



UNIVERSITÀ
DI PAVIA

Dipartimento di Scienze Economiche e Aziendali

Corso di Laurea magistrale in

Economia e gestione delle imprese

**Predictive analytics per il rischio
di default: confronto tra modelli
tradizionali e machine learning**

Relatore:

Prof. Matteo Navaroni

**Tesi di Laurea
di Luca Tentori
Matr. n. 542365**

Anno Accademico 2024-2025

Sommario

Introduzione.....	2
Capitolo 1	5
L'approccio tradizionale alla previsione del rischio di default.....	5
1.1 L'analisi di bilancio come strumento di previsione.....	5
1.2 L'approccio statistico tradizionale.....	9
1.2.1 L'analisi univariata di Beaver.....	10
1.2.2 L'analisi discriminante multivariata: lo Z-Score di Altman.....	11
1.2.3 La regressione logistica: il modello di Ohlson.....	16
1.3 L'approccio strutturale.....	24
1.3.1 Il modello di Merton.....	25
1.3.2 Il modello KMV.....	29
1.4 L'evoluzione verso i modelli Hazard: il contributo di Shumway	34
Capitolo 2	39
L'approccio basato su Intelligenza Artificiale e Machine Learning.....	39
2.1 Introduzione al Machine Learning	39
2.2 Il machine learning applicato alla previsione del default.....	42
2.2.1 I vincoli normativi.....	43
2.3 I principali algoritmi per la previsione del rischio default	44
2.3.1 La Random Forest	46
2.3.2 L'Extreme Gradient Boosting.....	53
2.3.3 Le Reti Neurali Artificiali	57
2.4 L'Explainable AI	63
Capitolo 3	66
Analisi empirica.....	66
3.1 Introduzione all'analisi.....	66
3.2 Metodologia.....	67
3.2.1 Descrizione del dataset.....	67
3.2.2 Data pre-processing.....	70
3.2.3 Metriche di valutazione	72
3.3 Altman Z'-Score (1983).....	74

3.4	Random Forest	81
3.5	XGBoost.....	87
3.6	Valutazione comparata delle prestazioni dei modelli	92
	Conclusione	95
	Bibliografia.....	97

Introduzione

Il presente elaborato si prefigge di analizzare le metodologie di previsione del rischio di default aziendale, confrontando l'approccio tradizionale, storicamente basato su analisi di bilancio e metodi statistici, con il più recente approccio fondato su algoritmi di Machine Learning. L'obiettivo è valutare l'efficacia dei diversi strumenti predittivi nella stima della probabilità di insolvenza, comprendendo punti di forza e limiti dei metodi più innovativi rispetto a quelli consolidati.

Con il termine default si intende il caso di inadempimento degli obblighi contrattuali derivanti dal debito, con riferimento al mancato pagamento, totale o parziale, di quanto dovuto al creditore. Tale situazione delinea uno stato di insolvenza o di distress aziendale, ossia un deterioramento della condizione economica, patrimoniale e finanziaria del debitore, tale da compromettere la sua capacità di eseguire la prestazione dovuta nei termini fissati dalle parti.

È evidente come un'ipotesi di default non comporti implicazioni unicamente sull'impresa interessata. Quest'ultima interagisce con una pluralità di individui e organizzazioni, gli stakeholder, ossia i portatori di interessi nei confronti di essa. Risulta essere pertanto fondamentale riuscire a prevedere ex ante un evento che possa ledere gli interessi di una molteplicità di attori, economici e non, e che, in alcuni casi, può portare a scenari catastrofici. Un caso emblematico è rappresentato dalla crisi finanziaria del 2008, dovuta in parte alla crisi dei mutui subprime (finanziamenti concessi ad attori con un profilo finanziario sotto gli standard di credit score richiesti usualmente) che ha portato al fallimento di Lehman Brothers, importante istituto di credito statunitense. Tale default, in un contesto di economia globale fortemente interconnessa, ha generato un effetto domino su altre imprese provocando una forte recessione su scala mondiale.

Ne consegue che l'esistenza e l'utilizzo di modelli per la stima della probabilità del default assumono particolare rilevanza non solo per le imprese, ma anche per gli investitori, i quali fanno largo uso di tali modelli per la gestione dei portafogli, per la valutazione dei titoli obbligazionari e per la migrazione tra diverse classi di rating creditizio. D'altra parte, anche le autorità sono interessate al monitoraggio del rischio di default, per via delle ripercussioni che potrebbe avere sull'economia, come il caso di

Lehman Brothers sopra citato. Infine, tale capacità di previsione è di fondamentale importanza nell'ambito bancario, facilitando per le banche la selezione dei potenziali mutuatari, la valutazione delle condizioni dei nuovi prestiti e la gestione dei rischi derivanti dalle attività di credito, oltre alla necessità di far fronte alle esigenze di compliance legate in particolare al quadro regolamentare definito dal Comitato di Basilea per la Vigilanza Bancaria (BCBS). Infatti, dalla pubblicazione degli accordi di Basilea II e dal più recente Basilea III, i modelli di previsione del default sono diventati molto popolari nel settore bancario.

Nel primo capitolo verrà trattato l'approccio tradizionale alla previsione del rischio di default, partendo dalla classica analisi di bilancio come strumento di previsione, passando per i principali modelli statistici che sono stati sviluppati dagli anni '60 del secolo scorso ad oggi: l'approccio univariato di Beaver, l'approccio multivariato di E. Altman con il modello Z-Score e l'utilizzo della regressione logistica con il modello di Ohlson. Si passerà quindi all'approccio strutturale con il modello di Merton e il modello KMV, che ne rappresenta un'implementazione empirica. Infine, verrà introdotto il modello *hazard* di Shumway, che costituisce un'evoluzione dei metodi statistici tradizionali.

Nel secondo capitolo verranno trattati i modelli basati su algoritmi di Machine Learning. L'esponentiale crescita della disponibilità di dati (c.d. Big Data) e il continuo progresso tecnologico correlato al paradigma dell'Intelligenza Artificiale, di cui il Machine Learning costituisce una branca, stanno spingendo un numero sempre maggiore di intermediari finanziari, ricercatori e autorità di vigilanza a considerare questi nuovi modelli nella previsione del rischio di default, in quanto potrebbero rappresentare una potenziale alternativa rispetto ai metodi statistici tradizionali.

Verranno analizzati i tre algoritmi più rilevanti che stanno trovando applicazione nell'analisi predittiva: XGBoost, Random Forest e le Reti Neurali. Tali modelli di Machine Learning, a differenza di quelli tradizionali, sono in grado di gestire relazioni complesse e non lineari tra le variabili su grandi quantità di dati, oltre alla capacità di eseguire analisi altamente accurate. A far da contraltare a questi vantaggi, tuttavia, vi è il problema della bassa interpretabilità dei modelli derivante dalla complessità strutturale degli algoritmi, criticità che risulta essere esacerbata dai requisiti di conformità previsti dagli accordi di Basilea e dai requisiti normativi introdotti dal recente AI Act dell'Unione

Europea. Per affrontare questo trade-off, gli attori ricorrono a metodologie di Explainable AI (XAI) con l'obiettivo di migliorare trasparenza e comprensione dei modelli cercando di mantenere l'elevata accuratezza che li contraddistingue.

Nel terzo capitolo verrà condotta l'analisi empirica. Verrà analizzato un dataset reale di 6911 imprese italiane nel periodo compreso tra il 2016 e il 2019, ottenuto tramite la banca dati Orbis Global, per confrontare le prestazioni dei modelli.

Nello specifico, il campione si costituisce di PMI del settore manifatturiero sotto forma di S.r.l e S.p.A. e il dataset comprenderà sia imprese fallite sia imprese sane. Verrà testato il modello tradizionale di Altman, lo Z'-Score del 1983, specifico per le imprese non quotate, per poi confrontarlo con due modelli di Machine Learning, la Random Forest e XGBoost.

Per effettuare tale valutazione, saranno utilizzate una serie di metriche standardizzate quali matrice di confusione, accuratezza, precision, recall, F1-Score e Area Under the Curve (AUC) della curva ROC.

Capitolo 1

L'approccio tradizionale alla previsione del rischio di default

1.1 L'analisi di bilancio come strumento di previsione

L'insieme delle informazioni che deriva dall'analisi di bilancio costituisce l'elemento principale su cui si fonda l'analisi predittiva del rischio di insolvenza. Il fallimento di un'impresa non rappresenta un evento improvviso, ma può essere considerato come il punto di arrivo di un processo di progressivo deterioramento delle condizioni di equilibrio economico e finanziario dell'impresa.¹

La crisi d'impresa si configura come una fase antecedente rispetto allo stato di insolvenza, ed è caratterizzata da squilibri economico-finanziari più o meno persistenti, che possono compromettere la prospettiva di continuità aziendale. Sovente risulta riconducibile all'incapacità dell'imprenditore o del management di prevedere determinati scenari o di reagire ad essi con azioni tempestive ed efficaci.

Essa non si manifesta in modo immediato, ma con una certa gradualità, dando luogo a uno stato latente della crisi che, in modi e tempi differenti e in relazione alle diverse fattispecie, emerge in una forma acuta rendendo necessari interventi di carattere straordinario. Quanto più ampio risulta il lasso temporale che intercorre tra la percezione dei primi segnali di difficoltà e l'implementazione di misure di risanamento, tanto maggiori ed evidenti saranno gli effetti della crisi. In assenza di un intervento adeguato, la situazione tende a divenire cronica e, spesso, irreversibile.²

Al fine di evitare la realizzazione di tale scenario, la prevenzione della crisi assume un'importanza fondamentale. In Italia, l'evoluzione normativa ha portato il Legislatore a passare dalla Legge Fallimentare del 1942, ritenuta ampiamente inadeguata nel far fronte alle esigenze attuali e alla complessità dell'economia moderna, al nuovo Codice della Crisi d'Impresa e dell'Insolvenza (CCII) del 2019, dove uno dei pilastri di questa riforma

¹ Quagli, A. (2016). Il concetto di crisi d'impresa come incontro tra la prospettiva aziendale e quella giuridica. *Crisi d'impresa e fallimento*.

² Dell'Atti, A. (2023). *La gestione delle crisi aziendali: cause, strategie e piani di risanamento*. Giappichelli.

è rappresentato dalla capacità di individuare e interpretare preventivamente i segnali della crisi.

Il nuovo assetto normativo sancisce così il superamento dell'impostazione *backward looking*. Infatti, si riconosce come l'esperienza passata cristallizzata nei prospetti contabili sia insufficiente in quanto non consente di integrare le aspettative circa l'evoluzione futura dell'azienda.

Si afferma inoltre una maggior responsabilizzazione degli attori coinvolti, prescrivendo l'obbligo per l'imprenditore di dotarsi di strumenti adeguati alla tempestiva rilevazione dello stato di crisi e per l'assunzione di idonee iniziative.

Il testo originario del CCII prevedeva la costituzione di indicatori di crisi volti a prevenire squilibri di carattere reddituale, patrimoniale o finanziario, rilevabili attraverso appositi indici che evidenziassero la sostenibilità aziendale dei debiti per almeno i sei mesi successivi e le prospettive di continuità per l'esercizio in corso. In tal senso la norma specificava che gli indicatori sarebbero stati elaborati dal Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili (CNDCEC), selezionando quelli che, valutati unitariamente, risultassero maggiormente significativi ai fini della previsione della crisi d'impresa.

Dal punto di vista logico il sistema previsto dal CNDCEC è gerarchico e l'applicazione degli indici avviene secondo una precisa sequenza. In primo luogo, si considera il patrimonio netto dell'impresa: se il valore è negativo sussiste una ragionevole presunzione dello stato di crisi, altrimenti si procede calcolando il *Debt Service Coverage Ratio* (DSCR). L'indice misura la sostenibilità del debito finanziario, intesa come la capacità a sei mesi di onorare i propri debiti mediante i flussi di cassa liberi previsti nello stesso arco temporale. Un rapporto inferiore alla soglia dell'unità denota una ragionevole presunzione dello stato di crisi d'impresa, ovvero l'incapacità di quest'ultima di far fronte al debito con i flussi di cassa, evidenziando una possibile difficoltà nell'adempimento delle proprie obbligazioni.

Il DSCR è utilizzabile solo in presenza di dati prognostici ritenuti affidabili dagli organi di controllo secondo il loro giudizio professionale. In mancanza di tale attendibilità, si prevede il passaggio all'analisi di cinque indici settoriali:

- Indice di sostenibilità degli oneri finanziari, in termini di rapporto tra gli oneri finanziari ed il fatturato;
- indice di adeguatezza patrimoniale, in termini di rapporto tra patrimonio netto e debiti totali;
- indice di ritorno liquido dell'attivo, in termini di rapporto da cash flow e attivo;
- indice di liquidità, in termini di rapporto tra attività a breve termine e passivo a breve termine;
- indice di indebitamento previdenziale e tributario, in termini di rapporto tra l'indebitamento previdenziale e tributario e l'attivo.

I cinque indici in esame acquisiscono significatività solo se valutati congiuntamente, in quanto, se presi singolarmente, forniscono solo viste parziali di eventuali indizi di crisi. La valutazione unitaria prescritta dal Legislatore richiede perciò il contestuale superamento di tutte le cinque soglie stabilite per tali indici.³

Nonostante la prematura scomparsa degli indicatori della crisi con l'approvazione del decreto legislativo nel 2022, si ritiene che il calcolo ed il costante monitoraggio di tali indicatori possa rappresentare un efficace strumento di controllo del proprio stato di salute e di allerta anticipata rispetto a situazioni di crisi.

Tra le metodologie più efficaci per la prevenzione della crisi rientra l'analisi basata sui flussi finanziari. Le informazioni derivanti da tale operazione sono utili per valutare l'attitudine delle varie aree aziendali a generare o assorbire liquidità e in quale misura. Più precisamente, uno strumento di previsione imprescindibile è rappresentato dal rendiconto finanziario prospettico, il quale permette di stimare le entrate e le uscite del periodo di analisi, determinando la dinamica della giacenza di cassa.

Risulta evidente come un'operazione di questo tipo richieda un costante monitoraggio, volto a verificare che le assunzioni che stanno alla base delle previsioni siano confermate, permettendo di intervenire e di affinare il modello a fronte di eventuali scostamenti.

Tuttavia, l'analisi per flussi prospettici non è esente da criticità. Nello specifico, la problematica principale risiede nella componente di soggettività che caratterizza le stime

³ Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili (CNDCEC). (2019). *Crisi d'impresa. Gli indici di allerta previsti dal Codice della crisi e dell'insolvenza*.

su cui si basa la predisposizione del rendiconto finanziario prospettico. A differenza dei dati di bilancio consuntivi, le previsioni economico-finanziarie implicite nel cash flow ne riducono il grado di attendibilità.

Si può pertanto individuare un *trade-off* tra oggettività e significatività. Tale elemento comporta la necessità di un utilizzo congiunto tra l'analisi per flussi prospettici, ritenuta maggiormente significativa di quella per indici, e quest'ultima che, fondandosi su dati consuntivi, risulta meno soggettiva.

In dottrina, si sono così affermati alcuni indicatori di bilancio ritenuti particolarmente significativi al fine dell'allerta, intesa principalmente come la preservazione dell'equilibrio finanziario. Quest'ultimo elemento trova riscontro anche nelle Linee Guida EBA-GL LOM (*Guidelines on loan origination and monitoring*), le quali impongono una visione *forward-looking* e un monitoraggio continuo delle dinamiche aziendali, tra cui rientra la determinazione di una serie di KPIs, molti dei quali adottabili anche per la prevenzione.

È necessario precisare che non tutti i tradizionali indicatori di bilancio sono ritenuti significativi al fine della prevenzione e, quindi, occorrerà valutare quali siano adatti allo specifico scopo conoscitivo, anche nell'ottica della costruzione di sistemi di analisi più complessi rispetto al singolo indicatore. Nella prassi, gli indici di bilancio vengono preferibilmente adottati congiuntamente, anche al fine di creare un sistema di prevenzione, inserito in un contesto più ampio di misurazione delle prestazioni aziendali.⁴

Tra gli indicatori più rilevanti, il DSCR risulta uno strumento molto utile, in quanto analizza in modo dinamico e prospettico la capacità di rimborso del debito di un'impresa. Esso è ampiamente utilizzato dagli istituti di credito, anche in ottemperanza a requisiti sempre più stringenti nella valutazione del credito, così come previsto dai principi contabili internazionali.⁵ Oltre a questo, vi sono ulteriori indicatori utili a intercettare i segnali di crisi, derivati dalla prassi professionale e dall'esperienza bancaria. Un primo indicatore è il rapporto tra la Posizione Finanziaria Netta e l'EBITDA, che permette di

⁴ Navaroni, M., & Panelli, M. (2024, dicembre 23). L'analisi di bilancio applicata all'impresa in crisi. *Diritto della Crisi*.

⁵ Con riferimento all'*International Financial Reporting Standards* n°9 (IFRS 9), il quale prescrive, per la valutazione del credito, un nuovo approccio basato sulle perdite attese (*Expected Credit Loss, ECL*), a fronte di quello basato sulle perdite *incurred*.

stimare in quanti anni l'impresa sarebbe in grado di ripagare i propri debiti finanziari, nel caso essa utilizzasse, al servizio del debito, tutti i flussi derivanti dall'attività caratteristica, oltre all'attuale disponibilità di risorse liquide. Un ulteriore strumento è rappresentato dal *debt-to-equity ratio* (DER), noto anche come indice di adeguatezza patrimoniale, che indica il rapporto tra i debiti e i mezzi propri dell'impresa. In ogni caso, l'analista ha a disposizione una pluralità di strumenti volti a valutare lo stato di salute dell'impresa, potendo integrare l'analisi con altri indicatori ritenuti significativi per il caso specifico.

1.2 L'approccio statistico tradizionale

L'uso degli indici di bilancio, quali indicatori dello stato di salute di un'impresa, è sempre stato soggetto a numerose critiche per la pluralità di distorsioni che potrebbero interessare l'informazione in essi contenuta. Nonostante questo evidente limite, diversi accademici hanno dimostrato empiricamente una certa capacità di taluni *ratios* di prevedere il rischio di default di un'impresa.⁶

I primi tentativi riguardanti l'utilizzo dei quozienti di bilancio a fini predittivi possono essere fatti risalire al periodo immediatamente successivo al crollo dell'indice di Wall Street, avvenuto nel 1929. Le prime analisi empiriche si basarono sulla distinta osservazione dei vari indicatori finanziari delle imprese negli anni immediatamente precedenti il default.⁷ Questi studi miravano a fornire uno strumento in grado di coadiuvare gli istituti di credito nelle decisioni circa le concessioni di credito alle imprese. Nonostante la loro pubblicazione abbia funto da preludio allo sviluppo dell'argomento, questi contributi hanno avuto carattere puramente esplorativo, in quanto mancavano di una base metodologica e di modelli previsionali.

Successivamente, a partire dalla seconda metà degli anni '60, si è sviluppato a livello internazionale un filone di ricerca che si è focalizzato sul ricorso a metodologie statistiche

⁶ Appetiti, S. (1984). *L'utilizzo dell'analisi discriminativa per la previsione delle insolvenze: ipotesi e test per un'analisi dinamica. Temi di discussione n. 27.* Banca d'Italia.

⁷ Smith, F. R. (1930). *A Test Analysis of Unsuccessful Industry Companies.* University of Illinois, Bureau of Business Research, n. 31; Fitzpatrick, P. J. (1932). A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Companies. *The Certified Public Accountant*, p. 598-605, 656-662, 727-731.

rigorose.⁸ Si intendeva innanzitutto ridurre la componente di soggettività che inevitabilmente poteva caratterizzare una tradizionale analisi di bilancio, rendendo la valutazione del rischio di default più sistematica.

1.2.1 L'analisi univariata di Beaver

Uno dei primi e più influenti tentativi di utilizzare un metodo quantitativo per la previsione del fallimento di un'impresa è rappresentato dal lavoro di William H. Beaver (1966). Prima di questo studio, l'analisi di bilancio era prevalentemente descrittiva e orientata ad un commento ex-post delle performance aziendali.

Beaver, a quel tempo professore di *accounting* presso l'Università di Chicago, analizzò le capacità predittive di trenta quozienti, cercando di capire quale *ratio* fosse il migliore in termini di performance predittiva.

L'approccio dell'autore si basò su una classificazione dicotomica delle imprese, selezionando quindi due gruppi di 79 imprese statunitensi, uno composto da imprese sane e l'altro da imprese fallite. Queste vennero selezionate per settore e dimensione, cercando di garantire una certa omogeneità del campione. Per ciascun rapporto, dopo aver osservato la differenza tra le medie dei due gruppi, venne individuato un *cut-off point*, ossia un punto di soglia ottimale in grado di minimizzare la percentuale di previsioni errate. Il confronto venne infine condotto anno per anno, su un orizzonte temporale di cinque anni precedenti il fallimento.⁹

L'analisi empirica condotta da Beaver ha dato evidenza che, tra i quozienti osservati, il *cash flow to total debt ratio* è risultato quello più efficace ai fini previsionali. Con questi risultati l'autore è giunto quindi alla conclusione che la possibilità di insolvenza è strettamente legata alle condizioni di liquidità dell'impresa, e che il concetto di liquidità che offre le migliori prestazioni è quello dinamico, in quanto fondato sulla capacità aziendale di generare flussi di cassa nel tempo.¹⁰

⁸ Altman, E. I., Danovi, A., & Falini, A. (2013). La previsione dell'insolvenza: l'applicazione dello Z Score alle imprese in amministrazione straordinaria. *Bancaria*, 4, p. 24-37.

⁹ Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, 4, p. 74-111.

¹⁰ Appetiti, *op.cit.*

La metodologia adottata si basava su un approccio univariato, in base al quale ciascun quoziente veniva analizzato singolarmente. Tuttavia, un'analisi di questo tipo è soggetta a possibili interpretazioni fuorvianti e può risultare confusa. Ad esempio, un'impresa con una scarsa redditività o una debole solvibilità potrebbe essere considerata a rischio fallimento; per contro, la presenza di un livello di liquidità superiore alla media potrebbe portare a una valutazione opposta. L'ambiguità potenziale relativa alla *performance* comparata di più imprese appare dunque evidente. È proprio in questo aspetto che risiede il principale limite dell'analisi univariata, in quanto la valutazione separata dei singoli indicatori non consente di cogliere le interazioni complesse e le correlazioni esistenti tra i vari indici di bilancio.¹¹

Nonostante questo, lo studio di Beaver è stato indiscutibilmente prezioso e pionieristico, portando sempre più studiosi a condurre ricerche dedicate alla previsione del default utilizzando un *corpus* di tecniche statistiche sempre più sofisticate. Questo ha portato, negli ultimi cinquant'anni, alla costruzione di decine di modelli di previsione, alcuni dei quali troveranno un ampio apprezzamento della letteratura economico-aziendale nonché degli operatori economici per la semplicità di utilizzo che li contraddistingue.

L'impossibilità di cogliere simultaneamente le relazioni tra le variabili esplicative dei modelli univariati portò allo sviluppo di tecniche statistiche multivariate. Il contributo più celebre e apprezzato è certamente quello di Altman (1968), il quale utilizzò l'analisi discriminante combinando una serie di indici di bilancio per prevedere il default delle imprese.¹²

1.2.2 L'analisi discriminante multivariata: lo Z-Score di Altman

Edward I. Altman, economista statunitense, riconoscendo le problematiche correlate all'approccio univariato, scelse di adottare l'analisi discriminante multivariata (MDA) come tecnica statistica appropriata per il suo studio. Sviluppò in tal senso il modello *Z-Score*, la cui formulazione originale risale al 1968.

¹¹ Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), p. 589-609.

¹² Piatti, D., Cincinelli, P., & Castellani, D. (2015). Ruolo dell'Efficienza nella Previsione del Default Aziendale. *Piccola impresa Small Business*, 2, p. 53-84.

In generale, l'obiettivo dell'analisi discriminante consiste nell'individuazione di un insieme di variabili in grado di discriminare, nel miglior modo possibile, le n unità statistiche tra due o più gruppi, che risultano noti a priori. Si perviene ad una regola decisionale, c.d. regola di classificazione, funzione di un numero limitato di variabili, attraverso la quale si attribuisce ciascuna delle unità a uno dei gruppi predefiniti.

Nel contesto aziendale, con tale tecnica statistica si sarebbe in grado di distinguere, con l'impiego di quozienti di bilancio, tra imprese sane, ovvero in grado di far fronte alle proprie obbligazioni, e non.¹³

Per il suo studio Altman predispose due gruppi di imprese manifatturiere statunitensi quotate in borsa, uno di 33 imprese fallite e uno di 33 imprese sane. Per il primo gruppo raccolse i dati per il periodo 1946-1965. Per il secondo gruppo l'autore si servì dei dati di imprese scelte casualmente tra quelle ancora esistenti nel 1966, simili per dimensione e settore di provenienza a quelle del primo gruppo.

Altman scelse gli indici che sarebbero stati poi adottati dal modello sulla base di due criteri distinti: la popolarità in letteratura e la rilevanza potenziale per lo studio. Selezione quattro indici di bilancio rappresentativi della liquidità, redditività, indebitamento e solvibilità delle imprese e un quinto indice di mercato. Infine, allo scopo di valutare la probabilità di insolvenza per ogni azienda calcolò un punteggio, lo *Z-Score*, composto da una funzione discriminante delle cinque variabili, ponderate per coefficienti.¹⁴

La funzione ricavata da Altman è la seguente:

$$Z = 1,2X_1 + 1,4X_2 + 3,3X_3 + 0,6X_4 + 0,99X_5$$

Dove:

$$X_1 = \frac{\textit{Capitale circolante netto}}{\textit{Totale attività}}$$

$$X_2 = \frac{\textit{Utile non distribuito}}{\textit{Totale attività}}$$

¹³ Appetiti, *op. cit.*

¹⁴ Altman (2013), *op. cit.*

$$X_3 = \frac{EBIT}{Totale\ attività}$$

$$X_4 = \frac{Patrimonio\ netto}{Totale\ passività}$$

$$X_5 = \frac{Ricavi}{Totale\ attività}$$

La prima applicazione del modello mostrò come le imprese con un punteggio inferiore a 1,8 presentassero un rischio molto elevato e quindi fossero ad elevata probabilità di default e, d'altra parte, quelle con uno *Z-Score* superiore a 3 fossero invece sane. I punteggi compresi tra 1,81 e 2,99 costituivano la zona grigia, a causa della suscettibilità a errori di classificazione (figura 1).

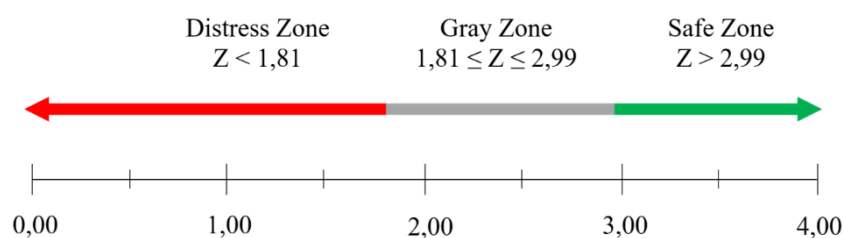


Figura 1. *Le aree di classificazione del modello.* Fonte: adattamento da Altman (2013)

Con questo studio l'autore ottenne risultati migliori rispetto a quelli ottenuti da Beaver. Infatti, il modello fu in grado di classificare correttamente il 95 per cento del campione, evidenziando un'elevata accuratezza predittiva. Gli errori di Tipo I, che si verificano quando un'impresa viene classificata come sana, pur essendo effettivamente fallita, furono solo del sei per cento. Quelli di Tipo II, relativi alle imprese classificate come fallite, ma che in realtà sono sane, furono del tre per cento. Tuttavia, le prestazioni del modello diminuivano sensibilmente dal secondo al quinto anno precedente il default mostrando il calo più marcato tra il secondo e il terzo anno.¹⁵

Lo studio fu accolto con pareri decisamente positivi da parte della comunità accademica e ricevette poche critiche. L'utilità del modello è indiscutibile a condizione che se ne

¹⁵ Appetiti, *op. cit.*

riconosca correttamente il campo d'azione quale metodo di *warning* (allarme) più che come strumento di previsione, poiché il risultato indica la maggiore vicinanza a un gruppo piuttosto che ad un altro.¹⁶ D'altro canto, è lo stesso autore, nel 1970, ad affermare che la finalità del modello consiste nell'individuazione di un andamento che accomuni le tendenze degli indicatori negli anni che precedono l'insolvenza, rispettivamente per le imprese sane e per quelle in crisi.¹⁷

La semplicità e la possibilità di utilizzo del modello senza che sia necessario disporre di particolari competenze statistiche hanno contribuito alla sua diffusione e alla sua fama. Infatti, come affermato dall'autore, «Sono sufficienti circa 30 minuti per valutare la possibilità di insolvenza di un'impresa, utilizzando semplicemente una calcolatrice e un bilancio già riclassificato».¹⁸ Questi vantaggi hanno portato lo stesso Altman a condurre nel tempo una serie di revisioni della formula.

La prima lo portò a sviluppare lo *Z'-Score* (1983), adeguando la funzione per le imprese non quotate. Non disponendo queste ultime del valore del capitale azionario, sostituì l'indice di mercato con uno di bilancio, mantenendo invariate le altre quattro variabili. I coefficienti furono poi ricalibrati e i *cut-off points* vennero aggiornati, considerando le imprese con un punteggio inferiore a 1,23 ad alto rischio di default e quelle con un punteggio superiore a 2,90 come sane. In questo caso la funzione individuata per dare luogo allo score è la seguente:¹⁹

$$Z' = 0,717X_1 + 0,847X_2 + 3,107X_3 + 0,420X_4 + 0,998X_5$$

Dove X_4 è il rapporto tra il patrimonio netto contabile e i debiti totali.

In questo modo lo studioso rese più universale il modello, rendendolo significativo anche per le PMI.

Negli anni successivi fu effettuata un'ulteriore rivisitazione della formula, spinta dalla necessità di applicare adeguatamente il modello a realtà aziendali differenti da quelle che operavano nel settore manifatturiero. Venne così introdotto lo *Z''-Score* (1995), al fine di

¹⁶ Teodori, C. (1989). *Modelli di previsione nell'analisi economico-aziendale*. Torino: Giappichelli.

¹⁷ Altman, E. I. (1970). Ratio Analysis and the Prediction of Firm Failure: A Reply. *The Journal of Finance*, 25(5), p. 1169-1172.

¹⁸ Altman, E. I., & La Fleur, J. K. (1985). I modelli di previsione delle insolvenze: le loro applicazioni alla gestione d'impresa. *Finanza, Marketing, Produzione*(4), p. 77-93.

¹⁹ Altman, E. I. (1983). *Corporate Financial Distress*. New York: Wiley Interscience.

ampliare la generalizzabilità rispetto ai modelli precedenti. La funzione venne sviluppata per le imprese non appartenenti al settore fino ad allora considerato oppure per quelle operanti nei paesi emergenti.²⁰

In tale formulazione gli autori mantennero le variabili adottate nel modello del 1983, ad eccezione del rapporto tra il fatturato e le attività totali, il quale venne eliminato, con l'obiettivo di depurare la funzione dalla possibile distorsione dovuta all'industria di riferimento. Anche in questo caso vennero aggiornati i valori dei coefficienti e furono aggiornate le soglie ottimali, considerando le imprese con un punteggio inferiore a 1,1 ad alto rischio di fallimento e quelle con un punteggio superiore a 2,6 come sane. Si definì formalmente lo *Z''-Score* secondo quanto segue:²¹

$$Z'' = 6,56X_1 + 3,26X_2 + 6,72X_3 + 1,05X_4$$

Nel calcolo del punteggio, i tre studiosi consigliarono l'aggiunta di una costante pari a 3,25, al fine di standardizzare i risultati, facendo corrispondere il valore zero allo stato di default.

Dall'applicazione del modello, Altman e Hotchkiss nel 2006 individuarono una corrispondenza tra i valori dello *Z''-Score* e le classi di rating creditizio assegnate dall'agenzia Standard & Poor's ai titoli obbligazionari (Tabella 1).²²

²⁰ Nello studio originale fu utilizzato come campione un insieme di imprese messicane.

²¹ Altman, E. I., Hartzell, J., & Peck, M. (1995). *Emerging Markets Corporate Bonds: A Scoring System*. Salomon Brothers Inc., New York.

²² Altman, E. I., & Hotchkiss, E. (2006). *Corporate Financial Distress and Bankruptcy*. New York: Wiley.

Tabella 1. *Corrispondenza tra il punteggio Z'' e il rating di Standard & Poor's*. Fonte: Altman & Hotchkiss (2006)

	Rating	Soglie	Rating	Soglie	
	AAA	> 8,15	BBB-	5,85	
	AA+	8,15	BB+	5,65	
	AA	7,60	BB	5,25	Gray
	AA-	7,30	BB-	4,95	Zone
Safe	A+	7,00	B+	4,75	
Zone	A	6,85	B	4,50	
	A-	6,65	B-	4,15	
	BBB+	6,40	CCC+	3,75	Distress
	BBB	6,25	CCC	3,20	Zone
			CCC-	2,50	
			D	< 1,75	

Un'ulteriore evoluzione del modello si ha con lo sviluppo dello Zeta Metrics System da parte di Altman e altri studiosi, in collaborazione con RiskMetrics Group. Il lavoro costituisce un'evoluzione degli studi di Altman, in quanto utilizza una metodologia differente per stimare il rating di credito e la probabilità di fallimento delle imprese. Il modello integra gli indici di bilancio unitamente alle osservazioni sulle variabili di mercato e sui dati macroeconomici, utilizzando la regressione logistica.²³ Trattandosi di una tecnica con natura proprietaria, i parametri del modello non sono esplicitati.

1.2.3 La regressione logistica: il modello di Ohlson

Il modello di regressione logistica, o *modello logit*, rappresenta un ulteriore metodo quantitativo che ha trovato applicazione nella stima della probabilità di default.

Esso consiste nella costruzione di un modello non lineare in grado di individuare le principali caratteristiche in base alle quali poter effettuare una determinata valutazione. Più precisamente si utilizza per studiare la relazione esistente tra una variabile dipendente

²³ Altman, E. I., & Rijken, H. A. (2010). *The Z-Metrics™ Methodology for estimating company credit ratings and default risk probabilities*. New York: RiskMetrics Group.

y e una o più variabili indipendenti x . La variabile y è una variabile le cui modalità rappresentano due o più alternative mutualmente esclusive.²⁴

Nel caso della previsione del fallimento, tipicamente si utilizza il modello logit per variabili dipendenti binarie o dicotomiche. In questo caso la variabile dipendente può assumere solo due valori che rappresentano le imprese sane o le imprese in default, i quali vengono codificati rispettivamente con 0 e 1:

$$y = \begin{cases} 0, & \text{impresa sana} \\ 1, & \text{impresa in default} \end{cases}$$

In questo modo si può stimare, sulla base della conoscenza dei valori assunti dalle variabili esplicative, la probabilità che si verifichi l'insolvenza.

La regressione logistica è particolarmente adatta a un'analisi di questo tipo in quanto è in grado di descrivere e testare ipotesi sulle relazioni tra una variabile di esito categoriale e una o più variabili, c.d. predittori, categoriali o continue.

Nell'ipotesi più semplice di regressione lineare con un solo predittore continuo X (a titolo esemplificativo, un indicatore economico-finanziario dell'impresa, come il *debt-to-equity ratio*) e una variabile di esito dicotomica Y (il verificarsi o meno del fallimento dell'impresa), il grafico di tali dati produrrebbe due linee parallele, ciascuna corrispondente a un valore dell'esito dicotomico. Poiché queste due linee parallele sarebbero difficili da descrivere con un'equazione di regressione ai minimi quadrati ordinari a causa della natura binaria degli esiti, si potrebbe invece procedere alla creazione di categorie per il predittore X e calcolare la media della variabile di esito Y per le rispettive categorie.²⁵

Il grafico risultante delle medie delle categorie apparirà così quasi lineare nella parte centrale, in modo simile a quanto ci si aspetterebbe di osservare in un normale grafico a dispersione, ma curvo alle estremità (figura 2).

L'andamento che si può evincere dal grafico è descritto dalla curva logistica, definita anche come sigmoide o *S-shaped curve*. Questa forma è difficile da descrivere mediante

²⁴ Neri, L. (2017). *Corso di analisi statistica per le imprese - Modello logit*. Università degli studi di Siena.

²⁵ Peng, C., Lee, K., & Ingersoll, G. (2002). An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, 96(1), p. 3-15.

un'equazione lineare per due ragioni. Da un lato, le estremità non seguono un andamento lineare; dall'altro, gli errori non sono distribuiti normalmente né sono costanti lungo l'intervallo di dati.²⁶

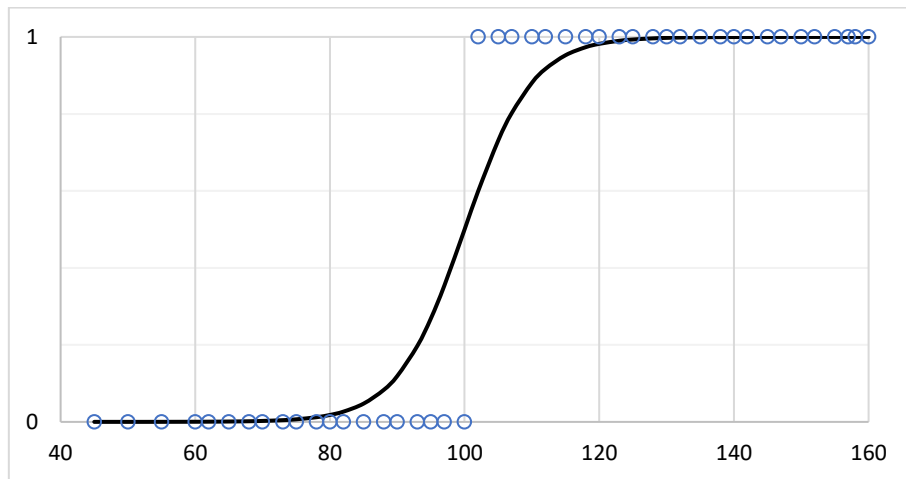


Figura 2. Relazione tra una variabile di esito dicotomica Y e un predittore continuo X nel modello logit. Fonte: elaborazione propria.

La regressione logistica risolve questi problemi applicando la “trasformazione” logit alla variabile dipendente, che può essere compiuta in due passi. In primo luogo, si passa dal concetto di probabilità p al concetto di *odds*, quest’ultimo definito come il rapporto tra la probabilità di fallimento e il suo complemento:

$$odds = \frac{p}{1 - p}$$

In secondo luogo, si calcola il logaritmo naturale dell’*odds*, che rappresenta il concetto matematico che sta alla base della regressione logistica, ovvero il logit o *log-odds*.

L’equazione del modello di regressione logistica semplice assume la seguente forma:

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

Dove:

²⁶ Peng, C., Manz, B., & Keck, J. (2001). Modeling Categorical Variables by Logistic Regression. *Journal of Health Behaviour*, 25(3), p. 278-284.

- p è la probabilità dell'evento, ovvero del default dell'impresa.
- α è l'intercetta.
- β è il coefficiente della regressione.
- X rappresenta il predittore, con riferimento alla variabile indipendente.

Applicando la funzione esponenziale a entrambi i membri, si ottiene un'equazione in grado di predire la probabilità che si verifichi l'evento:

$$p = \frac{e^{\alpha+\beta X}}{1 + e^{\alpha+\beta X}}$$

Nella prima equazione la relazione tra le due variabili è lineare, mentre nella seconda è non lineare. Da questo si può desumere la *ratio* sottesa al logit, in quanto operazione necessaria per rendere lineare la relazione tra una variabile di esito categoriale e il suo predittore.²⁷

Estendendo la logica della regressione logistica semplice a più predittori, è possibile costruire una regressione logistica complessa per Y , come segue:

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i$$

Quindi:

$$p = \frac{e^{\alpha+\beta_1 X_1+\beta_2 X_2+\dots+\beta_i X_i}}{1 + e^{\alpha+\beta_1 X_1+\beta_2 X_2+\dots+\beta_i X_i}}$$

Con $i = 1, 2, \dots, n$

Per stimare l'intercetta e i coefficienti del modello si impiega tipicamente il metodo della massima verosimiglianza (*Maximum Likelihood, ML*). Poiché le equazioni generate dalla massimizzazione della verosimiglianza non sono lineari nei parametri e non ammettono soluzione esplicita, le stime dei coefficienti si ottengono utilizzando procedure numeriche iterative.

I dati vengono inseriti nell'analisi come 0 o 1 per l'esito dicotomico, valori continui per i predittori continui e codifiche *dummy* (generalmente 0 e 1) per i predittori categoriali.

²⁷ Peng *et al.* (2002), *op. cit.*

L'interpretazione dei risultati viene svolta utilizzando l'*odds ratio* per entrambe le tipologie di predittori.²⁸

Il modello logit allenta alcune delle ipotesi restrittive richieste dall'analisi discriminante, quali la normalità multivariata e l'uguaglianza delle matrici di covarianza, e presenta il vantaggio significativo della semplice interpretazione dei risultati, poiché fornisce direttamente la probabilità di default (PD) per l'impresa analizzata. Queste caratteristiche hanno suscitato grande interesse da parte dei ricercatori e ne hanno favorito la diffusione come modello di *scoring* tra i professionisti.

Tuttavia, permangono alcune limitazioni: il modello non considera le interazioni non lineari o complesse tra variabili osservabili e il fallimento dell'impresa, la sensibilità ai valori anomali o ai dati mancanti e può essere difficile da applicare su grandi dataset.²⁹

Uno degli aspetti più gravosi è la selezione dei predittori per il modello di previsione del default, in quanto la teoria finanziaria non indica quali variabili dovrebbero essere incluse. La procedura *stepwise forward*³⁰, un metodo di ricerca ampiamente utilizzato per lo sviluppo di modelli di previsione, è utile quando il ricercatore intende valutare la relazione tra un elevato numero di variabili indipendenti per includerle nell'equazione. In questa procedura, la significatività delle statistiche di punteggio e la probabilità di una statistica del rapporto di verosimiglianza basata sulle stime di massima verosimiglianza parziale vengono utilizzate per determinare quali variabili inserire nel modello.³¹

La regressione logistica è in grado di gestire anche variabili di esito categoriali politomiche, ovvero il caso in cui la variabile dipendente può assumere più di due valori. Nell'ambito della previsione di fallimento di un'impresa, ci si è posti la domanda se fosse possibile passare dal classico approccio binario che prevede la categorizzazione delle

²⁸ Ibidem.

²⁹ Moscatelli, M., & et al. (2019). *Corporate default forecasting with machine learning. Temi di discussione n. 1256*. Banca d'Italia.

³⁰ La selezione di tipo *stepwise forward* (graduale in avanti) inizia con un modello senza predittori e procede aggiungendo un predittore alla volta, fino a quando risultano inclusi nel modello tutti quelli statisticamente significativi. Più precisamente, ad ogni passo la variabile che apporta il miglioramento più rilevante alla stima rispetto alle altre viene aggiunta al modello.

³¹ Hair, J., & et al. (1998). *Multivariate Data Analysis*. Englewood Cliffs (USA): Prentice Hall.

aziende che fanno parte del campione in imprese fallite e non fallite, all'ipotesi in cui le aziende potessero essere classificate in base a una pluralità di difficoltà finanziarie.

Alcuni ricercatori hanno svolto degli studi utilizzando quest'ultimo approccio, impiegando la regressione logistica multinomiale, al fine di determinare se l'accuratezza predittiva potesse essere migliorata attraverso un'ulteriore classificazione delle imprese in difficoltà finanziaria. Il modello logit dicotomico classifica sia gli eventi di lieve difficoltà finanziaria sia gli eventi di bancarotta e riorganizzazione aziendale come default, mentre il modello logit multinomiale distingue tali condizioni in diversi livelli di eventi di fallimento.

I risultati empirici degli studi hanno mostrato che gli errori di classificazione delle imprese del modello logit binario sono inferiori a quelli del modello multinomiale, suggerendo che l'approccio dicotomico offre delle prestazioni superiori nella previsione del fallimento. I risultati del confronto hanno mostrato che le imprese in lieve difficoltà finanziaria e quelle in bancarotta e riorganizzazione presentano caratteristiche simili.³²

Tra i primi studi più influenti che hanno utilizzato il modello logit per la previsione del fallimento vi è quello svolto da James A. Ohlson nel 1980. L'autore decise di adottare come metodologia econometrica l'analisi logit per evitare una serie di problemi associati all'analisi discriminante multivariata (MDA).

Come già noto, quest'ultimo approccio è stato la tecnica più popolare negli studi sul default che utilizzavano una pluralità di variabili esplicative, a partire dagli studi di Altman. Tuttavia, Ohlson evidenziava la presenza di una serie di problematiche.

Lo studioso criticava, in particolare, la presenza di determinati requisiti statistici imposti sulle proprietà distributive delle variabili esplicative, quali l'uguaglianza delle matrici di varianza-covarianza per entrambi i gruppi di imprese e la limitazione della possibilità di utilizzare variabili *dummy* per via del requisito di predittori normalmente distribuiti. La violazione di tali condizioni potrebbe essere poco rilevante se l'obiettivo fosse quello di sviluppare un modello discriminante. Sebbene questa possa essere considerata una valida argomentazione, è comunque chiaro che questo approccio limita l'ambito dell'indagine.

³² Tsai, B.-H. (2012). Comparison of Binary Logit Model and Multinomial Logit Model in Predicting Corporate Failure. *Better Advanced Press*, 2, p. 99-111.

In molti casi, è preferibile effettuare un'analisi econometrica più tradizionale, testando le variabili per significatività statistica e altri elementi.³³

Inoltre, l'output dell'applicazione di un modello basato sulla MDA consiste in un punteggio di difficile interpretazione intuitiva, poiché è fondamentalmente uno strumento di classificazione ordinale. Ne segue la possibile irrilevanza del punteggio per quei problemi decisionali in cui la semplice classificazione non è sufficiente a descrivere le conseguenze economiche.

Infine, vi sono dei problemi relativi alle procedure di "matching". Le imprese fallite e sane vengono abbinate secondo criteri quali dimensione e settore, e questi tenderebbero ad essere alquanto arbitrari. Ohlson sostiene l'irrilevanza di tali procedure, in quanto non porterebbero alcun contributo al modello. Si potrebbe invece utilizzare queste caratteristiche direttamente come variabili esplicative.³⁴

L'uso della regressione logistica, d'altra parte, consente di evitare sostanzialmente i problemi discussi in relazione alla MDA. Infatti, non vi è la necessità di formulare assunzioni sulle probabilità a priori di fallimento o sulla distribuzione delle variabili esplicative. La significatività statistica dei diversi predittori è ottenuta tramite la teoria asintotica (o dei grandi campioni). Ovviamente, come in qualsiasi analisi parametrica, è necessario specificare un modello, per cui esiste sempre il rischio di un'errata specificazione del modello di probabilità di base.

Per il suo studio, l'autore utilizzò osservazioni provenienti da 105 imprese fallite e 2058 imprese sane nel periodo compreso tra il 1970 e il 1976. Scelse unicamente società quotate che rientravano nel settore industriale, escludendo le *utilities* (servizi pubblici), le imprese del settore dei trasporti e quelle del settore dei servizi finanziari. Le aziende in questi settori sono strutturalmente eterogenee, presentano differenti contesti di fallimento e in alcuni casi i dati necessari per l'analisi sono difficili da reperire.³⁵ Ohlson raccolse questi dati al fine di formulare una formula finanziaria multifattoriale che diventerà famosa con la denominazione *Ohlson O-Score*.

³³ Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), p. 109-131.

³⁴ Ibidem.

³⁵ Ibidem.

La formula è il risultato di una combinazione lineare di nove fattori e si compone di una serie di indicatori aziendali ponderati per coefficienti, che possono essere agevolmente ottenuti o derivati dai bilanci resi pubblici dalle società quotate.³⁶

Di seguito è riportato il calcolo dell'*O-Score*:

$$\begin{aligned} O - Score = & -1,32 + 0,407 SIZE + 6,03 TLTA - 1,43 WCTA + 0,0757 CLCA \\ & - 2,37 NITA - 1,83 FUTL + 0,285 INTWO - 1,72 OENEG \\ & - 0,521 CHIN \end{aligned}$$

Dove:

$$SIZE = \ln\left(\frac{\text{Totale attività}}{\text{Prodotto nazionale lordo}}\right)$$

$$TLTA = \frac{\text{Totale passività}}{\text{Totale attività}}$$

$$WCTA = \frac{\text{Capitale circolante netto}}{\text{Totale attività}}$$

$$CLCA = \frac{\text{Passività correnti}}{\text{Attività correnti}}$$

OENEG = 1 se le passività totali > attività totali, altrimenti 0

$$NITA = \frac{\text{Reddito netto}}{\text{Totale attività}}$$

$$FUTL = \frac{\text{Funds from operations}}{\text{Totale passività}}$$

INTWO = 1 se reddito netto negativo nei due anni precedenti, altrimenti 0

$$CHIN = \frac{NI_t - NI_{t-1}}{|NI_t| + |NI_{t-1}|}$$

Con *NI* (Net Income) che rappresenta il reddito netto del periodo di riferimento.

Nella settima variabile, con *funds from operations* si intendono i fondi generati dalle operazioni aziendali, ottenuti dai formati storici dei prospetti finanziari disponibili

³⁶ Alam, I. (2022). *James Ohlson O-Score for Predicting Corporate Bankruptcy*. Jahangirnagar University.

all'epoca nel database *Standard & Poor's Compustat*. Tali fondi potrebbero essere oggi interpretati come un'approssimazione dei flussi di cassa operativi.³⁷

Al fine di individuare la probabilità di insolvenza di un'impresa, si applica la funzione logistica:³⁸

$$p = \frac{e^{O-Score}}{1 + e^{O-Score}}$$

Per il modello, l'accuratezza previsionale a un anno è pari al 96,12 per cento, calcolata sulla base di un *cut-off point* pari a 0,5. Dunque, l'impresa viene classificata come fallita se $P > 0,5$. Tuttavia, l'autore rileva che un punto di soglia pari a 0,038 minimizza la somma degli errori di Tipo I e Tipo II, riuscendo a classificare correttamente l'82,6 per cento delle imprese non fallite e l'87,6 per cento delle imprese fallite.³⁹

Da notare che Ohlson (1980) classifica gli errori in modo diverso rispetto ad Altman (1968, 1983). In questo caso, un errore di Tipo I si verifica quando un'impresa viene classificata come fallita, pur essendo effettivamente sana. Al contrario, le imprese classificate come sane ma che in realtà sono fallite rappresentano l'errore di Tipo II.

1.3 L'approccio strutturale

A partire dagli anni Settanta, la letteratura ha introdotto i modelli strutturali per la valutazione del rischio di default. Questo approccio si propone di descrivere, considerando le caratteristiche intrinseche di un emittente di titoli di debito, l'andamento del valore complessivo delle attività dell'impresa in esame. Tale valore è rappresentato da un processo stocastico, ovvero da un modello matematico che descrive l'evoluzione nel tempo di un fenomeno aleatorio dinamico. Di conseguenza, il valore delle attività aziendali, seguendo un processo di diffusione geometrica noto come moto browniano geometrico, può fluttuare nel tempo.

I modelli che si basano sull'approccio strutturale hanno la caratteristica di descrivere la struttura interna dell'emittente in modo tale da definire il verificarsi del default come

³⁷ Ohlson, *op. cit.*

³⁸ Mitchell, K., & Walker, M. (2008). Bankers on Boards, Financial Constraints, and Financial Distress. *SSRN Electronic Journal*.

³⁹ Ohlson, *op. cit.*

diretta conseguenza di un evento interno. Questi modelli vengono denominati *firm value models* in quanto il processo alla base del modello di valutazione è il valore dell'impresa.

Storicamente, questo approccio ha rappresentato il primo tentativo di fornire un modello quantitativo del rischio di credito con finalità di valutazione. In questo campo, il primo contributo è stato quello di Merton (1974) sulla base del modello di Black-Scholes (1973).⁴⁰

1.3.1 Il modello di Merton

Nell'ambito dei modelli in forma strutturale, il contributo di Robert C. Merton (1974) è ampiamente considerato il più celebre, in quanto pionieristico e tra i più studiati nella letteratura di riferimento. Questo modello si basa sulla teoria finanziaria, più precisamente sulla teoria delle opzioni. Queste ultime si riferiscono a una tipologia di strumenti finanziari derivati il cui valore non è autonomo, bensì deriva da quello di un bene tangibile o di natura finanziaria, che fa da riferimento e si chiama *sottostante*. Merton ipotizza che il valore dell'impresa possa essere inteso come un'opzione di acquisto su attività sottostanti.⁴¹

L'autore nel suo lavoro assume che la struttura del capitale di un'impresa sia costituita da:

- Il debito, con un ammontare pari a K , emesso sotto forma di *zero coupon bond* aventi una scadenza T e un valore al tempo t pari a $D(t, T)$;
- Il capitale proprio, sotto forma di azioni, avente valore oggi pari a $S(t)$.

Per ogni momento antecedente alla scadenza del debito, viene definito il valore totale delle attività dell'impresa V_T .⁴²

Il valore dell'*equity* e il valore dei titoli di debito sono collegati al valore dell'impresa mediante la seguente relazione:

$$V_T = S(t) + D(t, T)$$

⁴⁰ Matta, A. (2004). *Approcci matematico-statistici alla valutazione del rischio di credito*. Torino: Associazione Torino Finanza.

⁴¹ Merton, R. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, 29(2), p. 449-470.

⁴² Matta, *op. cit.*

L'assunto fondamentale nel modello di Merton è che il *default*, quindi l'insolvenza dell'impresa in relazione ai titoli di debito, possa avvenire esclusivamente alla loro scadenza. Infatti, essendo titoli privi di cedola, l'unico periodo in cui avviene il pagamento è la scadenza T .

Dalla struttura del capitale della società è possibile determinare il valore alla scadenza delle obbligazioni emesse. Da un lato, se il valore dell'impresa alla scadenza T è maggiore del debito, la società è solvente e i titoli valgono il loro valore nominale. Dall'altro, se V_T è minore del valore dei titoli di debito, allora l'impresa è insolvente e dichiara *default*. Gli obbligazionisti hanno il diritto di rivalersi sui beni della società, l'attivo, fino al raggiungimento di quanto spetta loro.⁴³

Il *payoff* alla scadenza per gli obbligazionisti sarà pertanto $D(T, T) = \min\{V_T, K\}$. Agli azionisti spetterà invece una remunerazione, eventuale e residuale, pari a $S(T) = \max\{V_T - K, 0\}$.⁴⁴ La figura 3 illustra graficamente la posizione dei due attori economici.

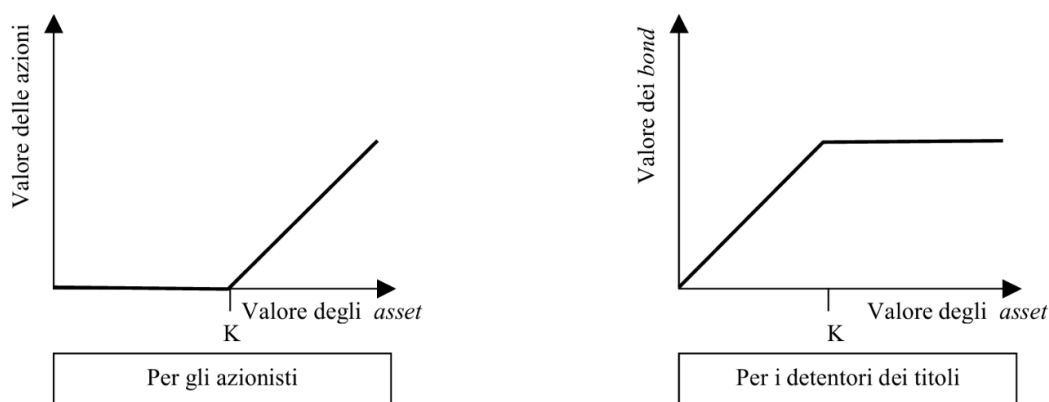


Figura 3. *Payoff per le due categorie di stakeholder*. Fonte: Matta (2004).

Dall'analisi di questi profili è possibile sostenere che il valore a scadenza per gli azionisti e per gli obbligazionisti dell'impresa sia rappresentato dal profilo a scadenza di un'opzione.

⁴³ Ibidem.

⁴⁴ Merton, *op. cit.*

Per i titolari delle azioni, la loro posizione può essere assimilata all'acquisto di un'opzione *call* di tipo europeo con prezzo di esercizio (*strike price*) pari a K . Se a scadenza il valore delle azioni della società al netto del debito è positivo, allora l'opzione viene esercitata per entrare in possesso del valore residuo dell'impresa, altrimenti l'opzione non viene esercitata e il valore residuo delle attività è destinato alla remunerazione dei creditori.

D'altra parte, la posizione degli obbligazionisti è rappresentata da un'ipotetica cessione di un'opzione *put* di tipo europeo. In tale modo, si ottiene il diritto a ricevere alla scadenza T un ammontare pari al valore nominale dell'obbligazione.⁴⁵

Sulla base delle ipotesi sinora introdotte, il valore di un *bond* risulta pari al valore degli *asset* al netto del valore di mercato delle azioni, che nel modello di Merton è calcolato utilizzando la formula di Black e Scholes (1973). In questo modello si assume che il valore dell'attività dell'impresa segua un processo di diffusione log-normale con volatilità costante.⁴⁶

Formalmente, il valore del titolo di debito emesso dall'impresa al tempo t è pari alla differenza tra valore di mercato degli *asset* al netto del valore di mercato di una opzione *call* con scadenza al tempo T e prezzo di esercizio pari al valore nominale K di tale obbligazione.⁴⁷

$$D(t, T) = V_T - C(K, T)$$

Dove il secondo termine è il valore di una *call option* secondo Black e Scholes (1973), ossia, in assenza di dividendi:

$$C(K, T) = V_T N(d_1) - Ke^{-r(T-t)} N(d_2)$$

In cui r rappresenta il tasso di interesse per attività *risk free* ed $N(\cdot)$ è la funzione di ripartizione di una variabile normale standard, per cui:

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy$$

⁴⁵ Matta, *op. cit.*

⁴⁶ Hull, J., Nelken, I., & White, A. (2004). Merton's model, credit risk and volatility skews. *Journal of Credit Risk*, 1(1), p. 3-28.

⁴⁷ Matta, *op. cit.*

Si definiscono i parametri intermedi d_1 e d_2 come segue:

$$d_1 = \frac{\ln\left(\frac{V_T}{K}\right) + \left(r + \frac{1}{2}\sigma_V^2\right)(T-t)}{\sigma_V\sqrt{T-t}}$$

$$d_2 = \frac{\ln\left(\frac{V_T}{K}\right) + \left(r - \frac{1}{2}\sigma_V^2\right)(T-t)}{\sigma_V\sqrt{T-t}}$$

Dove σ_V rappresenta la volatilità media del valore degli *asset*.⁴⁸

La probabilità di osservare il fallimento dell'impresa alla scadenza T secondo Merton equivale quindi alla probabilità che a tale data il valore della società sia inferiore al valore del debito K :⁴⁹

$$P[\tau = T] = P[V_T < K] = N(-d_2) = 1 - N(d_2)$$

Il differenziale (*spread*) creditizio del titolo di debito rispetto ad altre attività non rischiose si ricava facilmente per differenza, come $S(t, T) = D(t, T) - B(t, T)$ e assume la seguente forma:

$$S(t, T) = -\frac{1}{T-t} \ln \left[N(d_2) + \frac{1}{d} N(-d_1) \right]$$

In cui d rappresenta il *quasi-debt ratio*, ovvero una misura del grado di leva finanziaria dell'impresa, ed è espresso dal seguente rapporto:⁵⁰

$$d = \frac{K}{V_T} e^{-r(T-t)}$$

La stima dei parametri inseriti nel modello avviene sulla base delle informazioni possedute dagli operatori, più precisamente:

- L'ammontare del debito K si ottiene dalle specifiche voci del passivo di stato patrimoniale;
- Il tasso di interesse *risk free* r si stima a partire dai rendimenti dei titoli governativi oppure dalla struttura dei tassi *swap*;

⁴⁸ Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, 81(3), p. 637-654.

⁴⁹ Merton, *op. cit.*

⁵⁰ Matta, *op. cit.*

- I valori di V_T e della volatilità σ_V si possono desumere indirettamente dall'osservazione del valore di mercato delle azioni e della loro volatilità.

Nel momento in cui il modello viene applicato alle dinamiche del mondo reale diventa molto più articolato. Infatti, devono essere definite tutte le variabili rilevanti per l'analisi aziendale, quali i debiti verso le banche, le richieste dei soci azionisti e dei creditori.

Per tali ragioni, l'utilizzo pratico dell'approccio strutturale tende a non coincidere specularmente con il modello, ma piuttosto ad individuarne le linee guida e gli elementi più rilevanti.⁵¹

Il modello di Merton è stato introdotto durante un periodo di crescente complessità nei mercati finanziari e ha rappresentato un avanzamento significativo nella comprensione del rischio di default. Ancora oggi, il modello è molto apprezzato dagli operatori ed è utilizzato non solo dalle banche ma anche da società di investimento e negli *hedge fund*. Tuttavia, i modelli strutturali sono suscettibili ad alcune critiche, tra cui:

- La struttura del capitale dell'impresa è estremamente semplificata, infatti comprende un'unica categoria di debito con un'unica scadenza;
- I valori dei differenziali (*spread*) non sono coerenti con quanto osservato empiricamente. Nello specifico, quelli ottenuti applicando il modello sono molto bassi rispetto a quanto osservabile sul mercato;⁵²
- Il default non prevede alcun costo.⁵³

1.3.2 Il modello KMV

Una delle più rilevanti implementazioni a livello pratico dell'approccio strutturale proposto da Merton è rappresentata dal modello KMV, il quale prende il nome dai suoi sviluppatori Kealhofer, McQuown e Vasicek. L'impegno profuso dai tre studiosi, riunitisi in KMV Corporation, nacque dall'esigenza di superare i limiti dei modelli per il rischio di credito tradizionalmente utilizzati dagli operatori economici, che spesso non erano in grado di cogliere le interazioni complesse tra le attività e le passività di un'impresa, nonché il rischio intrinseco di default. La genesi del modello KMV può essere fatta

⁵¹ Ibidem.

⁵² Cossin, D., & Pirotte, H. (2001). *Advanced Credit Risk Analysis: Financial Approaches and Mathematical Models to Assess, Price, and Manage Credit Risk*. Wiley.

⁵³ Matta, *op. cit.*

risalire agli inizi degli anni '90, quando i fondatori dell'omonima società californiana cercarono di sviluppare un approccio più performante per la previsione del rischio di fallimento. Nel 2002 la società è stata acquisita da Moody's Analytics, creando la divisione Moody's KMV, che tuttora continua a sviluppare e perfezionare il modello.

Nel modello, il default si verifica nel momento in cui il valore di mercato degli *asset* della società scende al di sotto di una determinata soglia critica (*default point*). Nel modello vengono considerate diverse classi della struttura del capitale quali quelle a breve termine, a lungo termine, obbligazioni convertibili, azioni ordinarie e azioni privilegiate.

Diversamente da quanto fa Merton, in questo modello si stima la frequenza con cui le imprese vanno in fallimento sulla base della frazione delle società, aventi caratteristiche simili, che sono effettivamente fallite negli anni precedenti. Per fare ciò, gli operatori utilizzano un database proprietario.⁵⁴

Il processo di determinazione della probabilità di fallimento di un'impresa si articola in tre fasi.

Nella prima fase, partendo dai dati osservabili è possibile stimare il valore dell'impresa (*asset value*) V e la volatilità dei suoi *asset* σ_V . Infatti, per la maggior parte delle imprese quotate sono direttamente osservabili il prezzo del capitale proprio (*equity*) e la parte del debito che viene negoziata, come i titoli obbligazionari.⁵⁵

Il modello KMV stima i due valori risolvendo simultaneamente le seguenti equazioni:

$$S = f(V_T, \sigma_V, \lambda, c, r)$$

$$\sigma_S = g(V_T, \sigma_V, \lambda, c, r)$$

Dove:

- S indica il valore dell'*equity*.
- σ_S rappresenta la volatilità dell'*equity*.
- λ indica il *leverage ratio* della struttura del capitale.
- c indica la cedola mediamente pagata sul debito a lungo termine.

⁵⁴ Crosbie, P., & Bohn, J. (2002). *Modeling Default Risk*. San Francisco (USA): KMV LLC.

⁵⁵ HKUST Math Department. (2002). *KMV Model*. Y. K. Kwok. Hong Kong University of Science and Technology

- r è il tasso *risk free*.⁵⁶

Nella seconda fase si calcola la distanza dal default (*distance to default*) a partire dal valore degli *asset*, la volatilità degli *asset* e dal valore delle passività iscritte a bilancio, intese come somma dei debiti a breve e a lungo termine. Le imprese non vanno necessariamente in default nel momento in cui il valore degli *asset* raggiunge il valore delle passività iscritte a bilancio in quanto la natura a lungo termine di alcune di queste passività conferisce alle imprese un certo margine temporale prima del default.⁵⁷

Il modello KMV definisce il punto di default (DP) come il valore degli *asset* al quale l'impresa va in default. L'evidenza empirica osserva che tale valore si colloca tra il totale delle passività e le passività a breve termine. Esso è dato da:

$$DP = STD + \frac{1}{2}LTD$$

Dove STD indica il debito a breve e LTD quello a lungo termine.

Pertanto, un'impresa va in default quando il suo valore netto di mercato, *i.e.* la differenza tra il valore degli *asset* e il punto di default, raggiunge lo zero.⁵⁸

Successivamente, è possibile calcolare una misura sintetica del rischio di fallimento, espressa dalla *distance to default* (DD). Essa rappresenta il numero di deviazioni standard che separano il valore degli *asset* dal default:

$$DD = \frac{S(V_T) - DP}{\sigma_V} = \frac{\ln \frac{V_0}{DP} + \left(\mu - \frac{\sigma_V^2}{2}\right)T}{\hat{\sigma}_V \sqrt{T}}$$

Dove:

- V_0 è il valore attuale di mercato degli *asset*;
- V_T è il valore di mercato degli *asset* al tempo T ;
- μ è il rendimento netto atteso sul valore dell'impresa;
- $\hat{\sigma}_V$ è la volatilità degli *asset* su base annua.⁵⁹

⁵⁶ Ibidem.

⁵⁷ Crosbie & Bohn, *op. cit.*

⁵⁸ Ibidem.

⁵⁹ HKUST, *op. cit.*

Se la distribuzione futura della *distance to default* fosse nota, la probabilità di default sarebbe semplicemente pari alla probabilità che il valore degli *asset* rimanga al di sotto del punto di default. Nella pratica, la distribuzione della DD è di difficile misurazione. Inoltre, le consuete assunzioni della distribuzione lognormale, come quella proposta da Merton (1974), costituiscono una scelta metodologica poco appropriata.

Per una corretta misurazione del rischio, è infatti critico determinare la probabilità di ampie variazioni avverse nel rapporto tra il valore degli *asset* e il punto di default. Tali variazioni possono derivare da fluttuazioni nel valore degli *asset* o da cambiamenti nel grado di *leverage* dell'impresa, fattori che risultano spesso altamente correlati tra di loro.⁶⁰

Di conseguenza, il modello KMV misura prima la DD come numero di deviazioni standard di cui il valore degli *asset* dista dal default, e successivamente utilizza i dati empirici per determinare la probabilità di fallimento corrispondente.

Nella terza fase si calcola la probabilità di default. Nel modello KMV si utilizza la *Expected Default Frequency* (EDF), che rappresenta la probabilità che il valore degli *asset* scenda al di sotto del punto di default, da cui si può desumere la probabilità di fallimento per l'anno (o gli anni) a venire per le imprese in esame (Figura 4).

Tali informazioni possono essere consultate e analizzate all'interno di un software denominato Credit Monitor™. Quest'ultimo calcola i valori dell'EDF per un orizzonte temporale che va da uno a cinque anni, consentendo all'utente di osservare una struttura a termine del rischio relativo alla società analizzata. La misura EDF di KMV assume che il default sia definito come il mancato pagamento di qualsiasi scadenza prevista, che sia quota capitale o quota interessi.

Le informazioni sulle imprese storicamente fallite desumibili dal database a disposizione degli operatori vengono utilizzate per derivare empiricamente una distribuzione che mette in relazione la distanza dal default con la probabilità di default. La distribuzione empirica risultante dalle frequenze dei default presenta code molto più ampie (*Fat tails*) rispetto alla distribuzione normale. Ad esempio, una DD pari a quattro, ossia quattro deviazioni

⁶⁰ Crosbie & Bohn, *op. cit.*

standard, corrisponde ad un EDF di circa 100 punti base (figura 5). La probabilità equivalente derivante dalla distribuzione normale è essenzialmente nulla.⁶¹

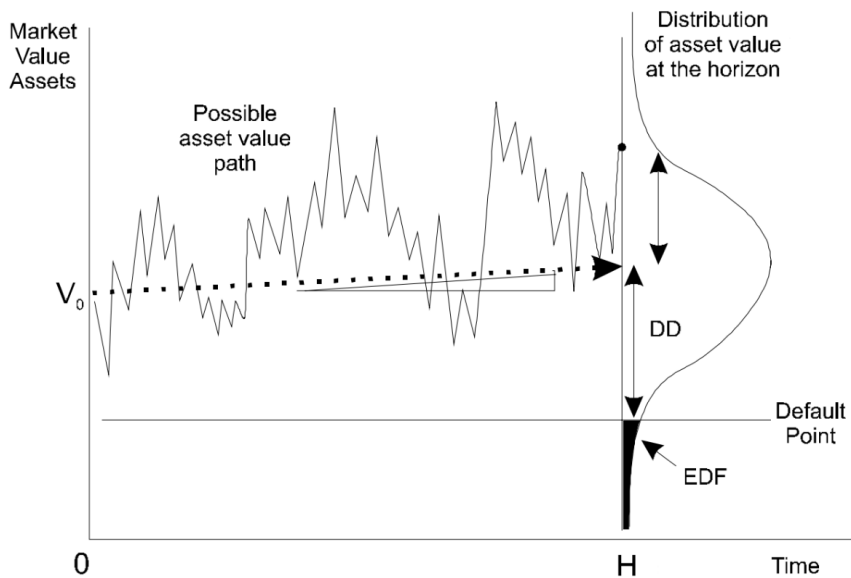


Figura 4. *Rappresentazione grafica della logica sottesa al modello.* Fonte: Crosbie & Bohn (2002).

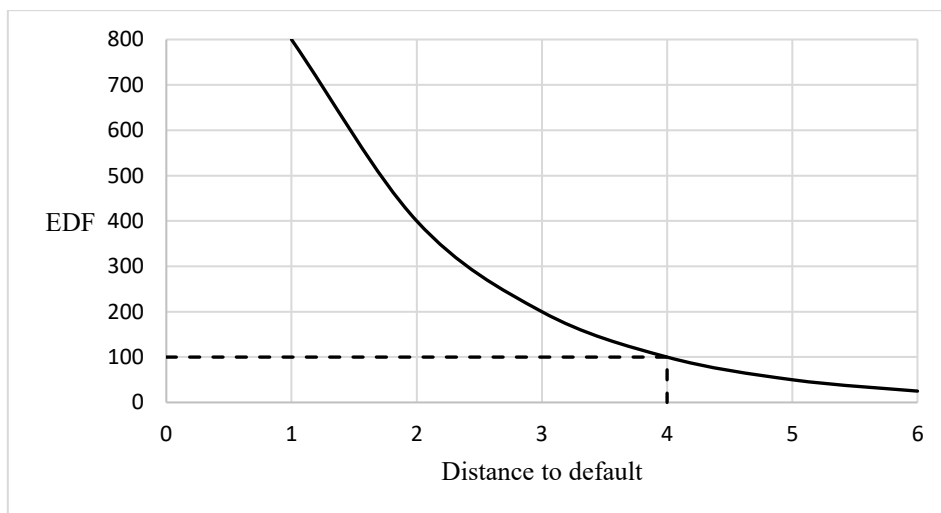


Figura 5. *Mappatura dell'Expected Default Frequency sulla base della Distance to default.* Fonte: elaborazione propria.

⁶¹ Ibidem.

Le notevoli qualità del modello KMV hanno spinto Moody's a continuare a supportarlo e svilupparlo negli anni. In particolare, si evidenziano alcuni vantaggi:

- Le variazioni dell'EDF tendono ad anticipare di almeno un anno l'eventuale declassamento dell'emittente da parte delle agenzie di rating;
- l'EDF fornisce una classificazione cardinale, e non ordinale, della qualità creditizia;
- le informazioni accurate e tempestive provenienti dal mercato azionario permettono un processo continuo di monitoraggio del credito, pratica difficile e onerosa da replicare con un'analisi più tradizionale;
- le revisioni annuali e le analisi creditizie convenzionali non sono in grado di mantenere lo stesso livello di monitoraggio che le EDF calcolate su base giornaliera o mensile sono capaci di offrire.

Di contro, il modello presenta alcuni profili di criticità, tra cui:

- Per alcuni parametri di input, KMV richiede una stima soggettiva;
- risulta difficile costruire EDF teoriche senza ricorrere alle ipotesi di normalità dei rendimenti degli *asset*;
- le EDF relative alle imprese non quotate possono essere calcolate solo attraverso analisi di comparabilità basate sui dati contabili;
- il modello non distingue le varie tipologie di obbligazioni a lungo termine in base alla loro *seniority*, alla presenza di garanzie (*collateral*), ai *covenants* o alla convertibilità.⁶²

1.4 L'evoluzione verso i modelli *Hazard*: il contributo di Shumway

Un notevole cambiamento metodologico nel campo della previsione dell'insolvenza avviene con l'introduzione dell'utilizzo degli *Hazard models* nella letteratura finanziaria da parte di T. Shumway (2001). L'autore critica in particolare l'utilizzo, nei decenni precedenti, di quelli che definisce come "modelli statici", ovvero modelli di classificazione per il fallimento a periodo singolo, che utilizzano i dati sul default relativi

⁶² HKUST, *op. cit.*

a più periodi. Questi modelli, ignorando il fatto che le realtà economiche si evolvono nel tempo, producono stime distorte sulla probabilità di default.

Poiché il fallimento è un evento raro, i ricercatori utilizzano dei campioni che coprono diversi anni per la costruzione di tali modelli. Tuttavia, questi possono utilizzare solo un singolo insieme di variabili esplicative per ogni impresa. Tipicamente l'applicazione dei modelli statici avviene osservando le caratteristiche di ogni azienda fallita nell'anno precedente l'evento, ignorando i dati sulle imprese sane che in futuro andranno in default. I ricercatori, scegliendo arbitrariamente quando osservare tali caratteristiche di ciascuna impresa, introducono un *bias di selezione* nelle loro stime.⁶³

Shumway, per ovviare a questo problema, decide di applicare un modello *Hazard* ai fini della previsione del default, il quale utilizza congiuntamente dati contabili e dati di mercato. Nello specifico, il modello assume la forma funzionale di *Hazard model* a tempo discreto e utilizza la funzione della regressione logistica. Tali modelli utilizzano delle variabili che si evolvono nel tempo. In tal senso, la probabilità che un'impresa fallisca in $t+1$ è condizionata alla sopravvivenza fino a t .⁶⁴

Analogamente a quanto previsto nel modello logit, la variabile dipendente è di tipo dicotomico, potendo assumere in $t+1$ lo stato di sopravvivenza o default.

L'autore per la costruzione del modello propone una procedura semplice utilizzando un software in grado di eseguire delle regressioni logistiche standard. Ogni anno in cui l'impresa sopravvive viene inclusa nel campione come un'impresa che non è fallita. D'altra parte, ogni impresa in default viene registrata come singola osservazione di fallimento. Le covariate che variano nel tempo vengono integrate attraverso l'uso dei dati annuali di ciascuna impresa per le relative osservazioni *panel*.⁶⁵

Shumway sottolinea inoltre la semplicità con la quale è possibile effettuare inferenze statistiche in un modello *hazard* sviluppato con un software logit. Poiché entrambi i

⁶³ Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74(1), p. 432-442.

⁶⁴ Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 40, p. 432-442.

⁶⁵ Shumway, *op. cit.*

modelli hanno la stessa funzione di verosimiglianza, essi presentano la medesima matrice di varianza-covarianza.⁶⁶ Tuttavia, le statistiche test prodotte dal programma non sono corrette per il modello *hazard*, in quanto assumono che il numero di osservazioni indipendenti utilizzate per la stima del modello sia pari al numero di “anni-impresa” presenti nei dati. Il calcolo corretto delle statistiche test richiede di aggiustare la dimensione del campione ipotizzata dal software per tenere conto della mancanza di indipendenza tra le osservazioni anno-impresa. Queste non possono essere indipendenti, poiché un’impresa non può fallire nel periodo t se è già fallita nel periodo $t-1$, e allo stesso modo, un’impresa che sopravvive fino al periodo t non può essere fallita in $t-1$.⁶⁷

Per il modello *hazard*, l’intera vita di ciascuna impresa rappresenta una singola osservazione. Ne segue che il valore corretto di n per le statistiche test è il numero di imprese nei dati utilizzati, non il numero di anni-impresa. La statistica test *chi-quadrato* restituita dal programma logit assume la seguente forma:

$$\frac{1}{n}(\hat{\mu}_k - \mu_0)' \Sigma^{-1} (\hat{\mu}_k - \mu_0) \sim \chi^2(k)$$

Dove k rappresenta il numero di momenti stimati testati rispetto a k ipotesi nulle (m_0). La correzione della statistica test restituita dal software logit per il modello *hazard* si ottiene dividendo tali valori per il numero di anni-impresa per ogni società.

I modelli logit che presentano più osservazioni per ciascun soggetto tipicamente tengono conto della mancanza di indipendenza tra le osservazioni, che è una caratteristica tipica dei dati panel.⁶⁸ Il logit utilizzato dall’autore risulta già penalizzato per la mancanza di indipendenza tra le osservazioni anno-impresa attraverso l’aggiustamento della dimensione del campione sopra citata. Poiché tale approccio non assume che le osservazioni siano indipendenti e identicamente distribuite, non sono necessari ulteriori aggiustamenti per correggere la dipendenza dei dati.⁶⁹

Interpretare il modello logit come un modello *Hazard* consente di rendere più intuitiva l’interpretazione dei coefficienti.

⁶⁶ Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, Mass.: Harvard University Press.

⁶⁷ Shumway, *op. cit.*

⁶⁸ Amemiya, *op. cit.*

⁶⁹ Shumway, *op. cit.*

Considerando θ come il vettore dei parametri e x come il vettore delle variabili esplicative ed effettuando la partizione di θ in θ_1 e θ_2 la funzione *Hazard* può essere espressa come:

$$\phi(t, x, \theta_1, \theta_2) = \frac{1}{1 + \exp(g(t)' \theta_1 + x' \theta_2)}$$

Qualora la funzione scelta per la variabile “età” dell’impresa, $g(t)$, sia il ln dell’età, il modello *hazard* si configura come un modello AFT (*accelerated failure time model*). In questo caso, i coefficienti possono essere interpretati attraverso l’equazione di regressione:⁷⁰

$$E[\ln(t) | x] = \frac{x' \theta_2}{\theta_1}$$

Tra le variabili più rilevanti utilizzate dall’autore troviamo:

$$NITA = \frac{\text{Reddito netto}}{\text{Totale attività}}$$

$$TLTA = \frac{\text{Totale passività}}{\text{Totale attività}}$$

Excess return ($r_{it-1} - r_{mt-1}$), che rappresenta la differenza del rendimento azionario dell’impresa e il rendimento di mercato, entrambi in $t-1$.

Relative size, inteso come il logaritmo del rapporto tra la dimensione di ciascuna impresa e la dimensione totale del mercato NYSE e AMEX.

Sigma, ovvero la deviazione standard idiosincratca dei rendimenti azionari di ciascuna impresa, dove con idiosincratco si intende il rischio che non può essere spiegato dai fattori di mercato, essendo specifico dell’impresa.

Shumway nel suo lavoro prova empiricamente la maggior flessibilità di utilizzo e le migliori *performance* del suo modello rispetto ai modelli statici. Inoltre, evidenzia tre ragioni principali per le quali si dovrebbe preferire un modello *Hazard*. In primo luogo, quando i periodi di campionamento sono lunghi, è essenziale distinguere tra le società che falliscono nel primo anno e quelle che falliscono dopo molti anni. I modelli statici non effettuano alcun aggiustamento per il periodo durante il quale l’impresa viene

⁷⁰ Ibidem.

considerata, mentre i modelli *Hazard* lo fanno automaticamente, evitando il *bias di selezione*. In secondo luogo, i modelli *Hazard* incorporano delle covariate che variano nel tempo. Se un'impresa peggiora nel periodo precedente al default, è essenziale che i dati finanziari riflettano tale peggioramento. Tali modelli includono sia osservazioni annuali, sia variabili macroeconomiche comuni a tutte le imprese, e tengono conto della variabile "età" dell'impresa. Infine, i modelli in esame forniscono delle previsioni fuori campione più efficienti, in quanto utilizzano una quantità maggiore di dati.⁷¹

⁷¹ Ibidem.

Capitolo 2

L'approccio basato su Intelligenza Artificiale e Machine Learning

2.1 Introduzione al Machine Learning

Con intelligenza artificiale (IA) si intende la capacità di un sistema informatico di replicare comportamenti umani quali il ragionamento, l'apprendimento, la pianificazione e la creatività.⁷²

Gli ultimi decenni sono stati caratterizzati da un notevole progresso tecnologico, soprattutto nel campo informatico, che ha favorito una significativa evoluzione di tali sistemi. Grazie a maggiori capacità di calcolo e alla disponibilità di grandi quantità di dati, i c.d. *big data*, è stata possibile l'applicazione pratica dell'IA su larga scala.

In realtà, il concetto alla base dell'IA non è estremamente recente. Già nel 1950, Alan Turing, matematico noto per il suo contributo alla decifrazione del codice tedesco Enigma durante la Seconda guerra mondiale, si interrogò sulla possibilità che una macchina potesse essere in grado di “pensare”.⁷³ Qualche anno più tardi, nel 1956, John McCarthy coniò il termine “intelligenza artificiale” in occasione della conferenza di Dartmouth. Nello stesso anno, i ricercatori Allen Newell e Herbert Simon, spinti dall'obiettivo di creare una macchina in grado di simulare ogni aspetto dell'apprendimento e dell'intelligenza umana, presentarono il primo programma di IA. Denominato “*Logic Theorist*”, esso era in grado di dimostrare teoremi partendo dai principi della matematica.⁷⁴

La continua ricerca in campo accademico e la crescente disponibilità di risorse computazionali consentirono negli anni '80 le prime applicazioni in ambito industriale. Tale evoluzione portò allo sviluppo di algoritmi con prestazioni sempre più elevate, i quali iniziarono a dispiegare concretamente le loro potenzialità sia nel tessuto produttivo sia nella sfera individuale.

⁷² ISO/IEC. (2023). *42001 Information technology, Artificial intelligence, Management system*. Ginevra.

⁷³ Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), p. 433-460.

⁷⁴ McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Dartmouth College.

Queste tecniche si differenziano ampiamente dall'approccio tradizionale della programmazione, nel quale un programmatore definisce un algoritmo che sia in grado, attraverso la codifica di una serie di operazioni, di trasformare deterministicamente dei dati di input in dati di output desiderati. Diversamente, nell'ambito dell'IA si costruisce un modello in grado di catturare le informazioni necessarie per derivare "conoscenza" in modo automatico a partire dai dati disponibili. In questo ambito, le competenze di sviluppo delle applicazioni vengono integrate con quelle tipiche del profilo di *data scientist*, figura professionale orientata all'analisi e all'interpretazione dei dati.

L'IA può essere sinteticamente ricondotta a due differenti approcci: l'approccio induttivo e l'approccio deduttivo. Nell'approccio induttivo, il sistema sintetizza la propria conoscenza sulla base dell'osservazione empirica dei dati, imparando da questi tramite un processo di generalizzazione; in quello deduttivo, a partire da una rappresentazione formale della conoscenza operata tramite linguaggi di *knowledge representation and reasoning* (KRR), il sistema produce nuova conoscenza dai dati in input in un processo di inferenza. L'approccio induttivo è tipicamente conosciuto come apprendimento automatico, o *Machine Learning* (ML). L'approccio deduttivo è noto come ragionamento automatico, o *Automated Reasoning* (AR).⁷⁵

Il *Machine Learning* (ML), o apprendimento automatico, è un sottoinsieme dell'intelligenza artificiale, che si basa su algoritmi in grado di apprendere dai dati in modo autonomo, senza che vi siano istruzioni esplicite o codificate.⁷⁶

La definizione più celebre deriva da Tom M. Mitchell, uno degli accademici più importanti in termini di contributi pionieristici nei metodi e nelle applicazioni del ML:

«Un programma apprende dall'esperienza *E* con riferimento ad alcune classi di compiti *T* e con misurazione della performance *P* se le sue performance nel compito *T*, come misurato da *P*, migliorano con l'esperienza *E*».⁷⁷

Ne deriva che un programma apprende se c'è un miglioramento delle prestazioni dopo lo svolgimento di un compito. In questo senso, il lavoro di Mitchell può essere interpretato

⁷⁵ Bonaccorsi, E., & Calabresi, F. (2022). *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano. Temi di discussione n. 721*. Banca d'Italia.

⁷⁶ Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

⁷⁷ Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.

come un'evoluzione concettuale rispetto a quanto sollevato da Turing, chiedendosi se le macchine fossero in grado di compiere azioni tipicamente attribuibili all'intelligenza umana.⁷⁸

Nella letteratura scientifica si identificano tre paradigmi di apprendimento, distinguibili sulla base della tipologia di *feedback* disponibile al sistema durante il processo di addestramento.

Nell'apprendimento per rinforzo il modello interagisce con un ambiente dinamico con lo scopo di raggiungere un obiettivo. I rinforzi si compongono di premi e punizioni, o dalla mancanza di essi.⁷⁹ Ad esempio, l'assenza di un premio al termine di una certa operazione indica al modello che è stato commesso un errore; al contrario, la presenza di un premio al termine dell'operazione indica al modello di aver agito correttamente. Spetta poi al modello determinare quali, tra le azioni compiute prima del rinforzo, sono state le principali responsabili del risultato ottenuto.

Nell'apprendimento non supervisionato il modello apprende schemi nei dati di input anche in assenza di un *feedback* esplicito, ovvero i dati forniti non vengono etichettati in alcun modo. Il compito più comune in questo ambito è il *clustering*, mediante il quale si cerca di individuare dei raggruppamenti di dati potenzialmente utili.

Nell'apprendimento supervisionato il modello analizza coppie di esempi composti da possibili input e i rispettivi output al fine di apprendere una funzione che permetta di mappare correttamente l'input sul rispettivo output. L'obiettivo è quindi quello di estrarre una regola generale che associ l'input all'output corretto.⁸⁰

Nella pratica, queste distinzioni non sono sempre così nette, portando al riconoscimento di un paradigma intermedio tra i due precedenti, definito apprendimento semi-supervisionato, in cui il *dataset* è parzialmente etichettato. L'adozione di questo approccio può essere motivata, ad esempio, dall'elevato costo per l'attività di etichettatura.

⁷⁸ Harnad, S. (2008). The Annotation Game: On Turing (1950) on Computing Machinery and Intelligence. *Parsing the Turing Test*, p. 23-66.

⁷⁹ Skinner, B. F. (1971). *Beyond Freedom and Dignity*. Knopf.

⁸⁰ Russel, S. J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3 ed.). Upper Saddle River: Prentice Hall.

2.2 Il machine learning applicato alla previsione del default

Uno degli ambiti in cui sono state esplorate le proprietà delle tecniche di ML è la valutazione del rischio di default. Rispetto all'approccio statistico tradizionale, basato sulla stima econometrica della probabilità di default, si evidenziano alcune differenze fondamentali. Nei modelli econometrici la determinazione delle variabili più significative e la specificazione delle forme funzionali delle relazioni vengono generalmente guidate dalla teoria economica, mentre nel ML gli algoritmi selezionano in maniera automatica le variabili rilevanti identificando relazioni anche non lineari e di difficile interpretazione tra le stesse.

Numerosi studi hanno confrontato la capacità delle tecniche di ML di prevedere il default delle imprese e dei privati con quella di approcci più tradizionali. In particolare, gli studi di Bazzana, Bee e Hussin (2024) mostrano che l'accuratezza dei modelli di ML nell'individuare le insolvenze è generalmente migliore rispetto a quella dei modelli econometrici.⁸¹ Il miglioramento in termini prestazionali delle tecniche di ML deriva in primo luogo dall'estensione della gamma di forme funzionali e relazioni tra le diverse variabili valutate dal modello. Un'analisi condotta da Alonso e Carbó (2020) evidenzia che l'utilizzo di tecniche di ML avanzate comporta un miglioramento delle previsioni del default rispetto ai modelli statistici tradizionali per lo più compreso tra due e dieci punti percentuali ma in alcuni casi anche superiore.⁸²

Un altro punto di forza del ML è quello di gestire ed elaborare grandi quantità di dati in volume (numero di osservazioni) e ricchezza (numero di variabili, tipi di dato), sfruttando la capacità di calcolo di elaboratori sempre più potenti. In generale, nella previsione del fallimento si è verificata una graduale estensione delle fonti di dati utilizzate, fenomeno che ha riguardato anche i modelli econometrici. Si spazia dai dati finanziari strutturati (indicatori patrimoniali ed economico-finanziari, di mercato) a dati non finanziari strutturati (dati di tipo socio-demografico ottenuti anche da fonti terze), dati non strutturati finanziari (analisi delle informazioni transazionali e di quelle derivate da *open*

⁸¹ Bazzana, F., Bee, M., & Hussin, A. (2024). Machine learning techniques for default prediction: an application to small Italian companies. *Risk Management*, 26(1), p. 1-25.

⁸² Alonso, A., & Carbó, J. M. (2020). *Machine Learning in Credit Risk: Measuring the Dilemma between Prediction and Supervisory Cost*. Working Paper No. 2032. Banco de España.

banking) e dati non strutturati non finanziari (dati di navigazione, *digital footprint*, informazioni conferite sui social network).⁸³

Il ricorso a modelli di ML, sfruttando le varie fonti di dati, permette di considerare variabili che non hanno una chiara interpretazione o relazione economica, che non verrebbero considerate in un modello tradizionale, grazie al peculiare processo di addestramento e calibrazione di questi modelli. Lo sfruttamento di informazioni alternative, o complementari, a quelle in uso in modelli tradizionali può contribuire a un miglioramento dell'accuratezza previsiva.⁸⁴

2.2.1 I vincoli normativi

L'adozione delle tecniche di ML nella previsione del default deve tuttavia confrontarsi con un quadro normativo rigoroso.

Le linee guida dell'EBA (Autorità Bancaria Europea) prevedono la responsabilità della governance per le tecnologie utilizzate, l'obbligo dell'integrità dei dati, e la tracciabilità dei processi automatizzati. Nel documento "*report on Big Data and Advanced Analytics*", vengono identificati alcuni pilastri fondamentali per un uso etico e sicuro dell'IA.

Tra questi, assume una rilevanza centrale il concetto di fiducia, inteso come il requisito che i modelli siano progettati e utilizzati in modo da prevenire e mitigare pregiudizi e discriminazioni da parte degli algoritmi, garantendo equità nei risultati. A tale principio si affiancano la trasparenza e la spiegabilità: gli istituti di credito hanno infatti l'obbligo di rendere intellegibile la logica sottostante a una decisione automatizzata, specialmente nell'ipotesi in cui essa impatti negativamente sul cliente, come nel caso del rifiuto di un finanziamento. Questo impianto si inserisce in una più ampia strategia di *data governance* che richiede il rigoroso rispetto della qualità dei dati e del GDPR (Regolamento UE 2016/679).

Le linee guida dell'EBA sulla concessione e monitoraggio dei prestiti pubblicate nel 2020 includono disposizioni specifiche per l'uso di modelli automatizzati nella valutazione del

⁸³Tobback, E., & Martens, D. (2019). Retail credit scoring using fine-grained payment data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(4), p. 1227-1246.; Berg, T., Burg, V., Gombovic, A., & Puri, M. (2020). On the Rise of Fintechs: Credit Scoring Using Digital Footprints. *The Review of Financial Studies*, 33(7), p. 2845-2897.

⁸⁴ Bazarbash, M. (2019). *Fintech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk*. IMF Working Paper No. 19/110. International Monetary Fund.

merito creditizio. Si richiedono una supervisione umana costante e test regolari per evitare che l'algoritmo porti a risultati errati nel tempo.

A completamento di tali profili, vi è la necessità di rispettare standard di *cybersecurity* come quelli previsti dal Regolamento DORA, per poter garantire la sicurezza e la protezione dei singoli consumatori e la stabilità del sistema finanziario nel suo complesso.

La novità più rilevante per l'utilizzo di tali sistemi è rappresentata dall'entrata in vigore nell'agosto 2024 del Regolamento UE 2024/1689 sull'intelligenza artificiale (AI Act), il primo quadro giuridico al mondo a trattare questa specifica materia.

In base all'importanza dei rischi che ciascun sistema di IA comporta per i diritti fondamentali dell'essere umano, è prevista una classificazione. Sebbene l'impiego dei modelli di previsione del default possa rientrare nelle esclusioni previste dall'AI Act qualora limitato a fini puramente prudenziali, la loro applicazione nell'ambito del *credit scoring* e della valutazione del merito creditizio ne determina la classificazione come sistemi ad alto rischio.

In risposta, l'EBA ha condotto nel 2025 un'attività di mappatura per armonizzare i requisiti dell'AI Act con le normative di settore. La posizione espressa dall'Autorità sottolinea la validità delle disposizioni vigenti (EBA/GL/2020/06), confermando che non è attualmente necessaria l'emanazione di nuovi orientamenti, ritenendo il *corpus* normativo esistente idoneo a disciplinare l'uso dell'IA nel settore creditizio.

L'AI Act troverà piena applicazione al termine di un periodo di attuazione graduale, previsto entro agosto 2027. In questo periodo di transizione, l'EBA intraprenderà una serie di attività specifiche a supporto dell'attuazione delle nuove disposizioni. Si intende promuovere un approccio di vigilanza armonizzato, in cui i requisiti specifici dell'AI Act si integrino sinergicamente con le linee guida settoriali, garantendo che l'innovazione tecnologica non comprometta la stabilità e la trasparenza dei processi di valutazione del merito di credito. È inoltre previsto un supporto tecnico continuativo alle discussioni del sottogruppo del Comitato per l'IA dedicato ai servizi finanziari.

2.3 I principali algoritmi per la previsione del rischio default

Con riferimento agli algoritmi utilizzati nella previsione del rischio di default, il paradigma metodologico di riferimento è quello dell'apprendimento supervisionato. Tale

approccio è ampiamente impiegato nei problemi di classificazione, nei quali si dispone di dataset etichettati, ossia caratterizzati da una preventiva assegnazione delle osservazioni a specifiche classi, che il modello apprende per poi estendere le proprie capacità predittive a nuovi dati. Rispetto all'apprendimento non supervisionato, volto a individuare strutture latenti nei dati in assenza di etichette o conoscenze a priori, le tecniche supervisionate consentono generalmente di ottenere dei modelli predittivi efficaci anche con volumi di dati relativamente più contenuti.

In tale ambito si richiamano due tra le principali famiglie di algoritmi, particolarmente diffuse anche in ragione delle loro caratteristiche strutturali. Le tecniche di *ensemble learning* si basano sull'impiego di più modelli distinti per lo svolgimento del compito di previsione, combinando successivamente le singole stime in una previsione aggregata, tipicamente ottenuta tramite media o voto di maggioranza. Un caso largamente adottato è quello in cui i modelli elementari sono rappresentati da alberi decisionali, che operano mediante una partizione progressiva dello spazio delle soluzioni. All'interno di questo paradigma si distinguono approcci basati sul *bagging*, che prevedono l'addestramento parallelo di modelli indipendenti su diverse porzioni dell'informazione disponibile, e approcci basati sul *boosting*, nei quali una sequenza di modelli viene addestrata in modo iterativo al fine di migliorare progressivamente la capacità predittiva complessiva. Due popolari esempi per gli approcci citati, nel contesto dell'apprendimento supervisionato, sono, rispettivamente, il *random forest* e il *gradient boosting*.

Le reti neurali artificiali costituiscono un'ulteriore famiglia di modelli di calcolo che si ispira al funzionamento delle reti neurali biologiche. Tali modelli si caratterizzano per strutture statistiche non lineari organizzate in forma di rete e risultano particolarmente adatti a rappresentare relazioni complesse tra variabili di input e output che difficilmente possono essere descritte mediante funzioni analitiche tradizionali. Una rete neurale artificiale riceve le informazioni attraverso uno strato di nodi di ingresso, ciascuno dei quali è connesso a un insieme di nodi interni organizzati in più livelli. Ogni nodo elabora i segnali in ingresso e trasmette il risultato ai livelli successivi, consentendo al modello di apprendere progressivamente rappresentazioni sempre più astratte dei dati.

2.3.1 La Random Forest

Tra gli algoritmi più diffusi e rilevanti nella letteratura sul ML vi è la Random Forest, sviluppata da Leo Breiman nel 2001. Tale metodologia si basa sugli alberi decisionali, i quali sono modelli facilmente interpretabili che prendono decisioni attraverso una serie di regole basate sulle caratteristiche dei dati, rappresentabili graficamente come un albero. Al fine di migliorare l'accuratezza predittiva, il metodo proposto da Breiman aggrega le previsioni di un insieme di alberi decisionali.⁸⁵

L'implementazione degli algoritmi di Random Forest richiede la definizione preventiva di tre principali iperparametri: la numerosità degli alberi nell'ensemble, la dimensione minima dei nodi e la quantità di variabili esplicative da campionare. Una volta configurato, questo modello si presta efficacemente alla risoluzione di problemi sia di regressione che di classificazione binaria.

Dal punto di vista strutturale, il Random Forest è costituito da una pluralità di alberi decisionali. Ogni componente dell'ensemble viene addestrato su una porzione di dati estratta dal set originale mediante campionamento con reinserimento, tecnica definita *bootstrap*. Durante questo processo, circa un terzo dei dati non viene utilizzato per l'addestramento dell'albero specifico; tale sottoinsieme, denominato *out-of-bag* (OOB), svolge un ruolo cruciale nella successiva fase di validazione. Ciascuno degli alberi decisionali è strutturalmente semplice e tende a presentare il problema dell'*overfitting*,⁸⁶ ovvero un adattamento eccessivo ai dati osservati che ne compromette la capacità di generalizzazione. Tuttavia, gli errori derivanti da tale fenomeno tendono a ridursi significativamente quando numerosi alberi vengono combinati per generare una foresta.

Un ulteriore livello di diversificazione viene introdotto attraverso il *feature bagging*: l'algoritmo non seleziona tutte le variabili disponibili per ogni nodo, ma ne sceglie solo un sottoinsieme casuale. Questa procedura mira a ridurre la correlazione tra i singoli alberi, incrementando la robustezza complessiva del modello.

⁸⁵ Breiman, L. (2001). *Random Forest*. University of California, Berkeley.

⁸⁶ Taddy, M. (2019). *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions*. McGraw-Hill.

La logica di previsione finale si adatta alla natura del problema analizzato. Nell'ambito della regressione, il modello calcola la media dei risultati prodotti dai singoli alberi. Nei compiti di classificazione, la previsione viene determinata attraverso un meccanismo di votazione a maggioranza, in cui prevale la classe selezionata più frequentemente dalla foresta. Quest'ultimo approccio risulta particolarmente efficace ai fini predittivi nel contesto della classificazione binaria tra aziende in stato di default e aziende sane.

Infine, i campioni OOB vengono impiegati al fine di ottenere una stima imparziale dell'errore di generalizzazione, consentendo di valutare le prestazioni del modello e la sua capacità predittiva complessiva senza la necessità di un dataset di test esterno separato.⁸⁷

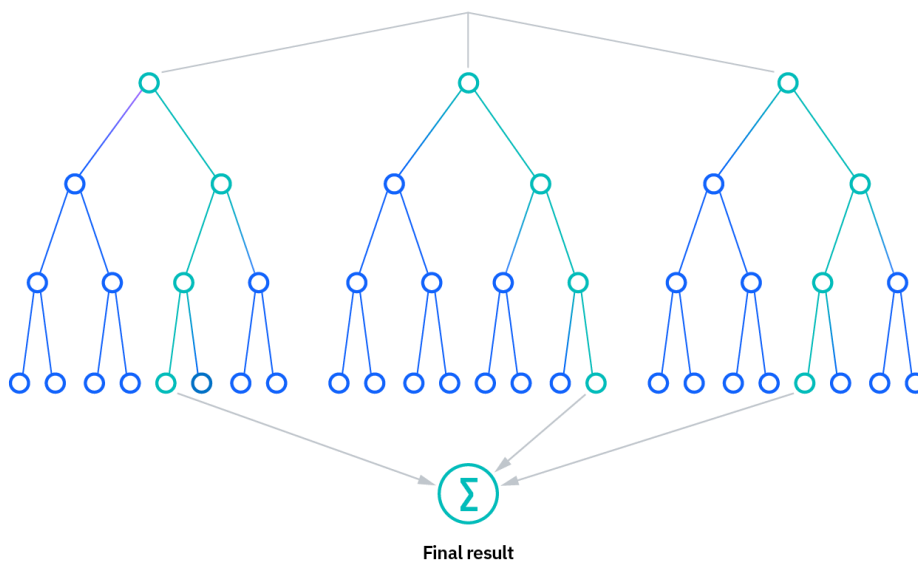


Figura 6. La struttura del modello Random Forest. Fonte: Huang B. (2024)

L'uso dell'algoritmo Random Forest per la previsione del default risponde alla necessità di gestire la complessità e la non-linearità dei dati di bilancio. A differenza dei modelli classici, come lo Z-score di Altman, questo approccio non presuppone una relazione lineare tra gli indici finanziari, permettendo di catturare segnali di crisi che sovente

⁸⁷ Alanis, E., Chava, S., & Shah, A. (2022). *Benchmarking Machine Learning Models to Predict Corporate Bankruptcy*. Research Paper n. 4249412. Georgia Tech Scheller College of Business.

emergono solo dalla combinazione specifica di più fattori (ad esempio un contestuale calo della liquidità e aumento della leva finanziaria).

Dal punto di vista analitico, il ricorso al *bagging* (*bootstrap aggregating*) risulta particolarmente efficace in presenza di dataset di default, tipicamente caratterizzati da un forte sbilanciamento, dato dalla netta preponderanza delle imprese sane rispetto a quelle fallite, e da un elevato "rumore" nei dati contabili. La procedura di addestramento della Random Forest può essere formalizzata come:

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x')$$

Dove $f_b(x')$ rappresenta la previsione del b -esimo albero e B il numero totale di alberi nell'*ensemble*.

L'output finale \hat{f} non è una semplice etichetta, ma la media delle previsioni degli alberi. In un contesto di rischio di default, questa media può essere interpretata come una stima della Probabilità di Default (PD), a condizione che le probabilità di classe risultino opportunamente calibrate.⁸⁸ La bontà del modello risiede nella sua capacità di ridurre la varianza: mentre un singolo albero potrebbe essere tratto in inganno da un'anomalia temporanea nei dati di bilancio⁸⁹, l'aggregazione di B alberi stabilizza la previsione, rendendola resiliente a fluttuazioni di mercato isolate.

Un ulteriore elemento di interesse nell'analisi del rischio di default è rappresentato dalla dispersione delle previsioni dei singoli alberi che compongono l'ensemble. Tale dispersione può essere interpretata come una misura dell'incertezza del modello ed è quantificabile attraverso la deviazione standard delle previsioni:

$$\sigma = \sqrt{\frac{\sum_{b=1}^B (f_b(x') - \hat{f})^2}{B - 1}}$$

In un modello per la previsione del rischio di default, σ assume un significato operativo: un'impresa con un alto score di rischio, inteso come \hat{f} elevata, accompagnato da una bassa

⁸⁸ Malley, J. D., Kruppa, J., Dasgupta, A., & Abhijit, M. (2012). Probability machines: consistent probability estimation using nonparametric learning machines. *Methods of Information in Medicine*, 51(1), p. 74-81.

⁸⁹ Con riferimento alle discrepanze, derivanti da errori, che in realtà non riflettono la situazione strutturale a lungo termine dell'impresa.

deviazione standard indica uno stato di crisi su cui tutti gli alberi concordano. Al contrario, una deviazione standard elevata suggerisce una posizione di incertezza predittiva, tipica di imprese con dati contraddittori che richiedono un approfondimento manuale.⁹⁰

Infine, nel contesto del rischio di insolvenza, l'errore OOB rappresenta un indicatore della capacità del modello di identificare correttamente imprese in stato di default non osservate in fase di addestramento, riducendo il rischio che il modello Random Forest si limiti a memorizzare i casi di default passati piuttosto che apprendere *pattern* strutturali della crisi d'impresa.⁹¹

Ogni albero all'interno della Random Forest cresce attraverso un processo di *scissione binaria ricorsiva*.⁹² Ad ogni nodo, l'algoritmo deve scegliere quale indice di bilancio utilizzare e quale valore soglia impostare (a titolo esemplificativo, *rapporto di Indebitamento* > 0.6). La scelta cade sulla variabile che massimizza la "purezza" dei nodi figli.

Nel caso di una classificazione binaria tra imprese sane e imprese in default, l'impurità di un nodo viene misurata tramite l'Indice di Gini G , definito come:

$$G = \sum_{k=1}^K p_k(1 - p_k) = 1 - \sum_{k=1}^K p_k^2$$

Dove K indica il numero di classi, nel presente contesto K è pari a 2: Impresa sana e impresa in default; e p_k rappresenta la proporzione di osservazioni appartenenti alla classe k all'interno del nodo.

L'indice di Gini varia tra 0 (massima purezza) e 0.5 (massima impurità o incertezza), dove:

- $G = 0$: Il nodo contiene solo aziende di una categoria (es. tutte fallite), indicando che il modello ha trovato una regola perfettamente discriminante;

⁹⁰ Taddy, *op cit.*

⁹¹ Baesens, B., Viaene, S., Van Gestel, T., & Stepanova, M. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), p. 27-635.

⁹² James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R* (2nd ed.). New York: Springer.

- $G = 0.5$: Il nodo contiene per una metà “imprese sane” e per l’altra “imprese fallite”; pertanto, la variabile selezionata non ha alcun potere discriminante.

L'obiettivo della Random Forest è minimizzare l'impurità totale ponderata dei nodi figli. Questo meccanismo riveste un ruolo centrale nell'analisi del rischio di default, in quanto consente al modello di effettuare una forma di selezione automatica delle variabili. In tal senso, gli indici di bilancio che producono una maggiore riduzione dell'impurità di Gini tendono a essere utilizzati più frequentemente e in posizioni più prossime alla radice degli alberi, potendo essere interpretati come le variabili con maggiore capacità discriminante tra imprese sane e insolventi.

Il criterio di Gini è inoltre alla base della misura di importanza delle variabili nota come *Mean Decrease Gini*. Ogni volta che un albero viene suddiviso utilizzando una determinata variabile (ad esempio, il DER), l'impurità del nodo padre risulta superiore alla somma delle impurità dei nodi figli. Tale differenza rappresenta il guadagno di purezza associato a quella variabile.

Aggregando questi guadagni su tutti gli *split* e su tutti gli alberi della foresta, si ottiene una misura dell'importanza relativa di ciascun indice di bilancio nella classificazione tra imprese sane e imprese in default. È opportuno sottolineare che tale metrica fornisce un'indicazione interna al modello della rilevanza delle variabili e non deve essere interpretata come una misura causale.⁹³

A supporto di questa analisi, risulta interessante approfondire un articolo di ricerca del 2017 condotto da Barboza, Kimura e dallo stesso Altman che 49 anni prima aveva sviluppato il primo Z-Score. Attraverso questa ricerca, gli autori contribuiscono in modo significativo al dibattito in letteratura sul confronto tra modelli di ML e tecniche statistiche tradizionali.

Nello specifico, gli studiosi analizzano diversi algoritmi di ML, tra cui la Random Forest, per poi confrontarli con i metodi tradizionali come la MDA e la regressione logistica, integrando delle nuove variabili esplicative rispetto al lavoro originale di Altman (1968).⁹⁴ Mediante un rigoroso lavoro di selezione, costruiscono un training set

⁹³ Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

⁹⁴ Carton, R., & Hofer, C. (2006). *Measuring organizational performance*. Edward Elgar Publishing.

raccogliendo i dati finanziari di imprese dai database Compustat e del Salomon Center (NYU) per un periodo compreso tra il 1985 e il 2005. Il set utilizzato si compone di 449 imprese in default selezionate raccogliendo i dati finanziari dei tre anni precedenti l'istanza di fallimento. D'altra parte, viene selezionato un insieme di pari dimensione di imprese sane, scelte casualmente tra le aziende che non hanno presentato l'istanza nel periodo di riferimento. Si noti come gli autori seguano la stessa metodologia di selezione di Altman (1968), considerando un insieme bilanciato di imprese, fallite e sane, anche al fine di evitare distorsioni legate allo sbilanciamento delle classi in fase di addestramento. Per quanto concerne il *validation set*, si considera un numero decisamente elevato di imprese, 133 in default e 13.167 sane, in un periodo compreso tra il 2006 e il 2013, estratte dagli stessi database. In tal senso gli studiosi seguono la metodologia opposta, più simile a quella utilizzata da Ohlson (1980): considerando che il default è un evento raro, appare più realistico analizzare un set nel quale vi è una netta preponderanza di imprese solventi.

Tabella 2. *Performance dei modelli*. Fonte: Barboza (2017)

Training sample								
Model	TP	TN	FP	FN	Type I Error (%)	Type II Error (%)	AUC (%)	ACC (%)
SVM-Linear	419	306	143	30	6.68	31.85	NA	80.73
SVM-RBF	421	376	73	28	6.24	16.26	NA	88.75
Boosting	434	430	19	15	3.34	4.23	NA	96.21
Bagging	448	447	2	1	0.22	0.45	NA	99.67
Random forest	449	449	0	0	0.00	0.00	NA	100.00
Neural networks	431	331	118	18	4.01	26.28	NA	84.86
Logit	414	329	120	35	7.80	26.73	NA	82.74
MDA	361	221	228	88	19.60	50.78	NA	64.81
Testing sample								
Model	TP	TN	FP	FN	Type I Error (%)	Type II Error (%)	AUC (%)	ACC (%)
SVM-Linear	123	9,389	3,778	10	7.52	28.69	67.2	71.52
SVM-RBF	105	10,505	2,662	28	21.05	20.22	85.17	79.77
Boosting	108	11,417	1,750	25	18.80	13.29	92.97	86.65
Bagging	110	11,284	1,883	23	17.29	14.30	92.48	85.67
Random forest	111	11,468	1,699	22	16.54	12.90	92.92	87.06
Neural networks	124	9,582	3,585	9	6.77	27.23	90.08	72.98
Logit	118	10,028	3,139	15	11.28	23.84	90.10	76.29
MDA	86	6,854	6,313	47	35.34	47.95	63.68	52.18

Come si evince dalla Tabella 2, il modello Random Forest mostra nella fase di *training* una totale assenza di falsi positivi e falsi negativi, comportando una piena accuratezza di classificazione. Questo risultato, come osservato dagli autori, risulta in parte atteso in

quanto il modello utilizza alberi decisionali, che tendono a adattarsi perfettamente ai dati di addestramento. Tuttavia, tale evidenza non garantisce la capacità di generalizzazione dell'algoritmo ed è necessario porre particolare attenzione al rischio di *overfitting*.

Come è possibile desumere dagli esiti della *validation*, la Random Forest rappresenta il modello più performante in termini di accuratezza (87,06 per cento) e la più bassa incidenza di errori di Tipo II (12,90 per cento), e, in relazione all'AUC, risulta secondo solo al *boosting* (92,92 per cento).

Nonostante il modello logit mostri performance discrete, presenta una percentuale elevata di errori di Tipo II che ne compromettono l'affidabilità.

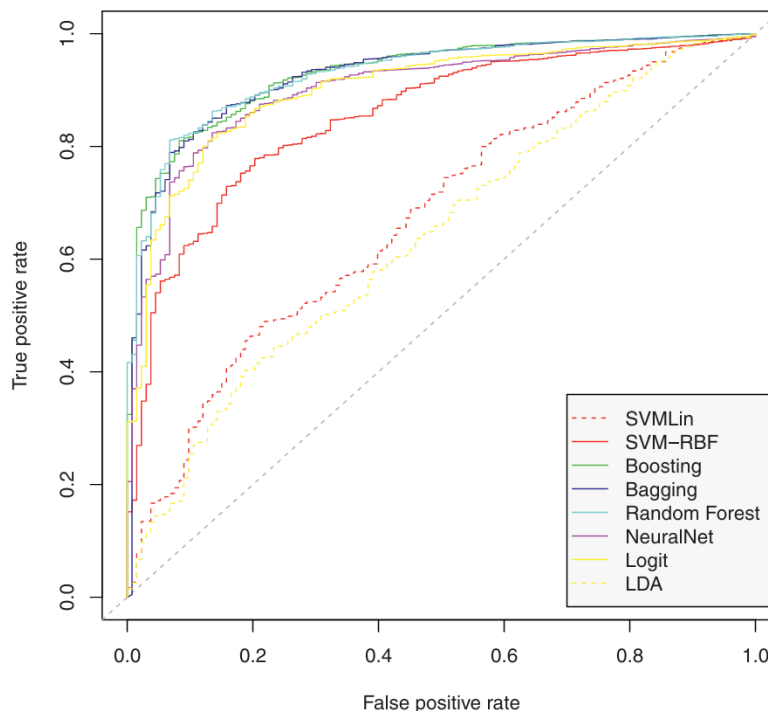


Figura 7. Curve ROC dei modelli utilizzati. Fonte: Barboza (2017)

Infine, con la Figura 7 è possibile apprezzare le curve ROC. Esse permettono di valutare graficamente la capacità discriminante di ciascun modello. I modelli di ML offrono prestazioni nettamente superiori rispetto ai modelli statistici tradizionali, ad eccezione del modello SVM lineare, a conferma che la struttura lineare utilizzata per separare le due

classi (default oppure sane) comporta scarse performance. La MDA, che impiega anch'essa un processo lineare, mostra un limite prestazionale analogo.⁹⁵

2.3.2 L'Extreme Gradient Boosting

Un ulteriore algoritmo d'interesse per la previsione del rischio di default è rappresentato dall'*eXtreme Gradient Boosting* (XGBoost) sviluppato nel 2016 da Chen e Guestrin presso l'Università di Washington. XGBoost si configura come una libreria di ML *open source* basata su alberi decisionali con *gradient boosting*, risultando particolarmente efficace in contesti di classificazione e regressione su dataset complessi.⁹⁶

Sotto il profilo concettuale, il *boosting* rappresenta una tecnica di *ensemble* che presenta delle forti differenze rispetto al *bagging*, tipico della Random Forest. Mentre quest'ultima addestra i singoli stimatori in modo indipendente e parallelo per poi mediane i risultati, gli algoritmi di *boosting* adottano un approccio additivo e sequenziale. Con tale tecnica, i modelli di apprendimento, definiti *weak learners*, vengono addestrati l'uno dopo l'altro con l'obiettivo specifico di correggere gli errori commessi dai predittori precedenti.⁹⁷

Il funzionamento dello XGBoost si basa sull'evoluzione del concetto di *Gradient Boosted Decision Trees* (GBDT). Questa metodologia adotta un approccio iterativo che prende avvio dall'utilizzo di un *base learner*, un primo albero decisionale semplice, il cui compito è fornire una stima iniziale delle osservazioni. Poiché tale modello iniziale presenta inevitabilmente un elevato errore sistematico, l'algoritmo procede in modo additivo: ogni albero successivo viene costruito non sui dati originari, ma sui residui generati dal modello precedente.

I residui rappresentano matematicamente la discrepanza tra i valori predetti e i valori effettivi. In questo contesto, l'obiettivo del modello non è più la semplice classificazione, ma la minimizzazione sistematica di una funzione di perdita (*Loss Function*).⁹⁸ Mentre nei problemi di regressione si ricorre spesso all'errore quadratico medio, nella previsione

⁹⁵ Barboza, *op cit.*

⁹⁶ Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (p. 785-794).

⁹⁷ Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), p. 1189-1232.

⁹⁸ Friedman, J. H. (2002). Stochastic Gradient Boosting. *Computational Statistics & Data Analysis*, 38(4), p. 367-378.

del default, trattandosi di un problema di classificazione binaria, si predilige la *Log-Loss*, definita come:

$$l = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i))$$

Dove y_i è l'etichetta reale dell'osservazione i e \hat{y}_i rappresenta la probabilità stimata di default. La funzione di perdita l misura quindi la discrepanza tra la previsione \hat{y}_i e il valore reale y_i .

Tale funzione penalizza in modo asimmetrico le previsioni ad elevata confidenza ma errate, rendendola particolarmente adatta alla previsione del rischio di fallimento, in cui errori gravi su osservazioni ad alta probabilità di default potrebbero avere implicazioni economiche rilevanti.⁹⁹

Dal punto di vista economico-finanziario, l'adozione del *gradient boosting* è giustificata dalla struttura intrinsecamente non lineare dei determinanti del default. Infatti, le variabili esplicative generalmente utilizzate presentano spesso effetti soglia e interazioni complesse che non possono essere catturate adeguatamente da modelli parametrici lineari come il modello logit. In queste condizioni, l'imposizione di una forma funzionale rigida introduce un errore di specificazione che può tradursi in errori sistematici di valutazione del rischio.¹⁰⁰ La flessibilità degli alberi decisionali, combinata con la natura additiva del *boosting*, consente di individuare le interazioni non lineari tra le variabili (ad esempio tra liquidità e redditività) senza la necessità di specificarle a priori, migliorando significativamente la capacità predittiva e la robustezza del modello di previsione.

Nonostante questi vantaggi, il *gradient boosting* classico presenta limiti rilevanti per le applicazioni alla previsione del default. In particolare, la costruzione sequenziale degli alberi comporta un rischio elevato di *overfitting* in assenza di adeguati meccanismi di regolarizzazione, che talvolta risulta essere superiore rispetto a quello riscontrato nella *random forest*. Inoltre, a causa della sua natura sequenziale, l'algoritmo risulta poco scalabile su dataset di grandi dimensioni. Tali criticità sono particolarmente rilevanti nel

⁹⁹ Xia, Y., Liu, C., LI, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyperparameter optimization for credit scoring. *Expert Systems with Applications*, 78, p. 225-241.

¹⁰⁰ Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge (USA): MIT Press.

contesto bancario, dove è fondamentale che le probabilità stimate siano stabili nel tempo e che il modello mantenga buone prestazioni anche su nuovi dati futuri.

Formalmente, l'algoritmo XGBoost viene addestrato minimizzando una funzione obiettivo regolarizzata che combina la funzione di perdita l con un termine di penalizzazione Ω della complessità degli alberi, secondo quanto segue:

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

Dove la penalizzazione Ω rientra nell'ambito più generale della regolarizzazione, un insieme di tecniche volte a ridurre l'*overfitting*. In genere, la regolarizzazione comporta una riduzione marginale dell'accuratezza dell'addestramento in cambio di un aumento della generalizzabilità del modello. XGBoost include tale metodologia come parte dell'obiettivo di apprendimento, a differenza del classico *gradient boosting*.

Ulteriori metodi permettono un aumento della scalabilità del modello, quali i calcoli svolti in parallelo, oltre ad implementare a sua volta un algoritmo per la gestione dei dati mancanti e dei dati sparsi (si pensi agli zeri all'interno dei dati di bilancio).

Queste caratteristiche rendono XGBoost particolarmente adatto ai fini della stima della probabilità di default, dove le relazioni non lineari e le complesse interazioni tra variabili richiedono modelli flessibili ma controllati per garantire risultati affidabili e interpretabili.

Per quanto concerne l'analisi in esame, XGBoost è senz'altro uno degli algoritmi più recenti e che più hanno beneficiato dei progressi esponenziali in campo tecnologico, grazie anche agli ingenti investimenti del settore privato. L'efficacia del modello nella previsione del rischio di default è stata inizialmente dimostrata in modo sistematico da Xia *et al.* (2017).

Gli autori confrontano un ampio insieme di modelli su molteplici dataset. In particolare, utilizzano cinque set di dati di credito piuttosto eterogenei per verificare le prestazioni dei vari modelli, tre dei quali ricavati da UCI *machine learning repository* rispettivamente di provenienza tedesca, australiana e taiwanese, e due basati sul *peer-to-peer* (P2P) *lending*, in quanto la documentazione relativa alle transazioni su diverse piattaforme di prestito risulta trasparente e liberamente accessibile. Per ogni set vengono inoltre definite

differenti dimensioni, variabili esplicative e bilanciamento tra casi di insolvenza e pagamenti regolari.

Gli studiosi riconoscono l'importanza degli iperparametri dei modelli di ML soprattutto per quanto concerne l'accuratezza, rendendo necessario un lavoro di ottimizzazione degli stessi. Tuttavia, tale processo è estremamente lungo e complesso, dipendendo in larga misura da giudizi soggettivi e da un approccio basato su *trial and error*.¹⁰¹ Gli autori optano così per l'ottimizzazione Bayesiana, la quale consiste nella costruzione di un modello probabilistico della funzione obiettivo che viene utilizzato per selezionare gli iperparametri più promettenti da valutare all'interno della funzione obiettivo reale.

I risultati empirici confermano la robustezza dello XGBoost. L'algoritmo mostra una capacità discriminante superiore, evidenziata da valori di AUC-H e Brier Score (BS) costantemente migliori rispetto ai benchmark. In particolare, nei dataset fortemente sbilanciati, il modello è in grado di identificare i debitori insolventi laddove modelli come SVM e AdaBoost falliscono. I modelli come gli alberi decisionali semplici, le SVM e le reti neurali mostrano una scarsa capacità predittiva. Questi modelli risultano sub-ottimali per il credit scoring e la valutazione del rischio di insolvenza moderno in relazione agli algoritmi di *ensemble*.

Tuttavia, la ricerca evidenzia come la regressione logistica mantenga una validità pratica significativa, soprattutto grazie alla sua semplicità, specialmente quando i dataset sono bilanciati o meno complessi.

Infine, la Random Forest e il GBDT emergono come gli unici reali competitor tra i metodi di *ensemble*.¹⁰²

Un ulteriore contributo di rilievo riguarda il superamento della natura di *black box* tipica dei modelli complessi. L'interpretabilità rappresenta infatti un requisito imprescindibile sia per ridurre la riluttanza del management verso i metodi di *ensemble*, sia per soddisfare i requisiti imposti dalle autorità di vigilanza.¹⁰³ Attraverso il calcolo di misure di *feature importance*, basate sul contributo marginale delle singole variabili alla funzione obiettivo,

¹⁰¹ Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, p. 281-305.

¹⁰² Xia, *op cit.*

¹⁰³ Finlay, S. (2011). Multiple classifier architectures and their application to credit risk assessment. *European Journal of Operational Research*, 210(2), p. 368-378.

XGBoost consente di identificare le variabili più rilevanti ai fini della valutazione del rischio di insolvenza. Tale approccio garantisce un adeguato livello di trasparenza senza compromettere le prestazioni, fornendo agli operatori economici schemi chiari a sostegno dei processi decisionali.

Queste conclusioni trovano un'ulteriore e più recente conferma nel contributo di Ben Jabeur, Carmona e Stef (2022). Gli autori, concentrandosi specificamente sulla previsione del default, propongono un modello ottimizzato di *Feature Selection* (FS-XGBoost), confermando la superiorità dell'algoritmo rispetto ai modelli classici di MDA e alla regressione logistica. L'efficacia dell'algoritmo viene attribuita alla sua capacità di catturare interdipendenze complesse e non lineari tra gli indicatori di bilancio. In linea con quanto sostenuto da Xia *et al.* (2017), anche in questo lavoro la selezione delle variabili riveste un ruolo centrale: l'identificazione dei fattori più informativi, ad esempio gli indici di liquidità e di redditività operativa, consente di costruire modelli non solo più accurati, ma anche più stabili e robusti in presenza di dataset sbilanciati.¹⁰⁴

2.3.3 Le Reti Neurali Artificiali

Un'ulteriore famiglia di algoritmi nel campo dell'apprendimento automatico è rappresentata dalle reti neurali, che rientrano oggi tra i sistemi maggiormente diffusi a livello globale. Le prime intuizioni scientifiche possono essere fatte risalire agli anni '40 e '50 del secolo scorso, e, nel tempo, sono stati sviluppati numerosi modelli, via via più complessi, fino a costituire la base di molti sistemi di IA odierni. Il loro utilizzo è estremamente vario e coinvolge una molteplicità di settori, tra cui la robotica, l'aerospaziale e la medicina. In campo finanziario le reti neurali vengono utilizzate per il rilevamento di frodi e anomalie, il *credit scoring* e la previsione della crisi d'impresa.

L'architettura della rete neurale trae ispirazione dal funzionamento del cervello umano. Le cellule del cervello umano, i neuroni, formano una rete complessa e fortemente interconnessa, attraverso la quale vengono trasmessi segnali elettrici che consentono l'elaborazione delle informazioni. Analogamente, una rete neurale artificiale si compone di neuroni artificiali che operano congiuntamente per risolvere un problema. I neuroni artificiali sono moduli di elaborazione, chiamati nodi, mentre le reti neurali artificiali sono

¹⁰⁴ Ben Jabeur, S., Carmona, P., & Stef, N. (2022). Bankruptcy Prediction using the XGBoost Algorithm and Variable Importance Feature Engineering. *Computational Economics*, 61(2), p. 715-741.

algoritmi o programmi che sfruttano la capacità di calcolo dei sistemi informatici per eseguire operazioni matematiche.¹⁰⁵

Ciò che distingue le reti neurali dagli altri modelli di ML è la loro struttura a strati e la loro capacità di eseguire trasformazioni non lineari.

Una rete neurale si compone di tre strati, o livelli, e di una pluralità di nodi (i neuroni artificiali). Lo strato di input, o *input layer*, riceve dall'esterno i dati iniziali, come i dati finanziari di un'impresa, li analizza e li invia al livello successivo. Gli strati nascosti, o *hidden layers*, possono essere molteplici e ricavano i dati dal livello precedente o da altri strati nascosti. A questo livello i dati vengono elaborati attraverso una trasformazione lineare tra input e output. Infine, nello strato di output, o *output layer*, viene aggiunta una funzione di attivazione al fine di pervenire al risultato finale, come la probabilità di default di un'impresa.

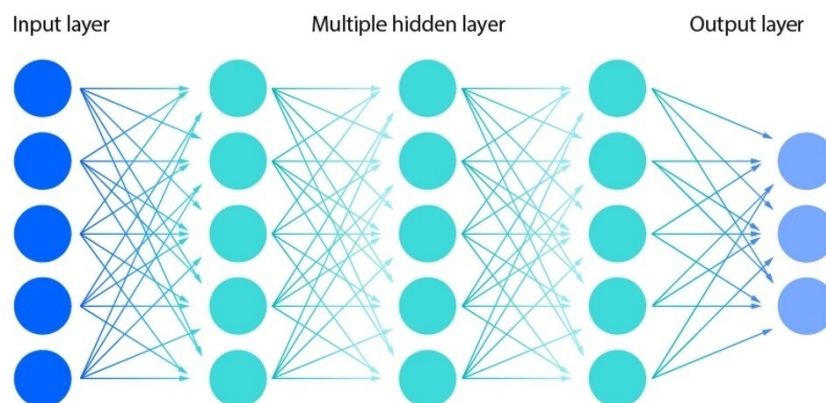


Figura 8. *Struttura di una rete neurale artificiale.* Fonte: Tochaiwat (2025)

Così come per gli altri algoritmi di ML, anche le reti neurali richiedono un processo di addestramento rigoroso per ottenere prestazioni accettabili nei test. Per addestrare una rete, un singolo neurone calcola:

¹⁰⁵ Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge (USA): MIT Press.

$$z = \sum_{i=1}^n w_i x_i + b$$

dove:

- x_i = caratteristica di input;
- w_i = peso;
- b = bias;
- z = somma pesata (trasformazione lineare);
- σ = funzione di attivazione (trasformazione non lineare);
- a = output.

Il simbolo σ rappresenta una funzione di attivazione nello strato di output che trasforma la combinazione lineare per adattarla alla decisione della funzione. Utilizzando questa architettura, le caratteristiche di input X vengono trasformate in un output Y , fungendo da modello predittivo di ML.

La potenza di una rete neurale deriva dalla sua capacità di apprendere i pesi e i *bias* (con riferimento ai parametri che regolano le soglie decisionali del modello) corretti a partire dai dati. Questo avviene minimizzando, come per il *boosting*, una funzione di perdita, utilizzando in questo caso un processo chiamato *backpropagation*: misurando il differenziale tra la previsione e il valore reale, l'errore viene propagato a ritroso attraverso la rete. Per ogni neurone, l'algoritmo calcola quanto ciascun peso e bias abbia contribuito all'errore.¹⁰⁶

Questo processo viene ripetuto più volte sul set di addestramento. Ogni passaggio aiuta la rete a “regolare” i propri parametri interni affinché le previsioni si avvicinino progressivamente ai valori corretti. Con il tempo, la rete converge verso un insieme di pesi e bias che minimizzano l'errore e che favoriscono una buona capacità di generalizzazione.

¹⁰⁶ Freeman, J. A., & Skapura, D. M. (1991). *Neural Networks: Algorithms, Applications, and Programming Techniques*. Reading (USA): Addison-Wesley Professional.

Tuttavia, le reti neurali sono caratterizzate da alcune criticità simili a quelle di altri modelli, la più rilevante delle quali è l'*overfitting*. Quando una rete neurale diventa eccessivamente complessa, con un numero troppo elevato di parametri, il modello tende ad adattarsi in modo eccessivo ai dati di addestramento. Tale problema richiede di prestare attenzione al *trade-off bias-varianza*: mentre una varianza elevata porta all'*overfitting*, un alto bias comporta il rischio opposto, l'*underfitting*, che si verifica quando l'algoritmo è troppo semplice per catturare accuratamente le relazioni esistenti nei dati.¹⁰⁷

Nel contesto della previsione del default, l'equilibrio tra queste due variabili è cruciale: un modello in *underfitting* potrebbe non essere in grado di distinguere eventuali segnali di crisi nei dati contabili, mentre uno in *overfitting* risulterebbe inefficace nell'applicazione a diverse realtà aziendali. Proprio per questo la fase di progettazione del modello è la più complessa e critica.

Poiché le reti neurali sono state introdotte prima di molti altri modelli di ML, come la RF o lo XGBoost, negli anni sono stati condotti molteplici studi circa l'applicazione delle reti neurali alla stima della probabilità di default. I primi possono essere fatti risalire agli inizi degli anni '90, tra i quali figura un'analisi condotta da Altman, Marco e Varetto (1994), i quali effettuano un confronto con la MDA. I risultati indicano un livello di accuratezza estremamente bilanciato tra i due modelli. Tuttavia, emergono una serie di criticità per le reti neurali, tra cui l'*overfitting* e la loro natura di *black box*, tale da suggerire un utilizzo congiunto tra i due modelli.¹⁰⁸

Lo sviluppo di nuove tecniche negli anni porta a una forte riduzione delle problematiche che caratterizzano il modello. In tal senso si colloca un importante studio svolto da Zhang *et al.* (1999). Utilizzando un set bilanciato di 220 imprese (110 in default e 110 sane), confrontano il modello logit con le reti neurali. Per superare le criticità note del modello, gli autori impiegano una tecnica di *cross-validation* attraverso cui esaminare la variabilità delle prestazioni delle reti neurali nella previsione del default su campioni differenti. Questo aumenta notevolmente la capacità di generalizzazione delle reti neurali. Oltre alla

¹⁰⁷ Kohavi, R., & Wolpert, D. H. (1996). Bias Plus Variance Decomposition for Zero-One Loss Functions. *Proceedings of the Thirteenth International Conference on Machine Learning*, (p. 275-283).

¹⁰⁸ Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance*, 18(3), p. 505-529.

maggior robustezza del modello, i risultati indicano che le reti neurali ottengono prestazioni significativamente superiori rispetto ai modelli di regressione logistica, sia in termini di accuratezza predittiva sia di capacità di classificazione (Tabella 3).¹⁰⁹

Tabella 3. *Percentuali di classificazione per i due modelli.* Fonte: Zhang et al. (1999).

Statistics	Overall	
	ANN	Logistic
Mean	86.64	78.55
<i>t</i> -statistic		10.3807
<i>p</i> -value		0.0005

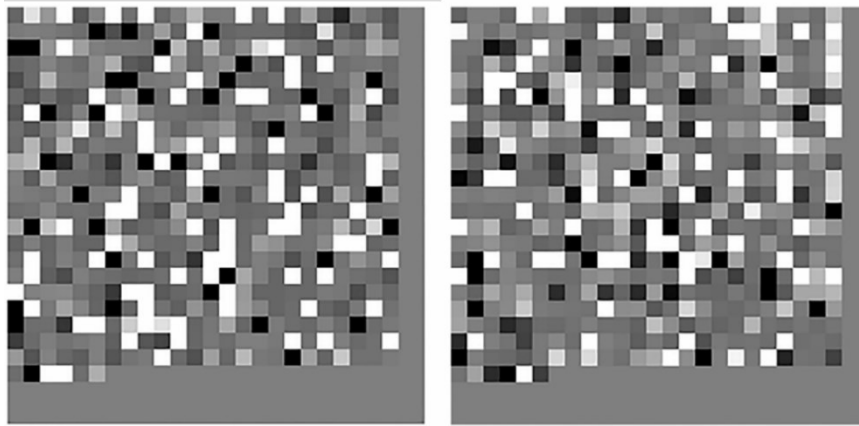
Un utilizzo recente e alquanto singolare delle reti neurali nella previsione del default viene proposto da Hosaka (2019). Nel suo studio l'autore utilizza le reti neurali convoluzionali (CNN), il cui schema di connessione fra i vari neuroni artificiali si ispira all'organizzazione della corteccia visiva biologica. Questo le rende particolarmente adatte per il riconoscimento delle immagini.¹¹⁰

Hosaka raccoglie i dati contabili di 102 società che sono state rimosse dalla borsa giapponese a causa del default e 2062 società sane e regolarmente quotate, in un orizzonte temporale di quattro anni. Con il suo metodo, calcola un insieme di indici finanziari e li rappresenta attraverso un'immagine in scala di grigi. Ogni pixel dell'immagine corrisponde a un determinato quoziente e il suo colore è correlato al valore del rapporto: se superiore alla media il pixel tende al bianco, altrimenti tende al nero. Le immagini generate rappresentano lo stato di salute di una data impresa e vengono utilizzate come input per la rete neurale.¹¹¹

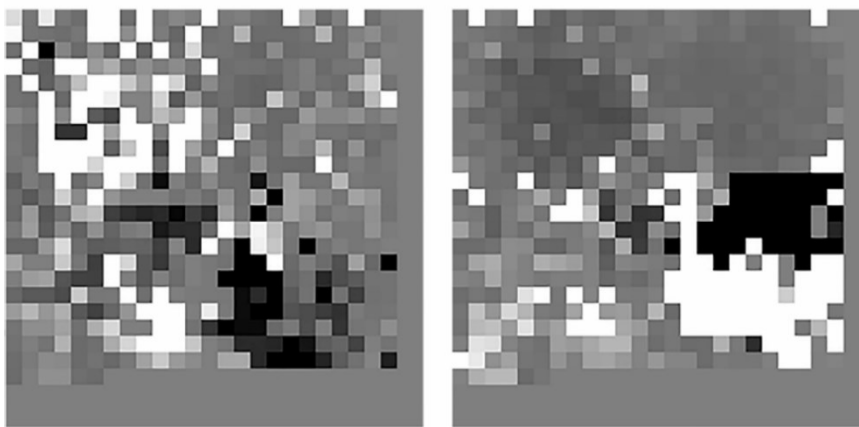
¹⁰⁹ Zhang, G., Hu, M., Patuwo, B., & Indro, D. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116(1), p. 16-32.

¹¹⁰ Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5), p. 555-559.

¹¹¹ Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, 117, p. 287-299.



a) Corrispondenza tra gli indici finanziari e i pixel con il metodo casuale.



b) Corrispondenza tra gli indici finanziari e i pixel con il metodo correlato.

Figura 9. *Rappresentazione degli indici finanziari sotto forma di immagine. A sinistra si trovano le immagini relative a un'impresa in default, e a destra quelle relative all'impresa sana. Fonte: Hosaka (2019)*

Con il metodo casuale gli indici finanziari vengono assegnati alle posizioni dei pixel in modo del tutto casuale, senza seguire una logica economica o statistica. In questo caso, grazie alla sua complessità, l'algoritmo può comunque individuare le relazioni tra le informazioni contenute nei pixel, anche se questi sono molto distanti. Con il metodo correlato invece i quozienti che hanno una forte correlazione tra loro vengono posizionati il più vicino possibile nell'immagine.

Infine, la CNN confronta le immagini cercando di individuare delle strutture comuni. Pertanto, il confronto tra le imprese non avviene in via convenzionale, ma tramite un processo di riconoscimento. La rete neurale impara a distinguere le caratteristiche visive

comuni alle imprese in default rispetto a quelle che contraddistinguono le imprese sane. Come output del processo l' algoritmo restituisce la probabilità di default.

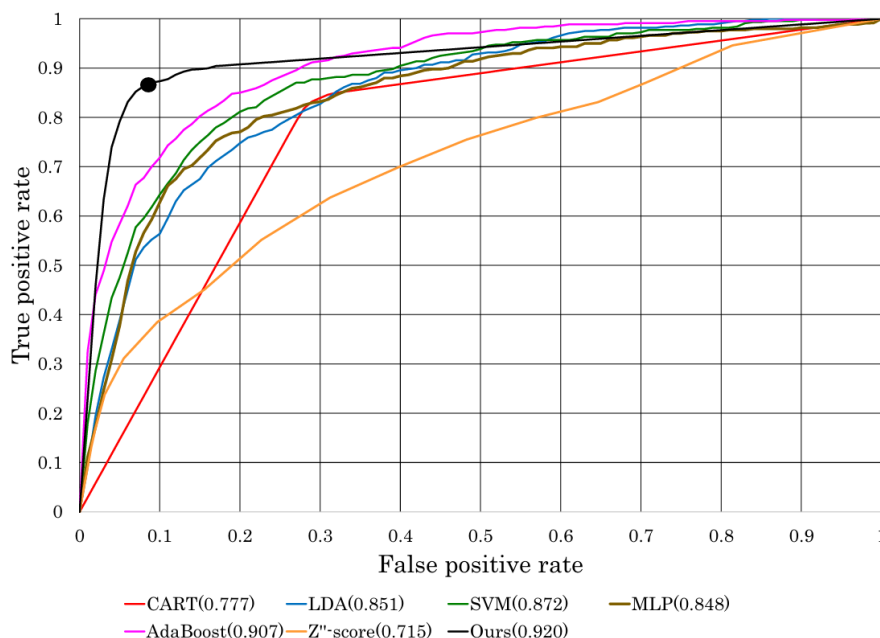


Figura 10. Confronto tra le curve ROC dei modelli analizzati. Fonte Hosaka (2019)

Come si evince dalla figura 10, le previsioni effettuate mediante la rete addestrata presentano delle prestazioni superiori rispetto agli altri metodi confrontati.

Tuttavia, il modello è caratterizzato da un importante limite. A differenza dei metodi convenzionali, con il metodo proposto dall'autore risulta difficile determinare quali tra gli indici finanziari utilizzati siano stati i più significativi. Pertanto, emergono forti dubbi circa la bontà del modello ai fini della previsione del default.

2.4 L'Explainable AI

L'applicazione del ML nel contesto della previsione del default introduce una serie di problematiche che tipicamente non emergono nell'utilizzo dei modelli tradizionali. Tra queste, un ruolo centrale è rivestito dalla spiegabilità dei risultati prodotti dai modelli, aspetto fondamentale affinché l'analista sia in grado di comprendere e comunicare le motivazioni alla base delle decisioni suggerite dall'algoritmo.

Con l'espressione eXplainable AI (XAI) si fa riferimento all'insieme di metodologie e strumenti che consentono di affiancare all'output principale di un modello di ML una serie di informazioni interpretative, volte a chiarirne il funzionamento e le logiche decisionali sottostanti.

Da un punto di vista concettuale, l'XAI può essere ricondotta a due principali dimensioni. La prima è quella dell'interpretabilità (*interpretability*), che mira a tracciare in modo quantitativo i meccanismi interni che governano il comportamento del modello. La seconda è rappresentata dalla spiegabilità (*explainability*), intesa come la formulazione di valutazioni qualitative sui risultati ottenuti, volte a giustificare e rendere comprensibili le logiche di funzionamento del modello.

In generale, nel panorama delle tecniche di ML si osserva come il livello di spiegabilità intrinseca di un modello tenda a ridursi all'aumentare della sua capacità predittiva, configurando un tipico *trade-off* tra performance e trasparenza. A titolo esemplificativo, le reti neurali artificiali sono in grado di raggiungere livelli di accuratezza molto elevati, ma risultano difficilmente interpretabili senza il ricorso a specifiche tecniche di XAI; al contrario, modelli come gli alberi decisionali presentano una struttura intrinsecamente più comprensibile, sebbene spesso a fronte di prestazioni inferiori.

Dal punto di vista applicativo, le tecniche di XAI possono essere distinte in *model-specific*, ovvero utilizzabili solo per determinate classi di modelli (come, ad esempio, le misure di *feature importance* negli alberi decisionali), e *model-agnostic*, applicabili indipendentemente dalla struttura del modello. Tra queste ultime rientrano metodologie ampiamente diffuse quali SHAP e LIME.

In particolare, SHAP (*Shapley Additive Explanations*) è una tecnica di spiegabilità *model-agnostic* che si basa sui c.d. *valori di Shapley*, utilizzati nella teoria dei giochi cooperativi per determinare il contributo marginale che ciascun partecipante ha fornito al risultato complessivo. In ambito ML, tali valori permettono di quantificare il contributo di ciascuna variabile esplicativa all'output del modello.

LIME (*Local Interpretable Model-Agnostic Explanations*), invece, si propone di spiegare le predizioni del modello approssimandone il comportamento in prossimità di una

specifica osservazione, attraverso la variazione controllata delle feature e l'analisi dell'impatto prodotto sulle previsioni.¹¹²

L'adozione di tecniche di XAI può avvenire in diverse fasi del ciclo di vita di un modello. In particolare, esse possono essere utilizzate prima della fase di modellizzazione, per favorire la comprensione dei dati impiegati nell'addestramento (*pre-modeling explainability*); durante lo sviluppo del modello, con l'obiettivo di migliorarne l'interpretabilità (*explainable modeling*); oppure successivamente all'addestramento, al fine di chiarire ex post le logiche decisionali del sistema (*post-modeling* o *post hoc explainability*).¹¹³

Alla luce di tali considerazioni, l'utilizzo di tecniche per la spiegabilità risulta un elemento essenziale per la valutazione dei modelli di ML applicati alla previsione del rischio di default, per massimizzare l'utilità e la trasparenza per i vari operatori coinvolti.

La letteratura scientifica concorda sul fatto che non esista un modello universalmente superiore. L'efficacia di un algoritmo è strettamente legata alle caratteristiche peculiari del dataset analizzato, alla qualità del *feature engineering* e alla precisione della taratura degli iperparametri.¹¹⁴

¹¹² Bagli, E., Visani, G., Chesani, F., & Poluzzi, A. (2020). Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models. *Journal of the Operational Research Society*, p. 1-18.

¹¹³ Moscatelli, M., Parlapiano, F., & Cascarino, G. (2022). *Explainable Artificial Intelligence: interpreting default forecasting models based on Machine Learning. Temi di discussione n. 674*. Banca.

¹¹⁴ Wolpert, D. H., & Macready, W. (1997). No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), p. 67-82.

Capitolo 3

Analisi empirica

3.1 Introduzione all'analisi

Il presente capitolo è dedicato all'analisi empirica e rappresenta il passaggio dalla rassegna della letteratura esposta nei primi due capitoli all'applicazione pratica.

L'obiettivo è confrontare l'approccio tradizionale ai fini della previsione del rischio di default con il nuovo approccio basato sul ML, utilizzando un dataset reale. Nello specifico, si confronta il modello Altman Z'-Score (1983), finalizzato all'applicazione a imprese non quotate, con i modelli basati sugli algoritmi di ML, Random Forest e XGBoost. L'analisi mira a valutare le loro prestazioni predittive, mantenendo, per i modelli di ML, la massima interpretabilità e trasparenza possibile, anche attraverso lo studio dei *valori di Shapley*. Si cerca in tal senso di far fronte alle esigenze di *compliance* legate al quadro regolamentare definito dalle norme di Basilea, dall'EBA e dall'attuale AI Act nei confini dell'Unione europea.

L'idea di fondo è quella di verificare se la flessibilità degli algoritmi possa offrire prestazioni più elevate in relazione al rilevamento dei segnali di default rispetto ai classici modelli statistici con parametri predeterminati.

In particolare, l'applicazione dei modelli viene svolta all'interno dell'ambiente Google Colab, una piattaforma cloud che consente l'accesso a risorse computazionali potenti. Il codice è basato sul linguaggio di programmazione Python 3, il quale permette l'accesso a un'ampia gamma di librerie. Per questo lavoro, sono stati importati i moduli *Pandas* e *NumPy* per la manipolazione dei dati, *Scikit-Learn* per la costruzione dei modelli di ML e *Matplotlib* e *Seaborn* per la creazione dei grafici. Per l'implementazione del modello XGBoost è stata importata l'omonima libreria.

L'architettura del capitolo è strutturata in modo da garantire una comprensione completa e trasparente di tutto il processo di analisi, anche attraverso la condivisione di codici Python e schermate dell'ambiente di sviluppo.

3.2 Metodologia

3.2.1 Descrizione del dataset

I dati utilizzati per l'analisi sono stati raccolti dal database Orbis Global di Bureau van Dijk, ora parte di Moody's Analytics. Il dataset comprende i dati finanziari di 6911 imprese italiane nel periodo compreso tra il 2016 e il 2019. Al fine di garantire quanta più omogeneità possibile, sono state considerate solamente le PMI del settore manifatturiero (codici NACE Rev.2 10-33) sotto forma di S.r.l. e S.p.A.

La scelta del periodo non è casuale, ma è motivata da alcune argomentazioni di carattere macroeconomico. In primo luogo, si è scelto di escludere il periodo successivo al 2019 al fine di evitare le distorsioni derivanti dalla pandemia di COVID-19 e la conseguente forte recessione che ha caratterizzato l'area Euro. La crisi sanitaria ha provocato una forte contrazione della domanda aggregata nel biennio 2020-2021, seguita da una rapida ripresa, resa possibile dal progressivo allentamento delle restrizioni. Di contro, il lato dell'offerta aggregata non è riuscito a replicare il forte aumento della domanda sia per la riduzione dei livelli di produzione a causa del periodo della pandemia sia per il forte aumento dei prezzi dell'energia conseguente allo scoppio della guerra in Ucraina nel 2022. Le crescenti tensioni tra l'Unione europea e la Russia hanno generato uno shock dell'energia, comportando un elevato aumento dell'inflazione che ha indotto la BCE ad adottare una politica monetaria restrittiva, dopo anni di orientamento espansivo.¹¹⁵

Parallelamente, si è voluto escludere il periodo relativo alla crisi 2007-2008 e alla crisi del debito sovrano europeo che hanno comportato una forte contrazione dell'attività economica e l'adozione di rigide politiche di *austerità*. A partire dal 2014, l'introduzione del Piano di Investimenti per l'Europa, il c.d. Piano Juncker, promosso dall'omonimo presidente della Commissione europea, ha contribuito a invertire la tendenza al calo degli investimenti che ha caratterizzato gli anni di crisi e ha facilitato l'accesso al credito per le PMI senza gravare sulle finanze pubbliche nazionali e sui livelli di debito, già troppo

¹¹⁵ Ascari, G., Bonomolo, P., Hoerberichts, M., & Trezzi, R. (2023). *The euro area great inflation surge*. Amsterdam: De Nederlandsche Bank (DNB).

elevati. Tale intervento ha contribuito a migliorare gradualmente il quadro economico all'interno dell'Unione europea.¹¹⁶

Pertanto, il periodo selezionato risulta caratterizzato da una relativa stabilità economica.

Considerando la rarità dell'evento del default in relazione al numero totale di imprese che compongono una data popolazione, il campione di 6911 imprese utilizzato si compone di 318 imprese fallite e 6593 imprese sane, coerentemente con lo sbilanciamento che caratterizza i vari campioni utilizzati nei paper scientifici analizzati nel secondo capitolo. Pertanto, il peso delle imprese in default della presente analisi è del 4,60 per cento.

I dati sono stati estratti dal database in Microsoft Excel. A seguito di una prima operazione di pulizia, sono state calcolate all'interno del foglio di lavoro 29 variabili incluse le cinque utilizzate nello score di Altman, indicate nella Tabella 4.

¹¹⁶ Commissione Europea. (2014). *Un piano di investimenti per l'Europa*. COM 913 final, Bruxelles.

Tabella 4. *Variabili utilizzate nell'analisi.* Fonte: elaborazione propria.

VARIABILE	DESCRIZIONE
CCN_TA	Capitale Circolante Netto / Totale Attività
UN_TA	Utile Netto / Totale Attività
EBIT_TA	Ebit / Totale Attività
PN_TP	Patrimonio Netto / Totale Passività
REV_TA	Ricavi / Totale Attività
AC_PC	Attività Correnti / Passività Correnti
LIQ_PC	Disponibilità Liquide / Passività Correnti
EBITDA_REV	Ebitda / Ricavi
EBT_TA	Utile Ante Imposte / Totale Attività
DER	Debiti / Patrimonio Netto
PCPF_TA	Passività Correnti E Fisse / Totale Attività
DFOR_TP	Debiti Vs Fornitori / Totale Passività
CCLI_REV	Crediti Vs Clienti / Ricavi
RIM_TA	Rimanenze / Totale Attività
IMM_TA	Immobilizzazioni / Totale Attività
Num_Dip	Numero Medio Di Dipendenti
LIQ_TA	Disponibilità Liquide / Totale Attività
CL_REV	Costo Del Personale / Ricavi
CCN_REV	Capitale Circolante Netto / Ricavi
D_TA	Debiti / Totale Attività
PC_PT	Passività Correnti / Totale Passività
REV_CCLI	Ricavi / Crediti Vs Clienti
REV_RIM	Ricavi / Rimanenze
EBITDA_TA	EBITDA / Totale Attività
LIQ_D	Disponibilità Liquide / Debiti
PN_D	Patrimonio Netto / Debiti
REV_Dip	Ricavi Per Dipendente
UN_Dip	Utile Netto Per Dipendente
ROE	Utile Netto / Patrimonio Netto

La selezione delle variabili è stata guidata da un duplice approccio. Da un lato, sono state considerate quelle più rilevanti secondo la letteratura, con particolare riferimento al lavoro di Ben Jabeur, Carmona e Stef (2022). Dall'altro, la raccolta è stata integrata e adattata sulla base della disponibilità dei dati nel database Orbis.

L'operazione si conclude con l'inserimento della variabile dicotomica con cui si indica lo *status* assunto dalle imprese: 0 per le sane e 1 per quelle in default (Figura 11).

	A	B	C	D	E	F	G	H	I
1	Num_Imp	Default	CCN_TA	UN_TA	EBIT_TA	PN_TP	REV_TA	AC_PC	LIQ_PC
2	1	1	-0.2245	-0.3576	-0.3115	-0.3082	0.9442	0.7695	0.1017
3	2	1	-0.3546	-0.2551	-0.2272	0.0275	0.8124	0.5058	0.0018
4	3	1	-0.1931	-0.3997	-0.3887	-0.2177	1.4821	0.8347	0.0639
5	4	1	-0.9118	0.0571	0.0584	-0.884	0.6478	0.3683	0.0149
6	5	1	-0.3158	-0.6147	-0.5527	-0.1797	0.9102	0.6388	0.0009
7	6	1	-0.0623	-0.2928	-0.26	-0.2904	1.5437	0.9237	0.0207
6906	6905	0	0.364	0.0019	0.0137	-0.045	1.1806	3.1831	1.1201
6907	6906	0	0.0218	0.0363	0.0755	0.083	1.2077	1.025	0.0485
6908	6907	0	0.3455	0.1117	0.1374	0.2538	1.1095	1.9228	0.15
6909	6908	0	0.4205	0.0879	0.1269	0.7312	1.1631	3.6159	0.9867
6910	6909	0	0.3435	0.1998	0.1978	0.3045	1.3044	1.5347	0.4036
6911	6910	0	0.2456	0.0868	0.1249	0.5022	0.7236	1.7171	0.4566
6912	6911	0	0.381	0.0948	0.1309	0.6304	1.0767	2.5903	0.5563

Figura 11. *File Excel completo*. Fonte: elaborazione propria.

Al fine di evitare il manifestarsi di uno squilibrio pregiudizievole tra le prestazioni dei vari modelli, soprattutto per lo Z'Score di Altman, i dati applicati vertono sulla previsione del fallimento a un anno. In particolare, è noto come il modello di Altman sia caratterizzato da una perdita notevole delle performance predittive già dal secondo anno che precede il default.

3.2.2 Data pre-processing

Per applicare gli algoritmi di ML, è necessario suddividere il dataset in *training set* e *testing set*. Il primo fa riferimento ai dati che vengono utilizzati dal sistema per addestrare i modelli mentre il secondo si riferisce ai dati sui quali verrà utilizzato il modello

addestrato, al fine di verificare la capacità di generalizzazione. Nella presente analisi, il 70 per cento del dataset viene utilizzato ai fini dell'addestramento e il 30 per cento per il test.¹¹⁷

Inoltre, nella suddivisione è necessario mantenere la stessa proporzione tra imprese sane e imprese in default, e questo viene fatto attraverso la strategia della stratificazione.¹¹⁸ Il modello di Altman, per garantire uniformità nell'analisi, viene applicato esclusivamente al *testing set*.

Per quanto concerne i valori mancanti, è stata riscontrata la loro totale assenza per le cinque variabili dello Z'-Score.

Per il modello Random Forest, si è optato per la sostituzione dei valori mancanti con la tecnica dell'imputazione mediante mediana.¹¹⁹ Questo calcolo è stato effettuato esclusivamente sul *training set*, al fine di evitare il fenomeno del “*data leakage*”. Tale metodo garantisce che il modello non riceva alcuna informazione, diretta o indiretta, riguardante la distribuzione dei dati di test durante la fase di apprendimento. In questo modo, si simula correttamente uno scenario reale: la mediana calcolata sul *training set* viene utilizzata per imputare i dati mancanti del *testing set*, assicurando che il sistema operi sulla base delle informazioni precedentemente acquisite. Nonostante il modello XGBoost preveda una soluzione nativa per la gestione dei valori mancanti, si è deciso di adottare anche per esso la strategia del calcolo della mediana, garantendo così che ciascun modello operi sulla stessa base di informazioni.

Considerata la natura del database utilizzato, la fase più importante del *data pre-processing* ha riguardato la gestione dei valori anomali. Nei database finanziari, è comune riscontrare degli indici di bilancio con valori fortemente elevati o distorti, spesso causati dal denominatore prossimo allo zero o da errori in fase di rilevazione dei dati. Per poter mitigare l'impatto di tali anomalie, è stata utilizzata la tecnica della *winsorizzazione* al 1° e al 99° percentile. Attraverso questo metodo i valori situati al di fuori di tale intervallo

¹¹⁷ La suddivisione è stata eseguita tramite la funzione `train_test_split`. Codice: `X_train_all, X_test_all, y_train, y_test = train_test_split(X_all, y, test_size=0.30)`.

¹¹⁸ A tal fine, viene impostato il parametro `stratify=y`.

¹¹⁹ Mediante il codice: `df = df.fillna(df.median(numeric_only=True))`.

(ovvero i valori estremi) vengono sostituiti con il valore corrispondente alla soglia. In questo modo non si elimina l'osservazione, ma se ne limita l'eccesso.

Nello specifico, tale metodologia si rende necessaria a causa della natura additiva del modello Z'-Score. Infatti, la presenza di un singolo valore estremo in una delle cinque variabili comporterebbe un'elevata distorsione del punteggio, invalidando la bontà del modello. Sebbene i modelli di ML riescano a gestire meglio i valori estremi, la *winsorizzazione* viene applicata in tutti i modelli per garantire un confronto *ceteris paribus*.

Con lo Z'-Score, la tecnica viene applicata all'intero set di dati, mentre con i modelli di ML viene calcolata sul *training set* e infine applicata al *testing set* per evitare, anche in questo caso, il *data leakage*.¹²⁰

Uno dei principali limiti dei modelli statistici tradizionali, come lo Z'-Score, risiede nella loro totale incapacità di adattarsi a set di dati fortemente sbilanciati. Di contro, gli algoritmi analizzati permettono una gestione dello sbilanciamento dei dati mediante l'utilizzo di pesi differenziati. Ad esempio, in XGBoost si inserisce il parametro *scale_pos_weight*, mentre in Random Forest si utilizza *class_weight*. Questi permettono di applicare una "penalità" agli errori commessi sui default, indirizzando il modello a dare priorità alla sensibilità predittiva verso le imprese a rischio.

Infine, per i modelli di ML si ricorre alla metrica della *feature importance*: non vengono utilizzate tutte le 29 variabili, ma, attraverso un calcolo, il modello seleziona e utilizza le 10 più rilevanti. Tale aspetto viene approfondito nel sottoparagrafo relativo all'analisi di ciascun modello.

3.2.3 Metriche di valutazione

Per la presente analisi vengono utilizzate metriche standardizzate per la valutazione dei modelli. Non potendo utilizzare esclusivamente l'accuratezza, che potrebbe risultare fuorviante a causa delle classi sbilanciate, è necessario ricorrere a ulteriori strumenti. Tra

¹²⁰ Applicando il seguente codice:

```
for col in X_train_all.select_dtypes(include=[np.number]).columns:  
    lower = X_train_all[col].quantile(0.01)  
    upper = X_train_all[col].quantile(0.99)  
    X_train_all[col] = X_train_all[col].clip(lower, upper)  
    X_test_all[col] = X_test_all[col].clip(lower, upper).
```

questi, il più comune è la matrice di confusione, la cui struttura si può evincere dalla Tabella 5.

Tabella 5. *Matrice di confusione*. Fonte: elaborazione propria.

	Previsione Sana (0)	Previsione Default (1)
Reale Sana (0)	TN	FP
Reale Default (1)	FN	TP

Dove:

- *True Negative* (TN): sono le imprese sane classificate correttamente;
- *True Positive* (TP): sono le imprese in default classificate correttamente;
- *False Negative* (FN): rappresentano le imprese in default che il modello ha erroneamente classificato come sane. Questo configura l'errore di Tipo II;
- *False Positive* (FP): indicano le imprese sane che il modello ha erroneamente classificato come in default. Questo configura l'errore di Tipo I.

Si noti come nel contesto del rischio di credito un errore di Tipo II, nel singolo caso, assuma una maggiore rilevanza rispetto a quello di Tipo I. Mentre quest'ultimo comporta un costo opportunità legato al mancato affare, il primo si traduce in una perdita totale o parziale di quanto prestato al debitore che si rivelerà insolvente.

Un altro tra gli strumenti più diffusi è la curva ROC (Receiver Operating Characteristic) attraverso cui si studia il *trade-off* tra TPR e FPR, dove:

- True Positive Rate (TPR) è la percentuale di positivi che sono stati classificati correttamente dal modello, espressa come:

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

- False Positive Rate (FPR) rappresenta la percentuale di negativi che sono stati erroneamente classificati come positivi, data da:

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

Attraverso il calcolo dell'AUC (*Area Under the Curve*) è possibile confrontare le prestazioni dei modelli.¹²¹ Questa metrica va da 0,5, il caso in cui il modello non ha capacità discriminante, al pari di una scelta casuale, fino a 1, dove il modello risulta essere perfettamente discriminante. Per definizione, l'AUC è l'integrale della curva ROC:

$$AUC = \int_0^1 f(x)dx$$

Infine, si utilizza l'F1-Score, che rappresenta la media armonica tra *precision* e *recall*:

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Dove *recall* è pari al TPR e *precision* si esprime con il seguente rapporto:

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

3.3 Altman Z'-Score (1983)

Il primo modello ad essere testato è lo Z'-Score (1983). Vengono selezionati i cinque indicatori finanziari per poi moltiplicarli per i coefficienti previsti dalla formula.¹²² Per quanto concerne la soglia per il default, si utilizza quella prevista dall'autore pari a 1,23. I valori inferiori a tale punteggio denotano un alto rischio di default e le imprese vengono classificate come tali, assumendo il valore 1. Tutte le altre, anche quelle appartenenti alla *gray zone*, vengono classificate come sane, assumendo il valore 0.¹²³

Infine, si ottengono i seguenti risultati:

¹²¹ Bradley, A. P. (1997). The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition*, 30, p. 1145-1159.

¹²² La formula viene applicata attraverso la seguente stringa:

```
test_df['Z_Score'] = 0.717*X1 + 0.847*X2 + 3.107*X3 + 0.420*X4 + 0.998*X5
Con: X1 = test_df['CCN_TA'] X2 = test_df['UN_TA'] X3 = test_df['EBIT_TA'] X4 =
test_df['PN_TP'] X5 = test_df['REV_TA'].
```

¹²³Per discriminare tra le due classi si applica il seguente codice:

```
test_df['Prediction'] = np.where(test_df['Z_Score'] < 1.23, 1, 0).
```

True Positives (TP): 90
 False Positives (FP): 433
 True Negatives (TN): 1546
 False Negatives (FN): 5

 Accuracy: 0.7888
 Precision: 0.1721
 Recall: 0.9474
 F1-Score: 0.2913
 AUC-ROC: 0.9553

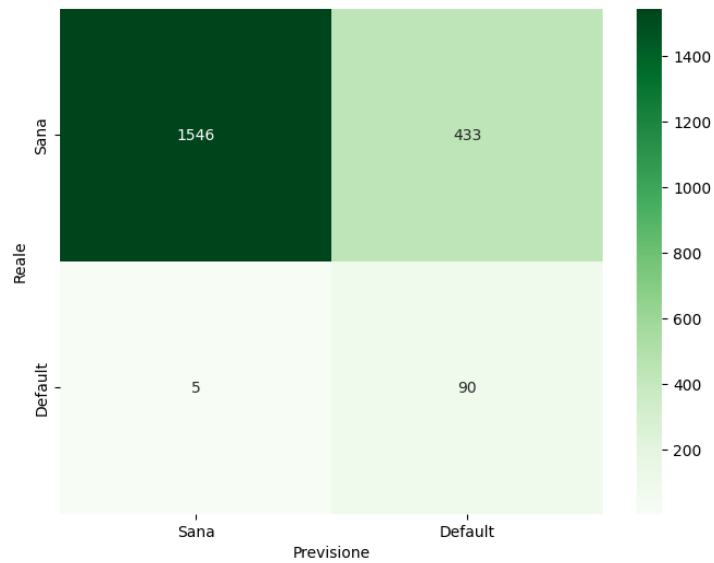


Figura 12. Matrice di confusione - Altman Z'-Score.

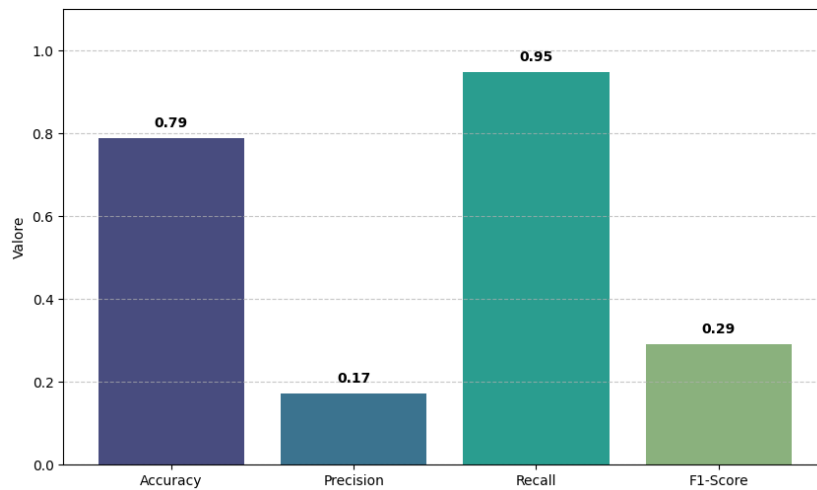


Figura 13. Prestazioni - Altman Z'-Score

Il modello individua 90 default su 95 totali, mostrando una forte capacità di intercettazione del rischio. Tuttavia, classifica erroneamente ben 433 imprese sane come default.

La *recall* è elevata, il modello riesce ad identificare il 94,74 per cento delle imprese fallite, riducendo i falsi negativi a un numero contenuto. Di contro, la *precision* è estremamente bassa: solo il 17,21 per cento delle imprese classificate come in default lo è effettivamente. Ne segue che il modello genera un numero elevato di falsi positivi, tendendo a sovrastimare il rischio, con un *false positive rate* pari a 21,88 per cento. Il modello pertanto risulta fortemente prudentiale e poco selettivo. L’F1-Score, essendo penalizzato dallo squilibrio tra *recall* e *precision*, si rivela piuttosto basso (29,13 per cento).

Sebbene l’accuratezza possa apparire discreta (78,88 per cento), tale metrica perde di significatività per via del dataset sbilanciato e per la netta preponderanza delle imprese sane.

Per comprendere meglio il comportamento del modello e i risultati ottenuti, è utile osservare come la soglia imposta da Altman si posiziona rispetto alla distribuzione dei punteggi (Figura 14):

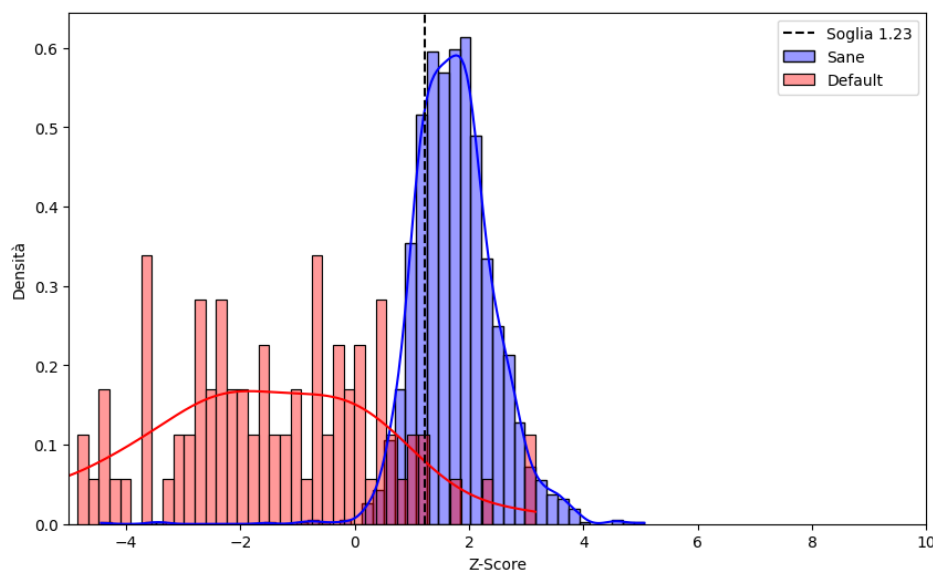


Figura 14. Distribuzione Z'-Score: Imprese sane vs default.

Il grafico permette di spiegare visivamente il motivo della forte differenza tra *recall* e *precision*. La soglia è infatti posizionata in modo tale da catturare la quasi totalità della distribuzione delle imprese in default, localizzata prevalentemente a sinistra della linea tratteggiata; questo spiega il numero basso di falsi negativi. Tuttavia, una parte significativa delle imprese sane ricade anch'essa a sinistra della soglia, comportando un numero elevato di falsi positivi.

L'analisi può essere approfondita visionando il seguente diagramma degli estremi e dei quartili (figura 15) per meglio comprendere come ogni variabile che compone la formula contribuisca alla sovrapposizione tra le classi.

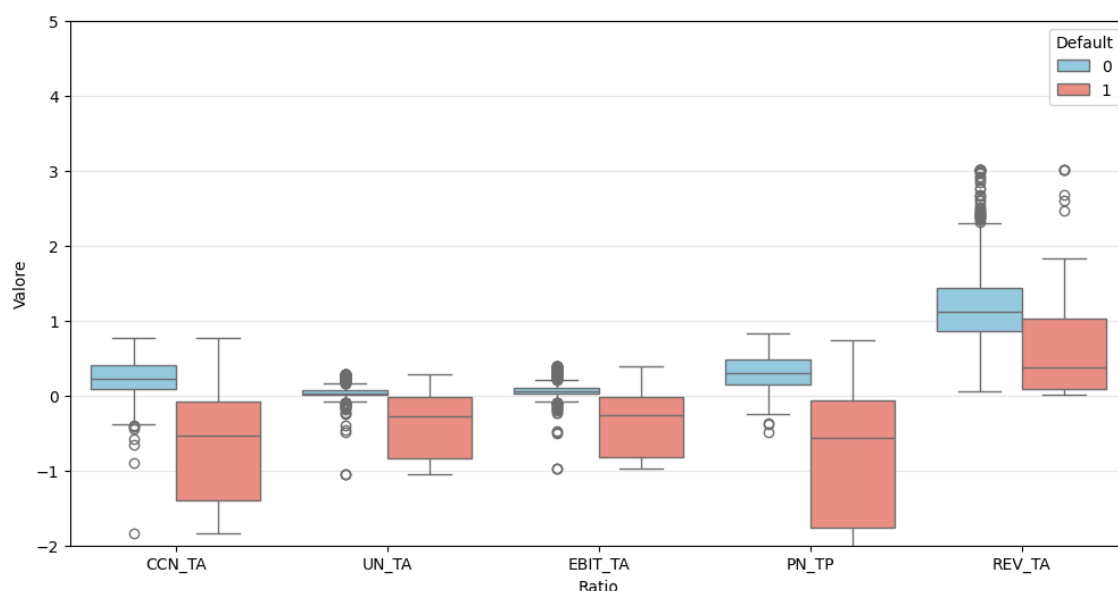


Figura 15. Confronto delle variabili dell'Altman Z'-Score.

Si noti come, in tutti gli indici finanziari, gli estremi (o “baffi”) inferiori delle imprese sane si sovrappongono in modo più o meno importante con i valori delle imprese in default. Ne segue che se un'impresa sana si trova nella fascia bassa di due o tre variabili contemporaneamente, il valore del suo Z'-Score tende a scendere sotto la soglia.

L'indice più critico è il REV_TA, o rotazione del capitale investito, dove i box¹²⁴ sono parzialmente sovrapposti. Questo denota, nella presente analisi, una capacità simile tra le

¹²⁴ I box rappresentano il 50 per cento dei valori centrali di una distribuzione.

imprese sane e in default di generare ricavi rispetto all'attivo, e tale elemento tende a confondere il modello.

La variabile più affidabile risulta PN_TP, che esprime il peso del patrimonio netto rispetto al totale del passivo, i cui box delle due classi sono distanti. Tuttavia, la sovrapposizione degli estremi comporta un margine di errore: esistono comunque imprese sane con un rapporto basso che il modello scambia per imprese in default.

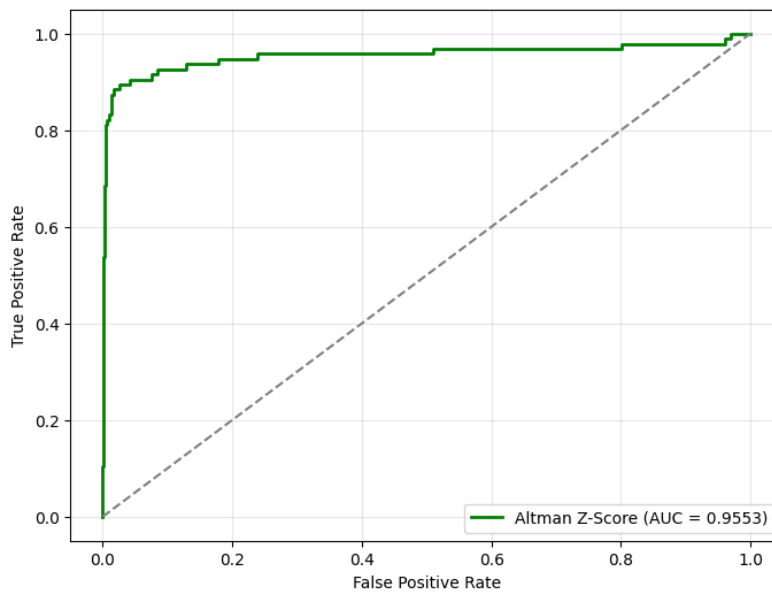


Figura 16. *Curva ROC - Altman Z'-Score.*

Il modello mostra un valore di AUC pari a 0,9553 evidenziando un'eccellente capacità discriminante in relazione al dataset utilizzato. Tuttavia, l'elevata AUC può risultare eccessivamente ottimistica in presenza di classi sbilanciate, rendendo necessaria l'integrazione dell'analisi con la curva *precision-recall*, che sposta il *focus* sulla qualità dei segnali di allerta in presenza di eventi rari.

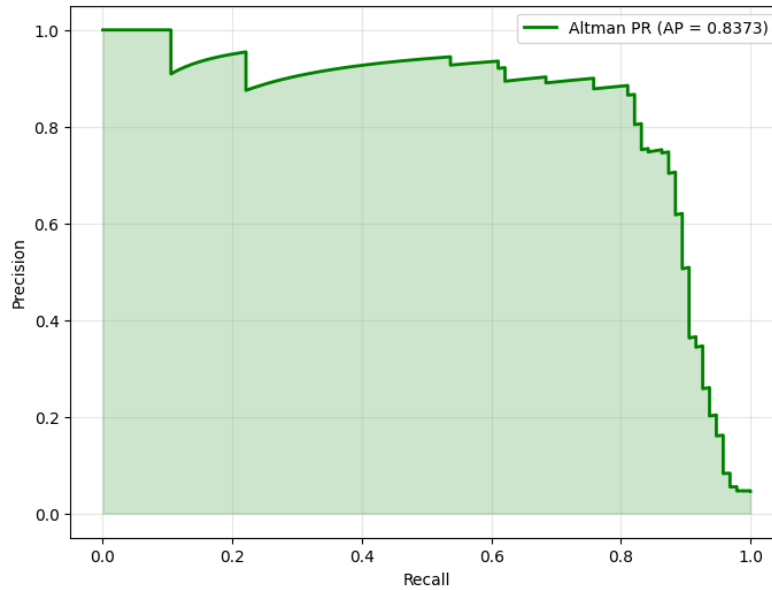


Figura 17. Curva Precision-Recall - Altman Z'-Score.

Come si può evincere dalla Figura 17, il valore dell'*average precision* è pari a 0,8373, evidenziando una robustezza complessiva molto positiva per un modello lineare tradizionale. Tuttavia, l'andamento della curva rivela un *trade-off* critico. Infatti, fino a un livello prossimo all'80 per cento di *recall*, il modello mantiene una *precision* elevata, mostrando una buona capacità di riconoscere i casi di default più evidenti. Oltre l'80 per cento, però, si osserva una forte flessione della curva. Per portare la *recall* a quasi il 95 per cento, la *precision* cala drasticamente.

Questo conferma visivamente quanto emerso dall'analisi dei box: la parziale sovrapposizione tra le distribuzioni delle classi rende impossibile l'identificazione della totalità dei default senza includere nel perimetro di rischio un numero elevato di imprese sane. In termini operativi, ciò implica che il tentativo di minimizzare gli errori di Tipo II tramite lo Z'-Score comporta un inevitabile aumento dei costi legati ai falsi positivi.

Il valore dell'AUC, pur posizionandosi nella fascia alta dei valori riscontrati in letteratura, in cui solitamente le applicazioni su campioni eterogenei evidenziano delle prestazioni più contenute, trova una solida giustificazione nella composizione del dataset utilizzato. A tal fine, è stata condotta un'indagine approfondita sui singoli indicatori del modello, attraverso il calcolo dell'AUC di ciascun indice:

```

ANALISI DI INTEGRITÀ:
-----
X1 (CCN_TA)      | AUC: 0.8961 | AP: 0.6962
X2 (UN_TA)      | AUC: 0.9262 | AP: 0.6840
X3 (EBIT/TA)    | AUC: 0.9298 | AP: 0.6782
X4 (PN/TP)      | AUC: 0.9497 | AP: 0.8277
X5 (REV/TA)     | AUC: 0.7731 | AP: 0.5333
-----
Distribuzione classi nel Test Set (Default %):
Default
0      0.954195
1      0.045805
Name: proportion, dtype: float64
-----
AUC-ROC TOTALE ALTMAN: 0.9553

```

Come si può desumere dal risultato, le prestazioni del modello sono trainate dagli indici EBIT_TA e PN_TP. Pertanto, si procede con un'analisi dei relativi valori, secondo quanto segue:¹²⁵

```

ANALISI DEI VALORI CRITICI:
-----
                EBIT_TA                PN_TP
                mean median   min    max   mean median   min    max
Default
0      0.069377  0.0502 -0.97971  0.38671  0.331756  0.2915 -2.15953  0.83434
1      -0.383618 -0.2661 -0.97971  0.38671 -0.847740 -0.5671 -2.15953  0.73750
-----
Aziende in Default con PN_TP <= 0: 73 su 95 (76.84%)

```

Dall'analisi emerge una netta discrepanza tra i valori delle due classi. Nello specifico, per quanto concerne l'indice PN_TP si osserva che la mediana delle imprese in default è di -0,5671 a fronte di un valore positivo di 0,2915 per le imprese sane. Questo fa sì che il modello sia in grado di distinguere efficacemente i due gruppi. Tale marcata divergenza trova la sua spiegazione più significativa nell'osservazione del campione delle imprese in default: come si desume dai risultati, il 76,84 per cento delle imprese della classe è giunto alla rilevazione con un patrimonio netto negativo o nullo. Questo fenomeno non deriva da un'anomalia statistica, ma trova riscontro nelle dinamiche che caratterizzano le PMI italiane. Infatti, come evidenziato dal CNDCEC, circa il 5,5 per cento delle imprese che avevano presentato almeno uno dei bilanci nel triennio 2017-2019 mostrava un patrimonio netto negativo pur essendo in condizioni di continuità aziendale.¹²⁶ Sebbene

¹²⁵ Si ricorda che i valori estremi sono uguali per via della tecnica di *winsorizzazione* applicata in fase di data pre-processing.

¹²⁶ CNDCEC. (2021). *Osservatorio sui bilanci delle srl 2019 trend 2017-2019*. Roma: Fondazione Nazionale dei Commercialisti.

il deficit patrimoniale possa risultare come un fenomeno statisticamente limitato se riferito alla generalità delle imprese attive, esso diventa un importante campanello d'allarme per il rischio di default nel campione analizzato.

3.4 Random Forest

L'analisi prosegue con il passaggio agli algoritmi di ML attraverso l'applicazione della Random Forest.

Il modello viene configurato con una serie di iperparametri specifici per bilanciare adeguatamente la capacità di previsione e la capacità di generalizzazione, secondo quanto segue:

- `n_estimators=300`: indica il numero di alberi che compongono la “foresta”. Un numero elevato di questi permette al modello di stabilizzare le previsioni e ridurre la varianza.
- `class_weight='balanced'`: rappresenta il parametro più rilevante per la presente analisi. Poiché il dataset è sbilanciato, questo comando istruisce l'algoritmo a dare un peso maggiore agli errori commessi sui casi di default. Concretamente impedisce alla Random Forest di ignorare i default solo perché sono rari, e questo comporta un miglioramento della *recall*.
- `max_depth=10`: indica la profondità massima di ciascun albero. L'obiettivo è quello di limitare la complessità del modello.
- `min_samples_split=5`: indica il numero minimo di osservazioni necessarie per creare un nuovo *split*, o biforcazione, in un nodo.
- `min_samples_leaf=2`: indica il numero minimo di campioni che devono trovarsi in una “foglia”, o nodo finale.¹²⁷

Successivamente si svolge la selezione delle variabili. Per valutare l'importanza relativa di queste nell'algoritmo, viene calcolata la *feature importance*, una metrica che indica quanto ciascun indicatore presente nel dataset contribuisce alla riduzione dell'impurità nei nodi dell'albero. Tale misura permette di comprendere quali variabili influenzano maggiormente le prestazioni nella previsione del rischio di default.

¹²⁷ In Colab, le stringhe di codice utilizzate coincidono con l'elenco degli iperparametri esposto e seguono il codice: `rf_final = RandomForestClassifier`

La selezione delle 10 variabili più rilevanti sulle 29 totali permette di concentrare l'analisi sui fattori chiave del rischio di default. Questo approccio consente di ottenere un buon compromesso tra prestazioni e interpretabilità, elemento centrale per l'ipotetico utilizzo in uno scenario reale.¹²⁸

L'operazione porta ai seguenti risultati:

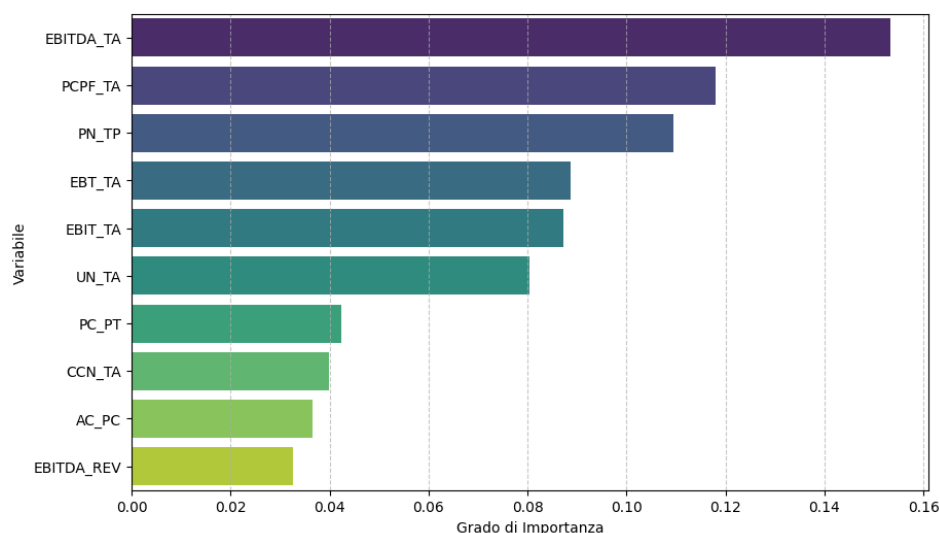


Figura 18. Le 10 variabili più importanti - Random Forest.

È interessante notare come la presenza simultanea di EBITDA, EBIT, EBT e dell'utile netto in relazione al totale dell'attivo denoti come l'algorithm ponga un rilevante peso sui margini di conto economico riclassificato e sul risultato netto.

Seguono il processo di addestramento, la classificazione in impresa sana o in default¹²⁹ e il calcolo della probabilità di default (PD).¹³⁰ Infine, emergono i seguenti risultati:

¹²⁸ L'operazione avviene eseguendo il seguente codice:

```
importances_df = pd.DataFrame({'Ratio': X_all.columns,
                              'Importanza': rf_selector.feature_importances_})
top_10_features = importances_df.sort_values(by='Importanza',
                                             ascending=False).head(10)['Ratio'].tolist()
```

¹²⁹ Come soglia si utilizza l'impostazione nativa dell'algorithm, pari a 0,5.

¹³⁰ Rispettivamente alla prima, seconda e terza riga del seguente codice:

```
rf_final.fit(X_train, y_train)
y_pred = rf_final.predict(X_test)
y_prob = rf_final.predict_proba(X_test)[: , 1].
```

True Positives (TP): 75
False Positives (FP): 14
True Negatives (TN): 1965
False Negatives (FN): 20

Accuracy: 0.9836
Precision: 0.8427
Recall: 0.7895
F1-Score: 0.8152
AUC-ROC: 0.9618

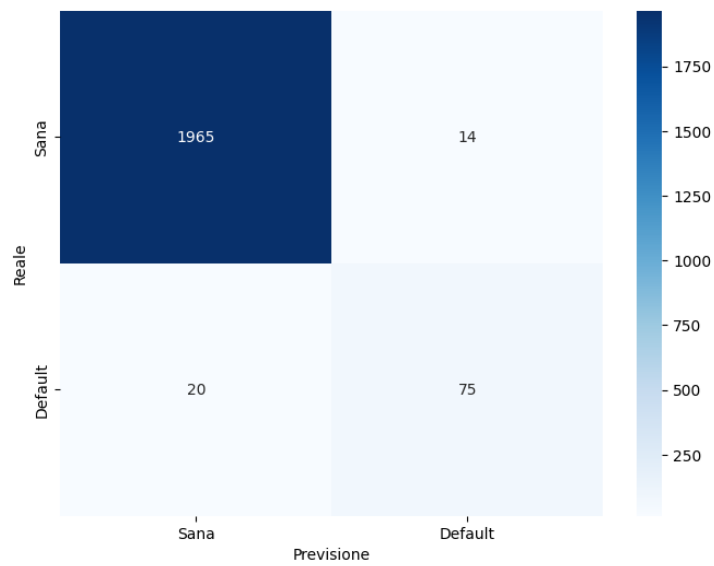


Figura 19. Matrice di confusione- Random Forest.

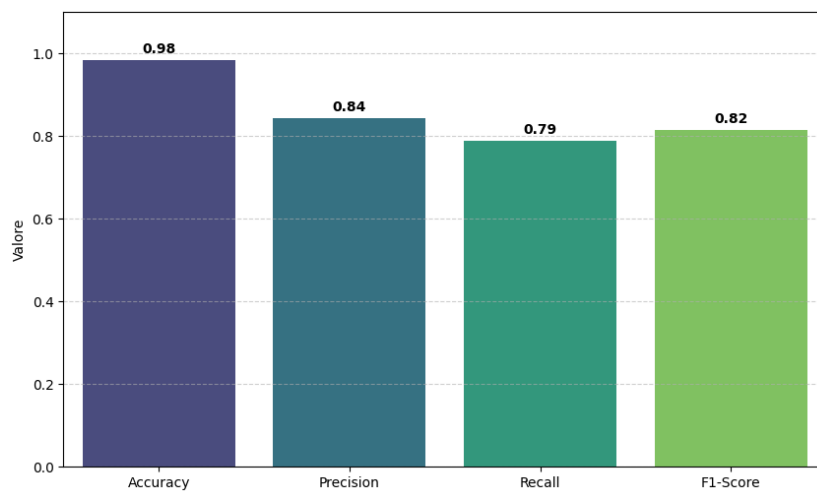


Figura 20. Prestazioni - Random Forest.

Il modello Random Forest individua correttamente 75 imprese su 95 che sono andate effettivamente in default e classifica solamente 14 imprese sane come in default.

La *recall* risulta ottima, il modello riesce ad identificare il 78,95 per cento delle imprese in default mantenendo al contempo una *precision* elevata: l'84,27 per cento delle imprese è stato classificato correttamente come in default. Il *false positive rate*, pari allo 0,71 per cento, risulta eccezionale e segnala la bontà del modello a non creare molteplici “falsi allarmi”. L'F1-Score, pari all'81,52 per cento, si conferma anche esso elevato.

Nonostante il numero di falsi negativi non si possa ritenere trascurabile, il limitato numero di falsi positivi indica un ottimo equilibrio, ad esempio in ambito bancario, tra la protezione dal rischio di credito e la necessità di non penalizzare eccessivamente imprese sane con segnalazioni errate. Per completezza si indica anche la percentuale di accuratezza, pari al 98,36 per cento.

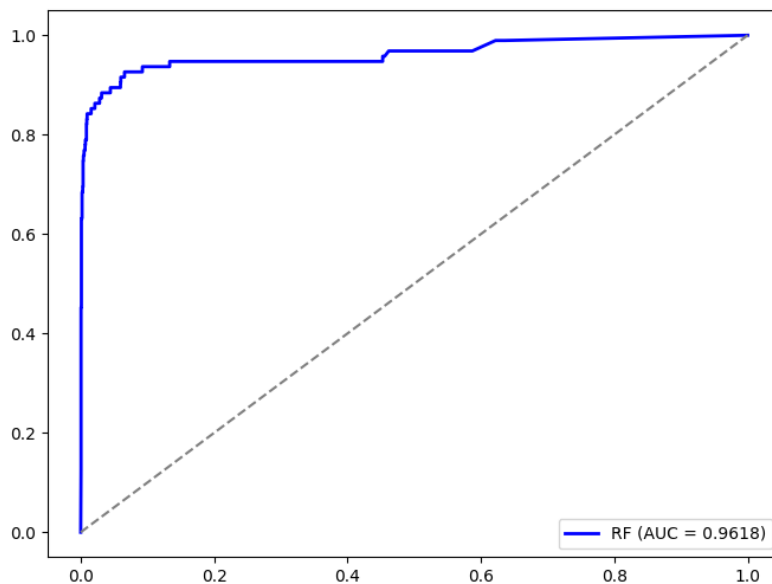


Figura 21. Curva ROC - Random Forest.

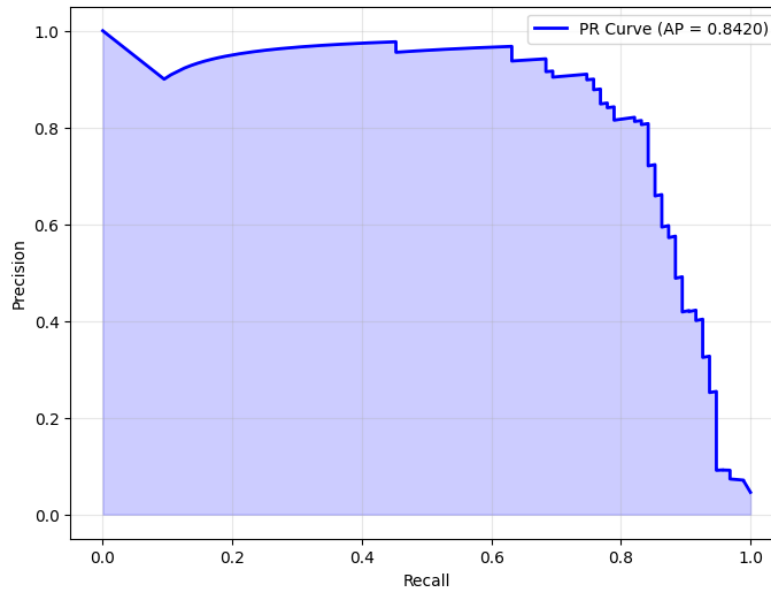


Figura 22. *Curva Precision-Recall - Random Forest.*

Il valore di AUC pari a 0,9618 denota un'eccellente capacità di classificazione. Questo dato indica che, con una probabilità del 96,18 per cento, il modello assegna correttamente un punteggio di rischio più alto a un'impresa destinata al default rispetto a un'impresa sana. Anche in questo caso è utile integrare l'analisi con la curva *precision-recall*, che presenta un'*average precision* di 0,8420, valore che conferma la validità del modello. Si noti come la curva mantiene livelli di *precision* molto elevati anche per valori di *recall* importanti. Per questo il modello riesce a individuare la maggioranza dei casi di default senza aumentare significativamente i falsi positivi raggiungendo i livelli di *precision* e *recall* riscontrati prima di incontrare il decadimento della curva nelle zone di massima *recall*.

Infine, per superare il limite della “*black box*”, tipica degli algoritmi di ML, l'analisi non si è limitata alla sola *feature importance*, ma è stata approfondita tramite la metodologia SHAP. Come si può desumere dalla Figura 23, i *valori di Shapley* offrono una visione più dettagliata del contributo di ciascuna variabile:¹³¹

¹³¹ Si noti come l'ordine delle variabili differisce leggermente tra la *feature importance* e SHAP. Questo deriva dalla differente capacità di ragionamento e valutazione dei due metodi ed è molto utile per poter sfruttare un duplice punto di vista.

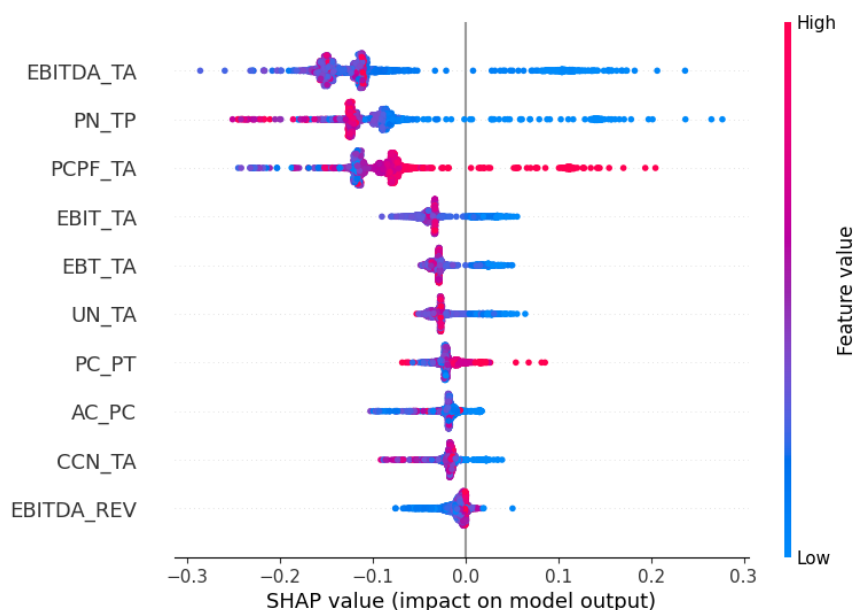


Figura 23. *SHAP Summary Plot - Random Forest.*

Dove:

- Sull'asse x si indica lo *SHAP value*, il quale indica l'intensità dell'impatto sulla previsione. I punti posizionati a destra dello zero indicano che il valore della variabile sta indirizzando il modello a classificare l'impresa come in default, mentre i punti a sinistra dello zero indicano un indirizzamento verso lo *status* di sana.
- La scala cromatica del *feature value* indica il valore assunto dal singolo ratio per ogni singola osservazione del campione. Il colore rosso indica un valore elevato, quello blu indica un valore basso.

L'indice EBITDA_TA si conferma come la variabile più influente. Anche il rapporto tra il patrimonio netto e il totale del passivo mantiene la sua forte rilevanza. Infatti, il grafico mostra che le imprese più solide presentano valori SHAP negativi, il che riduce la probabilità di default. D'altra parte, la dispersione dei punti blu verso destra conferma che il deficit patrimoniale è un fattore di rischio determinante, coerentemente con quanto appurato attraverso il modello di Altman.

Il *current ratio* e il rapporto tra il CCN e il totale dell'attivo, pur essendo meno dominanti rispetto ad altri indici, mostrano una distribuzione coerente con la teoria economica: una

carezza di liquidità (punti blu sulla destra) contribuisce in modo puntuale all'aumento del rischio di default.

3.5 XGBoost

L'ultimo modello ad essere testato è XGBoost, il più recente tra i modelli presentati.

L'algoritmo viene configurato con i seguenti iperparametri:

- `n_estimators=300`: indica il numero di alberi sequenziali. Tale valore consente al modello di apprendere le relazioni complesse riducendo progressivamente l'errore residuo.
- `scale_pos_weight=ratio`: rappresenta l'iperparametro fondamentale per la gestione del dataset sbilanciato. Funziona in modo analogo al `class_weight='balanced'` della Random Forest ma agisce scalando il peso della classe di default (1) rispetto all'altra (0). In tal modo l'algoritmo presta maggiore attenzione alla corretta identificazione delle imprese in default.
- `max_depth=5`: indica la profondità massima di ciascun albero. Rispetto alla Random Forest, XGBoost utilizza alberi meno profondi per evitare l'*overfitting*.
- `learning_rate=0.1`: esso determina l'entità del contributo di ogni nuovo albero alla correzione dell'errore residuo del modello.
- `eval_metric='logloss'`: specifica la funzione di perdita utilizzata per il monitoraggio delle prestazioni durante la fase di addestramento.¹³²

Analogamente a quanto fatto con la Random Forest, si procede con il calcolo della *feature importance* e l'analisi dei relativi risultati.¹³³

¹³² In Colab, le stringhe di codice utilizzate coincidono con l'elenco degli iperparametri esposto e seguono il codice: `xgb_final = XGBClassifier`.

¹³³ Il codice in Python è il medesimo utilizzato per la Random Forest.

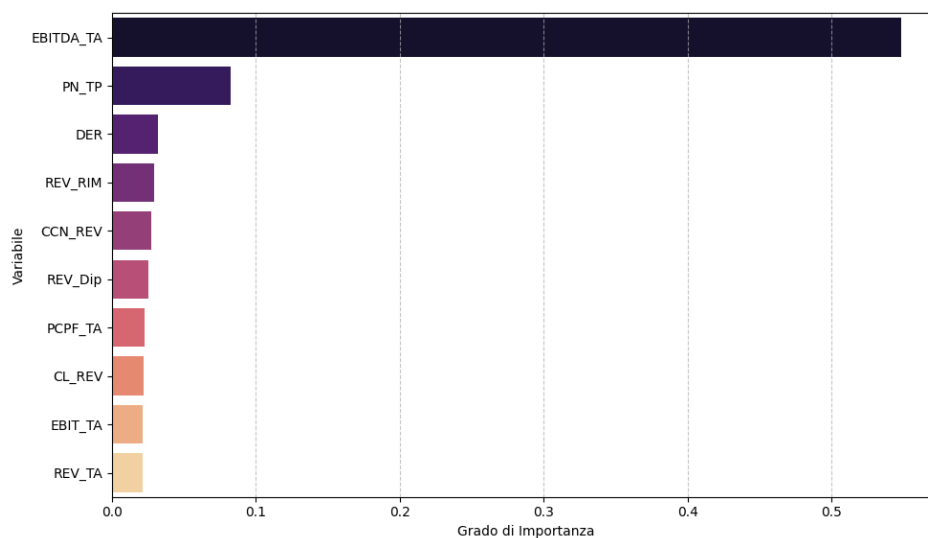


Figura 24. Le 10 variabili più importanti - XGBoost.

Come si evince dal grafico, il rapporto tra EBITDA e le attività totali si conferma come il predittore ampiamente più rilevante. Questo mostra che, per il modello in analisi, l'attitudine dell'impresa a generare flusso di cassa operativo in relazione alla dimensione delle proprie attività è la variabile discriminante principale per la classificazione.

Seguono il processo di addestramento, la classificazione in impresa sana o in default¹³⁴ e il calcolo della probabilità di default (PD). Emergono i seguenti risultati:¹³⁵

```

True Positives (TP): 76
False Positives (FP): 9
True Negatives (TN): 1970
False Negatives (FN): 19
-----
Accuracy:  0.9865
Precision: 0.8941
Recall:    0.8000
F1-Score:  0.8444
AUC-ROC:  0.9596

```

¹³⁴ Come soglia si utilizza l'impostazione nativa dell'algoritmo, pari a 0,5, garantendo uniformità con la Random Forest.

¹³⁵ Il codice in Python è il medesimo utilizzato per la Random Forest.

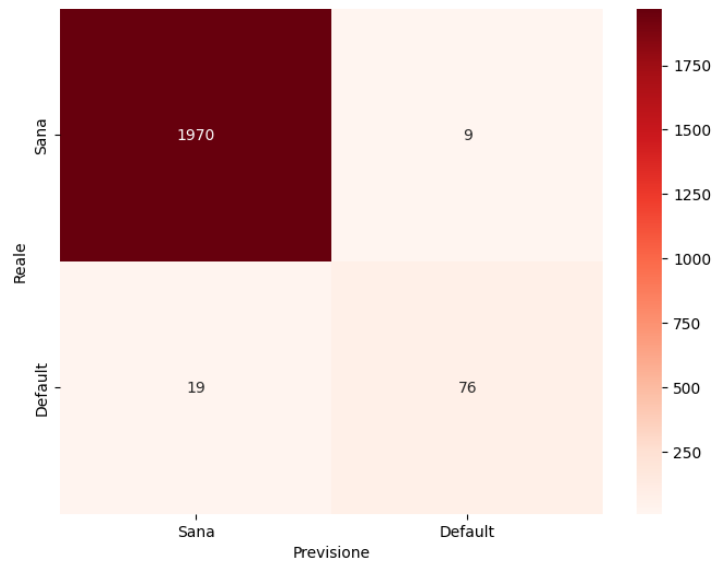


Figura 25. *Matrice di confusione - XGBoost.*

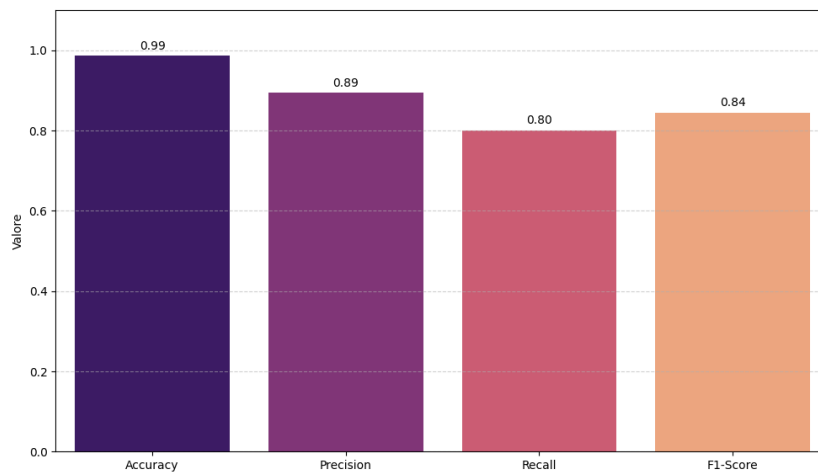


Figura 26. *Prestazioni - XGBoost.*

Il modello XGBoost individua 76 imprese su 95 che sono andate in default e classifica erroneamente solo 9 imprese sane come in default, garantendo un eccezionale *false positive rate* dello 0,41 per cento.

La *recall* e la *precision* risultano rispettivamente pari all'80 e al 89,41 per cento, comportando un F1-Score del 84,44 per cento. Tali valori rivelano delle prestazioni decisamente elevate. Per quanto concerne il numero di falsi negativi però, è opportuno effettuare una considerazione analoga a quella della Random Forest: il valore non si può

non ritenere significativo (20 per cento), ma, considerato il numero trascurabile di falsi positivi, l'algoritmo è in grado di garantire un ottimo equilibrio.

Infine, l'accuratezza è prossima all'unità, con una percentuale pari a 98,65.

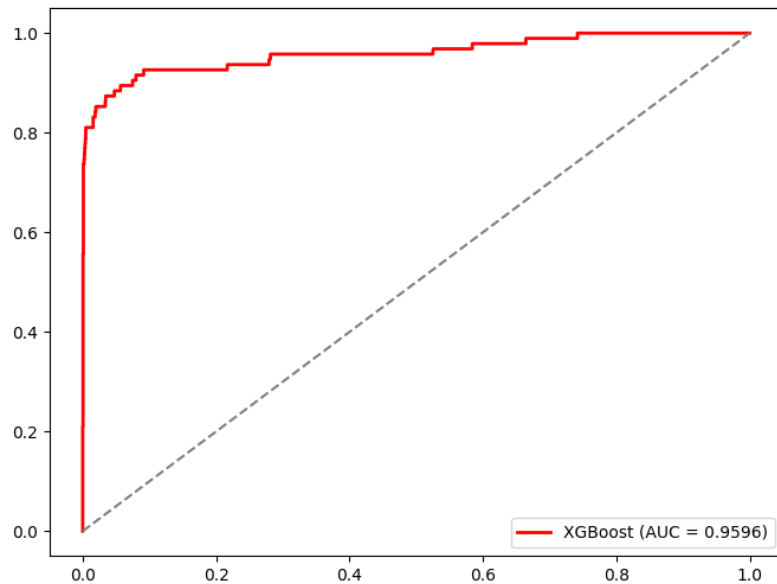


Figura 27. Curva ROC - XGBoost.

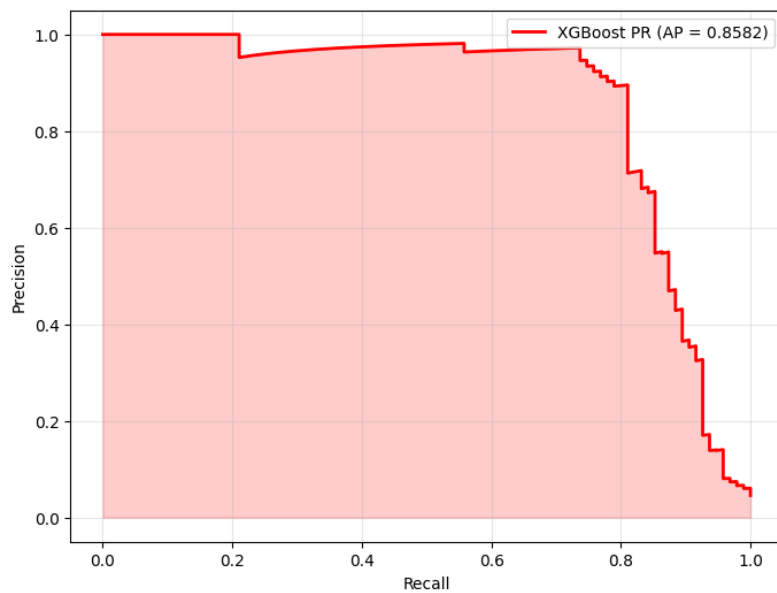


Figura 28. Curva Precision-Recall - XGBoost.

I valori di AUC e dell'*average precision* correlata alla curva *precision-recall*, rispettivamente a 0,9596 e 0,8582, risultano molto elevati, coerentemente con quanto analizzato in letteratura. In particolare, si noti come la *precision* rimanga prossima all'unità fino a livelli di *recall* pari all'80 per cento.

L'analisi del modello si conclude attraverso lo studio dei *valori di Shapley*:

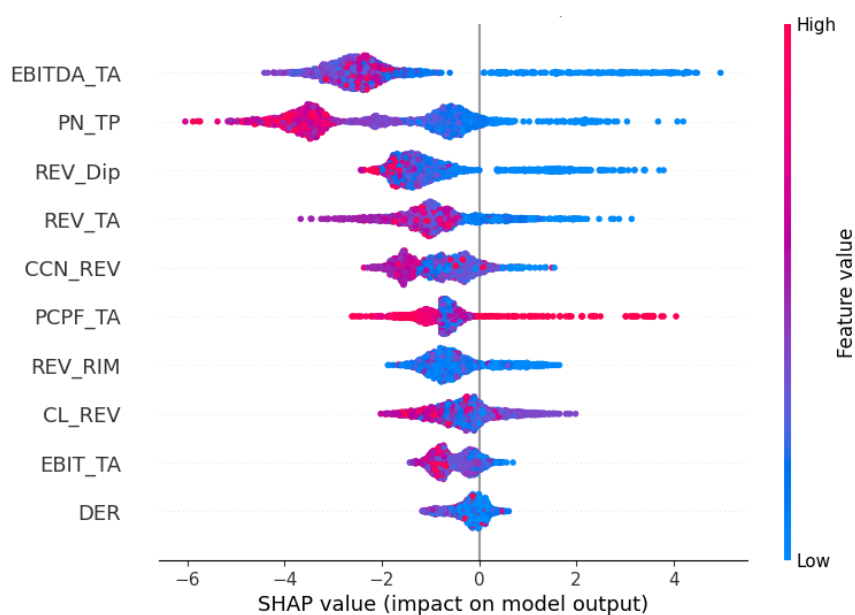


Figura 29. *SHAP Summary Plot - XGBoost.*

Oltre all'indice EBITDA_TA che si conferma nuovamente come indice più rilevante, è interessante analizzare la dispersione verso sinistra del rapporto PN_TP. Da ciò si può desumere che una buona capitalizzazione esercita un forte effetto di mitigazione del rischio di default, orientando la previsione verso la classificazione di impresa sana.

L'algoritmo attribuisce una certa importanza agli indici di rotazione e produttività che nella Random Forest risultavano meno centrali. Nello specifico, una bassa rotazione delle rimanenze (REV_RIM) o una bassa produttività per dipendente (REV_Dip), segnalati dalla pluralità di punti blu presenti a destra dello zero, vengono interpretati come segnali di inefficienza che precedono il default.

3.6 Valutazione comparata delle prestazioni dei modelli

Il confronto finale tra il modello tradizionale Z'-Score e gli algoritmi di ML Random Forest e XGBoost consente di evidenziare le principali differenze in termini di performance predittive e di equilibrio tra individuazione del rischio e contenimento dei falsi segnali.

In termini di capacità discriminante, tutti e tre i modelli mostrano valori di AUC decisamente elevati, come si può desumere dalla figura 30:

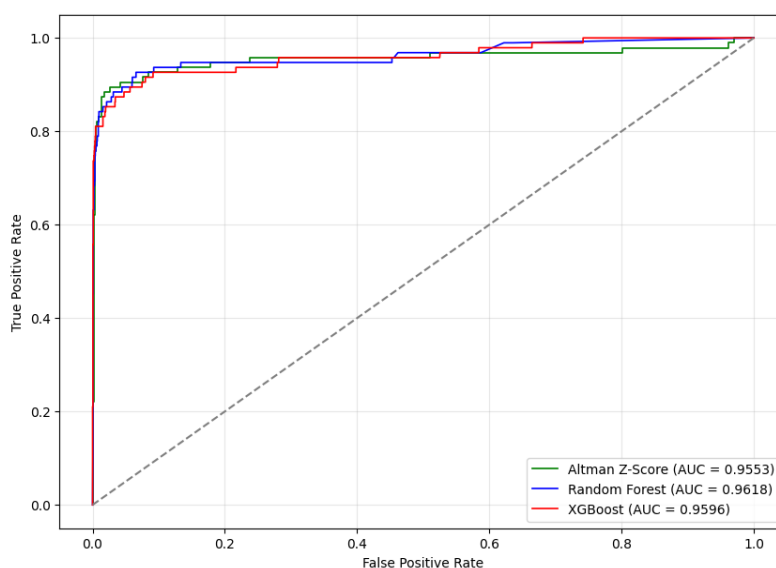


Figura 30. Confronto delle curve ROC dei tre modelli.

Le differenze risultano contenute, ma è comunque possibile notare come in molteplici punti la curva dei modelli di ML si mantenga sopra quella del modello Z'-Score. In ogni caso, i valori indicano che anche il modello tradizionale mantiene una forte capacità di separazione tra imprese sane e in default.

Emergono invece delle differenze più marcate dall'analisi della curva *precision-recall* (Figura 31):

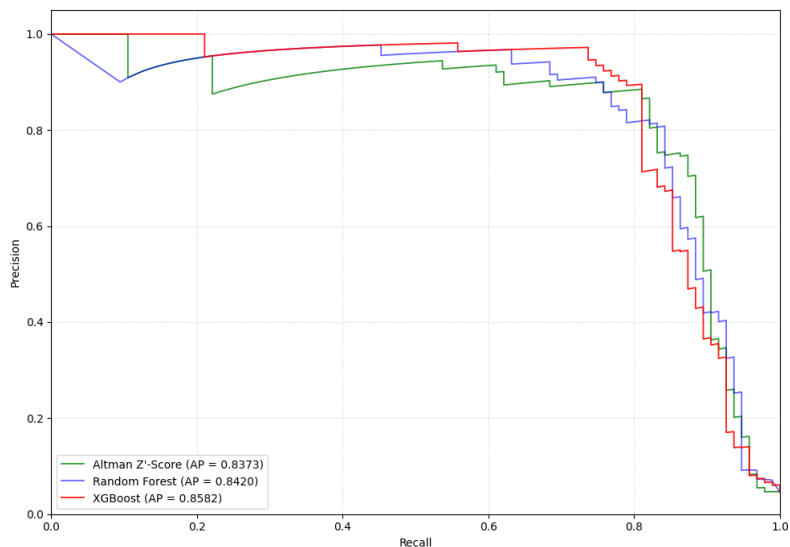


Figura 31. Confronto delle curve Precision-Recall dei tre modelli.

Lo Z'-Score si caratterizza per una *recall* eccezionale, intercettando la quasi totalità dei casi di default, ma al prezzo di una *precision* estremamente bassa. I modelli di ML, soprattutto XGBoost, riescono a mantenere una maggiore stabilità lungo la curva, confermata dalla più elevata *average precision*. Si noti come, prima del decadimento della curva, i livelli di *precision* degli algoritmi si mantengano decisamente più elevati rispetto al modello tradizionale.

Con la tabella 6 è possibile confrontare la totalità dei valori riscontrati nella presente analisi:

Tabella 6. Confronto degli esiti dell'analisi. Fonte: elaborazione propria in ambiente Colab.

	Metrica	Altman Z-Score	Random Forest	XGBoost
0	Accuracy	0.7888	0.9836	0.9865
1	Precision	0.1721	0.8427	0.8941
2	Recall	0.9474	0.7895	0.8000
3	F1-Score	0.2913	0.8152	0.8444
4	AUC-ROC	0.9553	0.9618	0.9596
5	Average Precision	0.8373	0.8420	0.8582
6	False Positive Rate	0.2188	0.0071	0.0045

Lo Z'-Score, come già appurato, può essere interpretato come un modello fortemente prudenziale, privilegiando la massima intercettazione del rischio di default ma a costo di un numero estremamente elevato di falsi positivi. I modelli di ML, invece, risultano più selettivi, riducendo al minimo di falsi positivi senza compromettere eccessivamente la capacità di identificare le imprese a rischio. La capacità di garantire un bilanciamento migliore denota la superiorità degli algoritmi rispetto al modello di Altman.

Come ampiamente riscontrato in letteratura, i modelli statistici tradizionali applicati alla previsione del rischio di default soffrono sistematicamente del problema rilevato dalla presente analisi con l'eccessiva segnalazione di imprese potenzialmente in default ma che in realtà risultano essere sane. Dal punto di vista operativo tale caratteristica comporta delle importanti implicazioni. In ambito creditizio, una forte segnalazione di falsi positivi sicuramente proteggerebbe l'istituto di credito da errori di Tipo II, ma comporterebbe un elevato mancato guadagno per i prestiti non concessi oltre all'erosione del rapporto con il cliente. Sotto il profilo della gestione d'impresa, in primo luogo, una segnalazione di questo tipo comporterebbe un assorbimento ingiustificato di risorse per la verifica del segnale di rischio. In secondo luogo, sussiste un rischio reputazionale e strategico: una diagnosi errata potrebbe orientare l'imprenditore verso procedure di risanamento non necessarie¹³⁶ innescando paradossalmente quella perdita di valore che si vorrebbe evitare a causa della diffidenza degli *stakeholder* verso situazioni di questo tipo.

Infine, un sistema caratterizzato da un'elevata frequenza di falsi positivi rischia di ridurre sensibilmente l'efficacia dello strumento stesso. Infatti, se i segnali erronei risultassero troppo frequenti gli operatori potrebbero sviluppare una tendenza a sottovalutarli, compromettendo la funzione predittiva del modello.

I risultati ottenuti supportano l'idea, sostenuta da una parte importante della letteratura, che le nuove tecniche di Intelligenza Artificiale apportino miglioramenti significativi alla previsione del rischio di default. Pertanto, i nuovi algoritmi di ML rappresentano una potenziale miglior alternativa ai modelli predittivi classici, in grado di fornire delle indicazioni probabilistiche più accurate. Tale aspetto si traduce in una più alta qualità del supporto alle decisioni assunte dagli analisti.

¹³⁶ Con riferimento alla Composizione Negoziata della Crisi (CNC), prevista dal nuovo Codice della Crisi d'impresa e dell'Insolvenza.

Conclusione

Il presente elaborato ha analizzato l'evoluzione nel tempo dei modelli tradizionali di previsione del rischio di default fino alle tecniche più recenti basate sull'applicazione dell'IA. Il confronto empirico tra lo Z'-Score di Altman, uno tra i modelli classici più famosi e apprezzati, e i modelli Random Forest e XGBoost nell'ambito del ML, ha evidenziato che tutti i modelli possiedono elevate capacità discriminanti per il dataset utilizzato. Tuttavia, il modello di Altman presenta un elevato numero di falsi positivi che ne compromettono l'efficacia nel caso di una reale applicazione.

Questo deriva dall'ineliminabile limite che caratterizza i modelli parametrici lineari. Infatti, l'incapacità di individuare le relazioni complesse tra le variabili, soprattutto in presenza di grandi quantità di dati con classi sbilanciate, ne abbassa le prestazioni. Al contrario, i modelli basati sul ML riescono a garantire un miglior bilanciamento dei segnali erronei portando a performance elevate.

Sotto il profilo dell'interpretabilità, limite intrinseco degli algoritmi, l'analisi della *feature importance* e dei *valori di Shapley* ha permesso di decodificare i processi decisionali della Random Forest e di XGBoost, individuando quali variabili abbiano contribuito alla previsione e l'impatto sulla stessa. Con questo approccio si è garantita la spiegabilità, trasformando un risultato meramente probabilistico in un'indicazione diagnostica comprensibile per l'operatore.

Le prospettive future indicano una crescita esponenziale nello sviluppo dei sistemi di ML, supportata dagli ingenti investimenti, dell'ordine dei miliardi di dollari, soprattutto da parte delle *Big Tech* statunitensi, nei vari ambiti di applicazione dell'IA. In un contesto globale caratterizzato da crescente incertezza e volatilità, l'integrazione di tali tecnologie nelle analisi del rischio di default rappresenta una necessaria evoluzione per garantire la resilienza del sistema finanziario. La capacità di elaborare dati non strutturati (come le informazioni derivanti dalla relazione sulla gestione, oppure quelle ricavate mediante la *sentiment analysis*, ampiamente utilizzata dalle banche d'investimento) permetterà agli algoritmi di superare la rigidità tipica degli schemi di bilancio, potendo permettere una valutazione del rischio in tempo reale.

Questa prospettiva di crescita dovrà essere accompagnata da una parallela evoluzione del quadro normativo, al fine di evitare potenziali discriminazioni da parte degli stessi modelli. È infatti essenziale che lo sviluppo dei sistemi non comprometta il principio di *accountability*, assicurando che ogni valutazione risulti comprensibile, verificabile e rispettosa degli elementi etici che stanno acquisendo sempre più rilevanza nella finanza moderna.

Inoltre, tale progresso tecnologico potrebbe contribuire a ridurre uno dei principali limiti dei modelli predittivi, rappresentato dalla presenza di errori e anomalie contabili, più o meno rilevanti, che potrebbero portare al c.d. fenomeno “*garbage in-garbage out*”.

I nuovi sistemi potrebbero infatti garantire un supporto più efficace per gli operatori, riducendo drasticamente gli errori attraverso un’ampia ridondanza di sistemi di controllo, nelle fasi di rilevazione e registrazione dei dati contabili o di revisione legale dei conti. Potrebbero inoltre integrare l’attività di controllo svolta dagli enti della Pubblica Amministrazione, come l’Agenzia delle Entrate, o supportare l’azione dell’autorità giudiziaria nel caso di frodi o manipolazione dei dati contabili.

Resta ferma, in ogni caso, la centralità del giudizio professionale. Qualsiasi sia l’utilizzo dei sistemi di ML, questi non devono essere intesi come un sostituto del decisore umano, ma come un sostegno che possa far crescere esponenzialmente l’efficienza. Con il c.d. approccio *human in the loop*, la sinergia tra le performance degli algoritmi e la bontà dell’esperienza del professionista diventa il principale garante per interpretare correttamente anche tutti quei casi complessi e caratterizzati da straordinarietà, come l’influenza derivante da uno shock esogeno, le cui caratteristiche non risultano presenti nei dati storici.

In conclusione, la prospettiva che si delinea è quella di un vero e proprio cambiamento di paradigma nell’analisi predittiva per il rischio di default. È opportuno precisare che questa trasformazione non implica necessariamente l’abbandono dei metodi tradizionali, i quali mantengono saldamente il loro ruolo di *benchmark*, grazie alla loro immediatezza e maggiore trasparenza. Pertanto, la prospettiva più concreta appare quella di un approccio ibrido in cui i modelli classici forniscono la base interpretativa e i sistemi di ML agiscono come strumenti altamente precisi.

Bibliografia

- Alam, I. (2022). *James Ohlson O-Score for Predicting Corporate Bankruptcy*. Jahangirnagar University. doi:10.13140/RG.2.2.13601.28005
- Alanis, E., Chava, S., & Shah, A. (2022). *Benchmarking Machine Learning Models to Predict Corporate Bankruptcy*. Research Paper n. 4249412. Georgia Tech Scheller College of Business.
- Alonso, A., & Carbó, J. M. (2020). *Machine Learning in Credit Risk: Measuring the Dilemma between Prediction and Supervisory Cost*. Working Paper No. 2032. Banco de España.
- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), p. 589-609.
- Altman, E. I. (1970). Ratio Analysis and the Prediction of Firm Failure: A Reply. *The Journal of Finance*, 25(5), p. 1169-1172.
- Altman, E. I. (1983). *Corporate Financial Distress*. New York: Wiley Interscience.
- Altman, E. I., & Hotchkiss, E. (2006). *Corporate Financial Distress and Bankruptcy*. New York: Wiley.
- Altman, E. I., & La Fleur, J. K. (1985). I modelli di previsione delle insolvenze: le loro applicazioni alla gestione d'impresa. *Finanza, Marketing, Produzione*(4), p. 77-93.
- Altman, E. I., & Rijken, H. A. (2010). *The Z-Metrics™ Methodology for estimating company credit ratings and default risk probabilities*. New York: RiskMetrics Group.
- Altman, E. I., Danovi, A., & Falini, A. (2013). La previsione dell'insolvenza: l'applicazione dello Z Score alle imprese in amministrazione straordinaria. *Bancaria*, 4, p. 24-37.
- Altman, E. I., Hartzell, J., & Peck, M. (1995). *Emerging Markets Corporate Bonds: A Scoring System*. Salomon Brothers Inc., New York.
- Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance*, 18(3), p. 505-529.
- Amemiya, T. (1985). *Advanced econometrics*. Cambridge (USA): Harvard University Press.
- Appetiti, S. (1984). *L'utilizzo dell'analisi discriminatoria per la previsione delle insolvenze: ipotesi e test per un'analisi dinamica*. Temi di discussione n. 27. Banca d'Italia.
- Ascari, G., Bonomolo, P., Hoerberichts, M., & Trezzi, R. (2023). *The euro area great inflation surge*. Amsterdam: De Nederlandsche Bank (DNB).

- Baesens, B., Viaene, S., Van Gestel, T., & Stepanova, M. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), p. 27-635.
- Bagli, E., Visani, G., Chesani, F., & Poluzzi, A. (2020). Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models. *Journal of the Operational Research Society*, p. 1-18.
- Barboza, F., Kimura, H., & Altman, E. I. (2017). Machine learning models and bankruptcy prediction. *Machine learning models and bankruptcy prediction*, 83, p. 405-417.
- Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 40, p. 432-442.
- Bazarbash, M. (2019). *Fintech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk*. IMF Working Paper No. 19/110. International Monetary Fund.
- Bazzana, F., Bee, M., & Hussin, A. (2024). Machine learning techniques for default prediction: an application to small Italian companies. *Risk Management*, 26(1), p. 1-25.
- Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, 4, p. 74-111.
- Ben Jabeur, S., Carmona, P., & Stef, N. (2022). Bankruptcy Prediction using the XGBoost Algorithm and Variable Importance Feature Engineering. *Computational Economics*, 61(2), p. 715-741.
- Berg, T., Burg, V., Gombovic, A., & Puri, M. (2020). On the Rise of Fintechs: Credit Scoring Using Digital Footprints. *The Review of Financial Studies*, 33(7), p. 2845-2897.
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, p. 281-305.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.
- Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, 81(3), p. 637-654.
- Bonaccorsi, E., & Calabresi, F. (2022). *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano. Temi di discussione n. 721*. Banca d'Italia.
- Bradley, A. P. (1997). The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition*, 30, p. 1145-1159.
- Breiman, L. (2001). *Random Forest*. University of California, Berkeley.
- Carton, R., & Hofer, C. (2006). *Measuring organizational performance*. Edward Elgar Publishing.

- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (p. 785-794).
- CNDCEC. (2021). *Osservatorio sui bilanci delle srl 2019 trend 2017-2019*. Roma: Fondazione Nazionale dei Commercialisti.
- Commissione Europea. (2014). *Un piano di investimenti per l'Europa*. COM 913 final, Bruxelles.
- Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili (CNDCEC). (2019). *Crisi d'impresa. Gli indici di allerta previsti dal Codice della crisi e dell'insolvenza*.
- Cossin, D., & Pirotte, H. (2001). *Advanced Credit Risk Analysis: Financial Approaches and Mathematical Models to Assess, Price, and Manage Credit Risk*. Wiley.
- Crosbie, P., & Bohn, J. (2002). *Modeling Default Risk*. San Francisco (USA): KVM LLC.
- Dell'Atti, A. (2023). *La gestione delle crisi aziendali: cause, strategie e piani di risanamento*. Giappichelli.
- Finlay, S. (2011). Multiple classifier architectures and their application to credit risk assessment. *European Journal of Operational Research*, 210(2), p. 368-378.
- Fitzpatrick, P. J. (1392). A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Companies. *The Certified Public Accountant*, p. 598-605, 656-662, 727-731.
- Freeman, J. A., & Skapura, D. M. (1991). *Neural Networks: Algorithms, Applications, and Programming Techniques*. Reading (USA): Addison-Wesley Professional.
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), p. 1189-1232.
- Friedman, J. H. (2002). Stochastic Gradient Boosting. *Computational Statistics & Data Analysis*, 38(4), p. 367-378.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge (USA): MIT Press.
- Hair, J., & et al. (1998). *Multivariate Data Analysis*. Englewood Cliffs (USA): Prentice Hall.
- Harnad, S. (2008). The Annotation Game: On Turing (1950) on Computing Machinery and Intelligence. *Parsing the Turing Test*, p. 23-66.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

- HKUST Math Department. (2002). *KMV Model*. (Y. K. Kwok, A cura di) Tratto il giorno 2025 Dicembre da Hong Kong University of Science and Technology: https://www.math.hkust.edu.hk/~maykwok/Web_ppt/KMV/KMV.pdf
- Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, 117, p. 287-299.
- Huang, B. (2024). The Investigation of Stock Price Prediction Based on Machine Learning. *Highlights in Science, Engineering and Technology*, 85, p. 991-996.
- Hull, J., Nelken, I., & White, A. (2004). Merton's model, credit risk and volatility skews. *Journal of Credit Risk*, 1(1), p. 3-28.
- ISO/IEC. (2023). *42001 Information technology, Artificial intelligence, Management system*. Ginevra.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R* (2nd ed.). New York: Springer.
- Kohavi, R., & Wolpert, D. H. (1996). Bias Plus Variance Decomposition for Zero-One Loss Functions. *Proceedings of the Thirteenth International Conference on Machine Learning*, (p. 275-283).
- Malley, J. D., Kruppa, J., Dasgupta, A., & Abhijit, M. (2012). Probability machines: consistent probability estimation using nonparametric learning machines. *Methods of Information in Medicine*, 51(1), p. 74-81.
- Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5), p. 555-559.
- Matta, A. (2004). *Approcci matematico-statistici alla valutazione del rischio di credito*. Torino: Associazione Torino Finanza.
- McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Dartmouth College.
- Merton, R. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, 29(2), p. 449-470.
- Mitchell, K., & Walker, M. (2008). Bankers on Boards, Financial Constraints, and Financial Distress. *SSRN Electronic Journal*.
- Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.
- Moscatelli, M., & et al. (2019). *Corporate default forecasting with machine learning*. *Temi di discussione n. 1256*. Banca d'Italia.
- Moscatelli, M., Parlapiano, F., & Cascarino, G. (2022). *Explainable Artificial Intelligence: interpreting default forecasting models based on Machine Learning*. *Temi di discussione n. 674*. Banca.

- Navaroni, M., & Panelli, M. (2024, dicembre 23). L'analisi di bilancio applicata all'impresa in crisi. *Diritto della Crisi*.
- Neri, L. (2017). *Corso di analisi statistica per le imprese - Modello logit*. Università degli studi di Siena.
- Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), p. 109-131.
- Peng, C., Lee, K., & Ingersoll, G. (2002). An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, 96(1), p. 3-15.
- Peng, C., Manz, B., & Keck, J. (2001). Modeling Categorical Variables by Logistic Regression. *Journal of Health Behaviour*, 25(3), p. 278-284.
- Piatti, D., Cincinelli, P., & Castellani, D. (2015). Ruolo dell'Efficienza nella Previsione del Default Aziendale. *Piccola impresa Small Business*, 2, p. 53-84.
- Quagli, A. (2016). Il concetto di crisi d'impresa come incontro tra la prospettiva aziendale e quella giuridica. *Crisi d'impresa e fallimento*.
- Russel, S. J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3 ed.). Upper Saddle River: Prentice Hall.
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74(1), p. 432-442.
- Skinner, B. F. (1971). *Beyond Freedom and Dignity*. Knopf.
- Smith, F. R. (1930). *A Test Analysis of Unsuccessful Industry Companies*. University of Illinois, Bureau of Business Research, n. 31.
- Taddy, M. (2019). *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions*. McGraw-Hill.
- Teodori, C. (1989). *Modelli di previsione nell'analisi economico-aziendale*. Torino: Giappichelli.
- Tobback, E., & Martens, D. (2019). Retail credit scoring using fine-grained payment data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(4), p. 1227-1246.
- Tochaiwat, K., & Suwanchaisakul, A. (2025). Land price prediction model for housing development in metropolitan area using deep learning technique: A case study of Thailand. *Al-Qadisiyah Journal for Engineering Sciences*, 18, p. 360-367.
- Tsai, B.-H. (2012). Comparison of Binary Logit Model and Multinomial Logit Model in Predicting Corporate Failure. *Better Advanced Press*, 2, p. 99-111.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), p. 433-460.
- Wolpert, D. H., & Macready, W. (1997). No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), p. 67-82.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge (USA): MIT Press.

Xia, Y., Liu, C., Liu, N., & Li, Y. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, 78, p. 225-241.

Zhang, G., Hu, M., Patuwo, B., & Indro, D. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116(1), p. 16-32.