



Early Warning System: prospettive e vincoli nell'adozione dei nuovi dati interni e pubblici (Open Data)

*di Fabrizio Cavallo, Emanuele Grimaldi, Dmytro Inozemtsev, Stefano Marielli, Luca Mignone,
Ciro Prisco, Marco Vatiere*

IPE Working Paper

N. 26

July 11, 2025

ISSN 2284-1229

Early Warning System: prospettive e vincoli nell'adozione dei nuovi dati interni e pubblici (Open Data)

di Fabrizio Cavallo, Emanuele Grimaldi, Dmytro Inozemtsev, Stefano Marielli, Luca Mignone, Ciro Prisco, Marco Vatiери¹

Abstract

Il presente lavoro propone lo sviluppo di un modello di *Early Warning System (EWS)* finalizzato all'identificazione precoce del *default* creditizio. Nella prima parte viene analizzato il quadro normativo di riferimento, con *focus* sulle linee guida europee e sull'evoluzione dei tassi di *default*. Si prosegue con l'esame degli strumenti utili alla progettazione di un sistema di allerta efficace, approfondendo modelli statistici e indicatori chiave. Infine, la fase applicativa prevede l'impiego di tipici dati gestionali di una banca, integrati con *open data* al fine di ampliare il set informativo utilizzato come input del modello. Quest'ultimo viene calibrato automaticamente mediante la stima dei pesi delle variabili esplicative e testato tramite esercizi di *backtesting* su posizioni già classificate in *default*. L'obiettivo finale è verificare in che modo l'integrazione di dati esterni impatti sulla capacità predittiva del modello rispetto all'uso della sola base informativa interna.

This paper proposes the development of an Early Warning System (EWS) model aimed at the early identification of credit default. In the first part, the regulatory framework is analyzed, with a focus on European guidelines and the evolution of default rates. It continues with an examination of the tools useful in designing an effective alert system, delving into statistical models and key indicators. Finally, the application phase involves the use of typical management data from a bank, integrated with open data in order to expand the information set used as input to the model. The latter is automatically calibrated by estimating the weights of the explanatory variables and tested through backtesting exercises on positions already classified as default. The ultimate goal is to test how the integration of external data impacts the predictive ability of the model compared to using only the internal information base.

¹ Allievi della XXIII edizione del Master in Finanza Avanzata dell'IPE Business School. Il paper è stato redatto nell'ambito del Project Work realizzato presso Reply Advantage.

Indice

Introduzione	4
1. Normativa e innovazione nei modelli di gestione del rischio di credito	5
1.1. Il framework normativo e contabile sul rischio di default e NPE	
1.2. Early Warning System e gestione proattiva del rischio	
1.3. Panoramica sui trend del credito e dei default	
2. Integrazione di dati e modelli per l'analisi del rischio	19
2.1 Struttura e utilizzo degli indicatori di rischio nei modelli predittivi	
2.1.1 Indicatori patrimoniali	
2.1.2 Indicatori reddituali	
2.1.3 Indicatori finanziari	
2.1.4.Indicatori di rischio	
2.1.5 Indicatori qualitativi	
2.1.6 Indicatori macroeconomici	
2.2 Approcci statistici e modelli applicabili nei sistemi di allerta	
2.2.1 Regressione Logistica	
2.2.2 Random Forest	
2.2.3 Altri modelli avanzati e approcci ibridi	
2.3 Leve di miglioramento e open data	
2.4 L'intelligenza artificiale nel contesto dei Sistemi di Allerta Precoce (EWS)	
3. Costruzione di un modello predittivo dei default	42
3.1 Analisi e arricchimento del patrimonio informativo mediante Open Data	
3.2 Preparazione delle feature e applicazione del modello	
3.3 Risultati ed interpretazioni	
Conclusioni	50
Bibliografia	51
Sitografia	52

Introduzione

Le criticità emerse nel sistema bancario europeo hanno evidenziato fragilità nella gestione del rischio di credito, traducendosi in un incremento dei crediti deteriorati (NPL - Non Performing Loans) che ha compromesso la redditività e la capacità di credito delle banche. In risposta, BCE ed EBA hanno rafforzato la normativa per migliorare la gestione degli NPL, riducendo significativamente il loro peso nei bilanci. Oggi, l'attenzione si è spostata dalla gestione ex post alla tempestiva individuazione di situazioni di rischio.

Questo lavoro si propone di sviluppare un modello di Early Warning per l'identificazione precoce del default creditizio. Un EWS è un sistema integrato di metodi e strumenti progettato per individuare rapidamente situazioni di rischio ed attivare misure di intervento efficaci per mitigarlo. A questo approccio operativo si affiancano spesso modelli statistici progettati per selezionare indicatori predittivi utili nella segnalazione anticipata delle crisi.

La prima parte del lavoro è dedicata all'inquadramento normativo, con particolare attenzione alle linee guida emanate dalle autorità europee in ambito credit risk e all'evoluzione storica dei tassi di default.

Successivamente, viene approfondito il processo di costruzione di un sistema di allerta efficace, attraverso l'analisi di modelli statistici e l'individuazione di indicatori rilevanti in grado di cogliere segnali anticipatori di deterioramento creditizio.

Infine, nella parte empirica è previsto l'impiego di un set di dati composto da informazioni gestionali interne, comunemente disponibili nei sistemi informativi bancari, integrate con *open data* provenienti da fonti esterne pubblicamente accessibili. L'obiettivo è arricchire la base informativa del modello, introducendo variabili potenzialmente rilevanti alla previsione dell'evento di default. Il modello sviluppato utilizza tecniche di stima automatica per attribuire un peso ottimale a ciascuna variabile esplicativa, in base alla sua capacità predittiva. L'efficacia del sistema è quindi valutata attraverso esercizi di *backtesting*, applicati retrospettivamente a posizioni creditizie che hanno registrato eventi di default, al fine di testare la precisione e l'affidabilità delle previsioni generate.

In sintesi, l'elemento centrale dello consiste nella valutazione dell'efficacia predittiva del modello in funzione della base informativa utilizzata, verificando se e quanto l'integrazione di dati esterni, provenienti da fonti pubbliche e open, possa potenziare la capacità di anticipare situazioni di rischio rispetto all'impiego esclusivo di informazioni interne. Questo confronto è

essenziale per comprendere il valore aggiunto degli *open data* nel migliorare gli strumenti di prevenzione e gestione del rischio creditizio.

1. Normativa e innovazione nei modelli di gestione del rischio di credito

La crisi finanziaria globale del 2008 ha avuto conseguenze profonde e durature sul sistema bancario europeo, mettendo in luce le vulnerabilità strutturali nella gestione del rischio di credito. Uno degli effetti più evidenti è stato l'accumulo di crediti deteriorati (Non-Performing Loans, NPL) nei bilanci bancari, che ha compromesso la redditività degli istituti, limitato la loro capacità di erogare nuovo credito e contribuito a un diffuso clima di sfiducia nel sistema finanziario.

Per rispondere a queste criticità, le autorità regolamentari, in primis la Banca Centrale Europea (BCE) e l'Autorità Bancaria Europea (EBA), hanno progressivamente introdotto un ampio corpus normativo volto a migliorare la gestione dei crediti deteriorati e a ridurre l'incidenza. Tra le principali iniziative si segnalano gli interventi volti a rafforzare la gestione dei crediti deteriorati da parte degli intermediari finanziari, attraverso indicazioni operative più dettagliate, criteri di classificazione più rigorosi e un progressivo irrigidimento dei requisiti prudenziali sugli accantonamenti. Queste misure hanno rafforzato la governance del rischio di credito, incentivato la pulizia dei bilanci e favorito un mercato secondario per la dismissione degli NPL.

Grazie a questo sforzo coordinato, il sistema bancario europeo ha registrato una significativa riduzione dell'NPL ratio, che oggi si attesta su livelli storicamente bassi. Tuttavia, l'esperienza passata ha insegnato che agire *ex post* non è sufficiente. L'attenzione si è quindi progressivamente spostata dalla gestione dei crediti già deteriorati alla prevenzione del deterioramento stesso, attraverso lo sviluppo e l'adozione di sistemi di allerta precoce (Early Warning Systems, EWS).

Questi strumenti permettono di individuare tempestivamente segnali di rischio e comportamenti anomali, consentendo interventi mirati prima che l'esposizione debba essere classificata come default. Prevenire è infatti meno oneroso che curare: un credito in default comporta costi elevati in termini di accantonamenti, minore redditività, peggioramento degli

indici di solidità patrimoniale e perdita di fiducia da parte del mercato. Inoltre, gestire crediti deteriorati richiede processi lunghi, dispendiosi e spesso con esiti incerti.

Nel presente capitolo si analizzerà quindi l'evoluzione del quadro regolamentare relativo al rischio di credito, con particolare attenzione al ruolo crescente della prevenzione attraverso strumenti predittivi, al fine di comprendere come il sistema bancario si stia attrezzando per evitare che si ripetano le crisi del passato.

1.1. Il framework normativo e contabile sul rischio di default e NPE

Come anticipato nella premessa, negli anni successivi alla crisi finanziaria del 2008 si è assistito a una profonda evoluzione del quadro regolamentare europeo in materia di rischio di credito, con l'obiettivo di ridurre l'incidenza delle esposizioni deteriorate (Non-Performing Exposures, NPE) nei bilanci bancari. La strategia normativa si è sviluppata su due direttrici principali: da un lato, la riduzione dello stock esistente di crediti deteriorati attraverso misure prudenziali e operative; dall'altro, la prevenzione del deterioramento del credito, mediante l'introduzione di strumenti predittivi e sistemi di allerta precoce (Early Warning Systems).

Una delle prime novità in ambito normativo è rappresentata dall'introduzione dello IFRS 9, lo standard contabile internazionale emanato dall'International Accounting Standards Board (IASB) entrato in vigore dal 1° gennaio 2018. Esso ha sostituito il precedente IAS 39, abbandonando l'approccio ex post delle "perdite sostenute" (incurred loss model) e introducendo un modello di perdita attesa (Expected Credit Loss – ECL). Questo nuovo modello di impairment adotta un approccio forward-looking per classificare le attività in base alle loro performance, richiedendo un monitoraggio proattivo dei crediti articolato in tre stadi²:

- **Stage 1:** riguarda le esposizioni che non hanno subito un aumento significativo del rischio di credito. Le perdite attese sono calcolate su un orizzonte di 12 mesi;
- **Stage 2:** include le esposizioni che hanno subito un aumento significativo del rischio (SICR) ma non sono ancora in default. In questo caso, le perdite attese sono calcolate sull'intera durata residua dello strumento (lifetime ECL);
- **Stage 3:** include i crediti in default o con evidenze oggettive di deterioramento. In questo stadio, le perdite sono calcolate come già avvenute, e vi è una stretta

² IFRS Foundation (2024). *International Financial Reporting Standard 9: Financial Instruments*.

connessione con la nozione di default oggetto di un approfondimento nelle sezioni successive del capitolo.

L'identificazione del passaggio tra uno stadio e l'altro è cruciale e deve basarsi su criteri oggettivi e soggettivi. Per adeguarsi a questi nuovi standard, molti player del settore hanno disegnato processi di staging in cui il passaggio di un'esposizione a stage 2 può essere anche determinato in relazione agli output dei sistemi di monitoraggio gestionale (es. watchlist, indicatori di "rischio alto") o con riferimento ad alcune evidenze prodotte direttamente dai sistemi di Early Warning.

L'IFRS 9 richiede quindi agli intermediari l'adozione di sistemi di monitoraggio sofisticati, capaci di individuare segnali precoci di deterioramento creditizio, integrando fonti informative eterogenee. Il principio di prudenza contabile si evolve così verso un approccio più dinamico, che mira a rafforzare la solidità patrimoniale delle istituzioni finanziarie e a contenere gli effetti pro-ciclici delle crisi.

Oltre all'evoluzione della disciplina contabile, nel corso degli anni le autorità europee hanno introdotto una serie di interventi normativi volti a rafforzare la gestione delle NPE:

1) Normative sulla classificazione e accantonamento

a) Linee guida EBA in merito alla definizione di default ai sensi dell'art. 178 del Regolamento (UE) n. 575/2013, uniformando a livello europeo i criteri per classificare un'esposizione in default, ponendo ad esempio il superamento di 90 giorni come soglia temporale comune. Ciò ha aumentato la comparabilità tra banche e paesi, eliminando le disparità interpretative che in passato consentivano approcci meno rigorosi nella rilevazione del rischio di credito. Una classificazione più tempestiva e coerente ha reso più trasparente l'effettiva qualità del portafoglio crediti e ha favorito interventi più rapidi da parte delle banche³.

b) Calendar Provisioning, un insieme di regole europee con l'obiettivo di migliorare la qualità degli attivi delle banche e ridurre le esposizioni non-performing (NPL/NPE) in modo sostenibile attraverso un piano graduale di accantonamento prudenziale. Questo regime impone oneri maggiori alle banche, spingendole da un approccio "wait and see" a uno proattivo e pervasivo nella gestione dei crediti deteriorati. Le autorità europee hanno introdotto tre

³ European Banking Authority. (2016). *Guidelines on the application of the definition of default under Article 178 of Regulation (EU) No 575/2013 (EBA/GL/2016/07)*.

principali framework per garantire una copertura prudenziale minima delle esposizioni deteriorate (NPE), in base alla loro anzianità e al tipo di garanzie:

- (i) **Regolamento 2019/630 (Pillar 1)**: Impone requisiti obbligatori e non negoziabili di copertura minima per le NPE originate dopo il 26 aprile 2019, con deduzione automatica dal capitale CET1 in caso di insufficiente accantonamento contabile. Le percentuali di copertura variano in base al "vintage" (periodo di permanenza a deteriorato) e al tipo di garanzia;
- (ii) **Addendum BCE (Pillar 2)**: Si applica alle NPE originate prima di aprile 2019 ma classificate deteriorate dopo. Prevede un dialogo tra banca e BCE per giustificare eventuali scostamenti dai target di copertura, simili a quelli del Pillar1;
- (iii) **Lettere SREP (Pillar Stock)**: Riguardano le NPE più datate (ante aprile 2018) e definiscono obiettivi di copertura differenziati per gruppi di banche, all'interno del processo di vigilanza SREP. Se la copertura contabile è inferiore a quella richiesta, il differenziale può comportare un impatto diretto sul capitale CET1.

c) CRR III introduce un limite minimo regolamentare ai requisiti patrimoniali (output floor), basato su metriche standardizzate (Regolamento (UE) 2024/1623). Questo meccanismo si applica anche ad alcune categorie di esposizione, come quelle immobiliari e legate ad attività di Acquisition, Development and Construction (ADC). Di conseguenza, anche in presenza di modelli interni avanzati e sistemi di Early Warning System (EWS) molto efficaci, permane un requisito minimo di capitale, che rappresenta un costo fisso non eliminabile. L'obiettivo è armonizzare e rendere più comparabili i livelli di capitale tra istituti, prevenendo l'eccessiva variabilità dovuta ai modelli interni. Tuttavia, per le banche maggiormente esposte a questi settori, ciò comporta una pressione aggiuntiva sulla redditività, influenzando le strategie di classificazione e gestione delle esposizioni a rischio⁴.

2) Normative sulla gestione attiva e dismissione NPE:

- a) Linee guida della BCE** sui crediti deteriorati del marzo 2017, suggerendo alle banche con elevati stock di NPE la predisposizione di piani strategici di riduzione, articolati su più anni e corredati da obiettivi quantitativi. Le banche sono state

⁴ Commissione Europea. (2024). *Regolamento (UE) 2024/1623 del Parlamento europeo e del Consiglio del 31 maggio 2024 che modifica il regolamento (UE) n. 575/2013 per quanto riguarda i requisiti prudenziali per gli enti creditizi e le imprese di investimento*. Gazzetta ufficiale dell'Unione europea.

spinte a dotarsi di strutture interne dedicate (NPE units), a migliorare la governance e ad adottare una gestione attiva del portafoglio deteriorato. Questo ha favorito la dismissione degli NPL, anche grazie a una maggiore trasparenza verso il mercato⁵;

- b) Regolamenti UE 2402/2017 e 2401/2017** riformando il quadro normativo delle cartolarizzazioni, rendendolo più chiaro e sicuro, anche per le operazioni basate su crediti deteriorati (NPE). L'obiettivo è stato quello di stimolare la nascita di un mercato secondario solido per gli NPL, favorendo la cessione di portafogli da parte delle banche a investitori specializzati. Le nuove regole hanno introdotto requisiti di trasparenza, standardizzazione e due diligence, migliorando la fiducia del mercato e riducendo il rischio sistemico. In questo contesto normativo rafforzato, strumenti come le cartolarizzazioni NPL assistite da garanzia statale (ad esempio il programma GACS in Italia) si sono dimostrati più efficaci nel supportare le banche nel deconsolidamento degli NPL, contribuendo al tempo stesso a rafforzare la solidità del sistema bancario e a liberare risorse per finanziare nuovo credito all'economia reale;
- c) Linee guida EBA (EBA/GL/2018/06)** sulla gestione delle NPE e delle esposizioni forborne, fornendo un quadro completo per una gestione integrata e sostenibile del rischio di credito, includendo l'adozione di sistemi di early warning, il monitoraggio delle esposizioni ristrutturata, la strutturazione di strategie a lungo termine e la necessità di un'adeguata governance. Queste linee guida hanno contribuito a rendere la riduzione degli NPL più sistemica e duratura, evitando soluzioni solo contingenti o "di facciata";
- d) Direttiva (UE) 2021/2167**, con l'obiettivo di sviluppare un mercato secondario efficiente per i crediti deteriorati, facilitando la cessione degli NPL. Introduce regole comuni per servicer e acquirenti specializzati, aumentando la trasparenza e la protezione del debitore, e riducendo le barriere alla vendita dei crediti deteriorati. Questo consente alle banche di smobilizzare più rapidamente gli NPL, migliorando i propri bilanci, la capacità di erogare nuovo credito e la gestione complessiva del rischio. Inoltre, l'attività dei servicer autorizzati garantisce una gestione più efficace dei crediti ceduti, contribuendo alla stabilità e resilienza del sistema finanziario europeo.

⁵ Banca Centrale Europea. (2017). *Linee guida per le banche sui crediti deteriorati (NPL)*.

3) Normative sulla prevenzione e concessione prudente:

- a) **Guidelines on Loan Origination and Monitoring (EBA/GL/2020/06)**⁶, introducendo criteri più rigorosi per la valutazione della sostenibilità del debito, l'analisi del merito creditizio, l'uso responsabile dei dati e il monitoraggio nel tempo. In questo modo l'EBA ha cambiato approccio passando da uno "reattivo" (basato sulla gestione dei crediti deteriorati esistenti) ad uno "proattivo" nella gestione del credito fin dalle prime fasi della sua vita, promuovendo dunque una concessione più prudente;
- b) **Codice della crisi d'impresa e dell'insolvenza (D.Lgs. 14/2019)**⁷, che una volta entrato pienamente in vigore nel luglio 2022, ha segnato una svolta nell'approccio alla gestione delle difficoltà aziendali, spostando il focus dalla risoluzione tardiva alla prevenzione e diagnosi precoce della crisi. Questo nuovo assetto normativo ha avuto effetti rilevanti anche sul fronte del rischio di credito, in quanto ha introdotto strumenti di allerta interna ed esterna che consentono di individuare tempestivamente segnali di squilibrio economico-finanziario.
- c) **Guidelines on the Management of ESG Risks**, che a partire dal 2025, impone alle banche europee di integrare i fattori ambientali, sociali e di governance (ESG) nei propri processi di concessione e monitoraggio del credito, in coerenza con l'approccio di risk-based supervision. Le linee guida dell'EBA definiscono aspettative precise sull'identificazione, valutazione e gestione dei rischi ESG, che devono essere considerati sia *ex ante* (in fase di concessione) sia *ex post* (nel monitoraggio e nella classificazione). Questo comporta l'integrazione di nuove fonti di dati non finanziari, la modifica dei modelli di valutazione della rischiosità e la revisione delle politiche creditizie. L'obiettivo è garantire che i rischi ESG non solo siano riconosciuti, ma anche pienamente incorporati nei framework di *risk management*, contribuendo alla resilienza a lungo termine degli intermediari.⁸

Le normative europee sugli NPE hanno progressivamente coperto in modo organico tutte le fasi del ciclo di vita del credito deteriorato: dalla classificazione del default, alla gestione attiva, fino alla cessione e derecognition, includendo anche strumenti di prevenzione ex ante come i sistemi di early warning e criteri più rigorosi nella concessione del credito.

⁶ European Banking Authority. (2020). *Guidelines on loan origination and monitoring (EBA/GL/2020/06)*.

⁷ Italia. (2019). *Codice della crisi d'impresa e dell'insolvenza (D.Lgs. 14/2019)*. Gazzetta Ufficiale della Repubblica Italiana.

⁸ European Banking Authority. (2025). *Guidelines on the management of ESG risks*.

Nel loro insieme, questi interventi hanno migliorato la qualità degli attivi bancari, promosso una maggiore trasparenza, rafforzato le coperture patrimoniali e favorito lo sviluppo di un mercato secondario più efficiente. Il risultato è stato un calo significativo degli NPL in Europa e un rafforzamento della capacità delle banche di sostenere l'economia reale, aspetto che sarà esaminato più nel dettaglio in seguito.

In questo scenario normativo evoluto, assume particolare rilievo la nuova definizione di default (**New DoD**⁹), che costituisce un elemento cardine nella convergenza tra normativa contabile e regolamentare. Approfondirla consente di comprendere meglio le implicazioni operative per le banche in termini di classificazione, accantonamento e gestione attiva delle esposizioni deteriorate.

Questa riforma è nata dall'esigenza di armonizzare le diverse prassi utilizzate nei Paesi dell'Unione Europea, con l'obiettivo finale di rendere confrontabili i parametri di rischio e il capitale regolamentare degli istituti di credito ubicati in differenti stati membri. Le nuove regole, pienamente in vigore dal 1° gennaio 2021, hanno ridefinito in modo uniforme su scala europea i parametri con cui le banche valutano la solvibilità dei propri debitori. Il punto di partenza delle nuove regole è l'articolo 178 del Regolamento (UE) N. 575/2013 (CRR). Secondo questa normativa, un debitore è considerato in default (cioè, "non performing") qualora si verifichi almeno uno dei seguenti due eventi o condizioni:

- Criterio qualitativo: la banca giudica improbabile che, senza il ricorso ad azioni quale l'escussione delle garanzie, il debitore adempia integralmente alle sue obbligazioni creditizie verso l'ente stesso, la sua impresa madre o una delle sue filiazioni. Questo criterio è di natura più qualitativa e si basa sul giudizio della banca circa la capacità del debitore di far fronte alle proprie obbligazioni. Qui entrano in gioco indicatori come l'avvio di una procedura di insolvenza (come la "liquidazione giudiziale", ex fallimento), la concessione di una ristrutturazione onerosa (forbearance), o la classificazione dell'esposizione come deteriorata secondo i principi contabili o di vigilanza. È essenziale che le banche siano in grado di rilevare tempestivamente questi segnali attraverso sistemi interni solidi ed efficaci.
- Criterio quantitativo: il debitore è in arretrato da oltre 90 giorni su una obbligazione creditizia rilevante verso l'ente, la sua impresa madre o una delle sue filiazioni. Questo

⁹ European Banking Authority. (2016). *Guidelines on the application of the definition of default under Article 178 of Regulation (EU) No 575/2013 (EBA/GL/2016/07)*.

criterio si basa su parametri numerici ben precisi. Un arretrato è ritenuto "significativo" o "rilevante" solo se supera simultaneamente due soglie:

- a) Un fattore assoluto: 100 euro per le esposizioni al dettaglio (retail) e 500 euro per le altre esposizioni.
- b) Un fattore relativo: pari all'1% e calcolato come il rapporto tra l'importo della obbligazione in arretrato e l'importo complessivo di tutte le posizioni dello stesso debitore.

Le novità introdotte dalla nuova normativa possono essere così distinte:

1) Criterio del past due nell'identificazione del default: Con la nuova normativa pur non essendoci modifiche sostanziali ai criteri qualitativo e quantitativo, cambiano significativamente parametri e modalità applicative della soglia di rilevanza dell'arretrato:

- a) Il conteggio dei 90 giorni consecutivi di arretrato parte solo dopo il superamento di entrambe le componenti della soglia di rilevanza (es. 100 euro e 1% per il retail), mentre nella definizione anteriore il conteggio partiva dal primo inadempimento;
- b) La frequenza di rilevazione passa da mensile a giornaliera, con conseguente maggiore tempestività nel rilevare lo stato di default;
- c) Il conteggio si azzerava solo se una delle soglie viene meno. È prevista la sospensione in caso di malfunzionamenti del sistema dei pagamenti;
- d) Non è più ammessa la compensazione tra esposizioni: importi scaduti non possono essere compensati con linee di credito inutilizzate, rendendo più rigorosa la classificazione in default;
- e) Si conferma il principio per cui il default su una esposizione si estende automaticamente a tutte le esposizioni del debitore verso la stessa banca.

2) Indicazioni di unlikeliness to pay: Le banche devono adottare un approccio rigoroso per identificare i segnali di unlikeliness to pay, ossia l'improbabilità che un debitore sia in grado di adempiere ai propri obblighi finanziari. Questo implica non solo il monitoraggio dei ritardi nei pagamenti o delle esposizioni deteriorate, ma anche l'analisi delle cause alla base di eventuali cessioni di crediti, distinguendo se siano motivate da problemi di solvibilità o da altre ragioni. È necessario quantificare le perdite economiche significative derivanti da queste operazioni (superiori al 5%) e valutare se si tratta di singole esposizioni o di portafogli.

Particolare attenzione deve essere posta alle ristrutturazioni finanziarie che comportano una riduzione sostanziale degli obblighi di pagamento (oltre l'1%, calcolata con il tasso di interesse originario), che rappresentano un chiaro indicatore di difficoltà. Tuttavia, anche nel caso di riduzioni inferiori a questa soglia, le banche sono invitate a considerare altri segnali di difficoltà, poiché qualsiasi forma di tolleranza può indicare un rischio crescente.

Le banche devono inoltre integrare i dati interni con informazioni esterne, come quelle provenienti da registri creditizi, e sviluppare sistemi di early warning che includano indicatori più sottili di rischio, come i legami economici e giuridici tra clienti. Questo consente di valutare l'effetto contagio, ovvero la possibilità che il default di un'impresa possa influenzare negativamente la capacità di pagamento di soggetti collegati, richiedendo una mappatura e un censimento delle relazioni tra debitori.

Infine, è fondamentale che le banche adottino una governance robusta e uniforme per l'applicazione della definizione di default, con politiche e procedure chiare valide per tutte le linee di business, entità legali e sedi geografiche, assicurando così coerenza e trasparenza nel processo di valutazione del rischio di credito.

3) Condizioni per il ritorno allo stato di non inadempimento e alla cura: per uscire dallo stato di default, secondo la nuova regolamentazione devono trascorrere almeno tre mesi dal momento in cui non sussistono più le condizioni che avevano portato alla classificazione in default. Il periodo minimo di osservazione è di almeno tre mesi, durante i quali l'impresa debitrice deve mantenere una situazione di regolarità costantemente monitorata dalla banca. Al termine di questo intervallo temporale, e dopo un'attenta valutazione del comportamento e della situazione finanziaria dell'impresa, la banca può procedere alla riclassificazione dell'impresa fuori dallo stato di default. Tale riclassificazione è ammessa solo se la banca ritiene che il miglioramento della qualità creditizia sia reale e duraturo. È inoltre essenziale che eventuali nuove esposizioni concesse al debitore rispettino i criteri di solvibilità necessari per permettere l'uscita dallo stato di default.

Esiste la possibilità per le banche di applicare diversi periodi di prova a diversi tipi di esposizioni (purché soddisfino il requisito minimo). In caso di vendita dell'esposizione in stato di default, la banca deve garantire che i requisiti del periodo di prova siano applicati a qualsiasi nuova esposizione verso lo stesso debitore. Le istituzioni sono inoltre incoraggiate a cercare i casi di inadempienze multiple (un'inadempienza subito

dopo aver riacquisitato lo stato di non inadempienza) e a tenerne conto nel giudicare la durata del periodo di prova nonché i possibili indicatori di incapacità di pagare.¹⁰

La definizione regolamentare e contabile del default non si limita alla classificazione delle esposizioni, ma promuove un approccio gestionale più tempestivo e consapevole. In questo contesto si inseriscono gli Early Warning Systems (EWS), strumenti chiave per intercettare e valutare i segnali precoci di deterioramento.

1.2. Early Warning System e gestione proattiva del rischio

Negli ultimi anni, il quadro normativo europeo ha promosso un approccio alla gestione del rischio più strategico e integrato, orientato non solo alla riduzione, ma anche alla prevenzione degli NPE (Non-Performing Exposures) tramite strumenti di allerta precoce. In risposta alla crisi finanziaria globale, le autorità di vigilanza hanno innalzato le aspettative sul controllo del rischio di credito, incentivando l'adozione di modelli predittivi in grado di rilevare tempestivamente segnali di deterioramento delle esposizioni.

Gli Early Warning Systems (EWS) rispondono concretamente alla necessità di prevenire il deterioramento del credito. Si fondano sull'analisi integrata di dati finanziari, comportamentali, settoriali e macroeconomici, con lo scopo di anticipare situazioni critiche e attivare tempestivamente misure correttive prima del default. Combinano metodologie quantitative e qualitative, spaziando dai modelli statistici tradizionali agli algoritmi di machine learning, e utilizzano indicatori chiave di rischio (Key Risk Indicators) per generare segnalazioni preventive. I principali strumenti saranno approfonditi nei paragrafi successivi.

Il passaggio da una logica reattiva a una preventiva è stato accompagnato da un'evoluzione normativa significativa. Tra i principali riferimenti vi sono le Guidelines on Loan Origination and Monitoring (LOM) dell'EBA, che delineano un approccio strutturato e sostenibile all'intero ciclo di vita del credito. Le LOM pongono particolare enfasi sulla qualità delle pratiche di concessione e monitoraggio, considerando gli EWS una componente imprescindibile della governance del rischio.

¹⁰ European Banking Authority. (2016). *Guidelines on the application of the definition of default under Article 178 of Regulation (EU) No 575/2013 (EBA/GL/2016/07)*.

L'EBA adotta un'accezione più ampia del concetto di Early Warning, includendo tra i segnali anticipati di deterioramento creditizio tutti gli elementi quali-quantitativi rilevabili con largo anticipo rispetto al peggioramento del merito creditizio.

In particolare, la Sezione 5.2 delle linee guida stabilisce che gli intermediari debbano adottare sistemi di allerta formalizzati, integrati nei processi di risk management e calibrati sulla natura e complessità del portafoglio.

In quest'ottica, il sistema di monitoraggio del credito deve configurarsi come un'attività continuativa, non limitata alle sole revisioni periodiche delle esposizioni (come previsto al capitolo 8.3 delle LOM), ma estesa anche al presidio costante dei segnali precoci di rischio (early warning indicators, cfr. capitolo 8.5). Questo presidio deve essere attuato su base regolare (giornaliera, settimanale, mensile) in funzione della natura delle variabili osservate. L'attivazione degli indicatori comporta una rivalutazione complessiva della posizione del cliente, comprensiva della revisione delle esposizioni e della rischiosità associata.

Gli EWS devono includere fonti informative interne ed esterne, aggiornamenti frequenti e meccanismi di escalation ben definiti. Gli indicatori utilizzati devono coprire dimensioni:

- Quantitative (es. ritardi nei pagamenti, peggioramento del rating interno)
- Qualitative (es. cambiamenti nella governance, tensioni settoriali)
- Comportamentali (es. movimenti anomali di conto o linee di credito)
- Alternative, come i dati da centrale rischi, open data o indicatori ESG, purché validati metodologicamente

Gli indicatori di preallerta dovrebbero prevedere soglie di attivazione coerenti con la propensione al rischio di credito definita nelle strategie e nelle politiche aziendali. A ciascun indicatore devono essere associate procedure di escalation, con chiara attribuzione delle responsabilità per le azioni di follow-up. Tali procedure includono anche l'utilizzo delle watch list, strutturate su livelli crescenti di attenzione (cfr. capitolo 8.5 delle LOM): dalle semplici liste di osservazione fino a cluster ad alta priorità, per i quali sono richieste azioni rapide e risolutive, al fine di evitare lo scivolamento dell'esposizione verso la classificazione a credito deteriorato (NPE).

L'Annex 3 delle EBA GL LOM fornisce una guida operativa sugli Early Warning Indicators (EWI), articolandoli in modo sistematico per supportare le istituzioni finanziarie nella selezione e calibrazione degli indicatori più rilevanti.

Inoltre, l'EBA supporta ulteriormente l'utilizzo degli EWI fornendo un elenco aggiornato di indicatori di rischio e una guida metodologica (EBA Methodological Guide) che ne descrive il

calcolo e l'interpretazione. Questa documentazione, pur non introducendo nuovi obblighi di segnalazione, consente alle autorità competenti e agli operatori di analizzare in modo coerente i dati bancari, rafforzando le capacità di valutazione preventiva del rischio di credito.

L'implementazione degli EWS non rappresenta un mero adempimento, ma uno snodo fondamentale per decisioni strategiche e prudenziali, come:

- la riclassificazione tra Stage 1 e Stage 2 secondo l'IFRS 9;
- la valutazione della condizione di "unlikelihood to pay" ai sensi del Regolamento CRR;
- l'attivazione di azioni di contenimento e ristrutturazione.

Questo approccio rafforza la resilienza del portafoglio crediti e promuove la stabilità complessiva del sistema bancario. La prospettiva proposta dall'EBA implica un cambiamento di paradigma: dagli strumenti statici a un monitoraggio dinamico, proattivo e data-driven, pienamente integrato nelle strategie creditizie.

A livello nazionale, il Codice della crisi d'impresa e dell'insolvenza, coerente con la Direttiva UE 2019/1023¹¹, ha rafforzato l'orientamento preventivo nella gestione del rischio, promuovendo la tempestiva rilevazione delle difficoltà aziendali. Il Codice, con l'aiuto del Consiglio Nazionale dei Dottori Commercialisti, introduce una serie di indicatori patrimoniali, reddituali e finanziari per le PMI, richiedendo agli organi amministrativi un'attenta attività di monitoraggio interno. Gli indicatori suggeriti dalle normative europee e nazionali possono essere ricondotti a diverse macrocategorie operative come ad esempio:

- **Economici:** riguardano il contesto macroeconomico e settoriale, nonché la sensibilità dell'impresa ai cicli di mercato.
- **Finanziari e patrimoniali:** analizzano la struttura del capitale, il livello di liquidità, l'indebitamento e la solidità dell'attivo.
- **Reddituali:** valutano la capacità dell'azienda di generare utili in modo stabile e sostenere i margini operativi nel tempo.
- **Rischio:** misurano l'esposizione a fonti specifiche di rischio, come il rischio di credito, di liquidità o di tasso d'interesse.
- **Performance:** esaminano l'efficienza gestionale e il grado di raggiungimento degli obiettivi strategici prefissati.

Mentre l'EBA LOM è orientata principalmente alla qualità e sostenibilità della concessione e del monitoraggio del credito, il Codice della crisi d'impresa pone l'accento sulla tempestiva

¹¹ Gazzetta Ufficiale della Repubblica Italiana (2019). *Codice della crisi d'impresa e dell'insolvenza in attuazione della legge 19 ottobre 2017, n. 155.*

rilevazione di segnali di crisi, con particolare attenzione alle PMI e alla responsabilità gestionale.

Nel seguito dell'elaborato saranno analizzati i modelli predittivi impiegati dalle banche, gli indicatori più rilevanti, le principali fonti informative utilizzate e le prospettive evolutive degli EWS, con particolare attenzione all'efficienza operativa e alla conformità normativa.

1.3. Panoramica sui trend del credito e dei default

Il contesto macroeconomico debole, combinato con l'incertezza geopolitica e la fine delle misure emergenziali introdotte durante la pandemia, costituisce lo sfondo delle previsioni contenute nell'Outlook Abi-Cerved, secondo cui il tasso di deterioramento del credito alle imprese è atteso in crescita nei prossimi anni¹². Dopo il dato del 2,4% registrato nel 2023, si stima un incremento al 3,5% nel 2024. Questo aumento riflette l'effetto ritardato di fattori avversi come l'inflazione elevata, il costo crescente del debito dovuto ai rialzi dei tassi di interesse da parte della Banca Centrale Europea e la persistente instabilità geopolitica (in particolare per la prosecuzione del conflitto russo-ucraino). Tuttavia, a partire dal 2025, la prevista ripresa economica, unita a una politica monetaria meno restrittiva, dovrebbe favorire una graduale riduzione del rischio di credito, con una stima del tasso di deterioramento in discesa al 3,2% nel 2025 e al 2,7% nel 2026. Questo livello, pur rimanendo superiore al dato del 2023, si colloca sotto la soglia del 2,9% registrata nel 2019, prima della pandemia.

Nel primo semestre del 2024, il tasso di default delle imprese italiane è rimasto stabile, con una media del 2,32%, in linea con la forchetta storica 2021-2023 (2,1-2,3%).

Tuttavia, il dato nasconde differenze tra settori, forme giuridiche e dimensioni. Di seguito si riporta il dettaglio per le diverse stratificazioni:

- **Forma giuridica**

- Le *ditte individuali* registrano un tasso del 2,42%;
- Le *società di persone* si attestano all'1,60%;
- Le *società di capitali* riportano un tasso del 2,46%. Per queste ultime si prevede un peggioramento, con un tasso atteso al 2,9% entro la fine del 2024 e al 3,5% entro il 2025, a causa delle persistenti fragilità economiche e settoriali.

¹² Cerved. "Outlook ABI-Cerved 2024-26. Crediti deteriorati delle imprese: tassi in aumento contenuto". 01 luglio 2024

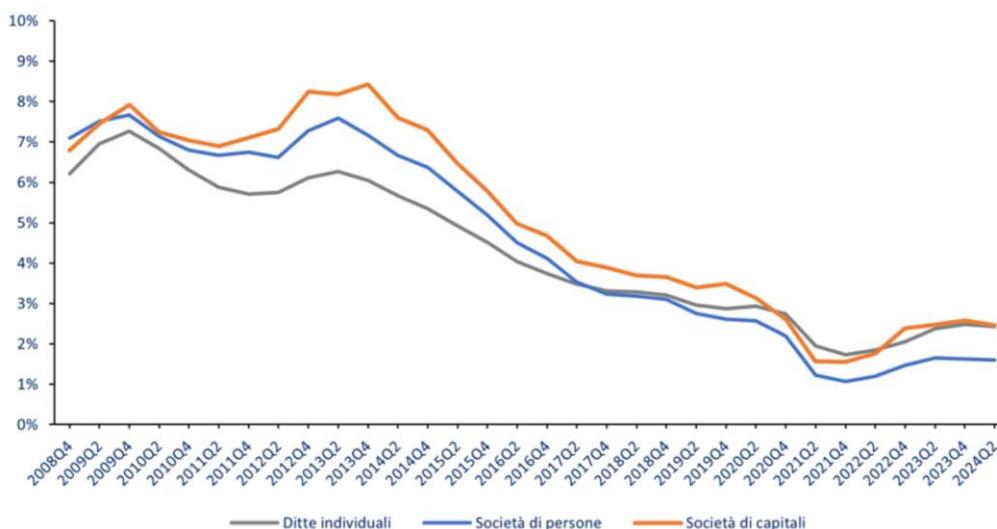


Figura 1: Andamento tassi di default per tipologia di impresa. Fonte: CRIF¹³

Anche il credito erogato alle imprese si è mantenuto su livelli stabili rispetto al primo semestre del 2023, mostrando un leggerissimo calo in termini di numero di finanziamenti (-0,1%) ma una lieve crescita in termini di importi complessivi (+0,9%). Questa dinamica riflette una prudente ma costante offerta di credito da parte degli intermediari, che tuttavia mostra forti differenze tra settori economici.

- **Settore**

- *Tessile e Abbigliamento* si conferma tra i più in difficoltà. Le società di capitali appartenenti a questo settore hanno fatto registrare uno dei più marcati aumenti del tasso di default rispetto a fine 2023, pari a +0,43 punti percentuali (passando da 2,9% a 3,33), a cui si accompagna un significativo calo degli importi erogati (-9,1%). Il deterioramento potrebbe riflettere pressioni su margini e tensioni commerciali internazionali.
- *Costruzioni*, sebbene mantenga un tasso di default stabile su livelli elevati (3,11%), ha subito una contrazione delle erogazioni pari al -6%, influenzata principalmente dalla fine del ciclo positivo legato agli incentivi ecobonus che avevano sostenuto la domanda nel triennio precedente e all'aumento dei costi di approvvigionamento.
- *Turismo*, che prosegue nella sua tendenza positiva. Le società di capitali del settore hanno beneficiato di un aumento degli importi erogati pari a circa l'11% e di un calo

¹³ Osservatorio CRIF sulle imprese, aprile 2024: Elaborazioni CRIF Ratings

del tasso di default di circa -0,4 punti percentuali, portandosi al 3,46% nel primo semestre 2024.

- *Trasporti e Logistica*, hanno risentito ancora del difficile contesto geopolitico, ma la stabilizzazione e il successivo calo dei prezzi dei carburanti nei primi mesi del 2024 hanno contribuito a contenere la rischiosità: il tasso di default per le società di capitali si è mantenuto stabile al 3,6%. Sul fronte del credito, il comparto ha registrato un significativo aumento degli importi erogati (+10,5%)¹⁴

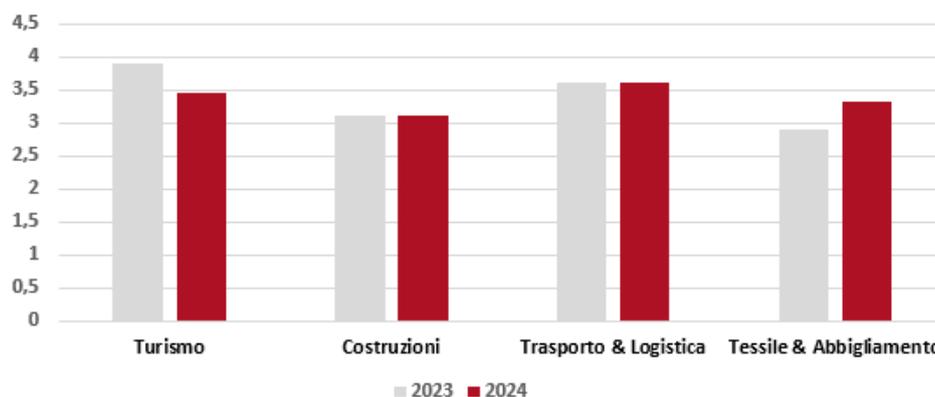


Figura 2: Confronto settoriale dei tassi di default. Fonte: CRIF

● Dimensione

Tutte le classi di imprese hanno mostrato una crescita dei tassi di deterioramento nel 2024, con incrementi omogenei:

- Le *microimprese* e le *medie imprese* hanno segnato un +1,0% (passando rispettivamente dal 2,7% al 3,7% e dall'1,3% al 2,3%);
- Le *piccole e grandi imprese* hanno registrato rialzi dello 0,8% e dello 0,9% (da 1,8% a 2,6% e da 1,1% a 2,0%).

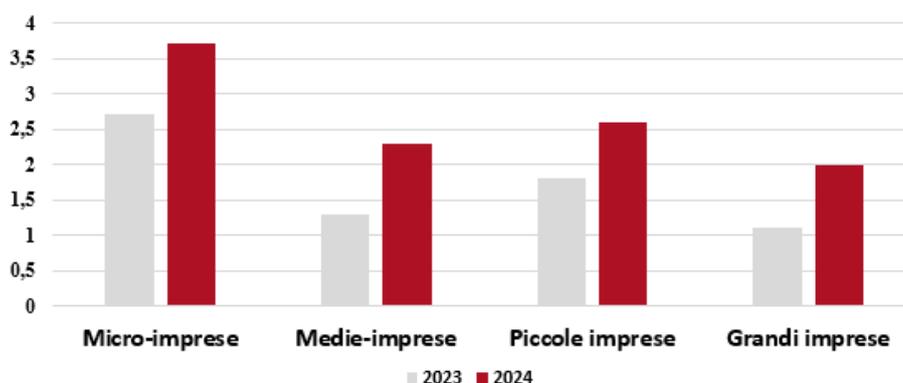


Figura 3: Tassi di deterioramento per categorie d'impresa. Fonte: CRIF

¹⁴ CRIF. "Il tasso di default delle imprese sale al 2,39% nel 2023, aumenta anche il credito erogato". 17 aprile 2024

Le Costruzioni restano il settore con i tassi di deterioramento più alti, seguite dai Servizi (3,6% nel 2024). Rispetto al 2019, Agricoltura, Industria e Servizi hanno superato i livelli pre-Covid, mentre solo le Costruzioni sono rimaste stabili. L'aumento diffuso dei tassi rappresenta un segnale d'allerta per i modelli di rischio e le politiche creditizie. Le previsioni per il 2025-2026 indicano una graduale ripresa economica, sostenuta dall'allentamento della politica monetaria, con un miglioramento atteso della qualità del credito: i tassi di deterioramento caleranno per medie e grandi imprese, mentre resteranno più elevati per micro e piccole.

In quest'ottica, l'adozione e lo sviluppo di sistemi di preallarme risultano centrali per identificare tempestivamente segnali di deterioramento e consentire interventi proattivi¹⁵. L'integrazione degli EWS nei processi di monitoraggio del credito consentirebbe di mitigare l'impatto dei futuri shock sistemici, offrendo a banche e policy maker una leva operativa fondamentale per preservare la stabilità del sistema finanziario e sostenere l'economia reale.

2. Integrazione di dati e modelli per l'analisi del rischio

Negli ultimi anni, il crescente ricorso a strumenti di **modellizzazione predittiva** ha trasformato radicalmente le modalità con cui le istituzioni finanziarie affrontano il rischio di credito. In particolare, gli **Early Warning System (EWS)** stanno assumendo un ruolo sempre più centrale nei processi di monitoraggio continuo con l'obiettivo di anticipare situazioni di deterioramento creditizio attraverso l'uso di specifici indicatori.

L'evoluzione tecnologica, sostenuta anche dalle indicazioni normative contenute nelle già citate Linee Guida EBA sulla concessione e il monitoraggio dei prestiti (LOM), ha incentivato le banche ad adottare soluzioni avanzate, basate su algoritmi di machine learning, modelli statistici non lineari e, più recentemente, sull'integrazione di **dati proprietari** con **Open Data**.

In particolare, la sezione 4.3.3 delle Linee Guida incoraggia esplicitamente l'uso di tecnologie innovative nei processi di valutazione del merito creditizio, riconoscendo il valore aggiunto che strumenti digitali evoluti possono offrire in termini di efficienza e accuratezza decisionale. Tuttavia, come chiarito nella successiva sezione 4.3.4, l'impiego di tali modelli richiede un presidio rigoroso: è necessario garantirne la validazione, documentarne in modo trasparente funzionamento e presupposti, assicurarne il monitoraggio continuo e mantenerne il controllo

¹⁵ ABI. "Outlook Abi-Cerved sui crediti deteriorati delle imprese". Gennaio 2023

attraverso unità indipendenti. Fondamentale, inoltre, è evitare un'eccessiva dipendenza da sistemi opachi ("black box"), preservando sempre un adeguato livello di comprensibilità e responsabilità umana nel processo decisionale.

In questo contesto, modelli come la **Regressione Logistica** (Logit) e le **Random Forest** vengono ampiamente utilizzati per il loro potenziale predittivo, ma non sono esenti da criticità, quali la difficoltà di interpretazione, la gestione del rischio di modellizzazione e la necessità di garantire trasparenza, tracciabilità e solidità nel tempo.

Di seguito verranno analizzate le suddette metodologie, mettendone in luce vantaggi e criticità, e approfondendo le potenzialità offerte dall'integrazione di Open Data e intelligenza artificiale per migliorare la capacità predittiva e la solidità dei modelli, superandone i limiti attuali.

2.1 Struttura e utilizzo degli indicatori di rischio nei modelli predittivi

Alla base del funzionamento degli EWS vi è l'impiego coordinato di indicatori quantitativi e qualitativi, che consentono di cogliere segnali deboli e discontinuità rispetto al normale andamento del debitore. Gli *indicatori quantitativi*, ricavati da fonti contabili, andamentali o settoriali, offrono misure oggettive di liquidità, solvibilità e sostenibilità dell'indebitamento. Gli *indicatori qualitativi*, invece, permettono di intercettare fattori più difficili da quantificare ma spesso rilevanti, come cambiamenti organizzativi, eventi reputazionali o segnali di deterioramento della trasparenza informativa.

L'efficacia degli EWS risiede proprio nella capacità di combinare queste dimensioni eterogenee in un framework integrato, dinamico e coerente con il profilo di rischio della controparte. La tempestività nella rilevazione e la capacità predittiva degli indicatori sono elementi essenziali per garantire la prontezza dell'azione correttiva, rafforzando così la stabilità del sistema creditizio.

Tale approccio è formalizzato a livello europeo nell'Annex III dell'*EBA LOM*¹⁶ attraverso l'identificazione di specifici indicatori, mentre a livello nazionale il *Codice della Crisi d'Impresa e dell'Insolvenza*¹⁷ ne promuove e incoraggia l'adozione.

Di seguito si propone una classificazione degli indicatori fondamentali per un sistema efficace di allerta precoce; si precisa tuttavia che l'elenco presentato non intende essere esaustivo, in

¹⁶ European Banking Authority (2020). *Guidelines on loan origination and monitoring*.

¹⁷ Gazzetta Ufficiale della Repubblica Italiana (2019). *Codice della crisi d'impresa e dell'insolvenza in attuazione della legge 19 ottobre 2017, n. 155*.

quanto esistono numerosi ulteriori indicatori variabili a seconda del contesto e delle specifiche necessità:

2.1.1 Indicatori patrimoniali

Gli indicatori patrimoniali consentono di valutare la solidità e la struttura finanziaria dell'impresa, fornendo informazioni chiave sulla sua capacità di sostenere nel tempo la propria attività.

2.1.1.1 Patrimonio netto:

$$\text{Patrimonio Netto} = \text{Totale Attivo} - \text{Totale Passivo}$$

Qualora il calcolo restituisca un valore negativo, significa che le passività eccedono le attività dell'impresa, segnalando una compromissione della continuità aziendale e un potenziale stato di insolvenza, in linea con quanto previsto dal Codice della Crisi.

2.1.1.2 Loan-to-Value Ratio (LTV):

$$LTV = \frac{\text{Ammontare del prestito}}{\text{Valore del bene posto a garanzia}}$$

Esprime il grado di copertura del credito tramite le garanzie reali: un valore elevato implica maggiore esposizione al rischio in caso di insolvenza del debitore.

Se superiore all'80% è generalmente considerato rischioso nei mutui ipotecari.

2.1.1.3 Equity Ratio:

$$\text{Equity Ratio} = \frac{\text{Patrimonio Netto}}{\text{Totale Attivo}}$$

Indica quanto del totale delle attività è finanziato con mezzi propri.

Se inferiore al 5% segnala una dipendenza eccessiva da capitale di terzi, come indicato dal Consiglio Nazionale dei Dottori Commercialisti¹⁸.

2.1.1.4 Indice di liquidità corrente (Current Ratio):

$$\text{Current Ratio} = \frac{\text{Attività correnti}}{\text{Passività correnti}}$$

Mostra la capacità dell'impresa di far fronte ai debiti a breve termine con le attività a breve.

Un valore inferiore a 1 segnala una potenziale difficoltà nel far fronte alle obbligazioni a breve.

¹⁸ Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili (2019). *Crisi d'impresa. Gli indici dell'allerta*.

2.1.2 Indicatori reddituali

Gli indicatori reddituali misurano la capacità dell'impresa di generare utili attraverso la gestione operativa e di sostenere nel tempo gli oneri finanziari. Questi indicatori forniscono una valutazione della redditività aziendale e della sostenibilità economica delle sue attività.

2.1.2.1 Interest Coverage Ratio (ICR):

$$ICR = \frac{EBIT}{\text{Oneri finanziari}}$$

Indica in che misura il risultato operativo è in grado di coprire gli oneri finanziari.

Un valore inferiore a 1,5 evidenzia una crescente vulnerabilità nella copertura degli interessi passivi.

2.1.2.2 Incidenza degli oneri finanziari sui ricavi:

$$\text{Indice di copertura degli oneri finanziari} = \frac{\text{Oneri finanziari}}{\text{Ricavi}}$$

Esprime il peso degli interessi passivi sui ricavi aziendali. Un'incidenza elevata riduce la redditività. Valori superiori al 3% possono riflettere uno stress da indebitamento eccessivo.

2.1.2.3 DSCR (Debt Service Coverage Ratio):

$$DSCR = \frac{EBITDA}{(\text{Oneri finanziari} + \text{Quote Capitale in scadenza})}$$

Misura la capacità dell'impresa di generare flussi di cassa sufficienti a coprire il servizio del debito, includendo sia la quota capitale sia gli interessi. Un valore inferiore a 1 indica che i flussi generati non sono sufficienti a far fronte agli impegni finanziari, segnalando una potenziale situazione di tensione finanziaria.

2.1.2.4 ROA (Return on Assets):

$$ROA = \frac{\text{Utile netto}}{\text{Totale attività}} \times 100$$

Misura la capacità dell'impresa di generare utile netto rispetto al totale delle attività impiegate. Indica l'efficienza con cui le risorse aziendali vengono utilizzate per produrre profitto. Un valore più elevato riflette una gestione efficace degli asset, mentre un valore basso può segnalare un impiego inefficiente delle risorse o margini operativi ridotti.

2.1.2.5 ROE (Return on Equity):

$$ROE = \frac{Utile\ netto}{Patrimonio\ netto} \times 100$$

Misura la redditività del capitale proprio investito nell'impresa, indicando quanto utile viene generato per ogni unità di patrimonio netto. È un indicatore chiave per valutare la capacità dell'azienda di remunerare gli azionisti. Un valore elevato segnala una buona capacità di generare profitti per i soci, mentre un valore basso può indicare una bassa efficienza nella gestione del capitale proprio o una struttura finanziaria poco ottimale.

2.1.3 Indicatori finanziari

Gli indicatori finanziari valutano la capacità dell'azienda di sostenere il proprio debito nel tempo e la liquidità disponibile per far fronte agli obblighi finanziari correnti. In particolare, misurano quanto l'impresa sia solida dal punto di vista finanziario, la sua capacità di onorare pagamenti a breve e lungo termine, e la stabilità complessiva della sua struttura patrimoniale.

2.1.3.1 Debt-to-Income (D/I)

$$DTI = \frac{Debito\ totale}{Reddito\ netto\ annuo}$$

Mostra il livello di indebitamento rispetto alle capacità reddituali.

Valori oltre il 40-45% indicano un indebitamento eccessivo rispetto alla capacità di rimborso.

2.1.3.2 Debt-to-Equity (D/E)

$$DTE = \frac{Debito\ totale}{Patrimonio\ Netto}$$

Indica il livello di leva finanziaria dell'impresa. Un rapporto elevato evidenzia una maggiore esposizione al rischio finanziario e una ridotta autonomia patrimoniale. Un valore superiore a 2 è generalmente associato a un'elevata leva finanziaria, mentre per le imprese consolidate un valore inferiore a 1 o compreso tra 1 e 2 è considerato più prudente.

2.1.4 Indicatori di rischio

Gli indicatori di rischio sono strumenti utilizzati per identificare tempestivamente possibili situazioni di deterioramento o vulnerabilità finanziaria di un'azienda o di un portafoglio crediti. Questi indicatori permettono di monitorare l'evoluzione del rischio.

2.1.4.1 Downgrade del rating

Un declassamento di due o più livelli (notch) nel corso di un anno costituisce un segnale importante per l'attivazione di azioni di monitoraggio più approfondite.

2.1.2.4 Ritardi nei pagamenti

Ritardi superiori a 30 giorni, pur non configurando automaticamente un default ai sensi dell'art. 178 CRR, sono comunque indicatori rilevanti ai fini del monitoraggio del rischio.

2.1.5 Indicatori qualitativi

Gli indicatori qualitativi rappresentano fattori non numerici che forniscono informazioni essenziali sul contesto, i comportamenti e le condizioni non immediatamente evidenti nei dati finanziari. Questi indicatori aiutano a identificare segnali precoci di rischio creditizio, valutando aspetti come la qualità della governance, i cambiamenti organizzativi, la reputazione e il contesto operativo del debitore. Di seguito una rassegna:

- **Instabilità del management:** può compromettere la continuità strategica e operativa dell'impresa.
- **Contenziosi legali significativi:** possono generare passività potenziali e incertezza finanziaria.
- **Perdita di contratti chiave:** indica un possibile calo dei ricavi futuri e della sostenibilità del business.
- **Scarsa cooperazione del debitore nella trasparenza informativa:** ostacola la valutazione del rischio e può celare situazioni critiche.
- **Mancata presentazione tempestiva dei documenti contabili:** segnala disorganizzazione o volontà di occultare problemi finanziari

2.1.6 Indicatori macroeconomici

Infine, per completezza e maggiore efficienza, gli EWS efficaci devono integrare parametri che riflettano la performance prospettica e il contesto operativo dell'impresa, includendo variabili macroeconomiche all'interno dei modelli interni.:

- **Variazione PIL:** una contrazione del PIL segnala una fase recessiva o di rallentamento economico, che può ridurre la capacità delle imprese di generare ricavi e rispettare gli obblighi di pagamento, aumentando così il rischio di default.
- **Tassi d'interesse:** un aumento dei tassi d'interesse accresce il costo del debito per le imprese, incidendo negativamente sulla loro sostenibilità finanziaria e quindi sul rischio di insolvenza.
- **Tasso di Inflazione:** un'inflazione elevata o volatile può erodere il potere d'acquisto e aumentare i costi operativi delle imprese, comprimendo i margini e influenzando negativamente la capacità di rimborso.
- **Tasso di disoccupazione:** un aumento della disoccupazione riduce il reddito disponibile delle famiglie, con possibili effetti negativi sui consumi e sulla domanda di beni e servizi, influenzando indirettamente la performance delle imprese e il rischio di default.
- **Variazione prezzi immobiliari:** il calo dei prezzi immobiliari può ridurre il valore delle garanzie reali a supporto dei prestiti, aumentando il rischio di perdita per le banche in caso di inadempienza.
- **Variazione vendite al dettaglio:** un calo delle vendite al dettaglio indica una diminuzione della domanda interna, che può segnalare difficoltà per le imprese nel generare ricavi e onorare i debiti.
- **Variazione Debito pubblico:** un aumento del livello di debito pubblico può riflettersi in politiche fiscali restrittive o instabilità economica, influenzando negativamente il contesto operativo delle imprese e aumentando il rischio complessivo di default.

Questi indicatori, adattati per controparte, settore e contesto geografico, aiutano a individuare segnali di deterioramento anche in assenza di evidenze contabili. Le banche sono inoltre tenute a implementare analisi di scenario e stress test per valutare l'impatto di shock macroeconomici su flussi di cassa e capacità di rimborso.

In conclusione, la costruzione di un sistema di allerta precoce realmente efficace richiede una selezione coerente di indicatori, validati statisticamente e calibrati rispetto al contesto settoriale e dimensionale dell'impresa. Non esiste una soglia unica o un indicatore infallibile, ma una rete di segnali che, se correttamente monitorati e interpretati, può offrire un vantaggio cruciale nella prevenzione dei default e nella salvaguardia della qualità del credito.

2.2 Approcci statistici e modelli applicabili nei sistemi di allerta

La valutazione del rischio di insolvenza di un'impresa può essere affrontata secondo due approcci principali: il primo si basa su modelli strutturali che prendono in considerazione il valore degli attivi aziendali; il secondo impiega tecniche statistiche, come l'analisi discriminante o modelli regressivi, costruiti su dati storici. Negli ultimi anni, l'applicazione del machine learning e dell'intelligenza artificiale ha contribuito a migliorare la selezione delle variabili rilevanti e l'accuratezza delle stime.

Il modello strutturale più conosciuto è quello proposto da Merton (1974), secondo cui il default si verifica quando il valore degli attivi aziendali scende al di sotto delle passività. Tuttavia, questo approccio richiede dati di mercato che spesso non sono disponibili per le piccole imprese, che costituiscono la maggior parte del tessuto imprenditoriale italiano e tale motivo, si tende a privilegiare l'impiego di modelli statistici, più adatti a contesti con dati limitati.

Conclusa l'analisi degli indicatori tipici previsti dalla normativa in materia di Early Warning System, l'attenzione si concentra ora sugli approcci statistici più diffusi per la calibrazione e l'applicazione dei modelli predittivi di default.

2.2.1 Regressione Logistica

Nel contesto degli Early Warning System (EWS), la regressione logistica è uno degli approcci statistici più diffusi per stimare la probabilità che si verifichi un evento critico. Si tratta di un metodo di classificazione binaria che, tramite la funzione logit, modella la probabilità di appartenenza a una delle due classi in base a variabili indipendenti. I parametri del modello sono stimati con il metodo della massima verosimiglianza, garantendo un'interpretazione probabilistica coerente.

2.2.1.1 Definizione del modello e metodologia di stima

Il modello logit costituisce una specifica del modello lineare generalizzato, in cui la funzione di collegamento adottata è la **funzione logistica**. Tale modello è particolarmente impiegato quando la variabile dipendente y_i è **dicotomica**, ovvero assume esclusivamente i valori 0 o 1, con probabilità associate rispettivamente pari a π e $1-\pi$.

I valori restituiti corrispondono:

- 1 se la banca è in difficoltà
- 0 altrimenti

In generale, y_i è distribuita secondo una **Bernoulli** con parametro π , e la relativa funzione di probabilità è:

$$P\{y_i\} = \pi^{y_i}(1 - \pi)^{1-y_i}$$

Dunque:

- Se $y_i = 1$ allora $\Pr(y_i = 1) = \pi$
- Se $y_i = 0$ allora $\Pr(y_i = 0) = 1 - \pi$

Una prima formulazione è quella del **modello di probabilità lineare**:

$$\pi_i = x_i' \beta$$

Dove:

- x_i è il vettore delle variabili esplicative,
- β è il vettore dei coefficienti, i quali rappresentano gli effetti marginali delle variabili indipendenti sulla probabilità di default

Una limitazione significativa di questo modello è che il valore stimato π_i potrebbe trovarsi **fuori dall'intervallo [0,1]**, a meno di imporre restrizioni sui coefficienti. Inoltre, le stime risultano affette da **eteroschedasticità** e gli errori non seguono una distribuzione normale.

Per ovviare a tali problematiche, si ricorre al **modello logit**, che applica una trasformazione non lineare alla probabilità π_i mediante la funzione logistica $\Lambda(\cdot)$, garantendo che il risultato sia sempre compreso tra 0 e 1:

$$\pi_i = \Lambda(x_i' \beta) = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)}$$

dove:

- Il numeratore rappresenta l'esponenziale del predittore lineare,
- Il denominatore normalizza il valore affinché la probabilità sia compresa nell'intervallo corretto.

La stima dei coefficienti β avviene tramite il metodo della **massima verosimiglianza**. Nel modello logit, a causa della natura non lineare della specificazione, i coefficienti β non possono essere interpretati direttamente come effetti marginali. Essi indicano solo il segno dell'effetto di ciascuna variabile sulla probabilità di default, mentre l'effetto quantitativo richiede un calcolo specifico degli effetti marginali. Inoltre, non viene esplicitata una variabile

di errore ε , poiché l'equazione è costruita direttamente su una trasformazione del valore atteso di y_i . Di conseguenza, non esiste una varianza residua in senso classico, e quindi non è definibile un coefficiente di determinazione R^2 analogo al modello lineare.

Tuttavia, è possibile calcolare uno **pseudo- R^2** , proposto da **McFadden**, come misura alternativa della bontà dell'adattamento del modello:

$$\widetilde{R^2} = 1 - \frac{l(\hat{\beta})}{l(\hat{y})}$$

Dove:

- il numeratore è il **log-verosimiglianza** del modello stimato,
- il denominatore è la **log-verosimiglianza** del modello contenente solo l'intercetta, ovvero costruito utilizzando la **media campionaria delle risposte** come unica informazione esplicativa.

Questo indice varia tra 0 e 1 in funzione della capacità del modello di spiegare le risposte osservate, ma non rappresenta la proporzione della varianza spiegata, come nel caso dell' R^2 classico. Valori tra 0,2 e 0,4 sono considerati buoni, mentre oltre 0,4 indicano un ottimo modello

2.2.1.2 Pregi del modello logit

- **Modella variabili binarie:** è ideale quando la variabile dipendente è dicotomica (es. default/non default). Permette di stimare la probabilità che un evento si verifichi;
- **Output interpretabile come probabilità:** i coefficienti possono essere trasformati in odds ratio; quindi, è facile interpretare l'effetto di ogni variabile indipendente sulla probabilità dell'evento;
- **Non linearità tra variabili indipendenti e dipendenti:** a differenza della regressione lineare, non serve che la relazione tra predittori e outcome sia lineare; la funzione logit gestisce la non linearità nella probabilità;
- **Buona performance con dati non perfettamente normali:** non assume normalità degli errori come nella regressione lineare; quindi, è più flessibile con dati reali;
- **Predizione e classificazione:** fornisce una stima della probabilità, che può essere usata per classificare osservazioni in due classi tramite una soglia (es. 0.5);
- **Facilmente estendibile:** può essere estesa a modelli multiclasse (multinomial logit) o modelli con effetti casuali (logit misto);
- **Robustezza:** funziona bene anche con piccoli campioni rispetto ad altri modelli più complessi.

In generale tra i principali vantaggi della regressione logistica si segnalano l'efficienza computazionale, che la rende adatta all'analisi di dataset di medie dimensioni, e la trasparenza interpretativa: i coefficienti stimati, infatti, permettono di identificare chiaramente il peso delle singole variabili indipendenti, aspetto cruciale in contesti come gli *Early Warning Systems* (EWS), dove è essenziale guidare decisioni operative o politiche attraverso evidenze quantitative. Un ulteriore punto di forza è la capacità di calibrare le probabilità in modo diretto, caratteristica sfruttata per definire soglie di intervento pratiche (ad esempio, classificando come "rischio elevato" tutti i casi con $P > 0,7$).

2.2.1.3 Limiti del modello logit

- **Dipendenza da specifica corretta del modello:** se alcune variabili importanti sono escluse o la forma funzionale è sbagliata, i risultati possono essere distorti;
- **Difficoltà con dati molto sbilanciati:** se una classe è molto rara rispetto all'altra (es. default molto raro), il modello può faticare a prevedere correttamente;
- **Non adatto a relazioni altamente complesse:** se la relazione tra variabili è molto complessa (interazioni multiple, effetti non additivi), la regressione logit semplice può non catturarla bene;
- **Presuppone indipendenza delle osservazioni:** se i dati sono raggruppati o correlati (es. clienti in filiali diverse), il modello base può produrre stime non corrette senza estensioni (logit misto);
- **Interpretazione complicata con molte variabili o trasformazioni:** quando il modello ha molte variabili o interazioni, interpretare gli odds ratio può diventare difficile;
- **Non gestisce bene variabili indipendenti altamente correlate:** la multicollinearità può causare instabilità nelle stime dei coefficienti;
- **Predizione non sempre precisa fuori dal campione:** il modello può soffrire di overfitting o bassa generalizzazione se non calibrato correttamente.

Tra i limiti si segnala dunque il presupposto di linearità nei logit, che impone una relazione lineare tra le variabili esplicative e il logaritmo degli odds; questo vincolo può rendere necessarie trasformazioni delle feature (come termini polinomiali o interazioni) per catturare relazioni non lineari. Inoltre, la regressione logistica è sensibile a correlazioni forti tra le variabili indipendenti: situazioni di multicollinearità, infatti, possono compromettere l'affidabilità delle stime dei coefficienti. Infine, in scenari con eventi target rari, il modello tende

a sottostimare la probabilità degli eventi meno frequenti, richiedendo l'applicazione di tecniche di bilanciamento dei dati come l'oversampling o l'aggiustamento dei pesi delle classi.

In conclusione, la regressione logistica rappresenta uno strumento solido e interpretabile per gli EWS, grazie alla sua semplicità e capacità predittiva. Tuttavia, la sua efficacia dipende da una corretta specificazione del modello e dall'attenzione a limiti strutturali come la non linearità e la gestione delle interazioni complesse.

2.2.2 Random Forest

Negli Early Warning System (EWS) le *Random Forest* emergono come strumenti predittivi efficaci per la loro flessibilità e robustezza. A differenza dei modelli lineari come la regressione logistica, questo approccio di machine learning adotta una logica ensemble, fondata sulla combinazione sinergica di centinaia di alberi decisionali. Tale architettura consente di bilanciare accuratezza predittiva e stabilità, superando limiti tipici dei singoli alberi, come l'overfitting.

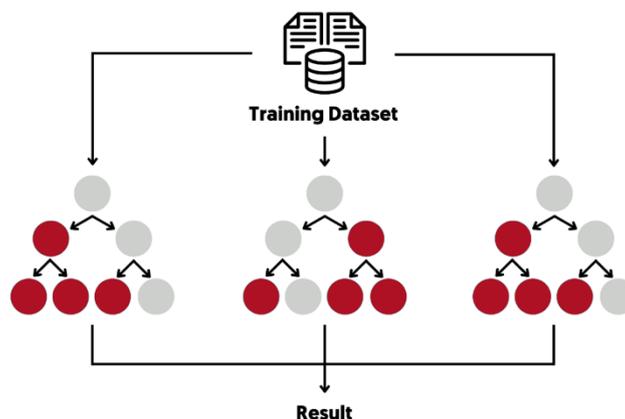


Figura 4: Random Forest

2.2.2.1 Definizione del modello e metodi di stima

Largamente impiegato in ambito predittivo grazie alla sua abilità nel catturare relazioni non lineari e gestire dataset ad alta dimensionalità, la struttura della Random forest si basa su un insieme di **alberi decisionali** generati in modo indipendente, la cui combinazione consente di incrementare la stabilità e l'affidabilità del modello rispetto all'utilizzo di un singolo albero.

Il processo di costruzione dell'ensemble si fonda su due elementi centrali:

1. **Bagging (Bootstrap Aggregating)**: ciascun albero viene addestrato su un sottoinsieme casuale estratto con reinserimento dal dataset originale. Questo meccanismo consente

di ridurre la varianza del modello complessivo, migliorandone la capacità di generalizzazione.

2. **Selezione casuale delle feature:** a ogni divisione interna degli alberi, viene selezionato in modo casuale un sottoinsieme di variabili candidate allo split, aumentando l'eterogeneità tra gli alberi e riducendo la correlazione interna dell'ensemble, con conseguente aumento dell'accuratezza predittiva.

Nei problemi di **classificazione binaria**, la decisione finale viene ottenuta tramite voto di maggioranza tra le previsioni dei singoli alberi.

2.2.2.2. Pregi della modello Random forest

Nel contesto dei **sistemi di allerta precoce**, il modello Random Forest offre diversi vantaggi applicativi:

- **Resistenza all'overfitting:** la varietà introdotta dal bagging e dalla selezione randomica delle variabili aiuta a limitare il rischio di sovradattamento, soprattutto con dati rumorosi o campioni limitati;
- **Gestione efficace di variabili eterogenee e non linearità:** il Random Forest è particolarmente adatto a elaborare simultaneamente variabili di diversa natura (es. finanziarie, macroeconomiche, qualitative) e a riconoscere schemi complessi, spesso non individuabili con modelli come la **logistic regression**;
- **Valutazione della rilevanza delle variabili (Feature Importance):** il modello produce metriche quantitative, come la riduzione dell'impurità di Gini o il calo di accuratezza media, utili per identificare le variabili più influenti (ad esempio, livello di indebitamento di breve termine o liquidità disponibile).

2.2.2.3 Limiti del modello Random forest

Tuttavia, il modello presenta alcune criticità, che si possono sintetizzare come segue:

- **Scarsa leggibilità:** trattandosi di un modello non parametrico e composto da molti alberi, risulta difficile spiegare in modo trasparente come vengono prese le decisioni, un aspetto cruciale in contesti regolamentari o decisionali;
- **Consumo elevato di risorse computazionali:** l'addestramento simultaneo di numerosi alberi su subset diversi richiede notevole potenza di calcolo, specialmente con grandi volumi di dati o con ottimizzazione complessa degli iperparametri;

- **Rischio residuo di overfitting:** se il dataset contiene molte variabili irrilevanti o altamente rumorose, il modello può ancora adattarsi troppo ai dati di training, compromettendo le performance sui nuovi dati;
- **Variabilità delle metriche di importanza:** in presenza di correlazioni elevate tra variabili esplicative, la stima dell'importanza può risultare distorta, portando a interpretazioni fuorvianti;
- **Mancanza di interpretabilità causale:** il Random Forest identifica associazioni statistiche, ma non permette di dedurre nessi causali tra le variabili, limitandone l'utilizzo in analisi orientate alla policy o alle cause;
- **Necessità di una calibrazione accurata:** l'efficacia del modello dipende fortemente dalla scelta appropriata degli iperparametri (es. numero di alberi, profondità massima, numero di variabili considerate a ogni split) e da un'accurata procedura di validazione incrociata.

Il Random Forest rappresenta dunque uno strumento avanzato per lo sviluppo di sistemi di allerta data-driven, grazie alla sua **elevata precisione predittiva**, alla **gestione efficiente della complessità** e alla **robustezza statistica**. Tuttavia, richiede una configurazione attenta, una gestione oculata delle risorse computazionali e una consapevolezza dei limiti interpretativi. La sua efficacia è massimizzata quando inserito in un framework analitico integrato, che preveda validazione esperta, aggiornamento dinamico dei dati e interazione con fonti informative esterne.

2.2.3 Altri modelli avanzati e approcci ibridi

Accanto ai modelli statistici parametrici, storicamente utilizzati negli Early Warning Systems, negli ultimi anni si è registrato un crescente interesse nell'integrare tali metodologie convenzionali con tecniche non parametriche provenienti dal campo del machine learning.

In questo contesto, alcuni modelli fanno uso di **reti neurali artificiali**, le quali apprendono le **relazioni non lineari** tra le variabili esplicative per individuare soggetti in difficoltà. Altri approcci sfruttano **metodi di pattern recognition**, che consistono nel suddividere lo spazio delle variabili esplicative in base a valori soglia significativi, al fine di classificare le performance aziendali e differenziare tra entità sane e in crisi.

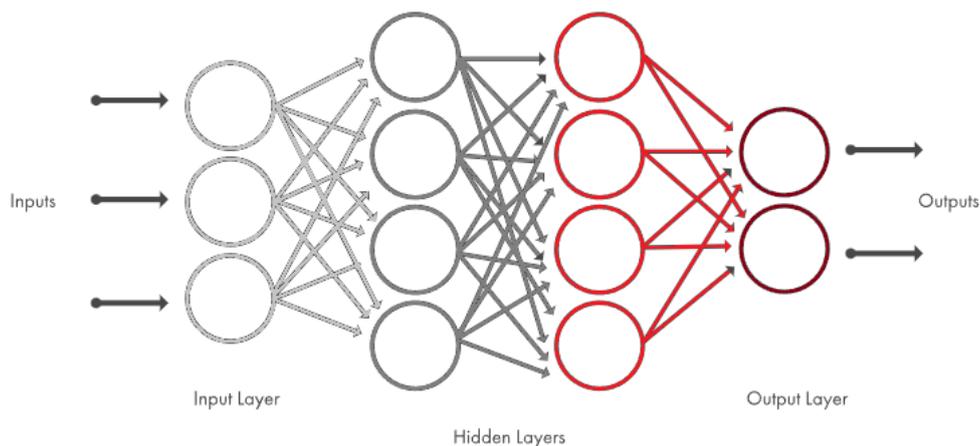


Figura 5: Reti neurali

Tra modelli sempre più avanzati, che vanno oltre le tradizionali tecniche statistiche si citano il **Gradient Boosting** e le reti neurali deep learning che si sono affermati grazie alla loro capacità di catturare relazioni complesse e non lineari nei dati. Altri approcci, come le **Support Vector Machines** o i modelli bayesiani, offrono soluzioni flessibili per gestire dati ad alta dimensionalità o con incertezza. Inoltre, metodi di anomaly detection e clustering permettono di identificare pattern nascosti e segnali anomali, ampliando le potenzialità predittive degli EWS. Spesso, queste tecniche vengono integrate in modelli ibridi, combinando i punti di forza di diverse metodologie per migliorare robustezza e accuratezza predittiva.

In sintesi, sia i modelli *Logit* sia quelli basati su *Random Forest* e altre tecniche più avanzate offrono approcci validi per lo sviluppo di EWS, tuttavia, ognuna presenta limiti che richiedono attenzione in fase di implementazione e una corretta e continua revisione degli input e dei dati inseriti per i calcoli. inoltre, la scelta del modello più adatto dipende anche dalla natura e dalla qualità dei dati disponibili, poiché ciascun metodo si presta meglio a specifiche condizioni.

Nel paragrafo seguente, verranno analizzate alcune leve di miglioramento, con un focus sull'integrazione degli open data come risorsa strategica per potenziare l'efficacia predittiva di questi sistemi.

2.3. Leve di miglioramento e open data

Per potenziare l'affidabilità e la precisione predittiva dei modelli finora descritti, è fondamentale esplorare diverse leve di miglioramento, a partire dalla qualità e dalla ricchezza della base informativa su cui essi si fondano.

Un'evoluzione significativa in questo ambito è rappresentata dall'integrazione degli open data, che ha introdotto cambiamenti rilevanti sia sul piano metodologico che su quello operativo nei sistemi di monitoraggio. Prima di approfondire come essi vengano impiegati nei sistemi di allerta precoce, è utile chiarirne brevemente il significato.

Secondo la definizione della *Open Knowledge Foundation*, gli open data sono dati liberamente accessibili, riutilizzabili e condivisibili da chiunque, per qualsiasi finalità, senza restrizioni rilevanti.¹⁹

I principi fondamentali che caratterizzano i dati aperti sono²⁰:

- **Disponibilità:** devono essere pubblicati online e accessibili senza restrizioni;
- **Accessibilità:** devono essere facilmente reperibili e leggibili da sistemi informatici;
- **Completezza:** devono offrire una rappresentazione esaustiva del fenomeno osservato;
- **Gratuità:** devono essere utilizzabili senza costi;
- **Riutilizzabilità:** devono poter essere impiegati liberamente, anche per scopi commerciali;
- **Interoperabilità:** devono essere compatibili con diversi sistemi, per facilitarne l'integrazione e l'analisi trasversale.

Per sfruttare a pieno il potenziale che gli open data possono offrire è necessario, prima, effettuare una riflessione critica su:

a) Limiti attuali e leve di miglioramento

I limiti delle fonti informative tradizionalmente utilizzate nei sistemi gestionali come la scarsa tempestività dei dati di bilancio, l'integrazione limitata tra sistemi, la presenza di dati mancanti o incompleti, problemi di qualità e accuratezza, mancanza di metadati dettagliati, limitata granularità, restrizioni di accesso per motivi di governance e difficoltà nell'analisi di dati non strutturati, rendono indispensabile integrare gli Early Warning Systems con dati open. Tuttavia, anche queste nuove fonti presentano criticità specifiche. tra cui l'eterogeneità e la frammentazione delle informazioni disponibili: i

¹⁹OKF – Open Knowledge Foundation. (2010). *Open Definition 2.1*.

²⁰OCSE (2021). *Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact*.

dati provengono da fonti diverse, con formati non uniformi, livelli di aggiornamento disomogenei e mancanza di standardizzazione, elementi che rendono difficile sviluppare modelli predittivi affidabili. Inoltre, le informazioni rilevanti per costruire indicatori di rischio sono distribuite su molteplici piattaforme, con differenti gradi di dettaglio. Altri limiti significativi riguardano la dimensione spesso molto ampia dei file dati, che ne rende complessa la gestione e l'elaborazione con strumenti tradizionali, oltre alla frequente assenza di metadati esaustivi che ostacola la corretta interpretazione e integrazione delle informazioni. Per massimizzare il potenziale degli open data negli EWS, è fondamentale migliorare l'interoperabilità dei sistemi informativi, disponendo di infrastrutture capaci di raccogliere, normalizzare e correlare dati da fonti diverse.²¹

Per affrontare queste criticità, le leve di miglioramento si articolano su tre direttrici principali, con particolare riferimento ai dati:

- **Qualità** - è strategico adottare standard comuni per la raccolta, classificazione e pubblicazione dei dati, favorendo l'interoperabilità tra fonti diverse. L'uso di formati aperti (open formats) e vocabolari controllati condivisi può agevolare l'integrazione dei dati nei modelli di previsione. È inoltre necessario rafforzare la governance dei dati pubblici, stabilendo scadenze più stringenti per la pubblicazione e verifiche di qualità;
- **Accessibilità** - si potrebbero potenziare le piattaforme pubbliche di open data, sia a livello nazionale (ad es. dati.gov.it, Istat, Banca d'Italia) che europeo (Eurostat, Open Data Portal dell'UE), ampliando la copertura informativa e semplificando le modalità di accesso²². Sviluppare piattaforme unificate per la raccolta e consultazione dei dati semplificherebbe l'accesso e il riutilizzo. Esempi concreti di fonti rilevanti includono dati dalle Camere di Commercio (bilanci, visure, protesti), centrali dei rischi e piattaforme delle autorità di vigilanza.

Tra le fonti open più rilevanti figurano le seguenti:

²¹European Data Portal. (2020). “*The economic impact of open data: Opportunities for value creation in Europe*”.

²²European Data Portal. (2022). “*Open data maturity report*”.

Fonte	Sito
Open CUP	https://www.opencup.gov.it/portale/web/opencup/accesso-agli-open-data
Open Coesione	OpenCoesione - Scarica gli Open data
Registro delle Startup e PMI Innovative- InfoCamere	Startup e PMI innovative
ISPRA	https://www.isprambiente.gov.it/it/banche-dati
CERVED Open Data	https://www.cerved.com/
ISTAT	https://esploradati.istat.it/databrowser/#/it
ANAC	https://dati.anticorruzione.it
Registro Nazionale Aiuti di Stato	https://www.rna.gov.it/open-data#opendataLink-79

Tabella 1: Fonti Open Data

L'analisi integrata di queste fonti, consente la costruzione di indicatori sintetici utili per l'allerta precoce;

- **Tempestività** - si può incentivare l'impiego di tecnologie di raccolta automatica dei dati (web scraping, API), l'adozione di obblighi normativi più stringenti in termini di reporting periodico e l'uso di dati alternativi o "non convenzionali" (es. social media, news, transazioni digitali) per arricchire gli indicatori di preallerta.²³

b) Benefici attesi:

Il rafforzamento del patrimonio informativo basato su open data e l'adozione di modelli predittivi evoluti comportano numerosi benefici attesi, sia in termini di efficacia operativa sia di efficienza nella gestione del rischio. L'integrazione sistematica di dati pubblici, aggiornati e interoperabili, consente di costruire indicatori più robusti e tempestivi, migliorando sensibilmente la capacità di intercettare in anticipo segnali di squilibrio economico-finanziario. Ciò consente alle imprese, agli enti pubblici e agli intermediari finanziari di attivare strategie correttive e di mitigazione prima che le criticità si aggravino, favorendo una gestione del rischio più proattiva. L'analisi predittiva basata su dataset più ampi, granulari e dinamici permette di superare i limiti degli approcci storici e reattivi, basati esclusivamente su serie storiche o dati consolidati,

²³World Bank. (2021). "Data for better lives: World development".

migliorando la precisione nella valutazione del rischio di credito²⁴. In particolare, si assiste a un incremento della sensibilità dei modelli predittivi nel rilevare segnali deboli (weak signals), che spesso anticipano situazioni di crisi, permettendo di guadagnare tempo prezioso per l'intervento. Nel settore pubblico, la maggiore disponibilità e trasparenza dei dati economico-finanziari può rafforzare la fiducia da parte di investitori, cittadini e organi di controllo, contribuendo a un miglioramento della reputazione degli enti e a una maggiore accountability. Inoltre, un'informazione più diffusa e accessibile agevola l'attività di valutazione ex-ante e monitoraggio ex-post da parte delle autorità competenti. Questi strumenti predittivi, basati su una combinazione di open data, tecniche avanzate di machine learning e indicatori di performance, risultano fondamentali per rafforzare la resilienza complessiva del sistema economico e finanziario, in coerenza con le Linee Guida dell'Autorità Bancaria Europea (EBA) sui Key Risk Indicator (KRI) e con gli standard internazionali in materia di prevenzione delle crisi.²⁵

c) Impatti e rischi:

L'adozione di Early Warning System (EWS) basati su open data, pur offrendo molteplici vantaggi, comporta anche alcune criticità e potenziali rischi che devono essere adeguatamente gestiti. Uno dei principali rischi è rappresentato dall'uso distorto o improprio dei dati, che può condurre a falsi allarmi o a diagnosi errate dello stato di salute di un'impresa o di un ente²⁶. In particolare, l'interpretazione automatica di segnali deboli, se non supportata da un'adeguata contestualizzazione, può generare falsi positivi, innescando azioni correttive non necessarie o allarmismi ingiustificati. Inoltre, l'eccessiva dipendenza da fonti esterne di valutazione, come i rating creditizi, può risultare fuorviante in quanto questi strumenti non sempre tengono conto della natura specifica del settore pubblico o della tempistica di aggiornamento dei dati. La qualità del sistema di allerta dipende quindi fortemente dalla pertinenza, freschezza e rappresentatività dei dati utilizzati. A livello sistemico, si aprono questioni importanti legate alla protezione dei dati personali e alla trasparenza algoritmica. L'impiego di

²⁴Floridi, L., & Cowls, J. (2019). "A Unified Framework of Five Principles for AI in Society". Harvard Data Science Review, 1(1)

²⁵EBA. (2021). "EBA updates list of risk indicators and analysis tools". European Banking Authority.

²⁶Žliobaitė, I. (2017). "Measuring discrimination in algorithmic decision making". Data Mining and Knowledge Discovery, 31(4), 1060–1089

algoritmi predittivi, soprattutto in ambito pubblico, solleva interrogativi etici circa la responsabilità delle decisioni automatizzate e il rispetto dei principi di equità, non discriminazione e accountability. Per mitigare tali rischi, è fondamentale garantire la tracciabilità e l'auditabilità dei modelli e degli algoritmi utilizzati, assicurando che ogni output sia verificabile e interpretabile anche da un punto di vista umano²⁷. A questo scopo, è auspicabile integrare le valutazioni quantitative con analisi qualitative e contestuali, che tengano conto di fattori non sempre misurabili ma rilevanti per una comprensione completa del rischio.

In conclusione, si può affermare che l'integrazione degli open data negli EWS è una leva strategica per anticipare situazioni di crisi. Per tradurre questo potenziale in benefici concreti, è indispensabile intervenire sui limiti attuali potenziando qualità, accessibilità e tempestività dei dati. Solo attraverso un utilizzo consapevole, strutturato e trasparente degli open data si potranno costruire modelli predittivi affidabili, in grado di sostenere una gestione più proattiva e resiliente del rischio.

2.4 L'Intelligenza Artificiale nel contesto dei Sistemi di Allerta Precoce (EWS)

L'Intelligenza Artificiale (IA), come definita dall'AI Act UE 2024, "è un sistema automatizzato con diversi livelli di autonomia, capace di generare output (previsioni, decisioni, raccomandazioni) che influenzano ambienti fisici o digitali"²⁸. Nei Sistemi di Allerta Precoce (EWS), l'IA – in particolare tramite il **machine learning** – consente di superare i limiti delle tradizionali scorecard, grazie alla capacità di apprendere dai dati, autocalibrarsi e riconoscere pattern complessi. I modelli ML analizzano crisi passate, definiscono soglie di alert e rilevano vulnerabilità latenti, rendendo gli EWS più accurati e proattivi.

Alla base di questa capacità vi sono diverse modalità di apprendimento, che definiscono come i modelli si allenano e migliorano nel tempo:

- **Supervisionato:** apprendimento da dati etichettati, utile per classificazioni e previsioni.

²⁷ Floridi, L., Cowls, J., Beltrametti, ... Vayena, E. (2018). "AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations". *Minds and Machines*, 28(4), 689–707.

²⁸ Gazzetta ufficiale dell'Unione europea. "Regolamento (ue) 2024/1689 del parlamento europeo e del consiglio". Articolo 3, Pag. 46. 13 giugno 2024

- **Non supervisionato:** individuazione autonoma di pattern o strutture latenti nei dati (es. clustering).
- **Per rinforzo:** apprendimento tramite interazione con l'ambiente e feedback sulle azioni svolte.

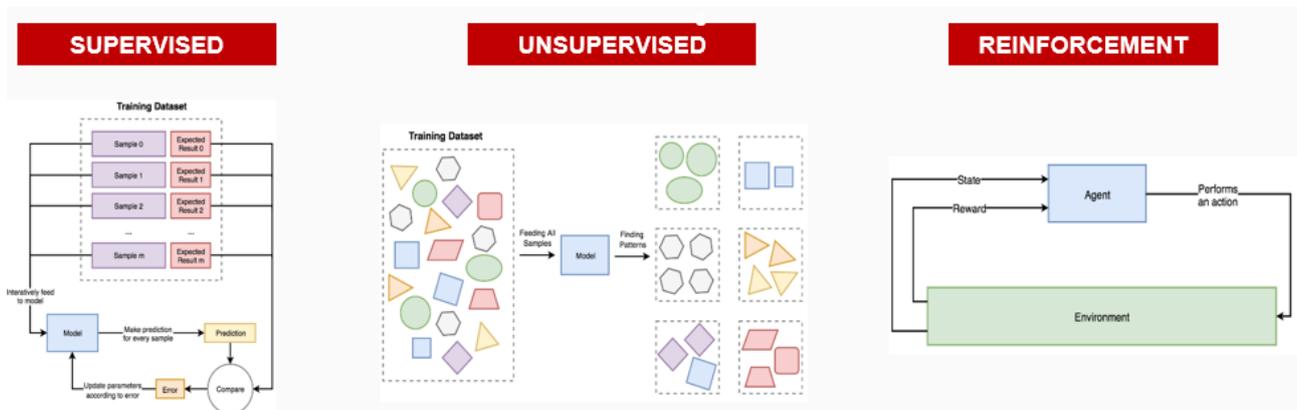


Figura 6: Modalità di apprendimento dei modelli

Oltre al ruolo analitico, l'intelligenza artificiale offre un contributo operativo concreto lungo tutto il ciclo di vita di un sistema di allerta precoce. Le tecnologie AI-based permettono di automatizzare attività complesse, velocizzare lo sviluppo e aumentare l'efficienza dei processi decisionali.²⁹

In particolare:

- **Gestione dei modelli predittivi:** l'AI facilita l'implementazione e la manutenzione di modelli avanzati (es. reti neurali, ensemble) grazie a strumenti come AutoML, che automatizzano selezione delle variabili, tuning degli iperparametri e validazione.³⁰
- **Architettura dei sistemi EWS AI-driven:** sistemi basati su pipeline intelligenti e framework modulari (come LEGO: Leverage, External Indicators, Governance, Ontology) permettono lo streaming in tempo reale e l'orchestrazione automatica dei modelli, rendendo l'EWS più reattivo e scalabile.³¹
- **Generazione automatica di codice e prototipazione:** un ambito in forte espansione è l'utilizzo dell'intelligenza artificiale per la generazione automatica del codice e la prototipazione veloce di componenti software. Questo approccio sta trasformando radicalmente il modo in cui vengono sviluppati i sistemi predittivi, inclusi gli EWS,

²⁹ European Banking Authority (EBA). Report on Big Data and Advanced Analytics. 2020

³⁰ Organisation for Economic Co-operation and Development. (2021). *The impact of artificial intelligence on the financial sector*.

³¹ PwC & Galytix. (s.d.). *Banks must act on their Early Warning Systems or risk ROE downturn*.

consentendo di passare più rapidamente dalla fase di design concettuale alla realizzazione operativa³². L'intelligenza artificiale generativa, in particolare i *Large Language Models (LLM)* come GPT, Codex o PaLM, è in grado di comprendere istruzioni in linguaggio naturale e tradurle in codice funzionante. Le principali applicazioni riguardano il **preprocessing** dei dati, l'**addestramento** dei modelli, la creazione automatica di **report** e dashboard, e la **prototipazione** veloce di logiche di alert. Questo approccio accelera il ciclo di sviluppo, riduce la dipendenza da competenze tecniche avanzate e promuove la standardizzazione delle best practice³³. Oltre alla scrittura, l'AI è in grado di interpretare, spiegare e ottimizzare codice esistente, migliorando **la trasparenza e l'affidabilità** dei sistemi³⁴. Questa capacità è particolarmente utile in contesti regolamentati, dove la leggibilità e l'auditabilità del codice sono essenziali. Nel contesto EWS, la generazione automatica di codice consente di realizzare prototipi agili, testare rapidamente nuove logiche di rischio e adattare i sistemi in risposta a mutate condizioni operative o normative.

- **Data analysis e interpretazione intelligente:** l'AI abilita analisi avanzate su dataset complessi e multimodali, combinando NLP, analisi semantica, clustering e anomaly detection. Questi strumenti sintetizzano insight da fonti eterogenee – come testi, transazioni o social network – individuando pattern deboli o emergenti non rilevabili con metodi statistici tradizionali.³⁵

In sintesi, l'adozione dell'Intelligenza Artificiale e del Machine Learning segna un'evoluzione chiave per i Sistemi di Allerta Precoce, superando i limiti dei modelli tradizionali grazie a una maggiore capacità predittiva e all'analisi di dati complessi.

Il capitolo seguente illustrerà un caso pratico sull'uso di modelli EWS basati su AI, analizzando l'impatto dell'integrazione di dati esterni sulla capacità predittiva rispetto all'impiego dei soli dati interni.

³² Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. Stanford Center for Research on Foundation Models.

³³ Bavishi, R., Jain, A., & Raychev, V. (2023). Code generation with language models: A survey. arXiv.

³⁴ Zhou, S., & Shen, S. (2023). Empirical evaluation of large language models for code generation and comprehension. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023).

³⁵ AIFIRM. "intelligenza artificiale: l'applicazione di machine learning e predictive analytics nel risk management". Febbraio 2019

3. Costruzione di un modello predittivo dei default

Questo capitolo presenta un modello predittivo del default, sviluppato a partire da dati interni dell'istituto di credito (anagrafica, rapporti, garanzie, contabilità) e successivamente arricchito con open data provenienti da fonti istituzionali come ANAC, ISPRA e il Registro Nazionale Aiuti di Stato. Attraverso tecniche statistiche e machine learning, il modello stima la probabilità di insolvenza dei debitori, supportando la prevenzione del rischio di credito e la gestione proattiva del portafoglio.

3.1 Analisi e arricchimento del patrimonio informativo mediante Open Data

L'efficacia di un modello predittivo dei default è strettamente legata alla qualità, profondità e varietà del patrimonio informativo su cui esso si basa. In particolare, la sola disponibilità di dati contabili e gestionali, pur rappresentando una solida base di partenza, non è più sufficiente a garantire livelli di accuratezza predittiva adeguati per i limiti mostrati nei capitoli precedenti. È in questo scenario che l'utilizzo strategico degli open data emerge come una risorsa chiave per arricchire il dataset di input del modello, ampliando lo spettro delle variabili esplicative e consentendo un'analisi più olistica e tempestiva del rischio.

In una prima fase sono state effettuate delle analisi volte a completare il set informativo utile al successivo confronto delle performance di modello.

Le informazioni a disposizione necessarie per il primo step di analisi sono le seguenti:

a) *Controparti*

- Numero: il dataset è composto da 7378 debitori a seguito di un'attività di assessment del patrimonio informativo
- Tipologia: il dataset si compone prevalentemente soggetti di natura corporate
- % Default: la percentuale di controparti in default all'interno del dataset è di circa il 9%

b) *Informazioni anagrafiche e strutturali*

- Area geografica: le controparti sono distribuite su tutto il territorio nazionale come segue:
 - Nord-Ovest - 30%
 - Nord-Est - 34%

- Centro - 23%
- Sud - 11%
- Isole - 3%
- Dimensione della controparte (size): la base dati è composta da controparti corporate, classificate in base alla loro dimensione misurata attraverso il totale dell'attivo (total asset).
 - Piccole imprese - 86,38%
 - Medie imprese - 10,54%
 - Grandi imprese - 3,04%

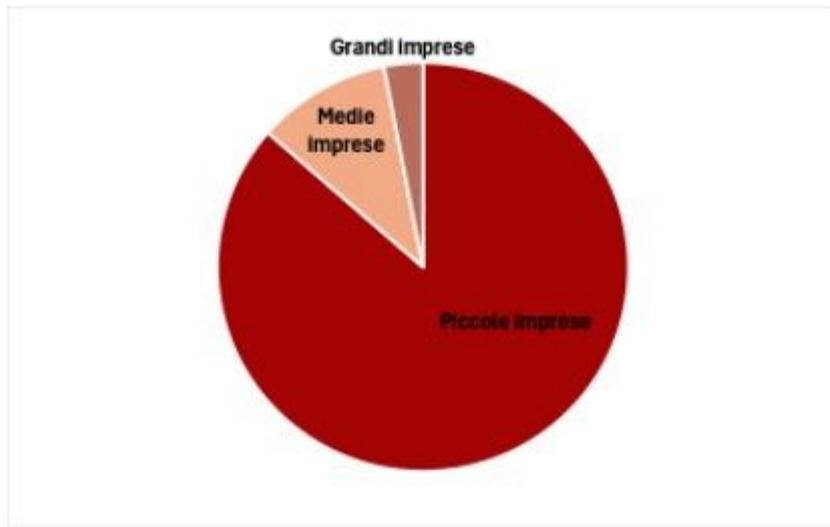


Figura 7: Dimensione della controparte

c) Informazioni di bilancio

- Patrimonio netto: misura la ricchezza propria dell'azienda, cioè la differenza tra attivo e passivo
- Totale attivo (asset): rappresenta l'insieme dei beni e diritti posseduti dall'impresa
- Ricavi: indicano il volume di affari e la capacità dell'impresa di generare reddito
- EBITDA: misura la redditività operativa, escludendo gli effetti di scelte finanziarie, fiscali e contabili, ed è un indicatore chiave per valutare la capacità di generare cassa

d) Informazioni sui rapporti creditizi

Descrivono le relazioni finanziarie tra l'ente erogatore del credito e la controparte, fondamentali per valutare il rischio associato:

- Tipologia di rapporto: informazione circa le forme tecniche es. mutui, c/c, leasing, factoring ecc
- Debito residuo: ammontare del debito ancora da rimborsare che si attesta a circa 4,17 mld
- Gross Book Value (GBV): valore nominale del debito, al netto di eventuali svalutazioni che ammonta a circa 6,3 mld
- Andamento storico: informazioni sulla regolarità nei pagamenti, eventuali ritardi o ristrutturazioni del debito

A partire dai dati di input, sono state effettuate diverse elaborazioni per aggregare le informazioni a livello di singolo debitore, ricodificare variabili con domini specifici, stratificare i dati in modo da ottenere una rappresentazione strutturata del dataset e calcolare valori fondamentali, come il totale degli interessi, necessari per la costruzione degli indicatori.

Una fase fondamentale del lavoro ha riguardato l'arricchimento del dataset originale attraverso l'integrazione di fonti pubbliche disponibili. Un'accurata attività di scouting ha inizialmente individuato circa 20 potenziali fonti Open Data in grado di ampliare il patrimonio informativo iniziale. Tuttavia, a seguito di un'attenta fase di selezione e rielaborazione, sono state effettivamente integrate nel dataset informazioni provenienti dalle seguenti fonti:

- **ANAC – Autorità Nazionale Anticorruzione**: utilizzata per estrarre dati su appalti pubblici, aggiudicazioni, e operatori economici. L'integrazione di questi dati ha permesso di rilevare comportamenti anomali o ricorrenti nei processi di affidamento e di monitorare il grado di esposizione al rischio corruzione delle aziende coinvolte
- **Open CUP**: fonte centrale per ottenere informazioni su progetti di investimento pubblico, codici unici di progetto (CUP), bandi e commesse. L'enrichment di questi dati ha consentito di migliorare la tracciabilità dei progetti nel tempo, mettendo in relazione la pianificazione e l'effettiva attuazione sul territorio
- **ISPRA – Idrogeologico**: ha fornito dati geospaziali su frane, alluvioni e altre situazioni di rischio idrogeologico. Questa componente ha rafforzato la dimensione territoriale del sistema EWS, rendendo possibile l'integrazione del rischio ambientale nel modello di analisi
- **Open Coesione**: ha messo a disposizione dati su progetti di natura ESG finanziati da fondi europei e nazionali, comprensivi di informazioni su beneficiari, importi, avanzamento e impatto. Questo ha reso possibile un'analisi territoriale del

finanziamento pubblico, utile per comprendere la distribuzione e la concentrazione degli investimenti.

Tutti i dati provenienti da queste fonti sono stati pre-processati, geocodificati ove necessario, e importati nel database centrale del sistema EWS. L'arricchimento ha permesso non solo un incremento del dettaglio informativo, ma anche la creazione di nuove feature utili all'addestramento del modello predittivo, con una visione più integrata tra ambiti tematici diversi (ambiente, appalti, finanza pubblica e coesione territoriale).

Completato il patrimonio informativo, sono stati calcolati numerosi indicatori, sia di natura contabile sia derivanti dall'integrazione di open data (si veda l'appendice per un elenco dettagliato). Sarà, tuttavia, il modello di machine learning a selezionare in modo autonomo le variabili più rilevanti, sulla base della loro significatività, come verrà illustrato nei paragrafi successivi.

3.2 Preparazione delle feature e applicazione del modello

Dopo aver completato e rielaborato le informazioni contenute nel dataset iniziale, è stato possibile avviare la seconda fase dell'analisi, finalizzata alla costruzione di un modello predittivo. Questa fase ha previsto l'implementazione di una regressione logistica (logit), un approccio ampiamente utilizzato nella letteratura economico-finanziaria per la stima della probabilità di eventi binari, come il rischio di default aziendale.

La scelta di adottare un modello logit è stata guidata dalla sua adattabilità ai dati disponibili. Infatti, i modelli logit tengono conto del tempo e permettono di integrare variabili aziendali, settoriali e macroeconomiche, seguendo contributi rilevanti che evidenziano l'importanza degli effetti settoriali e della congiuntura economica nel determinare il default, rispetto a modelli strutturali che necessitano del valore di mercato degli attivi e delle passività di un'impresa non sempre disponibili per PMI.

Il modello logit è stato sviluppato attraverso un codice Python generato e ottimizzato con il supporto dell'intelligenza artificiale. In particolare, l'IA è stata impiegata non solo per la stesura iniziale del codice, ma anche per guidarne l'evoluzione attraverso l'introduzione di metodologie statistiche robuste, suggerimenti per la selezione delle variabili, accorgimenti per migliorare la leggibilità e l'efficacia della visualizzazione dei risultati, e infine per garantire la riproducibilità dell'intero processo analitico.

Il codice ha previsto inizialmente l'importazione delle principali librerie statistiche e di machine learning, tra cui *statsmodels*, *sklearn* e *shap*, necessarie per la regressione logistica e la standardizzazione delle variabili. Successivamente, il dataset è stato pre-processato per rimuovere eventuali valori mancanti, ed è stata definita la variabile dipendente (DEFAULT) da prevedere.

La costruzione dei modelli si è basata su due insiemi di variabili indipendenti. Il primo includeva esclusivamente indicatori di bilancio tradizionali D/E, Equity Ratio, ROA, Interest Coverage Ratio, Forbearance, Debt Yield, ROE operativo. Il secondo modello ha invece esteso la base informativa, combinando gli indicatori di bilancio con una selezione di variabili provenienti da fonti open data. Tra queste si annoverano: la presenza di aiuti di stato, l'aggiudicazione di gare d'appalto pubbliche, l'adesione a progetti ESG, e due proxy di rischio ambientale (idrogeologico e idrico).

L'intelligenza artificiale ha fornito un contributo significativo anche nella gestione della multicollinearità, attraverso l'integrazione automatizzata dell'analisi del Variance Inflation Factor (VIF), e nella selezione automatica delle variabili statisticamente significative, rimuovendo progressivamente quelle con p-value superiori a 0,05. Inoltre, il codice ha incluso una valutazione comparata delle performance dei due modelli attraverso metriche diagnostiche standard — tra cui il coefficiente pseudo R^2 di McFadden, l'area sotto la curva ROC (AUC), la matrice di confusione, la sensibilità (recall), la specificità e la precisione predittiva (PPV).

Un altro aspetto rilevante introdotto con il supporto dell'IA riguarda l'analisi dell'importanza delle variabili predittive, effettuata attraverso la metodologia SHAP (SHapley Additive exPlanations), che consente di interpretare in maniera trasparente il contributo di ciascuna variabile alle previsioni del modello. Tale approccio, combinato con la rappresentazione grafica degli odds ratio e delle curve ROC, ha permesso una lettura approfondita dei risultati e ha rafforzato la validità interpretativa del modello costruito.

Il codice generato è stato runnato attraverso Deepnote, un ambiente di sviluppo open source per data science e machine learning, basato su notebook.

In sintesi, l'impiego dell'intelligenza artificiale ha permesso non solo di automatizzare le fasi più critiche del processo modellistico, ma anche di rafforzarne il rigore metodologico e la replicabilità scientifica, rendendo l'intero workflow efficiente, trasparente e facilmente adattabile a contesti applicativi differenti.

3.3 Risultati ed interpretazioni

Conclusa la fase di sviluppo ed esecuzione del modello, si è passati all'analisi e all'interpretazione dei risultati. L'indagine empirica ha messo in evidenza il valore aggiunto derivante dall'integrazione di Open Data e variabili ESG nei modelli predittivi del rischio di default aziendale.

Il confronto tra due specifiche modellistiche – una basata unicamente su variabili finanziarie tradizionali e l'altra arricchita con dati esterni e pubblicamente disponibili – ha evidenziato un miglioramento tangibile delle performance predittive.

In particolare, il McFadden R^2 è passato dal 31% al 34% (+3 punti percentuali), mentre la precision, ovvero la quota di veri positivi rispetto al totale dei casi classificati come positivi (TP / (TP + FP)), è aumentata dal 69% al 71% (+2 punti percentuali), segnalando un incremento significativo sia nella capacità esplicativa sia nella potenza discriminatoria del secondo modello.

Questo progresso è confermato anche dalla matrice di confusione, dove si registra una lieve diminuzione dei casi di default non correttamente classificati, determinando dunque un miglioramento della precision,

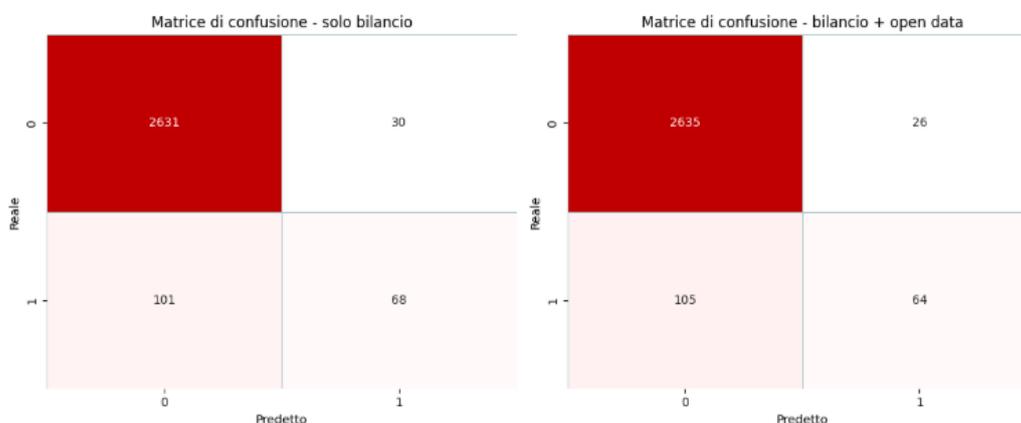


Figura 8: Matrici di confusione a confronto

L'AUC rappresenta l'area sotto la curva ROC e fornisce una misura sintetica dell'abilità del modello nel distinguere tra le classi. Nel caso analizzato, il modello arricchito con Open Data mostra un miglioramento nella capacità discriminatoria, con una probabilità elevata – pari all'89% – di identificare correttamente un'impresa in default rispetto a una non in default.

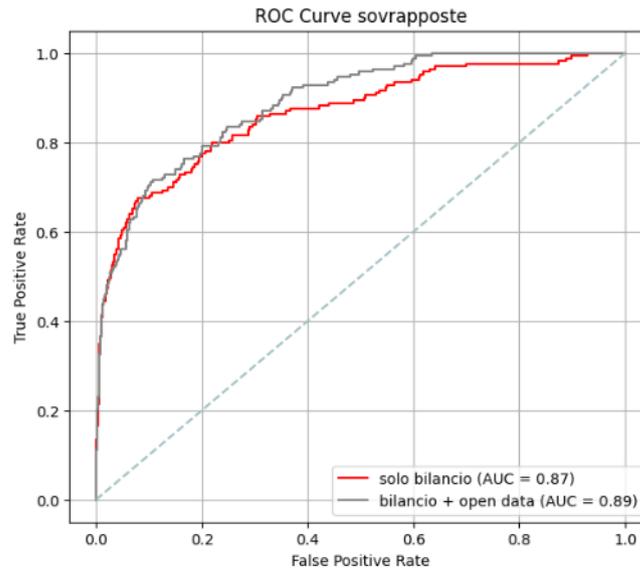


Figura 9: AUC-ROC curve a confronto

Dal punto di vista metodologico, questi risultati suggeriscono che le variabili Open, come la ricezione di aiuti di stato, la partecipazione a gare pubbliche, l'esposizione a rischi ambientali e l'impegno in progetti ESG, riescono a intercettare informazioni strutturali e di contesto che non sono normalmente catturate dai bilanci. Tali fattori, spesso legati alla resilienza territoriale, alla sostenibilità delle strategie aziendali e alla qualità della governance, possono influenzare in modo sostanziale la probabilità di insolvenza, soprattutto in scenari di crisi o instabilità. L'analisi di interpretabilità basata sugli indici SHAP ha confermato la rilevanza delle nuove variabili derivate da Open Data. Uno SHAP value rappresenta il contributo individuale di una feature alla differenza tra la previsione del modello e il valore medio delle previsioni (baseline). Deriva dal concetto dei valori di Shapley della teoria dei giochi.

Alcune di queste variabili – in particolare gli aiuti pubblici, i progetti ESG e il rischio idrogeologico – si sono posizionate tra le feature più importanti del modello, a fianco di indicatori finanziari tradizionali come il Debt Yield, Interest coverage ratio, Equity ratio e le misure di forbearance. Questa evidenza rafforza la tesi secondo cui gli Open Data non rappresentano solo un'integrazione informativa, ma possono costituire vere e proprie variabili core nei sistemi di scoring e nella valutazione del rischio.

Nel dettaglio, la variabile "aiuti di stato" è particolarmente rilevante perché riflette l'esito di un'istruttoria pubblica: per ricevere l'aiuto, l'impresa viene valutata da un ente che ne accerta lo stato economico-finanziario. La concessione stessa definisce un valore segnaletico importante, già validato da un processo esterno al modello. Questo la rende altamente predittiva rispetto al rischio di default.

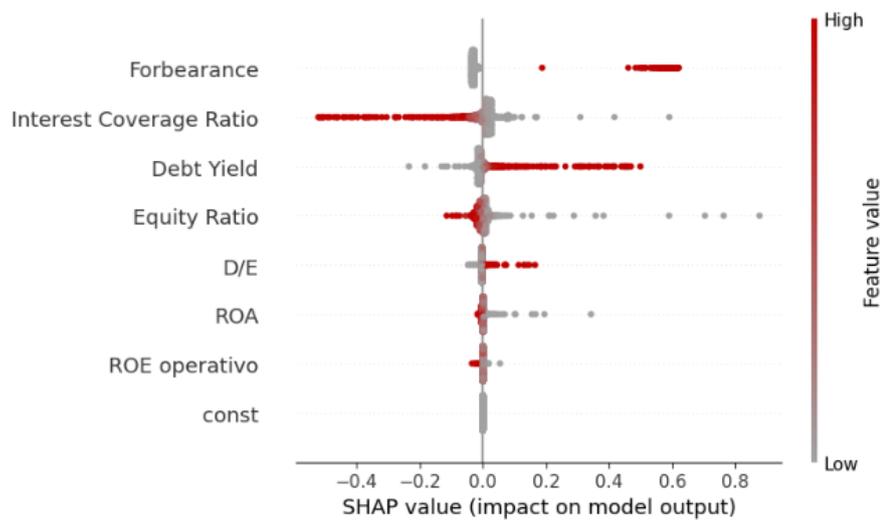


Figura 10: Shap plot – modello w/o Open Data

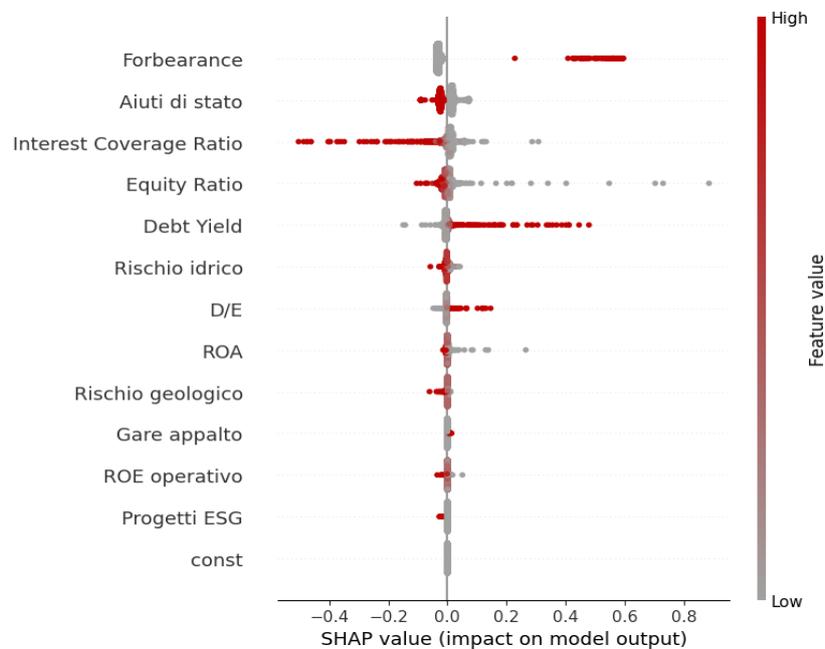


Figura 11: Shap plot – modello con Open Data

In conclusione, l'integrazione degli Open Data nei modelli predittivi rafforza l'efficacia classificatoria degli algoritmi e offre un vantaggio strategico nella valutazione del rischio. L'utilizzo sistematico di dati aperti emerge come una prospettiva promettente per l'evoluzione del credit scoring, del risk assessment e delle decisioni di policy.

Conclusioni

Alla luce del rafforzamento normativo promosso da BCE ed EBA in materia di rischio di credito e del crescente interesse verso strumenti di early warning, questo lavoro ha esplorato l'integrazione di open data nei modelli predittivi del rischio di default, per arricchire la comprensione dei fenomeni di deterioramento creditizio e migliorare la capacità discriminatoria dei modelli.

L'analisi condotta ha evidenziato con chiarezza il valore aggiunto che l'integrazione di fonti informative eterogenee può apportare alla modellazione del rischio di default aziendale. Attraverso un approccio strutturato basato sulla regressione logistica e sviluppato con il supporto dell'intelligenza artificiale, è stato possibile confrontare le performance predittive di due modelli: uno basato esclusivamente su indicatori di bilancio, e un secondo che ha integrato anche variabili provenienti da open data.

Nel primo caso, utilizzando unicamente dati di natura economico-finanziaria aziendale, il modello ha raggiunto un valore di R^2 di McFadden pari a 0,31, indicando un livello di spiegazione accettabile ma migliorabile del fenomeno di default. Tuttavia, l'inclusione di dati open — quali la ricezione di aiuti di stato, la partecipazione a gare d'appalto, la presenza di progetti ESG, e l'esposizione a rischi ambientali — ha permesso di catturare nuove dimensioni del rischio, legate alla trasparenza, alla sostenibilità e al contesto operativo delle imprese.

Il secondo modello, che combina queste nuove informazioni con gli indicatori di bilancio tradizionali, ha mostrato un miglioramento sostanziale in termini di capacità esplicativa e predittiva. Il valore del R^2 di McFadden è aumentato a 0,34, segnando un incremento significativo nella qualità del modello. Contestualmente, i casi di default correttamente individuati sono pari a 64, confermando l'efficacia dell'approccio integrato nel migliorare la precisione predittiva.

Questi risultati sottolineano l'importanza di superare una visione puramente finanziaria dell'analisi del rischio di credito, favorendo un'integrazione più ampia e multidimensionale delle fonti informative. L'apporto dell'intelligenza artificiale si è rivelato strategico non solo per la generazione automatizzata e ottimizzata del codice di analisi, ma anche per la strutturazione del workflow, la selezione delle variabili rilevanti e l'interpretabilità dei risultati attraverso tecniche avanzate come SHAP.

In conclusione, l'esperienza dimostra che l'adozione di strumenti data-driven e l'apertura verso fonti non convenzionali rappresentano una leva fondamentale per migliorare la capacità di

previsione dei modelli di rischio. Tale approccio risulta particolarmente promettente in un contesto in cui la disponibilità di dati aperti e trasparenti è in continua crescita, offrendo nuove opportunità per l'analisi creditizia e la valutazione della solidità delle imprese.

Bibliografia

ABI (2023). *Outlook Abi-Cerved sui crediti deteriorati delle imprese.*

AIFIRM (2019). *Intelligenza artificiale: l'applicazione di machine learning e predictive analytics nel risk management.*

Banca Centrale Europea (2017). *Linee guida per le banche sui crediti deteriorati.*

Bavishi, R., Jain, A., & Raychev, V. (2023). *Code generation with language models: A survey.*

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., & Liang, P. (2021). *On the opportunities and risks of foundation models.* Stanford Center for Research on Foundation Models.

Cerved (2024). *Outlook ABI-Cerved 2024-26. Crediti deteriorati delle imprese: tassi in aumento contenuto.*

Commissione Europea (2024). *Regolamento (UE) 2024/1623 del Parlamento europeo e del Consiglio del 31 maggio 2024 che modifica il regolamento (UE) n. 575/2013 per quanto riguarda i requisiti prudenziali per gli enti creditizi e le imprese di investimento.* Gazzetta ufficiale dell'Unione europea.

Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili (2019). *Crisi d'impresa. Gli indici dell'allerta.*

CRIF (2024). *Il tasso di default delle imprese sale al 2,39% nel 2023, aumenta anche il credito erogato.*

European Banking Authority (2016). *Guidelines on the application of the definition of default under Article 178 of Regulation (EU) No 575/2013 (EBA/GL/2016/07).*

European Banking Authority (2017). *Orientamenti sull'applicazione della definizione di default ai sensi dell'articolo 178 del Regolamento (UE) n. 575/2013.*

European Banking Authority (2020). *Guidelines on loan origination and monitoring.*

European Banking Authority (2020). *Report on Big Data and Advanced Analytics.*

European Banking Authority (2021). *EBA updates list of risk indicators and analysis tools.*

European Banking Authority (2025). *Guidelines on the management of ESG risks.*

- European Banking Authority (2025). *Predicting bank distress in Europe: Using machine learning and a novel definition of distress*. *EBA Staff Paper Series*, n. 21.
- European Data Portal (2020). *The economic impact of open data: Opportunities for value creation in Europe*.
- European Data Portal (2022). *Open data maturity report*.
- Floridi, L., & Cowls, J. (2019). *A Unified Framework of Five Principles for AI in Society*. *Harvard Data Science Review*, 1(1).
- Gazzetta ufficiale dell'Unione europea (2024). *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio. Articolo 3, Pag. 46. 13 giugno 2024*.
- Gazzetta Ufficiale della Repubblica Italiana (2019). *Codice della crisi d'impresa e dell'insolvenza in attuazione della legge 19 ottobre 2017, n. 155*.
- IFRS Foundation (2024). *International Financial Reporting Standard 9: Financial Instruments*.
- OCSE (2021). *Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact*.
- Open Knowledge Foundation (2010). *Open Definition 2.1*.
- Organisation for Economic Co-operation and Development (2021). *The impact of artificial intelligence on the financial sector*.
- CRIF (2024). *Elaborazioni CRIF Ratings*.
- PwC & Galytix (s.d.). *Banks must act on their Early Warning Systems or risk ROE downturn*.
- World Bank (2021). *Data for better lives: World development*.
- Žliobaitė, I. (2017). *Measuring discrimination in algorithmic decision making*. *Data Mining and Knowledge Discovery*, 31(4), 1060–1089.
- Zhou, S., & Shen, S. (2023). *Empirical evaluation of large language models for code generation and comprehension*. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*.

Sitografia

- ABI-Cerved, *Outlook ABI-Cerved 2024-26 sui crediti deteriorati delle imprese*, disponibile su: <https://www.abi.it/abi-cerved-picco-crediti-deteriorati-al-35-in-2024-poi-rientro-graduale-in-25-e-26/>, consultato il 13 maggio 2025.
- CRIF, *Credito alle imprese: in lieve crescita gli importi erogati nei primi 9 mesi del 2024 (+2,4%), ma i tassi di default potrebbero crescere*, disponibile su: <https://www.crif.it/area->

[stampa/credito-alle-imprese-in-lieve-crescita-gli-importi-erogati/](#), consultato il 13 maggio 2025.

CRIF, *Il tasso di default delle imprese sale al 2,39% nel 2023, aumenta anche il credito erogato*, disponibile su: <https://www.crif.it/area-stampa/il-tasso-di-default-delle-imprese-sale-al-2-39-nel-2023-aumenta-anche-il-credito-erogato/#:~:text=il%20credito%20erogato-.il%20tasso%20di%20default%20delle%20imprese%20sale%20al%202%2C39,aumenta%20anche%20il%20credito%20ero>, consultato il 13 maggio 2025.

Cerved, *Outlook ABI-Cerved 2024-26. Crediti deteriorati delle imprese: tassi in aumento contenuto*, disponibile su: <https://www.cerved.com/ricerca-e-analisi/a/ricerche-analisi-credito/outlook-abi-cerved-2024-2026-crediti-deteriorati-imprese>, consultato il 13 maggio 2025.