LUISS

Corso di Laurea Magistrale in Strategic Management
Cattedra Metodi Quantitativi in Strategic Management

# Credit Scoring e Machine Learning: il contributo dei Dati Alternativi in un'Analisi Empirica


Prof. Marco Pirra
RELATORE

Prof. Francesco Morelli
CORRELATORE

Eleonora Cocco 775581
CANDIDATA


Anno Accademico: 2024/2025

# Abstract

Il presente elaborato indaga il ruolo dei dati alternativi e delle tecniche di machine learning nella valutazione del merito creditizio, con l'obiettivo di verificare se, e in che misura possano migliorare l'accuratezza predittiva e favorire una maggiore inclusione finanziaria. Dopo aver ricostruito il quadro normativo europeo e italiano e passato in rassegna le principali metodologie per il credit scoring, la parte empirica utilizza un dataset integrato che combina variabili tradizionali per la valutazione del merito creditizio, con informazioni comportamentali e transazionali, testato attraverso diversi algoritmi di machine learning.

I risultati dell'analisi mostrano come i dati tradizionali restano la fonte più solida e predittiva, mentre l'apporto dei dati alternativi si è rivelato limitato e talvolta non significativo. Gli algoritmi complessi, come Random Forest e XGBoost, hanno raggiunto performance migliori rispetto ai modelli lineari, ma senza ottenere reali benefici dall'integrazione delle nuove variabili.

La conclusione è quindi prudente: il potenziale dei dati alternativi non è escluso in linea di principio, ma, alla luce delle evidenze raccolte, dipende in modo cruciale da qualità, certificazione, granularità e tempestività delle fonti. Poiché il dataset impiegato deriva da basi pubbliche non ufficiali, i risultati vanno letti con cautela. Le ricerche future dovranno verificare l'efficacia dei dati alternativi quando provengano da fonti certificate e ad alta risoluzione e valutarne l'impatto su accuratezza, inclusione e *fairness* in contesti reali.

# Indice

# Introduzione

Nell'ultimo decennio il settore del credito al consumo ha subito una profonda trasformazione, determinata dall'innovazione tecnologica e dal progressivo ampliamento delle fonti informative a disposizione degli operatori finanziari. Il tradizionale approccio alla valutazione del merito creditizio, basato su informazioni anagrafiche, storici di pagamento e variabili economico-finanziarie, è stato progressivamente affiancato dall'impiego di algoritmi di machine learning e dall'integrazione di dati alternativi: tracce digitali, abitudini di consumo, informazioni comportamentali e sociali.

Questa evoluzione si inserisce in un contesto normativo e sociale complesso. Da un lato, l'Unione Europea ha rafforzato i presidi di tutela dei consumatori attraverso strumenti come il GDPR e l'AI Act, imponendo criteri di trasparenza, equità e supervisione umana nei processi automatizzati. Dall'altro lato, permane l'esigenza di garantire l'inclusione finanziaria di soggetti spesso esclusi dai circuiti tradizionali, i cosiddetti *credit invisible*, che non dispongono di una storia creditizia sufficiente a essere valutati dai sistemi standard.

In questo quadro, si colloca la domanda di ricerca che ha guidato il presente lavoro: *l'integrazione di dati alternativi nei modelli di credit scoring, affiancati a tecniche avanzate di machine learning, è in grado di migliorare la capacità predittiva e, al tempo stesso, di favorire una maggiore inclusione finanziaria, senza introdurre nuove forme di bias e discriminazione?*

Rispondere a questa domanda significa affrontare un nodo cruciale dell'innovazione finanziaria contemporanea: comprendere se i dati non convenzionali possano rappresentare un reale strumento di equità, capace di ampliare l'accesso al credito a individui meritevoli ma tradizionalmente esclusi, oppure se, al contrario, essi costituiscano una fonte di rumore informativo, opacità e nuovi rischi per la tutela dei consumatori.

Il presente lavoro si articola lungo due direttrici principali. La prima, di natura teorico-normativa, si sviluppa nei primi tre capitoli: il Capitolo I ricostruisce il quadro regolatorio europeo e italiano sul merito creditizio e sull'intelligenza artificiale, con attenzione alla CCD I e II, al TUB, al GDPR e al recente AI Act. Il Capitolo II analizza le principali metodologie di credit scoring, dai modelli statistici tradizionali alle tecniche di machine

learning ed *ensemble*, con cenno agli strumenti di Explainable AI. Il Capitolo III approfondisce il ruolo dei Big Data e dei dati alternativi, evidenziandone opportunità e rischi, con particolare riferimento ai *bias* algoritmici e alle implicazioni per l'inclusione finanziaria.

La seconda direttrice, di carattere empirico, si concentra nel Capitolo IV e ha l'obiettivo di testare la domanda di ricerca alla base del lavoro. Il capitolo definisce il disegno sperimentale, applica un confronto strutturato tra diverse configurazioni informative e valuta i risultati secondo metriche di accuratezza ed equità, restituendo un quadro complessivo volto a trarre conclusioni utili.

# Capitolo I

## Quadro normativo italiano e europeo sul merito creditizio e sull'IA

Nel corso degli ultimi decenni, la nozione di credito al consumo ha subito un'evoluzione profonda, sia dal punto di vista economico-sociale che giuridico. Negli anni '60 e '70, l'economia europea era ancora fortemente caratterizzata da una "*cash society*", in cui le transazioni avvenivano prevalentemente in contanti e il ricorso al credito da parte delle famiglie era limitato a forme specifiche come la vendita a rate o il prestito personale. In quel contesto, il credito al consumo era uno strumento circoscritto, con un impatto marginale sul sistema economico e normativo.

Oggi, al contrario, il credito è diventato un elemento strutturale della vita economica e un vero e proprio motore di crescita. Secondo stime consolidate, tra il 50% e il 65% dei consumatori europei dispone di almeno un prodotto di credito al consumo, utilizzato per finanziare beni durevoli, servizi o spese personali. Circa il 30% usufruisce inoltre di agevolazioni di sconfino sul conto corrente, una pratica pressoché assente nei decenni passati. A livello macroeconomico, l'ammontare complessivo del credito al consumo nei paesi membri ha superato i 500 miliardi di euro, rappresentando oltre il 7% del PIL, con un tasso di crescita annuo intorno al 7%.

Sebbene il credito al consumo contribuisca al benessere delle famiglie e alla dinamica dei consumi, esso comporta anche rischi non trascurabili: da un lato per i finanziatori, che devono gestire l'esposizione all'insolvenza; dall'altro per i consumatori, che possono incorrere in situazioni di sovraindebitamento. Tali criticità hanno progressivamente evidenziato l'esigenza di rafforzare la regolazione del settore, per garantire un uso responsabile del credito e assicurare una tutela effettiva e uniforme dei consumatori all'interno del mercato unico.

Le differenze esistenti tra gli ordinamenti nazionali, così come tra le prassi bancarie e finanziarie, avevano storicamente prodotto forme di tutela disomogenee e opportunità di accesso al credito fortemente variabili tra gli Stati membri. Questa frammentazione normativa aveva alimentato distorsioni concorrenziali e limitato la possibilità per i consumatori di accedere a finanziamenti transfrontalieri in condizioni di trasparenza e parità. A partire da tale scenario, si è sviluppato un percorso di progressiva armonizzazione del diritto europeo, che ha già condotto, attraverso tappe normative

consolidate, all'adozione di strumenti legislativi atti a colmare quelle lacune e a costruire un impianto regolatorio più coerente e integrato.[1]

In parallelo a questo processo di consolidamento normativo, si è assistito all'adozione sempre più diffusa di sistemi di credit scoring automatizzato per la valutazione del merito creditizio.

L'introduzione di metodi innovativi ha profondamente trasformato il modo in cui viene valutata l'affidabilità creditizia. Tuttavia, se da un lato queste tecniche hanno aumentato la precisione e la capacità predittiva dei modelli di credit scoring, dall'altro ne hanno compromesso spesso la trasparenza. A differenza dei modelli tradizionali, più semplici da interpretare, i nuovi approcci, basati su tecniche avanzate di modellazione, risultano spesso opachi e difficili da spiegare. Al contempo, si è ampliato anche il ventaglio di dati ritenuti rilevanti: accanto alle informazioni raccolte dalle agenzie creditizie, molti operatori finanziari oggi ricorrono a fonti non tradizionali per valutare soggetti con una storia creditizia limitata o assente.

Questa evoluzione ha richiamato l'attenzione crescente delle autorità regolatorie, sia a livello nazionale che europeo, per via delle implicazioni legate all'utilizzo di dati non tradizionali e di modelli algoritmici poco trasparenti. L'uso di queste tecnologie solleva infatti questioni rilevanti in termini di privacy, trasparenza delle decisioni automatizzate e inclusione finanziaria.

Il dibattito è particolarmente attivo nei contesti dove le regole a tutela dei consumatori o i codici di condotta per gli operatori del credito, risultano poco chiari o ancora in fase di sviluppo, alimentando dubbi sull'affidabilità e sull'equità dei nuovi sistemi di credit scoring.

Alla luce di queste dinamiche, il presente capitolo si propone di analizzare le principali norme europee e italiane che regolano il mercato del credito al consumo e la valutazione del merito creditizio, con particolare attenzione agli strumenti legislativi recentemente adottati, al fine di delineare un quadro chiaro e aggiornato del contesto giuridico entro cui si sviluppano oggi i sistemi di credit scoring.[2]

---

[1] https://www.privacy.it/archivio/com2002-443def.html

[2] World Bank Group. (2019). *Credit scoring: Approaches, guidelines*. Washington, DC: World Bank Group.

## 1.1 Direttive europee sulla disciplina del credito al consumo

Ad oggi la disciplina del credito al consumo (CCD – Credit Consumer Directive) a livello europeo è regolata dalla Direttiva 2008/48/CE (CCD I), che ha sostituito la precedente Direttiva 87/102/CEE (CCD), ormai superata rispetto alle esigenze di un mercato del credito sempre più articolato.

La normativa del 1987, pur rappresentando il primo tentativo di armonizzazione tra gli ordinamenti nazionali, mostrava limiti evidenti: lasciava ampio margine di discrezionalità ai singoli Stati membri, non imponeva l'adozione di criteri comuni per il calcolo del TAEG (Tasso Annuo Effettivo Globale)[3], e non forniva strumenti informativi standardizzati per i consumatori.[4]

La Direttiva 2008/48/CE, adottata in un contesto economico più maturo, si inserisce in una logica di rafforzamento della tutela del consumatore che ha guidato buona parte della produzione normativa europea degli ultimi decenni. Il suo obiettivo primario è quello di riequilibrare i rapporti contrattuali tra consumatore e istituti finanziari, riducendo le asimmetrie informative e prevenendo situazioni di sovraindebitamento, attraverso un insieme di misure obbligatorie e vincolanti per gli operatori del credito.

Uno dei capisaldi della riforma è rappresentato dal rafforzamento del principio di trasparenza, che si traduce in una serie di obblighi più rigorosi per i creditori. In primo luogo, questi ultimi saranno tenuti a fornire al consumatore, prima della conclusione del contratto, un insieme di informazioni standardizzate, chiare e facilmente confrontabili. A ciò si affianca l'obbligo di valutare in modo adeguato la solvibilità del consumatore prima della concessione del credito, al fine di garantire un'erogazione responsabile e prevenire situazioni di sovraindebitamento.

Infine, la direttiva prevede che i creditori mantengano un dialogo attivo e costante con il consumatore: l'intermediario non può più limitarsi a un ruolo meramente esecutivo, ma

[3] Il TAEG rappresenta lo strumento principale di trasparenza nei contratti di credito al consumo. E' un indice armonizzato a livello comunitario che nelle operazioni di credito al consumo rappresenta il costo totale del credito a carico del consumatore, comprensivo degli interessi e di tutti gli altri oneri da sostenere per l'utilizzazione del credito stesso. Il TAEG è espresso in percentuale del credito concesso e su base annua. Deve essere indicato nella documentazione contrattuale e nei messaggi pubblicitari o nelle offerte comunque formulate. Definizione tratta da: Banca d'Italia, Glossario in TAEG.

[4] Parlamento Europeo e Consiglio dell'Unione Europea. (1987). *Direttiva 87/102/CEE del Consiglio, del 22 dicembre 1986, per il ravvicinamento delle disposizioni legislative, regolamentari e amministrative degli Stati membri in materia di credito al consumo.*

deve impegnarsi concretamente nell'individuare soluzioni sostenibili, in un'ottica di prevenzione dell'inadempimento e di gestione equilibrata del rapporto contrattuale.[5]

L'attuale Direttiva 2008/48/CE continuerà a disciplinare il credito ai consumatori ancora per un periodo limitato. Il 23 ottobre 2023 è stata ufficialmente approvata la Direttiva (UE) 2023/2225, che dovrà essere recepita dagli Stati membri entro il 20 novembre 2025 e troverà applicazione a partire dal 20 novembre 2026. Fino ad allora, il quadro normativo continuerà a fondarsi sulla direttiva attualmente in vigore.[6]

È importante soffermarsi sul contenuto della nuova Direttiva, poiché, seppur non ancora in vigore, rappresenta un passaggio centrale rispetto ai temi che verranno affrontati successivamente nell'elaborato. Costituisce infatti, una prima chiave di lettura utile per comprendere il contesto più ampio entro cui si inserisce l'evoluzione del credit scoring e l'uso crescente dell'intelligenza artificiale.

La Direttiva 2023/2225UE (CCD II) è il frutto di un lungo processo di revisione della Direttiva 2008/48/CE. Obiettivo inserito tra le principali iniziative che l'Unione Europea si era ripromessa di assumere nel contesto della Agenda dei consumatori presentata nel 2020.

Questa necessità di cambiamento deriva dalla consapevolezza da parte del legislatore europeo che la direttiva del 2008 si è dimostrata solo parzialmente efficace nel perseguimento degli obiettivi prefissati, portando a un livello di protezione dei consumatori inadeguato. Da qui l'esigenza non tanto di ridefinire i principi ispiratori, quanto di rafforzarli attraverso strumenti normativi più chiari, inclusivi e adatti al contesto attuale.[7]

L'adozione di una nuova direttiva sul credito al consumo non nasce solo dalla volontà di colmare le lacune emerse nella normativa precedente, ma anche dalla necessità di tenere il passo con i profondi cambiamenti tecnologici degli ultimi decenni. Nuovi prodotti,

---

[5] Parlamento Europeo e Consiglio dell'Unione Europea. (2008). *Direttiva 2008/48/CE del Parlamento europeo e del Consiglio, del 23 aprile 2008, relativa ai contratti di credito ai consumatori e che abroga la direttiva 87/102/CEE.*

[6] Parlamento Europeo e Consiglio dell'Unione Europea. (2023). *Direttiva (UE) 2023/2225 del Parlamento europeo e del Consiglio, del 18 ottobre 2023, relativa ai contratti di credito ai consumatori e che abroga la direttiva 2008/48/CE*.

[7] https://www.altalex.com/documents/2023/11/14/ccd-ii-prima-lettura-direttiva-credito-consumo-ue-2023-2225

canali e modalità di erogazione, in particolare tramite piattaforme online, hanno reso più semplice l'accesso ai finanziamenti, ma al contempo più complessa la valutazione del rischio e la tutela dell'utente finale.[8]

Inoltre, la rapida diffusione dell'intelligenza artificiale e delle sue applicazioni nel settore finanziario ha riportato l'attenzione sul processo di valutazione del merito creditizio, evidenziando le opportunità offerte dall'uso dei dati, ma anche i rischi connessi alla gestione di informazioni sensibili.[9]

Di seguito si riportano, senza pretesa di esaustività, alcune delle principali novità introdotte dalla CCD II.

L'articolo 2, dopo aver chiarito che la nuova disciplina si applica ai "contratti di credito", elenca in modo dettagliato le diverse tipologie contrattuali escluse dall'ambito di applicazione della CCD I.

Rientrano ora nel campo di applicazione anche i contratti di credito fino a 100.000 euro, non garantiti da ipoteca o da altra garanzia analoga sui beni immobili o da altro diritto connesso ai beni immobili, eliminando la precedente soglia minima pari ad euro 75.000 euro.[10] Sono previsti i prestiti di importo inferiore a 200 euro e i contratti di locazione o di leasing con opzione di acquisto. La direttiva include inoltre i servizi di credito tramite piattaforme di *crowdfunding*, quando il finanziamento è erogato direttamente ai consumatori o quando i prestatori facilitano la concessione di credito tra operatori professionali e consumatori. Tra le novità più rilevanti si segnala l'inclusione dei contratti basati sul modello "compro ora, pago dopo" (*buy now, pay later* – BNPL), sempre più diffusi nel mercato digitale.

L'articolo 5 stabilisce che gli Stati membri devono garantire che tutte le informazioni fornite ai consumatori in conformità alla presente direttiva siano messe a disposizione gratuitamente, a prescindere dal canale utilizzato per la comunicazione.

---

[8] https://www.dirittobancario.it/art/ccd-ii-la-nuova-direttiva-sul-credito-al-consumo/

[9] https://www.rivistadellaregolazionedeimercati.it/Article/Archive/index_html?ida=326&idn=23&idi=-1&idu=-1

[10] Con tale eccezione, il legislatore ha inteso confermare la specificità del credito garantito da ipoteca o da altra garanzia immobiliare, che continuerà ad essere disciplinato dalla Direttiva 2014/17/UE (Mortgage Credit Directive). Allo stesso tempo, ha ritenuto opportuno estendere la tutela rafforzata della CCD II anche ai contratti privi di garanzie immobiliari ma finalizzati alla ristrutturazione dell'abitazione, riconoscendone la particolare rilevanza per il consumatore.

L'esigenza di ottenere dal consumatore un consenso non solo informato ma anche pienamente consapevole sta alla base di una ulteriore innovazione della nuova direttiva.
L'articolo 8 a tal proposito, impone che ogni forma di pubblicità relativa ai contratti di credito contenga un chiaro avvertimento sul costo del finanziamento, mediante la formula esplicita "*Attenzione! Prendere in prestito denaro costa denaro"*. Anche se il messaggio può sembrare semplice e quasi scontato, la sua funzione è chiara: indurre il consumatore a riflettere su i possibili effetti collaterali del finanziamento e decidere coscienziosamente.
In aggiunta, l'articolo prescrive che devono essere fornite in modo facilmente leggibile o chiaramente udibile alcune informazioni essenziali, come la natura del tasso d'interesse (fisso o variabile), l'importo totale del credito, il TAEG e la durata del contratto. In circostanze specifiche e motivate, come nel caso della pubblicità radiofonica, se il mezzo non consente la presentazione visiva delle informazioni, queste dovrebbero essere ridotte per evitare un sovraccarico di informazioni e ridurre gli oneri superflui.[11]

Il Capo V della CCD II assume un ruolo introduttivo centrale rispetto alla questione dei dati alternativi. Rappresenta un salto qualitativo importante verso un modello di credit scoring più trasparente, responsabile e attento ai diritti del consumatore, in risposta alla crescente penetrazione dell'intelligenza artificiale nei processi decisionali del settore finanziario.
L'articolo 18 impone al creditore l'obbligo di effettuare, prima della stipula del contratto di credito, una valutazione approfondita del merito creditizio del consumatore, basata su dati pertinenti, completi e accurati riguardanti la situazione economica e finanziaria dello stesso. Gli elementi raccolti devono risultare necessari e proporzionati rispetto alla natura, alla durata, all'ammontare e ai rischi del contratto di credito per il consumatore.
Le informazioni utilizzate nella valutazione possono includere prove relative al reddito, ad altre fonti di rimborso, ad attività e passività finanziarie, nonché ad ulteriori impegni economici del consumatore. Tuttavia, sono escluse le categorie particolari di dati previste dall'art. 9, par. 1 del Regolamento (UE) 2016/679 (GDPR). Tali dati possono essere raccolti da fonti interne o esterne, incluso lo stesso consumatore, ed eventualmente mediante consultazione di banche dati ai sensi dell'articolo 19 della direttiva.

[11] https://www.tidona.com/la-nuova-disciplina-europea-sul-credito-al-consumo-2/

La direttiva esclude espressamente che i social network possano essere considerati dai creditori quali fonti di informazioni su cui basare la valutazione del merito creditizio.
Nei casi in cui tale valutazione si fondi su un trattamento automatizzato di dati personali, il consumatore ha il diritto di: i) ottenere dal creditore una spiegazione chiara e comprensibile dell'esito della valutazione; ii) esprimere la propria opinione; iii) sollecitare un riesame della valutazione stessa.
L'articolo 19 regola l'uso delle banche dati nella valutazione del merito creditizio. Stabilisce che solo i creditori soggetti alla vigilanza dell'autorità nazionale competente e pienamente conformi al Regolamento (UE) 2016/679 (GDPR) possano accedere a tali banche dati. Questo vincolo assicura che il trattamento dei dati personali avvenga nel rispetto degli standard europei in materia di protezione della privacy.
Parallelamente, i gestori delle banche dati sono tenuti ad adottare procedure adeguate per garantire l'aggiornamento e l'accuratezza delle informazioni contenute nei loro sistemi, per evitare che le decisioni creditizie si basino su dati obsoleti o inesatti.
Infine, nel caso in cui una richiesta di credito venga respinta a seguito della consultazione di una banca dati, il creditore ha l'obbligo di informare tempestivamente e gratuitamente il consumatore. La comunicazione deve includere non solo l'esito della consultazione, ma anche i dettagli della banca dati utilizzata e le categorie di dati su cui si è fondata la valutazione negativa.

Un'ulteriore innovazione introdotta dalla CCD II su cui è opportuno soffermarsi, riguarda il contenimento dei tassi e, più in generale, del costo totale del credito, tematica cui è dedicato il capo IX della disciplina.
In particolare, l'articolo 31 impone agli Stati membri l'obbligo di adottare misure efficaci volte a prevenire abusi e a garantire che i consumatori non siano soggetti a tassi debitori, TAEG o costi complessivi eccessivamente elevati. Il legislatore europeo impone quindi agli Stati membri di adottare meccanismi di contingentamento non solo dei tassi di interesse applicati al finanziamento, ma estende tale limitazione anche al profilo dei costi.
Sulla base del quadro normativo delineato, è ora opportuno approfondire in che modo tali previsioni siano state recepite nell'ordinamento italiano.

### 1.1.1 Valutazione del merito creditizio nel diritto italiano

In Italia, il merito creditizio e il ricorso a sistemi di credit scoring sono oggetto di una disciplina articolata, volta a garantire il prestito responsabile e la tutela sia della stabilità del sistema bancario che dei consumatori.

Il Testo Unico Bancario (TUB) costituisce la fonte primaria della normativa sul credito in Italia. In particolare, il Titolo VI del TUB è dedicato alla trasparenza e correttezza delle relazioni tra intermediari e clienti e include le disposizioni sul credito ai consumatori introdotte con il d.lgs. n. 141/2010[12], in recepimento della direttiva 2008/48/CE.

Il fulcro di questo cambiamento è rappresentato dal principio di "sana e prudente gestione" sancito all'articolo 5 del Testo Unico Bancario. Tradizionalmente inteso come principio guida per garantire la solidità degli intermediari finanziari, tale regola assume, in questa fase, una portata più ampia: non solo strumento di controllo prudenziale, ma anche clausola generale a tutela dell'equilibrio nei rapporti tra banca e clientela.

In un contesto in cui la complessità dei beni negoziati sui mercati finanziari è sempre in evoluzione, la trasparenza inizia, quindi, ad essere declinata sotto il profilo della correttezza delle condotte e diviene sinonimo di assistenza e vera e propria consulenza del cliente.

Questa nuova prospettiva trova formale riconoscimento nell'ampliamento dell'articolo 127 TUB, che attribuisce alle autorità creditizie, tra cui la Banca d'Italia, il compito di esercitare poteri di controllo non più limitati agli aspetti prudenziali, ma estesi anche alla trasparenza delle condizioni contrattuali e alla correttezza dei rapporti con la clientela. È un passaggio cruciale: per la prima volta, il legislatore equipara, sul piano degli obiettivi di vigilanza, la stabilità del sistema finanziario alla tutela del consumatore. Le due finalità diventano così interdipendenti, e la Banca d'Italia assume il ruolo di garante anche del rispetto delle norme di protezione della clientela bancaria, contribuendo in modo attivo alla regolazione sostanziale del mercato del credito.[13]

---

[12] Repubblica Italiana. (2010). *Decreto legislativo 13 agosto 2010, n. 141. Attuazione della direttiva 2008/48/CE sui contratti di credito ai consumatori, nonché modifiche del titolo VI del testo unico bancario (TUB)*. Gazzetta Ufficiale della Repubblica Italiana, Serie generale, n. 207 del 4 settembre 2010.

[13] Robustella, L. (2024). *Misure preventive e restitutorie nell'ambito della vigilanza di tutela.* In A. Urbani, R. Natoli, & D. Rossano (Eds.), Quaderni di Ricerca Giuridica della Consulenza Legale della Banca d'Italia – A 30 anni dal Testo unico bancario (1993–2023). (pp. 251–266). Banca d'Italia.

L'art. 124-bis TUB, rubricato "Verifica del merito creditizio", impone ai finanziatori di valutare sempre il merito creditizio del consumatore prima di concludere un contratto di credito al consumo. Ciò deve avvenire sulla base di informazioni adeguate e attuali, fornite se necessario dallo stesso consumatore e, ove occorra, acquisite consultando banche dati pertinenti.[14]

Inoltre, lo stesso articolo prevede che, qualora dopo la conclusione del contratto le parti concordino un aumento significativo dell'importo totale del credito, il finanziatore debba aggiornare le informazioni finanziarie a propria disposizione e rivalutare il merito creditizio del consumatore prima di procedere ad ampliare il fido. Questo implica che anche in caso di concessione di credito aggiuntivo, occorre una nuova verifica di solvibilità. L'art. 124-bis demanda infine alla Banca d'Italia il compito di emanare disposizioni attuative in materia, in conformità alle deliberazioni del CICR[15].

L'art. 125 TUB rubricato "Banche dati" disciplina il trattamento delle informazioni creditizie raccolte nei sistemi di informazione creditizia (SIC), siano essi pubblici o privati e recepisce la normativa europea in merito alla valutazione del merito creditizio.

La norma stabilisce l'accessibilità alle banche dati da parte dei finanziatori comunitari, imponendo condizioni di non discriminazione rispetto agli operatori già presenti sul territorio nazionale. In questo modo, si garantisce una parità di trattamento e si favorisce un'effettiva concorrenza nel mercato del credito.

Un passaggio particolarmente rilevante riguarda l'obbligo, in capo al finanziatore, di informare tempestivamente e gratuitamente il consumatore qualora una richiesta di credito venga respinta sulla base di dati contenuti in una banca dati creditizia. Oltre all'esito della valutazione, devono essere forniti gli estremi dell'archivio consultato e le categorie di dati utilizzate, affinché il consumatore possa prendere visione delle informazioni a lui riferite ed eventualmente attivarsi per correggerle in caso di inesattezze.

---

[14] Cecchinato, E. (2023). *Note sulla disciplina della verifica del merito creditizio: per una sua rilettura alla luce della buona fede precontrattuale.* Rivista di Diritto Bancario, (III).

[15] Il Testo unico bancario (d.lgs. 1° settembre 1993, n. 385 - TUB) attribuisce al Comitato interministeriale per il credito ed il risparmio (CICR) l'alta vigilanza in materia di credito e di tutela del risparmio. Nella regolamentazione dell'attività delle banche e degli altri intermediari finanziari disciplinati dal Testo unico bancario, il CICR delibera, su proposta della Banca d'Italia, principi e criteri per l'esercizio della vigilanza. Le deliberazioni in tema di trasparenza delle condizioni contrattuali concernenti le operazioni e i servizi bancari e finanziari sono assunte su proposta della Banca d'Italia d'intesa con la Consob.

Ulteriori garanzie si applicano nel caso di segnalazioni negative: l'intermediario, prima di registrare un'informazione pregiudizievole – come un ritardo nei pagamenti – è tenuto a darne preventiva comunicazione al cliente. Tale prassi, spesso attuata mediante un sollecito o una formale messa in mora, ha la finalità di tutelare il debitore, avvisandolo dell'imminente registrazione e delle possibili conseguenze.

La norma impone inoltre che le informazioni trasmesse alle banche dati siano corrette e costantemente aggiornate. Qualora vengano riscontrati errori, è onere dell'intermediario procedere senza indugio alla rettifica. Ancora, si richiede che il consumatore venga informato sugli effetti che eventuali annotazioni negative potrebbero avere sulla sua futura capacità di ottenere credito, rafforzando così la consapevolezza e la trasparenza del rapporto contrattuale.

Infine, l'articolo 125 richiama espressamente il rispetto delle disposizioni in materia di protezione dei dati personali, facendo riferimento sia al Codice Privacy (D.Lgs. 196/2003) che al Regolamento Generale sulla Protezione dei Dati (Reg. UE 2016/679), nonché ai codici deontologici che disciplinano il funzionamento dei SIC. In tal modo, il legislatore intende bilanciare le esigenze di affidabilità del sistema creditizio con i diritti fondamentali del consumatore in termini di correttezza, trasparenza e riservatezza.[16]

Un altro importante passo avanti nella regolamentazione del merito creditizio in Italia è stato compiuto con il recepimento della Direttiva 2014/17/UE[17] (CMD – Credit Mortgage Directive), che riguarda i contratti di credito relativi a beni immobili residenziali. Questa direttiva, introdotta dopo la crisi finanziaria dei mutui *subprime*, ha posto l'accento sulla necessità di una concessione del credito più attenta e responsabile, al fine di evitare che decisioni imprudenti possano destabilizzare l'intero sistema finanziario.

Il legislatore italiano ha recepito la direttiva attraverso gli articoli 120-quinquies e seguenti del TUB, rafforzando ulteriormente l'obbligo per i creditori di effettuare un'attenta valutazione della solvibilità dei consumatori prima di concedere un mutuo. In particolare, l'articolo 120-undecies TUB stabilisce che tale valutazione debba basarsi su

[16] Rabitti, M. (2023). *Credit scoring via machine learning e prestito responsabile.* Rivista di Diritto Bancario, (I).

[17] Parlamento Europeo e Consiglio dell'Unione Europea. (2014). *Direttiva 2014/17/UE del Parlamento europeo e del Consiglio, del 4 febbraio 2014, in merito ai contratti di credito ai consumatori relativi a beni immobili residenziali e recante modifica delle direttive 2008/48/CE e 2013/36/UE e del regolamento (UE) n. 1093/2010*.

informazioni complete, verificate, proporzionate e rilevanti rispetto alla situazione economica e finanziaria del richiedente. Il criterio fondamentale non è più l'esistenza di una garanzia immobiliare sufficiente, ma la reale capacità del consumatore di rimborsare il prestito secondo i termini del contratto.

È inoltre previsto che il valore dell'immobile non possa costituire l'elemento determinante per concedere il credito, né si può fare affidamento sull'eventuale aumento del valore dell'immobile nel tempo, salvo che il finanziamento sia destinato a costruzione o ristrutturazione. Questo approccio mira a evitare l'illusione che una garanzia immobiliare possa compensare l'assenza di una reale solidità finanziaria del mutuatario.

Il legislatore ha anche chiarito che, sebbene il creditore possa trasferire il rischio di credito a terzi (ad esempio attraverso cartolarizzazioni o assicurazioni), ciò non lo esonera dal rispetto dell'obbligo di effettuare una valutazione rigorosa del merito creditizio. In altre parole, il rischio non può essere "scaricato" altrove per giustificare l'erogazione di prestiti insostenibili.

Le informazioni utilizzate per la valutazione provengono primariamente dal consumatore, ma il creditore ha l'obbligo di verificarle con attenzione, anche attraverso documentazione indipendente se necessario. A tal fine, può consultare la Centrale Rischi della Banca d'Italia e i Sistemi di Informazioni Creditizie (SIC), strumenti fondamentali per accertare la situazione debitoria del richiedente.

Infine, la direttiva sottolinea anche l'importanza dell'educazione finanziaria, invitando gli Stati membri a promuovere iniziative che aiutino i consumatori a comprendere meglio i rischi connessi all'indebitamento e a prendere decisioni più consapevoli.[18]

Il quadro normativo ricostruito fin qui, a partire dalle direttive europee sul credito al consumo fino al recepimento interno da parte dell'ordinamento italiano, mette in luce un processo graduale di consolidamento delle tutele legate alla valutazione del merito creditizio. Tuttavia, nonostante i significativi passi avanti, resta evidente come la disciplina del credit scoring, soprattutto con riferimento all'impiego di dati alternativi e all'uso crescente di sistemi automatizzati, presenti ancora ampie aree di incertezza. È proprio in questa zona grigia che si apre la necessità di ulteriori strumenti normativi,

[18] Sirena, P. (2024). *Tutela dei clienti e regolazione del mercato trent'anni dopo l'emanazione del Testo Unico Bancario.* In A. Urbani, R. Natoli, & D. Rossano (Eds.), Quaderni di Ricerca Giuridica della Consulenza Legale della Banca d'Italia – A 30 anni dal Testo unico bancario (1993–2023). (pp. 123–137). Banca d'Italia.

capaci di accompagnare l'innovazione tecnologica senza sacrificare le garanzie sostanziali a tutela del consumatore.

### 1.2 GDPR per la tutela dei dati personali e profilazione automatizzata

Per comprendere i confini giuridici dell'utilizzo di dati non tradizionali nei processi di valutazione creditizia, è indispensabile integrare l'analisi con due testi normativi fondamentali: il Regolamento Generale sulla Protezione dei Dati (GDPR) e il recente Artificial Intelligence Act. Entrambi i documenti rappresentano pilastri centrali nella regolazione dell'economia digitale europea e forniscono indicazioni cruciali sui limiti, le responsabilità e le condizioni di liceità nell'utilizzo di tecnologie algoritmiche e fonti informative non convenzionali.

Il Regolamento generale sulla protezione dei dati personali (GDPR), Regolamento (UE) 2016/679[19], rappresenta la risposta normativa dell'Unione Europea alla necessità di tutelare in modo uniforme e moderno il diritto alla privacy nell'era digitale. Approvato nel 2016 ed entrato in vigore il 25 maggio 2018, il GDPR ha sostituito la previgente Direttiva 95/46/CE, superando le frammentarie attuazioni nazionali e imponendo standard comuni agli Stati membri.

In Italia, le disposizioni del GDPR sono direttamente applicabili e contestualmente, il legislatore italiano ha emanato il D.Lgs. 10 agosto 2018 n. 101[20], in vigore dal 19 settembre 2018, che ha adeguato il Codice Privacy nazionale (D.Lgs. 196/2003) alle nuove regole europee, introducendo alcune fattispecie di illeciti penali, accanto alle sanzioni pecuniarie già previste dal GDPR.

Ne è derivato un sistema aggiornato e più rigoroso, volto a rispondere alle sfide poste dallo sviluppo tecnologico e dalla globalizzazione dei flussi di dati, garantendo al contempo una tutela effettiva dei diritti fondamentali degli interessati in tutta l'UE.

---

[19] Parlamento Europeo e Consiglio dell'Unione Europea. (2016). *Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati)*.

[20] Repubblica Italiana. (2018). *Decreto legislativo 10 agosto 2018, n. 101. Disposizioni per l'adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati*. Gazzetta Ufficiale della Repubblica Italiana, Serie generale, n. 205 del 4 settembre 2018.

Le norme si applicano anche alle imprese situate fuori dall'Unione europea che offrono servizi o prodotti all'interno del mercato UE. Tutte le aziende, ovunque stabilite, dovranno quindi rispettare le nuove regole.[21]

Sul piano contenutistico, il GDPR ha introdotto principi e obblighi innovativi per titolari e responsabili del trattamento, rafforzando in parallelo le garanzie per gli interessati.
Di seguito una breve rassegna dei principi cardine del documento e alcune delle applicazioni più rilevanti.

- Principio di *accountability*: prevede che il titolare del trattamento garantisca e sia in grado di dimostrare la conformità alla normativa, adottando misure organizzative e di sicurezza adeguate;

- Privacy by design e by default[22]: tale concetto impone che, per impostazione predefinita, siano trattati solo i dati strettamente necessari alla finalità prevista, per il tempo indispensabile e accessibili solo da chi ne ha bisogno.

- Sanzioni amministrative pecuniarie: si introducono importi più elevati, ci sono diverse fattispecie e si va da una mera diffida amministrativa a sanzioni fino a 20 milioni di euro.

- Trasparenza e consenso: vengono definiti obblighi più stringenti di informativa e requisiti più chiari per un consenso valido, inoltre accanto ai diritti già previsti dalla normativa previgente, il GDPR ne ha ampliato la portata e ne ha introdotti di ulteriori.[23]

  Il primo adempimento previsto dal GDPR per le imprese italiane è l'adozione del registro dei trattamenti, obbligatorio per chi ha almeno 250 dipendenti, ma anche

[21] https://www.agendadigitale.eu/cittadinanza-digitale/gdpr-tutto-cio-che-ce-da-sapere-per-essere-preparati/#:~:text=In%20data%2019%20settembre%202018,pecuniarie%20gi%C3%A0%20previste%20dal%20GDPR

[22] https://protezionedatipersonali.it/privacy-by-design-e-by-default#:~:text=Il%20principio%20di%20privacy%20by%20default%20(protezione%20per%20impostazione%20predefinita,che%20possono%20accedere%20ai%20dati.

[23] https://privacy.it/2017/05/25/12-passi-per-prepararsi-al-gdpr/

per realtà più piccole se i trattamenti non sono occasionali, comportano rischi per i diritti degli interessati o riguardano dati sensibili o giudiziari. Si tratta di un documento fondamentale che descrive le attività di trattamento svolte dal titolare dell'impresa, utile per monitorare i trattamenti e da esibire in caso di controlli del Garante. Il registro, redatto in forma scritta anche elettronica, deve contenere le informazioni previste dall'art. 30 del GDPR, secondo le indicazioni fornite dal Garante nel 2018.[24] Inoltre, un ruolo centrale nella garanzia del rispetto delle norme in materia di protezione dei dati personali è affidato al responsabile della protezione dei dati, noto come Data Protection Officer (DPO). Questa figura, obbligatoria in determinati casi, è incaricata di sorvegliare l'applicazione del Regolamento all'interno di enti pubblici o imprese private, assicurandosi che le operazioni di trattamento siano conformi alle disposizioni normative. Il DPO agisce in piena autonomia e riferisce direttamente ai vertici dell'organizzazione; deve essere scelto in base a specifiche competenze giuridiche e tecniche in ambito privacy, oltre che in relazione alla complessità del contesto operativo. Pur non determinando le finalità del trattamento, la sua funzione è essenziale per promuovere una cultura della responsabilità e rafforzare la tutela dei diritti fondamentali degli interessati.

- Limiti al trattamento automatizzato: per la prima volta viene delineata espressamente una tutela contro decisioni basate unicamente su trattamenti automatizzati che producano effetti giuridici o incidano in modo significativo sulla persona (art. 22 GDPR);

Nella sua configurazione generale il GDPR pone un forte accento sulla tutela dell'individuo di fronte alla profilazione automatizzata, riconoscendo nuovi diritti e imponendo limiti chiari all'uso di algoritmi decisionali, in modo da assicurare trasparenza, correttezza e intervento umano nei trattamenti che potrebbero influire significativamente sulle persone.

[24] https://www.agendadigitale.eu/cittadinanza-digitale/gdpr-tutto-cio-che-ce-da-sapere-per-essere-preparati/#:~:text=In%20data%2019%20settembre%202018,pecuniarie%20gi%C3%A0%20previste%20dal%20GDPR

Il GDPR definisce infatti la profilazione come qualsiasi forma di trattamento automatizzato consistente nell'utilizzo di dati personali per valutare aspetti personali di una persona fisica, in particolare allo scopo di analizzare o prevedere elementi come il rendimento professionale, l'affidabilità, la situazione economica, il comportamento ecc. (art. 4, par. 1 n.4 GDPR).
Inoltre, l'art. 22 del Regolamento sancisce in via generale che "l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato [...] che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona", salvo che tale decisione sia necessaria per l'esecuzione di un contratto, autorizzata da una legge, oppure fondata sul consenso esplicito dell'interessato.[25] Anche quando il trattamento automatizzato rientri in queste eccezioni, il GDPR prevede comunque importanti cautele: il titolare del trattamento deve infatti adottare misure idonee a tutelare i diritti, le libertà e i legittimi interessi dell'interessato, garantendo quantomeno il diritto a ottenere intervento umano, a esprimere la propria opinione e a contestare la decisione automatizzata (art. 22, parr. 2-3). In aggiunta, gli obblighi di trasparenza impongono di informare preventivamente l'interessato dell'eventuale esistenza di processi decisionali automatizzati, fornendo "informazioni significative sulla logica utilizzata" e sulle conseguenze previste di tali processi (artt. 13-15 GDPR).[26]

Le garanzie offerte dal GDPR assumono particolare rilievo in un settore specifico come quello del credit scoring. La digitalizzazione del settore finanziario ha reso comune l'uso di algoritmi e grandi moli di dati per profilare i clienti al fine di valutare il merito creditizio, spesso attingendo non solo ai dati creditizi tradizionali, storico di rimborso di prestiti, utilizzo del credito, durata della storia creditizia, ecc., ma anche a dati "alternativi" provenienti da fonti eterogenee, utenze, comportamento online, social media, e così via. Questo processo, cruciale per l'efficienza e la rapidità di concessione del credito, comporta però rischi rilevanti per i diritti dei consumatori, tra cui possibili

[25] Gaggero, A., Valenza, M. (2024). *Intelligenza artificiale e merito creditizio: nuovi orizzonti normativi tra AI Act e GDPR*. Rivista di Diritto Bancario, (III).

[26] https://ristrutturazioniaziendali.ilcaso.it/Articolo/580#:~:text=,automatizzate%20e%20richiedere%20l%27intervento%20umano

violazioni della privacy, mancanza di trasparenza, errori o *bias* algoritmici discriminatori che potrebbero precludere l'accesso al credito in modo ingiusto.[27]

Nonostante l'importanza del tema per il corretto funzionamento dei mercati finanziari, le normative di settore, ad esempio le direttive europee sul credito al consumo, come abbiamo visto precedentemente, non contengono riferimenti puntuali alla profilazione automatizzata a fini di concessione del credito, né disciplinano in dettaglio l'utilizzo di dati non tradizionali in tali valutazioni.

In questo vuoto normativo, il GDPR fornisce parte fondamentale del quadro di riferimento per garantire che i sistemi di credit scoring siano utilizzati in modo lecito.

In primo luogo, ogni trattamento di dati personali svolto per fini di credit scoring deve fondarsi su un'idonea base giuridica ai sensi dell'art. 6 GDPR. Nella pratica, la base giuridica potrà essere, il consenso esplicito dell'interessato oppure, più spesso invocata dagli operatori di settore, il legittimo interesse del creditore nell'analizzare l'affidabilità del potenziale cliente, o ancora la necessità di eseguire misure precontrattuali su richiesta dell'interessato.[28]

L'art. 9 del GDPR riveste un ruolo cruciale nel contesto del credit scoring, in particolare per quanto riguarda l'utilizzo dei dati alternativi. Tale norma vieta in linea generale il trattamento di dati sensibili, come quelli che rivelano origine razziale o etnica, opinioni politiche, convinzioni religiose o filosofiche, appartenenza sindacale, dati genetici, biometrici, relativi alla salute o alla vita sessuale o all'orientamento sessuale, salvo specifiche deroghe tassative. Questo implica che, anche laddove tali dati siano tecnicamente accessibili e potenzialmente rilevanti per l'analisi creditizia, non possono essere trattati ai fini della valutazione del merito creditizio, se non ricorrendo a una delle eccezioni previste al par. 2. L'art. 9 rappresenta un vero limite giuridico all'estensione indiscriminata delle fonti informative, assicurando che la profilazione automatizzata non violi i diritti fondamentali dell'interessato.[29]

---

[27] Paal, B. (2023). Article 22 GDPR: *Credit Scoring Before the CJEU*. *GPLR*, 127-137.

[28] https://www.agendadigitale.eu/sicurezza/privacy/merito-creditizio-lai-riscrive-le-regole-le-norme-che-ci-tutelano/

[29] Busacca, A. (2018). *Le "categorie particolari di dati" ex art. 9 GDPR: Divieti, eccezioni e limiti alle attività di trattamento*. Ordine internazionale e diritti umani.

Ulteriore principio cardine è quello di minimizzazione dei dati: possono essere raccolti ed elaborati soltanto i dati pertinenti e strettamente necessari allo scopo di determinare lo score creditizio, escludendo informazioni eccedentarie o non rilevanti. Questo principio mira a limitare l'invasività delle analisi automatizzate, evitando che le banche o agenzie di rating attingano a quantità di dati sproporzionate rispetto alla finalità di valutazione del rischio. Inoltre, vige un obbligo di esattezza e aggiornamento dei dati (art. 5, co.1 lett. d GDPR), cruciale nel contesto creditizio: errori, dati obsoleti o non corretti nelle banche dati creditizie potrebbero condurre a valutazioni errate del merito creditizio, penalizzando indebitamente i consumatori.

In quest'ottica, il GDPR impone ai titolari anche di adottare misure affinchè i dati utilizzati nei processi decisionali siano accurati e, in caso di inesattezze, rettificati tempestivamente su richiesta dell'interessato. Quest'ultimo, a sua volta, ha il diritto di accesso ai propri dati e di ottenere la correzione di informazioni inesatte (art. 15 e 16 GDPR).

L'aspetto centrale sul quale il GDPR incide, è proprio quello delle decisioni interamente automatizzate nel credit scoring. Come già evidenziato, l'art. 22 GDPR stabilisce il diritto a non essere sottoposti a decisioni basate unicamente su trattamenti automatizzati che producano effetti giuridici o incidano significativamente sulla persona.

Va sottolineato che la mera esistenza di un intervento umano formale non deve diventare un escamotage per eludere la norma: la persona coinvolta nel processo decisionale deve avere effettivamente il potere di riesaminare e modificare la decisione, altrimenti si ricadrebbe comunque nell'alveo dell'assoluto processo automatizzato.[30]

In conclusione, pur non essendo una normativa specificamente dedicata al settore bancario, il GDPR rappresenta oggi un pilastro essenziale nella regolamentazione dei processi automatizzati di valutazione del merito creditizio. La sua applicazione si integra con le direttive europee e nazionali in materia di credito al consumo, con i principi del *responsible lending* e con i più recenti sviluppi normativi in ambito tecnologico, come l'Artificial Intelligence Act. Insieme, queste fonti concorrono a definire un quadro regolamentare in continua evoluzione, necessario per governare in modo equilibrato

[30] https://www.agendadigitale.eu/mercati-digitali/profilazione-e-verifiche-del-merito-creditizio-un-difficile-equilibrio/#:~:text=forme%20di%20profilazione%20incontrollate%2C%20che,a%20conseguenze%20imprevedibili%20ed%20ingiuste

l'impiego crescente di sistemi automatizzati e di dati alternativi nelle pratiche di credit scoring.

### 1.3 Artificial Intelligence Act e credit scoring

Il Consiglio dell'UE ha approvato unanimemente il 21 Maggio 2024, Il Regolamento (UE) 2024/1689[31], noto come Artificial Intelligence Act (AI Act), che rappresenta il primo quadro giuridico europeo organico sull'intelligenza artificiale.

L'AI Act istituisce norme armonizzate di carattere precauzionale per promuovere un'IA "affidabile" e mitigare i rischi legati ai diritti fondamentali e alla sicurezza.

Gli obiettivi dichiarati del regolamento sono molteplici: favorire un mercato unico dell'IA, aumentare la fiducia dei cittadini tramite sistemi affidabili e trasparenti che rispettino principi etici, prevenire e ridurre i rischi legati agli usi dell'IA e al contempo sostenere l'innovazione responsabile con incentivi e cooperazione fra Stati membri.

L'AI Act si inserisce in un più ampio quadro normativo europeo, integrando il Regolamento Generale sulla Protezione dei Dati (GDPR) e mantenendo coerenza con altre importanti iniziative legislative, quali il Digital Services Act (DSA), il Digital Markets Act (DMA), il Data Governance Act (DGA), la direttiva sull'intelligenza artificiale in materia di responsabilità civile, la direttiva aggiornata sulla responsabilità del prodotto e diversi regolamenti settoriali sulla sicurezza dei prodotti.[32]

Nel corpo normativo dell'AI Act, il concetto di "sistema di intelligenza artificiale" è definito in modo volutamente ampio e tecnologicamente neutro, così da garantire che la normativa resti attuale anche in un contesto in continua evoluzione. Secondo l'articolo 3, un sistema di IA è "un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni,

---

[31] Unione Europea. (2024). *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce norme armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica i regolamenti (CE) n. 300/2008, (UE) 2018/858, (UE) 2019/2144 e (UE) 2020/1056 del Parlamento europeo e del Consiglio e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 del Parlamento europeo e del Consiglio*.

[32] https://www.agendadigitale.eu/cultura-digitale/ai-act-ci-siamo-ecco-come-plasmera-il-futuro-dellintelligenza-artificiale-in-europa/#:~:text=dell%E2%80%99esercizio%20precedente%20per%20la%20fornitura,in%20risposta%20a%20una%20richiesta

contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali". Le parole chiave di questa definizione sono "deduzione" e "autonomia", che distinguono i sistemi di IA dai software tradizionali, in cui gli output sono predeterminati da rigide regole logiche.[33]

La portata inclusiva della definizione consente di ricomprendere in essa non solo i sistemi basati su tecniche di machine learning, ma anche quelli fondati su approcci differenti, purché capaci di adattarsi o generare risultati che impattano concretamente sull'ambiente circostante. In tal senso, la definizione europea si ispira all'ultima formulazione dell'OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico) e si distacca da modelli più restrittivi, come quelli incentrati esclusivamente sui sistemi esperti, riconoscendo che l'intelligenza artificiale, nel tentativo di emulare il pensiero umano, produce risultati che, seppur plausibili o ragionevoli, possono essere incerti e suscettibili ad errori. Di conseguenza, i sistemi di IA devono essere valutati e regolati in funzione del potenziale impatto, e non della sola struttura tecnica che li caratterizza.

### 1.3.1 Struttura generale dell'AI Act e classificazione per rischio

Strutturato in 12 titoli, 85 articoli e 9 allegati, l'AI Act si fonda su un approccio umano-centrico basato sul rischio: quanto più elevato è il rischio associato a un sistema di intelligenza artificiale, tanto più stringenti saranno gli obblighi normativi e le responsabilità per sviluppatori e utilizzatori. In tal senso, il regolamento prevede anche il divieto assoluto per le applicazioni ritenute eccessivamente pericolose.[34]

1. Il primo livello di rischio classifica i sistemi a "rischio inaccettabile": L'Unione Europea include in questo primo livello di rischio le tecnologie il cui impiego è vietato perché ritenuto lesivo dei diritti fondamentali e incompatibile con i principi fondanti dell'ordinamento europeo, come la dignità umana, la libertà individuale e lo stato di diritto.

[33] Maestri, E. (2024). *L'impatto della normativa europea sull'intelligenza artificiale: una garanzia di sicurezza o un ostacolo all'innovazione?* Agenda 17 – Webmagazine del Laboratorio Design of Science, Università degli Studi di Ferrara.

[34] https://www.credit-scoring.co.uk/blog/eu-ai act#:~:text=,factors%20that%20influence%20credit%20scores

Questi sistemi, infatti, impiegano tecniche manipolative, discriminatorie o intrusivamente sorveglianti, con effetti potenzialmente gravi sulla vita delle persone. Tra i casi espressamente vietati figurano le applicazioni di IA che manipolano il comportamento umano sfruttando vulnerabilità individuali, come l'età o la disabilità, che valutano i cittadini con meccanismi di "social scoring", che deducono emozioni in ambienti lavorativi o scolastici, o che classificano gli individui in base a dati biometrici sensibili come razza o orientamento sessuale.[35]

È vietato anche l'uso del riconoscimento facciale "in tempo reale" in spazi pubblici, salvo eccezioni strettamente regolamentate, come la ricerca di minori scomparsi. Il divieto non si fonda solo su valutazioni etiche, ma anche sulla constatazione che i potenziali danni di tali sistemi sarebbero troppo elevati e difficilmente mitigabili.

In caso di violazione, l'AI Act prevede sanzioni fino a 35 milioni di euro o al 7% del fatturato globale dell'impresa. Questo approccio mira a prevenire abusi strutturali e a mantenere l'IA allineata con i principi democratici dell'UE.[36]

2. Il secondo livello di rischio classifica i sistemi ad "alto rischio": L'Unione Europea definisce come ad alto rischio quei sistemi automatizzati che possono incidere significativamente sui diritti fondamentali o sulla sicurezza delle persone. Rientrano in questa categoria, ad esempio, le applicazioni impiegate in ambiti cruciali come sanità, giustizia, istruzione e servizi pubblici essenziali. In questa casistica rientra il credit scoring automatizzato, la cui capacità di incidere sull'accesso al credito rende necessario un controllo particolarmente rigoroso.[37]

[35] https://www.cybersecurity360.it/legal/ai-act-scattano-i-primi-divieti-chi-rischia-le-sanzioni-e-le-prossime-tappe/

[36] Neuwirth, R. J. (2023). *Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA).* Computer Law & Security Review, 48, 105798.

[37] https://www.agenda17.it/2024/06/16/dossier-ai-act-maggiori-i-rischi-piu-rigorose-le-regole-nella-nuova-legge-europea/

Tali sistemi devono rispettare precisi requisiti in termini di gestione del rischio, governance dei dati, documentazione tecnica, trasparenza, tracciabilità e supervisione umana. Ogni fase del loro ciclo di vita, dalla progettazione all'utilizzo, deve essere strutturata in modo da prevenire danni rilevanti per gli individui e garantire la correttezza e l'affidabilità dei risultati prodotti.

[38]


*Figura 1 - Framework per i sistemi ad alto rischio.*

3. Il terzo livello di rischio classifica i sistemi a "rischio limitato": questo livello include quelle applicazioni che, pur non comportando un impatto diretto e grave sui diritti fondamentali, possono influenzare le decisioni o la percezione dell'utente.

   A tali sistemi è imposto un obbligo essenziale di trasparenza: chi interagisce con contenuti generati in modo automatizzato, come chatbot conversazionali o deepfake audiovisivi, deve essere chiaramente informato della natura non umana dell'interlocutore o dell'origine artificiale del contenuto. Questo obbligo di informativa tutela il diritto dell'individuo a conoscere la natura dell'interazione, evitando manipolazioni occulte e garantendo un uso più consapevole delle tecnologie automatizzate. Sebbene non siano richiesti requisiti tecnici particolarmente onerosi, questa forma di trasparenza si

[38] Novelli, C., Casolari, F., et al. (2024). *AI risk assessment: A scenario-based, proportional methodology for the AI Act.* Digital Society, 3.

configura come un presidio minimo ma imprescindibile per salvaguardare la fiducia degli utenti.[39]

4. Il quarto livello di rischio classifica i sistemi a "rischio minimo o nullo": sono quei sistemi il cui utilizzo non comporta effetti significativi sui diritti fondamentali, sulla sicurezza o sul benessere delle persone. Si tratta di applicazioni che lasciano pieno margine di controllo all'utente e che, proprio per la loro neutralità, sono esentate da obblighi normativi. Questo approccio mira a favorire la sperimentazione e l'innovazione, incentivando lo sviluppo tecnologico in settori a bassa criticità.
   Ne sono esempi tipici i filtri estetici nelle app fotografiche, le funzioni di suggerimento nei videogiochi o altre applicazioni ludiche che non interferiscono con aspetti sensibili della vita individuale o collettiva.[40]

### 1.3.2 L'inquadramento del credit scoring nell'AI Act

All'interno della cornice normativa delineata dall'Artificial Intelligence Act (AI Act), i sistemi automatizzati di credit scoring rientrano espressamente tra le applicazioni considerate ad alto rischio, secondo quanto previsto dall'Allegato III del Regolamento. Questa classificazione riflette la rilevanza del ruolo che tali strumenti ricoprono nell'accesso a servizi essenziali come mutui, prestiti, contratti di locazione e altri strumenti finanziari. L'uso improprio, opaco o discriminatorio di un sistema di valutazione creditizia automatizzata rischia infatti di compromettere gravemente diritti fondamentali quali la parità di trattamento, l'autonomia individuale e l'accesso equo a beni e servizi essenziali.[41]
Conseguentemente, l'AI Act impone ai fornitori e utilizzatori di questi sistemi una serie articolata di obblighi, propri dei sistemi ad alto rischio. Tra questi, vi è innanzitutto l'obbligo di istituire, attuare e mantenere un sistema di gestione dei rischi durante l'intero

[39] Schuett, J. (2023). *Risk management in the Artificial Intelligence Act*. European Journal of Risk Regulation. Cambridge University Press.

[40] https://www.altalex.com/documents/news/2023/06/23/ai-act-ue-traccia-futuro-intelligenza-artificiale

[41] Mattassoglio, F. (2025). *La Corte di giustizia europea, algoritmi e credit scoring. L'apertura del vaso di Pandora delle società che si "limitano" a elaborare gli scoring*. Dialoghi di Diritto dell'Economia, Gennaio 2025. Università degli Studi di Milano-Bicocca.

ciclo di vita del sistema (art. 9), insieme alla necessità di utilizzare dataset di addestramento, convalida e test che siano pertinenti, rappresentativi, privi di *bias* e tecnicamente adeguati allo scopo previsto (art. 10). A supporto della conformità, è inoltre richiesto che sia redatta una documentazione tecnica dettagliata, da presentare alle autorità competenti per la verifica (art. 11), e che sia assicurata la tracciabilità mediante registrazione automatica (art. 12).

Particolare rilievo assumono gli obblighi di trasparenza (art. 13) e supervisione umana (art. 14), essenziali per garantire l'intervento umano nella comprensione, gestione e, se necessario, correzione degli output prodotti dal sistema. A ciò si aggiungono i requisiti di accuratezza, robustezza e sicurezza informatica (art. 15), che devono essere mantenuti per tutta la durata dell'impiego del sistema.

L'AI Act prevede inoltre, che il rispetto degli obblighi non dipenda solo dagli sviluppatori di algoritmi, ma da un'intera "catena del valore". Ciò significa che anche distributori, importatori, fornitori e utilizzatori finali del sistema IA possono essere considerati soggetti responsabili in base al regolamento. Ciascuno di essi deve pertanto istituire procedure interne idonee a gestire i rischi e a documentare le attività. L'art.16 del regolamento afferma che ogni operatore coinvolto è considerato fornitore del sistema di IA ad alto rischio, con obblighi specifici.[42]

Un'eccezione alla classificazione automatica come "alto rischio" è prevista solo qualora il sistema sia destinato a compiti procedurali limitati, supporti attività umane già concluse senza influenzarne significativamente gli esiti, oppure venga utilizzato per rilevare schemi decisionali senza profilazione diretta degli individui.

Tuttavia, tale esenzione non si applica ai sistemi di credit scoring, i quali, essendo fondati su processi di profilazione, come definiti all'art. 4 del GDPR, sono soggetti in ogni caso alla regolamentazione stringente dell'AI Act.

[42] Bianchini, G. (2024). *Il quadro in materia di credit scoring, tra l'AI Act e le sentenze SCHUFA*. NT+ Diritto – Il Sole 24 Ore.

### 1.3.3 Il caso SCHUFA tra GDPR e AI Act: riflessioni giuridiche su un caso pratico

Le inferenze tra credit scoring, profilazione e GDPR sono state recentemente chiarite dalla Corte di Giustizia dell'Unione Europea in merito al caso *SCHUFA*, società privata tedesca che fornisce a soggetti terzi, tipicamente istituti finanziari, informazioni relative all'affidabilità creditizia di persone fisiche e giuridiche basate su dati personali.

La Corte è stata chiamata a chiarire se il calcolo automatizzato di una probabilità di insolvenza, comunicato a una terza parte e decisivo per l'instaurazione o meno di un rapporto contrattuale, possa costituire una "decisione automatizzata" ai sensi dell'art. 22 GDPR.

Nel caso di specie, un soggetto si era visto negare un prestito da una banca a causa di un punteggio di affidabilità creditizia attribuitogli dalla società *SCHUFA*. Per capire come fosse stato calcolato tale punteggio, l'interessato aveva chiesto l'accesso ai dati personali detenuti dalla società, oltre alla cancellazione di alcune informazioni ritenute inesatte. In risposta, la società aveva comunicato il valore dello scoring (pari all'85,96%) e spiegato in termini generici il funzionamento del sistema di valutazione, ma si era rifiutata di rivelare i dati precisi e i criteri di ponderazione utilizzati, invocando il segreto commerciale. *SCHUFA* aveva inoltre sottolineato di limitarsi a fornire informazioni alle banche, senza influenzare direttamente le decisioni contrattuali assunte da queste ultime.

Il soggetto aveva dunque presentato reclamo al Commissario per la protezione dei dati e la libertà di informazione, chiedendo di imporre alla società in oggetto l'accesso completo ai dati, la rettifica delle informazioni errate e chiarimenti sulle logiche del trattamento.[43]

Tuttavia, l'autorità aveva ritenuto di non dover intervenire, dichiarando che la procedura seguita fosse conforme ai requisiti stabiliti dalla normativa tedesca sulla protezione dei dati personali (Bundesdatenschutzgesetz – BDSG).[44]

Le questioni centrali emerse nel caso SHUFA portate all'attenzione del giudice amministrativo si articolano attorno a due snodi giuridici fondamentali.

Il primo riguarda la definizione e i confini del concetto di "decisione unicamente automatizzata", che rappresenta il presupposto essenziale per l'applicazione dell'articolo

[43] Chemlali, L., & Benseddik, L. (2024). *Credit scoring under the GDPR: Insights from the CJEU's SCHUFA case*. *Journal of Data Protection & Privacy*, 7(2), 152–165.

[44] Silvano, C. (2024). *La nozione di "decisione completamente automatizzata" sotto la lente della Corte di Giustizia: il caso Schufa. CERIDAP* – Rivista interdisciplinare sul diritto delle amministrazioni pubbliche, Fascicolo 4/2024, ottobre-dicembre.

22 del GDPR. È infatti sulla base di tale qualificazione che si determina se un soggetto possa legittimamente essere sottoposto a una decisione che lo riguarda in modo significativo, senza un intervento umano sostanziale.

Il secondo tema riguarda invece chi effettivamente prende la decisione finale di concedere o meno un prestito. Nella situazione in esame, la valutazione automatica del merito creditizio è effettuata da *SCHUFA*, ma la decisione concreta, di concedere o meno un prestito, spetta a un altro soggetto, ovvero la banca. Questo "distacco" tra chi elabora il punteggio e chi prende la decisione rende più complesso stabilire se ci si trovi davvero di fronte a una "decisione automatizzata" ai sensi del GDPR.[45]

Il conflitto normativo evidenzia la complessità del quadro giuridico che regola il credit scoring automatizzato, in cui occorre bilanciare le esigenze di trasparenza e tutela dei diritti individuali con la protezione dei segreti industriali e la discrezionalità dei soggetti economici coinvolti.

La Corte di Giustizia in merito al caso ha chiarito che, per applicare l'art. 22 del GDPR, che vieta in linea di principio le decisioni basate unicamente su trattamenti automatizzati, devono essere soddisfatte tre condizioni cumulative: la presenza di una decisione, il fatto che essa sia fondata esclusivamente su un trattamento automatizzato, compresa la profilazione, e infine la produzione di effetti giuridici o comunque un impatto significativo sull'interessato.[46]

Nel caso specifico la Corte di Giustizia, il 7 dicembre 2023[47], ha ritenuto che tutte queste condizioni fossero presenti. In primo luogo, ha considerato che il calcolo dello scoring, pur essendo un dato probabilistico, possa configurare una vera e propria decisione, in quanto può incidere concretamente sulla possibilità di accedere a un contratto. In secondo luogo, ha riconosciuto che l'attività di SCHUFA costituisca profilazione, dato che il punteggio viene elaborato automaticamente in base a dati personali del soggetto valutato. Infine, ha rilevato che, se il punteggio trasmesso dalla società determina in modo decisivo

[45] Mattassoglio, F. (2025). *La Corte di giustizia europea, algoritmi e credit scoring. L'apertura del vaso di Pandora delle società che si "limitano" a elaborare gli scoring*. Dialoghi di Diritto dell'Economia, Gennaio 2025. Università degli Studi di Milano-Bicocca.

[46] Silvano, C. (2024). *La nozione di "decisione completamente automatizzata" sotto la lente della Corte di Giustizia: il caso Schufa. CERIDAP – Rivista interdisciplinare sul diritto delle amministrazioni pubbliche*, Fascicolo 4/2024, ottobre-dicembre.

[47] Corte di Giustizia dell'Unione Europea. (2023). *Sentenza del 7 dicembre 2023, C-634/21*. EUR-Lex.

la concessione o meno del credito da parte di una banca, allora produce un effetto giuridico rilevante sull'interessato.

La Corte ha inoltre precisato che eventuali deroghe previste dal diritto nazionale, come quella invocata da SCHUFA sulla base della legge federale tedesca (art. 31 BDSG), sono ammissibili solo se rispettano pienamente i requisiti posti dal GDPR, inclusa la previsione di garanzie adeguate a tutela dei diritti dell'interessato. Sarà quindi compito del giudice nazionale verificare, caso per caso, se la normativa interna sia effettivamente conforme a tali parametri.

In conclusione, il caso SCHUFA rappresenta un'importante pronuncia chiarificatrice sul perimetro applicativo dell'art. 22 GDPR, in particolare in relazione al credit scoring automatizzato. La Corte ha sottolineato la necessità di evitare che sistemi apparentemente neutri o "intermedi" eludano le garanzie offerte dalla normativa europea, riaffermando l'importanza di assicurare trasparenza, controllo umano e rispetto dei diritti fondamentali nelle valutazioni algoritmiche.

Tali esigenze trovano un ulteriore consolidamento nell'AI Act, che rafforza la responsabilità lungo l'intera catena di impiego dei sistemi ad alto rischio.

Il caso conferma quindi l'importanza di un quadro normativo integrato, capace di garantire che anche le tecnologie più evolute siano sviluppate e utilizzate nel rispetto dei principi di equità, *accountability* e tutela dei diritti fondamentali.

# Capitolo II

# Metodologie di credit-scoring

Nel corso degli anni, il credit scoring si è evoluto da semplici valutazioni manuali basate su informazioni anagrafiche e storici di pagamento a sistemi altamente automatizzati in grado di elaborare grandi quantità di dati eterogenei. In parallelo all'evoluzione normativa delineata nel primo capitolo, si è sviluppato un panorama metodologico variegato, in cui coesistono modelli statistici classici e tecniche più recenti di apprendimento automatico.
Il presente capitolo si propone di offrire una panoramica sulle principali metodologie utilizzate nella valutazione del merito creditizio. Verranno dapprima analizzati brevemente i modelli tradizionali, come la regressione logistica e l'analisi discriminante lineare, che per decenni hanno rappresentato lo standard nell'ambito delle scorecard creditizie. Queste tecniche, pur essendo ancora largamente impiegate per la loro solidità e interpretabilità, presentano limiti strutturali legati alla rigidità dei modelli, all'utilizzo di dati esclusivamente strutturati e alla scarsa capacità di adattamento a dinamiche di mercato più complesse.
Successivamente, l'attenzione sarà rivolta a modelli più sofisticati, appartenenti all'ambito del machine learning e del deep learning, come gli alberi decisionali e le reti neutrali. Questi strumenti offrono nuove potenzialità predittive, grazie alla capacità di cogliere pattern complessi e relazioni non lineari tra le variabili, ma pongono anche importanti sfide in termini di trasparenza, controllo e rischi di *bias* algoritmici.
Infine, verrà affrontato il tema dell'integrazione dei dati alternativi nei processi decisionali automatizzati, valutandone le implicazioni operative ed etiche. L'obiettivo è comprendere come le innovazioni tecnologiche e l'espansione delle fonti informative stiano ridefinendo il perimetro del credit scoring, ampliandone significativamente le opportunità ma anche i rischi per consumatori e operatori del credito.

## 2.1 Evoluzione storica del credit-scoring: scorecard logistiche e sistemi bureau

La valutazione del merito creditizio rappresenta una delle attività più centrali per le banche nella gestione delle decisioni relative alla concessione del credito. Tale processo prevede la raccolta, l'analisi e la classificazione di molteplici informazioni e variabili finanziarie, con l'obiettivo di supportare decisioni consapevoli in ambito creditizio. La

qualità del portafoglio prestiti si configura infatti come un fattore cruciale per la competitività, la stabilità e la redditività di un istituto bancario. Tra gli strumenti più efficaci per segmentare la clientela e individuare soggetti a rischio di insolvenza si colloca il credit scoring.[48]

Agli albori dell'informazione creditizia, gli operatori finanziari facevano affidamento sui credit report per decidere se concedere credito a consumatori, imprese o grandi organizzazioni. Tali report includevano informazioni demografiche, coperture assicurative e altri dati associati al soggetto richiedente, sia esso persona fisica o giuridica. Le basi teoriche del moderno credit scoring risalgono alla prima metà del Novecento, con l'introduzione dell'analisi lineare discriminante, una tecnica statistica ideata da Ronald A. Fisher negli anni '30. Un passaggio significativo avvenne nel 1941, quando David Durand[49], pioniere nell'applicazione dei modelli statistici alla finanza, intuì che tale approccio poteva essere adattato al settore creditizio per distinguere in modo oggettivo tra profili solidi e profili ad alto rischio di inadempienza.[50]

A partire dagli anni '60 e '70, il ricorso ai punteggi di credito si diffuse più rapidamente grazie alla nascita dei credit bureau e al progresso tecnologico e normativo. I credit bureau, anche noti come Sistemi di Informazioni Creditizie (SIC), sono database privati che raccolgono e gestiscono informazioni sulla posizione creditizia di soggetti beneficiari di finanziamenti. Questi archivi registrano sia dati positivi che negativi, relativi a prestiti personali, mutui, carte di credito, fidi bancari, finanziamenti finalizzati, richieste di credito rifiutate e altre informazioni utili a ricostruire lo storico creditizio di una persona fisica o giuridica. Il loro obiettivo è quello di valutare l'affidabilità creditizia di chi richiede credito, facilitando così le decisioni da parte di banche e società finanziarie. Le società aderenti ai SIC possono accedere a tali dati esclusivamente per finalità legate alla valutazione del merito e del rischio creditizio, alla verifica della puntualità nei pagamenti, nonché alla prevenzione di frodi, inclusi i tentativi di furto d'identità.

---

[48] Abdou, H. A., & Pointon, J. (2011). *Credit scoring, statistical techniques and evaluation criteria: A review of the literature.* Intelligent Systems in Accounting, Finance and Management.

[49] David Durand è stato uno dei primi studiosi a introdurre l'uso rigoroso di metodi statistici nell'ambito finanziario e in particolare del credito al consumo. Attivo tra gli anni '40 e '70, ha condotto studi empirici sull'analisi dei tassi di interesse, sul rischio di default nei prestiti al consumo e su altri fenomeni finanziari, contribuendo alla nascita del credit scoring moderno. Tratto da https://news.mit.edu/1996/durand

[50] World Bank Group. (2019). *Credit scoring: Approaches, guidelines*. Washington, DC: World Bank Group.

Nello stesso periodo, le Banche Centrali introdussero anche sistemi pubblici di informazione creditizia, paralleli a quelli privati, per monitorare il rischio a livello sistemico.[51]

I sistemi automatizzati per la valutazione dell'affidabilità creditizia si sono affermati solo in tempi relativamente recenti: fino agli anni '80, infatti, l'approvazione dei prestiti avveniva manualmente, sulla base del giudizio di analisti e funzionari, che valutavano ogni pratica singolarmente, ricorrendo talvolta a strumenti statistici.
Uno dei primi algoritmi di credit scoring fu sviluppato nel 1989 dalla Fair Isaac Corporation (FICO), segnando un passaggio decisivo verso la modellizzazione matematica del rischio e l'automazione delle decisioni creditizie. Il punteggio FICO si impose rapidamente come standard nel settore finanziario, offrendo una misura sintetica del rischio di inadempienza, calcolata su una scala da 300 a 850.

L'algoritmo prendeva in considerazione cinque aree chiave per valutare l'affidabilità creditizia di un cliente: cronologia dei pagamenti, livello di indebitamento, durata della storia creditizia, nuovi conti aperti e tipologia di credito utilizzata.

La Fair Isaac Corporation, fondata nel 1956, è tutt'oggi una delle aziende leader a livello mondiale nel campo dell'analisi predittiva e del credit scoring. Con sede negli Stati Uniti, la società ha rivoluzionato il modo in cui il rischio creditizio viene valutato, fornendo soluzioni avanzate basate su metodi statistici altamente sofisticati.

I punteggi FICO rappresentano ancora lo standard di riferimento globale per la misurazione dell'affidabilità creditizia, oltre il 90% dei principali istituti di credito negli Stati Uniti utilizza questi punteggi per guidare le proprie decisioni, incidendo su miliardi di valutazioni ogni anno.[52]

Parallelamente allo sviluppo dei primi modelli statistici, si affermarono per la prima volta le scorecard, ovvero gli strumenti operativi attraverso cui veniva applicato in modo sistematico il processo di valutazione del rischio di credito. È proprio in questo contesto che iniziarono a delinearsi i primi contorni teorici e pratici del concetto noto come credit scoring. L'innovazione delle scorecard risiedeva nella possibilità di sostituire o affiancare il giudizio discrezionale con uno strumento oggettivo, replicabile e fondato su dati storici

[51] Szepannek, G., & Lübke, K. (2021). *Facing the challenges of developing fair risk scoring models.* Frontiers in Artificial Intelligence, 4, Article 681915.

[52] https://www.myfico.com/credit-education/blog/history-of-the-fico-score

e modelli statistici, in particolare regressioni logistiche.[53] Nel tempo, hanno evoluto la loro funzione, passando da semplici strumenti predittivi a veri e propri dispositivi standardizzati, utilizzati su vasta scala da banche, società finanziarie e agenzie di credito. L'adozione diffusa delle scorecard a livello globale ha consolidato l'idea del merito creditizio come attributo oggettivamente misurabile e comparabile, favorendo la standardizzazione dei criteri valutativi.[54]

Con l'affermarsi di tali sistemi e innovazioni sopra discusse, i grandi credit bureau hanno iniziato a offrire propri sistemi di scoring, costruiti aggregando dati da fonti sia esterne che interne, per generare punteggi standardizzati di affidabilità creditizia.

Oggi, società italiane come Experian o CRIF forniscono punteggi calcolati a livello nazionale, che vengono spesso integrati nelle scorecard applicative degli enti erogatori. Le scorecard più efficaci, infatti, combinano dati dichiarati dal cliente con informazioni storiche provenienti da bureau, offrendo una visione più completa del comportamento creditizio. Questi strumenti hanno aperto la strada allo sviluppo di modelli più sofisticati, come quelli basati su machine learning, i quali, pur introducendo algoritmi predittivi avanzati, mantengono la logica di base orientata alla previsione del rischio.

## 2.2 Introduzione ai modelli di Machine learning

Prima di esaminare le tecniche di machine learning più utilizzate nel campo del credit scoring, è utile definire il concetto stesso di apprendimento automatico e approfondire le sue principali articolazioni, tra cui il deep learning. Disporre di una base concettuale consente infatti di comprendere in che modo queste tecniche possano essere applicate alla valutazione del merito creditizio, e con quali vantaggi e criticità.

Oggi, quando si parla di apprendimento automatico, o più comunemente conosciuto nella versione inglese, Machine Learning (d'ora in poi ML), si tende a considerarlo quasi come sinonimo di intelligenza artificiale in senso lato. Tuttavia, pur essendo una delle

[53] https://www.experian.it/business/analytics-and-decisioning/decision-analytics/application-scorecards#:~:text=Tutto%20ci%C3%B2%20%C3%A8%20possibile%20inizialmente,credito%20per%20gli%20istituti%20finanziari

[54] Poon, M. (2007). *Scorecards as devices for consumer credit: The case of Fair, Isaac & Company Incorporated*. The Sociological Review, 55(s2), 284–306.

metodologie più diffuse e utilizzate in molti settori, il ML è solo una parte dell'intelligenza artificiale e non la rappresenta nella sua totalità.

Il ML è una tecnica di analisi dei dati che permette ai sistemi di imparare e migliorare autonomamente attraverso l'esperienza, senza bisogno di essere esplicitamente programmati per ogni singolo compito. Questo approccio si basa su algoritmi che apprendono dai dati: durante l'addestramento, il sistema esamina esempi e identifica schemi e relazioni che non sono immediatamente evidenti.

Gli algoritmi di ML sono progettati principalmente per identificare schemi, classificare oggetti, prevedere eventi futuri e prendere decisioni sulla base di dati che non devono necessariamente essere strutturati. Questi algoritmi possono essere utilizzati singolarmente o combinati, in modo da ottenere risultati più precisi quando i dati sono complessi o difficili da prevedere, migliorando così l'affidabilità delle previsioni e delle decisioni.[55]

Il processo inizia con la raccolta e la preparazione dei dati, che possono essere numeri, immagini o testi di varia natura, provenienti da diverse fonti. La quantità e la qualità dei dati raccolti sono fattori determinanti per l'affidabilità e la precisione dei risultati ottenuti. Una volta raccolti, questi dati vengono utilizzati dai programmatori per addestrare i modelli di apprendimento automatico, che vengono configurati per rilevare schemi o fare previsioni.[56]

Un aspetto chiave di questa metodologia è la ripetitività: maggiore è l'esposizione dei modelli ai dati, maggiore è la loro capacità di adattarsi autonomamente. Con il tempo, i modelli possono essere affinati e ottimizzati, modificando i parametri per ottenere previsioni sempre più precise e accurate.

In generale, un sistema di ML può svolgere diverse funzioni, a seconda degli obiettivi che si vogliono raggiungere: può avere una funzione descrittiva, quando i dati vengono usati per spiegare eventi passati, una funzione predittiva, quando vengono impiegati per prevedere eventi futuri, o una funzione prescrittiva, quando il sistema analizza i dati e fornisce suggerimenti su quali azioni intraprendere in base alle informazioni acquisite.

---

[55] Gao, H. & Kou, G., 2024. *Machine learning in business and finance: a literature review and research opportunities*. *Financial Innovation*, 10(86).

[56] https://www.lum.it/machine-learning/

Il ML si articola in vari modelli di apprendimento, ognuno dei quali è applicabile a scenari diversi in base alla natura dei dati a disposizione e agli obiettivi da raggiungere. Questi modelli si distinguono principalmente in quattro categorie.

- Apprendimento Supervisionato: è probabilmente la tipologia di ML più diffusa. La supervisione si riferisce alla presenza di etichette nel dataset di addestramento. In questo processo, un supervisore fornisce alla macchina degli esempi pratici, in cui vengono specificate le variabili di input (x) e le variabili di output (y), contenenti il risultato corretto. La macchina apprende da questi esempi e utilizza le informazioni per costruire un modello predittivo in grado di fare previsioni su nuovi dati.

  Tra le applicazioni più comuni dell'apprendimento supervisionato ci sono la classificazione e i modelli di regressione. Nel primo caso ogni etichetta è una classe discreta che identifica il risultato atteso oppure un giudizio di valore. Nel secondo caso, dove il risultato è un valore continuo, alla macchina spetta il compito di trovare una relazione tra i valori di input, valori descrittivi, e di output.

- Apprendimento Non Supervisionato: a differenza dell'apprendimento supervisionato, nell'apprendimento non supervisionato i dati a disposizione non sono etichettati. L'algoritmo deve cercare autonomamente di scoprire pattern, strutture o correlazioni all'interno dei dati. Questo tipo di apprendimento è particolarmente utile quando non si ha una risposta "corretta" predefinita o quando è troppo costoso etichettare grandi quantità di dati.

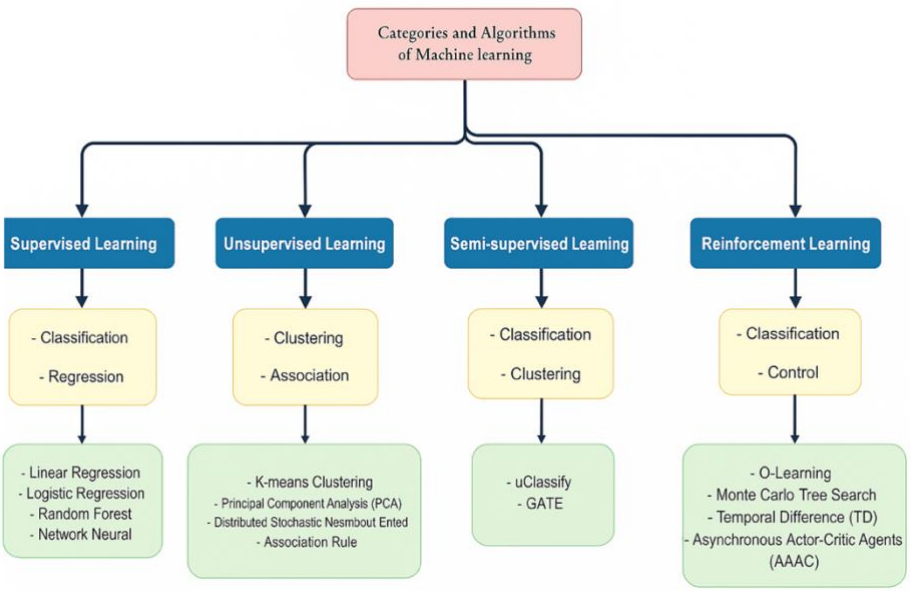
  Un esempio classico di apprendimento non supervisionato è il *clustering*, che raggruppa insieme dati simili, senza conoscere in anticipo a quali categorie appartengano.[57]

- Apprendimento Semi-Supervisionato: si colloca a metà strada tra i due metodi appena esposti. Si dispone di una piccola quantità di dati etichettati e una grande quantità di dati non etichettati. Anche in questo caso viene spesso adottato in

[57] https://www.ibm.com/it-it/topics/machine-learning

situazioni in cui è costoso o impraticabile etichettare tutti i dati. Il modello inizierà a lavorare con i pochi dati etichettati a sua disposizione e cercherà di applicare gli schemi che impara a un set di dati molto più vasto e non etichettato.
Sebbene promettente, questa tecnica non è esente da rischi: ad esempio, il modello potrebbe apprendere dai dati etichettati errori che, successivamente, riprodurrà su larga scala, portando a predizioni imprecise.

- Apprendimento per Rinforzo: è un approccio di ML che si distingue dagli altri per il suo processo per "tentativi ed errori". Invece di essere addestrato su un set di dati predefiniti, il modello apprende attraverso l'interazione con un ambiente, ricevendo un "premio" o una "penalità" in base alle sue azioni. Questo tipo di apprendimento è simile a come un essere umano apprenderebbe una nuova attività: si fa esperienza e si impara a migliorare nel tempo.




[58] *Figura 2 - Diversi modelli e algoritmi di Machine Learning*

Nel grafico appena riportato sono riportati i diversi modelli di apprendimento illustrati in precedenza, insieme alle principali funzioni che ciascun modello è progettato per svolgere.

[58] https://medium.com/@prasannathupati/types-of-algorithms-in-machine-learning-71b28a60680c

Ogni funzione è a sua volta associata agli algoritmi considerati più efficaci in quel contesto. Va precisato che molti algoritmi, con le dovute varianti, possono essere applicati a più funzioni e adattati a differenti modelli di apprendimento. Tuttavia, ai fini di una maggiore chiarezza espositiva, nelle prossime sezioni verranno descritte separatamente le diverse funzioni del ML e i principali ambiti applicativi, così da inquadrare correttamente gli algoritmi che verranno analizzati nei paragrafi successivi, selezionati in base alla loro coerenza con le esigenze specifiche della valutazione del merito creditizio.
Nei problemi di classificazione, il valore in output non è numerico continuo, ma appartiene a un insieme finito di categorie distinte, dette classi. L'obiettivo dell'algoritmo è quindi quello di assegnare ogni nuova osservazione a una di queste classi, sulla base delle caratteristiche in input. Per farlo, il modello si addestra utilizzando un dataset in cui le classi di appartenenza sono già note (dati etichettati), imparando a riconoscere schemi ricorrenti e relazioni tra le variabili. Una volta completata la fase di apprendimento, l'algoritmo sarà in grado di prevedere la categoria corretta per nuove osservazioni non viste in precedenza, replicando la logica appresa durante l'addestramento.
I modelli di regressione si distinguono per la presenza di valori di output continui, ossia numerici e non categoriali. L'obiettivo di questi modelli è prevedere una variabile quantitativa sulla base di un insieme di caratteristiche (*feature*) osservate, stimando così un valore numerico coerente con i dati di input.
I modelli di classificazione e regressione sono ampiamente utilizzati per distinguere tra clienti solvibili e non solvibili e per stimare la probabilità numerica di *default*, e sono alla base della maggior parte dei sistemi predittivi nel credit scoring.[59]
Un'altra funzione rilevante è il *clustering*, che ha l'obiettivo di identificare strutture latenti nei dati raggruppando le osservazioni in insiemi omogenei, detti cluster, senza che sia necessario specificare a priori le classi di appartenenza. Questa tecnica è frequentemente impiegata nella segmentazione della clientela o nell'analisi comportamentale.
Le regole di associazione (*association* o *ensemble*) permettono invece di scoprire relazioni ricorrenti tra insiemi di variabili, rilevando pattern frequenti all'interno dei dati.

[59] Nasteski, V. (2017). *An overview of the supervised machine learning methods.* Horizons: International Scientific Journal, 21(4).

Trovano applicazione, ad esempio, nelle strategie di cross-selling o nell'analisi combinata dei comportamenti di acquisto.

Infine, la funzione di controllo si basa sulla capacità del sistema di gestire sistemi dinamici, ovvero sistemi che evolvono nel tempo secondo regole determinate. L'obiettivo è progettare *controller* in grado di guidare il comportamento di tali sistemi verso uno stato desiderato, anche in presenza di incertezza o variabilità. Esempi tipici di sistemi dinamici comprendono robot, veicoli autonomi, aeroplani e anche processi complessi come reazioni chimiche, tutti contesti in cui è fondamentale regolare l'evoluzione del sistema in modo efficace e reattivo.

## 2.3 Metodi tradizionali per il credit scoring

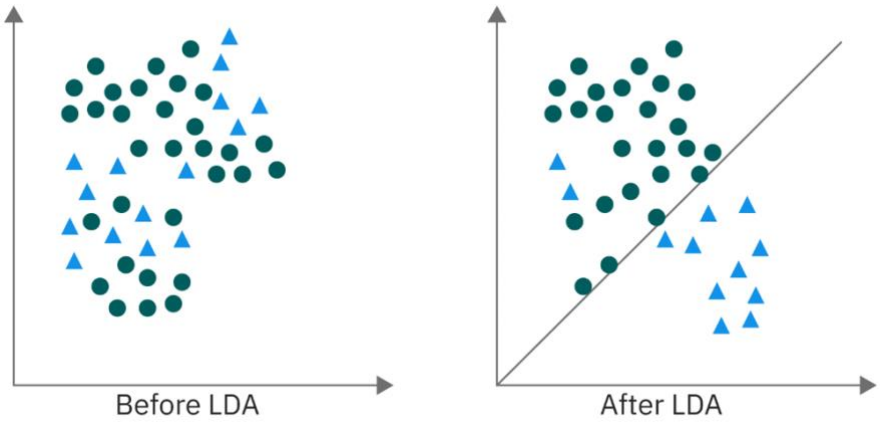
Nel I paragrafo del II capitolo si è accennato all'uso della regressione lineare e logistica, nonché dell'analisi discriminante lineare, tra i primi metodi introdotti e ancora oggi largamente impiegati nelle scorecard per la valutazione del merito creditizio. Per comprendere appieno la logica che ne sta alla base e per poter successivamente affrontare l'analisi di modelli più sofisticati e complessi, è utile proporre un quadro introduttivo e basilare sui principali algoritmi di regressione e classificazione statistica. Queste tecniche ancora oggi rappresentano alcune delle metodologie multivariate più diffuse e consolidate, in quanto consentono di individuare e descrivere l'esistenza e la natura di una relazione funzionale tra variabili all'interno di una popolazione, e sono ampiamente utilizzate per prevedere la probabilità di un determinato esito categorico a partire da più variabili esplicative.

### 2.3.1 L'Analisi Discriminante Lineare

L'Analisi discriminante lineare (LDA) ideata da Fisher, è stato uno dei primissimi algoritmi impiegati per il calcolo del credit scoring. La LDA consente di valutare la qualità creditizia delle imprese attraverso una funzione discriminante lineare che classifica i richiedenti in gruppi distinti, tipicamente inadempienti e non inadempienti, sulla base delle loro caratteristiche osservabili.[60]

[60] Antonogeorgos, G., Panagiotakos, D. B., Priftis, K. N., & Tzonou, A. (2009). *Logistic regression and linear discriminant analyses in evaluating factors associated with asthma prevalence among 10- to 12-*

Il modello mira a individuare una combinazione lineare delle variabili esplicative in grado di distinguere al meglio tra classi differenti. Più precisamente, l'LDA utilizza le dimensioni originali dello spazio dei dati per generare un nuovo asse discriminante, orientato in modo tale da massimizzare la distanza tra le medie delle classi e, contemporaneamente, minimizzare la varianza interna a ciascuna di esse. Questo processo consente di proiettare i dati, inizialmente distribuiti in uno spazio multidimensionale, lungo una sola direzione, semplificando la struttura del dataset e facilitando la separazione tra i gruppi.[61]




[62]

*Figura 3 - Rappresentazione grafica della LDA*

La distanza perpendicolare tra ciascun punto e la funzione discriminante, fornisce un'intuizione geometrica del funzionamento dell'analisi discriminante: i dati appartenenti a classi diverse risultano distribuiti in modo più netto e meno sovrapposto lungo l'asse ottenuto, migliorando la chiarezza e l'accuratezza della classificazione.
Questa tecnica è particolarmente utile anche nei casi con più di due classi, poiché è in grado di individuare più assi discriminanti ottimali, contrariamente ad approcci come la regressione logistica, che sono strutturalmente limitati alla classificazione binaria.

---

*years-old children: Divergence and similarity of the two statistical methods*. International Journal of Pediatrics.

[61] Gambella, C., Ghaddar, B., et al. (2021). *Optimization problems for machine learning: A survey.* European Journal of Operational Research, 290.

[62] https://www.ibm.com/it-it/think/topics/linear-discriminant-analysis#:~:text=L'analisi%20discriminante%20lineare%20(LDA)%20%C3%A8%20un%20approccio%20utilizzato,riduzione%20della%20dimensionalit%C3%A0%20dei%20dati.

Tuttavia, l'efficacia dell'LDA dipende dalla disuguaglianza tra le medie delle distribuzioni: in presenza di classi con medie molto simili o sovrapposte, la tecnica può fallire nel trovare un asse utile alla separazione. In tali situazioni, diventa necessario ricorrere a metodi più flessibili, come le tecniche di analisi discriminante non lineare, che permettono di catturare relazioni più complesse tra le variabili.

### 2.3.2 Regressione lineare

Nel modello di regressione si distingue fra la variabile dipendente Y e la variabile esplicativa X. La finalità del modello e spiegare il valore della variabile dipendente in funzione di quello assunto dalla variabile esplicativa. In generale il legame fra la Y e la X non è una relazione esatta. Nel modello di regressione si assume che il valore atteso della Y sia funzione del valore assunto dalla X, definendo così un legame in media fra le due variabili. In particolare, nel modello di regressione lineare, si assume un legame lineare e pertanto il valore atteso della Y si trova sulla retta di regressione.[63]

Il modello più semplice è la regressione lineare con un solo regressore:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

dove:

- $\beta_0$ costituisce l'intercetta del modello di regressione, ossia il valore atteso della Y quando X = 0.
- $\beta_1$ è il coefficiente angolare o la pendenza della retta di regressione ed esprime la variazione che subisce il valore atteso della Y per una variazione unitaria della variabile esplicativa.
- $\varepsilon$ rappresenta il termine di errore, è una variabile casuale che rappresenta lo scarto della y rispetto al suo valore atteso.

La struttura di tale modello è piuttosto intuitiva, ma può essere estesa e resa più articolata in funzione delle esigenze dell'analisi. In molti casi, per descrivere con maggiore accuratezza il comportamento della variabile dipendente Y, è opportuno includere più di una variabile esplicativa. In tali circostanze si parla di regressione lineare multipla.

[63] Monti, A. C. (2008). *Introduzione alla statistica*. Edizioni Scientifiche Italiane.

Se si indicano con k i regressori considerati, l'equazione della retta di regressione può essere espressa nella forma generalizzata:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

### 2.3.3 Regressione logistica

La regressione logistica, o modello logit, è un metodo statistico utilizzato per modellare una variabile binaria discreta, cioè che può assumere solo i valori 0 o 1. È impiegato quando l'obiettivo è stimare la probabilità che si verifichi un determinato evento (codificato come 1), dato un insieme di variabili esplicative. A differenza della regressione lineare, in cui la variabile dipendente può assumere qualunque valore reale nell'intervallo $(-\infty, +\infty)$, nel caso di una variabile dicotomica tale approccio non è adatto. La regressione logistica è progettata specificamente per gestire questo vincolo, in quanto restituisce una probabilità, cioè un valore compreso tra 0 e 1.
Il modello si basa sulla trasformazione logit della combinazione lineare dei predittori. La trasformazione logit è definita come il logaritmo naturale del rapporto tra la probabilità di successo $p$ e la probabilità di insuccesso $1 - p$, ovvero:

$$logit(p) = \ln\left(\frac{p}{1-p}\right)$$

Tale trasformazione restituisce una funzione sigmoide (a forma di "S") che garantisce che la probabilità stimata sia sempre compresa nell'intervallo $[0,1]$.
Il modello di regressione logistica assume la seguente forma:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

In questo modello, i log-odds (cioè i logaritmi del rapporto tra probabilità di successo e insuccesso) sono espressi come una combinazione lineare dei predittori.[64]

[64] Muhamedyev, R. I. (2015). Machine learning methods: An overview. *Computer Modelling & New Technologies, 19*(6).

I coefficienti $\beta$ del modello vengono stimati tramite il metodo della massima verosimiglianza (Maximum Likelihood Estimation, MLE). Questo approccio consiste nel trovare i valori dei parametri che massimizzano la funzione di verosimiglianza, ovvero la probabilità di osservare i dati campionari dati i parametri del modello. L'ottimizzazione avviene attraverso un processo iterativo, che consente di individuare i coefficienti che forniscono la migliore aderenza del modello ai dati.

Una volta stimati i parametri, è possibile calcolare la probabilità condizionata per ciascuna osservazione. Tali probabilità possono essere utilizzate per la classificazione binaria: se la probabilità prevista è inferiore a 0,5 si prevede l'esito 0, se è superiore a 0,5 si prevede l'esito 1.[65]

### 2.3.4 Alberi decisionali

Nel seguente paragrafo verrà illustrato il funzionamento e l'utilità di un altro modello statistico ampiamente impiegato nell'ambito del credit scoring, gli alberi decisionali. Sebbene presentino alcune analogie con le tecniche impiegate nella regressione gli alberi decisionali si configurano come una metodologia più flessibile e avanzata sotto diversi aspetti.

Le origini d questo modello risalgono al lavoro svolto da Belson negli anni '50, in particolare ai suoi studi condotti per analizzare le indagini di ascolto su scala nazionale per conto della British Broadcasting Corporation (BBC).

In un'epoca priva di calcolatori digitali, questi modelli si adattavano perfettamente alla tecnologia meccanica dell'epoca, grazie alla loro struttura sequenziale e alla possibilità di applicare procedure ricorsive senza ricorrere a calcoli matriciali. L'approccio si basava sull'identificazione di squilibri tra frequenze osservate e attese all'interno di sottoclassi predittive, un metodo che, pur semplice, costituì il fondamento per tutte le successive implementazioni degli alberi decisionali.

Un'importante innovazione proposta da Belson riguardava l'analisi ricorsiva dei nodi discendenti, che potevano essere suddivisi non solo con lo stesso predittore, ma anche con predittori diversi, dando luogo a strutture bilanciate o sbilanciate in funzione della

[65] https://www.ibm.com/it-it/think/topics/logistic regression#:~:text=La%20regressione%20logistica%20stima%20la,di%20dati%20di%20variabili%20indipendenti.&text=In%20questa%20equazione%20di%20regressione,x%20%C3%A8%20la%20variabile%20indipendente.

forza esplicativa delle variabili. Questo schema rimane alla base della logica di ramificazione selettiva utilizzata ancora oggi.

Negli anni successivi, Morgan e Sonquist ripresero e ampliarono questi concetti, mostrando come gli alberi decisionali potessero rappresentare una valida alternativa ai modelli di regressione, soprattutto per la capacità di individuare automaticamente le interazioni tra le variabili, senza la necessità di specificarle a priori. Nei loro esperimenti, un albero decisionale che suddivideva i dati in 21 gruppi spiegava oltre il 66% della varianza della variabile risposta, contro il 36% di una regressione con 30 termini, comprese le interazioni.

Nonostante lo scetticismo iniziale della comunità statistica, più incline a metodi con solide basi teoriche come la regressione, lo sviluppo successivo di questi modelli, grazie ai contributi di Kass, Hawkins e Breiman, ha consolidato gli alberi decisionali come strumenti efficaci e flessibili per l'analisi predittiva. La loro capacità di adattarsi a strutture dati complesse, unita a una logica interpretabile, li ha resi nel tempo uno dei metodi più apprezzati anche in ambito applicativo.[66]

Si tratta di modelli di apprendimento supervisionato non parametrici[67] utilizzati per problemi sia di classificazione sia di regressione (Classification And Regression Trees – CARTs), tali strutture decisionali sono note per la loro facilità di interpretazione da parte degli esseri umani.

Sono caratterizzati da una struttura gerarchica ad albero con un nodo radice da cui si diramano nodi interni e foglie terminali. Il modello si costruisce con un approccio ricorsivo *divide et impera*[68]: a partire dal nodo radice, i dati vengono suddivisi iterativamente in sottoinsiemi sempre più omogenei rispetto alla variabile target, creando

---

[66] de Ville, B. (2013). *Decision trees*. WIREs Computational Statistics, 5(6).

[67] I test non parametrici, spesso detti anche test "distribution-free", si basano su un numero ridotto di assunzioni: ad esempio, non richiedono che i dati seguano una distribuzione normale. Al contrario, i test parametrici si fondano su distribuzioni di probabilità specifiche e prevedono la stima di parametri chiave della distribuzione, come la media o la differenza tra medie, a partire dai dati del campione.
Il vantaggio dei test non parametrici è la loro maggiore flessibilità; tuttavia, questa semplicità ha un costo: in generale, sono meno potenti dei test parametrici, cioè hanno una minore probabilità di rifiutare correttamente l'ipotesi nulla quando l'alternativa è vera. Definizione tratta da Sullivan, L. (2020). *Nonparametric tests*. Boston University School of Public Health.

[68] In informatica gli algoritmi che vengono scomposti ai fini risolutivi vengono chiamati anche *Divide et impera.* Il termine si riferisce alla scomposizione di un problema grande in tanti problemi più piccoli al fine di risolverli singolarmente e ricomporli come un unico problema risolto. Tratto da https://startingfinance.com/approfondimenti/il-problem-solving/

a ogni passo nuovi nodi fino a raggiungere foglie che rappresentano risultati il più possibile puri.

[69]

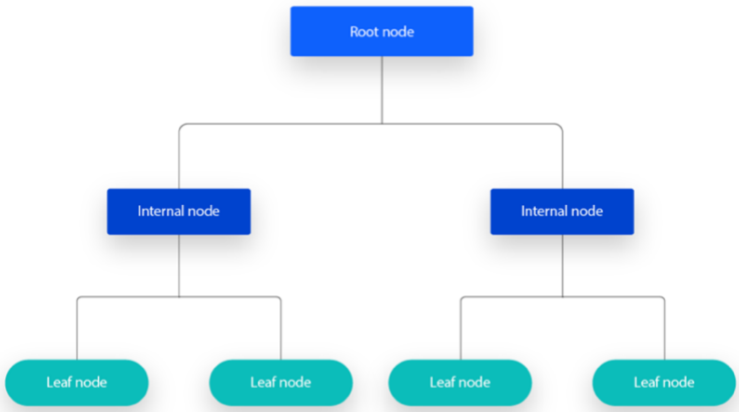



*Figura 4 - Struttura grafica di un albero decisionale*

A ciascun nodo, l'algoritmo seleziona la migliore condizione di split in base a metriche statistiche come *l'information gain*[70], derivante dal concetto di entropia, scegliendo la variabile e la soglia che massimizzano la riduzione dell'impurità nelle partizioni successive. Il risultato finale è un insieme di regole decisionali *if-then* che conduce a una previsione in ogni nodo foglia.

Rispetto alle tradizionali scorecard basate su regressione logistica, gli alberi decisionali rappresentano un'evoluzione per la maggiore flessibilità nel modellare i dati. Un modello logistico stima la probabilità di default come combinazione lineare di variabili indipendenti ed è apprezzato per la trasparenza nel mostrare l'effetto di ogni variabile. Un albero decisionale, invece, non assume una forma funzionale predefinita, ma segmenta lo spazio delle osservazioni attraverso condizioni sequenziali sulle variabili.

In ambito creditizio, applicare un albero decisionale alla costruzione di una scorecard significa ottenere una serie di regole di decisione interpretabili che suddividono il portafoglio clienti in gruppi di rischio omogenei. Ogni foglia corrisponde tipicamente a un certo profilo di rischio con una previsione associata (ad esempio, esito "buon pagatore"

[69] https://www.ibm.com/it-it/think/topics/decision-trees

[70] L'Information Gain (IG), ovvero guadagno di informazione, è una misura utilizzata negli alberi decisionali per valutare quanto un determinato attributo sia efficace nel suddividere un insieme di dati in classi distinte. In termini tecnici, quantifica la riduzione dell'entropia, cioè dell'incertezza associata alla variabile target (le etichette di classe), che si ottiene quando il valore di una specifica variabile predittiva è noto. Tratto da. Kononenko, I., & Kukar, M. (2007). *Measures for evaluating the quality of attributes.* In I. Kononenko & M. Kukar (Eds.), Machine learning and data mining.

vs “cattivo pagatore” o una probabilità di default stimata per quel segmento). Tali output possono essere convertiti in punteggi di credito analogamente a quanto avviene con la regressione logistica, ma ottenuti attraverso le regole adattative dell’albero invece che tramite una formula lineare predefinita.

Dal punto di vista dei benefici, gli alberi decisionali offrono notevoli vantaggi in termini di interpretabilità e flessibilità. Si tratta di modelli *white-box* il cui funzionamento è facilmente comprensibile: la struttura ad albero funge da diagramma di flusso che rende trasparenti i criteri alla base delle decisioni, permettendo di risalire intuitivamente al perché una certa istanza sia classificata come buona o cattiva.

Inoltre, richiedono una preparazione minima dei dati: possono gestire direttamente variabili di natura sia numerica sia categorica senza necessità di trasformazioni come la normalizzazione o la creazione di variabili *dummy*. Dal punto di vista predittivo, la capacità di suddividere l’universo dei clienti in modo mirato consente di cogliere pattern complessi nei dati, garantendo spesso una buona accuratezza nella stima del rischio di credito.

### 2.3.5 Limiti nei modelli tradizionali

Nonostante la loro diffusione e utilità, modelli di credit scoring, basati sui metodi statistici tradizionali appena discussi, presentano alcuni limiti intrinseci. In primo luogo, questi tipicamente utilizzano un insieme relativamente limitato di dati strutturati, spesso circoscritto a informazioni demografiche e storico-creditizie del cliente.[71]

Questo approccio funziona bene in contesti stazionari e con popolazioni di clientela note, ma può risultare meno efficace quando i dati diventano molto voluminosi, eterogenei o complessi.

I modelli logistici, infatti richiedono alcune assunzioni semplificative, ad esempio l’indipendenza tra variabili esplicative, che possono limitarne la capacità di cogliere pattern più sottili o interazioni non lineari presenti in basi dati di grande dimensione.[72]

[71] World Bank Group. (2019). *Credit scoring: Approaches, guidelines*. Washington, DC: World Bank Group.

[72] Markov, A., Seleznyova, Z., & Lapshin, V. (2022). *Credit scoring methods: Latest trends and points to consider*. The Journal of Finance and Data Science.

Inoltre, le scorecard tradizionali sono spesso progettate su dati storici di pagamento e indicatori finanziari classici, ciò comporta che un cliente con una scarsa storia creditizia (i cosiddetti thin files, ad esempio un giovane al primo prestito) risulti "*non scorable*" per mancanza di dati, impedendogli di ottenere credito anche se sarebbe meritevole.
Il requisito di uno storico minimo è una barriera intrinseca del credit scoring tradizionale basato su dati limitati: in assenza di informazioni pregresse, il modello non può generare alcun punteggio, escludendo dal mercato del credito soggetti potenzialmente affidabili ma *informationally opaque*. Un ulteriore limite riguarda la necessità di un forte intervento umano nella loro costruzione. Per sviluppare un modello efficace, è spesso richiesto che un esperto definisca manualmente le categorie delle variabili, selezioni i predittori più rilevanti e assegni dei pesi basandosi su esperienza o vincoli normativi. Questo introduce potenziali *bias* umani in fase di progettazione e rende i modelli meno adattivi ai cambiamenti: se mutano i comportamenti dei debitori o emergono nuovi fattori di rischio, uno sistema rigido potrebbe non catturare rapidamente le nuove tendenze senza un aggiornamento manuale del modello.
Infine, i modelli tradizionali, pur essendo relativamente semplici, non sono esenti da problemi di robustezza: se sviluppati su campioni poco rappresentativi o dati scarsamente qualitativi, possono manifestare *selection bias* o altri errori statistici che ne minano la possibilità di estendere e generalizzare il concetto**.**
Più in particolare, per quanto riguarda gli alberi decisionali discussi in precedenza, è opportuno evidenziare alcune limitazioni specifiche. Innanzitutto, questi modelli tendono facilmente all'*overfitting*: se l'albero cresce troppo in profondità o contiene rami poco significativi, può adattarsi eccessivamente ai dati storici, perdendo capacità predittiva su nuovi clienti. Questo problema viene spesso gestito attraverso tecniche di "potatura", che limitano o correggono la complessità dell'albero. Un ulteriore aspetto critico è la sensibilità alle variazioni nei dati: anche piccoli cambiamenti nel dataset possono generare strutture decisionali profondamente diverse, compromettendo la stabilità del modello. Per contenere questa instabilità, si utilizza spesso il *bagging*, che combina le previsioni di più alberi, ma tale approccio può portare a risultati ridondanti se gli alberi generati sono troppo simili tra loro.[73]

[73] Gunnarsson, B. R., vanden Broucke, S., Baesens, B., Óskarsdóttir, M., & Lemahieu, W. (2021). *Deep learning for credit scoring: Do or don't?* European Journal of Operational Research, 295(1).

## 2.4 Modelli di ensemble learning per il credit scoring

*L'ensemble learning* è un paradigma del ML secondo cui si addestrano più modelli (*base learners*) per risolvere lo stesso problema, combinandone poi le previsioni per ottenere una decisione finale più accurata e robusta. A differenza degli approcci tradizionali che si basano su un'unica ipotesi appresa dal dataset di addestramento, i metodi ensemble costruiscono un insieme di ipotesi diverse che vengono integrate in un'unica struttura predittiva più solida e affidabile.

Tra i primi contributi teorici si ricordano gli studi di Dasarathy e Sheela (1979), che proposero la suddivisione dello spazio delle caratteristiche mediante più classificatori, e di Hansen e Salamon (1990), i quali dimostrarono che la capacità di generalizzazione di una rete neurale artificiale poteva essere migliorata combinando un insieme di reti simili. Nello stesso periodo, Schapire (1990) introdusse il concetto che un classificatore forte poteva emergere dalla combinazione di modelli deboli, gettando le basi per la tecnica del *boosting*. A partire da questi lavori pionieristici, *l'ensemble learning* ha conosciuto un'evoluzione rapida, dando origine a un'ampia gamma di approcci applicativi e teorici.[74]

Uno dei motivi principali del successo dei modelli *ensemble* risiede nella loro superiore capacità di generalizzazione rispetto ai modelli isolati. Perché sia efficace, è necessario che i modelli base soddisfino due condizioni fondamentali: innanzitutto, devono possedere un livello di accuratezza almeno superiore a quello di un classificatore casuale, altrimenti il loro contributo all'*ensemble* sarebbe poco significativo; in secondo luogo, devono offrire diversità, ossia commettere errori su istanze differenti, così che possano compensarsi a vicenda nel processo di aggregazione. Quando questi requisiti sono rispettati, l'*ensemble* può superare i limiti predittivi dei modelli individuali e produrre risultati più stabili e affidabili.[75]

Nel ML applicato al credit scoring, questi metodi si sono dimostrati particolarmente utili in quanto permettono di migliorare sensibilmente la qualità delle previsioni, sfruttando appieno la complessità informativa dei dati disponibili. L'idea alla base è paragonabile al confronto tra più esperti prima di una decisione rilevante: ciascun modello base

[74] Abellán, J., & Castellano, J. G. (2017). *A comparative study on base classifiers in ensemble methods for credit scoring.* Expert Systems with Applications, 73.

[75] Wang, G., Hao, J., et al. (2011). *A comparative assessment of ensemble learning for credit scoring*. Expert Systems with Applications, 38(1).

contribuisce con la propria prospettiva e l'*ensemble* ne sintetizza i giudizi, giungendo a una valutazione complessiva più solida.

Numerose evidenze teoriche ed empiriche hanno dimostrato che un *ensemble* ben progettato consente di migliorare significativamente accuratezza e robustezza delle previsioni, qualità fondamentali nel contesto del credit scoring, dove anche lievi miglioramenti nelle metriche predittive possono generare benefici economici tangibili per istituzioni finanziarie e creditori.

L'utilizzo dei metodi *ensemble* si inserisce naturalmente nel processo di evoluzione del credit scoring: dalla regressione logistica e dall'analisi discriminante lineare, si è passati progressivamente a modelli di apprendimento più complessi e flessibili, capaci di adattarsi a basi dati di dimensioni maggiori, con caratteristiche eterogenee e relazioni non lineari. In questo scenario, tre approcci si sono distinti per diffusione ed efficacia.
[76]

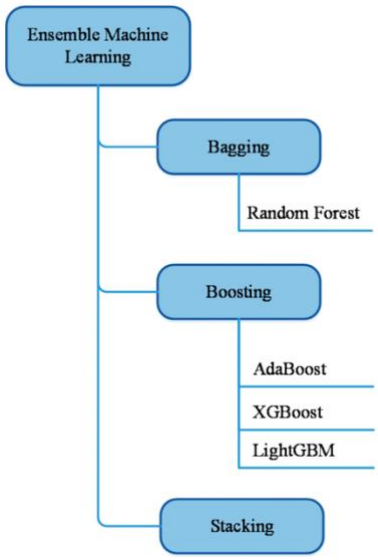



*Figura 5 - Rappresentazione grafica dei modelli ensemble*

Il primo è il *bagging*, una tecnica che aggrega parallelamente le previsioni di più modelli addestrati su versioni ri-campionate del dataset originale, riducendo così la varianza complessiva del sistema. Il secondo è il *boosting*, che addestra sequenzialmente i modelli base, focalizzandosi a ogni iterazione sugli errori commessi dal modello precedente: in questo modo si riesce a ridurre progressivamente il *bias* del sistema, migliorandone la

[76] Li, Y., & Chen, W. (2020). *A comparative performance assessment of ensemble learning for credit scoring*. Mathematics, 8(10).

precisione. Il terzo approccio è lo *stacking*, che combina modelli anche eterogenei, ad esempio, alberi, reti neurali e modelli lineari, attraverso un *meta-learner*, ossia un modello di secondo livello addestrato a integrare le predizioni dei modelli sottostanti.
Nel contesto del credit scoring, le tecniche di ensemble più diffuse e consolidate restano il *bagging*, di cui la *Random Forest* rappresenta l'applicazione più nota e utilizzata, e il *boosting*, spesso implementato tramite algoritmi di *gradient boosting* come *XGBoost*. Questi metodi si sono dimostrati particolarmente efficaci nella previsione del rischio di credito, superando frequentemente in accuratezza i modelli tradizionali.
Lo *stacking* è meno adottato nella pratica corrente a causa della sua maggiore complessità teorica e implementativa, per questo motivo i paragrafi successivi saranno dedicati all'analisi dei primi due approcci citati, approfondendone il funzionamento, i vantaggi specifici, le applicazioni pratiche nel credit scoring e i principali limiti operativi.[77]

### 2.4.1 Il bagging

Il *bagging* (*bootstrap aggregating*), introdotto da Breiman nel 1996, è uno dei primi e più noti algoritmi di *ensemble learning*. Si tratta di una tecnica intuitiva e semplice da implementare, ma al tempo stesso sorprendentemente efficace, in particolare per la sua capacità di ridurre la varianza dei modelli e contenere il rischio di *overfitting*. Sebbene venga comunemente associato agli alberi decisionali, può essere applicato con qualunque tipo di classificatore.
Il *bagging* è un metodo *ensemble* a struttura parallela (*parallel ensemble*), in cui più modelli base vengono costruiti simultaneamente. Ciascun modello apprende indipendentemente a partire da un sottoinsieme del dataset originale, estratto tramite campionamento *bootstrap*. Questa indipendenza tra i modelli consente di migliorare la stabilità e l'affidabilità del risultato finale, grazie alla combinazione delle predizioni individuali in un'unica uscita aggregata. Un ulteriore vantaggio della struttura parallela è la sua efficienza computazionale: i modelli possono essere addestrati in parallelo, riducendo sensibilmente il tempo complessivo di apprendimento.

[77] Tripathi, D., Shukla, A. K., et al. (2022). *Credit scoring models using ensemble learning and classification approaches: A comprehensive survey.* Wireless Personal Communications, 123.

[78]

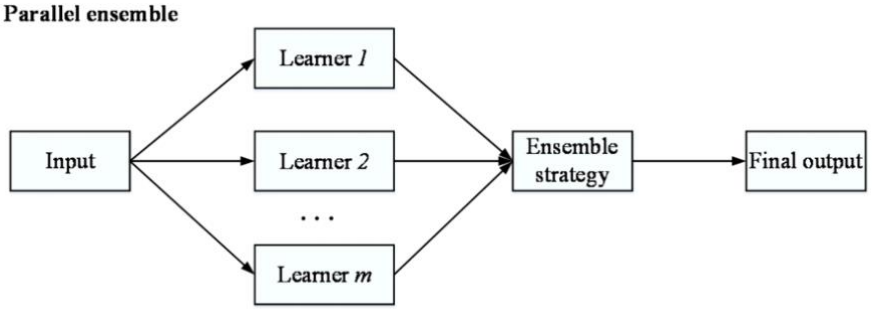



*Figura 6 - Rappresentazione grafica di un modello bagging*

Poiché ciascun classificatore viene costruito su un campione leggermente diverso, si ottiene una certa variabilità strutturale tra i modelli: le frontiere decisionali generate non sono identiche, il che introduce diversità nell'*ensemble*, elemento chiave per il buon funzionamento del metodo. Le predizioni dei modelli base vengono infine combinate tramite voto di maggioranza (per problemi di classificazione) o media (per regressione), producendo così una decisione finale più stabile e affidabile rispetto a quella ottenuta da un singolo modello.

Il *bagging* è particolarmente vantaggioso quando si lavora con dataset di dimensioni contenute. Per garantire che ogni sotto-campione sia rappresentativo, si estraggono porzioni consistenti del dataset (in genere dal 75% al 100%). Questo comporta una notevole sovrapposizione tra i diversi set di training, con molte osservazioni ripetute sia tra campioni differenti che all'interno dello stesso. Per assicurare che i modelli generati siano comunque diversi tra loro, è importante utilizzare modelli base instabili, ovvero modelli che siano sensibili anche a piccole variazioni nei dati di input. Gli alberi decisionali, in particolare, si prestano bene a questo scopo, poiché tendono a cambiare struttura anche per minime modifiche nel dataset.

Grazie a questa logica, il *bagging* consente di creare un *ensemble* di classificatori eterogenei che, aggregando previsioni leggermente differenti, compensano reciprocamente i propri errori e migliorano la robustezza complessiva del modello. Proprio per questa sua semplicità ed efficacia, il *bagging* rappresenta una delle strategie più consolidate per potenziare le prestazioni predittive, soprattutto in presenza di modelli soggetti ad alta varianza.

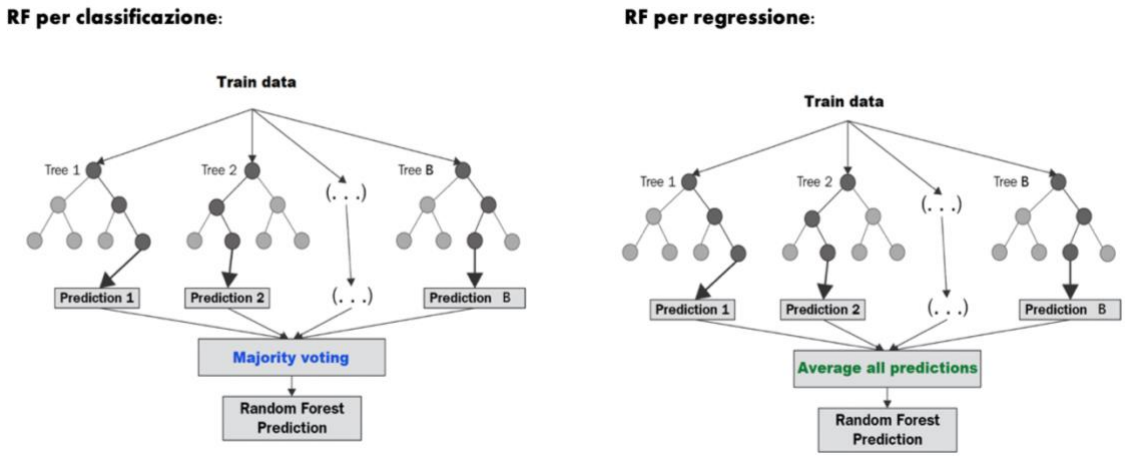
[78] Li, Y., & Chen, W. (2020). *A comparative performance assessment of ensemble learning for credit scoring.* Mathematics, 8(10).

Le *Random Forest* rappresentano un'evoluzione diretta degli alberi decisionali e una delle applicazioni più efficaci del metodo di *bagging*.

In questo approccio, l'*ensemble* è costituito da una moltitudine di alberi decisionali costruiti in parallelo, ognuno dei quali viene addestrato su un campione diverso del dataset originario ottenuto tramite *bootstrap*, ovvero estrazione casuale con reinserimento.

A ogni *split* all'interno di ciascun albero, la Random Forest seleziona casualmente solo un sottoinsieme di variabili tra quelle disponibili, limitando così le caratteristiche considerate nella costruzione di ciascun nodo. Questo introduce ulteriore diversità tra gli alberi, rendendo l'*ensemble* più robusto e meno soggetto a *overfitting*. I singoli alberi sono indipendenti tra loro e non correlati: ciò significa che errori sistematici di un albero difficilmente si ripresentano in altri, rafforzando la capacità correttiva del modello complessivo.

La decisione finale viene ottenuta combinando le predizioni di tutti gli alberi: per i problemi di classificazione, come nel credit scoring, si applica il voto di maggioranza, mentre per la regressione si utilizza la media dei risultati. Questo processo aggregato consente di migliorare sensibilmente l'accuratezza e la stabilità predittiva rispetto a un singolo albero.[79]




[80] *Figura 7 - Rappresentazione grafica di Random Forest*

[79] Ha Van Sang, Nguyen Ha Nam, & Nguyen Duc Nhan. (2016). *A novel credit scoring prediction model based on feature selection approach and parallel random forest.* Indian Journal of Science and Technology, 9(20).

[80] https://fhernanb.github.io/libro_mod_pred/rand-forests.html

La costruzione di ciascun albero all'interno di una *Random Forest* si basa su una metrica di impurità nota come indice di Gini, calcolato al fine di selezionare lo split ottimale a ogni nodo.

L'indice di Gini, calcolato in un nodo interno dell'albero decisionale, è definito come segue: per un attributo candidato allo split $X_i$, si considerano i possibili livelli $L_1, \dots L_k$. L'indice di Gini per questo attributo si calcola come:

$$G(X_i) = \sum_{k=1}^{K} \Pr(X_i = L_k)(1 - \Pr(X_i = L_k))$$

$$= 1 - \sum_{k=1}^{K} Pr(X_i = L_k)^2$$

Valori più bassi indicano maggiore purezza del nodo. Il numero massimo di livelli o profondità dei rami è regolato da un parametro $d$, definito durante la fase di addestramento.[81]

Nel contesto del credit scoring, le Random Forest si sono affermate come uno degli algoritmi più performanti per diversi motivi:

- funzionano in modo efficiente anche su dataset di grandi dimensioni;
- sono particolarmente resistenti al problema dell'*unbalanced data*, cioè quando la distribuzione tra clienti affidabili e insolventi è fortemente sbilanciata;
- gestiscono efficacemente dati mancanti, senza compromettere significativamente la qualità delle predizioni;
- sono poco sensibili alla multicollinearità, rendendo inutile la rimozione preventiva di variabili correlate.

Infine, le Random Forest supportano anche la stima della *feature importance*, cioè l'identificazione delle variabili più rilevanti nel determinare la previsione. Questo aspetto è di particolare utilità per i modelli di credit scoring, dove la trasparenza delle decisioni automatizzate e la possibilità di interpretare i fattori di rischio sono elementi cruciali per banche, regolatori e consumatori.

[81] Zhu, L., Qiu, D., Ergu, D., Ying, C., & Liu, K. (2019). *A study on predicting loan default based on the random forest algorithm.* In Procedia Computer Science, 162

### 2.4.2 Il boosting

Il *boosting* è un metodo di *ensemble learning* che, diversamente dal *bagging*, procede in modo sequenziale e adattativo. Anziché addestrare simultaneamente diversi modelli indipendenti su dataset ri-campionati, il *boosting* costruisce iterativamente una serie di modelli, ciascuno focalizzato sulla correzione degli errori commessi dal precedente.

82

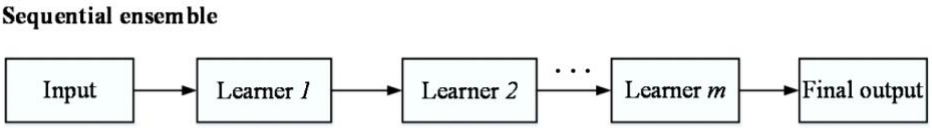



*Figura 8 - Rappresentazione grafica di un modello boosting*

Inizialmente il primo modello viene addestrato sul dataset completo con pesi uniformi; nelle iterazioni successive, tali pesi vengono aggiornati in modo da dare maggiore enfasi alle osservazioni classificate erroneamente. Di conseguenza, ciascun modello successivo pone maggiore attenzione ai casi più difficili da gestire, migliorando gradualmente l'accuratezza dell'*ensemble* finale. La combinazione delle previsioni dei singoli modelli avviene generalmente attraverso una somma pesata, in cui i pesi riflettono l'accuratezza dei modelli di base.

Un esempio storico di successo del *boosting* è *AdaBoost* (Freund e Schapire, 1996), che utilizza modelli base molto semplici, come gli alberi decisionali a profondità ridotta, aggiornando iterativamente i pesi secondo una formula analitica che minimizza progressivamente l'errore di classificazione. *AdaBoost* ha dimostrato che combinando molti modelli deboli è possibile ottenere prestazioni competitive con modelli singoli più complessi e accuratamente calibrati.[83]

Nel tempo, il *boosting* ha subito ulteriori evoluzioni, dando vita al *gradient boosting* (Friedman, 2001). Questo approccio interpreta la costruzione dei modelli come un problema di ottimizzazione numerica: ad ogni iterazione, si aggiunge un nuovo modello che apprende direttamente dai residui, cioè dagli errori commessi dal modello precedente, muovendosi nella direzione indicata dal gradiente della funzione obiettivo da

[82] Li, Y., e Chen, W. (2020). *A comparative performance assessment of ensemble learning for credit scoring.* Mathematics, 8(10).

[83] Abellán, J., Castellano, J. G. (2017). *A comparative study on base classifiers in ensemble methods for credit scoring.* Expert Systems with Applications, 73.

minimizzare. Tale funzione può essere selezionata in base alle specifiche esigenze applicative.

Un'implementazione avanzata del *gradient boosting*, divenuta popolare in ambito creditizio, è XGBoost (*eXtreme Gradient Boosting*, Chen e Guestrin, 2016). Questo modello introduce una serie di miglioramenti significativi rispetto al *gradient boosting* tradizionale, come una gestione efficiente della memoria, tecniche di regolarizzazione per limitare la complessità dei modelli base e ridurre il rischio di *overfitting*, e, aspetto distintivo, la possibilità di sfruttare l'elaborazione parallela nella costruzione degli alberi. Quest'ultima caratteristica, tipica dei metodi di *bagging*, consente a XGBoost di ridurre sensibilmente i tempi di addestramento, rendendolo particolarmente adatto a operare su grandi dataset con elevata accuratezza predittiva.

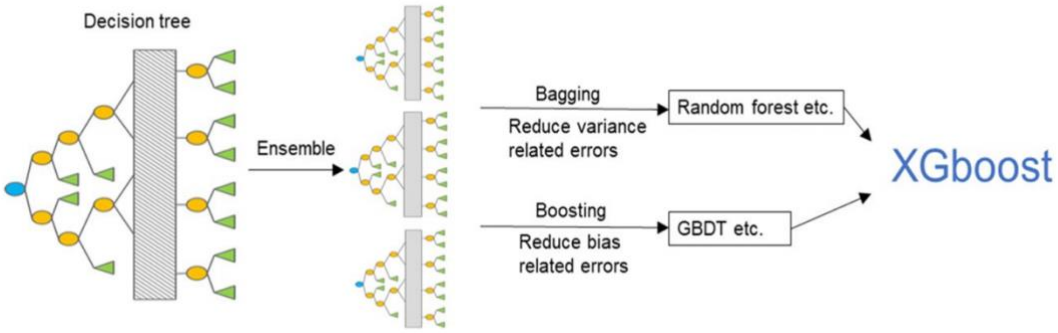



*Figura 9 - Rappresentazione schematica del modello XGBoost.*

[84]

Nel credit scoring, XGBoost si è affermato come uno degli strumenti più performanti. Grazie alla sua capacità di gestire interazioni complesse e relazioni non lineari tra variabili, questo algoritmo supera spesso i metodi tradizionali nella previsione del rischio di default, consentendo alle istituzioni finanziarie di migliorare contemporaneamente i tassi di approvazione dei prestiti e la qualità complessiva del portafoglio.[85] Tra le sue funzionalità più rilevanti per il credit scoring figurano la gestione automatica dei dati mancanti, *l'early stopping* per prevenire sovradattamento, e il *subsampling* dei dati per garantire maggiore robustezza.

[84] Jiang, J., Pan, H., et al. (2021*). Predictive model for the 5-year survival status of osteosarcoma patients based on the SEER database and XGBoost algorithm.* Scientific Reports, 11(1).

[85] Wang, H., Li, C., et al. (2019). *Does AI-based credit scoring improve financial inclusion?* Evidence from online payday lending. In 40th International Conference on Information Systems, ICIS *2019* (40th International Conference on Information Systems, ICIS 2019). Association for Information Systems.

Nonostante i suoi benefici, XGBoost richiede un'accurata calibrazione degli iperparametri: numero di alberi, profondità massima, termini di regolarizzazione, per ottimizzare le prestazioni predittive. Inoltre, come molti modelli *ensemble*, presenta una limitata interpretabilità. Ciò rende necessario il ricorso a tecniche di *Explainable AI* per soddisfare i requisiti regolamentari di trasparenza nelle decisioni creditizie.

### 2.4.3 Limiti nei modelli di ensemble learning

I metodi *ensemble* offrono indubbi vantaggi in termini di performance predittiva, ma presentano anche alcuni limiti e trade-off che assumono particolare rilievo nel settore del credito regolamentato. Uno dei punti deboli più discussi è la scarsa interpretabilità dei modelli *ensemble*. Tecniche come *Random Forest* o XGBoost sono spesso considerate "modelli opachi" o *black box*[86], in quanto non forniscono in modo diretto una giustificazione comprensibile delle decisioni prese. Questo contrasta con la necessità, sentita sia dai regolatori sia dalle istituzioni finanziarie, di poter spiegare le valutazioni di merito creditizio ai clienti e di documentare i criteri di decisione. In contesti bancari, normative come la Consumer Credit Directive 2023/2225UE e il GDPR richiedono che, se una decisione di affidamento viene presa tramite sistemi automatizzati, il consumatore abbia diritto a ottenere una "spiegazione significativa" circa la logica utilizzata dal modello, oltre che l'intervento umano in caso di contestazione.[87] Chiaramente fornire tale spiegazione è molto più complesso per un *ensemble* di centinaia di alberi o reti neurali rispetto a un modello lineare o a un singolo albero decisionale. Per questo motivo, alcune banche e FinTech adottano un approccio ibrido: sfruttano la potenza predittiva degli *ensemble* internamente, ma distillano poi le risultanze in punteggi più semplici o in regole decisionali comprensibili per la comunicazione verso l'esterno.[88]

---

[86] Nel contesto delle scienze, dell'informatica e dell'ingegneria, i termini *black box*, gray box e white box sono utilizzati per descrivere differenti livelli di opacità rispetto alla natura interna di un componente. In particolare, un componente black box non rivela alcuna informazione circa il proprio design interno, la struttura o l'implementazione; un white box è totalmente trasparente. Tra i due, i gray box offrono accesso parziale di grado variabile ai dettagli. Tratto da Adadi, A., e Berrada, M. (2018). *Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)*. *IEEE Access*, *6*, 52138–52160.

[87] Bonaccorsi di Patti, E., Calabresi, F., et al. (2022). *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano* (Questioni di Economia e Finanza, n. 721). Banca d'Italia.

[88] World Bank Group. (2019). *Credit scoring: Approaches, guidelines*. Washington, DC: World Bank Group.

In parallelo, la ricerca su *XAI* (*eXplainable AI*) sta producendo metodi per interpretare modelli complessi, e persino le big tech come Google e IBM stanno investendo per rendere gli algoritmi di ML più interpretabili e inclusivi, segno della rilevanza di questo tema anche al di fuori dell'ambito accademico.

Un secondo ordine di considerazioni riguarda i costi computazionali e di implementazione. I modelli *ensemble* richiedono infrastrutture dati e potenza di calcolo ben maggiori rispetto a modelli tradizionali. Mentre per grandi banche ciò non rappresenta un grosso ostacolo, in mercati emergenti o per piccoli intermediari finanziari le risorse necessarie per implementare i metodi più innovativi possono essere limitate.

Inoltre, l'adozione di modelli sofisticati comporta la necessità di un solido sistema di gestione e controllo del modello: le banche devono dotarsi di regolamenti interni per la validazione indipendente dei modelli di scoring, per il monitoraggio continuo delle loro prestazioni e per la gestione del rischio.[89]

Organismi come la BCE e la FED enfatizzano l'importanza di processi di gestione solidi quando si impiegano strumenti di AI nel credito, inclusa la verifica della solidità concettuale, dei dati usati, e dei possibili impatti su clienti e stabilità finanziaria. Nel caso di modelli *ensemble*, ciò implica test rigorosi per scongiurare *bias* discriminatori, ad e esempio, verificare che il modello non penalizzi sistematicamente gruppi protetti, e per garantire la robustezza *out-of-sample*, un *ensemble* troppo su misura dei dati storici potrebbe non generalizzare a nuovi contesti.

Per mitigare questo rischio si applicano tecniche come la *cross-validation*, si riservano *test set* per misurare l'*overfitting*, e si utilizzano regolarizzazioni come quelle integrate in XGBoost. In aggiunta, nei contesti regolamentati come quello bancario si devono considerare le implicazioni delle decisioni automatizzate sul piano etico e legale: un modello di credit scoring, anche se accurato, potrebbe essere giudicato inaccettabile se le sue decisioni non sono spiegabili o se crea disparità di trattamento non giustificabili. Ad esempio, l'European Banking Authority ha richiamato l'attenzione sul fatto che algoritmi di ML opachi potrebbero confliggere con i principi *di responsible lending* e con le norme

[89] Ammannati, L., Greco, G. L. (2021). *Piattaforme digitali, algoritmi e big data: Il caso del credit scoring.* Rivista Trimestrale di Diritto dell'Economia, 2, 1–32, 290-325.

anti-discriminatorie, invitando gli operatori ad usare particolare cautela e trasparenza nell'impiego di tali strumenti.[90]

## 2.5 Deep learning vs Machine learning

Quando si parla di deep leanring, molti tendono erroneamente a considerarlo sinonimo del ML o a includerlo indistintamente sotto l'ombrello dell'intelligenza artificiale. In realtà, il deep learning costituisce una sottocategoria del ML che utilizza reti neurali profonde (*deep neural networks*) con numerosi strati, capaci di emulare i complessi processi cognitivi umani.

La maggior parte delle applicazioni di intelligenza artificiale con cui interagiamo quotidianamente si fondano proprio su questa metodologia. Una rete neurale profonda (DNN) ha almeno tre livelli, uno di input, uno o più livelli nascosti intermedi, e uno di output, anche se le implementazioni pratiche spesso ne presentano molti di più. Queste reti sono addestrate su grandi volumi di dati e riescono così a riconoscere pattern, identificare relazioni e comprendere fenomeni, raggiungendo spesso capacità predittive superiori a quelle umane.

La peculiarità del deep learning rispetto ai metodi tradizionali risiede nella capacità di gestire dati grezzi e non strutturati come testi, immagini o audio, grazie a un processo automatico di estrazione delle caratteristiche (*feature extraction*), che minimizza l'intervento umano. Al contrario, nei metodi di ML tradizionale, questo passaggio richiede molto tempo ed esperienza specifica.

[91]

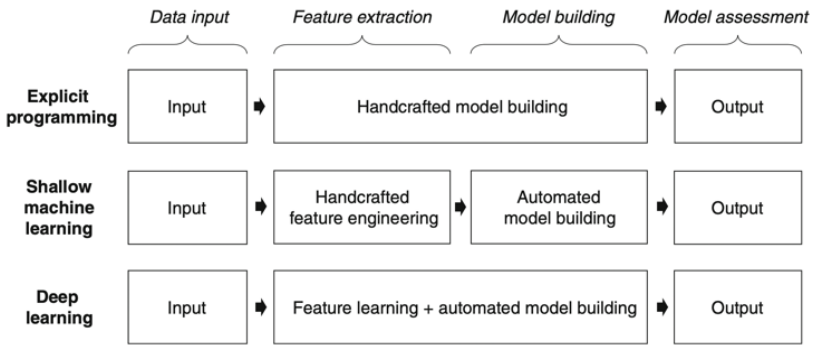



*Figura 10 - Processo di costruzione del modello analitico.*

[90] Chen, T., Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[91] Janiesch, C., Zschech, P., et al. (2021). *Machine learning and deep learning.* Electronic Markets, 31(3).

Come mostrato nella figura, infatti, nella programmazione esplicita (*explicit programming*) il processo di costruzione del modello è interamente manuale, con le caratteristiche e il modello definiti esplicitamente dal programmatore. Nei metodi tradizionali di ML, invece, la fase di estrazione delle caratteristiche avviene manualmente (*handcrafted feature engineering*), mentre il processo di costruzione del modello è automatizzato. Infine, nel deep learning sia l'estrazione delle caratteristiche che la costruzione del modello avvengono automaticamente attraverso la rete stessa, rendendo l'intero processo integrato e più efficiente rispetto ai metodi precedenti.[92]

### 2.5.1 Reti neurali artificiali

Le reti neurali artificiali (ANN - *Artificial Neural network*) si ispirano al funzionamento del cervello umano e consistono in rappresentazioni matematiche di unità di elaborazione chiamate neuroni artificiali, connessi da collegamenti simili alle sinapsi del cervello biologico.

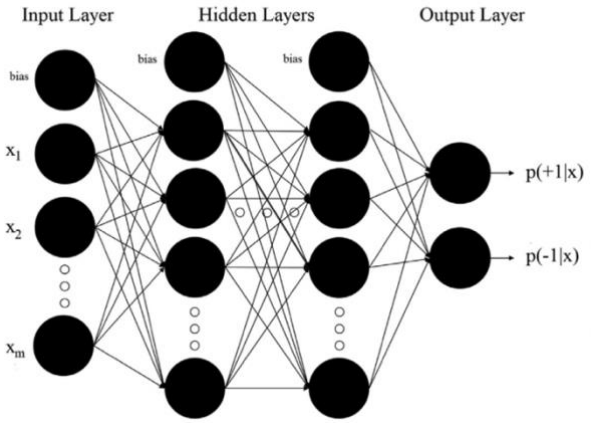
Durante il processo di apprendimento delle reti neurali, il modello cerca di migliorare progressivamente le proprie previsioni confrontandole con i risultati corretti. Quando la rete produce un errore, cioè una differenza tra il valore previsto e quello reale, questo viene "riportato indietro" attraverso la rete, a partire dallo strato finale fino a quelli iniziali. Questo meccanismo è detto retropropagazione dell'errore e consente di calcolare quanto ogni connessione ha contribuito all'errore, così da sapere dove e come intervenire per correggere. Una volta individuati gli elementi da correggere, si procede con un'altra tecnica chiamata discesa del gradiente. In pratica, il modello modifica i "pesi", cioè i valori numerici che regolano l'intensità delle connessioni tra i neuroni, in modo da ridurre l'errore commesso. Questi segnali vengono elaborati dai neuroni successivi solo se superano una determinata soglia definita dalla funzione di attivazione.[93]

Esistono vari tipi di reti neurali profonde, come le reti neurali convoluzionali (CNN - Convolutional Neural Network), particolarmente efficaci nell'elaborazione di immagini e dati spaziali, e le reti neurali ricorrenti (RNN - Recurrent Neural Network), adatte all'elaborazione di sequenze temporali e linguistiche. Le reti neurali profonde, inoltre,

[92] Goodfellow, I., Bengio, Y., et al. (2016). *Deep Learning*. Chapter 1. MIT Press.

[93] Kriegeskorte, N., Golan, T. (2019). *Neural network models and deep learning. Current Biology*, *29*(7).

sono organizzate in architetture annidate molto complesse, e spesso includono neuroni avanzati che utilizzano operazioni sofisticate, come convoluzioni o attivazioni multiple, per migliorare la capacità rappresentativa del modello. Ciò consente alle reti neurali profonde di apprendere automaticamente rappresentazioni complesse dai dati grezzi, riducendo al minimo la necessità di progettare manualmente le caratteristiche.




94

*Figura 11 – Rappresentazione grafica di una rete neurale profonda multistrato.*

Queste caratteristiche rendono il deep learning particolarmente efficace in domini che trattano dati di dimensione elevata e complessi, come immagini, video, testo, audio, in cui generalmente superano i metodi di ML. Tuttavia, per dati di bassa dimensione e set limitati, i modelli tradizionali possono ancora fornire risultati superiori e più interpretabili rispetto alle reti profonde. Inoltre, nonostante le notevoli prestazioni, il deep learning presenta sfide legate all'interpretabilità dei modelli (*black box*), sollevando questioni etiche e normative che ne limitano talvolta l'impiego in settori fortemente regolamentati, come quello finanziario.

## 2.6 Considerazioni conclusive sui modelli di credit scoring

Dopo aver esaminato in dettaglio le principali tecniche di ML applicate al credit scoring, è opportuno formulare alcune considerazioni conclusive sui modelli più efficaci e affidabili. A tal fine, si è scelto di fare riferimento a due contributi scientifici recenti e considerati tra i più completi in letteratura per quanto riguarda l'analisi comparativa delle

[94] Gunnarsson, B. R., Vanden Broucke, et al. (2021). *Deep learning for credit scoring: Do or don't?* European Journal of Operational Research, 295(1).

diverse metodologie di classificazione impiegate nella valutazione del merito creditizio: Gunnarsson et. Al. (2021)[95] e Hayashi (2022).[96]

Queste ricerche rappresentano un punto di riferimento aggiornato e metodologicamente solido, offrendo una panoramica sistematica sui vantaggi e i limiti di modelli statistici classici, metodi *ensemble* e architetture di deep learning.

Lo studio di Gunnarsson et al. (2021) evidenzia con chiarezza come i metodi *ensemble*, in particolare il modello XGBoost, risultino tra i più efficaci in termini di accuratezza predittiva, superando non solo approcci tradizionali come la regressione logistica e gli alberi decisionali, ma anche modelli di deep learning più complessi. In particolare, le architetture neurali profonde, come le *Deep Belief Network* (DBN) e le *Multilayer Perceptron* (MLP) a più strati nascosti, non apportano miglioramenti significativi rispetto a reti più semplici, e anzi comportano costi computazionali maggiori e una più alta complessità nella progettazione e nella fase interpretativa.

Complementare a questa analisi, il lavoro di Hayashi (2022) si concentra sulle tendenze emergenti del deep learning nel credit scoring, confermando che l'efficacia di tali modelli si manifesta principalmente in presenza di dati non strutturati o ad alta complessità, come immagini o testi.

Nei contesti più comuni del settore finanziario, dove prevalgono dati tabellari, le tecniche *ensemble*, e in particolare i modelli basati su *gradient boosting*, rimangono la soluzione preferibile grazie al loro equilibrio tra accuratezza, efficienza e flessibilità. XGBoost viene nuovamente riconosciuto come un benchmark consolidato, capace di modellare efficacemente relazioni non lineari e garantire risultati stabili su differenti set di dati.

Un aspetto trasversale sollevato da entrambi gli studi riguarda la scarsa trasparenza dei modelli complessi: sia le reti neurali profonde sia gli *ensemble* vengono classificati come "*black box*", ossia modelli la cui logica decisionale risulta opaca e difficilmente interpretabile. Questo elemento solleva questioni critiche, soprattutto in ambiti regolamentati come quello del credito, dove la spiegabilità delle decisioni assunte è un requisito fondamentale per assicurare equità, responsabilità e conformità normativa. Per questo motivo, il ricorso a tecniche di Explainable Artificial Intelligence (XAI), come

[95] Gunnarsson, B. R., Vanden Broucke, et al. (2021). *Deep learning for credit scoring: Do or don't?* European Journal of Operational Research, 295(1).

[96] Hayashi, Y. (2022). *Emerging trends in deep learning for credit scoring: A review*. Electronics, 11(11), 3181.

SHAP e LIME, diventa imprescindibile per integrare questi modelli all'interno di processi decisionali trasparenti e giustificabili.

Le evidenze disponibili suggeriscono che, nel panorama attuale del credit scoring, i metodi *ensemble* rappresentano la scelta più robusta, combinando performance predittiva elevate con un ragionevole grado di gestione della complessità. I modelli statistici classici, come la regressione logistica, pur non essendo i più accurati, rimangono fondamentali nei casi in cui la trasparenza decisionale sia prioritaria. Le reti neurali profonde, invece, appaiono meno indicate, salvo scenari caratterizzati da dati complessi o non strutturati.

Tuttavia, è necessario contestualizzare tali conclusioni all'interno della realtà italiana, significativamente diversa per livello di maturità tecnologica. Secondo lo studio condotto dalla Banca d'Italia (Bonaccorsi Di Patti et al., 2022)[97], il ricorso a tecniche di intelligenza artificiale e ML nel credit scoring da parte degli intermediari italiani è ancora limitato, sebbene in crescita. Gli algoritmi più utilizzati si basano su modelli *ensemble*, come Gradient Boosting e Random Forest, ma la loro adozione è spesso ancora in fase sperimentale o subordinata alla supervisione di analisti umani. Inoltre, l'uso di tecniche XAI è diffuso principalmente a fini interni e raramente destinato al cliente finale, nonostante gli obblighi imposti dal GDPR in materia di trasparenza e diritto all'informazione.

La seguente tabella confronta le principali metodologie di credit scoring analizzate nel Capitolo 2, secondo criteri fondamentali quali accuratezza predittiva, interpretabilità, complessità computazionale e robustezza. Essa fornisce una visione sintetica dei punti di forza e delle limitazioni dei vari modelli, evidenziando le differenze principali tra approcci statistici tradizionali, tecniche di ML e deep learning.

[97] Bonaccorsi di Patti, E., Calabresi, F., et al. (2022). *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano* (Questioni di Economia e Finanza, n. 721). Banca d'Italia.

| Modello | Tipo | Accuratezza | Interpretabilità | Capacità di gestire non-linearità | Richiesta dati strutturati | Robustezza a overfitting | Complessità computazionale | Uso prevalente |
|---|---|---|---|---|---|---|---|---|
| **Regressione Logistica** | Statistico tradizionale | Medio-Buona | Alta | Bassa | Necessaria | Buona | Bassa | Ampio uso in Italia |
| **LDA** | Statistico tradizionale | Media | Alta | Bassa | Necessaria | Debole con classi sovrapposte | Bassa | Applicazioni semplici |
| **Alberi Decisionali** | ML interpretabile | Media | Molto alta | Media | Non sempre necessaria | Debole | Bassa | Scorecard visive |
| **Random Forest** | Ensemble (bagging) | Alta | Media | Alta | Non necessaria | Alta | Media | Ampio uso in ML bancario |
| **XGBoost** | Ensemble (boosting) | Molto alta | Bassa | Molto alta | Non necessaria | Alta | Alta | Benchmark attuale |
| **Reti Neurali (ANN/ DNN)** | Deep Learning | Alta (in dati complessi) | Molto bassa | Altissima | Non necessaria | Variabile | Molto alta | Limitato uso reale |
| **Scorecard** | Derivazione da logistica | Media | Alta | Bassa | Necessaria | Buona | Bassa | Tradizionale/ regolato |

*Figura 12 - Confronto sinottico tra i principali modelli di credit scoring*

Dalla tabella emerge come i metodi *ensemble*, in particolare Random Forest e XGBoost, rappresentino attualmente la soluzione più robusta e accurata per il credit scoring, specialmente in contesti ad alta complessità. Tuttavia, in contesti fortemente regolamentati come quello italiano, i modelli tradizionali (es. regressione logistica e scorecard) continuano ad avere un ruolo centrale grazie alla loro semplicità, trasparenza e facilità di implementazione. Va inoltre osservato che le categorie interpretabilità e capacità di gestire non-linearità sono spesso inversamente proporzionali nella maggior parte dei modelli. XGBoost rappresenta attualmente il compromesso più efficace tra prestazioni predittive e flessibilità, sebbene con costi elevati in termini di interpretazione e *tuning*. Le reti neurali risultano vantaggiose solo con dati non strutturati (immagini, testi), poco comuni nel credit scoring bancario attuale.

Nel contesto italiano, dunque, le tecniche statistiche tradizionali, in particolare la regressione logistica, continuano a rappresentare la *best practice*, soprattutto per la loro interpretabilità e coerenza con i requisiti normativi. L'adozione diffusa di modelli complessi, come quelli *ensemble* richiede non solo un'evoluzione tecnologica, ma anche una strutturazione più solida di presidi di governance, monitoraggio, validazione e spiegabilità. In questo senso, la sfida è duplice: da un lato, integrare modelli predittivamente avanzati, dall'altro, garantire il rispetto dei principi di *accountability*, non discriminazione e trasparenza.

Proprio per questa ragione, il prossimo paragrafo sarà dedicato all'analisi delle tecniche di XAI, che rappresentano uno strumento imprescindibile per mitigare i rischi legati

all'utilizzo di modelli *black box*, bilanciando l'efficacia predittiva con l'esigenza di interpretabilità e affidabilità nei processi decisionali creditizi.

### 2.7 XAI: Principi, tecniche e applicazioni nel credit scoring

L'Explainable Artificial Intelligence (XAI) rappresenta uno dei temi centrali dell'attuale dibattito sull'utilizzo responsabile dell'intelligenza artificiale nei contesti decisionali critici, tra cui rientra a pieno titolo il settore del credito. Con il crescente impiego di algoritmi di ML nelle valutazioni del merito creditizio, si assiste a un aumento della complessità dei modelli utilizzati e, parallelamente, a una crescente opacità dei meccanismi decisionali. In questo scenario, l'adozione di strumenti XAI diventa fondamentale per rendere tali sistemi trasparenti, affidabili e conformi alle normative, oltre che eticamente accettabili.

La XAI si propone di garantire che i modelli di ML, spesso strutturati come *black box*, siano comprensibili non solo agli sviluppatori, ma anche agli utenti finali e ai decisori. Questo implica la possibilità di rispondere, in modo soddisfacente, alle domande: come è stata generata una predizione? e perché è stata emessa proprio quella decisione?[98]

Nella letteratura, si distinguono due concetti spesso sovrapposti ma distinti: interpretabilità e spiegabilità. La prima si riferisce alla capacità intrinseca di un modello di essere compreso nella sua struttura e funzionamento; la seconda riguarda la possibilità di fornire una spiegazione, anche ex post, dell'output generato da un sistema complesso. Nei contesti finanziari regolamentati, questa distinzione assume una rilevanza particolare, anche in virtù della normativa GDPR (art. 22), che impone il diritto a una spiegazione in caso di decisioni automatizzate che incidano significativamente sull'individuo.

La necessità di sistemi XAI diventa cruciale nei settori classificati ad alto rischio dall'Artificial Intelligence Act, in cui rientra il credit scoring, ovvero quei settori in cui un errore predittivo può avere conseguenze economiche, legali o sociali gravi.[99]

Uno dei problemi più discussi nella progettazione dei sistemi ML riguarda il compromesso tra accuratezza e interpretabilità. I modelli più performanti, come le reti

[98] Shiam, S. A., Hasan, M. M., et al. (2024). *Credit risk prediction using explainable AI.* Journal of Business and Management Studies, 6(2).

[99] Demaio, L. M., Vella, V., et al. (2020). *Explainable AI for interpretable credit scoring*. In D. C. Wyld et al. (Eds.), *arXiv preprint arXiv:2012.03749*.

neurali profonde o gli *ensemble boosting*, tendono a essere poco trasparenti. Viceversa, modelli interpretabili come la ragressione logistica o gli alberi decisionali sono spesso meno capaci di catturare relazioni complesse nei dati.

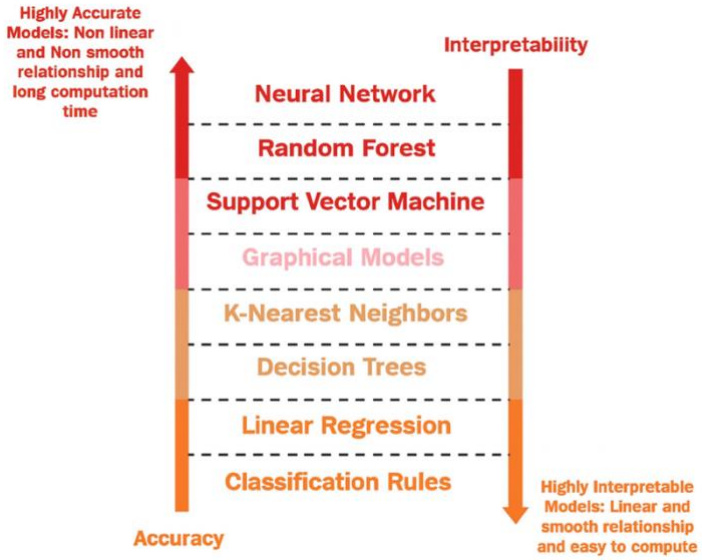



*Figura 13 - Trade-off tra accuratezza e interpretabilità.*

A questo punto sorge spontaneo chiedersi se sia preferibile adottare modelli altamente accurati oppure modelli fortemente interpretabili? La risposta più corretta è dipende. La scelta dipende infatti dal tipo di problema affrontato e dal pubblico a cui è destinata la decisione prodotta dal modello.

Nei contesti ad alto impatto decisionale, dove le conseguenze di una predizione errata possono essere particolarmente gravi, è opportuno privilegiare modelli più interpretabili, anche a costo di sacrificare parzialmente l'accuratezza.

Al contrario, se ci si trova in un contesto ben controllato, dove il problema è formalizzato con chiarezza e l'obiettivo primario è minimizzare l'errore predittivo, allora può essere giustificato l'utilizzo di modelli complessi e ad alta accuratezza, poiché un errore non comporta danni concreti o conseguenze rilevanti. Nel credit scoring, è fondamentale mantenere un equilibrio. Da un lato, la precisione è essenziale per ridurre i rischi di default; dall'altro, la trasparenza è imprescindibile per rispettare i vincoli normativi, garantire la correttezza delle decisioni e prevenire discriminazioni.

[100] Bhattacharya, A. (2022*). Foundational concepts of explainability techniques*. In Applied machine learning explainability techniques (pp. 1–26). Packt Publishing.

Proprio per questo motivo, diventa essenziale comprendere come attenuare questo trade-off e conciliare al meglio le esigenze di performance predittiva e di interpretabilità. Di seguito analizzeremo alcune tra le tecniche più diffuse dell'Explainable AI, oggi impiegate per rendere più comprensibili anche i modelli più complessi.

### 2.7.1 SHAP e LIME: due strumenti a confronto

Le tecniche di Explainable AI (XAI) possono essere classificate secondo diversi criteri, in funzione del tipo di spiegazione che forniscono e del modo in cui interagiscono con il modello. Una prima distinzione riguarda l'ambito della spiegazione: le tecniche locali, come SHAP e LIME, si focalizzano sull'analisi di singole predizioni, spiegando il comportamento del modello su casi specifici; le tecniche globali, invece, mirano a fornire una rappresentazione complessiva del funzionamento del modello su tutto il dataset. Un secondo criterio differenzia i modelli intrinsecamente interpretabili, come la regressione lineare o gli alberi decisionali semplici, che permettono di comprendere direttamente la relazione tra input e output, dai modelli complessi che necessitano di spiegazioni esterne (post-hoc), come nel caso degli algoritmi *ensemble* o delle reti neurali. Inoltre, esistono tecniche progettate per particolari classi di modelli (model-specific), come la visualizzazione delle split nei decision tree, e tecniche più flessibili e generalizzabili (model-agnostic), applicabili a qualsiasi tipo di algoritmo, indipendentemente dalla sua struttura interna. Infine, si distingue tra approcci centrati sul modello, che analizzano il comportamento appreso, e approcci centrati sui dati, che pongono l'attenzione sulla qualità, coerenza e robustezza del dataset impiegato.

Due delle tecniche model-agnostic post-hoc più diffuse e impiegate in ambito XAI sono SHAP (SHapley Additive exPlanations) e LIME (Local Interpretable Model-Agnostic Explanations).[101]

SHAP si basa sulla teoria dei giochi cooperativi e quantifica il contributo marginale di ciascuna variabile all'output del modello, considerando tutte le possibili combinazioni di *feature*. Questo consente di ottenere sia spiegazioni locali, relative a una singola predizione, sia spiegazioni globali riferite al comportamento complessivo del modello.

[101] Bonaccorsi di Patti, E., Calabresi, F., et al. (2022). *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano* (Questioni di Economia e Finanza, n. 721). Banca d'Italia.

LIME, al contrario, approssima il modello complesso in prossimità dell'istanza da spiegare tramite un modello lineare interpretabile. È più leggero in termini computazionali, ma meno efficace nel cogliere relazioni non lineari e complesse, e fornisce esclusivamente spiegazioni locali.[102]

Entrambi i metodi presentano alcune limitazioni strutturali. Innanzitutto, presuppongono l'indipendenza tra le variabili predittive, ipotesi che spesso non è verificata nei dati reali. Inoltre, sono sensibili alla collinearità tra le *feature*: la presenza di variabili fortemente correlate può compromettere l'affidabilità dei *ranking* forniti, riducendo la precisione dell'interpretazione. Un altro rischio è legato alla possibilità che entrambi gli strumenti riflettano o amplifichino *bias* impliciti nel modello ML sottostante, producendo spiegazioni apparentemente corrette ma in realtà distorte.[103]

Oltre a ciò, emerge un'ulteriore criticità legata alla comprensibilità delle spiegazioni per gli utenti non esperti: i valori numerici restituiti da SHAP o i coefficienti di LIME, se privi di un'adeguata contestualizzazione, possono essere mal interpretati. Per questo è fondamentale accompagnare tali metodi con visualizzazioni intuitive, descrizioni chiare e strumenti di supporto che esplicitino in modo semplice le ipotesi metodologiche. Inoltre, l'impiego combinato di tecniche diverse e il confronto tra più modelli rappresentano buone pratiche per rafforzare la robustezza delle spiegazioni e aumentare la fiducia degli stakeholder nei sistemi di decisione automatizzata.

In sintesi, l'analisi dei modelli di credit scoring ha mostrato come l'introduzione di algoritmi di ML e tecniche avanzate abbia migliorato la capacità predittiva rispetto agli approcci tradizionali, pur aumentando la complessità e riducendo la trasparenza delle decisioni. Questa evoluzione solleva interrogativi cruciali legati alla presenza di *bias* e alla difficoltà di interpretare modelli particolarmente sofisticati, che possono amplificare disuguaglianze anziché ridurle. Nel prossimo capitolo l'attenzione si sposterà sui dati alternativi, analizzando il loro potenziale per favorire l'inclusione finanziaria e al tempo stesso i rischi di nuove forme di discriminazione se integrati in sistemi di ML complessi.

---

[102] Nallakaruppan, M. K., Balusamy, B., et al. (2024). *An explainable AI framework for credit evaluation and analysis*. Applied Soft Computing, 153, 111307.

[103] Salih, A. M., Raisi-Estabragh, Z., Boscolo Galazzo, I., et al. (2025*). A perspective on explainable artificial intelligence methods: SHAP and LIME*. Advanced Intelligent Systems, 7(5), 2400304.

# Capitolo III

## Big Data, utilizzo dei dati alternativi e bias nel credit scoring

Il capitolo precedente ha tracciato l'evoluzione delle metodologie di credit scoring, dai modelli statistici tradizionali fino alle più recenti tecniche di ML, evidenziando come ogni approccio porti con sé specifici vantaggi e limiti.

Il presente capitolo si addentra nel cuore della trasformazione in atto, analizzando il ruolo dirompente dei Big Data e più in particolare dei dati alternativi nel ridefinire il perimetro della valutazione del merito creditizio. Verranno esplorate le diverse tipologie, le fonti e le caratteristiche. Verrà approfondito il loro duplice volto: da un lato, come potente strumento per promuovere l'inclusione finanziaria di soggetti altrimenti "invisibili" al sistema creditizio tradizionale, dall'altro, come fonte di nuove e complesse criticità legate alla qualità del dato, alla trasparenza dei processi decisionali e ai rischi di profilazione.

Infine, il capitolo affronterà una delle sfide più significative dell'era algoritmica, i *bias*. Si investigherà come i modelli di ML, nonostante la loro apparente oggettività matematica, possano non solo riflettere, ma addirittura amplificare le disuguaglianze e i pregiudizi esistenti nella società. Attraverso l'analisi delle diverse tipologie di *bias* si metterà in luce come dati apparentemente neutri possano generare risultati discriminatori, perpetuando cicli di esclusione finanziaria.

### 3.1 L'impatto dei Big Data nel credit scoring

L'introduzione dei Big Data nella gestione del rischio di credito ha modificato radicalmente i modelli predittivi tradizionali, consentendo alle istituzioni finanziarie di integrare fonti informative molteplici, eterogenee e continuamente aggiornate. Le informazioni oggi disponibili includono dati transazionali online, *digital footprint*, dati generati da dispositivi mobili e sensori, nonché contenuti testuali e verbali prodotti dagli utenti. Questa ricchezza informativa permette una valutazione più articolata, dinamica e personalizzata del rischio di credito.

Il valore aggiunto di queste nuove fonti risiede nella capacità di catturare segnali comportamentali e psicometrici che sfuggono all'analisi tradizionale. L'analisi dei testi nelle richieste di prestito, le abitudini di navigazione web o l'utilizzo delle app mobili possono fungere da *proxy* per tratti della personalità o comportamenti finanziari rilevanti,

come l'affidabilità, l'impulsività o la stabilità lavorativa. Tali segnali, una volta integrati nei modelli di scoring tramite algoritmi di ML, migliorano sensibilmente la capacità predittiva e la tempestività nell'individuazione di potenziali situazioni di rischio.

L'evoluzione da un approccio *expert-driven*, basato su giudizi soggettivi e dati limitati, a un modello *data-driven*, fondato sull'elaborazione di grandi volumi di dati eterogenei, rappresenta una discontinuità metodologica di grande rilievo. L'utilizzo dei Big Data consente una valutazione del rischio non più basata solo su indicatori statici, ma in grado di adattarsi in tempo reale ai mutamenti del contesto economico e comportamentale del debitore. Come discusso nel Capitolo I, questa evoluzione deve però rispettare i vincoli imposti dal GDPR e dall'AI Act, che impongono trasparenza, correttezza e possibilità di supervisione umana nei processi automatizzati di valutazione del merito creditizio.

I risultati emersi da recenti studi confermano l'efficacia di questo approccio: l'integrazione dei Big Data ha permesso, in alcune realtà finanziarie, di ridurre i tassi di default fino al 30% rispetto ai modelli tradizionali. Gli algoritmi di ML, inoltre, hanno mostrato una capacità superiore di elaborare grandi volumi di dati rispetto ai metodi statistici convenzionali, con una conseguente riduzione dei tassi di insolvenza fino al 25%.[104]

Tuttavia, l'adozione dei Big Data non è esente da sfide. L'eterogeneità delle fonti e la varietà dei formati comportano complessità significative in termini di qualità, integrazione e interoperabilità dei dati. Solo una parte minoritaria delle istituzioni finanziarie è riuscita a implementare efficacemente sistemi di gestione dei Big Data, mentre molte altre incontrano difficoltà legate alla compatibilità e alla velocità di elaborazione.[105] Queste difficoltà impongono investimenti rilevanti in infrastrutture IT e competenze specialistiche, ponendo l'accento sulla necessità di un approccio strategico all'adozione di queste tecnologie.

In parallelo, emergono con forza le questioni legate alla protezione dei dati personali e alla sicurezza informatica. La diffusione di dati sensibili provenienti da fonti non tradizionali ha accentuato i rischi di violazione della privacy e uso improprio delle informazioni. La crescente incidenza di attacchi informatici e le richieste imposte da

[104] Pérez Martín, A., Pérez-Torregrosa, A., et al. (2018). *Big data techniques to measure credit banking risk in home equity loans*. Journal of Business Research, 89, 118–123.

[105] VenkateswaraRao, M., Vellela, S. S., et al. (2023). *Credit investigation and comprehensive risk management system based big data analytics in commercial banking*. IEEE.

normative stringenti, come il GDPR e l'AIA impongono l'adozione di pratiche di gestione dei dati solide e trasparenti.

Accanto agli aspetti tecnologici e normativi, si registra un cambiamento profondo nelle competenze richieste agli operatori del settore. Al fianco dei tradizionali analisti finanziari, si affermano nuove figure professionali, come i *data scientist* e gli esperti di intelligenza artificiale, in grado di progettare, implementare e manutenere modelli analitici complessi. Questa trasformazione ha implicato una revisione delle strutture organizzative e un ripensamento dei percorsi formativi.

I Big Data si configurano oggi come un elemento centrale nel panorama del credit scoring. L'integrazione di dati massivi e non strutturati consente una visione più ampia e profonda del profilo creditizio, migliorando sensibilmente l'efficienza e la precisione delle valutazioni. Nonostante le sfide operative, normative e infrastrutturali, l'adozione dei Big Data rappresenta un passaggio inevitabile per le istituzioni finanziarie che intendano rafforzare la propria capacità competitiva e gestire il rischio in modo più efficace e proattivo.[106]

### 3.2 Opportunità per l'inclusione creditizia tramite dati alternativi

Negli ultimi anni, l'espansione dell'inclusione finanziaria e digitale ha compiuto progressi significativi, ma rimangono ampi margini di miglioramento. Secondo i dati del *Global Findex 2025*, a livello mondiale il 79% degli adulti possiede un conto finanziario e l'86% ha accesso a un telefono cellulare, strumenti chiave per l'integrazione nei circuiti finanziari formali. In diversi Paesi a medio e basso reddito, tra cui Brasile, Cina, India, Kenya e Mongolia, il tasso di bancarizzazione ha raggiunto o superato il 90%, segnalando un'espansione capillare dell'accesso ai servizi finanziari. Tuttavia, oltre 1,3 miliardi di adulti nel mondo restano ancora esclusi da tali servizi, e circa 850 milioni non possiedono nemmeno un telefono cellulare. Particolarmente colpite sono le donne del Sud Asia, con 325 milioni di casi. Questo scenario evidenzia come l'inclusione finanziaria debba affrontare non solo ostacoli tecnologici ed economici (come il costo dei dispositivi e l'analfabetismo digitale), ma anche barriere culturali e infrastrutturali.

[106] Nahar, J., Rahaman, M. A., et al. (2024). Big data in credit risk management: A systematic review of transformative practices and future directions. *International Journal of Management Information Systems and Data Science, 1*(04), 68–79.

L'inclusione formale, inoltre, non garantisce un uso attivo dei servizi bancari. Molti titolari di conto continuano a ricevere redditi in contanti, a non effettuare operazioni bancarie e a non disporre di strumenti di risparmio o assicurazione. A ciò si aggiunge un quadro di fragilità finanziaria: una larga parte della popolazione globale non si sente in grado di affrontare emergenze economiche impreviste, a causa della mancanza di reddito stabile o di capacità di accumulo. Questi limiti suggeriscono la necessità di strategie differenziate, capaci di promuovere non solo l'accesso, ma anche l'utilizzo consapevole e continuativo dei servizi finanziari, favorendo la costruzione di sicurezza economica nel breve e nel lungo periodo.

In questo contesto, i dati alternativi rappresentano una leva strategica per ampliare la portata del credito e includere soggetti finora esclusi. Un limite noto dei modelli di scoring tradizionali è infatti la scarsa capacità di valutare individui privi di una significativa storia creditizia. In Italia, come in molti altri Paesi, esiste una quota rilevante di cosiddetti *credit invisible* o *thin-file*: giovani al primo impiego, nuovi immigrati, consumatori che non hanno mai contratto prestiti o utilizzato strumenti di credito al consumo. Questi soggetti, pur potenzialmente affidabili, risultano spesso privi di un punteggio poiché i sistemi tradizionali non dispongono di informazioni sufficienti.

L'impiego di dati alternativi nasce proprio dall'esigenza di dare un "volto" creditizio a queste persone, sfruttando tracce informative diverse da quelle finanziarie classiche. Le opportunità offerte sono significative. Fonti di dati come le utenze domestiche (bollette di acqua, luce, gas), i pagamenti telefonici o le spese ricorrenti possono fungere da *proxy* del comportamento di pagamento di un individuo. Ad esempio, una storia di bollette pagate puntualmente può indicare abitudini finanziarie responsabili e capacità di far fronte a obblighi mensili. Non a caso, in molti Paesi tali dati "non finanziari" vengono già incorporati nelle valutazioni: uno studio condotto su enti regolatori in diverse giurisdizioni ha rilevato che i dati di pagamento di utenze e telecomunicazioni sono le fonti alternative più comunemente permesse nei framework di credit scoring (citati da 10 autorità su 14 intervistate). Ciò riflette la crescente percezione che tali informazioni abbiano una correlazione concreta con l'affidabilità creditizia, contribuendo a migliorare la capacità predittiva dei modelli senza introdurre eccessivi rischi.

Un'altra categoria di dati alternativi di grande interesse è rappresentata dai dati transazionali granulari, dati che descrivono singole operazioni o eventi, piuttosto che

aggregazioni o riassunti di tali eventi, ad esempio la movimentazione del conto corrente o le spese effettuate tramite strumenti di pagamento digitale. L'analisi dei flussi di cassa consente di valutare il profilo di reddito e spesa del consumatore in tempo reale, superando i limiti di istantanee statiche come l'ultima busta paga. Negli Stati Uniti, società fintech come Petal o Upstart hanno adottato modelli di scoring basati su flussi di cassa, dimostrando che l'inclusione di dati di conto corrente porta ad approvare prestiti a clienti "nuovi" senza aumentare il rischio di default. Nell'Unione Europea, l'entrata in vigore della Direttiva 2015/2366 relativa ai servizi di pagamento nel mercato interno (PSD2 - Payment Services Directive 2), che abilita la condivisione delle informazioni finanziarie del cliente con il suo consenso, ha aperto la strada a soluzioni di *open banking* anche per il merito creditizio: alcune banche e piattaforme stanno iniziando a considerare i dati transazionali dei conti come input aggiuntivo per valutare la solvibilità, soprattutto nei casi di clientela giovane o priva di storia creditizia.

Sul fronte dei dati comportamentali e psicometrici, vi sono esperienze pionieristiche soprattutto nei mercati emergenti. Ad esempio, la società LenddoEFL (operante in Asia e America Latina) ha sviluppato algoritmi che analizzano il profilo social e lo smartphone data del richiedente, come la rete di contatti, la frequenza di interazione sui social network o perfino il modo in cui il soggetto compila i moduli online, per dedurre indicatori di affidabilità ed atteggiamento verso il rischio. Analogamente, alcuni programmi di microfinanza in Africa hanno sperimentato test psicometrici (questionari sulla personalità, giochi di ruolo finanziari) come surrogati del credit scoring per chi non ha documenti reddituali formali. I risultati di queste iniziative, sebbene circoscritti, suggeriscono che opportuni modelli possono estrarre segnali utili anche da informazioni insospettabili: celebre il caso di ZestFinance, fintech statunitense che già intorno al 2010 impiegava oltre 10.000 variabili da attività online e offline (dati di navigazione, modalità di ricerca, ecc.) per valutare richieste di piccoli prestiti (payday loans), accordando credito a una quota di richiedenti che il sistema tradizionale avrebbe altrimenti respinto.

La tabella seguente riassume le principali tipologie di dati, tradizionali e alternativi, utilizzate nei modelli di valutazione del merito creditizio, indicando per ciascuna il relativo impiego applicativo.[107]

[107] Tabella adattata da World Bank Group. (2019). *Credit scoring: Approaches, guidelines*. Washington, DC: World Bank Group.

| Categoria di dati | Tipo di dato | Applicazione nel credit scoring |
|---|---|---|
| **Tradizionale** | Dati transazionali bancari | Registri di ritardi nei pagamenti su crediti attuali e passati, importi dei prestiti e finalità, storico creditizio |
| **Tradizionale** | Dati provenienti dai sistemi di informazioni creditizie | Numero di richieste di credito |
| **Tradizionale** | Dati commerciali | Bilanci, numero di prestiti per capitale circolante, e altri |
| **Alternativo** | Dati su utenze | Registri di pagamenti puntuali come possibile indicatore di affidabilità creditizia |
| **Alternativo** | Social media | Dati provenienti dai social media con possibili informazioni sullo stile di vita del consumatore |
| **Alternativo** | Applicazioni mobili | Sistemi di pagamento mobile con possibile visione sul comportamento del consumatore |
| **Alternativo** | Transazioni online | Dati transazionali granulari con possibili approfondimenti sui modelli di spesa |
| **Alternativo** | Dati comportamentali | Dati psicometrici, compilazione di moduli |

*Figura 14 - Tipologie di dati impiegati nel credit scoring.*

Studi accademici supportano la tesi che i dati alternativi possano migliorare l'inclusione finanziaria senza pregiudicare la qualità delle decisioni. Agarwal et al. (2019)[108] nel loro studio, evidenziano come l'uso di fonti informative non convenzionali abbia un impatto positivo soprattutto per i consumatori a basso reddito, minor livello di istruzione o residenti in aree poco servite dal sistema finanziario, categorie storicamente penalizzate da i metodi di scoring tradizionali.[109] Un maggiore afflusso di dati consente al modello di differenziare meglio il rischio all'interno di gruppi che prima venivano valutati in modo uniforme, tipicamente con esito negativo: ciò porta ad aumentare il tasso di approvazione dei prestiti per tali segmenti, riducendo al contempo i default, in virtù di una migliore capacità di selezione dei debitori meritevoli. Dunque, l'implementazione di sistemi di credit scoring potenziati da Big Data può tradursi in migliaia di nuovi consumatori che accedono al credito per la prima volta.

[108] Agarwal, S., Alok, S., et al. (2019). *Financial inclusion and alternate credit scoring: Role of big data and machine learning in fintech*. Indian School of Business.

[109] Ammannati, L., Greco, G. L. (2021). *Piattaforme digitali, algoritmi e big data: Il caso del credit scoring.* Rivista Trimestrale di Diritto dell'Economia, 2, 1–32, 290-325.

Tali benefici, tuttavia, non sono automatici: dipendono dalla capacità di utilizzare dati realmente pertinenti e predittivi del rischio, applicati in modo corretto e responsabile. In questo senso, la vera sfida, come ha messo in luce anche il dibattito europeo più recente, è riuscire a coniugare inclusione e responsabilità: modelli sempre più accurati devono procedere di pari passo con criteri di trasparenza e con una supervisione umana costante, così che l'innovazione si traduca in giustizia finanziaria e non in nuove forme di esclusione.

### 3.2.1 Criticità legate all'implementazione dei dati alternativi

L'utilizzo di dati alternativi, se da un lato offre indubbi vantaggi, dall'altro pone una serie di criticità che richiedono attenta valutazione. Le problematiche principali riguardano: (i) la qualità e affidabilità intrinseca di questi dati; (ii) i potenziali *bias* e rischi di profilazione indebita che possono emergere; (iii) la trasparenza ed equità dei modelli che li utilizzano;[110]

- Qualità e significatività del dato

Molti dati alternativi non nascono con finalità creditizie e presentano pertanto problemi di accuratezza, completezza o rilevanza. Ad esempio, le informazioni tratte dai social media possono essere autodichiarate, e quindi non verificate, oppure rappresentare solo in modo indiretto le capacità finanziarie di una persona. C'è poi un tema di stabilità temporale: alcune correlazioni rilevate in un dato periodo, come tra abitudini di navigazione internet e rischio di default, potrebbero svanire al mutare del contesto tecnologico o sociale, rendendo il modello obsoleto in breve tempo. Il rischio è di basare le decisioni su informazioni poco robuste, con possibili errori di valutazione. Non a caso, la letteratura segnala come l'utilità e l'oggettività di molti dati non strutturati (come testi, post sui social, video, ecc.) nel migliorare il credit scoring non sia ancora dimostrata con certezza. L'integrazione di queste fonti richiede dunque un'attenta validazione statistica:

[110] Pérez Martín, A., Pérez-Torregrosa, A., et al. (2018). *Big data techniques to measure credit banking risk in home equity loans*. Journal of Business Research, 89, 118–123.

occorre provare che aggiungano reale potere predittivo al modello, evitando al contempo fenomeni di overfitting o associazioni spurie.[111]

- Profilazione, *bias* e discriminazione

Una delle preoccupazioni maggiori riguarda la possibilità che l'uso di dati alternativi amplifichi *bias* preesistenti o introduca nuove forme di discriminazione. I dati non finanziari, per loro natura, possono contenere informazioni strettamente legate a caratteristiche personali protette: ad esempio, i contatti sui social media possono riflettere l'appartenenza etnica o il contesto socio-economico; la localizzazione GPS può rivelare la zona di residenza (facendo da surrogato per il reddito medio o la composizione etnica di quel quartiere); persino lo stile di scrittura in un form online potrebbe fornire indicazioni sul livello di istruzione. Se tali elementi entrano nello scoring senza adeguate cautele, si rischia di dar luogo a decisioni algoritmiche distorte, che penalizzano sistematicamente alcuni gruppi. Ad esempio, molti algoritmi di scoring basati su Big Data finiscono per cristallizzare le disuguaglianze esistenti, classificando come meno meritevoli individui provenienti da comunità svantaggiate e creando così un circolo vizioso di esclusione finanziaria. Questo accade perché l'algoritmo, addestrato su dati storici, dove certe minoranze hanno tassi di default più alti, tende a riprodurre quello schema, continuando a negare loro credito anche quando sarebbero in grado di ripagarlo. In letteratura vengono segnalati casi emblematici: ad esempio sistemi di credit scoring per carte di credito che assegnavano linee di fido inferiori alle donne rispetto agli uomini con identico profilo finanziario, oppure algoritmi che penalizzavano residenti di aree a basso reddito indipendentemente dalle effettive capacità individuali di rimborso.

In sostanza, simili esiti non costituiscono soltanto un problema etico, ma possono portare a una forma di profilazione automatizzata potenzialmente illecita ai sensi del quadro giuridico vigente. Il quadro normativo europeo, come discusso in precedenza, pone infatti forti limiti all'uso di algoritmi decisionali suscettibili di produrre effetti discriminatori (artt. 4 e 22 GDPR), richiedendo trasparenza, correttezza e intervento umano nei processi automatizzati.

[111] Johnson, K. N. (2019). *Written testimony before the United States House Committee on Financial Services Task Force on Financial Technology: Examining the use of alternative data in underwriting and credit scoring to expand access to credit*. Tulane University Law School.

Un ulteriore rischio è quello della profilazione occulta: combinando diversi dati alternativi, un modello può arrivare a dedurre informazioni sensibili, senza che queste siano presenti esplicitamente, ma solo perché altri dati le rappresentano indirettamente. Queste sono chiamate variabili surrogato, e il loro utilizzo può portare a discriminazioni, anche se non intenzionali. In tal modo si rischia di eludere il divieto di trattare categorie particolari di dati imposto dall'art. 9 GDPR, finendo per utilizzare indirettamente informazioni sensibili in violazione della normativa sulla protezione dei dati personali.
Tuttavia, la natura opaca degli algoritmi può rendere difficile accorgersi di *bias* indiretti: un modello formalmente "neutro", che non include variabili sensibili, può comunque produrre effetti disparitati se alcuni input fungono da *proxy* di caratteristiche protette. Ad esempio, il CAP di residenza può funzionare come un surrogato dell'origine etnica e causare disparità nei punteggi senza che ciò sia immediatamente evidente.[112]

- Privacy, trasparenza e conformità normativa

Come già trattato nel Capitolo I, l'impiego di dati alternativi nei modelli di credit scoring solleva rilevanti interrogativi sul piano della protezione dei dati e della trasparenza algoritmica. In particolare, l'uso di tecniche avanzate di ML, come reti neurali o Random Forest, rende spesso opachi i criteri che conducono all'assegnazione di un punteggio o al rifiuto di credito, dando origine al cosiddetto "effetto *black box*". Tale opacità limita la possibilità per l'utente di comprendere le decisioni subite e ostacola il controllo da parte delle autorità di vigilanza.
A livello normativo, il GDPR impone che siano utilizzati esclusivamente dati pertinenti e proporzionati alla finalità legittima perseguita, escludendo in modo esplicito qualsiasi uso di informazioni sensibili non correlate alla valutazione del rischio creditizio. Il Codice di condotta italiano per i Sistemi di Informazione Creditizia conferma tali restrizioni, vietando, ad esempio, l'impiego di dati tratti dai social network o relativi a opinioni politiche o sanitarie.
In linea con questo orientamento, il Regolamento europeo sull'Intelligenza Artificiale (AI Act) qualifica i sistemi automatizzati di credit scoring come "ad alto rischio", imponendo obblighi di trasparenza, tracciabilità e possibilità di intervento umano sulle decisioni

[112] Trinh, T. K., Zhang, D., et al. (2024). *Algorithmic fairness in financial decision-making: Detection and mitigation of bias in credit scoring applications*. Journal of Advanced Computing Systems (JACS).

automatizzate. Tuttavia, nella pratica, tali requisiti risultano difficili da applicare a causa dell'intrinseca complessità e non interpretabile struttura di molti modelli algoritmici.
Nel complesso, la sfida è trovare un equilibrio tra l'innovazione necessaria a migliorare l'accesso al credito e la tutela dei diritti fondamentali degli individui, garantendo processi trasparenti, equi e rispettosi della privacy.[113]
Il prossimo paragrafo esaminerà come il quadro attuale in Italia si sta posizionando rispetto a queste sfide, anche in confronto con altri Paesi, delineando lo stato dell'arte nell'uso di dati alternativi per il credit scoring.

### 3.2.2 Confronto internazionale e stato dell'arte in Italia nell'adozione dei dati alternativi

In Italia, l'impiego dei dati alternativi nel credit scoring è un tema ancora in fase sperimentale e oggetto di attenzione sia da parte degli operatori finanziari che delle autorità di vigilanza. Tradizionalmente, il mercato del credito al dettaglio italiano si è basato su un insieme relativamente ristretto di informazioni: dati anagrafici, reddito documentato, storico creditizio registrato dai SIC (Sistemi di Informazioni Creditizie), eventuali segnalazioni di sofferenze o insolvenze. La cultura bancaria italiana, fortemente ancorata a principi prudenziali, ha finora adottato un approccio cauto verso l'inclusione di fonti non tradizionali. Inoltre, la cornice normativa, in primis le stringenti regole sulla privacy, ha limitato di fatto, la possibilità di integrare nel processo decisionale dati estranei alla sfera strettamente finanziaria del cliente. Il GDPR ha infatti fissato limiti precisi al trattamento di categorie particolari di dati e alla profilazione automatizzata, imponendo agli operatori bancari un approccio più prudente rispetto a quello di altri Paesi.[114] Ad esempio, l'utilizzo di informazioni dai social media è vietato, e persino l'accesso a dati come il pagamento delle utenze non è prassi comune, sebbene non vi sia un divieto assoluto su quest'ultimo, alcune banche valutano positivamente l'esibizione volontaria di bollette pagate come prova di affidabilità, ma si tratta di iniziative caso per caso e non di un sistema automatizzato su larga scala.

[113] Nopper, T. K. (2020). *Alternative data and the future of credit scoring*. Data for Progress.

[114] Ammannati, L., Greco, G. L. (2021). *Piattaforme digitali, algoritmi e big data: Il caso del credit scoring.* Rivista Trimestrale di Diritto dell'Economia, 2, 1–32, 290-325.

Negli ultimi anni, tuttavia, anche in Italia si osservano segnali di cambiamento. FinTech e banche tradizionali hanno avviato programmi pilota per arricchire i modelli di scoring con dati alternativi selezionati. Un importante impulso viene dall'implementazione dell'open banking: grazie alla Direttiva PSD2, i consumatori possono consentire a terze parti, incluse banche concorrenti o società fintech, di accedere al loro storico transazionale. Questa apertura regolamentare, oltre a favorire nuovi modelli di business, si collega anche alle sfide di trasparenza algoritmica richiamate nel Capitolo II: l'uso dei dati PSD2 nei modelli di ML richiede infatti strumenti di XAI per garantire spiegazioni comprensibili agli utenti e alle autorità di vigilanza.

Alcuni intermediari hanno iniziato a sfruttare queste informazioni, come ad esempio l'analisi delle entrate e uscite sul conto corrente, dei pattern di spesa, degli addebiti ricorrenti, per migliorare la valutazione del merito creditizio di clienti con breve relazione bancaria o provenienti da altri istituti. La Banca d'Italia, nel suo Rapporto FinTech 2021, rileva che un certo numero di operatori sta investendo in sistemi di intelligenza artificiale per il credit scoring, anche se per ora si tratta soprattutto di progetti limitati a segmenti specifici, come il *peer-to-peer lending* o i prestiti istantanei di basso importo.[115]

Un esempio concreto in ambito nazionale è rappresentato da Fabrick, piattaforma lanciata nel 2018 dal Gruppo Banca Sella con l'obiettivo di promuovere un ecosistema collaborativo tra banche, fintech e corporate, sfruttando le potenzialità offerte dall'open banking e dalla Direttiva PSD2. In questo contesto, Fabrick ha sviluppato un modello di credit scoring che combina dati transazionali derivanti da conti correnti PSD2 con algoritmi di ML, al fine di ottenere valutazioni più tempestive e predittive affidabili.[116]

Anche Intesa Sanpaolo, tra i principali istituti bancari italiani, sta investendo nello sviluppo di modelli avanzati basati su Big Data e intelligenza artificiale per potenziare i propri sistemi di analisi del rischio di credito.

Un altro ambito di interesse è quello dei dati di utenze: benché non esista ancora una condivisione strutturata di questi dati con i credit bureau italiani, si esplorano convenzioni che permettano, nel rispetto della privacy, di acquisire informazioni sull'affidabilità nei pagamenti delle bollette. Ciò potrebbe avvenire, ad esempio, mediante accordi con le

[115] Sciarrone Alibrandi, A., Borello, G., et al. (2019). *Marketplace lending: Verso nuove forme di intermediazione finanziaria?* (Quaderni FinTech, n. 5). Commissione Nazionale per le Società e la Borsa (CONSOB).

[116] https://www.fabrick.com/it-it/corporate/storia/

utility per segnalare i clienti morosi gravi, similmente a quanto già accade per le bollette telefoniche: alcune compagnie telefoniche, in base ad accordi di settore, comunicano ai SIC le morosità persistenti dei contratti di abbonamento, rendendo di fatto questi dati parte del patrimonio informativo considerato in fase di istruttoria. anche se il loro peso nel punteggio finale è solitamente marginale e confinato all'indicazione di un evento negativo.

Per quanto riguarda i dati pubblici, l'Italia dispone di registri consolidati che sono già integrati nei sistemi creditizi. Non c'è invece evidenza di utilizzo di dati non convenzionali come registri di immigrazione o precedenti penali ai fini del credito, prassi che, peraltro, sarebbero altamente problematiche sul piano etico e legale.

Anche a livello europeo la tendenza è analoga: Paesi come la Germania mantengono un'impostazione conservativa, il caso SCHUFA, analizzato nel I capitolo, ne è una dimostrazione pratica.

In Francia, la valutazione del merito di credito per il consumo è tradizionalmente focalizzata sul tasso di indebitamento e sulla presenza di eventi negativi nel registro della Banque de France (FICP), l'uso di dati alternativi non risulta praticato dagli istituti in maniera significativa, anche perché il contesto normativo tende a privilegiare la privacy, la Francia è stata fra le prime a recepire in modo restrittivo le norme GDPR limitando la profilazione dei consumatori.

Discorso diverso per gli Stati Uniti, dove l'approccio è più orientato al mercato e all'innovazione, pur entro i limiti posti dalle leggi antidiscriminatorie. Negli USA le principali agenzie di credito (Experian, Equifax, TransUnion) hanno iniziato da qualche anno a offrire prodotti basati su dati alternativi: ad esempio, Experian Boost consente ai consumatori di autorizzare l'inclusione nel proprio credit report di pagamenti di utenze e canoni di locazione, incrementando immediatamente il credit score di chi ha una storia limitata ma paga regolarmente bollette e affitti. Anche FICO, il più diffuso punteggio di credito, ha lanciato versioni "estese" come UltraFICO, che integrano dati del conto corrente, come risparmio medio, scoperti, attività transazionali, per valutare i richiedenti privi di storia creditizia.

Queste innovazioni sono state accolte con interesse anche da parte delle autorità: il Consumer Financial Protection Bureau (CFPB) ha sostenuto sperimentazioni controllate, come quella di Upstart, monitorando da vicino il suo modello di credit scoring alternativo

per verificare che non introducesse *bias* razziali. 33. I risultati pubblicati hanno mostrato che il modello di Upstart approvava una percentuale significativamente più alta di candidati appartenenti a minoranze rispetto al modello tradizionale, con tassi di default equivalenti, un dato incoraggiante per i fautori dei dati alternativi.

Negli USA, l'inclusione dei dati alternativi avviene comunque entro i limiti fissati dal Fair Credit Reporting Act (FCRA), che impone l'obbligo di fornire spiegazioni comprensibili in caso di decisioni negative, e dall'Equal Credit Opportunity Act (ECOA), che proibisce l'uso di attributi discriminatori. Analogamente, anche in Europa la responsabilità d'uso è al centro del dibattito, seppur incardinata in un impianto normativo più rigido.[117]

In Asia, e in particolare in Cina, l'impiego di dati alternativi è molto più avanzato. Un esempio emblematico è rappresentato da *Sesame Credit (Zhima Credit)*, sistema sviluppato dal gruppo Ant, affiliato ad Alibaba, che valuta la solvibilità degli utenti sulla base di dati comportamentali e transazionali generati nell'ecosistema digitale del gruppo. Sebbene facoltativo, è ampiamente diffuso e rivolto in particolare a soggetti esclusi dal sistema bancario tradizionale, ma ha suscitato critiche per l'opacità dei criteri utilizzati e i rischi per la privacy.

La normativa cinese si sta progressivamente rafforzando con interventi come la *Personal Information Protection Law*, ma restano lacune in termini di enforcement. Anche in India si sperimentano modelli simili: piattaforme come *Flipkart* utilizzano le tracce digitali dei consumatori per offrire credito a soggetti privi di storico finanziario, contribuendo all'inclusione, ma sollevando interrogativi sul controllo algoritmico e la protezione dei dati.[118]

Dal quadro delineato emerge come l'approccio italiano, così come quello di molti Paesi europei, resta improntato alla cautela: l'adozione di dati alternativi procede sotto stretto controllo regolamentare, con attenzione prioritaria alla tutela della privacy e alla prevenzione di discriminazioni. Tuttavia, il confronto con modelli più dinamici, come quelli sviluppati negli Stati Uniti o in alcuni mercati emergenti, evidenzia un potenziale ancora inespresso. La sfida sarà cogliere queste opportunità senza compromettere i presìdi

[117] World Bank Group. (2019). *Credit scoring: Approaches, guidelines*. Washington, DC: World Bank Group.

[118] Centraco, G., Dore, G. M. D. (2024). Il Sistema di Credito Sociale cinese: tecnologia come strumento di sorveglianza e persuasione. *OrizzonteCina*, *15*(1), 61–75.

di trasparenza e correttezza: una prospettiva che richiede non solo un'attenta selezione delle fonti informative, ma anche strumenti tecnici e procedurali capaci di garantire affidabilità, spiegabilità e equità nei modelli decisionali.

### 3.3 Bias algoritmico e machine learning: Ruolo dei dati e del machine learning nell'amplificazione delle disuguaglianze

Nel contesto del credit scoring automatizzato, l'impiego di modelli di ML rappresenta un'innovazione promettente ma allo stesso tempo una potenziale minaccia in termini di equità e non discriminazione. Gli algoritmi predittivi vengono addestrati su dati storici per valutare la probabilità che un richiedente onori o meno un prestito. Tuttavia, se i dati storici riflettono decisioni passate influenzate da stereotipi o esclusioni sistemiche, il modello finisce per riprodurre o addirittura amplificare tali distorsioni, generando discriminazioni nei confronti di gruppi protetti, come donne, minoranze etniche, o lavoratori non tradizionali.

In questo quadro, il concetto di *fairness*, traducibile in italiano come equità, imparzialità o anche giustizia algoritmica, assume un ruolo centrale. La *fairness* è un concetto complesso e multidimensionale, la cui definizione varia a seconda del contesto culturale e dell'ambito applicativo. Non esiste una definizione universale, ma è fondamentale che ogni istituzione finanziaria adotti una propria nozione operativa di questo concetto, bilanciandola con l'efficienza e la redditività nelle proprie attività creditizie. Nella pratica, la *fairness* viene formalizzata principalmente attraverso tre approcci: misure statistiche, misure basate sulla similarità e ragionamenti causali. I primi due sono i più rilevanti per l'ambito del credit scoring. Tra queste tecniche rientrano a loro volta gli strumenti di XAI, discussi nel II Capitolo, che permettono di rendere esplicite le relazioni tra variabili e outcome.

Le misure statistiche si basano su una cosiddetta matrice di confusione, che mette a confronto ciò che l'algoritmo prevede con ciò che succede realmente. In questo schema, se una persona viene considerata "affidabile" e poi effettivamente ripaga il prestito, si parla di vero positivo**.** Se invece viene ritenuta affidabile ma in realtà non restituisce il denaro, si parla di falso positivo**.** Esistono poi anche i veri negativi, quando l'algoritmo prevede correttamente che una persona non restituirà il prestito, e i falsi negativi**,** quando l'algoritmo sbaglia e considera inaffidabile una persona che in realtà lo sarebbe stata.

Le definizioni di *fairness* basate su questo modello cercano di capire se l'algoritmo commette errori più spesso su certi gruppi rispetto ad altri. Per esempio, se le donne hanno una percentuale molto più alta di falsi negativi rispetto agli uomini, significa che l'algoritmo è meno equo nei loro confronti. In sostanza, l'obiettivo di queste misure è fare in modo che gli errori siano distribuiti in modo equilibrato tra i diversi gruppi**,** così che nessuno venga sistematicamente svantaggiato.[119]

Le misure basate sulla similarità si fondano sul principio che individui simili dovrebbero essere trattati in modo simile, anche se appartengono a gruppi differenti. In ambito creditizio, ciò significa che due richiedenti con profili finanziari comparabili (reddito, stabilità lavorativa, spese, ecc.) dovrebbero ricevere lo stesso esito, indipendentemente da caratteristiche come genere o origine etnica. L'importanza di questi concetti emerge in modo evidente considerando che i *bias* algoritmici possono portare a vere e proprie forme di discriminazione, cioè situazioni in cui alcuni gruppi ricevono vantaggi sistematici, come tassi di approvazione più alti, mentre altri subiscono svantaggi sistematici, come negazioni più frequenti di prestito, a parità di condizioni.

È importante notare che il *bias*, in senso lato, può riferirsi a qualsiasi forma di preferenza. Tuttavia, in questo contesto si parla di *bias* ingiusto o discriminazione algoritmica, ossia l'emergere di disparità sistematiche nel trattamento delle persone da parte di sistemi automatizzati. Questo tipo di *bia*s, può manifestarsi in relazione a caratteristiche personali quali razza, genere, età, orientamento sessuale, nazionalità, religione o lingua.

Nello specifico, Kelly e Mirpourian (2021)[120] si concentrano sul *gender bias*, ovvero la tendenza degli algoritmi a produrre risultati sistematicamente sfavorevoli per le donne, anche quando queste presentano un rischio creditizio pari a quello degli uomini. La causa non risiede tanto nella macchina in sé, quanto nel fatto che gli algoritmi sono costruiti da persone, le quali, anche inconsapevolmente, portano con sé pregiudizi cognitivi che finiscono per "trascriversi" nel codice.

I *bias* possono dunque entrare nei sistemi in due modi principali: (1) per come l'algoritmo è progettato e costruito e (2) a causa dei dati usati per addestrarlo, che possono essere incompleti, distorti o riflettere pratiche passate discriminatorie.

---

[119] Trinh, T. K., Zhang, D., et al. (2024). *Algorithmic fairness in financial decision-making: Detection and mitigation of* bias *in credit scoring applications*. Journal of Advanced Computing Systems (JACS).

[120] Kelly, S., Mirpourian, M. (2021, February). *Algorithmic bias, financial inclusion, and gender: A primer on opening up new credit to women in emerging economies*. Women's World Banking.

Alla luce di tutto ciò, appare evidente come il concetto di *fairness* non sia solo un principio etico, ma una necessità tecnica e operativa per garantire che i modelli di credit scoring automatizzato non contribuiscano a consolidare dinamiche di esclusione e disuguaglianza.

### 3.3.1 Tipologie di bias e conseguenze sull'inclusione finanziaria

Nel credit scoring automatizzato, i *bias* algoritmici si manifestano in forme diverse, spesso sovrapposte, che si sviluppano lungo l'intero ciclo di vita del modello: dalla raccolta dei dati fino alla progettazione e applicazione del sistema decisionale. Questi *bias* non solo minano la precisione del modello, ma soprattutto compromettono l'equità e l'inclusione finanziaria, penalizzando gruppi già svantaggiati. Di seguito si analizzano tre forme ricorrenti: *bias* di campionamento, *labeling bias* e *bias* da variabili *proxy*.

- *Bias* di campionamento (*Sampling Bias*)

Il *bias* di campionamento si verifica quando alcune porzioni della popolazione sono sovra o sottorappresentate nei dati utilizzati per addestrare il modello. Questo porta a un apprendimento sbilanciato da parte dell'algoritmo, che generalizza le decisioni sulla base del gruppo più presente.

La sottorappresentazione porta a una maggiore incidenza di errori predittivi, come falsi negativi, nei confronti di gruppi minoritari, con conseguenze gravi: richieste respinte ingiustamente, linee di credito limitate e tassi d'interesse più elevati. Questo *bias* è particolarmente problematico per soggetti giovani, lavoratori informali o migranti, le cui informazioni sono spesso scarse o incomplete nei dataset.[121]

- *Bias* di etichettatura (*Labeling Bias*)

Il *bias* di etichettatura riguarda le modalità con cui i dati vengono classificati o etichettati dagli sviluppatori del modello. Le etichette sono fondamentali per "insegnare" al sistema quali sono i comportamenti da apprendere, ma possono contenere distorsioni implicite. Ad esempio, se le professioni dei richiedenti vengono etichettate come "medico" o "infermiera" invece di "operatore sanitario", si introducono variabili fortemente correlate

[121] Moldovan, D. (2023). *Algorithmic decision making methods for fair credit scoring*. IEEE Access.

al genere. Il modello finisce così per considerare il tipo di lavoro come indicatore indiretto del sesso del richiedente, portando a disparità sistematiche nel trattamento di uomini e donne, anche a parità di condizioni economiche. Il *labeling bias* è particolarmente pericoloso perché spesso passa inosservato: le etichette sembrano tecnicamente corrette, ma codificano stereotipi culturali o professionali, che l'algoritmo assimila e trasforma in decisioni concrete.

- *Bias* da *proxy* o da variabili surrogato (*Outcome Proxy Bias*)

Un altro tipo di distorsione è il cosiddetto *bias da proxy*, che emerge quando il modello utilizza una variabile come sostituto per prevedere un determinato esito, senza che ci sia una chiara giustificazione causale. Per esempio, l'uso dell'indirizzo di residenza per stimare la probabilità di insolvenza può introdurre un grave errore sistemico. Se il tasso di default è storicamente più alto in certi quartieri a basso reddito, l'algoritmo potrebbe dedurre che tutti gli individui provenienti da quella zona siano ad alto rischio, anche se molti di loro sono in realtà finanziariamente affidabili. Questa correlazione aggregata maschera la reale diversità individuale e penalizza chi vive in aree economicamente svantaggiate. L'uso di *proxy* non è di per sé scorretto, ma lo diventa quando la variabile surrogata è collegata ad attributi sensibili come genere, etnia o condizione socioeconomica, portando a decisioni discriminatorie.

Questi tre tipi di *bias* non operano in modo isolato. Spesso si combinano, aggravando le distorsioni nel processo di valutazione creditizia. I risultati sono concreti e misurabili: maggiore incidenza di richieste respinte per i gruppi protetti, linee di credito ridotte e condizioni peggiorative anche a fronte di profili simili a quelli dei gruppi maggioritari.[122]

### 3.3.2 Strategie per la mitigazione dei bias e promozione dell'equità

Per garantire un credit scoring equo e inclusivo, la letteratura propone approcci strutturati per individuare e mitigare i *bias*. È opinione condivisa che occorra dapprima misurare le disparità con apposite metriche di *fairness*, poi intervenire con tecniche adeguate. In fase diagnostica si usano indicatori come l'impact ratio e metriche di parità tra gruppi.

[122] Trinh, T. K., Zhang, D., et al. (2024). *Algorithmic fairness in financial decision-making: Detection and mitigation of bias in credit scoring applications*. Journal of Advanced Computing Systems (JACS).

Una prima linea d'intervento sono le tecniche di *pre-processing,* che agiscono sui dati prima dell'addestramento. Tali metodi possono correggere informazioni discriminatorie, riequilibrare i dataset o rivedere etichette errate derivanti da decisioni passate ingiuste.
Le tecniche di in-processing agiscono direttamente durante la fase di addestramento del modello, inserendo al suo interno regole che lo obbligano a comportarsi in modo equo. Un esempio è *l'adversarial debiasing*, una tecnica che "costringe" l'algoritmo a non basarsi su caratteristiche discriminatorie (come il genere o l'etnia), anche indirettamente. Un altro approccio consiste nell'imporre vincoli di equa distribuzione (*fair allocation*), ossia garantire che, a parità di condizioni, anche i gruppi protetti ricevano un trattamento proporzionale a quello del gruppo di riferimento.
Infine, le tecniche di *post-processing* operano sui risultati del modello già addestrato, ad esempio ricalibrando le soglie decisionali o correggendo leggermente le probabilità di default per ridurre il divario tra gruppi.
Ognuna di queste strategie presenta vantaggi e limiti. Il *pre-processing* affronta il problema alla radice ma può eliminare informazioni utili; *l'in-processing* garantisce modelli più fair ma è più complesso; il *post-processing* è semplice da applicare ma rischia di distorcere le previsioni.
La soluzione più efficace è spesso combinare più interventi. È fondamentale però affiancare queste tecniche a una governance istituzionale solida. Le realtà più virtuose coinvolgono tutta l'organizzazione nella promozione dell'equità, con *policy* interne e controlli periodici.
Mitigare i *bias* è quindi essenziale per rendere il credit scoring automatizzato uno strumento di reale inclusione finanziaria.[123]
Nei tre capitoli precedenti sono stati approfonditi il quadro normativo, l'evoluzione metodologica del credit scoring e i complessi aspetti legati ai dati, ai *bias* e all'inclusione finanziaria, offrendo una panoramica completa delle opportunità e delle criticità presenti nella letteratura e nella pratica. A partire da questa base teorica, il capitolo successivo si concentrerà sulla parte sperimentale, con l'obiettivo di dimostrare come l'utilizzo di dati alternativi possa migliorare la predittività dei modelli e, al contempo, favorire una maggiore inclusione finanziaria.

[123] Kelly, S., Mirpourian, M. (2021, February). *Algorithmic bias, financial inclusion, and gender: A primer on opening up new credit to women in emerging economies*. Women's World Banking.

# Capitolo IV

# Ricerca empirica: metodologia e sperimentazione

## 4.1 Introduzione e Obiettivi

Il presente capitolo illustra la fase sperimentale dello studio, volta a indagare in che misura l'integrazione di dati alternativi ai dati creditizi classici possa migliorare le capacità predittive dei modelli di credit scoring e, allo stesso tempo, favorire una maggiore inclusione finanziaria. L'ipotesi di fondo è che un set informativo più ampio e diversificato permetta di ridurre le esclusioni ingiustificate e affinare la valutazione del rischio.

Tuttavia, come vedremo nella fase sperimentale, le evidenze non confermano l'ipotesi iniziale: l'inclusione di dati alternativi non ha prodotto miglioramenti significativi e sistematici rispetto ai modelli basati esclusivamente su variabili tradizionali. I vantaggi osservati, limitati a casi circoscritti, sono deboli e non replicabili. Tale esito risulta costante in tutte le analisi e sotto tutte le ipotesi e configurazioni considerate. Le possibili cause sono discusse nelle sezioni successive.

Per rispondere alla domanda di ricerca sono stati utilizzati due dataset distinti di origine pubblica (Kaggle): uno con sole variabili tradizionali e uno con variabili alternative di tipo comportamentale e transazionale.

Dopo le opportune operazioni di pulizia e armonizzazione, i due dataset sono stati combinati in un unico dataset integrato di tutte le variabili. I dettagli tecnici del processo di unificazione e *pre-processing* sono presentati nella sezione successiva.

L'analisi è strutturata in nove configurazioni modellistiche, ottenute combinando tre algoritmi di ML con tre versioni del dataset: solo variabili tradizionali, solo variabili alternative e una versione completa. Gli algoritmi selezionati sono Regressione logistica, Random Forest e XGBoost. L'impostazione consente un confronto diretto del contributo informativo dei dati alternativi rispetto ai tradizionali e al set completo.

Oltre alla valutazione delle performance predittive, un aspetto centrale della sperimentazione riguarda la questione della trasparenza dei modelli complessi.

Random Forest e XGBoost, pur mostrando buoni risultati in termini di accuratezza, sono infatti percepiti come *black box*. Per superare questa limitazione, nella fase finale

dell'analisi sono stati applicati i metodi di Explainable AI (SHAP e LIME), con l'obiettivo di rendere interpretabili le decisioni degli algoritmi e di verificare se tali strumenti possano garantire un utilizzo affidabile e giustificabile dei modelli avanzati nel contesto del credit scoring.

Gli obiettivi specifici della sperimentazione sono:

1. Quantificare l'incremento di performance (accuracy, precision, recall, F1-score, AUC-ROC) derivante dall'inclusione di variabili alternative.
2. Valutare l'impatto sull'inclusione finanziaria, con particolare attenzione al recall della classe positiva (clienti affidabili) come indicatore di riduzione delle esclusioni ingiustificate.
3. Analizzare la classificazione di profili borderline, ovvero soggetti inizialmente considerati non affidabili dai modelli tradizionali e potenzialmente rivalutati positivamente con dati alternativi.
4. Esaminare la trasparenza e la stabilità dei modelli, confrontando soluzioni interpretabili (logistica) e complesse ma performanti (XGBoost e Random Forest), e verificando se l'applicazione di SHAP e LIME consenta di superare la natura di *black box* di tali algoritmi.
5. Individuare un modello ottimale che migliori le previsioni senza introdurre disparità nei confronti delle fasce di clientela più vulnerabili.

Attraverso questo disegno sperimentale sarà possibile verificare se l'uso combinato di dati alternativi, tecniche avanzate di ML e strumenti di explainable AI rappresenti un progresso significativo per le politiche di concessione del credito, sia in termini di efficienza predittiva sia di equità.

### 4.2 Unificazione dei dataset

Per valutare l'apporto informativo dei dati alternativi è stato costruito un dataset integrato combinando due dataset pubblici di Kaggle. La fusione si fonda su tre variabili chiave comuni: age_gap (classe di età), income_level (fascia di reddito) e education_level (livello di istruzione); così da accoppiare profili omogenei per età, reddito e istruzione. È stata effettuata un'Inner join su queste variabili per mantenere solo le osservazioni presenti in entrambe le fonti di dati, garantendo comparabilità e coerenza del campione.

Prima della fusione è stata effettuata una riduzione delle variabili, basata su criteri soggettivi, per eliminare campi irrilevanti rispetto agli obiettivi di credit scoring e inclusione finanziaria, passando da 112 a 67 variabili utili. Una scrematura analitica ulteriore, basata su tecniche statistiche, è prevista nelle sezioni successive.

Il processo di unificazione ha seguito questi passaggi essenziali:

1. Armonizzazione dei nomi delle variabili (formato snake_case) per coerenza.
2. Aggiunta di prefissi per distinguere l'origine dei dati: trad_ (dati tradizionali) e alt_ (dati alternativi).
3. Pulizia delle tre variabili chiave: rimozione delle osservazioni con valori mancanti o incoerenti.
4. Inner join su age_gap, income_level, education_level.
5. Campionamento casuale di 10.000 osservazioni dal risultato dell'Inner join per garantire eterogeneità e stabilità computazionale.

L'obiettivo è disporre di una base unica che unisca la solidità dei dati creditizi tradizionali con la ricchezza dei segnali comportamentali, così da testare se, e quanto, i dati alternativi migliorino accuratezza predittiva, identificazione dei profili meritevoli *thin file* e livelli di inclusione.

I due dataset utilizzati provengono da Kaggle e rappresentano fonti informative molto diverse tra loro.

Il Loan Default Dataset (Yasser H.)[124] include 20.000 osservazioni e 34 variabili tipiche dei sistemi di credit scoring tradizionali, credit score, rapporto debito/reddito, cronologia dei pagamenti, importo, durata e tasso del prestito. La forza di questa base dati è la stretta connessione con la capacità di rimborso, il limite è la scarsa varietà informativa, che tende a penalizzare i soggetti "*credit invisible*" e non cattura segnali comportamentali.

Il Large Retail Data Set for EDA (Utkalk K.)[125] è molto più ampio (1.000.000 di osservazioni, 78 variabili) e nasce per analisi commerciali e comportamentali. Contiene informazioni demografiche (età, genere, stato civile, istruzione, fascia di reddito), indicatori di acquisto (frequenza, canale, valore medio, loyalty) e segnali digitali (visite online, uso dell'app, engagement). Pur non essendo concepito per il credit scoring, offre

[124] https://www.kaggle.com/datasets/yasserh/loan-default-dataset

[125] https://www.kaggle.com/datasets/utkalk/large-retail-data-set-for-eda

proxy utili su abitudini e transazioni che permettono di esplorare l'apporto dei dati alternativi in termini di accuratezza e inclusione.
Il dataset integrato ottenuto dalla fusione delle due fonti è la base delle analisi successive, in cui i modelli verranno confrontati nelle versioni "tradizionale", "alternativa" e "completa".

### 4.3 Selezione e scrematura delle variabili

Il dataset unificato, ottenuto dall'integrazione delle fonti tradizionali e alternative, risultava inizialmente composto da 65 variabili predittive (escludendo l'ID e la variabile target), suddivise tra variabili numeriche e categoriche, prive di valori mancanti. Nel contesto di questa analisi, la variabile target rappresenta l'esito della decisione creditizia, ossia se il prestito richiesto dal cliente è stato approvato o meno. In particolare, è stata utilizzata la variabile *trad_loan_approved*, presente nel dataset unificato, codificata in formato binario come segue:

- 1 = Prestito approvato
- 0 = Prestito rifiutato

Questa variabile funge da elemento di riferimento per tutte le analisi statistiche e di ML condotte.
Il suo ruolo è quello di consentire la formazione dei modelli predittivi, i quali imparano a individuare le relazioni tra le caratteristiche (variabili indipendenti) e l'esito dell'approvazione del prestito (variabile dipendente).
La scelta di questa target è coerente con l'obiettivo dello studio, ovvero valutare in che misura l'integrazione di variabili tradizionali e alternative possa migliorare l'accuratezza delle previsioni e favorire l'inclusione finanziaria, garantendo al contempo l'interpretabilità dei modelli.
Un numero così elevato di predittori, se da un lato consente un'ampia copertura informativa del profilo del richiedente, dall'altro introduce rischi significativi:

- Ridondanza informativa, con variabili altamente correlate tra loro;
- Multicollinearità, che può compromettere la stabilità delle stime nei modelli lineari;

- Overfitting, soprattutto con algoritmi complessi come Random Forest e XGBoost;
- Maggiore complessità interpretativa, particolarmente rilevante in un contesto in cui la trasparenza e la spiegabilità (XAI) sono elementi centrali.

Per affrontare tali criticità è stato adottato un processo di riduzione dimensionale basato su un duplice approccio: criteri statistici, per identificare le variabili con maggiore potere discriminante rispetto alla variabile target, e considerazioni teoriche, per garantire la rappresentatività di entrambe le tipologie di dati e un potenziale contributo all'inclusione finanziaria.

L'analisi delle variabili numeriche è stata condotta calcolando la correlazione assoluta di Pearson rispetto alla variabile target binaria *trad_loan_approved*.

Sono stati considerati rilevanti i predittori con $|r| \geq 0.10$; quelli con $0.05 \leq |r| < 0.10$ sono stati mantenuti solo in presenza di una solida giustificazione economico-finanziaria, mentre per $|r| < 0.05$ si è proceduto in genere all'esclusione. In questo quadro, income mostra l'associazione più alta ($|r| \approx 0.60$), seguita da trad_loan_amount e trad_net_worth; indicatori di storia creditizia come trad_length_of_credit_history e trad_credit_score, pur con correlazioni moderate ($|r| \approx 0.15$ - $0.20$), sono stati mantenuti per il loro significato economico. In pochi casi sono state preservate variabili con $|r| < 0.10$ quando garantivano copertura informativa non ridondante.

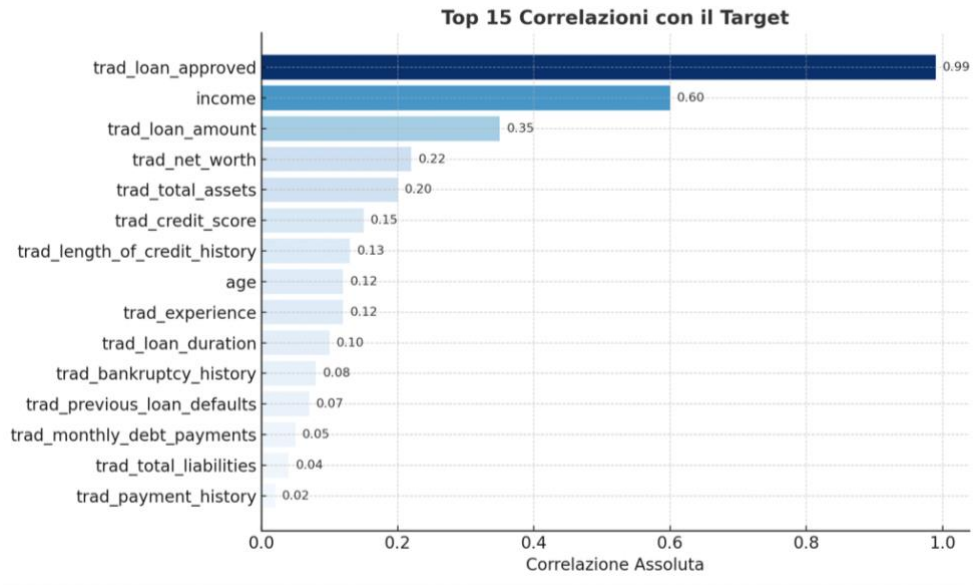



*Figura 15 - Top 15 Correlazioni con il Target*

Per le variabili categoriali è stato applicato il test di Chi-quadro rispetto alla variabile target. Il Chi-quadro confronta le frequenze osservate con quelle attese in assenza di relazione. Nel nostro campione tutte le variabili in figura sono significative ($p < 0.01$). Spiccano trad_debt_to_income_ratio e trad_utility_bills_payment_history, seguite da trad_credit_card_utilization_rate; elevata anche la capacità discriminante degli indicatori di costo del credito (trad_interest_rate, trad_base_interest_rate). Le variabili alternative alt_unit_price e alt_avg_purchase_value risultano anch'esse significative, con contributo informativo più contenuto, utile soprattutto per profili con scarsa o assente storia creditizia.

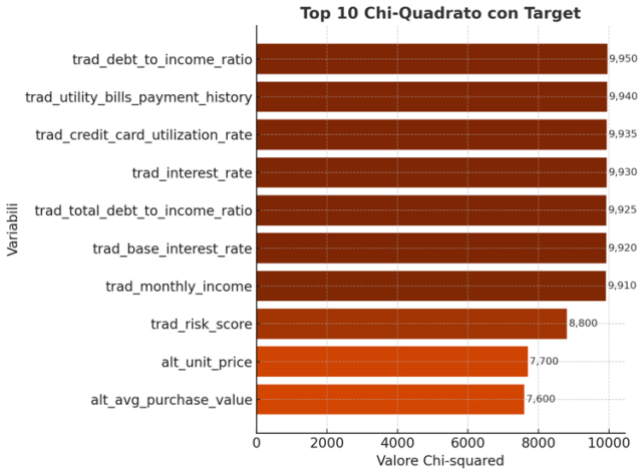



*Figura 16 - Top 10 Chi-Quadro con Target*

La selezione delle variabili non si è basata esclusivamente sui valori statistici, ma anche sulla rilevanza economico-finanziaria. Tra i dati tradizionali sono stati mantenuti income, trad_total_debt_to_income_ratio, trad_risk_score, trad_credit_score, trad_length_of_credit_history, trad_bankruptcy_history, trad_net_worth *e* trad_experience. Tra i dati alternativi, sono stati conservati alt_avg_discount_used, alt_promotion_channel, alt_churned, alt_social_media_engagement *e* alt_payment_method, in grado di descrivere comportamenti di spesa, preferenze nei canali di interazione e grado di fidelizzazione.

L'integrazione di variabili tradizionali e alternative ha portato alla definizione di 24 variabili concettuali (16 numeriche e 8 categoriali), considerate il miglior compromesso tra potere predittivo e diversità informativa. Le variabili categoriali sono state trasformate in *dummy variables* per garantirne la compatibilità con i modelli di ML, con un aumento apparente da 24 a 35 variabili senza introdurre nuovi predittori. Il processo di selezione è

passato dalle 65 variabili iniziali alle 24 concettuali individuate teoricamente, fino alle 35 pronte per il modello dopo la codifica e il controllo della collinearità. Questa riduzione progressiva ha migliorato l'efficienza computazionale, ridotto il rischio di overfitting, aumentato l'interpretabilità e mantenuto un equilibrio informativo tra dati tradizionali e alternativi.

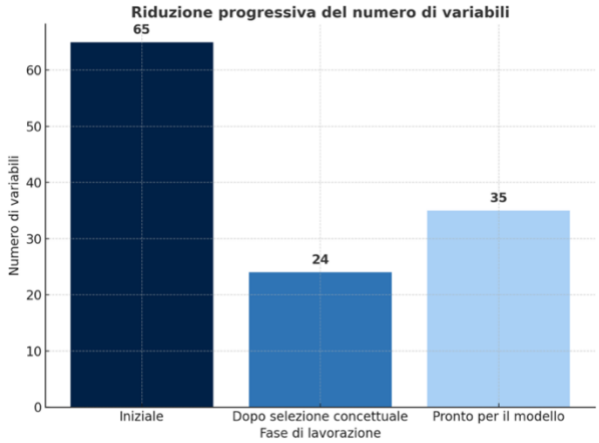



*Figura 17 - Riduzione progressiva del numero di variabili*

La seguente tabella riassume le 24 variabili concettuali effettivamente utilizzate nella fase sperimentale, selezionate in base a criteri statistici e considerazioni teoriche.

| Nome variabile | Descrizione | Tipo |
|---|---|---|
| income | Reddito annuo complessivo del cliente | Numerica |
| trad_risk_score | Punteggio di rischio creditizio calcolato su parametri tradizionali | Numerica |
| trad_interest_rate | Tasso di interesse applicato al prestito | Numerica |
| trad_loan_amount | Importo totale del prestito richiesto | Numerica |
| trad_net_worth | Patrimonio netto stimato (attività – passività) | Numerica |
| trad_credit_score | Punteggio creditizio (es. FICO Score) del cliente | Numerica |
| trad_experience | Anni di esperienza lavorativa del cliente | Numerica |
| age | Età del cliente | Numerica |
| trad_total_debt_to_income_ratio | Rapporto tra debito complessivo e reddito annuo | Numerica |
| trad_length_of_credit_history | Anni di storia creditizia del cliente | Numerica |
| trad_bankruptcy_history | Numero di precedenti fallimenti registrati | Numerica |
| alt_avg_discount_used | Sconto medio percentuale utilizzato negli acquisti (fonte dati alternativi) | Numerica |
| alt_online_purchases | Numero totale di acquisti effettuati online | Numerica |
| alt_in_store_purchases | Numero totale di acquisti effettuati in negozio fisico | Numerica |
| alt_avg_transaction_value | Valore medio delle transazioni effettuate | Numerica |
| alt_days_since_last_purchase | Giorni trascorsi dall'ultimo acquisto | Numerica |
| income_level | Fascia di reddito dichiarata (Low, Medium, High) | Categoriale |
| trad_employment_status | Stato occupazionale (Employed, Self-Employed, Unemployed) | Categoriale |
| education_level | Livello di istruzione (High School, Bachelor's, Master's, PhD) | Categoriale |
| age_gap | Classe di età (Young Adult, Middle Aged, Older Adult) | Categoriale |
| alt_promotion_channel | Canale promozionale principale (Online, Social Media, In-store) | Categoriale |
| alt_churned | Indicatore di cessazione della relazione commerciale (Yes/No) | Categoriale |
| alt_social_media_engagement | Livello di interazione con i contenuti social del fornitore (Low, Medium, High) | Categoriale |
| alt_payment_method | Metodo di pagamento preferito (Carta di credito, Bonifico, Portafoglio digitale, Altro) | Categoriale |

*Figura 18 - Tabella riassuntiva delle variabili concettuali utilizzate per la fase sperimentale*

### 4.4 Analisi descrittiva delle variabili

Una prima analisi preliminare è stata condotta su un insieme selezionato di variabili ritenute strategiche nel processo di concessione del credito. Gli obiettivi sono duplici: da un lato comprendere la forma e la variabilità delle distribuzioni (identificando eventuali valori anomali e concentrazioni); dall'altro, verificare in via esplorativa la capacità discriminante di tali variabili rispetto all'esito della richiesta di finanziamento.

In primo luogo, si esamina la variabile Income (reddito annuo complessivo). L'istogramma (*Figura 19*) mostra una distribuzione fortemente asimmetrica a destra, con la maggior parte delle osservazioni al di sotto di €100.000 e una asimmetria positiva verso redditi fino a €500.000. La presenza di alcuni outlier nella fascia alta, sebbene influenzi media e deviazione standard, riflette un fenomeno tipico nelle popolazioni reali dove pochi individui possiedono redditi molto superiori alla media. Il boxplot di Income suddiviso per esito del prestito conferma un divario marcato tra i gruppi: i richiedenti con prestito approvato presentano redditi medi nettamente più elevati, con anche una maggiore dispersione interna. Ciò indica che, sebbene vengano talvolta approvati anche clienti con redditi modesti, la probabilità di approvazione cresce sensibilmente all'aumentare della solidità economica del richiedente. Questo comportamento è coerente con le prassi di valutazione creditizia, in cui un reddito elevato è associato a un rischio percepito inferiore e quindi a maggiori chance di ottenere il credito.

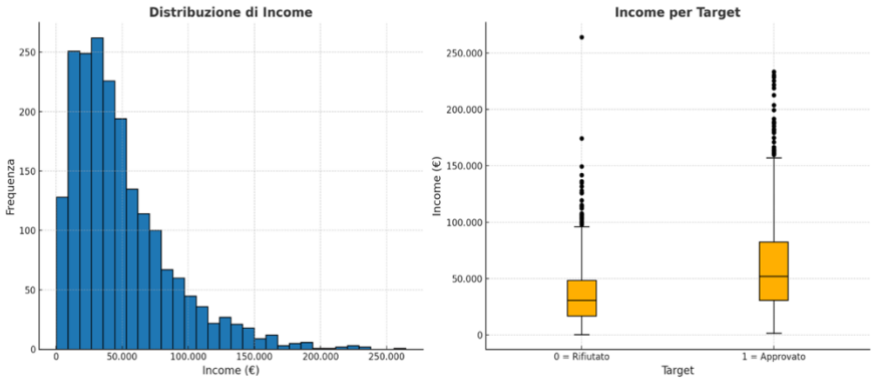



*Figura 19 - Distribuzione del reddito e differenze per target di approvazione*

Anche la distribuzione dell'importo del prestito richiesto (Trad Loan Amount) risulta fortemente asimmetrica verso destra, con una concentrazione di richieste sotto i €50.000 e un numero limitato di importi molto elevati, oltre €100.000. L'analisi separata per esito della domanda rivela un pattern interessante: i prestiti approvati presentano un importo

medio inferiore rispetto a quelli rifiutati. Ciò suggerisce una politica di concessione prudente da parte dell'ente finanziatore: importi richiesti più contenuti implicano un'esposizione al rischio minore e quindi criteri di approvazione meno restrittivi. Questa variabile fornisce un'informazione complementare a quella del reddito: a parità di condizioni, un alto reddito associato a una richiesta di importo moderato aumenta sensibilmente la probabilità di approvazione del prestito (*Figura 20*).

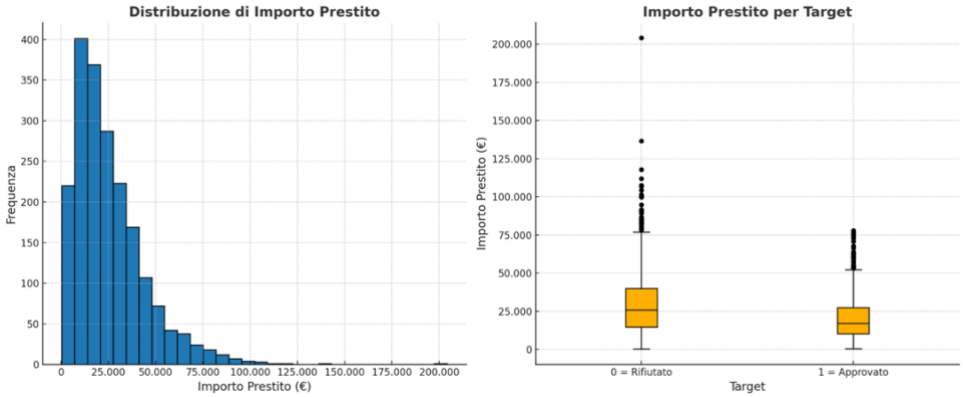



*Figura 20 - Distribuzione dell'importo richiesto e differenze per target di approvazione*

Passando al punteggio creditizio tradizionale (Trad Credit Score), la distribuzione appare quasi simmetrica, centrata attorno a 580–600 punti, un intervallo tipico di profili medi nella scala FICO. Il confronto tra i gruppi di esito mostra che i clienti approvati dispongono in media di punteggi leggermente più alti, ma con una sovrapposizione significativa rispetto ai punteggi dei clienti rifiutati. Questo suggerisce che il credit score, pur rilevante nel processo decisionale, non funge da discriminante assoluto: anche clienti con punteggi creditizi nella media possono ottenere l'approvazione, probabilmente compensando con redditi elevati o altri indicatori favorevoli del merito creditizio (*Figura 21).*

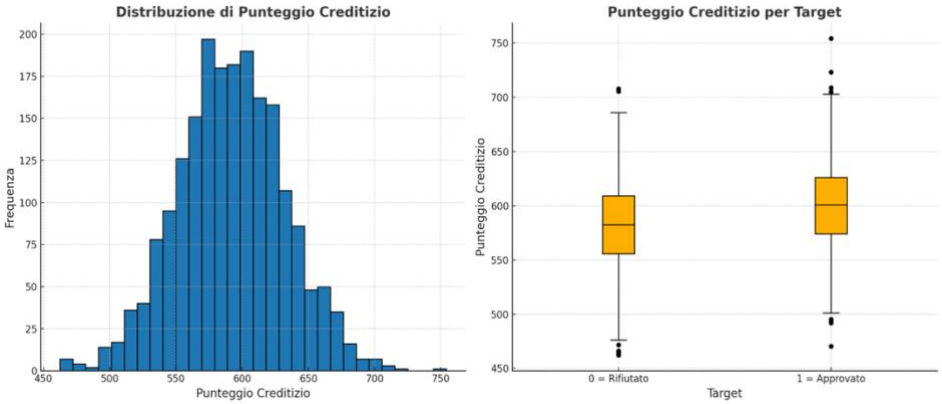



*Figura 21 - Distribuzione del punteggio creditizio e differenze per target di approvazione*

La statistica descrittiva di ulteriori variabili integrate nel modello arricchisce la comprensione del dataset, includendo sia indicatori tradizionali sia metriche alternative legate al comportamento d'acquisto.

| Variabile | Media | Mediana | Dev.Std | Varianza | Q1 | Q3 | Min | Max |
|---|---|---|---|---|---|---|---|---|
| age | 44.31 | 44.0 | 11.72 | 137.28 | 36.0 | 53.0 | 18.0 | 80.0 |
| alt_avg_discount_used | 0.25 | 0.25 | 0.14 | 0.02 | 0.13 | 0.37 | 0.0 | 0.5 |
| alt_avg_transaction_value | 258.92 | 258.92 | 80.21 | 6432.96 | 136.62 | 382.47 | 10.03 | 499.91 |
| alt_days_since_last_purchase | 183.08 | 184.0 | 109.11 | 11901.78 | 92.0 | 273.0 | 0.0 | 364.0 |
| alt_in_store_purchases | 49.02 | 49.0 | 28.83 | 830.49 | 24.0 | 74.0 | 0.0 | 99.0 |
| alt_online_purchases | 49.19 | 48.0 | 28.85 | 832.43 | 25.0 | 74.0 | 0.0 | 99.0 |
| income | 62209.96 | 50743.0 | 46237.48 | 2137918574.97 | 33204.0 | 78312.5 | 15000.0 | 485341.0 |
| trad_bankruptcy_history | 0.05 | 0.0 | 0.23 | 0.05 | 0.0 | 0.0 | 0.0 | 1.0 |
| trad_credit_score | 578.52 | 585.0 | 51.28 | 2629.95 | 548.0 | 615.0 | 362.0 | 707.0 |
| trad_experience | 21.96 | 21.0 | 11.55 | 133.3 | 12.0 | 28.0 | 1.0 | 61.0 |
| trad_interest_rate | 0.24 | 0.23 | 0.04 | 0.0 | 0.21 | 0.26 | 0.11 | 0.41 |
| trad_length_of_credit_history | 14.87 | 15.0 | 7.21 | 51.7 | 8.0 | 22.0 | 2.0 | 29.0 |
| trad_loan_amount | 24961.18 | 21862.0 | 17850.57 | 318077654.89 | 15526.0 | 30988.5 | 3674.0 | 184732.0 |
| trad_net_worth | 72681.96 | 32379.5 | 125262.19 | 15698270170.39 | 86823.25 | 86823.25 | 1004.0 | 2603208.0 |
| trad_risk_score | 406944.29 | 402182.0 | 121845.76 | 14849000000.0 | 48.0 | 58.0 | 28.8 | 93.2 |
| trad_total_debt_to_income_ratio | 0.29 | 0.29 | 0.29 | 0.08 | 0.17 | 0.32 | 0.02 | 0.37 |

*Figura 22 - Sintesi statistica delle variabili numeriche del dataset*

Per quanto riguarda le variabili socio-demografiche, il livello di istruzione e l'età mostrano tendenze significative. La maggior parte dei richiedenti nel dataset presenta un diploma di scuola superiore, seguita da coloro con laurea triennale, master e, in misura più ridotta, dottorato di ricerca (PhD).

La probabilità di approvazione tende ad aumentare all'aumentare del grado di istruzione, raggiungendo i valori più elevati tra i possessori di master e dottorato, mentre i diplomati presentano tassi di approvazione inferiori. Ciò suggerisce un legame tra istruzione avanzata e maggiore affidabilità percepita dal punto di vista creditizio. Allo stesso modo, la fascia d'età più rappresentata nel campione è quella "Middle Aged" (35–54 anni), seguita da "Older Adult" e "Adult", mentre i "Young Adult" risultano minoritari. I tassi di approvazione sono più alti tra i clienti "Older Adult", probabilmente grazie a una maggiore stabilità lavorativa e a uno storico creditizio più lungo e più bassi tra i "Young Adult" che tendono ad avere meno esperienza creditizia e garanzie lavorative.

Le variabili comportamentali e di marketing mostrano un contributo predittivo molto limitato: canali promozionali, modalità di pagamento e indicatore di abbandono (Alt

Churned) presentano distribuzioni simili tra approvati e rifiutati e non forniscono segnali discriminanti. Per questo se ne riporta solo un cenno sintetico, in linea con la loro scarsa rilevanza in analisi univariata.

Dal punto di vista economico-lavorativo, la segmentazione per fasce di reddito e stato occupazionale conferma e amplia le evidenze emerse. La maggioranza dei clienti si colloca nelle fasce di reddito Medio o Basso, mentre gli High income (fascia ad alto reddito) sono meno numerosi. I tassi di approvazione differiscono nettamente: i richiedenti ad alto reddito ottengono l'approvazione in una proporzione decisamente superiore, quelli a basso reddito hanno probabilità molto più basse, e la fascia media mostra un tasso intermedio, confermando la forte relazione positiva tra solidità reddituale e esito favorevole, già osservata con la variabile continua Income. Per quanto riguarda lo status occupazionale, la maggior parte dei richiedenti risulta lavoratore dipendente, con una quota più piccola di autonomi e pochi disoccupati. I tassi di approvazione sono leggermente più alti per i lavoratori autonomi rispetto ai dipendenti e ai non occupati, ma le differenze non sono particolarmente marcate. In generale, avere un impiego è una condizione quasi necessaria per l'accesso al credito, ma nel dataset analizzato l'essere lavoratore autonomo sembra dare un lieve vantaggio in termini di approvazione rispetto ad altre categorie.

Un grafico riepilogativo (*Figura 23*) evidenzia congiuntamente la distribuzione dei clienti e i rispettivi tassi di approvazione per le principali variabili categoriali descritte sopra. Complessivamente, questo confronto conferma il peso strategico delle caratteristiche socio-economiche nel determinare l'esito delle richieste, mentre gli indicatori comportamentali e di marketing, almeno in analisi univariata, forniscono un valore aggiunto limitato.

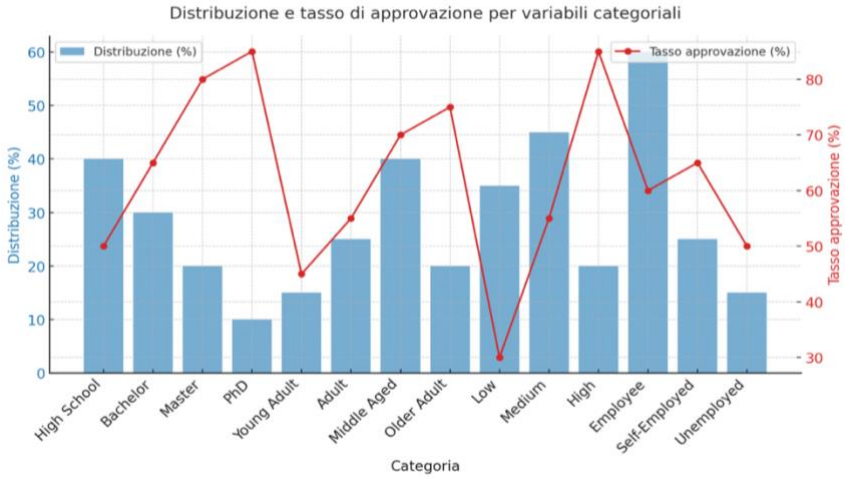



*Figura 23 - Distribuzione e tasso di approvazione per variabili categoriali*

Le evidenze emerse dall'analisi descrittiva confermano che il dataset riflette logiche creditizie tradizionali plausibili e attese:

- Un reddito elevato, soprattutto se accompagnato da richieste di prestito contenute, favorisce l'approvazione del finanziamento;
- Il punteggio creditizio tradizionale del cliente è importante ai fini decisionali, ma non costituisce da solo un criterio esclusivo (clienti con score medi possono essere approvati se altri fattori sono solidi);
- Le variabili alternative offrono insight aggiuntivi sulle abitudini di spesa, la propensione al risparmio e la fedeltà dei clienti, arricchendo la valutazione del rischio soprattutto per i soggetti con una storia creditizia limitata o assente.

Questi risultati preliminari indicano dunque che l'integrazione di dati tradizionali e dati alternativi potrebbe migliorare la capacità predittiva dei modelli di credit scoring, oltre a supportare strategie di inclusione finanziaria più mirate verso clienti altrimenti poco profilabili.

### 4.4.1 Analisi bivariata e correlazioni

Oltre alle distribuzioni univariate, è stata esaminata la matrice di correlazione e svolta un'analisi bivariata su alcune coppie di variabili per identificare le relazioni più

significative. Le correlazioni più elevate riscontrate, con le rispettive interpretazioni, sono riportate di seguito.

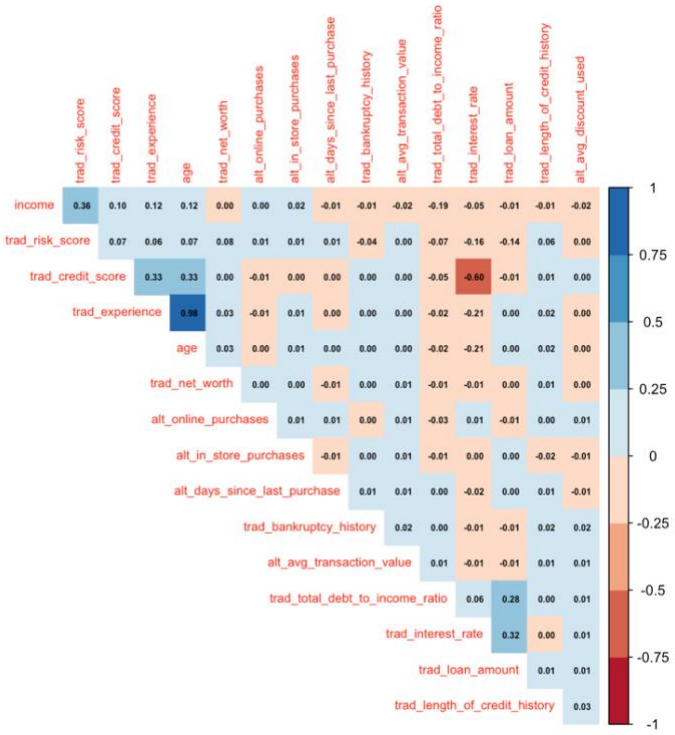



*Figura 24 - Matrice di Correlazione*

- trad_experience vs age (ρ=+0.98): correlazione quasi perfetta; veicolano la stessa informazione. Nei modelli lineari è preferibile non includerle insieme per evitare multicollinearità.
- trad_credit_score vs trad_risk_score (ρ=+0.33): correlazione moderata e coerente; punteggi di credito più alti si associano a rischio stimato migliore.
- income vs trad_risk_score (ρ=+0.36): moderata positiva; redditi più elevati si accompagnano a profili meno rischiosi.
- trad_credit_score vs trad_interest_rate (ρ = –0.60): correlazione negativa forte, la più elevata in valore assoluto tra quelle considerate rilevanti per l'analisi, escludendo la coppia *trad_experience vs age*, potenzialmente affetta da collinearità. Un legame che merita un approfondimento specifico. Il diagramma di dispersione (*Figura 25*) evidenzia un chiaro andamento inverso: all'aumentare del *credit score*, il tasso di interesse applicato diminuisce sensibilmente.

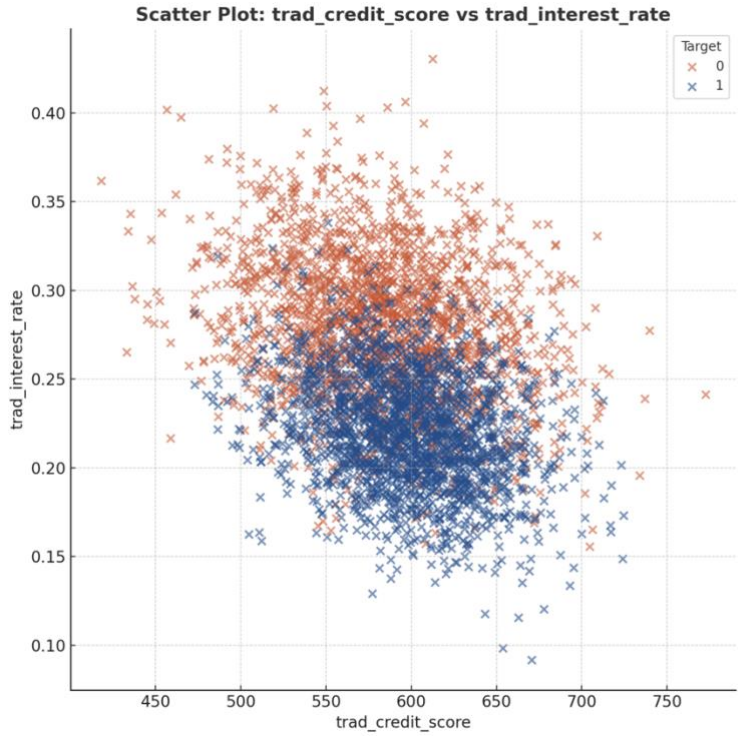



*Figura 25 - Diagramma di Dispersione: relazione tra punteggio creditizio e tasso si interesse*

La densità maggiore si colloca tra 500–650 punti di credit score; oltre ~650 i tassi si stabilizzano su valori più bassi (~15–20%). L'andamento è coerente con un pricing risk-based: all'aumentare dello score i tassi scendono e cresce la quota di richieste approvate. L'evidenza visiva conferma quindi il ruolo di queste due variabili come driver centrali sia nella previsione dell'esito di approvazione, sia nella definizione delle condizioni contrattuali.

- trad_total_debt_to_income_ratio vs trad_loan_amount (ρ=+0.28): debole-moderata; importi più alti si riflettono in un Debt to Income maggiore.
- trad_interest_rate vs trad_risk_score (ρ=−0.16): negativa debole; profili meno rischiosi ottengono tassi leggermente più bassi.

Di contro, molte variabili alternative (es. *alt_online_purchases*, *alt_avg_transaction_value*, *alt_days_since_last_purchase*) mostrano correlazioni prossime allo zero con gli indicatori tradizionali: non evidenziano legami lineari forti, ma potrebbero agire tramite effetti non lineari, che verranno valutati con Random Forest e XGBoost.

### 4.4.2 Gestione degli outlier in vista delle analisi predittive

Dopo aver rilevato degli outlier in alcune variabili chiave, in particolare Income e Trad Loan Amount, sono state adottate due versioni della base dati: una versione "pulita" per la regressione logistica e una versione "completa", con outlier, per Random Forest e XGBoost, meno sensibili a tali valori. La scelta, non prevista ex ante, è stata definita in esito alla verifica sui valori anomali, mantenendo stesse variabili e medesime partizioni train/validation per garantire la confrontabilità.
La seguente tabella sintetizza i risultati dell'analisi condotta, una valutazione sistematica della presenza di valori anomali nelle variabili chiave.

| Variabile | N_outlier | Perc_outlier | Soglia_inferiore | Soglia_superiore |
|---|---|---|---|---|
| trad_risk_score | 1055 | 10.55 | 33.0 | 73.0 |
| trad_net_worth | 822 | 8.22 | -108375.5 | 203942.5 |
| trad_total_debt_to_income_ratio | 580 | 5.8 | -0.31 | 0.97 |
| trad_bankruptcy_history | 540 | 5.4 | 0.0 | 0.0 |
| income | 487 | 4.87 | -34458.75 | 145975.25 |
| trad_loan_amount | 376 | 3.76 | -7667.75 | 54182.25 |
| trad_credit_score | 158 | 1.58 | 447.5 | 715.5 |
| trad_interest_rate | 123 | 1.23 | 0.12 | 0.34 |
| age | 11 | 0.11 | 10.5 | 78.5 |
| trad_experience | 10 | 0,1 | -11.5 | 56.5 |
| trad_length_of_credit_history | 0 | 0.0 | -13.0 | 43.0 |
| alt_avg_discount_used | 0 | 0.0 | -0.23 | 0.73 |
| alt_online_purchases | 0 | 0.0 | -48.5 | 147.5 |
| alt_in_store_purchases | 0 | 0.0 | -51.0 | 149.0 |
| alt_avg_transaction_value | 0 | 0.0 | -232.15 | 751.24 |
| alt_days_since_last_purchase | 0 | 0.0 | -179.5 | 544.5 |

*Figura 26 - Tabella di sintesi per la gestione di Outlier*

Nel complesso la quota di outlier è ridotta (<5%), fanno eccezione Income e Trad Loan Amount, con code marcate e incidenze superiori al 15%. Nella versione destinata alla logistica sono state applicate trasformazioni logaritmiche e la rimozione dei valori estremi per attenuarne l'impatto sulle stime. In questo modo si limita la distorsione nei modelli lineari e, al contempo, si preserva l'intera variabilità informativa dove gli algoritmi sono meno sensibili agli outlier, garantendo confronti omogenei tra i modelli.

### 4.5 Confronto predittivo tra modelli di Machine learning nel credit scoring

Dopo la fase di analisi descrittiva, che ha evidenziato la forte capacità discriminante di alcune variabili tradizionali e il potenziale informativo di alcune variabili alternative, si passa alla fase cruciale dello studio: la valutazione predittiva.

Questa fase è progettata in coerenza con gli obiettivi specifici già illustrati nel paragrafo 4.1, con l'intento di verificare se l'integrazione di variabili alternative possa incrementare le prestazioni dei modelli e migliorare l'identificazione di soggetti meritevoli, e di valutare se algoritmi di ML più complessi, siano in grado di sfruttare meglio tali informazioni rispetto alla regressione logistica, modello interpretabile di riferimento.

Nella fase di sperimentazione sono stati messi a confronto tre algoritmi rappresentativi di approcci differenti: la regressione logistica, modello lineare interpretabile utilizzato come benchmark; la Random Forest, un *ensemble* di alberi decisionali capace di cogliere relazioni non lineari e interazioni complesse; e infine XGBoost, algoritmo di *boosting* ad alte prestazioni particolarmente efficace in scenari predittivi complessi.

Ciascun modello è stato applicato a tre diverse configurazioni di input: una versione tradizionale, basata unicamente su variabili creditizio-finanziarie; una configurazione alternativa, che utilizza esclusivamente variabili comportamentali e transazionali; e una versione unificata, che combina le due tipologie di informazioni.

La regressione logistica è stata addestrata su un dataset trasformato e ottimizzato per modelli parametrici, mentre Random Forest e XGBoost hanno operato sul dataset completo, sfruttando la loro robustezza nella gestione di variabili numerose, ridondanti e potenzialmente rumorose.

Per la valutazione delle prestazioni dei modelli sono state utilizzate cinque metriche statistiche di riferimento: accuracy, recall, precision, F1-score e AUC-ROC.

L'accuracy, (TP + TN) / (TP + TN + FP + FN), misura la proporzione complessiva di osservazioni correttamente classificate sul totale delle osservazioni. Dove TP sono i veri positivi, TN i veri negativi, FP i falsi positivi e FN i falsi negativi. Sebbene sia un indicatore intuitivo e di facile interpretazione, può risultare poco informativo in presenza di classi sbilanciate, in quanto valori elevati possono essere ottenuti anche da modelli che trascurano del tutto la classe meno frequente.

Il recall (o sensibilità), TP / (TP + FN), rappresenta la capacità del modello di identificare correttamente i casi positivi. Un valore elevato indica che il modello riesce a rilevare gran

parte degli elementi appartenenti alla classe positiva, mentre un valore basso segnala una quota significativa di falsi negativi.
La precision, TP / (TP + FP), esprime la proporzione di osservazioni realmente positive tra quelle che il modello classifica come tali. Un'alta precisione implica un basso numero di falsi positivi, mentre un valore basso indica che molte delle previsioni positive del modello sono in realtà errate.
L'F1 score, 2 × (Precision × Recall) / (Precision + Recall), combina in un unico indicatore precision e recall calcolando la loro media armonica. Questa metrica è particolarmente utile quando è necessario trovare un compromesso tra la capacità di individuare correttamente i positivi e quella di minimizzare gli errori di classificazione.
Infine, l'AUC-ROC (Area Under the Curve – Receiver Operating Characteristic) fornisce una misura sintetica della capacità discriminante di un modello su tutte le possibili soglie decisionali. Un valore pari a 0.5 indica prestazioni equivalenti a una scelta casuale, valori intorno a 0.8 sono considerati buoni, mentre valori pari o superiori a 0.9 sono ritenuti eccellenti.

### 4.5.1 Regressione Logistica

Il primo modello testato è stata la regressione logistica, pur essendo un modello lineare semplice, rappresenta uno standard consolidato nel settore per l'interpretabilità dei coefficienti e per la capacità di fornire una base di confronto con approcci più sofisticati. I risultati mostrano come il modello tradizionale ottenga le migliori prestazioni complessive, con un'accuratezza del 75.1% e un'AUC di 0.735. L'unificato si colloca su valori molto simili (74.7% di accuratezza e 0.733 di AUC), indicando che l'aggiunta delle variabili alternative non modifica in modo sostanziale la capacità predittiva. Al contrario, il modello basato unicamente su variabili alternative raggiunge performance nettamente inferiori: il recall è nullo, l'F1-score pari a zero e l'AUC appena sopra la casualità (0.510). Questi dati sono sintetizzati nella *Figura 27*.

| Modello Logit | Accuracy | Recall | Precision | F1 | AUC |
|---|---|---|---|---|---|
| Tradizionale | 0.751 | 0.201 | 0.607 | 0.302 | 0.735 |
| Alternativo | 0.732 | 0.000 | 0.000 | 0.000 | 0.510 |
| Unificato | 0.747 | 0.195 | 0.586 | 0.293 | 0.733 |

*Figura 27 - Confronto delle performance dei modelli Logit*

La curva ROC conferma questa lettura: le curve dei modelli tradizionale e unificato si collocano nettamente al di sopra della diagonale casuale, mentre quella del modello alternativo vi si avvicina in modo preoccupante, segnalando un'incapacità quasi totale di discriminazione.

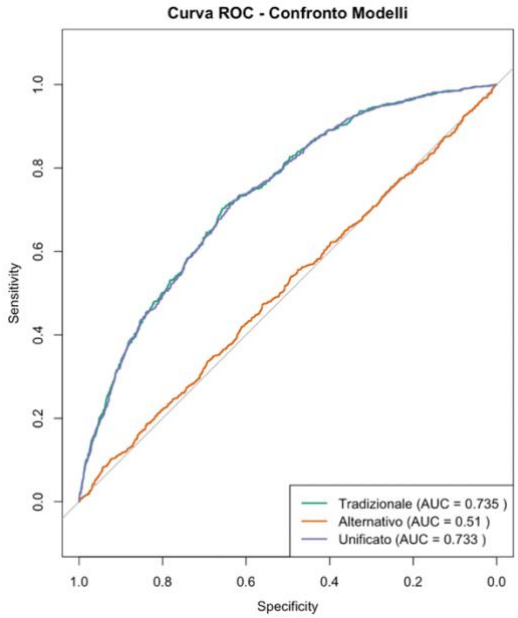



*Figura 28 - Confronto della curva ROC dei modelli Logit*

Le variabili tradizionali sono i predittori più solidi: credit score, storia creditizia e patrimonializzazione aumentano la probabilità di approvazione, mentre bancarotta e importi richiesti elevati la riducono. Le variabili alternative risultano non significative (odds ratio ≈ 1) e la loro inclusione non cambia le prestazioni della logistica. Eventuali effetti non lineari dei dati alternativi saranno valutati con modelli non lineari nelle sezioni successive.

### 4.5.2 Random Forest

Dopo la regressione logistica come benchmark lineare, il passo successivo è l'applicazione del modello Random Forest, con l'obiettivo di verificare se un modello più complesso colga relazioni non lineari e valorizzi meglio le informazioni disponibili.
I risultati riportati in Figura 29 mostrano un netto miglioramento rispetto alla logistica: con le variabili tradizionali l'accuracy è 0.84 e l'AUC 0.865, il recall passa a 52.5%, nella logistica non superava il 20%. Il modello unificato migliora leggermente rispetto alla logistica (recall 23.6%, AUC 0.774), ma non supera il tradizionale. L'approccio alternativo resta debole, AUC 0.606, recall prossimo allo zero.

| Modello RF | Accuracy | Recall | Precision | F1 | AUC |
|---|---|---|---|---|---|
| **Tradizionale** | 0.840 | 0.525 | 0.788 | 0.630 | 0.865 |
| **Alternativo** | 0.741 | 0.0064 | 0.833 | 0.0127 | 0.606 |
| **Unificato** | 0.773 | 0.236 | 0.687 | 0.351 | 0.774 |

*Figura 29 - Confronto delle performance dei modelli Random Forest*

L'analisi dell'importanza delle variabili (*Figura 30*) conferma il ruolo centrale di income, loan amount, net worth e degli indicatori di storia creditizia. L'importanza è calcolata tramite Mean Decrease Gini (MDG), che misura il contributo medio di ogni variabile alla riduzione dell'impurità nei nodi: in aggregato, le variabili citate generano gli split più informativi, mentre molte variabili alternative forniscono contributi minori.

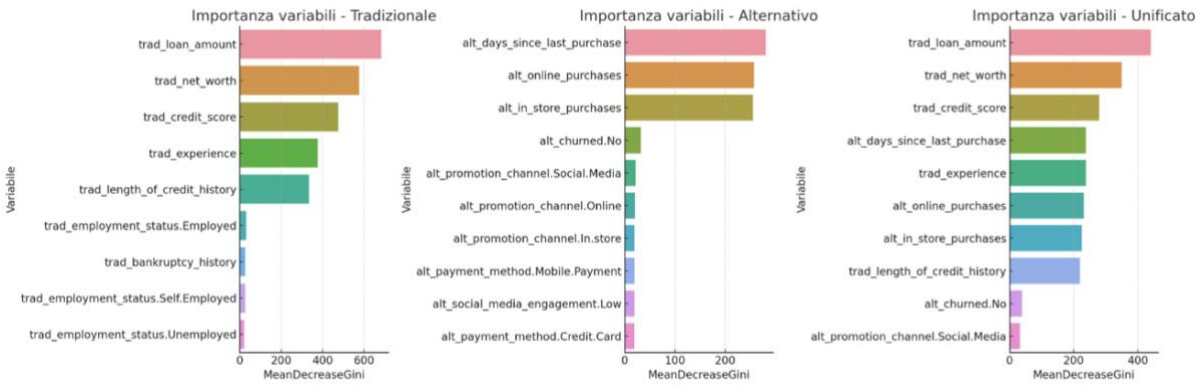



*Figura 30 - Analisi dell'importanza delle variabili nei modelli Random Forest*

Infine, la curva ROC evidenzia la netta superiorità della configurazione tradizionale, seguita dal modello unificato che resta discreto ma meno performante, mentre quello alternativo si avvicina alla casualità. Nel complesso, la Random Forest conferma quanto già osservato con la logistica. La differenza principale rispetto al benchmark lineare è che l'algoritmo ad alberi riesce a sfruttare meglio le informazioni creditizie, innalzando recall e AUC, ma senza modificare la gerarchia tra le diverse tipologie di input.

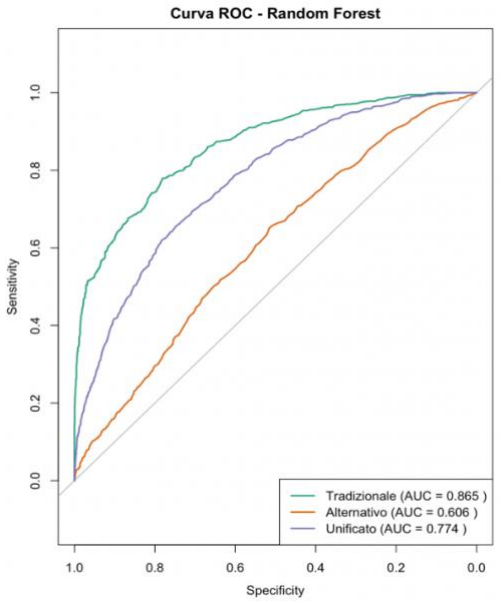



*Figura 31- Confronto della curva ROC dei modelli Random Forest*

### 4.5.3 XGBoost

In continuità con quanto osservato per la regressione logistica e la Random Forest, l'algoritmo XGBoost è stato applicato nelle tre configurazioni, per verificare se il *boosting* riuscisse a sfruttare meglio le informazioni disponibili. Le prestazioni sintetizzate in *Figura 32* mostrano un quadro coerente: il tradizionale è il più affidabile (Accuracy 0.786, Precision 0.696, Recall 0.361, AUC 0.806), l'unificato si colloca poco sotto (Accuracy 0.768, F1 0.419, AUC 0.768), segnalando che l'aggiunta di variabili non convenzionali non porta un guadagno sostanziale e può introdurre rumore, l'alternativo resta debole (Accuracy 0.715, Recall 0.081, AUC 0.577).

| Modello XGBoost | Accuracy | Recall | Precision | F1 | AUC |
|---|---|---|---|---|---|
| Tradizionale | 0.786 | 0.361 | 0.696 | 0.476 | 0.806 |
| Alternativo | 0.715 | 0.081 | 0.363 | 0.132 | 0.577 |
| Unificato | 0.768 | 0.312 | 0.637 | 0.419 | 0.768 |

*Figura 32 - Confronto delle performance dei modelli XGBoost*

La ROC conferma quanto visto fin ora: tradizionale sopra la diagonale e prevalente, unificato inferiore ma ancora discreto, alternativo prossimo alla casualità. In sintesi, XGBoost migliora la discriminazione sul setup tradizionale (AUC > 0.80) senza modificare la gerarchia tra le tre configurazioni; per la lettura comparativa complessiva si rimanda al paragrafo 4.6.

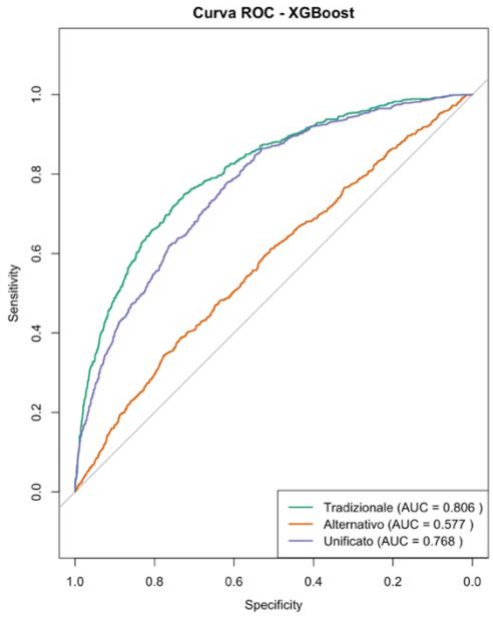



*Figura 33 - Confronto della curva ROC dei modelli XGBoost*

## 4.6 Interpretabilità dei modelli tramite XAI

Le analisi condotte nei paragrafi precedenti hanno evidenziato come modelli avanzati di ML, quali Random Forest e XGBoost, siano in grado di migliorare le metriche predittive rispetto ad approcci più semplici. Tuttavia, fermarsi al solo miglioramento delle prestazioni predittive non è sufficiente: sia Random Forest sia XGBoost restano modelli *black box*, caratterizzati da un processo decisionale complesso e non immediatamente comprensibile. Nel contesto del credit scoring, ciò rappresenta un limite sostanziale,

poiché regolatori, istituti di credito e clienti richiedono decisioni trasparenti, giustificabili e verificabili.

Per questo motivo si è scelto di introdurre tecniche di Explainable AI (XAI), al fine di valutare se strumenti come SHAP e LIME, applicati al dataset unificato, siano in grado di rendere interpretabili modelli complessi e quindi effettivamente applicabili in scenari reali di concessione del credito.

Un primo passo è stato quello di osservare le *feature importance*, ossia una misura aggregata dell'importanza relativa di ciascuna variabile nel modello. La *Figura 34* mostra l'importanza delle variabili per Random Forest e per XGBoost. In entrambi i casi emergono chiaramente alcune variabili dominanti, come il reddito (income), l'ammontare del prestito richiesto (loan amount) e il patrimonio netto (net worth). Questi grafici forniscono una visione sintetica e immediata delle variabili più influenti, ma restano limitati: mostrano infatti soltanto un peso medio, senza chiarire se l'impatto di una variabile sia positivo o negativo e senza dare informazioni sulla variabilità degli effetti tra i diversi individui del dataset.

Dai grafici emerge che le due rappresentazioni adottano metriche differenti. Per Random Forest è stata utilizzata la *Permutation Importance*, calcolata come perdita di AUC quando una variabile viene casualmente permutata (*One minus AUC loss*): in questo modo si misura in che misura la variabile contribuisce alla capacità predittiva complessiva del modello. Per XGBoost, invece, si è preferito ricorrere ai valori medi assoluti di SHAP (*Mean |SHAP|*), che quantificano direttamente l'impatto medio delle variabili sulle predizioni e permettono anche successive analisi locali.

La scelta di metriche diverse risponde a ragioni pratiche: Random Forest non offre in modo diretto un calcolo dei valori SHAP efficiente, mentre XGBoost è invece particolarmente adatto a questo approccio, poiché i valori SHAP vengono calcolati in maniera esatta e relativamente rapida grazie alla struttura ad alberi potenziati. Per questo, nel caso della Random Forest, si è fatto ricorso a una misura più tradizionale come la permutazione, mentre per XGBoost si è potuto sfruttare appieno la potenza dello strumento SHAP.

Le due metriche non sono quindi perfettamente sovrapponibili, ma convergono nell'identificare lo stesso nucleo di variabili decisive. Questo rafforza l'affidabilità del risultato e conferma, in linea con quanto discusso nei paragrafi precedenti, la netta

superiorità delle variabili creditizie tradizionali rispetto a quelle alternative. È importante sottolineare come queste evidenze siano perfettamente coerenti con quanto discusso nei paragrafi precedenti, dove l'analisi delle performance aveva già mostrato la netta superiorità delle variabili creditizie tradizionali rispetto a quelle alternative.

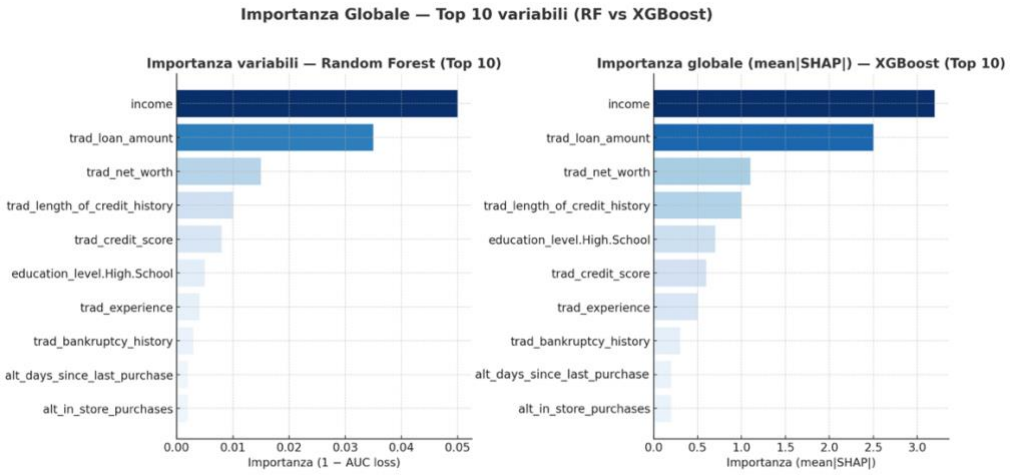



*Figura 34 - Importanza Globale 10 variabili RF vs XGBoost*

Per andare oltre la sola importanza aggregata si è fatto ricorso a SHAP, che permette di spiegare in dettaglio il contributo marginale di ciascuna variabile alla predizione. È utile distinguere due livelli di analisi: quella globale, che sintetizza come le variabili agiscono mediamente sul comportamento del modello in tutto il campione, e quella locale, che si concentra invece sulla decisione presa per un singolo individuo.

Nel caso di XGBoost, per l'analisi globale si è scelto di utilizzare un *beeswarm plot*.

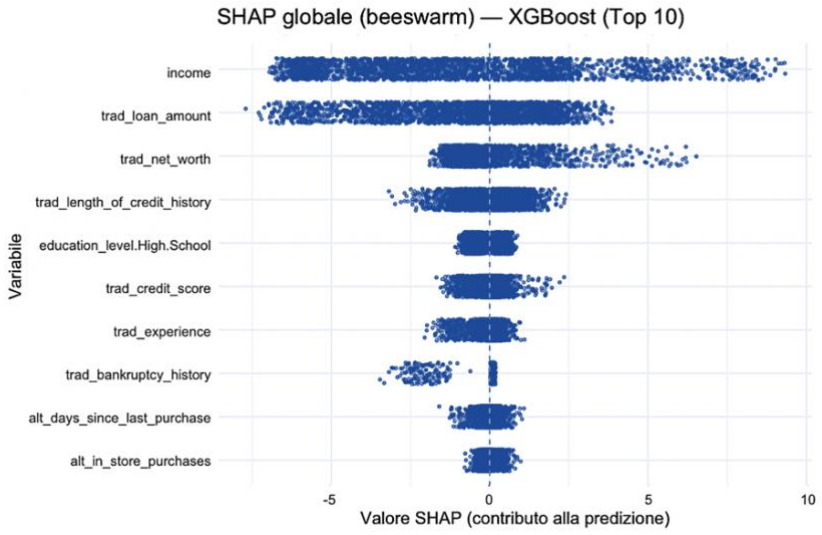



*Figura 35 - SHAP globale (beeswarm) XGBoost*

Questo tipo di grafico rappresenta in modo congiunto l'importanza, la direzione e la distribuzione degli effetti. Ogni punto corrisponde a un individuo del dataset e mostra quanto una variabile abbia inciso sulla sua predizione. I punti collocati a destra dell'asse verticale indicano un impatto positivo, cioè una spinta verso la classificazione come affidabile; quelli a sinistra rappresentano invece un impatto negativo, che spinge verso l'esito opposto. L'ampiezza della nuvola di punti riflette la variabilità degli effetti tra i diversi individui. In questo modo *il beeswarm* arricchisce la lettura rispetto alla semplice *feature importance*, offrendo una visione tridimensionale: peso, direzione e distribuzione. Per mantenere il grafico leggibile si è scelto di rappresentare soltanto le dieci variabili più influenti. Dai risultati emerge con chiarezza che variabili come il reddito e il patrimonio elevato spingono le decisioni del modello verso esiti positivi, mentre prestiti troppo consistenti o credit score ridotti tendono a penalizzare la classificazione.
Accanto a questa prospettiva globale, è stata condotta un'analisi locale con riferimento al modello Random Forest, rappresentata attraverso un *waterfall plot* (*Figura 36*).

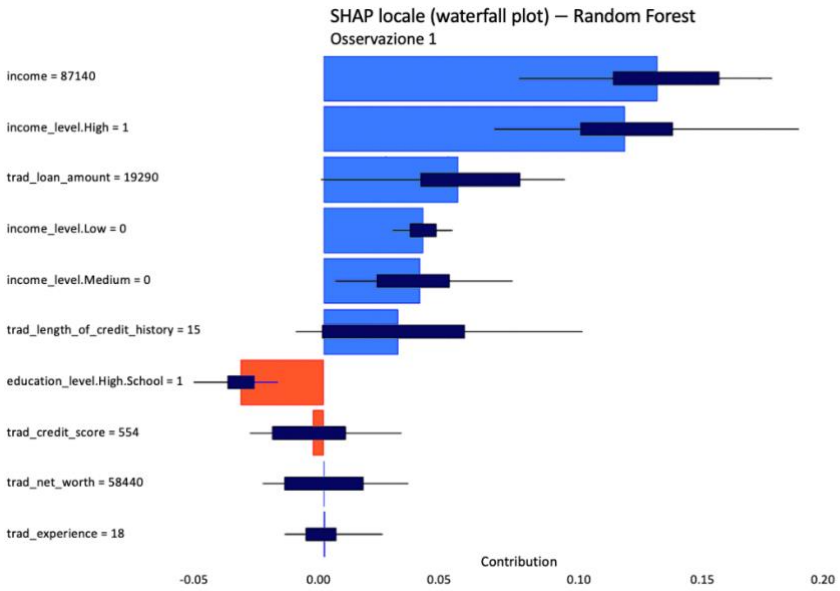



*Figura 36 - SHAP locale (waterfall plot) Random Forest*

Questo grafico consente di ricostruire, passo dopo passo, come il modello sia giunto alla decisione per un singolo individuo. La logica è semplice: si parte dalla *baseline*, ossia dalla probabilità media di affidabilità calcolata sul dataset, e si mostrano i contributi delle variabili specifiche del cliente. Le barre verdi rappresentano i fattori che aumentano la

probabilità di classificazione positiva, mentre le barre rosse indicano quelli che spingono in direzione contraria. Nel caso riportato, il reddito elevato e una lunga storia creditizia hanno spostato la predizione verso un esito positivo, mentre il basso livello di istruzione ha inciso negativamente. Questo tipo di rappresentazione è particolarmente utile perché traduce un output complesso in una sequenza di passaggi chiari e comprensibili, comunicabili anche a stakeholder non tecnici.

L'analisi è stata completata con LIME, che si differenzia da SHAP per l'approccio seguito. Se SHAP fornisce spiegazioni matematicamente rigorose, LIME privilegia la leggibilità immediata. Per spiegare la predizione su un individuo, genera osservazioni sintetiche simili a quella del cliente analizzato e vi adatta un modello lineare interpretabile. In questo modo le decisioni del modello complesso vengono tradotte in regole semplici del tipo "se il reddito è superiore a una certa soglia, allora la probabilità di affidabilità aumenta". Le *Figure 37 e 38* riportano quattro casi studio analizzati rispettivamente con Random Forest e con XGBoost, mostrando per ciascuno la classe assegnata, la probabilità stimata e *l'explanation fit*, cioè quanto bene il modello locale riesca a replicare il comportamento di quello complesso. Gli esempi confermano la coerenza con i risultati ottenuti da SHAP: redditi elevati favoriscono le decisioni positive, mentre prestiti troppo alti o credit score bassi guidano le classificazioni negative.

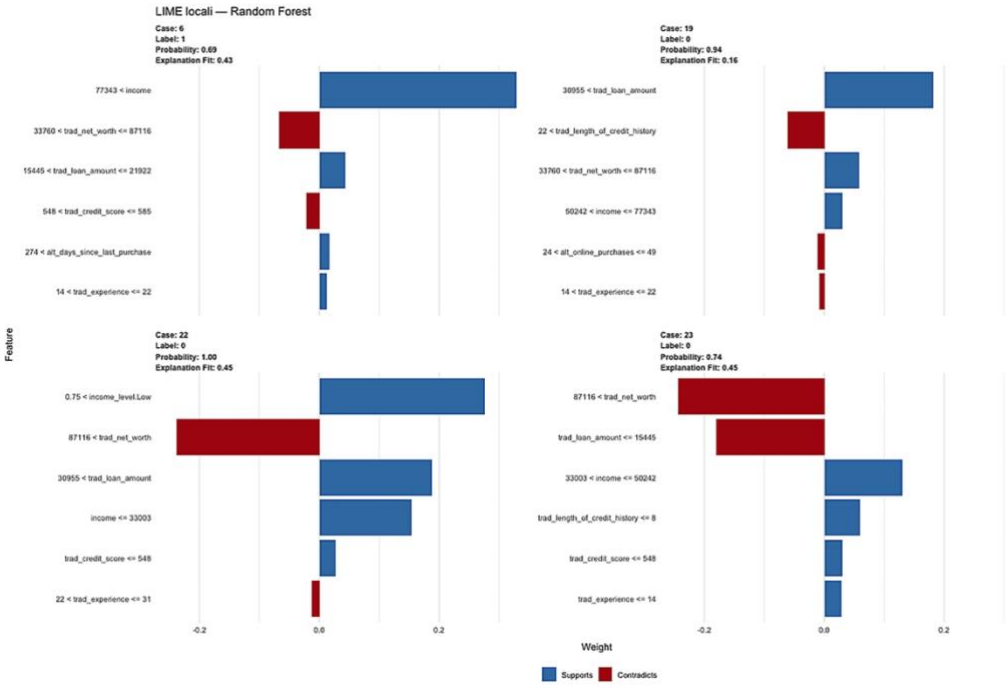



*Figura 37 - Lime locali Random Forest*

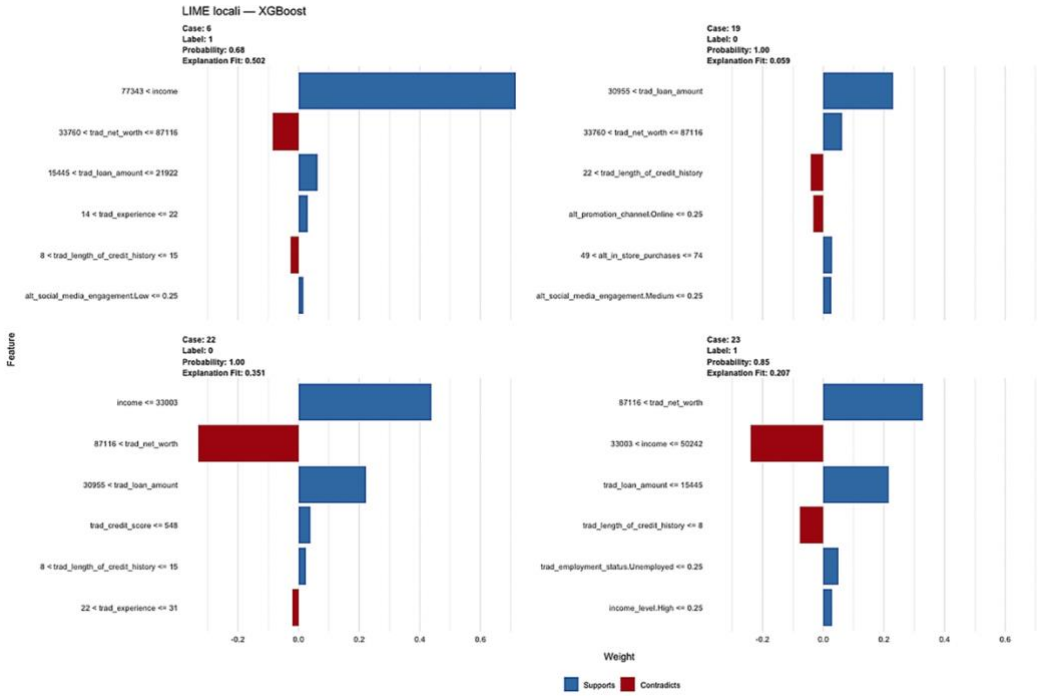



*Figura 38 - Lime locali XGBoost*

In conclusione, l'integrazione di SHAP e LIME ha permesso di superare l'opacità dei modelli. Le *feature importance* hanno fornito un primo quadro sintetico, il *beeswarm plot* ha arricchito la visione globale introducendo la direzione e la distribuzione degli effetti; il *waterfall plot* ha reso trasparenti le decisioni sui singoli individui, infine LIME ha tradotto la logica complessa in regole intuitive e comunicabili. Per facilitare il confronto tra i due approcci, la tabella sottostante riassume in modo compatto le differenze operative tra SHAP e LIME. Questa sintesi orienta la scelta dello strumento in base al contesto d'uso.

Complessivamente, Random Forest e XGBoost non sono più soltanto modelli performanti, ma strumenti trasparenti e giustificabili, pienamente applicabili a un settore regolamentato come quello del credit scoring.

| Confronto modelli XAI | SHAP | LIME |
|---|---|---|
| **Principio** | Metodo matematico basato sui valori di Shapley. | Modello semplice costruito intorno a un singolo caso. |
| **Output** | Quanto e in che direzione ogni variabile ha influenzato la previsione (per ogni persona) e offre anche un quadro globale. | Una spiegazione locale del caso: regole semplici di tipo condizionale. |
| **Vantaggi** | Più rigoroso e coerente. | Più rapido e intuitivo. |
| **Limiti** | Può essere più lento; serve scegliere un sotto-campione coerente con il dataset. | Può essere variabile tra esecuzioni; dipende da parametri locali. |
| **Applicazione** | Audit, policy, documenti ufficiali; modelli ad alberi (es. XGBoost). | Spiegare un caso alla volta a un pubblico non tecnico; demo rapide. |

*Figura 39 - Confronto Operativo tra modelli XAI*

## 4.7 Risultati e analisi critica della ricerca

L'analisi sperimentale condotta non è esente da limiti che è importante esplicitare, sia per inquadrare correttamente la portata delle conclusioni, sia per orientare eventuali sviluppi futuri. Un primo limite riguarda la natura stessa del dataset: i dati a disposizione non provenivano da un archivio unico e organico, ma dall'unione di due fonti differenti, rispettivamente una con variabili tradizionali e una con dati alternativi di natura comportamentale e transazionale. L'integrazione è stata possibile grazie a sole tre variabili comuni e ha richiesto un allineamento che non garantisce perfetta coerenza in termini di definizione, modalità di raccolta o distribuzione temporale. È plausibile che tale passaggio abbia introdotto rumore statistico, in particolare nella sezione relativa alle variabili alternative.

Un secondo limite riguarda la natura dei dati utilizzati. Pur essendo pubblicamente disponibili su piattaforme come Kaggle, essi non provengono da archivi ufficiali né da fonti certificate da autorità di vigilanza o da istituti finanziari. Si tratta dunque di basi dati adatte a fini sperimentali e didattici, ma che non garantiscono lo stesso livello di affidabilità, accuratezza e rappresentatività di dataset istituzionali.

Le prestazioni osservate, pertanto, vanno intese come evidenze significative in un contesto sperimentale, ma non come prova definitiva da proiettare senza cautele nel mondo bancario.

Infine, un ulteriore limite riguarda le tipologie dei modelli analizzati. Lo studio si è concentrato su tre approcci, regressione logistica, Random Forest e XGBoost, che, sebbene rappresentino alcuni tra i modelli più consolidati e performanti nel credit scoring, non esauriscono l'intero spettro delle metodologie oggi disponibili. Reti neurali profonde, tecniche di *representation learning* o modelli ibridi non sono stati inclusi, e ciò restringe l'orizzonte della sperimentazione.

Nel complesso, le analisi effettuate risultano sufficienti a delineare un quadro solido e significativo delle opportunità e dei rischi legati all'integrazione di dati alternativi e algoritmi avanzati nel processo di valutazione del merito creditizio.

Passando ai risultati, l'analisi comparativa ha messo in luce differenze nette tra i modelli e, soprattutto, ha evidenziato il ruolo contrastante dei dati alternativi. La regressione logistica si conferma un riferimento per il credit scoring: semplicità, trasparenza dei coefficienti e robustezza statistica ne fanno lo strumento preferito da regolatori e operatori. Ogni coefficiente ha un significato immediato e comunicabile, in linea con i requisiti di spiegabilità del quadro normativo europeo e nazionale. Sul piano delle performance, tuttavia, il modello lineare mostra limiti: la capacità predittiva è inferiore rispetto agli algoritmi più avanzati, in particolare per AUC e recall.

Random Forest e XGBoost, al contrario, hanno dimostrato maggiore efficacia nel discriminare i soggetti affidabili da quelli insolventi grazie alla gestione di non linearità e interazioni tra variabili; il recall sensibilmente più alto indica una migliore attitudine a intercettare i profili rischiosi.

Il risultato più rilevante della sperimentazione riguarda però l'inclusione dei dati alternativi: la loro aggiunta non ha migliorato le performance dei modelli complessi e, in alcuni casi, le ha peggiorate.

Questo esito non smentisce il potenziale dei dati alternativi in assoluto, ma segnala che la loro utilità dipende in modo cruciale da qualità, fonte e contesto di raccolta. In questo studio, la natura non certificata e la ridotta granularità verosimilmente ne hanno limitato il contributo; in scenari reali, con dati ufficiali, ricchi di dettaglio e tempestivi, il loro ruolo nel ridurre l'asimmetria informativa e favorire l'inclusione finanziaria potrebbe essere ben diverso.

Poiché i modelli più sofisticati hanno mostrato maggiore accuratezza e una migliore capacità di individuare i debitori rischiosi, è stato necessario renderli più comprensibili

con tecniche di Explainable AI. L'opacità di questi algoritmi, infatti, può limitarne l'uso in ambiti regolamentati, dove servono trasparenza ed equità. Le analisi con SHAP e LIME, impiegate a scopo illustrativo, hanno permesso di "aprire" i modelli e capire quali variabili guidano le classificazioni: le variabili tradizionali restano centrali, mentre i dati alternativi hanno avuto un peso marginale nel dataset considerato. I risultati si confermano sia a livello globale sia a livello locale su singole osservazioni ad alto rischio, e si sono mantenuti stabili sui campioni *out-of-sample*. In questo senso, la XAI non solo ha rafforzato la fiducia nei modelli complessi, ma ha anche supportato verifiche di equità (assenza di proxy sensibili evidenti) e la tracciabilità delle decisioni, offrendo basi per la revisione umana e per una comunicazione motivazionale verso il cliente, in linea con i principi di *fairness*, *accountability* e trasparenza richiesti nelle valutazioni di merito creditizio.

L'analisi dei risultati, insieme alle cautele metodologiche evidenziate, definiscono i confini entro cui i risultati possono essere letti e comparati, senza estenderne la portata oltre il perimetro sperimentale. Il capitolo si chiude dunque, con questa evidenza descrittiva, rimandando alle Conclusioni la discussione interpretativa e le eventuali implicazioni.

# Conclusioni

La domanda di ricerca da cui questo lavoro ha preso avvio era la seguente: i dati alternativi, integrati con le tecniche di ML, migliorano davvero la valutazione del merito creditizio e contribuiscono a ridurre le esclusioni ingiustificate?

L'analisi teorica e la sperimentazione empirica hanno mostrato che, nelle condizioni considerate, il loro contributo non è stato decisivo. Ciò non significa che il potenziale dei dati non convenzionali sia da escludere, ma che la loro efficacia dipende in modo cruciale dalla qualità, dalla provenienza e dal contesto di applicazione.

Le considerazioni finali della sperimentazione non mirano a contrapporre dati tradizionali e dati alternativi, ma a stabilire in quali condizioni e per quali soggetti ciascuna tipologia informativa offre un contributo effettivo. Per i richiedenti con storico bancario e reddituale solido, le informazioni tradizionali restano il punto di riferimento più affidabile e predittivo. Per i *credit invisible*, giovani, lavoratori atipici, nuovi residenti, i dati alternativi possono invece costituire la prima evidenza informativa, a condizione che siano granulari, certificati e tracciabili. In assenza di questi requisiti, l'apporto alternativo si attenua fino a diventare nullo.

Ne deriva un principio operativo semplice: integrare sì, ma condizionatamente. L'unione fra set tradizionale e alternativo è sensata quando le variabili aggiuntive superano soglie minime di affidabilità e pertinenza, diversamente, non migliora le performance e rischia di confonderle.

Il passaggio davvero abilitante è l'accesso a fonti alternative qualificate, utenze e servizi essenziali, flussi di pagamento certificati, informazioni contributive e fiscali, dati derivanti da *open banking* con canali consolidati, capaci di offrire continuità temporale, responsabilità del trattamento e verificabilità.

Questo cambiamento non comporta uno sforzo solo tecnologico. Richiede infrastrutture di dati trasparenti e interconnesse, standard di qualità condivisi a livello Europeo, controlli ex ante sulla provenienza e monitoraggio continuo di dati e modelli, e richiede un quadro regolatorio ed etico che traduca privacy, equità e trasparenza in prassi verificabili: accordi di condivisione, audit periodici, metriche di *fairness* dichiarate. Nel contesto italiano, ad oggi particolarmente prudente in materia, il percorso potrebbe essere realistico se le istituzioni abilitassero modalità d'uso chiare per fonti affidabili e gli

operatori elevassero competenze e governance: non un allentamento delle tutele, ma il loro aggiornamento ad archivi informativi più ricchi.

Un secondo elemento che emerge con chiarezza, riguarda l'impiego dei modelli di ML più avanzati e la questione della loro spiegabilità. Gli algoritmi che hanno mostrato le migliori performance nella sperimentazione risultano spesso opachi. Nel merito creditizio, però, la legittimazione sociale e giuridica dipende dalla possibilità di motivare le decisioni. In questa prospettiva, le tecniche di Explainable AI, applicate nello studio, offrono l'opportunità di utilizzare algoritmi complessi e difficilmente interpretabili rendendoli al tempo stesso fruibili e compatibili con le esigenze del contesto creditizio.

Il contributo di questo lavoro è duplice. Da un lato, pone ordine nelle aspettative: chiarisce che i dati alternativi non rimpiazzano i tradizionali, ma li affiancano con ruoli diversi.

Dall'altro, evidenzia la priorità operativa per sviluppi futuri: senza fonti ufficiali e governate, l'innovazione informativa non si traduce automaticamente né in migliore previsione né in reale inclusione. L'obiettivo non è aumentare genericamente le approvazioni, ma ridurre le esclusioni ingiustificate a parità di rischio, monitorando effetti differenziati tra categorie di clientela e correggendo eventuali distorsioni.

Ne derivano due obiettivi concreti per il futuro: per i regolatori, definire standard di qualità e di equità, promuovere canali sicuri di condivisione e prevedere processi di spiegabilità. Per gli operatori, adottare un modello "a geometria variabile" dei dati, con criteri trasparenti su cosa può rientrare o meno nel calcolo del credit scoring e da dove proviene.

Per concludere, se qualità delle fonti, governance dei modelli e controllo delle decisioni sono garantite, i dati alternativi possono diventare una leva di inclusione, soprattutto per i *credit invisible*, e l'innovazione potrebbe radicarsi anche in Italia senza sacrificare l'equilibrio tra efficienza, tutela e fiducia.

# Bibliografia


Abdou, H. A., Pointon, J. (2011). *Credit scoring, statistical techniques and evaluation criteria: A review of the literature.* Intelligent Systems in Accounting, Finance and Management.

Abellán, J., Castellano, J. G. (2017). *A comparative study on base classifiers in ensemble methods for credit scoring.* Expert Systems with Applications, 73, 1–12.

Adadi, A., Berrada, M. (2018). *Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)*. IEEE Access, 6, 52138–52160.

Agarwal, S., Alok, S., et al. (2019). *Financial inclusion and alternate credit scoring: Role of big data and machine learning in fintech.* Indian School of Business.

Ammannati, L., Greco, G. L. (2021). *Piattaforme digitali, algoritmi e big data: Il caso del credit scoring.* Rivista Trimestrale di Diritto dell'Economia*, 2*, 290–325.

Antonogeorgos, G., Panagiotakos, D. B., et al. (2009). *Logistic regression and linear discriminant analyses in evaluating factors associated with asthma prevalence among 10- to 12-years-old children: Divergence and similarity of the two statistical methods.* International Journal of Pediatrics.

Bhattacharya, A. (2022). *Foundational concepts of explainability techniques.* In *Applied machine learning explainability techniques* (pp. 1–26). Packt Publishing.

Bianchini, G. (2024). *Il quadro in materia di credit scoring, tra l'AI Act e le sentenze SCHUFA.* NT+ Diritto – Il Sole 24 Ore.

Bonaccorsi di Patti, E., Calabresi, F., et al. (2022). *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano*. Questioni di Economia e Finanza, n. 721. Banca d'Italia.

Busacca, A. (2018). *Le "categorie particolari di dati" ex art. 9 GDPR: Divieti, eccezioni e limiti alle attività di trattamento.* Ordine internazionale e diritti umani.

Cecchinato, E. (2023). *Note sulla disciplina della verifica del merito creditizio: per una sua rilettura alla luce della buona fede precontrattuale*. Rivista di Diritto Bancario (III).

Centraco, G., Dore, G. M. D. (2024). Il Sistema di Credito Sociale cinese: tecnologia come strumento di sorveglianza e persuasione. *OrizzonteCina, 15*(1), 61–75.

Chen, T., Guestrin, C. (2016). *XGBoost: A scalable tree boosting system.* In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Chemlali, L., Benseddik, L. (2024). *Credit scoring under the GDPR: Insights from the CJEU's SCHUFA case.* Journal of Data Protection & Privacy, 7(2), 152–165.
Corte di Giustizia dell'Unione Europea. (2023). Sentenza del 7 dicembre 2023, C-634/21. EUR-Lex.

De Ville, B. (2013). *Decision trees.* WIREs Computational Statistics, 5(6).

Demaio, L. M., Vella, V., et al. (2020). *Explainable AI for interpretable credit scoring.* arXiv preprint arXiv:2012.03749.

Gao, H., Kou, G. (2024). *Machine learning in business and finance: A literature review and research opportunities.* Financial Innovation, 10(86).

Gambella, C., Ghaddar, B., et al. (2021). *Optimization problems for machine learning: A survey.* European Journal of Operational Research, 290.

Gaggero, A., Valenza, M. (2024). *Intelligenza artificiale e merito creditizio: nuovi orizzonti normativi tra AI Act e GDPR.* Rivista di Diritto Bancario (III).

Goodfellow, I., Bengio, Y., et al. (2016). *Deep learning*. MIT Press.

Gunnarsson, B. R., Vanden Broucke, S., et al. (2021). *Deep learning for credit scoring: Do or don't?* European Journal of Operational Research, 295(1).

Hayashi, Y. (2022). *Emerging trends in deep learning for credit scoring: A review.* Electronics, 11(11), 3181.

Janiesch, C., Zschech, P., et al. (2021). *Machine learning and deep learning.* Electronic Markets, 31(3).

Jiang, J., Pan, H., et al. (2021). *Predictive model for the 5-year survival status of osteosarcoma patients based on the SEER database and XGBoost algorithm.* Scientific Reports, 11(1).

Johnson, K. N. (2019). *Written testimony before the United States House Committee on Financial Services Task Force on Financial Technology: Examining the use of alternative data in underwriting and credit scoring to expand access to credit.* Tulane University Law School.

Kelly, S., Mirpourian, M. (2021, February). *Algorithmic bias, financial inclusion, and gender: A primer on opening up new credit to women in emerging economies.* Women's World Banking.

Kononenko, I., Kukar, M. (2007). *Measures for evaluating the quality of attributes.* In I. Kononenko, M. Kukar (Eds.), Machine learning and data mining.

Kriegeskorte, N., Golan, T. (2019). *Neural network models and deep learning.* Current Biology, 29(7).

Li, Y., Chen, W. (2020). *A comparative performance assessment of ensemble learning for credit scoring.* Mathematics, 8(10).

Maestri, E. (2024). *L'impatto della normativa europea sull'intelligenza artificiale: una garanzia di sicurezza o un ostacolo all'innovazione?* Agenda 17 – Webmagazine del Laboratorio Design of Science, Università degli Studi di Ferrara.

Markov, A., Seleznyova, Z., et al. (2022). *Credit scoring methods: Latest trends and points to consider.* The Journal of Finance and Data Science.

Mattassoglio, F. (2025). *La Corte di giustizia europea, algoritmi e credit scoring. L'apertura del vaso di Pandora delle società che si "limitano" a elaborare gli scoring.* Dialoghi di Diritto dell'Economia, Università degli Studi di Milano-Bicocca.

Moldovan, D. (2023). *Algorithmic decision making methods for fair credit scoring.* IEEE Access.

Monti, A. C. (2008). *Introduzione alla statistica*. Edizioni Scientifiche Italiane.

Muhamedyev, R. I. (2015). *Machine learning methods: An overview.* Computer Modelling & New Technologies, 19(6).

Nahar, J., Rahaman, M. A., Alauddin, M., et al. (2024). *Big data in credit risk management: A systematic review of transformative practices and future directions.* International Journal of Management Information Systems and Data Science, 1(04), 68–79.

Nallakaruppan, M. K., Balusamy, B., et al. (2024). *An explainable AI framework for credit evaluation and analysis.* Applied Soft Computing, 153, 111307.

Nasteski, V. (2017). *An overview of the supervised machine learning methods.* Horizons: International Scientific Journal, 21(4).

Neuwirth, R. J. (2023). *Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA).* Computer Law & Security Review, 48, 105798.

Nopper, T. K. (2020). *Alternative data and the future of credit scoring.* Data for Progress.

Novelli, C., Casolari, F., et al. (2024). *AI risk assessment: A scenario-based, proportional methodology for the AI Act.* Digital Society, 3.

Paal, B. (2023). Article 22 GDPR: *Credit Scoring Before the CJEU*. *GPLR*, 127-137.

Parlamento Europeo e Consiglio dell'Unione Europea. (1987). *Direttiva 87/102/CEE del Consiglio, del 22 dicembre 1986, per il ravvicinamento delle disposizioni legislative, regolamentari e amministrative degli Stati membri in materia di credito al consumo*.

Parlamento Europeo e Consiglio dell'Unione Europea. (2008). *Direttiva 2008/48/CE del Parlamento europeo e del Consiglio, del 23 aprile 2008, relativa ai contratti di credito ai consumatori e che abroga la direttiva 87/102/CEE*.

Parlamento Europeo e Consiglio dell'Unione Europea. (2014). *Direttiva 2014/17/UE del Parlamento europeo e del Consiglio, del 4 febbraio 2014, in merito ai contratti di credito ai consumatori relativi a beni immobili residenziali e recante modifica delle direttive 2008/48/CE e 2013/36/UE e del regolamento (UE) n. 1093/2010*.

Parlamento Europeo e Consiglio dell'Unione Europea. (2016). *Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati)*.

Parlamento Europeo e Consiglio dell'Unione Europea. (2023). *Direttiva (UE) 2023/2225 del Parlamento europeo e del Consiglio, del 18 ottobre 2023, relativa ai contratti di credito ai consumatori e che abroga la direttiva 2008/48/CE*.

Pérez Martín, A., Pérez-Torregrosa, A., et al. (2018). *Big data techniques to measure credit banking risk in home equity loans*. Journal of Business Research, 89, 118–123.

Poon, M. (2007). *Scorecards as devices for consumer credit: The case of Fair, Isaac & Company Incorporated.* The Sociological Review, 55(s2), 284–306.

Rabitti, M. (2023). *Credit scoring via machine learning e prestito responsabile.* Rivista di Diritto Bancario (I).

Repubblica Italiana. (2010). *Decreto legislativo 13 agosto 2010, n. 141. Attuazione della direttiva 2008/48/CE sui contratti di credito ai consumatori*. Gazzetta Ufficiale della Repubblica Italiana, Serie generale, n. 207 del 4 settembre 2010.

Repubblica Italiana. (2018). *Decreto legislativo 10 agosto 2018, n. 101. Disposizioni per l'adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2016/679...*. Gazzetta Ufficiale della Repubblica Italiana, Serie generale, n. 205 del 4 settembre 2018.

Robustella, L. (2024). *Misure preventive e restitutorie nell'ambito della vigilanza di tutela.* In A. Urbani, R. Natoli, D. Rossano (Eds.), *Quaderni di Ricerca Giuridica della Consulenza Legale della Banca d'Italia – A 30 anni dal Testo unico bancario (1993–2023)* (pp. 251–266). Banca d'Italia.

Salih, A. M., Raisi-Estabragh, Z., et al. (2025). A perspective on explainable artificial intelligence methods: SHAP and LIME. *Advanced Intelligent Systems, 7*(5), 2400304.

Schuett, J. (2023). Risk management in the Artificial Intelligence Act. European Journal of Risk Regulation. Cambridge University Press.

Sciarrone Alibrandi, A., Borello, G., et al. (2019). *Marketplace lending: Verso nuove forme di intermediazione finanziaria?* (*Quaderni FinTech, n. 5*). Commissione Nazionale per le Società e la Borsa (CONSOB).

Shiam, S. A., Hasan, M. M., (2024). *Credit risk prediction using explainable AI.* Journal of Business and Management Studies, 6(2).

Silvano, C. (2024). La nozione di "decisione completamente automatizzata" sotto la lente della Corte di Giustizia: il caso Schufa. *CERIDAP – Rivista interdisciplinare sul diritto delle amministrazioni pubbliche, 4*(ottobre-dicembre).

Sirena, P. (2024). *Tutela dei clienti e regolazione del mercato trent'anni dopo l'emanazione del Testo Unico Bancario.* In A. Urbani, R. Natoli, D. Rossano (Eds.), Quaderni di Ricerca Giuridica della Consulenza Legale della Banca d'Italia – A 30 anni dal Testo unico bancario (1993–2023) (pp. 123–137). Banca d'Italia.

Sullivan, L. (2020). *Nonparametric tests.* Boston University School of Public Health.

Szepannek, G., Lübke, K. (2021). *Facing the challenges of developing fair risk scoring models.* Frontiers in Artificial Intelligence, 4, 681915.

Tripathi, D., Shukla, A. K., et al. (2022). *Credit scoring models using ensemble learning and classification approaches: A comprehensive survey.* Wireless Personal Communications, 123.

Unione Europea. (2024). *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce norme armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale).*

VenkateswaraRao, M., Vellela, S. S., et al. (2023). *Credit investigation and comprehensive risk management system based big data analytics in commercial banking.* IEEE.

Wang, G., Hao, J., et al. (2011). *A comparative assessment of ensemble learning for credit scoring.* Expert Systems with Applications, 38(1).

Wang, H., Li, C., et al. (2019). *Does AI-based credit scoring improve financial inclusion? Evidence from online payday lending.* In 40th International Conference on Information Systems *(ICIS 2019)*. Association for Information Systems.

World Bank Group. (2019). *Credit scoring: Approaches, guidelines*. Washington, DC: World Bank Group.

# Sitografia


Agenda Digitale. (2018, settembre 19). GDPR: tutto ciò che c'è da sapere per essere preparati. https://www.agendadigitale.eu/cittadinanza-digitale/gdpr-tutto-cio-che-ce-da-sapere-per-essere-preparati/#:~:text=In%20data%2019%20settembre%202018,pecuniarie%20gi%C3%A0%20previste%20dal%20GDPR

Agenda Digitale. (2025, luglio 15). AI Act: ci siamo, ecco come plasmerà il futuro dell'intelligenza artificiale in Europa. https://www.agendadigitale.eu/cultura-digitale/ai-act-ci-siamo-ecco-come-plasmera-il-futuro-dellintelligenza-artificiale-in-europa/#:~:text=dell%E2%80%99esercizio%20precedente%20per%20la%20fornitura,in%20risposta%20a%20una%20richiesta

Agenda Digitale. (2025, febbraio 18). Merito creditizio: l'AI riscrive le regole, le norme che ci tutelano. https://www.agendadigitale.eu/sicurezza/privacy/merito-creditizio-lai-riscrive-le-regole-le-norme-che-ci-tutelano/

Agenda Digitale. (2024, febbraio 5). Profilazione e verifiche del merito creditizio: un difficile equilibrio. https://www.agendadigitale.eu/mercati-digitali/profilazione-e-verifiche-del-merito-creditizio-un-difficile-equilibrio/#:~:text=forme%20di%20profilazione%20incontrollate%2C%20che,a%20conseguenze%20imprevedibili%20ed%20ingiuste

Agenda Digitale. (2023, agosto 8). Privacy by design e by default. https://protezionedatipersonali.it/privacy-by-design-e-by-default#:~:text=Il%20principio%20di%20privacy%20by%20default%20(protezione%20per%20impostazione%20predefinita,che%20possono%20accedere%20ai%20dati

Agenda 17. (2024, giugno 16). Dossier AI Act: maggiori i rischi, più rigorose le regole nella nuova legge europea. https://www.agenda17.it/2024/06/16/dossier-ai-act-maggiori-i-rischi-piu-rigorose-le-regole-nella-nuova-legge-europea/

Altalex. (2023, giugno 23). AI Act: l'UE traccia il futuro dell'intelligenza artificiale. https://www.altalex.com/documents/news/2023/06/23/ai-act-ue-traccia-futuro-intelligenza-artificiale

Altalex. (2023, novembre 14). CCD II, prima lettura: direttiva credito al consumo (UE 2023/2225). https://www.altalex.com/documents/2023/11/14/ccd-ii-prima-lettura-direttiva-credito-consumo-ue-2023-2225

Cybersecurity360. (2024). AI Act: scattano i primi divieti, chi rischia le sanzioni e le prossime tappe. https://www.cybersecurity360.it/legal/ai-act-scattano-i-primi-divieti-chi-rischia-le-sanzioni-e-le-prossime-tappe/

Credit-Scoring.co.uk. (2025 luglio 10). EU AI Act and credit scoring. https://www.credit-scoring.co.uk/blog/eu-ai%20act#:~:text=,factors%20that%20influence%20credit%20scores

Dirittobancario.it. (2023). CCD II: la nuova direttiva sul credito al consumo. https://www.dirittobancario.it/art/ccd-ii-la-nuova-direttiva-sul-credito-al-consumo/

Experian Italia. (s.d.). Application scorecards. https://www.experian.it/business/analytics-and-decisioning/decision-analytics/application-scorecards#:~:text=Tutto%20ci%C3%A8%20possibile%20inizialmente,credito%20per%20gli%20istituti%20finanziari

Fabrick. (s.d.). La nostra storia. https://www.fabrick.com/it-it/corporate/storia/

Fhernanb. (s.d.). Random forests (appunti didattici). https://fhernanb.github.io/libro_mod_pred/rand-forests.html

IBM. (2023 novembre 27). Linear discriminant analysis (LDA). https://www.ibm.com/it-it/think/topics/linear-discriminant-analysis#:~:text=L'analisi%20discriminante%20lineare%20(LDA)...

IBM. (2025 maggio 14). Logistic regression. https://www.ibm.com/it-it/think/topics/logistic%20regression#:~:text=La%20regressione%20logistica%20stima.

IBM. (s.d.). Machine learning. https://www.ibm.com/it-it/topics/machine-learning

Kaggle. (s.d.). Loan Default Dataset (Yasser H) [Dataset]. https://www.kaggle.com/datasets/yasserh/loan-default-dataset

Kaggle. (s.d.). Large Retail Data Set for EDA (utkalk) [Dataset]. https://www.kaggle.com/datasets/utkalk/large-retail-data-set-for-eda

Medium. (2017). Types of algorithms in machine learning. https://medium.com/@prasannathupati/types-of-algorithms-in-machine-learning-71b28a60680c

MIT News. (1996). David Durand obituary. https://news.mit.edu/1996/durand

MyFICO. (s.d.). History of the FICO Score. https://www.myfico.com/credit-education/blog/history-of-the-fico-score

Privacy.it. (2002). Comunicazione 2002/443/DEF. https://www.privacy.it/archivio/com2002-443def.html

Privacy.it. (2017, 25 maggio). *Prepararsi al GDPR: 12 passi da fare ora*. https://www.privacy.it/2017/05/25/12-passi-per-prepararsi-al-gdpr/

Ristrutturazioni Aziendali (IlCaso.it). (s.d.). La nuova disciplina europea sul credito al consumo. https://ristrutturazioniaziendali.ilcaso.it/Articolo/580#:~:text=,automatizzate%20e%20richiedere%20l%27intervento%20umano

Rivista della Regolazione dei Mercati. (s.d.). Credito al consumo e regolazione UE. https://www.rivistadellaregolazionedeimercati.it/Article/Archive/index_html?ida=326&idn=23&idi=-1&idu=-1

Starting Finance. (s.d.). Il problem solving: divide et impera. https://startingfinance.com/approfondimenti/il-problem-solving/

# Appendice

Codici R

Fase 1: Scrematura delle variabili

```
library(tidyverse)
library(readxl)
library(caret)
library(ggplot2)

# Caricamento del dataset
df <- read_excel(file.choose())  # Seleziona
UNIFIED_DATASET_10000.xlsx

# Impostazione variabile target
df$target <-
as.numeric(df$trad_loan_approved)

# Variabili selezionate (16 numeriche + 8
categoriche)
selected_num <- c(
  "income",
  "trad_risk_score",
  "trad_interest_rate",
  "trad_loan_amount",
  "trad_net_worth",
  "trad_credit_score",
  "trad_experience",
  "age",
  "trad_total_debt_to_income_ratio",
  "trad_length_of_credit_history",
  "trad_bankruptcy_history",
  "alt_avg_discount_used",
  "alt_online_purchases",
  "alt_in_store_purchases",
  "alt_avg_transaction_value",
  "alt_days_since_last_purchase"
)

selected_cat <- c(
  "income_level",
  "trad_employment_status",
  "education_level",
  "age_gap",
  "alt_promotion_channel",
  "alt_churned",
  "alt_social_media_engagement",
  "alt_payment_method"
)

# Dataset ridotto alle variabili selezionate
df_selected <- df %>%
  select(all_of(selected_num),
all_of(selected_cat), target) %>%
  mutate(across(all_of(selected_cat),
as.factor))

# Encoding sulle categoriche selezionate
cat_data <- df_selected %>%
select(all_of(selected_cat))
dummies <- dummyVars(~ ., data = cat_data)
cat_encoded <- predict(dummies, newdata =
cat_data) %>% as.data.frame()

# Unione numeriche + dummy + target
df_model <- cbind(df_selected %>%
select(all_of(selected_num)), cat_encoded,
target = df_selected$target)

# Funzione “fast” per rimozione collinearità
remove_high_corr <- function(data, cutoff =
0.9) {
  keep <- rep(TRUE, ncol(data))
  for (i in 1:(ncol(data)-1)) {
    if (keep[i]) {
      for (j in (i+1):ncol(data)) {
        if (keep[j]) {
          corr_val <-
suppressWarnings(cor(data[[i]], data[[j]], use
= "pairwise.complete.obs"))
          if (!is.na(corr_val) && abs(corr_val) >
cutoff) {
            keep[j] <- FALSE
          }
        }
      }
    }
  }
  return(data[, keep])
}

num_features <- df_model %>%
select(where(is.numeric), -target)
```

```
dataset_no_corr <-
remove_high_corr(num_features, cutoff =
0.9)

# Ricreiamo dataset finale con target
dataset_finale <- cbind(dataset_no_corr,
target = df_model$target)

# Salvataggio dataset finale
write.csv(dataset_finale,
"dataset_finale_modello_COMPLETO_light_
fast.csv", row.names = FALSE)

# Riepilogo numeri
n_iniziale <- ncol(df) - 2
n_concettuale <- length(selected_num) +
length(selected_cat)
n_finale <- ncol(dataset_finale) - 1

cat("Variabili iniziali:", n_iniziale, "\n")
cat("Variabili concettuali selezionate:",
n_concettuale, "\n")
cat("Variabili finali post-encoding e
rimozione collinearità:", n_finale, "\n")

# Grafico riduzione variabili
df_fasi <- data.frame(
  Fase = factor(c("Iniziale", "Dopo selezione
concettuale", "Pronto per il modello"),
                levels = c("Iniziale", "Dopo
selezione concettuale", "Pronto per il
modello")),
  Numero_variabili = c(n_iniziale,
n_concettuale, n_finale)
)

ggplot(df_fasi, aes(x = Fase, y =
Numero_variabili, fill = Fase)) +
  geom_col(width = 0.6) +
  geom_text(aes(label = Numero_variabili),
vjust = -0.5, size = 5) +
  theme_minimal() +
  labs(title = "Riduzione progressiva del
numero di variabili",
       y = "Numero di variabili", x = "Fase di
lavorazione") +
  scale_fill_manual(values = c("#00A0D6",
"#F28E2B", "#4E79A7"))

ggsave("riduzione_variabili_light_fast.png",
width = 8, height = 5)

# Grafico Top 15 correlazioni
num_cols <- df %>%
select(where(is.numeric)) %>% select(-target)
%>% names()
corr_results <- sapply(num_cols,
function(col) cor(df[[col]], df$target, method
= "pearson"))
corr_df <- data.frame(variable =
names(corr_results), abs_corr =
abs(corr_results)) %>%
  arrange(desc(abs_corr))

ggplot(corr_df[1:15,], aes(x =
reorder(variable, abs_corr), y = abs_corr)) +
  geom_col(fill = "steelblue") +
  coord_flip() +
  labs(title = "Top 15 correlazioni con target",
x = "", y = "Correlazione Assoluta")

ggsave("top15_correlazioni_light_fast.png",
width = 8, height = 5)

# Grafico Top 10 chi-quadro
cat_cols <- df %>%
select(where(is.character)) %>% names()
chi2_results <- lapply(cat_cols, function(col)
{
  tbl <- table(df[[col]], df$target)
  test <- suppressWarnings(chisq.test(tbl))
  data.frame(variable = col, chi2 =
test$statistic, p_value = test$p.value)
})
chi2_df <- do.call(rbind, chi2_results) %>%
arrange(desc(chi2))

ggplot(chi2_df[1:10,], aes(x =
reorder(variable, chi2), y = chi2)) +
  geom_col(fill = "darkorange") +
  coord_flip() +
  labs(title = "Top 10 chi-quadro con target", x
= "", y = "Chi-squared")

ggsave("top10_chi2_light_fast.png", width =
8, height = 5)
```

Fase 2: Inner join

```
# Importazione dataset
trad <- read_excel(file.choose())  # Scegli
TRADITIONAL_DATA.xlsx
alt  <- read_excel(file.choose())  # Scegli
ALTERNATIVE_DATA.xlsx

# Uniforma i nomi in snake_case
trad <- trad %>% clean_names()
alt  <- alt %>% clean_names()

# Variabili chiave
Chiavi <- c("age_gap", "income_level",
"education_level")

# Aggiunta prefisso "trad_" e "alt_" alle altre
variabili
trad <- trad %>%
  rename_with(~ paste0("trad_", .x), .cols = -
all_of(chiavi))

alt <- alt %>%
  rename_with(~ paste0("alt_", .x), .cols = -
all_of(chiavi))

# Pulizia: rimuove righe con NA nelle chiavi
trad_clean <- trad %>%
filter(across(all_of(chiavi), ~ !is.na(.)))
alt_clean  <- alt  %>%
filter(across(all_of(chiavi), ~ !is.na(.)))

# Inner join su variabile chiave
merged_data <- inner_join(trad_clean,
alt_clean, by = chiavi)

# Campionamento casuale di 10000
osservazioni
set.seed(123)
merged_sample <- merged_data %>%
sample_n(10000)

# Salvataggio
write.xlsx(merged_sample,
"UNIFIED_DATASET_10000.xlsx")
```

Fase 3: Analisi descrittiva

```
library(dplyr)

# Statistiche di sintesi
summary_stats_long <- num_vars %>%
  pivot_longer(cols = everything(), names_to
= "Variabile", values_to = "Valore") %>%
  group_by(Variabile) %>%
  summarise(
    media = mean(Valore, na.rm = TRUE),
    mediana = median(Valore, na.rm = TRUE),
    dev_std = sd(Valore, na.rm = TRUE),
    varianza = var(Valore, na.rm = TRUE),
    q1 = quantile(Valore, 0.25, na.rm =
TRUE),
    q3 = quantile(Valore, 0.75, na.rm =
TRUE),
    min = min(Valore, na.rm = TRUE),
    max = max(Valore, na.rm = TRUE)
  )

# Visualizza tabella
print(summary_stats_long)

# Salva in CSV
write.csv(summary_stats_long,
"statistiche_sintesi_dataset_interpretativo.csv
", row.names = FALSE)


library(ggplot2)

# Distribuzione e forme

# Lista di variabili continue strategiche
vars_cont <- c("income",
"trad_loan_amount", "trad_credit_score",
"alt_avg_transaction_value")

# Istogrammi
for (v in vars_cont) {
  p <- ggplot(df_interpret, aes_string(x = v)) +
    geom_histogram(fill = "steelblue", color =
"white", bins = 30) +
    theme_minimal() +
    labs(title = paste("Distribuzione di", v),
```

```
      x = v, y = "Frequenza")

  print(p)
}

# Boxplot confrontati per target
for (v in vars_cont) {
  p <- ggplot(df_interpret, aes_string(x =
"target", y = v, fill = "target")) +
    geom_boxplot() +
    theme_minimal() +
    labs(title = paste("Boxplot di", v, "per
target"),
         x = "Target (0 = rifiutato, 1 =
approvato)", y = v) +
    scale_fill_manual(values = c("#F28E2B",
"#4E79A7"))

  print(p)
}

library(dplyr)
library(ggplot2)
library(tidyr)

# Confronto tra gruppi

# Lista di variabili numeriche da confrontare
vars_confronto <- c("income",
"trad_net_worth", "trad_credit_score",
"trad_loan_amount")

# Calcolo statistiche per gruppi target
confronto_stats <- df_interpret %>%
  group_by(target) %>%
  summarise(across(all_of(vars_confronto),
                   list(media = ~mean(.x, na.rm =
TRUE),
                        mediana = ~median(.x, na.rm
= TRUE)),
                   .names = "{.col}_{.fn}"))

# Trasformazione per grafico
confronto_long <- confronto_stats %>%
  pivot_longer(cols = -target,
               names_to = c("Variabile",
"Statistica"),
               names_sep = "_",
               values_to = "Valore")

# Grafico a torta delle medie per variabile e
target
ggplot(filter(confronto_long, Statistica ==
"media"),
       aes(x = Variabile, y = Valore, fill =
as.factor(target))) +
  geom_col(position = "dodge") +
  theme_minimal() +
  labs(title = "Confronto delle medie per
variabile e target",
       x = "Variabile", y = "Media",
       fill = "Target (0 = Rifiutato, 1 =
Approvato)") +
  scale_fill_manual(values = c("#F28E2B",
"#4E79A7")) +
  coord_flip()

# Salvataggio tabella in CSV
write.csv(confronto_stats,
"confronto_statistiche_target.csv", row.names
= FALSE)

library(dplyr)
library(ggplot2)

#Analisi variabili categoriche

# Lista variabili categoriche da analizzare
vars_cat <- c("income_level",
"trad_employment_status", "education_level",
              "age_gap", "alt_promotion_channel",
"alt_churned",
              "alt_social_media_engagement",
"alt_payment_method")

# Funzione per creare grafici di distribuzione
per ciascuna variabile categorica
for (var in vars_cat) {

  # Distribuzione semplice
  p1 <- ggplot(df_interpret, aes_string(x =
var)) +
    geom_bar(fill = "#4E79A7") +
    theme_minimal() +
    labs(title = paste("Distribuzione di", var), x
= var, y = "Frequenza") +
```

```
  theme(axis.text.x = element_text(angle =
45, hjust = 1))
 print(p1)

 # Distribuzione per target
 p2 <- ggplot(df_interpret, aes_string(x = var,
fill = "as.factor(target)")) +
  geom_bar(position = "fill") +
  theme_minimal() +
  labs(title = paste("Distribuzione di", var,
"per Target"),
       x = var, y = "Proporzione",
       fill = "Target (0 = Rifiutato, 1 =
Approvato)") +
  scale_fill_manual(values = c("#F28E2B",
"#4E79A7")) +
  theme(axis.text.x = element_text(angle =
45, hjust = 1))
 print(p2)
}

# Test chi quadro

chi2_results <- data.frame(
 Variabile = character(),
 Chi2 = numeric(),
 p_value = numeric(),
 stringsAsFactors = FALSE
)

for (var in vars_cat) {
 tab <- table(df_interpret[[var]],
df_interpret$target)
 test <- chisq.test(tab)
 chi2_results <- rbind(chi2_results,
                       data.frame(Variabile = var,
                                  Chi2 =
round(test$statistic, 2),
                                  p_value =
round(test$p.value, 4)))
}

# Ordina per valore Chi²
chi2_results <- chi2_results %>%
arrange(desc(Chi2))

# Visualizza risultati
print(chi2_results)

# Salva in CSV
write.csv(chi2_results,
"chi2_variabili_categoriche.csv", row.names
= FALSE)


# Relazioni tra variabili numeriche

# Trova automaticamente la colonna target
(factor con 2 livelli)
target_name <-
names(df_interpret)[sapply(df_interpret,
function(x) is.factor(x) & length(levels(x)) ==
2)]
print(paste("Colonna target individuata:",
target_name))

# Seleziona le variabili numeriche escludendo
il target
num_vars <- df_interpret %>%
 select(where(is.numeric), -
all_of(target_name))

#  Calcola matrice di correlazione
corr_mat <- cor(num_vars, use =
"complete.obs")

# Visualizza matrice con corrplot
library(corrplot)
library(RColorBrewer)

corrplot(corr_mat, method = "color", col =
brewer.pal(n = 8, name = "RdBu"),
         type = "upper", order = "hclust",
         addCoef.col = "black", tl.cex = 0.7,
number.cex = 0.5,
         diag = FALSE, mar = c(0,0,1,0))

#  Salvataggio matrice di correlazione
write.csv(corr_mat,
"matrice_correlazione_variabili_numeriche.cs
v", row.names = TRUE)

# Scatter Plot mirati

library(ggplot2)
```

```
# Alcune coppie interessanti da analizzare
coppie_variabili <- list(
  c("income", "trad_loan_amount"),
  c("trad_credit_score", "trad_interest_rate"),
  c("trad_net_worth",
"trad_total_debt_to_income_ratio"),
  c("alt_avg_transaction_value",
"alt_online_purchases")
)

# Ciclo per generare scatter plot colorati per
target
for (coppia in coppie_variabili) {
  var_x <- coppia[1]
  var_y <- coppia[2]

  p <- ggplot(df_interpret, aes_string(x =
var_x, y = var_y, color = target_name)) +
    geom_point(alpha = 0.6) +
    theme_minimal() +
    labs(title = paste("Scatter Plot:", var_x,
"vs", var_y),
         x = var_x, y = var_y, color = "Target")
+
    theme(plot.title = element_text(size = 14,
face = "bold"))

  print(p)

  # Salvataggio di ogni grafico
  ggsave(paste0("scatter_", var_x, "_vs_",
var_y, ".png"), plot = p, width = 7, height = 5)
}
```

Fase 4: Analisi predittiva

```
# REGRESSIONE LOGISTICA -
TRADIZIONALE, ALTERNATIVO,
UNIFICATO

# Librerie necessarie
library(tidyverse)
library(caret)
library(pROC)

# Estrazione coefficienti e salvataggio CSV
extract_coefficients <- function(model, label)
{
  cat("\n===== Interpretazione modello:",
label, "=====\n")
  coeffs <- summary(model)$coefficients
  odds_ratio <- exp(coeffs[, 1])

  results <- data.frame(
    Variabile = rownames(coeffs),
    Coefficiente = round(coeffs[, 1], 4),
    Std_Error = round(coeffs[, 2], 4),
    Z_value = round(coeffs[, 3], 4),
    P_value = signif(coeffs[, 4], 4),
    Odds_Ratio = round(odds_ratio, 4)
  )

  print(results)

  # Salvataggio CSV
  file_name <- paste0("coeff_",
tolower(label), ".csv")
  write.csv(results, file_name, row.names =
FALSE)
  cat("File salvato come:", file_name, "\n")

  return(results)
}


# Valutazione modello logistico con metriche

valuta_logistica <- function(dataset, label,
target_col = "target") {
  y <- dataset[[target_col]]
  X <- dataset %>% select(-all_of(target_col))

  # Train/Test split
  set.seed(123)
  idx <- createDataPartition(y, p = 0.7, list =
FALSE)
  X_train <- X[idx, ]
  y_train <- y[idx]
  X_test <- X[-idx, ]
  y_test <- y[-idx]

  # Modello
  model <- glm(y_train ~ ., data =
as.data.frame(X_train), family = binomial)
```

```
  # Predizioni
  prob <- predict(model, newdata =
as.data.frame(X_test), type = "response")
  pred <- ifelse(prob > 0.5, 1, 0)

  # Confusion matrix
  cm <- confusionMatrix(as.factor(pred),
as.factor(y_test), positive = "1")

  # ROC e AUC
  roc_obj <- roc(y_test, prob)
  auc_val <- as.numeric(auc(roc_obj))

  # Metriche
  metrics <- tibble(
    Modello = label,
    Accuracy =
as.numeric(cm$overall["Accuracy"]),
    Recall =
as.numeric(cm$byClass["Sensitivity"]),
    Precision = as.numeric(cm$byClass["Pos
Pred Value"]),
    F1 = as.numeric(2 * ((cm$byClass["Pos
Pred Value"] * cm$byClass["Sensitivity"]) /
                         (cm$byClass["Pos Pred
Value"] + cm$byClass["Sensitivity"]))),
    AUC = auc_val
  )

  return(list(model = model, metrics = metrics,
roc = roc_obj))
}

# Caricamento dati
file_path <- file.choose() # seleziona
dataset_pulito_regressione.csv
data <- read.csv(file_path, sep = ",",
stringsAsFactors = FALSE)

target <- "target"

# Analisi tradizionale
trad_vars <- data %>%
select(starts_with("trad_"), target)
res_trad <- valuta_logistica(trad_vars,
"Tradizionale")
coeff_trad <-
extract_coefficients(res_trad$model,
"Tradizionale")

# Analisi alternativa
alt_vars <- data %>%
select(starts_with("alt_"), target)
res_alt <- valuta_logistica(alt_vars,
"Alternativo")
coeff_alt <-
extract_coefficients(res_alt$model,
"Alternativo")

# Analisi unificata
unif_vars <- data %>%
select(starts_with("trad_"),
starts_with("alt_"), target)
res_unif <- valuta_logistica(unif_vars,
"Unificato")
coeff_unif <-
extract_coefficients(res_unif$model,
"Unificato")

# Tabella comparativa metriche + salvataggio
tabella_metriche <-
bind_rows(res_trad$metrics, res_alt$metrics,
res_unif$metrics)
write.csv(tabella_metriche,
"metriche_comparative.csv", row.names =
FALSE)

cat("\n File CSV salvati nella cartella di
lavoro:\n",
    "- coeff_tradizionale.csv\n",
    "- coeff_alternativo.csv\n",
    "- coeff_unificato.csv\n",
    "- metriche_comparative.csv\n")

print(tabella_metriche)

# Grafico roc unico per tutti i modelli
plot(res_trad$roc, col = "#1b9e77", lwd = 2,
main = "Curva ROC - Confronto Modelli")
plot(res_alt$roc, col = "#d95f02", lwd = 2,
add = TRUE)
plot(res_unif$roc, col = "#7570b3", lwd = 2,
add = TRUE)
```

```
legend("bottomright",
       legend = c(
         paste("Tradizionale (AUC =",
round(res_trad$metrics$AUC, 3), ")"),
         paste("Alternativo (AUC =",
round(res_alt$metrics$AUC, 3), ")"),
         paste("Unificato (AUC =",
round(res_unif$metrics$AUC, 3), ")")
       ),
       col = c("#1b9e77", "#d95f02",
"#7570b3"),
       lwd = 2)


# RANDOM FOREST - TRADIZIONALE,
ALTERNATIVO, UNIFICATO

# Librerie
library(tidyverse)
library(caret)
library(pROC)
library(randomForest)

# Valutazione Random Forest con metriche +
importanza
valuta_rf <- function(dataset, label, target_col
= "target") {
  y <- as.factor(dataset[[target_col]])
  X <- dataset %>% select(-all_of(target_col))

  # Train/Test split
  set.seed(123)
  idx <- createDataPartition(y, p = 0.7, list =
FALSE)
  X_train <- X[idx, ]
  y_train <- y[idx]
  X_test <- X[-idx, ]
  y_test <- y[-idx]

  # Modello Random Forest
  model <- randomForest(x = X_train, y =
y_train, ntree = 500, mtry =
floor(sqrt(ncol(X_train))), importance =
TRUE)

  # Predizioni probabilità e classi
  prob <- predict(model, X_test, type =
"prob")[, "1"]
  pred <- predict(model, X_test, type =
"response")

  # Confusion matrix
  cm <- confusionMatrix(pred, y_test, positive
= "1")

  # ROC e AUC
  roc_obj <-
roc(as.numeric(as.character(y_test)), prob)
  auc_val <- as.numeric(auc(roc_obj))

  # Metriche
  metrics <- tibble(
    Modello = label,
    Accuracy =
as.numeric(cm$overall["Accuracy"]),
    Recall =
as.numeric(cm$byClass["Sensitivity"]),
    Precision = as.numeric(cm$byClass["Pos
Pred Value"]),
    F1 = as.numeric(2 * ((cm$byClass["Pos
Pred Value"] * cm$byClass["Sensitivity"]) /
                         (cm$byClass["Pos Pred
Value"] + cm$byClass["Sensitivity"]))),
    AUC = auc_val
  )

  # Importanza variabili
  var_imp <-
as.data.frame(importance(model))
  var_imp <- tibble(
    Variabile = rownames(var_imp),
    MeanDecreaseGini = var_imp[,
"MeanDecreaseGini"]
  ) %>% arrange(desc(MeanDecreaseGini))

  # Salva CSV importanza
  file_name <- paste0("var_importanza_RF_",
tolower(label), ".csv")
  write.csv(var_imp, file_name, row.names =
FALSE)

  return(list(model = model, metrics = metrics,
roc = roc_obj, importanza = var_imp))
}

# Caricamento dati
```

```
file_path <- file.choose() # seleziona
dataset_pieno_RF_XGB.csv

target <- "target"

# Analisi tradizionale
trad_vars <- data %>%
select(starts_with("trad_"), target)
res_trad <- valuta_rf(trad_vars,
"Tradizionale")

# Analisi alternativa
alt_vars <- data %>%
select(starts_with("alt_"), target)
res_alt <- valuta_rf(alt_vars, "Alternativo")

# Analisi unificata
unif_vars <- data %>%
select(starts_with("trad_"),
starts_with("alt_"), target)
res_unif <- valuta_rf(unif_vars, "Unificato")


# Tabella comparativa metriche + salvataggio
tabella_metriche <-
bind_rows(res_trad$metrics, res_alt$metrics,
res_unif$metrics)
write.csv(tabella_metriche,
"metriche_RF_comparative.csv", row.names
= FALSE)

cat("\n File CSV salvati:\n",
    "- var_importanza_RF_tradizionale.csv\n",
    "- var_importanza_RF_alternativo.csv\n",
    "- var_importanza_RF_unificato.csv\n",
    "- metriche_RF_comparative.csv\n")

print(tabella_metriche)

# Grafico roc unico per tutti i modelli
plot(res_trad$roc, col = "#1b9e77", lwd = 2,
main = "Curva ROC - Random Forest")
plot(res_alt$roc, col = "#d95f02", lwd = 2,
add = TRUE)
plot(res_unif$roc, col = "#7570b3", lwd = 2,
add = TRUE)

legend("bottomright",
       legend = c(
         paste("Tradizionale (AUC =",
round(res_trad$metrics$AUC, 3), ")"),
         paste("Alternativo (AUC =",
round(res_alt$metrics$AUC, 3), ")"),
         paste("Unificato (AUC =",
round(res_unif$metrics$AUC, 3), ")")
       ),
       col = c("#1b9e77", "#d95f02",
"#7570b3"),
       lwd = 2)


# XGBOOST - TRADIZIONALE,
ALTERNATIVO, UNIFICATO

# Librerie necessarie
library(tidyverse)
library(caret)
library(pROC)
library(xgboost)

# Valutazione XGBoost con metriche +
importanza
valuta_xgb <- function(dataset, label,
target_col = "target") {

  # Convert target in numerico 0/1
  y <-
as.numeric(as.factor(dataset[[target_col]])) - 1
  X <- dataset %>% select(-all_of(target_col))

  # One-hot encoding
  X_mat <- model.matrix(~ . - 1, data = X)

  # Train/Test split
  set.seed(123)
  idx <- createDataPartition(y, p = 0.7, list =
FALSE)
  X_train <- X_mat[idx, ]
  y_train <- y[idx]
  X_test  <- X_mat[-idx, ]
  y_test  <- y[-idx]

  # DMatrix
  dtrain <- xgb.DMatrix(data = X_train, label
= y_train)
```

```
  dtest  <- xgb.DMatrix(data = X_test, label =
y_test)

  # Parametri XGBoost
  params <- list(
    objective = "binary:logistic",
    eval_metric = "auc",
    max_depth = 6,
    eta = 0.1,
    subsample = 0.8,
    colsample_bytree = 0.8
  )

  # Addestramento
  model <- xgb.train(
    params = params,
    data = dtrain,
    nrounds = 200,
    watchlist = list(train = dtrain, test = dtest),
    verbose = 0
  )

  # Predizioni
  prob <- predict(model, X_test)
  pred <- ifelse(prob > 0.5, 1, 0)

  # Confusion matrix
  cm <- confusionMatrix(as.factor(pred),
as.factor(y_test), positive = "1")

  # ROC e AUC
  roc_obj <- roc(y_test, prob)
  auc_val <- as.numeric(auc(roc_obj))

  # Metriche
  metrics <- tibble(
    Modello = label,
    Accuracy =
as.numeric(cm$overall["Accuracy"]),
    Recall =
as.numeric(cm$byClass["Sensitivity"]),
    Precision = as.numeric(cm$byClass["Pos
Pred Value"]),
    F1 = as.numeric(2 * ((cm$byClass["Pos
Pred Value"] * cm$byClass["Sensitivity"]) /
                         (cm$byClass["Pos Pred
Value"] + cm$byClass["Sensitivity"]))),
    AUC = auc_val
  )

  # Importanza variabili
  var_imp <- xgb.importance(model = model)
  file_name <-
paste0("var_importanza_XGB_",
tolower(label), ".csv")
  write.csv(var_imp, file_name, row.names =
FALSE)

  return(list(model = model, metrics = metrics,
roc = roc_obj, importanza = var_imp))
}

# Caricamento dati
file_path <- file.choose() # seleziona
dataset_pieno_RF_XGB.csv

target <- "target"
# Analisi tradizionale

trad_vars <- data %>%
select(starts_with("trad_"), target)
res_trad <- valuta_xgb(trad_vars,
"Tradizionale")

# Analisi alternativa

alt_vars <- data %>%
select(starts_with("alt_"), target)
res_alt <- valuta_xgb(alt_vars, "Alternativo")

# Analisi unificata
unif_vars <- data %>%
select(starts_with("trad_"),
starts_with("alt_"), target)
res_unif <- valuta_xgb(unif_vars,
"Unificato")

# Tabella comparativa metriche + salvataggio
tabella_metriche <-
bind_rows(res_trad$metrics, res_alt$metrics,
res_unif$metrics)
write.csv(tabella_metriche,
"metriche_XGB_comparative.csv",
row.names = FALSE)

cat("\n File CSV salvati:\n",
```

```
   "-
var_importanza_XGB_tradizionale.csv\n",
   "-
var_importanza_XGB_alternativo.csv\n",
   "- var_importanza_XGB_unificato.csv\n",
   "- metriche_XGB_comparative.csv\n")

print(tabella_metriche)

# Grafico roc unico per tutti i modelli
plot(res_trad$roc, col = "#1b9e77", lwd = 2,
main = "Curva ROC - XGBoost")
plot(res_alt$roc, col = "#d95f02", lwd = 2,
add = TRUE)
plot(res_unif$roc, col = "#7570b3", lwd = 2,
add = TRUE)

legend("bottomright",
       legend = c(
         paste("Tradizionale (AUC =",
round(res_trad$metrics$AUC, 3), ")"),
         paste("Alternativo (AUC =",
round(res_alt$metrics$AUC, 3), ")"),
         paste("Unificato (AUC =",
round(res_unif$metrics$AUC, 3), ")")
       ),
       col = c("#1b9e77", "#d95f02",
"#7570b3"),
       lwd = 2)
```

Fase 5: Applicazione tecniche XAI; SHAP E LIME

```
#RF + XGB con SHAP (globale e locale) +
LIME
# con beeswarm XGBoost Top 10 leggibile

# Librerie
packages <- c("randomForest", "xgboost",
"DALEX", "lime", "shapviz", "ggplot2")
newpkgs <- packages[!packages %in%
installed.packages()[,"Package"]]
if(length(newpkgs))
install.packages(newpkgs)

library(randomForest)
library(xgboost)
library(DALEX)
library(lime)
library(shapviz)
library(ggplot2)

# Caricamento dataset
csv_path <- file.choose()   # scegli
manualmente il file CSV
df <- read.csv(csv_path, stringsAsFactors =
FALSE)

target <- "target"

# Converte target in numerico binario 0/1 se
necessario
if (is.factor(df[[target]]) ||
is.character(df[[target]])) {
  df[[target]] <- ifelse(df[[target]] %in%
c("1","Yes","Default","Positive","Positivo","
Si","Sì"), 1, 0)
}

# Fattorizza variabili categoriche
char_cols <- names(df)[vapply(df,
is.character, TRUE)]
if (length(char_cols) > 0) df[char_cols] <-
lapply(df[char_cols], as.factor)

# Train/test split
set.seed(123)
train_idx <- sample(seq_len(nrow(df)), size =
0.7*nrow(df))
train <- df[train_idx, ]
test  <- df[-train_idx, ]

X_train_df <- train[ , !(names(train) %in%
target), drop = FALSE]
y_train    <- train[[target]]
X_test_df  <- test[ , !(names(test) %in%
target), drop = FALSE]
y_test     <- test[[target]]

# Modelli
# Random Forest
rf_model <- randomForest(
  x = X_train_df, y = as.factor(y_train),
  ntree = 200
)
```

```
# XGBoost
X_train_mm <- model.matrix(~ . -1, data =
X_train_df)
X_test_mm  <- model.matrix(~ . -1, data =
X_test_df)
dtrain <- xgb.DMatrix(X_train_mm, label =
y_train)
dtest  <- xgb.DMatrix(X_test_mm,  label =
y_test)

xgb_model <- xgboost(
  data = dtrain,
  nrounds = 100,
  objective = "binary:logistic",
  verbose = 0
)

# Explainers DALEX
expl_rf  <- DALEX::explain(
  model = rf_model,
  data  = X_train_df,
  y     = y_train,
  label = "Random Forest"
)

expl_xgb <- DALEX::explain(
  model = xgb_model,
  data  = X_train_mm,
  y     = y_train,
  label = "XGBoost"
)

# Feature Importance
vip_rf <- DALEX::model_parts(expl_rf)
png("01_feature_importance_rf.png",
width=1000, height=700)
print(plot(vip_rf) + ggtitle("Importanza
variabili — Random Forest"))
dev.off()

# XGB feature importance con SHAP
shap_xgb <- predict(xgb_model, X_test_mm,
predcontrib=TRUE)
if ("BIAS" %in% colnames(shap_xgb))
shap_xgb <- shap_xgb[, -ncol(shap_xgb)]
mean_abs_shap <- colMeans(abs(shap_xgb))
imp_xgb_df <-
data.frame(Feature=names(mean_abs_shap),
mean_abs_SHAP=mean_abs_shap)
imp_xgb_df <- imp_xgb_df[order(-
imp_xgb_df$mean_abs_SHAP), ]

png("02_feature_importance_xgb.png",
width=1000, height=700)
print(
  ggplot(head(imp_xgb_df,20),
aes(x=reorder(Feature, mean_abs_SHAP),
y=mean_abs_SHAP)) +
    geom_col() + coord_flip() +
    labs(title="Importanza globale
(mean|SHAP|) — XGBoost", x="Feature",
y="mean|SHAP|")
)
dev.off()
write.csv(imp_xgb_df,
"table_importance_xgb.csv",
row.names=FALSE)

# SHAP — Random Forest (locale)
pp_rf <- DALEX::predict_parts(
  explainer = expl_rf,
  new_observation = X_test_df[1, , drop =
FALSE],
  type = "shap",
  N = 50
)

png("03_shap_local_rf.png", width = 1000,
height = 700)
print(plot(pp_rf) + ggtitle("SHAP locale —
RF (osservazione 1)"))
dev.off()


# SHAP Beeswarm XGBoost (Top 10)

library(ggplot2)

# Seleziona le 10 variabili più importanti
top_features <-
names(sort(colMeans(abs(shap_xgb)),
decreasing = TRUE))[1:10]

# Crea dataset lungo per ggplot
```

```
df_plot <- data.frame(
  Feature = rep(top_features, each =
nrow(shap_xgb)),
  SHAP    = as.vector(shap_xgb[,
top_features])
)

# Ordina le feature in base all'importanza
media
df_plot$Feature <- factor(
  df_plot$Feature,
  levels = rev(top_features)  # così la più
importante sta in alto
)

# Grafico beeswarm
png("04_shap_beeswarm_xgb_top10.png",
width = 1400, height = 900, res = 150)
ggplot(df_plot, aes(x = SHAP, y = Feature)) +
  geom_jitter(height = 0.25, alpha = 0.5, size
= 1) +
  geom_vline(xintercept = 0, linetype =
"dashed", color = "grey50") +
  labs(
    title = "SHAP globale (beeswarm) —
XGBoost (Top 10)",
    x = "Valore SHAP (contributo alla
predizione)",
    y = "Variabile"
  ) +
  theme_minimal(base_size = 14)
dev.off()


# LIME - RF
model_type.randomForest <- function(x, ...)
"classification"
predict_model.randomForest <- function(x,
newdata, ...) {
  as.data.frame(predict(x, newdata,
type="prob"))
}
expl_lime_rf <- lime(X_train_df, rf_model)
explanation_rf <- lime::explain(
  X_test_df[1:3, ,drop=FALSE],
  explainer=expl_lime_rf,
  n_features=6,
  n_labels=1
)
png("05_lime_rf.png", width=1200,
height=800)
print(plot_features(explanation_rf) +
ggtitle("LIME locali — Random Forest"))
dev.off()

# LIME - XGB
model_type.xgb.Booster <- function(x, ...)
"classification"
predict_model.xgb.Booster <- function(x,
newdata, ...) {
  p <- predict(x,
xgb.DMatrix(data=as.matrix(newdata)))
  data.frame(`0`=1-p, `1`=p,
check.names=FALSE)
}
expl_lime_xgb <-
lime(as.data.frame(X_train_mm), xgb_model)
explanation_xgb <- lime::explain(

as.data.frame(X_test_mm[1:3,,drop=FALSE])
,
  explainer=expl_lime_xgb,
  n_features=6,
  n_labels=1
)
png("06_lime_xgb.png", width=1200,
height=800)
print(plot_features(explanation_xgb) +
ggtitle("LIME locali — XGBoost"))
dev.off()

# Log
cat("\nFile creati:\n",
    "- 01_feature_importance_rf.png\n",
    "- 02_feature_importance_xgb.png\n",
    "- 03_shap_local_rf.png\n",
    "- 04_shap_beeswarm_xgb_top10.png\n",
    "- 06_lime_rf.png\n",
    "- 07_lime_xgb.png\n",
    "- table_importance_xgb.csv\n")
```