LUISS

Dipartimento di Impresa e Management

Laurea Triennale in “Economia e Management”

Cattedra di Intelligenza Artificiale Per I Processi Decisionali

Machine Learning e Trattamento Dei Dati Finanziari: Prospettive, Regolamentazione e Un Caso D’uso Esemplificativo

Relatore

Prof. Vittorio Carlei

Candidato

Gabriele Rizzo

266261

Anno Accademico 2023/2024

Sommario

Introduzione

1.1. Contesto generale: L'importanza dell'Intelligenza Artificiale nei processi decisionali

1.2. Obiettivi della ricerca e struttura della tesi

Capitolo 1: Scenario Normativo – Fondamenti del Regolamento FiDA e Implicazioni per il Settore Finanziario

1.1. Introduzione al Regolamento FiDA

1.2. Accesso e utilizzo dei dati finanziari: Nuovi paradigmi

1.3. Ruolo dei Prestatori di Servizi di Informazione Finanziaria (PSIF)

1.4. Impatti Potenziali sull'Ecosistema Finanziario Europeo

Capitolo 2: Scenario Tecnologico e Opportunità di Mercato nell'Utilizzo dell'IA sui Dati Finanziari

2.1. Criticità attuali nei processi di gestione dei dati finanziari

2.2. Benefici dell'Intelligenza Artificiale nel sistema finanziario odierno

2.3. Confronto tra approccio tradizionale e modello IA

Capitolo 3: Caso d'Uso Esemplificativo – Sviluppo e Implementazione di un Algoritmo di Machine Learning per l'Analisi dei Dati Finanziari

3.1. Principi teorici alla base dell'algoritmo

3.2. Scelta delle variabili e pre-processing dei dati

3.3. Struttura e componenti: applicazione del modello supervisionato (GBC) in Python

3.4. Valutazione delle prestazioni e Test del modello

3.5. Obiettivi e Opportunità dell'implementazione dello Use Case

Capitolo 4: Conclusioni – Considerazioni Finali e Sviluppi Futuri

Riferimenti Bibliografici

Introduzione

L'intelligenza artificiale rappresenta una delle tecnologie più evolute dei tempi moderni e ha cambiato completamente il processo decisionale. L'IA, con la sua capacità di analizzare grandi insiemi di dati e quindi di fornire approfondimenti con un carattere predittivo, ha rivoluzionato le tradizionali dicotomie decisionali in ambito aziendale, sanitario, finanziario e di pubblica amministrazione.

La crescente dipendenza da queste tecnologie è dovuta al suo potenziale nel migliorare la velocità, l'accuratezza e l'efficacia nel prendere decisioni, dando così alle organizzazioni la possibilità di navigare con precisione in ambienti che stanno diventando sempre più complessi. Con l'avanzare dei progressi dell'IA, il suo ruolo nel supportare e migliorare il giudizio umano diventa di importanza fondamentale e fornisce nuove soluzioni a problemi finora intrattabili.

L'integrazione dell'IA nel processo decisionale ha tuttavia segnato una deviazione significativa dalla modalità tradizionale che si affida solo all'intuizione e all'esperienza umana. Secondo Agrawal et al. (2018), l'IA sta cambiando radicalmente il processo decisionale grazie alla sua avanzata capacità di previsione[1]. Questi sistemi analizzano modelli e tendenze all'interno di dati estesi, il tutto supportato da sofisticati algoritmi e tecniche di machine learning; di conseguenza, forniscono informazioni utili ai *decision-makers*. Questa è l'essenza del potere predittivo: ridurre l'incertezza, non solo a livello di organizzazione, ma anche fornendo alle organizzazioni la possibilità di proiettare i risultati futuri e ottimizzare l'allocazione delle risorse, in modo da poter prendere decisioni strategiche per la crescita e l'efficienza.

La maggior parte degli impatti dell'IA si avverte nel campo della gestione. Judijanto et al. (2022) sottolineano che le pratiche di gestione supportate dall'IA aumentano notevolmente l'efficienza del processo decisionale attraverso l'automazione del follow-up e le raccomandazioni basate sui dati[2]. A questo proposito, l'IA è in grado di elaborare grandi quantità di informazioni in un periodo di tempo molto breve, il che significa che i manager avranno più possibilità di sviluppare il pensiero strategico e quindi di migliorare

[1] Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction, judgment, and complexity: a theory of decision-making and artificial intelligence. In *The economics of artificial intelligence*. NBER.

[2] Judijanto, L., Asfahani, A., & Bakri, A. A. (2022). AI-supported management through leveraging artificial intelligence for effective decision making. *Journal of Artificial Intelligence*.

le prestazioni aziendali. Il passaggio verso un processo decisionale guidato dall'IA non solo semplifica i processi operativi, ma inculca anche una cultura dell'innovazione in cui le intuizioni basate sui dati caratterizzano la pianificazione e l'esecuzione strategica. I Big Data hanno ulteriormente evidenziato il ruolo dell'IA nel processo decisionale e con l'aumento del volume e della complessità di questi, le forme di analisi tradizionali sono decisamente divenute inadeguate. In effetti, come indicano Duan et al. (2019), l'IA è uno strumento fondamentale per la gestione e la comprensione dei Big Data, in quanto converte le informazioni grezze in informazioni preziose[3]. Gli algoritmi di machine learning possono analizzare insiemi giganteschi di dati, riconoscere fatti pertinenti e scoprire modelli che sarebbero quasi impercettibili per l'occhio umano. Si tratta quindi di un elemento cruciale per ottenere un vantaggio competitivo, dal momento che le organizzazioni possono raccogliere dati utili che le aiutino a prendere decisioni adeguate, supportate da analisi in tempo reale.

Nonostante i numerosi vantaggi, l'adozione dell'IA nel processo decisionale può sollevare alcune questioni etiche. Le questioni relative alla privacy dei dati, alla distorsione (*bias*) degli algoritmi e alla trasparenza degli *AI-driven decision-making mechanisms* devono essere gestite con attenzione, in modo che l'impiego di queste tecnologie sia responsabile ed etico.

In un contesto moderno come quello odierno sempre più organizzazioni dipendono dall'Intelligenza artificiale per prendere decisioni su questioni fondamentali o semplici task di interazione con i clienti, siccome questo rapporto è in costante crescita sorge una necessità sempre maggiore di sviluppare e adottare solidi quadri etici al fine di promuovere la consapevolezza, l'equità e la trasparenza di questi nuovi sistemi.

Le potenzialità e le prospettive di impiego dell'Intelligenza Artificiale sono in continua evoluzione: gli aspetti principalmente analizzati nel presente elaborato si riferiscono all'applicazione dell'IA nei processi decisionali all'interno del settore finanziario. In particolare, l'obiettivo principale della ricerca è esplorare come l'IA possa ottimizzare l'utilizzo dei dati finanziari, migliorare l'accuratezza delle previsioni e rendere più efficiente la gestione delle informazioni; di conseguenza, costituisce il pilastro più

[3] Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management*.

importante per l'innovazione e l'efficienza in vari settori. Questo studio si propone di dimostrare come l'adozione di algoritmi di machine learning possa non solo incrementare l'efficacia delle decisioni finanziarie, ma anche rispondere alle nuove sfide normative e di mercato che emergono nel contesto contemporaneo.

La struttura della tesi è organizzata per fornire una comprensione graduale dell'argomento. Il primo capitolo esplora lo scenario normativo, con un focus sul Regolamento FiDA e le sue implicazioni per l'industria finanziaria, inclusi i Prestatori di Servizi di Informazione Finanziaria (PSIF). Il secondo capitolo analizza lo scenario tecnologico e le opportunità di mercato, partendo dalle criticità attuali nella gestione dei dati finanziari per poi approfondire l'integrazione dell'Intelligenza Artificiale nei processi esistenti. Viene illustrato come queste tecnologie possano migliorare l'efficienza e la rapidità nel decision-making, confrontando i metodi tradizionali con i modelli basati sull'IA, supportato da casi di studio di successo che evidenziano i benefici concreti.

Il terzo capitolo introduce un caso d'uso esemplificativo, sviluppando un algoritmo di machine learning per l'analisi dei dati finanziari, e mostra come questo possa ottimizzare i processi decisionali sia da un punto di vista dell’azienda che da un punto di vista del cliente, in linea con l'evoluzione normativa del Regolamento FiDA. Infine, il quarto capitolo offre considerazioni finali e possibili sviluppi futuri.

Capitolo 1: Scenario Normativo

1.1 Introduzione al Regolamento FiDA

Il Regolamento FiDA[4], proposto dalla Commissione Europea nel giugno 2023, mira a stabilire un quadro normativo chiaro e armonizzato per l'accesso ai dati finanziari all'interno dell'Unione Europea. Questo regolamento è concepito per facilitare la trasformazione digitale nel settore finanziario, promuovendo l'uso dei dati per migliorare l'efficienza e l'innovazione. Una delle motivazioni principali dietro l'introduzione di questo regolamento è il riconoscimento del ruolo fondamentale che i dati finanziari giocano in un'economia basata sui dati, e l'importanza di bilanciare l'ampio utilizzo di tali dati con alti standard di sicurezza e tutela della privacy.

Il Regolamento FiDA è strettamente connesso con la più ampia strategia europea per i dati e si inserisce nel contesto della finanza digitale, un ambito individuato dalla Commissione come prioritario già nel 2020. L'intento è di consentire ai consumatori e alle imprese un controllo più efficace sui propri dati finanziari, aumentando la fiducia nella condivisione di tali informazioni e stimolando l'adozione di servizi e prodotti finanziari innovativi basati sui dati. Attualmente, l'accesso ai dati finanziari in Europa è limitato e frammentato, con i clienti spesso incapaci di condividere facilmente le loro informazioni al di fuori dei conti di pagamento. Questo ha portato a una mancanza di personalizzazione dei servizi finanziari e ha ostacolato l'innovazione, impedendo alle imprese, in particolare alle piccole e medie imprese (PMI), di beneficiare pienamente delle opportunità offerte dalla digitalizzazione.

Il regolamento cerca di superare queste limitazioni introducendo norme specifiche che facilitano la condivisione dei dati finanziari in modo sicuro e controllato. Esso si basa sui principi stabiliti dalla direttiva sui servizi di pagamento (PSD2), che ha già permesso una certa apertura dei dati relativi ai conti di pagamento, e mira a estendere questa apertura a una gamma più ampia di informazioni finanziarie. L'obiettivo principale del regolamento è consentire un accesso più ampio ai dati finanziari dei clienti, permettendo una

[4] Proposta di Regolamento del Parlamento europeo e del Consiglio relativo a un quadro per l'accesso ai dati finanziari e che modifica i regolamenti (UE) n. 1093/2010, (UE) n. 1094/2010, (UE) n. 1095/2010 e (UE) 2022/2554 - COM/2023/360 final.

condivisione sicura ed efficiente di tali informazioni da parte di istituzioni finanziarie e nuovi attori di mercato, come i Prestatori di Servizi di Informazione Finanziaria. L'obiettivo finale è di creare un ambiente di mercato più competitivo e innovativo, in cui i consumatori possano accedere a servizi finanziari su misura e le imprese possano sviluppare nuovi modelli di business basati sui dati.

1.2 Accesso e utilizzo dei dati finanziari: Nuovi paradigmi

Il Regolamento FiDA chiarisce immediatamente, nei primi due commi dell'art. 1, gli obiettivi e l'oggetto della normativa, ossia da un lato la regolazione dell'accesso e dell'utilizzo di talune categorie di dati dei consumatori nell'ambito finanziario al di fuori del settore dei conti di pagamento[5], dall'altro la creazione di un nuovo soggetto giuridico denominato "Prestatore di Servizi di Informazione Finanziaria" (o "*Financial Information Service Provider*" o PSIF)[6], il tutto senza pregiudicare l'applicazione di altri atti giuridici dell'Unione in materia di accesso ai dati del cliente, e di condivisione degli stessi, salvo quanto espressamente previsto nel medesimo Regolamento[7].

Le categorie di dati oggetto del Regolamento sono descritte puntualmente dall'articolo 2, par. 1, che ne delinea quindi l'ambito di applicazione oggettivo, preordinato alla inclusione dei dati che apportano "*un elevato valore aggiunto per l'innovazione finanziaria nonché un basso rischio di esclusione finanziaria per i consumatori*"[8]. In particolare, anche se non prive di eccezioni, sei sono le categorie di dati che rientrano nella disciplina proposta[9].

La lettera a) dell'art. 2 del Regolamento FiDA comprende, in primo luogo, gli elementi indefettibili dei rapporti finanziari quali i contratti di credito ipotecario, i prestiti e i conti, da questi esclusi, come già indicato, i conti di pagamento, il tutto con l'obiettivo di

[5] Nel regolamento si evince espressamente come esso sia preordinato ad evitare sovrapposizioni con le disposizioni relative ai conti di pagamento dettate dalla direttiva (UE) 2015/2366, evitando così disallineamenti e incertezze normative. V. Considerando 12 del Regolamento FiDA.
[6] I PSIF sono disciplinati, in particolare, dal Titolo V del Regolamento FiDA.
[7] V. Art. 2, par. 4, del Regolamento FiDA, che nient'altro fa che esplicitare l'applicazione del *principio lex specialis derogat generali* nei limiti della incompatibilità tra normative orizzontali e verticali.
[8] V. Considerando 9 del Regolamento FiDA.
[9] Art. 2, par. 1 del Regolamento FiDA.

individuare quei dati che "*possono consentire ai clienti di avere una migliore visione d'insieme dei loro depositi e di soddisfare meglio le loro esigenze di risparmio sulla base dei dati sul credito*"[10]. La lettera b) contenuta nella medesima disposizione estende l'applicazione ad ulteriori dati relativi a servizi talvolta accessori, quali risparmi, investimenti in strumenti finanziari, prodotti di investimento assicurativi, cripto-attività, beni immobili e altre attività finanziarie correlate[11].

Le lettere c) e d) includono poi i dati dei clienti sui diritti pensionistici, relativi ai vari prodotti pensionistici sia aziendali, professionali[12] o individuali[13], la cui accessibilità e condivisione si ritiene fondamentale perché, per un verso, dovrebbe contribuire allo sviluppo di strumenti di tracciamento delle pensioni che forniscano ai risparmiatori una panoramica completa dei loro diritti[14] e, per altro verso, consentirebbe agli operatori di poter offrire prodotti e servizi assicurativi rilevanti per le esigenze del cliente, come la protezione di abitazioni, veicoli e altri beni[15].

---

[10] V. Considerando 12 del Regolamento FiDA, il quale precisa che *"I dati del cliente contenenti i dettagli del saldo, delle condizioni o delle operazioni relativi a mutui ipotecari, prestiti e risparmi possono consentire ai clienti di avere una migliore visione d'insieme dei loro depositi e di soddisfare meglio le loro esigenze di risparmio sulla base dei dati sul credito. Il presente regolamento dovrebbe riguardare i dati del cliente al di là dei conti di pagamento definiti nella direttiva (UE) 2015/2366. I conti di credito coperti da una linea di credito che non possono essere utilizzati per l'esecuzione di operazioni di pagamento a terzi dovrebbero rientrare nell'ambito di applicazione del presente regolamento. Il presente regolamento dovrebbe essere quindi inteso come riguardante l'accesso ai dettagli del saldo, delle condizioni o delle operazioni relativi a contratti di credito ipotecario, prestiti e conti di risparmio nonché i tipi di conti che non rientrano nell'ambito di applicazione della direttiva (UE) 2015/2366".*

[11] Nonché i benefici economici derivanti da tali attività, compresi i dati raccolti ai fini della valutazione dell'appropriatezza e dell'adeguatezza a norma dell'articolo 25 della direttiva 2014/65/UE relativa ai mercati degli strumenti finanziari. La finalità di includere tali dati nella presente disciplina ben si coglie nel Considerando 11, nel quale è riportato che "*Consentire ai clienti di condividere i loro dati sui loro investimenti correnti può incoraggiare l'innovazione nell'offerta di servizi di investimento al dettaglio*".

[12] V. Considerando 15. La norma sul punto fa richiamo ai prodotti paneuropei a norma del regolamento (UE) 2019/1238 e alla direttiva 2009/138/CE e alla direttiva (UE) 2016/2341 del Parlamento europeo e del Consiglio.

[13] La norma, sul punto, fa rinvio ai prodotti pensionistici individuali paneuropei, a norma del regolamento (UE) 2019/1238.

[14] V. Considerando 15 ove è altresì precisato che "*i dati sui diritti pensionistici riguardano in particolare i diritti pensionistici maturati, i livelli previsti delle prestazioni pensionistiche, i rischi e le garanzie per gli aderenti e i beneficiari*", giudicati dal legislatore un mercato particolarmente promettente che incontra attualmente numerose difficoltà proprio in considerazione della difficoltà nella circolazione dei dati.

[15] V. Considerando 14 in cui si precisa che "*I dati del cliente sui prodotti assicurativi che rientrano nell'ambito di applicazione del presente regolamento dovrebbero includere sia informazioni sui prodotti assicurativi, quali dettagli sulla copertura assicurativa, sia dati specifici relativi alle attività assicurate dei consumatori che sono raccolti al fine di verificare le richieste e le esigenze*".

La lettera e) si sofferma invece sui prodotti di assicurazione (non vita), escludendo esplicitamente i prodotti di assicurazione malattia e le valutazioni ad esse connesse[16], dato il loro carattere strettamente personale che comprometterebbe la privacy dei consumatori in maniera non proporzionata[17].

In ultimo, la lettera f) consente di ricomprendere nella normativa in analisi anche i dati utilizzati per la valutazione del merito creditizio di imprese raccolti nell'ambito di procedure di richiesta di prestito o di *rating* del credito. La scelta di non estendere questa categoria di dati a quella relativa alla valutazione del merito dei consumatori è voluta e valutata a valle di una considerazione sui rischi per i diritti fondamentali della persona[18], in particolare i diritti garantiti dagli articoli 7 e 8 della Carta dei diritti fondamentali dell'Unione, laddove risulta più opportuno sfruttare tali dati unicamente nell'ambito dei soggetti giuridici dalle micro alle grandi imprese[19].

L'art. 2, del Regolamento FiDA disciplina poi l'ambito di applicazione soggettivo della normativa in esame disponendo che, ad eccezione delle entità di cui all'articolo 2, paragrafo 3, lettere da a) a e), del Regolamento (UE) 2022/2554, sono assoggettati alla normativa gli enti creditizi, gli istituti di pagamento, compresi i prestatori di servizi di informazione sui conti e gli istituti di pagamento a cui è stata concessa un'esenzione a norma della direttiva (UE) 2015/2366, gli istituti di moneta elettronica, le imprese di investimento, i prestatori di servizi per le cripto-attività, gli emittenti di token collegati ad attività, i gestori di fondi di investimento alternativi, le società di gestione di organismi d'investimento collettivo in valori mobiliari, le imprese di assicurazione e di riassicurazione, gli intermediari assicurativi e intermediari assicurativi a titolo accessorio, gli enti pensionistici aziendali o professionali, le agenzie di *rating* del credito e i fornitori di servizi di crowdfunding, purché agiscano in qualità di titolari[20] o utenti[21] dei dati sopra illustrati.

---

[16] Ovvero i dati raccolti a norma degli articoli 20 e 30 della direttiva (UE) 2016/97 sulla distribuzione assicurativa.

[17] V. Considerando 9.

[18] Espressi nella relazione del Consiglio, punto 3 paragrafo 6 e ribaditi nel Considerando 18.

[19] V. Considerando 16.

[20] Tale è, ai sensi dell'art. 3, n. 5), del Regolamento Fida "*un ente finanziario diverso da un prestatore di servizi di informazione sui conti che raccoglie, conserva e altrimenti tratta i dati di cui all'articolo 2, paragrafo 1".*

[21] Tale è, ai sensi dell'art. 3, n. 6), del Regolamento Fida "*una delle entità di cui all'articolo 2, paragrafo 2, che, previa autorizzazione di un cliente, ha accesso legittimo ai dati del cliente di cui all'articolo 2, paragrafo 1*".

Il Titolo II del Regolamento FiDA è dedicato alla predisposizione degli obblighi spettanti ai diversi attori coinvolti in materia e innesca, dunque, quel processo virtuoso di condivisione e circolazione dei dati auspicata dal legislatore europeo[22], attualmente ostacolato dalla disomogeneità della materia[23].I primi due articoli (artt. 4 e 5 del Regolamento) disciplinano gli obblighi in capo ai titolari dei dati[24], essenziali per garantire le finalità della normativa, declinando e rafforzando nello specifico ambito finanziario i diritti di accesso, di portabilità e di condivisione dei dati come già declinati, rispettivamente, nel GDPR e nella PSD2. In particolare, l'art. 4 obbliga i titolari a mettere a disposizione del cliente i dati appartenenti ad una delle categorie già sopra indicate, senza indebito ritardo, gratuitamente[25], in maniera continuativa e in tempo reale ogni qualvolta siano richiesti per via elettronica. L'art. 5, poi, predispone l'obbligo dei titolari di mettere a disposizione, dietro richiesta del cliente, gli stessi dati ad ogni utente dei dati[26] debitamente autorizzato[27].

La condivisione dei dati dei clienti, da parte dei titolari, nei confronti degli utenti soggiace ai medesimi canoni espressi precedentemente, ad eccezione della gratuità: è infatti previsto, al secondo comma dell'art. 5, la possibilità per il titolare di richiedere un compenso[28] all'utente qualora siano forniti dati nelle modalità prescritte dal Regolamento[29]. Sono poi specificati al terzo comma ulteriori regole atte a disciplinare lo scambio di queste informazioni, a garanzia dell'integrità delle stesse e della tutela del cliente; in particolare è richiesto che i dati siano condivisi in un formato basato su norme generalmente riconosciute e che abbiano almeno la stessa qualità di quelli posseduti dal titolare[30] e devono inoltre essere comunicati su un canale sicuro e solo a seguito di una

[22] V. Considerando 11.
[23] V. Considerando 6 e 8.
[24] Ovvero a norma dell'art. 3 n. 5 "*un ente finanziario diverso da un prestatore di servizi di informazione sui conti che raccoglie, conserva e altrimenti tratta i dati di cui all'articolo 2, paragrafo 1*".
[25] In deroga all'esplicito diritto di rimborso del titolare alle spese amministrative di cui allo stesso art. 15 comma 3 GDPR.
[26] Ovvero a norma dell'art. 3 n. 6 "*una delle entità di cui all'articolo 2, paragrafo 2, che, previa autorizzazione di un cliente, ha accesso legittimo ai dati del cliente di cui all'articolo 2, paragrafo 1*".
[27] Qualora i dati siano classificabili come personali l'utente dovrà disporre di una lecita e valida base per il trattamento a norma del regolamento (UE) 2016/679 (v. Considerando 10).
[28] Anche al fine di garantire che i titolari abbiano un interesse nel fornire interfacce di elevata qualità (v. Considerando 29).
[29] Con particolare riferimento al sistema di condivisione di cui agli artt. 9 e 10.
[30] V. Considerando 24.

autorizzazione del cliente che potrà essere monitorata e gestita tramite un pannello di gestione regolato dall'articolo 8 dello stesso Regolamento.

Anche l'utente dei dati è sottoposto a obblighi non dissimili, delineati dall'articolo 6. Primariamente è disposto un obbligo di autorizzazione preventiva da parte dell'autorità competente[31] ovvero, qualora l'ente rientri nella nuova categoria dei prestatori di servizi di informazione finanziaria, una concessione specifica a norma dell'articolo 14. È rafforzato il principio di limitazione[32] imponendo all'utente l'uso esclusivo dei dati per le finalità e alle condizioni per le quali il cliente ha concesso l'autorizzazione[33], la quale potrà essere revocata *ad nutum*, qualora il rapporto si basi sul consenso, o conformemente agli eventuali obblighi contrattuali alla base del trattamento. In accordo con il principio di *privacy by design*[34], nell'ottica dell'*accountability*, è richiesta all'utente dei dati la predisposizione di "*adeguate misure tecniche, giuridiche e organizzative*" nonché "*un livello adeguato di sicurezza per la conservazione*" al fine di minimizzare i danni dovuti al fisiologico accentramento informatico di dati tanto rilevanti[35].

Il Titolo III mira ad una disciplina volta ad un "*Uso responsabile dei dati*", regolando innanzitutto il loro perimetro di utilizzo all'articolo 7. È ribadito il generale principio di minimizzazione e limitazione delle finalità[36] per quei dati del cliente che costituiscono dati personali, tuttavia la concreta individuazione dei livelli di tutela è rimandata ad orientamenti elaborati dall'Autorità Bancaria Europea, per i prodotti e i servizi connessi al punteggio di affidabilità creditizia del consumatore, e all'Autorità europea delle assicurazioni e delle pensioni aziendali e professionali per i prodotti e i servizi connessi alla valutazione del rischio e alla determinazione dei prezzi per un consumatore nel caso di prodotti di assicurazione vita e malattia. Queste ultime "*cooperano strettamente*" con il comitato europeo per la protezione dei dati[37] al fine di "*fornire un quadro proporzionato sulle modalità di utilizzo dei dati personali relativi a un consumatore*"[38].

---

[31] Designata dagli stati membri a norma dell'art. 17 e regolata dal Titolo VI del Regolamento.
[32] Di cui all'art. 5 lett. b del GDPR.
[33] È esplicitamente vietato l'utilizzo dei dati del cliente per finalità pubblicitarie, ad eccezione del marketing diretto conformemente al diritto dell'Unione e nazionale.
[34] Di cui all'art. 25 GDPR.
[35] Il Considerando 32 ricorda a tal proposito le prescrizioni in materia di ciberresilienza dettate dal Regolamento (UE) 2022/2554, con l'esplicita volontà di includervi i prestatori di servizi di informazione finanziaria.
[36] Di cui all'art. 5, comma 1, lett. b e c del GDPR.
[37] Istituito dal regolamento (UE) 2016/679.
[38] Il Considerando 19 fa presente come una corretta individuazione dei perimetri in questi due specifici settori deve presupporre "*la coerenza tra l'ambito di applicazione del presente regolamento, che esclude i*

La concreta attuazione della disciplina fin qui esposta ha come prodotto un "*pannello di gestione delle autorizzazioni*"[39], uno strumento sviluppato dal titolare dei dati[40] con lo scopo di fornire un'interfaccia front-end atta a monitorare e gestire le autorizzazioni fornite dal cliente agli utenti dei dati[41]. L'articolo 8 comma 2 riporta le funzionalità essenziali di questa interfaccia, la quale:

a) *offre al cliente una panoramica di ogni autorizzazione in corso concessa agli utenti dei dati, tra cui:*
    i. *il nome dell'utente dei dati cui è stato concesso l'accesso;*
    ii. *il conto, prodotto finanziario o servizio finanziario del cliente cui è stato concesso l'accesso;*
    iii. *la finalità dell'autorizzazione;*
    iv. *le categorie di dati condivisi;*
    v. *il periodo di validità dell'autorizzazione;*
b) *consente al cliente di revocare l'autorizzazione concessa a un utente dei dati;*[42]
c) *consente al cliente di ripristinare un'autorizzazione revocata;*
d) *comprende un registro delle autorizzazioni revocate o scadute per un periodo di due anni.*"

Questo pannello di gestione dovrà essere facilmente reperibile e le informazioni ivi contenute dovranno essere chiare, accurate e facilmente comprensibili per il cliente, nonché aggiornate in tempo reale. Il titolare e l'utente dei dati dovranno collaborare per mettere le informazioni a disposizione tramite il pannello di gestione, determinando obblighi di informazione reciprochi. Il titolare dovrà infatti informare l'utente di ogni

dati che fanno parte della valutazione del merito creditizio di un consumatore nonché i dati relativi all'assicurazione vita e malattia dello stesso, e l'ambito di applicazione degli orientamenti, che formulano raccomandazioni sul modo in cui i tipi di dati provenienti da altri ambiti del settore finanziario che rientrano nell'ambito di applicazione del presente regolamento possono essere utilizzati per fornire tali prodotti e servizi."

[39] Talvolta già presente nella prassi per vari altri tipi di autorizzazioni, ma mai positivizzato e comunque diffuso in modo non uniforme dati i suoi costi di attuazione (V. Considerando 7).

[40] Il Considerando 22 fa salva la possibilità che piccole e medie imprese che agiscono in qualità di titolari dei dati siano autorizzate a istituire congiuntamente un'interfaccia per programmi applicativi, in modo da ridurre i costi a carico di ciascuna, ovvero avvalersi di fornitori esterni di tecnologia che gestiscono interfacce di programmazione in modalità "*pay-per-call*".

[41] "*Anche quando i dati personali sono condivisi sulla base del consenso o sono necessari per l'esecuzione di un contratto*" (V. Considerando 22).

[42] Si veda il Considerando 10 per una panoramica sulle dinamiche giuridiche dell'autorizzazione e della revoca.

modifica relativa alla sua autorizzazione, mentre ogni utente dovrà condividere con il titolare la finalità di questa, il suo periodo di validità e le categorie di dati interessate.
La collaborazione tra titolari e utenti dei dati è, dunque, elemento essenziale per la corretta tutela dei diritti dei clienti, la sua organizzazione[43] pertanto non è liberamente demandata alle dinamiche del mercato ma disciplinata tramite l'individuazione puntuale di "*sistemi di condivisione dei dati finanziari*" al Titolo IV.
Entro 18 mesi dall'entrata in vigore del Regolamento sarà fatto obbligo per titolari e utenti di aderire ad uno o più di questi sistemi di condivisione dei dati finanziari con le caratteristiche di *governance* e con il contenuto previste dall'articolo 10. In particolare, i soggetti previsti sono non solo utenti e titolari "*che rappresentano una quota significativa del mercato del prodotto o del servizio in questione*", ma anche organizzazioni dei clienti e associazioni dei consumatori. Il sistema di governo interno prescritto è rigorosamente paritario sia con riguardo alla rappresentanza nei processi decisionali interni,[44] adottando un sistema maggioritario per teste e talvolta per categorie[45], che in merito all'accesso[46], non dovendo comportare oneri ulteriori rispetto a quanto previsto dallo stesso regolamento e da altre norme dell'Unione applicabili e garantendo che le adesioni dei nuovi membri sia sempre possibile e alle stesse condizioni applicate in qualsiasi momento ai membri esistenti.
Le norme oggetto di questo sistema comprendono regole di trasparenza, obblighi di comunicazione, nonché "*norme comuni per i dati e le interfacce tecniche per consentire ai clienti di richiedere la condivisione dei dati*"[47], stabiliscono, inoltre, un modello per determinare il compenso massimo[48] che il titolare dei dati ha il diritto di addebitare per la

[43] Leggasi "*l'interazione contrattuale e tecnica*" di cui al Considerando 25.
[44] L'art. 10 lett. a sub-lett. i chiarisce che la quota di mercato non influenza i diritti del soggetto e "*ciascuna parte gode di una pari ed equa rappresentanza nei processi decisionali interni del sistema e di pari peso nelle procedure di voto*".
[45] L'art. 10 lett. e impone al sistema di prevedere "*un meccanismo attraverso il quale è possibile modificarne le norme, a seguito di un'analisi d'impatto e del consenso della maggioranza, rispettivamente, di ciascuna comunità di titolari dei dati e di utenti dei dati*".
[46] Lo stesso art. 10 lett. c impone che le norme "*garantiscano che la partecipazione al sistema sia aperta a qualsiasi titolare e utente dei dati sulla base di criteri oggettivi e che tutti i membri siano trattati in modo equo e paritario*".
[47] Queste ultime "*elaborate dagli aderenti al sistema o da altre parti od organismi*" (art. 10 lett. g), coerentemente ad una linea che incoraggia la standardizzazione degli aspetti tecnici per motivi sia economici che di sicurezza (si veda, tra molti, i considerando 5, 23, 24, 28).
[48] Nella considerazione già accennata che "*agevolare l'accesso ai dati a fronte di un compenso garantirebbe un'equa distribuzione dei relativi costi tra i titolari dei dati e gli utenti dei dati lungo la catena del valore dei dati*" (Considerando 29).

messa a disposizione dei dati attraverso il pannello di gestione. Sono elencate sei caratteristiche di questo modello di determinazione del compenso atte a garantire parità di trattamento, efficienza e uno scambio ordinato dei dati: si richiede una correlazione diretta tra la messa a disposizione dei dati e l'imputazione del compenso dovuto, basandosi su una "*metodologia obiettiva, trasparente e non discriminatoria*" che in ogni caso sia orientata "*verso i livelli più bassi prevalenti sul mercato*", da adattare periodicamente tenendo in considerazione i progressi tecnologici. Qualora l'utente sia un soggetto particolarmente debole come PMI o microimpresa, il compenso non può superare "*i costi direttamente connessi alla messa a disposizione dei dati al destinatario dei dati e imputabili alla richiesta*"[49].[50]

Il sistema così strutturato deve essere notificato alle autorità competenti del luogo di stabilimento dei tre titolari dei dati più significativi[51] e ogni soggetto comunica ogni sua nuova adesione all'autorità competente dello Stato membro in cui è stabilito. Il benestare dell'autorità competente[52] entro un mese dalla notifica comporterà la comunicazione all'Autorità Bancaria Europea che aggiornerà il proprio registro assicurando il riconoscimento del sistema a livello "eurounitario".

L'articolo 11 in chiusura di Titolo richiama lo strumento dell'atto delegato di cui all'art. 290 del Trattato sul Funzionamento dell'Unione Europea, per evitare che l'inerzia del mercato renda ineffettiva la tutela prevista dal Regolamento. È dunque affidato alla Commissione[53] il potere di integrare in futuro il Regolamento "*nel caso in cui non sia stato realizzato alcun sistema di condivisione dei dati finanziari per una o più categorie di dati*" ovvero "*non vi sia alcuna prospettiva realistica che tale sistema sia istituito entro un lasso di tempo ragionevole*"[54]. I limiti di questo potere risiedono nel conferimento della delega che consta di tre specifici punti, nel dettaglio l'articolo 11: "*a) norme comuni per i dati e, se del caso, le interfacce tecniche per consentire ai clienti di richiedere la*

[49] Il legislatore ritiene infatti che "*la proporzionalità per i partecipanti al mercato più piccoli dovrebbe essere garantita limitando rigorosamente il compenso ai costi sostenuti per agevolare l'accesso ai dati*" (Considerando 29).

[50] V. Art. 10 lett. i-j – RESPONSABILITÀ.

[51] E "*qualora i tre titolari dei dati più significativi siano stabiliti in Stati membri diversi o qualora vi sia più di un'autorità competente nello Stato membro di stabilimento dei tre titolari dei dati più significativi, il sistema è notificato a tutte le autorità in questione, le quali concordano tra loro quale autorità effettuerà la valutazione di cui al paragrafo 6*". (art. 10 par. 4).

[52] Atto a garantire la sussistenza delle caratteristiche di governance di cui all'art. 10 par. 1.

[53] Nel rispetto dei principi stabiliti nell'accordo interistituzionale "Legiferare meglio" del 13 aprile 2016 e le ulteriori accortezze di cui al Considerando 27.

[54] V. Art. 11.

condivisione dei dati a norma dell'articolo 5, paragrafo 1; b) un modello per determinare il compenso massimo che il titolare dei dati ha il diritto di addebitare per la messa a disposizione dei dati; c) la responsabilità delle entità coinvolte nella messa a disposizione dei dati del cliente."

Il Titolo V è relativo alla disciplina dei Prestatori di Servizi di Informazione Finanziaria, la cui figura, però, verrà discussa nella sezione *1.3*.

Passando ad ultimo in rassegna la disciplina delle autorità competenti, l'articolo 17 affida ad ogni Stato membro il compito di designare un'autorità tra quelle che dispongono delle risorse necessarie, in particolare in termini di personale dedicato, al fine di adempiere agli obblighi previsti dal presente Regolamento[55], comunicandolo senza indugio alla Commissione. I poteri riconosciuti dall'articolo 18 comprendono poteri di accesso finalizzati al controllo di tutte le informazioni necessarie, nonché veri e propri poteri istruttori di ispezione e la facoltà di richiede misure cautelari quali il congelamento o il sequestro di beni. Possono inoltre imporre ai prestatori di servizi di hosting di sopprimere, disabilitare o limitare l'accesso a un'interfaccia online ovvero ordinare alle autorità di registrazione del dominio di cancellare un nome di dominio completo, tenendo in debita considerazione il danno complessivo effettivo o potenziale della violazione. La loro azione può esprimersi anche mediante collaborazione con altre autorità, delegando poteri ad altre autorità od organismi ovvero ricorrendo all'autorità giudiziaria. È fatta salva per l'autorità la possibilità di chiudere un'indagine relativa a una presunta violazione tramite un accordo transattivo ovvero adottando speciali procedure accelerate nazionali.

Le sanzioni previste sono sostanzialmente di carattere amministrativo[56], possono consistere in sospensioni provvisorie o definitive dell'autorizzazione, interdizione temporanea[57] dei membri dell'amministrazione o, per le persone fisiche, sanzioni amministrative pecuniarie fino a 25 000 EUR per violazione e fino a un totale di 250 000 EUR all'anno, raddoppiate per le persone giuridiche. Le autorità competenti hanno il diritto di imporre sanzioni per la reiterazione dell'inadempimento in caso di persistente inosservanza di qualsiasi decisione, ingiunzione, misura provvisoria, richiesta, obbligo o altra misura amministrativa adottata. Le circostanze da tenere in considerazione per

[55] V. Art. 17 comma 2.

[56] Fatte salve quelle che sono previste a norma del diritto penale nazionale, che andranno notificate alla Commissione qualora rilevanti.

[57] Che in caso di reiterazione può arrivare fino a dieci anni.

determinare le sanzioni sono indicate dall'articolo 22 e comprendono la gravità e la durata della violazione, il grado di responsabilità, la capacità finanziaria della persona fisica o giuridica responsabile nonché l'ingiusto guadagno derivato e l'impatto della violazione sugli interessi dei clienti[58].

La Commissione effettuerà entro quattro anni dall'entrata in vigore del Regolamento una valutazione sul suo impatto da presentare al Parlamento europeo, al Consiglio e al Comitato economico e sociale europeo, anche al fine di individuare ulteriori categorie di dati da ricomprendere nel suo ambito di applicazione, aggiornare le pratiche contrattuali dei titolari e degli utenti dei dati e tenere traccia dell'evoluzione nel funzionamento dei sistemi di condivisione dei dati finanziari. A questa valutazione sarà corredata, sempre dalla Commissione, una più ampia analisi atta a delineare l'effetto combinato del presente Regolamento con la direttiva (UE) 2015/2366 relativa ai servizi di pagamento.

1.3 Ruolo dei Prestatori di Servizi di Informazione Finanziaria

I Prestatori di Servizi di Informazione Finanziaria (PSIF) sono una nuova figura introdotta dal regolamento europeo sull'accesso ai dati finanziari, con l'obiettivo di facilitare il processo di condivisione delle informazioni, assicurando che il flusso di dati tra diverse istituzioni finanziarie e terze parti avvenga in modo sicuro e senza interruzioni, contribuendo a migliorare l'efficienza complessiva del mercato finanziario europeo. Operano dunque come intermediari autorizzati e si inseriscono in un contesto normativo armonizzato a livello europeo, che include regolamenti come il GDPR (Regolamento generale sulla protezione dei dati) e il DORA (Atto sulla resilienza operativa digitale), con un forte focus sulla sicurezza dei dati. Essi sono tenuti a rispettare rigorosi requisiti di protezione dei dati personali e a adottare misure avanzate di sicurezza cibernetica, assicurando che le informazioni sensibili siano trattate in modo conforme alle normative. I PSIF sono chiamati a svolgere un ruolo essenziale nell'ecosistema finanziario europeo e, agendo come custodi dell'informazione, promuovono un accesso sicuro, regolamentato

[58] Questi i caratteri principali tra i dodici individuati dall'art. 22 par. 1.

e trasparente ai dati finanziari con l'obiettivo di sostenere l'innovazione e migliorare la personalizzazione dei servizi finanziari per consumatori e aziende.

Il Titolo V del regolamento è dedicato all'istituzione dei Prestatori di Servizi di Informazione Finanziaria, soggetti ai quali è riconosciuto un regime speciale di accesso ai dati che evita la richiesta del cliente per via elettronica[59], bastando l'autorizzazione dell'autorità competente a garantire l'accesso.

La domanda di autorizzazione a norma dell'art. 12 par. 2 deve illustrare dettagliatamente gli assetti tecnici[60] e organizzativi[61] dell'ente e delle sue attività, deve essere indicato il tipo di accesso ai dati previsto, la natura giuridica e lo statuto del richiedente con puntuale indicazione degli amministratori e delle persone responsabili della gestione, che dovranno attestare la loro onorabilità e il possesso di conoscenze e di esperienza adeguate. Il paragrafo 3 impone la sottoscrizione di un'assicurazione per responsabilità civile professionale[62] per la copertura della responsabilità derivante dall'accesso o l'uso non autorizzato o fraudolento dei dati.[63] L'Autorità Bancaria Europea presenterà entro nove mesi dall'entrata in vigore del Regolamento alla Commissione dei progetti di norme tecniche atte a delineare una metodologia di valutazione comune per la concessione dell'autorizzazione e integrare i requisiti della domanda.[64]

Qualora siano accertati i presupposti l'autorizzazione sarà rilasciata dalla stessa ABE entro tre mesi ed iscritta in un registro appositamente creato, se il PSIF non è stabilito all'interno dell'Unione sarà necessaria, inoltre, l'individuazione di un rappresentante all'interno nel territorio comunitario che si assuma la responsabilità per la ricezione, il rispetto e l'applicazione del Regolamento[65]. L'accesso transfrontaliero intraeuropeo è invece disciplinato dall'articolo 28, che richiede al primo accesso una notifica all'autorità

[59] In deroga, dunque, alle modalità descritte dall'art. 5 par. 1.

[60] Sono ricordati in via generale gli obblighi del Regolamento (UE) 2022/2554, Capo II, in materia di resilienza operativa digitale (art. 12 par. 2 comma 3) e altre disposizioni specifiche dello stesso in materia di operazioni critiche e servizi TIC, in particolare al Capo III relativo alla gestione degli incidenti e ai reclami (art. 12 par. 2 lett. c, d, e).

[61] I dispositivi di governance, gli assetti di controllo, l'organizzazione strutturale compresi eventuali accordi di esternalizzazione nonché un piano aziendale contenente la stima provvisoria del bilancio per i primi 3 esercizi (art. 12 par. 2 lett. b, c, g).

[62] O garanzia analoga.

[63] In alternativa è possibile iniziare ugualmente l'attività con un capitale iniziale minimo di 50.000€, che potrà essere sostituito successivamente dall'assicurazione.

[64] Tenendo conto degli elementi dell'art. 12 par. 4 comma 2.

[65] A patto che il paese terzo non sia inserito nell'elenco delle giurisdizioni non cooperative a fini fiscali ai sensi della pertinente politica dell'Unione.

competente dello Stato membro di origine con riferimento al tipo di dati a cui si desidera accedere, lo Stato membro o gli Stati membri in cui intende accedere e il sistema di condivisione finanziaria a cui aderisce, in modo da permettere una corretta comunicazione tra le diverse autorità.

L'autorizzazione può essere revocata qualora non ce ne si avvalga per un periodo superiore a 12 mesi, ovvero quando le dichiarazioni fornite dovessero risultare false, incomplete o non più sussistenti e in ogni caso qualora la gestione costituisse "*un rischio per la protezione dei consumatori e la sicurezza dei dati.*"

1.4 Impatti Potenziali sull'Ecosistema Finanziario Europeo

Il Regolamento FiDA ha il potenziale di trasformare profondamente l'ecosistema finanziario europeo, rendendolo più integrato, aperto e orientato alla condivisione dei dati. L'obiettivo è dunque quello di sviluppare un ambiente dove i dati siano una risorsa centrale per migliorare i servizi offerti dagli operatori del mercato e rafforzare la competitività europea.

Uno degli effetti più evidenti del Regolamento FiDA sarà l'evoluzione verso una finanza sempre più data-driven; grazie all'accessibilità dei dati e all'integrazione di tecnologie come il machine learning e l'intelligenza artificiale, si punterà ad una standardizzazione e interoperabilità delle informazioni finanziarie tramite l'introduzione di interfacce comuni che esemplificheranno la circolazione dei dati tra i vari sistemi finanziari europei, andando così a ridurre i costi operativi migliorandone però l'efficienza. Tale migliore flusso che si viene a generare è dovuto in parte anche alla riduzione delle barriere normative, resa possibile dall'introduzione del "passaporto europeo" per i Prestatori di Servizi di Informazione Finanziaria, che permetterà loro di operare liberamente in tutti gli Stati membri dell'Unione Europea.

Esempi concreti di come questi nuovi paradigmi stimoleranno la concorrenza si possono ricavare facilmente dal settore bancario: l'ingresso di nuovi attori (come le FinTech) è agevolato e potranno quindi competere direttamente con le banche tradizionali, le quali verranno catapultate in uno scenario in cui saranno spinte a innovare e migliorare i propri servizi digitali, con l'opportunità, anche per le piccole banche, di sviluppare piattaforme

di online banking più efficienti e competitive; oppure ancora si pensi alle società di consulenza finanziaria: l'accesso a dati più dettagliati permetterà di fornire consulenze personalizzate con maggiore precisione, elevando l'elaborazione di piani d'investimento su misura con conseguente aumento della soddisfazione dei clienti e consolidando il loro ruolo come partner strategici nella gestione del patrimonio.

Capitolo 2: Scenario Tecnologico

2.1 Criticità attuali nei processi di gestione dei dati finanziari

Nel contesto odierno, la gestione dei dati finanziari rappresenta una sfida complessa per le istituzioni di tutto il mondo. L'aumento della mole di dati da elaborare, l'esigenza di conformarsi a normative sempre più stringenti e la crescente richiesta di soluzioni tecnologiche avanzate per garantire l'integrità e la sicurezza delle informazioni sono tra le principali problematiche che queste realtà devono affrontare. Una gestione efficace dei dati non è solo cruciale per assicurare la continuità operativa, ma anche per preservare la competitività e la reputazione delle organizzazioni stesse. Questo processo implica non soltanto la conservazione e l'elaborazione dei dati, ma anche la loro migrazione, integrazione e protezione dai rischi potenziali, come la perdita di informazioni o le violazioni della sicurezza. Per affrontare tali sfide, è necessario adottare un approccio olistico e coordinato che comprenda una pianificazione strategica adeguata e la disponibilità di risorse altamente qualificate.

Una delle difficoltà più rilevanti nella gestione dei dati finanziari deriva dalla complessità delle fonti e dei sistemi, spesso eterogenei, che rendono difficile centralizzare e standardizzare le informazioni. Molte istituzioni stanno adottando il *data fabric* come soluzione per integrare e gestire i dati in modo più efficiente. Secondo uno studio di InterSystems (2022), questa tecnologia rappresenta "*il futuro della gestione dei dati*", consentendo di unificare informazioni frammentate attraverso diversi silos organizzativi. Questo non solo migliora la qualità e la credibilità delle informazioni, ma riduce anche i rischi di incoerenza e perdita di dati[66]. Un sistema di gestione dei dati frammentato non solo influisce negativamente sull'efficienza operativa, ma può anche compromettere la capacità di prendere decisioni strategiche basate su informazioni affidabili. Come evidenziato da Cannone (2021), l'ottimizzazione della gestione dei dati passa principalmente attraverso la riduzione della duplicazione delle informazioni e la frammentazione dei sistemi[67]. L'adozione di tecnologie centralizzate, come il *data fabric*,

[66] InterSystems. (2022). *The top data and technology challenges in financial services*.
[67] Cannone, V. (2021, February 4). *Top data management challenges faced by financial firms*. The Golden Source.

consente alle organizzazioni di avere una visione coerente dei propri dati, migliorando così la capacità di analisi e decision-making.

La migrazione dei dati rappresenta un'altra area critica e problematica. Trasferire dati da un sistema a un altro comporta rischi significativi, come la perdita di informazioni o la compromissione della loro integrità. In accordo a un reportage di Hussain (2024) uno dei problemi più frequenti durante la migrazione è la perdita di dati dovuta all'incompatibilità tra sistemi legacy e nuove tecnologie[68]. Questo non solo può compromettere la qualità dei dati, ma può anche causare gravi conseguenze operative, come l'interruzione delle attività aziendali e il mancato rispetto delle normative vigenti. Ridurre questi rischi richiede l'adozione di strumenti automatizzati di verifica e una stretta collaborazione tra i dipartimenti IT e compliance per garantire che i processi di migrazione siano sicuri ed efficienti. La qualità dei dati è strettamente connessa alla migrazione e rappresenta una componente essenziale per il successo di qualsiasi processo decisionale. Dati inaccurati o incompleti possono condurre a decisioni mal informate, con potenziali conseguenze finanziarie disastrose. Secondo Intone (2024), una delle principali sfide nel mantenere elevati standard di qualità dei dati è la mancanza di framework efficaci per garantirne l'integrità e l'affidabilità[69]. Un framework di gestione della qualità consente di rilevare, correggere e prevenire errori, migliorando la precisione e l'attendibilità delle informazioni utilizzate per le decisioni aziendali.

Oltre alla qualità, un'altra area critica nella gestione delle informazioni finanziarie è la sicurezza dei dati. Le istituzioni finanziarie gestiscono ogni giorno enormi quantità di dati sensibili, il che le rende obiettivi primari per attacchi informatici. Secondo un articolo di Herzberg (2024), le principali minacce includono violazioni della sicurezza, furti di informazioni e attacchi ransomware[70]. Il crescente numero di attacchi ha messo in luce la necessità di misure di sicurezza più scrupolose, come l'implementazione di soluzioni di crittografia avanzate, l'uso di firewall e una formazione adeguata del personale sui rischi informatici. Tuttavia, garantire la protezione dei dati richiede più di un semplice investimento in tecnologia: è necessario un cambiamento culturale all'interno delle

[68] Hussain, E. (2024, September 2). *12 challenges in financial data migration and how IT leaders tackle them*. Data Ladder.
[69] Intone. (2024, April 30). *Challenges and solutions in implementing financial data quality management frameworks*.
[70] Herzberg, B. (2024, January 16). *5 biggest challenges of data security in the financial service industry*. Due.com.

organizzazioni. La sicurezza deve diventare una priorità a tutti i livelli, promuovendo un monitoraggio continuo delle infrastrutture e applicando rigorosi protocolli di accesso ai dati.

In definitiva, per prosperare in un contesto come quello attuale, le istituzioni devono integrare tecnologie avanzate con una cultura aziendale focalizzata sulla gestione responsabile dei dati. Solo così sarà possibile affrontare concretamente le sfide sempre più innovative proposte da un ambiente in costante evoluzione.

2.2 Benefici dell'Intelligenza Artificiale nel sistema finanziario odierno

L'integrazione dell'intelligenza artificiale nei processi finanziari esistenti rappresenta un passo naturale nella trasformazione digitale del settore. Contrariamente a quanto si potrebbe pensare, l'IA non sostituisce i modelli operativi tradizionali, ma li potenzia significativamente, migliorando la produttività, la precisione e la velocità nel prendere decisioni critiche. Le soluzioni basate sull'IA consentono alle istituzioni finanziarie di processare enormi volumi di dati istantaneamente, permettendo di interpretare e rispondere con rapidità alle mutevoli condizioni di mercato. Secondo quanto riportato da Feng (2024), l'introduzione dell'IA nei servizi finanziari ha un impatto notevole sulla resa operativa, grazie alla sua capacità di automatizzare attività ripetitive e complesse che, gestite manualmente, richiederebbero considerevoli risorse di tempo e manodopera impegnando più del necessario le risorse umane che potrebbero invece concentrarsi su compiti più strategici e complessi[71]. Un esempio concreto dell'automazione è la gestione delle richieste di credito. L'intelligenza artificiale non solo accelera il processo decisionale, approvando o rifiutando le transazioni in pochi minuti, ma aumenta anche l'accuratezza delle verifiche, riducendo il rischio di errori umani. Questa automazione migliora significativamente l'esperienza cliente, consentendo tempi di risposta rapidi e maggiore trasparenza. Un ulteriore vantaggio risiede nella capacità dell'IA di rendere i

[71] Feng, S. (2024). *Integrating artificial intelligence in financial services: Enhancements, applications, and future directions*. Semantic Scholar.

sistemi finanziari più dinamici e resilienti, permettendo alle istituzioni di adattarsi prontamente ai cambiamenti del mercato. Oltre a migliorare l'efficienza operativa, l'adozione dell'IA esercita un impatto considerevole sui processi decisionali, in particolare nella gestione dei rischi. Nasir e Gasmi (2024) evidenziano come l'intelligenza artificiale stia rivoluzionando la valutazione dei rischi finanziari, attraverso lo sviluppo di analisi predittive sempre più affidabili. Prima dell'introduzione di queste tecnologie, la valutazione dei rischi associati a specifiche operazioni si basava su modelli statici e analisi manuali, un approccio valido ma limitato nella capacità di elaborare grandi quantità di dati in tempi brevi. Con l'IA, le istituzioni finanziarie possono ora integrare una gamma più ampia di variabili nei loro modelli di rischio, migliorando la precisione delle previsioni. Ad esempio, un algoritmo di machine learning può analizzare sia dati storici che in tempo reale su eventi macroeconomici, identificando con precisione trend e schemi nascosti che potrebbero sfuggire ai metodi tradizionali. Inoltre, viene sottolineato che le tecnologie intelligenti riducono significativamente il margine di errore umano. I sistemi computazionali avanzati sono in grado di apprendere autonomamente dai dati passati e di adattarsi in modo dinamico alle nuove condizioni di mercato[72]. Questo processo di apprendimento continuo permette di aggiornare costantemente i modelli di rischio, riflettendo in tempo reale i cambiamenti nel panorama economico globale.

Oltre al perfezionamento della gestione del rischio, l'IA potenzia il processo decisionale in modo più ampio, soprattutto nel contesto finanziario. Come evidenziato da Stone e Suja (2024), gli algoritmi avanzati, combinati con l'analisi predittiva, permettono di identificare pattern e tendenze nascoste che possono avere un impatto significativo sull'intera strategia aziendale. Questo tipo di analisi è fondamentale per rilevare opportunità di investimento o segnali di rischio imminente. Ad esempio, nell'ambito della gestione patrimoniale, gli algoritmi di apprendimento automatico possono analizzare miliardi di transazioni per identificare opportunità di investimento nascoste, permettendo ai gestori di portafoglio di agire responsabilmente e ridurre i rischi di perdite improvvise. L'intelligenza artificiale, inoltre, gioca un ruolo cruciale nella prevenzione delle frodi, rilevando comportamenti anomali e segnalando attività sospette. Attraverso l'analisi di transazioni bancarie e modelli comportamentali, l'IA migliora la conformità normativa e

[72] Nasir, W., & Gasmi, S. (2024, Agosto). *Revolutionizing Risk Assessment in Finance through AI-Driven Decision-Making*. ResearchGate.

mitiga i rischi legati alle truffe finanziarie[73]. L'impiego dell'intelligenza artificiale offre un chiaro vantaggio competitivo alle aziende che ne fanno uso. In accordo con quanto indicato da Nanda et al. (2024), le organizzazioni finanziarie che implementano modelli intelligenti riescono a ridurre i tempi operativi, migliorare l'efficienza complessiva e ottimizzare le risorse, tagliando i costi associati ai processi complessi e intensivi[74]. Inoltre, l'IA consente di migliorare significativamente il servizio clienti, grazie a tecniche di indagine avanzate che permettono di personalizzare le offerte in base ai comportamenti e alle preferenze degli utenti, aumentando così la soddisfazione dei consumatori. Le aziende possono utilizzare sistemi di targeting per prevedere i bisogni dei clienti, sviluppando nuovi prodotti e servizi in modo più rapido ed efficace, mantenendo così un vantaggio strategico rispetto alla concorrenza. L'integrazione dell'intelligenza artificiale non solo porta a miglioramenti operativi immediati, ma rappresenta anche un investimento strategico a lungo termine. Grazie alla capacità di adattarsi rapidamente alle dinamiche di mercato e di innovare costantemente, le istituzioni finanziarie possono garantire una crescita sostenibile e mantenere una posizione di leadership in un settore altamente competitivo. L'IA non solo ottimizza i processi interni, ma offre la flessibilità necessaria per affrontare crisi economiche, fluttuazioni del mercato o cambiamenti normativi, assicurando che le istituzioni rimangano competitive anche in tempi di incertezza economica.

In conclusione, l'utilizzo di tecnologie cognitive nel settore finanziario non solo offre benefici in termini di efficienza operativa e gestione del rischio, ma favorisce l'innovazione connessa all'automazione e la capacità decisionale.

2.3 Confronto tra approccio tradizionale e modello IA

Per cogliere appieno le differenze tra i due approcci, è fondamentale esaminare esempi concreti corroborati da dati che possano agevolare la comprensione. A tal proposito in

[73] Stone, R., & Suja, S. (2024, Settembre). *Leveraging Predictive Analytics and Big Data in Machine Learning for Real-Time Financial Decision-Making*. ResearchGate.

[74] Nanda, A., Mohapatra, N., Rautela, R., & Acharya, D. (2024, Agosto 15). *Role of Artificial Intelligence (AI) Driven Business Excellence for Effective Corporate Operations and Competitive Advantage: An Empirical Study*.

questo paragrafo verrà illustrato un confronto esauriente attraverso l'analisi comparativa redatta da Melnyk et al. (2022). Questo paper esplora in profondità il cambiamento in atto nel settore bancario, confrontando i modelli bancari tradizionali con le soluzioni innovative delle FinTech, come accennato nella conclusione del Capitolo 1. L'obiettivo principale è capire come le ambizioni delle banche commerciali stiano evolvendo in parallelo all'espansione del mercato delle imprese digitali finanziarie e quale sia il futuro possibile delle due entità, con particolare attenzione a una possibile simbiosi strategica tra i modelli.

Le banche tradizionali, come evidenziato nel paper, operano secondo modelli consolidati e spesso regolamentati che prevedono la raccolta di depositi e la concessione di prestiti. Tali istituzioni, per decenni, hanno goduto di un quasi monopolio sulla gestione delle informazioni dei clienti, determinando chi potesse ricevere prestiti e gestire i risparmi. Tuttavia, la crisi finanziaria globale del 2008-2009 ha indebolito la fiducia nei confronti delle banche, creando lo spazio per l'ingresso di startup finanziarie digitali, che sono riuscite a offrire nuovi servizi finanziari a costi inferiori e con maggiore efficienza. D'altro canto, le FinTech rappresentano un'innovazione tecnologica, che può comprendere sistemi di trasferimento di denaro online, piattaforme di prestito *peer-to-peer* e crowdfunding. Queste soluzioni, integrate con algoritmi IA e machine learning, promuovono un modello di business completamente nuovo, capace di trasformare radicalmente il mercato tradizionale[75]. Secondo dati forniti da CB Insights, già nel 2019, esistevano circa 5.000 startup FinTech nel mondo, il doppio rispetto a tre anni prima. Inoltre, gli investimenti in capitale di rischio nel settore delle tecnologie finanziarie hanno raggiunto i 2,9 miliardi di dollari nel 2019, un aumento rispetto ai 2,3 miliardi dell'anno precedente[76].

L'adozione di queste soluzioni innovative ha già mostrato impatti tangibili attraverso l'ascesa di numerose aziende emergenti, che stanno rimodellando il panorama del settore finanziario. Due esempi rilevanti discussi nell'indagine sono le startup Dave e Betterment.

Dave, un'applicazione che consente agli utenti di evitare le commissioni bancarie per gli scoperti di conto, ha trasformato il suo modello di business diventando un *non-bank*,

[75] Melnyk, M., Kuchkin, M., & Blyznyukov, A. (2022). Commercial Banks: Traditional Banking Models Vs. FinTechs Solutions. *Financial Markets, Institutions and Risks.*

[76] CB Insights. (2019). 2019 Fintech Trends to Watch.

utilizzando la piattaforma Synapse per lanciare il proprio conto corrente e carta di debito. Questo cambiamento ha permesso all'azienda di generare entrate significative, passando da 19 milioni di dollari nel 2018 a 100 milioni di dollari nel 2021, con una valutazione complessiva di oltre 1 miliardo di dollari[77]. Allo stesso modo, Betterment, un'altra azienda digitale finanziaria con sede a New York, utilizza algoritmi per gestire 18 miliardi di dollari di asset dei clienti e ha recentemente lanciato un conto di risparmio ad alto rendimento, attirando oltre 1 miliardo di dollari in depositi in sole due settimane[78]. Secondo McKinsey & Company (2021), l'ascesa di queste nuove realtà rappresenta una minaccia significativa per le banche tradizionali, con la previsione che entro il 2025 possano sottrarre fino al 40% delle entrate complessive del settore[79].

Dunque, la ricerca parte dall'assunto che le banche commerciali e le startup tecnologiche finanziarie non possano continuare a operare in modo indipendente, e gli esperti sostengono che il futuro del settore finanziario potrebbe non vedere né una completa dominanza delle nuove società né un ritorno esclusivo ai modelli bancari tradizionali. Piuttosto, il modello più plausibile sarà una simbiosi strategica tra questi due attori. Questo modello prevede che le banche tradizionali sfruttino le innovazioni tecnologiche delle aziende digitali per ridurre i costi e migliorare l'efficienza operativa, mentre le imprese tecnologiche finanziarie beneficiano dell'esperienza e della regolamentazione offerta dalle banche. Un esempio di successo è la collaborazione tra il venture fund Life.SREDA e la banca russa PJSC Tatfondbank, che ha portato alla creazione di un acceleratore FinTech in Russia e ha generato oltre 37,3 milioni di dollari tra il 2017 e il 2020[80]. Questa sinergia va interpretata come un'opportunità per entrambi i settori di beneficiare delle rispettive competenze, creando un nuovo ecosistema finanziario più robusto.

Un altro aspetto chiave della simbiosi tra banche tradizionali e aziende finanziarie digitali è l'integrazione di tecnologie in via di sviluppo come le *Application Programming Interfaces* (API). L'impiego di API semplifica l'interoperabilità tra i servizi delle startup finanziarie e le infrastrutture delle banche in modo rapido e sicuro. Secondo Thomson Reuters (2021), l'adozione delle API favorisce la creazione di ecosistemi finanziari

[77] Dave Official Website, (2022).
[78] Betterment Official Website, (2022).
[79] McKinsey & Company. (2021). The 2021 McKinsey Global Payments Report.
[80] Life Sreda Singapore Official Website, (2022).

digitali, ovvero comunità di soggetti che interagiscono scambiandosi informazioni per ampliare la loro conoscenza, il cosiddetto "*know-how*". Questo approccio è particolarmente vantaggioso in un mercato sempre più digitalizzato, dove i clienti si aspettano operazioni rapide, sicure e personalizzate[81]. E, a proposito di clienti, uno dei maggiori benefici derivanti dalla cooperazione di questi operatori è rappresentato dai vantaggi per i consumatori. Grazie all'introduzione di tecnologie avanzate portate dalle FinTech, i consumatori possono ora accedere a una gamma più ampia di servizi finanziari e con maggiore facilità. Ad esempio, le piattaforme di *robo-advisory* consentono agli utenti di gestire i loro portafogli di investimento in modo autonomo e personalizzato, utilizzando algoritmi di intelligenza artificiale per ottimizzare i rendimenti. Queste piattaforme, che hanno già attratto miliardi di dollari di asset gestiti grazie alla loro capacità di offrire servizi di investimento automatici e convenienti, permettono ai consumatori di accedere a consulenze ottimizzate con maggiore trasparenza e controllo sui loro investimenti, consentendo di monitorare costantemente la performance del portafoglio e di apportare modifiche in tempo reale[82].

Un'altra innovazione dirompente che sta rivoluzionando il settore finanziario è l'utilizzo della blockchain. Questa tecnologia di registro distribuito migliora drasticamente la trasparenza, la sicurezza e l'efficienza delle transazioni, consentendo di tracciare e verificare ogni operazione in modo immutabile. Riduce la necessità di intermediari, aumentando al contempo la protezione e la salvaguardia delle operazioni finanziarie. Numerose FinTech stanno già sfruttando questa innovazione per fornire servizi inediti, come piattaforme di pagamento decentralizzate e smart contract che automatizzano determinate transazioni in modo sicuro e trasparente. Tuttavia, l'adozione su larga scala della blockchain comporta investimenti significativi in infrastrutture e richiede un aggiornamento delle normative per garantire la conformità e l'integrità del sistema. Oltre ai benefici operativi, la blockchain offre nuove opportunità per la democratizzazione dei mercati finanziari. Grazie a questa tecnologia, è possibile sviluppare piattaforme che permettono a un numero sempre maggiore di utenti di accedere a servizi e investimenti che, fino a poco tempo fa, erano riservati esclusivamente agli investitori istituzionali o ai clienti di fascia alta. Questo processo di democratizzazione potrebbe portare a una

[81] Reuters, T. (2021). Banking evolution: how to take on the challenges of FinTech.
[82] Betterment Official Website, (2022). op. cit.

maggiore inclusione finanziaria, ampliando l'accesso alle opportunità di investimento e creando nuove possibilità per consumatori in tutto il mondo.

In sintesi, il settore finanziario si sta evolvendo verso un modello ibrido, dove la collaborazione tra banche tradizionali e startup tecnologiche finanziarie diventa indispensabile per sopravvivere all'ambiente esterno. Il sostegno reciproco, unito all'integrazione di tecnologie emergenti come le blockchain, promette di rivoluzionare ulteriormente il mercato così come lo conosciamo oggi.

Capitolo 3: Caso d'Uso Esemplificativo

3.1 Principi Teorici alla base dell'algoritmo

I dati finanziari rappresentano una sfida complessa nel campo del Machine Learning a causa della loro natura altamente dinamica, non lineare e spesso rumorosa. La necessità di estrarre informazioni pertinenti e fare previsioni accurate richiede l'utilizzo di algoritmi avanzati e robusti. In questo contesto, il *Gradient Boosting Classifier* si distingue come uno strumento potente, combinando l'efficacia dei metodi ensemble con tecniche di ottimizzazione basate sul gradiente.

Nel campo del machine learning, esistono due categorie principali di algoritmi: supervisionati e non supervisionati. I modelli supervisionati sono addestrati su un dataset etichettato, nel quale ogni esempio include sia le caratteristiche di input (*features*) sia l'uscita desiderata (*label*). Durante l'addestramento, l'algoritmo apprende a mappare gli input alle uscite corrette, basandosi su esempi predefiniti, in modo da poter effettuare previsioni su dati nuovi e non visti. Gli algoritmi supervisionati sono particolarmente efficaci quando si dispone di una grande quantità di dati etichettati e si ha una chiara idea dell'obiettivo finale, come nel caso della classificazione o della regressione. Al contrario, i modelli non supervisionati lavorano su dati non etichettati. In questo contesto, l'algoritmo cerca di individuare strutture nascoste o relazioni intrinseche nei dati stessi, senza una guida esplicita su cosa cercare. Tecniche come il clustering o l'analisi delle componenti principali (PCA) rientrano in questa categoria, aiutando a scoprire raggruppamenti o ridurre la dimensionalità dei dati. Questi algoritmi sono utili quando si desidera esplorare pattern nascosti senza avere un obiettivo predittivo chiaro.

Nel contesto dell'analisi finanziaria, l'apprendimento supervisionato è spesso preferibile. Questo perché disponiamo di dati storici etichettati, come registrazioni di default creditizi o variazioni di prezzo delle azioni, che possono guidare il modello nell'apprendimento di pattern predittivi. Utilizzando un sistema come il *Gradient Boosting Classifier*, il modello viene addestrato su coppie di input-output già note, ovvero dati di input con esiti già

etichettati, al fine di fare previsioni più accurate su nuovi dati. Questo approccio è particolarmente vantaggioso perché ci permette di sfruttare la conoscenza preesistente per addestrare l'algoritmo e migliorare la sua capacità di generalizzazione. Inoltre, l'apprendimento supervisionato consente di valutare facilmente le prestazioni del modello attraverso metriche specifiche, facilitando l'ottimizzazione e l'interpretazione dei risultati. I metodi ensemble rappresentano una famiglia di algoritmi che combinano le previsioni di più modelli base, noti come *weak learners* (apprenditori deboli), al fine di costruire un modello complessivamente più robusto e accurato, definito *strong learner* (apprenditore forte). La logica alla base dei metodi ensemble risiede nel concetto che la combinazione di modelli deboli, ognuno dei quali ha una moderata capacità predittiva, possa ridurre gli errori individuali e migliorare la generalizzazione complessiva del modello finale. In questo modo, si può ottenere una previsione finale più affidabile e performante rispetto a quella ottenuta utilizzando un singolo modello complesso, poiché la diversità delle previsioni dei modelli deboli permette di compensare le loro limitazioni individuali.

Esistono diverse tecniche di ensemble, ognuna con caratteristiche e finalità specifiche:

- *Bagging* (*Bootstrap Aggregating*): È una tecnica che prevede la creazione di più modelli, ciascuno addestrato su un sottocampione casuale del dataset originale tramite campionamento con sostituzione (con ripetizione). Ogni modello viene addestrato su un sottocampione differente, e le previsioni finali sono ottenute aggregando le previsioni dei singoli modelli, ad esempio tramite media o votazione di maggioranza. Un esempio famoso di algoritmo basato su questa tecnica, nota come *bagging*, è il *Random Forest*, che utilizza alberi di decisione come *weak learners*. Questa strategia consente di ridurre la varianza del modello e migliorare la sua capacità di generalizzare su dati non visti.

- *Boosting*: In questa tecnica, i modelli vengono costruiti in sequenza, e ogni nuovo modello cerca di correggere gli errori dei modelli precedenti, concentrandosi sulle osservazioni difficili da prevedere. A differenza del *bagging*, in cui i modelli vengono addestrati in parallelo, il *boosting* costruisce modelli uno dopo l'altro, con ogni nuovo modello che assegna maggiore importanza ai dati che sono stati

mal classificati dai modelli precedenti, permettendo di migliorare progressivamente la capacità predittiva complessiva.

- *Stacking*: Questa tecnica prevede la combinazione di diversi modelli eterogenei (ad esempio, reti neurali, alberi decisionali, ecc.), addestrando un meta-modello che apprende come combinare al meglio le previsioni di ciascuno dei modelli base.

Il *Gradient Boosting* si categorizza anzitutto come modello ensemble di *boosting* che sfrutta la capacità degli apprenditori deboli per creare un modello finale più robusto attraverso un processo sequenziale.
Di seguito come si sviluppa il processo iterativo nel dettaglio:

1. Inizializzazione: Il modello inizia con una previsione semplice, ad esempio calcolando la media delle etichette di output nel caso di un problema di regressione o una stima di probabilità uniforme nel caso di classificazione.

2. Addestramento Sequenziale: Ad ogni iterazione, viene addestrato un nuovo *weak learner*, che viene adattato per correggere gli errori commessi dai modelli precedenti. Ciò avviene assegnando un peso maggiore alle istanze difficili da prevedere.

3. Aggiornamento dei Pesi: Dopo ogni iterazione, i pesi delle osservazioni vengono aggiornati in modo che le osservazioni erroneamente classificate ricevano un peso maggiore, mentre quelle classificate correttamente vedranno il loro peso ridotto. Questo garantisce che i modelli successivi si concentrino sugli esempi più difficili.

4. Combinazione Finale: Alla fine del processo, le previsioni dei modelli individuali sono combinate, generalmente tramite una somma ponderata, per produrre una previsione complessiva.

Questo processo continua fino al raggiungimento di un criterio di arresto, come un numero massimo di iterazioni o una soglia di errore minima.

La denominazione "*Gradient*" rappresenta una variante del *boosting* che applica i principi dell'ottimizzazione del gradiente per migliorare l'accuratezza del modello. A differenza di altri algoritmi di *boosting*, che si concentrano sull'aggiornamento dei pesi delle osservazioni, il *Gradient Boosting* si basa sulla minimizzazione della funzione di perdita in modo iterativo.
L'elemento distintivo del *Gradient Boosting* è l'utilizzo della discesa del gradiente per ottimizzare una funzione di perdita differenziabile. Ad ogni iterazione, l'algoritmo aggiunge un nuovo modello che approssima il gradiente negativo della funzione di perdita rispetto alle previsioni attuali. In questo modo, anziché ricalcolare i pesi delle istanze come avviene, ad esempio, in *AdaBoost*, il *Gradient Boosting* si concentra sulla riduzione diretta della discrepanza tra le previsioni e i valori reali, migliorando gradualmente le performance del modello.

Il processo del *Gradient Boosting* può essere formalizzato attraverso una serie di passaggi matematici che riflettono le fasi operative dell'algoritmo:

1. **Inizializzazione del Modello**

Il modello viene inizializzato con una previsione semplice, spesso utilizzando la media dei valori target nel caso di regressione o la probabilità a priori delle classi nel caso di classificazione. Matematicamente, si determina una stima iniziale costante che minimizza la funzione di perdita sul dataset:

$$F_0(x) = arg\min_{\gamma} = \sum_{i=1}^{n} L(y_i, \gamma)$$

Dove $L(y_i, \gamma)$ è la funzione di perdita che misura la discrepanza tra i valori target y_i e la previsione iniziale γ.

2. **Calcolo dei Residui**

Per ogni iterazione $m = 1, 2, \ldots, M$

In ogni iterazione successiva, l'obiettivo è ridurre gli errori residui, cioè la differenza tra le previsioni correnti e i valori target. Per questo, si calcola il gradiente negativo della funzione di perdita rispetto alle previsioni attuali, che rappresenta la direzione di massima riduzione dell'errore:

$$r_{i,m} = -\left(\frac{\partial L\left(y_i,\ F_{m-1}(x_i)\right)}{\partial F_{m-1}(x_i)}\right)$$

Dove $r_{i,m}$ è il gradiente negativo della fx. di perdita alla m-esima iterazione per l'osservazione i

Questo gradiente negativo funge da nuovo target per l'apprenditore debole successivo.

3. **Adattamento ai Residui**

Ogni nuovo *weak learner* è addestrato per predire i residui calcolati, ossia il gradiente negativo della funzione di perdita. Questo permette all'algoritmo di correggere progressivamente gli errori precedenti, concentrandosi sulle istanze più problematiche:

$$h_m(x) = arg\min_h \ = \sum_{i=1}^{n} \left(r_{i,m} - h(x_i)\right)^2$$

Dove $h_m(x)$ è il nuovo modello che meglio approssima il gradiente negativo, rappresenta il weak learner addestrato alla m-esima iterazione.

4. **Aggiornamento del Modello**

Le previsioni del nuovo modello sono combinate con quelle dei modelli precedenti per migliorare l'accuratezza complessiva. Questo viene realizzato sommando le previsioni attuali a una frazione del gradiente calcolato, controllata dal parametro di *learning rate* v:

$$F_m(x) = F_{m-1}(x) + vh_m(x)$$

Il *learning rate* v determina l'influenza del nuovo modello nell'aggiornamento totale, permettendo un apprendimento più graduale e controllato.

5. Funzione di Perdita

La funzione di perdita $L(y, F(x))$ quantifica la discrepanza tra le previsioni del modello $F(x)$ e i valori target y. Nel contesto del *Gradient Boosting* per la classificazione, una funzione di perdita comune è la deviance binomiale o *log-loss*, che misura la log-verosimiglianza negativa delle previsioni corrette:

$$L\big(y, F(x)\big) = -\left[y \log\big(p(x)\big) + (1-y)\log(1-p(x))\right]$$

Dove $p(x)$ è la probabilità prevista che l'osservazione appartenga alla classe positiva. La scelta della funzione di perdita guida la direzione in cui il modello deve muoversi per migliorare le previsioni, influenzando il calcolo del gradiente negativo.

Questo processo iterativo continua per un numero predefinito di iterazioni M o fino al raggiungimento di un criterio di convergenza. Ad ogni passo, il modello si avvicina sempre più alla minimizzazione della funzione di perdita, migliorando l'accuratezza delle previsioni.

Il *Gradient Boosting* può essere interpretato come una procedura di discesa del gradiente applicata in uno spazio funzionale. Ad ogni iterazione, il modello si muove nella direzione che riduce maggiormente la funzione di perdita, aggiornando le previsioni attraverso l'aggiunta di un modello che approssima il gradiente negativo.
Un elemento chiave di questo sistema è l'utilizzo di alberi di decisione come *weak learners*. Questi modelli sono particolarmente adatti per via della loro semplicità, flessibilità e capacità di catturare interazioni complesse tra variabili. Inoltre, sono in grado di gestire dati numerici, categoriali e mancanti senza richiedere un'eccessiva pre-elaborazione, modellando relazioni non lineari e selezionando automaticamente le caratteristiche più rilevanti, ignorando quelle meno informative. Questa capacità li rende ideali per una vasta gamma di problemi, essendo in grado di adattarsi a vari scenari

applicativi, come la classificazione, la regressione e la rilevazione di anomalie, dove l'obiettivo è prevedere una categoria o un valore numerico specifico.

Tuttavia, una delle sfide principali del *Gradient Boosting* è il rischio di *overfitting*, ossia la tendenza del modello a adattarsi troppo ai dati di training, inclusi rumori e difformità. Per contrastare questo fenomeno, il modello utilizza diverse tecniche di regolarizzazione, che aiutano a mantenere la capacità di generalizzazione del modello. Tra queste, un *learning rate* più basso permette un apprendimento più graduale e stabile, limitando l'impatto di ogni nuovo modello. Allo stesso modo, limitare il numero di iterazioni (ossia il numero di alberi) riduce la complessità complessiva del modello. Anche il controllo sulla profondità massima degli alberi aiuta a limitare l'eccessiva complessità, riducendo il rischio di catturare variazioni specifiche del set di training. Tecniche aggiuntive come il *subsampling* (addestrare ogni albero su un campione casuale del dataset) e *l'early stopping* (interrompere l'addestramento quando le prestazioni sul set di validazione iniziano a peggiorare) contribuiscono ulteriormente a migliorare la robustezza del modello.

Grazie a queste caratteristiche, il *Gradient Boosting* è noto per le sue elevate prestazioni in una vasta gamma di applicazioni. La sua capacità di ridurre sia il *bias* che la varianza permette di ottenere previsioni altamente accurate, rendendolo una scelta privilegiata in molte competizioni di machine learning e in contesti dove la precisione predittiva è fondamentale. La possibilità di personalizzare la funzione di perdita (grazie alla proprietà della scalabilità) consente inoltre di ottimizzare il modello per diversi obiettivi di business.

Questa versatilità rende il modello scelto particolarmente adatto anche all'analisi dei dati finanziari, un ambito caratterizzato da relazioni complesse e variabili non lineari. L'algoritmo è capace di cogliere queste relazioni senza richiedere trasformazioni o manipolazioni eccessive dei dati. Inoltre, le tecniche di regolarizzazione aiutano a gestire rumore e volatilità, caratteristiche comuni nei dati finanziari, mantenendo l'efficacia del modello senza farlo sovra-adattare a fluttuazioni temporanee.

In sintesi, il *Gradient Boosting Classifier* è uno strumento potente e versatile, che ai nostri fini si dimostra efficace nel fornire indicazioni sull'importanza delle variabili, una funzione particolarmente utile in contesti come quello finanziario, dove comprendere i driver delle previsioni è cruciale. Esempi applicativi includono la previsione del default

creditizio, in cui l'algoritmo classifica i clienti in base al rischio di insolvenza, la previsione della compatibilità tra cliente e prodotto finanziario, aiutando le istituzioni a consigliare prodotti personalizzati, e l'analisi dei movimenti di mercato, dove l'algoritmo può utilizzare indicatori storici e tecnici per prevedere fluttuazioni nei prezzi e volumi di scambio.

3.2 Scelta delle variabili e pre-processing dei dati

L'obiettivo principale di questo progetto è sviluppare un algoritmo capace di analizzare le caratteristiche sia dei prodotti finanziari che degli utenti, al fine di fornire previsioni accurate. Idealmente, per le società di consulenza finanziaria, FinTech e banche, sarebbe auspicabile attingere ai sistemi di condivisione dei dati finanziari menzionati nel Regolamento FIDA. Questi sistemi, una volta pienamente operativi, permetteranno l'accesso a dati finanziari di alta qualità, integrità e trasparenza, favorendo una migliore tutela dei diritti dei clienti e un'efficace collaborazione tra titolari e utenti dei dati. Tuttavia, al momento attuale, è ancora prematuro poter reperire dati finanziari corretti e completi da questi sistemi, poiché la loro implementazione è in fase di approvazione. Pertanto, per costruire il modello e disporre di una base dati su cui lavorare e addestrare l'algoritmo supervisionato GradientBoostingClassifier, si è deciso di generare casualmente questi dati finanziari (sia per i prodotti che per gli utenti) entro certi limiti, utilizzando Python in combinazione con il software Excel. Così facendo è stato possibile controllare le caratteristiche dei dati e garantire la qualità necessaria per l'addestramento del modello. Le variabili sono state scelte in accordo con il Regolamento FiDA, che pone l'accento su dati rilevanti per l'innovazione finanziaria e la minimizzazione dei rischi di esclusione. Per quanto riguarda la parte di script dedicata alla costruzione artificiale delle informazioni, sono state sfruttate librerie come *pandas* e *numpy*, con lo scopo di creare due dataset distinti: uno per gli utenti e uno per i prodotti finanziari, con particolare attenzione alla scelta delle caratteristiche (Figura 1).

Features degli Utenti	*Features dei Prodotti Finanziari*
Depositi TOT: tra 25.000 e 300.000 euro. **Quota da Investire:** % dei depositi totali. **Orizzonte Temporale:** tra 12 e 120 mesi. **Età:** Distribuzione normale tra 22 e 65 anni. **Debiti in Corso:** tra 0% e 60%. **Avversione al Rischio:** interi casuali da 1 a 10. **Obiettivo Finanziario:** • Accumulo di capitale • Generazione del reddito • Conservazione del capitale • Fondo di emergenza • Pianificazione della pensione **Esperienza di Investimento**: • Principiante • Intermedio • Avanzato	**Nome del Prodotto:** Etichettato come "Prodotto 1", "Prodotto 2", ecc. **Scadenza:** tra 12 e 120 mesi. **Finalità del Prodotto:** Scelta casualmente tra gli stessi obiettivi finanziari degli utenti. **Tipo di Prodotto:** • Azioni • Obbligazioni • Fondi di investimento • Fondi pensione • Titoli di Stato **Presenza Cedola/Dividendo:** • No • Fissa • Variabile **Requisiti Minimi di Investimento:** tra 100 e 1.000 euro. **Rendimento Atteso:** tra 1% e 7%. **Indice di Rischio:** interi casuali da 1 a 6.

Figura 1 – Features degli Utenti e dei Prodotti (Elaborazione personale)[83]

Prima di alimentare il dataset nel modello di machine learning, è stato necessario effettuare una fase di pre-processing per adattare i dati al formato richiesto dall'algoritmo in esame. Infatti, lo step successivo alla generazione dei dati è la codifica delle relative caratteristiche, al fine di fornire al programma un insieme di dati pulito e ordinato ma soprattutto adeguato su cui svolgere la fase di training.

[83] Elaborazione personale basata sui dati raccolti per il progetto.

Anzitutto occorre distinguere tra variabile numerica e categorica. Una variabile numerica è una variabile che assume valori numerici e consente operazioni matematiche come somme o medie. Può essere discreta (valori interi come il numero di figli) o continua (valori su un intervallo continuo come l'altezza). Una variabile categorica, invece, rappresenta categorie o gruppi distinti. I suoi valori sono etichette o nomi, come il colore degli occhi o il genere. Non ha senso eseguire operazioni matematiche su queste variabili, ma si possono contare le frequenze o analizzare le proporzioni tra le categorie.
Quindi per trasformare i dati da "*Raw*" a "*Normalized*" serve tenere in considerazione questi aspetti, ed in questo caso la codifica effettuata prevede:

- Variabili numeriche: è stata applicata una normalizzazione utilizzando lo strumento 'MinMaxScaler', che trasforma i dati scalando i valori in un intervallo compreso tra 0 e 1. Questo permette di uniformare le scale delle diverse variabili numeriche, facilitando l'apprendimento del modello.
- Variabili categoriche: è stata utilizzata la codifica one-hot. Questo metodo trasforma le categorie in vettori binari o nuove variabili binarie (*dummy variables*), in cui ogni categoria viene rappresentata da una colonna che assume valore 1 se l'osservazione appartiene a quella categoria, e 0 altrimenti. Ad esempio, la variabile "Obiettivo Finanziario" è stata trasformata in cinque vettori: "Obb_Accumulo_capitale", "Obb_Generazione_reddito", ecc. Se un utente ha come obiettivo "accumulo di capitale", la colonna corrispondente avrà valore 1 per quell'utente e 0 per gli altri obiettivi.

Nonostante le limitazioni dovute alla mancanza di dati reali dai sistemi di condivisione previsti dal Regolamento FIDA, la generazione di un dataset sintetico ha permesso di procedere nello sviluppo e nell'addestramento del modello. Il pre-processing dei dati, come volevasi dimostrare, è una fase fondamentale nel processo di Machine Learning, in quanto assicura che tutte le informazioni rilevanti siano correttamente interpretate dal modello. Questa fase risulta cruciale non solo per garantire l'uniformità dei dati, ad esempio rendendo le variabili confrontabili tramite la standardizzazione delle scale, ma anche per ottimizzare l'accuratezza del training. Dati ben pre-processati non solo migliorano l'efficacia dell'apprendimento del modello, ma contribuiscono in maniera significativa alla sua capacità di generalizzare su dati non visti, incrementando così le prestazioni complessive.

3.3 Struttura e componenti: applicazione del modello supervisionato (GBC) in Python

Dopo aver creato e codificato i dati relativi agli utenti e ai prodotti, il passo successivo è il training del modello. A tal fine, sono stati predisposti due nuovi fogli Excel: "Raccomandazioni" e "Train". Nel foglio "Raccomandazioni", le coppie utente-prodotto non codificate sono state inserite per una valutazione manuale della loro compatibilità. A ciascuna coppia è stato assegnato un valore di 1 se il prodotto risultava compatibile con l'utente, e 0 altrimenti.

Questa fase è stata sviluppata basandosi sulle conoscenze acquisite durante il percorso di studi e attraverso il confronto con docenti d'eccellenza, ma, data la natura sperimentale del progetto e la non disponibilità di dati autentici, le raccomandazioni sono state determinate manualmente attraverso un criterio personale e soggettivo.

Una volta completato questo step di preparazione per l'addestramento, 100 coppie utente-prodotto codificate, insieme alle relative raccomandazioni "umane", sono state inserite nel foglio "Train". Questo dataset servirà per addestrare il modello ad apprendere un criterio di classificazione, così che possa replicare, nel tempo, il giudizio di un consulente finanziario esperto.

Con questa doverosa precisazione, essendo questi i file principali da cui il modello attinge, possiamo passare alla spiegazione dello script di training del modello.

Il primo passo per costruire un programma è l'importazione delle librerie.

- **Pandas**: utilizzata per la manipolazione dei dati e il caricamento dei dataset Excel.
- **Scikit-learn** (sklearn): libreria fondamentale per l'implementazione di algoritmi di machine learning e strumenti per la modellazione.
  - 'train_test_split': per dividere il dataset in training set e test set.
  - 'GridSearchCV': per l'ottimizzazione degli iperparametri del modello.
  - 'StratifiedKFold': per la *cross-validation* stratificata.

  - ‘GradientBoostingClassifier’: algoritmo di apprendimento supervisionato utilizzato per la classificazione.
  - ‘classification_report’, ‘accuracy_score’, ‘roc_auc_score’: metriche per la valutazione delle prestazioni del modello.
- **Imblearn** (imbalanced-learn): libreria per gestire dataset sbilanciati.
  - ‘SMOTE’: tecnica di *oversampling* per bilanciare le classi minoritarie.
  - ‘Pipeline’: per creare una sequenza di trasformazioni dei dati e applicare il modello.
- **Joblib**: per salvare e caricare modelli addestrati.

```
import pandas as pd
from sklearn.model_selection import train_test_split, GridSearchCV, StratifiedKFold
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.metrics import classification_report, accuracy_score, roc_auc_score
from imblearn.over_sampling import SMOTE
from imblearn.pipeline import Pipeline as ImbPipeline
import joblib
```

Dopodiché si procede con il caricamento del dataset utilizzando ‘pd.read_excel’, specificando il percorso del file e il nome del foglio “TRAIN”. Questo dataset contiene le coppie utente-prodotto codificate e le raccomandazioni assegnate manualmente.
Successivamente si separano le caratteristiche (X) dall'obiettivo (y), dove l'obiettivo è la colonna “RACCOMANDAZIONE”:

- X = Contiene tutte le colonne del DataFrame ‘df’ tranne la colonna “RACCOMANDAZIONE”, che rappresenta le caratteristiche (features) utilizzate per l'addestramento del modello.
- y = Contiene la colonna “RACCOMANDAZIONE”, che è la variabile target o etichetta che il modello deve predire (1 per consigliato, 0 per non consigliato).

Proseguendo si suddivide il dataset in training set e test set utilizzando ‘train_test_split’. Si utilizza il parametro ‘stratify=y’ per mantenere la proporzione originale delle classi nelle due suddivisioni, essenziale quando si lavora con dataset sbilanciati.

```
# Load data
df = pd.read_excel('/Users/gabrielerizzo/Downloads/Aigab/dativ1.5.xlsx', sheet_name='TRAIN')
```

```
# Separare features and target
X = df.drop(['RACCOMANDAZIONE'], axis=1)
y = df['RACCOMANDAZIONE']

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)
```

Si è scelto di utilizzare la classe 'ImbPipeline' per creare una pipeline che consenta di concatenare diversi processi, gestendo sia *l'oversampling* che la classificazione in modo integrato. Questo approccio permette di combinare le trasformazioni dei dati e il modello in un unico flusso di lavoro, centralizzando così il core del progetto. È una scelta fondamentale per garantire che *l'oversampling* venga applicato esclusivamente al training set all'interno di ogni *fold* durante la *cross-validation*, evitando così il rischio di *leakage* di dati dal test set al modello.

La pipeline include due step (lista di *tuple* che definisce i passaggi),

1. 'oversampling': applica 'SMOTE' per bilanciare le classi nel training set.
   - Nel dataset, potrebbe esserci uno sbilanciamento tra le classi (ad esempio, più raccomandazioni negative che positive). Per affrontare questo problema, si utilizza SMOTE (*Synthetic Minority Over-sampling Technique*), una tecnica di *oversampling* che genera esempi sintetici della classe minoritaria. Funziona creando nuovi campioni sintetici lungo la linea che collega i campioni minoritari esistenti e i loro vicini più prossimi. Questo aiuta a bilanciare il dataset senza semplicemente duplicare i campioni esistenti.
   - 'random_state=42': per garantire la riproducibilità nella generazione dei campioni sintetici.

2. 'classifier': applica il modello di classificazione GradientBoostingClassifier.
   - Visto e considerato quanto già detto nel paragrafo 3.1; è un algoritmo di ensemble basato sul *boosting*, che costruisce il modello predittivo finale combinando una sequenza di *weak learners* (nel nostro caso alberi di decisione poco profondi). Ad ogni iterazione, l'algoritmo cerca di

minimizzare una funzione di perdita aggiustando gli errori commessi dai modelli precedenti (Friedman, 2001)[84].

- Poiché la raccomandazione viene affrontata come un problema di classificazione multi-classe binario, il modello scelto si dimostra efficace per la sua capacità di gestire dati complessi e non lineari.
- 'random_state=42': per garantire la riproducibilità nei risultati del modello.

```
pipeline = ImbPipeline(steps=[
    ('oversampling', SMOTE(random_state=42)),
    ('classifier', GradientBoostingClassifier(random_state=42))
])
```

Per ottimizzare le prestazioni del modello, si definisce una griglia di iperparametri da esplorare tramite 'GridSearchCV'. Si crea un dizionario 'param_grid' per specificare i valori da testare:

- 'classifier__n_estimators': numero di alberi (da 1 a 20.000).
- 'classifier__learning_rate': tasso di apprendimento (valori tra 0.01 e 0.5).
- 'classifier__max_depth': profondità massima degli alberi (valori tra 3 e 15).

L'uso del doppio underscore serve per accedere ai parametri dell'estimatore all'interno della pipeline. Mentre per valutare le combinazioni di iperparametri, si utilizza la *cross-validation* tramite 'StratifiedKFold'. Questo metodo divide il dataset in k *fold* stratificati, assicurando che la proporzione delle classi sia mantenuta in ciascuno di essi. In questo modo è garantito che ogni *fold* sia rappresentativo dell'intera popolazione, promuovendo una valutazione più affidabile su dataset sbilanciati.

In combinazione con 'StratifiedKFold' si utilizza 'GridSearchCV' per eseguire una ricerca esaustiva su tutte le possibili combinazioni di iperparametri definite in 'param_grid'. Il parametro 'scoring='roc_auc'' indica che si utilizzerà l'AUC della curva ROC come metrica per valutare le prestazioni del modello. Con 'grid_search.fit(X_train, y_train)' invece si addestra il modello utilizzando i dati di training e si cerca la migliore combinazione di iperparametri.

[84] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics.*

```
# Definizione della griglia di parametri per GridSearchCV
param_grid = {
    'classifier__n_estimators': [1, 20000],
    'classifier__learning_rate': [0.01, 0.1, 0.2, 0.3, 0.5],
    'classifier__max_depth': [3, 5, 7, 10, 15]
}

# Setup della cross-validation
cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

# Ricerca grid con cross-validation
grid_search = GridSearchCV(pipeline, param_grid, cv=cv, scoring='roc_auc', n_jobs=-1, verbose=3)
grid_search.fit(X_train, y_train)
```

In seguito, si verifica la selezione del modello migliore:

- ‘grid_search.best_estimator_’: recupera il modello con la combinazione di iperparametri che ha ottenuto le migliori prestazioni durante la ricerca su griglia.
- ‘best_model’: viene assegnato il modello migliore trovato, che verrà utilizzato per le predizioni successive.

Dopodiché vengono esaminate le prestazioni del modello:

- ‘y_pred’: contiene le predizioni effettuate dal modello sul test set ‘X_test’.
- ‘accuracy_score(y_test, y_pred)’: calcola l'accuratezza delle predizioni confrontandole con le etichette reali ‘y_test’.
- ‘roc_auc_score(y_test, y_pred)’: calcola l'Area Under the ROC Curve, una misura della capacità del modello di distinguere tra le classi.
- ‘classification_report(y_test, y_pred)’: fornisce un report dettagliato con precisione, recall, *F1-score* per ciascuna classe.

```
# Modello migliore
best_model = grid_search.best_estimator_

y_pred = best_model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print("ROC-AUC:", roc_auc_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

Per poter riutilizzare il modello in futuro senza doverlo riaddestrare, si salva l'oggetto modello utilizzando 'joblib.dump'. Viene fornita anche un'opzione interattiva per scegliere se salvare o meno il modello.

```
# Funzione per salvare il modello
def save_model(model, save_path='models/financial_product_advisor.joblib'):
    joblib.dump(model, save_path)
    print(f"Model saved at {save_path}.")

# Salvataggio interattivo del modello
if input("Do you want to save the model? (yes/no): ").lower() in ['yes', 'y']:
    save_model(best_model)
else:
    print("Model not saved.")
```

3.4 Valutazione delle prestazioni e Test del modello

Dopo aver addestrato il modello ma prima di provare ad effettuare una predizione, è essenziale valutare le sue prestazioni mediante metriche di performance appropriate, in grado di fornire informazioni dettagliate sulla capacità del modello di predire correttamente le raccomandazioni così da poter identificare agevolmente eventuali aree di miglioramento.

Le metriche utilizzate, brevemente accennate nel paragrafo precedente sono,

- Accuratezza (*Accuracy*): rappresenta la proporzione di predizioni corrette sul totale delle predizioni effettuate. Un'accuratezza elevata indica che il modello è in grado di classificare correttamente la maggior parte delle coppie utente-prodotto. Tuttavia, l'accuratezza da sola potrebbe non essere sufficiente, specialmente in presenza di classi sbilanciate.
    - Formula 'Accuracy' $= \frac{TP+TN}{TP+FP+FN+TN}$
- *Area Under the ROC Curve* (ROC-AUC): misura la capacità del modello di distinguere tra le classi positive (prodotti raccomandati) e negative (prodotti non raccomandati). Un ROC-AUC vicino a 1 indica un'eccellente capacità del modello

di discriminare tra le classi. Un valore di 0.5 suggerisce una classificazione casuale.

- Report di Classificazione: Fornisce metriche dettagliate per ciascuna classe, tra cui precisione, *recall* e *F1-score*.
  - Precisione: rappresenta la proporzione di predizioni positive corrette sul totale delle predizioni positive effettuate. Una precisione elevata indica che il modello commette pochi errori quando predice che un prodotto è raccomandabile.
    - Formula 'Precision' $= \frac{TP}{(TP+FP)}$
  - *Recall*: La proporzione di esempi positivi correttamente identificati sul totale degli esempi positivi reali. Un recall elevato significa che il modello riesce a identificare la maggior parte dei prodotti effettivamente raccomandabili.
    - Formula 'Recall' $= \frac{TP}{(TP+FN)}$
  - *F1-score*: non è altro che la media armonica tra Precisione e Recall. Fornisce un equilibrio tra precisione e recall, utile quando è necessario considerare sia i falsi positivi che i falsi negativi.
    - Formula 'F1-score' $= 2 * \left[\frac{(Precision*Recall)}{(Precision+Recall)}\right]$

Prima di procedere con la fase di "*testing*" del modello è necessario menzionare brevemente la matrice di confusione per capire meglio il perché di quelle formule e cosa sono gli elementi che le compongono. La matrice in questione non è altro che una tabella che consente di visualizzare le prestazioni del modello mostrando le predizioni corrette e incorrette suddivise per classe.

L'interpretazione della matrice si riduce a:

- Vero Positivi (*TP*): Numero di prodotti raccomandabili correttamente identificati.
- Falsi Positivi (*FP*): Numero di prodotti non raccomandabili che il modello ha erroneamente classificato come raccomandabili.
- Vero Negativi (*TN*): Numero di prodotti non raccomandabili correttamente identificati.

- Falsi Negativi (*FN*): Numero di prodotti raccomandabili che il modello ha erroneamente classificato come non raccomandabili.

Analizzando la matrice di confusione, è possibile individuare il tipo di errori che il modello commette più frequentemente.

Le metriche così ottenute indicano la capacità del modello di generalizzare su dati non visti durante l'addestramento. Complessivamente si può dire che:

- Se la precisione è alta ma il *recall* è basso, significa che il modello è molto preciso quando predice la classe positiva, ma potrebbe non identificare tutti gli esempi positivi. Se il *recall* è alto ma la precisione è bassa, il modello cattura la maggior parte degli esempi positivi, ma con molti falsi positivi. Un equilibrio tra le due metriche è spesso desiderabile, a seconda del contesto applicativo.
- Un'accuratezza elevata combinata con un alto ROC-AUC suggerisce che il modello è efficace nel distinguere tra prodotti raccomandabili e non raccomandabili.
- Un *F1-score* elevato indica un buon equilibrio tra precisione e recall.

È importante considerare il contesto applicativo per decidere quali metriche ottimizzare. Ad esempio, in un sistema di raccomandazione, potrebbe essere preferibile massimizzare il recall per garantire che tutti i prodotti adatti siano suggeriti all'utente, anche a costo di includere qualche prodotto meno pertinente.

Dopo aver valutato le prestazioni del modello, il passo successivo consiste nel testarlo su nuovi dati per verificare la sua capacità di fornire raccomandazioni utili in situazioni reali. A tal fine, è stato sviluppato un secondo script Python che permette di effettuare predizioni su nuovi esempi di coppie utente-prodotto. Di seguito, verrà presentato e spiegato dettagliatamente lo script utilizzato per questo scopo.

Come prima cosa si importano le librerie necessarie:

- **Joblib**: Utilizzata per caricare il modello precedentemente salvato. È una libreria efficiente per la serializzazione di oggetti Python, come i modelli di machine learning.

- **Pandas**: Per la manipolazione dei dati, in particolare per leggere il nuovo dataset contenente le coppie utente-prodotto da valutare.

Dopodiché occorre selezionare e caricare il modello addestrato. In questo caso, il modello si chiama 'v6acc80.joblib' ed è salvato nella cartella 'models'.
Si utilizza 'joblib.load()' per caricare il modello dal percorso specificato. Il modello viene così caricato in memoria e può essere utilizzato per effettuare predizioni.

Successivamente bisogna fornire al programma i nuovi dati da valutare, a tal proposito si utilizza 'pd.read_excel()' per leggere il foglio Excel denominato 'EXAMPLE', che contiene le nuove coppie utente-prodotto codificate su cui il modello effettuerà le predizioni[85].

```
import joblib
import pandas as pd

# Percorso al file del modello salvato
modello_path = 'models/v6acc80.joblib'

# Carica il modello
modello = joblib.load(modello_path)

df_dati = pd.read_excel('/Users/gabrielerizzo/Downloads/Aigab/dativ1.5.xlsx', sheet_name='EXAMPLE')
```

Ora che sono stati caricati sia il modello che il nuovo dataset, si procede con le predizioni. Per farlo si adopera il metodo '.predict()' del modello caricato per predire la classe (raccomandare o non raccomandare) per ciascuna coppia utente-prodotto nel dataset di esempio. Il risultato, 'predizioni', è un *array* contenente le etichette previste (in questo caso, 1 per prodotto consigliabile e 0 per prodotto non consigliabile).

```
# Effettua le predizioni
predizioni = modello.predict(df_dati)
```

[85] *Nota*: Nel file 'EXAMPLE' è compilata esclusivamente la prima riga (una coppia utente-prodotto), poiché l'intento è fornire un esempio del funzionamento e si presuppone che 'predizioni' contenga un singolo valore.

In sintesi, il metodo '.predict()' esegue una somma ponderata delle predizioni di tutti gli alberi nel modello, applica una funzione di attivazione per ottenere una probabilità e assegna la classe basandosi su questa probabilità. Questo processo permette di integrare le decisioni di molteplici alberi deboli per ottenere una predizione robusta e accurata, sfruttando la potenza dell'algoritmo di *Gradient Boosting* nel catturare pattern complessi nei dati[86].

3.5 Obiettivi e Opportunità dell'implementazione dello Use Case

Il caso d'uso illustrato rappresenta un esempio completo di come preparare, addestrare, valutare e testare un modello di machine learning per la classificazione. L'obiettivo principale era creare un sistema capace di emulare un consulente esperto nel suggerire ai clienti prodotti finanziari adeguati alle loro esigenze specifiche, analizzando le caratteristiche sia degli utenti che dei prodotti. A causa dell'indisponibilità di dati finanziari reali, si è generato artificialmente un dataset, ponendo le basi per future implementazioni con dati autentici. Con l'entrata in vigore del Regolamento FiDA, si prevede la creazione di sistemi di condivisione dei dati finanziari che consentiranno una più fluida circolazione dei dati supportata da integrità, sicurezza e trasparenza. Con la speranza che, in futuro, con l'implementazione del Regolamento, l'algoritmo possa operare su dati concreti e migliorare la capacità di individuare pattern nell'affinità tra utenti e prodotti.

È importante sottolineare che il modello sviluppato non è inteso come una soluzione completamente pronta, ma come un primo passo verso la costruzione di architetture informatiche più complesse, con un numero maggiore di parametri perfezionati. L'obiettivo a lungo termine è generare uno strumento che supporti, potenzi o migliori il sistema di gestione e ottimizzazione dei dati finanziari.

[86] *Nota*: Come già anticipato, questo esempio presuppone che 'predizioni' contenga un singolo valore. Se 'predizioni' contiene più valori (ad esempio, perché sono state valutate più coppie), è necessario iterare sull'array per interpretare ciascuna predizione.

Nel caso discusso nella tesi, il fine ultimo potrebbe essere interpretato come un sistema di guida dei clienti verso il prodotto più "raccomandabile" per loro, valutando e computando le caratteristiche specifiche di ciascun utente e prodotto. Questo va inteso sempre nell'ambito in cui, a seguito dell'approvazione del *Financial Data Access*, si stabilisca un clima di fiducia generale tra clienti, banche, tool e governo europeo, nonché una rigida tutela dei diritti dei clienti perlopiù legati alla privacy.
In un contesto simile dove l'accesso e la circolazione delle informazioni sono altamente sviluppati,
esistono due possibili scenari per l'implementazione di questo modello:

1. Strumento di supporto decisionale: il tool viene utilizzato prevalentemente come potenziamento e supporto al consulente umano, consentendo una prima selezione dei prodotti più idonei al profilo del cliente. In questo scenario, il consulente rimane centrale nel processo decisionale, avvalendosi delle raccomandazioni generate dal modello per offrire un servizio più personalizzato ed efficiente.
2. Automazione del processo di consulenza: il tool sostituisce completamente il consulente esperto in alcune fasi, velocizzando una serie di processi e rendendo la consulenza finanziaria più accessibile. Questo potrebbe essere particolarmente utile per clienti che preferiscono un'esperienza digitale o per istituzioni che mirano a ottimizzare le risorse.

Per le istituzioni finanziarie, l'adozione di un sistema di raccomandazione ben addestrato potrebbe portare numerosi vantaggi. In particolare, per le FinTech, data la loro natura digitale, un modello simile potrebbe influire positivamente su tutta l'organizzazione, andando ad impattare molto sulla cultura interna della società, tra i benefici si possono considerare:

- Per il settore FinTech e finanziario in generale:
  - Maggiore personalizzazione dei servizi.
  - Aumento dell'efficienza operativa.
  - Incremento delle vendite: proponendo prodotti più adatti alle esigenze dei clienti, si possono aumentare le opportunità di *cross-selling* e *up-selling*.
- Per i clienti:

- Decisioni finanziarie più informate: grazie a raccomandazioni basate su dati approfonditi e analisi avanzate.
- Risparmio di tempo: riducendo la necessità di consultazioni estese altrimenti svolte mediante un operatore umano.
- Riduzione dello stress finanziario: affidandosi a un sistema che tiene conto delle proprie esigenze e propensioni, in un contesto regolamentato e trasparente.

Volendo ampliare l'orizzonte delle opportunità, si può considerare un contesto applicativo più ampio, auspicabilmente accompagnato dal tema di credibilità prima menzionato. Cioè le potenzialità dovute all'implementazione di un algoritmo simile potrebbero cambiare radicalmente le abitudini di gestione del patrimonio sia per gli intermediari finanziari che per i consumatori. Attualmente, le banche si finanziano in gran parte attraverso capitale di credito, una soluzione che comporta obblighi finanziari rigidi verso i clienti. L'introduzione di modelli che raccomandano prodotti finanziari più adatti alle esigenze individuali potrebbe incentivare i clienti a considerare una gamma più ampia di strumenti, inclusi quelli legati al capitale di rischio. Questo potrebbe portare a una maggiore partecipazione attiva dei clienti nel rischio e nel successo delle banche, diventando essi stessi investitori e quindi soci. In tal modo, le banche avrebbero l'opportunità di ridurre la loro dipendenza dal capitale di credito, costruendo una struttura finanziaria più equilibrata e resiliente, soprattutto durante periodi di instabilità economica.

Il concetto chiave è che, in caso di congiunture economiche sfavorevoli, il rischio di incorrere in perdite aumenta proporzionalmente al peso del capitale di credito rispetto al capitale di rischio. Per questo motivo, sarebbe vantaggioso per gli intermediari finanziari orientarsi verso un modello in cui una parte significativa degli asset gestiti sia sottoscritta come capitale di rischio.

Per quanto riguarda i consumatori, un cambiamento nelle abitudini di investimento potrebbe manifestarsi nella scelta di titoli più diversificati. Anche gli investitori tradizionalmente più avversi al rischio potrebbero essere spinti a esplorare nuove opportunità di investimento, sapendo di poter fare affidamento su strumenti tecnologicamente avanzati e su un quadro normativo solido. Questo non solo

migliorerebbe la percezione pubblica del sistema finanziario, attirando nuovi clienti, ma consentirebbe anche di ottimizzare le relazioni con le classi di clienti esistenti.
Convincere i clienti a spostare una parte maggiore dei loro depositi verso strumenti a più alto rendimento, come il capitale di rischio, ridurrebbe l'incidenza dei depositi inattivi, che non generano ricavi significativi per la banca. Sebbene le banche debbano sempre mantenere riserve liquide sufficienti per far fronte alle richieste dei clienti, una porzione di queste riserve spesso rimane inutilizzata. Riducendo questo "costo opportunità" attraverso una gestione più dinamica e strategica degli investimenti, le banche potrebbero massimizzare i profitti, bilanciando al meglio le loro riserve liquide e gli strumenti ad alto rendimento.

Capitolo 4: Considerazioni Finali e Sviluppi Futuri

L'intelligenza artificiale sta trasformando profondamente il settore finanziario, offrendo nuove opportunità per ottimizzare l'efficienza, la precisione e la rapidità nei processi decisionali. Come dimostrato nel corso di questa tesi, l'impiego di modelli avanzati di machine learning può rivoluzionare l'analisi dei dati finanziari, rispondendo efficacemente alle sfide normative e tecnologiche degli ultimi anni. Il Regolamento FiDA, discusso nei capitoli introduttivi, rappresenta un progresso significativo verso una regolamentazione più inclusiva e trasparente, che facilita l'adozione di strumenti di intelligenza artificiale in grado di gestire i dati finanziari con maggiori livelli di sicurezza e affidabilità.

Il caso d'uso sviluppato ha evidenziato come l'implementazione di un modello supervisionato, come il *Gradient Boosting Classifier*, non solo migliori l'accuratezza delle previsioni finanziarie, ma consenta anche una maggiore personalizzazione dei servizi, adattandoli alle esigenze specifiche dei clienti. Questa sinergia tra innovazione tecnologica e conformità normativa sottolinea come l'IA non sia più una semplice opzione per il settore finanziario, bensì un pilastro strategico indispensabile per affrontare le complessità di un contesto sempre più dinamico e regolamentato.

Un esempio emblematico di questa tendenza è l'adozione su larga scala di soluzioni basate sull'intelligenza artificiale da parte di istituzioni finanziarie di primo piano come JPMorgan Chase. Recentemente, l'azienda ha introdotto modelli avanzati di linguaggio naturale (*LLM Suite*) per ottimizzare l'analisi dei dati, migliorare la gestione patrimoniale e supportare i dipendenti nella generazione di idee e sintesi. Jamie Dimon, CEO di JPMorgan, ha sottolineato come l'IA stia gradualmente trasformando ogni aspetto del settore finanziario, aprendo nuove opportunità ma anche sollevando sfide legate alla gestione della forza lavoro e alla sostenibilità (Tessa, 2024)[87].

Nel contesto di questa evoluzione, l'elaborato pone l'accento sull'importanza di bilanciare l'innovazione tecnologica con una solida cultura aziendale, garantendo trasparenza e

[87] Tessa, M. (2024, 26 Luglio). *Jp Morgan sposa l'IA, svolgerà il ruolo di analista*. Wall Street Italia.

responsabilità. Questo è essenziale per mitigare i rischi connessi a *bias*, sicurezza e privacy. Le prospettive future indicano una crescente integrazione dei modelli di IA nei processi decisionali finanziari, favorendo una maggiore sinergia tra tecnologia e regolamentazione, e spingendo il settore verso una finanza sempre più orientata ai dati e all'innovazione.

In conclusione, l'adozione di tecnologie basate sull'intelligenza artificiale non rappresenta soltanto un vantaggio competitivo per le istituzioni finanziarie, ma costituisce una necessità strategica per navigare in un panorama normativo in rapida evoluzione, contribuendo a migliorare l'efficienza, l'inclusività e l'innovazione nel settore finanziario.

Riferimenti Bibliografici

Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction, judgment, and complexity: a theory of decision-making and artificial intelligence. In *The economics of artificial intelligence: An agenda* (pp. 89-110). University of Chicago Press. NBER. Recuperato da: link

Betterment Official Website. (2022). Recuperato da: link

Cannone, V. (2021, Febbraio 4). Top data management challenges faced by financial firms. *The Golden Source*. Recuperato da: link

CB Insights. (2019). *2019 Fintech Trends to Watch*. Recuperato da: link

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management, 48*, 63-71. Recuperato da: link

Dave Official Website. (2022). Recuperato da: link

Feng, S. (2024). Integrating artificial intelligence in financial services: Enhancements, applications, and future directions. *Semantic Scholar*. Disponibile da: link

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189-1232. Recuperato da: link

Herzberg, B. (2024, Gennaio 16). 5 biggest challenges of data security in the financial service industry. *Due.com*. Recuperato da: link

Hussain, E. (2024, Settembre 2). 12 challenges in financial data migration and how IT leaders tackle them. *Data Ladder*. Recuperato da: link

InterSystems. (2022). *The top data and technology challenges in financial services*. InterSystems. Recuperato da: link

Intone. (2024, Aprile 30). Challenges and solutions in implementing financial data quality management frameworks. *Intone*. Recuperato da: link

Judijanto, L., Asfahani, A., Bakri, A. A., Susanto, E., & Kulsum, U. (2022). AI-Supported Management through Leveraging Artificial Intelligence for Effective Decision Making. *Journal of Artificial Intelligence and Development*, *1*(1), 59-68. Recuperato da: link

Life Sreda Singapore Official Website. (2022). Recuperato da: link

McKinsey & Company. (2021, Ottobre). *The 2021 McKinsey Global Payments Report*. Recuperato da: link

Melnyk, M., Kuchkin, M., & Blyznyukov, A. (2022). Commercial Banks: Traditional Banking Models Vs. FinTechs Solutions. *Financial Markets, Institutions and Risks*, *6*(2), 122-129. Recuperato da: link

Nasir, W., & Gasmi, S. (2024, Agosto). Revolutionizing Risk Assessment in Finance through AI-Driven Decision-Making. *ResearchGate*. Recuperato da: link

Nanda, A., Mohapatra, N., Rautela, R., & Acharya, D. (2024, Agosto 15). Role of Artificial Intelligence (AI) Driven Business Excellence for Effective Corporate Operations and Competitive Advantage: An Empirical Study. Recuperato da: link

Reuters, T. (2021). Banking evolution: how to take on the challenges of FinTech. Recuperato da: link

Stone, R., & Suja, S. (2024, Settembre). Leveraging Predictive Analytics and Big Data in Machine Learning for Real-Time Financial Decision-Making. *ResearchGate*. Recuperato da: link

Tessa, M. (2024, 26 Luglio). Jp Morgan sposa l'IA, svolgerà il ruolo di analista. *Wall Street Italia*. Recuperato da: link

Direttive:
Direttiva (UE) 2016/2341 del Parlamento europeo e del Consiglio, del 14 dicembre 2016, relativa alle attività e alla supervisione degli enti pensionistici aziendali o professionali. *Gazzetta ufficiale dell'Unione europea*. Recuperato da: link

Direttiva (UE) 2016/97 del Parlamento europeo e del Consiglio, del 20 gennaio 2016, sulla distribuzione assicurativa. *Gazzetta ufficiale dell'Unione europea*. Recuperato da: link

Direttiva (UE) 2015/2366 del Parlamento europeo e del Consiglio, del 25 novembre 2015, relativa ai servizi di pagamento nel mercato interno. *Gazzetta ufficiale dell'Unione europea*. Recuperato da: link

Direttiva 2014/65/UE del Parlamento europeo e del Consiglio, del 15 maggio 2014, relativa ai mercati degli strumenti finanziari. *Gazzetta ufficiale dell'Unione europea*. Recuperato da: link

Direttiva 2009/138/CE del Parlamento europeo e del Consiglio, del 25 novembre 2009, relativa all'accesso ed esercizio delle attività di assicurazione e riassicurazione (Solvibilità II). *Gazzetta ufficiale dell'Unione europea*. Recuperato da: link

Regolamenti:
Proposta di Regolamento del Parlamento europeo e del Consiglio relativo a un quadro per l'accesso ai dati finanziari e che modifica i regolamenti (UE) n. 1093/2010, (UE) n.

1094/2010, (UE) n. 1095/2010 e (UE) 2022/2554 - COM/2023/360 final. Recuperato da: link

Regolamento (UE) 2022/2554 del Parlamento europeo e del Consiglio, del 14 dicembre 2022, relativo alla resilienza operativa digitale per il settore finanziario (DORA). *Gazzetta ufficiale dell'Unione europea*. Recuperato da: link

Regolamento (UE) 2019/1238 del Parlamento europeo e del Consiglio, del 20 giugno 2019, relativo a un prodotto pensionistico individuale paneuropeo (PEPP). *Gazzetta ufficiale dell'Unione europea*. Recuperato da: link

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 e alla libera circolazione di tali dati (GDPR). *Gazzetta ufficiale dell'Unione europea*. Recuperato da: link

Altro:
Carta dei diritti fondamentali dell'Unione europea, 2000/C 364/01. *Gazzetta ufficiale dell'Unione europea*. Recuperato da: link

Accordo interistituzionale "Legiferare meglio", del 13 aprile 2016. *Gazzetta ufficiale dell'Unione europea*. Recuperato da: link