LUISS

Dipartimento di
Economia e Finanza

Cattedra di Economia e Gestione degli Intermediari Finanziari
(corso progredito)

SVILUPPO DI UN PRESCORE ADATTIVO PER MIGLIORARE L'EFFICIENZA E L'EFFICACIA DEL PROCESSO DI VALUTAZIONE DEL CREDITO: IL CASO STUDIO FIN.PROMO.TER

Prof. Domenico Curcio
RELATORE

Prof. Giancarlo Mazzoni
CORRELATORE

Roberta Romano
Matr. 780451
CANDIDATO

Anno Accademico 2024/2025

Sommario

Introduzione

Il processo di valutazione del credito riveste un ruolo centrale nell'operatività degli intermediari finanziari, rappresentando uno degli elementi cruciali per la gestione del rischio e l'allocazione efficiente delle risorse. L'importanza di tale processo si è intensificata negli ultimi anni a causa dell'evoluzione normativa, dell'adozione di strumenti tecnologici avanzati e della crescente concorrenza nel settore bancario e finanziario.

Questa tesi si propone di analizzare in modo approfondito il processo di valutazione del credito negli intermediari finanziari, con un *focus* particolare sul ruolo strategico del *prescore*, uno strumento essenziale per garantire efficienza e accuratezza nelle decisioni di credito. Il contesto attuale è caratterizzato dall'adozione di normative stringenti come Basilea III, che richiedono agli intermediari di adottare modelli robusti di gestione del rischio, e dall'incremento dell'uso di algoritmi di intelligenza artificiale e *machine learning* per migliorare la capacità predittiva e ridurre i costi.

La rilevanza del tema è evidenziata dall'aumento delle esposizioni creditizie e dalla necessità di ottimizzare i processi per evitare insolvenze e perdite significative. L'obiettivo principale di questo lavoro è quello di offrire un'analisi completa delle fasi che compongono il ciclo di vita di una pratica di credito, approfondendo il ruolo che il *prescore* svolge all'interno di questo processo e analizzando le inefficienze esistenti, i rischi connessi e le possibili strategie di miglioramento.

La tesi si articola in diverse sezioni che comprendono una revisione della letteratura esistente, l'analisi di un caso studio reale, l'utilizzo di dati finanziari raccolti tramite API e altre fonti, e la proposta di soluzioni per ottimizzare il processo di valutazione del credito. Particolare attenzione sarà dedicata all'impatto delle tecnologie digitali e all'utilizzo di modelli di *machine learning*, che hanno trasformato il modo in cui gli intermediari valutano il rischio creditizio, rendendo il processo più veloce, accurato e meno costoso. Il lavoro si inserisce in un dibattito più ampio sulla modernizzazione dei processi bancari, contribuendo con un'analisi empirica basata su dati reali e fornendo suggerimenti pratici per l'implementazione di modelli di *prescore* adattivi. L'obiettivo finale è quello di offrire un quadro chiaro delle sfide e delle opportunità legate alla valutazione del credito, supportando gli intermediari finanziari nell'adozione di *best practice* per migliorare la loro competitività e solidità.

Questa tesi si avvale di una vasta gamma di fonti, tra cui testi accademici, articoli di riviste scientifiche e documentazione tecnica, che offrono una visione completa del tema. L'analisi empirica basata su un caso studio reale garantirà un contributo significativo sia dal punto di vista teorico che pratico, fornendo strumenti utili per migliorare l'efficienza e l'efficacia del processo di valutazione del credito.

Capitolo 1: Valutazione del rischio di credito negli intermediari finanziari

1.1 Introduzione al rischio di credito

Il rischio di credito rappresenta la possibilità che un debitore non riesca a rispettare gli obblighi finanziari contrattuali, compromettendo il recupero del capitale prestato e degli interessi associati. Questo rischio può assumere diverse forme, includendo sia l'insolvenza manifesta che un semplice deterioramento della qualità creditizia della controparte, che può comunque influenzare negativamente il valore delle attività finanziarie detenute dall'istituto di credito.

Il rischio di credito può essere suddiviso in varie categorie, ciascuna caratterizzata da specifici elementi di incertezza:

1. Rischio di insolvenza. Si verifica quando il debitore non è più in grado di rispettare i propri obblighi contrattuali, comportando una perdita per l'ente finanziatore. Questa perdita è calcolabile come il prodotto tra l'esposizione al momento del default (EAD, *Exposure at Default*) e il tasso di perdita in caso di insolvenza (LGD, *Loss Given Default*).
2. Rischio di migrazione o di *downgrading*. Si riferisce al rischio che il merito creditizio del debitore peggiori nel tempo, riducendo il valore dell'attivo creditizio e potenzialmente aumentando il costo del capitale.
3. Rischio di *spread*. Riguarda la possibilità che, a causa di una maggiore avversione al rischio da parte degli investitori, gli spread richiesti per il finanziamento aumentino, riducendo il valore delle esposizioni anche in assenza di cambiamenti nel *rating* della controparte.
4. Rischio di recupero. Si manifesta quando gli importi effettivamente recuperabili da un debitore insolvente risultano inferiori alle aspettative, spesso a causa di svalutazioni o tempi di recupero più lunghi del previsto.
5. Rischio di esposizione. Rappresenta la possibilità che l'esposizione effettiva aumenti in prossimità del *default*, ad esempio a causa dell'utilizzo delle linee di credito non ancora sfruttate.
6. Rischio di pre-regolamento o di sostituzione. Tipico delle transazioni OTC (*over-the-counter*), si verifica quando la controparte fallisce prima della data di

regolamento, costringendo a sostituire il contratto con uno a condizioni meno favorevoli.

7. Rischio Paese. Emerge quando le controparti estere non riescono a rispettare i propri obblighi a causa di eventi politico-legislativi, come controlli valutari, sanzioni economiche o moratorie.

La gestione del rischio di credito richiede un'attenta considerazione dei fattori macroeconomici, delle caratteristiche specifiche del debitore e delle condizioni contrattuali. In particolare, è fondamentale distinguere tra perdita attesa (*Expected Loss*, EL) e perdita inattesa (*Unexpected Loss*, UL), due concetti chiave per valutare correttamente il profilo di rischio.

Perdita attesa (EL, expected loss) - Rappresentata dal valore medio della distribuzione delle perdite. Questa perdita viene considerata già nella fase di *pricing* del prestito e compensata attraverso l'applicazione di uno *spread*. In altre parole, se la perdita si verifica come previsto, il rendimento netto sarà in linea con quanto anticipato dal creditore. Quindi, sebbene la perdita attesa sia contabilmente significativa, non rappresenta un rischio in senso stretto.

La stima della perdita attesa di un'esposizione creditizia richiede a sua volta di stimare tre parametri[1]:

$$EL = \overline{EAD} \cdot PD \cdot \overline{LGD}$$

1. L'***Exposure at Default*** (EAD). Rappresenta l'importo che ci si aspetta di perdere in caso di *default*. Include la quota di credito utilizzata (*Drawn Portion*, DP) e la quota inutilizzata ma potenzialmente utilizzabile (*Undrawn Portion*, UP), moltiplicata per un *Credit Conversion Factor* (CCF), che stima la percentuale di fido inutilizzato che potrebbe essere impiegato in prossimità dell'insolvenza dell'insolvenza[2].

$$EAD = DP + UP \cdot CCF$$

2. ***Probability of Default*** (PD). Misura la probabilità che la controparte diventi insolvente entro un determinato periodo di tempo.

[1] Si ipotizza, per semplicità, che i tre fattori di rischio riassunti (rischio di esposizione, rischio di *default*, rischio di recupero) siano indipendenti. Se così non fosse, la stima della PD e dei valori attesi di EAD e LGD non sarebbe sufficiente per ricavare la perdita attesa, poiché occorrerebbe conoscere anche le covarianze tra i diversi fattori di rischio.

[2] Un'analisi empirica relativa al mercato statunitense Asarnow e Marker (1995) mostra che i dati relativi al CCF (talvolta noto anche come LEQ, cioè *loan equivalent*) sono compresi in un *range* che va dal 40 per cento al 75 per cento circa. Sulla stima dei CCF e del rischio di esposizione, cfr. anche Araten e Jacobs (2001) e Moral (2006).

3. La ***Loss Given Default*** (LGD), la quale quantifica la porzione dell'esposizione che non sarà recuperabile in caso di default e si calcola come il complemento a uno del *Recovery Rate* (RR), ovvero il tasso di recupero delle somme recuperate al netto dei costi amministrativi.

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

Il recovery *rate* è il valore attuale delle somme recuperate (ER) nei vari tempi al netto dei costi amministrativi (AC) espresso in percentuale dell'EAD.

Perdita inattesa (UL, unexpected loss) - Il rischio di credito, inteso come la possibilità che la perdita effettiva risulti superiore rispetto alla stima iniziale, è strettamente collegato al concetto di perdita inattesa. In termini generali, la perdita inattesa può essere descritta come la deviazione della perdita dal suo valore medio, ovvero dalla perdita attesa (EL). La distinzione tra perdita attesa e perdita inattesa assume particolare importanza quando si analizza un portafoglio di impieghi. Infatti, mentre la perdita attesa complessiva del portafoglio corrisponde alla somma delle perdite attese dei singoli impieghi, la variabilità totale delle perdite nel portafoglio è generalmente inferiore alla somma delle singole variabilità, soprattutto se la correlazione tra gli impieghi è ridotta.

Le fluttuazioni dei mercati finanziari globali, influenzate da fattori macroeconomici e dall'efficienza complessiva del sistema finanziario, contribuiscono alla definizione di rischio generico o sistemico, in quanto non è possibile eliminarlo. Questo tipo di rischio è legato alla variabilità indotta dai cambiamenti dei mercati, che hanno un impatto generalizzato su tutti i titoli. Durante i periodi di recessione economica, è probabile che anche gli investimenti più diversificati risentano della situazione sfavorevole.

In contrapposizione, il rischio specifico è associato a caratteristiche particolari dell'investimento, variabili caso per caso. Mentre il rischio sistemico non può essere eliminato, il rischio specifico può essere mitigato attraverso la diversificazione del portafoglio, riducendo l'esposizione a singoli investimenti e compensandola con altri titoli. Tuttavia, raggiungere una diversificazione effettiva non è sempre semplice, poiché i tradizionali concetti di "frammentazione" e "concentrazione"[3] del rischio non sempre rappresentano accuratamente la realtà. Ad esempio, strumenti di misurazione come

[3] numero di crediti presenti in portafoglio e al loro peso in relazione al valore dell'intero portafoglio.

l'indice di concentrazione di Hischmann-Herfindhal[4] e l'indice delle quattro maggiori imprese forniscono una stima della concentrazione, ma non tengono conto della correlazione tra gli elementi del portafoglio. Pertanto, due portafogli con lo stesso grado di concentrazione potrebbero presentare livelli di diversificazione differenti se le controparti si muovono in modo scarsamente correlato, evidenziando così l'importanza di considerare il grado di correlazione con il mercato.

1.2 Il Ruolo strategico del *prescore* nella prevenzione del rischio di credito

Per gli istituti finanziari la necessità di instaurare rapporti stabili e duraturi con la propria clientela rappresenta un elemento strategico per garantire vantaggi competitivi e migliorare la redditività nel lungo periodo. Difatti, i clienti fidelizzati tendono ad essere meno influenzati dalle offerte dei concorrenti, garantendo flussi di reddito più stabili e presentando, nel tempo, un rischio operativo inferiore rispetto alla media di mercato. Negli ultimi anni, la crescente competenza finanziaria dei clienti e la sempre maggiore concorrenza nel settore bancario hanno spinto le istituzioni finanziarie a focalizzarsi maggiormente sulla costruzione di relazioni di lungo termine, attraverso iniziative specifiche mirate a consolidare la fedeltà della clientela e a ottimizzare l'offerta dei prodotti presenti in portafoglio. In questo contesto, i sistemi di *prescreening* rappresentano strumenti fondamentali per identificare i clienti ad alto potenziale, ovvero quelli che mostrano una bassa probabilità di rischio e una forte propensione all'acquisto dei prodotti offerti. Questi sistemi si basano sull'analisi dei dati comportamentali accumulati dai clienti nel corso delle loro attività finanziarie, permettendo di prevedere con maggiore precisione i profili di rischio e le opportunità di *cross-selling*.

L'efficacia di un sistema di *prescreening* dipende da diverse componenti chiave, tra cui:

- Sistema di *scoring*: collega le caratteristiche dei clienti a una variabile obiettivo che rappresenta l'evento da prevedere (ad esempio, l'insolvenza o il rischio di *default*), consentendo di assegnare un punteggio (*score*) che misura la probabilità di eventi critici nel comportamento di pagamento. Questo processo permette di

[4] L'indice Herfindahl-Hirschman è un indice che misura la concentrazione di mercato di un settore. L'indice viene ottenuto sommando le quote di mercato delle imprese del settore al quadrato: $HHI = \sum_{i=0}^{n} Q_i^2$ dove Q indica la quota di mercato dell'impresa i. Più è basso il valore dell'indice più il mercato è competitivo. Il valore massimo indica la condizione di monopolio.

suddividere il portafoglio in segmenti omogenei, facilitando l'adozione di strategie di gestione differenziate.

- Profilazione della clientela: analizza le caratteristiche sociodemografiche e comportamentali dei clienti per definire gruppi omogenei con esigenze e potenzialità simili.
- Strategie di contatto: determinano le modalità e i tempi ottimali per interagire con i clienti, massimizzando l'efficacia delle iniziative commerciali.
- Processo di *test & learn*: consente di valutare l'efficacia delle strategie adottate, identificando le più performanti e adattando continuamente l'approccio in base ai risultati ottenuti.

Il sistema di *scoring*, in particolare, è un modello statistico multivariato che associa a ciascun cliente un punteggio rappresentativo della probabilità di insolvenza, calcolato in base a variabili economico-finanziarie rilevanti. Questo punteggio viene utilizzato per classificare i clienti in categorie di rischio, rendendo possibile l'ottimizzazione delle strategie di gestione del portafoglio. L'integrazione dei punteggi provenienti da diversi modelli può essere visualizzata in mappe di segmentazione, che costituiscono strumenti essenziali per identificare le opportunità di crescita e le aree di rischio nel portafoglio clienti.

Nonostante le tecniche fondamentali alla base dei modelli di *scoring* siano state sviluppate già negli anni Trenta, grazie ai lavori pionieristici di Fisher (1936) e Durand (1941), è stato solo a partire dagli anni Sessanta, con i contributi di Beaver (1967), Altman (1968) e altri ricercatori[5], che questi strumenti hanno trovato ampia applicazione nel contesto finanziario.

La costruzione di un modello di *scoring* richiede l'impiego di un campione rappresentativo, composto da controparti che in passato hanno manifestato comportamenti finanziari affidabili o, al contrario, segnali di insolvenza. È fondamentale evitare selezioni arbitrarie all'interno del campione, poiché ciò potrebbe introdurre distorsioni nei dati di *input*, compromettendo l'affidabilità del modello. Una volta definito il campione, si procede con l'identificazione delle variabili quantitative che meglio separano le controparti insolventi da quelle sane.

[5] Il principale autore che, negli anni Settanta, ha studiato l'applicazione alla previsione delle insolvenze dell'analisi discriminante è Edward Altman.

Successivamente, queste variabili vengono integrate in una funzione di *scoring* che assegna un punteggio (*score*) a ciascuna controparte. La logica alla base del punteggio è che le controparti finanziariamente solide tendono a ottenere punteggi più elevati, mentre quelle con una storia di insolvenza tendono a posizionarsi su livelli di score più bassi. Questo consente di definire una soglia di rischio (*cut-off*), al di sotto della quale le richieste di credito possono essere automaticamente respinte o sottoposte a una revisione manuale, in base alla propensione al rischio dell'istituto. Una soglia più elevata implica una maggiore avversione al rischio, mentre una soglia più bassa indica una maggiore tolleranza verso i clienti a rischio.

Un altro aspetto critico nella costruzione di un modello di *scoring* è la gestione dei dati anomali o atipici, che potrebbero alterare significativamente i risultati del modello. Inoltre, poiché molti dei dati utilizzati derivano dai bilanci delle imprese, è importante considerare il ritardo temporale che tali informazioni possono introdurre, dato che spesso si riferiscono a eventi avvenuti diversi mesi prima.

Per garantire la robustezza del modello nel tempo, è essenziale monitorare periodicamente le sue *performance*, verificandone l'efficacia discriminante rispetto ai cambiamenti economici e normativi. Quando l'accuratezza del modello inizia a ridursi, è necessario procedere con una sua ristima, aggiornando le variabili e i parametri utilizzati.

Infine, mentre in molti casi l'obiettivo del modello è semplicemente quello di distinguere tra controparti solvibili e insolventi (variabile binaria *default/no-default*), in altre applicazioni può essere richiesto un monitoraggio più complesso, che tenga conto anche delle variazioni nella qualità del credito nel tempo, piuttosto che limitarsi alla sola misurazione del rischio di insolvenza.

Approccio	**Metodologia**
Metodi discrezionali	Si basa sull'esperienza sulla comprensione del valutatore per decidere se concedere o rifiutare il credito.
Sistemi esperti *(es. comitati di concessione prestiti)*	Utilizzano un approccio di gruppo per giudicare il caso o formalizzare decisioni discrezionali attraverso procedure e sistemi di concessione prestiti.
Modelli analitici	Utilizzano una serie di metodi analitici, generalmente basati su dati quantitativi, per prendere una decisione.

Modelli statistici *(es. credit scoring)*	Si basano sull'inferenza statistica per derivare relazioni appropriate utili al processo decisionale.
Modelli comportamentali	Osservano il comportamento nel tempo per derivare relazioni appropriate per prendere una decisione.
Modelli di mercato	Si basano sulle informazioni contenute nei prezzi di mercato finanziario come indicatori di solvibilità finanziaria.

Tabella 1 Differenti approcci al processo di valutazione del credito

I diversi metodi riportati nella Tabella 1 richiedono dati e/o informazioni dall'ambiente aziendale (ad esempio, rapporti aziendali, notizie, bilanci, prezzi di mercato dei titoli dell'azienda, cronologia dei pagamenti e così via). Gli approcci analitici possono essere raggruppati in modo approssimativo in (1) modelli di conoscenza, che hanno un certo grado di soggettività (ad esempio, l'uso del giudizio di esperti da parte di un analista), (2) modelli ad effetto, che combinano alcuni elementi di soggettività e analisi sistemica (l'analisi per rapporti rientrerebbe in questa categoria) e (3) modelli statistici, che possono essere considerati più sistemici nell'approccio (i modelli di *credit scoring* sono di questo tipo). I risultati dell'analisi vengono utilizzati nello spazio decisionale, vale a dire per giungere a una decisione in merito alla concessione o meno di un credito.

1.2.1 Metodi di valutazione del credito qualitativi e quantitativi

Esistono diversi approcci alla valutazione del credito. Possiamo classificare l'elenco della Tabella 1 in quattro categorie, che sono, in una certa misura, sovrapposte. Possiamo considerare questi come (1) sistemi esperti, (2) sistemi di *rating*, (3) modelli di *credit scoring* e (4) modelli basati sul mercato. In pratica, gli analisti del credito utilizzano una combinazione di metodi per valutare le imprese e prevedere la loro futura affidabilità creditizia.

Il primo metodo sono **i sistemi esperti**. Questi vanno dal semplice giudizio dell'analista del credito a modelli più formali. Tali modelli sono stati costruiti nel tempo sulla base dell'esperienza collettiva dei singoli e delle organizzazioni nel processo di credito e si riflettono in una serie di procedure operative. Sebbene non siano rigorosi, possono essere utili in situazioni complesse e come liste di controllo quando si effettua una valutazione.

Un settore importante per l'utilizzo dei sistemi esperti è l'analisi finanziaria. Questo è il processo di esame del bilancio di un'azienda al fine di comprendere la natura, l'attività e

i rischi inerenti all'attività. Quando viene formalizzato, l'uso di sistemi esperti tende ad essere combinato con il secondo approccio, **i sistemi di *rating***, in cui la qualità creditizia di un'impresa o di un individuo è classificata in un insieme di casi che si ritiene abbiano lo stesso grado di merito creditizio.

Ad esempio, Dun & Bradstreet, l'agenzia di informazioni creditizie specializzata nell'analisi delle piccole e medie imprese, pubblica una valutazione del credito composita basata sulle dimensioni dell'impresa (definita come patrimonio netto) come *proxy* per la capacità finanziaria e suddivide le imprese all'interno di una data fascia di patrimonio netto in qualità "alta", "buona", "discreta" e "limitata". Queste classifiche composite sono alimentate da una serie di fattori che hanno un impatto sulla qualità del credito, inclusi elementi come il numero di dipendenti, la cronologia dei pagamenti dell'azienda e così via.

L' approccio del **modello analitico** si basa sull'utilizzo dell'informazione finanziaria e si avvale di relazioni contabili che, nel loro insieme, forniscono un quadro della qualità creditizia dell'entità.

I **sistemi di *rating*** si sviluppano nella terza categoria dei metodi di valutazione del credito, nota come modelli di *credit scoring*. Questi modelli di punteggio forniscono un sistema di valutazione che viene formalizzato in un modello matematico o statistico e tutti i crediti vengono valutati utilizzando gli stessi dati e la stessa metodologia. In quanto tali, sono più rigorosi e trasparenti nel loro approccio rispetto ai sistemi di *rating*, sebbene siano progettati per fornire lo stesso livello di supporto decisionale.

L'ultima categoria di modelli di valutazione del credito è costituita dai **modelli basati sul mercato**. Si tratta di modelli formali, come per i modelli di *credit scoring*, ma le informazioni utilizzate per determinare la qualità del credito derivano dai prezzi dei mercati finanziari. Questi modelli utilizzano l'elaborazione delle informazioni che avviene nei mercati finanziari per modellare la probabilità di *default*. Ad esempio, se gli investitori nutrono riserve sulla futura affidabilità creditizia di una determinata società quotata, tendono a vendere le azioni. Questo ha l'effetto di ridurre il prezzo delle azioni poiché queste preoccupazioni si traducono nel prezzo di mercato. Un modello di valutazione del credito in grado di catturare questo effetto utilizza la comprensione combinata e l'elaborazione delle informazioni di tutti gli investitori sul mercato. Pertanto,

quest'ultimo tipo di modello utilizza un set di informazioni più ampio rispetto ai primi tre.

1.2.2 Processo di valutazione del credito

Quando arriva una nuova richiesta di credito, questa viene analizzata utilizzando una delle modalità descritte nella sezione precedente. Qualsiasi decisione di credito sarà incentrata sui rischi aziendali, finanziari e strutturali, se applicabili, e terrà in considerazione le seguenti caratteristiche chiave sul credito:

- L'ambiente competitivo dell'azienda
- Rischi del settore, in particolare tecnologia, requisiti normativi, barriere all'ingresso per i concorrenti e possibili sostituti.
- Struttura del capitale, che comprenderebbe, a seconda delle finalità dell'analisi del credito, il livello del requisito di spesa in conto capitale, eventuali passività fuori bilancio, questioni contabili e fiscali, leva finanziaria, struttura di rimborso del debito, capacità di servizio del debito e base degli interessi.
- Flessibilità finanziaria, che include elementi quali esigenze finanziarie, piani e alternative, capacità di attingere ai mercati dei capitali e *covenant* sul debito. La flessibilità finanziaria può includere questioni quali le relazioni bancarie, le linee di credito impegnate, la capacità di debito attuale e futura e altre questioni.

L'analisi dell'ambiente competitivo e dei rischi del settore, insieme all'analisi del potere contrattuale dei fornitori e degli acquirenti, fornirà un quadro chiaro delle forze che modellano la concorrenza del settore. Questo è spesso basato sull'analisi delle cinque forze di Porter (1985)[6]. Le opportunità e le minacce esterne di un'azienda possono essere prese in considerazione, riconoscendo il ciclo di vita del prodotto dell'azienda e del settore e le prospettive future[7].

Le informazioni non finanziarie utilizzate dagli analisti del credito includono:

- rapporti di un prestatore sulla reputazione e l'esperienza dei clienti ricevuti dalle risorse del mercato locale;

[6] Le cinque forze sono la rivalità industriale, il potere dei fornitori, il potere dei clienti, la minaccia dei sostituti e la minaccia dei nuovi entranti.

[7] I settori ad alta tecnologia sono soggetti a importanti cambiamenti di mercato che possono destabilizzare le imprese e portare a ingenti svalutazioni del valore economico delle attività che potrebbero essere state utilizzate come garanzia.

- relazioni redatte da: agenzie di valutazione del credito (se disponibili), analisti finanziari, fornitori di informazioni o agenzie e uffici di *rating* del credito;
- relazioni sui rischi ambientali;
- Promemoria di chiamata e *report* preparati da rapporti in loco e responsabili delle relazioni.

Per un'analisi più sofisticata, potrebbe essere possibile indagare la probabilità di migrazione del credito (cioè, il rischio di un declassamento del credito in un determinato periodo di tempo) e il rischio di insolvenza.

Le misure della frequenza di *default* attesa (EDF) forniscono benefici nel processo di valutazione:

- offrire un maggior grado di accuratezza nella valutazione del rischio di credito;
- quantificare il rischio per una determinazione del prezzo appropriata e quindi migliorare la redditività (ovvero, avere una stima della frequenza di *default* attesa facilita l'uso del prezzo del rischio);
- concentrare le risorse di analisi del credito nelle aree in cui possono apportare il massimo valore;
- fornire indicatori di allerta precoce di un grave deterioramento del credito.

1.2.3 Limiti delle decisioni umane e *bias* cognitivi

Quando il *credit scoring* venne introdotto, molti esperti del settore finanziario erano scettici riguardo alla sua efficacia rispetto al giudizio umano. Tuttavia, Falkenstein et al. (2000) evidenziano diversi difetti delle decisioni basate sull'intuito:

1. Le persone tendono a sovrastimare la propria conoscenza (Alpert & Raiffa, 1982)
2. La loro fiducia aumenta con l'importanza del compito, portandole a ricordare più facilmente i successi rispetto ai fallimenti (Barber & Odean, 1999).
3. Il *feedback* per le decisioni umane è spesso aneddotico e non strutturato (Nisbet et al., 1982).
4. Sebbene capaci di identificare fattori rilevanti, le persone non riescono sempre a integrarli in modo ottimale (Meehl, 1954).
5. Nei *test* su 60 bilanci finanziari, i funzionari di prestito hanno ottenuto il 74% di risposte corrette, meno efficace di un semplice rapporto "passività su attività" (Libby, 1975)

Uno dei vantaggi del *credit scoring* è la riduzione del *bias* umano, un aspetto critico in un mondo in cui errori di valutazione possono portare a gravi danni reputazionali o cause legali. Tuttavia, il processo decisionale oggettivo non è completamente immune da generalizzazioni e stereotipi, che sono parte della natura umana.

Capitolo 2: Fondamenti teorici per l'ottimizzazione del *prescore*

2.1 Tecnologie emergenti per il miglioramento dei modelli di *prescore*

I metodi tradizionali per stimare la solidità finanziaria dei mutuatari e prevederne l'eventuale insolvenza ricorrono da tempo a dati strutturati e modelli statistici, allo scopo di appurare l'affidabilità creditizia dei soggetti richiedenti. Benché le formule specifiche siano in larga parte protette da riservatezza, un esempio emblematico è rappresentato dal sistema FICO (*Fair Isaac Corporation*)[8] e da altri meccanismi analoghi, i quali convertono in uno *score* vari fattori come la cronologia dei pagamenti, il livello d'indebitamento e il reddito. Questo punteggio identifica il grado di rischio associato a un determinato mutuatario e orienta gli istituti di credito, sia sulla decisione di erogare o meno il prestito, sia sulle condizioni applicabili. In parallelo, la previsione di un potenziale *default* si basa su variabili prestabilite, come l'utilizzo delle linee di credito e il rapporto tra debito e reddito, nonché su ulteriori indicatori finanziari che anticipano l'eventuale inadempienza.

Sebbene molto diffusi, questi modelli mostrano alcuni limiti strutturali. Un primo problema è la dipendenza da un insieme ristretto di parametri, che non sempre coglie la complessità del rischio finanziario di un individuo o di un'impresa. Inoltre, si tratta di approcci statici, ancorati a dati storici e regole fisse, risultando così poco duttili se il comportamento del mutuatario o l'andamento dell'economia mutano in modo rapido. Ne consegue una scarsa reattività alle variazioni, con il rischio ulteriore di incorporare pregiudizi o di sottovalutare i potenziali clienti dei mercati emergenti, i più giovani o coloro che non hanno uno storico creditizio ben documentato, rendendo il processo meno inclusivo.

Nel corso degli ultimi anni, l'uso dell'intelligenza artificiale (IA) è divenuto un elemento di grande trasformazione nel comparto finanziario, specialmente per quanto riguarda la valutazione del merito di credito. I modelli IA, infatti, possono vagliare enormi quantità di informazioni eterogenee, incluse fonti non convenzionali come attività *social*, transazioni immediatamente tracciate e *pattern* comportamentali *online*. Rispetto alle soluzioni tradizionali, i sistemi di intelligenza artificiale si basano su algoritmi adattivi e

[8] Il sistema FICO è uno dei modelli di *credit scoring* più utilizzati negli Stati Uniti, sviluppato da *Fair Isaac Corporation* nel 1956. È basato su un algoritmo proprietario che tiene conto di diversi fattori, tra cui la cronologia dei pagamenti, il livello di indebitamento e la durata del credito.

sono in grado di apprendere continuamente dai dati, generando analisi del rischio più articolate e dinamiche. Ne derivano vantaggi quali la riduzione di possibili distorsioni, il miglioramento dell'accessibilità per soggetti meno visibili ai circuiti convenzionali e, in generale, una maggiore efficienza nella stima del merito creditizio.

I meccanismi di *scoring* classici e la previsione di *default*, d'altra parte, non riescono a rispondere in maniera tempestiva alle esigenze dei moderni scenari finanziari, caratterizzati da processi sempre più complessi e volatili. L'urgenza di introdurre modelli che sfruttino tecniche IA scaturisce proprio dal limite dei metodi convenzionali, i quali si basano su schemi rigidi spesso non allineati all'evoluzione del profilo dei mutuatari. Se i dati sono ben articolati e sottoposti ad algoritmi adeguati, si aprono prospettive più ampie sia nell'assegnare finanziamenti meritevoli, sia nell'evitare che soggetti potenzialmente affidabili restino esclusi dai servizi di credito.

La strategia per ovviare a tali criticità risiede nell'adozione di sistemi alimentati dall'intelligenza artificiale. Questi raccolgono e processano un numero maggiore di informazioni, servendosi di algoritmi avanzati di apprendimento automatico per interpretare condotte finanziarie complesse. Man mano che i *dataset* crescono, tali soluzioni affinano ulteriormente la loro accuratezza, migliorando contestualmente la capacità di reagire ai cambiamenti nel profilo dei debitori e nell'ambiente economico. Grazie, inoltre, alla possibilità di generare valutazioni in tempo reale, gli istituti finanziari possono formulare decisioni più rapide e coerenti in materia di credito.

Un aspetto cruciale riguarda infine l'inclusione finanziaria: l'impiego di dati cosiddetti non tradizionali, tra cui evidenze di pagamento relative a bollette o transazioni via mobile[9], consente ai modelli IA di intercettare individui con un passato creditizio inesistente o limitato. Questo approccio incrementa l'accesso ai prestiti per fasce di popolazione tradizionalmente trascurate, fornendo al contempo nuovi elementi di affidabilità per chi eroga credito, soprattutto in contesti emergenti.

[9] L'uso di dati non tradizionali nel *credit scoring* è una pratica crescente, resa possibile dall'integrazione di tecnologie come l'intelligenza artificiale e l'apprendimento automatico, che permettono di analizzare fonti di dati non convenzionali, migliorando l'inclusività finanziaria.

2.1.1 Evoluzione del *credit scoring*

Il *credit scoring* ha registrato un'evoluzione di rilievo, passando da sistemi basati su regole rigide a metodologie più complesse che integrano l'intelligenza artificiale (IA) nei processi di valutazione del merito creditizio[10]. In questa sezione, vengono esaminate le tappe fondamentali che hanno segnato il suddetto cambiamento, a partire dagli approcci manuali iniziali, con tutti i loro limiti insiti, fino all'impiego dell'IA nel contesto del *credit scoring*.

Le origini del punteggio di credito risalgono alla metà del Novecento, quando i finanziatori introdussero procedure manuali basate su criteri fissi per stabilire la solvibilità, valutando aspetti come il reddito, la carriera lavorativa e i debiti in essere. Il risultato di tali processi era una stima numerica che orientava i creditori[11], ma questi modelli non offrivano la flessibilità necessaria a tenere il passo con l'evoluzione dei mercati.

Con il progredire dei sistemi finanziari, è emerso chiaramente come i tradizionali punteggi di credito fondati su regole[12] fossero statici, inadatti a recepire dati in tempo reale e incapaci di catturare la complessità dei comportamenti finanziari individuali. Tale approccio "unico per tutti" ha finito per produrre valutazioni imprecise, penalizzando o trascurando alcune categorie di mutuatari. Inoltre, tali modelli presentavano carenze nell'includere individui con storie creditizie non convenzionali, limitando l'inclusione finanziaria.

L'avvento dell'IA ha quindi rivoluzionato la concezione del punteggio di credito. Grazie a un insieme di algoritmi di apprendimento automatico, è diventato possibile superare i difetti dei sistemi rigidi, in quanto tali algoritmi sono in grado di estrarre e analizzare enormi quantità di dati, identificando schemi e formulando previsioni con livelli di accuratezza inediti[13]. In questo senso, l'IA ha introdotto ulteriori variabili e prospettive di rischio in precedenza trascurate.

Tecnologie basate su reti neurali, alberi decisionali o modelli di insieme (*ensemble*) si sono rivelate capaci di adattarsi dinamicamente al contesto, apprendendo da flussi di dati

[10] Mendhe et al., 2024
[11] Ampountolas et al., 2021
[12] Mhlanga, 2021
[13] Kamyab et al., 2021

aggiornati in tempo reale[14]. Ciò ha eliminato la rigidità delle soluzioni tradizionali, migliorando la precisione delle valutazioni del merito creditizio. Di conseguenza, l'adozione di tali strumenti ha permesso di estendere l'ambito di analisi al di là delle fonti consuete, comprendendo anche tracce provenienti dalle interazioni *online* e dai *social media*, con un sensibile aumento del grado di inclusività del sistema.

2.1.2 Analisi predittiva

L'analisi predittiva e i sistemi di apprendimento automatico introducono un approccio dinamico alla determinazione del punteggio di credito. Tali modelli si adattano continuamente a schemi in trasformazione, offrendo una previsione del rischio creditizio più reattiva e precisa. Gli algoritmi di *machine learning*, tra cui regressione, alberi decisionali e reti neurali, consentono di cogliere le relazioni complesse all'interno dei *dataset*. Una di queste abilità di analisi predittiva è la valutazione del rischio, riconsiderata rispetto ai modelli statici. Grazie a dati in tempo reale, tali procedure possono aggiornare i punteggi di credito in modo sincrono in funzione di fattori variabili, fornendo rappresentazioni più al passo con l'andamento del credito.

L'adozione di modelli di apprendimento profondo per definire il punteggio di credito ha ulteriormente sottolineato l'importanza della spiegabilità. In primo luogo, *l'Explainable Artificial Intelligence* (XAI) riguarda il chiarimento delle logiche decisionali di un modello d'IA, così da renderlo comprensibile. Nel *credit scoring*, emergono responsabilità in caso di contestazioni o imprevisti, assieme alla necessità di rispettare la normativa e di mantenere la fiducia di utenti e istituti finanziari. L'IA spiegabile si rivela indispensabile per fronteggiare questioni di distorsione e correttezza nel processo di valutazione.

Mostrando in che modo i modelli giungono alle decisioni, gli *stakeholder* sono in grado di individuare e ridurre i pregiudizi, favorendo un trattamento equo tra gruppi demografici diversi. Tale chiarezza rafforza prassi etiche di *credit scoring*.

L'intelligenza artificiale interpretabile mette i consumatori in condizione di capire i parametri che incidono sui loro punteggi di credito. Questa visibilità permette alle persone di attuare misure proattive per migliorare il proprio profilo creditizio e alimenta fiducia nei meccanismi di calcolo del punteggio. L'analisi predittiva per il *credit scoring* integra

[14] Barja-Martinez et al., 2021

sempre più fonti di dati non convenzionali oltre ai classici dati finanziari. Tra di esse rientrano informazioni come il canone d'affitto, la regolarità nei pagamenti delle bollette o persino i comportamenti sui social media. L'inclusione di dati aggiuntivi offre un quadro più ampio delle abitudini finanziarie di un soggetto, soprattutto per chi abbia una storia di credito limitata.
Ricorrere a fonti di dati alternative produce valutazione del rischio più consistenti, specie per quei soggetti con fascicoli creditizi "sottili". Grazie a una maggiore varietà informativa, l'analisi predittiva elabora stime più affidabili, limitando la dipendenza dai tradizionali indicatori di credito.
Pur offrendo vantaggi, il ricorso a fonti di dati atipiche pone anche diverse criticità, come la tutela della riservatezza dei dati, il contrasto a possibili disparità e la necessità di rispettare gli obblighi di legge. Mantenere un equilibrio tra innovazione e valori etici diventa essenziale nell'uso di dati alternativi nei sistemi predittivi di *credit scoring*.

In definitiva, l'analisi predittiva ha assunto un ruolo chiave nel rinnovare i parametri del *credit scoring*. La sua capacità di pronosticare il rischio di credito, fornire un'IA trasparente e incorporare dati eterogenei favorisce un metodo più articolato, limpido e inclusivo nell'accertamento dell'affidabilità creditizia. Con il progresso tecnologico, l'analisi predittiva continuerà probabilmente a rivestire una funzione sempre più rilevante per perfezionare i modelli di *credit scoring* in termini di esattezza, correttezza e fruibilità per i consumatori.

2.2 Modelli di *scoring* creditizio: tecniche in uso

Le tecniche di *credit scoring* sono strumenti quantitativi utilizzati dagli istituti finanziari per valutare l'affidabilità creditizia dei potenziali debitori e facilitare le decisioni relative alla concessione di prestiti. Questi sistemi si basano su modelli statistici e algoritmi automatizzati che, analizzando una serie di dati storici e finanziari, generano punteggi numerici (*score*) o classificazioni sintetiche per descrivere il livello di rischio associato a ciascun richiedente. I punteggi risultanti forniscono una stima della probabilità di insolvenza, riflettendo la capacità del cliente di rispettare i propri impegni finanziari nel tempo. Tale processo si svolge tipicamente nella fase di valutazione della richiesta di finanziamento, quando l'intermediario deve decidere se approvare o meno un mutuo, un

prestito o un'altra forma di credito. In questo contesto, il *credit score* rappresenta una misura sintetica del rischio di insolvenza, esprimendo il grado di incertezza sulla capacità del debitore di restituire il capitale ricevuto. Per comprendere appieno l'importanza del credit scoring, è essenziale prima definire chiaramente il concetto di rischio di credito, su cui si basa l'intera valutazione del merito creditizio.

Per gli intermediari, è fondamentale dotarsi di modelli di analisi in grado di "*clusterizzare*" i finanziamenti in gruppi di esposizioni "buone" o "cattive" in termini di merito creditizio. La sfida risiede nell'asimmetria informativa tra chi concede il credito e chi lo riceve. Spesso, i prestiti sono numerosi e di piccolo importo, quindi l'analisi puntuale di ogni singola posizione richiederebbe risorse non sostenibili. In questo contesto, i modelli di *credit scoring* rappresentano una soluzione di sintesi quantitativa, evolutisi con il tempo verso forme sempre più avanzate grazie a tecniche di intelligenza artificiale come il *machine learning* e il *deep learning*. Nel seguito, si descrivono alcuni degli approcci più diffusi.

2.2.1 Modelli univariati

Le origini dell'analisi dei bilanci a fini valutativi risalgono già agli anni Venti, quando si iniziò a utilizzare singoli indicatori economico-finanziari per comprendere lo stato di salute di un'impresa. In particolare, l'approccio univariato analizza uno alla volta gli indici di bilancio considerati significativi per individuare potenziali "punti deboli" dell'impresa: si coglie così sia una fotografia dello stato attuale sia eventuali segnali di criticità future.

Un celebre studio in questo filone è quello di Beaver (1967)[15]. Egli analizzò 158 imprese, metà insolventi (o in grave difficoltà) e metà sane, appartenenti allo stesso settore e di simile dimensione in termini di attivo. L'obiettivo era verificare quale singolo indicatore di bilancio offrisse la migliore capacità predittiva dell'insolvenza. Beaver notò che il rapporto tra *cash flow* e debiti totali risultava particolarmente efficace nello spiegare il successivo *default*. Nel campione analizzato, la corretta identificazione delle imprese che sarebbero fallite a un anno di distanza fu dell'87%, percentuale che scese al 78% a cinque anni dall'evento. Dai risultati emerse che le informazioni più utili erano quelle legate a

[15] Beaver, W. H. (1967), 'Financial Ratios as Predictors of Failure', Journal of Accounting Research, vol. 4, pp. 71-111.

struttura finanziaria e capacità di generare cassa, mentre gli indicatori di liquidità a breve mostravano minor potere diagnostico.

L'approccio univariato può essere rappresentato graficamente disponendo sull'asse delle ascisse un indicatore (ad esempio il ROE) e su quello delle ordinate la frequenza di ciascun valore. Si individua poi una soglia (*cut-off*) che separa, con il minor margine d'errore possibile, le imprese insolventi da quelle sane.

2.2.2 Modelli multivariati

A differenza dei metodi univariati, l'approccio multivariato considera congiuntamente più variabili ritenute utili per valutare l'affidabilità di un'impresa, assegnando a ciascuna un peso. Ciò permette di ottenere un unico valore di sintesi (*score*) per descrivere la solvibilità del soggetto. La principale differenza tra i vari modelli risiede nella metodologia adottata per stimare i coefficienti associati alle variabili. Tra gli approcci multivariati, uno dei più noti è l'analisi discriminante, introdotta da Fisher e poi ripresa da molti altri studiosi.

2.2.3 Analisi discriminante lineare

L'analisi discriminante lineare, introdotta da Fisher nel 1936[16]. è una tecnica statistica utilizzata per separare campioni appartenenti a due o più categorie distinte. Nel contesto del *credit scoring*, questa metodologia è spesso impiegata per distinguere tra imprese solide e imprese insolventi, perseguendo due obiettivi principali:

- Identificazione delle variabili discriminanti, ossia stabilire quali caratteristiche finanziarie o operative meglio separano i gruppi di imprese con comportamenti finanziari differenti.
- Classificazione dei nuovi casi, il che consiste nell'assegnare nuove osservazioni a una delle categorie predefinite, in base alle loro caratteristiche.

L'approccio di Fisher si basa sulla costruzione di una combinazione lineare delle variabili di *input*, che genera uno *score* per ciascuna impresa, utilizzato come criterio di classificazione. Questa combinazione è espressa dalla seguente formula:

$$S_j = \sum_{i=1}^{n} \alpha_i \cdot X_{ij} \qquad (2.1)$$

[16] R.A. Fisher, "*The use of multiple measurements in taxonomic problems*", Annals of eugenics, 1936

Dove:

- S_j è lo *score* dell'impresa j-esima;
- a_i rappresenta il coefficiente associato alla i-esima variabile;
- X_{ij} è la variabile descrittiva i-esima per l'impresa j-esima;
- n è il numero totale di variabili considerate.

L'obiettivo dell'analisi discriminante è identificare il vettore dei coefficienti α che massimizza la distanza tra i gruppi, minimizzando al contempo la varianza all'interno dei gruppi stessi. Questo viene ottenuto massimizzando il rapporto tra la distanza tra le medie dei gruppi e la varianza complessiva, secondo la seguente relazione:

$$\frac{\mathrm{E}(\mathrm{X}'_{\mathrm{i}1}-\alpha)-\mathrm{E}(\mathrm{X}'_{\mathrm{j}2}-\alpha)^2}{\mathrm{VAR}(\mathrm{X}'\cdot\alpha)} = \frac{(\bar{\mathrm{S}}_1-\bar{\mathrm{S}}_2)}{\sigma^2_{\mathrm{S}}} \tag{2.2}$$

Dove:

- $\mathrm{X}'_{\mathrm{i}1}$ rappresenta il vettore delle variabili per l'impresa i-esima nel gruppo insolvente;
- $\mathrm{X}'_{\mathrm{j}2}$ rappresenta il vettore delle variabili per l'impresa j-esima nel gruppo solvibile;
- S_n rappresenta la media degli *score* del gruppo n-esimo (n = 1,2), noto anche come "centroide".

La regola di classificazione risultante si basa sul confronto degli score, secondo il criterio:

$$|S_j - S_1| < |S_j - S_2|, ovvero\ S_j < \frac{1}{2}\cdot(S_1 + S_2), per\ S_1 < S_2 \tag{2.3}$$

Se questa condizione è soddisfatta, l'impresa j-esima viene assegnata al gruppo delle imprese insolventi, altrimenti viene classificata tra le imprese sane. Il termine $\frac{1}{2}\cdot(S_1 + S_2)$ rappresenta dunque il *cut-off* utilizzato per separare le due categorie.

È importante notare che, nonostante la potenza di questa metodologia, essa presenta alcune limitazioni. In particolare, l'efficacia del modello dipende dalla capacità di individuare una funzione lineare che riduca al minimo l'errore di classificazione, ovvero l'area di sovrapposizione tra le due popolazioni. Maggiore è questa sovrapposizione, maggiore sarà l'incertezza nel processo di classificazione. Infine, il modello di Fisher è considerato di tipo non parametrico, poiché non richiede ipotesi specifiche sulla forma delle distribuzioni delle variabili in *input*.

2.2.4 *Z-Score* di Altman

Nella sua celebre ricerca del 1968, Edward Altman[17] applicò l'analisi discriminante lineare a un campione composto da 33 imprese statunitensi fallite tra il 1945 e il 1965 e 33 imprese sane, sviluppando un modello che è divenuto uno dei più noti strumenti per la previsione dell'insolvenza aziendale.

Il modello proposto, noto come ***Z-score***, è descritto dalla seguente equazione:

$$Z = 0{,}012 \cdot \frac{Capitale\ circolante}{Attivo\ netto} + 0{,}014 \cdot \frac{Riserve\ di\ utili}{Attivo\ netto} + 0{,}033 \cdot \frac{Utile\ ante\ interessi\ e\ tassi}{Attivo\ netto} + 0{,}006 \cdot \frac{Valore\ di\ mercato\ del\ patrimonio\ netto}{Debiti\ totali} + 0{,}999 \cdot \frac{Ricavi}{Attivo\ netto} \quad (2.4)$$

Le variabili utilizzate nel modello rappresentano diverse dimensioni della *performance* aziendale, tra cui la liquidità, la redditività e l'efficienza operativa. I *test* di significatività condotti da Altman indicano che le componenti relative alla redditività e all'efficienza complessiva dell'impresa sono i principali fattori discriminanti, mentre la liquidità risulta avere un impatto più limitato, in linea con quanto osservato da Beaver nelle sue ricerche precedenti.

La classificazione delle imprese avviene confrontando lo *Z-score* calcolato con un valore di soglia, o ***cut-off***, fissato a 2,675. Questo valore rappresenta il punto medio tra i centroidi delle due popolazioni (imprese sane e insolventi). Secondo questa regola, le imprese con un punteggio superiore a 2,675 sono considerate finanziariamente sane, mentre quelle con un punteggio inferiore a questa soglia sono considerate a rischio di insolvenza. In altre parole, un valore più elevato dello *Z-score* indica una minore probabilità di fallimento e un rischio finanziario più contenuto.

Altman ha inoltre identificato una "zona grigia" (*grey zone*), compresa tra 1,81 e 2,99[18], all'interno della quale aumenta l'incertezza del modello e il rischio di errori di classificazione. Questa zona rappresenta un'area di ambiguità, dove è più difficile prevedere con precisione la solvibilità dell'impresa.

---

[17] Altman, E. I. (1968), 'Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy', Journal of Finance, vol. 23, no. 4, pp. 589-609.

[18] La *grey zone* di Altman rappresenta l'intervallo in cui la classificazione delle imprese è meno certa, riflettendo una maggiore difficoltà nel distinguere tra imprese finanziariamente sane e a rischio di insolvenza.

La classificazione dei livelli di rischio secondo lo Z-score di Altman può essere sintetizzata come segue:

- Z < 1,81 elevato rischio di insolvenza;
- 1,81< Z < 2,99 rischio medio (zona grigia);
- Z > 2,99 basso rischio di insolvenza.

Dal punto di vista predittivo, il modello di Altman ha dimostrato elevate capacità nel prevedere l'insolvenza nel periodo immediatamente precedente al fallimento dell'impresa, come evidenziato nella Tabella 2. Tuttavia, l'accuratezza del modello tende a diminuire se applicato a periodi più lontani dall'evento di insolvenza: ad esempio, a due anni di distanza, la capacità predittiva scende all'82%, rendendolo meno preciso rispetto al modello di Beaver per previsioni a medio termine.

Gruppo effettivo	**Classificazione ottenuta**		**Totale**
	Anomala	Sana	
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 2 Analisi predittiva del modello di Altman al tempo t-1

La tabella indica che il modello di Altman presenta una notevole capacità predittiva quando utilizzato per valutare il rischio di insolvenza a un anno dall'evento critico. In particolare, l'analisi delle imprese effettivamente insolventi mostra un tasso di errore di classificazione del 6%, mentre per le imprese finanziariamente solide l'errore risulta ancora più contenuto, attestandosi al 3%. Complessivamente, il modello è in grado di classificare correttamente circa il 95% delle imprese, confermandone l'elevata affidabilità in previsione a breve termine.

2.2.5 Approfondimento al calcolo del *cut-off*

La determinazione del *cut-off* in un modello di classificazione come quello di Altman richiede la definizione di alcuni concetti fondamentali. In primo luogo, è necessario considerare le probabilità a priori e a posteriori. Le probabilità a priori rappresentano la probabilità che un'impresa appartenga a un determinato gruppo (ad esempio, sano o insolvente) prima che siano osservate le sue caratteristiche finanziarie. Una volta che

queste informazioni sono disponibili, si possono calcolare le probabilità a posteriori utilizzando il Teorema di Bayes, che aggiorna le stime iniziali alla luce delle nuove evidenze:

$$\mathrm{P(n}|X_i) = \frac{p(n)q_n}{p(X_i)} = \frac{p(n)q_n}{p(s)q_s + p(\alpha)q_\alpha} \quad (2.5)$$

Dove:

- q_n è la probabilità a priori dell'appartenenza al gruppo n-esimo;
- $p(n)$ è la probabilità che l'impresa osservata appartenga al gruppo n-esimo in base alle sue caratteristiche finanziarie;
- $\mathrm{p(n}|X_i)$ è la probabilità a posteriori dell'appartenenza al gruppo n-esimo, dato il vettore delle variabili X_i.

In base a questa relazione, un'impresa verrà classificata come finanziariamente solida se soddisfa la seguente condizione:

$$q_s \cdot \mathrm{p(s)} > q_\alpha \cdot \mathrm{p}(\alpha) \quad (2.6)$$

Se questa disuguaglianza non è verificata, l'impresa verrà assegnata al gruppo delle controparti a rischio.

Un aspetto critico è rappresentato dagli errori di classificazione, che si dividono in errori di prima e seconda specie. L'errore di prima specie, noto anche come Tipo I, si verifica quando un'impresa effettivamente insolvente viene erroneamente classificata come sana. Questo tipo di errore è particolarmente costoso, poiché comporta una perdita diretta del capitale erogato. L'errore di seconda specie, o Tipo II, avviene invece quando un'impresa finanziariamente sana è classificata come insolvente, con conseguente perdita di opportunità di profitto, dato che l'istituto potrebbe rifiutare finanziamenti potenzialmente redditizi.

Poiché questi due tipi di errore hanno impatti finanziari differenti, è essenziale tenerne conto nella definizione del *cut-off* del modello. La matrice di confusione (o *misclassification rate*), come illustrato nella Tabella 3, è uno strumento utile per valutare l'accuratezza del modello e bilanciare adeguatamente i costi associati agli errori di classificazione.

Gruppo effettivo	Classificazione ottenuta		Totale
	Anomala	Sana	
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 *Misclassification Rate* -confronto tra comportamento reale e previsto delle imprese

Il peso relativo dei due tipi di errore (prima e seconda specie) è generalmente diverso, poiché le conseguenze economiche associate sono assai differenti. In particolare, classificare erroneamente un'impresa sana come rischiosa può comportare la perdita di potenziali opportunità di profitto, come nel caso di un finanziamento rifiutato a un'impresa solvibile. Tuttavia, l'errore opposto, ovvero considerare affidabile un'impresa che successivamente risulta insolvente, ha un impatto molto più rilevante, poiché comporta una perdita diretta del capitale erogato, oltre agli interessi attesi.

Il costo atteso dell'errore di prima specie può essere minimizzato utilizzando la seguente relazione:

$$\frac{p(s)}{p(A)} = \frac{q_\alpha \cdot C_1}{q_s \cdot C_2} \tag{2.7}$$

In questa formula:

- p(s) rappresenta la probabilità a posteriori che l'impresa appartenga al gruppo delle imprese sane;
- p(A) è la probabilità a posteriori di appartenere al gruppo delle imprese anomale;
- q_α e q_s sono le probabilità a priori delle rispettive popolazioni;
- C_1 e C_2 rappresentano i costi associati agli errori di prima e seconda specie, rispettivamente.

Per il calcolo del *cut-off* è importante considerare che il costo degli errori varia a seconda del gruppo e influenza solo il termine costante dell'equazione discriminante.

$$Cut - off = ln \frac{q_\alpha \cdot C_1}{q_s \cdot C_2} \tag{2.8}$$

Dove:

- q_α e q_s sono le probabilità a priori che un'impresa appartenga al gruppo delle insolventi o delle sane, rispettivamente;

- C_1 e C_2 rappresentano i costi associati agli errori di prima e seconda specie, rispettivamente.

In molti casi pratici, i termini q_α e q_s possono essere trattati come le proporzioni dei due gruppi all'interno del campione utilizzato per costruire il modello. Tuttavia, se il campione è bilanciato, come avviene spesso nelle analisi empiriche, le probabilità a priori e i costi degli errori possono essere omessi, facendo sì che la funzione discriminante si riduca a una versione centrata sullo zero, equivalente alla funzione lineare proposta da Fisher.

Il valore atteso del costo degli errori di classificazione, dato l'utilizzo del modello, è calcolato con la seguente relazione:

$$E(C) = q_\alpha \cdot C_1 \cdot \frac{M_{\alpha,s}}{N_\alpha} + q_s \cdot C_2 \cdot \frac{M_{s,\alpha}}{N_s} \qquad (2.9)$$

Dove:

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

Infine, è importante notare che l'analisi discriminante lineare presenta alcune somiglianze con la regressione lineare, poiché i coefficienti ottenuti attraverso questo metodo corrispondono, a meno di un rapporto costante, a quelli stimati con il metodo dei minimi quadrati ordinari. Tuttavia, la regressione logistica rappresenta una valida alternativa, in quanto consente di superare alcune limitazioni dell'analisi discriminante, come l'assunzione di normalità multivariata delle variabili indipendenti.

2.2.6 Modelli di regressione

Il modello logistico rappresenta un approccio alternativo all'analisi discriminante lineare per stimare la probabilità di insolvenza di un'impresa. In questo contesto, si utilizza una variabile dipendente dicotomica, Y, che assume solo due possibili valori, a seconda che l'impresa appartenga al gruppo delle imprese sane o a quello delle imprese anomale:

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

Le variabili indipendenti del modello sono tipicamente indicatrici di bilancio che si presume abbiano una relazione causale con la probabilità di insolvenza. Tuttavia, se si utilizza un modello lineare semplice, come il *Linear Probability Model (LPM)*, possono emergere alcune difficoltà tecniche. In particolare, l'LPM soffre di eteroschedasticità, poiché la varianza degli errori non è costante, e presenta il problema della non normalità degli errori. Inoltre, l'LPM può produrre valori stimati della probabilità che non rientrano nell'intervallo [0,1], generando quindi risultati non interpretabili dal punto di vista probabilistico.

Per superare queste limitazioni, il modello logistico utilizza una funzione non lineare per collegare le variabili indipendenti alla probabilità di insolvenza, garantendo che i valori stimati siano sempre compresi tra 0 e 1. La relazione fondamentale del modello logistico è data da:

$$p = F(\alpha + \beta X) \tag{2.10}$$

Dove:

- p è la probabilità di insolvenza dell'impresa,
- X è il vettore delle variabili indipendenti,
- α è il termine costante,
- β rappresenta il vettore dei coefficienti,
- F è la funzione di distribuzione cumulativa logistica.

La funzione logistica cumulativa è definita come:

$$F(\alpha + \beta X) = \int_{-\infty}^{\alpha + \beta X} f(h)dh = \frac{1}{1+e^{-(\alpha + \beta X)}} \tag{2.11}$$

Questa funzione è strettamente crescente e garantisce che l'*output* sia sempre compreso tra 0 e 1. La corrispondente funzione di densità è data da:

$$f(h) = \frac{e^h}{(1+e^h)^2} \tag{2.12}$$

che descrive la probabilità istantanea che l'evento si verifichi in funzione del valore dell'*input*.

Riorganizzando l'equazione logistica, si ottiene l'*odds ratio*, ovvero il rapporto tra la probabilità di insolvenza e la probabilità del suo complemento (l'impresa non insolvente):

$$e^{-(\alpha + \beta X)} = \frac{1-p}{p} \tag{2.14}$$

Applicando il logaritmo naturale a questa relazione, si arriva alla forma lineare del modello logistico:

$$ln\frac{1-p}{p} = \alpha + \beta X \quad (2.15)$$

Questa espressione rappresenta la *log-odds*, ovvero il logaritmo del rapporto tra la probabilità che l'impresa diventi insolvente e la probabilità che rimanga solvibile.

La principale differenza tra il modello logistico e il *Linear Probability Model* è proprio questa trasformazione logaritmica, che consente di evitare i problemi di interpretazione dei risultati caratteristici dell'LPM.

In modo equivalente, questa relazione può essere scritta come:

$$ln\frac{p(A)}{p(B)} = \alpha + \beta X \quad (2.16)$$

Dove p(A) e p(B) rappresentano le densità di probabilità delle due categorie considerate (ad esempio, imprese sane e insolventi). Se si applica il teorema di Bayes, come illustrato nella formula (2.5), l'impresa sarà assegnata alla categoria A se:

$$ln\frac{p(A)}{p(B)} > ln\frac{q(B)}{q(A)}, ovvero\ \alpha + \beta X > ln\frac{q(B)}{q(A)} \quad (2.17)$$

Nel caso più semplice, in cui vi sia perfetta incertezza (ovvero, quando la probabilità di insolvenza è esattamente 0,5), l'esponente della funzione logistica cumulativa è pari a zero, determinando un *cut-off* centrale. Questo rappresenta il punto in cui le due categorie hanno uguale probabilità di verificarsi, rendendo la classificazione più incerta.

2.2.7 Le reti neutrali

Le reti neurali artificiali rappresentano un approccio innovativo all'analisi predittiva, diverso dai modelli statistici tradizionali come l'analisi discriminante o la regressione lineare. A differenza di questi ultimi, le reti neurali sono spesso descritte come "scatole nere"[19] perché le relazioni tra *input* e *output* non sono immediatamente osservabili, rendendo difficile interpretare come il modello arrivi a una determinata previsione. Questo approccio è ispirato al funzionamento del cervello umano, dove l'apprendimento avviene attraverso l'elaborazione parallela di segnali da parte di neuroni interconnessi. Le prime ricerche sulle reti neurali risalgono agli anni '50 e '60, con il Perceptron di Frank Rosenblatt, un modello a singolo livello capace di apprendere *pattern* semplici. Tuttavia, lo sviluppo di questa tecnologia subì un rallentamento dopo le critiche di Marvin Minsky

[19] A differenza dei modelli lineari, le reti neurali non permettono una facile interpretazione delle relazioni tra input e output, poiché il processo decisionale è distribuito attraverso molte connessioni interne complesse (Lipton, Z. C. (2018). *The Mythos of Model Interpretability*. Communications of the ACM, 61(10), 36-43.).

e Seymour Papert nel 1969, che evidenziarono i limiti del Perceptron nel risolvere problemi non lineari. Solo negli anni '80, con l'introduzione dei neuroni nascosti e dell'algoritmo di retro-propagazione[20] dell'errore, le reti neurali hanno ripreso a svilupparsi rapidamente, sostenute dai progressi tecnologici nel calcolo parallelo.

Le reti neurali possiedono caratteristiche che le rendono particolarmente efficaci in contesti complessi. Innanzitutto, elaborano informazioni in parallelo, migliorando l'efficienza computazionale. Inoltre, distribuiscono le informazioni attraverso l'intera rete, aumentando la resilienza agli errori. Sono tolleranti a dati rumorosi o incompleti, poiché riescono a generalizzare le informazioni apprese durante l'addestramento. Questo è possibile grazie all'uso di funzioni di attivazione non lineari, che consentono di modellare relazioni complesse tra *input* e *output*, conferendo loro la capacità di approssimare qualsiasi funzione continua con sufficiente complessità della rete. Questa proprietà, nota come "approssimazione universale", è uno dei motivi principali per cui le reti neurali sono così ampiamente utilizzate in applicazioni come la visione artificiale, il riconoscimento del linguaggio naturale e l'analisi di dati ad alta frequenza.

Strutturalmente, una rete neurale è composta da più livelli. Lo strato di *input* riceve i dati grezzi, gli strati nascosti eseguono l'elaborazione interna e lo strato di *output* fornisce i risultati finali. La scelta del numero di neuroni e strati nascosti è cruciale per garantire la capacità di generalizzazione della rete. Se il numero di neuroni è troppo elevato, il modello rischia di sovra attarsi ai dati di addestramento (*overfitting*[21]), mentre un numero insufficiente di neuroni limita la capacità del modello di apprendere le relazioni complesse presenti nei dati. Le funzioni di attivazione più comuni includono la sigmoide[22], che limita i valori tra 0 e 1, e la tangente iperbolica[23], che produce valori tra -1 e 1, entrambe utili per problemi di classificazione. La ReLU (*Rectified Linear Unit*)[24]

[20] Concetto è stato formalizzato nel teorema dell'approssimazione universale, che dimostra come una rete con almeno uno strato nascosto possa approssimare qualsiasi funzione continua (Cybenko, 1989).

[21] L'*overfitting* si verifica quando un modello si adatta troppo ai dati di addestramento, perdendo capacità di generalizzazione sui dati nuovi (Goodfellow et al., 2016).

[22] La sigmoide mappa i valori tra 0 e 1, e come la sua derivata risulti utile nel calcolo del gradiente per la retro-propagazione.

[23] Questa funzione è simile alla sigmoide ma con valori nell'intervallo -1 a 1, rendendola più adatta per dati con valori sia positivi che negativi.

[24] Funzione che evita il problema del *vanishing gradient* e accelera l'addestramento (Nair e Hinton, 2010).

è particolarmente popolare nelle reti profonde, poiché riduce i problemi di saturazione e velocizza l'addestramento.

L'addestramento delle reti neurali avviene generalmente attraverso l'apprendimento supervisionato, utilizzando l'algoritmo di *backpropagation*, che aggiorna iterativamente i pesi delle connessioni per minimizzare l'errore di previsione. Questo processo si basa sulla regola del gradiente, che guida l'ottimizzazione dei pesi in direzione opposta al gradiente dell'errore, migliorando la precisione del modello. Tuttavia, la scelta del tasso di apprendimento e l'inizializzazione dei pesi sono aspetti critici per evitare oscillazioni e garantire una rapida convergenza. In alcuni casi, si utilizza un termine aggiuntivo chiamato *momentum*[25], che stabilizza ulteriormente l'aggiornamento dei pesi, riducendo la probabilità di convergenza verso minimi locali.

La struttura della rete può variare notevolmente a seconda dell'applicazione. Alcune reti sono pienamente connesse, dove ogni neurone è collegato a tutti i neuroni del livello successivo, mentre altre adottano connessioni a salti[26], che consentono collegamenti diretti tra livelli non adiacenti, migliorando la capacità di apprendere strutture complesse. In alternativa, le connessioni ripetute[27] permettono ai neuroni degli strati intermedi di ricollegarsi agli strati di *input*, come nelle reti ricorrenti, particolarmente efficaci nel trattamento di sequenze temporali.

In sintesi, le reti neurali offrono una notevole flessibilità e potenza predittiva, ma richiedono una progettazione attenta per evitare problemi come l'*overfitting* e garantire una buona generalizzazione. La scelta dell'architettura e dei parametri di addestramento è cruciale per il successo dell'analisi, rendendo questa tecnologia particolarmente adatta a problemi complessi dove i metodi statistici tradizionali risultano inefficaci.

2.3 Tecniche di *Machine Learning*

L’avvento dell’era digitale, insieme alla diffusione capillare di *Internet*, ha generato una straordinaria crescita nella quantità di dati prodotti e resi fruibili, tanto da portare alcuni

[25] Il *momentum* è un termine aggiuntivo che permette di ridurre le oscillazioni nel percorso di apprendimento, accelerando la convergenza (Polyak, 1964).

[26] Connessioni utilizzate per reti più complesse, come le reti residue (ResNet) che hanno dimostrato di essere efficaci in applicazioni come il riconoscimento delle immagini (He et al., 2015).

[27] Tipiche delle reti ricorrenti (RNN), ampiamente utilizzate nell'elaborazione del linguaggio naturale e nelle serie temporali (Hochreiter e Schmidhuber, 1997).

studiosi a definire la nostra epoca come una vera e propria "società dei dati". A questo fenomeno si accompagna l'ascesa della disciplina nota come *data science*, un'area di ricerca in cui si integrano più competenze (matematica, statistica, informatica, *machine learning* e ricerca operativa) per estrarre valore da *dataset* di notevoli dimensioni.
All'interno di questo orizzonte, il *data mining* rappresenta il cuore del processo di analisi, consentendo di individuare *pattern*, regole e correlazioni nascoste nei dati. L'estrazione di conoscenza segue una sequenza di passi formalizzata come KDD (*Knowledge Discovery from Data*). Tale processo, necessario a strutturare la gran mole di informazioni in forma utile, comprende le seguenti fasi:

1. Selezione: si individuano i dati e le variabili di interesse all'interno di un database più ampio.
2. Pre-trattamento: in questo stadio si ripuliscono i dati, eliminando o rettificando valori errati, gestendo i *missing values* ed eventuali *outlier*, per evitare distorsioni nell'analisi.
3. Trasformazione: i dati, già ripuliti, vengono ulteriormente "lavorati" per adattarli alle tecniche di estrazione. Si possono effettuare operazioni di aggregazione, normalizzazione, standardizzazione, discretizzazione e così via, allo scopo di enfatizzare gli aspetti rilevanti.
4. *Data Mining*: è la fase centrale in cui si applicano gli algoritmi di *machine learning* o di natura statistica per individuare schemi nascosti (*pattern*). A seconda dello scopo, si utilizzano modelli descrittivi (identificare relazioni e gruppi) o predittivi (prevedere una determinata variabile).
5. Interpretazione-Valutazione: i risultati vengono studiati per verificarne l'utilità e la coerenza con le esigenze della ricerca. Se necessario, si può iterare il processo, tornando alle fasi precedenti, oppure consolidare il patrimonio di conoscenza appena acquisito.

Tra le varie tecniche di analisi dati, il *machine learning* si focalizza su procedure che, a partire da un insieme di esempi, apprendono regole o funzioni generali, applicabili a nuove osservazioni. In questo ambito, si distinguono sostanzialmente tre tipi di apprendimento:

- Supervisionato: quando nel *dataset* è presente una variabile-obiettivo (o etichetta), sulla quale viene "allenato" il modello

- Non supervisionato: quando non vi sono etichette e si cerca di scoprire *pattern*, correlazioni o *cluster* direttamente dai dati
- Semi-supervisionato: quando solo una parte dei dati risulta etichettata, mentre la restante è priva di etichetta.

Il presente paragrafo si sofferma sulle tecniche supervisionate e non supervisionate, essendo queste le due categorie di gran lunga più impiegate nella valutazione del merito di credito e nell'analisi dei rischi.

2.3.1 Apprendimento supervisionato

Un algoritmo di apprendimento supervisionato utilizza un *dataset* in cui ogni osservazione è associata a una variabile etichetta (o *label*), che rappresenta l'informazione di riferimento per il processo di apprendimento. Questa etichetta indica per ciascun esempio nel *dataset* la categoria di appartenenza (ad esempio, *spam*/non *spam*, vince/perde), mentre le restanti variabili (dette *feature*) forniscono le informazioni utilizzate per definire una regola di classificazione. L'obiettivo dell'algoritmo è apprendere, a partire da questi dati etichettati, un modello che possa successivamente classificare nuovi dati privi di etichetta, sulla base delle conoscenze acquisite durante la fase di addestramento.

A seconda del tipo di *output* prodotto, si distinguono diversi tipi di classificazione. Se l'etichetta può assumere solo due valori distinti, il modello è detto binario (ad esempio, 0/1 o vero/falso). Se invece sono possibili più classi, si parla di classificazione multi-classe (ad esempio, bambino/ragazzo/anziano). In alcuni casi, è possibile che un'osservazione appartenga simultaneamente a più classi indipendenti, dando origine a una classificazione *multilabel*, dove la variabile etichetta è composta da più valori simultanei.

In contrasto con gli algoritmi di classificazione, che producono valori discreti per l'etichetta, gli algoritmi di regressione sono progettati per predire valori numerici continui. In questo caso, l'etichetta non rappresenta una categoria fissa, ma una quantità numerica che il modello cerca di stimare a partire dalle *feature* disponibili.

Un aspetto cruciale nella costruzione dei modelli di classificazione è la gestione degli errori di previsione. Nei modelli binari, l'etichetta può assumere solo due stati, e questo comporta la possibilità di commettere due tipi di errori:

- Falsi Positivi (FP). Quando il modello classifica erroneamente come positivo un esempio che in realtà appartiene alla classe negativa.
- Falsi Negativi (FN). Quando il modello classifica erroneamente come negativo un esempio che in realtà appartiene alla classe positiva.

La valutazione delle prestazioni di un classificatore può essere effettuata utilizzando diverse metriche, ciascuna delle quali è adatta a specifici contesti applicativi. Alcune delle più comuni includono:

- Accuratezza, la quale misura la percentuale di osservazioni correttamente classificate rispetto al totale degli esempi. È utile quando le classi sono bilanciate.
- *Recall* (Sensibilità o Tasso di Rilevamento) che calcola la proporzione di veri positivi sul totale dei positivi effettivi, secondo la formula:

$$Recall = \frac{Veri\ positivi}{Veri\ positivi + Falsi\ Negativi}$$

  Questa metrica è particolarmente rilevante quando si desidera minimizzare i falsi negativi, come nel caso delle diagnosi mediche, dove non rilevare un caso positivo può avere gravi conseguenze.
- Precisione. Misura la proporzione di veri positivi tra tutti gli esempi classificati come positivi dal modello, secondo la formula:

$$Precisione = \frac{Veri\ positivi}{Veri\ positivi + Falsi\ Positivi}$$

  Questa metrica è preferibile quando è importante evitare falsi positivi, come nelle applicazioni di sicurezza, dove è meglio evitare allarmi falsi che possono generare costi elevati.

Infine, esiste una metrica combinata, nota come F-misura (o *F1-score*), che considera contemporaneamente precisione e *recall*, bilanciando i due aspetti in un'unica misura. La formula è:

$$F_1 = 2 \cdot \frac{Precisione\ \cdot Recall}{Precisione + Recall}$$

L'F1-*score* è particolarmente utile quando si vuole ottenere un compromesso tra precisione e *recall*, evitando che uno dei due valori prevalga eccessivamente sull'altro, riducendo l'efficacia complessiva del modello.

Si passa ora ad analizzare in dettaglio gli algoritmi di apprendimento supervisionato più usati.

2.3.1.1 Alberi decisionali

Un albero decisionale può essere concepito come un sistema che associa a n variabili in ingresso (X_1, X_2,…,X_n) una o più variabili in uscita (ad esempio, una classificazione o una decisione da intraprendere). In pratica, ciascuna variabile in *input*, spesso detta attributo, deriva dall'ambiente che stiamo analizzando o dal problema da risolvere; le variabili in *output*, invece, esprimono il risultato finale del processo decisionale. Nel caso in cui l'albero presenti un'elevata profondità, i risultati intermedi di un nodo superiore coincidono con gli ingressi del nodo immediatamente inferiore, influenzando passo dopo passo il percorso che conduce alla decisione conclusiva.

La struttura dell'albero è rappresentata graficamente come un diagramma "capovolto": il nodo radice è collocato in alto e, scendendo verso il basso, si incontrano diversi nodi interni, ciascuno dei quali esegue un *test* su una determinata proprietà (o attributo). A ogni nodo, si possono dipartire due o più rami, a seconda del numero di valori o intervalli previsti dalla variabile considerata. Il processo decisionale procede in maniera sequenziale: dal nodo radice, si valutano gli attributi uno dopo l'altro, eseguendo via via i *test* impostati. Ogni *test* esclude alcune ramificazioni e ne conserva altre, restringendo così progressivamente lo "spazio delle ipotesi", ossia le possibili strade percorribili. Il risultato finale viene raggiunto in corrispondenza dei nodi foglia, cioè i nodi più bassi nella struttura, in cui il sistema produce la decisione o la classificazione definitiva (ad esempio, "approvare il credito" vs. "rifiutare il credito").

Quando le variabili in *input* sono discrete (categoriche), l'albero decisionale si orienta verso un compito di classificazione, mentre se gli attributi o la variabile *target* sono continui, si ottiene una procedura di regressione. La gestione delle variabili continue risulta più articolata poiché, nel caso di un *test* condizionale, occorre stabilire soglie o partizioni (ad esempio, "se il valore di X supera 10, segui un certo ramo, altrimenti un altro"). Questo approccio può essere integrato con la cosiddetta logica sfumata (*fuzzy logic*), in cui i *test* non si limitano a valutazioni binarie ma considerano gradi di verità, rendendo gli alberi decisionali più duttili di quanto si potrebbe pensare a un primo sguardo.

Uno dei principali punti di forza di tali alberi risiede nell'estrema semplicità interpretativa: dal momento che ogni passaggio equivale a un *test* condizionale su una variabile, chi utilizza il modello può ricostruire agevolmente la catena di ragionamenti

che porta alla soluzione. Se applichiamo un albero di decisione in ambito medico, ad esempio per determinare una diagnosi, un professionista potrà verificare passo dopo passo quale valore di soglia o quale attributo abbia spinto verso una certa conclusione, valutando quindi se il risultato appare sensato e coerente. Questa trasparenza rende gli alberi particolarmente utili in contesti in cui le decisioni devono essere giustificate a esseri umani, come l'approvazione di un prestito, la definizione di un *budget* o, appunto, procedure diagnostiche in cui l'errore può avere conseguenze rilevanti.

Di contro, gli alberi decisionali presentano alcuni limiti. Se il problema risulta molto complesso, lo spazio delle ipotesi cresce in modo esponenziale all'aumentare delle variabili: per un semplice albero booleano (ovvero con attributi che possono assumere valori vero/falso), sono possibili 2^n combinazioni di valori per nnn attributi. In tali circostanze, la costruzione di un albero completo richiederebbe risorse computazionali ingenti e si rischierebbe di incorrere in *overfitting* (l'albero "impara a memoria" i dati e fatica a generalizzare). È anche importante notare che l'albero decisionale, strutturato come sequenza di *test*, non sempre è in grado di catturare relazioni molto complesse o funzioni non facilmente scomponibili. Ad esempio, alcune funzioni di parità o di maggioranza risulterebbero di difficilissima espressione tramite una gerarchia di *test* binari.

Nonostante tali restrizioni, gli alberi decisionali conservano un valore strategico in ambito di *machine learning* e *data mining* poiché:

1. La conoscenza del dominio può essere incorporata con relativa facilità, inserendo *test* specifici sugli attributi più rilevanti.
2. La lettura e la validazione umana del modello rimangono immediate, riducendo il rischio di errori dovuti a "scatole nere" non interpretabili.
3. L'integrazione con metodi di potatura (*pruning*) o con tecniche di *ensemble* (quali *Random Forest*) consente di affrontare meglio i problemi di sovradimensionamento e di controllare l'*overfitting*.
4. L'aggiornamento incrementale del modello può essere gestito in modo modulare, inserendo nuovi rami o sostituendoli, senza ricostruire interamente la struttura.

Quando occorre ridurre il rischio di *overfitting*, risulta frequente la tecnica nota come potatura: dopo aver generato l'albero, spesso molto profondo, si eliminano selettivamente alcuni sotto-rami con scarso contributo predittivo. Così facendo, si ottiene una struttura

più snella, che generalmente conserva una buona efficacia su dati mai visti prima. Se invece si mira ad accrescere la robustezza e l'accuratezza, si ricorre a metodi di *bagging* (come le *Random Forest*, che costruiscono molti alberi su campioni differenti del *dataset* e combinano le previsioni dei singoli alberi tramite votazione) o a metodi di *boosting* (dove una sequenza di alberi viene addestrata in modo che ciascun nuovo albero corregga gli errori dei precedenti).

In sintesi, gli alberi decisionali continuano a occupare una posizione di riguardo nel panorama del *machine learning*, perché bilanciano in modo efficace trasparenza, semplicità di implementazione e potere predittivo. Seppur non sempre ottimali per problemi caratterizzati da un'enorme complessità o da funzioni non lineari particolarmente articolate, restano uno strumento altamente versatile, soprattutto quando la possibilità di spiegare le decisioni ai non addetti ai lavori rappresenta una priorità.

2.3.1.2 Naive Bayes

Il *Naive* Bayes è un algoritmo di classificazione basato sui principi della probabilità condizionata, fondato sul teorema di Bayes. Il suo funzionamento consiste nel calcolare la probabilità che un dato oggetto appartenga a ciascuna delle possibili classi, utilizzando le caratteristiche osservate come *input*. L'algoritmo assegna quindi all'oggetto l'etichetta con la probabilità più alta. L'aggettivo "*naive*" deriva dall'assunzione semplificata secondo cui tutte le caratteristiche dell'oggetto in esame sono indipendenti tra loro, condizionatamente alla classe di appartenenza. Questa ipotesi, pur essendo raramente verificata nella pratica, semplifica notevolmente i calcoli e rende l'algoritmo molto efficiente.

Per comprendere il funzionamento del Naive Bayes, è necessario richiamare il concetto di probabilità condizionata. Dati due eventi A e B, dove A rappresenta la classe e B le variabili attributo (cioè, le caratteristiche dell'oggetto), la probabilità condizionata P(A|B) indica la probabilità che l'evento A si verifichi, dato che è noto che si è verificato l'evento B. Questa è definita come:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Da questa relazione è possibile derivare il Teorema di Bayes, che mette in relazione P(A|B) e P(B|A) come segue:

$$P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}$$

Questo risultato è particolarmente significativo, poiché consente di calcolare la probabilità a posteriori P(A|B) di un'ipotesi A (ad esempio, l'appartenenza a una determinata classe) data l'osservazione di una determinata evidenza B.

Una delle assunzioni chiave del modello Naive Bayes è che le caratteristiche dell'evento B siano stocasticamente indipendenti tra loro, condizionatamente alla classe A. Questa ipotesi di indipendenza semplifica notevolmente il calcolo delle probabilità congiunte, rendendo l'algoritmo computazionalmente molto efficiente, anche se spesso non del tutto realistico.

I termini chiave utilizzati nel teorema di Bayes includono:

- P(A), detta probabilità a priori, rappresenta la probabilità iniziale dell'ipotesi A, ossia la probabilità che una determinata classe sia presente nell'intero insieme di dati.
- P(B), nota come evidenza, è la probabilità di osservare le caratteristiche B indipendentemente dalla classe. Questa quantità è costante per tutte le classi e può essere ignorata nel processo di classificazione, poiché non influisce sulla scelta della classe più probabile.
- P(B|A), detta verosimiglianza, rappresenta la probabilità che l'evidenza B si verifichi, dato che l'evento A è vero. Questo termine può essere stimato utilizzando diverse distribuzioni di probabilità, a seconda della natura dei dati.

Una variante comune del Naive Bayes è il Gaussian Naive Bayes, che assume che le caratteristiche siano distribuite secondo una distribuzione normale (Gaussiana). In questo caso, è necessario stimare media e varianza per ciascuna caratteristica, utilizzando tecniche come la stima della massima verosimiglianza (*Maximum Likelihood Estimation, MLE*). Questo approccio consente di massimizzare contemporaneamente la verosimiglianza dei dati e la probabilità a posteriori P(A|B), sfruttando il fatto che P(A) è una costante fissata dal *dataset*.

In sintesi, il Naive Bayes è un algoritmo estremamente semplice ma potente, particolarmente efficace quando le ipotesi di indipendenza condizionale sono almeno approssimativamente valide.

2.3.1.3 K-Nearest Neighbors

L'algoritmo denominato *K-Nearest Neighbors* (abbreviato in K-NN) rappresenta una tecnica di riconoscimento di *pattern* fondata sul criterio di vicinanza nello spazio dei dati. Una delle sue peculiarità è l'estrema semplicità costruttiva, poiché in fase di addestramento si limita a memorizzare le istanze già classificate senza necessitare di un modello esplicito. Inoltre, il KNN è considerato non parametrico, in quanto non fa assunzioni sulla forma o la distribuzione dei dati da analizzare. Ciò lo rende particolarmente idoneo in situazioni in cui non si dispone di informazioni pregresse sulle relazioni o sulle caratteristiche dei dati, ammesso che sia comunque possibile definire una misura di similitudine o distanza tra le osservazioni in un opportuno spazio normato.

A livello operativo, si parte da un *dataset* in cui ogni istanza risulta etichettata con la classe di appartenenza (oppure con un valore numerico, nel caso di un problema di regressione). Prima di utilizzare il KNN, di norma si procede ad allineare le scale di misura delle variabili, così da impedire che un attributo con valori molto grandi condizioni in modo sproporzionato la misura di distanza. A tal fine, si possono adoperare procedure di normalizzazione o standardizzazione.

Quando si deve classificare un nuovo elemento, l'algoritmo calcola la distanza tra quest'ultimo e ciascuna delle istanze note nel *dataset*, selezionando i k punti più vicini (da cui il termine "*nearest neighbors*"). La distanza può essere definita in modi diversi, tra cui la metrica euclidea, la distanza di Manhattan o altri criteri a seconda delle esigenze. Una volta individuati i k vicini più prossimi, si guarda alle rispettive etichette: nella classificazione, la classe prevalente tra questi vicini viene assegnata alla nuova istanza. Se ci si trova in un contesto di regressione, invece, il KNN ricava solitamente la media (o un diverso indicatore, come la mediana) dei valori associati ai k punti più vicini, per stabilire la stima da assegnare all'osservazione in esame.

Uno dei parametri centrali del KNN è appunto k, il numero di vicini da considerare. Da un lato, scegliere un valore troppo basso (ad esempio k=1) può rendere il modello eccessivamente sensibile al rumore locale, perché la classificazione sarà dominata dal singolo punto più prossimo, che potrebbe costituire un'eccezione. Dall'altro lato, incrementare eccessivamente k spinge il modello a basarsi su un campione ampio di vicini, rischiando di classificare l'istanza seguendo la maggioranza complessiva del *dataset*, con possibile perdita di precisione nei casi in cui esistano sottogruppi o

distinzioni sottili. Inoltre, incrementando k aumenta il costo computazionale legato al calcolo delle distanze, poiché l'algoritmo resta *instance-based* e, a ogni nuova richiesta di classificazione, deve confrontare la nuova istanza con tutte (o gran parte) delle osservazioni memorizzate.

Per orientare la scelta di k, alcuni autori suggeriscono regole empiriche, ad esempio assumere $k \approx \sqrt{N}$, dove N è il numero delle istanze nel set di addestramento. In ogni caso, la selezione ottimale può dipendere dalla natura specifica del problema e dal livello di rumore nel *dataset*. Una buona pratica consiste nell'utilizzare un approccio di validazione incrociata (*cross-validation*) per testare diversi valori di k e valutare quale produca i risultati migliori su dati di validazione.

È importante sottolineare che, sebbene il KNN non richieda un vero e proprio processo di addestramento, esso può rivelarsi dispendioso in sede di predizione, poiché l'analisi di una nuova istanza implica il calcolo delle distanze verso l'intero archivio di dati, che potrebbe essere molto consistente. Questa complessità diventa particolarmente significativa quando si opera in spazi ad alta dimensionalità (fenomeno noto come *curse of dimensionality*). In tali ambienti, le distanze tendono a confondersi, riducendo la discriminabilità tra i punti e, di conseguenza, l'efficacia della classificazione basata sul concetto di "vicinanza". Per attenuare tali problematiche, è prassi valutare tecniche di riduzione della dimensionalità o di selezione delle caratteristiche (*feature selection*).

Riassumendo, il KNN risulta di facile interpretazione e non impone ipotesi forti su come siano distribuiti i dati, rivelandosi, di fatto, un solido punto di partenza in molte applicazioni di classificazione e regressione. Tuttavia, occorre bilanciare con attenzione la scelta di k e tenere conto dei costi di calcolo in fase di predizione, specie se i dati presentano una dimensionalità elevata o se l'archivio di riferimento ha dimensioni molto grandi.

2.3.1.4 Support-vector machines

Le *Support-Vector Machines* rappresentano una famiglia di algoritmi di apprendimento supervisionato concepiti per risolvere sia problemi di classificazione sia di regressione, ponendo al centro la ricerca di un "iperpiano ottimale" in grado di separare lo spazio delle osservazioni in maniera netta. In particolare, quando l'obiettivo è classificare, l'SVM cerca di identificare il confine che massimizza il margine tra i dati appartenenti alle

diverse classi. Più precisamente, si considerano come fondamentali i cosiddetti vettori di supporto, ossia i punti del dataset più vicini al confine: la distanza tra i vettori di supporto e l'iperpiano viene definita "margine". L'idea consiste nel selezionare quel confine che, fra tutti quelli possibili, è in grado di mantenere la massima distanza da entrambi i gruppi di punti, così da aumentare la robustezza del modello nei confronti di piccole variazioni o di possibili rumori nei dati.

Formalmente, il problema può essere espresso tramite la minimizzazione di una funzione soggetta a dei vincoli che garantiscano la corretta classificazione dei punti. Viene definito un modello "primitivo" (o *primal*), che è spesso riformulato nella sua versione "*dual*", sfruttando il teorema di Karush-Kuhn-Tucker *(KKT)*. I punti che svolgono un ruolo essenziale nella soluzione sono appunto i vettori di supporto, i cui coefficienti nella formulazione duale risultano non nulli. A livello geometrico, sono i punti più vicini all'iperpiano di separazione, ma che continuano a essere classificati correttamente. Il vantaggio di prendere in considerazione solo tali punti è rendere più efficiente il calcolo, poiché il modello non necessita di memorizzare l'intero *dataset* in fase di predizione.

Un aspetto di rilievo è che non sempre i dati risultano linearmente separabili nello spazio originario. In tal caso, l'SVM adotta quello che viene comunemente chiamato *kernel trick*. In pratica, si proietta il *dataset* in uno spazio di dimensioni potenzialmente superiori (chiamato "spazio delle *feature*"), dove la separazione lineare diventi più facile da ottenere. La scelta della funzione *kernel* (ad esempio lineare, polinomiale, gaussiana RBF, sigmoide) influisce notevolmente sulle performance complessive, poiché definisce la modalità con cui i punti vengono mappati nel nuovo spazio. Nella regressione, un approccio concettualmente simile consente di definire una "fascia" entro cui le predizioni dell'SVM non vengono penalizzate, concentrando l'ottimizzazione sulle deviazioni maggiori di una soglia epsilon.

Dal punto di vista operativo, occorre specificare alcuni iperparametri. Tra i più importanti, figurano:

- Il parametro di regolarizzazione C, che governa il compromesso tra trovare un margine ampio e ridurre al minimo le violazioni ai vincoli di separazione (ovvero gli errori di classificazione). Un valore elevato di C spinge il modello a classificare correttamente anche i punti più isolati, rischiando di ridurre il margine e di

incorrere in *overfitting*; un valore troppo basso, invece, amplia sì il margine, ma può tradursi in una minor precisione se i dati non risultano ben separabili.

- I parametri del *kernel*, come il grado del polinomio se si usa un *kernel* polinomiale, o il parametro γ\gammaγ (gamma) nella funzione *Radial Basis Function* (RBF). Questi incidono sulla flessibilità del confine di separazione: valori eccessivamente alti potrebbero modellare troppo finemente la frontiera intorno ai dati di *training*, con un alto rischio di *overfitting*, mentre valori troppo bassi potrebbero lasciare il modello incapace di cogliere la complessità presente nello spazio delle *feature*.

Sebbene l'SVM si sia inizialmente diffuso in ambito di classificazione binaria, esistono procedure per gestire anche la classificazione multiclasse. Alcune strategie ricorrenti sono:

- La decomposizione *One-vs-One*, in cui si addestra un classificatore SVM per ogni possibile coppia di classi, e in fase di decisione si sceglie quella con più voti vincenti.
- La decomposizione *One-vs-Rest*, dove si costruisce un classificatore per ciascuna classe, distinguendo ogni volta gli elementi di quella classe da tutti gli altri. L'appartenenza definitiva si ottiene, di volta in volta, confrontando i vari punteggi forniti dai classificatori.

Un ulteriore aspetto da considerare è la possibilità di ottenere stime probabilistiche. Di *default*, l'SVM fornisce semplicemente la decisione su quale lato dell'iperpiano cada un punto e, nei problemi di classificazione, si limita a restituire un'etichetta positiva o negativa. Tuttavia, in alcune applicazioni (ad esempio, valutazione del rischio di credito) può essere interessante disporre di una probabilità associata all'evento di *default*. Per questo motivo, si utilizzano tecniche di calibrazione, come il *Platt scaling* o il metodo di *isotonic regression*, per convertire la distanza dal confine di separazione in una stima di probabilità. Questo passaggio, sebbene introduca un livello di complessità in più, risulta spesso utile nei sistemi di supporto alle decisioni che richiedono una misura di incertezza o una soglia di accettazione modulabile.

Da un punto di vista vantaggi/svantaggi, le SVM:

- Punti di forza: offrono un elevato potere predittivo (soprattutto se è ben scelto il *kernel*), gestiscono bene *dataset* con molte dimensioni (fintanto che la scelta di

parametri è accurata) e mostrano buona robustezza al rumore e all'*overfitting* (specialmente con margini ampi).

- Limiti: includono la scarsa interpretabilià rispetto ad alberi decisionali o modelli lineari (soprattutto in presenza di *kernel* non lineari) e la necessità di selezionare correttamente i parametri. Inoltre, per i modelli di classificazione *multiclass*, occorre costruire schemi combinati (*One*-vs-*One* o *One*-vs-*Rest*), e il calcolo della probabilità richiede un livello di calibrazione ulteriore.

In conclusione, *le Support-Vector Machines* si collocano tra i modelli di riferimento per molti problemi di classificazione e regressione, grazie alla capacità di tracciare confini complessi con un rigoroso controllo della complessità del modello. Pur richiedendo una certa esperienza nella configurazione di *kernel* e iperparametri, esse hanno dimostrato negli anni un'affidabilità significativa in numerosi scenari, compresa l'analisi del rischio di credito, dove la corretta identificazione delle soglie di separazione fra buoni e cattivi pagatori riveste un ruolo di cruciale importanza.

2.3.1.5 Modelli di regressione

Il modello di regressione lineare è uno dei metodi più semplici ed efficaci per prevedere valori continui a partire da variabili indipendenti. Quando si considera una sola variabile indipendente, si parla di regressione lineare semplice, mentre nel caso di più variabili indipendenti si utilizza il termine regressione lineare multipla. Al contrario, quando l'obiettivo è classificare le osservazioni in categorie discrete, si ricorre alla regressione logistica, che, nonostante il nome, appartiene alle tecniche di classificazione ed è strettamente connessa alla funzione sigmoide.

Uno degli approcci più comuni per stimare i parametri di un modello di regressione lineare è il metodo dei minimi quadrati ordinari (*Ordinary Least Squares, OLS*). Questo metodo si basa sulla minimizzazione della somma dei quadrati degli errori di previsione, ovvero la differenza tra i valori osservati e quelli stimati dalla retta di regressione. La funzione obiettivo che si vuole minimizzare è:

$$\beta = argmin_b \sum_{i=1}^{n} (y_i - x_i \cdot \mathrm{b})^2$$

dove:

- y_i rappresenta il valore osservato della variabile dipendente,

- x_i è il vettore delle variabili indipendenti per l'osservazione iii,
- b è il vettore dei coefficienti da stimare,
- n è il numero totale delle osservazioni.

Il termine *argmin* indica l'insieme dei valori di b per cui la funzione raggiunge il minimo. Gli stimatori OLS hanno la proprietà di essere non distorti (*bias* nullo), il che significa che, in media, il valore stimato è corretto. Tuttavia, possono presentare una varianza elevata in presenza di dati multicollineari o altamente correlati, il che riduce l'accuratezza delle previsioni e aumenta la probabilità di *overfitting*. Per affrontare questo problema, sono state sviluppate diverse tecniche di regolarizzazione, che aggiungono una penalità ai coefficienti del modello per ridurre la varianza, anche a costo di introdurre un leggero *bias*. Le tecniche di regolarizzazione più comuni includono:

- Regressione *Ridge*

  Questa introduce una penalità quadratica nella funzione obiettivo, aggiungendo un termine proporzionale al quadrato dei coefficienti. L'equazione del modello *Ridge* diventa:

  $$\beta = argmin_b \sum_{i=1}^{n} (y_i - x_i \cdot \mathrm{b})^2 + \lambda \cdot \sum_{k=1}^{K} b_k^2$$

  dove:

  - λ è un parametro di regolarizzazione che controlla l'intensità della penalità,
  - K è il numero totale di coefficienti nel modello.

  Quando λ è molto piccolo, l'effetto della penalità è trascurabile e il modello si avvicina alla regressione OLS classica. All'aumentare di λ, i coefficienti vengono ridotti, limitando l'impatto dei predittori meno rilevanti e migliorando la robustezza del modello. Tuttavia, un valore di λ eccessivamente alto può portare a un modello troppo semplificato, con coefficienti vicini allo zero.

  Un vantaggio significativo della regressione *Ridge* è la sua capacità di gestire situazioni in cui il numero di variabili indipendenti è superiore al numero di osservazioni, riducendo il rischio di *overfitting*. Tuttavia, uno svantaggio è che questa tecnica non effettua una selezione esplicita delle variabili, mantenendo tutti i predittori nel modello, anche quelli meno rilevanti.

- Regressione LASSO

La regressione LASSO (*Least Absolute Shrinkage and Selection Operator*) introduce una perturbazione lineare, diversa dalla penalità quadratica della Ridge, secondo la formula:

$$\beta_\lambda = argmin_b \sum_{i=1}^{n} (y_i - x_i \cdot \mathrm{b})^2 + \lambda \cdot \sum_{k=1}^{K} |b_k|$$

In questo caso, l'effetto della penalizzazione è tale da ridurre gradualmente i coefficienti dei predittori man mano che il parametro λ aumenta. Quando λ è sufficientemente grande, alcuni coefficienti vengono compressi fino a diventare esattamente zero, eliminando così le variabili meno rilevanti dal modello. Questo processo permette di ottenere una selezione automatica dei predittori, semplificando la struttura del modello e migliorandone l'interpretabilità.

Se λ è vicino a zero, la regressione produce risultati simili a quelli ottenuti con il metodo dei minimi quadrati ordinari (OLS), poiché la penalizzazione è trascurabile. Tuttavia, all'aumentare di λ, la penalizzazione diventa più significativa, forzando molti coefficienti a zero e mantenendo solo quelli con un contributo sostanziale alla previsione. Questo approccio è particolarmente vantaggioso quando il modello è caratterizzato da pochi predittori importanti e molti altri con contributi trascurabili. A differenza della regressione Ridge, che tende a ridurre tutti i coefficienti in modo uniforme senza annullarli completamente, la regressione LASSO è più adatta a situazioni in cui alcuni predittori hanno un impatto significativo e altri sono marginali.

- *Elastic Net*

  Combina le penalità *Ridge* e LASSO in un'unica funzione obiettivo:

$$\beta_{\mathrm{enet}} = \frac{argmin_b \sum_{i=1}^{n}}{2n} + \lambda \left(\frac{1-\alpha}{2} \sum_{k=1}^{K} b_k^2 + \alpha \sum_{k=1}^{K} |b_k|\right)$$

  Dove α è un parametro che controlla il bilanciamento tra la penalità *Ridge* (α=0) e Lasso (α=1\).

  Questo approccio è particolarmente utile in contesti in cui i predittori sono altamente correlati, sfruttando i vantaggi di entrambi i metodi. L'*Elastic Net* è noto per la sua capacità di gestire problemi di multicollinearità e di selezionare gruppi di variabili correlate, migliorando la stabilità del modello rispetto al LASSO puro.

In sintesi, la scelta tra *Ridge*, LASSO ed *Elastic Net* dipende dalla natura dei dati e dall'obiettivo del modello. Mentre *Ridge* è più adatto a situazioni con molti predittori di pari importanza, il LASSO è preferibile quando si desidera ottenere modelli semplici e interpretabili. L'*Elastic Net* rappresenta un compromesso efficace quando si desidera combinare questi due approcci.

2.3.1.6 Modelli di *Ensemble Learning*

Il metodo *Ensemble Learnings* è una strategia di apprendimento automatico che combina più modelli differenti per ottenere un sistema di classificazione complessivamente più accurato e robusto rispetto ai singoli modelli presi isolatamente. Questa tecnica sfrutta la diversità dei modelli per massimizzarne le prestazioni, utilizzando i punti di forza di ciascuno e riducendo al minimo le loro debolezze. In pratica, l'*ensemble* crea una sinergia tra i modelli di base, detti classificatori deboli (*weak learners*), che, se combinati in modo appropriato, danno vita a un classificatore forte (*strong learner*).

Il principio fondamentale di questa metodologia è che, pur essendo i singoli modelli suscettibili a errori, questi errori tendono a bilanciarsi reciprocamente se i modelli sono sufficientemente eterogenei. Questo avviene perché ogni modello ha una propria prospettiva sui dati e, di conseguenza, commette errori in modo diverso. Quando i risultati dei singoli classificatori vengono aggregati, gli errori casuali di ciascun modello si distribuiscono in maniera più uniforme, annullandosi parzialmente a vicenda, mentre le previsioni corrette tendono a concentrarsi intorno alla soluzione giusta.

Questo meccanismo porta a due principali vantaggi:

- Maggiore capacità di generalizzazione, in quanto riduce il rischio di *overfitting*, in quanto le previsioni sono meno influenzate da fluttuazioni casuali nei dati di addestramento, migliorando così l'accuratezza sui dati non visti.
- Maggiore robustezza agli *outlier*, dal momento che questi ultimi tendono a influenzare alcuni modelli più di altri, un *ensemble* può attenuare l'impatto di questi valori anomali, rendendo le previsioni complessive più affidabili.

Inoltre, l'efficacia dell'*Ensemble Learning* dipende dalla capacità dei modelli di base di commettere errori in modo indipendente. Se tutti i classificatori sono fortemente correlati, è probabile che commettano errori simili, riducendo i benefici dell'*ensemble*. Per questo

motivo, è essenziale utilizzare modelli diversi per massimizzare la diversità e ottenere un miglioramento significativo delle prestazioni complessive.

Ci sono tre diversi tipi di *ensemble*.

Bagging

Si parte da un insieme di addestramento iniziale contenente N osservazioni, e si creano M campioni *bootstrap*, ognuno di ampiezza N, estraendo con reinserimento dal *dataset* originale. Ciascun campione viene usato per istruire un modello debole (spesso un albero decisionale). In fase di predizione, tutti i modelli così ottenuti partecipano alla decisione: se si tratta di classificazione, la classe finale è scelta mediante votazione a maggioranza; se si tratta di regressione, si effettua la media delle uscite numeriche.

L'intuizione alla base del *bagging* è che, generando *dataset* parzialmente diversi grazie al *sampling* con rimpiazzo, i singoli modelli addestrati producano errori scarsamente correlati. Ciò aiuta a contenere il problema dell'*overfitting* tipico di un singolo modello molto profondo (ad esempio un albero). Nell'ottica del *bagging*, gli alberi di tipo *Decision Tree* vengono talvolta resi ulteriormente eterogenei mascherando una frazione degli attributi a ogni nodo, come avviene in una *Random Forest*. Quest'ultima, infatti, seleziona casualmente, a ogni *split*, un certo sottoinsieme di variabili candidate, facendo sì che i vari alberi differiscano non soltanto perché "vedono" dati di training diversi, ma anche perché valutano *subset* di attributi differenti. Questa duplice casualizzazione (*bootstrap* dei *record* e *random subspace* degli attributi) incrementa ancora l'indipendenza tra i singoli modelli, migliorando l'accuratezza di previsione sull'insieme di *test*.

Gradient Boosting

A differenza del *bagging*, che costruisce i modelli in parallelo su diversi campioni *bootstrap*, il *Boosting* realizza invece un'aggregazione sequenziale di modelli, in cui ciascun nuovo *weak learner* si concentra sugli errori commessi dagli stadi precedenti. Il risultato finale è un composito "a cascata" di modelli, con l'obiettivo di correggere via via le imperfezioni residue.

La versione originaria, parte da un insieme di pesi uniformi assegnati a tutte le osservazioni. Dopo l'addestramento del primo classificatore debole, le istanze classificate erroneamente vengono "potenziate" (ossia, si incrementano i loro pesi) e si normalizzano i pesi totali affinché la loro somma resti invariata. Al passo successivo, il nuovo modello si addestra tenendo conto di tali pesi aggiornati; questo iter prosegue per un numero

stabilito di *round* o finché l'errore scende sotto una soglia prefissata. Nel modello finale, i classificatori sono combinati con un peso proporzionale alla loro accuratezza. Se il classificatore h_m ottiene un errore ϵ_m, il peso associato può essere una funzione monotona decrescente in ϵ_m. Così, i modelli più accurati ricevono una rilevanza maggiore nella previsione conclusiva.

Una delle evoluzioni più diffuse del concetto di *boosting* è il *Gradient Boosting*, che riformula il problema come un'ottimizzazione graduale di una funzione obiettivo, in cui a ogni iterazione si aggiunge un nuovo modello che approssima il gradiente degli errori residui. Questa idea è stata implementata in librerie molto popolari, come *XGBoost*, *LightGBM* e *CatBoost*, ognuna con strategie di compressione dei dati e ottimizzazioni specifiche per velocizzare l'addestramento e migliorare la gestione di *feature* categoriali o mancanti.

Stacking

Lo *Stacking* (o *Stacked Generalization*) è una tecnica di *ensemble* in cui i modelli base (di tipologie anche molto diverse fra loro) vengono allenati in parallelo sul medesimo *dataset* e i rispettivi *output* fungono da *input* per un modello di secondo livello, detto *meta-learner* o *meta-classifier*. In altre parole, invece di votare o mediare semplicemente le previsioni dei modelli deboli, si introduce un ulteriore algoritmo che "impara" come combinare nel modo più proficuo i risultati dei diversi *base learners*.

2.3.2 Apprendimento non supervisionato

Nell'apprendimento non supervisionato, non esistono etichette di riferimento. L'obiettivo, pertanto, è individuare strutture latenti, correlazioni, *pattern* o raggruppamenti (*cluster*) direttamente dai dati. Questa tipologia di apprendimento trova ampio utilizzo, ad esempio, nella segmentazione della clientela e nell'individuazione di *pattern* ricorrenti in serie finanziarie o *log* di transazioni.

2.3.2.1 *Clustering*

Il *clustering* comprende un ventaglio di tecniche di analisi multivariata che si prefiggono di suddividere un insieme di dati in gruppi (*cluster*), nei quali gli elementi risultino tra loro il più possibile simili (omogenei) e, al contempo, significativamente diversi dagli elementi appartenenti ad altri *cluster*. In molti approcci, questa "similarità", è tradotta nel

concetto di distanza entro uno spazio multidimensionale, in modo che l'appartenenza o l'esclusione da un gruppo dipendano da quanto ciascun punto risulti distante dal centro o dalle altre entità del medesimo *cluster*.

L'analisi di *clustering*, a differenza delle metodologie supervisionate, non fa riferimento a etichette predefinite (*labels*). L'obiettivo è piuttosto di scoprire autonomamente relazioni, strutture o raggruppamenti insiti nei dati, facendo leva su misure di prossimità. I vari algoritmi di *clustering* si distinguono principalmente per la strategia con cui formano i gruppi e per le assunzioni implicite circa la forma e la distribuzione dei *cluster*. Una prima classificazione fa riferimento all'organizzazione gerarchica e a quella partizionale.

Metodi gerarchici

Questi metodi costruiscono una struttura ad albero che riflette diversi livelli di raggruppamento, dall'intero *dataset* considerato come un singolo *cluster* sino a un frazionamento in *cluster* sempre più piccoli. A loro volta, si dividono in:

- Approcci *bottom-up* (metodi aggregativi): iniziano considerando ogni punto come un *cluster* a sé; quindi, procedono unendo via via i *cluster* più simili, fino a ottenere un numero prestabilito di gruppi (o sino a quando la distanza tra i *cluster* diventa eccessiva).
- Approcci *top-down* (metodi divisivi): partono da un unico grande *cluster* comprendente tutti i dati, scindendolo gradualmente in sottogruppi via via più ridotti.

  Una volta determinata la gerarchia, la selezione del livello al quale "tagliare" l'albero fornisce la suddivisione finale in un certo numero di *cluster*.

Metodi partizionali

Qui i dati vengono direttamente suddivisi in k *cluster* (o in un numero variabile di *cluster*) ottimizzando un criterio di similarità interna o di dissimilarità verso l'esterno. Esempi

noti di algoritmi partizionali sono il K-*means*[28], il DBSCAN[29] o l'E-M[30], ciascuno caratterizzato da un particolare modo di definire la distanza, la densità o la probabilità di appartenenza a un *cluster*.

Un ulteriore criterio di distinzione deriva dalla natura esclusiva o non esclusiva del *clustering*:

- *Clustering* esclusivo (*hard clustering*): ciascun elemento è assegnato a un solo *cluster*, così che i gruppi risultanti non abbiano elementi in comune.
- *Clustering* non esclusivo (*soft* o *fuzzy clustering*): un elemento può appartenere a più *cluster*, seppur con gradi di appartenenza differenti, tipicamente gestiti attraverso la logica *fuzzy*.

Nell'ambito bancario e finanziario, il *clustering* si dimostra particolarmente utile per identificare segmenti di clientela, gruppi di prodotti o aree geografiche con caratteristiche omogenee, consentendo di elaborare strategie personalizzate di *marketing* o di gestione del rischio.

Occorre infine evidenziare che:

- Il risultato del *clustering* dipende in modo significativo dalla scelta della misura di distanza (euclidea, Manhattan, coseno, Mahalanobis, ecc.), che può influenzare notevolmente la formazione dei gruppi.
- La normalizzazione o la standardizzazione delle variabili gioca spesso un ruolo cruciale per evitare che una *feature* con valori molto grandi domini il calcolo delle distanze.
- È importante valutare la validità dei *cluster* ottenuti. Diversi indici, come l'indice di *silhouette*, l'indice di Davies-Bouldin o l'errore di coesione/separazione, consentono di misurare quanto i *cluster* risultino compatti e ben separati. Anche approcci più empirici, come l'interpretazione manuale o la verifica con esperti di dominio, possono aiutare a giudicare la bontà della partizione.

---

[28] Il *K-means* è un algoritmo di *clustering* partizionale introdotto da MacQueen nel 1967, noto per la sua semplicità ed efficienza computazionale, ma sensibile alla scelta del numero di *cluster* e ai valori anomali.

[29] Il DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), sviluppato da Ester et al. nel 1996, è particolarmente efficace nell'identificare *cluster* di forma arbitraria e nel gestire i dati rumorosi.

[30] L'algoritmo E-M (*Expectation-Maximization)* è invece un metodo iterativo per stimare i parametri di distribuzioni probabilistiche, utilizzato frequentemente per *clustering* di dati continui e misti."

Così, il *clustering* si configura come uno strumento chiave per l'analisi esplorativa e la segmentazione dei dati, nonché un punto di partenza per metodologie ibride che integrano in un secondo momento la supervisione umana o il confronto con etichette reali. Nel *credit risk management*, si presta, ad esempio, a porre in luce sottogruppi di clienti con abitudini di spesa o profili di redditività e insolvenza potenzialmente simili, un'informazione preziosa tanto per la fase di *risk assessment* quanto per la definizione di politiche commerciali adeguate a ciascun *cluster*.

2.3.2.2 K-means

Il *K-means* rappresenta uno dei metodi di *clustering* più diffusi e applicati. È un algoritmo di tipo *partizionale* che suddivide i dati in k raggruppamenti (*cluster*) predefiniti, cercando di minimizzare la distanza interna a ciascun *cluster*. Nel caso più classico, ogni *cluster* è descritto dal proprio centroide, ovvero un punto ipotetico (non necessariamente compreso nel *dataset*) che funge da "baricentro" del gruppo.

La procedura di base può essere illustrata nei passaggi che seguono:

1. Scelta di k: si stabilisce preliminarmente quanti *cluster* si desiderano individuare. La selezione di questo parametro è spesso non banale e richiede conoscenza pregressa o l'impiego di metodi per valutare la bontà delle diverse soluzioni (ad esempio, metodo del gomito, *silhouette*, ecc.).
2. Inizializzazione dei centroidi: si posizionano casualmente k punti nello spazio delle *feature*. Questi punti costituiscono una prima stima dei centroidi. Un'alternativa più accurata è data da tecniche come *k-means++*, che mirano a distribuire inizialmente i centroidi in modo più intelligente, riducendo la probabilità di convergere a soluzioni di scarsa qualità.
3. Assegnazione ai *cluster*: per ciascuna osservazione del *dataset*, si calcola la distanza dal centroide e la si assegna al gruppo corrispondente al centroide più vicino. La distanza può essere valutata con metriche diverse (euclidea, Manhattan, Minkowski, ecc.), a seconda delle peculiarità del problema e dei dati.
4. Ricalcolo dei centroidi: una volta che tutti i punti del *dataset* sono stati assegnati a uno dei k gruppi, si aggiornano i centroidi, in genere calcolando la media di tutte le osservazioni appartenenti allo stesso *cluster*. In questo modo, il nuovo centroide può trovarsi in una posizione differente rispetto a quello iniziale.

5. Iterazione: i due passaggi precedenti (assegnazione ai *cluster* e aggiornamento dei centroidi) vengono ripetuti fino a che i centroidi non mostrano più modifiche significative (oppure finché si raggiunge un numero massimo di iterazioni stabilito). A convergenza, i *cluster* risultano stabilizzati: eventuali variazioni risulterebbero minime o assenti.

Il *K-means* tende quindi a creare regioni "sferiche" attorno ai centroidi, pertanto, se i dati presentano strutture molto irregolari, non lineari o con densità fortemente variabile, questo algoritmo può non coglierne appieno la complessità. Inoltre, richiede l'impostazione a priori del parametro k, il che non è sempre intuitivo.

Un ulteriore limite di *K-means* è la sua sensibilità ai valori anomali (*outlier*): bastano pochi punti distanti per spostare i centroidi in zone non rappresentative della struttura principale del *dataset*. Per ridurre tale effetto, talvolta si effettua un pretrattamento dei dati per individuare e isolare gli *outlier* o si utilizzano varianti dell'algoritmo che scelgono effettivamente un punto del *dataset* come rappresentante del *cluster* anziché la media dei membri.

Malgrado tali accorgimenti necessari, *K-means* resta un metodo assai popolare per la sua combinazione di semplicità concettuale ed efficienza computazionale. Una volta che i *cluster* sono definiti, ciascun nuovo dato può essere classificato in tempo rapido, calcolando semplicemente la distanza dai centroidi. Inoltre, nei casi in cui i raggruppamenti mostrino effettivamente forme tendenzialmente sferiche e una dimensione paragonabile, *K-means* si rivela molto efficace, restituendo risultati interpretabili e facilmente utilizzabili in analisi successive, come l'individuazione di profili di rischio omogenei o la personalizzazione di strategie commerciali.

2.3.2.3 DBSCAN

DBSCAN è un metodo di *clustering* che sfrutta il concetto di densità per individuare raggruppamenti di forma non necessariamente regolare e isolare gli elementi rumorosi dal resto dei dati. A differenza di altri algoritmi, non occorre specificare a priori quanti *cluster* si desidera ottenere. Il processo di scoperta dei gruppi si basa invece su due parametri fondamentali:

- epsilon (eps, ε): per un punto p, indica l'ampiezza di una "sfera" entro cui ricercare i punti vicini a p;

- min_pts: definisce il numero minimo di punti richiesto a p per essere considerato un "*core point*". Di norma, questo parametro non va impostato a un valore inferiore a (dimensione dello spazio + 1). In un *dataset* bidimensionale, per esempio, min_pts sarà almeno 3. Aumentarlo riduce l'incidenza di *cluster* spuri, ma occorre anche rivedere il valore di eps per bilanciare correttamente la rilevazione dei gruppi.

All'interno del *dataset*, i punti vengono classificati in tre categorie:

1. ***Core point***: ha almeno min_pts vicini entro la distanza eps. Tale nucleo genera un *cluster*.
2. ***Border point***: non possiede il numero minimo di vicini per essere un *core point*, ma rientra nell'intorno di un *core point* già definito, per cui viene aggregato al suo *cluster*.
3. ***Noise point***: si colloca fuori da qualunque intorno di un *core point*, risultando isolato e considerato un valore anomalo.

Uno snodo cruciale di DBSCAN è la cosiddetta "connessione secondo densità": se un insieme di *core point* è unito da percorsi che rispettano il vincolo di prossimità, quei punti rientrano in un'unica regione, mentre i punti rumorosi rimangono esclusi.

Questo algoritmo presenta due vantaggi principali: da un lato, riconosce *cluster* di forma arbitraria (non per forza sferica), dall'altro esclude gli *outlier* senza richiedere una soglia di similarità o il numero di gruppi da formare. La criticità sta nel gestire *dataset* che mostrano variazioni notevoli di densità: l'utilizzo di un singolo valore di eps e min_pts potrebbe non catturare correttamente *cluster* con livelli di densità molto diversi fra loro.

2.3.2.4 E-M

L'algoritmo E-M (*Expectation-Maximization*) fa parte dei metodi di *clustering* non esclusivi e postula che i dati provengano da un insieme di distribuzioni combinate: ogni *cluster* genera osservazioni in base a una propria legge probabilistica, e nel complesso si ottiene una distribuzione multimodale.

Il fine di E-M è risalire, a partire dalle osservazioni, ai parametri che caratterizzano ciascuna distribuzione, assumendo che la forma di queste ultime sia nota e identica per ogni gruppo. Spesso, come caso comune, si impiegano più distribuzioni gaussiane, di cui bisogna stimare i parametri.

La procedura di stima sfrutta la massima verosimiglianza (MLE), in cui la verosimiglianza indica quanto sia probabile che i dati osservati derivino da un determinato modello. Per motivi di stabilità numerica, si massimizza il logaritmo della verosimiglianza.

L'algoritmo procede a iterazioni, alternando due passaggi fondamentali:

- ***Expectation***: per ciascuna osservazione, vengono calcolate le probabilità di appartenenza alle varie distribuzioni;
- ***Maximization***: medie e varianze di ogni distribuzione si ricalcolano secondo il criterio di massima verosimiglianza.

Queste fasi si ripetono finché non si raggiunge la convergenza dei parametri. In pratica, E-M può essere considerato una generalizzazione probabilistica del *K-Means*.

2.3.2.5 Regole di associazione

Le regole di associazione sono tecniche di *data mining* utilizzate per identificare *pattern* ricorrenti e connessioni nascoste tra insiemi di elementi in grandi volumi di dati. L'obiettivo principale è quello di rilevare relazioni statisticamente significative tra gruppi di oggetti, indipendentemente dalle preferenze individuali, concentrandosi invece sulle combinazioni di elementi che tendono a verificarsi insieme in diverse transazioni.

A differenza di altri algoritmi di *machine learning*, che si concentrano sulla previsione di una variabile *target*, le regole di associazione mirano a scoprire connessioni intrinseche tra insiemi di elementi, spesso identificati come *item set*, presenti in un insieme più grande di transazioni. L'obiettivo non è quello di estrarre le preferenze di un singolo utente, ma piuttosto di identificare le correlazioni più comuni all'interno dell'intero *dataset*.

Nonostante la loro versatilità, queste tecniche presentano alcune limitazioni. Uno dei problemi principali è il costo computazionale associato all'elaborazione di grandi *dataset*. Anche quando si utilizzano algoritmi ottimizzati come Apriori, il numero di combinazioni possibili può diventare enorme in presenza di inventari estesi o soglie di supporto molto basse. Questo rende il processo di ricerca delle regole estremamente dispendioso in termini di tempo e risorse computazionali.

Inoltre, abbassare eccessivamente la soglia di supporto per catturare *pattern* rari può aumentare il rischio di identificare associazioni spurie, ovvero relazioni che appaiono statisticamente significative solo per caso e non riflettono una vera connessione tra gli

elementi. Questo può portare a risultati fuorvianti se le regole identificate non vengono opportunamente validate.

Per ridurre questi rischi e garantire che le regole di associazione siano effettivamente generalizzabili, è consigliabile adottare un approccio in due fasi:

- Addestramento iniziale, il quale consiste nell'identificazione delle regole su un *set* di dati di addestramento.
- Validazione, ossia la verifica delle regole su un *set* di dati di *test* separato, per confermare che i *pattern* scoperti siano applicabili anche a nuovi dati non osservati in precedenza.

Questa strategia aiuta a limitare l'inclusione di correlazioni spurie e a migliorare l'affidabilità delle regole individuate, rendendole più utili per l'implementazione pratica.

2.4 Approcci relazionali e transazionali nello sviluppo del *prescore*, con richiamo al tema di modelli dinamici

L'evoluzione dei modelli di valutazione del rischio di credito è fortemente intrecciata con il dibattito sull'efficacia comparata tra prestiti di natura "relazionale" e quelli di tipo "transazionale". Le recenti evidenze empiriche, tra cui quelle illustrate nel documento di Banca d'Italia "*Artificial Intelligence* and *Relationship Lending*"[31], mostrano come l'impiego di tecnologie di intelligenza artificiale (IA), soprattutto nella fase di *credit scoring*, incida in modo decisivo sui meccanismi di formazione del merito creditizio e sulla stabilità delle relazioni banca-impresa. Nel tentativo di integrare queste risultanze all'interno di un modello di *prescore* più evoluto e dinamico, è opportuno soffermarsi su come la componente relazionale e quella transazionale possano essere adeguatamente coniugate.

All'interno della letteratura economico-finanziaria, i prestiti relazionali (*relationship lending*) si basano sull'idea che la banca accumuli conoscenza "*soft*" sul cliente nel corso di un rapporto protratto nel tempo. Tale informazione è spesso qualitativa: comprende, ad esempio, giudizi sul *management*, analisi dell'andamento di progetti in cui l'azienda è coinvolta, valutazioni dirette sull'affidabilità dell'imprenditore e sul suo storico di interazione con l'istituto di credito. Questa conoscenza, difficilmente codificabile con parametri numerici standardizzati, rappresenta però un vantaggio informativo in termini

[31] Gambacorta, Sabatini e Schiaffi, 2025

di riduzione delle asimmetrie: la banca può anticipare segnali di deterioramento o, all'opposto, cogliere nuove opportunità di crescita del cliente molto prima che emergano dai meri dati contabili.

Per contro, i prestiti transazionali (*transactional lending*) si basano quasi esclusivamente su dati "*hard*", ossia variabili quantitative e relativamente omogenee[32] e su metodologie statistiche o algoritmiche di valutazione (*credit scoring*). Il ricorso massiccio a tali tecniche ha permesso, nel tempo, di standardizzare e velocizzare molte operazioni di concessione del credito, grazie all'applicazione di modelli matematici sempre più raffinati. La letteratura classica sul *credit scoring*, da Altman (1968) fino agli studi più recenti in ambito di *machine learning*, ha illustrato i vantaggi di un approccio basato su informazioni codificate, specie in un panorama finanziario dove il volume di richieste di finanziamento è cresciuto esponenzialmente.

L'interrogativo cruciale riguarda il modo in cui queste due prospettive possano convivere e, soprattutto, come possano interagire in un modello di *prescore* dinamico. Da un lato, l'approccio relazionale non può prescindere dalla fiducia costruita con il cliente, fiducia che si consolida con la continuità del rapporto; dall'altro, i dati transazionali ampliano enormemente la base informativa e permettono un monitoraggio costante e aggiornato delle condizioni aziendali. Il dibattito è rimasto a lungo polarizzato, ma gli studi sul campo dimostrano che la vera sfida sia conciliare questi due approcci in uno schema ibrido, nel quale la tecnologia (intelligenza artificiale, *big data*, algoritmi di *machine learning*) supporti tanto l'acquisizione di dati "*hard*" quanto la gestione di informazioni più "*soft*" scaturite dal rapporto di lungo periodo.

Il grafico sottostante (Figura 1) mostra l'andamento delle linee di credito (*credit lines*), dei prestiti a termine (*term loans*) e del totale degli impieghi, evidenziando la dinamica "ciclica" del credito (specie durante la fase Covid).

[32] bilanci d'impresa, flussi di cassa, andamenti delle vendite, storico dei rimborsi

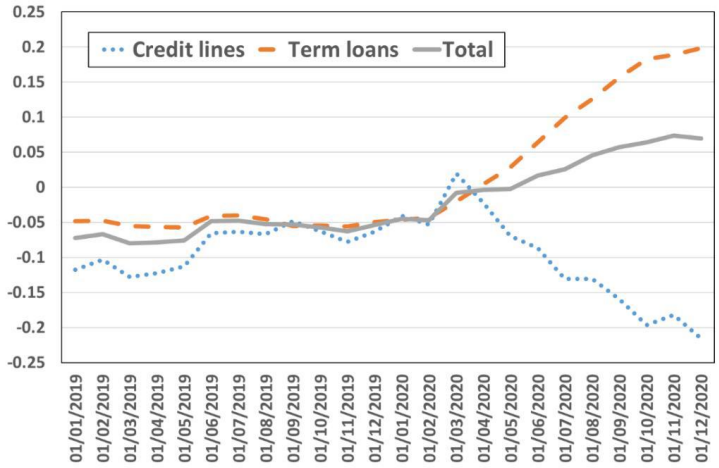


Figura 1 Crescita dei prestiti alle società non finanziarie in Italia[33]

2.4.1 Evidenze empiriche: come l'IA modifica la relazione banca-impresa

Nel loro lavoro, Gambacorta, Sabatini e Schiaffi (2025) esplorano come gli investimenti delle banche italiane in strumenti di intelligenza artificiale per il *credit scoring* interagiscano con la tradizionale relazione di lungo termine tra banca e impresa. Partendo da un *dataset* unico, che connette (i) i dati di bilancio e di affidamento delle imprese con (ii) i dati sulle banche italiane che hanno introdotto tecniche di AI per la valutazione del merito creditizio, gli autori mettono in luce diversi aspetti.

In primo luogo, il lavoro conferma risultati consolidati in letteratura (ad esempio, i benefici del "rapporto di fiducia" nelle fasi di crisi). In momenti di *shock* macroeconomici o settoriali, un vincolo relazionale solido offre la possibilità alla banca di continuare a sostenere un'impresa nonostante il peggioramento di alcuni parametri transazionali di breve periodo.

L'IA possiede la capacità di mitigare l'estrazione di rendita (*rent extraction*) nei periodi "normali". Quando non si è in presenza di situazioni di crisi, la banca che si affidi eccessivamente alle informazioni "morbide" potrebbe imporre tassi più alti o condizioni meno vantaggiose al cliente, proprio perché fa leva sulla conoscenza privilegiata del suo profilo di rischio. L'uso di modelli di *machine learning*, addestrati su *dataset* ampi e

[33] Fonte: BCE. Elaborazioni su dati AnaCredit. Note: variazioni percentuali su 12 mesi. I dati non sono rettificati per gli effetti della cartolarizzazione. Questa cifra include linee di credito e prestiti con scadenza originaria fino a un anno (linea tratteggiata). I prestiti a termine (linea tratteggiata) sono sostituiti da prestiti con scadenza originaria superiore a un anno.

comprensivi di molte variabili "dure" (pagamenti, movimenti di conto, bilanci infrannuali, etc.), sembra ridurre tale spazio di discrezionalità e migliorare la concorrenza sul costo del credito.

Inoltre, il ricorso all'IA, se da un lato può affinare la diagnosi sulle singole posizioni, talvolta innalzandone la "soglia di accettazione", dall'altro favorisce (soprattutto nelle banche di dimensione medio-grande) un allargamento a nuovi segmenti di clientela, poiché i dati transazionali e gli algoritmi di IA consentono di valutare con ragionevole affidabilità anche imprese con storico creditizio limitato.

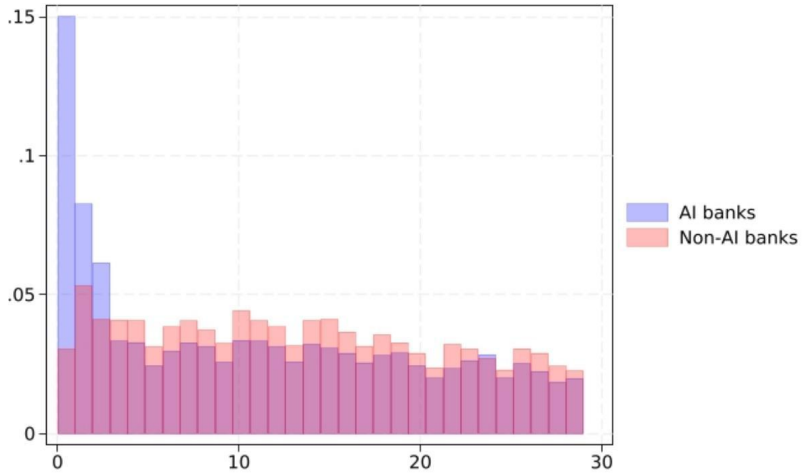


Figura 2 Distribuzione della lunghezza della relazione per le banche AI e non AI[34]

Un aspetto fortemente sottolineato dagli autori è la necessità di non considerare IA e relazioni di lungo periodo come due poli opposti e separati. Al contrario, un sistema avanzato di *credit scoring* può integrare punteggi costruiti su dati "relazionali" (ad esempio, segnalando la stabilità del rapporto, i *feedback* provenienti da filiali territoriali, i giudizi qualitativi su *management* e *governance*) e inserirli in algoritmi di *machine learning* che usano anche dati "duri". In questo modo, la dimensione qualitativa tipica del *relationship lending* si trasforma in *input* formalizzati, pur conservando la natura informativa preziosa.

[34] questa distribuzione visualizza in modo immediato come le banche "innovative" (AI) e quelle "non innovative" si differenziano riguardo alla durata media delle relazioni. È ideale per illustrare "empiricamente" come l'IA si associ a rapporti più o meno consolidati e per suggerire l'idea che l'IA possa agire da fattore di discontinuità rispetto al "classico" *relationship lending* di lungo periodo.

2.4.2 L'approccio microeconometrico: dal metodo di Khwaja e Mian (2008) all'analisi dell'impatto dell'IA

Gli autori, adottando un impianto metodologico di tipo microeconometrico, hanno unito le informazioni relative alle singole banche e quelle riferite alle singole imprese in un campione *panel*, così da osservare l'evoluzione dei rapporti di affidamento su più periodi temporali consecutivi. In particolare, hanno tratto ispirazione dal metodo proposto da Khwaja e Mian (2008), che si focalizza sulla necessità di scomporre la componente di domanda di credito (riconducibile alle esigenze peculiari dell'impresa) da quella di offerta (rimandabile alle scelte e alle politiche della banca). Questa distinzione è cruciale quando si vuole stabilire, con sufficiente rigore statistico, come fattori quali l'adozione dell'intelligenza artificiale per il *credit scoring* o la durata della relazione tra banca e impresa influiscano sul volume dei prestiti e sui tassi applicati.

L'aspetto centrale del modello è rappresentato dall'introduzione di un duplice insieme di effetti fissi (*fixed effects*).

Bank-time fixed effects: questi consentono di catturare, in ogni finestra temporale (che di solito è trimestrale o semestrale, a seconda della granularità dei dati), tutte le caratteristiche specifiche dell'intermediario. In tal modo, si bloccano eventuali oscillazioni connesse, ad esempio, a variazioni di liquidità della banca, scelte strategiche interne (come una rimodulazione del portafoglio crediti) o condizioni di mercato che possano incidere su quell'istituto in particolare.

Firm-time fixed effects: parallelamente, per ogni impresa e in ogni periodo si "filtra" qualunque *shock* di domanda, come un improvviso aumento del fabbisogno di liquidità per ragioni commerciali, un calo delle vendite o un investimento imprevisto. Così facendo, le differenze nell'erogazione di credito non dipendono da quanto l'impresa "chiede" in un determinato momento (domanda), ma da come ogni banca risponde a quello stesso "stimolo".

Questa impostazione regge sulla premessa che l'impresa, nello stesso periodo, possa avere rapporti con più banche, rendendo possibile il confronto tra i diversi comportamenti di offerta creditizia. Di conseguenza, se un'azienda X necessita di maggiore liquidità nel secondo trimestre di un dato anno, tutte le banche con cui X collabora saranno sottoposte alla medesima variazione di domanda. Ciò che lo studio di Gambacorta, Sabatini e Schiaffi evidenzia, allora, è in che modo queste banche rispondano in maniera

differenziata, in relazione (tra le altre cose) al fatto che utilizzino o meno tecnologie di IA per il *credit scoring* e alla lunghezza del rapporto già esistente con la stessa impresa.

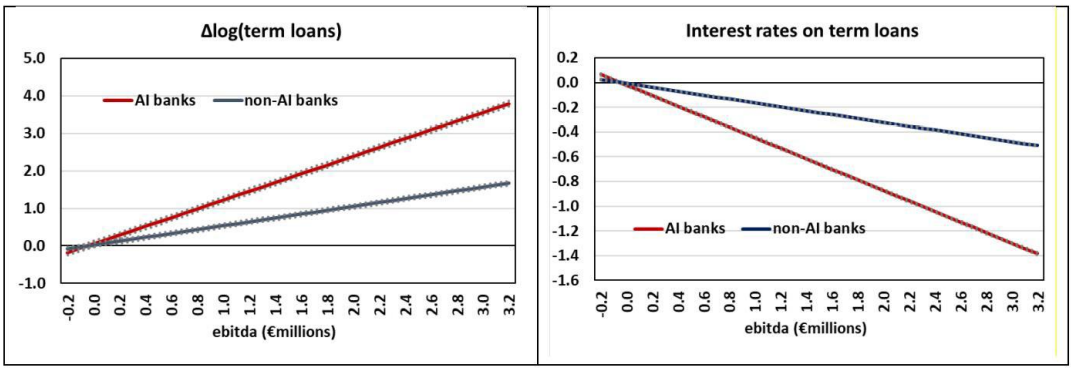


Figura 3 Diversa reattività delle banche AI e non AI alle variazioni dell'EBIDTA[35]

Un ulteriore elemento di rilevanza, non ancora discusso nei paragrafi precedenti, è la gestione delle dinamiche intertemporali. L'uso di un *dataset panel* consente infatti di non limitarsi a un singolo scatto fotografico (*cross-section*) ma di osservare il susseguirsi di più intervalli temporali. È in questo contesto che la variabile "relazione di lungo termine" (*relationship length*) assume pieno significato: essa non si riduce semplicemente a un numero di mesi o anni trascorsi dall'avvio dell'affidamento, ma è declinata come una continuità effettiva del rapporto, trimestre dopo trimestre. Il valore di questa prospettiva dinamica è determinante, poiché permette di verificare, ad esempio, se i vantaggi informativi accumulati negli anni dalla banca su una specifica impresa siano impiegati diversamente quando la banca stessa implementa algoritmi di IA, rispetto a quando la valutazione del rischio resta affidata a tecniche statistiche convenzionali o al giudizio basato prevalentemente su *soft information*.

Un altro aspetto spesso trascurato, ma cruciale in un'analisi microeconometrica, riguarda la pulizia e preparazione del *dataset*. Gli autori spiegano come la costruzione dei *bank-time fixed effects* e dei *firm-time fixed effects* richieda un consistente lavoro di allineamento delle date di riferimento, soprattutto se le fonti informative (ad esempio, i

[35] Fonte: Elaborazioni su dati Banca d'Italia e Cerved. Note: I modelli includono anche effetti fissi bancari. L'EBITDA si concentra sulla performance operativa di un'azienda misurando gli utili prima dell'impatto delle decisioni finanziarie e contabili. Fornisce informazioni sulla redditività operativa dell'azienda. Le variazioni dell'EBITDA riflettono le variazioni nella *performance* aziendale di base, come la crescita delle vendite, la gestione dei costi e l'efficienza operativa complessiva. Errori *standard* robusti raggruppati a livello aziendale. Bande di confidenza al livello del 95%.

dati su bilanci aziendali) sono disponibili con cadenza annuale mentre gli andamenti delle linee di credito possono essere mensili o trimestrali. Ciò presuppone uno sforzo di armonizzazione delle finestre temporali, in modo tale che ogni osservazione rifletta in modo coerente la combinazione "banca-impresa-tempo".

Infine, è rilevante sottolineare come questa metodologia permetta anche di indagare, in un secondo momento, le ricadute reali sulle imprese, andando a misurare l'effetto dell'adozione dell'IA (e della componente relazionale) sulla capacità dell'impresa di sostenere investimenti, occupazione o altre scelte di carattere strategico. In virtù della grande granularità offerta dagli effetti fissi su banca e impresa, gli autori possono spingersi a valutare come l'erogazione del credito, in diversi momenti del ciclo economico, si traduca in differenze di *performance* aziendale.

In conclusione, l'approccio microeconometrico prescelto da Gambacorta, Sabatini e Schiaffi (2025) non è solo una scelta "tecnica", bensì risulta indispensabile per conseguire l'obiettivo di isolare il ruolo dell'intelligenza artificiale e quello della relazione nel processo di *credit scoring* e nell'allocazione del credito. Il ricorso esteso ai *fixed effects* incrociati permette di distinguere, con elevato grado di credibilità, le decisioni di offerta riconducibili all'innovazione tecnologica e all'esperienza accumulata nei rapporti di lungo corso, da eventuali fattori di contesto (domanda di credito generale, condizioni settoriali, strategie "macro" della banca). In tal modo, la conclusione a cui gli autori giungono, e cioè che IA e *relationship lending* si configurano come strumenti complementari e sinergici in un *prescore* evoluto, trova un solido fondamento nella robustezza statistica e nella precisione esplicativa del modello.

2.4.3 Prospettive di sviluppo

L'analisi approfondita del *paper* "*Artificial intelligence and relationship lending*" di Gambacorta, Sabatini e Schiaffi (2025), dimostra che approcci relazionali e transazionali non si pongano in competizione, bensì in complementarità. Il vero salto qualitativo avviene quando:

- Il dato "*hard*" (transazionale) diventa lo scheletro quantitativo solido su cui basare stime puntuali di rischio;
- La relazione "*soft*" (di lungo periodo) agisce come fattore di conoscenza profonda del cliente e di stabilizzazione del rapporto, specie in fasi di crisi;

- L'intelligenza artificiale funge da catalizzatore, velocizzando l'elaborazione dei dati, migliorando la precisione dello *scoring* e favorendo l'aggiornamento continuo del giudizio.

In quest'ottica, il *prescore* dinamico costituisce l'architettura ideale: un sistema che, a intervalli regolari o sulla base di eventi specifici (e.g. deterioramento di un indice finanziario, cambio delle condizioni macro, informazioni dirette riportate dal gestore), ricalcola il profilo di rischio dell'impresa, integrando la base storica con nuovi *input*. Questa prospettiva "*always on*" eleva la funzione del *prescore* da semplice strumento statico a vero e proprio "modello vivo", sensibile al contesto in cui opera l'azienda, alle sue scelte strategiche e alla qualità del rapporto con l'istituto di credito.

Le implicazioni operative riguardano tanto le banche quanto le imprese:

- Per le banche, l'adozione di IA e la parallela cura della "relazione" (tramite un approccio consulenziale e un *network* di *account manager* dedicati) divengono *asset* fondamentali per costruire vantaggi competitivi e gestire il rischio in maniera proattiva.
- Per le imprese, risulta essenziale mantenere un dialogo trasparente e continuativo, fornendo dati attuali e affidabili, al fine di "nutrire" in maniera corretta i modelli di *scoring*. La fiducia da parte della banca potrebbe tradursi in condizioni migliorative, sia in termini di tassi sia di disponibilità di credito, soprattutto nei momenti di congiuntura sfavorevole.

Un ulteriore aspetto riguarda l'attenuarsi del confine tra "relazionale" e "transazionale". Gli studi evidenziano come molte informazioni "*soft*" (frequenza dei contatti in filiale, affidabilità storica, puntualità nei pagamenti, ecc.) possano essere codificate in variabili utilizzabili dai modelli IA. Di conseguenza, anche un'impresa priva di uno storico bancario di lungo periodo può essere valutata in modo più equo, grazie a fonti di dati alternative e aggiornate in tempo reale (contratti, fatturazione elettronica, *trend* di incassi). Ciò non solo migliora la capacità predittiva dei modelli di rischiosità, ma favorisce la maggiore inclusione di soggetti e imprese "non storici", riducendo gli sbalzi tipici del *relationship lending* in tempi di crisi. In prospettiva, la coesistenza di modelli dinamici basati su IA e di forme residue di approccio relazionale (specie per segmenti di clientela più ristretti) porterà a una fase di convergenza, con nuove opportunità di mercato per gli intermediari e una maggiore stabilità del sistema del credito.

Infine, sotto il profilo macroprudenziale, emerge l'opportunità di vigilare attentamente sulla qualità dei dati utilizzati e sulle metodologie di *machine learning* adottate, affinché gli algoritmi non generino discriminazioni o non sopravvalutino determinati parametri a scapito della stabilità di lungo periodo. La sfida per il futuro immediato è dunque affinare la sinergia tra "metodo relazionale" e "valutazione transazionale algoritmica", così che il sistema del credito possa godere di una maggiore solidità strutturale e, al contempo, di un'elevata capacità di inclusione finanziaria e sostegno all'economia reale.

L'evidenza portata dall'articolo della Banca d'Italia, considerata nel suo complesso, fornisce una base empirica e metodologica solida per sostenere che i modelli dinamici di *prescore*, costruiti su un'integrazione bilanciata di informazioni relazionali e transazionali, rappresentano la migliore frontiera per ottimizzare il processo di valutazione del merito creditizio. Questo genera benefici tangibili per gli stessi intermediari (minori perdite attese, maggiore efficienza allocativa), per le imprese (maggiore stabilità delle fonti di finanziamento, riduzione delle distorsioni legate a situazioni contingenti) e, non ultimo, per la stabilità complessiva del sistema finanziario.

2.5 Implicazioni dell'utilizzo dell'IA nel *credit scoring*

L'intelligenza artificiale (IA) ha profondamente modificato l'ambito del *credit scoring*, offrendo metodologie avanzate e analisi predittive caratterizzate da livelli superiori di precisione ed efficienza[36]. Con l'adozione sempre più diffusa di approcci incentrati sull'IA, le ricadute pratiche risultano considerevoli, poiché l'inclusione viene favorita e le strategie di gestione del rischio subiscono una vera e propria metamorfosi, influenzando insieme i processi decisionali e la fiducia dei fruitori. La presente sezione intende approfondire tali dinamiche, illustrando i molteplici effetti che la tecnologia esercita nel *credit scoring*.

L'IA applicata al *credit scoring* può sostenere l'inclusione finanziaria, estendendo i servizi di prestito a categorie di individui solitamente trascurate dai modelli convenzionali, che si basano su *set* informativi limitati ed escludono coloro con scarsa o nessuna storia creditizia. Grazie alla capacità di esaminare fonti di dati alternative, le tecniche d'intelligenza artificiale rendono più completa l'analisi, ampliando la platea di soggetti potenzialmente idonei al credito.

[36] Patel, 2023

In pratica, il ricorso all'IA nel punteggio di credito comporta l'inclusione di segmenti un tempo ignorati. Soggetti con percorsi lavorativi irregolari, un vissuto creditizio ridotto o appartenenti a comunità svantaggiate possono beneficiare di soluzioni in cui l'intelligenza artificiale prende in esame molteplici indicatori, plasmando un sistema finanziario più accogliente. Se ideati e condotti secondo principi etici, gli algoritmi di IA sono in grado di mitigare eventuali discriminazioni[37]. Privilegiando variabili oggettive e pertinenti, si mira infatti a valutazioni prive di distorsioni, fronteggiando possibili *bias* storicamente presenti nei sistemi tradizionali.

La capacità dell'IA di processare ampi volumi di informazioni in tempo reale incrementa sensibilmente l'accuratezza con cui viene stimato il rischio di credito. Attraverso l'individuazione di *pattern* e correlazioni sottili, gli algoritmi di apprendimento automatico forniscono una visuale più sofisticata della solvibilità di un richiedente[38]. Tale miglioramento consente agli intermediari finanziari di affinare le proprie strategie di *risk management* e di formulare scelte di concessione del credito più consapevoli. I vecchi sistemi di *scoring*, d'altronde, facevano ricorso a regole statiche, spesso inadatte alle oscillazioni dei mercati o alle peculiarità del singolo cliente. Al contrario, i modelli di IA assicurano una valutazione dinamica, reagendo e adeguandosi costantemente a scenari mutanti. In tal modo, gli enti finanziari riescono a rispondere in maniera tempestiva alle variazioni dell'economia e ai differenti profili debitori.

Le soluzioni di IA che incorporano tecniche di *machine learning* sono inoltre in grado di rilevare precocemente rischi e tendenze sfuggenti all'osservazione dei metodi classici, permettendo di adottare contromisure preventive e rafforzando la stabilità globale del portafoglio dei finanziamenti. Allo stesso tempo, un sistema di *scoring* basato sull'IA semplifica la fase decisionale, producendo valutazioni più celeri e puntuali. L'automazione di compiti ricorrenti, come l'elaborazione e l'analisi del rischio, velocizza l'iter di erogazione, garantendo alla clientela un accesso più rapido ai prestiti e migliorando l'esperienza complessiva.

Progettati all'insegna di trasparenza ed equità, i modelli di IA sostengono una maggiore oggettività e uniformità nelle determinazioni finali. L'adozione di dati e parametri verificabili riduce l'ingerenza di inclinazioni soggettive, limitando l'influenza dei

[37] Langenbucher, 2020
[38] Bhilare et al., 2024

pregiudizi umani nelle scelte di credito. Tale coerenza contribuisce a generare un senso di fiducia nei consumatori, i quali sperimentano un sistema percepito come più giusto e imparziale[39]. Ciononostante, alcuni algoritmi di IA, essendo non sempre interpretabili, sollevano interrogativi sulla spiegabilità del processo. I mutuatari potrebbero avvertire perplessità riguardo a decisioni adottate mediante tecniche di apprendimento automatico, senza che venga fornito un chiarimento sui parametri utilizzati. Per rafforzare la fiducia, gli istituti finanziari devono dunque privilegiare la spiegabilità, assicurandosi che le persone coinvolte possano comprendere con chiarezza i criteri su cui si fondano i giudizi di credito basati sull'intelligenza artificiale.

2.5.1 Sfide e limitazioni

L'adozione dell'Intelligenza Artificiale (IA) ha trasformato il panorama del punteggio di credito, introducendo modelli avanzati e metodologie predittive in grado di fornire risultati più accurati ed efficienti. Nonostante questi benefici, l'integrazione dell'IA nel *credit scoring* presenta una serie di criticità da non sottovalutare, riconducibili in particolare ai limiti di interpretabilità, alla tutela della *privacy*, alle vulnerabilità in materia di sicurezza e al delicato compromesso tra efficienza predittiva ed equità.

Un ostacolo rilevante risiede nell'opacità che caratterizza i cosiddetti modelli "*black-box*". Le soluzioni basate su algoritmi complessi, specie quelle fondate su tecniche di *deep learning*, offrono spesso prestazioni elevate in termini di accuratezza, ma restano poco trasparenti nelle motivazioni sottostanti le scelte decisionali[40]. Questa mancanza di chiarezza solleva interrogativi in merito alla correttezza, alla responsabilità sociale dell'ente erogante e al rischio di introdurre, anche in modo non intenzionale, forme di pregiudizio algoritmico. L'opacità dei sistemi di IA pone inoltre difficoltà per i destinatari, in quanto i mutuatari possono trovarsi disorientati di fronte a algoritmi impenetrabili, senza comprendere quali fattori influiscano sul proprio merito creditizio. L'assenza di trasparenza, in tal senso, può erodere la fiducia nella valutazione algoritmica e frenare l'accettazione diffusa di metodi di *credit scoring* basati sull'intelligenza artificiale, specie quando i soggetti coinvolti non riescono a decifrarne la logica, che incide sul loro benessere economico.

[39] Adeleke et al., 2019
[40] Kim et al., 2020

Per attenuare il problema della scarsa interpretabilità occorre investire nella progettazione di modelli esplicabili. L'implementazione di metodologie che consentano di mostrare in modo chiaro come l'IA arrivi a un dato verdetto creditizio favorisce la fiducia degli utenti. In tal senso, è utile affiancare la trasparenza alla precisione predittiva, affinché i mutuatari possano identificare i principali fattori che concorrono al calcolo del proprio merito creditizio[41]. L'area del *credit scoring*, d'altro canto, richiede l'analisi di dati personali e finanziari spesso altamente sensibili. L'impiego di algoritmi di IA, che necessitano di ingenti quantità di informazioni per generare previsioni affidabili, comporta dunque notevoli responsabilità in termini di protezione dei dati. Risulta perciò essenziale adottare solide misure di sicurezza, scongiurando accessi non autorizzati e violazioni, nonché garantire un utilizzo etico e trasparente di informazioni potenzialmente riservate. L'attenzione crescente alle disposizioni normative in tema di trattamento dei dati personali, come il GDPR e ulteriori vincoli regionali, si traduce in un quadro complesso per la gestione dell'IA nel *credit scoring*. Gli istituti di credito, in particolare, devono attenersi a *standard* rigorosi di *governance* dei dati, comunicare in modo chiaro le finalità del trattamento e ottenere un consenso esplicito, prima di ricorrere a tali dati per la valutazione della solvibilità.

Un'ulteriore criticità costante riguarda la possibilità che emergano distorsioni algoritmiche. Se i *set* di dati storici alla base del *training* di un sistema di IA manifestano *bias* pregressi, è plausibile che i modelli finiscano per perpetuarli o perfino accentuarli, con conseguenze discriminatorie nelle decisioni di credito. Per assicurare un livello adeguato di equità, è indispensabile adottare procedure di monitoraggio regolare, strumenti di rilevamento delle distorsioni e meccanismi correttivi delle eventuali deviazioni che si palesino durante l'intero ciclo di vita del modello. Il raggiungimento di un equilibrio tra *performance* predittiva ed equità si rivela spesso problematico. I tradizionali sistemi di *scoring*, ancorati a dati storici, possono inglobare *bias* in modo inconsapevole, generando discriminazioni tra fasce di utenza diverse. Occorre dunque prestare la massima attenzione agli indicatori di equità in sede di progettazione dei modelli e svolgere verifiche costanti per far emergere e rettificare eventuali disomogeneità valutative.

[41] Vincent et al., 2021

Da un lato, privilegiare l'accuratezza può tradursi in un calo della giustizia distributiva, mentre dall'altro, perseguire la massima equità potrebbe compromettere la capacità del modello di segnalare con tempestività i rischi di insolvenza. Di conseguenza, le istituzioni finanziarie devono definire linee guida e presupposti etici stringenti a supporto della creazione e dell'implementazione dei sistemi di IA dedicati al *credit scoring*. Ai fini dell'equità, è cruciale integrare metodiche di intelligenza artificiale spiegabile. Garantire la trasparenza del processo decisionale all'interno del modello consente di identificare e correggere tempestivamente schemi di pregiudizio, sostenendo la responsabilizzazione degli operatori. Infine, una spiegabilità adeguata offre agli *stakeholder* la possibilità di comprendere e contestare le conclusioni algoritmiche, promuovendo un meccanismo di punteggio creditizio più inclusivo e corretto[42].

[42] Percy et al., 2021

Capitolo 3: Analisi del *prescore* di Fin.promo.ter

Fin.Promo.Ter. Scpa è un intermediario finanziario vigilato dalla Banca d'Italia. Confcommercio lo ha costituito nel 1999 con l'obiettivo di sviluppare le attività imprenditoriali nel commercio, nel turismo e nei servizi. È stato istituito con la riforma Bersani (D.lgs. 114/98) come supporto alle piccole e medie imprese (PMI) nell'accesso al credito attraverso garanzie e finanziamenti agevolati.

Dal 18 maggio 2016, Fin.promo.ter è stata ufficialmente iscritta nel Registro Unico degli Intermediari Finanziari ai sensi dell'articolo 106 del Testo Unico Bancario (TUB) come *partner* affidabile e dotato di solido patrimonio per le imprese e per le istituzioni bancarie aderenti. Attraverso la sua diffusa presenza territoriale e in collaborazione con il sistema Confidi, può offrire garanzie dirette, finanziamenti con condizioni favorevoli e controgaranzie, rendendo quindi agevole un accesso al credito per le imprese in modo trasparente ed efficiente. Fin.promo.ter si caratterizza per un modello di *business* innovativo e sinergico, fondato sulla *partnership* con primari istituti finanziari, tra cui il Fondo di Garanzia per le PMI del Ministero dello Sviluppo Economico e il Medio Credito Centrale (MCC). L' intermediario offre supporto concreto alle imprese attraverso:

- Garanzie dirette sugli affidamenti concessi dalle banche *partner*, migliorando le condizioni di accesso al credito;
- Credito diretto per piccole imprese e *startup*, con procedure snelle e tassi agevolati;
- Controgaranzie a favore dei Confidi soci, rafforzando il sostegno alle PMI operanti nei settori chiave dell'economia.

3.1 Struttura e metodologia del *prescore* esistente

Nel settore della concessione del credito, la capacità di selezionare in modo rapido ed efficace le richieste di finanziamento rappresenta uno dei fattori chiave per garantire che la gestione delle risorse aziendali sia efficiente e per mantenere sotto controllo il rischio di insolvenza. Fin.promo.ter, in qualità di intermediario finanziario vigilato, ha introdotto un sistema di *prescoring* per migliorare la fase iniziale del processo di valutazione, riducendo il numero di pratiche da sottoporre a valutazione approfondita e concentrando l'attenzione degli analisti su quelle con maggiore probabilità di approvazione.

A differenza di un sistema di *rating* tradizionale, che attribuisce un punteggio alla solvibilità di un'impresa sulla base di modelli probabilistici avanzati, il *prescore* di Fin.promo.ter è uno strumento deterministico[43], volto ad effettuare un primo filtro sulle richieste di finanziamento, scartando automaticamente quelle che non rispettano gli *standard* minimi stabiliti per rispettare i criteri di rischio dell'istituto. Il *prescore*, dunque, non esprime un giudizio per ciò che attiene alla solidità creditizia del richiedente, ma ha la funzione di selezionare le pratiche che possono proseguire nel processo di istruttoria, evitando di allocare risorse su domande che presentano anomalie evidenti o carenze documentali.

L'adozione del *prescore* consente di ottenere molteplici benefici operativi. In primo luogo, il sistema permette di ridurre il volume delle pratiche che necessitano di valutazione manuale, diminuendo i tempi di lavorazione e migliorando l'efficienza del processo decisionale. In secondo luogo, esso garantisce un approccio più rigoroso e sistematico nella selezione delle richieste, contribuendo a migliorare la qualità complessiva del portafoglio crediti. Infine, attraverso l'automazione della fase di pre-analisi, Fin.promo.ter è in grado di canalizzare le richieste idonee verso i prodotti più adatti alle loro caratteristiche, ottimizzando così l'esperienza del cliente e favorendo una maggiore coerenza nelle decisioni di concessione del credito.

Di seguito viene riportato il diagramma di flusso rappresentativo del *prescore* di Fin.promo.ter. I componenti di quest'ultimo, verranno approfonditi successivamente nel corso del capitolo.

[43] Per una panoramica storica e concettuale dei modelli di *scoring* si veda Altman (1968); Crook, Edelman & Thomas (2007).

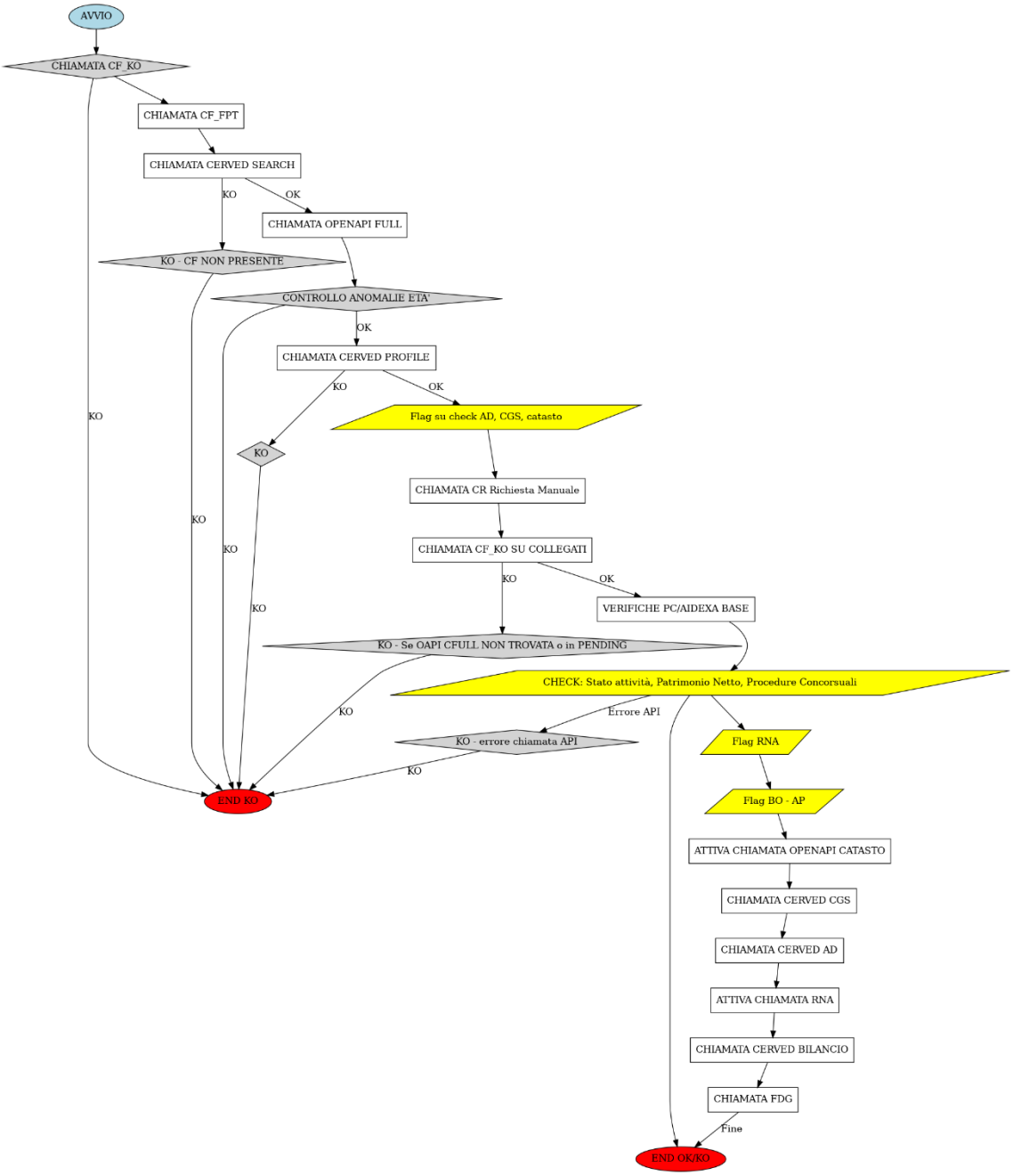


Figura 4 Diagramma di flusso *prescore*

3.1.1. Metodologia di calcolo e variabili analizzate

L'algoritmo che determina il *prescore* di Fin.promo.ter ha le sue fondamenta su di una combinazione di regole di esclusione e pesi ponderati assegnati a una serie di variabili chiave. Il modello utilizzato opera secondo una logica deterministica, in cui ogni richiesta viene analizzata sulla base di parametri oggettivi e prestabiliti.

Dal punto di vista matematico, il valore del *prescore* è calcolato attraverso la seguente equazione:

$$P = w_1X_1 + w_2X_2 + w_3X_3 + \cdots + w_nX_n$$

Dove:

- P rappresenta il punteggio del *prescore* assegnato alla domanda di finanziamento.
- X_1, X_2, …, X_n sono le variabili prese in considerazione nel modello, tra le quali dati finanziari, patrimoniali, storici e settoriali.
- W_1, W_2, …, W_n sono i pesi assegnati a ciascuna variabile sulla base della loro significatività nel determinare la probabilità di idoneità della pratica

Il valore ottenuto viene poi confrontato con delle **soglie predefinite**, che determinano se la pratica può essere accettata o meno. Determinati fattori, quali l'esistenza di procedure concorsuali o la presenza di patrimonio netto negativo, impediscono automaticamente la prosecuzione dell'elaborazione della richiesta, mentre altri, possono generare un ***alert*** che richiede un esame più attento da parte di un operatore.

Uno degli aspetti distintivi del *prescore* di Fin.promo.ter è la sua capacità di adattarsi alle esigenze dell'istituto, consentendo la modifica dei pesi e delle soglie in base all'evoluzione del contesto economico e alle strategie di rischio adottate dalla società.

L'accuratezza del *prescore* è influenzata dalla qualità e completezza dei dati analizzati. Il modello impiegato da Fin.promo.ter si basa su un ampio *set* di variabili, classificabili nelle seguenti categorie principali:

1. Dati anagrafici e settoriali
    - Codice fiscale e Partita IVA: verifica della correttezza e validità dell'identificativo fiscale dell'azienda.
    - Codice ATECO: categorizzazione dell'attività economica e valutazione del rischio relative al settore di appartenenza della società richiedente.
    - Data di costituzione e inizio attività: parametro essenziale per determinare l'anzianità aziendale.
2. Dati finanziari e patrimoniali
    - Bilanci aziendali: analisi di parametri fondamentali come Margine Operativo Lordo (MOL), patrimonio netto e livello di indebitamento.

- *Cerved Group Score* (CGS): indicatore sintetico della probabilità di *default*, in una scala da 1 a 100, acquisito da *Cerved*.

  Le principali informazioni elementari considerate nel modello per il calcolo dello *score* sono:

  a) Bilanci aziendali (intesi anche come fonte di informazioni sulle attività e sulle strategie aziendali)
  b) Informazioni anagrafiche dell'impresa (età, settore, area geografica, ecc.)
  c) Informazioni catastali
  d) Dati sul patrimonio immobiliare dell'impresa e sui principali amministratori/soci
  e) Eventi negativi sull'impresa e sui soggetti ad essa connessi (protesti, pregiudizievoli, CIGS, …)
  f) Data base proprietario sulle transazioni commerciali (*Payline*)
  g) Informazioni qualitative provenienti dal contatto diretto con l'impresa o i suoi *partner* commerciali (es. stato di attività, andamento dei rapporti, ecc.)
  h) Scenari di settore storici e previsionali
  i) Rassegna stampa nazionale e locale

- Rapporto debito/capitale: misura della sostenibilità finanziaria dell'azienda.

3. Dati creditizi e bancari
   - Segnalazioni nella Centrale dei Rischi (CR): verifica della presenza di sofferenze bancarie o sconfini.
   - Storicità bancaria: analisi della relazione che intercorre tra il soggetto richiedente e il sistema creditizio.
   - *Anomaly Detection Score*: indicatore che rileva la presenza di segnali generalmente associati a frode, acquisito da *Cerved*.
4. Dati patrimoniali
   - Presenza di immobili: il possesso di *asset* immobiliari può fungere da garanzia implicita nella valutazione della solidità dell'azienda.

- Verifica catastale: controllo della titolarità dei beni dichiarati tramite *OpenAPI*.

L'analisi congiunta di queste variabili[44] consente al sistema di effettuare una prima valutazione, individuando eventuali criticità e garantendo un processo decisionale più efficiente.

3.1.2 *Output* del *prescore* e interpretazione dei risultati

L'*output* del *prescore* di Fin.promo.ter rappresenta il punto finale del processo di valutazione preliminare delle richieste di finanziamento. Dopo aver raccolto ed elaborato i dati provenienti dalle diverse fonti informative, il sistema restituisce un esito sintetico, che determina se la pratica può passare alla fase successiva dell'istruttoria creditizia o se, al contrario, deve essere respinta automaticamente.

L'esito del *prescore* può rientrare in una delle seguenti tre categorie principali.

***PRESCORE* OK**

Questo esito viene assegnato alle richieste che soddisfano tutti i requisiti minimi previsti dai criteri di valutazione di Fin.promo.ter. Ciò implica che la pratica non presenta anomalie significative e che il soggetto richiedente possiede le caratteristiche necessarie per accedere ai prodotti finanziari offerti dall'istituto.

Un esito positivo del *prescore* non implica necessariamente l'approvazione del finanziamento, ma consente alla pratica di accedere alla fase successiva dell'istruttoria, nella quale verranno effettuate ulteriori analisi qualitative e quantitative da parte degli analisti. Nella fase di istruttoria approfondita, vengono esaminati in dettaglio aspetti come la sostenibilità del finanziamento, la documentazione a supporto della richiesta e l'eventuale presenza di garanzie supplementari.

Inoltre, le pratiche con *prescore* OK vengono classificate in diverse fasce di priorità, a seconda del livello di affidabilità rilevato dal sistema. Ad esempio, un'impresa con un *Cerved Group Score* (CGS) particolarmente elevato e un livello di indebitamento contenuto potrà essere trattata con maggiore priorità rispetto a un'azienda che, pur superando il filtro del *prescore*, presenta indicatori finanziari più deboli.

***PRESCORE* KO**

[44] La letteratura suggerisce che una buona selezione delle variabili deve massimizzare la capacità discriminante e ridurre l'*overfitting* del modello (Thomas et al., 2002).

L’esito KO (*Knock Out*) viene assegnato alle richieste che non soddisfano i criteri minimi di ammissibilità. Il sistema segnala una o più anomalie gravi che impediscono alla pratica di proseguire, come ad esempio:

- Patrimonio netto negativo, segnale di potenziale instabilità finanziaria dell’azienda.
- Presenza di procedure concorsuali in corso, come fallimenti, liquidazioni, amministrazioni straordinarie o concordati preventivi.
- Assenza di bilanci aggiornati, nel caso di società di capitali.
- Stato attività non ammissibile, ad esempio se l’azienda risulta cessata, sospesa o in fase di scioglimento.
- Rapporto tra debiti e capitale troppo elevato, segnale di un rischio di sovraindebitamento.

Se una richiesta registra un esito sfavorevole, viene esclusa automaticamente dal processo di valutazione, senza possibilità di proseguire nella fase di istruttoria. Tuttavia, l’azienda richiedente ha la facoltà di ripresentare una nuova richiesta in futuro, qualora la sua situazione finanziaria e patrimoniale dovesse migliorare.

Le pratiche che ricevono un KO automatico vengono registrate all’interno del sistema di Fin.promo.ter, con l’obiettivo di evitare duplicazioni e garantire la tracciabilità delle decisioni di esclusione. Inoltre, in alcune circostanze, i clienti possono essere informati sulle motivazioni del rifiuto, in modo da comprendere quali aspetti devono essere migliorati per poter accedere a un finanziamento in futuro.

***PRESCORE* KO con possibilità di deroga**

In determinati casi, la richiesta può presentare delle criticità, ma non in modo così grave da impedirne automaticamente l’accettazione. Quando questo accade, il sistema assegna un esito KO con possibilità di deroga, riservando la decisione finale agli operatori di Fin.promo.ter.

Le richieste con KO in deroga possono essere sottoposte a una valutazione manuale da parte di un analista, il quale è incaricato di accertare se le anomalie segnalate dal *prescore* possano essere compensate da altri elementi positivi. Ad esempio:

- Un’impresa con un bilancio in perdita potrebbe comunque essere ammessa se dispone di garanzie reali che riducono il rischio di insolvenza.

- Un'azienda con un rapporto debito/capitale elevato potrebbe ottenere una valutazione positiva se mostra una crescita significativa del fatturato negli ultimi anni.
- Un'impresa con un CGS inferiore alla soglia di accettabilità potrebbe comunque essere presa in considerazione se ha una storia creditizia positiva e nessuna segnalazione in Centrale dei Rischi.

In questi casi, l'analista ha la possibilità di richiedere ulteriore documentazione al richiedente (come piani industriali, dichiarazioni fiscali o relazioni dettagliate sul *business*) e di esaminare il contesto specifico dell'azienda prima di emettere una decisione definitiva.

L'esito del *prescore* viene registrato e visualizzato all'interno della piattaforma gestionale di Fin.promo.ter, Airtable, dove gli operatori possono monitorare lo stato di avanzamento delle richieste e intervenire in caso di necessità. Il sistema è progettato per garantire la massima trasparenza e tracciabilità, permettendo di risalire a tutte le informazioni che hanno portato alla decisione finale.

Nello specifico, la piattaforma consente di:

- Visualizzare le specifiche delle analisi effettuate, incluse le variabili chiave che hanno determinato l'esito della pratica.
- Applicare filtri per la gestione delle pratiche, categorizzando le richieste in base allo stato dell'istruttoria (es. "*PRESCORE OK* – In attesa di istruttoria", "*PRESCORE KO* – Non finanziabile", "*PRESCORE KO* con deroga – In revisione manuale").
- Predisporre *report* analitici per monitorare le statistiche di approvazione e identificare eventuali *trend* nel tempo.

Un ulteriore vantaggio della piattaforma di gestione è la possibilità di condurre un'analisi retrospettiva sulle pratiche respinte. Questo consente agli analisti di Fin.promo.ter di individuare eventuali aree di miglioramento nel modello di *prescoring*, verificando se ci siano stati casi in cui il sistema ha erroneamente scartato aziende potenzialmente finanziabili.

3.2 Mappatura delle regole di calcolo e del processo

Il *prescore* di Fin.promo.ter è progettato per offrire una prima valutazione, rapida ma rigorosa, delle richieste di credito. Utilizzando un insieme di soglie e *test* su parametri fondamentali (anagrafici, bilancistici, indicatori esterni e *policy* settoriali), esso consente di:

- Bloccare le richieste appartenenti ad imprese che non soddisfano i requisiti minimi;
- Ridurre i costi di analisi, evitando di proseguire con verifiche superflue;
- Focalizzare l'attenzione sulle richieste che meritano un approfondimento (segnate come *ALERT*) o che, allo stato attuale, appaiono pienamente idonee (*OK*).

Le sezioni che seguono illustrano perché questo sistema è strutturato in un determinato modo, con particolare attenzione alle scelte delle soglie, e come si snodano le varie fasi (dall'acquisizione dei dati alla classificazione finale), tenendo in considerazione i ruoli delle diverse funzioni aziendali coinvolte.

La logica del *prescore* ha origine da una duplice esigenza: da un lato, accelerare il processo iniziale di valutazione del rischio (per non sovraccaricare l'ufficio Crediti di pratiche chiaramente non idonee), dall'altro garantire che decisioni complesse come l'erogazione di finanziamenti siano prese sulla base di metriche omogenee, evitando eccessiva discrezionalità.

Il sistema, dunque, opera come un filtro: se emergono violazioni di parametri ritenuti inderogabili (ad esempio, la rilevazione di un codice ATECO esplicitamente escluso dalle *policy*), la richiesta viene immediatamente respinta (KO). Grazie a questo meccanismo a soglia, si evita di coinvolgere altri servizi (*Cerved*, *OpenAPI*) per un caso già destinato all'esclusione, e si minimizzano i costi correlati (ad esempio, l'addebito per la chiamata API del *rating* esterno).

Il processo può essere visto come una catena di blocchi (*step*) successivi, in cui si analizzano variabili $\{x_1, x_2, \dots, x_n\}$ e si confrontano con soglie $\{t_i\}$. Se $x_i \leq t_i$ (o, a seconda del caso, $x_i \geq t_i$), si genera KO immediato e il flusso si interrompe.

$$D = \begin{cases} KO \quad se\ \exists i: x_i \leq t_i\ (vincolo\ inderogabile), \\ ALERT \quad se\ t_j < x_j > t'_j\ (valore\ borderline) \\ OK \quad altrimenti \end{cases}$$

Per rendere questo sistema più intuitivo, lo si paragona spesso a un semaforo:

- Semaforo Rosso (KO)
  Significa che la pratica ha violato almeno un parametro fondamentale e perciò non può procedere. Dopo il primo KO, non vengono più fatti ulteriori controlli.
- Semaforo Giallo (*ALERT*)
  Si utilizza quando una variabile non è drasticamente fuori soglia, ma è prossima al limite o non è disponibile in modo completo. Un tipico esempio è la *ratio* Debiti/EBITDA leggermente sopra la soglia *standard*, ma compensata da altri fattori.
- Semaforo Verde (OK)
  Quando i valori superano tutti i requisiti predefiniti, la richiesta passa alla fase successiva senza necessità di ulteriori approfondimenti.

Il concetto di *stop* anticipato è fondamentale per comprendere il funzionamento della catena di blocchi. Una volta identificata una violazione inderogabile, il *prescore* cessa ogni ulteriore interrogazione, evitando di richiedere dati o costose API (Cerved Bilanci, *OpenAPI* Catasto) relativi a una posizione già "perduta", oppure di sostenere il carico computazionale e di tempo derivante dalla consultazione di più fonti informatiche o dalla preparazione di *report* finanziari aggiuntivi.

In pratica, la catena di blocchi è ordinata in modo che i *test* meno costosi, come la verifica della tabella CF_KO o la lettura di alcuni dati anagrafici, vengano svolti per primi. Se uno di questi risulta KO, si evita di lanciare i successivi *step* più dispendiosi, come *rating* CGS e analisi di bilancio dettagliato.

3.2.1 Definizione delle regole di calcolo

Le variabili utilizzate nel *prescore* emergono sia dalla letteratura finanziaria, che dimostra come indici quali Debiti/EBITDA e ROE siano buoni predittori di *default*, sia dall'esperienza diretta di Fin.promo.ter nell'analisi del rischio. Alcuni dati dimostrano, ad esempio, che imprese con un MOL (Margine Operativo Lordo) negativo presentano una probabilità di *default* significativamente più elevata rispetto a quelle con MOL positivo. Allo stesso modo, un livello di indebitamento elevato (Debiti/EBITDA) è spesso associato a tensioni di liquidità.

Ogni prodotto (Piccolo Credito, Plus, Aidexa, ecc.) deve rispondere a criteri di ammissibilità specifici. Questa differenziazione trova riscontro nella selezione dei parametri:

- Politica di esclusione settoriale. Se la *policy* di Fin.promo.ter stabilisce di non operare in determinati codici ATECO considerati ad alto rischio (ad esempio, "fabbricazione di armi e munizioni", "stabilimenti balneari", "servizi finanziari"), diventa imprescindibile includere la variabile "ATECO" tra le regole del *prescore*.
- Richiesta di bilanci trasparenti. Se un'impresa richiede un prodotto "*Plus*" o "*Small Corporate*", è necessario fornire dati di bilancio più completi, inclusa la distinzione tra Debiti a breve termine (BT) e Debiti a medio-lungo termine (MLT), per calcolare correttamente ed accuratamente gli indici Debiti/EBITDA o PFN/EBITDA.

Le variabili scelte devono essere facilmente reperibili presso i *database* integrati (Cerved, *OpenAPI*, MCC, RNA) o ricavabili dai bilanci. Per questo motivo, parametri molto complessi ma raramente disponibili (es. analisi di flussi di tesoreria su base mensile o *rating* di carattere qualitativo) non vengono inclusi nella fase preliminare del *prescore*, poiché rallenterebbero il processo e comporterebbero costi eccessivi rispetto al valore informativo che potrebbero fornire.

La selezione si concentra sulle principali aree di rischio:

- Struttura patrimoniale (Patrimonio Netto, Capitalizzazione)
- Solvibilità (Debiti/EBITDA, Sostenibilità della rata)
- Rischio esterno (*Cerved Group Score*, *Anomaly Detection*)
- Rischio settoriale (ATECO)
- Rischio anagrafico (età amministratore, anzianità impresa)

Ciascuna contribuisce a dipingere un quadro più chiaro del rischio, permettendo di avere una visione più completa. Ad esempio, un'azienda con buona capitalizzazione ma un punteggio CGS molto basso potrebbe nascondere anomalie di altra natura, che si rivelano attraverso il *rating* esterno.

Nel *prescore*, ciascuna regola si traduce in una formula (rapporto, soglia numerica, confronto booleano) che, se violata in modo non derogabile, porta a *KO* immediato, come precisato nella sezione sulla Struttura Generale: soglie e *stop* anticipato. Di seguito, le formule più comunemente impiegate.

Le verifiche effettuate dal *prescore* si articolano su due livelli:

- Regole generali: comprendono i controlli su *blacklist* (Tabella CF-KO), stato attività (cessata, fallita, ecc.), *rating* interni e parametri essenziali (p.es. PN negativo, attività avviata da meno di 3 anni se non ammesso da *policy*). Tali criteri si applicano indipendentemente dal tipo di prodotto richiesto.
- Requisiti di prodotto: definiscono soglie aggiuntive, più rigorose, a seconda della categoria di finanziamento (Piccolo Credito, Plus, Aidexa, *Small Corporate*, ecc.). Ad esempio, un CGS < 30 può rappresentare un *KO* per il "Piccolo Credito" ma un valore-soglia differente può essere fissato per la linea "*Plus*" o "*Aidexa* XG".

Questa distinzione è necessaria per modulare la severità del *prescore* in base al profilo di rischio e alla struttura del prodotto, mantenendo tuttavia un insieme di controlli unificati e inderogabili in caso di anomalie conclamate.

3.2.1.1 Indicatori bloccanti di bilancio

I quattro indicatori sottostanti sono ritenuti cruciali e non negoziabili per la sostenibilità finanziaria di un'impresa. Se uno di essi presenta valori oltre la soglia, viene considerato indice di rischio troppo elevato per continuare l'esame.

$$Capitalizzazione = \frac{Patrimonio\ netto}{Totale\ Attivo}$$

- Se $PN < 0$, si genera KO senza ulteriori interrogazioni.
- Se $\text{Cap} \leq 5\%$ la posizione è KO. La *ratio* cattura quanto l'impresa sia "sostenuta" dai propri mezzi, evitando eccessiva dipendenza dal credito esterno.

$$\frac{Debiti\ verso\ banche}{EBITDA}$$

- OK se < 5x
- KO se ≥ 5x

$$\text{Ebitda MARGIN} = \frac{\text{EBITDA}}{\text{Fatturato}}$$

- OK se > 3%
- KO se ≤ 3%

$$\frac{\text{Oneri finanziari}}{\text{EBITDA}}$$

- OK se < 35%
- KO se ≥ 35%

$$Sostenibilità\ rata = \frac{MOL}{\left(\frac{Debili\ MLT + Importo\ finanaziamento}{d}\right) + Oneri\ finaziari}$$

Dove d è la durata media (in anni) del finanziamento. Se la formula restituisce un valore < 1, si presume che il reddito operativo annuo non basti per coprire la quota capitale annuale e gli oneri finanziari, generando di regola un KO.

3.2.1.2 Indicatori "*Alert*" di bilancio

Le soglie aggiuntive di seguito riportate servono a intercettare casi *borderline*, dove l'impresa potrebbe non essere del tutto compromessa (da cui la classificazione "*ALERT*"), ma evidenzia una situazione di debito potenzialmente preoccupante. Per il Piccolo Credito, però, la tolleranza risulta minore, ragion per cui tali indicatori diventano a tutti gli effetti bloccanti e non soggetti a eventuali deroghe.

$$Livello\ di\ indebitamento = \frac{Debiti\ MLT}{EBITDA} > 6$$

$$\frac{Debiti\ Tributari}{Fatturato} \cdot 100\% > 60\%$$

$$\frac{Totale\ Debiti}{Fatturato} \cdot 100\% > 100\%$$

3.2.1.3 Indicatori esterni

- *Cerved Group Score* (CGS): restituisce una valutazione sintetica del rischio di *default*. Soglia tipica: CGS < 30 → KO per "Piccolo Credito".
- Anomaly Detection (AD): se AD > 4, molte linee considerano la posizione troppo rischiosa, con KO immediato. Questo perché un AD elevato segnala possibili profili di frode o incoerenze sostanziali nei dati aziendali.

3.2.1.4 ATECO vietati

Più che una formula, in questo caso si effettua un controllo di appartenenza: se il codice ATECO (primario o secondario) dell'azienda è compreso in una lista interna di esclusioni (es. 25.4 - Armi, 64-65-66 - Servizi finanziari, etc.), l'esito è KO.

3.2.2 Fasi del processo di elaborazione

Fase 1: Raccolta e *preprocessing*

La prima fase del *prescore* è dedicata all'acquisizione dei dati interni e alla consultazione di alcune fonti esterne. Da un lato, l'algoritmo passa in rassegna liste interne come CF_KO (anagrafiche con eventi particolarmente negativi o già scartate in passato) e CF_FPT (clienti in portafoglio). Dall'altro, ha la possibilità di accedere a banche dati esterne (*Cerved*, *OpenAPI*, Registro Nazionale Aiuti, Fondo di Garanzia), laddove ciò sia necessario per calcolare gli indici previsti.

Un aspetto cruciale di questa fase è la verifica immediata di eventuali motivi di esclusione che richiedono poco sforzo di controllo. Ad esempio, se l'azienda figura in CF_KO (per pregressi insoluti o segnalazioni negative consolidate), la logica del *prescore* stabilisce un KO senza necessità di avviare analisi aggiuntive come il calcolo delle *ratio* di bilancio. Se la visura anagrafica mostra che l'impresa è cessata o ha uno "stato di attività" non compatibile (come una procedura fallimentare in corso), si registra immediatamente un KO, risparmiando costi per eventuali chiamate ai servizi di *rating* (CGS, AD *Score*, etc.). L'obiettivo principale, in questa sede, è evitare di compiere passaggi più complessi (come la valutazione dei bilanci o l'estrazione di dati catastali) quando la pratica si trova già in una condizione di evidente inaccettabilità. Siffatto approccio, che è stato definito *stop* anticipato, assicura una gestione "a piramide": i controlli basilari e meno costosi vengono eseguiti prima e, se non superati, l'analisi non prosegue. In questo modo, si ottimizzano tempi, risorse e costi legati all'acquisizione di informazioni più complesse.

Fase 2: Applicazione delle regole di calcolo

Una volta passati i controlli preliminari della fase di raccolta, il sistema inizia a elaborare gli indicatori e a confrontarli con le soglie stabilite.

Per quanto riguarda il calcolo delle variabili, vengono determinate le *ratio* fondamentali (ad esempio Debiti/EBITDA, oneri finanziari/EBITDA, capitalizzazione, ecc.) e si recuperano i punteggi esterni come CGS e AD *Score*). Se un bilancio non fornisce sufficiente dettaglio, ad esempio se non effettua la distinzione tra debiti a medio e lungo termine e debiti a breve termine, il sistema potrebbe etichettare alcune *ratio* come "N.A." (non calcolabili), generando un *ALERT* da discutere con l'analista del credito.

Passando alla verifica delle soglie, per ogni variabile si controlla se il valore rilevato supera o è inferiore alla soglia corrispondente. In questa fase, si fa una distinzione tra regole "inderogabili" (che producono KO immediato) e regole che generano un *ALERT*. Infine, al termine del confronto tra indicatori e soglie, si attribuisce a ciascuna regola un segnale (ROSSO/KO, GIALLO/*ALERT*, VERDE/OK). La presenza anche di un singolo semaforo rosso in ambito "inderogabile", può essere sufficiente per bloccare la pratica. Se al contrario non emergono semafori rossi, ma qualcuno giallo, la posizione va esaminata in modo più approfondito.

Questo passaggio del processo è il cuore della valutazione: qui il *prescore* "traduce" i dati numerici in un giudizio sintetico, basato sulle politiche di credito e sulle formule discusse in precedenza.

Fase 3: Classificazione del merito creditizio

Al termine del calcolo, il sistema assegna alla pratica uno dei tre esiti globali:

1. KO: se almeno un parametro inderogabile è violato, oppure se uno dei parametri "*ALERT*" supera il limite in un prodotto che non ammette tolleranza, come nel caso del "Piccolo Credito". In questa circostanza, il processo si ferma qui e la richiesta non prosegue per ulteriori istruttorie.
2. *ALERT*: se nessuna regola inderogabile è violata, ma uno o più indicatori *borderline* suggeriscono di procedere con cautela. Anche se la pratica non viene bloccata automaticamente, richiede l'intervento dell'analista per decidere se sia necessario richiedere ulteriori documenti, attivare un garante esterno o concedere una deroga.
3. OK: se tutte le regole risultano superate con ampio margine, senza anomalie o aree grigie. In tal caso, la pratica può essere inviata direttamente all'ufficio che gestisce l'istruttoria formale, riducendo notevolmente i tempi di lavorazione.

Questa classificazione, a differenza della fase 2 che si focalizza su singoli indicatori, fornisce un giudizio finale sul merito creditizio preliminare, tenendo in considerazione la priorità delle regole e l'eventuale cumulazione di semafori gialli.

Fase 4: Validazione e *output*

Conclusa la classificazione, le informazioni vengono archiviate su *Airtable* o un sistema analogo. Questo processo di salvataggio comprende:

- Esito globale (*KO/ALERT/OK*)

- Indicatori salienti (Debiti/EBITDA, *rating* CGS, AD *Score*, data di bilancio)
- *Link* ai documenti (bilanci scaricabili, visure, catasto)
- Note: vi sono spesso campi per "Note Esterne" (condivise con segnalatori esterni) e "Note Interne" (accessibili solo allo *staff* di Fin.promo.ter).

Se l'esito è OK, la pratica avanza verso l'istruttoria, dove un analista approfondirà la posizione verificando documentazione aggiuntiva (Centrale Rischi, contratti, etc.) e preparerà la delibera creditizia.

In caso di *ALERT*, l'ufficio *prescore* o i *credit analyst* contattano il richiedente per richiedere eventuali integrazioni (ad esempio, un bilancio aggiornato, la prova di un garante, ulteriori dettagli su come abbattere la *ratio* Debiti/Fatturato). Solo dopo aver ricevuto un riscontro positivo si potrà trasformare l'*ALERT* in un OK.

Se l'esito è KO, la pratica viene chiusa. Tuttavia, l'operatore può inserire le motivazioni (es. "Patrimonio Netto < 0" oppure "codice ATECO escluso") in un campo note visibile a mediatori o Confidi, in modo da fornire un riscontro immediato sul motivo per cui la richiesta è stata rifiutata.

Nel processo di elaborazione, il *prescore* effettua automaticamente una serie di interrogazioni verso *provider* esterni e banche dati ufficiali per raccogliere informazioni aggiuntive sulla controparte. Più nello specifico:

- *Cerved*: consente di recuperare dati anagrafici, bilanci depositati in XBRL e punteggi di rischio (es. *Group Score*, *Anomaly Detection*).
- *OpenAPI* Catasto: fornisce informazioni sulla consistenza immobiliare della controparte (o dei soci/amministratori rilevanti). Se questa informazione è necessaria per il prodotto, può influenzare il punteggio o i criteri di KO (ad es. "assenza totale di immobili" può innescare un *alert*).
- MCC (Fondo di Garanzia): verifica l'ammissibilità o il *plafond* residuo. Nel caso di esito non ammissibile o *plafond* esaurito, si procede con KO immediato. Questa integrazione con MCC è particolarmente rilevante nelle linee di fido che richiedono copertura pubblica.
- Registro Nazionale degli Aiuti (RNA): fornisce le "visure *de minimis*" e le evidenze su eventuali aiuti di Stato già ricevuti dalla controparte.
- Tabella CF-KO/CF-FPT: gestita internamente per segnalare soggetti in *blacklist* (CF-KO) o già presenti in portafoglio (CF-FPT).

Ciascuna chiamata si attiva in una fase specifica del flusso (come illustrato nella Fig./Diagramma X) e, in caso di esito negativo (KO) o di errore, il processo si interrompe e il punteggio non viene calcolato. In questo modo, il sistema riesce a mappare i dati in modo sequenziale, filtrando rapidamente i soggetti non ammissibili o con anomalie evidenti.

Inoltre, le linee guida operative rivelano l'opzione di avviare un processo di "deroga" o *override* quando si verificano una o più situazioni di KO (*knockout*) automatico, ma il richiedente fornisce elementi compensativi (garanzie, co-obbligati, nuovi bilanci non ancora caricati in anagrafe, ecc.). In questo caso, il flusso prevede:

1. Arresto temporaneo del *prescore*.
2. Avvio di richiesta di deroga, con relativa istruttoria manuale da parte del gruppo rischio/commerciale.
3. Se la deroga è approvata, si procede con l'aggiornamento parziale dei dati o della documentazione aggiuntiva; in caso contrario, la pratica resta in KO.

Simile meccanismo di *override* si integra, infine, con la fase di monitoraggio continuo, nella quale eventuali nuove evidenze (un bilancio aggiornato, la rimozione di una negatività) consentono di ricalcolare il *prescore*.

La decisione di inserire ogni pratica e relativo esito in un sistema unico (*Airtable* o similari) porta a diversi benefici:

1. Tracciabilità totale: in qualsiasi momento, il *management* o i *risk analyst* possono "navigare" tra le pratiche e prendere visione dello storico dei motivi di rifiuto, le soglie violate, i valori limite che hanno portato a un *ALERT*, o i commenti interni degli operatori.
2. Possibilità di ricalibrazione: se alcune regole vengono riviste (ad esempio, si sposta la soglia di Debiti/EBITDA da 5 a 6), è possibile rianalizzare l'archivio delle pratiche per stimare l'impatto di tale modifica. Se risultassero molte pratiche respinte con valori compresi tra 5 e 6, potrebbe indicare eccessiva cautela, richiedendo un adeguamento della politica.
3. Supporto a decisioni future: tenere traccia di note e documenti associati a ciascuna pratica semplifica la futura analisi di performance del modello (ad esempio, per valutare quante pratiche KO si sono poi rivelate effettivamente in *default* su altri canali, o quante *ALERT* non hanno ottenuto la documentazione necessaria).

4. Miglioramento continuo: se emergono indicatori "residuali" che si rivelano poco discriminanti (cioè, generano troppi falsi positivi o falsi negativi), si possono sostituire o ridefinire; se, viceversa, alcuni parametri non compaiono nel *prescore* ma l'analisi storica dimostra la loro grande capacità predittiva, è possibile aggiungerli.

3.2.3 Ruoli e responsabilità

Nell'ambito del *prescore*, è molto importante che ciascun attore interno all'organizzazione abbia ben chiari i compiti legati al proprio ruolo. Con una chiara suddivisione delle responsabilità[45], il processo di valutazione preliminare del merito creditizio risulta più rapido e lineare, assicurando anche un solido controllo sui rischi associati.

Il ***Credit Analyst*** rappresenta la figura di riferimento ogni qualvolta il sistema genera un *ALERT*. Il ruolo dell'analista, dunque, è di:

- Valutare la sanabilità dell'anomalia: se, ad esempio, il Debiti/EBITDA è leggermente superiore alle soglie *standard*, il *Credit Analyst* può decidere se una garanzia aggiuntiva o un aggiornamento del bilancio possano riportare i parametri entro limiti accettabili.
- Integrare documentazione e informazioni aggiuntive: quando un *ALERT* emerge a causa di dati mancanti o insufficienti (es. mancanza di *disclosure* MLT/BT nel bilancio), l'analista può chiedere all'impresa di fornire un bilancio diverso o attestare la presenza di garanzie.
- Relazionarsi con l'impresa: se esistono dubbi interpretativi (ad esempio, scostamenti importanti tra un esercizio e l'altro), l'analista è la figura che può chiedere chiarimenti all'azienda, direttamente o tramite il mediatore/confidi, per capire se l'*ALERT* può essere superato o se è necessario chiudere la pratica con un KO.

In caso di dubbi più tecnici (es. definizione delle soglie o reinterpretazione di un indicatore particolare), il *Credit Analyst* può consultare il *Risk Manager*, che detiene la competenza in materia di calibrazione generale delle regole.

[45] *La separazione delle funzioni è prescritta dalle normative di vigilanza, come stabilito dalla Circolare di Banca d'Italia n. 285/2013.*

Il ***Risk Manager*** è il custode della politica di rischio di Fin.promo.ter e definisce:

- Le soglie per ciascuna variabile (ad es. Debiti/EBITDA ≤ 5, CGS ≥ 30, ecc.), basandosi su criteri di prudenza e su dati storici di insolvenza. In caso di troppi KO o *ALERT* in un determinato periodo, può decidere di ricalibrare tali *cut-off point*.
- Le priorità tra i vari test: se determinati parametri sono ritenuti inderogabili (ad es. Patrimonio Netto < 0), si stabilisce che generino KO immediato evitando di sprecare risorse su pratiche già non conformi.

Il *Risk Manager* svolge anche analisi retrospettive (*backtesting*) confrontando, per esempio, le pratiche rifiutate con i successivi andamenti reali. Ciò serve a capire se una certa soglia genera troppi falsi negativi (richieste rifiutate ma che, in realtà, sarebbero state affidabili) o troppi falsi positivi (pratiche approvate che, in seguito, mostrano insolvenze). Sulla base di queste osservazioni, può suggerire aggiustamenti e miglioramenti continui del *prescore*, collaborando con la funzione *Compliance* e il *Team IT/Data Science*.

Il ***Team IT / Data Science*** è responsabile dell'implementazione tecnica del *prescore*, gestendo sia la parte infrastrutturale (integrazioni con le API di *Cerved*, *OpenAPI*, MCC, ecc.) sia la componente algoritmica (logica di "*stop* immediato" e catena di controllo delle soglie). Tra le principali attività di questo *team* rientrano:

- Sviluppo e manutenzione del *software*: garantire che il calcolo di *ratio* e *rating*, unito all'estrazione dei dati, funzioni correttamente e senza ritardi eccessivi.
- Gestione dei *feedback*: eventuali problematiche segnalate da analisti o *risk manager* (es. incongruenze su un indicatore, scostamenti nei dati di bilancio, calcoli non congrui) vengono trasformati in correzioni o miglioramenti dello *script prescore*.
- Aggiornamento della mappatura delle regole: in caso di variazioni di soglia o di introduzione di nuovi indicatori, il *Team IT/Data Science* aggiorna l'algoritmo, assicurandosi che sia allineato con i sistemi di *front-end* (*Airtable* o strumenti analoghi).

Questo *team*, inoltre, funge da collegamento tecnico per i servizi esterni; ad esempio, deve monitorare i costi e i volumi delle chiamate API (*Cerved*, catasto, ecc.) e ottimizzare

la sequenza di controlli in modo da evitare richieste superflue quando si è già in presenza di un KO certo.

La ***Compliance*** si occupa principalmente di due aree fondamentali:

- Riservatezza dei dati e conformità alle normative (GDPR, antiriciclaggio, requisiti Banca d'Italia). Il *prescore*, in quanto sistema che manipola dati sensibili e informazioni provenienti da banche dati esterne, deve essere progettato e gestito in modo da garantire la sicurezza e l'uso lecito delle informazioni trattate.
- Equità e non discriminazione. Se il *prescore* prevede l'esclusione di determinati settori o categorie di impresa, la *Compliance* si assicura che tale scelta sia basata su chiare direttive strategiche e su motivazioni di rischio oggettive, evitando comportamenti discriminatori (ad esempio, negare credito in base a fattori irrilevanti dal punto di vista dell'analisi creditizia).

Inoltre, la *Compliance* può controllare eventuali segnalazioni di clienti o mediatori che ritengono che una soglia sia troppo restrittiva, verificando se essa rispetta gli *standard* condivisi con le autorità di vigilanza.

Questa struttura organizzativa, dunque, risulta determinante per integrare il *prescore* in modo armonico all'interno dell'operatività di Fin.promo.ter, garantendo non solo efficienza, ma anche affidabilità e trasparenza nelle decisioni di credito.

3.3 *Performance* storica del *prescore*: analisi degli ultimi 12 mesi

L'analisi delle *performance* storiche del *prescore* ha come obiettivo principale lo studio ed il comportamento effettivo del sistema di prevalutazione implementato da Fin.promo.ter durante l'anno 2024, evidenziando il modo in cui le variabili utilizzate influenzano la classificazione delle pratiche ed il processo di selezione delle richieste di finanziamento. Poiché il *prescore* non costituisce una valutazione definitiva di merito creditizio, bensì uno strumento di preclassificazione, è di fondamentale importanza comprendere con precisione quali dati vengono acquisiti, come vengono elaborati e rispetto a quali regole viene assegnato un esito positivo ("*Prescore* OK") o negativo ("*Prescore* KO")[46].

[46] Tale segmentazione è coerente con le prassi operative descritte da Siddiqi (2006) in ambito di *performance monitoring*.

3.3.1 Metodologia di estrazione e classificazione delle pratiche

La base di dati utilizzata per l'analisi è stata estratta direttamente da *Airtable*, il sistema gestionale adottato da Fin.promo.ter per il processo di *prescore*. Ogni *record* presente nella piattaforma rappresenta una singola pratica creditizia (*lead*) e contiene variabili strutturate che seguono il suo percorso di valutazione.

Per circoscrivere il perimetro temporale dell'analisi, sono state filtrate tutte le pratiche registrate tra il 1° gennaio e il 31 dicembre 2024. Al fine di organizzare in maniera più ordinata l'analisi, le pratiche sono state raggruppate in base allo stato della pratica (es. "Erogata", "*Prescore* KO", "*Prescore* scaduto") e per motivazione di KO, laddove presente. Questa categorizzazione ha consentito di costruire, in *Excel*, tabelle *pivot* utili a rappresentare la distribuzione degli esiti e determinare le cause principali di esclusione automatica.

Un ulteriore passaggio ha riguardato l'integrazione dei dati di *default*, contenuti nel foglio di lavoro denominato: "*Aggiornamento rate_2023-2024*". In questo *dataset* sono riportate informazioni aggiornate sull'andamento delle pratiche erogate nel 2024. A partire da tali dati, è stata creata una variabile binaria di *default*, attribuendo:

- valore 1 alle pratiche per cui si è verificato un evento di *default* nell'anno 2024;
- valore 0 a tutte le altre pratiche erogate, non risultate in *default*.

Questa variabile è stata poi confrontata con le informazioni relative al *prescore* per valutare, nei passaggi successivi, il potere predittivo degli indicatori disponibili.

3.3.2 Struttura dei filtri applicati e costruzione delle tabelle *pivot*

A partire dai dati estratti ed organizzati secondo i criteri temporali e tipologici illustrati nella precedente sezione, è stata condotta un'attività di analisi volta a valutare l'efficienza del sistema di *prescreening* nella selezione automatica delle pratiche creditizie. L'analisi si è articolata su più fronti, con l'intento di esaminare il comportamento storico del *prescore* non solo in termini generali, ma anche nella sua capacità di prevedere il rischio di *default* e di discriminare tra pratiche con profili informativi differenti.

Le elaborazioni sono state realizzate mediante l'utilizzo di *Microsoft Excel*, sfruttando strumenti quali tabelle *pivot*, formule di aggregazione, segmentazioni per fascia di rischio e, laddove necessario, costruzione manuale di matrici classificatorie. La selezione delle variabili e delle modalità di aggregazione è stata guidata da una logica strettamente

coerente con la struttura operativa del sistema di valutazione automatica, tenendo in considerazione la natura delle variabili, della scala di misura e della loro rilevanza nel processo decisionale. Alcune tabelle hanno avuto carattere descrittivo e classificatorio, mentre altre si sono concentrate sull'efficacia predittiva rispetto al *default*. In uno specifico caso, inoltre, si è proceduto alla costruzione di una curva ROC per stimare la capacità discriminatoria dell'*AD Score*.

Di seguito sono riportate le singole analisi condotte, con una descrizione delle logiche adottate nella costruzione delle tabelle e dei grafici, nonché delle finalità metodologiche alla base di ciascun foglio di lavoro.

Distribuzione Esiti *Prescore*

Nel foglio "Distribuzione Esiti *Prescore*" è stata condotta un'analisi descrittiva dei principali esiti del processo di *prescore*. L'obiettivo era quello di distinguere in modo chiaro e significativo le pratiche che hanno superato la fase di valutazione preliminare da quelle che sono state bloccate automaticamente. L'elaborazione si è articolata in due passaggi principali, realizzati manualmente in *Excel*. La distribuzione delle pratiche per esito è stata sintetizzata tramite una **tabella *pivot***, in cui si è contato il numero di pratiche ("Conteggio di CF") associate a ciascun stato. Di seguito si riporta la tabella risultante.

Status pratica	**Conteggio di CF**
24 – Erogata	610
52 - Declinata Fin.promo.ter (DELIBERA FINPROMOTER)	27
53 - Declinata Finanziatore (DELIBERA NON FINPROMOTER)	29
54 - Istruttoria negativa	692
Totale complessivo	**1358**

Tabella 4 Distribuzione delle pratiche per stato di avanzamento

- **Etichette di riga**: stato di ciascuna pratica in base al valore contenuto nella colonna "*Status NEW*". Le modalità considerate includono "Erogata", "Declinata Fin.promo.ter", "Declinata Finanziatore" e "Istruttoria negativa". Questa classificazione riflette fedelmente la struttura degli esiti operativi secondo la codifica interna adottata da Fin.promo.ter.

- **Valori**: per ciascuna categoria, è stato effettuato un conteggio assoluto, tale scelta, come misura aggregata, è giustificata dal fatto che l'interesse era puramente frequenziale: si intendeva quantificare il numero di pratiche ricadenti in ciascun esito senza effettuare sintesi numeriche inappropriate.
- **Filtri**: non sono stati applicati filtri direttamente in *Excel*, poiché il perimetro temporale dell'analisi era già stato delimitato a monte nella fase di estrazione da *Airtable* (01/01/2024–31/12/2024).

Partendo da questa prima classificazione, è stata costruita una seconda tabella riepilogativa in cui gli esiti originari sono stati raggruppati in due macrocategorie**:**

- **Erogate**: comprendente esclusivamente le pratiche con esito "Erogata";
- **Bloccate dopo *prescore***: aggregazione di tutte le pratiche con esito "Declinata Finpromoter", "Declinata Finanziatore" e "Istruttoria negativa".

Questa rielaborazione ha avuto lo scopo di mettere in evidenza, in maniera più funzionale all'analisi, la distinzione tra le pratiche che hanno superato il filtro automatico e quelle escluse in fase iniziale. La semplificazione in due macrocategorie è riportata nella seguente tabella riepilogativa, quest'ultima mostra anche la percentuale di incidenza di ciascun gruppo sul totale analizzato.

ESITO	**NUMERO PRATICHE**	**%**
Erogate	610	45%
Bloccate dopo il *prescore*	748	55%

Tabella 5 Sintesi degli esiti del *prescore*

Per rappresentare graficamente la composizione percentuale tra queste due categorie, è stato creato un grafico a torta. La scelta di questa rappresentazione è stata effettuata per comunicare, in modo immediato, la proporzione relativa tra pratiche ammesse alla fase successiva e quelle escluse dal sistema automatizzato. Il grafico a torta è particolarmente adatto a questo scopo, in quanto facilita la percezione visiva del peso specifico delle due classi rispetto al totale. La percentuale di ciascun gruppo è stata calcolata direttamente nella tabella di supporto, rapportando il numero di pratiche per categoria al totale complessivo. Di seguito è riportato il grafico risultante.

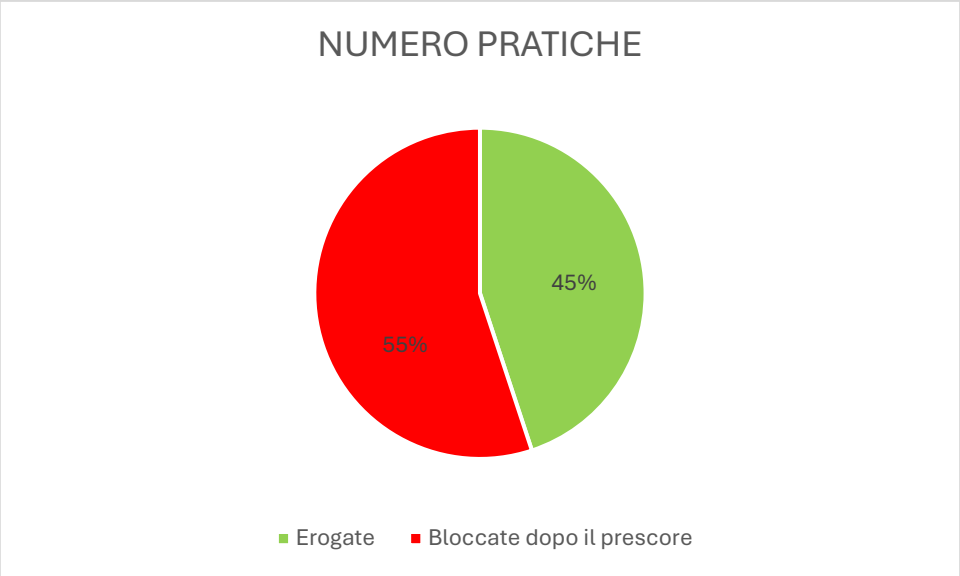


Grafico 1 Distribuzione Percentuale delle Pratiche per Esito del *Prescore*

Predittività esito KO

Nel foglio "Predittività esito KO" è stata costruita una tabella a doppia entrata, finalizzata a misurare la capacità discriminante del sistema di prescore rispetto al rischio di *default* effettivo. Questo è stato fatto incrociando le motivazioni di KO automatico con i risultati delle pratiche che sono state successivamente erogate. L'impostazione è stata realizzata manualmente in Excel, suddivisa in due sezioni analitiche complementari.

- **Etichette di riga**: nella prima colonna sono state riportate le diverse **motivazioni di KO** indicate dal sistema di prescore (es. "KO Ateco", "KO Consistenza patrimoniale", "KO Rating profilo aziendale (AD Score)", ecc.). L'inserimento di queste etichette consente di analizzare il comportamento del sistema in funzione dei criteri di esclusione automatica effettivamente applicati alle pratiche.
- **Valori e colonne**: nella prima sezione della tabella sono state conteggiate, per ciascuna motivazione di KO, le pratiche che sono comunque giunte allo stato di "Erogata", con distinzione tra:
    - *Default* = 0 (pratiche sane)
    - *Default* = 1 (pratiche deteriorate)

  I conteggi sono stati inseriti manualmente utilizzando i dati contenuti nel foglio "Aggiornamento *rate*_2023-2024". I casi in cui una pratica con motivo di KO viene successivamente erogata senza andare in *default* sono stati evidenziati come falsi positivi, poiché indicano una possibile esclusione errata da parte del *prescore*.

- **Filtri**: il perimetro di analisi è stato limitato alle sole pratiche erogate nel 2024, con informazioni disponibili circa il verificarsi o meno di *default*. Questa scelta assicura una coerenza logica tra la motivazione di KO e la possibilità di osservare un effettivo outcome economico-finanziario.

La tabella sottostante mostra l'incrocio tra le motivazioni di KO e l'esito in termini di *default* per le pratiche che, nonostante abbiano ricevuto almeno un KO, sono state successivamente erogate.

Status pratica	24 - Erogata	
	Default	
Motivazioni KO	**0**	**1**
KO Ateco	1	
KO Consistenza patrimoniale	1	
KO Altro	1	1
KO Assenza affidamenti	1	
KO *Rating* profilo aziendale (AD *score*)	3	
KO Anzianità impresa	4	
KO *Rating* ec-fin (CGS)	9	1
KO Indicatori di bilancio	23	5
Totale complessivo	**43**	**7**

Tabella 6 Distribuzione delle motivazioni di KO per pratiche erogate

Un ulteriore approfondimento ha riguardato la verifica della presenza di rate scadute tra queste pratiche, con l'obiettivo di osservare eventuali segnali premonitori di deterioramento. La tabella sottostante sintetizza i risultati.

Conteggio di CF	***Status* pratica**
Rate scadute	**24 - Erogata**
0	43
1	7
Totale complessivo	**50**

Tabella 7 Distribuzione delle pratiche erogate in base al numero di rate scadute

Per rappresentare graficamente i risultati è stato costruito un grafico a barre orizzontali affiancate, che raffigura, per ciascuna motivazione di KO, la suddivisione tra pratiche erogate andate in *default* e pratiche erogate che sono rimaste regolari.

Di seguito si riporta il grafico prodotto, costruito sulla base della tabella precedente.

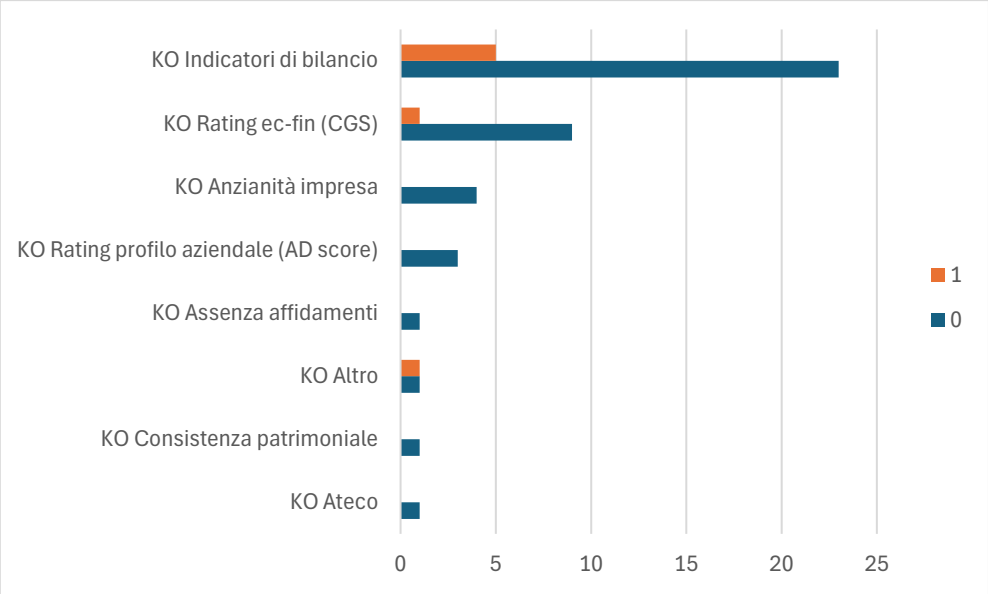


Grafico 2 Distribuzione delle Motivazioni di KO per Pratiche Erogate

Questa forma di visualizzazione è stata preferita in quanto consente di apprezzare, per ogni motivazione di esclusione automatica, la distribuzione interna tra comportamenti virtuosi e deteriorati, mettendo in evidenza la quota di errori di classificazione eventualmente generati dal sistema.

Variabili Rilevanti

Nel foglio "Variabili Rilevanti" è stata condotta un'analisi quantitativa con l'obiettivo di comprendere in che misura alcune variabili economico-finanziarie, associate alle pratiche erogate, fossero correlate con l'esito di *default*. L'analisi si è articolata in due momenti distinti: la costruzione di una tabella *pivot* e il calcolo successivo, separato, dei coefficienti di correlazione.

- **Etichette di riga**: nella tabella *pivot* sono state disposte, per ciascuna riga, le diverse variabili analitiche (es. CGS, Eurisc, Debiti/Fatturato, Deb/EBITDA, Ebitda *margin*, Sostenibilità, Capitalizzazione, ecc.), in quanto rappresentano dimensioni rilevanti nella determinazione del rischio secondo la logica del *prescore*. L'inserimento in riga permette un confronto diretto tra gruppi di pratiche con e senza *default*.
- **Colonne**: nella sezione delle colonne è stata inserita la variabile binaria "*Default*", per distinguere le pratiche andate in *default* (1) da quelle non andate in *default* (0).

Questa impostazione consente un confronto tra i valori medi e le deviazioni *standard* delle variabili finanziarie condizionate all'esito di *default*.

- **Valori**: per ciascuna combinazione di variabile e gruppo di *default*, sono state calcolate, due misure statistiche: la media e la deviazione *standard*. Le variabili per le quali sono state calcolate tali misure includono: CGS, Eurisc, Debiti tributari/Fatturato, Debiti/Fatturato, Ebitda *margin*, Livello di indebitamento, Sostenibilità, Deb/EBITDA e Capitalizzazione. La media è stata scelta per identificare eventuali differenze nei livelli centrali delle variabili tra pratiche sane e deteriorate. La deviazione *standard* è stata utilizzata per valutare la dispersione dei valori, utile per individuare variabili più stabili o più volatili nei due gruppi.
- **Filtri**: sono stati esclusi i *record* con valori nulli o anomali nelle variabili analizzate, per garantire l'affidabilità dei risultati statistici. Inoltre, l'analisi è stata condotta unicamente sulle pratiche erogate nel 2024 per le quali fosse noto l'esito di *default*.

Successivamente alla costruzione della tabella *pivot*, è stata elaborata una seconda analisi, esterna alla *pivot*, che ha previsto il calcolo manuale del coefficiente di correlazione lineare[47] tra ciascuna variabile e la variabile di *default*. Il calcolo è stato eseguito su coppie di colonne contenenti, rispettivamente, i valori numerici della variabile esplicativa e i corrispondenti valori binari della variabile *target* (*default* = 1 o 0).

Per rappresentare visivamente l'andamento delle correlazioni è stato costruito un grafico a dispersione (*scatter plot*). In quest'ultimo, ogni punto rappresenta una variabile posizionata sull'asse delle ascisse in ordine arbitrario, mentre l'ordinata riflette il valore della correlazione. Questa rappresentazione è stata scelta per la sua efficacia nel mettere in evidenza, in modo intuitivo, l'intensità e il segno delle relazioni lineari tra variabili e rischio di *default*.

Di seguito è riportato il grafico risultante.

[47] funzione CORRELAZIONE

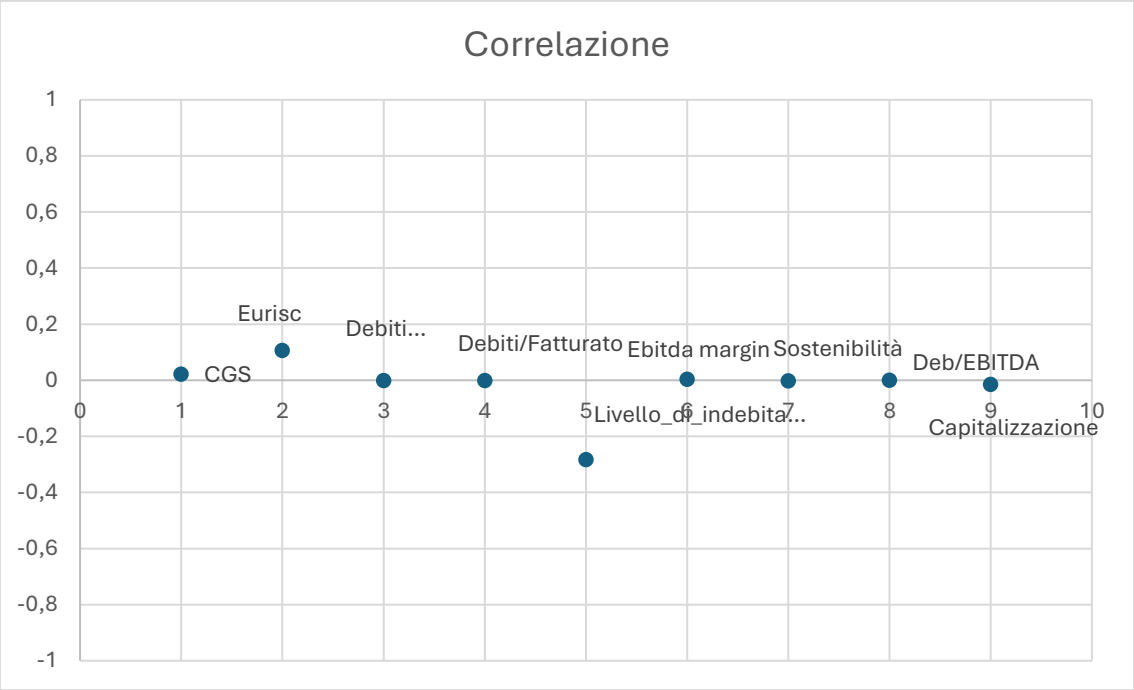


Grafico 3 Analisi della Correlazione tra Variabili Economico-Finanziarie e Rischio di *Default*

Tasso di *Default* per Classe CGS e Eurisc

Nel foglio "Tasso di *Default*-CGS e Eurisc" è stata condotta un'analisi incrociata con l'obiettivo di valutare in che misura le combinazioni tra due indicatori di *scoring*, il *rating* CGS e il *rating* Eurisc, si associno a differenti probabilità di *default* per le pratiche effettivamente erogate. L'analisi è volta a verificare se l'integrazione tra la valutazione economico-finanziaria (CGS) e l'affidabilità creditizia storica (Eurisc) consenta una segmentazione più efficace del rischio rispetto all'osservazione dei singoli indicatori.

L'elaborazione è stata effettuata mediante la costruzione di una tabella *pivot* principale, affiancata da due tabelle *pivot* marginali, riferite rispettivamente al solo *rating* CGS e al solo *rating Eurisc*. Sulla matrice principale è stata inoltre applicata una formattazione condizionale con mappa di calore, per evidenziare visivamente i livelli di rischio più elevati all'interno della tabella. Le celle con valori più alti sono state colorate con tonalità progressivamente più intense, rendendo facile individuare le combinazioni di punteggio più critiche.

- **Etichette di riga**: nella tabella *pivot* principale è stato inserito il *rating* Eurisc, suddiviso in tre categorie qualitative: "alto", "medio" e "basso". L'inserimento di

Eurisc, come variabile di riga consente, di analizzare l'andamento del rischio di *default* al variare della qualità creditizia derivante dai dati di Centrale Rischi[48].

- **Colonne**: sono stati riportati i livelli del *rating* CGS, "buono", "medio" e "scarso", secondo la classificazione interna Fin.promo.ter. Questa disposizione consente di confrontare, per ciascuna classe Eurisc, le differenze di rischio associate a diverse valutazioni economico-finanziarie.
- **Valori**: è stata calcolata la media della variabile "*Default*", che assume valore 1 in caso di deterioramento e 0 altrimenti. Questo calcolo ci offre una stima del tasso medio di *default* osservato in ciascuna combinazione di *rating*, permettendo così un confronto diretto tra i vari incroci CGS–Eurisc.
- **Filtri**: è stato applicato un filtro sul campo "*Status NEW*" per considerare esclusivamente le pratiche con esito "24 - Erogata". Tale selezione è stata necessaria per circoscrivere l'analisi alle sole pratiche che hanno generato un'esposizione effettiva e per le quali risulta osservabile un eventuale evento di *default*.

Status pratica	24 - Erogata		
	Eurisc		
CGS	buono	medio	scarso
Alto	4%	8%	11%
Medio	0%	4%	10%
Basso	0%	14%	25%

Tabella 8 Analisi Congiunta del tasso di *default* per livelli di *rating* CGS ed Eurisc

Per completare l'analisi principale, sono state costruite due tabelle *pivot* aggiuntive. La prima mostra il tasso medio di *default* per ciascun livello di CGS, aggregando su tutte le fasce Eurisc; la seconda mostra l'analogo tasso per ciascuna fascia Eurisc, aggregando tutte le categorie di CGS. Entrambe le tabelle sono state strutturate seguendo le stesse logiche metodologiche descritte in precedenza.

- **Etichette di riga**: nella prima tabella marginale sono stati disposti i livelli del *rating* CGS; nella seconda, le fasce del *rating* Eurisc.

[48] La Centrale Rischi della Banca d'Italia è una banca dati pubblica che raccoglie informazioni su esposizioni creditizie e situazioni di rischio, utilizzata dagli intermediari per valutare l'affidabilità dei richiedenti.

- **Valori**: in entrambi i casi, è stata calcolata la media della variabile "*Default*" per stimare il tasso medio di *default* per ciascun indicatore considerato singolarmente.
- **Filtri**: anche per le tabelle marginali è stato mantenuto il filtro sullo stato "Erogata", in linea con la scelta metodologica generale.

Media di *Default*	**Eurisc**		
***Status* pratica**	**buono**	**medio**	**scarso**
24 – Erogata	3%	7%	11%

Tabella 9 Tasso medio di *default* per livelli di *rating* Eurisc nelle pratiche erogate

Media di *Default*	**Eurisc**		
***Status* pratica**	**alto**	**medio**	**basso**
24 – Erogata	7%	5%	18%

Tabella 10 Tasso medio di *default* per livelli di *rating* CGS nelle pratiche erogate

A partire dalla tabella principale è stato costruito un grafico a colonne verticali affiancate, nel quale ciascuna categoria del *rating* Eurisc è rappresentata lungo l'asse delle ascisse, mentre per ogni fascia sono riportate tre colonne verticali corrispondenti ai livelli del rating CGS. L'asse delle ordinate riporta i tassi medi di *default* espressi in forma percentuale. Di seguito è riportato il grafico risultante.

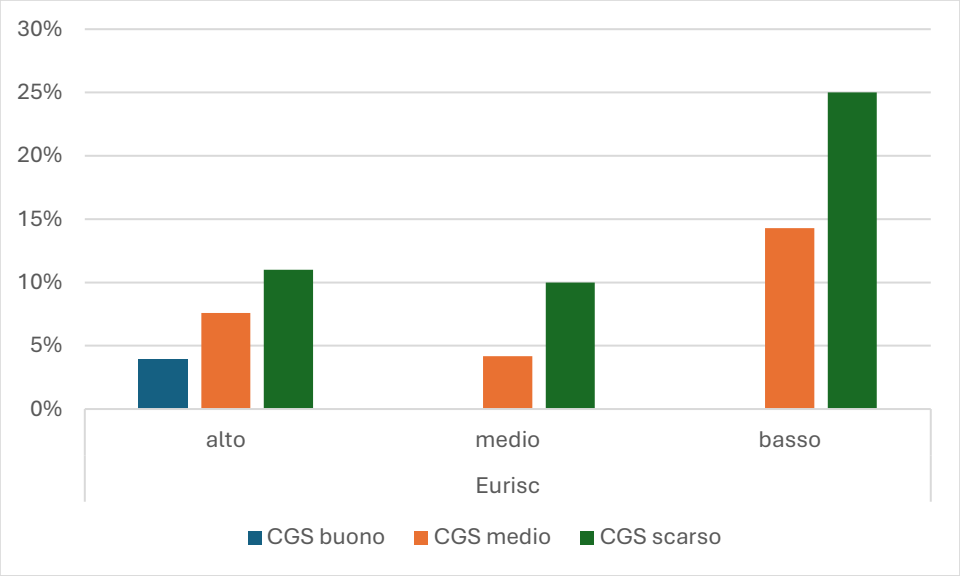


Grafico 4 Distribuzione CGS e Eurisc

Questa visualizzazione offre un modo intuitivo per rappresentare l'evoluzione del rischio al variare delle combinazioni tra i due punteggi. Il grafico mostra un chiaro incremento del tasso di *default* al peggiorare simultaneo dei *rating* CGS ed Eurisc. In particolare, la fascia "Eurisc basso" congiunta a "CGS scarso" presenta un'incidenza di *default*

superiore al 25%, mentre la combinazione “Eurisc alto” e “CGS buono” registra valori inferiori al 5%. Tale evidenza conferma l’utilità dell’approccio combinato per identificare precocemente pratiche ad alto rischio.

Analisi ROC

Nel foglio "Analisi ROC" è stata condotta un’analisi finalizzata a verificare la capacità discriminatoria dell’AD *Score* nella previsione del *default*, attraverso la costruzione manuale di una tabella classificatoria e della relativa curva ROC. A differenza delle analisi precedenti, in questo caso non è stata utilizzata alcuna tabella *pivot*: tutti i calcoli sono stati effettuati manualmente su una matrice costruita ad hoc in *Excel*.

È stata, innanzitutto, creata una colonna "*Default* Predetto", nella quale è stato attribuito il valore 1 (*default* previsto) a tutte le pratiche con AD *Score* maggiore o uguale a 5, mentre è stato assegnato il valore 0 (nessun *default* previsto) a tutte le altre. Questa soglia è stata scelta in base a una valutazione soggettiva, considerando la distribuzione osservata dell'AD *Score* e l'intenzione di identificare un livello di rischio che fosse sufficientemente alto da giustificare un’allerta. Questa colonna è stata poi confrontata con la colonna "*Default* Reale", che indica se la pratica è effettivamente andata o meno in *default* nel periodo osservato.

A partire dal confronto tra "*Default* Predetto" e "*Default* Reale" sono stati calcolati i seguenti indicatori di classificazione, fondamentali nella costruzione della curva ROC:

- **TP (*True Positives*)**: numero di pratiche per le quali il sistema ha previsto correttamente il *default* (*Default* Predetto = 1 e *Default* Reale = 1);
- **FP (*False Positives*)**: pratiche che il sistema ha erroneamente previsto come rischiose, ma che in realtà non sono andate in *default* (*Default* Predetto = 1 e *Default* Reale = 0);
- **FN (*False Negatives*)**: pratiche che il sistema ha classificato come non rischiose, ma che sono poi effettivamente andate in *default* (*Default* Predetto = 0 e *Default* Reale = 1);
- **TN (*True Negatives*)**: pratiche che sono state correttamente identificate come non rischiose (*Default* Predetto = 0 e *Default* Reale = 0).

Queste quattro categorie formano la base della matrice di confusione, fondamentale per ogni valutazione di classificazione binaria. A partire da esse, sono stati calcolati:

- **TPR (*True Positive Rate*)**: definito come TP / (TP + FN), rappresenta la sensibilità del modello, ovvero la sua capacità di individuare correttamente i casi di *default*;
- **FPR (*False Positive Rate*)**: definito come FP / (FP + TN), misura invece il tasso di falsi positivi, cioè la percentuale di pratiche sane che il modello sbaglia a classificare come rischiose.

I valori di TPR e FPR sono stati calcolati per diverse soglie di *AD Score*, generando così una serie di coordinate (FPR, TPR) utilizzate per costruire graficamente la curva ROC.

La seguente tabella riassume i valori di TP, FP, FN, TN, TPR e FPR calcolati per soglie decrescenti di *AD Score*, in modo da costruire l'intero profilo della curva ROC.

AD *Score*	***Default* previsto**	**TP (*True Positives*)**	**FP (*False Positives*)**	**FN (*False Negatives*)**	**TN (*True Negatives*)**	**TPR (*True Positive Rate*)**	**FPR (*False Positive Rate*)**
6	1	0	2	42	555	0	0,00359
5	1	0	30	42	527	0	0,05386
4	0	3	64	39	493	0,07143	0,11490
3	0	6	123	36	434	0,14286	0,22083
2	0	20	271	22	286	0,47619	0,48654
1	0	42	557	0	0	1	1

Tabella 11 Matrice di confusione per la valutazione dell'accuratezza del modello *AD Score*

La curva è stata rappresentata tramite un grafico a dispersione con una linea connettiva, dove ogni punto rappresenta una coppia (FPR, TPR) associata a una soglia specifica di *AD Score*. Il grafico è stato costruito collegando i punti in sequenza, rendendo così visibile l'andamento della curva ROC nel suo complesso. Il tratto crescente della curva riflette la maggiore capacità del modello di identificare correttamente i *default* (aumento del TPR), al costo però di un incremento dei falsi positivi (aumento del FPR).

Questa visualizzazione consente di analizzare il comportamento del classificatore al variare della soglia di discriminazione: nei punti iniziali (in basso a sinistra), il sistema è molto selettivo, con pochi falsi positivi ma anche una bassa sensibilità; nei punti più in alto a destra, la sensibilità aumenta, ma a scapito di una minore specificità. Il grafico include inoltre la bisettrice, ovvero la diagonale da (0,0) a (1,1), che rappresenta la

performance attesa di un classificatore puramente casuale. Maggiore è la distanza della curva ROC da questa linea, migliore è la capacità predittiva del modello.

In questo contesto, la costruzione della curva ROC ha rappresentato uno strumento essenziale per valutare la qualità dell'*AD Score* come variabile discriminante, consentendo di andare oltre l'analisi statica delle soglie e fornendo un quadro dinamico dell'accuratezza predittiva del sistema.

Di seguito si riporta il grafico risultante.

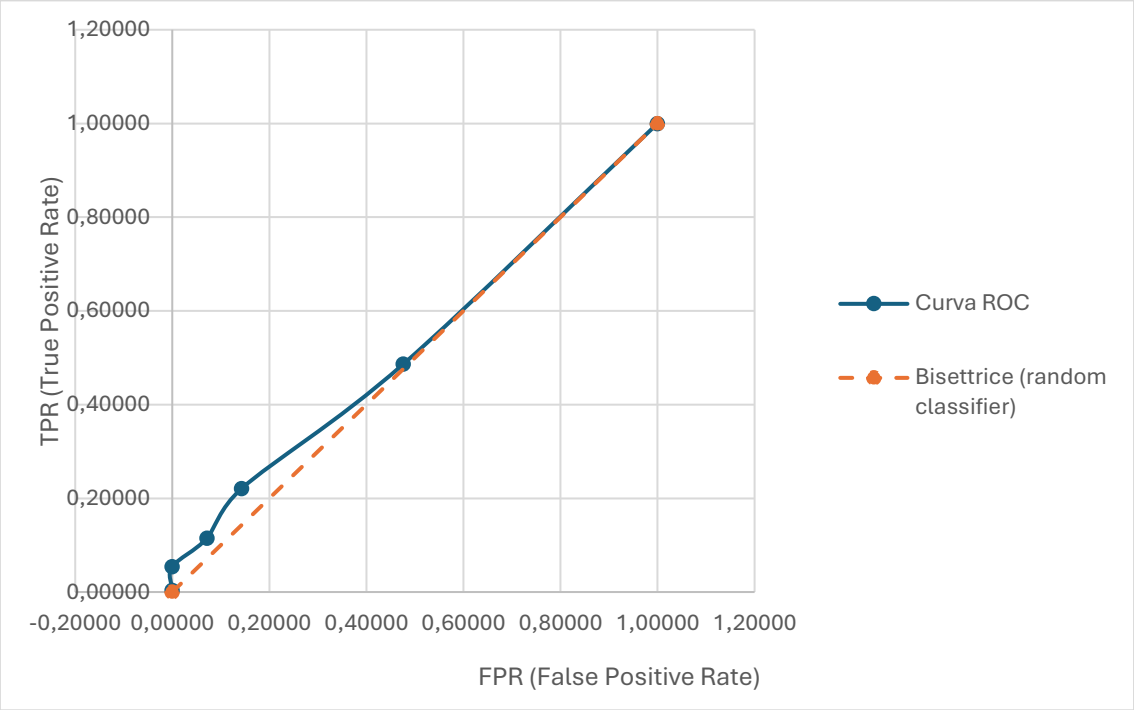


Grafico 5 Curva ROC

L'analisi ROC consente dunque di valutare il compromesso tra sensibilità e specificità, fornendo sia un'indicazione visiva che quantitativa dell'efficacia dell'*AD Score* nel distinguere tra pratiche che andranno in *default* e pratiche regolari. Una curva che si colloca stabilmente al di sopra della bisettrice indica un buon potenziale discriminatorio del modello, mentre una curva vicina alla diagonale evidenzia una bassa capacità predittiva.

3.3.3 Le variabili chiave del *prescore*: fonti, significato e ruolo

In questa sezione si approfondiscono le variabili utilizzate per costruire le tabelle *pivot* presentate nel paragrafo precedente, selezionate in quanto centrali nella logica di funzionamento del sistema *Finprescore* e presenti nei cruscotti decisionali operativi.

L’obiettivo non è solo fornire una descrizione tecnica delle variabili, ma chiarirne anche la funzione logica all’interno del modello di prevalutazione e le motivazioni alla base della loro inclusione nel processo analitico.

Come illustrato nei paragrafi precedenti, la logica del *prescore* si fonda su un sistema di regole deterministiche che attiva automaticamente *flag*, soglie e filtri basati su *input* informativi provenienti da fonti esterne come *Cerved*, *OpenAPI*, RNA, CR e MCC, oltre a dati elaborati nel database *Airtable*. In questo contesto, alcune variabili giocano un ruolo chiave nella classificazione delle pratiche, mentre altre forniscono segnali indiretti riguardo al rischio o all'affidabilità del richiedente. Le variabili oggetto di questo approfondimento non sono state trattate in precedenza e meritano una spiegazione dedicata.

Una prima variabile di rilievo è **“*Status* pratica”**, che identifica lo stato amministrativo conclusivo della pratica all’interno del flusso operativo. Essa può assumere valori quali *Prescore KO,* Erogata, Da riprocessare o altri stati intermedi. Sebbene possa apparire come una variabile puramente descrittiva, in realtà rappresenta il punto cruciale del sistema decisionale, riassumendo l'esito della richiesta in base alle regole implementate. Nel contesto dell’analisi, questa variabile è stata impiegata come chiave di aggregazione per confrontare l’andamento delle pratiche e misurare la frequenza relativa degli esiti, sia a livello aggregato che all’interno di sottogruppi segmentati per *score* o caratteristiche patrimoniali.

Un’altra variabile centrale è **“Motivazioni KO”**, che rappresenta una componente chiave per la tracciabilità e la trasparenza del processo. Si tratta di una variabile testuale compilata automaticamente dal sistema sulla base dei controlli falliti. A differenza di un semplice indicatore binario di rifiuto, questa variabile permette di individuare con precisione la specifica regola (o combinazione di regole) che ha generato l’esclusione, offrendo un supporto essenziale per l’analisi qualitativa dei *driver* di KO. Nell’ambito dell’analisi svolta, è stata utilizzata per costruire una classificazione gerarchica delle motivazioni prevalenti, misurando la loro incidenza relativa sul totale dei casi respinti. All’interno del *dataset* utilizzato per la valutazione delle *performance* del *prescore,* la variabile **“*Default*”** assume un ruolo cardine poiché rappresenta l’esito osservabile dell’evento di rischio che il modello di *prescreening* mira a prevedere o prevenire. Si tratta di una variabile binaria (*dummy*), costruita *ex post* sulla base dell’andamento

effettivo delle pratiche di finanziamento, con valore pari a 1 nel caso in cui la posizione abbia manifestato una condizione di deterioramento creditizio secondo i criteri adottati internamente da Fin.promo.ter, e pari a 0 in caso contrario. L'identificazione di una posizione in *default* si fonda su criteri osservabili come il mancato rimborso delle rate, la presenza di posizioni scadute e/o sconfinanti da oltre 90 giorni, o l'intervenuta classificazione dell'esposizione tra le sofferenze in Centrale Rischi[49]. In alcuni casi, possono essere inclusi anche elementi qualitativi provenienti dal monitoraggio *post*-erogazione (es. *report* negativi da parte dei gestori, segnalazioni da aree operative). Dal punto di vista metodologico, la variabile "*Default*" viene utilizzata come *target* (o variabile dipendente) nelle analisi statistiche e predittive. Questo approccio permette di valutare quanto bene le variabili indipendenti, come indicatori finanziari e *rating* esterni, riescano a discriminare tra le diverse situazioni, e quindi, quanto sia efficace il sistema di *prescore* nel segnalare in anticipo le posizioni a rischio. Le analisi condotte evidenziano differenze significative nelle medie delle variabili rilevanti tra i gruppi "*default*" e "non-*default*", confermando l'affidabilità dei segnali anticipatori individuati dal modello. L'utilizzo di una definizione rigorosa e stabile del *default* risulta essenziale non solo per la validazione interna del *prescore*, ma anche per garantire la coerenza con le logiche di *risk management* e con gli *standard* previsti dalle normative di vigilanza in materia di *credit risk modelling*[50].

Infine, è stata prestata particolare attenzione ad alcune variabili economico-finanziarie costruite a partire dai bilanci disponibili tramite *Cerved*. In particolare, il **"Livello di indebitamento"**, la **"Sostenibilità della rata"**, la **"Capitalizzazione",** il rapporto **"Debiti Tributari/Fatturato"**, il rapporto **"Debiti/Fatturato"**, il rapporto **"Debiti/EBITDA"** e l'**"EBITDA *margin*"**. Questi indicatori, talvolta oggetto di soglie fisse nel sistema (es. *KO se EBITDA margin < 3%*), non solo costituiscono strumenti tecnici per la valutazione del merito creditizio, ma forniscono anche *insight* aggiuntivi sulla capacità dell'impresa di reggere un nuovo impegno finanziario. Essi sono stati

[49] La classificazione a sofferenza nella Centrale Rischi di Banca d'Italia rappresenta uno degli indicatori ufficiali di deterioramento creditizio, utilizzato anche per finalità di vigilanza prudenziale.
[50] Cfr. EBA/GL/2017/16, "*Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures*", *European Banking Authority, 2017*.
https://www.eba.europa.eu/regulation-and-policy/credit-risk/guidelines-on-pd-lgd-estimation-and-treatment-of-defaulted-exposures

utilizzati in modo esplorativo per segmentare le pratiche in fasce di rischio e confrontare il comportamento del *prescore* su *cluster* con caratteristiche contabili differenti.
A scopo riepilogativo, si riporta, in **Appendice A**, una tabella sintetica che descrive le principali variabili utilizzate nell'analisi, la loro fonte, la funzione nel processo di valutazione e il criterio per cui sono state selezionate.
La combinazione di variabili interne, indicatori esterni e *score* derivati consente dunque di esplorare in profondità la coerenza, l'efficacia e il bilanciamento della logica *prescore*. La loro inclusione nelle tabelle *pivot* permette non solo una visione aggregata dei risultati, ma anche una lettura trasversale delle dinamiche decisionali.

3.4 Limiti, inefficienze e impatti economici del modello attuale

L'analisi critica del modello di *prescreening*, attualmente adottato da Fin.promo.ter, ha messo in evidenza diverse vulnerabilità strutturali e metodologiche. Queste non sono immediatamente evidenti da una semplice lettura degli *output* aggregati, ma diventano chiare quando si adotta un approccio integrato che confronta i risultati delle decisioni automatiche con i *trend* osservati *ex post* dei finanziamenti concessi. In particolare, l'analisi ha rivelato quattro ordini di criticità, che si configurano come determinanti nel comprometterne l'efficacia operativa e predittiva: (i) una propensione eccessiva all'esclusione delle pratiche, portando a una significativa perdita di opportunità commerciali potenzialmente valide; (ii) una logica deterministica che si basa sull'applicazione rigida di soglie predefinite, incapace di adattarsi ai cambiamenti nei profili di rischio o alle interazioni tra variabili; (iii) una scarsa coerenza statistica tra le variabili esplicative utilizzate nel modello e il verificarsi dell'evento di *default*, che ne riduce la capacità informativa effettiva; (iv) infine, l'assenza di un meccanismo di *feedback* che consenta al sistema di apprendere dai risultati passati e migliorare progressivamente le proprie *performance* decisionali.
Queste limitazioni non solo inficiano la precisione del modello nel distinguere tra pratiche sane e deteriorate, ma comportano anche impatti economici e organizzativi tangibili. Le esclusioni improprie determinano un incremento dei falsi positivi[51], riducendo il bacino erogabile e causando rallentamenti nel processo di valutazione a causa della necessità di

[51] Per falsi positivi si intendono pratiche escluse in fase di *prescreening* che, in base all'andamento osservato, non hanno manifestato condizioni di rischio o *default*, risultando quindi "rifiuti errati".

interventi manuali per rivalutare posizioni potenzialmente meritevoli. Inoltre, l'utilizzo di indicatori scarsamente discriminanti[52] introduce una componente di rumore informativo, che può indurre a decisioni subottimali, sia sul fronte dell'accettazione che del rifiuto.

In sintesi, l'analisi rivela che l'attuale modello *prescore* è costruito su un *framework* concettuale statico e non reattivo, inadatto a catturare la complessità e il dinamismo dei profili di rischio presenti nel mercato del credito alle imprese. Queste considerazioni motivano la necessità di una revisione profonda della struttura del *prescore*, utilizzando strumenti più flessibili basati su approcci quantitativi e adattivi, che consentano di rafforzare la capacità predittiva e la coerenza decisionale del sistema.

Sovra esclusione sistemica e squilibrio nella selezione

L'elevata percentuale di pratiche escluse dal flusso successivo alla prevalutazione, circa il 55% del totale analizzato, suggerisce l'esistenza di una logica di filtraggio tendenzialmente difensiva. Sebbene una certa prudenza sia giustificabile in fase preliminare, la portata delle esclusioni sembra sproporzionata rispetto all'effettiva incidenza di inadempienze tra le pratiche erogate. Il dato più rilevante a tal proposito proviene dall'analisi incrociata tra esiti finali e rischio osservato: delle 50 pratiche con almeno una motivazione di KO ma successivamente ammesse all'erogazione, ben 43 non hanno manifestato segnali di deterioramento. Questa evidenza configura un tasso di falsi positivi pari all'86%, che si traduce, nella pratica, in un'occasione commerciale non colta e in un inefficiente utilizzo del canale di *origination*.

Esiti erronei e distorsione delle regole deterministiche

Un ulteriore elemento di criticità riguarda l'assenza di ponderazione nelle esclusioni automatiche. Alcune motivazioni di KO, in particolare quelle relative agli indicatori di bilancio e ai *rating* economico-finanziari, hanno un impatto significativo nel generare blocchi, ma allo stesso tempo mostrano una scarsa capacità di identificare con precisione i soggetti realmente rischiosi. Il fatto che 23 pratiche escluse per "KO indicatori di bilancio" e 9 per "KO CGS" siano state successivamente ammesse e siano rimaste performanti, evidenzia una rigidità delle soglie che impedisce di considerare la

[52] Un indicatore si definisce scarsamente discriminante quando presenta una bassa capacità di separare efficacemente soggetti a rischio da soggetti affidabili. Questo si misura con strumenti statistici come l'AUC-ROC, l'*information value* o il Gini *index*.

multidimensionalità dei profili aziendali. L'applicazione di regole assolute, senza un bilanciamento tra criteri contrapposti o una valutazione congiunta delle variabili, produce una logica binaria che non riflette la complessità delle situazioni reali.

Correlazioni deboli tra predittori e rischio di *default*

L'analisi statistica dei legami tra variabili e *default* ha confermato la scarsa capacità esplicativa del *set* informativo utilizzato. I valori di correlazione tra le variabili più frequentemente impiegate, come CGS, Eurisc, Debiti/Fatturato, Sostenibilità della rata, Capitalizzazione, e l'evento di *default* risultano prossimi allo zero, o comunque troppo deboli per essere considerati affidabili in ottica decisionale. In alcuni casi, il segno stesso della correlazione è controintuitivo (es. CGS e Capitalizzazione). L'unica eccezione parziale riguarda il rapporto Debiti/EBITDA, che presenta una modesta correlazione positiva, ma comunque non sufficiente a giustificare l'utilizzo isolato nel processo di *scoring*. Questi risultati mettono in discussione l'efficacia del *prescore* nella sua forma attuale e suggeriscono l'opportunità di riformulare il modello su basi statisticamente solide.

Capacità predittiva congiunta degli *score* CGS e Crif

L'analisi incrociata delle fasce qualitative attribuite ai due *score* principali, CGS e Eurisc, ha evidenziato una rilevante capacità predittiva congiunta, attualmente non sfruttata dal modello. In particolare, la combinazione di un CGS basso con un Crif scarso si associa a un tasso di *default* del 25%, valore decisamente superiore alla media di portafoglio. Anche la coppia CGS basso e Crif medio si distingue per un'incidenza di *default* elevata (14%), mentre il rischio sembra essere piuttosto contenuto nei quadranti in cui almeno uno dei due *score* si mantiene su livelli favorevoli.

Un dato particolarmente rilevante emerge anche dalla combinazione CGS alto + Crif scarso, che registra un *default* all'11%, suggerendo che il Crif sia in grado di intercettare segnali di fragilità non il CGS non riesce a rilevare. Al contrario, la zona di minima rischiosità si colloca nell'intersezione tra CGS alto e Crif buono o medio, dove l'incidenza di default oscilla tra il 4% e l'8%.

Queste evidenze metto in luce il limite di un impianto valutativo che considera i predittori in modo disgiunto e rigido, senza riconoscere l'importanza delle loro interazioni. Un

modello adattivo, basato su tecniche di segmentazione e analisi multivariata[53], potrebbe invece rivelare configurazioni di rischio latente che sfuggono a un approccio basato su soglie

Prestazione insoddisfacente del classificatore AD *Score*

L'analisi ROC ha fornito un'ulteriore prova della debolezza del sistema nella sua funzione predittiva. La curva ottenuta, costruita per soglie decrescenti di AD *Score*, mostra una traiettoria che si discosta solo lievemente dalla bisettrice, ovvero dalla linea teorica di un classificatore casuale. In particolare, per la soglia ≥5, attualmente utilizzata per determinare l'esclusione automatica, il tasso di individuazione corretta dei *default* (TPR) risulta nullo, mentre il tasso di falsi allarmi (FPR) è già pari al 5,4%. Solo abbassando la soglia a valori molto bassi (≥2) il TPR si avvicina al 48%, ma con un FPR superiore al 48%, generando un compromesso inaccettabile in termini di efficacia operativa.

Questa dinamica suggerisce che l'*AD Score*, nella configurazione attuale, non è in grado di distinguere in modo significativo tra soggetti ad alto e basso rischio.

Le inefficienze del modello non si limitano agli aspetti predittivi, ma si estendono anche al piano gestionale. La rigidità delle regole e l'alto numero di *over ride* richiesti comportano un aumento del carico operativo per gli analisti, con conseguente rallentamento dei tempi di delibera e aumento dei costi di processo. Inoltre, la mancanza di un meccanismo di aggiornamento o di auto-correzione rende il sistema vulnerabile a cambiamenti nel contesto economico o nel profilo della clientela *target*. L'assenza di adattività implica che il modello, una volta calibrato, resti statico, perdendo progressivamente la capacità di affrontare nuove configurazioni di rischio.

Le evidenze raccolte nel corso dell'analisi mettono in luce una fragilità strutturale del *prescore* attuale, riconducibile a un'impostazione deterministica non più adeguata alla complessità del rischio creditizio contemporaneo. Le soglie fisse, le esclusioni binarie e la debole significatività dei predittori concorrono a delineare un modello statico, poco sensibile e potenzialmente distorsivo. A ciò si aggiunge l'incapacità del sistema di valorizzare le interazioni tra variabili, come nel caso dei punteggi CGS e Crif, che, se

[53] Si fa riferimento a modelli statistici e algoritmici in grado di individuare *pattern* nascosti nelle interazioni tra variabili, come ad esempio *clustering*, *decision tree*, *random forest* o modelli *logit* con interazioni.

considerati congiuntamente, evidenziano una capacità di segmentazione e predittiva ben superiore rispetto al loro impiego isolato.

Il superamento di questi limiti passa necessariamente attraverso l'adozione di una logica adattiva, capace di apprendere dai dati, di aggiornarsi dinamicamente e di integrare le variabili non solo singolarmente, ma anche nelle loro relazioni più significative. Su queste premesse si fonda il progetto di revisione metodologica che sarà oggetto del Capitolo 4.

Capitolo 4: Sviluppo e valutazione di un modello di *prescore* adattivo per Fin.promo.ter

L'analisi condotta e presentata nel Capitolo 3 ha messo in luce le caratteristiche ed i limiti del sistema attualmente utilizzato da Fin.promo.ter per attribuire un punteggio preliminare alle pratiche di finanziamento (*prescore*). Quest'ultimo si basa su un approccio deterministico e poco flessibile. In questo contesto, il presente capitolo si propone di sviluppare e valutare un modello predittivo alternativo, realizzato con tecniche di *machine learning*, capace di adattarsi dinamicamente alle informazioni contenute nei dati disponibili.

L'introduzione di un approccio adattivo apre a nuove opportunità nella valutazione ex ante del rischio, rendendo il processo decisionale più reattivo alle specifiche configurazioni delle pratiche analizzate. Il modello proposto, pur nella sua semplicità strutturale, si basa su una logica di apprendimento dai dati che mira a superare i limiti dell'attuale sistema statico. Ci si interroga, in particolare, su quali caratteristiche delle pratiche risultino più informative, su quanto i modelli siano in grado di distinguere i diversi esiti attesi e su come queste nuove evidenze possano affiancare, o potenzialmente sostituire, le metriche ad oggi impiegate.

Nel corso del capitolo, saranno presentati i dati utilizzati, le scelte metodologiche adottate e i modelli costruiti, con un *focus* particolare alla loro capacità predittiva e al valore informativo delle variabili coinvolte. L'analisi empirica non si limita alla misurazione della *performance*, ma include anche aspetti legati all'affidabilità probabilistica delle previsioni, offrendo così una lettura più articolata e consapevole del processo di *scoring*.

4.1 Dati e metodologia

Questa sezione offre una panoramica dettagliata della base dati, dei criteri di selezione, delle trasformazioni effettuate e della *pipeline* di preparazione dei dati utilizzata per sviluppare i modelli predittivi.

Tutti i riferimenti a moduli *Python* (es. *airtable.py*, *columns.py*, *preprocessing.py*) si riferiscono a *script* sviluppati *ad hoc* in ambiente *Python* per l'elaborazione proprietaria dei dati interni alla piattaforma Fin.promo.ter.

L’obiettivo è duplice: da un lato, garantire la trasparenza del nostro approccio empirico, e dall’altro, assicurare che le analisi possano essere replicate in contesti operativi o in futuri sviluppi della piattaforma Fin.promo.ter.

Per raggiungere questo scopo, è stata utilizzata un'infrastruttura *Python* modulare, composta dai moduli *airtable.py*, *columns.py*, *config.py* e *preprocessing.py*, ognuno con compiti specifici nel processo di acquisizione, pulizia, normalizzazione e strutturazione del *dataset*.

4.1.1 Origine e struttura dei dati

L'analisi empirica è stata realizzata su un *dataset* che raccoglie dati microeconomici riguardanti le pratiche di credito gestite da Fin.promo.ter nel periodo dal 1° gennaio 2024 al 31 dicembre 2024. La scelta di questo intervallo in linea con gli obiettivi di questo studio, che mira a valutare le performance del sistema di *prescoring* adottato dalla società durante l'intero anno solare 2024. Inoltre, questo intervallo si allinea con quello utilizzato nelle analisi statistiche e descrittive del Capitolo 3.

I dati sono stati estratti dalla piattaforma gestionale *Airtable*, che Fin.promo.ter utilizza come sistema informativo centrale per gestire il ciclo di vita delle pratiche. Nella base di dati, la tabella denominata “*Leads*” contiene informazioni su ogni singola richiesta di finanziamento: dati anagrafici, documenti, esiti delle deliberazioni, punteggi di *scoring* e dati economico-finanziari provenienti da bilanci o fonti esterne. La struttura logica di *Airtable* permette una gestione semi-strutturata di tipo relazionale, dove ogni riga rappresenta una pratica e ogni campo corrisponde a un attributo informativo. La dimensione grezza del *dataset* iniziale supera le 400 colonne, molte delle quali non sono utilizzabili per scopi predittivi a causa di ridondanza, mancanza di standardizzazione o rilevanza esclusivamente operativa.

L'interrogazione della tabella *Leads* è stata effettuata tramite l'interfaccia API fornita da Airtable, utilizzando la libreria *pyairtable*[54] in ambiente *Python*, all'interno del modulo *airtable.py*. È stata sviluppata una funzione chiamata *download_records_trainer()* al fine di filtrare i record utili all'addestramento del modello predittivo, escludendo quelli che si trovano in uno stato non idoneo, come ad esempio:

[54] *pyairtable* è una libreria *Python open source* che consente di interfacciarsi con il database *Airtable* tramite API *RESTful*.

- pratiche mai avviate o in attesa di documentazione ("ATTESA AVVIO", "08 - Raccolta documenti"),
- pratiche archiviate o ritirate prima della valutazione ("51 - Ritirata"),
- pratiche scadute o non completate ("30 - Scaduto", "40 - Da riprocessare")

Il criterio di selezione che adottato è pensato per assicurare che ogni *record* rappresenti effettivamente una pratica con un iter decisionale concluso, indipendentemente dall'esito, sia esso positivo o negativo. Il risultato di questo lavoro è un *dataset* che include circa 2.000 pratiche, distribuite mensilmente, con una buona varietà in termini di settore economico, forma giuridica e localizzazione geografica delle imprese richiedenti.

Un altro aspetto metodologico importante riguarda la gestione delle informazioni sull'identità del richiedente. A questo scopo, è stata utilizzata la Partita IVA come chiave identificativa unica. Questa è fondamentale sia per il monitoraggio delle pratiche nel tempo, sia per costruire la variabile *target* T2 (rischio di *default*), che richiede un collegamento tra più pratiche riferite allo stesso soggetto. Inoltre, la Partita IVA viene utilizzata come secondo livello di indicizzazione nel *MultiIndex* del *dataframe Pandas*[55], consentendo un'aggregazione coerente e stabile dei dati longitudinali e transazionali. Infine, è importante sottolineare che, per garantire l'integrità e la riservatezza dei dati, l'intero processo di estrazione, normalizzazione e trasformazione è stato effettuato in un ambiente protetto e locale, senza alcuna trasmissione a servizi *cloud* esterni o database remoti, in conformità con le *policy* aziendali e le normative sulla protezione dei dati (GDPR)[56].

La funzione di estrazione *download_records_trainer()*, che realizza queste operazioni di selezione e costruzione del *dataset* iniziale, è documentata in dettaglio con codice e commento riga per riga in **Appendice B**, a supporto della piena trasparenza metodologica.

4.1.2 Struttura delle variabili esplicative e costruzione dei *target*

Creare un *dataset* per addestrare modelli predittivi supervisionati richiede una selezione attenta delle variabili esplicative e una definizione metodologica solida delle variabili obiettivo. In contesti operativi come quello che stiamo analizzando, dove l'informazione

[55] Il *MultiIndex* è una struttura di indicizzazione gerarchica in *Pandas* che consente di indicizzare i dati su più livelli (es. Lead ID e Partita IVA).

[56] Il GDPR (*General Data Protection Regulation* – Regolamento UE 2016/679) è il regolamento europeo sulla protezione dei dati personali, in vigore dal 25 maggio 2018.

è molto eterogenea e i processi aziendali sono complessi, è fondamentale che le scelte riguardanti le *feature* e i *target* non si basino solo su criteri statistici, ma anche su considerazioni pratiche, economiche e operative.

La selezione delle variabili esplicative si è fondata su un principio fondamentale: includere solo variabili disponibili prima della decisione creditizia, evitando qualsiasi forma di data *leakage*[57]. Per questo motivo, è stata condotta un'analisi approfondita delle colonne nella tabella *Leads*, con l'obiettivo di isolare gli attributi: rilevanti per prevedere il comportamento creditizio dell'impresa; completi (con una copertura elevata su quasi tutte le pratiche); strutturati e coerenti sia dal punto di vista sintattico che semantico.
Dal punto di vista tecnico, il modulo *columns.py* ha facilitato la definizione delle variabili informative attraverso le strutture *COLUMNS_GET* (colonne da estrarre), *COLUMNS_DIGIT* (numeriche continue) e *COLUMNS_INT* (valori interi).
Queste liste vengono poi utilizzate nel modulo *airtable.py* nella funzione *download_records_trainer()*, per costruire il primo livello di selezione delle *feature*.
Le 25 variabili selezionate sono state suddivise in tre macrocategorie:

- **Variabili contabili**. Espresse in euro o come rapporti (es. Fatturato, EBITDA, Debiti/EBITDA, Patrimonio netto). La loro interpretazione economica è immediata e ben consolidata nella letteratura sul *credit scoring*[58]
- **Indicatori sintetici**. Rappresentano i risultati di modelli già esistenti, come l'AD *Score*, il CGS Valore o la probabilità di *default* stimata. L'inserimento di queste caratteristiche ha due scopi: da un lato, permette di confrontare le performance predittive dei modelli sviluppati con gli strumenti già in uso; dall'altro, aiuta a verificare quanto queste informazioni siano utili rispetto ad altre variabili indipendenti.
- **Attributi categoriali**. Si riferiscono principalmente alla forma giuridica dell'impresa (ad esempio, SRL, SPA, ditta individuale). Anche se non sono variabili quantitative, possono fornire informazioni preziose sul profilo giuridico e fiscale dell'azienda, influenzando il suo comportamento creditizio.

[57] Il *data leakage* si verifica quando informazioni future o post-decisione influenzano erroneamente la fase di addestramento, portando a stime sovra ottimistiche.
[58] Altman, 1968; Basel Committee, 2001

Sono state escluse tutte le variabili che seguono una decisione (come l'esito MCC, gli importi deliberati, le note successive) e tutte le colonne con alta cardinalità o codifica testuale irregolare, poiché non compatibili con modelli quantitativi supervisionati.

Dopo aver identificato le *feature* predittive, è stato fondamentale sviluppare uno o più *target* che fossero in linea con l'oggetto dell'analisi. In questo contesto, si è optato per un approccio a doppio *target* binario, seguendo una strategia duale che distingue tra:

- valutazione *ex ante* del rischio accettativo, ovvero la probabilità che una pratica venga approvata in base alle informazioni iniziali (T1),
- valutazione *ex post* del rischio comportamentale, cioè la probabilità che, anche se approvata, la controparte possa diventare insolvente o inadempiente (T2).

Questa impostazione permette di sviluppare due modelli distinti ma complementari: il primo mira a ottimizzare il tasso di successo e l'efficienza della rete commerciale, mentre il secondo è focalizzato sulla riduzione del rischio economico reale.

***Target* T1 – Probabilità di approvazione della pratica**

Il primo *target* (T1) si basa sullo stato finale della pratica. È definito come segue:

$$T1_i = \begin{cases} 1 \; se \; Status\_NEW_i = \text{“}24 \; - \; Erogata\text{”} \\ 0 \; altrimenti \end{cases}$$

Tale indicatore è stato progettato per stimare la probabilità che una nuova pratica venga accettata. Si tratta quindi di un *target* utile per supportare la rete commerciale e l'area istruttoria, fornendo una misura di "approvabilità" del *lead*.

***Target* T2 – Rischio di *default* comportamentale**

Il secondo *target* (T2) si fonda sulla verifica della presenza della Partita IVA del richiedente in una lista interna di soggetti insolventi (*UNPAID_LIST*), stilata dall'area amministrativa. L'obiettivo è quello di catturare eventi avversi che si manifestano dopo l'erogazione. È definito come segue:

$$T2_i = \begin{cases} 1 \; se \; P.IVA_i \in UNPAID\_LIST \\ 0 \; altrimenti \end{cases}$$

Costruire T2 partendo dalla Partita IVA, piuttosto che dalla singola pratica, consente di aggregare storicità e ricorrenza di comportamenti scorretti, attribuendo la responsabilità a livello di soggetto giuridico.

L'utilizzo combinato di T1 e T2 permette di creare una mappa bivariata della qualità creditizia, distinguendo, per esempio:

- pratiche con alta probabilità di approvazione ma alto rischio, candidati per deroghe consapevoli;
- pratiche con bassa probabilità di approvazione ma basso rischio, rivalutazioni strategiche;
- pratiche con doppio alto rischio, da escludere;
- pratiche con doppio basso rischio, prioritarie per la rete.

Questo schema decisionale è particolarmente utile per sviluppare sistemi di supporto alla decisione (*Decision Support Systems*) e può alimentare logiche di automazione parziale o di pre-validazione da parte di agenti esterni.

L'intera logica per la costruzione dei *target* è implementata nel modulo *preprocessing.py*, all'interno della funzione *prepare_data_for_training()*[59]. Un estratto dettagliato e commentato del codice utilizzato è fornito in **Appendice C**, con spiegazioni tecniche riga per riga per facilitare la replicabilità del metodo.

4.1.3 Trattamento dei dati mancanti, codifica e standardizzazione delle variabili

Creare un *dataset* informativo solido per addestrare modelli predittivi richiede di prestare attenzione a tre aspetti fondamentali: i valori mancanti, la natura categorica di alcune variabili e la diversità delle scale di misura.

In un *database* gestionale come quello di Fin.promo.ter, è normale imbattersi in valori mancanti, causati da documentazione incompleta, pratiche ancora in fase di chiusura o informazioni non obbligatorie per tutte le aziende. Le tecniche di imputazione utilizzate variano a seconda del tipo di variabile:

- Variabili numeriche: per queste, si è adottata una strategia di imputazione univariata basata sulla media, utilizzando l'oggetto *SimpleImputer(strategy='mean')*[60] del modulo *sklearn.impute*. Questo approccio permette di mantenere la distribuzione della variabile senza ridurre la dimensione del campione.

[59] La funzione *prepare_data_for_training()* ha lo scopo di costruire un *dataset* numerico e completo per l'addestramento di modelli predittivi, a partire da dati grezzi estratti da *Airtable*. Include la definizione dei *target*, la separazione delle *feature*, l'imputazione dei valori mancanti, la codifica delle variabili categoriche, lo *scaling* delle *feature* numeriche e la serializzazione degli oggetti per uso futuro

[60] La media è una scelta comune per variabili numeriche simmetriche e prive di *outlier*; in presenza di valori estremi, la mediana può risultare preferibile.

- Variabili categoriche: per campi come la forma giuridica (OAPI-Forma giuridica), è stata utilizzata la moda, ovvero il valore più frequente *(strategy='most_frequent')*.

La decisione di non ricorrere a imputazioni multivariate (come *l'IterativeImputer*) è stata guidata dal principio di parsimonia e dalla volontà di ridurre al minimo l'inferenza arbitraria su dati non osservati.

Le variabili testuali, sebbene ricche di informazioni, non possono essere utilizzate direttamente nei modelli predittivi senza prima essere convertite in formato numerico. Per questo scopo, è stato adottato il metodo del *One-Hot Encoding*, utilizzando *OneHotEncoder()* di *scikit-learn* con i seguenti parametri:

- *drop='first'*[61]: rimuove la prima categoria per ogni variabile, evitando così la collinearità perfetta;
- *handle_unknown='ignore'*: consente al modello di gestire correttamente le modalità che non ha mai visto durante l'addestramento.

La seguente trasformazione, pur aumentando la dimensionalità del *dataset*, permette di mantenere intatta l'informatività delle variabili categoriali, rendendole adatte per modelli lineari, alberi decisionali e metodi *ensemble*.

Per garantire che le *feature* di natura diversa siano confrontabili e per migliorare la stabilità numerica degli algoritmi, si è resa necessaria la standardizzazione *z-score* tramite *StandardScaler()*[62] di *sklearn.preprocessing*. Questa tecnica trasforma ogni variabile secondo la formula:

$$Z = \frac{x - \mu}{\sigma}$$

dove μ è la media e σ la deviazione *standard* della variabile.

Tale trasformazione garantisce che tutte le *feature* abbiano media nulla e deviazione *standard* unitaria, condizione utile nei modelli lineari e nei modelli basati su distanza.

L'intera *pipeline* di *preprocessing* descritta è stata implementata nella funzione *prepare_data_for_training()* del modulo *preprocessing.py*, e viene documentata dettagliatamente nell'**Appendice C**. Tale funzione rappresenta il cuore dell'infrastruttura

[61] Il parametro *drop='first'* evita la collinearità perfetta tra le *dummy* generate eliminando la prima modalità di ciascuna variabile categorica.

[62] *StandardScaler* standardizza ogni variabile sottraendo la media e dividendo per la deviazione *standard*, rendendo comparabili variabili su scale diverse.

di pulizia e trasformazione dati, e ne garantisce la replicabilità su nuovi *dataset* in fase di predizione operativa.
Per una rappresentazione schematica delle principali fasi della *pipeline* di *preprocessing*, si rimanda all'**Appendice D**.

4.1.4 Struttura finale del *dataset* e controlli di coerenza

Alla fine della fase di *preprocessing*, il *dataset* è pronto per l'addestramento dei modelli predittivi supervisionati. Il *DataFrame* risultante ha una struttura ben definita, senza valori nulli, con *feature* standardizzate e completamente numeriche, e presenta un *MultiIndex* organizzato su due livelli: *Lead ID* (l'identificativo unico della pratica) e Partita IVA (l'identificativo fiscale dell'impresa).
La base informativa finale conta circa 2.000 osservazioni, relative al periodo dal 1° gennaio al 31 dicembre 2024, e il numero di colonne, dopo il *One-Hot Encoding* delle variabili categoriche, supera le 40 *feature* totali. Queste comprendono sia le 25 variabili originali (selezionate tramite *COLUMNS_GET*) sia le variabili *dummy* create per ciascuna modalità delle variabili categoriche. Il numero esatto può variare a seconda della cardinalità delle variabili testuali nel campione analizzato.
Prima di passare alla fase di modellazione, è stato essenziale garantire che il *dataset* fosse logicamente coerente e statisticamente valido. A tal proposito, sono stati effettuati i seguenti controlli:

- Assenza di colonne a varianza nulla: le *features* costanti, che non forniscono alcuna informazione utile, sono state automaticamente eliminate tramite un filtro che verifica se *X_train.var()* è uguale a 0;
- Verifica dell'assenza di duplicati: è stato eseguito un controllo per verificare che ogni riga fosse unica rispetto al *MultiIndex*, per eviare la presenza di pratiche duplicate;
- Esclusione di colonne con valori unici su quasi tutte le righe: le colonne che presentavano lo stesso valore nel 99% dei casi sono state scartate, poiché non aggiungono valore esplicativo;
- Controllo del bilanciamento delle classi *target*: si è proceduto al calcolo del numero e della percentuale di istanze per ciascuna classe di *target* T1 (pratica erogata) e T2 (impresa inadempiente). Questo bilanciamento è cruciale per

scegliere la metrica di valutazione dei modelli (accuratezza vs precisione/richiamo) e per decidere se adottare tecniche di bilanciamento (come *oversampling SMOTE*[63], sottocampionamento o pesatura delle classi).

Esempio di bilanciamento con valori simulati:

Target | Classe 1 | Classe 0 | Percentuale Positivi

T1 | 820 | 1.180 | 41%

T2 | 220 | 1.780 | 11%

Ciascun gruppo di variabili categoriche è stato analizzato per accertarsi che il numero di *dummy* generate corrispondesse correttamente alle modalità distinte meno una (n-1). Inoltre, è stata effettuata un'analisi descrittiva utilizzando *X.describe()* per valutare il *range*, la simmetria e i valori estremi. Alcune osservazioni con valori anomali o incoerenti (ad esempio, Debiti/EBITDA > 100) sono state escluse.

I controlli di qualità e coerenza sono stati realizzati attraverso funzioni diagnostiche su *Pandas* e *scikit-learn*. Un estratto del codice utilizzato per questi filtraggi preliminari è disponibile in **Appendice E**.

4.2 Sviluppo e implementazione dei modelli predittivi

Dopo aver completato la fase di preparazione e trasformazione del *dataset*, come descritto nel paragrafo precedente, si passa allo sviluppo dei modelli predittivi per valutare il rischio creditizio. In questa sezione, si ha il proposito di presentare gli algoritmi scelti, le strategie di addestramento e validazione, e l'architettura generale del sistema di modellazione automatizzato implementato. L'obiettivo è garantire che l'intero processo sia replicabile, trasparente e modulare, in modo da poterlo integrare facilmente nei processi aziendali futuri.

4.2.1 Scelta dell'algoritmo e configurazione del processo modellistico

L'implementazione dei modelli predittivi si colloca in una fase intermedia tra la preparazione del *dataset* (descritta nel paragrafo 4.1) e la valutazione dei risultati, che sarà trattata nei paragrafi successivi. In questo contesto, la scelta degli algoritmi è stata guidata da vincoli metodologici ed operativi. È stato fondamentale adottare modelli

[63] *SMOTE* (*Synthetic Minority Over-sampling Technique*) è una tecnica di bilanciamento che genera esempi sintetici della classe minoritaria interpolando tra i suoi vicini.

capaci di bilanciare accuratezza, robustezza e comprensibilità, specialmente considerando una possibile integrazione nei processi decisionali aziendali. Da questo punto di vista, per poter valutare la probabilità di successo della pratica (T1) e il rischio di *default* comportamentale (T2), sono stati utilizzati due modelli di *machine learning* supervisionati: la regressione logistica[64], che offre un'interpretazione lineare, e la *Random Forest*[65], nota per la sua abilità nel gestire non linearità e interazioni tra variabili.

L'intera *pipeline* è stata progettata per assicurare replicabilità e modularità: ogni fase, dall'addestramento alla validazione, è implementata attraverso funzioni specifiche, documentate nelle appendici tecniche. In particolare, è stato introdotto un sistema di *naming* standardizzato per i *file* generati, che facilita l'identificazione chiara di modello, *target* e tipo di *output*. In particolare, l'addestramento e la valutazione su base *hold-out* sono stati implementati tramite la funzione *train_and_evaluate()* (**Appendice F**), che automatizza l'intero processo: separazione dei dati, *fitting*, generazione delle metriche e salvataggio strutturato dei risultati.

Questo approccio è vantaggioso sotto due aspetti fondamentali:

1. Versionamento dei modelli[66]

   Ciò include la registrazione delle modifiche apportate, come l'aggiunta o la rimozione di variabili, l'aggiornamento dei dati di addestramento o la modifica dei parametri. Questa pratica consente di:

   - Confrontare le prestazioni tra diverse versioni del modello
   - Ripristinare versioni precedenti se necessario
   - Documentare l'evoluzione del modello per fini di *audit* o conformità

2. Analisi retrospettiva dei risultati

   Implica l'esame dei risultati ottenuti dai modelli per comprendere cosa ha funzionato bene e cosa potrebbe essere migliorato. Questo processo aiuta a:

   - Identificare eventuali problemi o anomalie nei risultati
   - Apprendere dalle esperienze passate per ottimizzare i modelli futuri

[64] La regressione logistica consente di stimare la probabilità di un evento binario e consente un'interpretazione diretta dei coefficienti in termini di *log-odds*, rendendola particolarmente adatta per contesti decisionali.

[65] La *Random Forest* è un *ensemble* di alberi decisionali addestrati su campioni *bootstrap* e *feature* casuali. È robusta al rumore e particolarmente efficace in presenza di interazioni tra variabili.

[66] Il versionamento è una pratica mutuata dall'ingegneria del software, utile per garantire tracciabilità, riproducibilità e auditabilità dei modelli predittivi.

- Migliorare continuamente il processo di sviluppo dei modelli

Un altro aspetto distintivo della configurazione adottata è la netta separazione tra i dati di addestramento e quelli di *test*. L'addestramento, infatti, è stato condotto su un campione stratificato, suddiviso in un rapporto 70/30 tra il *set* di addestramento e quello di *test*[67], per garantire che le classi fossero rappresentative nella fase di validazione *out-of-sample*, riducendo il rischio di *overfitting* e migliorando la generalizzabilità dei risultati. Il rispetto di questi criteri è stato verificato tramite controlli automatici sulle distribuzioni delle variabili nei due sottoinsiemi.

Inoltre, la configurazione è stata progettata per essere facilmente estendibile, dal momento che la struttura del codice permette di integrare senza sforzo nuovi algoritmi, parametri o trasformazioni. Questo rende la soluzione scelta non solo efficace per il contesto attuale, ma anche pronta a crescere con le evoluzioni future del *prescore*.

4.2.2 Validazione incrociata e analisi dell'importanza delle variabili

Un elemento fondamentale nello sviluppo di modelli predittivi affidabili è la valutazione della loro robustezza dinanzi a variazioni nei dati di *input*. Per questo motivo, è stata implementata una validazione incrociata a *5-fold* stratificata[68], che permette di stimare in modo più preciso la varianza delle metriche di *performance* rispetto a un singolo *split train/test*. La stratificazione assicura anche che ogni *fold* mantenga la proporzione originale delle classi, un requisito essenziale quando si lavora con *dataset* sbilanciati come quelli analizzati in questo studio.

L'intero processo è stato racchiuso nella funzione *`perform_cross_validation()`* (**Appendice H**), che automatizza l'esecuzione dei modelli sui vari *fold*, raccoglie le metriche di *output* e le organizza in un formato facilmente interpretabile. I risultati della *cross-validation* non sono solo uno strumento di controllo statistico, ma anche un punto di partenza per confrontare i modelli in termini di stabilità e affidabilità predittiva. La scelta delle metriche (ROC AUC, *accuracy*, *F1-score*[69]) riflette l'intento di valutare in modo dettagliato la capacità discriminativa, la precisione e l'equilibrio dei modelli.

---

[67] La stratificazione assicura che la proporzione tra classi (es. erogato/non erogato) sia mantenuta in entrambi i sottoinsiemi, evitando distorsioni nella stima delle *performance*.

[68] La validazione incrociata suddivide il campione in k sottoinsiemi (*fold*): ciascun *fold* è usato come *test set* mentre gli altri costituiscono il *training set*.

[69] L'AUC ROC misura la capacità discriminativa complessiva; l'*F1-score* è indicato per classi sbilanciate; il *Brier Score* quantifica la calibrazione delle probabilità predette.

In parallelo alla validazione, è stata condotta un'analisi sistematica per capire quanto siano importanti le variabili predittive. Questo non solo aiuta ad interpretare i dati, ma permette anche di scoprire eventuali ridondanze informative. La regressione logistica è stata costruita sui coefficienti normalizzati, i quali offrono una visione chiara dell'influenza marginale di ciascuna variabile sull'*output*. D'altra parte, per la *Random Forest*, è stata tenuta in cosiderazione l'importanza calcolata sulla base della riduzione dell'impurità[70]. Anche se questa metrica è meno interpretabile in termini economici, fornisce comunque indicazioni preziose sulla struttura interna del modello.
La funzione *`plot_feature_importance()`* (**Appendice G**) ha aiutato ad automatizzare la selezione e la visualizzazione delle variabili più significative. I risultati sono stati salvati sia in formato tabellare (.csv) che grafico (.png), per rendere più semplice la consultazione e l'integrazione nelle analisi future. È significativo sottolineare che identificare le *feature* rilevanti non significa necessariamente selezionarle automaticamente durante il processo di addestramento, piuttosto, è un passo preliminare per eventuali interventi di semplificazione o ottimizzazione in futuro, specialmente per quanto riguarda la riduzione della dimensionalità.
Questa fase si conclude con l'individuazione delle variabili più predittive, che saranno analizzate in dettaglio nel paragrafo 4.3.

In un contesto operativo, sviluppare nuovi modelli predittivi richiede necessariamente di confrontarsi con gli strumenti già in uso nell'azienda. Nel caso di Fin.promo.ter, l'*AD Score* è stato a lungo utilizzato come sistema interno per una valutazione preliminare del rischio. Anche se questo punteggio è stato migliorato nel tempo, presenta alcune rigidità strutturali che ne limitano l'adattamento a situazioni particolari o a cambiamenti nei dati di *input*. Per valutare l'efficacia dei modelli sviluppati, è stato organizzato un confronto diretto con l'*AD Score*, utilizzando l'analisi delle curve ROC. Questo approccio permette di mettere in evidenza eventuali differenze nella capacità discriminativa tra le varie soluzioni. La parte tecnica è stata gestita dalla funzione *`plot_ad_comparison()`* (**Appendice I**), che consente di sovrapporre le curve ROC dei modelli predittivi e dell'*AD*

[70] L'importanza è calcolata come media della riduzione dell'impurità (Gini) provocata da ciascuna variabile lungo tutti gli alberi dell'ensemble.

Score su un campione comune, rendendo più semplice l'analisi visiva e il calcolo comparato delle AUC.

Il confronto è stato realizzato solo con le osservazioni per cui erano disponibili sia *l'AD Score* che la probabilità prevista dai modelli sviluppati. Questo ha garantito un campione di valutazione omogeneo, evitando *bias* dovuti a dati mancanti o disallineamenti temporali. La curva ROC è stata scelta come metrica di confronto perché sintetizza in modo efficace il compromesso tra sensibilità e specificità, un aspetto particolarmente importante in contesti dove il costo degli errori può variare. Anche se l'obiettivo di questa fase non è quello di arrivare a conclusioni definitive, il confronto rappresenta un passo cruciale per integrare le nuove soluzioni nell'ecosistema decisionale esistente. I risultati ottenuti e le implicazioni delle eventuali differenze nelle prestazioni predittive saranno analizzati in dettaglio nel paragrafo successivo, supportati da tabelle comparative e rappresentazioni grafiche.

Tutti i risultati generati sono stati salvati nella cartella *`results`*, organizzati per modello e *target*. Nei prossimi paragrafi, ci concentreremo su un'analisi approfondita delle *performance* predittive.

4.3 Valutazione della *performance* predittiva e delle variabili

La valutazione delle *performance* predittive dei modelli adattivi che sono stati sviluppati è un passaggio fondamentale nell'analisi, poiché serve a capire quanto siano efficaci le soluzioni proposte nella classificazione anticipata delle pratiche di credito. In questo paragrafo, si offre un'analisi dettagliata, rigorosa e comparativa delle prestazioni dei modelli di regressione logistica e *random forest*, focalizzandoci su due obiettivi principali: il buon esito della pratica (T1) e la possibilità di eventi di *default* nei mesi successivi all'erogazione (T2). L'analisi si sviluppa attorno a cinque dimensioni chiave: capacità discriminante, accuratezza probabilistica, robustezza valutata tramite validazione incrociata, importanza delle variabili predittive e confronto con il *prescore* attualmente in uso.

4.3.1 Valutazione delle metriche discriminanti: ROC, AUC, *Accuracy* e *F1-score*

Il primo livello di analisi si concentra sulle metriche discriminanti, ovvero sulla capacità del modello di distinguere in modo accurato le pratiche positive da quelle negative. Per

entrambi i *target*, sono state calcolate le curve ROC e le aree sottese (AUC)[71]. Nel caso di T1, la regressione logistica ha ottenuto un AUC di 0,702, mentre la *Random Forest* ha registrato un valore di 0,684. Questi risultati, raffigurati nella Figura 5 sottostante, mostrano che entrambi i modelli offrono una separazione moderata tra le due classi, con un leggero vantaggio per il modello lineare.

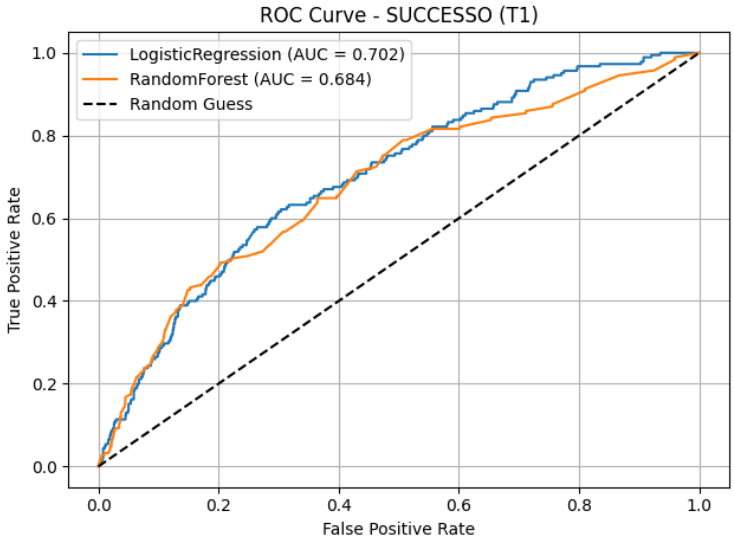


Figura 5 Curva ROC – *Target* T1: confronto tra *Logistic Regression* e *Random Forest*.

Per quanto riguarda T2, invece, la situazione si inverte (Figura 6): la *Random Forest* raggiunge un AUC di 0,661, superiore a quello della regressione logistica (0,615). Questo risultato suggerisce che, di fronte a un *target* molto sbilanciato e caratterizzato da strutture non lineari, un modello basato su *ensemble* si dimostra più efficace nel cogliere le relazioni complesse tra le variabili.

[71] La curva ROC rappresenta il *trade-off* tra tasso di veri positivi (TPR) e falsi positivi (FPR); l'AUC ne misura l'efficacia complessiva (valori prossimi a 1 indicano elevata discriminazione).

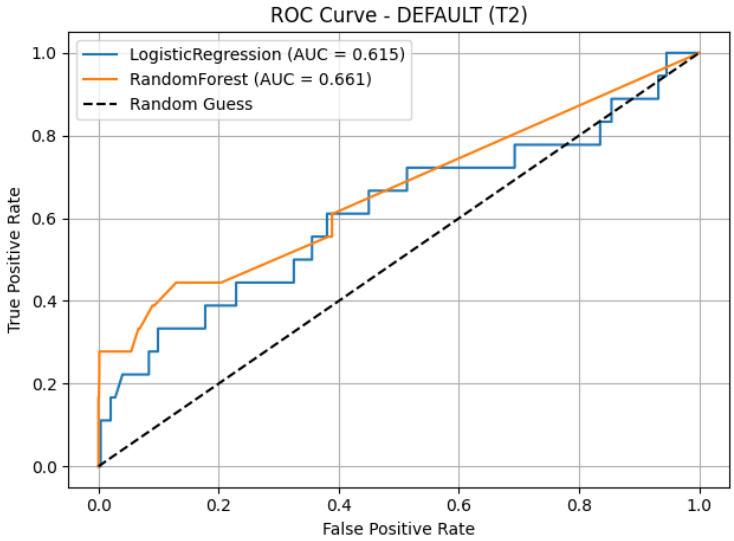


Figura 6 Curva ROC – *Target* T2: confronto tra *Logistic Regression* e *Random Forest*

Tuttavia, l'AUC è una metrica che non dipende dalla soglia di classificazione e non riflette le *performance* reali in scenari pratici. Per questo motivo, sono stati calcolati anche indicatori che dipendono dalla soglia, in particolare l'*accuracy* e l'*F1-score*[72]. Anche se l'*accuracy* risulta alta per entrambi i *target* (superiore all'86% per T1 e vicina al 99% per T2), questa metrica è fortemente influenzata dalla predominanza della classe negativa. L'*F1-score*, che bilancia precisione e *recall*, evidenzia invece le difficoltà dei modelli nel riconoscere correttamente la classe positiva: per T1, la regressione logistica ha un *F1-score* di 0,0125, mentre la *Random Forest* arriva a 0,1560. Per il *target* T2, la differenza è ancora più evidente: la regressione logistica si ferma a 0,0000, mentre la *Random Forest* raggiunge 0,2762.

Analisi delle *confusion matrix* e gestione dello sbilanciamento

L'analisi delle *confusion matrix* è fondamentale per valutare con precisione le *performance* dei modelli, in particolare per quanto riguarda la classificazione corretta e gli errori rispetto ai target T1 e T2. Questi strumenti sono essenziali per comprendere le dinamiche dei falsi positivi (FP), falsi negativi (FN), veri positivi (TP) e veri negativi (TN), mettendo in luce gli effetti reali dell'adozione di una certa soglia di classificazione in situazioni operative concrete.

[72] L'F1-*score* è la media armonica tra *precision* (accuratezza sui positivi predetti) e *recall* (capacità di individuare i positivi reali).

Per ciò che concerne il *target* T1, il quale si riferisce al buon esito della pratica, la regressione logistica ha identificato correttamente 1.152 casi di insuccesso (TN), con 3 FP e un solo vero positivo (TP), a fronte di 184 FN. Questo risultato è rappresentato nella Figura 7 sottostante, che evidenzia come il modello penalizzi notevolmente la classe positiva.

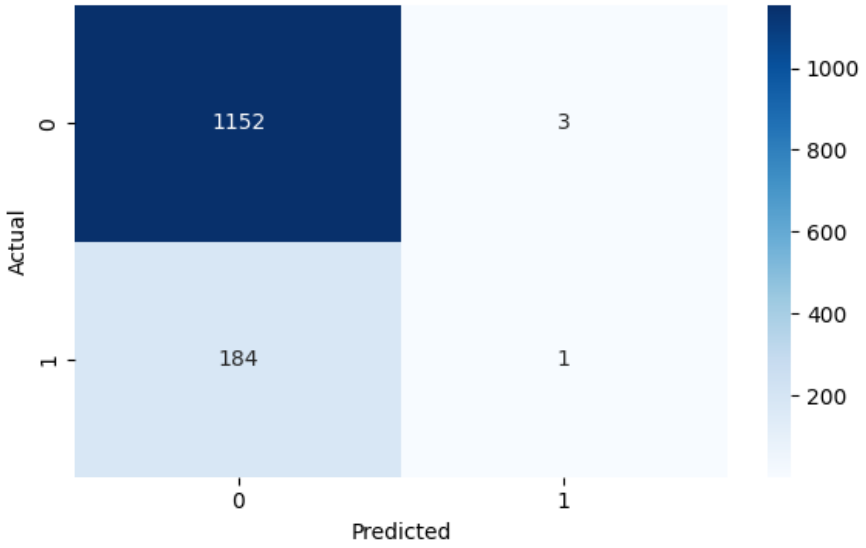


Figura 7 ***Confusion matrix – Target* T1: *Logistic Regression*.**

D'altra parte, la *Random Forest*, sebbene commetta un numero maggiore di falsi positivi (26), riesce a catturare 12 successi reali, riducendo il numero di falsi negativi a 173 (Figura 8). Anche se il miglioramento in termini di *recall* è ancora limitato in valore assoluto, è significativo dal punto di vista operativo, specialmente considerando l'alta asimmetria nei costi associati agli errori di tipo I e II.

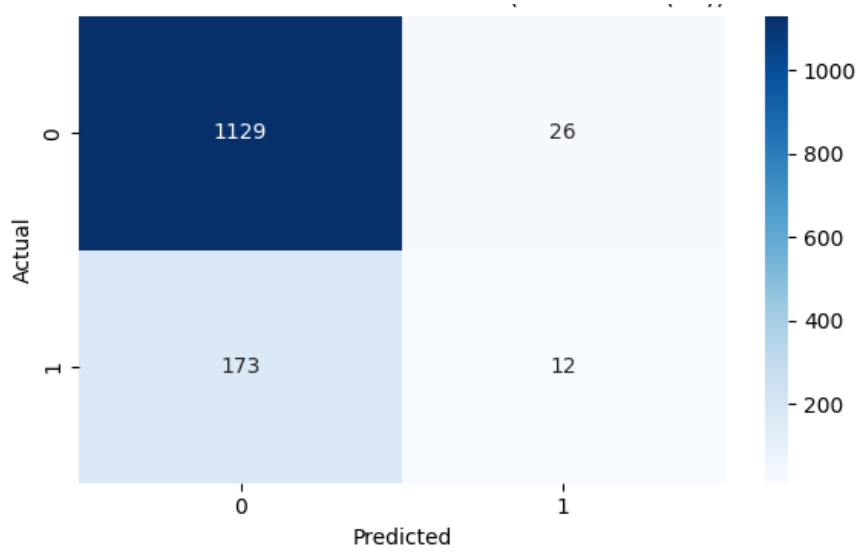


Figura 8 ***Confusion matrix – Target* T1: *Random Forest*.**

Passando al target T2, che riguarda i casi di *default* dopo l'erogazione, la regressione logistica classifica tutti i casi come negativi, producendo 1.319 TN, 3 FP, 18 FN e nessun TP (Figura 9).

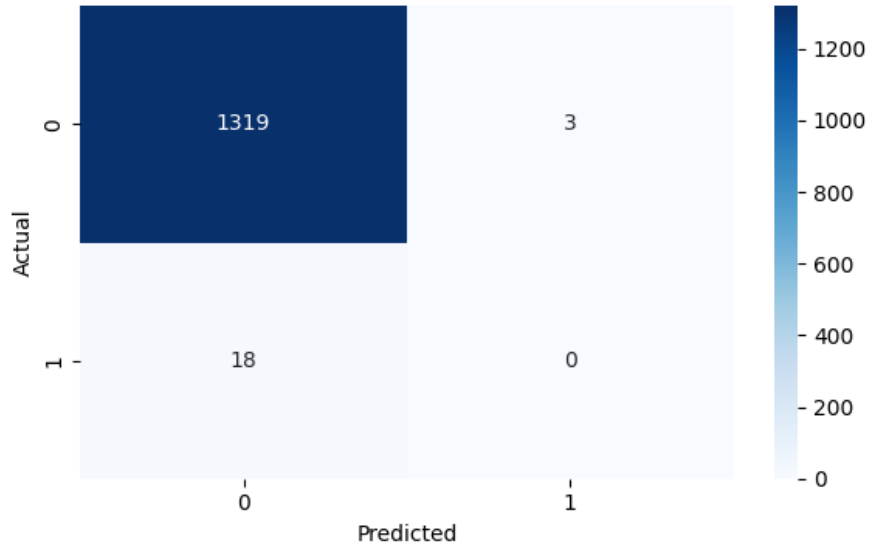


Figura 9 ***Confusion matrix – Target* T2: *Logistic Regression*.**

Questo risultato conferma l'*F1-score* nullo già osservato nelle metriche sintetiche. La *Random Forest*, nonostante operi in condizioni di forte sbilanciamento, riesce a identificare 2 veri positivi, riducendo parzialmente il numero di FN (Figura 10).

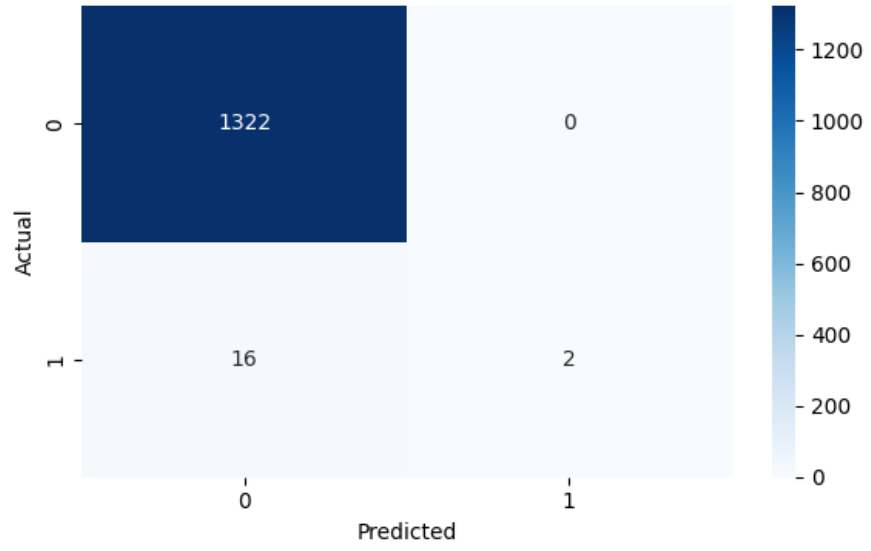


Figura 10 ***Confusion matrix – Target* T2: *Random Forest*.**

I risultati ottenuti confermano quanto la *Random Forest* sia più adattabile nella gestione di problemi di sbilanciamento, grazie alla sua abilità di modellare relazioni non lineari e di combinare decisioni su variabili diverse. Tuttavia, mettono in luce anche le sfide

strutturali che i modelli affrontano in contesti in cui la classe positiva è estremamente rara, come nel caso T2[73]. In sintesi, le *confusion matrix* offrono una visione chiara e dettagliata delle performance dei modelli. L'analisi congiunta delle Figure 7 - 10 permette di comprendere le scelte implicite dei classificatori, valutandone l'efficacia non solo in termini statistici, ma anche in un'ottica applicativa per la selezione e gestione delle pratiche.

Sulla base delle evidenze fin qui discusse, si procede ora al confronto con il sistema di *scoring* attualmente adottato, al fine di valutarne la capacità predittiva relativa rispetto ai modelli sviluppati

4.3.2 Confronto con il sistema di *scoring* attualmente in uso

Un confronto diretto è stato realizzato tra i modelli predittivi sviluppati e il sistema di *scoring* attualmente in uso dall'intermediario. In questo contesto, l'*AD Score* è una delle componenti più significative, ma non l'unica. Si tratta di un indicatore sintetico già esistente che gioca un ruolo importante nella valutazione complessiva delle pratiche. Tuttavia, il sistema attuale si basa anche su altre fonti informative e criteri sia qualitativi che quantitativi, come i dati CRIF, CGS e le valutazioni soggettive degli analisti. Pertanto, il confronto presentato in questa sede non ha l'intento di ridurre l'intero sistema di *scoring* alla sola variabile *AD Score*, ma piuttosto di utilizzare quest'ultima come punto di riferimento per misurare il valore aggiunto di un modello predittivo costruito in modo multivariato e basato sui dati.

Per quanto riguarda il *target* T1, la regressione logistica si dimostra decisamente più efficace rispetto all'*AD Score* lungo l'intero dominio della curva ROC, come illustrato di seguito nella Figura 11. Questo risultato evidenzia la capacità del modello di catturare segnali informativi distribuiti su un ampio numero di variabili, superando così le limitazioni di un indicatore sintetico.

[73] Lo sbilanciamento delle classi si verifica quando una delle due classi *target* è molto meno rappresentata dell'altra, rendendo difficile per il modello apprendere correttamente i *pattern* della classe minoritaria.

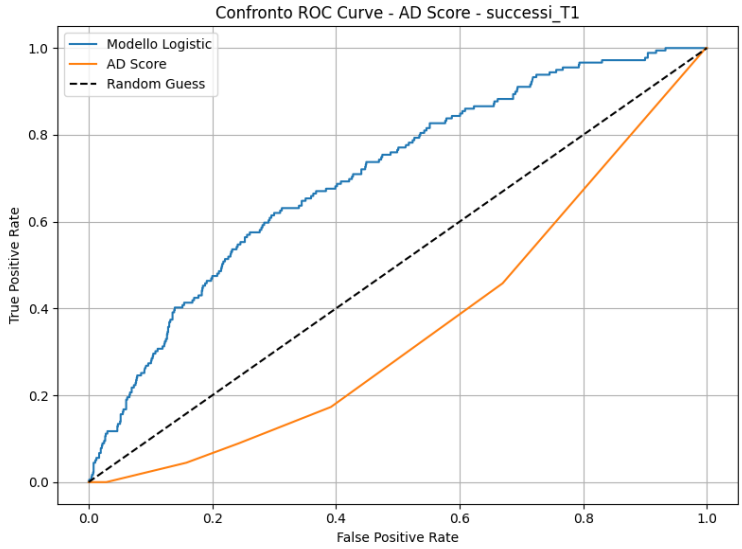


Figura 11 Curva ROC – *Target* T2: confronto tra modello *Logistic* e *AD Score*.

La rappresentazione grafica mette in risalto visivamente il guadagno netto in termini di tasso di *True Positive* a fronte di *False Positive*, suggerendo un miglioramento significativo nella qualità della selezione ex ante.

Al contempo, per il *target* T2, la *Random Forest* mostra una netta superiorità rispetto all'*AD Score*, come evidenziato dalla Figura 12.

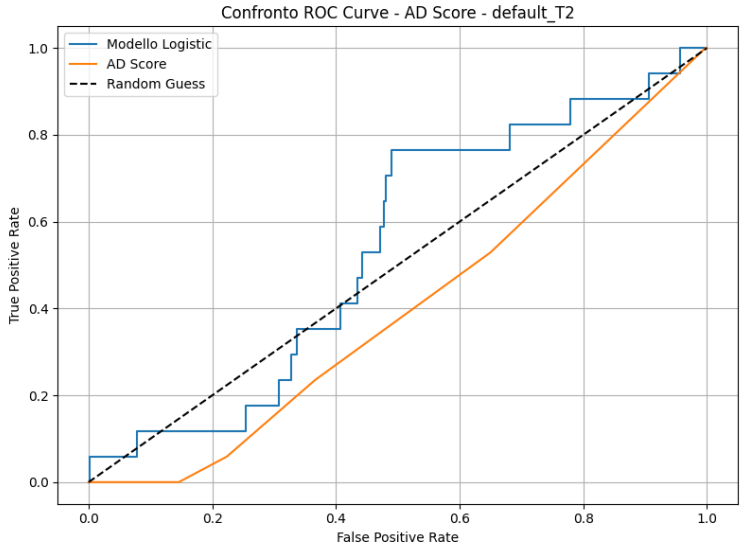


Figura 12 Curva ROC – *Target* T1: confronto tra modello *Logistic* e *AD Score*.

In questo caso, la differenza è particolarmente marcata a soglie basse, dove il tasso di individuazione dei *default* migliora notevolmente. La maggiore flessibilità della *Random Forest* nel cogliere interazioni non lineari tra variabili e la sua capacità di gestire lo sbilanciamento rendono questo confronto particolarmente significativo.

Le evidenze presentate sono supportate anche dai *file* roc_curve_confronto_successo.csv e roc_curve_confronto_default.csv, che mostrano il vantaggio in termini di TPR mantenendo costante il FPR lungo l'intero intervallo di soglie. In sostanza, pur riconoscendo che il sistema di *scoring* operativo si basi su diversi fattori, l'*AD Score* si rivela un utile punto di partenza per un'analisi comparativa. Le analisi dimostrano che i modelli adattivi, che utilizzano tecniche di *machine learning* supervisionato, offrono prestazioni costantemente superiori rispetto a questa metrica, sfruttando in modo più efficace l'intero patrimonio informativo a disposizione.

4.3.3 Affidabilità delle probabilità predette

Oltre alla capacità di discriminazione, è stata sottoposta ad analisi anche la qualità delle probabilità previste, cioè quanto le stime fornite dai modelli rispecchino realmente la probabilità empirica degli eventi. Per fare questo, si è resa necessaria la stima della curva di calibrazione per la *Random Forest* applicata al *target* T2 (Figura 13). La curva mostra una buona aderenza alla retta ideale nei *bin* centrali (con intervalli di probabilità predetta tra 0,3 e 0,6), mentre presenta scostamenti più evidenti alle estremità, probabilmente a causa della scarsa numerosità di osservazioni in quelle fasce estreme.

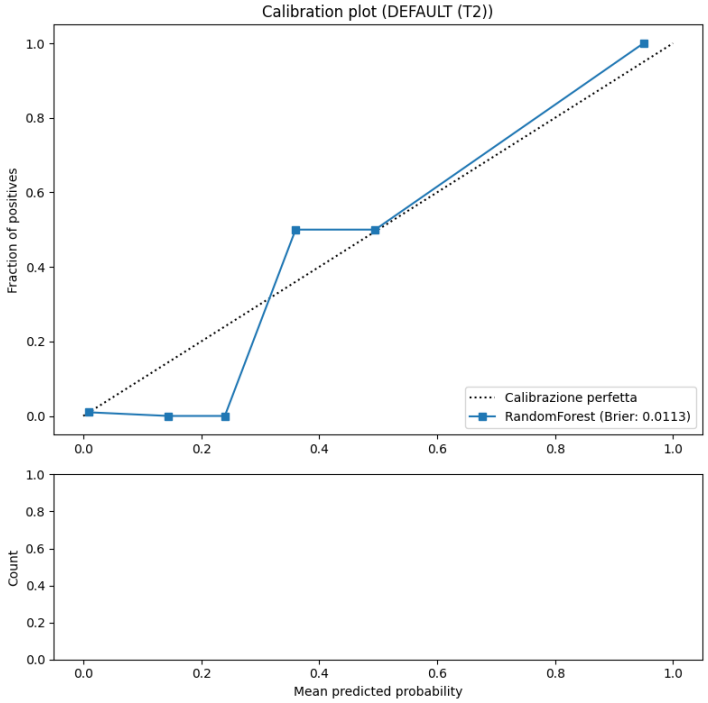


Figura 13 Curva di calibrazione – *Target* T2: *Random Forest*

Dal punto di vista interpretativo, questo suggerisce che, per le osservazioni nelle fasce centrali, la *Random Forest* riesce a fornire probabilità calibrate, cioè vicine alla frequenza osservata degli esiti. In altre parole, quando il modello assegna una probabilità del 40% a un certo evento, questo tende a verificarsi effettivamente in circa il 40% dei casi. Questa coerenza è cruciale nei contesti decisionali in cui si desidera intervenire con soglie flessibili, adattate a specifiche condizioni di rischio o vincoli di portafoglio.

Il *Brier score*[74], che è pari a 0,0113, conferma quantitativamente la buona calibratura delle probabilità previste. Tale metrica, che misura la distanza quadratica media tra le probabilità stimate e i valori binari osservati, penalizza le previsioni errate che si discostano di più da 0 o 1. Un valore basso indica una distribuzione predittiva ben centrata, il che è particolarmente utile in ambito operativo, dove le soglie di intervento possono variare nel tempo o in base a *policy* interne.

A differenza dell'AUC, che misura la capacità discriminante a prescindere dalle soglie, il *Brier score* integra sia la calibrazione sia la discriminazione, rappresentando quindi un indicatore più completo quando l'obiettivo non è solo classificare correttamente, ma anche fornire stime probabilistiche attendibili. In particolare, per le attività di monitoraggio, *pricing* del rischio e *stress testing*, la qualità probabilistica della previsione può risultare determinante per la robustezza complessiva del sistema di valutazione.

Per assicurarsi che i risultati ottenuti non fossero influenzati dalla specifica suddivisione dei dati di *training* e *test*, è stata adottata una procedura di validazione incrociata stratificata a 5 *fold*. Questo metodo permette di valutare quanto i modelli siano generalizzabili e stabili anche di fronte a campioni estratti in modo casuale. I risultati medi, presentati nella Tabella 12, confermano la tendenza già osservata nei *test set*: la *Random Forest* supera la regressione logistica in tutti gli scenari, con un vantaggio particolarmente evidente nel caso T2.

Target	Modello	AUC (*test*)	*Accuracy* (*test*)	F1-*score* (*test*)	AUC (CV)	F1-*score* (CV)	*Brier score*
T1 – Successo	*Logistic Regression*	0,702	0,8605	0,0125	0,6813 ± 0,0112	0,0125 ± 0,0154	–

[74] Il *Brier score* misura la distanza quadratica tra le probabilità predette e gli esiti osservati (0 o 1): valori più bassi indicano una calibrazione migliore.

T1 – Successo	*Random Forest*	0,684	0,8551	0,1560	0,7071 ± 0,0128	0,1560 ± 0,0226	–
T2 – *Default*	*Logistic Regression*	0,615	0,9857	0,0000	0,6602 ± 0,0707	0,0000 ± 0,0000	–
T2 – *Default*	*Random Forest*	0,661	0,9888	0,2762	0,7949 ± 0,0682	0,2762 ± 0,1021	0,0113

Tabella 12 Metriche di *performance* e validazione incrociata per i modelli predittivi

Entrando nei dettagli, per T2 la *Random Forest* raggiunge un AUC medio di 0,7949 con una deviazione *standard*[75] di 0,0682, rispetto a 0,6602 (± 0,0707) della regressione logistica. Anche l'*F1-score* medio della *Random Forest* (0,2762) è nettamente superiore a quello nullo della regressione. Per T1, la differenza è meno marcata ma comunque significativa (AUC: 0,7071 contro 0,6813; *F1-score*: 0,1560 contro 0,0125). Questi risultati non solo evidenziano una maggiore efficacia predittiva, ma anche una coerenza superiore nelle prestazioni su diversi campioni, un aspetto cruciale per l'affidabilità di un sistema di *scoring* in produzione.

In aggiunta, è stata messa in evidenza la solidità delle prestazioni del modello al variare delle configurazioni iperparametriche durante il *tuning*: la *Random Forest*, sebbene più complessa dal punto di vista computazionale, ha dimostrato una buona stabilità anche nei *run* meno performanti, riducendo il rischio di una varianza elevata tra i *fold*. Questo rafforza l'idea che il modello sia robusto e che i risultati ottenuti non siano il risultato di un *overfitting* accidentale o di una selezione casuale dei dati.

Un altro aspetto da considerare è l'analisi delle matrici di correlazione tra le previsioni nei vari *fold*, che ha messo in luce una notevole coerenza nelle assegnazioni probabilistiche delle classi. Questo indica che i modelli, in particolare la *Random Forest*, tendono a riprodurre schemi di classificazione simili anche su diversi *subset* del campione. Questo è un elemento che rafforza la replicabilità del modello in situazioni di *scoring* reali e durante gli aggiornamenti periodici del *training set*.

In generale, i risultati della *cross-validation* non solo confermano la solidità delle *performance* medie, ma forniscono anche un'indicazione sulla sostenibilità dell'uso

[75] La deviazione *standard* rappresenta la variabilità delle metriche ottenute nei diversi *fold* di validazione, ed è indicativa della stabilità del modello su campioni differenti.

operativo dei modelli in contesti dinamici, che possono subire evoluzioni nel portafoglio clienti e nelle condizioni macroeconomiche.

4.3.4 Interpretabilità del modello e importanza delle variabili

Un ulteriore aspetto esaminato è l'importanza delle variabili predittive nei modelli[76]. Le stime della loro rilevanza sono illustrate nelle Figure 14 – 17 di seguito riportate.

Per ciò che concerne il caso T1, i modelli concordano sull'importanza di variabili legate ai profitti (CERVED-EBITDA, CERVED-MOL), alla struttura (PC_7-EBITDA_Valore, PC_8-Debiti/Fatturato_Valore) e agli *scoring* esterni (CGS PD). Un aspetto particolarmente significativo è il ruolo della forma giuridica, che ha un impatto notevole nella regressione logistica. Questi risultati sono mostrati nella Figura 14 per la regressione logistica e nella Figura 16 per la *Random Forest*.

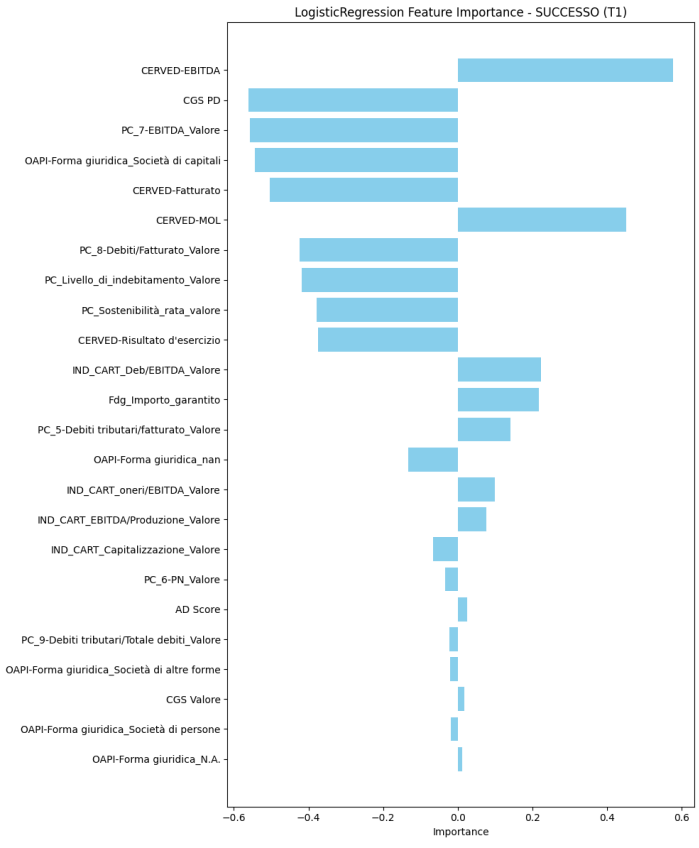


Figura 14 ***Feature importance – Target* T1: *Logistic Regression*.**

[76] La *feature importance* riflette il contributo relativo di ciascuna variabile alla predizione: nei modelli ad albero è basata sul miglioramento dell'impurezza o sulla riduzione dell'errore.

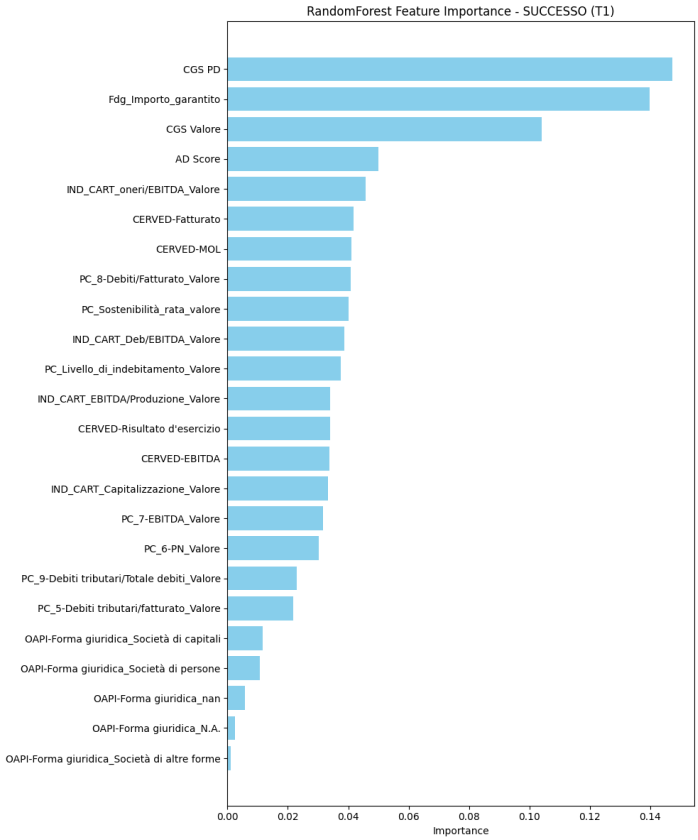


Figura 16 ***Feature importance – Target* T1: *Random Forest.***

Nel *target* T2, la *Random Forest* mette in evidenza in modo marcato gli indicatori relativi all'indebitamento e alla sostenibilità dei rimborsi (PC_Livello_di_indebitamento_Valore, PC_Sostenibilità_rata_valore), insieme al Fdg_Importo_garantito, il che conferma la capacità del modello di integrare aspetti patrimoniali e di garanzia. La regressione logistica, in linea con quanto osservato nel T1, attribuisce importanza anche alla forma giuridica e al CGS PD, come mostrato nelle Figure 15 e 17.

Figura 15 ***Feature importance – Target* T2: *Logistic Regression***

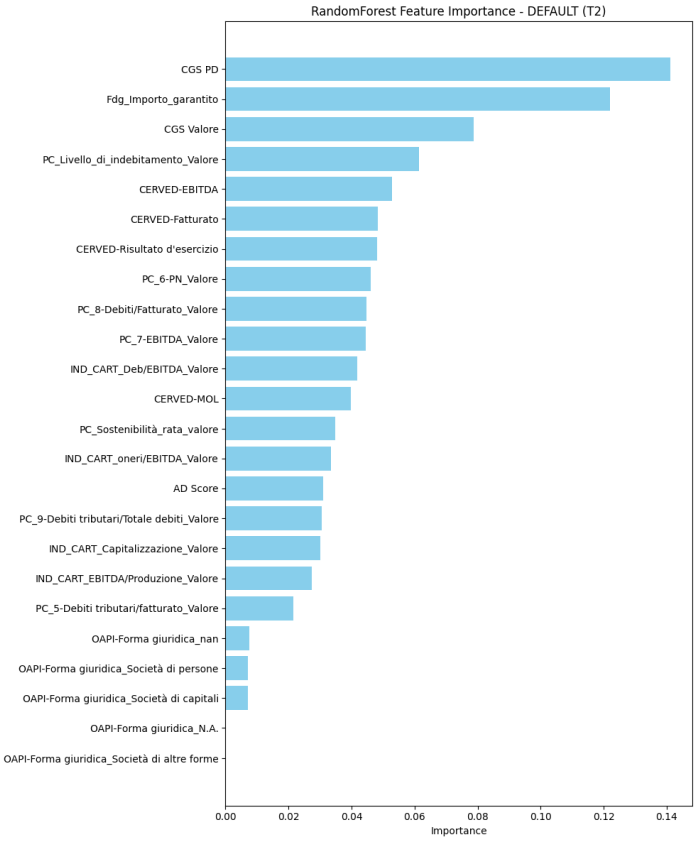


Figura 17 ***Feature importance – Target*** **T2:** ***Random Forest.***

È interessante notare che, sebbene l'*AD Score* sia presente tra le *feature*, risulta poco rilevante in entrambi i modelli. Questo suggerisce che, nel contesto multivariato, il suo contributo predittivo sia ampiamente assorbito da variabili più dettagliate e informative. Inoltre, la convergenza delle due tecniche nella selezione delle *feature* principali indica una buona coerenza tra metodi lineari e non lineari, rafforzando l'affidabilità interpretativa delle soluzioni proposte.

Combinando l'importanza delle caratteristiche con approfondimenti economico-finanziari, si riesce a migliorare la comprensione dei fattori che influenzano la classificazione. Questo è fondamentale per l'implementazione pratica del modello e per eventuali miglioramenti nelle strategie decisionali e nelle politiche interne.

Nel complesso, la valutazione effettuata mette in luce la solidità metodologica dei modelli sviluppati, la loro capacità di fornire previsioni sia accurate che ben calibrate, e la loro interpretabilità attraverso un insieme coerente di variabili significative. I risultati ottenuti nei *test* e nella validazione incrociata, insieme a un confronto ben strutturato con il sistema attualmente in uso, confermano il potenziale applicativo dell'approccio adattivo, in linea con le più recenti indicazioni della letteratura sul *credit scoring*.

4.4 Discussione dei risultati

L'analisi empirica condotta ha permesso di confrontare le *performance* predittive di due modelli di classificazione supervisionata: la *Logistic Regression* e la *Random Forest*. Questi modelli sono stati applicati a due obiettivi distinti: il successo della richiesta di credito (SUCCESSO - T1) e la possibilità di un *default* nel medio termine (*DEFAULT* - T2). I risultati sono stati esaminati non solo in termini di *performance* discriminanti, ma anche considerando l'affidabilità probabilistica, la stabilità dei modelli e l'importanza economica delle variabili predittive. Questo approccio segue le migliori pratiche suggerite nella letteratura accademica sul *credit scoring*, che sottolinea l'importanza di bilanciare accuratezza, interpretabilità e robustezza operativa[77].

Per quanto riguarda il *target* SUCCESSO (T1), la *Logistic Regression* ha raggiunto un ROC AUC di 0.6813 in *cross-validation* (con una deviazione *standard* di 0.0112), mostrando un'accuratezza dell'85.1% e un *F1-score* di 0.0125. D'altra parte, la *Random*

[77] Anderson, 2007; Lessmann et al., 2015

Forest ha mostrato un miglioramento nell'*F1-score* (0.1560) e un AUC di 0.7071, mantenendo un'accuratezza quasi invariata (85.5%). La maggiore efficacia della *Random Forest* nel rilevare i positivi è anche messa in luce dalla matrice di confusione, che mostra un numero maggiore di *True Positive* rispetto al modello lineare.

Passando al *target DEFAULT* (T2), la differenza tra i due modelli diventa ancora più evidente. La *Logistic Regression* ha ottenuto un ROC AUC di 0.6602, con un *F1-score* nullo e un'accuratezza alta ma fuorviante (98.6%), a causa dello sbilanciamento delle classi. Al contrario, la *Random Forest* ha raggiunto un AUC di 0.7949 e un F1-*score* di 0.2762, dimostrando una chiara superiorità nella capacità di identificare i soggetti a rischio. Questo risultato è in linea con quanto evidenziato in recenti studi empirici[78], i quali sottolineano l'efficacia degli *ensemble methods* nella gestione di dati ad alta eterogeneità e sbilanciamento.

Le curve ROC, sul *test set*, danno conferma che la *Random Forest* supera T2 (AUC = 0.661 contro 0.615), mentre per T1 si nota un leggero vantaggio della *Logistic Regression* (AUC = 0.702 contro 0.684). Anche se ci sono piccole variazioni tra i due modelli, questi risultati dimostrano che la scelta dell'algoritmo migliore dipende dalla natura e dall'obiettivo del *target* da prevedere. Il confronto con l'AD *Score*, attualmente utilizzato da Fin.promo.ter come una delle componenti di valutazione preliminare ed impiegato in questa analisi esclusivamente come *benchmark* comparativo, mostra un miglioramento significativo in entrambi i casi: i modelli sviluppati offrono una capacità discriminante superiore su tutto il *range* delle soglie operative.

Dal punto di vista probabilistico, la curva di calibrazione della *Random Forest* per T2 mostra una buona aderenza alla retta ideale nei *bin* centrali[79], il che indica stime ben calibrate. Il *Brier Score* di 0.0113 conferma la qualità predittiva in termini probabilistici, in linea con gli *standard* raccomandati nel *risk management*[80]. La capacità di fornire stime affidabili è fondamentale per l'adozione di modelli predittivi in contesti operativi, dove le decisioni devono riflettere soglie flessibili.

---

[78] Brown & Mues, 2012

[79] Nel contesto delle curve di calibrazione, i "*bin*" sono intervalli di probabilità (es. 0.0–0.1, 0.1–0.2, ..., 0.9–1.0) all'interno dei quali vengono raggruppate le predizioni del modello per confrontare la probabilità stimata con la frequenza osservata dell'evento. I "*bin* centrali" si riferiscono ai gruppi intermedi (es. 0.3–0.7), dove si concentra la maggior parte delle osservazioni e dove il modello dovrebbe esprimere la massima affidabilità nella stima delle probabilità.

[80] Stefanini et al., 2021

L'analisi dell'importanza delle variabili ha messo in luce un insieme di variabili sistematicamente rilevanti per entrambi i *target* e per entrambe le tecniche. La variabile "CGS - PD" si rivela la più informativa, seguita da indicatori economico-finanziari come "FdG Importo garantito", "CERVED-EBITDA", "PC_7-EBITDA_Valore" e "IND_CART_Deb/EBITDA_Valore". La convergenza tra modelli lineari e non lineari nella selezione delle *feature* conferma la solidità statistica delle informazioni identificate, suggerendo l'esistenza di determinanti strutturali comuni per l'esito e il rischio delle pratiche di credito. Questo aspetto è in linea con l'evidenza empirica che sottolinea come la stabilità delle variabili predittive sia un requisito cruciale per l'adozione sostenibile di modelli nel settore bancario[81].

Per concludere quindi, i risultati ottenuti confermano che l'adozione di modelli predittivi avanzati, in particolare *Random Forest*, potrebbe consentire a Fin.promo.ter di rafforzare significativamente la capacità di valutare in modo proattivo la qualità delle pratiche di credito, con impatti tangibili sia in termini di accuratezza che di affidabilità probabilistica. Inoltre, anche la *Logistic Regression* si dimostra un'alternativa valida e particolarmente adatta per *target* più equilibrati, grazie alla sua trasparenza, semplicità implementativa e maggiore spiegabilità. In questa cornice, si propone l'esplorazione di architetture ibride, come modelli a cascata o approcci *ensemble*, che combinino la forza predittiva degli algoritmi non lineari con la leggibilità dei modelli lineari. Tali soluzioni potrebbero garantire un compromesso efficiente tra efficacia operativa e rispetto delle esigenze normative, in linea con i principi dell'*Explainable AI*[82].

Questa riflessione finale termina il percorso di analisi tecnica del presente lavoro e conduce a considerazioni di sintesi, più ampie e trasversali, che saranno sviluppate nel capitolo seguente.

[81] Baesens et al., 2003

[82] *L'Explainable Artificial Intelligence* (XAI) si riferisce all'insieme di metodi e tecniche che rendono comprensibili e trasparenti i risultati generati da modelli complessi, facilitando l'interpretazione da parte di utenti, regolatori e decisori aziendali (cfr. Doshi-Velez & Kim, 2017).

Conclusioni

Il presente elaborato si propone di analizzare in modo critico il sistema di *prescreening* attualmente utilizzato da *Fin.promo.ter*, suggerendo un modello alternativo, adattivo e basato sui dati, capace di migliorare le performance operative e predittive. La fase di *prescore*, sebbene si collochi nelle primissime fasi del ciclo di vita di una pratica creditizia, ha un impatto significativo sull'efficienza complessiva del processo di istruttoria, influenzando direttamente la selezione delle posizioni che verranno sottoposte a un'analisi più approfondita.

Il lavoro si è sviluppato in diverse fasi. Dopo aver fornito un inquadramento teorico e procedurale del ciclo del credito, con particolare attenzione al ruolo e alle criticità del *prescore*, è stata condotta un'analisi empirica sul sistema attualmente in uso presso *Fin.promo.ter*. Simile sistema ha rivelato notevoli lacune in termini di capacità discriminante, rigidità dei filtri applicati e scarsa trasparenza nei criteri di valutazione.

Alla luce di queste evidenze, si è proceduto a costruire e testare un modello predittivo alternativo, utilizzando tecniche di apprendimento automatico supervisionato. Attraverso l'analisi di un *dataset* interno contenente dati anagrafici, qualitativi e quantitativi relativi a pratiche esaminate da *Fin.promo.ter* nel corso dell'anno 2024, sono stati sviluppati due modelli: una regressione logistica e una *Random Forest*. Questi ultimi sono stati calibrati su obiettivi differenti ma complementari: l'esito della pratica (approvazione o rifiuto) e il successivo *default*.

I risultati ottenuti mostrano chiaramente che adottare un modello adattivo può portare a miglioramenti notevoli rispetto all'approccio attuale, sia per quanto riguarda l'accuratezza predittiva che per la chiarezza nella scelta delle variabili chiave. In particolare, la *Random Forest* ha mostrato di avere una capacità discriminante superiore, mentre la regressione logistica ha offerto una maggiore interpretabilità, sottolineando l'importanza di un approccio integrato e contestualizzato.

Pur trattandosi di un caso specifico, questa ricerca potrebbe offrire spunti di riflessione potenzialmente rilevanti anche in ambito accademico.

In primo luogo, arricchisce la letteratura sui sistemi di *scoring* per gli intermediari non bancari, un settore che è ancora poco esplorato rispetto alla vasta produzione scientifica

dedicata al sistema bancario tradizionale. L'uso di modelli predittivi su dati reali durante la fase di valutazione rappresenta un approccio metodologico originale.
In secondo luogo, la tesi si inserisce nel filone di studi che esplorano l'uso delle tecniche di *machine learning* nei processi di valutazione creditizia, mettendo in luce sia i vantaggi che le sfide operative e interpretative, specialmente per gli enti regolati che devono bilanciare l'efficacia degli algoritmi con i requisiti di trasparenza e responsabilità.
In ultimo, questo lavoro contribuisce al dibattito sull'Intelligenza *Explainable AI* (XAI) nel contesto delle decisioni automatizzate, dimostrando che anche i modelli più complessi possono essere resi comprensibili attraverso strumenti analitici adeguati, facilitando così l'adozione di soluzioni basate sui dati anche in ambiti regolamentati.

Dal punto di vista manageriale, i risultati ottenuti offrono spunti operativi significativi per *Fin.promo.ter*.
Difatti, l'introduzione di un *prescore* adattivo permette una selezione più efficace delle pratiche fin dall'inizio, riducendo il carico di lavoro legato all'analisi di posizioni non meritevoli e aumentando le probabilità di successo per quelle che superano codesta analisi preliminare. Questo porta a un miglioramento dell'efficienza operativa e a una gestione più razionale delle risorse aziendali.
La possibilità di integrare variabili qualitative, anagrafiche e relazionali nel processo di valutazione aiuta a superare la rigidità di sistemi basati solo su indicatori storici, ampliando le prospettive di analisi e consentendo una personalizzazione maggiore, senza compromettere la coerenza metodologica.
Inoltre, l'uso di modelli replicabili e giustificabili può facilitare una comunicazione più fluida tra l'area operativa e le funzioni di controllo, rafforzando la *governance* interna del rischio e allineando meglio il sistema di *scoring* con le politiche decisionali.
Infine, l'adozione di un sistema adattivo e aggiornabile offre una maggiore flessibilità in risposta ai cambiamenti nel contesto macroeconomico, finanziario o normativo, aumentando la resilienza della struttura operativa di *Fin.promo.ter* nel medio termine.

Come in ogni ricerca empirica, anche questo studio ha le sue limitazioni, che ne definiscono i confini e offrono spunti per futuri approfondimenti.

In primo luogo, il *dataset* utilizzato fa riferimento a un singolo anno solare. L'assenza di una prospettiva pluriennale limita la possibilità di valutare l'efficacia predittiva del modello in presenza di *shock* macroeconomici o variazioni strutturali nel portafoglio analizzato. Un'estensione temporale futura consentirebbe di testare la stabilità delle relazioni stimate e l'evoluzione della capacità classificatoria del modello nel tempo.
In secondo luogo, la mancanza di una validazione esterna limita la generalizzabilità del modello. Anche se l'obiettivo della tesi è focalizzato sul caso di Fin.promo.ter, una futura sperimentazione su portafogli diversi o su segmenti differenti (come prodotti garantiti o categorie di clienti) potrebbe fornire ulteriori indicazioni sull'efficacia e sulla trasferibilità dell'approccio.
In terzo luogo, l'architettura del modello adottata, sebbene robusta, potrebbe essere ulteriormente sviluppata. L'introduzione di modelli *bayesiani*, reti neurali spiegabili o tecniche di apprendimento incrementale potrebbe migliorare le *performance* predittive, specialmente in contesti caratterizzati da un'alta dinamicità informativa.
La possibilità di integrare il modello in un sistema di monitoraggio continuo e aggiornamento automatico rappresenta una promettente frontiera applicativa, che permetterebbe a Fin.promo.ter di passare da un approccio di *scoring* statico a un sistema dinamico, capace di reagire in tempo reale agli eventi informativi rilevanti.

In conclusione, il percorso di ricerca presentato in questa tesi si è focalizzato sull'analisi critica e sul potenziale miglioramento del sistema di *prescreening* utilizzato da *Fin.promo.ter*, mediante l'implementazione di tecniche modellistiche avanzate. Pur senza pretendere di essere esaustivo, il presente studio ha dimostrato come un uso consapevole di strumenti predittivi, sviluppati su dati interni e adattati alle necessità operative dell'intermediario, possa costituire un valido supporto nella selezione iniziale delle pratiche di credito.
Nonostante tratti di un caso studio limitato, l'analisi proposta ha cercato di fornire un punto di riflessione concreto su come i dati e i modelli possano essere utilizzati per migliorare la qualità operativa, contribuendo a rendere l'attività quotidiana di un intermediario finanziario come *Fin.promo.ter* più consapevole, solida e reattiva.

Bibliografia

Altman, Edward I. 1968. “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy.” Journal of Finance, 589–609.

Alpert, Marc, and Howard Raiffa. 1982. A Progress Report on the Training of Decision Makers. Cambridge: Cambridge University Press.

Anderson, Raymond. 2007. The Credit Scoring Toolkit. Oxford: Oxford University Press.

Banca d'Italia. 2025. Artificial Intelligence and Relationship Lending. Gambacorta L., Sabatini F., Schiaffi S. Questioni di Economia e Finanza (Occasional Papers). Roma: Banca d’Italia.

Beaver, William H. 1967. “Financial Ratios as Predictors of Failure.” Journal of Accounting Research, 71–111.

Breiman, Leo. 2001. “Random Forests.” Machine Learning, 5–32.

Brier, Glenn W. 1950. “Verification of Forecasts Expressed in Terms of Probability.” Monthly Weather Review, 1–3.

Brown, Ian, and Cristian Mues. 2012. “An Experimental Comparison of Classification Algorithms for Imbalanced Credit Scoring Data Sets.” Expert Systems with Applications, 3446–3456.

Documenti interni Finpromoter. n.d. Analisi dei modelli di prescore e dati storici aziendali. Documentazione interna non pubblicata.

Doshi-Velez, Finale, and Been Kim. 2017. “Towards a Rigorous Science of Interpretable Machine Learning.” arXiv preprint arXiv:1702.08608.

Dun & Bradstreet. n.d. Composite Credit Scoring Models for SMEs. Dun & Bradstreet Reports.

Falkenstein, Eric G., Alfred Boral, and Lawrence V. Carty. 2000. RiskCalc for Private Companies: Moody’s Default Model. Moody’s Investors Service.

Fisher, Ronald A. 1936. “The Use of Multiple Measurements in Taxonomic Problems.” Annals of Eugenics, 179-188.

Hand, David J., and William E. Henley. 1997. “Statistical Classification Methods in Consumer Credit Scoring.” Journal of the Royal Statistical Society: Series A, 523–541.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. An Introduction to Statistical Learning. New York: Springer.

Kuhn, Max, and Kjell Johnson. 2013. Applied Predictive Modeling. New York: Springer.

Kumar, Pankaj, and Amit Singh. 2023. Machine Learning in Credit Risk Assessment. Singapore: Springer.

Lee, Kwan, and Marco Rossi. 2023. “Efficiency Gains from Automated Credit Assessment.” European Journal of Operational Research, 120–135.

Lessmann, Stephan, Bart Baesens, Hian-Ven Seow, and Lyn C. Thomas. 2015. “Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring.” European Journal of Operational Research, 621–636.

Lundberg, Scott M., and Su-In Lee. 2017. “A Unified Approach to Interpreting Model Predictions.” Advances in Neural Information Processing Systems (NeurIPS), 4765–4774.

Nguyen, Hanh, and Minh Tran. 2025. Optimizing Credit Scoring with AI Techniques. Amsterdam: Elsevier.

Niculescu-Mizil, Alexandru, and Rich Caruana. 2005. “Predicting Good Probabilities with Supervised Learning.” Proceedings of the 22nd International Conference on Machine Learning (ICML), 625–632.

Patel, Ramesh, and Amit Gupta. 2023. Optimizing Credit Evaluation Processes. London: Palgrave Macmillan.

Resti, Andrea, and Andrea Sironi. 2021. Rischio e valore nelle banche. Milano: EGEA.

Ribeiro, Marco T., Sameer Singh, and Carlos Guestrin. 2016. “Why Should I Trust You? Explaining the Predictions of Any Classifier.” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.

Thomas, Lyn C., Jonathan Crook, and David Edelman. 2002. Credit Scoring and Its Applications. Oxford: Oxford University Press.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. Deep Learning. Cambridge: MIT Press.

Chen, Tianqi, and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Zheng, Alice, and Amanda Casari. 2018. Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. Sebastopol: O'Reilly Media.

Dietterich, Thomas G. 2000. Ensemble Methods in Machine Learning. In Multiple Classifier Systems, 1–15. Springer.

Appendici

Appendice A – Tabella descrittiva delle variabili esplicative

Variabile	Fonte	Descrizione	Funzione nell'analisi
Motivazioni KO *Prescore*	*Airtable*	Testo automatico con indicazione della regola di esclusione attivata (es. KO bilancio, KO *rating*).	Classificazione qualitativa dei motivi di esclusione.
Status NEW	*Airtable*	Etichetta finale assegnata alla pratica (es. *Prescore* KO, Erogata, Da riprocessare).	Aggregazione delle pratiche per esito finale.
AD *Score*	*Cerved* API	Indicatore di rischio frode (1=basso rischio, 6=alto rischio). KO se ≥5.	Calcolo media *score* per esito, analisi per fasce.
CGS *Score*	*Cerved* API	Indicatore di probabilità di *default* (0–100). KO se <30.	Calcolo media *score* per esito, analisi per fasce.
Eurisc	CRIF	Punteggio di affidabilità creditizia rilevato da CRIF tramite sistema *Eurisc*. La scala è crescente: un valore più elevato segnala un rischio creditizio maggiore.	Classificazione preliminare per rischio esterno percepito; utile nella segmentazione dei profili a rischio elevato.
Livello di indebitamento	Bilancio *Cerved*	Rapporto tra debiti finanziari e mezzi propri.	Classificazione delle pratiche per fasce di indebitamento.
Sostenibilità rata	Bilancio *Cerved*	Capacità di rimborsare l'impegno finanziario con il MOL disponibile.	Classificazione delle pratiche in base alla capacità di sostenere la rata.

Debiti/EBITDA	Bilancio *Cerved*	Indicatore di leva finanziaria. Alto valore segnala debito eccessivo.	Valutazione del peso del debito rispetto alla redditività.
EBITDA *margin*	Bilancio *Cerved*	Margine operativo lordo sul fatturato. Sotto soglia → KO.	Analisi della redditività operativa.
Debiti/Fatturato	Bilancio *Cerved*	Quota dei debiti totali rispetto al fatturato.	Valutazione della struttura finanziaria e della leva.
Debiti tributari/Fatturato	Bilancio *Cerved*	Quota dei debiti tributari rispetto al fatturato.	Verifica della sostenibilità fiscale.
Capitalizzazione	Bilancio *Cerved*	Quota di patrimonio netto rispetto al totale attivo.	Valutazione della solidità patrimoniale.

Appendice B - Funzione *download_records_trainer()*

```
# Costruzione dei target
```

Questa sezione inizializza due variabili target binarie: T1 identifica le pratiche erogate; T2 identifica soggetti già insolventi.

```
df['T1'] = (df['Status NEW'] == '24 - Erogata').astype(int)
```

Assegna T1 = 1 a tutte le pratiche con esito '24 - Erogata'.

```
df['T2'] = df.index.get_level_values(1).isin(UNPAID_LIST).astype(int)
```

Assegna T2 = 1 se la partita IVA della pratica è inclusa nella lista di inadempienti UNPAID_LIST.

```
# Separazione e imputazione
```

Si separano le variabili numeriche e categoriche per applicare trattamenti differenti ai valori mancanti.

```
X_num = df.select_dtypes(include=['int64', 'float64'])
```

Seleziona tutte le variabili numeriche.

```
X_cat = df.select_dtypes(include='object')
```

Seleziona tutte le variabili categoriche.

```
imputer_num = SimpleImputer(strategy='mean')
```

Definisce un imputatore che sostituisce i valori mancanti con la media per le numeriche.

```
X_num_imputed = pd.DataFrame(imputer_num.fit_transform(X_num),
columns=X_num.columns)
```

Applica l'imputazione media alle variabili numeriche.

```
imputer_cat = SimpleImputer(strategy='most_frequent')
```

Definisce un imputatore che sostituisce i valori mancanti con la moda per le categoriche.

```
X_cat_imputed = pd.DataFrame(imputer_cat.fit_transform(X_cat),
columns=X_cat.columns)
```

Applica l'imputazione più frequente alle variabili categoriche.

```
# Codifica e scaling
```

Le variabili categoriche vengono codificate e tutte le feature normalizzate.

```
encoder = OneHotEncoder(drop='first', handle_unknown='ignore',
sparse_output=False)
```

Codifica le variabili categoriche usando one-hot encoding, evitando collinearità.

```
X_cat_encoded = encoder.fit_transform(X_cat_imputed)
```

Applica la codifica one-hot ai dati categorici.

```
scaler = StandardScaler()
```

Crea uno standardizzatore per portare le variabili su scala standard.

```
X_scaled = scaler.fit_transform(pd.concat([X_num_imputed,
pd.DataFrame(X_cat_encoded)], axis=1))
```

Concatena le variabili numeriche e categoriali codificate, quindi le scala.

```
# Salvataggio oggetti
```

Si salvano gli oggetti di trasformazione per poterli riutilizzare in produzione.

```
joblib.dump(imputer_num, 'models_trained/imputer_num.pkl')
```

Salva l'imputatore delle numeriche.

```
joblib.dump(imputer_cat, 'models_trained/imputer_cat.pkl')
```

Salva l'imputatore delle categoriche.

```
joblib.dump(encoder, 'models_trained/encoder_cat.pkl')
```

Salva l'encoder delle categoriche.

```
joblib.dump(scaler, 'models_trained/scaler.pkl')
```

Salva lo scaler delle feature.

Appendice C - Funzione *prepare_data_for_training()*

```
# Costruzione dei target
```

Questa sezione inizializza due variabili target binarie: T1 identifica le pratiche erogate; T2 identifica soggetti già insolventi.

```
df['T1'] = (df['Status NEW'] == '24 - Erogata').astype(int)
```

Assegna T1 = 1 a tutte le pratiche con esito '24 - Erogata'.

```
df['T2'] = df.index.get_level_values(1).isin(UNPAID_LIST).astype(int)
```

Assegna T2 = 1 se la partita IVA della pratica è inclusa nella lista di inadempienti UNPAID_LIST.

```
# Separazione e imputazione
```

Si separano le variabili numeriche e categoriche per applicare trattamenti differenti ai valori mancanti.

```
X_num = df.select_dtypes(include=['int64', 'float64'])
```

Seleziona tutte le variabili numeriche.

```
X_cat = df.select_dtypes(include='object')
```

Seleziona tutte le variabili categoriche.

```
imputer_num = SimpleImputer(strategy='mean')
```

Definisce un imputatore che sostituisce i valori mancanti con la media per le numeriche.

```
X_num_imputed = pd.DataFrame(imputer_num.fit_transform(X_num),
columns=X_num.columns)
```

Applica l'imputazione media alle variabili numeriche.

```
imputer_cat = SimpleImputer(strategy='most_frequent')
```

Definisce un imputatore che sostituisce i valori mancanti con la moda per le categoriche.

```
X_cat_imputed = pd.DataFrame(imputer_cat.fit_transform(X_cat),
columns=X_cat.columns)
```

Applica l'imputazione più frequente alle variabili categoriche.

```
# Codifica e scaling
```

Le variabili categoriche vengono codificate e tutte le feature normalizzate.

```
encoder = OneHotEncoder(drop='first', handle_unknown='ignore',
sparse_output=False)
```

Codifica le variabili categoriche usando one-hot encoding, evitando collinearità.

```
X_cat_encoded = encoder.fit_transform(X_cat_imputed)
```

Applica la codifica one-hot ai dati categorici.

```
scaler = StandardScaler()
```

Crea uno standardizzatore per portare le variabili su scala standard.

```
X_scaled = scaler.fit_transform(pd.concat([X_num_imputed,
pd.DataFrame(X_cat_encoded)], axis=1))
```

Concatena le variabili numeriche e categoriali codificate, quindi le scala.

```
# Salvataggio oggetti
```

Si salvano gli oggetti di trasformazione per poterli riutilizzare in produzione.

```
joblib.dump(imputer_num, 'models_trained/imputer_num.pkl')
```

Salva l'imputatore delle numeriche.

```
joblib.dump(imputer_cat, 'models_trained/imputer_cat.pkl')
```

Salva l'imputatore delle categoriche.

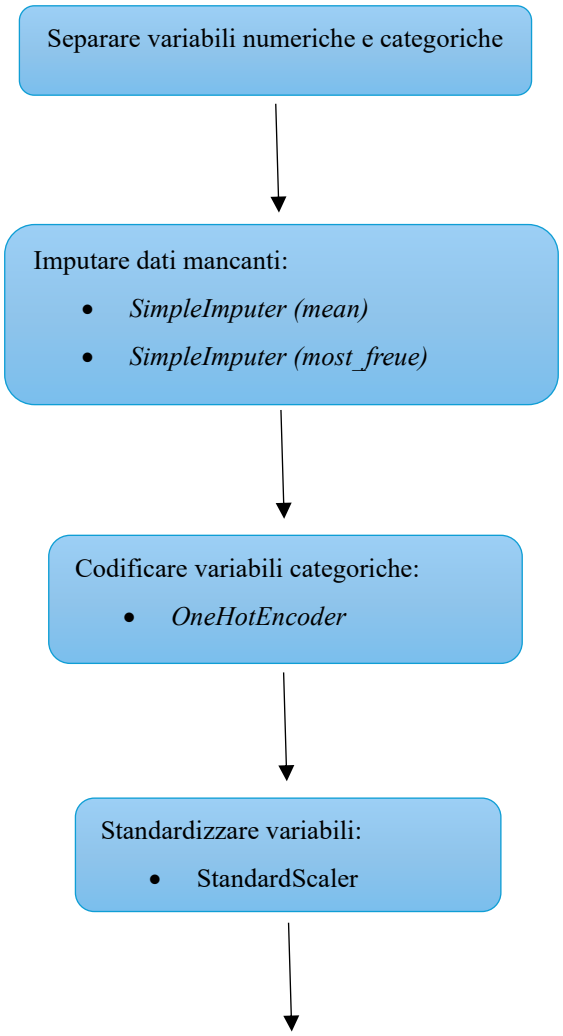
```
joblib.dump(encoder, 'models_trained/encoder_cat.pkl')
```

Salva l'encoder delle categoriche.

```
joblib.dump(scaler, 'models_trained/scaler.pkl')
```

Salva lo scaler delle feature.

Appendice D - Diagramma di flusso della *pipeline di preprocessing* dei dati



Output: *X_scaled*, T1, T2

Appendice E – Controlli di qualità e coerenza del *dataset*

```
# Identificazione delle colonne a varianza nulla
```

Rimuove le feature che non variano tra le osservazioni.

```
low_var_cols = X.var()[X.var() == 0].index.tolist()
```

Elenco delle colonne con varianza nulla.

```
# Verifica di colonne costanti o a cardinalità singola
```

Individua le colonne con un solo valore distinto.

```
constant_cols = [col for col in X.columns if X[col].nunique() <= 1]
```

Cattura feature prive di variabilità.

```
# Verifica duplicati sull'indice MultiIndex
```

Controlla se ci sono pratiche duplicate nel dataset.

```
duplicates = X.index.duplicated().sum()
```

Conta i duplicati su Lead ID e PIVA.

```
# Calcolo del bilanciamento delle classi target
```

Verifica distribuzione di T1 e T2.

```
print('Distribuzione T1:', y_T1.value_counts(normalize=True))
```

Distribuzione percentuale per T1.

```
print('Distribuzione T2:', y_T2.value_counts(normalize=True))
```

Distribuzione percentuale per T2.

```
# Analisi descrittiva delle feature numeriche
```

Riepilogo statistico di tutte le feature.

```
X.describe()
```

Genera media, deviazione standard, min, max, quartili.

Appendice F - Funzione *train_and_evaluate()*

```
# Addestramento e valutazione su base hold-out
```

```
def train_and_evaluate(X, y, target_name, save_dir):
```

Gestisce il training e la valutazione su un campione 70/30 stratificato.

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, stratify=y, random_state=42)
```

Suddivide i dati mantenendo proporzioni tra le classi.

```
for name, model in {
```

Definisce i due modelli da addestrare:

```
    'LogisticRegression': LogisticRegression(max_iter=1000,
random_state=42),
```

Regressione logistica.

```
    'RandomForest': RandomForestClassifier(random_state=42)
```

Random Forest.

```
}.items():

    model.fit(X=X_train, y=y_train)
```

Addestra il modello sul training set.

```
    joblib.dump(model, os.path.join(PATH_MODELS_TRAINED,
f'{name}_{target_name}.pkl'))
```

Salva il modello in formato pickle.

```
    y_pred = model.predict(X_test)
```

Effettua previsioni di classe sul test set.

```
    y_proba = model.predict_proba(X_test)[:, 1]
```

Calcola la probabilità per la classe positiva.

```
    acc = accuracy_score(y_test, y_pred)
```

Calcola l'accuratezza della previsione.

```
    prec = precision_score(y_test, y_pred, zero_division=0)
```

Precisione: quanto sono rilevanti i positivi predetti.

```
    rec = recall_score(y_test, y_pred, zero_division=0)
```

Recall: quanto sono catturati i veri positivi.

```
    f1 = f1_score(y_test, y_pred, zero_division=0)
```

F1-score: media armonica tra precision e recall.

```
    auc = roc_auc_score(y_test, y_proba) if len(np.unique(y_test)) > 1
else np.nan
```

AUC ROC: misura complessiva della discriminazione.

Appendice G – Funzione *plot_feature_importance()*

```
# Visualizzazione e salvataggio delle feature rilevanti

def plot_feature_importance(importance, feature_names, model_name,
target_name, top_n, save_path, csv_path):
```

Funzione per mostrare e salvare le variabili più importanti.

```
top_idx = np.argsort(importance)[::-1][:top_n]
```

Ordina le feature per importanza decrescente e seleziona le top_n.

```
top_features = [(feature_names[i], importance[i]) for i in top_idx]
```

Abbina ogni feature selezionata al suo valore di importanza.

```
df_feat = pd.DataFrame(top_features, columns=['Feature',
'Importance'])
```

Crea un DataFrame tabellare delle feature selezionate.

```
df_feat.to_csv(csv_path, index=False)
```

Esporta il contenuto in un file CSV.

```
plt.figure(figsize=(10, 6))
```

Inizializza un grafico di dimensioni 10x6.

```
plt.barh(df_feat['Feature'], df_feat['Importance'])
```

Crea un grafico a barre orizzontali delle importanze.

```
plt.gca().invert_yaxis()
```

Ordina le feature dalla più alla meno importante.

```
plt.title(f'Importanza delle feature - {model_name} ({target_name})')
```

Aggiunge un titolo dinamico al grafico.

```
plt.tight_layout()
```

Ottimizza la disposizione del grafico.

```
plt.savefig(save_path)
```

Salva il grafico come immagine.

```
plt.close()
```

Chiude la figura corrente.

Appendice H - Funzione *perform_cross_validation()*

```
# Validazione incrociata dei modelli

def perform_cross_validation(X, y, target_name, save_dir):
```

Gestisce una validazione incrociata a 5 fold stratificata.

```
models_cv = {
```

Definisce i modelli da testare:

```
    'LogisticRegression': LogisticRegression(max_iter=1000,
random_state=42),
```

Modello di regressione logistica.

```
    'RandomForest': RandomForestClassifier(random_state=42)
```

Modello Random Forest.

```
}

cv_strategy = StratifiedKFold(n_splits=5, shuffle=True,
random_state=42)
```

Strategia di validazione incrociata stratificata su 5 fold.

```
for name, model in models_cv.items():
```

Ciclo per applicare la validazione a ciascun modello.

```
    cv_scores = cross_validate(
```

Applica la validazione e raccoglie le metriche richieste.

```
        model, X, y, cv=cv_strategy,

        scoring=['roc_auc', 'accuracy', 'f1'],
```

Metriche calcolate: AUC, accuracy, F1.

```
        return_train_score=False, n_jobs=-1)
```

Non salva score sul training set, usa tutti i core disponibili.

Appendice I - Funzione *plot_ad_comparison()* e curva di calibrazione

```
# Calibrazione della probabilità predetta

if name == 'RandomForest' and target_name == 'T2':
```

Condizione per eseguire la calibrazione solo su RF e target T2.

```
brier = brier_score_loss(y_test, y_proba)
```

Calcola il Brier Score, metrica di calibrazione della probabilità.

```
prob_true, prob_pred = calibration_curve(y_test, y_proba, n_bins=10,
strategy='uniform')
```

Costruisce la curva di calibrazione su 10 bin uniformi.

```
plot_calibration_curve(prob_true, prob_pred, model_name, target_name,
brier_score=brier, save_path=...)
```

Genera e salva il grafico di calibrazione.

```
# Confronto tra modelli predittivi e AD Score
```

```
def plot_ad_comparison(ad_scores, model_probs, y_true):
```

Funzione per confrontare AD Score e modelli predittivi usando la curva ROC.

```
    fpr_ad, tpr_ad, _ = roc_curve(y_true, ad_scores)
```

Calcola la curva ROC per l'AD Score.

```
    fpr_model, tpr_model, _ = roc_curve(y_true, model_probs)
```

Calcola la curva ROC per il modello predittivo.

```
    plt.plot(fpr_ad, tpr_ad, label='AD Score')
```

Disegna la curva ROC dell'AD Score.

```
    plt.plot(fpr_model, tpr_model, label='Modello')
```

Disegna la curva ROC del modello.

```
    plt.xlabel('False Positive Rate')
```

Etichetta l'asse x.

```
    plt.ylabel('True Positive Rate')
```

Etichetta l'asse y.

```
    plt.title('Confronto ROC')
```

Titola il grafico.

```
    plt.legend()
```

Aggiunge la legenda per distinguere le curve.

```
    plt.show()
```

Mostra il grafico.