connessi.

1.9.2.1 Rischio di tasso

La Società risulta esposta al rischio di tasso di interesse sul portafoglio bancario (rischio di Secondo Pilastro), ma non al rischio di tasso di interesse sul portafoglio di negoziazione (rischio di Primo Pilastro ai sensi di Basilea 2) che, essendo incorporato nel rischio di posizione, rientra nel rischio di mercato. Più nel dettaglio, l'esposizione al rischio di tasso di interesse per la società è connesso alla natura del suo portafoglio crediti composto prevalentemente da attività a tasso fisso, a fronte di un indebitamento caratterizzato da fonti di finanziamento a tasso variabile. Le politiche di gestione di questo tipo di rischio sono volte a proteggere il margine finanziario dagli impatti delle variazioni dei tassi d'interesse e intervengono allineando il profilo di scadenza delle passività al profilo di scadenza del portafoglio attivo. A questo scopo la Società, nell'ambito della Policy di Gruppo che disciplina la gestione del rischio di tasso di interesse di banking book a livello consolidato e a livello individuale per le società di natura bancaria e finanziaria esposte a tale tipologia di rischio, utilizza contratti derivati mirati a neutralizzare i rischi di aumento dei tassi d'interesse. In accordo con quanto evidenziato nella Policy di efficacia delle coperture di Gruppo, lo strumento di copertura utilizzato è tipicamente il contratto di interest rate swap (IRS) con nominale ammortizzato in coerenza con la struttura ed il profilo temporale di ammortamento dell'oggetto della copertura (capitale erogato relativo ad una o più produzioni mensili), l'utilizzo di strumenti derivati OTC (Over the Counter) esclusivamente in una logica di copertura dell'esposizione al rischio di tasso di interesse sul portafoglio bancario derivante da asimmetrie nelle scadenze e nei tempi di ridefinizione del tasso di interesse delle attività e delle passività di bilancio. In considerazione dell'elevata numerosità delle operazioni che compongono l'attivo, la copertura avviene per masse, posizionandosi sulle fasce di scadenza degli impieghi e allineandone la copertura. A conclusione di ogni contratto, la Capogruppo rilascia una dichiarazione in cui attesta che l'IRS stipulato possiede la finalità di copertura.

La gestione del rischio di tasso di interesse di banking book, oltre che attraverso l'impiego di derivati come sopra descritto, può essere realizzata mediante operazioni di natural hedge che prevedono la sottoscrizione di finanziamenti a tasso fisso con durate comprese tra i 3 ed i 4 anni. Come previsto nella Policy di gestione del rischio di banking book del Gruppo Banca Sella, la Società oltre a essere assoggettata a misurazione da parte della Capogruppo del rischio di tasso secondo la metodologia semplificata indicata da Banca d'Italia nell'allegato C della circolare di Banca d'Italia n.285/2013, è altresì soggetta ad un monitoraggio, a soli fini gestionali, del rischio di tasso con metodologia VAR.

Il rischio di tasso è monitorato mensilmente, attraverso la valutazione del VAR calcolato come massima perdita potenziale del valore di mercato del portafoglio che potrebbe registrarsi con orizzonte temporale di 90 giorni e con un intervallo statistico di confidenza del 99%. Il modello di calcolo utilizzato per tale monitoraggio prevede la ripartizione di attivo, passivo e derivati di copertura per fasce di scadenza tassi e la conseguente valutazione del relativo livello di VAR (90 giorni, 99%). La soglia limite del VAR, stabilita dalla Capogruppo per il 2017 per la società, prevedeva un valore limite pari a 3.600.000 di euro: nel corso dell'esercizio 2017 il valore medio annuo del VAR è stato pari a Euro 2.865.000, mentre il coefficiente di rischiosità (rapporto tra esposizione al rischio tasso di

interesse e il patrimonio di vigilanza) è stato mediamente pari al 17,37% (soglia di attenzione di Banca d'Italia pari al 20%). La metodologia adottata considera tre livelli di esposizione al rischio, definiti come “basso”, “medio” e “alto” e fissati in funzione della dotazione patrimoniale della singola società. Il livello di esposizione “medio” rappresenta un segnale di allerta; in questo caso la Società deve valutare, d'intesa con la Capogruppo, se il contesto macroeconomico richieda o meno una riduzione del rischio mediante copertura. Un livello di esposizione “alto” implica invece un immediato intervento sulle posizioni di copertura. L'esposizione in termini di VAR e in termini di coefficiente di rischio è calcolata dalla Capogruppo, monitorati mensilmente e discussi nell'ambito del Comitato ALM (asset and liability management) di Gruppo. In caso di superamento dei limiti operativi imposti, vengono prontamente informati i membri del Comitato ALM di Gruppo, la Direzione della società, la Direzione della Capogruppo e vengono definite azioni mirate a ridefinire la posizione di rischio.

1.9.2.2 Rischio di cambio

La Società non detiene in portafoglio titoli azionari né obbligazionari e non assume posizioni che possano determinare rischio di cambio poiché utilizza esclusivamente l'euro quale moneta di conto e non effettua alcuna operazione finanziaria o commerciale in altre valute.

1.9.3 Rischio di prezzo

La società non è soggetta al rischio di prezzo in quanto non ha un portafoglio di negoziazione riferito ad azioni ed obbligazioni.

1.9.4 Rischi operativi

In considerazione dell'attività svolta dalla società, le tipologie di rischi operativi si concentrano principalmente sulle tipologie sotto indicate:

- a) attività di processo
- b) frodi interne e esterne
- c) legale e compliance
- d) sistemi informativi
- e) reputazione
- f) strategico
- g) outsourcing
- h) immobiliare
- i) riciclaggio

1.9.4.1 Attività di processo

Al fine di presidiare il rischio in oggetto la società ha provveduto e provvede nel continuo alla redazione ed aggiornamento delle norme interne che regolamentano l'esecuzione delle singole attività che compongono i processi aziendali, affinché ne sia garantita l'efficacia e

l'efficienza. A ciascun processo sono associati controlli di primo livello volti alla mitigazione dei rischi rivenienti dall'esecuzione delle singole attività che compongono i processi aziendali.

1.9.4.2 Frodi interne o esterne

La società ha predisposto un corpo di norme cui tutto il personale deve attenersi e pone in essere un'attività di costante osservazione su comportamenti e segnalazioni che non è di per sé riconducibile a schemi predeterminati. Dal punto di vista organizzativo e regolamentare, la società ha posto particolare attenzione all'operatività dei dipendenti massimizzando la separatezza dei ruoli nell'esecuzione delle attività e limitando l'autonomia di esecuzione delle operazioni da parte di un unico soggetto.

Per quanto riguarda l'esposizione al rischio di frodi perpetrate da soggetti esterni, clienti o esercenti convenzionati, la società è dotata, oltre che di un presidio specifico di antifrode, anche di procedure e controlli, automatici e manuali, volti a individuare e segnalare comportamenti potenzialmente fraudolenti. Inoltre, così come definito dal Decreto del Ministero dell'Economia e delle Finanze, 19 maggio 2014, n. 95, recante il Regolamento di attuazione del D.Lgs n. 64 del 2011, la società ha aderito al Sistema Centralizzato Informatico Prevenzione Furto d'Identità (di seguito SCIPAFI). SCIPAFI è il sistema pubblico di prevenzione che consente il riscontro dei dati contenuti nei principali documenti d'identità, riconoscimento e reddito, con quelli registrati nelle banche dati degli enti di riferimento, attualmente quelle dell'Agenzia delle Entrate, Ministero dell'Interno, Ministero delle Infrastrutture e dei Trasporti, INPS e INAIL. Questo riscontro si configura quindi come efficace strumento di prevenzione per i "furti d'identità" sia totali che parziali.

1.9.4.3 Compliance

SPC, ai sensi della circolare Banca d'Italia n. 288 del 3 aprile 2015 e come previsto dalla Policy di compliance del Gruppo Banca Sella, ha istituito al proprio interno una funzione responsabile incaricata della gestione del rischio di non conformità.

Le attività svolte dalla funzione di compliance della società rientrano nell'ambito del processo di Capogruppo denominato "Adeguamento normativo".

Obiettivi del processo sono:

- identificare nel corso del tempo le norme applicabili e la misurazione/valutazione del loro impatto su processi e sulle procedure aziendali;
- suggerire modifiche organizzative e procedurali finalizzate ad assicurare un adeguato presidio dei rischi di compliance identificati;
- verificare l'efficacia degli adeguamenti organizzativi (relativamente a strutture, processi, procedure, attività operative) suggeriti per la prevenzione del rischio di conformità.

Il processo di adeguamento normativo è composto dalle seguenti attività:

1. "Pianificazione";
2. "Risk assessment";
3. "Analisi di impatto";
4. "Consulenza";
5. "Reporting".

L'obiettivo è individuare e quantificare i rischi di una possibile non conformità e valutare, in un momento logico successivo, i presidi realizzati dalla struttura aziendale per mitigarne gli effetti. Le attività realizzate dalla funzione di compliance confluiscono nel "reporting", mediante il quale la funzione informa una serie di soggetti - in primo luogo l'Organo con funzione di supervisione strategica, l'Organo con funzione di gestione, l'Organo con funzione di controllo - riguardo alle risultanze della attività realizzata nel corso del periodo di riferimento.

La funzione esercita un presidio diretto riguardo le normative di maggior impatto per la società, "normative core", vale a dire le norme più rilevanti ai fini del rischio di non conformità, quali quelle che riguardano l'esercizio dell'attività di concessione del credito, la gestione dei conflitti di interesse, la trasparenza nei confronti della clientela e, più in generale, la disciplina posta a tutela del consumatore. In questo ambito la funzione della compliance è responsabile dell'intero processo di adeguamento normativo.

Relativamente alle altre norme che impattano sulla Società definite "normative non core", norme per le quali, anche in ragione delle competenze tecnico-specialistiche richieste in determinate materie, siano previste forme di presidio specializzato, sono individuati responsabili di presidio specialistico per la gestione del rischio di non conformità.

1.9.4.4 Sistemi informativi

Il vasto ricorso al trattamento elettronico dei dati ed il ruolo centrale che tale sistema svolge nel consentire la conduzione ed il monitoraggio dell'attività aziendale, comporta attenzione all'affidabilità ed all'efficienza per garantire indennità dai rischi in oggetto cui si potrebbe incorrere per errori umani o per cattivo utilizzo delle risorse e delle applicazioni.

In riferimento ai rischi di intrusione, si evidenzia che le risorse informatiche (web server) preposte alle attività legate al mondo internet, sono ubicate nella zona protetta (DMZ) presso il CED della Capogruppo, la quale provvede a fornire i servizi di security (firewall, monitoraggio del sistema) e di posta elettronica.

Per i servizi che la società ha deciso di acquisire in outsourcing, la scelta di ricorrere a società di settore specializzate di valide e ben conosciute capacità, può, di per sé, rappresentare una minore occasione di rischio per la mancata possibilità o per la maggiore difficoltà di accedere fisicamente ai programmi software ed agli impianti tecnologici installati presso la società medesima.

È da sottolineare il fatto che tutti gli scambi di informazioni con le società fornitrici di servizi in outsourcing sono regolamentate da specifiche policy ed oggetto di controlli periodici allocati presso le strutture aziendali in diretto contatto con tali società.

1.9.4.5 Reputazione

È costante impegno della società, e per essa di tutti gli organi aziendali, conservare sempre alta e vigile l'attenzione circa il mantenimento della buona immagine e reputazione.

La Direzione Generale si assume lo specifico compito di individuare le azioni più idonee per conseguire ed accrescere tali caratteristiche e di vigilare, sia ex ante avvalendosi delle funzioni preposte sia ex post con l'internal auditing, sulla corretta operatività nel rispetto delle azioni individuate.

In particolare, occorre che tutto il personale risulti indenne da qualsiasi coinvolgimento in fatti illeciti, che siano rispettate le condizioni pattuite, che i comportamenti dei singoli

addetti verso i clienti siano caratterizzati dalla dovuta gentilezza, cortesia e disponibilità, che non si commettano errori operativi e, laddove si siano commessi, si provveda ad informare la funzione competente e farsi carico di una sollecita sistemazione.

Anche la scelta delle controparti con cui la società intende stabilire rapporti è condizionata da un attento esame di sana e affidabile moralità delle stesse, essendo un principio di Gruppo la rinuncia a qualsiasi relazione ed affare con chi non sia in possesso di tali requisiti.

Va altresì citato il contributo dato dall'esame dei reclami dei clienti, all'importanza che la società attribuisce a detti reclami ed ai tempi di risoluzione.

La società è dotata di un'unità organizzativa dedicata, ufficio reclami, per la gestione delle rimostranze della clientela. L'attività dell'ufficio è disciplinata da specifica procedura che recepisce le indicazioni della Capogruppo in tema di gestione dei reclami.

1.9.4.6 Strategico

Per quanto attiene al rischio strategico ossia il rischio di erosione di utile/capitale derivante da scarsa/erronea reattività a variazioni del contesto competitivo, in considerazione della complessità del rischio in esame, si forniscono sinteticamente le linee guida strategiche cui la Società fa riferimento:

- consolidamento/crescita dimensionale al fine del raggiungimento di adeguate economie di scala;
- ottimizzazione delle risorse finanziarie impiegate;
- diversificazione dei prodotti collocati nonché dei canali distributivi;
- innovazione tecnologica ed ottimizzazione operativa per incrementare i livelli di servizio;
- ricerca di eventuali opportunità di partnership con player specializzati su determinati prodotti specifici.

Oltre ai punti sopra elencati, periodicamente sono svolti ulteriori analisi ad oggetto il rischio strategico:

- Benchmarking e analisi di marketing. È costantemente esaminato il posizionamento della Società rispetto ai competitor, principalmente attraverso l'analisi delle variabili di marketing mix. Particolare rilevanza assume anche il confronto periodico con le evidenze fornite dall'associazione di categoria di appartenenza (Assofin). Tali approfondimenti sono periodicamente sottoposti all'attenzione dell'Organo con funzione di supervisione strategica (Consiglio di Amministrazione), della Direzione Generale e dei Responsabili di tutti gli uffici;
- Financial controlling. Il processo di budgeting e la periodica verifica degli scostamenti, rispetto agli obiettivi prefissati (volumi, pricing, livelli di spesa, risk figures, ecc.) rappresentano strumenti gestionali di rilievo ai fini del presidio del rischio in esame. Il documento che ratifica e quantifica le linee guida strategiche è il business plan redatto ed approvato dal Consiglio di Amministrazione. Tali linee guida e conseguenti obiettivi operativi sono definiti in accordo con la Capogruppo normalmente definite/aggiornate una volta all'anno;
- Risk Appetite Framework. La società ha definito il proprio RAF coerentemente con gli obiettivi di rischio, il business model e gli indirizzi strategici approvati dal Consiglio di Amministrazione. Il monitoraggio degli indicatori e delle relative soglie

in esso contenuti rappresenta un valido supporto per la valutazione on going del raggiungimento dei propri obiettivi strategici.

1.9.4.7 Outsourcing

Circa il rischio in oggetto si segnala che le attività di controllo sull'operato dei fornitori ed il rispetto dei livelli di servizio concordati sono in capo ai singoli uffici aziendali owner di processo in cui intervengono tali attori esterni.

Si evidenzia, inoltre, che la Società ha stipulato con la Capogruppo un contratto quadro di collaborazione ed assistenza al fine di adempiere alle incombenze derivanti all'appartenenza al Gruppo e di ottenere specifici supporti al perseguimento delle linee guida e degli obiettivi del piano strategico di Gruppo (contratto quadro di collaborazione e assistenza tecnico-operativa, organizzativa, istituzionale).

Più in dettaglio, attraverso la collaborazione e l'ausilio dei servizi offerti, la Società intende rafforzare i propri presidi dei vari profili di rischio, principalmente di carattere operativo, ma anche di credito e di mercato, che potrebbero verificarsi nell'esercizio delle attività tipiche della società.

Gli obiettivi preposti sono di attenuare e contenere i rischi entro limiti coerenti con le strategie di sviluppo fissate dalla società, come appartenente al Gruppo e, nel contempo, assicurare il mantenimento delle condizioni di equilibrio economico, finanziario e patrimoniale della società, allo scopo di prevenire eventuali squilibri finanziari e gestionali, rafforzare le strutture e l'impostazione dei processi interni, ottenere supporto specialistico nell'espletamento degli adempimenti connessi e conseguenti ad obblighi normativi e regolamentari, ivi compresi quelli disposti dalle Autorità di vigilanza .

1.9.4.8 Immobiliare

Stante il fatto che la società è proprietaria dell'immobile ove è allocata la sede legale e operativa sussiste il rischio che questo subisca una riduzione del proprio valore, a fronte di diverse cause.

Ai fini di mitigare il rischio immobiliare la società ricorre principalmente a polizze assicurative, stipulate presso primarie compagnie di assicurazione, aderendo a convenzioni di Gruppo.

1.9.4.9 Riciclaggio di Finanziamento al Terrorismo

La società ha definito presidi organizzativi e di controllo per assicurare il rispetto della normativa in materia di prevenzione del rischio di riciclaggio e di finanziamento al terrorismo.

Circa il processo di adeguata verifica del cliente/fidejussore e dell'eventuale titolare effettivo, le procedure interne definiscono le attività che gli incaricati devono seguire al fine di garantire conformità a quanto previsto dalla normativa di riferimento.

In particolare, per le operazioni di credito personale, la società ha predisposto uno specifico questionario, denominato colloquio di benvenuto, che ha lo scopo di arricchire le informazioni di base necessarie alla valutazione del merito creditizio approfondendo la conoscenza del cliente in fase istruttoria al fine di acquisire il maggior numero di

informazioni in merito al soggetto, alla natura e alla finalità del rapporto che desidera instaurare con la società.

Sempre relativamente al processo di adeguata verifica, il sistema informativo adottato dalla società prevede controlli automatici relativi alla completezza e correttezza formale delle informazioni circa l'anagrafica e gli estremi di identificazione del cliente/fidejussore e dell'eventuale titolare effettivo. A supporto degli automatismi di sistema sono previsti anche controlli manuali.

La società è inoltre dotata di un applicativo, denominato BLAM (Black List Analysis Management), che recepisce gli elenchi (liste UE e US Treasure) di nominativi dei cosiddetti soggetti designati, ovvero le persone fisiche, le persone giuridiche, i gruppi e le entità designati come destinatari del congelamento di risorse economiche. Il data base BLAM è alimentato anche da nominativi forniti dal GBS (liste FBI e World Check, quest'ultima per l'individuazione delle persone politicamente esposte).

Il data base è integrato nel sistema informativo della società e la consultazione è contestuale all'inserimento delle richieste di finanziamento.

La Società ha sviluppato con la Capogruppo e secondo l'approccio basato sul rischio di riciclaggio, l'attribuzione alla clientela di un profilo omogeneo, basato su 4 classi, in particolare condividendo il proposito di uniformare i criteri di assegnazione delle classi di rischio MEDIO e ALTO dei profili antiriciclaggio; sulla base di tale profilatura si attiva il meccanismo di funzionamento del Comitato Accettazione Clientela.

Con riferimento agli obblighi di registrazione e di segnalazione dei dati aggregati, la società conserva i documenti e registra le informazioni che ha acquisito per assolvere gli obblighi di adeguata verifica della clientela e i dati contabili relativi alle singole operazioni effettuate dalla clientela sopra la soglia prevista da normativa, affinché possano essere utilizzati per la valutazione del rischio di riciclaggio e per qualsiasi indagine su eventuali operazioni di riciclaggio o di finanziamento del terrorismo o per corrispondenti analisi effettuate dalla UIF (Unità di Informazione Finanziaria) o da qualsiasi altra autorità competente.

La corretta tenuta dell'Archivio Unico Informatico è un presupposto essenziale per assicurare l'affidabilità dei dati elaborati in forma aggregata; il Responsabile Antiriciclaggio si avvale dell'Area Amministrazione e Fiscalità e, in particolare, dell'Ufficio Vigilanza per la tenuta dell'Archivio Unico Informatico, nonché per la puntuale gestione dei flussi di segnalazione dei dati aggregati, ricevendo tempestivamente evidenza delle eventuali anomalie riscontrate.

Per quanto attiene la segnalazione di operazioni sospette, la struttura di riferimento alla quale indirizzare la segnalazione è individuata nel Delegato Aziendale di Gruppo.

Le attività operative per l'inoltro delle segnalazioni al Delegato Aziendale di Gruppo sono formalizzate in specifica procedura, unitamente alle principali casistiche di possibili operazioni sospette al fine di agevolare gli incaricati nell'individuazione.

1.9.5 Rischio di liquidità

Il rischio di liquidità si manifesta in genere come incapacità di far fronte ai propri impegni di pagamento, causato dall'impossibilità di reperire provvista (funding liquidity risk) ovvero dalla presenza di limiti allo smobilizzo delle attività (market liquidity risk).

Tale rischio può anche manifestarsi come perdita rispetto alla realizzazione al fair value derivante dalla forzata cessione di attività (forced sale).

SPC, in quanto società facente parte del Gruppo Banca Sella è soggetta al rispetto delle linee guida di Governo del rischio di liquidità, di cui il Gruppo Banca Sella si è dotato che delineano l'insieme dei principi, delle metodologie, delle norme e dei processi necessari a prevenire l'insorgere di situazioni di crisi di liquidità e che prevedono che il Gruppo sviluppi approcci prudenziali nella gestione del rischio stesso con l'obiettivo di mantenere il profilo di esposizione al rischio su livelli estremamente contenuti.

La Policy per la gestione ed il controllo del rischio di liquidità del Gruppo Banca Sella indica i seguenti principi cardine alla base del modello di governance del rischio:

- conformità dei processi e delle metodologie di gestione e monitoraggio del rischio di liquidità con le indicazioni della vigilanza prudenziale
- condivisione delle decisioni e chiarezza delle responsabilità tra organi direttivi, di controllo ed operativi.

Il monitoraggio della dotazione di liquidità è attuato sulla crescita degli attivi ed il conseguente presidio della liquidità strutturale è volto a contenere il mismatch di scadenze che si genera fra attivo e passivo.

La gestione della liquidità aziendale è assicurata dall'ufficio tesoreria di SPC che garantisce l'utilizzo ottimale delle risorse finanziarie e opera con l'obiettivo di mantenere stabile e diversificata la struttura delle fonti finanziarie, di garantire il rispetto delle esigenze di liquidità e di minimizzare il costo della raccolta, in armonia con i principi e gli obiettivi di Gruppo.

La tipologia di rischio di liquidità prevalente nell'ambito dell'attività svolta dalla società è il funding liquidity risk, che si verifica qualora le controparti bancarie si rendono indisponibili per le usuali transazioni di raccolta, ovvero chiedano in contropartita una remunerazione significativamente superiore rispetto all'operatività svolta in condizioni ordinarie.

L'impatto di tale tipologia di rischio sull'operatività della società risulta tuttavia attenuato dalla prevalenza del funding proveniente dalla Capogruppo sul totale della raccolta, nonché da una accorta politica di progressiva diversificazione delle fonti di finanziamento che includono anche operazioni di cartolarizzazione collocate presso investitori istituzionali.

2 Rischio di credito – Regulation

2.1 Introduzione

Il seguente capitolo ha lo scopo di descrivere l'evoluzione della normativa sul rischio di credito, riportando i punti di debolezza delle disposizioni pubblicate che hanno reso necessari continui aggiornamenti e modifiche, portando alla definizione di principi che permettessero

2.1 Definizione

Il rischio di credito è definito come la possibilità che una variazione inattesa del merito creditizio di una controparte nei confronti della quale esiste un'esposizione, generi una corrispondente variazione inattesa del valore di mercato della posizione creditoria; ne deriva, pertanto, che:

- rischio di credito non significa solo possibilità di insolvenza di una controparte, in quanto anche il semplice deterioramento del merito creditizio deve considerarsi una manifestazione del rischio predetto;
- un'attenta valutazione delle controparti da affidare, seguendo principi e regole che determinano i criteri di concessione degli affidamenti, è attività imprescindibile per la corretta allocazione degli impieghi.

Tale rischio di credito non si limita alle sole posizioni iscritte a bilancio, ma comprende anche quelle fuori bilancio, ossia gli strumenti derivati.

2.2 Componenti del Rischio di credito

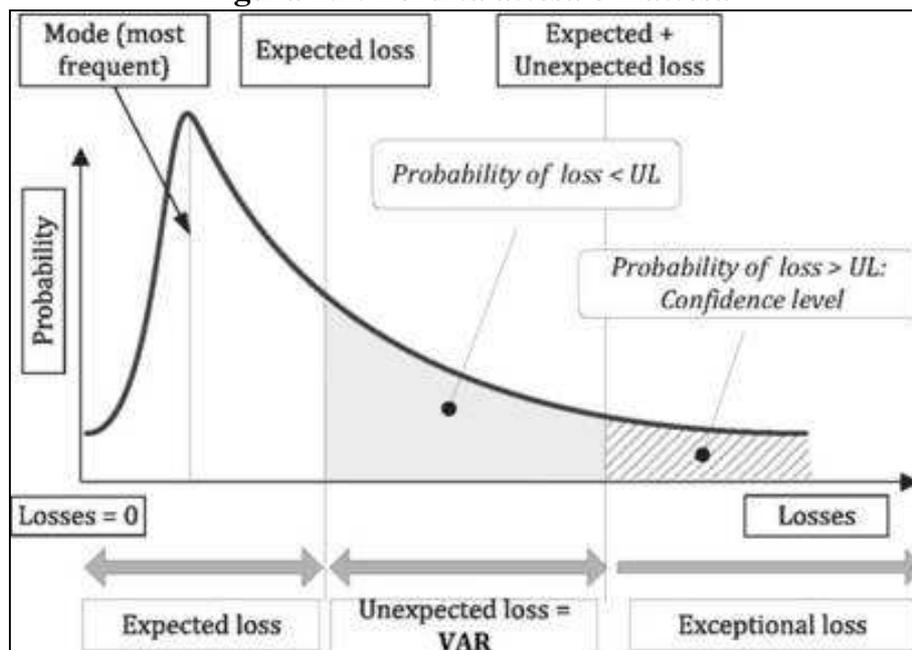
La cessione del credito espone l'istituto finanziario al rischio di subire perdite parziali o totali dell'ammontare concesso, tali perdite possono avere una duplice natura: perdite attese (Expected Loss – EL) oppure perdite inattese (Unexpected Loss – UL).

Le perdite attese costituiscono le perdite che la banca si aspetta di subire rispetto alle sue esposizioni in un determinato periodo di tempo e vanno a far parte degli accantonamenti prudenziali, in termini statistici rappresenta il valore medio della distribuzione delle perdite. Essendo quindi una componente quantificabile essa non costituisce il vero rischio di un'esposizione, ma piuttosto viene considerata come una voce di costo.

La vera e propria fonte di rischio è costituita dalle perdite inattese, che rappresentano la variabilità della perdita attesa, quindi in termini statistici la deviazione standard (o scarto quadratico medio). Tale componente può essere limitata mediante la diversificazione del portafoglio, ossia assumendo esposizioni con una bassa correlazione tra loro.

La figura 2.1 rappresenta, rispetto all'ammontare totale delle perdite, la componente attesa (EL) e quella inattesa (UL), la somma delle perdite attese e inattese viene coperta da accantonamenti e dal calcolo del VaR in base a un determinato livello di confidenza, la parte che rimane fuori sono le perdite effettive, cioè non coperte.

Figura 2.1: Perdita attesa e inattesa



2.2.1 Perdita attesa

La perdita attesa rappresenta, in termini percentuali, la perdita che in media si manifesta entro un intervallo temporale di un anno su ogni esposizione (o pool di esposizioni) esistente in portafoglio.

Le perdite attese possono essere ricavate mediante la formula seguente (2.1):

$$EL = AE \cdot PD \cdot LGD \quad (2.1)$$

Dove:

- AE (adjusted exposure) rappresenta l'esposizione al momento dell'insolvenza;
- PD (probabilità di default) rappresenta la probabilità di insolvenza della controparte;
- LGD (Loss Given Default) rappresenta il tasso di perdita in caso di insolvenza.

2.2.2 Perdita inattesa

Per il calcolo della perdita inattesa viene utilizzata la formula seguente (2.2):

$$UL = \sqrt{PD \cdot (1 - PD) \cdot LGD^2 + PD \cdot \sigma_{LGD}^2} \quad (2.2)$$

Dove σ_{LGD}^2 è la varianza del tasso di perdita.

Il presupposto su cui si basa il calcolo delle perdite inattese è che vi sia indipendenza statistica tra le componenti PD e LGD, ma l'evidenza dimostra che esista una correlazione. I fattori a favore di tale correlazione riguardano:

- ciclo economico: nei periodi di crisi economica si registrano difficoltà maggiori per le imprese, con conseguente aumento delle PD. Siccome i crediti verso altre imprese fanno parte delle attività recuperabili dalle imprese insolventi, in questi periodi di difficoltà possono essere più difficilmente incassabili.

- Aumento dei tassi di interesse: comporta una difficoltà per le imprese, riducendone il valore delle attività finanziarie e delle garanzie su di esse.
- Periodi lunghi delle procedure giudiziarie: a seconda se il periodo in cui si verifica il recupero sia migliore o peggiore rispetto al momento della stima si può avere un tasso di recupero (RR) più alto o più basso.

2.3 Tipologie di rischio di credito

Il rischio di credito si suddivide in diverse tipologie:

- a) Rischio di insolvenza: rappresenta la probabilità che la controparte di una certa posizione diventi insolvente. In questo caso la perdita percepita dal creditore corrisponde alla differenza tra l'importo del credito e quanto viene effettivamente recuperato;
- b) Rischio di migrazione (downgrading): consiste nella possibilità che si deteriori il merito creditizio di una determinata controparte, comportando una perdita di valore della posizione;
- c) Rischio di recupero: rappresenta la possibilità che il tasso di recupero (RR) di una determinata esposizione si riveli inferiore a quanto previsto, ad esempio in seguito a ritardi nei processi giudiziari;
- d) Rischio di esposizione: consiste nella possibilità che l'ammontare delle esposizioni verso una controparte aumenti in corrispondenza del verificarsi del default;
- e) Rischio di spread: riguarda la probabilità che, a parità di rating e probabilità di default, aumenti lo spread e quindi il grado di copertura richiesto;
- f) Rischio paese: consiste nel rischio di possedere esposizioni nei confronti di controparti con sede legale in paesi con profili di rischio elevati (es. cause di natura politica o legislativa);
- g) Rischio di concentrazione: consiste nel rischio dovuto a detenere portafogli con un basso grado di diversificazione;
- h) Rischio di sostituzione (o pre-regolamento): parimenti al rischio di insolvenza, ma per le esposizioni in derivati.

2.4 Disposizioni normative sul rischio di credito

All'origine della regulation degli intermediari, ogni paese stabiliva i propri criteri normativi, comportando difformità e incoerenze a livello regolamentare e disparità concorrenziali. Al fine di porre rimedio e creare uniformità normativa nei diversi Paesi, venne istituito nel 1974 il Comitato di Basilea composto dai rappresentanti delle banche centrali e dalle autorità di vigilanza. Il primo accordo si raggiunse nel 1988 passato sotto il nome di Basilea 1, con l'obiettivo di rafforzare la solvibilità e la solidità dei sistemi bancari riducendo le disparità competitive. Tale accordo riguardava solo le banche internazionali, concentrandosi sulla necessità di assicurare i requisiti minimi di capitale. Fu richiesto alle banche di detenere un capitale di vigilanza almeno dell'8% delle attività ponderate per il rischio, dove i coefficienti di ponderazione erano definiti sulla base delle tipologie di rischio insite in ciascuna attività, come segue:

- 0% per contante e valore assimilati, obbligazioni e crediti rilasciate da governi OCSE;

- 10% per i crediti verso enti pubblici;
- 20% per esposizioni verso banche multilaterali di sviluppo, crediti garantiti da titoli rilasciati da tali banche, esposizioni verso banche non facenti parte di governi OCSE con scadenza massima di un anno;
- 50% per mutui ipotecari su immobili residenziali;
- 100% per tutti gli altri crediti.

Secondo tale approccio il rischio di credito veniva calcolato come segue (2.3):

$$\text{Requisito per rischio di credito} = 8\% * RWA (\text{Risk Weighted Assets}) \quad (2.3)$$

Questo approccio però si dimostrò inefficace dal punto di vista della tutela del rischio degli investimenti delle banche.

Nacque l'esigenza di ampliare e migliorare i criteri di monitoraggio dei rischi al fine di garantire un rafforzamento delle norme sull'adeguatezza del capitale, che portarono nel giugno 2004 al rilascio della versione definitiva del Nuovo Accordo sul Capitale, noto come Basilea 2. Il sistema proposto si basa su tre pilastri:

- I. Requisiti minimi di capitale (capital requirement): identifica i requisiti patrimoniali che le banche devono detenere, andando a considerare anche il rischio operativo e calcolando il rischio di credito in maniera più complessa (2.4):

$$\text{Requisito per il rischio di credito} = \frac{\text{Capitale di vigilanza}}{\text{Rischio di credito} + \text{mercato} + \text{operativo}} \geq 8\% \quad (2.4)$$

- II. Controlli dell'autorità di vigilanza (supervisory review): definisce il processo di adeguatezza patrimoniale delle banche attraverso il controllo delle autorità di vigilanza, con la possibilità di imporre requisiti più elevati rispetto a quelli previsti;
- III. Disciplina di mercato (Market discipline): vengono imposti alle banche una serie di obblighi di informativa e di trasparenza.

Nell'ambito di tale riforma la principale novità è l'introduzione di tre approcci per la ponderazione del rischio di credito, metodo standard e metodo basato sui modelli interni, il quale a sua volta si suddivide in base e avanzato, con l'obiettivo di incoraggiare le banche a valutare e monitorare internamente i propri rischi. Tali modelli, per essere applicati, dovranno essere sottoposti all'autorizzazione delle Autorità di Vigilanza al fine di garantirne la validità della metodologia adottata.

I limiti di Basilea 2 iniziarono a emergere con la grande crisi finanziaria degli anni 2007-2009, i quali denotarono la mancanza di controllo sulla prociclicità, la mancanza di monitoraggio sull'utilizzo dei modelli interni da parte delle banche, l'incapacità di prevedere crisi sistemiche e un controllo inefficace sul rapporto di indebitamento delle banche. Tale contesto di crisi fece sì che il Comitato di Basilea, tra il luglio 2009 e il settembre 2010, pubblicasse quello che è l'attuale quadro normativo, conosciuto come Basilea 3. Va tuttavia riportato che, nel dicembre 2017 il Comitato di Basilea ha rilasciato il documento di revisione finale dei principi di Basilea 3 con il nome di Basilea 4, al fine di correggere delle misure troppo semplicistiche relative alle valutazioni dei rischi e a promuovere l'omogeneità per una migliore comparazione tra i diversi Paesi. Il nuovo quadro normativo prenderà efficacia dal 1° gennaio 2022.

2.5 La crisi finanziaria del 2007-2009

La crisi a cui si fa riferimento è quella cominciata negli Stati Uniti nel 2007 con lo scoppio della bolla immobiliare. Tale crisi trae i suoi presupposti dell'aumento in modo significativo dell'erogazione dei mutui subprime¹⁰ a partire dal 2003, ossia la concessione di mutui a clienti che, in condizioni normali, non sarebbero stati considerati meritevoli di tale credito, perché incapaci di fornire sufficienti garanzie; inoltre, la diffusione dei mutui subprime fu stimolata dalle dinamiche del mercato immobiliare statunitense e allo sviluppo delle cartolarizzazioni. La bolla immobiliare negli Stati Uniti fu generata a partire dal 2000 fino alla metà del 2006 in seguito all'aumento costante dei prezzi delle abitazioni e vide nella politica monetaria accomodante della Federal Reserve (FED) un fattore favorevole alla sua diffusione. Tale politica monetaria fu introdotta dal 2000 fino alla prima metà del 2004 e prevedeva l'applicazione di bassi tassi di interesse al fine di riavviare il ciclo economico, in risposta alla crisi dovuta alla bolla internet e all'attacco dell'11 settembre 2001. La conseguenza dei tassi di interessi bassi fu la maggiore concessione di mutui, che aumentò la domanda di abitazione e dei relativi prezzi. Gli istituti creditizi, in seguito alla concessione di un mutuo, operavano la cartolarizzazione dello stesso, ossia veniva convertito in titolo e trasferito a soggetti terzi (le società veicolo) recuperando immediatamente parte del credito. Tale processo consentiva alle banche di "cedere" il rischio di insolvenza, comportando una minore attenzione alla valutazione dell'affidabilità dei clienti e aumentando quindi la concessione di mutui. Le società veicolo a loro volta provvedevano a ricollocare tali mutui cartolarizzati agli investitori di titoli a breve termine.

Osservando una ripresa dell'economia statunitense la FED, ad inizio 2004, cominciò ad innalzare i tassi di interesse, con la conseguenza di mutui più costosi. I casi di insolvenza di mutuatari incapaci di far fronte alle rate più onerose aumentarono, mentre la domanda di immobili iniziò a contrarsi, con conseguente perdita di valore dei titoli derivati da tali mutui che divennero illiquidabili, costringendo le società veicolo a richiedere fondi alle banche che li avevano emessi e che ne avevano garantito la liquidità. Tuttavia, non tutte le banche furono in grado di far fronte a tale richiesta, perché non disponevano della liquidità necessaria e nessun istituto finanziario era disposto a fare loro credito. Tale sfiducia degli istituti di credito verso le banche sviluppò una crisi di liquidità, a seguito della quale le banche registrarono pesanti perdite per le esposizioni verso le società veicolo e verso quei soggetti che avevano investito nei titoli cartolarizzati. Alcuni tra i maggiori istituti di credito statunitense, considerati fino a quel momento *too big to fail*, evitarono il fallimento grazie all'intervento del Tesoro e della FED, mentre la banca di investimento Lehman Brothers, non ricevendo aiuti statali, il 15 settembre 2008 avviò le procedure fallimentari. Il default della Lehman Brothers innescò una serie di preoccupazioni sulla solidità delle banche d'affari, aumentando la percezione del rischio di controparte dagli operatori, determinando una contrazione della liquidità sul mercato e un aumento dei tassi a breve termine, nonostante le banche centrali avessero già provveduto a massicce iniezioni di liquidità. La crisi dei mutui subprime si diffuse in breve tempo anche nell'economia europea, provocando caduta del reddito e dell'occupazione, restrizione del credito bancario a famiglie e imprese, crollo dei mercati azionari e dei prezzi delle abitazioni, con progressivo

¹⁰ I subprime sono prestiti o mutui erogati a clienti definiti "ad alto rischio". Sono chiamati prestiti subprime perché a causa delle loro caratteristiche e del maggiore rischio a cui sottopongono il creditore sono definiti di qualità non primaria, ossia inferiore ai debiti primari (prime) che rappresentano dei prestiti erogati in favore di soggetti con una storia creditizia e delle garanzie sufficientemente affidabili. Definizione della Borsa Italiana.

deterioramento delle aspettative di famiglie e imprese riportando effetti negativi sui consumi e sugli investimenti.

La risposta del governo americano alla crisi avvenne tramite un piano di salvataggio del sistema finanziario e dei grandi istituti di credito statunitensi, prevedendo operazioni di nazionalizzazione e di acquisto di titoli privati. Durante il biennio 2007-2009, il programma di acquisto di titoli cartolarizzati Tarp (Troubled Asset Relief Program) raggiunse complessivamente 7.700 miliardi di dollari e comportò immissione di liquidità sul mercato bancario a tassi prossimi allo zero dalla FED a sostegno delle banche e compagnie di assicurazione.

La crisi ha messo in discussione la capacità di tenuta di quasi tutti i comparti della regolamentazione del sistema finanziario, da quella sui requisiti di capitale a quella sui principi contabili, per l'attitudine a creare un sistema di incentivi distorto e deresponsabilizzante. È emersa, inoltre, la necessità di rivedere l'approccio tradizionalmente improntato all'autodisciplina in alcuni settori del mercato finanziario (tra i quali quello relativo ad agenzie di rating, fondi speculativi e mercati cosiddetti over the counter), e di dettare standard più vincolanti in materia di governance delle imprese, soprattutto per ciò che riguarda le politiche di remunerazione dei manager e di gestione dei rischi. Gli eventi occorsi hanno messo in evidenza, infine, la necessità di una riforma degli assetti istituzionali della supervisione finanziaria in Europa e negli Usa.

2.6 La risposta del comitato di Basilea

Il Comitato di Basilea per la vigilanza bancaria e il suo organo direttivo, il Gruppo dei Governatori e dei Capi della vigilanza, hanno elaborato un programma di riforme che si richiama agli insegnamenti tratti dalla crisi e al mandato di riforma del settore bancario conferito dal G20 in occasione del vertice di Pittsburgh. La profondità e la gravità della crisi sono state amplificate da una serie di debolezze presenti nel settore bancario, come un'eccessiva leva finanziaria, un capitale inadeguato e di bassa qualità e insufficienti riserve di liquidità. La crisi è stata accentuata da un processo prociclico di deleveraging¹¹ e dalle interconnessioni tra istituzioni finanziarie sistemicamente rilevanti. In risposta, il Comitato ha elaborato riforme che mirano a rafforzare la capacità del sistema bancario di assorbire shock derivanti da tensioni finanziarie ed economiche, indipendentemente dalla loro origine, riducendo così il rischio di contagio dal settore finanziario all'economia reale.

Le riforme potenziano la regolamentazione microprudenziale, ossia a livello di singole banche, e contribuiscono in tal modo ad aumentare la resistenza dei singoli istituti bancari in periodi di stress. Le nuove regole hanno anche una dimensione macroprudenziale, in quanto affrontano i rischi sistemici che possono accumularsi nel settore bancario, così come l'amplificazione prociclica di tali rischi nel tempo. Entrambi gli approcci di vigilanza, micro e macroprudenziale, sono chiaramente interconnessi, poiché una migliore tenuta a livello di singole banche riduce il rischio di shock di portata sistemica.

Nel loro insieme, i nuovi standard globali per affrontare i rischi a livello di singole istituzioni e di sistema vanno sotto il nome di "Basilea 3". Gli elementi fondanti del nuovo schema, approvati e annunciati dal Comitato di Basilea e dai Governatori e Capi della vigilanza tra il luglio 2009 e il settembre 2010, sono i seguenti:

¹¹Con deleveraging si intende a riduzione della leva finanziaria, il disinvestimento che si realizza rimborsando il debito pregresso con la liquidità disponibile o vendendo asset finanziari in portafoglio per fare cassa.

- innalzamento della qualità del capitale regolamentare per aumentare la capacità delle banche di assorbire le perdite in un'ottica sia di continuità aziendale (going concern) sia di liquidazione (gone concern);
- estensione della copertura dei rischi nello schema patrimoniale, in particolare per le attività di trading, le cartolarizzazioni, le esposizioni a veicoli fuori bilancio e al rischio di controparte connesso a strumenti derivati;
- fissazione di più elevati requisiti patrimoniali minimi, in particolare mediante l'aumento dal 2 al 4,5% del livello minimo per il capitale di qualità più elevata (common equity) e l'introduzione di un capital conservation buffer del 2,5%, che porterà il requisito totale per il common equity al 7%;
- introduzione di un indice di leva finanziaria (leverage ratio) armonizzato a livello internazionale, che funge da supporto ai coefficienti patrimoniali ponderati per il rischio, al fine di contenere l'eccessivo accumulo di leva nel sistema;
- innalzamento degli standard per il processo di controllo prudenziale (secondo pilastro) e l'informativa al pubblico (terzo pilastro), insieme a ulteriori linee guida in materia di prassi di valutazione, prove di stress, gestione del rischio di liquidità, governo societario e retribuzioni;
- introduzione di due standard globali minimi di liquidità, costituiti dal liquidity coverage ratio focalizzato sul breve periodo e dal net stable funding ratio, indicatore strutturale di più lungo periodo;
- promozione dell'accantonamento di risorse patrimoniali in eccesso nelle fasi cicliche espansive cui poter attingere nei periodi di tensione, attraverso la costituzione di un capital conservation buffer e di un buffer anticiclico finalizzato a proteggere il settore bancario in periodi di crescita eccessiva del credito.

In sintesi, il nuovo schema normativo può essere racchiuso in 3 parti, dette pilastri, di cui il primo introduce un requisito patrimoniale per fronteggiare i rischi tipici dell'attività bancaria e finanziaria (di credito, di controparte, di mercato e operativi); il secondo richiede alle banche di dotarsi di una strategia e di un processo di controllo dell'adeguatezza patrimoniale, attuale e prospettica, affidando all'Autorità di vigilanza il ruolo di verifica di affidabilità e coerenza dei relativi risultati e applicare appropriate misure correttive laddove sia necessario; il terzo introduce obblighi di informativa al pubblico riguardanti l'adeguatezza patrimoniale, l'esposizione ai rischi e le caratteristiche generali dei relativi sistemi di gestione e controllo.

2.6.1 Provvedimenti Microprudenziali

Le riforme proposte dal Comitato di Basilea sono volte principalmente a rafforzare la regolamentazione prudenziale in materia di patrimonio e di liquidità, accompagnate da misure atte a migliorare le prassi di vigilanza, gestione del rischio, governo societario, trasparenza e l'informativa al pubblico.

2.6.1.1 Patrimonio

Il tema sulla qualità del patrimonio posseduto dalle banche è stato un punto cardine all'interno delle riforme, in quanto la crisi ha evidenziato definizioni incoerenti tra le varie giurisdizioni, denotando l'incapacità di poter valutare e confrontare compiutamente la

qualità del capitale nelle diverse istituzioni. L'accordo sulla definizione di patrimonio si raggiunse nel luglio 2010, prevedendo una più elevata qualità del capitale al fine di garantire l'assorbimento delle perdite e di conseguenza una maggior solidità delle banche di fronte a fasi di tensione.

Uno dei principali fattori della nuova definizione è il common equity, inteso come componente di migliore qualità del patrimonio di una banca, al quale è stato attribuito un maggior peso. Il Comitato ha prescritto che le deduzioni regolamentari vengano applicate a tale componente, anziché al patrimonio di base (TIER-1) o a quello supplementare (TIER-2). Inoltre, le perdite su crediti e le svalutazioni sono imputate agli utili non distribuiti, che fanno parte del common equity. Il requisito minimo di tale componente di capitale è stato innalzato dal 2% al 4,5%, inoltre, tenendo conto del capital conservation buffer il requisito totale per il common equity sale al 7%. Questo ha costretto le banche a ricostruire le proprie basi di common equity nel momento più difficile, in piena crisi.

In sede del Comitato di Basilea si è ritenuto che tutti gli strumenti del patrimonio di vigilanza debbano essere in grado di assorbire una perdita almeno in situazioni di liquidazione, a tal fine si è predisposto che gli strumenti del TIER-1 non compresi nel common equity e il TIER-2 siano in grado di far fronte alle perdite nel caso in cui la banca emittente raggiunga il punto di non sopravvivenza.

Il componente di capitale TIER-1 viene definito nell'art.25 del CRR come "somma del capitale primario di classe 1 e del capitale aggiuntivo di classe 1 dell'ente", quindi del Common Equity Tier 1 (CET 1) e Additional Tier, a cui è stato imposto dal Comitato una soglia minima del 6% di copertura dei Risk Weighted Assets (RWA). Le definizioni del CET 1 e Additional Tier sono contenute rispettivamente negli artt. 26 e 51 del CRR.

Per quanto riguarda il TIER-2 è invece normato all'art. 62 del CRR come "strumenti di capitale e prestiti subordinati, quando sono rispettate le condizioni di cui all'articolo 63" e previa deduzione degli elementi indicati negli artt. 66-70. La copertura fornita da TIER-1 e TIER-2 deve essere almeno dell'8% dei RWA.

2.6.1.2 Copertura dei rischi

Insieme al miglioramento della qualità e del livello del patrimonio, il nuovo schema regolamentare deve anche assicurare tutti i rischi rilevanti emersi durante la crisi. Ad esempio, alcune banche detenevano ingenti volumi di prodotti creditizi complessi e illiquidi nei propri portafogli di negoziazione, senza aver predisposto parte del capitale a copertura del rischio assunto.

Per far fronte a tali situazioni, nel luglio 2009, il Comitato di Basilea ha introdotto una serie di modifiche al fine di migliorare lo schema patrimoniale, ad esempio aumentando i requisiti minimi per le cartolarizzazioni complesse. Per riflettere meglio il rischio dei prodotti generati dalla cartolarizzazione sono state elevate le ponderazioni di rischio ad esse connesse e sono stati innalzati i requisiti patrimoniali per alcune esposizioni verso veicoli fuori bilancio. Vengono inoltre richieste alle banche analisi creditizie più rigorose delle posizioni in cartolarizzazioni con rating esterni.

All'interno del programma di riforme del Comitato è stata posta attenzione all'incremento del patrimonio di vigilanza a fronte del portafoglio di negoziazione. A tale proposito sono state sensibilmente rafforzate le norme riguardanti i requisiti patrimoniali per le esposizioni nel trading book, mediamente le banche dovranno detenere capitale aggiuntivo da tre a quattro volte superiore a quanto previsto precedentemente, tra i quali figurano un requisito

per il VAR in condizioni di stress, un incremental risk charge (IRC) per i rischi di migrazione e di insolvenza e un incremento dei requisiti per i prodotti creditizi strutturati. Altro punto di attenzione del Comitato è stato il deterioramento della qualità creditizia delle controparti, per cui è stato innalzato il requisito prudenziale per il rischio di controparte ed è stato introdotto l'impiego di input in condizioni di stress per determinare il requisito a fronte del rischio di insolvenza e nuovi standard patrimoniali per proteggere le banche contro il rischio di degrado del merito di credito.

2.6.1.3 Innalzamento del livello di capitale

La definizione più stringente di common equity, il rafforzamento dei requisiti patrimoniali a fronte delle attività di trading, del rischio di controparte e di altre operazioni connesse ai mercati dei capitali considerate nel loro insieme rappresentano un innalzamento del requisito patrimoniale minimo per assicurare che le banche siano in grado di resistere a situazioni di stress come quelle sperimentate durante l'ultima crisi. A maggior garanzia della solidità finanziaria le autorità di vigilanza potranno imporre buffer patrimoniali aggiuntivi in periodi di eccessiva espansione del credito.

Per quanto riguarda il patrimonio di base (TIER-1) sarà elevato dal 4% al 6% (senza considerare il conservation buffer).

2.6.1.4 Contenimento della leva finanziaria

Ad integrazione dei requisiti patrimoniali basati sul rischio è stato introdotto un indice di leva finanziaria non basata sul rischio (non-risk-based leverage ratio), al fine di frenare la crescita eccessiva del grado di leva del sistema, fungere da presidio aggiuntivo contro i tentativi di "aggirare" i requisiti basati sul rischio e concorrerà a limitare il rischio di modello. Questa decisione è derivata dalla segnalazione di molte banche di robusti coefficienti ponderati per il rischio per la componente patrimoniale TIER-1, mentre allo stesso tempo venivano accumulati elevati rapporti di indebitamento dovuto agli effetti degli assets off-balance sheet.

La struttura e la calibrazione dell'indice di leva sono stati definiti nel luglio 2010 dall'organo direttivo del Comitato, comprendendo le posizioni in bilancio, fuori bilancio e quelle in derivati, per questi ultimi il trattamento sarà armonizzato tra i vari regimi contabili utilizzando la definizione regolamentare di compensazione (netting).

2.6.1.5 Liquidità

L'adozione di rigorosi requisiti patrimoniali deve essere affiancata dal rafforzamento della liquidità al fine di garantire la stabilità del settore bancario, in quanto l'inadeguatezza dei relativi standard è fonte di tensioni a livello sia di singoli istituti, sia di sistema. La crisi ha comportato un esaurimento delle fonti di finanziamento che hanno continuato a scarseggiare per molto tempo, per questo motivo sono stati introdotti standard minimi di liquidità volti a gestire squilibri strutturali di lungo periodo e rafforzare la capacità di tenuta delle banche. Al fine di monitorare la liquidità delle banche sono stati introdotti due indici, liquidity coverage ratio (LCR) e il net stable funding ratio (NSFR), rispettivamente di breve e di lungo periodo. L'indice LCR richiede alle banche di detenere attività liquide di elevata qualità (High Quality Liquid Assets – HQLA) sufficienti a fronteggiare uno scenario di

stress nella raccolta dai supervisori nel periodo di un mese, mentre il NSFR è destinato a segnalare squilibri di liquidità e promuovere una maturity structure di attività e passività nel lungo periodo, ricoprendo l'intero bilancio e fornendo incentivi per le banche a utilizzare fonti di approvvigionamento stabili.

Il comitato ha inoltre sviluppato una serie di criteri comuni da utilizzare da parte delle autorità per monitorare i profili del rischio di liquidità sia a livello di singola banca sia di sistema.

2.6.1.6 Gestione e supervisione del rischio

In un contesto di rapida innovazione finanziaria il rafforzamento dei requisiti di capitale e di liquidità deve accompagnarsi a una migliore gestione e supervisione del rischio. A tal proposito il Comitato nel luglio 2009 ha riesaminato il processo di controllo prudenziale del secondo pilastro, con l'obiettivo di correggere significativi punti di debolezza emersi nella gestione del rischio delle banche durante la crisi finanziaria. Tale controllo interessa il governo societario e la gestione dei rischi a livello di impresa, la rilevazione del rischio collegato alle esposizioni fuori bilancio e alle operazioni di cartolarizzazione, la gestione delle concentrazioni di rischio, gli incentivi per una migliore gestione nel lungo periodo dei rischi e dei rendimenti da parte delle banche e prassi di remunerazione corrette.

Il Comitato ha inoltre rafforzato le direttive prudenziali nei seguenti punti chiave:

- Gestione del rischio di liquidità: nel settembre 2008 ha pubblicato il documento “Principles for Sound Liquidity Risk Management and Supervision”¹², come linee guida articolate in 17 principi per la gestione e la supervisione del rischio di liquidità. Il documento recepisce gli insegnamenti derivati dalla crisi e sono basati su best practice per la gestione del rischio di liquidità nelle istituzioni bancarie. Si evidenzia la necessità che gli organi di vigilanza valutino l'adeguatezza di tali sistemi di gestione delle banche e del livello di liquidità, suggerendo misure da adottare in caso venga riscontrata l'inadeguatezza di entrambi. Viene inoltre posto rilievo alla collaborazione tra supervisori e altre parti interessate, come le banche centrali, specialmente nei periodi di tensione.
- Prassi di valutazione: nell'aprile 2009 è stato pubblicato il documento “Supervisory guidance for assessing banks' financial instrument fair value practices”¹³ con lo scopo di migliorare l'accertamento prudenziale delle prassi di valutazione delle banche, riguardando tutte le posizioni valutate al fair value in qualsiasi momento, non solo durante fasi di stress.
- Prove di stress: nel maggio 2009 è stato pubblicato il documento “Principles for sound stress testing practices and supervision”¹⁴, dove vengono riportati i principi per la corretta governance, elaborazione e attuazione di programmi di stress testing presso le banche, al fine di rimediare alle carenze nelle prove di stress delle banche poste in evidenza dalla crisi.
- Prassi di remunerazione corrette: il documento “Compensation Principles and Standards Assessment Methodology”¹⁵ pubblicato nel gennaio 2010 contiene gli approcci di vigilanza atti a promuovere prassi di remunerazione corrette presso le

¹² Si riporta il link per il testo completo: <https://www.bis.org/publ/bcbs144.pdf>

¹³ Si riporta il link per il testo completo: <https://www.bis.org/publ/bcbs145.pdf>

¹⁴ Si riporta il link per il testo completo: <https://www.bis.org/publ/bcbs155.pdf>

¹⁵ Si riporta il link per il testo completo: <https://www.bis.org/publ/bcbs166.pdf>

banche e a favorire condizioni di parità concorrenziale. Questa metodologia si pone la finalità di aiutare i supervisori a valutare la conformità di un istituto con i criteri descritti nel documento “Principles for Sound Compensation Practices”¹⁶ del Financial Stability Board (FSB) e con i relativi standard di applicazione. Inoltre, l’FSB ha raccomandato il Comitato a elaborare un rapporto, il “Range of Methodologies for Risk and Performance Alignment of Remuneration”¹⁷ del maggio 2011, sulla gamma di metodologie per allineare gli schemi retributivi al rischio e alla performance e sulla loro efficacia alla luce dell’esperienza recente.

- Governo societario: al fine di porre rimedio alle carenze nella corporate governance bancaria, messa in rilievo durante la crisi finanziaria, il Comitato ha pubblicato, nell’ottobre 2010, una serie di principi atti a promuovere prassi corrette di governo societario nelle organizzazioni bancarie, tali principi sono contenuti nel documento “Principles for enhancing corporate governance”¹⁸. Le autorità di vigilanza assumono il compito di elaborare linee guida o norme in che prescrivano strategie, politiche e procedure di governo societario in linea con tali principi.
- Collegi di supervisori: il Comitato ha pubblicato nell’ottobre 2010 il documento “Good Principles on Supervisory Colleges”¹⁹, con l’obiettivo di porre rimedio alla difficoltà di porre in atto una supervisione consolidata da parte delle autorità di vigilanza del paese di origine e del paese ospitante nei confronti di gruppi bancari internazionali. Tali principi non hanno solo lo scopo di rafforzare la collaborazione e il coordinamento a livello microprudenziale, ma sono altresì volti a promuovere la stabilità finanziaria sul piano macroprudenziale.

2.6.1.7 Disciplina del mercato

Nel contesto della crisi è emerso come numerose banche avessero lacune e incoerenze nell’informativa in materia di esposizione al rischio e di componenti del patrimonio di vigilanza. Per provvedere a questa carenza informativa il Comitato, nel luglio 2009, ha deciso di rivedere i criteri previsti dal terzo pilastro concernenti le esposizioni a cartolarizzazione e la sponsorizzazione dei veicoli fuori bilancio. Per migliorare la trasparenza riguardo le informazioni sulle componenti del patrimonio, è stato richiesto alle banche di rendere pubbliche tutte le componenti della base patrimoniale, le deduzioni applicate e un riscontro completo di conformità con i conti finanziari. Il Comitato e l’FSB hanno elaborato una proposta sui requisiti di informativa del terzo pilastro in materia di remunerazioni volta ad assicurare che le banche forniscano informazioni chiare, esaurienti e tempestive in merito alle proprie prassi retributive, senza creare oneri eccessivi o imporre divulgazione di informazioni sensibili o riservate, con l’obiettivo più generale di rafforzare la disciplina di mercato e consentire un’adeguata valutazione da parte degli operatori così da favorire condizioni di parità concorrenziale nel settore bancario.

2.6.2 Provvedimenti Macroprudenziali

¹⁶ Si riporta il link per il testo completo: http://www.fsb.org/wp-content/uploads/r_0904b.pdf

¹⁷ Si riporta il link per il testo completo: <https://www.bis.org/publ/bcbs194.pdf>

¹⁸ Si riporta il link per il testo completo: <https://www.bis.org/publ/bcbs176.pdf>

¹⁹ Si riporta il link per il testo completo: <https://www.bis.org/publ/bcbs177.pdf>

La maggiore solidità delle singole banche ottenuto tramite un approccio a livello di singola istituzione è condizione necessaria, ma non sufficiente per un sistema bancario più robusto e per promuovere la stabilità finanziaria. Per ottenere tali risultati sono necessarie misure di più vasta portata, consentendo così di mitigare la prociclicità e rafforzare la tenuta dell'intero sistema bancario. In questo contesto rientrano le iniziative rivolte ad affrontare i rischi derivanti dalle interconnessioni tra banche globali con rilevanza sistemica²⁰, le problematiche legate alle liquidazioni bancarie nazionali e internazionali e il moral hazard²¹ associato alle istituzioni considerate troppo grandi per fallire (too big to fail). La maggiore sensibilità nei confronti dell'innovazione finanziaria e del perimetro regolamentare, unitamente alla rinnovata attenzione a favore di un processo più omogeneo e tempestivo di attenuazione e a una supervisione più rigorosa costituiranno uno strumento di salvaguardia contro i rischi derivanti dal settore non bancario.

2.6.2.1 Misure per contenere la prociclicità

L'introduzione di un indice di leva finanziaria per contenere l'accumulo di un eccessivo grado di leva nel sistema in periodi di espansione del credito e l'utilizzo di input in condizioni di stress per il calcolo del VAR e del rischio di controparte costituiscono misure in grado di ridurre la prociclicità. Per gestire l'eccessiva prociclicità derivante dai requisiti patrimoniali minimi, il Comitato ha vagliato diversi approcci, a cui si aggiunge una proposta per rendere operativa una metodologia di accantonamento basata sulle perdite attese quale contributo ai progetti di riforma dello IASB (International Accounting Standards Board) in materia.

2.6.2.2 Buffer patrimoniali

Punto cardine del nuovo schema di regolamentazione patrimoniale è la costituzione di riserve aggiuntive nelle fasi espansive cui poter attingere nei periodi di tensione. Come già introdotto nel paragrafo 1.2, il Comitato ha disposto che le banche detengano un capital conservation buffer del 2,5% dei RWA costituito da common equity. Questo strumento ha l'obiettivo di garantire una solida vigilanza e governance bancaria affrontando il problema dei comportamenti collettivi, che hanno impedito ad alcune banche di ridurre le distribuzioni di bonus discrezionali e di più alti dividendi, anche in presenza di un deterioramento della situazione patrimoniale. È stato inoltre costituito uno strumento di misura ad hoc, definito capital adequacy ratio, come indice composto da TIER-1, TIER-2 e capital conservation buffer, il quale deve essere almeno il 10,5% dei RWA. Le istituzioni finanziarie che non mantengono il buffer sopra il livello minimo devono affrontare restrizioni sui pagamenti di dividendi, riacquisti di azioni proprie e bonus.

Come estensione al capital conservation buffer è stato approvato dall'organo direttivo del Comitato un buffer anticiclico di capitale, detto countercyclical buffer, costituito anch'esso da common equity o da altri strumenti di capitale in grado di far fronte pienamente alle perdite. Tale buffer è variabile da 0% al 2,5% ed ha lo scopo di proteggere il settore

²⁰ La [direttiva UE/2013/36](#) esplicita i criteri su cui si basa la metodologia per individuare le O-SII (Other Systemically Important Institutions). In Italia tale direttiva è stata recepita nella [circolare n. 285/2013 della Banca d'Italia](#).

²¹ L'azzardo morale (moral hazard) è il rischio che può sopravvenire in una situazione ex-post alla erogazione di un finanziamento e deriva dal comportamento scorretto di un'impresa nell'utilizzare i prestiti erogati per attività più rischiose rispetto a quelle dichiarate.

bancario da fasi di eccessiva espansione del credito totale ed al contempo sarebbe possibile attingere a tale risorse quando, a giudizio delle autorità, il capitale liberato consentirebbe di assorbire le perdite che rappresentano una minaccia per la stabilità finanziaria.

2.6.2.3 Accantonamenti

Per far fronte a specifiche preoccupazioni in merito alla prociclicità il Comitato ha pubblicato, nell'agosto del 2009, una serie di principi volti a fornire assistenza allo IASB in merito agli accantonamenti e le valutazioni al fair value. In tali principi viene definito il processo delle rettifiche di valore al fine di evitare errori nella rilevazione a conto economico in contesti caratterizzati da incertezza economica. Peraltro, tali accantonamenti devono essere consistenti e fondati su metodologie solide, riflettendo il valore delle perdite attese di una banca corrispondenti alla vita residuale del suo portafoglio crediti.

2.6.2.4 Rischio sistemico e interconnessioni

L'eccessiva interconnessione tra banche sistematicamente rilevanti ha fatto sì che lo shock al sistema finanziario e all'economia si amplificasse. Il Comitato, insieme all'FSB, ha definito per tali istituti finanziari sistematicamente rilevanti delle misure aggiuntive relative alla capacità di assorbimento delle perdite quali requisiti aggiuntivi di capitale (capital surcharge), contingent capital²² e bail-in debt²³, con l'obiettivo di assicurare la continuità operativa (gone concern) e di ridurre la probabilità che le banche raggiungano il punto di non sopravvivenza e, qualora ciò avvenga, di garantire sufficienti risorse disponibili per gestire la risoluzione o la ristrutturazione delle istituzioni bancarie. Il Comitato vuole anche attenuare i rischi derivati dalle esposizioni reciproche tra singole istituzioni finanziarie e contenere il rischio sistemico e l'interconnessione attraverso diversi requisiti patrimoniali:

- Incentivi patrimoniali per le banche ad avvalersi di controparti centrali per i derivati over the counter;
- Requisiti patrimoniali più alti per le attività di trading e in derivati, nonché per cartolarizzazioni complesse e le esposizioni fuori bilancio;
- Requisiti patrimoniali più alti a fronte di esposizioni all'interno del settore finanziario;
- Introduzione di requisiti di liquidità che penalizzino l'eccessivo ricorso alla raccolta interbancaria a breve per finanziare impieghi a più lunga scadenza.

2.6.3 Applicazione degli standard

Il programma esposto fin qui vuole essere la risposta risolutiva del Comitato alla crisi finanziaria, mirando a rafforzare la solidità delle singole banche e del sistema bancario. Il Comitato si pone inoltre come obiettivi promuovere la condivisione delle problematiche di vigilanza e migliorare la qualità della supervisione bancaria a livello mondiale. Attraverso lo Supervision and Implementation Group (SIG) istituito nel gennaio 2009 il Comitato di

²² Strumenti di debito (CoCoBond) che si convertono in azioni, quindi in capitale, al verificarsi di specifici eventi, al fine di alleggerire l'esposizione debitoria della banca.

²³ Processo di assorbimento delle perdite e ricapitalizzazione dell'istituto finanziario mediante la svalutazione delle sue azioni o crediti, finanche all'azzeramento.

Basilea vuole favorire lo scambio di informazioni sugli approcci di vigilanza relativi all'adozione delle norme e dei principi elaborati. Il SIG ha sviluppato il documento "Regulatory Consistency Assessment Programme" (RCAP) allo scopo di promuovere l'omogeneità e la completezza degli standard pubblicati dal Comitato. La trasposizione delle norme di Basilea 3 nelle normative nazionali viene monitorata su base semestrale in base alle informazioni fornite dalla giurisdizione di ciascun membro, inoltre valuta la coerenza e la completezza degli standard adottati, compresa la rilevanza di eventuali scostamenti rispetto al quadro normativo. Nella tabella 2.2 seguente si riportano le fasi di applicazione delle normative elaborate.

Tabella 2.2: Fasi di applicazione

	2011	2012	2013	2014	2015	2016	2017	2018	2019
Indice di leva (leverage ratio)	Monitoraggio regolamentare		Fase di sperimentazione 1° gennaio 2013 - 1° gennaio 2017 Informativa dal 1° gennaio 2015					Migrazione al primo pilastro	
Requisito minimo per il common equity			3,5%	4,0%	4,5%	4,5%	4,5%	4,5%	4,5%
Capital conservation buffer						0,625%	1,250%	1,875%	2,5%
Requisito minimo per il common equity più capital conservation buffer			3,5%	4,0%	4,5%	5,125%	5,75%	6,375%	7,0%
Introduzione delle deduzioni dal CET 1 (compresi gli importi eccedenti il limite per DTA, MSR e investimenti in istituzioni finanziarie)				20,0%	40,0%	60,0%	80,0%	100,0%	100,0%
Requisito minimo per il patrimonio di base (tier 1)			4,5%	5,5%	6,0%	6,0%	6,0%	6,0%	6,0%
Requisito minimo per il capitale totale			8,0%	8,0%	8,0%	8,0%	8,0%	8,0%	8,0%
Requisito minimo per il capitale totale più capital conservation buffer			8,0%	8,0%	8,0%	8,625%	9,25%	9,875%	10,5%
Strumenti di capitale non più computabili nel non-core tier 1 e nel tier 2	Esclusione su un arco di 10 anni con inizio dal 2013								
Indice di copertura della liquidità (liquidity coverage ratio)	Inizio periodo di osservazione				Introduzione standard minimo				
Coefficiente dei fondi di approvvigionamento stabili (net stable funding ratio)		Inizio periodo di osservazione						Introduzione standard minimo	

2.7 Il recepimento della normativa post Basilea 3 nel quadro normativo europeo

Il 27 giugno 2013 sono stati pubblicati nella Gazzetta ufficiale dell'Unione europea (GUUE) i testi del Regolamento (UE) n. 575/2013 (CRR) e della Direttiva 2013/36/UE24 (CRD IV). Tali regolamenti vanno a sostituire integralmente la Direttiva 2006/48/CE (CRD), relativa all'accesso all'attività degli enti creditizi e al suo esercizio, e la Direttiva 2006/49/CE (CAD), relativa all'adeguatezza patrimoniale delle imprese di investimento e degli enti creditizi; inoltre, introducono nel quadro normativo dell'Unione europea i principi guida summenzionati definiti dal Comitato di Basilea per la vigilanza bancaria, noto come Basilea

²⁴ Per visionare i testi sulla Gazzetta ufficiale dell'Unione europea: <https://eur-lex.europa.eu/legal-content/IT/ALL/?uri=OJ:L:2013:176:TOC>

3, al fine di favorire un sistema bancario più solido e resistente agli shock finanziari, nonché a fornire una disciplina unica (single rulebook) e armonizzata applicabile alle istituzioni finanziarie del Mercato Unico. Primo dei tre pilastri sui quali si reggerà l'unione bancaria è il Meccanismo di vigilanza unico (Single Supervisory Mechanism - SSM), il quale attraverso il single rulebook potrà esercitare le sue funzioni di vigilanza in maniera omogenea nei paesi dell'area euro e degli altri Stati membri aderenti.

La direttiva 2013/36/UE contiene disposizioni in materia di: autorizzazione all'esercizio dell'attività bancaria, libertà di stabilimento e libera prestazione dei servizi, cooperazione fra autorità di vigilanza, processo di controllo prudenziale, metodologie per la determinazione delle riserve di capitale (buffer), disciplina delle sanzioni amministrative, regole su governo societario e remunerazioni.

Il Regolamento (UE) n. 575/2013 ha efficacia diretta sugli stati membri e contiene norme in materia di fondi propri, requisiti patrimoniali minimi, rischio di liquidità, leva finanziaria (leverage) ed informativa al pubblico.

Costituiscono parte del single rulebook europeo anche gli standard tecnici vincolanti, rappresentati da Regulatory Technical Standard (RTS) e Implementing Technical Standard (ITS), che hanno efficacia diretta sugli Stati membri e dalle linee guida dell'Autorità Bancaria Europea (European Banking Authority – EBA). All'EBA è affidato il compito di mettere in pratica la normativa primaria e definire le regole del reporting delle autorità di vigilanza.

Il Regolamento e la Direttiva prevedono ambiti di discrezionalità nazionali, al fine di permettere agli Stati membri e alle Autorità di vigilanza di tenere conto di alcune specificità nelle diverse giurisdizioni. Per quanto attiene all'Italia, Banca d'Italia ha deciso di mantenere le scelte normative presenti nelle direttive 2006/48/CE e 2006/49/CE che riguardavano tali ambiti di discrezionalità, mentre per le nuove introduzioni normative ha fatto riferimento alle best practices maturate negli anni precedenti a livello internazionale tenendo conto delle peculiarità mercato italiano nel contesto europeo.

2.7.1 Il framework normativo

In coerenza con la revisione della legislazione finanziaria internazionale sono state introdotte nuove regole nelle seguenti materie:

- a) Capitale: sono aumentati i requisiti di qualità e livello minimo regolamentare del patrimonio di vigilanza;
- b) Rischio di controparte: prevedendo l'introduzione di due strumenti: il credit valuation adjustment (CVA), che consiste in un requisito patrimoniale ad hoc per coprire le perdite derivanti dagli aggiustamenti al valore di mercato dei contratti derivati OTC a seguito delle variazioni del merito del credito; uno specifico trattamento verso le controparti centrali (Central Counterparty – CCP)²⁵;
- c) Leverage ratio: è stato introdotto l'obbligo di reporting in materia di leva finanziaria per banche e imprese di investimento (SIM), prevedendo a partire dal 1° gennaio 2015 un obbligo di disclosure del proprio coefficiente di leva finanziaria;

²⁵ La controparte centrale è il soggetto che, in una transazione, si interpone tra due contraenti evitando che questi siano esposti al rischio di inadempimento della propria controparte contrattuale e garantendo il buon fine dell'operazione.

- d) Liquidità: a partire dal 1° gennaio 2014 è iniziato il periodo di monitoraggio di un anno durante il quale banche e SIM hanno provveduto a specifici obblighi segnalatici per la rilevazione della propria esposizione al rischio di liquidità;
- e) Buffer di capitale: disposizioni summenzionate previste dal Comitato di Basilea;
- f) Obblighi di disclosure: gli artt. 89 e 90 del CRD IV disciplinano gli obblighi di disclosure con frequenza annuale in merito alle informazioni relative all'utile/perdita prima delle imposte, all'ammontare delle imposte stesse e ai contributi pubblici ricevuti;
- g) Ambiti di applicazione per le imprese di investimento: ai sensi dell'art. 4, comma 1 del CRR rientrano nel campo di applicazione di tali normative le imprese di investimento assoggettate alla disciplina della Direttiva 2004/39/CE, ad eccezione di quelle che non sono autorizzate alla prestazione del servizio accessorio di "locazione di cassette di sicurezza e amministrazione di strumenti finanziari per conto dei clienti", che prestano esclusivamente uno o più dei seguenti servizi di investimento "ricezione e trasmissione di ordini riguardanti uno o più strumenti finanziari", "esecuzione di ordini per conto dei clienti", "gestione di portafogli" e che non sono autorizzati a detenere fondi o titoli appartenenti ai loro clienti e che, per tale motivo, non possono mai trovarsi in situazione di debito con tali clienti.

2.8 Approcci metodologici

Ai sensi dell'art. 107 del CRR gli istituti creditizi al fine del calcolo degli importi delle esposizioni ponderate per il rischio possono avvalersi dell'approccio standardizzato o "se autorizzati dalle autorità competenti ai sensi dell'articolo 143, il metodo basato sui rating interni". In altre parole, alle banche viene richiesto, obbligatoriamente, di valutare la qualità creditizia delle proprie esposizioni al fine di determinare il RWA ad esse associato.

In tale contesto il ruolo di primaria importanza è rivestito dai sistemi di scoring, i quali basano la loro valutazione sul concetto di probabilità di default della controparte. Ai fini della trattazione risulta importante distinguere i concetti di scoring e di rating, in quanto non possono essere considerati termini sostitutivi; con scoring si intende un'analisi basata su informazioni quantitative pregresse elaborate mediante un modello statistico ed assegnato ad una controparte e si riferisce esclusivamente alla sua probabilità di default espresso mediante un punteggio, mentre il rating aggiunge al giudizio di scoring dati qualitativi guardando anche al futuro della controparte, dando quindi un giudizio potenzialmente differente rispetto a quello del solo score.

L'approccio standard (Standard Approach – SA) trae le sue origini dal sistema definito in seguito all'Accordo sul Capitale del 1988, la cui sensibilità al rischio è stata accresciuta nel tempo grazie ad una sempre maggiore segmentazione dei portafogli di esposizioni e all'utilizzo di rating calcolati da agenzie di credito delle esportazioni (Export Credit Agency – ECA) e da agenzie specializzate (External Credit Assessment Institution – ECAI) riconosciute dalle Autorità di vigilanza. La legislazione europea riconosce nelle imprese di medie e piccole dimensioni un ruolo chiave nell'economia, a tal proposito riconosce nei loro confronti un trattamento prudenziale di favore (art. 501 del CRR).

Il metodo basato sui rating interni, Internal Rating Based (IRB) permette alle banche di effettuare internamente ponderazioni sul rischio rispetto ai debitori. Il metodo IRB si distingue a sua volta in due tipologie, IRB di base (B-IRB) e IRB avanzato (A-IRB). La normativa fornisce le linee guida e i principi a cui devono attenersi gli intermediari in

termini di componenti di rischio, definizione di default, classi di attività e regole di ponderazione, indicando requisiti organizzativi e qualitativi da rispettare al fine di garantire la validità dei modelli di rating elaborati. Tali requisiti attengono alle funzioni di controllo di convalida interna del sistema di rating, delle caratteristiche necessarie (integrità, unicità e replicabilità), alla struttura del sistema di rating, alla determinazione dei parametri di rischio e alle prove di stress.

In entrambe le metodologie il calcolo dell'esposizione per il rischio viene effettuato mediante la formula (2.5):

$$\text{Importo dell'esposizione ponderata per il rischio} = RW \cdot \text{valore dell'esposizione} \quad (2.5)$$

Con l'unica distinzione a riguardo al coefficiente di ponderazione (RW), in quanto nell'approccio standard si utilizza quello pubblicato dalle agenzie di rating riconosciute da dall'Autorità di Vigilanza, mentre per l'approccio IRB viene calcolato internamente dall'istituto finanziario.

2.8.1 Standard Approach – SA

L'adozione del metodo standardizzato nel calcolo dei requisiti patrimoniali ponderati per il rischio richiede la suddivisione dei crediti in classi, come indicato nell'art. 112 del CRR, distinte in base alla natura della controparte, quindi dalle caratteristiche tecniche o dalle modalità di svolgimento del rapporto e consistono in:

- a) Esposizioni verso amministrazioni centrali o banche centrali;
- b) Esposizioni verso amministrazioni regionali o autorità locali;
- c) Esposizioni verso organismi del settore pubblico;
- d) Esposizioni verso banche multilaterali di sviluppo;
- e) Esposizioni verso organizzazioni internazionali;
- f) Esposizioni verso enti;
- g) Esposizioni verso imprese;
- h) Esposizioni al dettaglio;
- i) Esposizioni garantite da ipoteche su beni immobili;
- j) Esposizioni in stato di default;
- k) Esposizioni associate a un rischio particolarmente elevato;
- l) Esposizioni sotto forma di obbligazioni garantite;
- m) Elementi che rappresentano posizioni verso cartolarizzazione;
- n) Esposizioni verso enti e imprese con una valutazione del merito di credito a breve termine;
- o) Esposizioni sotto forma di quote o di azioni in organismi collettivi di investimento (OIC);
- p) Esposizioni in strumenti di capitale;
- q) Altre posizioni.

A ciascun portafoglio identificato viene applicato uno specifico coefficiente di ponderazione fornito da ECAI o ECA riconosciute dalle autorità di vigilanza sulla base di valutazioni del merito creditizio come indicato negli artt. 135 e 137 del CRR. L'elenco delle ECAI

riconosciute è pubblicato sul sito web dell’Autorità Bancaria Europea (ABE)²⁶. Le banche sono tenute a comunicare a Banca d’Italia l’elenco delle ECAI e delle ECA delle quali intendono avvalersi per le valutazioni, una volta definita la valutazione che intende utilizzare per una determinata esposizione, la banca è tenuta ad adottare tale valutazione per tutte le esposizioni appartenenti a quella classe in modo continuo e coerente nel tempo. Qualora una banca non intenda adottare i rating rilasciati da ECAI o ECA, oppure se nessuna ECAI o ECA prescelta dalla banca esprime un giudizio in merito a una determinata esposizione, tale banca deve applicare un coefficiente di ponderazione del 100%. Nella tabella 2.3 viene riportato il mapping dei rating rilasciati da Moody’s, Standard & Poor’s e Fitch.

Tabella 2.3: Mapping dei rating rilasciati da Moody’s, Standard & Poor’s e Fitch²⁷

	Moody’s	S&P	Fitch	Meaning
Investment Grade	Aaa	AAA	AAA	Prime
	Aa1	AA+	AA+	High Grade
	Aa2	AA	AA	
	Aa3	AA-	AA-	
	A1	A+	A+	Upper Medium Grade
	A2	A	A	
	A3	A-	A-	
	Baa1	BBB+	BBB+	Lower Medium Grade
	Baa2	BBB	BBB	
Baa3	BBB-	BBB-		
Junk	Ba1	BB+	BB+	Non Investment Grade Speculative
	Ba2	BB	BB	
	Ba3	BB-	BB-	
	B1	B+	B+	Highly Speculative
	B2	B	B	
	B3	B-	B-	
	Caa1	CCC+	CCC+	Substantial Risks
	Caa2	CCC	CCC	Extremely Speculative
	Caa3	CCC-	CCC-	In Default w/ Little Prospect for Recovery
	Ca	CC	CC+	
		C	CC	
			CC-	In Default
D	D	DDD		

Il calcolo degli importi delle esposizioni ponderati per il rischio viene effettuato partendo dal valore di bilancio delle attività di rischio, a cui viene moltiplicato il coefficiente ponderato per il rischio in base alla classe di appartenenza. Per le esposizioni fuori bilancio il valore dell’esposizione viene calcolato tramite l’applicazione di fattori di conversione di cui all’art. 166 del CR, in generale valgono tali indicazioni:

- a) 100% nel caso di elemento a rischio pieno;
- b) 50% nel caso di elemento a rischio medio;

²⁶ Si riporta link dell’elenco delle agenzie certificate: <https://www.esma.europa.eu/supervision/credit-rating-agencies/risk>

²⁷ Fonte: <https://www.haconsultancies.com/moodys-bahrain-credit-rating-2018/>

- c) 20% nel caso di elemento a rischio medio-basso;
- d) 0% nel caso di elemento a rischio basso.

Con riguardo ai fattori di rischio (RW) da applicare devono essere conformi agli artt. 114-134 della Sezione 2. Di seguito vengono forniti i valori di RW da applicare per classe di esposizione.

2.8.1.1 Esposizioni verso amministrazioni centrali e banche centrali

Alle esposizioni nei confronti di amministrazioni centrali e banche centrali vengono attribuiti coefficienti di ponderazione sulla base del rating corrispondente applicato dalle ECAI come riportato in tabella 2.4.

Tabella 2.4: Ponderazione applicata per esposizioni verso amministrazioni centrali o banche centrali

Classe di merito di credito	Ponderazione	Rating Standard & Poor's	Rating Moody's
1	0%	Da AAA a AA-	Da Aaa a Aa3
2	20%	Da A+ a A-	Da A1 a A3
3	50%	Da BBB+ a BBB-	Da Baa1 a Baa3
4	100%	Da BB+ a BB-	Da Ba1 a Ba3
5	100%	Da B+ a B-	Da B1 a B3
6	150%	CCC+ e inferiori	Caa1 e inferiori

I coefficienti di ponderazione per le esposizioni verso amministrazioni centrali e banche centrali di Stati membri dell'Unione Europea sono posti a zero.

2.8.1.2 Esposizioni verso amministrazioni regionali o autorità locali

A tali esposizioni viene applicato il medesimo fattore di ponderazione previsto per gli enti. La Banca d'Italia può però consentire l'utilizzo di coefficienti di ponderazione pari a quelli applicati per le amministrazioni centrali, a condizione che non vi sia nessuna significativa differenza di rischio.

2.8.1.3 Esposizioni verso organismi del settore pubblico

Per gli organismi ai quali è applicato un giudizio di rating da un'ECAI prescelta vengono trattate conformemente all'art. 120, in cui si differenziano le esposizioni con durata fino a tre mesi o superiore secondo quanto riportato rispettivamente nelle tabelle 2.4 e 2.5. Nel caso in cui ad un organismo non sia applicata nessuna classe di rating da un'ECAI prescelta, allora viene assegnata la ponderazione corrispondente alla classe di merito di credito nella quale sono classificate le esposizioni verso l'amministrazione centrale di appartenenza.

2.8.1.4 Esposizioni verso banche multilaterali di sviluppo

Anche per tali esposizioni vengono trattate come esposizioni verso enti, senza però adottare il trattamento preferenziale per le esposizioni di breve periodo. Alle banche indicate al paragrafo 2 dell'art. 117 viene applicato un fattore di ponderazione pari a 0%.

2.8.1.5 Esposizioni verso organizzazioni internazionali

Fanno parte di questa classe le esposizioni nei confronti della comunità europea, della Banca dei Regolamenti Internazionali e del Fondo Monetario Internazionale, applicando un fattore di ponderazione dello 0%.

2.8.1.6 Esposizioni verso enti

Per gli enti provvisti di rating i fattori di ponderazione applicati si distinguono in base alla durata dell'esposizione, utilizzando trattamenti più favorevoli per le esposizioni fino a tre mesi. Nella tabella 2.4 vengono riportati i fattori di ponderazione per le esposizioni fino a tre mesi, mentre nella tabella 2.5 per le esposizioni superiori a tre mesi.

Tabella 2.4: Ponderazioni per esposizioni di durata fino a tre mesi

Classe di merito di credito	1	2	3	4	5	6
Fattore di ponderazione del rischio	20%	20%	20%	50%	50%	150%

Tabella 2.5: Ponderazioni per esposizioni di durata superiore a tre mesi

Classe di merito di credito	1	2	3	4	5	6
Fattore di ponderazione del rischio	20%	50%	50%	100%	100%	150%

Per le esposizioni verso enti privi di valutazione del merito di credito fornita da un'ECAI prescelta si applicano i fattori di ponderazione relativi alla classe di merito dell'amministrazione centrale di appartenenza dell'ente secondo la tabella 2.6.

Tabella 2.6: Ponderazioni per esposizioni verso enti privi di rating

Classe di merito di credito	1	2	3	4	5	6
Fattore di ponderazione del rischio	20%	50%	100%	100%	100%	150%

2.8.1.7 Esposizioni verso imprese

Per le imprese alle quali è applicato un giudizio di rating da un'ECAI prescelta si applica il fattore di ponderazione relativo alla classe di merito di credito conformemente alla tabella 2.7.

Tabella 2.7: Ponderazioni per esposizioni verso imprese con classe di rating

Classe di merito di credito	1	2	3	4	5	6
Fattore di ponderazione del rischio	20%	50%	100%	100%	150%	150%

Alle esposizioni per cui non è disponibile una classe di rating si applica una ponderazione del 100% o la ponderazione delle esposizioni verso l'amministrazione centrale del paese nel quale l'impresa ha sede.

2.8.1.8 Esposizioni al dettaglio

Nei casi di esposizioni nei confronti di persone fisiche o di piccole o medie imprese si applica un coefficiente di ponderazione del 75%.

2.8.1.9 Esposizioni garantite da ipoteche su beni immobili

A tale classe di esposizioni viene applicato un fattore di ponderazione del 100%, ad eccezione delle parti dell'esposizione assegnate ad un'altra classe. Se le esposizioni, o una parte di esse, sono pienamente e totalmente garantite da ipoteche su immobili residenziali, anche se in forma di leasing, il fattore di ponderazione applicato è del 35%, mentre se si tratta di immobili non residenziali (uffici o locali per commercio anche sotto forma di leasing) si applica un fattore del 50%.

2.8.1.10 Esposizioni in stato di default

Alle esposizioni che rientrano in questa classe viene applicato un coefficiente di ponderazione sulla base alla rettifica applicata, per cui se tale rettifica è inferiore al 20% della parte non garantita il coefficiente applicato è 150%, se la rettifica è almeno il 20% si applica una ponderazione del 100%. Se il valore rimanente dopo la rettifica è pienamente e totalmente garantito da ipoteche su immobili residenziali o non, si attribuisce un fattore di ponderazione del 100%.

2.8.1.11 Esposizioni associate a un rischio particolarmente elevato

Il fattore di ponderazione del rischio applicato alle esposizioni di questa classe è del 150%. Le esposizioni che sono associate a un rischio particolarmente elevato sono definite da specifiche pubblicazioni da parte dell' ABE.

2.8.1.12 Esposizioni sotto forma di obbligazioni garantite

Per le esposizioni che rientrano in questa classe per le quali è disponibile una valutazione del merito creditizio di un'ECAI prescelta, seguono le indicazioni in tabella 2.8 per i coefficienti di ponderazione da adottare:

Tabella 2.8: Ponderazioni per esposizioni sotto forma di obbligazioni garantite con classe di rating

Classe di merito di credito	1	2	3	4	5	6
Fattore di ponderazione del rischio	10%	20%	20%	50%	50%	100%

Riguardo alle esposizioni per cui non è disponibile un giudizio di rating di un'ECAI prescelta si applicano le seguenti condizioni:

- a) se le esposizioni verso l'ente sono ponderate per il rischio al 20 %, all'obbligazione garantita è attribuito un fattore di ponderazione del 10 %;

- b) se le esposizioni verso l'ente sono ponderate per il rischio al 50 %, all'obbligazione garantita è attribuito un fattore di ponderazione del 20 %;
- c) se le esposizioni verso l'ente sono ponderate per il rischio al 100 %, all'obbligazione garantita è attribuito un fattore di ponderazione del 50 %;
- d) se le esposizioni verso l'ente sono ponderate per il rischio al 150 %, all'obbligazione garantita è attribuito un fattore di ponderazione del 100 %.

2.8.1.13 Elementi che rappresentano posizioni verso cartolarizzazione

Il coefficiente di ponderazione per tale classe di esposizione è riportato in tabella 2.9:

Tabella 2.9: Ponderazioni per esposizioni verso cartolarizzazione

Classe di merito di credito	1	2	3	4*	5-6
Posizioni verso la cartolarizzazione	20%	50%	100%	350%	1250%
Posizioni verso la ricartolarizzazione	40%	100%	225%	650%	1250%

*Solo per le variazioni del merito di credito diverse dalle valutazioni del merito di credito a breve termine

2.8.1.14 Esposizioni verso enti e imprese con una valutazione del merito di credito a breve termine

Alla presente classe di esposizioni viene applicato un fattore di ponderazione secondo la tabella 2.10.

Tabella 2.10: Ponderazioni per esposizioni con valutazioni del merito a breve termine

Classe di merito di credito	1	2	3	4	5	6
Fattore di ponderazione del rischio	20%	50%	100%	150%	150%	150%

2.8.1.15 Esposizioni sotto forma di quote o di azioni in organismi collettivi di investimento (OIC)

Se per le esposizioni appartenenti a questa classe è previsto un giudizio del merito di credito da parte di un'ECAI prescelta si applicano i fattori di ponderazioni della tabella 2.11, altrimenti viene applicata una ponderazione del 100%.

Tabella 2.11: Ponderazioni per esposizioni sotto forma di quote o azioni con giudizio di rating

Classe di merito di credito	1	2	3	4	5	6
Fattore di ponderazione del rischio	10%	20%	20%	50%	50%	100%

2.8.1.16 Esposizioni in strumenti di capitale

Alle esposizioni in strumenti di capitale viene assegnato un fattore di ponderazione del 100%, salvo i casi previsti dagli art. 48 paragrafo 4 e 128 che ricevono un fattore del 250%.

2.8.1.17 Altre posizioni

Il coefficiente di ponderazione applicato per le altre esposizioni è del 100%.

2.8.2 Metodo basato sui rating interni

L'alternativa all'approccio standardizzato è quella di utilizzare il metodo basato sui rating interni, in cui viene lasciato all'istituto finanziario il compito di calcolare l'RWA, previa autorizzazione delle autorità di vigilanza.

Con sistema di rating si fa riferimento "all'insieme di metodi, processi, controlli, meccanismi di raccolta dati e sistemi informativi che fungono da supporto alla valutazione del rischio di credito, all'attribuzione delle esposizioni a classi o aggregati (pool) di rating e alla stima quantitativa dei default e delle perdite per un dato tipo di esposizione", come da definizione all'art. 143 del CRR.

Va sottolineato che l'autorizzazione concessa dalle Autorità di Vigilanza non consiste nel concedere libertà all'ente di calcolare internamente i modelli di rating, ma concede all'ente l'autorizzazione all'utilizzo del modello di rating per cui ha fatto richiesta.

2.8.2.1 Applicazioni e requisiti

La distinzione tra IRB di base e IRB avanzato sta nel fatto che nel primo caso si utilizzano stime interne solo per il calcolo della PD e gli altri parametri sono presi da valori regolamentari, mentre nel secondo le banche utilizzano proprie stime, oltre che per la PD, anche di LGD, EAD e, ove richiesto, della M.

I sistemi IRB si distinguono anche sulla base del grado di autonomia affidato al modello, in base alla maggiore o minore rilevanza si possono identificare:

- Sistemi incentrati sulla componente automatica, dove non sono previsti interventi discrezionali di una persona (cd. overrides);
- Sistemi integrati con informazioni non standardizzate o comunque non considerate dal modello;
- Sistemi incentrati sulla valutazione soggettiva di una persona.

La scelta del livello di autonomia del sistema viene lasciato all'istituto che ne fa uso, tenendo in considerazione le proprie caratteristiche in termini di dimensioni, specificità operative, assetti organizzativi e segmenti di portafoglio interessati (large corporate, corporate, retail), in quanto a specifiche caratteristiche si adatta meglio una determinata tipologia di modello. Infatti, la valutazione dei crediti di minore dimensione riferiti al segmento retail prevede prevalentemente informazioni di tipo quantitativo e standardizzate, mentre i segmenti corporate e large corporate presentano informazioni di natura più complessa e non di tipo quantitativo.

Nell'applicazione di un sistema di rating l'ente deve garantire i seguenti requisiti:

1. Documentazione sulla struttura del sistema: riportando l'assetto e i dettagli operativi, la segmentazione di portafogli e i criteri di assegnazione del rating, le definizioni adottate di inadempienza e di perdita, la metodologia del modello statistico con situazioni in cui non risulta efficace e il processo di validazione interno e le persone che hanno la possibilità di modificare il rating e i limiti previsti, infine le attività di controllo sul processo (controlli di linea, validazione e revisione);

2. Completezza delle informazioni: in termini di analisi di tutti i dati utilizzati valutandone completezza, rilevanza e pertinenza.
3. Replicabilità: gli istituti devono permettere ai soggetti interessati (strutture di controllo e revisori esterni) e a Banca d'Italia di comprendere agevolmente l'assegnazione dei rating, nonché di poter ripercorrere il processo adottato al fine di testarne la capacità discriminante e la correttezza dei criteri di classificazione applicati. Al fine di garantire tale requisito è richiesto di tenere traccia delle decisioni assunte durante l'iter elaborativo.
4. Integrità del processo di attribuzione del rating: tale requisito si presenta per quei modelli in cui viene introdotto l'intervento di una persona nell'attribuzione definitiva del rating; per queste situazioni gli istituti devono adottare cautele per far sì che l'attribuzione del credito non sia condizionato da possibili interessi della persona che li effettua. Si devono quindi adottare soluzioni organizzative per cui vi sia netta separazione tra il processo di delibera e quello di attribuzione del rating.
5. Omogeneità: il sistema deve garantire che a debitori od operazioni che comportano rischi analoghi siano classificati nella medesima classe di rating. Gli istituti dovranno dunque predisporre istruzioni e linee guida in merito al trattamento delle informazioni di natura oggettiva e in materia di override, al fine di prevenire difformità di assegnazione. Tali istruzioni e linee guida devono essere periodicamente verificate al fine di valutare eventuali scostamenti dal punteggio automatico generato.
6. Univocità: ad ogni controparte verso cui si intrattiene un rapporto deve essere assegnato un unico rating. A tale scopo bisogna prevedere opportune cautele in relazione a clienti che intrattengono rapporti con diverse aree dell'istituto.

2.8.2.2 Sistema dei controlli

Al fine di assicurare che i requisiti richiesti per l'adozione dei modelli interni di rating devono essere previsti dei controlli interni che ne garantiscano l'effettiva validità in termini organizzativi, operativi e affidabilità dei sistemi informativi e procedure di report utilizzati. Le funzioni di controllo si articolano su tre livelli:

1. Controllo di primo livello: sono controlli automatici effettuati dalle strutture organizzative coinvolte nel processo di attribuzione del rating disciplinati da appositi protocolli operativi. Hanno l'obiettivo di verificare la correttezza, completezza e coerenza delle informazioni utilizzate dal modello e il corretto svolgimento del processo di attribuzione del rating. Rientrano anche le verifiche sulla pertinenza delle informazioni oggettive aggiuntive introdotte e la validità e coerenza degli override intervenuti.
2. Controlli di secondo livello: sono inerenti al processo di validazione del sistema, comprendendo strumenti e procedure volti a valutare l'accuratezza delle stime di tutte le componenti rilevanti del rischio e a valutarne il regolare funzionamento in termini di capacità predittiva e performance. La validazione non si esaurisce nella comparazione, mediante tecniche statistiche, delle misure di rischio con le stime iniziali, ma riguarda anche analisi riguardanti tutte le componenti del sistema, quali processi operativi, presidi di controllo, la documentazione, le infrastrutture informatiche e la loro complessiva coerenza. Le attività di validazione comprendono:
 - a. Processo di sviluppo del modello;

- b. Analisi di performance del sistema, calibrazione dei parametri e di benchmarking;
 - c. Prove di stress;
 - d. Verifica dei requisiti quantitativi e organizzativi relativi al sistema di rating;
 - e. Valutazione dell'adeguatezza e dell'affidabilità delle strutture informatiche e dei dati.
3. Controllo di terzo livello: tali controlli comprendono:
- a. Verifica della funzionalità del complessivo sistema di controlli al sistema di rating;
 - b. Verifica del rispetto delle norme e procedure previste;
 - c. Analisi delle attività di collegamento tra procedure operative, gestionali ed i sistemi di rating;
 - d. Revisione del processo di validazione, con lo scopo di verificarne l'adeguatezza;
 - e. Verifiche sull'effettivo utilizzo del sistema di rating.
- Tali controlli sono svolti dall'internal audit, il quale non deve essere coinvolto nella progettazione e nello sviluppo del sistema di rating, fatto salvo un ruolo di consulenza.

2.8.2.3 Stima della PD

In base all'art. 178 comma 1 del CRR “si considera intervenuto un default in relazione a un particolare debitore allorché si verificano entrambi gli eventi sotto indicati o uno di essi:

- a) l'ente giudica improbabile che, senza il ricorso ad azioni quale l'escussione delle garanzie, il debitore adempia integralmente alle sue obbligazioni creditizie verso l'ente stesso, la sua impresa madre o una delle sue filiazioni;
- b) il debitore è in arretrato da oltre 90 giorni su una obbligazione creditizia rilevante verso l'ente, la sua impresa madre o una delle sue filiazioni. Le autorità competenti possono sostituire il periodo di 90 giorni con uno di 180 giorni per le esposizioni garantite da beni immobili residenziali o da beni immobili non residenziali di PMI nella classe delle esposizioni al dettaglio, nonché per le esposizioni verso organismi del settore pubblico. Il periodo di 180 giorni non si applica ai fini dell'articolo 127.”

I modelli di rating restituiscono una PD individuale compresa tra 0% e 100%, raggruppate in classi omogenee di rischio, così imposte dalla normativa:

- almeno 7 classi per i bonis, 1 classe per i default;
- PD crescente;
- Non eccessiva concentrazione nelle classi.

Il processo di stima della PD associata ad una determinata classe di rating avviene in due stadi: nel primo l'istituto assegna ciascuna controparte ad una classe di rating in base a criteri espliciti e formalizzati; nel secondo viene determinata una PD da associare a ciascuna classe di rating e viene applicata alle controparti comprese.

Gli istituti si possono avvalere di una delle seguenti tecniche:

1. Esperienza interna di default: la stima della PD avviene mediante dati di default desunti dalla propria esperienza. Se i dati posseduti sono limitati si deve applicare un margine di cautela;
2. Mapping con dati esterni: la classificazione delle esposizioni avviene sulla base di scale di rating impiegate da ECAI e si applicano i relativi tassi di default.

3. Modelli statistici: la stima della PD avviene mediante l'utilizzo di appropriati modelli statistici, quali ad esempio regressione lineare, discriminante e logistico.

I dati utilizzati per la stima della PD possono avere natura interna, esterna, condivisa o una loro combinazione, sulla base di un orizzonte temporale di almeno due anni per il metodo base, che aumenta ogni anno fino ad arrivare ad almeno cinque anni, mentre per il metodo avanzato l'orizzonte minimo previsto è di almeno cinque anni.

Sono imposti i seguenti limiti normativi alla PD di classe applicabile:

- Amministrazioni centrali e banche centrali: nessun floor;
- Intermediari vigilati, imprese, esposizioni al dettaglio: 0,03%;
- Strumenti di capitale: 0,09% - 0,4% - 1,25% a seconda della tipologia di strumento di capitale.

La stima della PD avviene per ciascuna classe sulla base delle medie di lungo periodo dei tassi di default annui osservati.

2.8.2.4 Stima della Loss Given Default (LGD)

L'LGD rappresenta la frazione dell'esposizione che l'istituto non recupera a seguito del default della controparte. Si può esprimere in formula come (2.6):

$$LGD = 1 - RR \quad (2.6)$$

Dove RR (Recovery Rate) rappresenta invece il tasso di recupero.

Sull'LGD incidono diversi fattori, come ad esempio le caratteristiche tecniche del finanziamento, le caratteristiche del soggetto finanziato, i fattori interni all'istituto e fattori esterni di tipo macroeconomico. Come per la PD anche l'LGD può essere ottenuto in due modi: i) utilizzare valori regolamentari; ii) stimarlo internamente.

È possibile stimare l'LGD mediante la formula (2.7):

$$LGD = 1 - \frac{\sum_{t=1}^n \frac{(ER_t - AC_t)}{(1+i)^t}}{EAD} \quad (2.7)$$

Dove:

- ER (expected recovery) rappresenta la stima dell'importo che si pensa di poter recuperare;
- AC (administrative costs) rappresentano i costi sostenuti per il processo di recupero;
- i è il tasso di attualizzazione:
 - tasso di interesse di trasferimento di fondi (costo del funding della banca);
 - tasso contrattuale del finanziamento andato in default;
 - tasso congruo per il rischio, tenuto conto dei rischi su ER;
 - tasso risk-free: nell'ipotesi in cui si assuma che il rischio e il premio per l'avversione ad esso siano già inclusi in altri parametri;
- t rappresenta il tempo espresso in anno, o frazione di esso, per il recupero del credito;
- n è il periodo stimato per realizzare il recupero.

La stima della LGD risulta però problematica, in quanto si hanno dati poco numerosi ed i valori medi dei RR sono volatili e poco significativi, comportando il rischio che l'effettivo recupero differisca significativamente da quello stimato.

L'approccio statistico considerato come best practice per il calcolo dell'LGD è la distribuzione BETA, caratterizzata da due parametri che possono essere messi in relazione con la media e la volatilità dei RR osservati.

2.8.2.5 Stima dell'Exposure at Default (EAD)

L'EAD rappresenta la stima del valore del credito nel momento in cui si verifica il default e può essere di due tipi:

1. Valore certo: quindi si conosce il valore esatto del finanziamento concesso (mutui);
2. Valore incerto: non si può quantificare l'importo a priori, ma solo al verificarsi del default (linee di credito, crediti revolving e derivati).

L'EAD si compone di due fattori: l'ammontare del fido utilizzato ed una stima della quota non ancora utilizzata, tale seconda parte viene stimata mediante elaborazioni statistiche o modelli econometrici. Per il calcolo dell'EAD sono riconosciuti tre approcci:

1. Credit Conversion Factor (CCF): rapporto tra la parte non utilizzata della linea di credito che si stima che possa essere utilizzata in caso di default e la parte attualmente non utilizzata. In altri termini il CCF non è altro che la quota del margine oggi inutilizzato che si prevede che verrà utilizzato dal debitore al momento dell'insolvenza. Il CCF è indicato anche come UGD = Usage Given Default;
2. Momentum Approach: rapporto tra l'utilizzato al momento del default ed il margine disponibile alla stessa data;
3. Exposure Multiplier: rapporto tra utilizzato al momento del default ed utilizzo corrente.

In generale la formula per il calcolo dell'EAD è la seguente (2.8):

$$EAD = QUOTA UTILIZZATA + QUOTA NON UTILIZZATA \cdot UGD \quad (2.8)$$

2.9 Verso Basilea 4

A partire dal 2015 il Basel Committee on Banking Supervision (BCBS) ha iniziato a pubblicare revisioni al framework normativo di Basilea 3, al fine di rivedere i concetti di trasparenza delle misure di rischio ed incrementarne la sensibilità, nonché a promuovere l'omogeneità e la comparabilità tra gli istituti dei diversi Paesi. Una critica che è stata rivolta al precedente framework riguarda l'eccessiva semplificazione dei risk drivers considerati nel calcolo dei rating esterni, in quanto i due soli parametri, ricavi e leverage, non tenevano in considerazione benchmark settoriali e geografici, perdendo informazioni sul rischio di credito.

I punti cardine della riforma a Basilea 3²⁸ sono:

- Revisione delle metodologie di valutazione del rischio di credito:
 - Standard Approach (SA): introduzione di due diligence all'interno delle banche sulla valutazione dell'attendibilità e della coerenza dei rating esterni;

²⁸ Si rimanda al link <https://www.bis.org/bcbs/publ/d424.pdf> per una visione completa delle riforme a Basile 3.

utilizzo di appropriati risk drivers per l'identificazione degli RWA da applicare;

- Approccio Internal Rating Based (IRB): limitazione di tale approccio solo per specifiche classi di esposizioni; introduzione di un output floor, che va a limitare il beneficio del risparmio di capitale, in aggiunta all'input floor già presente; richiesta di maggiore dettaglio nella stima dei parametri di rischio;
- Revisione del rischio di mercato:
 - passaggio dall'approccio Value at Risk (VaR) all'Expected Shortfall; Aggiunta del default risk charge (DRC);
 - Utilizzo di un unico calcolo calibrato sulla base di uno scenario di stress;
- Revisione del rischio operativo:
 - Utilizzo di un nuovo approccio standardizzato: introduzione
 - Introduzione del Business Indicator (BI) e Business Indicator Components (BIC)²⁹ nel modello di calcolo;
- Nuovo modello per il calcolo del rischio di controparte:
 - Introduzione del coefficiente alpha;
 - Calcolo del replacement cost (RC) include i concetti di accordi con e senza margine;
 - Introduzione del Potential Future Exposure (PFE);
- Revisione del framework relativo alle cartolarizzazioni:
 - Utilizzo di criteri semplici, trasparenti e standardizzati (STS);
 - Introduzione della gerarchia di modelli;
 - Previsto RWA incrementali per livelli più bassi del merito creditizio;
- Introduzione di un leverage ratio buffer per le banche di importanza sistemica (G-SIB).

Le riforme introdotte con Basilea 4 avranno decorrenza dal 1° gennaio 2022.

²⁹ Si rimanda al documento “Standardised Measurement Approach for operational risk” del 3 giugno 2016 per ulteriori approfondimenti.

3 Rischio di credito: Modelli di scoring

3.1 Introduzione

I modelli di scoring vengono utilizzati per la previsione dell'insolvenza di una controparte per supportare la decisione di concedere o meno un finanziamento. Lo scopo di tali modelli è quello di affidare uno score alla clientela, che ne rifletta l'affidabilità creditizia e permetta di assegnare una specifica classe di rating. Il processo di assegnazione del punteggio si basa sulla valutazione di opportune variabili di tipo quantitativo, ad esempio economico-finanziarie per un'impresa o socio-demografiche per un soggetto, che contribuiscono a definirne lo stato di salute dal punto di vista dei creditori. Nel corso del capitolo vengono riportati i diversi approcci metodologici proposti, riportandone le formule di calcolo ed effettuando un confronto tra le metodologie. Si descrive infine il processo che porta alla realizzazione di un modello.

3.2 Requisiti di un modello

Per poter costruire un modello di scoring è necessario disporre di un campione che comprenda controparti che in passato abbiano assunto comportamenti affidabili o insolventi. È importante non applicare delle selezioni arbitrarie su tale campione, altrimenti si potrebbe introdurre una distorsione nei dati in input. A questo punto si identificano le variabili quantitative che meglio discriminano le controparti che hanno registrato l'insolvenza, rispetto a quelle sane. A tale scopo si ritiene adeguato un orizzonte temporale di un anno. Identificati il campione e le variabili si procede con l'applicazione di una funzione in cui vengono inserite le variabili che genera uno score per ogni controparte. In base a come si posizionano le controparti sane ed insolventi nella scala di punteggio, si andrà ad assegnare un grado di rischiosità maggiore o minore ai punteggi più alti o più bassi; ad esempio se le aziende sane si collocano per la maggior parte in punteggi alti, mentre quelle insolventi in punteggi bassi, allora per le controparti con punteggi bassi si definisce una rischiosità maggiore, mentre per punteggi alti una rischiosità minore. L'assegnazione della rischiosità viene fatta sulla base di classi di rischio, quindi suddividendo la scala di punteggio in fasce dalla rischiosità più alta alla più bassa. È inoltre possibile inserire un valore di soglia, cosiddetto cut-off, al di sotto del quale le richieste di credito vengono respinte, o quantomeno ricontrattate da una persona. La scelta della soglia dipende dalla propensione al rischio, in quanto minore sarà la propensione, più la soglia sarà impostata in modo da filtrare i clienti rischiosi, accettando solo quelli considerati più sicuri dal modello, viceversa se la propensione al rischio è maggiore. Bisogna altresì prevedere una gestione adeguata di eventuali dati anomali in fase di input, che potrebbero falsare lo score generato. Gli obiettivi che un modello di scoring deve perseguire sono la stabilità delle performance a livello previsionale e nelle diverse fasi del ciclo economico, inoltre anche mutamenti nei criteri decisionali delle banche o nella normativa possono modificare considerevolmente le relazioni tra le variabili. È dunque necessario effettuare una verifica periodica delle performance ed una ristima quando si riduce l'efficacia discriminante del modello. Particolare attenzione va posta ai dati in input del modello, oltre alla suddetta gestione per i dati anomali si deve tenere in considerazione che nell'utilizzo di dati presi dai bilanci si rileva un ritardo rispetto alla reale situazione di quel momento dell'impresa, in quanto si

riferiscono ad eventi accaduti parecchi mesi prima. Alcuni modelli più complessi fanno anche uso di variabili di tipo qualitativo, ad esempio riguardanti il management dell'azienda, al suo prestigio, che pongono il problema su quale sia il più corretto modo di considerarli. Un problema che spesso si verifica è quello relativo alla scarsità dei dati utilizzati dal modello, dovuto a carenze dei sistemi informativi.

3.3 Approcci metodologici

3.3.1 Modello univariato

Sin dagli anni '20 si iniziarono ad utilizzare gli indicatori di bilancio per valutare la situazione economico-finanziaria di un'impresa. Attraverso l'analisi dei singoli indicatori di bilancio, l'approccio univariato rileva i fattori identificabili come "punti deboli" dell'impresa; in tal modo, accanto alla valutazione dello stato attuale di un'azienda, questa metodologia consente altresì di prevedere gli elementi che possono gravare sullo sviluppo futuro della stessa. Rispetto a questo approccio fu celebre la ricerca svolta da Beaver nel 1967³⁰ volta a esaminare la capacità predittiva di alcuni singoli indicatori al fine di prevedere l'insolvenza di un'impresa. Egli analizzò 158 imprese, di cui metà sane e metà anomale (intese come insolventi o prossime al fallimento) prese all'interno dello stesso settore e con le stesse dimensioni in termini di attivo netto. Il risultato della ricerca dimostra che vi sia una relazione tra l'insolvenza di un'impresa e il suo rapporto tra cash flow e debiti totali, mentre gli altri indicatori analizzati riportano risultati inferiori e meno significativi; ottenne una valutazione corretta per l'87% delle imprese che ad un anno si sono rivelate insolventi, tale capacità predittiva si va man mano riducendo andando indietro negli anni, arrivando al 78% se valutate cinque anni prima dell'insolvenza. Dal punto di vista degli indicatori la ricerca condotta da Beaver ha dimostrato che le informazioni utili a identificare precocemente il rischio d'insolvenza o di fallimento di un'impresa sono i dati contabili, le variabili legate alla capacità di generazione di cassa e quelle concernenti la struttura finanziaria; quelle indicanti, invece, la liquidità a breve termine hanno minore capacità diagnostica e una più imprecisa rilevanza informativa.

L'espressione grafica del modello è riportata nella figura 3.1, in cui sull'asse X viene riportata la variabile di analisi (ad esempio ROE), sull'asse Y la frequenza di quel valore ed infine il valore soglia (cut-off) definito come punto ottimo di separazione tra le due categorie, sane o anomale:

³⁰ Willian H. Beaver, "Financial Ratio as Predictors of Failure", Journal of Accounting Research, Vol.4, Issue Empirical Research in Accounting, 1966

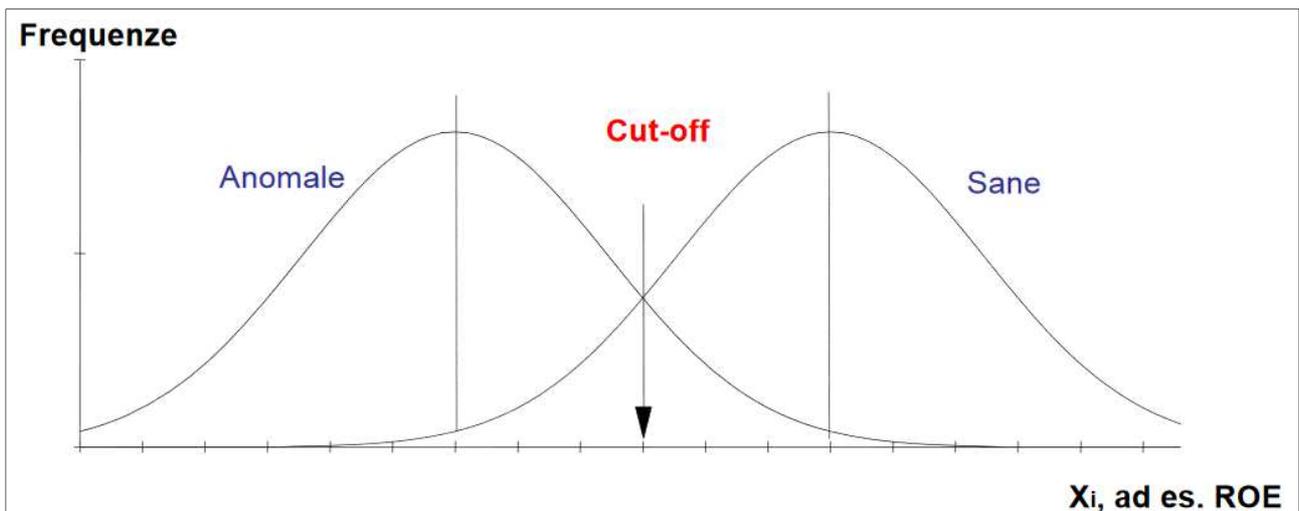


Figura 3.1: Grafico analisi univariata³¹

Tale studio però si dimostrò limitato, perché si considerano separatamente le variabili, comportando tra gli altri il rischio di giudicare equivalenti imprese in realtà differenti.

3.3.2 Modelli multivariati

L'approccio multivariato si basa su una combinazione delle variabili di interesse e ne svolge un'analisi simultanea al fine di valutare lo stato di salute di un'impresa dal punto di vista dei creditori. Nel considerare simultaneamente più variabili, l'approccio multivariato ne assegna dei pesi, consentendo di ottenere un'unica valutazione di sintesi relativa ad un'impresa. La differenza tra le varianti dei modelli multivariati risiede nell'assegnazione dei pesi alle variabili. Tra i diversi approcci il più celebre risulta senz'altro l'analisi discriminante introdotta da Fisher.

3.3.3 Analisi discriminante lineare

L'analisi discriminante lineare fu proposta da Fisher nel 1936³². Tale analisi si fonda sulla classificazione del campione in due (o più) categorie distinte, nell'ambito della trattazione la distinzione è tra imprese sane e insolventi. Con tale modello si vogliono perseguire due obiettivi: il primo è quello di costruire una regola di classificazione che consenta di individuare le variabili che meglio permettono di distinguere le categorie; il secondo è quello di poter classificare un nuovo oggetto, in questo caso una nuova impresa, in una delle categorie individuate al fine di prevederne il comportamento.

Ciascuna categoria è definita da una serie di variabili e ciascuna impresa viene assegnata alla categoria più prossima sulla base della maggiore o minore distanza complessiva e opportune ponderazioni sulle singole variabili. La classificazione avviene mediante una combinazione lineare delle variabili prese in considerazione per le due categorie, andando ad adottare la funzione che permette di massimizzare la distanza tra le medie delle due popolazioni a parità di variazioni. A titolo esplicativo si riporta nella figura 3.2 il grafico del

³¹ Fonte: http://disa.uniroma3.it/wp-content/uploads/2016/09/408_2635.pdf

³² R.A. Fisher, "The use of multiple measurements in taxonomic problems", Annals of eugenics, 1936

modello discriminante lineare con due variabili (X_1 e X_2) dove le due popolazioni A e B rappresentano le imprese sane e insolventi e la funzione discriminante rappresenta il cut-off, quindi il punteggio soglia che distingue le due popolazioni:

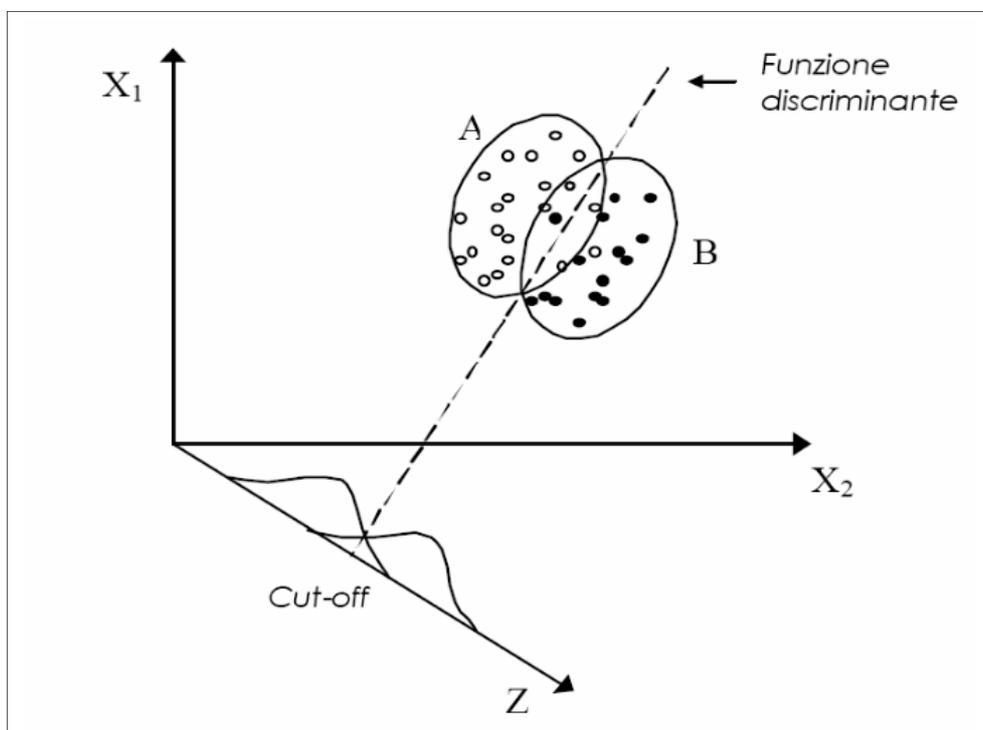


Figura 3.2: Grafico analisi discriminante a due variabili³³

L'approccio di Fisher consiste nell'elaborare una combinazione lineare tra le variabili considerate ottenendo un unico risultato, lo score, che determina la classificazione dell'impresa; tale combinazione lineare viene rappresentata dalla formula seguente (3.1):

$$S_j = \sum_{i=1}^n \alpha_i \cdot X_{ij} \quad (3.1)$$

Dove:

- S_j rappresenta lo score della j-esima impresa;
- α_i rappresenta il coefficiente della variabile X_i ;
- X_{ij} rappresenta la variabile descrittiva i-esima dell'impresa;
- n rappresenta il numero di variabili osservate.

L'analisi discriminante individua il vettore di coefficienti α che massimizza la distanza tra le due categorie mediante la formula (3.2):

$$\frac{E(X'_{i1} \cdot \alpha) - E(X'_{j2} \cdot \alpha)^2}{VAR(X' \cdot \alpha)} = \frac{(\bar{S}_1 - \bar{S}_2)}{\sigma_S^2} \quad (3.2)$$

Dove:

- X'_{i1} rappresenta il vettore delle variabili indipendenti rilevato per l'azienda i-esima appartenente alla prima categoria, indicativo delle imprese insolventi;

³³ Fonte: "Rischio e Valore nelle banche – misura, regolamentazione, gestione" – A. Resti, A. Sironi, Egea, 2008 – link: <https://www.lumsa.it/sites/default/files/UTENTI/u740/Capitolo%2011.pdf>

- X'_{j2} rappresenta il vettore delle variabili indipendenti rilevato per l'azienda j-esima appartenente alla seconda categoria, indicativo delle imprese sane;
- S_n rappresenta la media dei valori della funzione discriminante per le osservazioni della categoria n-esima (con $n = 1, 2$), secondo la relazione $S_{in} = X'_{in} \cdot \alpha$ detto anche centroide.

La regola di classificazione lineare viene quindi espressa in termini di distanze tra gli score, per cui l'impresa j-esima sarà assegnata alla prima categoria se vale la (3.3):

$$|S_j - S_1| < |S_j - S_2|, \text{ ovvero } S_j < \frac{1}{2} \cdot (S_1 + S_2), \text{ per } S_1 < S_2 \quad (3.3)$$

Altrimenti viene assegnato alla seconda categoria. Il termine $\frac{1}{2} \cdot (S_1 + S_2)$ rappresenta quindi il cut-off per la classificazione.

Tornando al riferimento della figura (3.2) si può dire che il piano X_1, X_2 rappresenta l'insieme delle imprese osservate, come già detto le aree delle popolazioni dei due gruppi di imprese (sane e insolventi) si sovrappongono, per cui il compito della funzione discriminante è quello di individuare la retta che meglio distingua le due categorie, in altre parole che commetta il minor numero di errori di classificazione possibile. Maggiore è l'area di sovrapposizione dei due gruppi, maggiore sarà la l'incertezza e quindi gli errori, di classificazione. Infine, va fatto notare che il modello di Fisher è di tipo non parametrico, in quanto non fornisce nessuna ipotesi sulla forma delle distribuzioni delle variabili.

3.3.4 Z-Score di Altman

Nella sua ricerca Edward Altman applicò l'analisi discriminante lineare ad una popolazione di 33 imprese statunitensi fallite nel periodo 1945-65 ed una popolazione di 33 società sane. Il suo risultato fu pubblicato nel 1968 ed il modello ottenuto fu il seguente:

$$\begin{aligned} Z = & 0,012 \cdot \text{Capitale circolante/Attivo netto} + \\ & 0,014 \cdot \text{Riserve di utili/Attivo Netto} + \\ & 0,033 \cdot \text{Utile ante interessi e tasse/Attivo Netto} + \\ & 0,006 \cdot \text{Valore di Mercato del Patrimonio netto/ Debiti Totali} + \\ & 0,999 \cdot \text{Ricavi/Attivo Netto} \end{aligned} \quad (3.4)$$

Secondo i test di significatività le variabili che risultano più importanti ai fini della classificazione sono la redditività e l'efficienza complessiva, mentre la liquidità risulta scarsamente significativa; questo risultato è in linea con quanto ottenuto da Beaver nella sua ricerca.

La classificazione proposta da Altman avviene confrontando lo score ottenuto dall'impresa con il cut-off, ottenuto come media dei centroidi delle due popolazioni, posto a 2,675. Al di sopra di tale cut-off l'impresa è considerata sana, al di sotto è considerata insolvente. In altre parole, maggiore è il punteggio dello Z-score, minore è la probabilità di insolvenza e quindi il rischio di quell'impresa. Altman individua inoltre una grey zone (zona d'ignoranza) nella quale si accentuano gli errori di classificazione, tale zona è compresa nell'intervallo 1,81 e 2,99.

Dal punto di vista del rischio il risultato del modello di Altman può essere letto come:

- $Z < 1,81$ rischio alto;

- $1,81 < Z < 2,99$ rischio medio;
- $Z > 2,99$ rischio basso.

La capacità predittiva di questo modello è molto buona se valutata nell'anno precedente alla registrazione dell'insolvenza dell'impresa come mostrato nella tabella (3.4), mentre perde di efficacia se si applica agli anni precedenti, ad esempio a due anni dall'insolvenza il potere predittivo cala all'82% e va riducendosi, risultando perfino meno accurato del modello di Beaver.

		Classificazione ottenuta		Totale
		Anomala	Sana	
Gruppo effettivo	Anomala	94%	6% Errore di I specie	Imprese realmente insolventi
	Sana	3% Errore di II specie	97%	Imprese realmente sane
Totale		Imprese classificate insolventi	Imprese classificate sane	Totale imprese considerate

Tabella 3.4: Analisi predittiva del modello di Altman al tempo t-1 in %

La tabella sopra riportata mostra come la capacità predittiva del modello di Altman sia particolarmente significativa se applicata un anno prima dell'effettiva insolvenza dell'impresa. Infatti, analizzando la riga relativa alle imprese effettivamente anomale, quindi insolventi, lo score commetterebbe un errore di classificazione del 6%, mentre per quanto riguarda le imprese sane l'errore commesso sarebbe solo del 3%. In media si avrebbe una corretta classificazione delle imprese del 95%.

3.3.5 Affinamento del modello di Altman

A seguito di varie critiche ricevute in merito al modello proposto nel 1968, Altman, insieme ad altri autori (Haldeman and Narayanan), pubblicò un nuovo modello nel 1977³⁴. Il nuovo modello, la cui formula non è stata resa nota, è stato basato sullo studio di 53 società fallite e 53 società sane e andando a considerare 7 variabili:

1. ROA: misurato come rapporto tra utili ante interessi e tasse e l'attivo netto totale;
2. Stabilità degli utili: calcolata con una misura normalizzata dello scarto quadratico medio della stima intorno al trend decennale del ROA;
3. Servizio del debito: valutato con il rapporto tra utili ante interessi e tasse e gli oneri finanziari totali;
4. Redditività cumulata: misurata dal rapporto tra riserve di utili ed Attivo netto;
5. Liquidità corrente;
6. Capitalizzazione: calcolata come rapporto tra valore di mercato del patrimonio netto (media dei prezzi delle azioni degli ultimi cinque anni) e valore totale del debito;

³⁴ E.I. Altman, R. Haldeman, P. Narayanan, "ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations", Journal of Banking and Finance, June 1977

7. Dimensione: misurata dal logaritmo dell'attivo netto.

3.3.6 Approfondimento al calcolo del cut-off

Per arrivare al calcolo del cut-off del modello di Altman è necessario introdurre tre concetti fondamentali: i) probabilità a priori e posteriori; ii) errore di prima e di seconda specie; iii) costo degli errori. La combinazione di questi elementi fornisce il livello di cut-off applicato da Altman nel suo modello.

3.3.6.1 Probabilità a priori e posteriori

Con probabilità a priori si intende la probabilità che, prima dell'osservazione del vettore X_i utilizzato per la classificazione, si estragga un'impresa appartenente ad uno dei due gruppi. Quindi la probabilità che, presa una nuova impresa, essa venga assegnata ex-ante ad uno dei due gruppi. Applicando il teorema di Bayes³⁵, si ottiene la probabilità a posteriori, ossia la probabilità che, osservate le caratteristiche dell'impresa X_i , questa appartenga al gruppo n-esimo (s=sane o a=anomale) attraverso la seguente formula (3.5):

$$P(n|X_i) = \frac{P(n) q_n}{P(X_i)} = \frac{P(n) q_n}{P(s) q_s + P(a) q_a} \quad (3.5)$$

Dove:

- q_n rappresenta la probabilità a priori;
- $P(n)$ rappresenta la probabilità, nota, che la popolazione n-esima generi l'impresa osservata;
- $P(n|X_i)$ rappresenta la probabilità a posteriori.

In tale ottica un'impresa verrà assegnata al gruppo delle imprese sane se vale la relazione (3.6):

$$q_s \cdot p(S) > q_a \cdot p(A) \quad (3.6)$$

Altrimenti sarà assegnata al gruppo delle imprese anomale.

Un simile criterio, semplificando le matrici di varianza e covarianza uguali tra le popolazioni, consente di ottenere una funzione discriminante lineare nella quale il cut-off di assegnazione è spostato della quantità $\ln\left(\frac{q_a}{q_s}\right)$ rispetto alla funzione originale di Fisher.

Nell'ipotesi limite in cui le probabilità a priori siano uguali ($q_a = q_s$) non si ottengono scostamenti dalla funzione discriminante lineare e si possono inserire tali probabilità lasciando inalterati gli altri criteri di semplificazione.

3.3.6.2 Errori di prima e seconda specie e costo dell'errore

In fase di analisi di un modello assume un ruolo centrale il concetto di errore, inteso come la classificazione errata di un'impresa. Nelle figure (3.1) e (3.2) le aree delle imprese sane e

³⁵ Detta anche probabilità condizionata, il teorema di Bayes viene adottato per calcolare la probabilità di una causa che ha scatenato l'evento verificato. La formula generica del teorema di Bayes è: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

anomale non sono perfettamente distinte, ma si sovrappongono e il cut-off applicato non sarà in grado di classificare correttamente tutte le osservazioni. Tali errori prendono il nome di errore di prima e di seconda specie, rispettivamente il primo indica un'impresa anomala classificata come sana e il secondo si riferisce alla classificazione di un'impresa sana come anomala. Di seguito si riporta in tabella 3.2 la matrice di confusione (Misclassification Rate – MR) in cui viene rappresentata l'intersezione tra la classificazione delle imprese prevista dal modello e il reale comportamento dell'azienda:

		Classificazione ottenuta		Totale
		Anomala	Sana	
Situazione reale	Anomala	Corretta classificazione Anomale	Errore di prima specie	Imprese realmente insolventi
	Sana	Errore di seconda specie	Corretta classificazione Sane	Imprese realmente sane
Totale		Imprese classificate insolventi	Imprese classificate sane	Totale imprese considerate

Tabella 3.2: Misclassification Rate -confronto tra comportamento reale e previsto delle imprese

Il peso di uno o dell'altro errore è evidentemente diverso, perché se classificare anomala un'impresa sana può non essere un errore grave, se consideriamo la cessione di un finanziamento si avrebbe un mancato guadagno, classificare invece sana un'impresa che poi si rileva insolvente ha una rilevanza maggiore, in quanto comporterebbe una perdita dell'investimento o del capitale concesso in finanziamento con i relativi interessi. Definendo C_1 il costo legato all'errore di prima specie e C_2 il costo legato all'errore di seconda specie, il costo atteso dell'errore di prima specie si minimizza mediante la formula (3.7):

$$\frac{p(S)}{p(A)} > \frac{q_a \cdot C_1}{q_s \cdot C_2} \quad (3.7)$$

Se si considerano uguali le matrici di varianza e covarianza delle due popolazioni la funzione discriminante lineare si sposta di una quantità $\ln \frac{q_a \cdot C_1}{q_s \cdot C_2}$ rispetto alla funzione di Fisher.

La funzione discriminante lineare si può quindi considerare come una tipologia di classificazione generica in cui si ignorano sia le probabilità a priori sia i costi di errate classificazioni, poiché vengono considerate valide le ipotesi di normalità multivariata delle distribuzioni delle variabili, il criterio d'uguaglianza delle matrici di varianza e covarianza tra le popolazioni e vengono classificate come identiche le probabilità a priori e i costi di classificazione.

3.3.6.3 Assegnazione del cut-off

L'assegnazione del cut-off è stata effettuata sulla base delle probabilità a priori e dei costi degli errori di classificazione. Il costo degli errori di classificazione è diverso a seconda del gruppo ed ha effetto solo sul termine noto. Il punteggio di cut-off è ottenuto dalla formula (3.8):

$$Cut - off = \ln \frac{q_a \cdot C_1}{q_s \cdot C_2} \quad (3.8)$$

Dove:

- q_a e q_s sono le probabilità a priori che l'impresa sia sana o insolvente;
- C_1 e C_2 sono i costi degli errori di prima e seconda specie.

Semplificando le cose, i termini q_a e q_s possono essere considerati come le proporzioni delle due popolazioni che, nel caso delle applicazioni pratiche, dove si effettua un'analisi su un campione pareggiato di aziende, non si specificano le probabilità a priori né i costi legati agli errori e si ottiene una funzione discriminante con cut-off centrato sullo 0 e tale funzione converge verso la funzione lineare di Fisher.

Il valore atteso del costo degli errori dovuti all'adozione del modello è dato dalla formula (3.9):

$$E(C) = q_a \cdot C_1 \cdot \frac{M_{a,s}}{N_a} + q_s \cdot C_2 \cdot \frac{M_{s,a}}{N_s} \quad (3.9)$$

Dove:

- N_a e N_s rappresentano le numerosità dei campioni delle imprese anomale e sane;
- $M_{a,s}$ e $M_{s,a}$ rappresenta il numero delle imprese classificate erroneamente;
- $q_a = 2\%$;
- $q_s = 98\%$;
- $C_1 = 70$;
- $C_2 = 2$.

Di conseguenza il valore del cut-off sarà pari a:

$$Cut - off = \ln \frac{0,02 \cdot 70}{0,98 \cdot 2} = -0,03364$$

Il modello di Altman ha classificato correttamente il 96,2% delle imprese anomale e l'89,7% di quelle sane al tempo t-1 rispetto al reale evento di insolvenza; il deterioramento dell'efficienza della classificazione negli anni precedenti risulta meno marcato rispetto al precedente modello, infatti andando a 5 anni dall'evento di avvenuta insolvenza la correttezza predittiva risulta circa del 70%.

L'analisi discriminante lineare ha una stretta relazione con la regressione lineare, infatti i coefficienti della funzione lineare sono pari a quelli della regressione con i minimi quadrati ordinari a meno di un rapporto costante. La regressione logistica costituisce un'alternativa all'analisi discriminante lineare nella valutazione del rischio di credito.

3.4 Regressione logistica

Attraverso questo modello è possibile stimare, mediante una variabile qualitativa dicotomica dipendente che descrive l'appartenenza all'insieme delle imprese sane o anomale:

$$Y = \begin{cases} 0 & \text{se impresa sana} \\ 1 & \text{se impresa anomala} \end{cases}$$

e come variabili indipendenti gli indicatori di bilancio.

Una versione della regressione è la linear probability model, che interpreta la y come probabilità di appartenenza ad un cluster. In tale modello si incontrano diverse difficoltà, ad esempio la varianza degli errori della stima non è costante, comportando un problema di eteroschedasticità³⁶ e la non normalità degli errori. Un altro problema riguarda il fatto che la stima della variabile Y non restituisce valori tra 0 e 1, ma prevede possibili valori negativi o molto maggiori di 1, comportando difficoltà interpretative dal punto di vista probabilistico.

Nella figura (3.3) seguente viene riportato un esempio di regressione lineare stimata mediante modello logistico e lineare tra la variabile dipendente (0;1) ed un indicatore di bilancio (X); si può notare come per il modello lineare le stime escano dall'intervallo ammissibile delle probabilità, mentre il modello logistico meglio approssimi la natura binaria della variabile target.

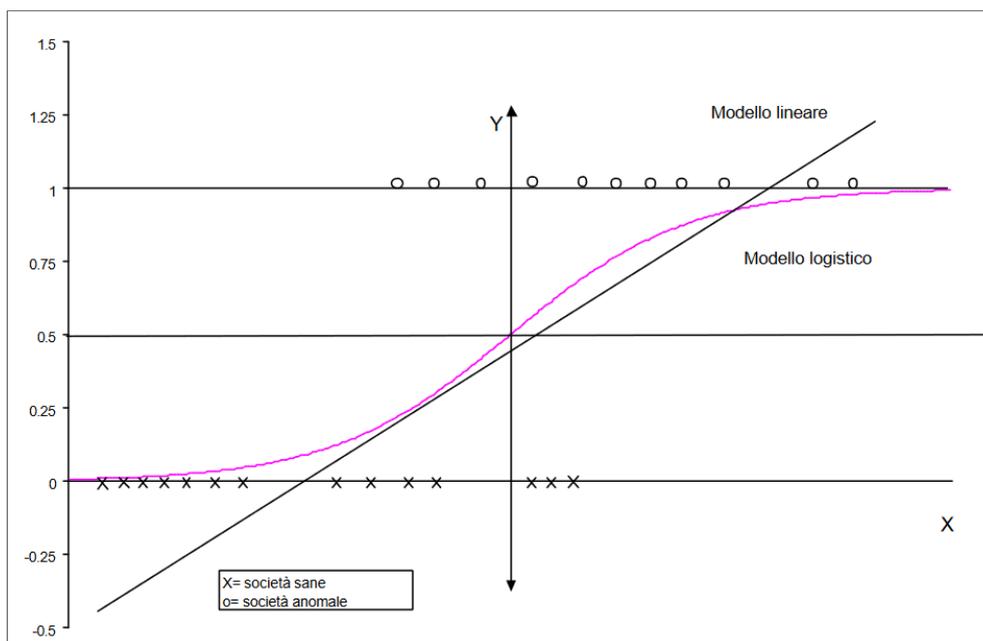


Figura 3.3: Confronto tra modello lineare e logistico³⁷

Il modello si basa sull'assunzione che esistano una serie di variabili osservabili che abbiano una relazione causale con la probabilità di un'impresa di diventare insolvente. Il risultato non sarà pertanto una probabilità dell'impresa di diventare insolvente, bensì una rappresentazione dicotomica di tale probabilità.

³⁶ Se $\text{var}(u|X=x)$ è costante, ossia se la varianza della distribuzione di u condizionata a X non dipende da X , allora u è detto omoschedastico. In caso contrario, u è eteroschedastico.

³⁷ Fonte: F. Varetto, corso di "Economia degli intermediari finanziari", A.A. 2017/2018, Politecnico di Torino.

Indicando con p la probabilità d'insolvenza di un'impresa, con X il vettore delle variabili indipendenti e con α e β il termine costante e i coefficienti si ha la seguente formula (3.10):

$$p = F(\alpha + \beta X) \quad (3.10)$$

ove F indica la funzione standard cumulativa logistica, in formula (3.11):

$$F(\alpha + \beta X) = \int_{-\infty}^{\alpha + \beta X} f(h) dh = \frac{1}{1 + e^{-(\alpha + \beta X)}} \quad (3.11)$$

In cui $f(h)$ indica la funzione di densità logistica (3.12):

$$f(h) = \frac{e^h}{(1 + e^h)^2} \quad (3.12)$$

Il modello stabilisce, dunque, la forma della distribuzione della probabilità d'insolvenza dalla relazione (3.13):

$$p = \frac{1}{1 + e^{-(\alpha + \beta X)}} \quad (3.13)$$

Si ha (3.14):

$$e^{-(\alpha + \beta X)} = \frac{1 - p}{p} \quad (3.14)$$

Dove il termine a destra rappresenta l'odd-ratio, dato dal rapporto tra la probabilità dell'evento ed il suo complemento.

Applicando il logaritmo naturale all'odd-ratio si ottiene (3.15):

$$\ln \frac{p}{1 - p} = \alpha + \beta X \quad (3.15)$$

La differenza tra linear probability model e modello logistico sta nel fatto che, nel primo è p ad essere messo in relazione con $\alpha + \beta X$, mentre nel secondo è il logaritmo dell'odd-ratio. La relazione (3.15) è equivalente a dire (3.16):

$$\ln \frac{p(A)}{p(B)} = \alpha + \beta X \quad (3.16)$$

Dove $p(A)$ e $p(B)$ sono le densità di probabilità delle popolazioni A e B. Applicando il teorema di Bayes, come in formula (3.5), si assegna l'oggetto alla categoria A se:

$$\ln \frac{p(A)}{p(B)} > \ln \frac{q_B}{q_A}, \text{ ovvero } \alpha + \beta X > \ln \frac{q_B}{q_A} \quad (3.17)$$

Si può osservare che, secondo tale relazione, il valore di cut-off, nel caso più semplice, si ha quando vi è perfetta incertezza in termini probabilistici ($p = 0,5$), cui corrisponde un valore nullo dell'esponente della funzione logistica cumulata.

3.5 Confronto tra analisi discriminante e regressione logistica

L'analisi discriminante ipotizza implicitamente che le imprese osservabili siano tratte da due universi distinti dati; la rilevazione delle variabili di bilancio sulle imprese può essere di aiuto per trovare le caratteristiche rilevanti e per individuare da quale universo esse provengono effettivamente. L'analisi discriminante cerca pertanto di prevedere l'appartenenza a un gruppo, dopo aver osservato le variabili ritenute rilevanti per caratterizzare le diversità tra i due universi. Inoltre, in termini di interpretazioni del risultato del modello e delle variabili per l'analisi discriminante non si stima un modello esplicativo dell'insolvenza ma si cerca di combinare insieme diverse variabili per avere un segnale unico, complesso, dell'appartenenza probabile ad un gruppo, dato a priori: mentre per quanto riguarda le variabili, quindi gli indicatori che compongono la funzione discriminante, la loro interpretazione va svolta individualmente in base al loro contributo marginale che danno al segnale complessivo.

Il modello logistico invece, come anche la regressione multipla, ipotizzano che le imprese siano tratte casualmente da un unico universo cui appartengono e cercano di stimare una caratteristica specifica di tali imprese: il grado di salute ovvero la probabilità (logistica, normale, lineare) di insolvenza/fallimento. Tale caratteristica è immaginabile come una variabile latente continua, di cui sono osservabili solo due essenziali determinazioni estreme (0;1). Questo modello quindi ipotizza che vi sia una relazione causale tra le variabili osservate sui dati contabili e la variabile dipendente; ciò significa che questi modelli implicitamente suppongono una relazione di causa-effetto tra i fenomeni economici sintetizzati dalle variabili di bilancio e lo stato di salute dell'impresa. Non viene stimata quindi l'appartenenza dell'impresa ad un gruppo, ma il grado dello stato di difficoltà economico-finanziaria in cui versa l'impresa. Per quanto concerne l'interpretazione degli indicatori che rappresentano le variabili esogene sono funzionali a spiegare la situazione dell'impresa dal punto di vista del creditore: gli indicatori risultano quindi utili in quanto forniscono indicazioni sulle componenti economico-finanziarie del modello che identificano lo stato di crisi dell'impresa e la sua evoluzione e/o il suo deterioramento.

I risultati empirici hanno dimostrato che:

- Se i dati sono normali e le matrici di varianza e covarianza dei due gruppi sono uguali l'analisi discriminante lineare è ottima per campioni piccoli, ma la performance è molto simile;
- Se le matrici di varianza e covarianza sono diverse il modello logistico è leggermente migliore;
- Per distribuzioni non normali il modello logistico è nettamente superiore.

3.6 Fasi del processo di stima di un modello

Il processo di stima è sostanzialmente diviso nelle seguenti fasi:

1. Selezione del campione;
2. Selezione delle variabili di analisi;
3. Stima del modello;
4. Calibrazione e Master Scale;
5. Backtesting.

3.6.1 Selezione del campione

La prima fase prevede la selezione di un numero sufficiente di oggetti (imprese), che vengono poi suddivise in gruppi (sane o anomale) identificati dalla variabile binaria Y. È importante avere a disposizione un elevato numero di imprese insolventi, al fine di garantire risultati più significativi possibili. Paradossalmente quindi una banca che ha concesso più crediti a clienti insolventi sarà avvantaggiata rispetto a una che ne ha concessi di meno, in quanto potrà ottenere un modello previsionale più efficace.

3.6.2 Selezione delle variabili di analisi

Questa fase consiste nell'identificare un set di indicatori in grado di fornire il maggior numero di informazioni rilevanti relative all'impresa. Le fonti informative possono avere diversa natura:

- Anagrafiche: identificano le informazioni personali della controparte specifica (nome, cognome, sesso, età, informazioni residenziali ecc.);
- Sociodemografiche e geo-settoriali: informazioni di carattere sociologico, demografico e geografico che possono caratterizzare delle controparti specifiche o una popolazione in esame;
- Andamentali interne: rappresentano tutte le informazioni del rapporto tra il cliente e la banca: indicano pertanto il rapporto che una controparte possiede nei confronti di una singola banca (tipo di rapporto, accordato, utilizzato, numero garanzie, giorni di sconfinò);
- Andamentali esterne: informazioni del rapporto tra il cliente e il sistema bancario: indicano pertanto il rapporto che una controparte possiede nei confronti di tutte le banche, tali dati sono forniti ad esempio dalla Centrale dei Rischi (CR) di Banca d'Italia o da CRIF;
- Bilancio: informazioni ricavabili dai bilanci di un'azienda, tramite questi dati è possibile individuare indicatori sintetici in grado di esplicitare la situazione patrimoniale-finanziaria-economica della controparte;
- Qualitative: informazioni qualitative del cliente in ambito business, settoriale e manageriale che non rientrano nelle casistiche precedenti, ma che hanno importanza nel processo di concessione del fido e nella valutazione della probabilità di default del cliente.

Il numero delle variabili dovrà essere ridotto al fine di ottenere un modello che sia stabile e generalizzabile, infatti l'uso di un numero eccessivo di variabili comporterebbe il rischio di overfitting, ossia il modello si adatta troppo ai dati utilizzati, per cui perde di significatività quando generalizzato. Nella scelta delle variabili bisogna tenere conto di 3 fattori:

1. Significatività dell'associazione con Y;
2. Interesse che le variabili rivestono nello studio del fenomeno;
3. Assenza di collinearità³⁸ tra di esse.

In particolare, la presenza di collinearità tra le variabili può inficiarne la significatività e l'attendibilità statistica.

³⁸ Situazione in cui i regressori, cioè le variabili, di un modello di regressione sono caratterizzati da una forte dipendenza lineare. In altre parole, una variabile è esprimibile mediante una funzione lineare di una, o più (multicollinearità), variabili.

3.6.3 Stima del modello

Una volta identificate le fonti è necessario selezionare il portafoglio clienti su cui sviluppare il modello segmentando le controparti in base a caratteristiche simili. La corretta identificazione del portafoglio è necessaria per stimare un modello coerente con il portafoglio a cui esso deve essere applicato. Drivers di segmentazione possono essere il settore di attività economica (SAE) fornito da Banca d'Italia, il tipo di operatività, il fatturato, la rischiosità, l'area geografica.

A partire dal segmento è necessario identificare la popolazione segmentando il portafoglio a date di riferimento prefissate a distanza almeno annuale per ridurre al minimo le correlazioni. A tal fine è possibile adottare due approcci: a finestre fisse, ossia considerando una popolazione a 12 mesi dall'ingresso in default, o a finestre variabili, ossia considerando una popolazione a date di riferimento fisse indipendentemente dalla data di ingresso in default.

Il campione viene successivamente suddiviso in modo da creare un campione di stima, su cui verranno calcolati i parametri del modello e un campione di validazione, per la verifica attraverso l'utilizzo di un campionamento stratificato mantenendo una distribuzione dei driver principali simili. La best practice per la divisione del campione è cross validation, che prevede di utilizzare il 70% per il campione di stima e il 30% per il campione di validazione. In caso di numerosità limitata del campione è possibile ricorrere alle seguenti logiche per identificare il campione di validazione:

1. Out of time: verifica del modello su dati osservati su un periodo di tempo diverso da quello utilizzato nel campione di stima;
2. Out of sample: verifica del modello con un campione di imprese diverse (statisticamente indipendenti) dal campione di stima, sullo stesso orizzonte temporale;
3. Out of universe: verifica del modello con un campione di imprese diverse e su un orizzonte temporale differente.

3.6.3.1 Analisi dei missing

Un aspetto importante da considerare in fase di stima è quantificare il fenomeno di mancanza informativa e approfondirne le cause. Un indicatore si definisce missing quando non è presente l'informazione. Le cause di tale assenza informativa possono derivare dalla perdita di informazioni in fase di archiviazione dei dati, oppure dall'imputazione non corretta da parte del gestore, o ancora in caso di assenza di informazione per una o più variabili sottostanti. Un elevato numero di missing può pregiudicare la significatività dell'indicatore, pertanto risulta necessario valutarne il mantenimento nel modello. Sono possibili diverse tecniche per la gestione dei missing:

1. Eliminazione del record;
2. Forzatura del valore (es. valore medio);
3. Clusterizzazione dell'indicatore inserendo i missing all'interno di una categoria;
4. Indicatori con un numero di missing superiore ad una certa soglia vengono scartati.

3.6.3.2 Analisi degli outliers

Ulteriore punto di attenzione sono gli outliers, anche detti valori estremi, che vengono trattati in modo da ridurre la distorsione in fase di analisi. Si riporta nella figura (3.4) un esempio grafico di outlier:

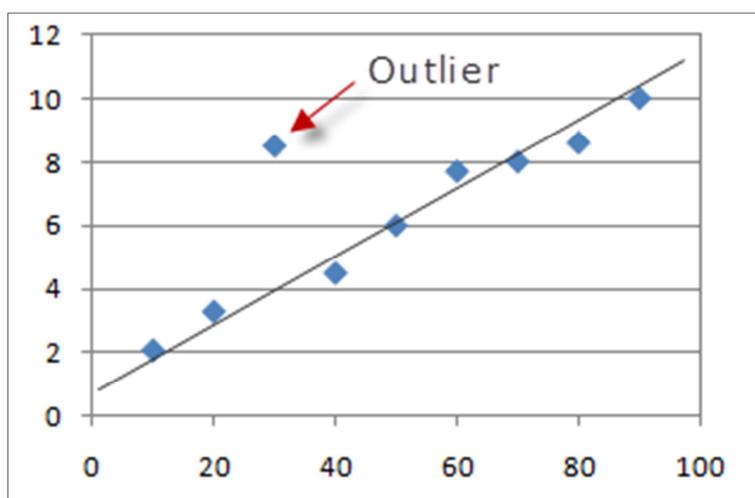


Figura 3.4: Rappresentazione grafica di un outlier³⁹

Rientrano nella categoria di outliers valori dell'indicatore molto distanti dal valore atteso, ovvero quando si hanno valori non rappresentativi della distribuzione dell'indicatore e che vanno ad inficiare le stime. L'analisi degli outliers deve comprendere anche un approfondimento sulle cause che hanno portato alla manifestazione del valore anomalo. La gestione degli outliers viene fatta attraverso il taglio delle code che prevede due approcci:

1. Approccio semplice: si procede a tagliare le code in maniera simmetrica e aprioristica, scegliendo una soglia oltre la quale effettuare il taglio;
2. Approccio complesso: si definisce una funzione di ponderazione da associare ad ogni valore della distribuzione: questa funzione assume diversi valori a seconda della valorizzazione in input e si procede al taglio in base ai valori assunti dalla funzione di ponderazione.

3.6.3.3 Discretizzazione degli indicatori

Gli indicatori continui e discreti infiniti vengono clusterizzati ottenendo indicatori discreti finiti maggiormente interpretabili. In tal senso gli approcci possibili sono i seguenti:

1. Analisi grafica della variabile continua: si analizza la distribuzione dell'indicatore in funzione della rischiosità e si inseriscono all'interno dello stesso bucket i valori che presentano caratteristiche di rischio simili. Tale discretizzazione avviene in modo diverso per variabili ordinali rispetto a quelle non ordinali, in quanto per le prime avviene mediante un algoritmo iterativo che si ripete finché tutti i bucket non sono significativamente differenti in termini di rischio, mentre per le seconde le eventuali aggregazioni sono valutate caso per caso e tenendo conto dell'interpretabilità dell'aggregazione;

³⁹ Fonte: <https://www.mathopenref.com/outlier.html>

2. Analisi per ventili: l'ampiezza della distribuzione dell'indicatore viene divisa per ventili che andranno a formare i bucket, nel caso in cui bucket adiacenti presentino caratteristiche simili vengono accorpati insieme;
3. Tukey test: si accorpano nello stesso bucket i valori per cui i test riportano medie del tasso di decadimento statisticamente simili.

Le prime due vengono utilizzati particolarmente nel processo di stima della PD, mentre il terzo approccio viene adottato maggiormente per modelli di EAD. Va in ogni caso rispettata la regola secondo cui valori uguali debbano finire nello stesso bucket.

Successivamente alla prima discretizzazione mediante uno dei tre approcci possono essere effettuati ulteriori trattamenti dei dati:

- Aggregazione per frequenza – variabili ordinali: consiste nell'aggregare i bucket scarsamente popolati con il più piccolo tra i bucket contigui. Questa aggregazione viene effettuata iterativamente al fine di eliminare bucket poco robusti dal punto di vista numerico rendendo la variabile più stabile nel tempo.
- Aggregazione per ordinamento – variabili ordinali: consiste nell'aggregazione di bucket che mostrano un'inversione del trend di rischiosità rispetto all'andamento dei bucket contigui. Anche tale trattamento viene effettuato iterativamente al fine di eliminare inversioni di trend e incoerenze sui segni attesi.

3.6.3.4 Analisi univariata

Una volta definita e costruita la long list di indicatori si conduce l'analisi univariata preliminare per selezionare i driver maggiormente significativi.

Un primo approccio possibile è la regressione lineare mediante la quale si può cogliere la relazione di linearità sussistente tra la variabile target e il regressore, ossia la variabile, sotto esame. Il metodo di stima dei parametri classico è quello dei minimi quadrati ordinari (OLS - Ordinary Least Squares), in cui i parametri sono scelti minimizzando lo scarto quadratico medio. La stima dei coefficienti mediante il metodo OLS è data dalla soluzione di un problema di minimizzazione (3.18):

$$\min_{b_i} \sum_{j=1}^n [Y_j - \sum_{i=1}^k \beta_i X_j]^2 \quad (3.18)$$

Dove:

- Y_j è la variabile dipendente;
- β_i rappresenta la pendenza del regressore, quindi la variazione che comporta sulla variabile dipendente una variazione unitaria dell' i -esimo indicatore;
- X_j è il valore dell'indicatore j -esimo.

In tale approccio la stima viene effettuata tramite una relazione lineare tra la variabile dipendente e le variabili indipendenti come riportato nella formula (3.19):

$$\bar{Y} = \sum_{i=1}^k \beta_i X_j \quad (3.19)$$

Dove \bar{Y} rappresenta la stima della variabile dipendente.

In seguito, si effettuano:

1. Un controllo sulla coerenza dei segni degli indicatori: in quanto verranno tenuti solo quelli il cui segno è coerente con la relazione che hanno con la variabile dipendente, in particolare:
 - Un segno positivo sarà indice di un aumento del rischio proporzionale con l'aumento dell'indicatore;
 - Un segno negativo indica una relazione inversa in termini di rischio.
2. Il test t: al fine di stimare la significatività di un indicatore, attraverso il test di significatività del coefficiente ad esso associato. Tale selezione può essere effettuata attraverso l'interpretazione del P-value associato a ciascun coefficiente. Il P-value rappresenta la probabilità che il test statistico convalidi l'ipotesi nulla H_0 , ossia l'ipotesi secondo cui la variabile non è statisticamente significativa per il modello, rispetto all'ipotesi alternativa H_1 secondo cui la variabile è statisticamente significativa. Utilizzando i P-value si può effettuare una valutazione per singola variabile dell'influenza sulla variabile dipendente. Il test viene effettuato sui coefficienti delle variabili, testando l'ipotesi nulla che essi siano pari a 0. Affinché sia rigettata l'ipotesi nulla il P-value deve essere inferiore al valore di α scelto, dove α rappresenta il livello di significatività, ossia la probabilità di accettare o rigettare l'ipotesi nulla. Se si rigetta l'ipotesi nulla si può quindi dire che la variabile testata ha un'influenza statisticamente significativa sulla variabile dipendente. Secondo tale principio sono da scartare le variabili che presentano P-value maggiori del valore α scelto.
3. Calcolo dell'indice R^2 : esprime la frazione della varianza di Y spiegata dalla regressione. Tale indice è calcolato mediante la formula (3.20):

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (3.20)$$

Dove:

- ESS (Explained Sum of Square) rappresenta la somma delle deviazioni tra valori predetti dal modello e la loro media al quadrato;
- TSS (Total Sum of Square) rappresenta la somma delle deviazioni tra valori osservati e la media loro media.

Se la regressione viene fatta su un singolo indicatore, R^2 rappresenta il quadrato del coefficiente di correlazione tra X e Y.

L'altro approccio utilizzabile è quello della regressione logistica, che, come già mostrato in figura (3.3), meglio approssima la variabile target e viene più largamente adottato per la stima dei modelli di PD. In tale approccio la stima dei parametri viene effettuata mediante il metodo della massima verosimiglianza, ossia attraverso la costruzione di una funzione di verosimiglianza, nel seguente modo:

$$y_i^* = \beta' x_i + u_i \quad (3.21)$$

Dove:

- y_i^* rappresenta la variabile non osservabile;
- x_i è il vettore degli indicatori;
- β' è il vettore dei coefficienti, singolarmente indica la variazione del logit per variazione unitaria dell'indicatore i-esimo;
- u_i il termine d'errore.

La probabilità di $y_i=1$ è:

$$Pr(y_i = 1) = Pr(u_i > -\beta'x_i) = 1 - F(-\beta'x_i) \quad (3.22)$$

Ove F è la distribuzione cumulata dell'errore u.
La funzione di verosimiglianza è la seguente (3.23):

$$L = \prod_{y_i=0} F(-\beta'x_i) \prod_{y_i=1} F(1 - \beta'x_i) \quad (3.23)$$

In cui la funzione F dipende dall'ipotesi sulla distribuzione di u_i . Ipotizzando che u_i segua una distribuzione cumulata logistica si ottiene il modello LOGIT, pertanto si hanno le formule (3.24) e (3.25):

$$F(-\beta'x_i) = \frac{1}{1+e^{\beta'x_i}} = Pr(y_i = 0|x_i) \quad (3.24)$$

$$1 - F(-\beta'x_i) = \frac{e^{\beta'x_i}}{1+e^{\beta'x_i}} = Pr(y_i = 1|x_i) \quad (3.25)$$

Risolvendo quindi la (3.23) si ha la (3.26):

$$L = \prod_{i=1}^n \left(\frac{1}{1+e^{\beta'x_i}} \right)^{(1-y_i)} \left(\frac{e^{\beta'x_i}}{1+e^{\beta'x_i}} \right)^{y_i} \quad (3.26)$$

Ipotizzando invece che la funzione F segua una distribuzione cumulata di tipo normale si ha il modello PROBIT (o NORMIT), in cui l'errore $u_i \sim N(0;\sigma^2)$, si ha (3.27):

$$F(-\beta'x_i) = \int_{-\infty}^{-\beta'x_i/\sigma} \frac{1}{(2\pi)^{1/2}} e^{(-\frac{r^2}{2})} dr \quad (3.27)$$

Poiché LOGIT e PROBIT hanno distribuzioni molto simili tranne che nelle code e la prima è più facilmente manipolabile matematicamente, si preferisce usare questa.

La probabilità condizionale che $y_i=1$ ha una relazione non lineare rispetto all'osservazione della variabile x_i , mentre il logaritmo dell'odd, formula (3.28) è funzione lineare delle variabili (3.29):

$$odd = \frac{Pr(y_i|x_i)}{1-Pr(y_i|x_i)} \quad (3.28)$$

$$\ln(odd) = \text{logit}(Pr(y_i|x_i)) = \ln \left[\frac{Pr(y_i|x_i)}{1-Pr(y_i|x_i)} \right] = \beta'x_i \quad (3.29)$$

Da cui deriva la funzione di log-verosimiglianza seguente (3.30):

$$\ln(L) = \sum_{i=1}^n \{y_i\beta'x_i - \ln[1 + e^{\beta'x_i}]\} \quad (3.30)$$

La stima dei parametri si ottiene derivando la funzione di log-verosimiglianza rispetto ad ogni parametro e ponendo tali equazioni uguali a zero.

Anche in tale approccio è prevista l'analisi di significatività dei beta, quindi degli indicatori e del calcolo degli indici di performance al fine di scartare indicatori statisticamente non significativi. Si possono applicare 3 approcci:

1. Test della significatività dei beta: attraverso l'uso di test come likelihood ratio, Wald test e score test si possono testare le significatività dei coefficienti beta e di conseguenza dell'indicatore ad esso associato;
2. Indice di Gini (o Accuracy ratio): è la misura per descrivere la disuguaglianza di una distribuzione. L'indice di Gini si basa sulla curva di Lorenz della distribuzione ed è legata all'area compresa tra la linea ideale e il modello casuale. In figura (3.5) sono riportate:
 - a. l'Ideal CAP, ovvero la curva che rappresenta il modello ideale, dove vengono identificate perfettamente tutte le imprese insolventi;
 - b. Random CAP assegna a tutte le imprese la stessa probabilità, per cui non ha nessun valore predittivo;
 - c. Actual CAP è il modello effettivo.

In termini grafici l'indice è dato dal rapporto tra l'area sottesa dalla curva del modello effettivo (A) e l'area sottesa dal modello ideale (A+B). Tale valore varia tra zero ed uno, tanto più è vicino all'uno, più il modello si adatta bene, quindi ha una buona capacità predittiva. Generalmente sono considerati buoni modelli quelli che registrano valori dell'indice superiori al 50%.

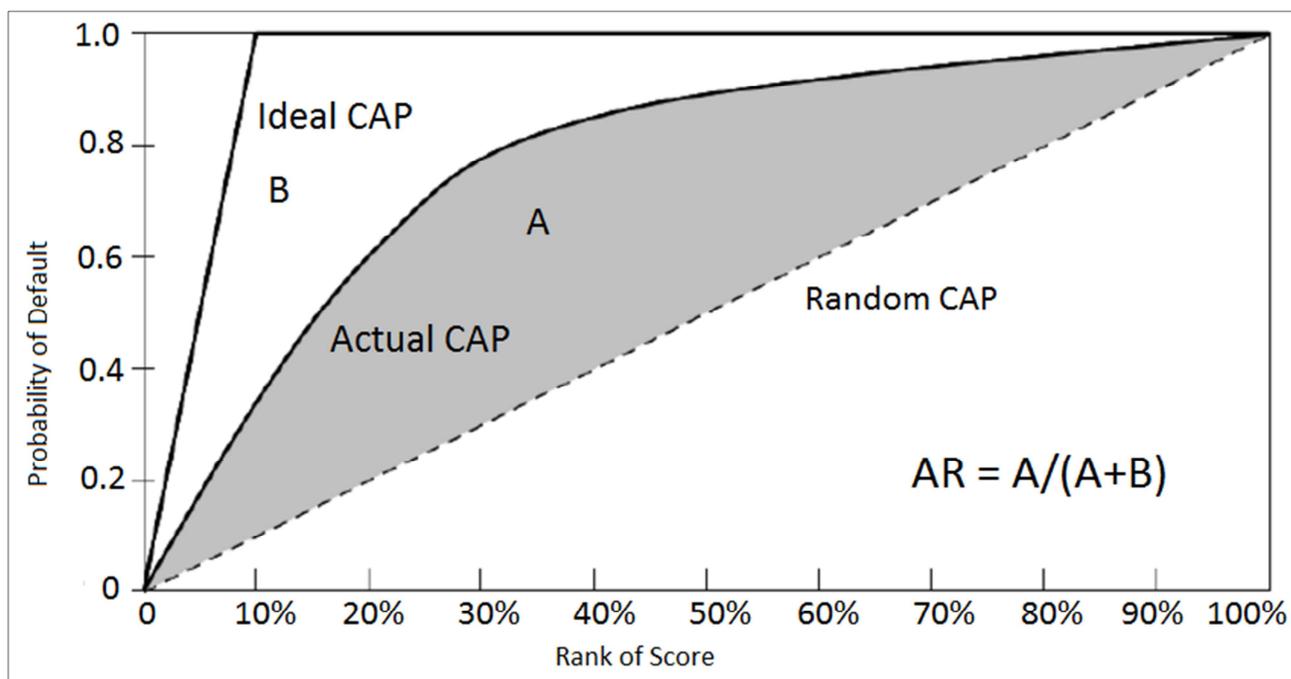


Figura 3.5: Accuracy Ratio⁴⁰

3. Curva ROC⁴¹: rappresenta la sensibilità e (1-specificità), come true positive rate (vero positivo) e false positive rate (falso positivo) e studia i rapporti tra allarmi veri e falsi allarmi.

⁴⁰ Fonte: https://www.researchgate.net/figure/Cumulative-accuracy-profile-CAP-curve_fig2_45133229

⁴¹ Tom Fawcett, "An introduction to roc analysis. Pattern Recognition Letters", 27(8):861 – 874, 2006

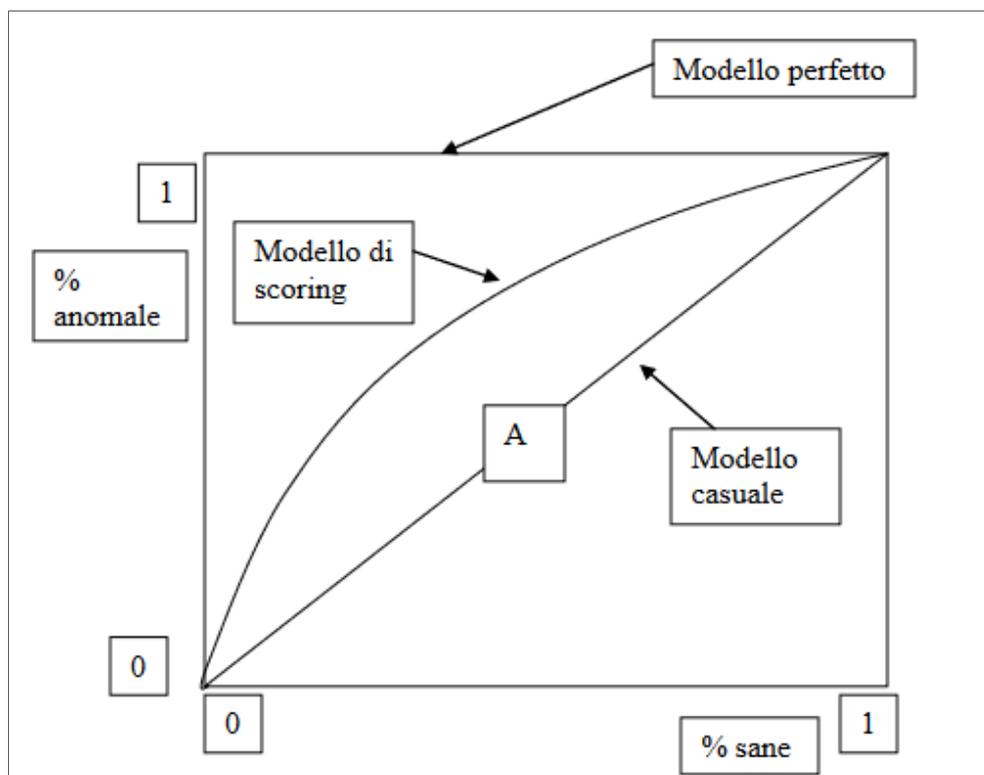


Figura 3.6: Curva ROC⁴²

Nella figura (3.6) la diagonale rappresenta un sistema di scoring casuale, mentre il segmento superiore che delimita il grafico individua il sistema perfetto. L'area A è compresa tra il segmento inferiore e la curva del modello di scoring effettivo. L'area tra il segmento inferiore ed il modello casuale vale 0.5. Se il sistema di scoring fosse perfetto l'area A sarebbe uguale ad uno. Nei sistemi reali l'area sottesa è compresa tra zero ed uno e sono tanto più accurati tanto più tendono ad uno. Il grafico della curva ROC (3.6) è molto simile a quello dell'accuracy ratio (3.5), infatti tra i due indici vi è una relazione data dalla formula (3.31):

$$ACCURACY = 2 \cdot ROC - 1 = 2 \cdot (ROC - 0,5) \quad (3.31)$$

In generale sono considerati buoni modelli quelli che registrano valori superiori al 75%.

La curva ROC (Receiver Operating Characteristic) è una curva utilizzata per la valutazione di un classificatore binario e lega la probabilità di ottenere un vero positivo tra le imprese effettivamente insolventi alla probabilità di ottenere un falso positivo (errore di prima specie) tra le imprese sane al variare della soglia del test scelta (valore di cut-off). Gli errori di classificazione sono identificabili sul grafico come riportato nella figura seguente (3.7):

⁴² Fonte: F. Varetto, corso di "Economia degli intermediari finanziari", A.A. 2017/2018, Politecnico di Torino.

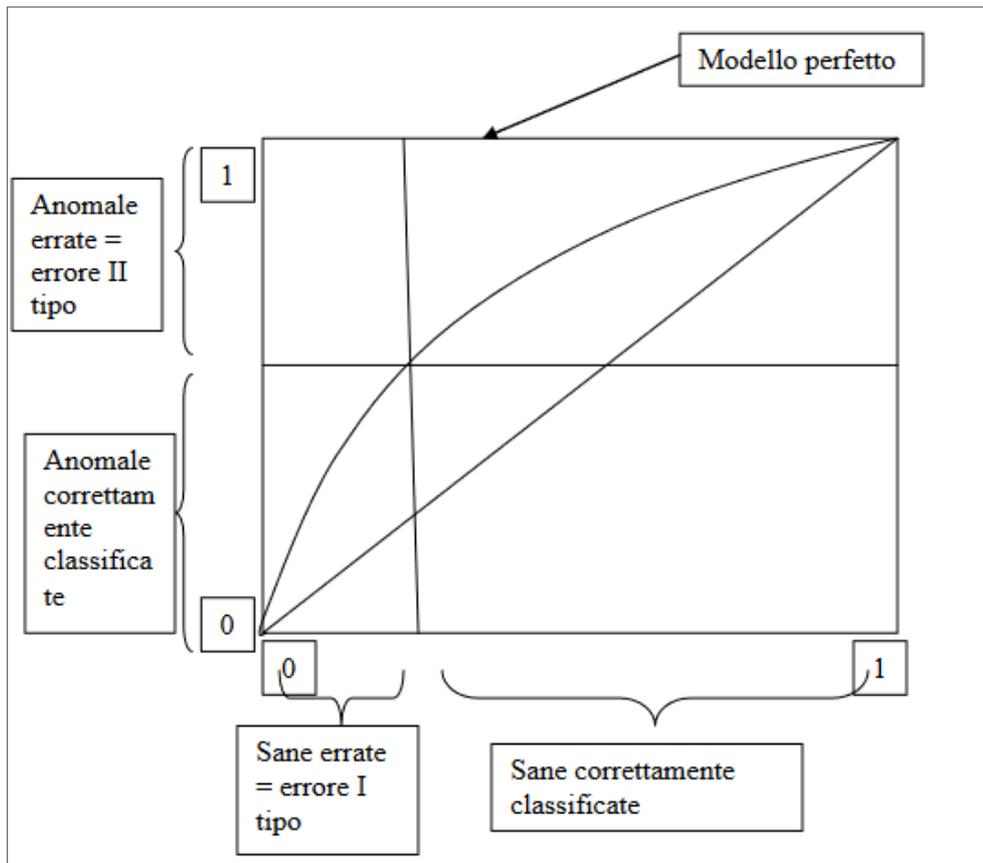


Figura 3.7: Curva ROC con indicazione degli errori⁴³

3.6.3.5 Analisi di correlazione

Successivamente all'analisi univariata si procede con il calcolo della matrice di correlazione i cui elementi sono calcolati a seconda del tipo di indicatore. È possibile effettuare un passaggio intermedio tra l'analisi univariata e l'analisi di correlazione in cui viene effettuata una trasformazione delle variabili nel loro Weight of Evidence (WOE) per ogni classe attraverso la seguente formula (3.32):

$$WOE_i = \ln \left(\frac{\frac{S_i}{\sum_{i=1}^K S_i}}{\frac{A_i}{\sum_{i=1}^K A_i}} \right) \quad (3.32)$$

Dove:

- S_i rappresenta la frequenza assoluta di osservazioni sane nella classe i ;
- A_i rappresenta la frequenza assoluta di osservazioni anomale nella classe i ;
- K è il numero di classi della variabile.

Attraverso questo metodo si ha il vantaggio di trasformare tutte le variabili (ordinali e non ordinali) in variabili numeriche con lo stesso ordine di grandezza, mantenendo lo stesso ordinamento delle classi rispetto alla rischiosità.

Le metriche utilizzabili per l'analisi di correlazione sono:

⁴³ Fonte: F. Varetto, corso di "Economia degli intermediari finanziari", A.A. 2017/2018, Politecnico di Torino.

1. Correlazione di Spearman: è una misura del grado di relazione tra due variabili per le quali non si fa altra ipotesi della misura ordinale e viene applicato ad indicatori discreti a cui è applicabile un ordinamento in termini di rischiosità. In primo luogo, viene effettuata una trasformazione delle variabili in modo da ordinarli dapprima in modo crescente e poi associarli ad un rango. A questo punto viene applicata la formula di Spearman (3.33):

$$\rho = 1 - \left(\frac{6 \sum_{i=1}^n d_i^2}{(n^2-1)n} \right) \quad (3.33)$$

dove d_i rappresenta la differenza di posizionamento tra i ranghi e n è la numerosità. Il risultato può variare tra -1 e +1, dove per valori vicini a -1 si ha correlazione negativa, per valori intorno allo 0 non si ha correlazione, per valori vicini a +1 si ha una correlazione positiva. La best practice prevede che vengano scartati quegli indicatori che hanno correlazione superiore a +0,5 o inferiore a -0,5.

2. Correlazione di Pearson: indice relativo a due variabili statistiche che ne esprime il grado di relazione di linearità, viene applicato per indicatori continui. L'indice si ricava dalla formula (3.34):

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad (3.34)$$

Dove:

- $S_{xy} = \sum_{i=1}^n (x - \bar{x})(y - \bar{y})$ in cui x e y sono i valori osservati mentre \bar{x} e \bar{y} sono i valori medi di quell'indicatore;
- $S_{xx} = \sum_{i=1}^n (x - \bar{x})^2$;
- $S_{yy} = \sum_{i=1}^n (y - \bar{y})^2$.

Il risultato può assumere valori tra 0 e 1, in cui un valore vicino a 0 esprime una relazione debole tra le variabili, mentre un valore vicino a 1 esprime una forte relazione tra le variabili. La best practice prevede di scartare i valori superiori a 0,5.

3. Indice V di Cramer: indica il grado di relazione tra due indicatori mediante la normalizzazione della variabile chi quadro associata alla distribuzione a confronto, viene applicato per indicatori discreti. Tale indicatore viene calcolato attraverso i seguenti passi:

$$\chi^2 = \sum_{i=1}^H \sum_{j=1}^K \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*} \quad (3.35)$$

$$\Phi^2 = \frac{\chi^2}{n} \quad (3.36)$$

$$V = \sqrt{\frac{\Phi^2}{\min\{(H-1), (K-1)\}}} \quad (3.37)$$

Dove:

- n_{ij} è la frequenza effettiva con cui si osservano i valori i -esimo e j -esimo;
- n_{ij}^* è la frequenza teorica attesa di osservare i valori i -esimo e j -esimo;
- $(n_{ij} - n_{ij}^*)$ è detta contingenza;
- H e K sono le numerosità dei bucket degli indicatori analizzati.

Tale indicatore assume valori tra 0 e 1 e, come per il coefficiente di Spearman, un valore prossimo allo 0 indica una relazione debole tra gli indicatori, mentre un valore prossimo all'1 indica una relazione forte tra gli indicatori. Anche in questo caso la best practice prevede di scartare gli indicatori con un valore dell'indice superiore a 0,5.

3.6.3.6 Analisi multivariata

L'analisi multivariata ha lo scopo di identificare il set di indicatori che massimizza la capacità predittiva del modello e minimizza la correlazione. Questa fase può essere condotta attraverso l'approccio stepwise, ossia mediante una selezione delle variabili indipendenti allo scopo di identificare un set di predittori che abbiano la migliore relazione con la variabile dipendente. Esistono due metodi di selezione delle variabili:

1. Forward (in avanti): inizia con un modello vuoto, nel quale nessuna variabile è selezionata; nel primo step viene aggiunta la variabile con l'associazione maggiormente significativa (AR maggiore) sul piano statistico. Ad ogni step successivo viene aggiunta la variabile con la maggiore associazione statisticamente significativa tra le rimanenti e si effettua l'analisi di correlazione con quelle già selezionate ed il processo prosegue sino a quando non vi è più variabile con associazione statisticamente significativa con la variabile dipendente.
2. Backward (all'indietro): parte considerando tutte le variabili nel modello e, ad ogni step, rimuove quella che ha l'associazione statisticamente meno significativa con la variabile dipendente.

Il modello stepwise va avanti e indietro, aggiungendo od eliminando variabili, fino ad ottenere un modello con i soli indicatori maggiormente significativi.

3.6.3.7 Risultato del modello

Arrivati a questo punto abbiamo ottenuto il set di informazioni consistente e significativo per effettuare l'analisi del rischio della controparte. È possibile ottenere per ogni controparte lo score indicativo del rischio ad esso assegnato e, sulla base di questo, individuare un ordinamento.

3.6.4 Test del modello

Una volta sviluppato il modello sul campione di stima, va testato sul campione di validazione, al fine di valutare la bontà dello score assegnato e l'assenza di overfitting, quindi che i risultati non siano troppo adattati al campione su cui è stato effettuato lo sviluppo. Come anticipato le caratteristiche del campione di validazione sono essenzialmente 2, un periodo di tempo differente (out of time) rispetto a quello del campione di sviluppo e un campione di controparti differente da quelle utilizzate nel campione di sviluppo (out of sample). Nei casi in cui la dimensione del campione non fosse sufficientemente ampia è preferibile effettuare una validazione in-sample con un metodo bootstrap, andando a selezionare in maniera casuale dal campione di sviluppo un numero opportuno di sottocampioni (indicativamente almeno 10 sottocampioni con dimensione pari ad almeno il 30% del campione di sviluppo). Si possono effettuare prove di backtesting anche su campioni di validazione diversi, ad esempio considerando solo un campione out of

time, solo out of sample o anche entrambe contemporaneamente. Inoltre, è anche possibile svolgere più volte la stima e la validazione del modello modificando la composizione dei campioni adottati, ad esempio, considerando la suddivisione del campione su base temporale, quindi il primo 70% per lo sviluppo e il restante 30% per la fase di validazione, oppure per estrazione casuale dal dataset sempre nello stesso rapporto. Questo approccio può essere utile per andare a verificare che siano coerenti i risultati ottenuti al variare della composizione dei campioni.

3.6.5 Calibrazione del modello

La calibrazione del modello avviene mediante l'aggiustamento dell'intercetta, effettuato nel passaggio dallo score campionario allo score dell'universo. Questo passaggio è reso necessario in quanto in fase di sviluppo si lavora su un campione pareggiato di controparti, mentre nella realtà i default che si registrano sono molto meno del 50% del totale della popolazione; di conseguenza si dovrà shiftare lo score verso lo 0, cioè verso un'impresa sana. Per farlo si andrà a sottrarre alla costante una certa quantità ottenuta dalla formula (3.38):

$$\text{costante corretta} = \text{costante} - \ln\left(\frac{\pi_{sane}}{\pi_{anomale}}\right) \quad (3.38)$$

Dove:

- π_{sane} è la percentuale di imprese sane nella popolazione;
- $\pi_{anomale}$ è la percentuale di imprese anomale nella popolazione.

Se il campione non è pareggiato tale correzione verrà fatta in due step, nel primo si riporta la composizione del campione ad un rapporto 50 e 50, poi si procede con la correzione della formula (3.38). Inoltre, tale aggiustamento è differente per regressione logistica (3.39) e analisi discriminante (3.40).

$$\text{costante corretta logistica} = \alpha + \ln\left(\frac{n_{sane}}{n_{anomale}}\right) + \ln\left(\frac{\pi_{anomale}}{\pi_{sane}}\right) \quad (3.39)$$

$$\text{costante corretta analisi discriminante} = \alpha + \ln\left(\frac{n_{anomale}}{n_{sane}}\right) + \ln\left(\frac{\pi_{sane}}{\pi_{anomale}}\right) \quad (3.40)$$

L'altro fattore di cui bisogna tenere conto in fase di calibrazione del modello è il diverso costo associato ai due tipi di errori, per cui si può decidere di rifiutare il credito ad un cliente quando il costo atteso dall'errore di primo tipo è superiore al costo atteso derivato dal secondo tipo di errore, quindi riprendendo la formula (3.9) relativa al costo atteso degli errori e assumendo le stesse considerazioni, la costante corretta diventa (3.41):

$$\text{costante corretta} = \alpha + \log\frac{\pi_{anomale} \cdot C_1}{\pi_{sane} \cdot C_2} \quad (3.41)$$

3.6.6 Master Scale

Calcolati gli score si andrà a costruire la scala di rating raggruppando i valori di score per intervalli (calibrati sulla base delle distribuzioni). Sono disponibili diverse metodologie per la costruzione di scale di rating:

- Approccio basato su analisi statistiche (cluster analysis o kernel analysis) che raggruppano in un numero finito e piccolo di cluster gli score, con l'obiettivo di minimizzare la variabilità intracuster e massimizzare quella intercluster, ossia raggruppare all'interno dello stesso cluster tutti gli score, quindi le controparti ad essi associate, che hanno caratteristiche simili. Tale metodo, però, prevede spesso l'interazione di un esperto che effettui degli aggiustamenti. Per ogni classe viene calcolata la PD media e viene identificato il range di rischiosità, ossia PD massima e minima di classe. È importante osservare se la PD media per classe abbia andamento monotono crescente;
- Utilizzo di scale definite a priori, ossia facendo riferimento a scale esterne per le PD e per gli estremi di ciascuna classe. Adottando tale approccio si ha un vantaggio sia in termini di trasparenza comunicativa, perché adotta classi regolamentari, sia in termini di benchmarking.

L'associazione a ciascuna classe di una PD avviene attraverso l'assegnazione del tasso di default teorico ottenibile mediante 2 approcci: through the cycle o point in time.

L'approccio through the cycle (figura 3.8) si basa sul concetto di central tendency (tendenza centrale), ossia la frequenza di default osservata sul portafoglio su un orizzonte temporale di lungo periodo, nella figura l'orizzonte è di 7 anni. Ha il vantaggio di essere più stabile nel lungo periodo, ma è meno reattivo in caso di variazioni improvvise del mercato o normative.

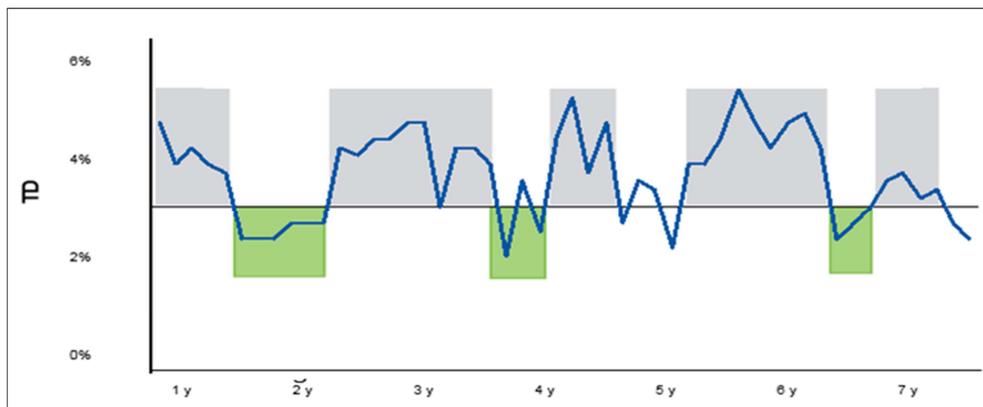


Figura 3.8: TD through the cycle – lungo periodo

L'approccio point in time (figura 3.9) invece si basa su un orizzonte temporale di breve periodo, nella figura l'orizzonte è di 12 mesi. Il vantaggio di tale approccio è quello di essere più reattivo nel breve periodo, ma risulta di conseguenza molto volatile.

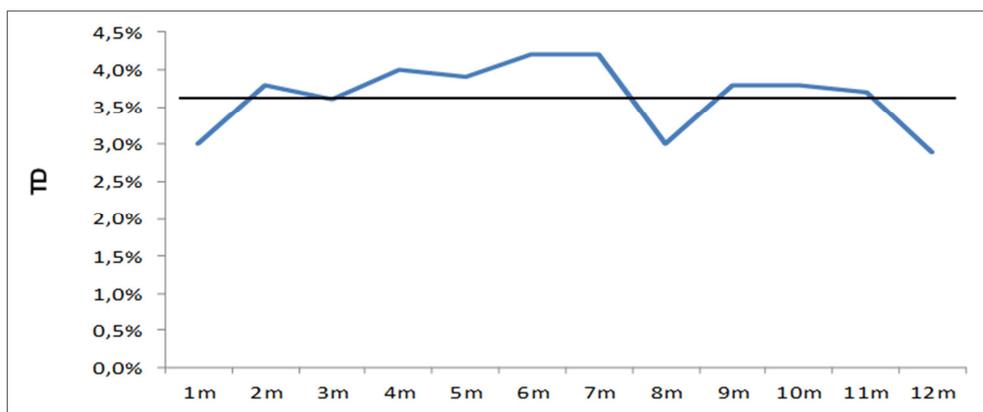


Figura 3.9: TD point in time – breve periodo

La PD ottenuta dal modello rispecchia la rischiosità del campione su cui viene fatta la stima, mentre la suddivisione in classi di rating assegna la stessa PD a tutte le esposizioni che vi fanno parte, passando quindi da uno score individuale ad un tasso di default calcolato per classi di rischio.

3.6.7 Backtesting del modello

La validazione dei modelli si basa su aspetti di natura quantitativa e qualitativa, al fine di valutare la bontà del modello stimato basandosi sull'analisi dei dati passati ed è classificabile nelle seguenti attività:

1. Analisi del potere discriminante: valutare la capacità del modello di rating di ordinare le controparti in base alla loro effettiva qualità creditizia. A tale scopo vengono utilizzati i già visti indice di Gini, curva ROC e Miscalssification rate (errori di I e II specie) a cui si aggiunge il test di Kolmogorov-Smirnov, il quale misura la massima deviazione verticale esistente tra la distribuzione cumulata di frequenza delle controparti in bonis e quella dei soggetti in default (figura 3.10):



Figura 3.10: frequenza cumulata delle controparti in bonis (verde) e in default (rosso)

2. Test di Calibrazione: valuta la differenza tra le PD stimate e i tassi di default osservati per classe di rating e verifica se tale differenza abbia natura casuale o sistemica. Il modello di rating è ben calibrato se i tassi di default osservati per

ciascuna classe si discostano marginalmente dalle PD stimate per la medesima classe. I principali test statistici per la calibrazione sono:

- Test Chi-Quadro: determina se sia presente una relazione tra distribuzione osservata e quella attesa;
- Traffic light test: mette a confronto i tassi di default osservati con le PD stimate per ogni classe di rating e può essere effettuato su un portafoglio per un determinato periodo t (single period), oppure su più orizzonti temporali (multi-period);
- Test binomiale: è utilizzato per verificare se il valore di PD di una classe di rating è in linea con il tasso di default medio osservato. Tale test vuole testare l'ipotesi nulla H_0 per cui la probabilità di default prevista di una determinata classe di rating è corretta, contro l'ipotesi alternativa H_1 per cui tale probabilità non sia corretta, quindi sia sotto/sovrastimata. Si definisce test binomiale ad una coda quando viene già stabilito quale tra i due gruppi debba essere il maggiore/minore. Si definisce test binomiale a due code quando non vengono effettuate considerazioni in merito a quale gruppo debba essere maggiore o minore;
- Reliability diagram: strumento grafico che permette di mettere in relazione i tassi di default osservati con i tassi di default previsti per ogni classe. In figura 3.11 viene riportato un esempio di tale diagramma, in cui la curva blu è detta calibration curve e tanto più questa sarà prossima alla bisettrice, tanto meglio calibrato sarà il modello stimato:

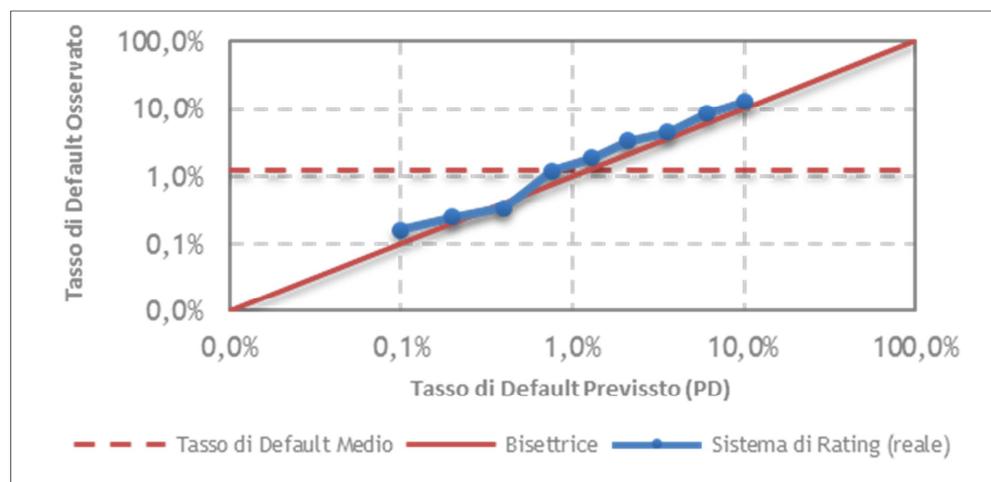


Figura 3.11: Reliability diagram

- Analisi di stabilità: verifica della volatilità del rating e individuazione di cambiamenti nella popolazione con il passare del tempo. A tale scopo sono utilizzati:
 - Population stability index (PSI)⁴⁴: effettua un confronto tra la distribuzione osservata della popolazione tra le classi di rating del modello stimato in periodi temporali differenti, misurando l'eventuale shift della distribuzione nei due periodi. Il modello è stabile se conserva il suo potere discriminante nel tempo ed in particolare per valori del PSI minori del 10%;

⁴⁴ Bilal Yurdakul, "Statistical Properties of Population Stability Index (PSI)". PhD thesis, 05 2018

- Matrici di transizione: indicano la probabilità che la controparte in una certa classe di rating rimanga nella stessa classe o si muova in un'altra. In figura 3.12 viene riportata la matrice di transizione ad un anno di Fitch:

Corporate Finance Transition Matrix: 1990-2017

(%)	AAA	AA	A	BBB	BB	B	CCC to C	D	WD*
AAA	87,87	5,45	0,25	-	-	-	-	0,12	6,31
AA	0,11	85,44	9,04	0,36	0,02	0,02	-	0,05	4,81
A	0,01	1,84	88,28	5,34	0,41	0,05	0,04	0,05	4,19
BBB	0,01	0,11	2,96	87,12	3,39	0,38	0,12	0,14	5,77
BB	-	0,03	0,11	7,45	76,33	5,89	1,14	0,70	8,34
B	-	-	0,23	0,29	7,56	75,87	4,43	2,02	9,60
CCC to C	-	-	-	0,23	1,75	17,78	48,19	21,05	10,99

*WD – Withdrawn

Figura 3.12: Fitch matrice di transizione ad un anno⁴⁵

- Herfindahl Hirschmann Index (HHI): misura il grado di concentrazione della popolazione nelle classi di rating previste dal modello. Un valore basso di questo indicatore indica che la popolazione è ben distribuita nelle classi di rating. Analiticamente corrisponde alla sommatoria dei quadrati delle esposizioni (EAD) appartenenti a una certa classe sulla somma delle medesime esposizioni al quadrato (3.42):

$$H = \frac{(\sum EAD_i^2)}{(\sum EAD_i)^2} \quad (3.42)$$

⁴⁵ Fonte: <https://www.fitchratings.com/site/re/10037356>

4 Descrizione del progetto

4.1 Introduzione

L'aumento sempre maggiore dei dati a disposizione abbinato alla disponibilità di tecniche di machine learning e software sempre più performanti, ha reso disponibili nuove opportunità che possono essere sfruttate dal Risk Management in alternativa o accostate ai metodi tradizionali di stima dei fattori di rischio basati sulla statistica tradizionale. Al contrario dei modelli tradizionali, i modelli di machine learning sono nella maggior parte non parametrici, fattore molto importante perché permette di cogliere caratteristiche e pattern di dati che non possono essere colti sottostando ai vincoli e alle assunzioni parametriche sui dati. Inoltre, queste tecniche sono state create appositamente per gestire al meglio grandi moli di dati e per facilitare l'adattamento e la ricalibrazione dovuta a un loro continuo aggiornamento. È stata dunque rapida l'applicazione di queste nuove tecniche ai problemi di credit scoring e di stima della componente PD.

Nel seguente capitolo vengono riportati lo scopo del progetto, quindi quali obiettivi si pone di raggiungere ed attraverso quali metodologie, del portafoglio su cui verrà applicato il modello e delle metodologie di stima esplorate.

4.2 Scopo del progetto

Il presente lavoro di tesi è stato basato sul progetto Credit Risk Assessment di Sella Personal Credit con la consulenza di un team dell'azienda Moxoff⁴⁶, al fine di sviluppare un modello di valutazione del rischio di credito mediante l'utilizzo di modelli e metodi di data science e introdurre un set di variabili che attualmente SPC non raccoglie e non archivia, ma che potrebbero rivelarsi utili per il processo di valutazione della clientela. L'assunto alla base di tali approcci di machine learning è che i computer possano essere in grado di imparare ed eseguire delle operazioni osservando i dati che gli vengono messi a disposizione, di conseguenza maggiore è la disponibilità di informazioni tanto maggiore sarà l'efficacia dell'apprendimento. L'applicazione di algoritmi di machine learning e di robotic process automation permettono di valorizzare al massimo il set informativo a disposizione traendone il massimo beneficio, arrivando ad acquisire un vantaggio competitivo ottimizzando i propri processi decisionali e il proprio modello operativo. In particolare, gli obiettivi che rientrano nell'ambito del progetto sono:

- Esplorare le tecniche di machine learning che possono essere utilizzate per l'analisi dei dati;
- Approfondire le tecniche di machine learning che si rivelano essere più appropriate per la valutazione del rischio, che quindi producono dei risultati che hanno maggior potere previsionale;
- Ampliare il set informativo aggiungendo dei dati che apportano un maggior potere predittivo.

⁴⁶ Link al sito: <https://www.moxoff.com/>

Il punto cardine su cui si basa lo studio è quello di ridurre le interrogazioni ai credit bureau, pertanto il modello sarà sviluppato a partire dai soli dati posseduti da SPC non arricchiti con altri esterni.

Il progetto si pone inoltre lo scopo di minimizzare l'interazione manuale da parte del valutatore, al fine di ridurre valutazioni soggettive e l'introduzione di possibili errori. Il modello sviluppato sarà altresì in grado di percepire modifiche significative nella popolazione richiedente e quindi proporre delle modifiche ai regressori utilizzati nel modello di stima. Questo meccanismo di auto adattamento avviene mediante un riaddestramento effettuato automaticamente.

4.3 Definizione dei clienti

In altre parole, l'obiettivo è quello di distinguere i clienti in fase di valutazione tra buoni e cattivi pagatori sulla base delle informazioni a disposizione e dei comportamenti osservati sul portafoglio pregresso.

La distinzione tra buoni e cattivi pagatori adottata nell'ambito del progetto riprende quella utilizzata da Sella Personal Credit, basata sulla serie storica degli insoluti del prestito di una pratica di un cliente. Il ciclo di vita del prestito di un cliente può essere espresso come una successione di N numeri, in cui N è il numero di rate previste dal piano di ammortamento del prestito. Ogni valore della successione fa riferimento ad una specifica rata ed è valorizzata in base al numero di insoluti presenti in quel preciso momento di vita del prestito, andando a definire la serie storica degli insoluti. Al fine di rendere più chiaro il concetto viene riportato di seguito un caso di esempio.

Sia $\{0,0,1,1,2,2,1,2,\dots\}$ la successione relativa alla serie storica degli insoluti di un cliente, si può osservare che:

1. Nei primi due periodi il cliente paga correttamente le rate, quindi vengono valorizzati a 0 i primi due campi;
2. Al terzo periodo il cliente non paga la rata, per cui si valorizza a 1 il terzo campo della serie;
3. Al quarto periodo viene pagata la rata di quel mese, ma non viene sanato l'insoluto precedente, per cui rimarrà valorizzato 1;
4. Al quinto periodo si verifica nuovamente un insoluto, per cui il campo viene incrementato a 2 tenendo conto dell'insoluto del terzo periodo e di quello corrente;
5. Al sesto periodo viene pagata correttamente la rata relativa a quel mese, ma non gli insoluti precedenti;
6. Al settimo periodo viene pagata correttamente la rata e viene sanato uno dei due insoluti precedenti, riportando il valore a 1;
7. All'ottavo periodo si verifica nuovamente un insoluto, riportando il valore a 2;
8. E così via.

La classificazione di un cliente viene fatta sulla base del numero massimo di insoluti che registra in tutto il ciclo di vita del prestito, quindi, in virtù di quanto suddetto, in base al valore più alto che viene riportato nella serie storica degli insoluti. Secondo la definizione di SPC viene indicato come:

- Buon pagatore: un cliente per cui non si registra mai più di una rata consecutiva impagata durante il ciclo di vita del prestito;
- Cattivo pagatore: un cliente che nell'arco della vita del prestito raggiunge almeno tre rate insolte consecutive, in altre parole se nel suo storico insoluti è presente almeno

una volta un valore superiore o uguale a tre. Tale classificazione viene assegnata indipendentemente dal fatto che alla fine del ciclo di vita del prestito il cliente abbia sanato tutti gli insoluti e abbia concluso correttamente il piano di ammortamento;

- Indeterminato: un cliente che nell'arco della vita del prestito registra al massimo due insoluti consecutivi.

In ambito del progetto si semplifica la definizione riportando due sole classificazioni, distinguendo solo buoni e cattivi pagatori secondo quanto segue:

- Buon pagatore: un cliente che non supera mai più di due insoluti consecutivi nel ciclo di vita del prestito. In questa definizione rientrano quindi quei clienti che raggiungono le due rate impagate, ma che riescono a regolarizzare i loro pagamenti entro il termine del prestito;
- Cattivo pagatore: rimane la stessa precedente, ossia un cliente il quale raggiunge le tre rate insolte consecutive nella serie storica degli insoluti.

È importante considerare, ad ogni modo, che gli indeterminati sono una casistica del tutto residuale, pertanto il loro accorpamento nel cluster dei buoni pagatori non pregiudica in alcun modo le caratteristiche complessive del campione.

Nelle tabelle 4.1 seguenti si riporta il confronto tra criteri di classificazione adottati da Sella Personal Credit e da Moxoff.

Tabelle 4.1: Classificazione clienti di Sella Personal Credit e di Moxoff

Classificazione cliente SPC	Numero massimo di insoluti	Classificazione cliente Moxoff	Numero massimo di insoluti
Buon pagatore	≤ 1	Buon pagatore	≤ 2
Indeterminato	≤ 2	Cattivo pagatore	≥ 3
Cattivo pagatore	≥ 3		

4.4 Metodologie di misurazione del rischio di credito

Come detto nell'introduzione negli ultimi anni le tecniche di machine learning hanno trovato ampia applicazione ai problemi di credit scoring e stima della componente PD. Nella prima fase esplorativa del progetto sono state valutate diverse tecniche di machine learning sulla base di quanto riscontrato nella letteratura.

4.4.1 Reti Neurali (Artificial Neural Network - ANN)

Uno di questi primi approcci è la rete neurale, un modello che mette in comunicazione le variabili esplicative con quella target tramite diversi strati di variabili latenti, detti strati nascosti, costituite da combinazioni delle variabili di input trasformate. Le ANN permettono di modellizzare sistemi senza la necessità di dover effettuare aprioristiche assunzioni sul tipo di funzioni da usare, cosa che invece è necessaria per la maggior parte degli approcci statistici tradizionali. La struttura delle ANN può essere rappresentata mediante un grafo in cui i nodi rappresentano i neuroni, i quali vengono disposti su più livelli (layers) ed interconnessi tra loro in un'unica direzione da uno strato all'altro. Di seguito viene schematizzata l'architettura di una ANN, figura 4.2:

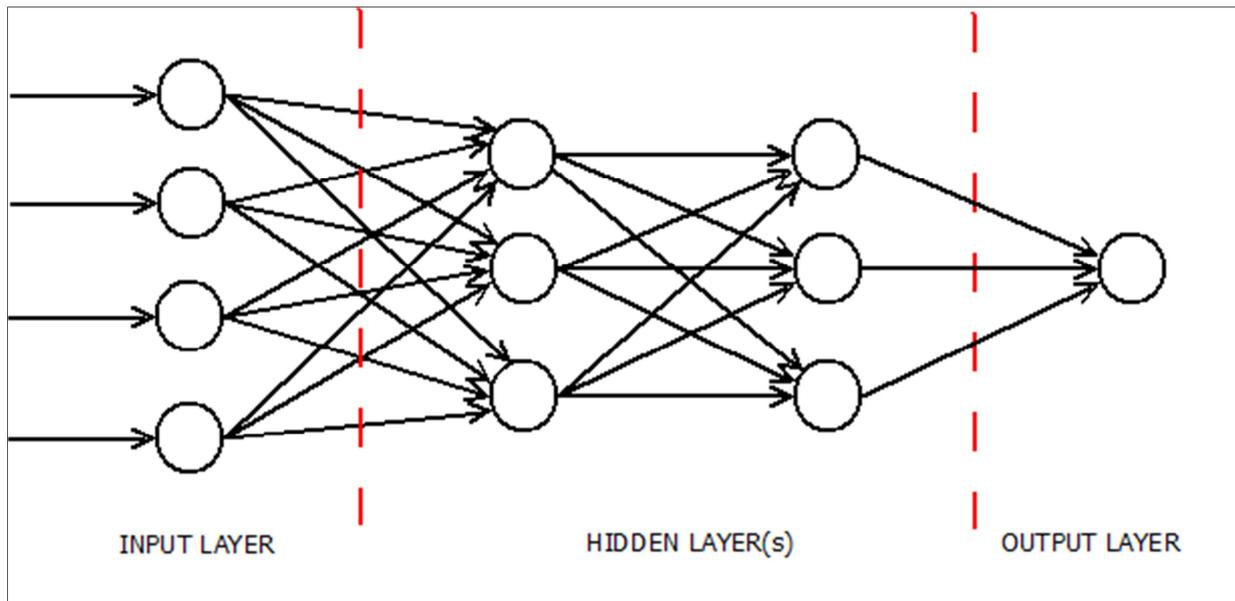


Figura 4.2: Architettura di una rete neurale

Tale architettura è definita Multi-Layered perceptron network (MLP) e presenta uno strato d'ingresso del segnale (input layer), uno o più strati nascosti che processano l'informazione (hidden layers) e un layer di uscita (output layer), che rende fruibile l'informazione processata, definita risposta al segnale d'ingresso. I neuroni presenti nell'input layer agiscono come dei buffer per distribuire i segnali d'ingresso (ovvero le variabili indipendenti del problema, x_i , $i = 1, 2, \dots, n$) verso i neuroni presenti nell'hidden layer (anche detti perceptroni). Ciascun perceptrone j elabora i suoi segnali d'ingresso x_i sommandoli, dopo averli moltiplicati per un peso associato a ciascuna connessione entrante in esso, $\omega_{j,i}$. Il neurone elabora il suo output y_j applicando una funzione f sul risultato di tale somma, secondo (4.1):

$$y_j = f\left(\sum_{i=1}^n \omega_{j,i} \cdot x_i\right) \quad (4.1)$$

La funzione di attivazione f può tipicamente essere costituita da un gradino, una lineare, una sigmoide o una tangente iperbolica.

Il segnale elaborato dai neuroni disposti nell'output layer è simile a quello dei perceptroni dello strato nascosto, con l'unica differenza che f è una funzione lineare, poiché essa deve rendere l'informazione direttamente fruibile verso l'esterno. Le risposte ai segnali ottenute vengono confrontate con il dato osservato, la cui differenza costituisce l'errore di apprendimento della rete, che deve essere minimizzato mediante l'algoritmo di backpropagation. Lo scopo di tale processo è quello di modificare i valori dei pesi associati alle connessioni tra neuroni, col fine di ridurre al di sotto di una prefissata soglia di tolleranza, definita performance (ad esempio mean squared error – MSE o la sum squared error – SSE), il gap generato tra il modello ed i dati osservati. Tale processo iterativo di aggiustamento dei pesi viene chiamato addestramento della rete ed il campione sul quale l'ANN viene calibrata è detto training set. L'algoritmo di backpropagation fornisce la variazione del peso da applicare alla connessione tra un neurone di partenza i -esimo ed uno di arrivo j -esimo secondo la formula (4.2):

$$\Delta\omega_{i,j} = \eta\delta_j x_i \quad (4.2)$$

Dove η è un parametro definito learning rate e δ_j è un fattore che dipende dalla funzione di attivazione del neurone di destinazione j -esimo. Tale fattore dipende dal contributo del j -esimo neurone al calcolo del gradiente della funzione obiettivo e cambia a seconda che il neurone stesso appartenga ad uno strato nascosto o a quello di output. Il processo parte dall'output layer e va a ritroso nei diversi livelli presenti nel grafo (backpropagation). Le ANN continuano il processo iterativo di aggiornamento dei pesi sin tanto che un criterio di stop viene incontrato dall'algoritmo. Tipici criteri di stop sono il raggiungimento di un numero massimo di iterazioni, il raggiungimento di una soglia di tolleranza tra i valori forniti dal modello e quelli osservati (target performance SSE o MSE), il raggiungimento di un tempo massimo entro il quale i pesi devono aggiornarsi. Il processo appena descritto costituisce la fase di addestramento della rete e viene definito apprendimento supervisionato, in quanto viene effettuato confrontando gli output con un vettore target contenente i dati osservati. Tale processo è schematizzato nella figura 4.3 seguente:

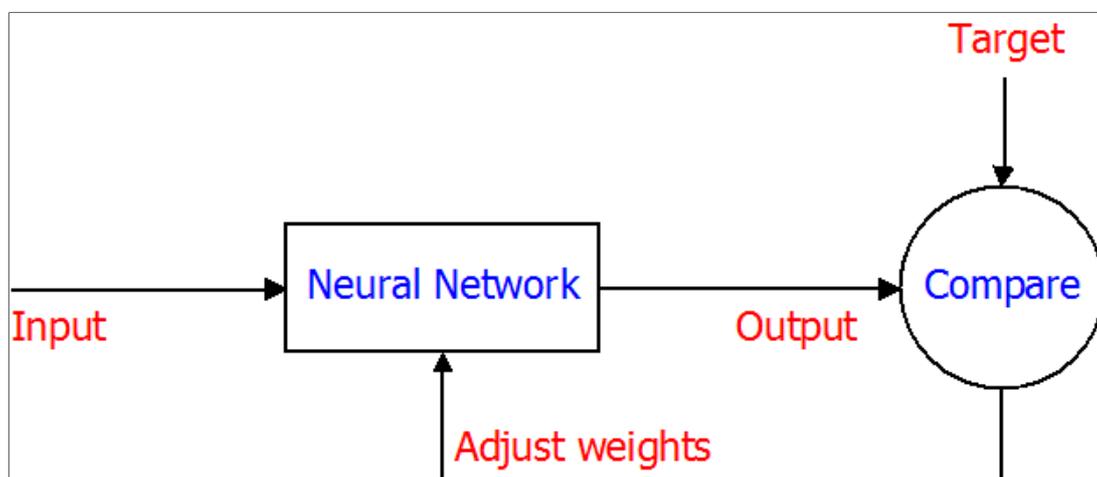


Figura 4.3: Processo di apprendimento supervisionato di una rete neurale

Tuttavia, questo processo comporta un rischio di overfitting dei dati: a mano a mano che il processo di apprendimento avanza, la differenza tra il target e l'output si riduce. Come conseguenza, può verificarsi che la relazione stimata permetta un fitting perfetto dei dati campionati, ma che non sia abbastanza "generale", cioè che si generi il rischio di ottenere risultati poveri in termini di fitting quando la rete è utilizzata per processare dati nuovi. Un modo per ovviare a tale problema è suddividere il campione di training in tre sottocampioni: set di addestramento, di validazione e di test. La procedura di minimizzazione avviene sul set di training, mentre il set di validazione è usato in maniera indiretta, cioè il MSE è monitorato anche su di esso ma la minimizzazione è effettuata solo sul set di addestramento. Normalmente, all'inizio della fase di training, l'errore diminuisce sia sul campione di addestramento, sia su quello di validazione, ma quando la rete inizia ad andare in overfitting, l'errore sul campione di validazione inizia a crescere. Da questo punto in poi, l'algoritmo di ottimizzazione aggiornerebbe i pesi solo per realizzare una corrispondenza esatta tra il campione di output e il target corrispondente, cioè si avrebbe un overfitting dei dati. Per evitare tale fenomeno, i pesi salvati sono quelli che minimizzano il MSE sul campione di validazione, anziché su quello di training. Infine, il campione di test è utilizzato per misurare la performance della rete su un set di dati completamente nuovi. La rete addestrata è applicata a nuovi input, cioè sono processati nuovi ingressi utilizzando i pesi

calcolati nella fase di apprendimento. Gli output così ottenuti sono poi confrontati con i target per calcolare una misura di performance.

La valenza delle reti neurali riportata negli studi⁴⁷ effettuati ne evidenzia la capacità di adattarsi ai dati grazie a una vasta gamma di impostazioni, come il numero di strati nascosti e di nodi all'interno di ogni singolo strato ed i valori dei pesi dei parametri con cui inizializzare la macchina da addestrare. Altre ricerche⁴⁸ hanno dimostrato come le reti neurali siano in grado di produrre stime di PD molto più accurate rispetto alla best practice di mercato, la regressione logistica. Nella figura 4.4 seguente viene riportato un generico modello di rete neurale in cui vi sono 24 neuroni nell'input layer, corrispondenti alle 24 variabili di input, un hidden layer (strato nascosto) composto da h neuroni ed un unico output layer che restituisce come esito se il cliente è un buono (Good) o cattivo (Bad) pagatore:

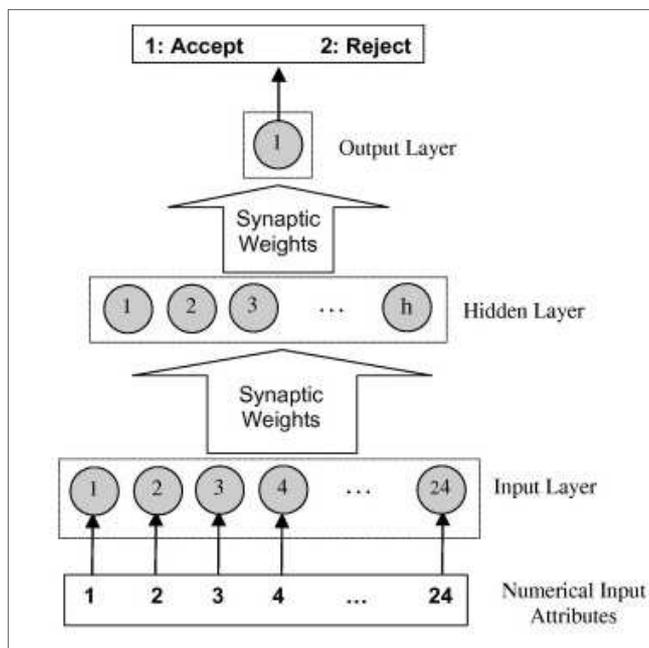


Figura 4.4: Esempio di rete neurale per la valutazione del rischio di credito

4.4.2 Alberi decisionali (CART)

Un altro modello tipico del machine learning sono gli alberi decisionali (CART – Classification And Regression Tree), con questa tecnica si è in grado di ottenere stime accurate tramite divisioni consecutive e dipendenti dello spazio dei dati basate su valori soglia delle singole variabili esplicative. In ognuna di queste divisioni vengono individuate porzioni di dati accomunate da valori della variabile target molto simili e stimati dall' algoritmo. Questo metodo prende il suo nome dalla rappresentazione grafica delle regole che lo compongono, in quanto le divisioni dello spazio operate dal modello possono appunto essere rappresentate da un albero binario le cui singole componenti, chiamate nodi, incorporano i valori soglia dai quali viene originata la scissione delle osservazioni. Convenzionalmente la prima scissione viene chiamata radice, mentre le terminazioni dell'albero rappresentano i sottoinsiemi di dati e vengono chiamate foglie.

⁴⁷ Khashman A., "Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes", 2010

⁴⁸ Lessmann S., Baesens B., Seow H. V., Thomas L.C., "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research", European Journal of Operational Research 247: 124-136, 2015

I CART, nonostante la loro struttura semplice, risultano essere uno strumento particolarmente potente, oltre che intuitivo e di facile interpretazione, in grado di ottenere elevate performance predittive. Si riporta nella figura 4.5 un esempio di modello CART basato su due variabili indipendenti che ha come esito se il cliente è un buono (Good) o un cattivo (Bad) pagatore⁴⁹:

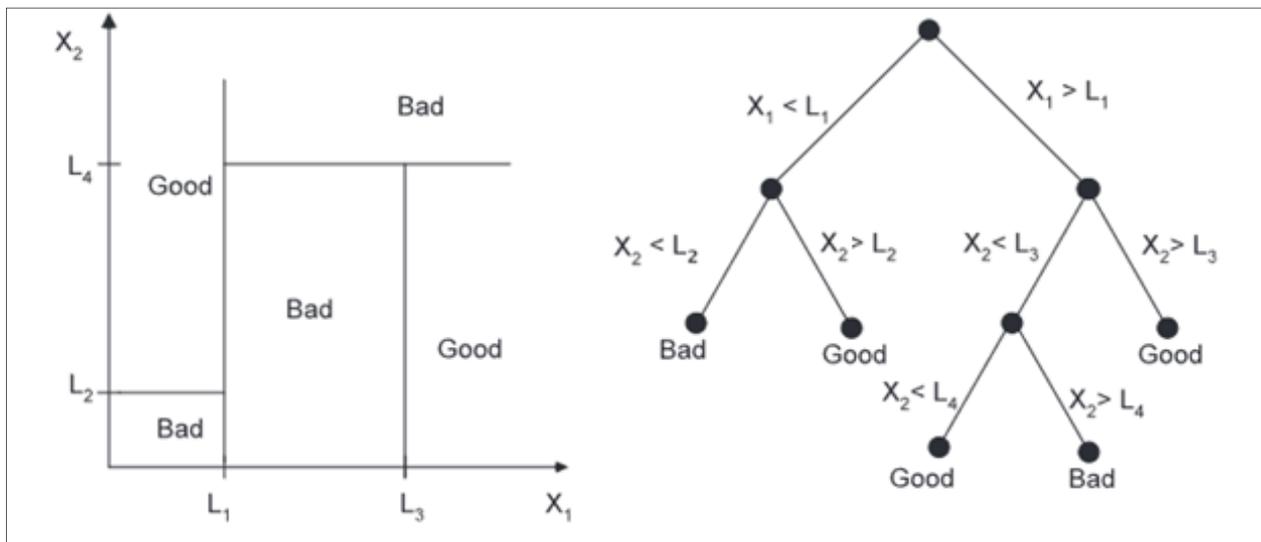


Figura 4.5: Esempio di modello CART con due variabili indipendenti (X_1 e X_2) e due possibili risultati (Good o Bad)

Nella figura sono rappresentati a sinistra la scissione a scalini in base al valore delle variabili X_1 e X_2 , mentre a destra la formalizzazione ad albero della funzione.

Il modello, per essere costruito, necessita di un set di dati di addestramento: deve essere quindi composto da record in cui la variabile target è nota. L'algoritmo del CART, dunque, sviluppa l'albero tramite una procedura step-by-step che comporta divisioni binarie e sequenziali dello spazio dei dati. Al primo passaggio, viene considerata per ogni variabile di soglia che permette una divisione in grado di minimizzare la devianza tra i dati appartenenti alla stessa divisione, ma che al contempo massimizzi quella tra i due insiemi di dati formati. Questi vincoli vengono incorporati nella formula della funzione obiettivo (4.3):

$$D = \sum_{i=1}^K \left\{ \sum_{j \in S_i} (y_j - \hat{\beta}_i)^2 \right\} = \sum_i D_i, \quad (4.3)$$

Dove y_i e $\hat{\beta}_i$ sono rispettivamente i valori della variabile target e il parametro ad essi associato presenti all'interno di uno dei K sottospazio dei dati S_i . A questo punto per entrambe le nuove divisioni viene ripetuto il processo precedente, cercando tra i dati dei sottoinsiemi il valore di soglia per la variabile in grado di ottenere la divisione ottimale. Questo processo viene iterato fino a quando non si verificano determinate condizioni che ne determinano l'arresto. Una di queste cause può essere, ad esempio, la creazione di un sottospazio di dati aventi la stessa categoria della variabile target o che possiede gli stessi valori, o ancora quando una qualsiasi suddivisione non comporterebbe un aumento della funzione obiettivo significativo o, più semplicemente, nel caso in cui la divisione creata contenga una singola osservazione. A questo punto, l'algoritmo otterrà un albero

⁴⁹ Khandani A. E., Kim J., LO A.W., "Consumer credit-risk models via machine-learning algorithms", Journal of Banking & Finance, 2010

particolarmente fitto e complesso, composto da molti rami e foglie; il risultato sarà scarsamente interpretabile, per l'elevato numero di tagli e per la tendenza al sovradattamento dei dati (overfitting). Quest'ultimo difetto è molto rilevante, poiché influenza direttamente le stime future e, pertanto, la bontà del modello. In questo caso è necessario ridurre l'albero tramite una procedura automatica chiamata "potatura": una tecnica che, partendo dal modello completamente sviluppato, elimina sequenzialmente i rami non utili ai fini della stima o con la minore carica informativa. Definendo la funzione di perdita come segue (4.4):

$$C_\alpha(K) = \sum_{i=1}^K D_i + \alpha K, \quad (4.4)$$

Dove K è la dimensione dell'albero considerato in ogni singolo passo e il parametro α associato al costo computazionale del modello, verrà ad ogni passo rimossa la foglia la cui eliminazione comporta il minore aumento della funzione obiettivo $\sum_{i=1}^K D_i$. La procedura continua fino a quando il valore di $C_\alpha(K)$ sarà stabilizzato.

Grazie alla sua alta semplicità logica che permette di comunicare facilmente le regole alla base della sua struttura, mettendo in evidenza quali sono i principali driver implicati nella stima, l'algoritmo degli alberi decisionali risulta uno dei modelli più informativi. Correlato a questo fatto, gli alberi risultano essere un ottimo metodo automatico di riduzione della dimensionalità dei dati, selezionando soltanto le variabili più importanti ai fini dell'approssimazione dei dati. Un altro vantaggio di questo modello è la sua ridotta complessità computazionale anche quando la mole di osservazioni e il numero di variabili è molto alto, permettendo una selezione automatica delle variabili migliori e consentendo di superare il problema di dataset incompleti.

4.4.3 Tecniche Random Forest

Una diretta evoluzione dei modelli CART sono le tecniche Random Forest. Esse derivano da una combinazione di più modelli ad albero costruiti in modo diverso e vengono sviluppati utilizzando il campionamento bootstrap⁵⁰ di dati senza potatura. Questa aggregazione può migliorare la stima ottenuta da ogni singolo CART sia in termini di accuratezza sia in termini di stabilità. Le tecniche di Random Forest prevedono i seguenti passi⁵¹:

1. Si consideri un set di dati X , definito training set, costituito da una variabile di risposta e da molte variabili predittive in N campioni. Il numero complessivo delle variabili predittive è M , con M che è sostanzialmente maggiore di N ;
2. Si costituisce un campione bootstrap X^* costituito da N campioni dal set originale X . In media, un terzo di tutti i campioni viene omesso andando a costituire il campione Out-Of-Bag (OOB);
3. Per la "crescita" di ogni albero viene utilizzato un campione X^* , dividendo in modo ricorsivo i dati in sottoinsiemi distinti, in modo che ogni nodo genitore porti a due nodi figli. Per la divisione dei dati si seleziona una regola di divisione sulla base di un criterio su una singola feature (insieme di variabili). Per ogni nodo la scelta della

⁵⁰ Introdotta da Efron nel 1979, è una tecnica statistica di ricampionamento con reimmissione. Efron, B. and Tibshirani, R., "An Introduction to the Bootstrap". Chapman and Hall, New York, London, 1993

⁵¹ Breiman L., "Random Forests", Machine Learning Vol. 45:5-32, 2001

feature migliore su cui effettuare la divisione non è fatta sull'intero insieme delle variabili, ma su un sottoinsieme $mtry$ ottenuto dalla (4.5):

$$mtry = \lceil \sqrt{M} \rceil \quad (4.5)$$

dove $\lceil \cdot \rceil$ denota l'approssimazione al numero intero più grande. In assenza di questo accorgimento, detto feature bagging, molti alberi sceglierebbero con elevata probabilità le stesse variabili, ossia le più discriminanti;

4. L'albero t cresce fino alla sua massima estensione senza effettuare il processo di "potatura";
5. Si ripetono gli step da 1 a 4 fino a far crescere un numero specifico di alberi;
6. Infine, grazie al bagging, le prestazioni possono essere stimate mediante la tecnica OOB che non richiede validation set separato, in quanto ogni osservazione viene prevista dagli alberi per i quali essa è stata scartata, ossia fa parte del set Out-Of-Bag. L'accuratezza previsionale viene stimata sottraendo la frazione di errore OOB, ossia la differenza tra valore osservato e previsto, dal suo massimo, che è 1.

In letteratura le Random Forest ottengono risultati estremamente consistenti nelle stime probabilistiche⁵² e sono spesso state oggetto di confronto con i metodi parametrici classici testandoli sulla statistica tradizionale. Inoltre, come nel caso dei CART, le Random Forest sono molto utili per superare problemi di alta dimensionalità dei dati, potendo ottenere un sottoinsieme di variabili determinato da una particolare misura di importanza.

4.5 Osservazioni

I modelli presenti in letteratura sono molti e diversi e tutti hanno particolari caratteristiche che li rendono unici, così come difetti che ne limitano le performance e l'applicazione; ad esempio, nonostante i CART siano un metodo molto accurato e intuitivo di stima in grado anche di selezionare automaticamente le variabili più importanti, trovano parecchie difficoltà ad approssimare funzioni matematiche con dei picchi irregolari. Allo stesso modo, le Reti Neurali ottengono altissime prestazioni ma possono raggiungere risultati molto diversi sullo stesso dataset cambiando anche di poco le impostazioni iniziali con cui viene addestrata la macchina. Inoltre, altri modelli più complessi possono richiedere un'eccessiva calibrazione che spesso porta a problemi di sovradattamento.

Molto importante è comunque sottolineare che, al di là delle prestazioni ottenute, il modello CART risulta essere molto più interpretabile da un punto di vista economico, in contrapposizione con Random Forest e Reti Neurali che non permettono una buona comprensione dei risultati e dei legami tra le variabili di input e quella di output.

⁵² Malley J., Kruppa J., Dasgupta A., Malley K., Ziegler A., Probability machines. Consistent probability estimation using nonparametric learning machines. *Methods Inf. Med* 51 (1): 74–81, 2012

5 La metodologia adottata

5.1 Introduzione

Nel seguente capitolo viene descritta la metodologia adottata nelle fasi di sviluppo del modello di machine learning.

Il modello è stato inizialmente sviluppato sui soli clienti accettati, per i quali è osservabile la performance (buono o cattivo pagatore).

Lo sviluppo del modello di machine learning si articola nelle seguenti fasi:

- Esplorazione ed analisi preliminare del dataset;
- Definizione del pattern;
- Preelaborazione dei dati;
- Definizione del modello predittivo;
- Training;
- Validazione.

A questo punto è possibile definire i punteggi da adottare per le soglie e la conseguente definizione della scala di rating.

Le attività di sviluppo sono state effettuate su ambiente PyCharm in linguaggio Python.

5.2 Portafoglio su cui si applica il modello

Le analisi sono state svolte su un campione estratto per i tre macro-prodotti principali di SPC, ossia consumo (CO), carte di credito (CA) e cessione del quinto (CQ). L'analisi si è concentrata sulle pratiche aperte a partire dal 1° gennaio 2011 fino al 30 agosto 2018, poiché sono cambiate le analisi svolte per la selezione dei clienti e di conseguenza i dati raccolti sono diversi, allora per avere un dataset coerente si è effettuata una prima restrizione sui dati in base alla data di raccolta. La tabella 5.1 riporta il numero di pratiche aperte nel periodo di tempo indicato e divise per macro-prodotto:

Tabella 5.1: Numero di pratiche per macro-prodotto

Prodotto	Numero di pratiche
CO	2,582,166
CA	581,951
CQ	36,623

Il campione di dati selezionato per effettuare le analisi è stato il consumo, in quanto il più ricco sia in termini di numerosità di record, sia in termini di informazioni contenute per ogni pratica. Nella prima fase di analisi (esplorazione dei dati e definizione del modello predittivo), la prima selezione è stata effettuata sulle sole pratiche chiuse, quindi quelle che presentavano uno dei seguenti stati della pratica:

1. Chiusura regolare;
2. Chiusura anticipata: ossia il prestito si è concluso prima della reale scadenza contrattuale;

3. **Messa in mora:** rientrano in tale gestione le posizioni per cui il piano di ammortamento è terminato, ma per cui sono presenti almeno 7 insolvenze oppure risulta una rata insoluta da più di sette mesi;
4. **Decadenza del beneficio del termine:** rientrano in tale gestione le posizioni per cui il piano di ammortamento non è ancora terminato, ma si sono presentate almeno 7 rate insolte non necessariamente consecutive;
5. **Passaggio a perdita:** rientrano in tale gestione le pratiche per cui si ritiene irrecuperabile il credito e non è possibile eseguire la cessione del credito.

Lo stato è un dato presente nel set informativo di ogni pratica, quindi non è stato necessario ricavarlo.

A seguito di questa selezione il set di pratiche CO ottenute è di 1,653,000 record. Una seconda selezione, come già detto, è stata fatta sulla data di richiesta del finanziamento, scartando le pratiche precedenti al 1° gennaio 2011, ottenendo 735,394 pratiche. Per semplicità le pratiche con dati missing sono state scartate, ottenendo un campione di 732,566 pratiche.

Per questo campione di pratiche è noto l'andamento delle insolvenze presentate e di conseguenza se il cliente è un buono o cattivo pagatore.

5.3 Modello predittivo

Il modello predittivo sviluppato appartiene alla categoria dei modelli di machine learning mix di Random Forest e Neural Network, il quale assegna un punteggio di score ad ogni cliente. Questa tipologia di modello è intrinsecamente multivariata, ossia lo score assegnato ad una pratica è ottenuto sulla base di tutto il set di informazioni disponibile, definendo automaticamente un peso per ogni informazione.

Nella fase di training il modello apprende i parametri necessari per la classificazione di un dato. La fase di validazione consiste nel valutare i dati ottenuti dal modello previsionale mediante un confronto con i dati reali. Il framework tipico per i processi di machine learning per le fasi di training e validazione due campioni distinti:

- I. **Training set (campione di sviluppo):** che contiene l'80% dei dati, su cui viene fatto l'addestramento del modello;
- II. **Validation set (campione di validazione):** che contiene il restante 20% dei dati, su cui viene effettuata la validazione del modello ottenuto.

I due campioni devono essere disgiunti, in modo da effettuare la fase di validazione su uno scenario reale con dati totalmente nuovi. Se il modello venisse testato sui dati contenuti nel test di allenamento, su cui sono regolati i parametri del modello, i risultati della validazione risulterebbero distorti e si incorrerebbe nel problema dell'overfitting.

Il pattern riportato è stato definito anche in base ai risultati ottenuti dall'esecuzione del modello, che ha dimostrato come alcune variabili che nelle analisi dimostravano un maggior potere discriminante non apportavano un significativo aumento delle performance.

5.3.1 Analisi delle variabili

In prima istanza si effettua un'analisi del set informativo a disposizione, con lo scopo di andare ad individuare le variabili che apportano il maggior contenuto informativo in termini di classificazione dei clienti. Pertanto, su tutto il set di variabili si effettuano analisi per ogni valore, oppure per intervalli di valori, relative al numero ed alla percentuale di buoni o cattivi pagatori.

Altre variabili di interesse possono essere derivate dal set a disposizione mediante, ad esempio, aggregazione dei dati od operazioni tra diversi campi.

Sulla base delle variabili individuate si definiscono differenti combinazioni che andranno a costituire i pattern forniti in input al modello. Tale processo, seppur possa essere parzialmente automatizzato, risulta molto oneroso in termini di tempo, per cui le combinazioni utilizzate sono state limitate ad un certo numero di variabili.

5.3.1.1 Definizione del pattern

I dati in input ad un modello di machine learning consistono in un vettore di numeri. A questo proposito, il primo passo è quello di codificare tutte le informazioni disponibili per una pratica in un pattern numerico. Il modello di machine learning assegna un punteggio ad ogni pattern in base ai risultati di performance ottenuti. Il risultato di questa fase sarà dunque una classificazione dei pattern dal punteggio più alto a quello più basso, che verranno utilizzati nelle successive fasi.

5.3.1.2 Preelaborazione dei dati

La preelaborazione dei dati avviene come segue:

- I dati categoriali sono codificati con la descrizione one-hot, la quale consiste nella rappresentazione tramite un gruppo di bit corrispondenti al numero delle categorie tutti posti a 0, tranne quello relativo alla categoria di riferimento che viene posto a 1, o dummy variables. DI seguito si riporta in tabella 5.2 Un esempio di one-hot encoding;

Tipo Prodotto	One-hot
PA	100
PF	010
PP	001

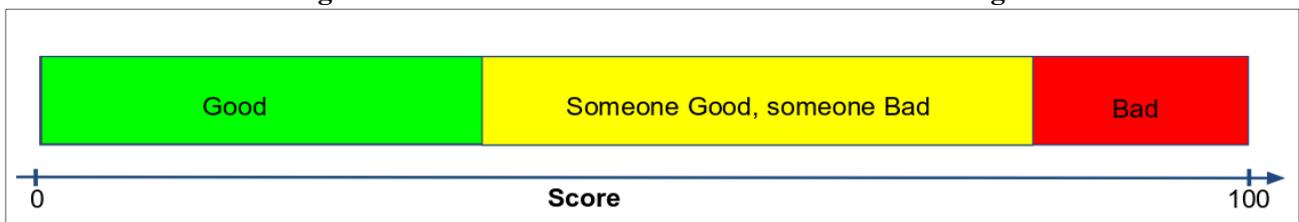
Tabella 5.2: One-hot encoding

- I dati numerici sono oggetti a normalizzazione, ovvero ridimensionati a una distribuzione centrata in 0 e varianza uguale a 1;
- I dati missing vengono trattati a seconda del tipo di variabile:
 - Per i dati categoriali si pone il valore codificato a 0;
 - Per i dati numerici si pone uguale alla media degli altri valori per quella variabile.

5.3.2 Modello di regressione e definizione della soglia

Il modello predittivo sviluppato è un modello di regressione che assegna a ciascuna cliente uno score compreso tra 0 e 1, che viene poi moltiplicato per 100 in modo da ottenere una scala da 0 a 100. Tale score assegnato è attinente alla probabilità di default, che in questa trattazione coincide con il concetto di cattivo pagatore; più alto è lo score, maggiore è la probabilità di default del cliente. Ordinando i clienti rispetto allo score assegnato dal modello, se il pattern di addestramento utilizzato è significativo, si otterrà una distribuzione simile a quella in figura 5.3, in cui nella parte dell'intervallo più prossima allo 0 (zona verde) sono classificati i buoni pagatori, in quella più prossima al 100 (zona rossa) vengono classificati i cattivi pagatori e in mezzo (zona gialla) si avrà un'area indeterminata, in cui è presente un mix di buoni e cattivi pagatori.

Figura 5.3: Ordinamento dei clienti sulla scala di scoring



Da questo momento il modello è in grado di identificare i buoni e i cattivi pagatori in base alla posizione dello score rispetto alla soglia; di seguito un cattivo pagatore sarà indicato come positivo, mentre un buon pagatore sarà indicato come negativo. Sulla base di questa definizione si identificano 4 scenari possibili:

1. Vero negativo: è un buon pagatore classificato correttamente;
2. Falso negativo: è un cattivo pagatore classificato come buono dal modello;
3. Falso positivo: è un buon pagatore classificato come cattivo dal modello;
4. Vero positivo: è un cattivo pagatore classificato correttamente.

La distinzione tra buoni e cattivi pagatori viene effettuata mediante l'applicazione di una soglia di cut-off alla scala di scoring individuata. Il livello di tale soglia può essere scelto arbitrariamente ed in base a dove essa viene posta si possono ottenere diverse strategie commerciali a seconda della minore o maggiore propensione al rischio.

5.3.2.1 Strategia "low risk"

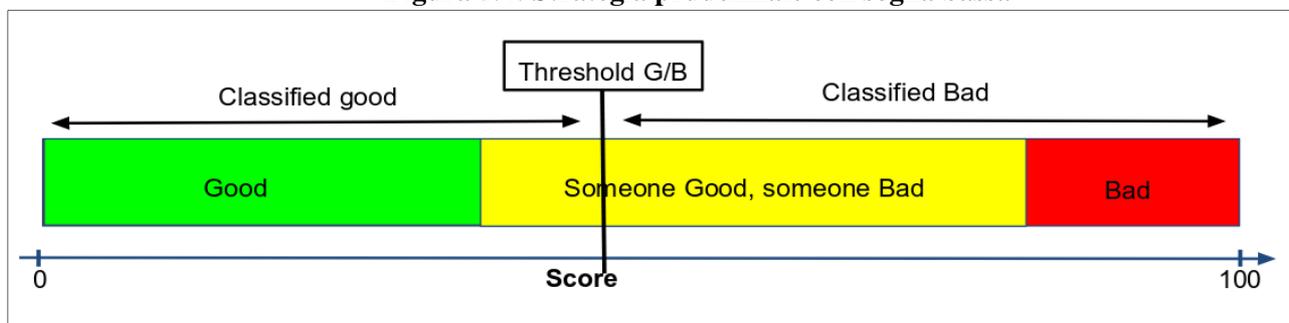
L'obiettivo di questa strategia è di identificare un set di clienti per cui la richiesta di contribuzione con dati dei credit bureau può essere evitata con un basso rischio, quindi applicando una soglia bassa che garantisce di ridurre il numero di cattivi pagatori classificati come buoni.

La variabile di output su cui è stato addestrato il modello deriva dal numero di insoluti consecutivi osservati nel ciclo di vita di un prestito:

- 0 se il numero di insoluti è 0;
- 0.25 se il numero di insoluti è 1;
- 0.5 se il numero di insoluti è 2;
- 0.75 se il numero di insoluti è 3;
- 1 se vi sono stati 4 o più insoluti.

Con una soglia bassa (figura 5.4), quindi più spostata verso i buoni pagatori, un numero elevato di pagatori è classificato come "cattivo", quelli alla destra della soglia. In questo caso il numero di clienti accettati è piccolo, ma molto pulito: c'è un piccolo numero di clienti classificati buoni che risultano essere cattivi (basso numero di falsi negativi), ma un alto numero di buoni pagatori classificati come cattivi (alto numero di falsi positivi).

Figura 5.4: Strategia prudentiale con soglia bassa



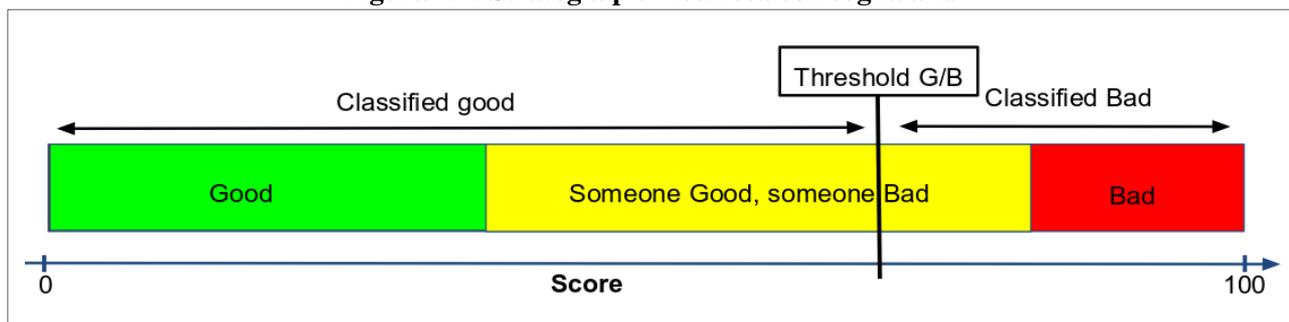
5.3.2.2 Strategia "risk score"

L'obiettivo della strategia "risk score" è quello di identificare i cattivi pagatori riducendo il numero di errori. La variabile di output su cui è stato addestrato il modello deriva dal numero di insoluti consecutivi osservati nel ciclo di vita di un prestito:

- 0 se il numero di insoluti è minore di 3;
- 1 se il numero di insoluti è maggiore o uguale a 3.

Con una soglia alta (figura 5.5) un elevato numero di pagatori è classificato come buono. In questo caso il numero di clienti accettati è elevato, ma molti clienti classificati come buoni potrebbero rivelarsi cattivi (basso numero di falsi positivi, ma elevato numero di falsi negativi).

Figura 5.5: Strategia più rischiosa con soglia alta

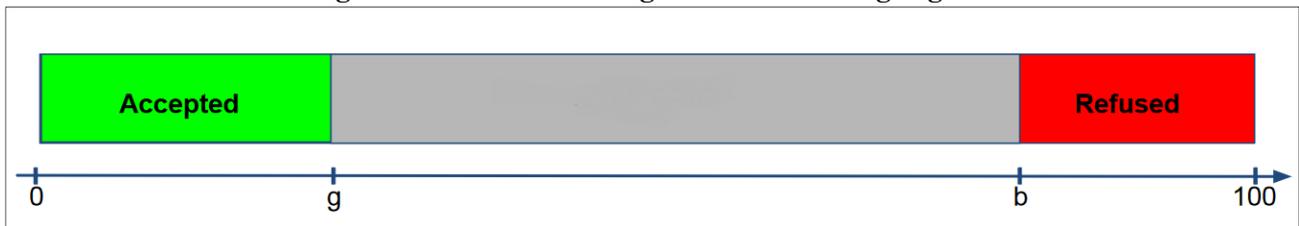


La prima strategia viene adottata quando si vuole assumere un atteggiamento più prudentiale, ossia accettare solo quei clienti che hanno una probabilità di default molto bassa. Invece, la seconda strategia viene adottata quando si vuole assumere una strategia più aggressiva e si è più propensi al rischio, accettando anche dei clienti con una probabilità di default media o alta.

5.3.2.3 Strategia con 2 soglie

Al fine di ottenere una distinzione con un livello di accuratezza maggiore tra i buoni e cattivi pagatori è possibile individuare sulla scala di rating due soglie g e b (figura 5.6), rispettivamente di accettazione e di rifiuto, la prima indicativa di quei clienti che possono essere classificati come buoni pagatori con un alto livello di confidenza, a sinistra della soglia g , e la seconda per i cattivi pagatori, a destra della soglia b .

Figura 5.6: Scala di scoring con indicate le soglie g e b



La variabile di output su cui viene allenato il modello è una classificazione binaria dei clienti in buoni e cattivi pagatori come segue:

- Buon pagatore: meno di 3 rate consecutive insolte nel corso di vita della pratica;
- Cattivo pagatore: 3 o più rate consecutive insolte nel corso di vita della pratica.

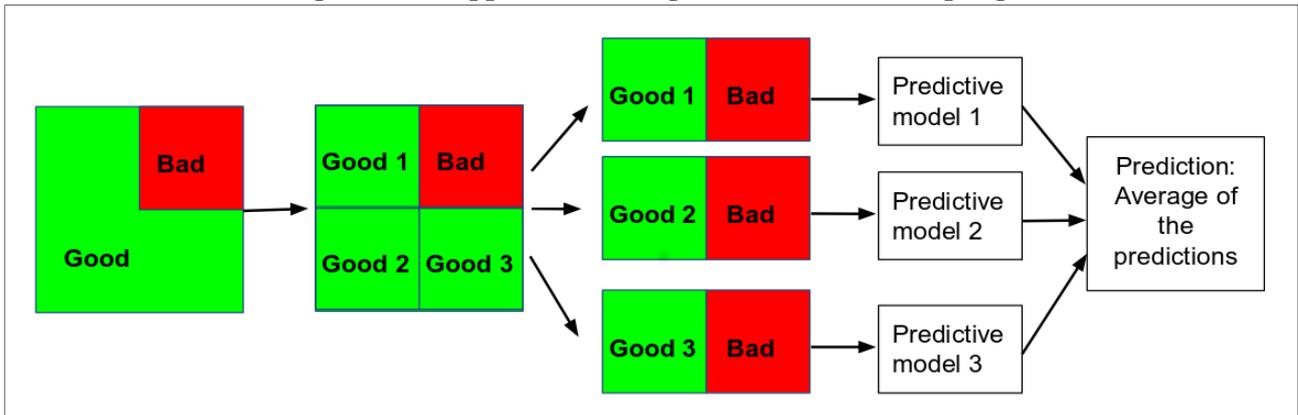
Poiché per alcuni clienti le informazioni fornite dai credit bureau sono di primaria importanza per avere un quadro completo della situazione creditizia dei clienti, le soglie saranno scelte per evitare un numero elevato di risultati falsi negativi o falsi positivi. L'intervallo di score compresa tra le soglie g e b in grigio rappresenta la sezione di clienti per cui il modello non è stato in grado di effettuare un'assegnazione ad una delle due categorie e per cui è richiesta un'integrazione con i dati dei credit bureau.

Analogamente a uno scenario di casi reali, le soglie vengono definite sul training set. In un caso reale, le soglie di accettazione e di rifiuto potrebbero essere regolate per ottimizzare una funzione obiettivo. Ad esempio, si potrebbe valutare il costo medio di un falso negativo e di un falso positivo e le soglie potrebbero essere impostate per massimizzare la differenza tra i costi risparmiati dal minor numero di chiamate ai credit bureau e il denaro perso a causa di falsi negativi e falsi positivi. In questo caso sarebbe il modello che automaticamente andrebbe ad individuare il livello di soglie che raggiungerebbe tale risultato.

5.3.3 Definizione del modello predittivo

Nella definizione del modello sono state considerate solo le pratiche per cui l'intero ciclo di vita è conosciuto, ma questo campione di pratiche è fortemente sbilanciato tra buoni e cattivi pagatori e ciò può essere un problema per il training del modello. Al fine di superare questa problematica viene adottata la metodologia dell'undersampling descritta di seguito con l'aiuto della figura 5.7:

Figura 5.7: Rappresentazione grafica dell'undersampling



Sia G l'insieme delle pratiche buone (relative a buoni pagatori) e B l'insieme delle pratiche anomale (relative a cattivi pagatori). Siano g e b rispettivamente il numero di pratiche buone e il numero di quelle anomale, si ha che $g > b$. I passi dell'undersampling sono i seguenti:

1. Dividere G in sottoinsiemi G_1, G_2, \dots, G_n in modo da ottenere per ognuno di essi approssimativamente un numero b di pratiche buone;
2. Ogni sottoinsieme G_i viene utilizzato insieme a tutto l'insieme B come training set per un algoritmo predittivo (random forest, neural network, mix di random forests e neural networks);
3. A questo punto ci sono n sotto modelli predittivi allenati che possono essere usati separatamente per valutare una pratica;
4. Il punteggio assegnato a una pratica è la media dei punteggi assegnati dall'insieme di sotto modelli predittivi.

Tale approccio ha i seguenti vantaggi:

- L'addestramento del modello viene eseguita su tutti i dati disponibili (dell'intero training set);
- Il modello impara molte volte, cioè da diverse prospettive, il pattern corrispondente a una pratica anomala, quindi non si adatta troppo su una singola tipologia di pattern riducendo il rischio di overfitting;
- Diversi algoritmi predittivi o diversi modelli con hyper-parameters⁵³ possono essere usati insieme per ottenere una previsione.

Nel campione del presente caso di studio le pratiche anomale sono circa 37.000, costituendo il 5% del totale del campione di pratiche analizzate. Adottando la tecnica dell'undersampling sono stati ottenuti 19 sotto modelli da altrettanti sottoinsiemi, ognuno

⁵³ Sono parametri il cui valore viene settato prima che inizi il processo di allenamento del modello, sono quindi dei valori costanti. Hutter, Frank, Holger Hoos, and Kevin Leyton-Brown, "An efficient approach for assessing hyperparameter importance.", International Conference on Machine Learning, 2014

della numerosità delle pratiche anomale, in modo da lavorare con un campione pareggiato di pratiche buone e anomale.

5.3.4 Fit del modello

Sono stati testati una serie di possibili algoritmi di base e dei loro hyper-parameters, ottenendo come combinazione più performante un mix di random forest e neural network, che è stata scelta per lo sviluppo del modello. I criteri adottati per la scelta sono stati rispettivamente AUROC, Gini index, PSI (Population Stability Index), FOR (false omission rate) e FDR (false discovery rate)⁵⁴.

Ad ognuno dei sotto modelli viene applicato l'algoritmo tra random forest e neural network che meglio si adatta. In termini di tempo impiegato per la fase di fit del modello si osserva che l'algoritmo neural network è decisamente più rapido rispetto al random forest, impiegando rispettivamente una media di 4 secondi ad esecuzione contro i 35 secondi del random forest per ogni sotto modello.

In particolare, rispettivamente per l'algoritmo random forest gli hyper-parameters che possono essere impostati sono:

- `n_estimators`: il numero di alberi della foresta;
- `criterion`: la funzione per misurare la qualità di una divisione (Gini impurity o information gain⁵⁵);
- `max_depth`: la profondità massima di un albero, ossia il numero massimo di ramificazioni;
- `min_samples_split`: numero minimo di dati in un nodo per effettuare una divisione;
- `min_sample_leaf`: numero minimo di dati presenti in un nodo foglia;
- `min_weight_fraction_leaf`: la frazione ponderata minima della somma totale dei pesi (di tutti i campioni di input) richiesta in un nodo foglia;
- `max_features`: il numero massimo di variabili da considerare nella ricerca della divisione migliore;
- `max_leaf_nodes`: numero massimo di nodi foglia;
- `min_impurity_decrease`: la divisione del nodo avviene se genera una riduzione dell'impurità maggiore o uguale al valore indicato;
- `min_impurity_split`: soglia del valore di impurità al di sotto del quale si arresta il processo di divisione dei nodi;
- `bootstrap`: permette l'utilizzo del campionamento bootstrap;
- `oob_score`: se utilizzare un campione out-of-bag;
- `warm_start`: se inizializzare il nuovo modello con i dati elaborati precedentemente, oppure se costruire una nuova foresta.

Per l'algoritmo neural network gli hyper-parameters sono:

- `hidden_layer_sizes`: numero di neuroni nello strato nascosto;
- `activation`: funzione di attivazione;
- `solver`: funzione di ottimizzazione dei pesi delle variabili;
- `alpha`: L2 penalty, introduzione di una penalità ai pesi per evitare l'overfitting;

⁵⁴ Si veda Capitolo 7 paragrafo 3

⁵⁵ Per ulteriori dettagli su Gini impurity e information gain: https://www.bogotobogo.com/python/scikit-learn/scikit_machine_learning/Decision_Tree_Learning/Information_Gain/IG_Impurity_Entropy_Gini_Classification_Error.php

- `learning_rate`: programma il learning rate per l'aggiornamento dei pesi (costante, decrescente)
- `learning_rate_init`: valore iniziale di learning rate;
- `max_iter`: numero massimo di iterazioni se non viene raggiunto prima un criterio di stop;
- `shuffle`: se mescolare i campioni ad ogni iterazione;
- `tol`: valore di tolleranza applicato;
- `warm_start`: se inizializzare il nuovo modello con i dati elaborati precedentemente;
- `early_stopping`: se utilizzare l'arresto anticipato per interrompere la fase di training quando le performance non migliorano;
- `validation_fraction`: proporzione del training set da tenere da parte per la fase di validazione;
- `n_iter_no_change`: numero massimo di epoche (iterazioni) in cui non si raggiunge il valore `tol` di miglioramento.

Gli algoritmi eseguono intrinsecamente un'analisi multivariata effettuando un mix di tutte le informazioni disponibili ricevute in input. Essi ponderano automaticamente i campi di input, assegnando un'importanza appropriata a ciascuna variabile, dedotta dalla fase di addestramento.

Inoltre, i modelli sono indipendenti dalla natura dei dati in input (a condizione che i dati siano codificati correttamente in un formato numerico). Il modello può essere quindi esteso con nuove informazioni provenienti da nuove fonti, previo il riaddestramento del modello. Lo svantaggio principale di questo tipo di modelli è che i motivi che portano a una particolare previsione non possono essere facilmente controllati (black box paradigm⁵⁶).

5.3.5 Training e validazione

Le fasi di training e validazione del modello sono state effettuate in due metodologie differenti:

- I. Campionamento casuale dell'80% delle pratiche per il campione di sviluppo e il restante 20% per il campione di validazione (cross validation);
- II. Utilizzo delle pratiche dal 2011 al 31 dicembre 2015 per il campione di sviluppo (85% dei dati disponibili) e le rimanenti pratiche dal 2016 al 2018 per il modello di validazione.

Il primo approccio è più significativo da un punto di vista statistico, mentre il secondo approccio è più facilmente interpretabile e meglio si adatta ad un caso reale, in quanto simula lo scenario in cui il modello viene addestrato fino alla fine del 2015 e viene utilizzato per valutare le nuove pratiche aperte dal 2016.

La fase di training per l'algoritmo di random forest consiste nello sviluppare un certo numero di alberi definito tra gli hyper-parameters e secondo i parametri indicati, che identificheranno per ogni variabile la suddivisione più appropriata per distinguere tra un buono e cattivo pagatore. Considerando ad esempio la variabile numero di rate del prestito l'algoritmo individua per quale valore si ottiene la miglior distinzione tra buoni e cattivi pagatori. L'output osservato sarà la media tra tutti gli alberi sviluppati a partire da un campione. Nella fase di validazione vengono applicati al nuovo campione le suddivisioni

⁵⁶ Nella teoria dei sistemi un modello black box permette solo di osservare come tale sistema reagisce (output) a una sollecitazione esterna (input), senza però sapere cosa avviene al suo interno.

individuate nella fase di training, ottenendo per ogni cliente uno score indice della probabilità di essere un buon o cattivo pagatore.

Per l'algoritmo di neural network nella fase di addestramento a partire dall'output osservato, ossia buono o cattivo pagatore, vengono assegnati dei pesi alle variabili che vanno a definire lo score del cliente, il processo con cui il modello assegna questi punteggi è però sconosciuto a causa del black box paradigm. In fase di validazione l'algoritmo così ottenuto viene applicato al nuovo campione ottenendo come output un punteggio per ogni cliente indice del suo grado di rischio.

L'utilizzo di modelli previsionali nei processi decisionali di SPC comporta la necessità di rilevare e monitorare anche il Rischio Modello, definito come il rischio di subire impatti negativi di natura finanziaria e/o reputazionale a causa di decisioni assunte sulla base di output di modelli che possano presentare errori o a causa di un cattivo/improprio utilizzo dei modelli stessi.

“Il termine modello si riferisce ad un metodo, sistema o approccio quantitativo che utilizza teorie, tecniche o ipotesi di natura statistica, economica, finanziaria o matematica al fine di trasformare dati di input in stime quantitative di natura previsionale.”⁵⁷

I modelli statistici adottati da SPC vengono sottoposti all'analisi del Comitato Modelli di Gruppo (Banca Sella), il quale esprime una valutazione dell'adeguatezza dei modelli:

- prima della messa in produzione nel caso di nuovo modello;
- su base periodica per i modelli esistenti;

ed indirizza le azioni correttive necessarie alla minimizzazione del rischio modello.

Il Comitato, dopo avere visionato le risultanze dei monitoraggi, esprime uno dei seguenti giudizi:

Giudizio	Effetti
Favorevole	Il modello presenta livelli di performance adeguati e può continuare ad essere utilizzato/può essere messo in produzione.
Favorevole con riserva	Il modello presenta evidenze di non adeguata performance che tuttavia non inficiano significativamente la misurazione di rischio. Il modello può continuare ad essere utilizzato, ma devono essere posti in atto approfondimenti che consentano di comprendere con maggior dettaglio le cause del deterioramento della performance e il possibile piano di azione (risviluppo, ricalibrazione) con le relative tempistiche.
Non favorevole	Il modello presenta evidenze di non adeguata performance che inficiano significativamente la misurazione della rischiosità e conseguentemente le decisioni di business connesse a tali misurazioni. Il modello non può più essere applicato/non può essere messo in produzione, devono essere attivate immediatamente le azioni di contingency identificate dal Comitato (es. deve essere attivato un immediato piano di risviluppo del modello).

Tabella 5.8: Test statistici potere discriminante

L'attività di validazione si basa sui seguenti principi:

- **Obiettività:** utilizzo di metodologie e test caratterizzati dall'assenza di elementi di soggettività;

⁵⁷ Federal Reserve “Supervisory guidance on model risk management”, 4 aprile 2011

- Comparabilità: utilizzo di metriche comuni a diversi controlli, confrontabili nel tempo e con le soglie presenti nella letteratura di riferimento;
- Efficienza: approfondimenti per step successivi soprattutto in caso di evidenze negative;
- Trasparenza: processo chiaro in tutti i suoi aspetti, per tutti gli attori coinvolti.

I modelli devono essere oggetto di validazione sia prima del passaggio in produzione (c.d. ‘prima validazione’), sia successivamente alla messa in opera del modello (monitoraggio e validazioni periodiche on-going) con una frequenza basata anche sulla dimensione del complessivo valore finanziato del business il cui rischio viene misurato dal modello. Nella seguente tabella sono riportati i diversi tipi di validazione che possono essere effettuati:

Tipo di validazione	Descrizione	Periodicità
Prima validazione	Validazione in fase di sviluppo del modello.	Prima della messa in produzione.
Monitoraggio (on-going)	Validazione periodica con analisi degli indicatori rilevanti.	Mensile o semestrale.
Validazione completa (on-going)	Validazione periodica con analisi degli indicatori rilevanti e dei test statistici finalizzati ad indagare la necessità di una eventuale ricalibrazione dei parametri del modello.	Annuale o biennale.

Tabella 5.9: Tipologie di validazione

Per quanto concerne i modelli di fornitori esterni essi possono essere utilizzati e verificati se e solo se l’outsourcer fornisce una piena disclosure sulla metodologia utilizzata e sui risultati dei backtesting effettuati.

5.3.5.1 Reject inference

La Reject Inference è uno strumento fondamentale nei modelli di accettazione, poiché permette di definire una performance “fittizia” per i clienti che sono stati rifiutati, e per i quali dunque non è possibile osservare la performance “vera”.

In ogni modello statistico, l’assunzione chiave che si fa è che il campione utilizzato per sviluppare il modello sia generalmente rappresentativo della popolazione su cui verrà applicato. Nel caso dei modelli di accettazione, questa ipotesi non è vera: approvando i probabili clienti buoni e rifiutando quelli cattivi, il dataset di sviluppo con i soli clienti accettati è intrinsecamente distorto.

Poiché il modello sarà utilizzato per essere applicato all’intera popolazione che richiede un fido (c.d. popolazione “through the door”) per decidere se approvare o rifiutare la richiesta, la distorsione diventa un problema critico. Le metodologie di Reject Inference sono un modo per tenere conto e correggere questa distorsione di campionamento.

Attraverso l’utilizzo di tecniche di reject inference la procedura di validazione è stata rieseguita anche sul campione di pratiche respinte dal modello di Sella Personal Credit, anche se per tali pratiche il ciclo di vita dei pagamenti e quindi il comportamento reale non

è disponibile. Quindi la validazione sui respinti avviene testando il modello su questi clienti e verificando quanti sarebbero stati accettati, supponendo che le pratiche rifiutate dal modello di SPC appartengano effettivamente a cattivi pagatori, per cui il modello sviluppato dovrebbe rifiutare la maggior parte di queste.

5.3.6 Definizione delle classi e assegnazione delle PD

Il punteggio di score ottenuto dal modello viene convertito in una probabilità di default del cliente, consentendo di stimare la probabilità che un cliente diventi un cattivo pagatore in base al punteggio assegnato. La probabilità di default viene stimata secondo i seguenti passi:

1. Si svolgono test indipendenti di cross validation;
2. Per ogni test:
 - a. Si divide la scala dello score [0,100] in 20 gruppi di intervalli uguali, ma tale suddivisione può essere arbitraria;
 - b. Si calcola la percentuale di cattivi pagatori che fanno parte di ogni gruppo;
3. La probabilità di default di un gruppo è data dalla media delle probabilità ottenute nei test effettuati per ogni gruppo e viene calcolata sul campione di validazione, non su quello di sviluppo, questo per far sì che i risultati siano consistenti quando vengono applicati su nuovi dati mai visti dal modello.

La media delle probabilità di default dei test di cross validazione viene effettuata per appiattare le fluttuazioni che si verificano sui risultati ottenuti selezionando in modo casuale i campioni di sviluppo e di validazione.

5.3.7 Analisi di performance del modello

Essa deve essere effettuata sia prima, sia successivamente alla messa in produzione del modello. In fase di prima validazione le analisi devono essere generalmente condotte, quando la numerosità delle osservazioni lo consente, su un campione di popolazione out-of-sample (cioè relativo ad osservazioni che, pur riferite allo stesso periodo temporale, non fanno parte del campione di sviluppo) e su un campione out-of-time (cioè relativo ad osservazioni riferite ad un periodo temporale differente rispetto al campione di sviluppo). Nei casi in cui la dimensione del campione non fosse sufficientemente ampia è preferibile effettuare una validazione in-sample con un metodo bootstrap, andando a selezionare in maniera casuale dal campione di sviluppo un numero opportuno di sottocampioni (indicativamente almeno 10 sottocampioni con dimensione pari ad almeno il 30% del campione di sviluppo). Successivamente alla messa in produzione del modello, le analisi devono essere effettuate su campioni out-of-time (cioè relative ad osservazioni riferite ad un periodo temporale differente rispetto al campione di sviluppo). La scelta delle tecniche/test utilizzati e la periodicità delle verifiche dipende dalla metodologia alla base del modello, dalla sua complessità, dalla disponibilità dei dati e dalla dimensione del complessivo valore finanziato del business il cui rischio viene misurato dal modello.

5.3.7.1 Potere discriminante

Il potere discriminante di un modello di scoring denota la capacità di discriminare ex-ante tra I debitori che entreranno in default e coloro che invece saranno buoni pagatori alla fine

del periodo in esame. Un modello di scoring ha un buon potere discriminante se i default sono concentrati nelle classi peggiori. I test statistici utilizzati per il potere discriminante sono riportati in tabella 5.10:

Test statistico	Finalità
AR – Accuracy Ratio (Gini Coefficient)	Confronta la capacità discriminante del modello con quella di un modello perfetto che cattura il 100% dei default nelle classi di rating peggiori.
AUROC – Area Under the ROC	Testa, per un dato livello di score (Cut-off), la capacità del modello di identificare correttamente I default reali, contrapponendoli ai falsi default segnalati.
MR – Misclassification rate	Indica la percentuale di casi classificati erroneamente dal modello.

Tabella 5.10: Test statistici potere discriminante

5.3.7.2 Stabilità

La stabilità di un modello denota la capacità di mantenere costante nel tempo il proprio potere discriminante. I principali test statistici da utilizzare per le analisi di stabilità e concentrazione sono riportati nella tabella 5.11:

Test statistico	Finalità
PSI – Population Stability Index	Confronta la distribuzione della popolazione osservata nel modello con quella del campione di sviluppo per testarne la stabilità nel tempo.
Check di monotonicità	Verifica l'andamento della PD osservata all'interno delle classi di assegnazione.

Tabella 5.11: Test statistici stabilità

Nella tabella 5.12 vengono riportati, per ambito di analisi, i test statistici utilizzati e le soglie di accettazione secondo le policy di SPC:

Ambito di analisi	Test statistici	Soglie Accettazione	
Performance – potere discriminante	Accuracy Ratio o Gini Coefficient	ROSSO	$AR < 35\%$
		GIALLO	$35\% \leq AR < 50\%$
		VERDE	$AR \geq 50\%$
	AUROC - Area under Receiver Operating Characteristic	ROSSO	$ROC < 67.5\%$
		GIALLO	$67.5\% \leq ROC < 75\%$
		VERDE	$ROC \geq 75\%$
	MR (Misclassification Rate)	ROSSO	$MR \geq 30\%$
		GIALLO	$10\% \leq MR < 30\%$
		VERDE	$MR < 10\%$
Stabilità	PSI – Population Stability Index	ROSSO	$PSI \geq 25\%$
		GIALLO	$10\% \leq PSI < 25\%$
		VERDE	$PSI < 10\%$
	Check di monotonicità	ROSSO	Monotonicità decrescente
		GIALLO	Monotonicità crescente
		VERDE	Monotonicità strettamente crescente

Tabella 5.12: Test statistici e soglie

Per ogni modello, gli esiti dei test statistici vengono aggregati al fine di ottenere una valutazione finale su una scala a tre livelli (adeguato, adeguato con riserva, non adeguato).

5.4 Osservazioni

Come suddetto prima della fase di fit e training del modello si ottengono molteplici combinazioni di variabili (pattern) che possono essere dati in input e testate, ottenendo per ognuna di esse diversi risultati dal punto di vista delle performance. Il pattern prescelto per il modello di machine learning sarà quello che permette di ottenere valori di performance migliori sul campione di validazione. I modelli di machine learning, essendo indipendenti dalla natura dei dati in input, hanno inoltre il vantaggio di poter integrare nuove variabili semplicemente effettuando un riaddestramento del modello.

Le fasi di elaborazione dei pattern, fine tuning degli algoritmi e degli hyper-parameters sono molto onerose in termini di tempo, nonostante possano essere parzialmente resi automatici, pertanto vengono testate un certo numero di possibili combinazioni impostando dei limiti di tempo.

Il modello ottenuto è stato addestrato su tutto il portafoglio del consumo di SPC, assumendo quindi un carattere generale e non specifico per un determinato prodotto o sottoprodotto. Questo comporta che, andando ad applicare il modello ad uno specifico prodotto o

sottoprodotto le performance effettivamente osservate potrebbero essere più basse di quelle riscontrate in fase di sviluppo.

Al fine di rendere ancora più accurato e performante il modello, si possono eseguire le medesime analisi su un determinato prodotto o insieme di sottoprodotti dalle caratteristiche simili, considerando come criteri di scelta di la numerosità della categoria, il costo di sviluppo ed il beneficio economico potenzialmente generato in termini di risparmio sui costi per le chiamate alle banche dati. Dunque, si potranno ottenere molteplici e differenti modelli, ognuno allenato e testato su una specifica categoria.

L'ulteriore beneficio che generano i modelli di machine learning, oltre quello economico, è di essere in grado di effettuare automaticamente la calibrazione dei parametri, pertanto qualora si verificassero delle modifiche nella popolazione richiedente, per cui si osservasse una riduzione delle performance, il modello automaticamente provvede a ricalibrare i parametri e riportare i valori delle performance sopra le soglie di accettazione.

6 Esplorazione dei dati – Analisi ed opportunità

6.1 Introduzione

Il capitolo tratta la fase esplorativa dei dati a disposizione di SPC relativi ad ogni prestito. Nel contesto dell'esplorazione sono state svolte analisi per ogni variabile allo scopo di individuarne un sub set di interesse sul quale effettuare lo sviluppo del modello.

L'altro aspetto trattato nell'ambito del capitolo riguarda l'introduzione di nuove variabili che attualmente non vengono utilizzate da SPC nella valutazione del merito creditizio oppure che non vengono raccolte. I dati introdotti provengono da fonti non convenzionali, in particolare dai digital media e possono essere utilizzati insieme a quelli tradizionali per ottenere una caratterizzazione più accurata dei clienti apportando un vantaggio competitivo sul mercato.

6.2 Analisi delle variabili

I dati su cui è stata svolta l'analisi contengono informazioni anagrafiche relative al cliente e al coobbligato (età, lavoro, residenza, nascita, ecc.), le informazioni ottenute dai credit bureau, che in fase di sviluppo non sono stati tenuti in considerazione, e i dati relativi al prodotto venduto (ammontare del finanziamento, numero di rate, tasso applicato, ecc). Altri dati di interesse, come ad esempio il numero di prestiti attivi o chiusi con SPC, sono stati derivati mediante un raggruppamento dei rapporti per cliente.

L'analisi esplorativa si è focalizzata sulla selezione di un subset di variabili di interesse, andando ad analizzare per ogni valore, o intervallo di valori, il numero di buoni e cattivi pagatori. Di seguito si elencano le variabili sottoposte all'analisi preliminare:

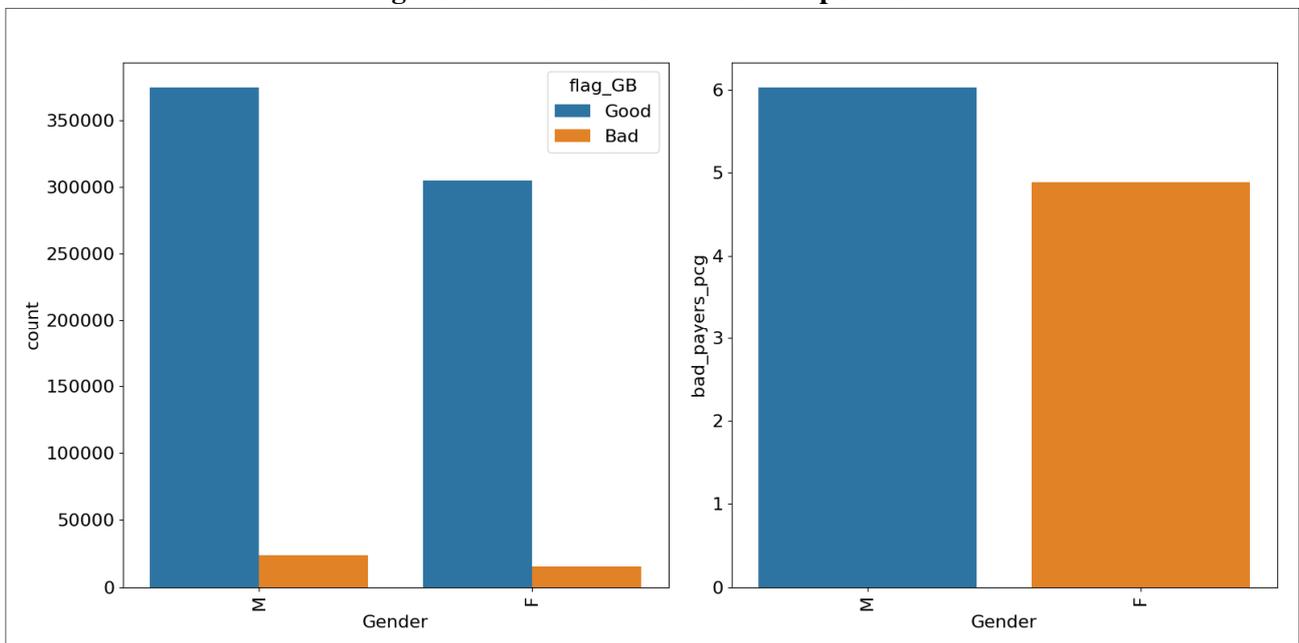
- Informazioni relative al cliente:
 - Regione di residenza;
 - Area Nielsen di residenza;
 - Regione di nascita;
 - Area Nielsen di nascita;
 - Stato civile;
 - Sesso;
 - Tipologia di abitazione;
 - Stipendio mensile;
 - Forma giuridica;
 - Professione;
 - Età;
 - Anzianità lavorativa.
- Informazioni sul prestito:
 - Tipo di prestito (PA, PF, PP);
 - Prodotto;
 - Numero di rate;
 - Importo rata;
 - Rete vendita;
 - Filiale;
 - Provincia punto vendita;

- Prezzo del prodotto;
- Valore del finanziamento;
- Presenza di un coobbligato;
- Tassi (TAN e TAEG);
- Metodo di pagamento (RID, bollettino postale);
- (Importo rata) / (Stipendio mensile).
- Rapporti presenti e passati:
 - Numero di pratiche CO attive;
 - Numero di pratiche CO chiuse;
 - Se già cliente CO identificato come buono o cattivo pagatore;
 - Numero di carte attive;
 - Numero di carte chiuse;
 - Se già cliente CA identificato come buono o cattivo pagatore;
 - Numero di pratiche CQ attive;
 - Numero di pratiche CQ chiuse;
 - Se già cliente CQ identificato come buono o cattivo pagatore.

Tutti i dati elencati fanno riferimento al momento di richiesta del prestito, ad esempio l'età del cliente è riferita al momento in cui viene fatta la richiesta, così come il numero di pratiche chiuse o attive fanno riferimento al momento della richiesta del prestito e così via per tutte le altre variabili.

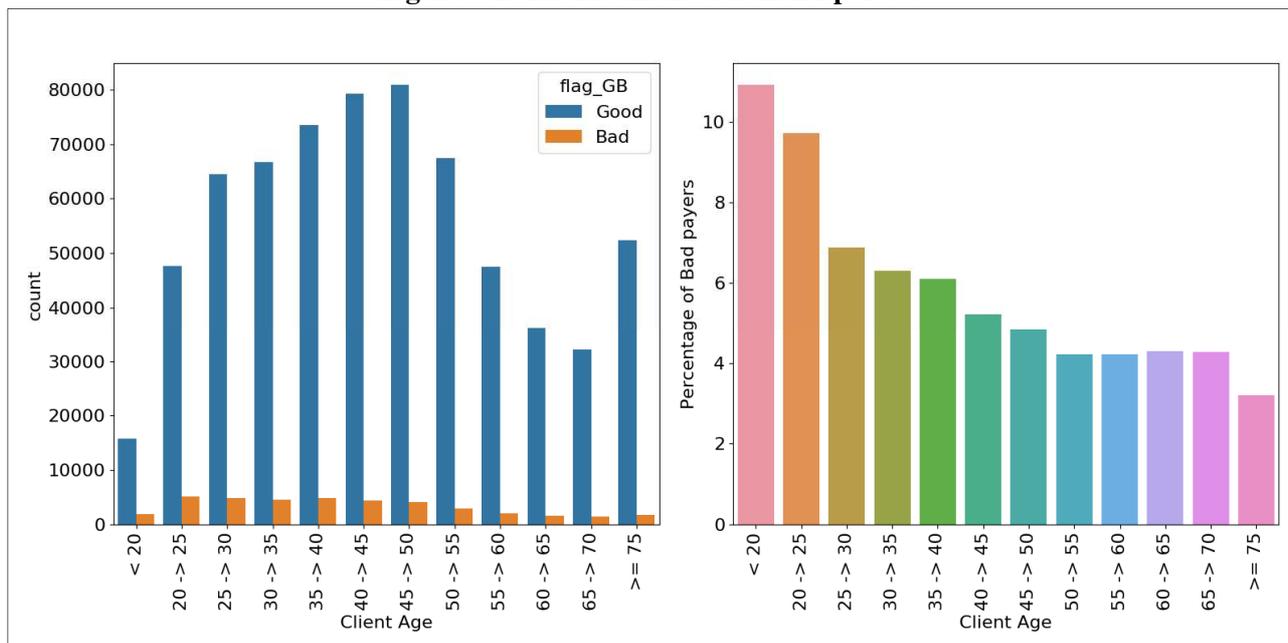
Di seguito vengono mostrati alcuni esempi di variabili che hanno riportato i risultati più significativi:

Figura 6.1: Distribuzione dei clienti per sesso



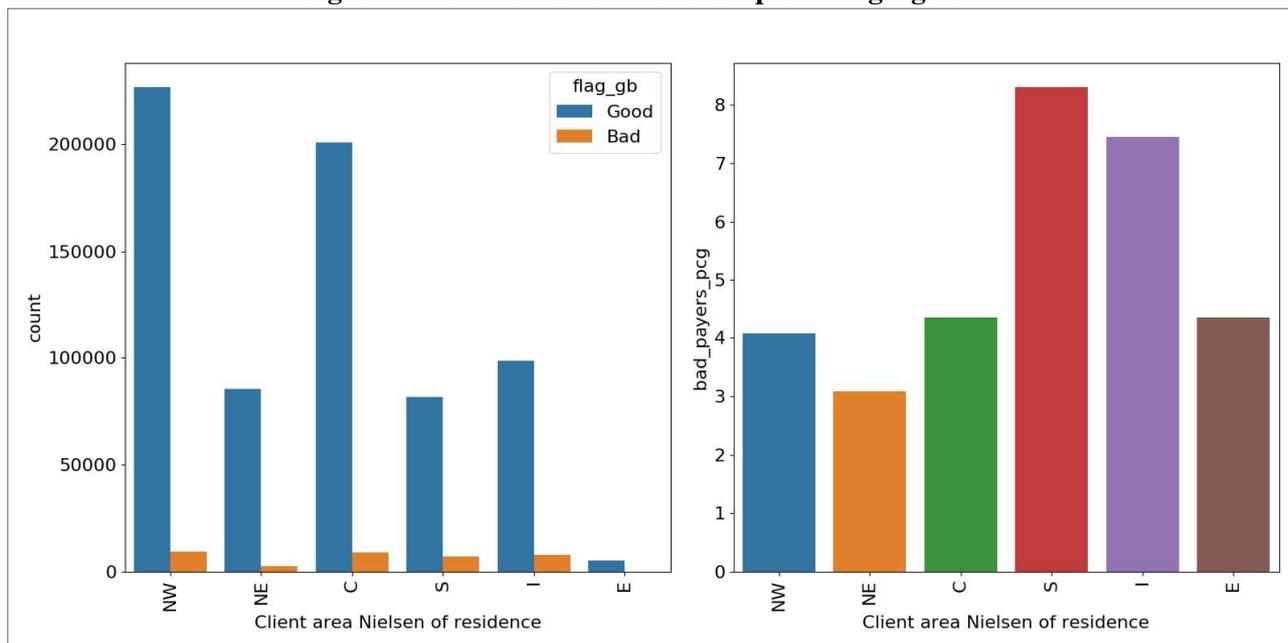
La figura 6.1 mostra a sinistra il numero di buoni e cattivi pagatori, mentre a sinistra mostra la percentuale di cattivi pagatori entrambi ripartiti per sesso. Tale grafico mostra che gli uomini costituiscono la maggior parte del portafoglio di prestiti al consumo e che essi abbiano una maggiore percentuale di cattivi pagatori rispetto alle donne. Il risultato ottenuto ci dice pertanto che, concedere un prestito ad una cliente donna è meno rischioso rispetto a concederlo ad un uomo.

Figura 6.2: Distribuzione dei clienti per età



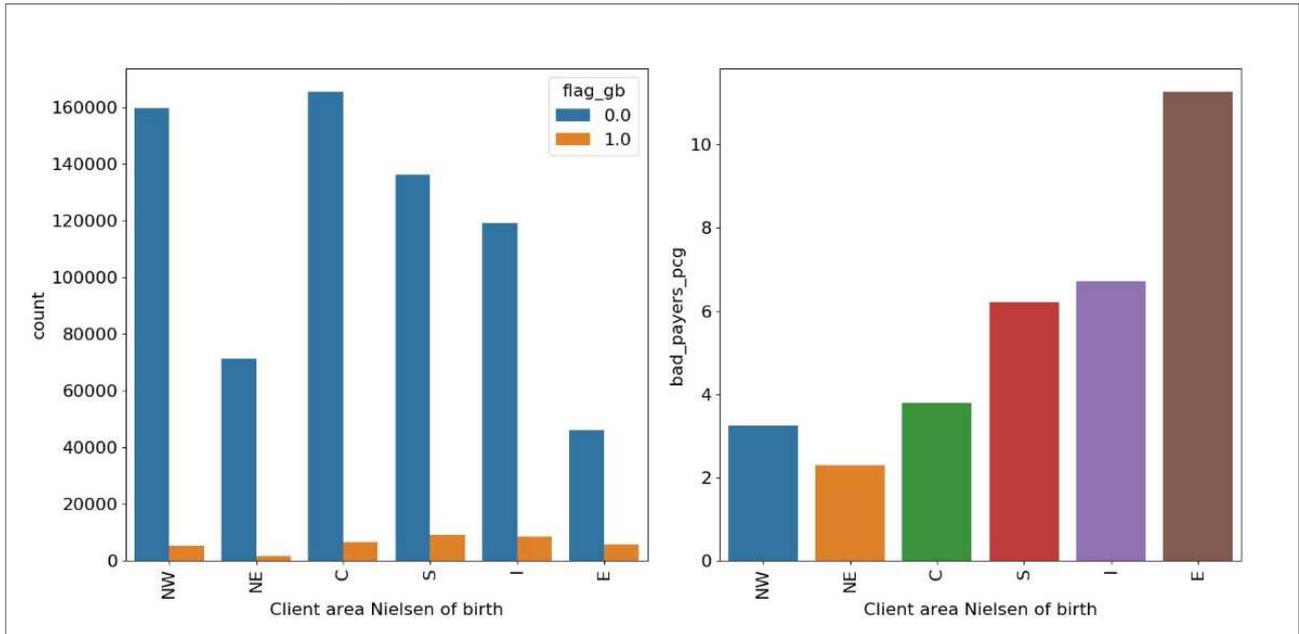
La figura 6.2 riporta a sinistra il numero di clienti distinti tra buoni e cattivi pagatori, mentre a destra la distribuzione percentuale dei cattivi pagatori entrambe suddivise per fasce di età. Il grafico mostra che la concentrazione del portafoglio di prestiti al consumo è nella fascia tra 40 e 50 anni, mentre dal punto di vista del rischio si ha un trend decrescente dove i clienti giovani riportano la percentuale più alta di cattivi pagatori, mentre più ci si sposta a destra, cioè verso clienti più vecchi, minore è la percentuale di cattivi pagatori. Il risultato ottenuto dice pertanto che, concedere un prestito a un cliente giovane (minore di 25 anni) è più rischioso rispetto a concederlo a un cliente più anziano (maggiore di 55 anni).

Figura 6.3: Distribuzione dei clienti per area geografica



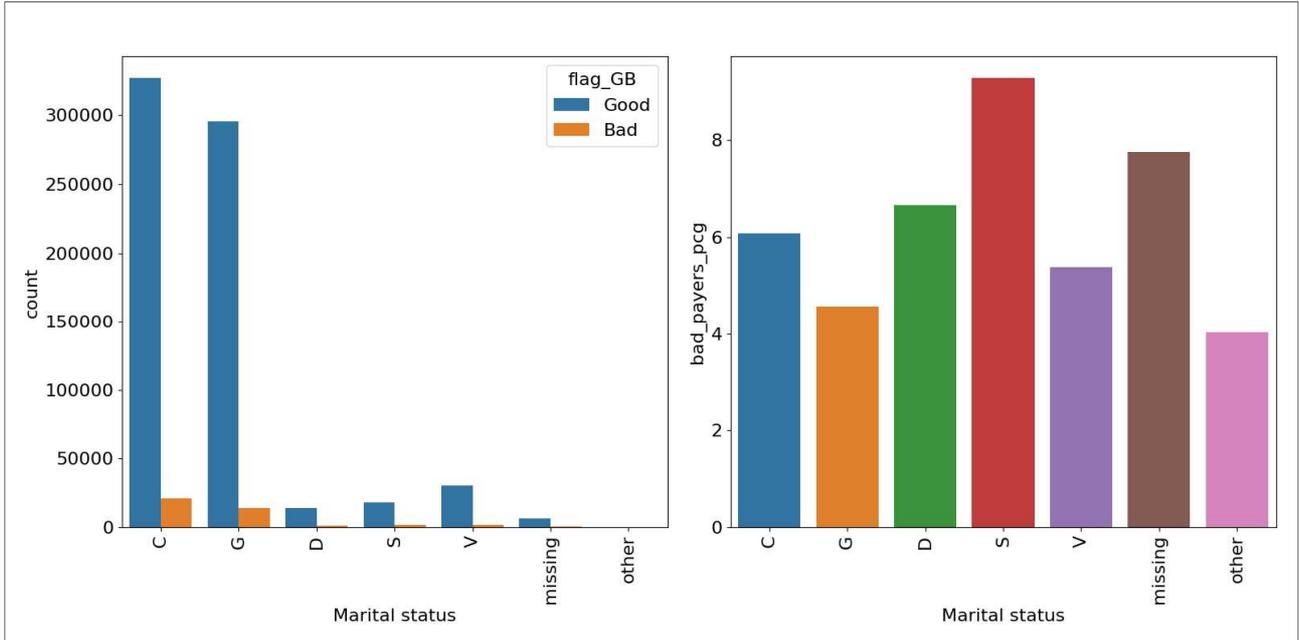
La figura 6.3 riporta a sinistra il numero di clienti distinti tra buoni e cattivi pagatori, mentre a destra la percentuale di cattivi pagatori divisi per area geografica di residenza (NW = nord-ovest; NE = nord-est; C = centro; S = sud; I = isole; E = estero). La figura ci dice che il portafoglio di prestiti al consumo si concentra principalmente tra clienti residenti al nord-ovest o al centro, mentre la percentuale di cattivi pagatori è maggiore al sud o nelle isole. Il risultato ottenuto ci dice pertanto che, concedere un prestito al nord-ovest oppure al nord-est oppure al centro Italia è meno rischioso di concedere un prestito ad un cliente residente all'estero oppure al sud o nelle isole.

Figura 6.4: Distribuzione dei clienti per area di nascita



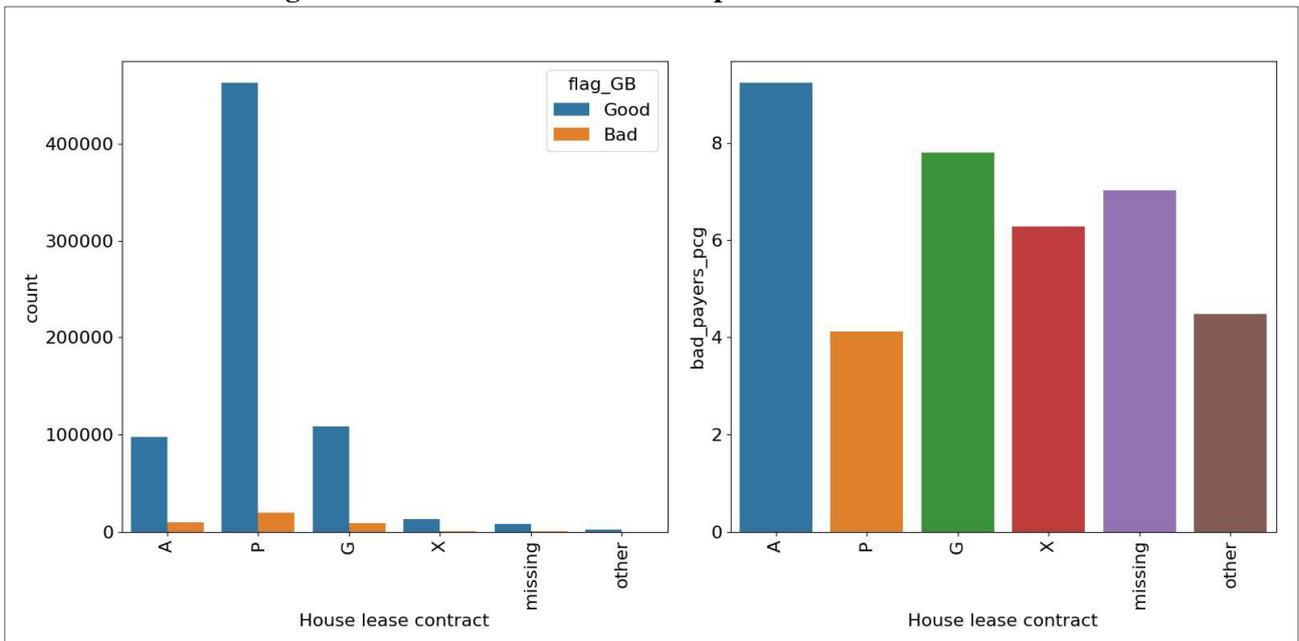
La figura 6.4 riporta il numero di buoni e cattivi pagatori suddivisi per area di nascita. L'informazione fornita dal grafico a sinistra è che il portafoglio del consumo si concentra sui clienti nati nelle zone nord ovest e centro, mentre la figura a destra evidenzia un picco di cattivi pagatori nei clienti nati all'estero, seguito dai clienti nati al sud e nelle isole, mentre si osservano tassi più bassi al nord. Pertanto, si può assumere che concedere un prestito ad un cliente nato al nord Italia è meno rischioso rispetto a concederlo ad un cliente nato all'estero o al sud o nelle isole. Questo risultato è peraltro in linea con quanto riportato nel grafico 6.3, ad eccezione per l'estero, ma questo perché il numero di prestiti concessi a clienti residenti all'estero è estremamente ridotto.

Figura 6.5: Distribuzione dei clienti per stato civile



La figura 6.5 riporta a sinistra il numero di buoni e cattivi pagatori divisi per stato civile (C = celibe/nubile; G = coniugato/a; D = divorziato/a; S = separato/a; V = vedovo/a; missing; altro), mentre a destra la percentuale di cattivi pagatori per categoria. La figura ci dice che il portafoglio di prestiti al consumo è maggiormente concentrato su clienti celibi/nubili oppure coniugati e che la percentuale di cattivi pagatori è più alta per i clienti separati o divorziati oppure missing. Tale risultato ci dice pertanto che è meno rischioso concedere un prestito a un cliente coniugato piuttosto che concederlo a un cliente separato.

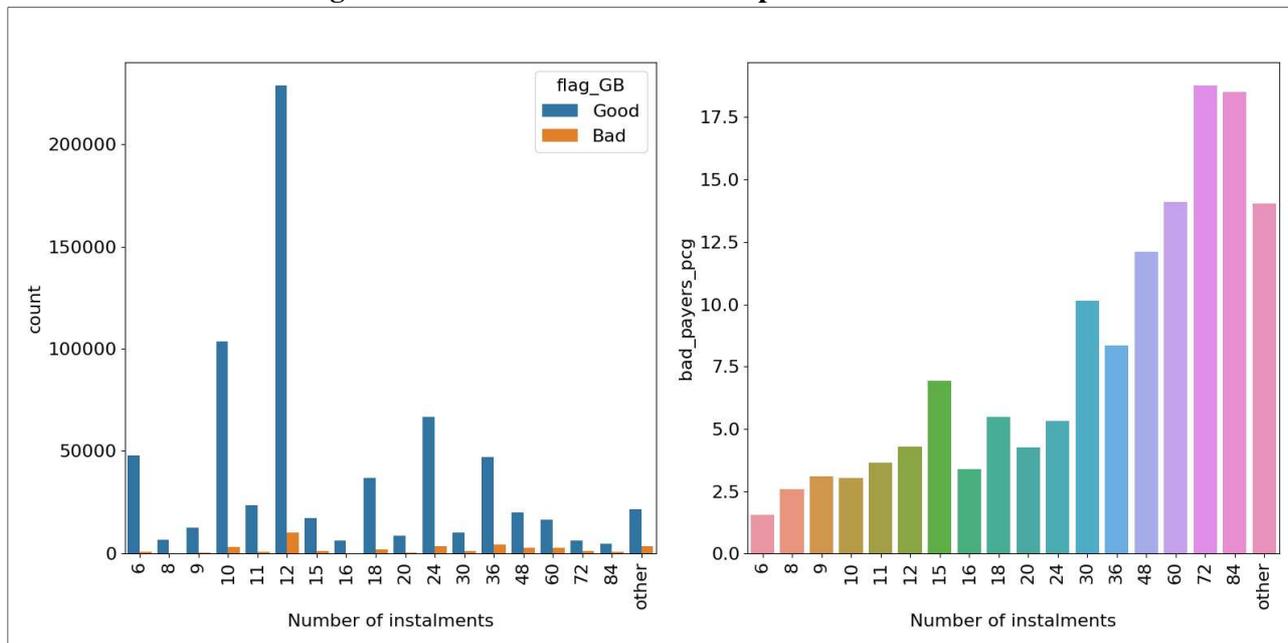
Figura 6.6: Distribuzione dei clienti per contratto di locazione



La figura 6.6 riporta a sinistra il numero di buoni e cattivi pagatori divisi per tipologia di contratto di locazione (A = affitto; P = proprietà; G = con i genitori; X = caserma/clero;

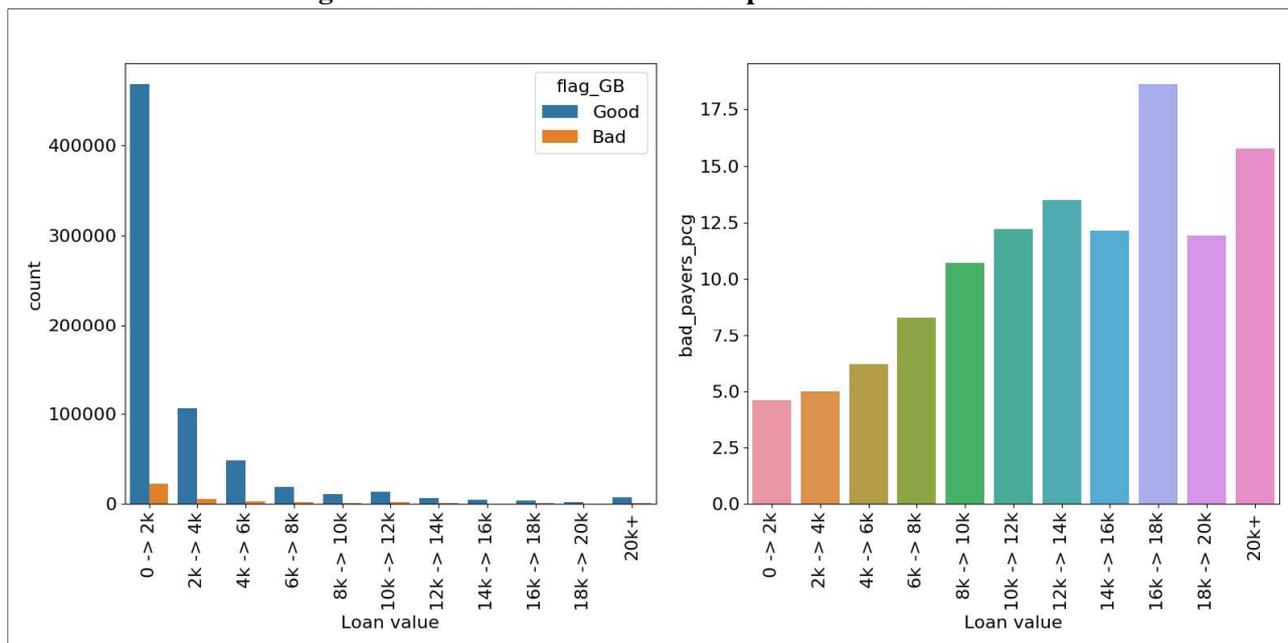
missing; altro), mentre a destra viene riportata la percentuale di cattivi pagatori per categoria. La figura mostra che il portafoglio di prestiti al consumo è maggiormente concentrato sui clienti con casa di proprietà, mentre i clienti che vivono in affitto risultano quelli con percentuale di cattivi pagatori più alta. Si può pertanto dire che, concedere un prestito a un cliente che possiede una casa di proprietà è meno rischioso di concedere un prestito a un cliente in affitto.

Figura 6.7: Distribuzione dei clienti per numero di rate



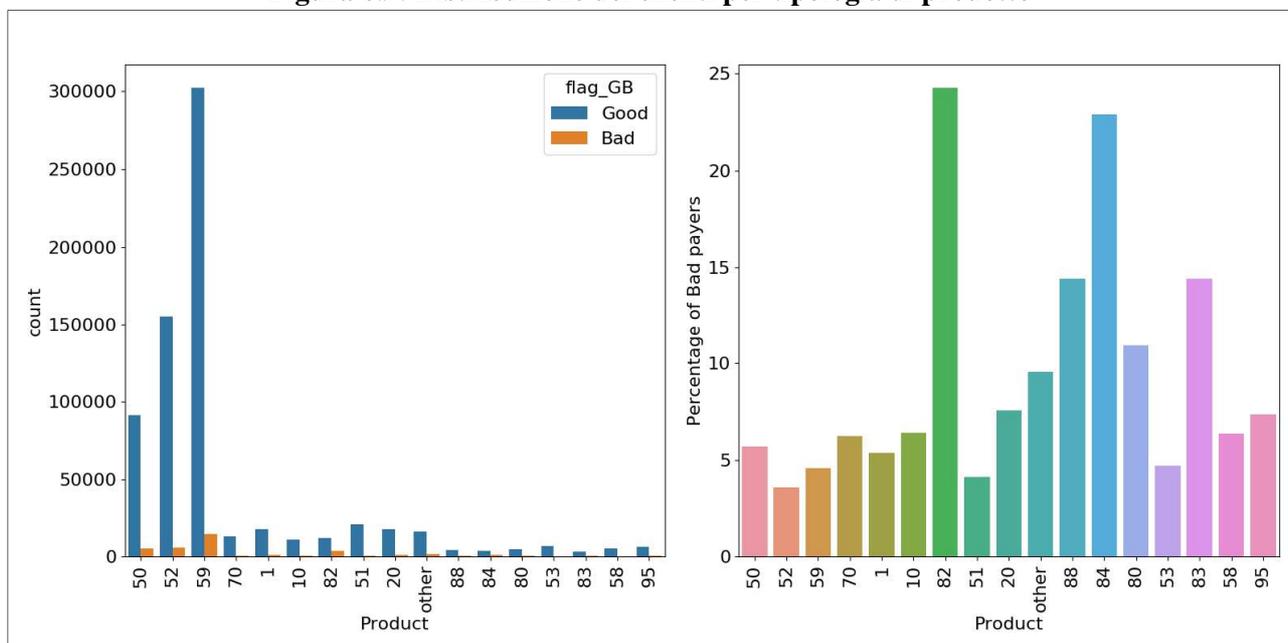
La figura 6.7 riporta a sinistra il numero di clienti classificati per numero di rate del prestito e divisi tra buoni e cattivi pagatori, mentre a destra riporta la percentuale di cattivi pagatori per classe. Il portafoglio di prestiti al consumo è maggiormente concentrato su prestiti della durata di 12 mesi, mentre in termini di rischiosità si delinea un trend crescente con il numero di rate, anche se non sempre rispettato, come ad esempio per il picco sui 15 mesi. Pertanto, si può dedurre che all'aumentare del numero di rate del prestito aumenterà la rischiosità ad esso associata.

Figura 6.8: Distribuzione dei clienti per valore finanziato



La figura 6.8 riporta a sinistra il numero di cattivi e di buoni pagatori suddivisi per fasce di valore finanziato, mentre a destra la percentuale di cattivi pagatori per fascia. Il portafoglio di prestiti al consumo è maggiormente concentrato su prestiti di importo minore a 2000€, mentre in termini di cattivi pagatori si delinea un trend crescente con l'aumentare dell'importo finanziato, con un picco per i prestiti tra 16.000 e 18.000 €. Pertanto, il risultato ci dice che il rischio aumenta con l'aumentare del valore del finanziamento.

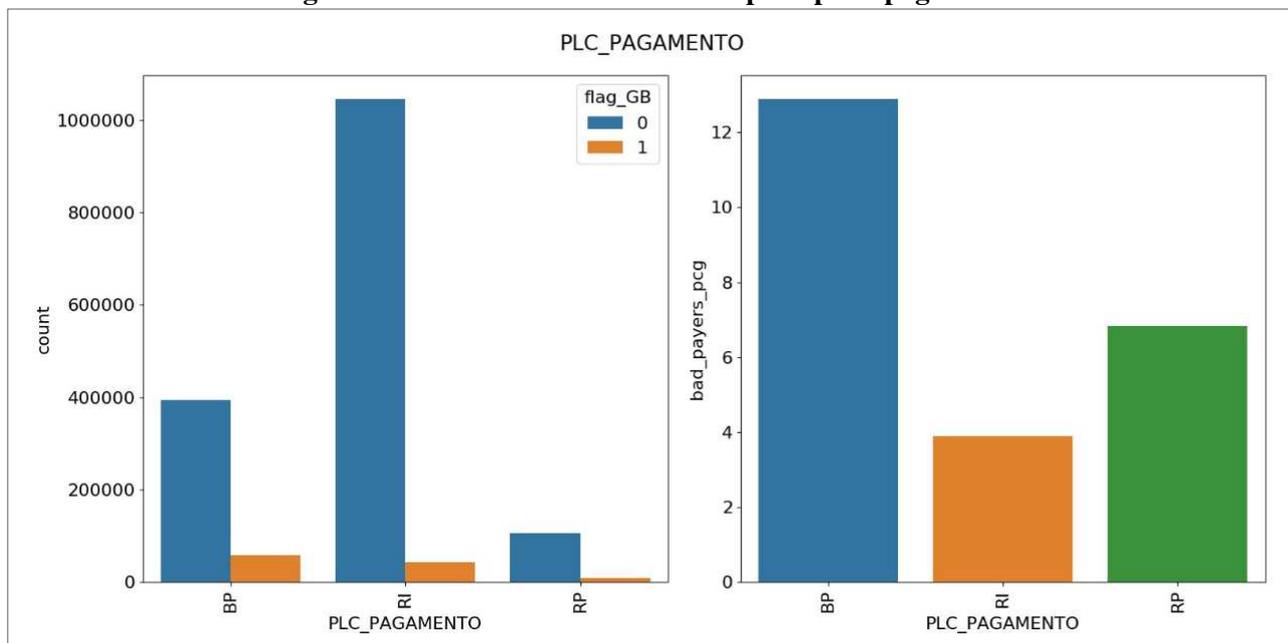
Figura 6.9: Distribuzione dei clienti per tipologia di prodotto



La figura 6.9 riporta il numero di buoni e cattivi pagatori suddivisi per tipologia di prodotto, ogni numero corrisponde a un determinato prodotto appartenente alla macro-categoria consumo, mentre a destra viene riportata la percentuale di cattivi pagatori per prodotto. Il portafoglio di prestiti al consumo è maggiormente concentrato sul prodotto 59, mentre in

termini di percentuale di cattivi pagatori i valori più alti si riscontrano per i prodotti 82 e 84. Pertanto, il risultato ci dice che è i prodotti per cui si ha un rischio più basso sono il 51, 52 e 53, mentre i prodotti più rischiosi sono l'82 e l'84.

Figura 6.10: Distribuzione dei clienti per tipo di pagamento



La figura 6.10 riporta a sinistra il numero di buoni e cattivi pagatori suddivisi per tipologia di pagamento (BP= bollettino postale, RI = addebito RID, RP= RID postale), mentre a destra il tasso di cattivi pagatori. Dalla figura a sinistra si osserva che il portafoglio di prestiti al consumo è maggiormente concentrato sui pagamenti di tipo RID, mentre la figura a sinistra mostra una maggiore concentrazione di cattivi pagatori per i clienti che pagano tramite bollettino postale. Pertanto, concedere un prestito ad un cliente che paga tramite bollettino postale risulta più rischioso rispetto ad un cliente che paga con un addebito RID.

Questa fase esplorativa dei dati è stata effettuata per ogni variabile disponibile, ma sono state riportate solo quelle di maggior interesse. L'insieme delle variabili che hanno rilevato un maggior potere discriminante nella fase di analisi costituisce il pattern che andrà in input al modello.

6.3 Profilo di un buon pagatore

A causa del "black box paradigm" è difficile identificare il profilo di un cliente che possa essere considerato sicuro, ma tramite un'analisi a posteriori dei risultati la features di buon pagatore che si è osservata maggiormente riporta le seguenti caratteristiche:

- Meno di 12 rate;
- Metodo di pagamento RID;
- Non giovane (il rischio decresce con l'aumentare dell'età).

Le caratteristiche menzionate sono le più frequenti riscontrate per un buon pagatore, ma non sono obbligatorie o sufficienti perché un qualsiasi cliente lo sia, inoltre, i risultati delle analisi cambiano a seconda delle soglie scelte.

6.4 Dati innovativi

L'utilizzo sempre più diffuso di internet rende disponibili molte informazioni sugli utenti che ne fanno uso, quali social networks, dati sulla navigazione online, e-mail e sui device utilizzati. Queste informazioni, definite “digital footprint”, si sono rivelate di interesse in studi sul credit scoring⁵⁸ soprattutto se combinate con i dati convenzionali. Nella tabella 6.11 sono riportati i dati innovativi che sono stati proposti:

Variabile	Descrizione
Tipo device	Tipologia di device utilizzato. (Desktop, Tablet o Mobile)
Sistema operativo	Windows, iOS, Android, Macintosh
E-mail provider	Gmail, Hotmail, Tiscali, Yahoo, ecc.
Geo localizzazione	Posizione del cliente che utilizza l'app o il sito di SPC
Indirizzo e-mail	Composizione dell'indirizzo e-mail del cliente. (Es. nome.cognome oppure nickname)
Tempi utilizzo App e sito	Orari di utilizzo dell'App o del sito di SPC
Clickstream su App e sito	Customer journey (pagine visitate, interazioni, tempo trascorso sulle pagine)
Social networks	Informazioni da LinkedIn (dati lavorativi e collegamenti) e Facebook (informazioni personali, amicizie ed attività)

Tabella 6.11: Dati innovativi dai digital media

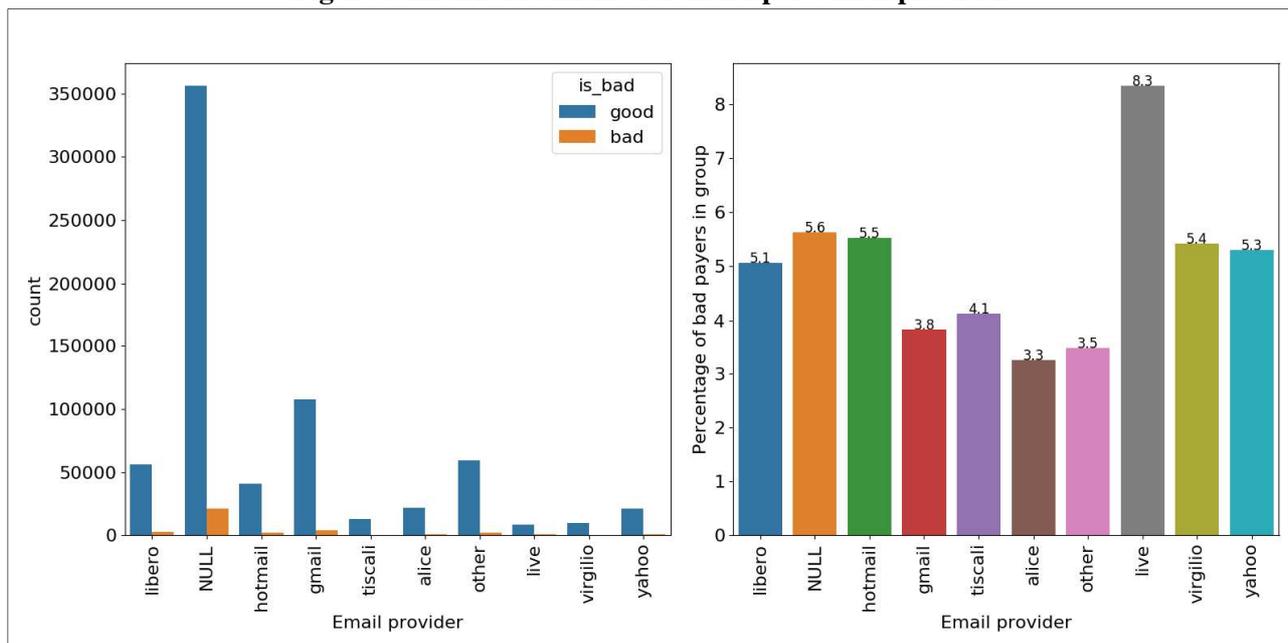
Oltre alle informazioni provenienti dai canali digitali si potrebbero sfruttare informazioni che SPC già dispone sui propri clienti, come i dati sulle carte di credito. Le informazioni sulle carte di credito riguardano due aspetti:

- Bilancio della carta: quante volte si è osservato un bilancio negativo;
- Punti vendita frequentati: correlazione tra punti vendita frequentati e prodotti acquistati e rischio.

Per ognuna delle variabili identificate si svolge la medesima analisi proposta per le variabili già disponibili. Una prima esplorazione è stata effettuata sui provider e-mail dei clienti per i quali è disponibile questa informazione riportata nella figura 6.12, dove a sinistra è riportato il numero di buoni e cattivi pagatori per service provider, mentre a destra la percentuale di cattivi pagatori per ogni categoria.

⁵⁸ Tobias Berg, Valentin Burg, Ana Gombović, and Manju Puri, “On the rise of fintechs – credit scoring using digital footprints”, Working Paper 24551, National Bureau of Economic Research, April 2018

Figura 6.12: Distribuzione dei clienti per e-mail provider



Tale analisi ha riscontrato che vi siano providers con un livello di rischio più alto rispetto ad altri. In particolare, gli indirizzi mail con dominio “gmail” o relativi a telecom, quali “alice” e “tiscali”, hanno una percentuale di cattivi pagatori più bassa rispetto ad altri providers come “live”, “yahoo” e “virgilio”, per i quali si sono osservati tassi di cattivi pagatori più alti. Pertanto, l’analisi suggerisce che, sulla base del portafoglio in essere di SPC, concedere un prestito ad un cliente con dominio e-mail “gmail” o appartenente a telecom è meno rischioso rispetto a concederlo ad un cliente con differente service provider.

6.5 Osservazioni

L’analisi svolta sulle variabili appartenenti al set informativo di SPC non sono in realtà nulla di nuovo, in quanto il profilo ideale di buon pagatore era già stato identificato da analisi interne precedenti. Le vere opportunità che possono essere ottenute sono date dall’integrazione dei dati innovati, che possono identificare da un lato nuove opportunità di business e dall’altro intercettare e limitare rischi nell’ambito della valutazione del merito creditizio dei clienti.

I limiti che si evidenziano nell’introduzione dei dati innovativi sono i seguenti:

1. Predisporre strumenti adatti alla raccolta e all’archiviazione delle nuove informazioni, richiedendo un investimento economico;
2. Normativi, in quanto alcuni dati richiedono uno specifico consenso da parte del cliente;
3. Creare una base dati sufficientemente ampia per poter svolgere le analisi, comportando un costo in termini di tempo per la raccolta dei dati;
4. Identificare le variabili sarà necessario rieffettuare il training del modello integrato con esse, ma testare nuovamente svariati pattern richiede un dispendio di tempo oneroso;
5. Utilizzando un set troppo ampio di variabili si rischia che il modello soffra di overfitting sui dati e di conseguenza riscontri variazioni nella popolazione in ingresso molto frequenti.

7 Applicazione del modello

7.1 Introduzione

Nel seguente capitolo viene riportato lo sviluppo del modello ottenuto seguendo la strategia con 2 soglie, una per identificare i buoni pagatori ed una per i cattivi pagatori, selezionate in base a delle soglie di accettazione predefinite in termini di errore di classificazione commesso (falsi negativi e falsi positivi). La procedura di campionamento mediante il quale è stato effettuato lo sviluppo è quella dell'undersampling con campioni casuali per training set e test set, composti rispettivamente dall'80% e dal 20% dei dati. I risultati ottenuti tramite campionamento effettuato in base al periodo temporale per training set e test set risultano simili a quelli riportati successivamente.

La procedura di cross validazione è stata ripetuta 4 volte effettuando 4 campionamenti casuali distinti ed indipendenti.

Si riportano anche i risultati ottenuti disaggregati relativi ai prodotti del consumo PA, PF e PP e sui singoli sottoprodotti al fine di verificare se l'applicazione del modello possa generare reali benefici ed opportunità economiche.

Di seguito si riporta direttamente il pattern finale, in quanto la fase di analisi è stata svolta una sola volta prima di sviluppare il modello, mentre la definizione del pattern è stata rieseguita più volte per stabilire le features che generassero i migliori risultati.

7.2 Pattern di variabili selezionate

A seguito dell'analisi esplorativa dei dati sono state selezionate le variabili candidate a formare il pattern in input al modello. Successivamente sono state testate diverse combinazioni di queste variabili, di seguito si riportano nella tabella 7.1 le variabili utilizzate per codificare il pattern:

Argomento	Variabile	Tipo variabile
Cliente	Regione di residenza	categoria
Cliente	Area Nielsen di residenza	categoria
Cliente	Regione di nascita	categoria
Cliente	Area Nielsen di nascita	categoria
Cliente	Stato civile	categoria
Cliente	Contratto di locazione	categoria
Cliente	Forma giuridica	categoria
Cliente	Sesso	categoria
Cliente	Impiego	categoria
Cliente	Età	numerico
Cliente	Anzianità lavorativa	numerico
Cliente	Stipendio mensile	numerico
Cliente	Numero di pratiche CO chiuse	numerico
Cliente	Comportamento nelle precedenti pratiche CO	numerico
Cliente	Numero di pratiche attive CO	numerico
Cliente	Numero di carte (CA) chiuse	numerico
Cliente	Comportamento nelle precedenti carte	numerico
Cliente	Numero di carte attive	numerico
Cliente	Numero di cessioni del quinto (CQ) chiuse	numerico
Cliente	Comportamento nelle precedenti cessioni del quinto	numerico
Cliente	Numero di cessioni del quinto attive	numerico
Pratica	Tipo prodotto (PP, PA, PF)	categoria
Pratica	Prodotto	categoria
Pratica	Rete di vendita	categoria
Pratica	Filiale	categoria
Pratica	Provincia della filiale	categoria
Pratica	Prezzo del prodotto	numerico
Pratica	Valore finanziato	numerico
Pratica	Numero di rate	numerico
Pratica	Importo rata	numerico
Pratica	Metodo di pagamento ("SDD", bonifico postale)	categoria
Pratica	Presenza coobbligato (S/N)	categoria
Pratica	Rapporto tra stipendio mensile e import rata	numerico

Tabella 7.1: Variabili codificate del pattern

La colonna tipo variabile indica se il valore di quella variabile è di tipo categoriale o numerico, per esempio il sesso che può assumere valori "M" o "F" ha carattere categoriale,

mentre l'età del cliente ha carattere numerico. Il pattern così ottenuto è oggetto di una preelaborazione dei dati contenuti.

Le variabili così identificate sono poi state preelaborate secondo i criteri riportati nel paragrafo 5.3.1.2 e date in input al modello.

7.3 Modello di regressione, fit e definizione della soglia

Individuato il pattern in input al modello si definiscono le caratteristiche del modello, ossia la tipologia di campionamento da applicare, quale algoritmo o algoritmi utilizzare ed i relativi parametri e le soglie da applicare per la classificazione degli output.

Il campione è stato suddiviso casualmente attraverso la metodologia dell'undersampling in due set distinti:

- 80% dei dati per il set di sviluppo;
- 20% dei dati per il set di validazione.

Gli algoritmi applicati per lo sviluppo sono stati random forest e neural network testando diverse combinazioni dei parametri riportati al paragrafo 5.3.4.

Nella scelta delle soglie di accettazione e rifiuto si è tenuto conto che il costo dato da un falso negativo è più alto di un falso positivo, pertanto si applicano i seguenti criteri:

- False omission rate (FOR), cioè la percentuale di falsi negativi, intorno allo 0,1%. Tale valore viene calcolato dal negative predictive value (NPV):

$$NPV = \frac{n \text{ veri negativi}}{n \text{ veri negativi} + n \text{ falsi negativi}} \quad (7.1)$$

$$FOR = 1 - NPV \quad (7.2)$$

- False discovery rate (FDR), cioè la percentuale di falsi positivi, circa del 50%. Tale valore è calcolato dal positive predictive value (PPV):

$$PPV = \frac{n \text{ falsi positivi}}{n \text{ veri positivi} + n \text{ falsi positivi}} \quad (7.3)$$

$$FDR = 1 - PPV \quad (7.4)$$

I risultati riportati successivamente sono stati ottenuti adottando come soglie $g = 12$ e $b = 85$. Al fine di ridurre la percentuale di falsi negativi e quindi ridurre il FOR è sufficiente abbassare il valore della soglia g , con la conseguenza che si riduce il numero di clienti accettati, però tale valore è più "pulito", ossia ci saranno pochi cattivi pagatori. Al contrario, per ridurre il numero di falsi positivi si alza il valore della soglia b , riducendo il numero di clienti respinti, ma allo stesso tempo riducendo il numero di clienti rifiutati che sarebbero stati buoni. Per tutti quei clienti che si posizionano in mezzo tra le due soglie il modello non è in grado di effettuare una classificazione abbastanza accurata, pertanto si richiede la contribuzione dei dati dai credit bureau.

7.4 Training e test

Parametrizzato il modello si passa alla fase di training, in cui il modello effettua un'analisi multivariata assegnando uno score sulla base di tutto il set di informazioni disponibile, definendo automaticamente un peso per ogni informazione. In particolare, a ciascun campione ottenuto mediante la tecnica dell'undersampling viene applicato ed allenato un algoritmo tra random forest e neural network. Avendo ottenuto dalla fase di undersampling 19 sotto campioni, pertanto si otterranno 19 sotto modelli distinti per ogni campione. Lo score assegnato ad ogni cliente sarà la media tra gli score ottenuti per ognuno dei 19 sotto modelli.

La fase di test consiste nell'applicare gli algoritmi derivati dalla fase di training ai nuovi dati, quelli del test set, ottenendo anche per essi uno score medio. Si svolgono a questo punto 4 test indipendenti di cross validazione, rieseguente le fasi di training e validazione mediante 4 campionamenti casuali indipendenti.

Il risultato dell'applicazione del modello è riportato nella matrice di confusione 7.2 seguente:

Comportamento reale	Previsioni del modello		
	Accettati	Contribuzione Bureau	Rifiutati
G	35,222	102,041	1,926
B	116	5,420	1,788

Tabella 7.2: Matrice di confusione dei risultati sul campione di validazione

Sulle righe della tabella sono riportati i risultati reali osservati, per cui se il cliente è stato un buon o cattivo pagatore, la prima riga riporta i buoni pagatori G = "good", mentre la seconda riporta i cattivi pagatori B = "bad". Le colonne riportano i risultati predetti dal modello in base alle soglie prescelte ($g = 12$, $b = 85$), per cui la prima colonna riporta i clienti accettati, la terza quelli rifiutati mentre la seconda contiene quei clienti per cui non è stato possibile effettuare una classificazione e per cui viene richiesto il contributo ai credit bureau. In ogni cella viene riportato il numero di clienti dato dall'intersezione tra risultato osservato e risultato previsto dal modello, ad esempio la prima cella 35,222 (prima riga e prima colonna) riporta il numero di clienti che si sono rivelati buoni pagatori e che contemporaneamente il modello ha giudicato come buoni clienti. Un punto di attenzione va alle celle evidenziate in giallo che rappresentano le errate classificazioni del modello. In particolare, la cella relativa a G - Rifiutati riporta il numero di clienti per cui il modello avrebbe previsto un comportamento "bad", ma per cui si è osservato un comportamento "good" (falsi positivi), mentre la cella relativa a B - Accettati riporta il numero di clienti per cui il modello avrebbe previsto un comportamento "good", ma per cui si è osservato un comportamento "bad" (falsi negativi).

A seguito delle soglie scelte ed all'applicazione del modello sul test set, facendo riferimento ai risultati della tabella sopra si osserva che:

- I falsi negativi sono 116 su 35,338 pratiche accettate, ottenendo un FOR = 0,33%;
- I veri positivi sono 1,788 su 3,714 pratiche rifiutate, ottenendo un FDR = 52%;
- Per 39,052 pratiche su 107,461, cioè il 26,65%, non è stata necessaria la chiamata ai credit bureau.

In base alle soglie scelte sono molto poche le pratiche rifiutate dal modello, appena il 3,46% del totale delle pratiche, identificando correttamente il 24,4% delle pratiche anomale totali

osservate (7,324). Questo risultato è dovuto al fatto che il campione di validazione è un sottoinsieme delle pratiche accettate dal modello di Sella Personal Credit, quindi tale campione è intrinsecamente sbilanciato su pratiche buone.

Si riporta nella tabella 7.3 la disaggregazione dei risultati predittivi del modello per prodotto PA, PF e PP:

Prodotto	Comportamento reale	Previsioni del modello (test set)		
		Accettati	Contribuzione Bureau	Rifiutati
PA	G	664	9789	237
PA	B	4	478	241
PF	G	34332	84909	955
PF	B	112	4205	659
PP	G	226	7343	734
PP	B	0	737	888

Tabella 7.3: Disaggregazione dei risultati del test set per prodotto (PA, PF e PP)

Il segmento più ampio ed interessante è sicuramente quello relativo al prodotto PF, che compone da solo oltre l'85% di tutto il portafoglio del campione di validazione. In entrambe le tabelle sono stati evidenziati in giallo gli errori di classificazione, nella prima colonna i falsi negativi, nella terza colonna i falsi positivi.

Un'ulteriore disaggregazione dei risultati viene riportata nella tabella 7.4 in cui i risultati sono presentati per sottoprodotto, limitandosi però a quelli con più occorrenze.

Descrizione prodotto	Comportamento reale	Previsioni del modello		
		Accettati	Contribuzione Bureau	Rifiutati
Auto	G	610	8538	202
	B	3	398	214
PF - Bianco	G	5939	15191	327
	B	22	860	272
PF - Bruno	G	28321	68575	608
	B	89	3307	370
Vendita fuori sede privati	G	61	955	18
	B	1	31	14
prestiti clienti promozione	G	27	939	43
	B	0	92	48
prestiti pers. nuovi clienti	G	9	1854	413
	B	0	284	488
Consolidamento nuovi clienti	G	2	607	111
	B	0	71	138
PP – Clienti Banca	G	39	1251	12
	B	0	72	26

Tabella 7.4: Risultati del modello divisi per sottoprodotto

I valori di false omission rate e false discovery rate possono variare a seconda della segmentazione considerata, assumendo valori più alti o più bassi rispetto a quelli riportati precedentemente, ma tali valori dipendono dalla soglia applicata, pertanto si possono individuare differenti soglie per prodotto, sottoprodotto o aggregazioni di questi, al fine di variare le strategie di business, aumentando o riducendo i clienti accettati e quelli rifiutati.

7.4.1 Reject Inference

Per correggere la distorsione introdotta nel modello dallo sviluppo sulle sole pratiche accettate, si introduce una metodologia di reject inference, ossia il modello viene applicato alla popolazione rifiutata da SPC per cui non si conosce il ciclo di vita del prestito, assumendo che tutte le pratiche respinte dal modello di SPC appartengano effettivamente a cattivi pagatori. Questa seconda validazione è stata effettuata sulle 197,712 appartenenti al portafoglio CO che sono state respinte dal 2011 e nella tabella 7.5 sono riportati i risultati ottenuti distinti per prodotto e suddivisi per anno.

Previsione sulle pratiche respinte da Sella Personal Credit					
Tipo prodotto	Anno richiesta	Accettati	Contribuzione Bureau	Rifiutati	Errore commesso
PA – prestito auto	2011	0	1170	330	0.00%
	2012	3	1105	244	0.22%
	2013	1	1435	242	0.06%
	2014	0	1984	223	0.00%
	2015	4	1780	149	0.21%
	2016	3	1471	119	0.19%
	2017	4	1606	129	0.23%
	2018	3	1071	69	0.26%
PF – prestito finalizzato	2011	121	7663	1901	1.25%
	2012	227	8036	2003	2.21%
	2013	440	10383	2138	3.39%
	2014	275	9841	1981	2.27%
	2015	665	17235	1917	3.36%
	2016	526	12380	1239	3.72%
	2017	474	11093	1146	3.73%
	2018	252	7321	681	3.05%
PP – prestito personale	2011	2	4330	3249	0.03%
	2012	5	7375	5359	0.04%
	2013	18	12533	6923	0.09%
	2014	30	10012	3850	0.22%
	2015	21	6810	2076	0.24%
	2016	29	6780	1468	0.35%
	2017	6	6173	1655	0.08%
	2018	7	4364	1554	0.12%

Tabella 7.5: Risultati della reject inference divisi per tipo prodotto e anno richiesta

Complessivamente i risultati ottenuti dalla reject inference constano in:

- 3,116 pratiche accettate, corrispondenti all'1,58% dei respinti da SPC;
- 40,645 pratiche rifiutate, corrispondenti al 20,55% di pratiche rifiutate da SPC;

- Le restanti pratiche prevedono la chiamata ai credit bureau per ottenere informazioni aggiuntive.

Si noti come, in questo caso, i risultati sono sbilanciati verso le pratiche respinte, tale risultato è conforme a quanto ottenuto dal campione di validazione, a sua volta invece, più orientato verso le pratiche accettate.

La colonna “Errore commesso” è stata ottenuta come il rapporto delle pratiche accettate sul totale delle pratiche in quell’anno, quindi sulla somma di tutta la riga. Facendo un’analisi per tipo prodotto si può notare che per il prestito finalizzato si registrano i valori più alti di errore, mentre per prestito auto e prestito personale si hanno mediamente gli stessi valori. Tale valore di errore è quello massimo che avrebbe compiuto il modello, in quanto l’assunto alla base della metodologia di reject inference adottata è che tutte le pratiche respinte dal modello di SPC appartengano effettivamente a cattivi pagatori, però tale assunto non è sicuramente vero, vi sono infatti altre tecniche in cui vengono considerate dalla popolazione respinta alcune pratiche appartenenti a buoni pagatori. Prendendo la metodologia del parceling, essa genera per ogni osservazione un numero casuale R estratto da una distribuzione uniforme $[0, 1]$ e lo confronta con la corrispondente PD stimata, ponendo la regola:

- $R > PD \rightarrow \text{Good}$
- $R \leq PD \rightarrow \text{Bad}$

In altre parole, ponendo ad es. $PD = 0,9$, l’osservazione avrà una probabilità del 90% di essere definita “Bad” e 10% di essere definita “Good”.

7.5 Definizione delle classi e assegnazione delle PD

Lo score di output del modello è indicativo del rischio associato ad un determinato cliente e quindi alla probabilità che esso sia un cattivo pagatore. Per ottenere la probabilità di default del cliente a partire dallo score si effettuano i seguenti step:

- si effettuano 4 test di cross validazione;
 - per ognuno dei quali si divide la scala dello score $[0,100]$ in 20 gruppi di intervalli uguali;
 - si calcola la percentuale di cattivi pagatori per ogni gruppo;
- La probabilità di default di un gruppo è data dalla media delle probabilità ottenute nei 4 test effettuati per ogni gruppo.

Di seguito si riportano i risultati ottenuti dai 4 test effettuati.

Test 1

Training set			Probability of default	Test set			Probability of default
Gruppo score	Numero buoni pagatori	Numero cattivi pagatori	Training set	Gruppo score	Numero buoni pagatori	Numero cattivi pagatori	Test set
00 -> 05	42484	16	0,04%	00 -> 05	10505	16	0,15%
05 -> 10	73417	76	0,10%	05 -> 10	18303	68	0,37%
10 -> 15	61266	186	0,30%	10 -> 15	15360	87	0,56%
15 -> 20	49515	252	0,51%	15 -> 20	12437	137	1,09%
20 -> 25	42354	316	0,74%	20 -> 25	10693	141	1,30%
25 -> 30	38984	362	0,92%	25 -> 30	9579	177	1,81%
30 -> 35	35700	558	1,54%	30 -> 35	8794	227	2,52%
35 -> 40	31112	719	2,26%	35 -> 40	7671	262	3,30%
40 -> 45	27559	844	2,97%	40 -> 45	6817	280	3,95%
45 -> 50	24556	960	3,76%	45 -> 50	6042	274	4,34%
50 -> 55	22523	1220	5,14%	50 -> 55	5486	331	5,69%
55 -> 60	21692	1437	6,21%	55 -> 60	5311	398	6,97%
60 -> 65	23760	2088	8,08%	60 -> 65	6015	596	9,02%
65 -> 70	21041	2659	11,22%	65 -> 70	5343	692	11,47%
70 -> 75	16767	3118	15,68%	70 -> 75	4399	725	14,15%
75 -> 80	11123	3227	22,49%	75 -> 80	2947	651	18,09%
80 -> 85	6777	2939	30,25%	80 -> 85	1831	577	23,96%
85 -> 90	3546	2622	42,51%	85 -> 90	951	492	34,10%
90 -> 95	1829	2487	57,62%	90 -> 95	481	487	50,31%
95 -> 100	667	3294	83,16%	95 -> 100	224	706	75,91%

Tabella 7.6: Risultati per training e validazione del test 1

Test 2

Training set			Probability of default	Test set			Probability of default
Gruppo score	Numero buoni pagatori	Numero cattivi pagatori	Training set	Gruppo score	Numero buoni pagatori	Numero cattivi pagatori	Test set
00 -> 05	40242	12	0,03%	00 -> 05	10075	19	0,19%
05 -> 10	67680	55	0,08%	05 -> 10	16651	47	0,28%
10 -> 15	62917	168	0,27%	10 -> 15	15739	103	0,65%
15 -> 20	48355	210	0,43%	15 -> 20	11737	112	0,95%
20 -> 25	42728	273	0,63%	20 -> 25	10696	152	1,40%
25 -> 30	38728	347	0,89%	25 -> 30	9695	151	1,53%
30 -> 35	35991	543	1,49%	30 -> 35	8937	230	2,51%
35 -> 40	31834	640	1,97%	35 -> 40	7862	257	3,17%
40 -> 45	27849	765	2,67%	40 -> 45	6926	276	3,83%
45 -> 50	24668	970	3,78%	45 -> 50	6165	291	4,51%
50 -> 55	22758	1124	4,71%	50 -> 55	5682	315	5,25%
55 -> 60	22136	1393	5,92%	55 -> 60	5445	402	6,88%
60 -> 65	23456	1951	7,68%	60 -> 65	5875	526	8,22%
65 -> 70	22082	2593	10,51%	65 -> 70	5666	725	11,34%
70 -> 75	18617	3249	14,86%	70 -> 75	4725	705	12,98%
75 -> 80	12381	3280	20,94%	75 -> 80	3298	643	16,32%
80 -> 85	7662	3049	28,47%	80 -> 85	1934	650	25,15%
85 -> 90	3963	2771	41,15%	85 -> 90	1186	506	29,91%
90 -> 95	1995	2540	56,01%	90 -> 95	579	560	49,17%
95 -> 100	712	3365	82,54%	95 -> 100	234	736	75,88%

Tabella 7.7: Risultati per training e validazione del test 2

Test 3

Training set			Probability of default	Test set			Probability of default
Gruppo score	Numero buoni pagatori	Numero cattivi pagatori	Training set	Gruppo score	Numero buoni pagatori	Numero cattivi pagatori	Test set
00 -> 05	45294	15	0,03%	00 -> 05	11288	21	0,19%
05 -> 10	61192	58	0,09%	05 -> 10	15301	38	0,25%
10 -> 15	66288	168	0,25%	10 -> 15	16636	94	0,56%
15 -> 20	51627	270	0,52%	15 -> 20	12878	114	0,88%
20 -> 25	44379	322	0,72%	20 -> 25	10913	144	1,30%
25 -> 30	41222	400	0,96%	25 -> 30	10207	188	1,81%
30 -> 35	35663	603	1,66%	30 -> 35	8847	226	2,49%
35 -> 40	30501	742	2,37%	35 -> 40	7602	244	3,11%
40 -> 45	26471	818	3,00%	40 -> 45	6650	285	4,11%
45 -> 50	23993	944	3,79%	45 -> 50	5815	319	5,20%
50 -> 55	22514	1181	4,98%	50 -> 55	5583	329	5,56%
55 -> 60	21938	1439	6,16%	55 -> 60	5352	399	6,94%
60 -> 65	23392	2084	8,18%	60 -> 65	5858	556	8,67%
65 -> 70	21675	2736	11,21%	65 -> 70	5463	653	10,68%
70 -> 75	17249	3249	15,85%	70 -> 75	4451	791	15,09%
75 -> 80	10484	3080	22,71%	75 -> 80	2780	683	19,72%
80 -> 85	6572	2805	29,91%	80 -> 85	1782	522	22,66%
85 -> 90	3723	2638	41,47%	85 -> 90	1003	461	31,49%
90 -> 95	1831	2468	57,41%	90 -> 95	522	516	49,71%
95 -> 100	679	3345	83,13%	95 -> 100	243	756	75,68%

Tabella 7.8: Risultati per training e validazione del test 3

Test 4

Training set			Probability of default	Test set			Probability of default
Gruppo score	Numero buoni pagatori	Numero cattivi pagatori	Training set	Gruppo score	Numero buoni pagatori	Numero cattivi pagatori	Test set
00 -> 05	37287	12	0,03%	00 -> 05	9233	13	0,14%
05 -> 10	71589	63	0,09%	05 -> 10	17647	60	0,34%
10 -> 15	62469	163	0,26%	10 -> 15	15549	105	0,67%
15 -> 20	48465	205	0,42%	15 -> 20	12107	113	0,92%
20 -> 25	42176	281	0,66%	20 -> 25	10319	153	1,46%
25 -> 30	39312	374	0,94%	25 -> 30	9919	168	1,67%
30 -> 35	35841	500	1,38%	30 -> 35	8913	228	2,49%
35 -> 40	31407	642	2,00%	35 -> 40	7868	247	3,04%
40 -> 45	27962	753	2,62%	40 -> 45	6920	254	3,54%
45 -> 50	24910	960	3,71%	45 -> 50	6317	308	4,65%
50 -> 55	22696	1170	4,90%	50 -> 55	5628	335	5,62%
55 -> 60	21289	1339	5,92%	55 -> 60	5267	399	7,04%
60 -> 65	23713	1935	7,54%	60 -> 65	5908	532	8,26%
65 -> 70	22681	2653	10,47%	65 -> 70	5767	693	10,73%
70 -> 75	18543	3265	14,97%	70 -> 75	4738	741	13,52%
75 -> 80	12306	3269	20,99%	75 -> 80	3133	697	18,20%
80 -> 85	7515	2992	28,48%	80 -> 85	1974	620	23,90%
85 -> 90	3927	2792	41,55%	85 -> 90	1101	533	32,62%
90 -> 95	1892	2417	56,09%	90 -> 95	521	467	47,27%
95 -> 100	778	3509	81,85%	95 -> 100	274	744	73,08%

Tabella 7.9: Risultati per training e validazione del test 4

I risultati dei 4 test effettuati, riportati tabelle precedenti, riscontrano risultati piuttosto simili in termini di probabilità di default e numerosità delle classi. Si osserva un andamento strettamente crescente della PD all'aumentare della classe sia per training set sia per test set, rispettando l'assunto secondo cui all'aumentare dello score aumenta il rischio associato a quel cliente, di conseguenza il numero di cattivi pagatori di ogni classe aumenta con lo score, viceversa il numero di buoni pagatori diminuisce. Si osserva una perdita delle performance nel passaggio da training set a test set rappresentato dall'aumento della PD nelle classi di score basso ed una riduzione delle PD nelle classi di score alto. Inoltre, a causa di questa perdita di performance l'aumento della PD con le classi, seppur coerente anche nel test set, risulta meno marcato rispetto al training set, presentando perfino un'inversione di trend, per cui fino all'intervallo di score 65 – 75 il test set riporta una PD maggiore, mentre per valori più alti di score è il training set a registrare valori più alti. Nelle tabelle sono stati evidenziati i gruppi in cui si verifica l'inversione della PD. Tale comportamento è meglio visibile nelle figure riportate di seguito, in cui si osserva come la colonna relativa alla PD del test set (in rosso) sia più alta in tutti i grafici fino all'intervallo 65 – 75 dove si verifica l'inversione del trend e la colonna del training set (in blu) diventa più alta.

Figura 7.10: Grafico delle PD del test 1

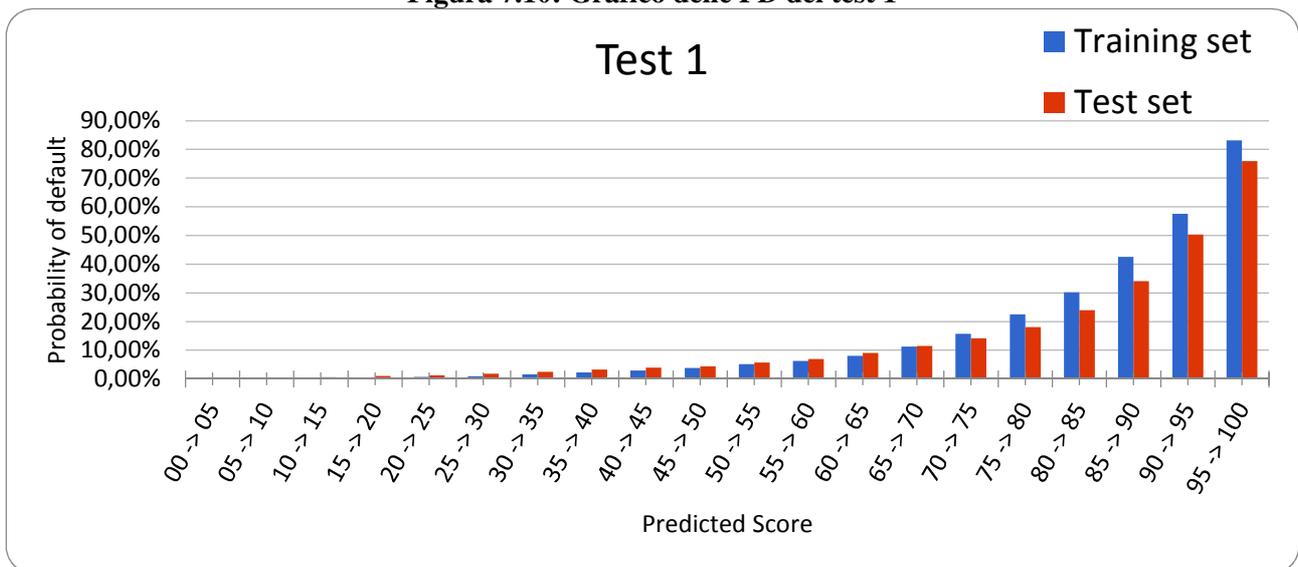


Figura 7.11: Grafico delle PD del test 2

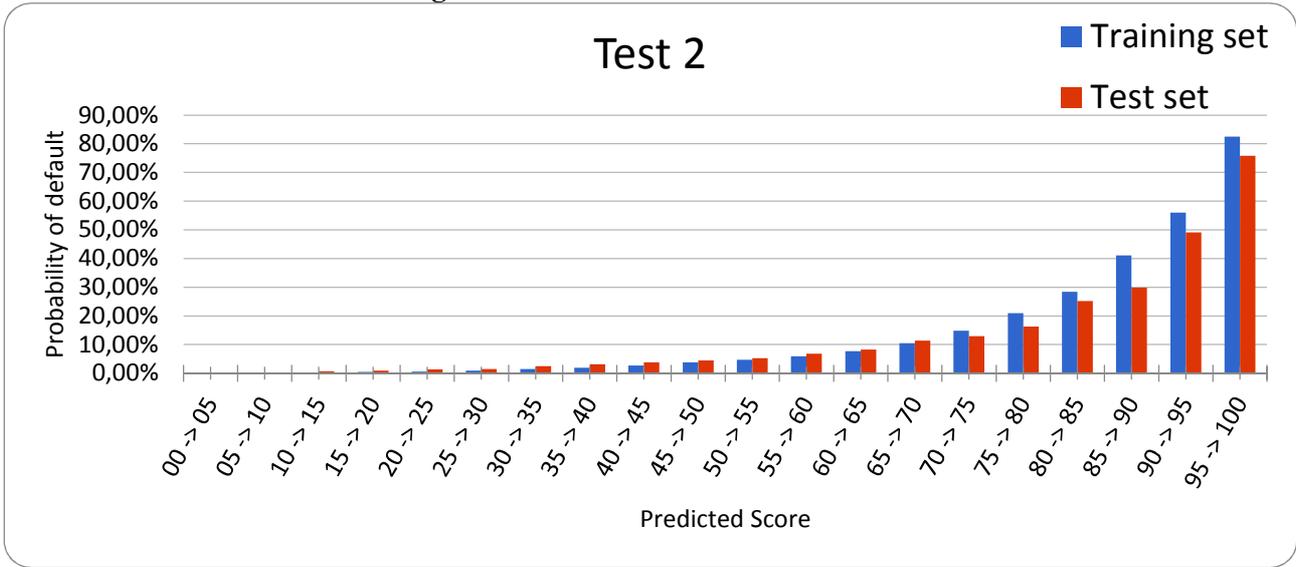


Figura 7.12: Grafico delle PD del test 3

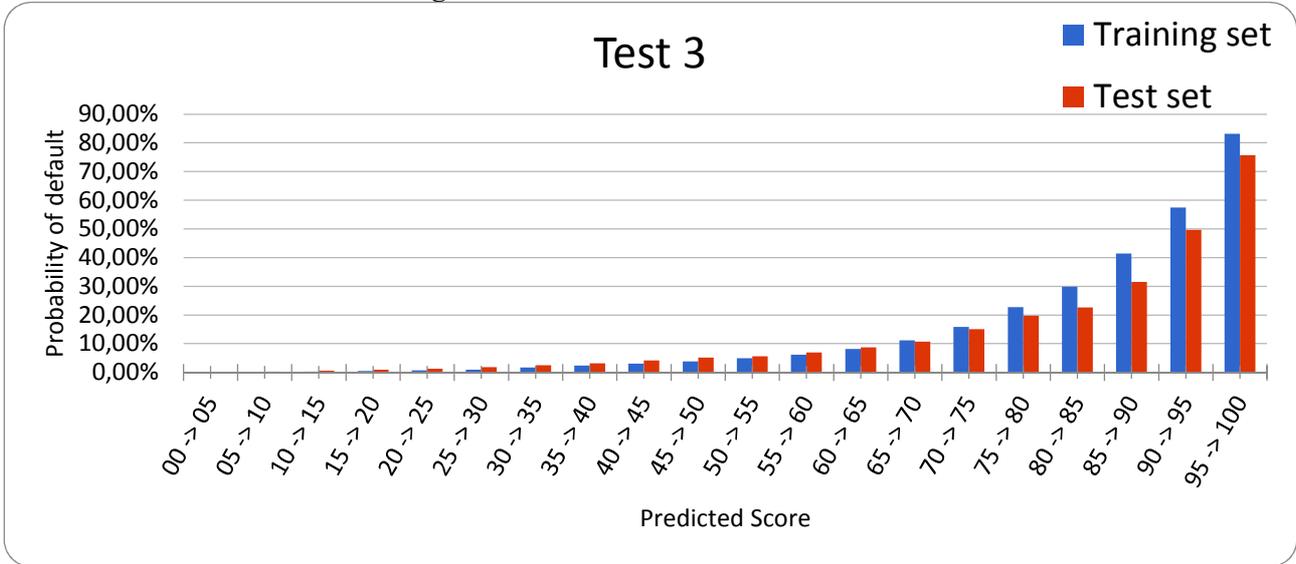
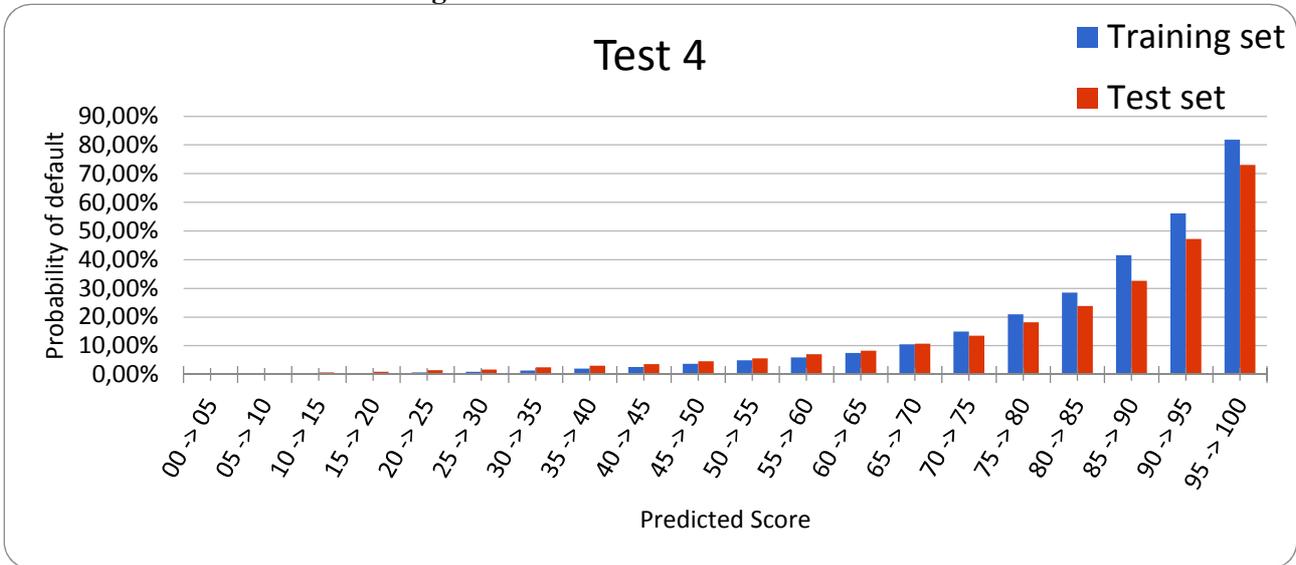


Figura 7.13: Grafico delle PD del test 4



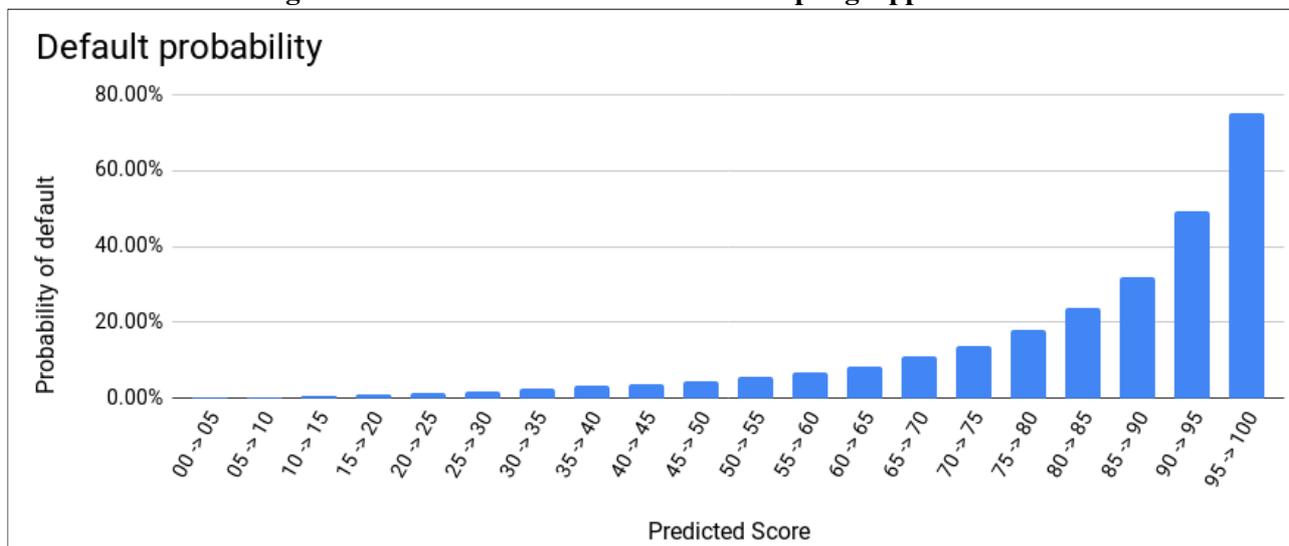
La PD assegnata ad ogni gruppo è data la media delle PD dei 4 test effettuati allo scopo di appiattare le fluttuazioni che si verificano selezionando casualmente i campioni di validazione. I risultati dei 4 test e la media complessiva sono riportati nella tabella 7.14 di seguito:

Predicted group	Probability of default				Average value
	Test 1	Test 2	Test 3	Test 4	
00 → 05	0.15%	0.19%	0.19%	0.14%	0.17%
05 → 10	0.37%	0.28%	0.25%	0.34%	0.31%
10 → 15	0.56%	0.65%	0.56%	0.67%	0.61%
15 → 20	1.09%	0.95%	0.88%	0.92%	0.96%
20 → 25	1.30%	1.40%	1.30%	1.46%	1.37%
25 → 30	1.81%	1.53%	1.81%	1.67%	1.71%
30 → 35	2.52%	2.51%	2.49%	2.49%	2.50%
35 → 40	3.30%	3.17%	3.11%	3.04%	3.16%
40 → 45	3.95%	3.83%	4.11%	3.54%	3.86%
45 → 50	4.34%	4.51%	5.20%	4.65%	4.67%
50 → 55	5.69%	5.25%	5.56%	5.62%	5.53%
55 → 60	6.97%	6.88%	6.94%	7.04%	6.96%
60 → 65	9.02%	8.22%	8.67%	8.26%	8.54%
65 → 70	11.47%	11.34%	10.68%	10.73%	11.05%
70 → 75	14.15%	12.98%	15.09%	13.52%	13.94%
75 → 80	18.09%	16.32%	19.72%	18.20%	18.08%
80 → 85	23.96%	25.15%	22.66%	23.90%	23.92%
85 → 90	34.10%	29.91%	31.49%	32.62%	32.03%
90 → 95	50.31%	49.17%	49.71%	47.27%	49.11%
95 → 100	75.91%	75.88%	75.68%	73.08%	75.14%

Tabella 7.14: PD per ogni gruppo di score nei 5 test di cross-validation e PD media

Nella figura 7.15 è riportata la rappresentazione grafica della distribuzione della PD media per gruppo di score della tabella precedente:

Figura 7.15: Distribuzione delle PD medie per gruppo di score



7.6 Validazione del modello

La fase di validazione del modello è stata eseguita in accordo con le policy interne di SPC. Nelle tabelle seguenti sono riportate le metriche di performance calcolate per training set e test set, il misclassification rate è calcolato sulla base delle soglie adottate e considerando solo le pratiche classificate tra accettati e rifiutati:

Metrica	Sviluppo	Esito	Validazione	Esito
Accuracy Ratio	89,52%	VERDE	84,90%	VERDE
AUROC	79,04%	VERDE	69,81%	VERDE
Misclassification Rate			5,23%	VERDE

Tabella 7.16: Analisi di performance test 1

Metrica	Sviluppo	Esito	Validazione	Esito
Accuracy Ratio	89,57%	VERDE	84,81%	VERDE
AUROC	79,14%	VERDE	69,62%	VERDE
Misclassification Rate			2,58%	VERDE

Tabella 7.17: Analisi di performance test 2

Metrica	Sviluppo	Esito	Validazione	Esito
Accuracy Ratio	89,44%	VERDE	85,26%	VERDE
AUROC	78,87%	VERDE	70,51%	VERDE
Misclassification Rate			8,77%	VERDE

Tabella 7.18: Analisi di performance test 3

Metrica	Sviluppo	Esito	Validazione	Esito
Accuracy Ratio	89,60%	VERDE	84,69%	VERDE
AUROC	79,20%	VERDE	69,37%	VERDE
Misclassification Rate			10,10%	GIALLO

Tabella 7.19: Analisi di performance test 4

Nella tabella seguente sono riportati i gli esiti dei test di stabilità, in cui il PSI è stato calcolato confrontando le distribuzioni nei diversi gruppi tra i campioni di sviluppo e di validazione, per il check sulla monotonicità si faccia riferimento alle tabelle da 7.6 a 7.9 ed ai grafici da 7.10 a 7.13.

Metrica	Validazione	Esito
PSI	2,94%	VERDE
Check monotonicità	-	VERDE

Tabella 7.20: Analisi stabilità - test 1

Metrica	Validazione	Esito
PSI	1,91%	VERDE
Check monotonicità	-	VERDE

Tabella 7.21: Analisi stabilità - test 2

Metrica	Validazione	Esito
PSI	1,66%	VERDE
Check monotonicità	-	VERDE

Tabella 7.22: Analisi stabilità - test 3

Metrica	Validazione	Esito
PSI	2,00%	VERDE
Check monotonicità	-	VERDE

Tabella 7.23: Analisi stabilità - test 4

Secondo le policy di SPC, il giudizio finale della validazione del modello è adeguato.

7.7 Osservazioni

I risultati ottenuti in questa fase esplorativa di applicazione del machine learning e predictive analysis al credit scoring sono molto interessanti e offrono notevoli spunti per effettuare ulteriori approfondimenti.

Altre applicazioni del modello sono state testate per saggiarne le performance in condizioni differenti, ad esempio utilizzando una singola soglia, più alta o più bassa a seconda della strategia che si intende adottare, oppure utilizzando valori differenti per la strategia a due soglie. In tutti i casi analizzati l'applicazione è stata effettuata su tutto il portafoglio del consumo, ottenendo risultati altrettanto interessanti ed offrendo notevoli opportunità per approfondimenti su specifici prodotti o sottoprodotti.

Un caso applicativo di particolare interesse è stato ottenuto utilizzando come campione per il training set l'insieme di pratiche aperte prima del 2016 (circa l'85% del campione totale) e per il test set le pratiche successive. Mediante tale metodologia si è voluto simulare

l'applicazione del modello ottenuto dall'addestramento sulle pratiche antecedenti il 2016 ed applicarlo sulle nuove pratiche aperte dal 2016 al 2018.

Il limite di questa esplorazione è legato al fatto che è stata effettuata su tutto il portafoglio del consumo, considerando di pari entità l'impatto generato da un cattivo pagatore con un finanziamento più alto rispetto ad uno con un finanziamento più basso e non effettuando distinzioni per tipologia di prodotto acquistato o di prestito richiesto. Inoltre, in tale trattazione non si tiene conto del fatto che un cliente classificato come cattivo abbia eventualmente regolarizzato la sua posizione, oppure che uno definito buono abbia presentato spesso degli insoluti. Ulteriori approfondimenti anche su queste considerazioni possono affinare ancora di più il modello e renderlo specifico per determinati segmenti di prodotti e per differenti strategie di business.

8 Conclusioni

I temi di intelligenza artificiale e machine learning negli ultimi anni stanno assumendo un ruolo sempre di maggiore importanza per molte aziende in svariati settori. Questa ampia applicabilità del machine learning è dovuta al fatto che i computer possono imparare ad eseguire delle operazioni basandosi sull'osservazione dei dati ed imparando da essi, indipendentemente dal tipo di attività o di dati. Questo apprendimento automatico è favorito dalla sempre maggiore disponibilità di dati in termini di volumi e varietà e da strumenti di calcolo sempre più potenti. Nel contesto bancario ed in particolare nel risk management il machine learning trova applicazione nelle seguenti aree:

- Rischio di credito: applicazione alla stima della PD;
- Rischio di mercato: monitoraggio del rischio di portafoglio;
- Rischio operativo: in termini di intercettazione delle frodi e antiriciclaggio;
- Data quality.

Nell'ambito dell'elaborato è stata trattata l'applicazione di tecniche di machine learning alla valutazione del credit scoring, attraverso il quale si vuole fornire un giudizio sintetico sul grado di solvibilità del cliente, cd. probability of default (PD).

L'elaborato riporta il proof of concept del progetto, ovvero la realizzazione di una bozza di modello con lo scopo di verificare che l'applicazione del machine learning al caso reale di SPC per la valutazione del rischio di credito potesse portare reali benefici. Per questo motivo, l'applicazione non è stata fin da subito effettuata su determinati prodotti o messo a confronto con uno dei modelli interni attualmente in uso, ma è stata fatta una prima esplorazione di applicazione del modello al segmento più ampio del portafoglio, il consumo CO, questo perché, come suddetto, l'apprendimento della macchina sarà tanto migliore tanto maggiore è il volume dei dati a disposizione. Quanto detto però evidenzia un limite degli algoritmi di machine learning, ovvero che se la disponibilità di dati non è sufficientemente ampia il modello generato avrà uno scarso potere previsionale ed inoltre tenderà a adattarsi troppo ai pochi dati disponibili. Pertanto, l'applicazione di tecniche di data science anche agli altri portafogli CA (carte di credito, linee di credito) e CQ (cessione del quinto dello stipendio e della pensione) potrebbe non rivelarsi altrettanto proficua ed in maniera analoga se lo sviluppo venisse fatto su un segmento del consumo con pochi dati a disposizione.

Il PoC è stato sviluppato a partire dal dataset a disposizione di SPC testando differenti algoritmi di machine learning e combinazioni di questi, al fine di individuare un modello adatto all'ambito del credit scoring. Le possibili combinazioni di tecniche e le differenti parametrizzazioni di queste sono molteplici, ma l'applicazione di ognuna di queste richiede molto tempo, pertanto non tutte le possibili alternative sono state testate ed il risultato proposto è quello che ha riportato i migliori risultati in termini di performance.

Il modello identificato ha riportato risultati promettenti, aprendo una strada per ulteriori approfondimenti ed affinamenti degli algoritmi ottenuti. Gli sviluppi successivi interesseranno i segmenti più ampi del portafoglio del consumo, per i quali i prossimi passi previsti si articolano in:

- Trattamento di data quality: gestione di dati errati o missing;
- Ottimizzazione dell'algoritmo;
- Fine tuning degli hyper-parameters;
- Inclusione dei dati innovativi.

Il modello così ottenuto verrà poi testato attraverso un periodo di parallel running con i modelli interni di SPC, in cui i due modelli opereranno simultaneamente per un certo periodo di tempo. Questo periodo di affiancamento serve a garantire che il nuovo modello funzioni correttamente in tutti i suoi aspetti e nel caso in cui si identificassero errori di funzionamento si può mantenere l'operatività attraverso il modello precedente mentre si risolvono i problemi riscontrati. Alla fine del periodo di parallel running il modello sostituirà totalmente quello precedente. Il modello verrà integrato nei sistemi di SPC congiuntamente ad un software, che permetterà di gestire il modello senza dover apportare modifiche al codice Python, quindi senza richiedere conoscenze specifiche di machine learning o di sviluppo di codice e l'introduzione di una dashboard dedicata al monitoraggio e controllo.

L'introduzione del modello porterà benefici sia in termini economici diretti, facendo risparmiare il costo dovuto all'interrogazione delle banche dati dei credit bureau, sia indiretti, riducendo il numero di cattivi pagatori a cui viene concesso un prestito.

9 Bibliografia

- William H. Beaver**, “Financial Ratio as Predictors of Failure”, *Journal of Accounting Research*, Vol.4, Issue Empirical Research in Accounting, 1966
- R.A. Fisher**, “The use of multiple measurements in taxonomic problems”, *Annals of eugenics*, 1936
- E.I. Altman, R. Haldeman, P. Narayanan**, “ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations”, *Journal of Banking and Finance*, June 1977
- Tom Fawcett**, “An introduction to roc analysis. *Pattern Recognition Letters*”, 27(8):861 – 874, 2006
- Bilal Yurdakul**, “Statistical Properties of Population Stability Index (PSI)”. PhD thesis, 05 2018
- Khashman A**, “Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes”, 2010
- Lessmann S., Baesens B., Seow H. V., Thomas L.C.**, “Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research”, *European Journal of Operational Research* 247: 124-136, 2015
- Khandani A. E., Kim J., LO A.W.**, “Consumer credit-risk models via machine-learning algorithms”, *Journal of Banking & Finance*, 2010
- Efron, B. and Tibshirani, R.**, “An Introduction to the Bootstrap”. Chapman and Hall, New York, London, 1993
- Breiman L.**, “Random Forests”, *Machine Learning* Vol. 45:5–32, 2001
- Malley J., Kruppa J., Dasgupta A., Malley K., Ziegler A.**, Probability machines. Consistent probability estimation using nonparametric learning machines. *Methods Inf. Med* 51 (1): 74–81, 2012
- Hutter, Frank, Holger Hoos, and Kevin Leyton-Brown**, "An efficient approach for assessing hyperparameter importance.", *International Conference on Machine Learning*, 2014
- Tobias Berg, Valentin Burg, Ana Gombović, and Manju Puri**, “On the rise of fintechs – credit scoring using digital footprints”, Working Paper 24551, National Bureau of Economic Research, April 2018