



UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA “GIORGIO FUÀ”

Corso di Laurea triennale in ECONOMIA E COMMERCIO

**L'EVOLUZIONE DEL CREDITO: L'INTELLIGENZA
ARTIFICIALE NEI SISTEMI DI CREDIT SCORING
DELLE BANCHE**

**CREDIT'S EVOLUTION: ARTIFICIAL
INTELLIGENCE IN BANKS' CREDIT SCORING
SYSTEMS**

Relatore:
Prof.ssa. Mazzoli Camilla

Rapporto Finale di:
Paoletti Nicolò

Anno Accademico 2022/2023

INDICE

1	INTRODUZIONE	1
2	L'INTELLIGENZA ARTIFICIALE	2
2.1	<i>L'INTELLIGENZA ARTIFICIALE: DEFINIZIONI E ORIGINI</i>	2
2.2	<i>INTELLIGENZA ARTIFICIALE FORTE E DEBOLE</i>	6
2.3	<i>IL MACHINE LEARNING E I DIVERSI APPRENDIMENTI</i>	7
2.4	<i>L'IA SPIEGABILE</i>	10
2.5	<i>AMBITI APPLICATIVI E FUTURO DELL'IA</i>	11
3	IL RISCHIO DI CREDITO E IL CREDIT SCORING	14
3.1	<i>IL RISCHIO DI CREDITO</i>	14
3.2	<i>GLI ACCORDI DI BASILEA</i>	17
3.2.1	<i>Basilea I</i>	18
3.2.2	<i>Basilea II</i>	20
3.2.3	<i>Basilea III</i>	21
3.3	<i>MODELLI DI RATING E DI SCORING NELLA MISURAZIONE DEL RISCHIO DI CREDITO</i>	23
4	L'UTILIZZO DI TECNOLOGIE IA NEI MODELLI IRB	25
4.1	<i>L'IA NELLA MISURAZIONE DEL RISCHIO DI CREDITO</i>	25
4.2	<i>POTENZIALI BENEFICI</i>	26

4.3	<i>RISULTATI DISTORTI</i>	27
4.4	<i>UNA QUESTIONE ETICA TRA PRIVACY E SPIEGABILITÀ</i>	29
4.5	<i>QUADRO NORMATIVO E VISIONE DELLE ISTITUZIONI INTERNAZIONALI</i>	32
5	CONCLUSIONI	37
	SITOGRAFIA	39
	BIBLIOGRAFIA	39

1 INTRODUZIONE

Il fenomeno dell'intelligenza artificiale non è mai stato così discusso come ora. Gli sviluppi sulla guida autonoma di Tesla, l'avvento improvviso di ChatGPT, sono solo alcuni degli esempi che hanno favorito la diffusione esplosiva tra il pubblico di queste nuove tecnologie. La curiosità nel comprendere i rischi e i pericoli dell'utilizzo delle tecniche di intelligenza artificiale oltre che le potenzialità di tali tecnologie con una mole di dati mai così ampia, mi hanno portato ad approfondire i possibili impatti sul settore finanziario, in particolare nella valutazione del merito di credito. L'obiettivo, infatti, è stato quello di evidenziare i benefici dell'utilizzo di tecniche IA nella misurazione dello scoring creditizio delle banche in un quadro che le vede districarsi tra accuratezza ed etica. Il lavoro è stato articolato in tre capitoli: nel primo capitolo si cerca di fornire un quadro generale sul fenomeno dell'intelligenza artificiale, sottolineando le sue origini e i percorsi di sviluppo più intensi. Nel secondo capitolo si è invece approfondito l'ambiente all'interno del quale il lavoro si concentra, il rischio di credito sopportato dalle banche, con una lente sulla misurazione del merito creditizio. Il terzo capitolo rappresenta il centro del lavoro, in cui verranno analizzati rischi e benefici dell'introduzione di tecniche IA nel contesto sopra indicato, con un'enfasi specifica su questioni etiche e normative, che rappresenteranno il fulcro delle future discussioni e dei successivi sviluppi delle tecnologie IA nell'ambito finanziario e non solo.

2 L'INTELLIGENZA ARTIFICIALE

2.1 *L'intelligenza artificiale: definizioni e origini*

L'intelligenza artificiale potrebbe essere generalmente intesa come l'abilità di una macchina di mostrarsi pensante o, in senso strettamente scientifico, la disciplina che studia la possibilità che elaboratori e sistemi informatici possano simulare le caratteristiche tipicamente umane del ragionamento e dell'apprendimento. Una definizione alternativa è data dall'"Enciclopedia della Scienza e dalla Tecnica" Treccani: "L'intelligenza artificiale studia i fondamenti teorici, le metodologie e le tecniche che consentono di progettare sistemi hardware e sistemi di programmi software atti a fornire all'elaboratore elettronico prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana". Un'ulteriore definizione, più istituzionale, è proposta invece dalla Commissione Europea che dice: "Per intelligenza artificiale si intendono quei sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere obiettivi specifici".

La nascita ufficiale dell'Intelligenza Artificiale come vera e propria disciplina scientifica risale all'estate 1956 al Dartmouth College di Hanover nel New Hampshire dove in occasione del seminario "*Dartmouth Summer Research Project on Artificial Intelligence*" organizzato dallo statunitense John McCarthy, figura di

spicco nel campo dell'informatica del tempo, si riunirono i maggiori studiosi e ricercatori dell'ambito dei sistemi intelligenti allo scopo di confrontare i propri studi e progressi sul tema. Nonostante quest'evento segni in maniera formale quella che è la nascita dell'IA (Intelligenza Artificiale), le sue autentiche origini, in termini di vera e propria ricerca su macchine pensanti, possono essere ricondotte a circa qualche decennio prima grazie al contributo della cibernetica e dello sviluppo di primi calcolatori elettronici, considerati da tutti come l'unico strumento che potesse avere la possibilità di riprodurre il ragionamento umano. Uno dei primi e importanti, se non fondamentali, contributi che poi hanno portato alla nascita dell'IA fu di Warren McCulloch e Walter Pitts nel 1943. I due nel loro studio proposero il primo modello di rete neurale, che ancora oggi è considerato uno standard di riferimento. La vera e propria svolta però arrivò grazie ad uno dei padri fondatori dell'informatica, A. M. Turing, che nel 1950 pubblica un articolo, *"Computer Machinery and Intelligence"* nel quale si chiede quando e se una macchina può considerarsi pensante. La soluzione da lui proposta è un test che prende il suo nome (Test di Turing) e consiste in un gioco a tre partecipanti: un uomo *A*, una donna *B* e una terza persona *C*. Quest'ultimo viene tenuto separato dagli altri due e, tramite una serie di domande, deve stabilire qual è l'uomo e quale la donna. Dal canto loro, anche *A* e *B* hanno dei compiti: *A* deve ingannare *C* mentre *B* deve aiutarlo. Le risposte alle domande di *C* devono essere trasmesse in maniera tale che *C* non disponga di alcun indizio (Turing consiglia una stampa a

video delle risposte). Il test si basa sul presupposto che *A* venga sostituito da una macchina. In tal caso, se *C* non si accorgesse di nulla, la macchina dovrebbe essere considerata intelligente, dal momento che sarebbe indistinguibile da un essere umano. La macchina cioè dovrebbe essere considerata come dotata di una “intelligenza” pari a quella dell’uomo. Questo criterio viene ritenuto tutt’oggi alla base dell’intelligenza artificiale (Treccani). Il decennio che porta agli anni ’60 fu caratterizzato da grande entusiasmo ed enormi aspettative sugli studi e programmi proposti. È il decennio in cui venne presentato il GPS (General Problem Solver, 1957) di A. Newell e H. Simon, un programma ideato per risolvere problemi universali estendendo gli ambiti di applicazione degli elaboratori al di fuori di quello puramente aritmetico. Sempre in questo periodo, il padre dell’IA John McCarthy, diede alla luce il LISP (List Processing), rimasto il linguaggio di riferimento per l’AI nei decenni successivi, ispirando anche la nascita di altri linguaggi. Fu però un periodo in cui alle enormi aspettative si contrapposero grandi insuccessi per via del fatto che le macchine e i programmi creati non erano ancora in grado di risolvere problemi di dimensioni reali. A partire dagli anni ’70 visto il periodo di insuccessi, si spostò la concentrazione dei ricercatori, anziché sulla costruzione di un sistema con intelligenza generale e assoluta ispirata a quella umana, sullo sviluppo di sistemi capaci di risolvere dei problemi in aree limitate, che avessero cioè una conoscenza limitata in uno specifico campo. Furono realizzati i primi sistemi esperti. In un sistema esperto, il complesso delle conoscenze presenti

in un esperto umano di uno specifico settore è codificato e rappresentato in una forma che l'elaboratore utilizza per risolvere problemi simili a quelli affrontati dall'esperto umano (M. Somalvico, F. Amigoni, V. Schiaffonati). In seguito, si evidenziarono enormi limiti nei sistemi esperti soprattutto nella specificazione chiara del dominio e nella difficoltà del sistema di capire se il problema rientrasse nel proprio dominio di appartenenza. Nella seconda metà degli anni '80 si registrò da parte di numerosi gruppi di ricerca, la ripresa nello studio delle reti neurali, abbandonata durante gli anni '60 dopo numerosi articoli che le definivano inadeguate per poter risolvere problemi reali. Questa riscoperta portò a definire nuovamente l'algoritmo di 'back-propagation'¹ che consentiva alla rete stessa di poter apprendere dai suoi errori. Con l'avvento del nuovo millennio e la disponibilità di una vastissima quantità di dati, i sistemi di intelligenza artificiale hanno subito un'importante accelerazione nel loro sviluppo. In tal senso, non è più necessario, per i sistemi basati sulla conoscenza, dover creare manualmente un'esperienza al programma, ma è sufficiente legarsi ad un database adeguatamente ampio. Ad oggi gli studi avanzano prevalentemente verso una direzione più "realistica" del termine intelligenza artificiale, andando a concentrare gli sforzi, non tanto su quella che potrebbe essere una macchina a tutti gli effetti pensante e in grado di poter ragionare autonomamente, ma soprattutto sullo sviluppo di sistemi

¹ È un algoritmo di apprendimento che confronta il valore in uscita del sistema con il valore desiderato. Sulla base della differenza così calcolata, l'algoritmo modifica i pesi sinaptici della rete neurale, facendo convergere progressivamente il set dei valori di uscita verso quelli desiderati.

che possano essere di supporto al lavoro umano. La via principe, per ora, è costruire un'intelligenza che non vada a sostituire o riprodurre quella dell'uomo, bensì che lo supporti e lo aiuti nella risoluzione di problemi che richiedono delle capacità di calcolo ed elaborazione superiori.

2.2 *Intelligenza artificiale forte e debole*

Tutti quelli che sono gli studi e le ricerche sull'intelligenza artificiale sono riconducibili a due principali classificazioni, i livelli dell'IA: IA forte e IA debole. Per intelligenza artificiale debole (o stretta) si intendono quei sistemi informatici costruiti per svolgere compiti specifici senza però riuscire a replicare quelle particolari capacità intellettuali dell'uomo. È il tipo di IA che viene utilizzata quotidianamente e soprattutto che realisticamente esiste.

L'intelligenza artificiale forte (o generale) potremmo definirla puramente teorica poiché tutt'ora non esiste. Questi sistemi non si limitano alla semplice emulazione dei ragionamenti umani, ma hanno una loro coscienza. In tal caso, non è più necessario che l'uomo programmi e addestri la macchina. Da ciò si può evincere che a supporto di questo livello dell'IA ci sono i sistemi esperti precedentemente citati.

2.3 *Il machine learning e i diversi apprendimenti*

La distinzione tra IA forte e debole ci è molto d'aiuto nella comprensione di quelli che sono i diversi "sottocampi", se così possiamo definirli, dell'intelligenza artificiale. Uno dei parenti più stretti dell'intelligenza artificiale è il machine learning (apprendimento automatico). Il machine learning può essere definito come quella branca dell'IA che, utilizzando metodi statistici, permette a sistemi e macchine di apprendere dai dati che elaborano, consentendo agli stessi di migliorare le proprie capacità e performance nel tempo, così da arrivare ad eseguire un compito senza venir precedentemente programmati per farlo. Gli studi effettuati nel campo del ML (Machine Learning) fino ad oggi, hanno portato ad avere differenti tipologie di apprendimento, che differiscono l'un l'altro soltanto per le caratteristiche degli algoritmi e i requisiti richiesti (ad es. l'ampiezza dei dati necessari per poter addestrare in maniera efficace la macchina). In base a queste differenziazioni e ai possibili interventi umani, si possono definire quattro approcci di apprendimento automatico: supervisionato, non supervisionato, semi-supervisionato e con rinforzo.

Nell'apprendimento supervisionato all'algoritmo vengono forniti preventivamente sia dati in input che i rispettivi dati output da un essere umano, al fine di identificare una regola generale che colleghi i dati in ingresso con i dati obiettivo, in modo da addestrare l'algoritmo a risolvere problemi specifici e utilizzarlo per casi simili.

L'apprendimento non supervisionato invece, è un approccio del ML in cui all'algoritmo viene fornito soltanto un dataset senza alcuna indicazione del risultato atteso. Sarà compito della macchina classificare i dati in suo possesso, organizzarli e apprendere da questi. In tal caso, la macchina dovrà catalogare i dati in maniera intelligente così da riuscire a comprendere quali siano i risultati da raggiungere a seconda delle differenti situazioni che possono presentarsi.

L'approccio semi-supervisionato rappresenta una sorta di compromesso rispetto agli approcci precedentemente elencati. In questo particolare approccio vengono utilizzati dei dataset etichettati più piccoli, come per il caso con supervisione, che guidano l'algoritmo nella comprensione di informazioni di un dataset non etichettato di maggior dimensione, come accade nell'approccio senza supervisione.

L'approccio con rinforzo è sicuramente il più complesso poiché prevede che la macchina sia dotata di strumenti di supporto in grado di aiutarla a comprendere le caratteristiche dell'ambiente circostante, cosicché ottenere i dati in input. Il funzionamento del sistema ruota attorno alla massimizzazione di una funzione profitto, in tal senso la macchina apprende le informazioni durante l'utilizzo, tramite prove e errori che la aiutano ad adattarsi nella maniera più ottima all'ambiente intorno ad essa (ad es. le auto senza pilota supportate da numerosi sensori e telecamere).

Una sottocategoria del machine learning è il deep learning, letteralmente apprendimento approfondito. Il suo funzionamento si basa sulla creazione di

modelli di apprendimento su più livelli (cosiddette reti neurali). Ad ogni livello della rete corrisponde una fase dell'apprendimento che diventa via via sempre più complessa per ogni livello che si raggiunge. Secondo una definizione scientifica l'azione dei sistemi deep learning è quella di apprendere dei dati che non siano forniti dall'uomo. Un algoritmo di deep learning utilizza la propria rete neurale, ispirata alla struttura del cervello umano, per verificare l'accuratezza dei risultati raggiunti e andare poi ad imparare da essi, grazie l'utilizzo di algoritmi prevalentemente di calcolo statistico.



Figura 1. Differenza tra sistemi di Machine Learning e Deep Learning

Un'ulteriore sottocategoria del machine learning è l'ensemble learning (apprendimento d'insieme), famiglia di algoritmi applicabili trasversalmente che utilizza molteplici modelli contemporaneamente allo scopo di ottenere delle performance predittive migliori rispetto a quelle dei modelli presi singolarmente. La previsione finale sarà ottenuta come la previsione media di ciascun modello che compone il sistema. Esistono diversi metodi di ensemble, tra i più importanti ricordiamo il *Bagging*, in cui tutti i modelli utilizzati nel sistema hanno lo stesso peso ed importanza; e il *Boosting*, ogni modello in questo caso ha un certo peso

definito in base all'errore di accuratezza che ciascuno di essi commette durante l'apprendimento.

2.4 L'IA spiegabile

L'IA spiegabile (eXplainable Artificial Intelligence, XAI) è una disciplina piuttosto recente, nell'ambito del machine learning, che si pone l'obiettivo di far chiarezza in merito a quello che succede all'interno degli algoritmi IA ovvero comprendere le motivazioni alla base di un determinato risultato raggiunto dagli algoritmi. Per avere un quadro più chiaro sulla XAI, la potremmo definire come un insieme di strumenti volti a integrare il risultato dell'algoritmo attraverso una serie di spiegazioni che possano giustificare le decisioni prese dal modello. Si cerca in tal senso, di interrompere o meglio, limitare, quel legame tra il termine 'black box' e i modelli di IA che trasmettono ai normali cittadini sfiducia e timore verso queste tecnologie, vista la mancata comprensibilità di come queste agiscano (si pensi alle auto a guida autonoma). È necessario evidenziare però, che questa spiegabilità dei modelli deve essere resa concreta e fruibile. Da un punto di vista concettuale difatti, la XAI deve garantire interpretabilità e spiegabilità. L'interpretabilità dei modelli è intesa come la possibilità di tracciare quantitativamente e in maniera logica la previsione o la scelta effettuata dal modello. La spiegabilità invece, sposta l'attenzione sul piano qualitativo, cioè è necessario essere in grado di poter

formulare giustificazioni su come il modello sia arrivato a compiere determinate scelte o previsioni.

2.5 *Ambiti applicativi e futuro dell'IA.*

Al giorno d'oggi, i sistemi basati su tecnologie IA si sono ampiamente diffusi, e senza esserne consapevoli, vengono utilizzati praticamente nella vita di tutti i giorni. Per chiarire nel migliore dei modi tale diffusione si può far riferimento alla classificazione definita dall'Osservatorio Digital Innovation del Politecnico di Milano, che ha individuato ben otto differenti tipologie di applicazioni:

- *Intelligent Data Processing*, si tratta di algoritmi utilizzati per estrapolare informazioni da dataset e compiere azioni in conseguenza. Molto utilizzati nell'analisi dei dati per fornire previsioni e nel rilevamento di frodi;
- *Virtual Assistant/Chatbot*, sono sistemi che utilizzano tecnologie NLP² in modo da erogare dei servizi attraverso delle interazioni testuali o vocali. Sono spesso utilizzati per offrire assistenza all'interno dei siti web delle aziende;
- *Reccomendation*, sono tecnologie che integrano i sistemi di vendita con applicazioni che guidano il consumatore verso la finalizzazione di un

² Si intendono algoritmi di intelligenza artificiale in grado di analizzare, rappresentare e quindi comprendere il linguaggio naturale (Osservatorio Digital Innovation).

acquisto. Questi sistemi riescono ad influenzare il processo decisionale del cliente.

- *Image Processing*, sono tecnologie che permettono il riconoscimento automatico di oggetti, persone e animali. Supportano molteplici applicazioni soprattutto nell'ambito della video sorveglianza o nel controllo qualità industriale.
- *Autonomous Vehicle*, tecnologie IA che supportano i sistemi di guida autonoma dei veicoli, oltre che la navigazione marittima e aerea.
- *Intelligent Object*, tali algoritmi consentono ai sistemi e oggetti in cui vengono integrati, di prendere decisioni e compiere azioni in base alle condizioni dell'ambiente circostante, attraverso l'associazione con sensori intelligenti.
- *Language Processing*, algoritmi integrati ad applicazioni basate sulla comprensione del testo e della comunicazione verbale.
- *Autonomous Robot*, tecnologie IA che vengono integrate sui sistemi di funzionamento dei robot, consentendo ad essi di poter compiere azioni senza la necessità che un utente impartisca comandi, grazie alla loro capacità di interagire con l'ambiente.

Esempi pratici di queste tecnologie nella quotidiano li possiamo ritrovare negli assistenti vocali (come Siri, Amazon Alexa o Google Assistant) o nel colosso dell'intrattenimento Netflix, il quale utilizza l'IA per comprendere quali siano i

gusti degli abbonati in base al loro comportamento sulla piattaforma, allo scopo di realizzare previsioni sui trend. Uno degli ambiti più “colpiti” dalla diffusione delle tecnologie IA è sicuramente quello sanitario, soprattutto per quel che riguarda l’assistenza chirurgica e la diagnostica dei dati, grazie all’utilizzo di sistemi robotici e informatici. Non si può inoltre, non citare la tendenza del momento, ChatGPT, di OpenAI. Un Chatbot basato sull’IA con cui si può tranquillamente conversare e chiedere qualsiasi cosa, ottenendo una risposta in forma scritta a video.

Tutt’oggi l’obiettivo di riuscire ad utilizzare l’intelligenza artificiale nella quotidianità è stato ampiamente raggiunto, riuscendo a rendere la disciplina una solida realtà e non più soltanto un’ipotesi. D’altro canto, il futuro dell’IA è fortemente legato all’impatto sul tessuto sociale ed economico, e alla paura che questi sistemi possano sostituire del tutto l’uomo. Di certo, nuove figure professionali saranno necessarie per la diffusione dell’uso di tecnologie IA, ma il futuro è certamente non ancora ben chiaro, ma potenzialmente la basi per una nuova rivoluzione industriale ci sono tutte.

3 IL RISCHIO DI CREDITO E IL CREDIT SCORING

3.1 Il rischio di credito

Il rischio di credito è inteso come la possibilità che un soggetto debitore non sia più in grado di far fronte, in tutto o in parte, ai propri obblighi, quali il rimborso del capitale e/o il pagamento degli eventuali interessi. In termini prettamente bancari, tale rischio riguarda la mancata restituzione del capitale prestato alla clientela. È necessario però, specificare che tale rischio non si riferisce alla sola possibilità di insolvenza, ma considera anche il semplice deterioramento del merito creditizio della controparte. È normale conseguenza, perché possa configurarsi un rischio di credito, l'aleatorietà nella variazione del merito di credito.

L'esercizio della funzione di intermediazione creditizia espone quindi, gli istituti bancari al rischio di perdite, totali o parziali, sui capitali dati a prestito. Perdite che possono configurarsi come attese o inattese. Da tale distinzione si osservano quelle che sono le due principali componenti del rischio di credito.

Definiamo perdita attesa (o Expected loss, EL), il valore medio di perdita che una banca si attende di registrare con riferimento ad un'esposizione o ad un pool di esposizioni, in un arco temporale di un anno. La perdita attesa viene espressa come il prodotto di tre valori stimati:

- La probabilità di insolvenza della controparte (Probability of default, PD).

Questa grandezza rappresenta la probabilità che un soggetto, nei confronti

del quale la banca sia esposta, non sia in grado di restituire il capitale preso in prestito nell'arco di tempo considerato (solitamente 12 mesi). Si tratta in tal senso della misurazione del merito creditizio, che può dipendere da condizioni economico-finanziarie, sia attuali che prospettiche, nonché eventuali fattori non prettamente quantitativi come la qualità del management aziendale nel caso in cui la controparte sia un'impresa.

- L'esposizione al momento del default (Expected exposure at default, EAD). Tale valore è la stima dell'esposizione della banca al momento dell'insolvenza del debitore, e va valutata sia per le attività di rischio per cassa sia per quelle fuori bilancio. Per quest'ultime l'EAD viene determinata attraverso un fattore di conversione creditizia (*Credit conversion factor*) che rappresenta il rapporto tra la parte non utilizzata del credito che verrà utilizzata poi nel momento dell'insolvenza e la parte attualmente utilizzata (Banca d'Italia).
- Il tasso di perdita in caso di default (Loss given default, LGD). La LGD può essere definita come il valore atteso, basato su serie storiche, della parte dell'esposizione che la banca nel momento di insolvenza della controparte, non riesce a recuperare. In un certo senso, non è altro che una stima che pesa le garanzie della banca sull'esposizione. Come spiega la Banca d'Italia, la perdita va determinata:

- facendo riferimento al concetto di perdita economica, che tiene conto dei flussi recuperati e dei costi diretti e indiretti collegati al recupero dei crediti, che devono essere attualizzati utilizzando un opportuno tasso di interesse;
- in relazione alle diverse tipologie di operazioni e non in relazione al singolo obbligato.

Il metodo di calcolo più comunemente utilizzato per quantificare la perdita effettiva è quello che considera l'esposizione al rischio e il tasso di recupero (*Recovery rate*). Quest'ultimo non è altro che la misura del capitale recuperabile nel momento in cui il debitore diventi insolvente.

$$LGD = 1 - RR$$

I tassi di recupero sono dei fattori aleatori che dipendono da diversi parametri come le caratteristiche dell'operazione, le caratteristiche del soggetto debitore, oltre che da elementi esterni alla procedura di finanziamento.

In conclusione, le perdite attese possono essere quantificate nella maniera seguente:

$$EL = PD \times LGD \times EAD$$

La perdita inattesa invece, misura il grado di variabilità del tasso di perdita attorno al proprio valore atteso. In altre parole, questa evidenzia il rischio che la perdita si dimostri superiore a quella inizialmente stimata. Riprendendo la definizione data

dalla normativa prudenziale la UL è la perdita eccedente la EL a un livello di confidenza del 99,9% su un orizzonte temporale di un anno. È necessario, perciò, utilizzare metodi statistici che possano definire la distribuzione delle perdite attese precedentemente calcolate e identificare nelle code della distribuzione normale lo 0,1%.

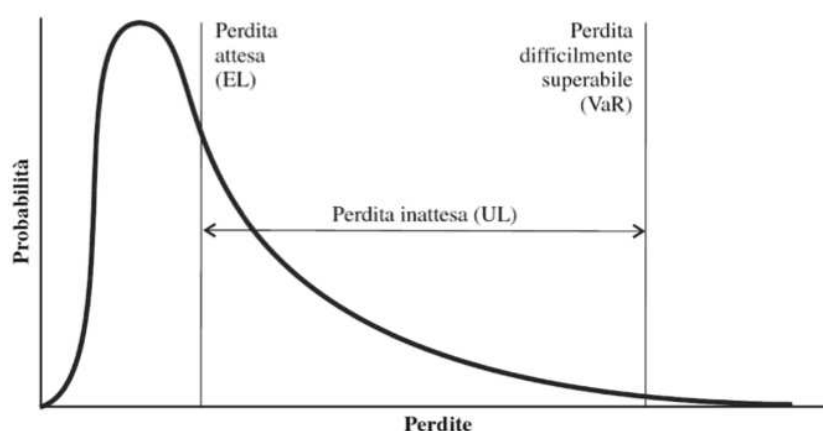


Figura 2. Esempio distribuzione normale delle perdite attese

3.2 *Gli accordi di Basilea*

Gli accordi di Basilea rappresentano delle linee guida nell'ambito dei requisiti patrimoniali e prudenziali degli intermediari creditizi, concordati ed emanati dal Comitato di Basilea. Lo statuto stesso dice: "Il Comitato di Basilea per la vigilanza bancaria (CBVB) è il principale organismo di definizione degli standard internazionali per la regolamentazione prudenziale del settore bancario. Esso fornisce una sede per la cooperazione in materia di vigilanza bancaria. Il suo


mandato consiste nel rafforzare la regolamentazione, la vigilanza e le prassi bancarie a livello mondiale allo scopo di accrescere la stabilità finanziaria”. Fu istituito nel 1974 da un gruppo di dieci paesi in occasione della crisi della banca tedesca Herstatt Bank. Le regole emanate dal comitato non sono direttamente applicabili ma necessitano, per esserlo, di un atto di recepimento da parte del relativo ordinamento. Dalla costituzione fino ad oggi sono stati raggiunti tre principali accordi, ed un quarto in discussione.

3.2.1 Basilea I

Il primo accordo di Basilea fu stipulato nel 1988, e segnò il passaggio dalla cosiddetta vigilanza “strutturale”, basata su autorizzazioni e controlli amministrativi, a una vigilanza di tipo “prudenziale” (Banca d’Italia). In altre parole, la solidità degli istituti non venne più perseguita con l’azione delle autorità di vigilanza, bensì con l’obbligo di rispettare un rapporto minimo tra il capitale detenuto e l’ammontare delle attività sottoposte al rischio. Fu in tale occasione che si iniziò a parlare di patrimonio di vigilanza, inteso come margine per fronteggiare i rischi derivanti proprio dall’esercizio della funzione di intermediazione creditizia. Tale patrimonio di vigilanza era composto, inizialmente, da due classi di capitale: capitale primario di classe 1, Tier 1, (al cui interno veniva prevista un’ulteriore categoria, l’Additional Tier 1 Capital) e il capitale di classe 2, Tier 2. Il Tier 1 è costituito dal capitale in forma di azioni ordinarie, gli utili non distribuiti e dalle

riserve di utili. Il Tier 2 è composto invece da fondi che non compaiono nei rendiconti e da tutta una serie di strumenti ibridi, che lo rendono meno affidabile del Tier 1. L'accordo originale prevedeva che il patrimonio di vigilanza da accumulare dovesse essere maggiore o uguale all'8% delle attività ponderate per classi di rischio.

Rischio crescente



0%	20%	50%	100%
Cassa	Crediti vs Banche Multilaterali di sviluppo	Mutui residenziali con garanzie reali	Crediti verso Imprese Private
Crediti vs Banche Centrali pesi OCSE	Crediti vs Banche Paesi OCSE	Leasing su immobili	
Titoli di Stato	Crediti vs Enti settore pubblico		Crediti vs Banche e Paesi non OCSE

Figura 3. Tabella ponderazioni del rischio Basilea I

Due erano però i limiti dell'accordo:

- il patrimonio da accantonare veniva definito prendendo in considerazione soltanto il rischio di credito;
- le ponderazioni si basavano esclusivamente sulla natura giuridica della controparte, spingendo gli istituti a finanziare soltanto quelle categorie di soggetti che permettevano di accantonare una quantità di capitale minore.

Il primo limite fu superato con una revisione dell'accordo che introduceva l'obbligo di accantonare capitale non soltanto a fronte del rischio di credito, ma anche a fronte del rischio di mercato³. Modifica che anticipava il secondo accordo.

3.2.2 Basilea II

Nel 2004 con il secondo accordo si cercò di eliminare definitivamente quelli che erano i limiti di Basilea 1. Il nuovo accordo era basato su tre pilastri: i requisiti patrimoniali minimi, il controllo prudenziale dell'adeguatezza patrimoniale e la disciplina di mercato. Il primo pilastro cercava di dare miglior attuazione alle regole definite dal precedente accordo, non andando a modificare quello che era il coefficiente minimo, ma bensì affiancando a quelli che erano fino ad allora gli unici rischi considerati (credito e di mercato), il rischio operativo⁴. Sotto questo aspetto, l'accordo introdusse un'importante novità per quanto riguarda il rischio di credito. Fu data la possibilità alle banche di stimare la probabilità di insolvenza dei clienti, oltre che con l'utilizzo di rating prodotti dalle agenzie, attraverso sistemi di rating e scoring interni. Si passò per cui da una ponderazione a rischio crescente in base alla categoria di appartenenza della controparte, ad una ponderazione basata sui rating. Il secondo pilastro invece introdusse l'obbligo per le banche di dotarsi di un

³ Rischio relativo alla variabilità del valore di mercato della attività e passività detenute dalla banca.

⁴ Rischio di perdite derivanti dall'inadeguatezza o dalla disfunzione di procedure, risorse umane e sistemi interni, oppure da eventi di origine esterna

sistema di controllo e valutazione dei rischi, facendo sì che esse dispongano di un capitale adeguato a sostenere i rischi connessi. Affidando alle autorità di vigilanza il mero compito di valutarne la correttezza. Infine, il pilastro della disciplina di mercato si poneva l'obiettivo di rafforzare la trasparenza e la stabilità del sistema stabilendo degli obblighi di comunicazione al pubblico per ciò che riguarda il patrimonio e l'esposizione ai rischi.

Con la crisi finanziaria del 2007 vennero alla luce i limiti di questo secondo accordo. Il primo fra tutti è legato al nuovo sistema di ponderazione per rating che ha provocato lo sviluppo di enormi conflitti d'interesse sia per quanto riguarda i sistemi di rating interni, sia per la stretta collaborazione con le agenzie. Altro principale limite fu quello della pro-ciclicità dei coefficienti che determinava in fase recessive un aumento del rischio con conseguente restrizione del credito concesso.

3.2.3 Basilea III

Come risposta alla crisi finanziaria mondiale il comitato emanò nuove linee guida, racchiuse nel terzo accordo. Basilea III ha portato con sé enormi modificazioni, tra le più importanti e significative troviamo:

- l'aumento dei requisiti di capitale con la previsione di due buffer patrimoniali aggiuntivi, il capital conservation buffer e il counter – cyclical buffer. Tali strumenti avevano l'obiettivo di limitare quello che era il

- problema della prociclicità del secondo accordo, poiché alle banche vengono richiesti accantonamenti maggiorati superiori ai requisiti minimi.
- il miglioramento della qualità del capitale. Il patrimonio di base (Tier 1), come definisce lo stesso comitato, deve essere costituito in misura preponderante da azioni ordinarie e riserve di utili (common equity).
 - l'introduzione per la prima volta di un vincolo alla leva finanziaria, il Leverage ratio, che impone alle banche dei limiti all'indebitamento, calcolato come rapporto tra il Tier 1 e le attività non ponderate al rischio (con aggiunta degli impegni fuori bilancio). Un rapporto che deve stabilizzarsi sopra la 3%.
 - la definizione di due standard minimi di liquidità, il Liquidity Coverage Ratio, che impone alle banche di detenere attività liquide non vincolate in modo tale da far fronte a crisi di liquidità di breve termine; e il Net Stable Funding Ratio che cerca di incentivare gli istituti a finanziarsi tramite fonti stabili che possano garantire equilibrio in termini di liquidità su un arco temporale di un anno.

Le riforme di Basilea III, che porteranno al nuovo accordo, sono tutt'ora in attesa di attuazione. Tra le principali modifiche contenute nel nuovo accordo possiamo evidenziare la limitazione della discrezionalità delle banche nell'utilizzo dei sistemi interni di rating e l'introduzione di un sistema standardizzato per la stima del rischio operativo. Tra le più recenti discussioni che con molta probabilità entreranno nel

nuovo accordo, c'è la considerazione del rischio climatico. Rischio molto più difficile da stimare a causa delle problematiche nel valutare gli impatti climatici sul sistema finanziario e come questi possano tramettersi all'economia reale.

3.3 Modelli di rating e di scoring nella misurazione del rischio di credito

Come introdotto dall'accordo di Basilea 2, la misurazione del rischio di credito si basa tutt'oggi sui rating delle controparti. Come stabilito dallo stesso accordo, e dalle modifiche nei seguenti, le banche possono decidere di utilizzare direttamente rating prodotti dalle agenzie del settore, si pensi a Moody's o Standard&Poor's, ovvero utilizzare rating prodotti internamente. Il rating non è altro che una valutazione espressa in una scala alfanumerica sul merito creditizio di un soggetto, che sia un'impresa o un paese. È perciò una valutazione che necessita di un enorme quantità di dati, passando da dati quantitativi in senso stretto a dati qualitativi legati alle politiche dell'impresa o alla qualità di chi la dirige, e non concentrandosi solamente sul passato del soggetto ma tenendo in considerazione anche le prospettive future. L'utilizzo dell'Internal Ratings based approach definito dagli accordi è rivolto a quelle banche che dimostrino di avere sistemi informativi a standard elevati che permettano di fare valutazioni accurate. Gli approcci interni sono di due tipologie: un approccio base in cui la banca determina internamente soltanto la probabilità di insolvenza ottenendo dalle autorità di vigilanza le stime

mancanti; e un approccio avanzato in cui la stima dei parametri viene effettuata totalmente dalla banca.

Lo scoring non è altro che uno strumento aggiuntivo, se non sostitutivo, del rating, utilizzato per raggiungere lo stesso scopo, valutare il merito di credito della controparte (generalmente per concessioni di credito al consumo) attraverso l'utilizzo di un algoritmo prefissato. Si evince che lo scoring utilizza esclusivamente dati numerici, che fanno riferimento alla sola situazione economica, finanziaria e patrimoniale passata del soggetto. Tra i principali modelli è corretto ricordare lo Z-score di Altman, che si prefigge di prevedere la probabilità che un'azienda fallisca nell'arco di due anni. Il modello originale è pensato per le società quotate, ma sono stati definiti dei sotto-modelli adattabili alle diverse realtà societarie. La formula utilizza ben cinque indicatori:

- il CCN in rapporto con il totale attivo (A)
- utile non distribuito sul totale attivo (B)
- l'utile operativo sul totale attivo (C)
- il valore di mercato dell'equity sul totale indebitamento (D)
- i ricavi di vendita sul totale attivo (E).

L'equazione è la seguente:

$$Z = 1,2A + 1,4B + 3,3C + 0,6D + 0,999E$$

Grazie al modello l'impresa verrà dichiarata finanziariamente sana per valori maggiori a 3, al contrario verrà definita una situazione di disequilibrio

finanziario per valori minori al 1,8. Per valori intermedi ai precedenti si evidenzia una situazione di incertezza tradotta in un precario equilibrio finanziario.

4 L'UTILIZZO DI TECNOLOGIE IA NEI MODELLI IRB

4.1 L'IA nella misurazione del rischio di credito

L'aumento della disponibilità di dati accompagnata all'evoluzione tecnologica, ha offerto un'enorme opportunità all'intero settore finanziario. I modelli econometrici tradizionali utilizzati nella misurazione del rischio di credito talvolta non sono in grado di tenere traccia di questi Big Data, che al contrario sono elemento essenziale per l'efficiente funzionamento dei modelli di ML. Questo cambio di direzione che si sta registrando nei modelli IRB delle banche, porta con sé l'abbandono di approcci guidati dalla teoria economica come quelli tradizionali, per far posto ad algoritmi che selezionano in maniera automatica le variabili rilevanti. Dall'introduzione di tecnologie di intelligenza artificiale all'interno delle attività degli intermediari finanziari possono discendere sia vantaggi come nuove opportunità di accesso al credito a soggetti che trovano difficoltà, sia il rischio di amplificare quelli che sono i problemi già esistenti, e rimanendo sull'esempio precedentemente citato, si potrebbero amplificare i rischi che precludono l'accesso al credito di particolari categorie di soggetti.

4.2 *Potenziali benefici*

Secondo quanto riportato dagli studi della Banca d'Italia e di altre istituzioni finanziarie internazionali, i benefici dell'utilizzo delle tecnologie ML sono principalmente legati alla disponibilità di un'enorme mole di dati. Tra i primi vantaggi notati c'è sicuramente un aumento dell'accuratezza della stima dei parametri necessari alla valutazione del merito creditizio. Una maggiore accuratezza legata fortemente al costo in termini di spiegabilità del modello, che verrà definita più avanti. Altro beneficio è legato proprio alla possibilità degli algoritmi ML di gestire ed elaborare in maniera super efficiente una grande quantità di dati. Sfruttando l'enorme potenza di calcolo di questi nuovi modelli è possibile valutare il merito di credito attraverso l'accesso ad un numero maggiore di fonti di dati, che possono andare da quelli strettamente finanziari, fino a dati non strutturati non finanziari tipo i dati di navigazione dei soggetti o i dati provenienti dai diversi social network in cui il soggetto è iscritto. Proprio l'utilizzo di fonti alternative consente di sfruttare informazioni complementari a quelle che già vengono utilizzate nei modelli tradizionali, in modo da migliorare l'accuratezza previsiva del modello stesso, come precedentemente indicato, e soprattutto ridurre le eventuali asimmetrie informative insite nel rapporto tra richiedente e banca (o altro istituto). Il risultato è la possibile estensione della valutazione del merito creditizio anche a soggetti che in precedenza non possedevano dati sufficienti per la concessione del credito, favorendo l'inclusione finanziaria. Da sottolineare però, che ciò potrebbe

virare anche nel senso opposto intensificando il rischio di esclusione finanziaria come verrà esposto nei paragrafi seguenti. È inoltre opinione diffusa che l'intelligenza artificiale porti con sé una riduzione dei costi operativi, visto che l'utilizzo di queste tecnologie riduce il tempo impiegato per valutare e successivamente definire la concessione o meno del credito, aumentando potenzialmente la frequenza delle concessioni. Legato alla riduzione dei costi è l'incremento della concorrenza nel mercato del credito, come evidenzia Banca d'Italia, poiché potrebbero inserirsi soggetti diversi dagli intermediari finanziari tradizionali con lo scopo di offrire finanziamenti a costi più bassi, grazie proprio all'utilizzo di fonti dati innovative.

4.3 Risultati distorti

Tra le potenziali problematiche nell'utilizzo di algoritmi IA c'è sicuramente la possibile presenza di distorsioni. Nell'ambito del rischio di credito tali distorsioni comportano situazioni di non corretta differenziazione e discriminazione dei clienti. Con non corretta differenziazione si intende, la mancata capacità di un modello di ordinare correttamente i clienti per livello di merito creditizio, non raggiungendo un livello ottimale di efficienza allocativa di risorse (Banca d'Italia). Tale distorsione potrebbe registrare come conseguenza diretta la possibilità che la banca (o altra istituzione) scelga in maniera non corretta i clienti a cui concedere prestiti, oltre alla possibilità di distorsioni nel pricing. La discriminazione indica invece, la

presenza di pregiudizi o trattamenti sfavorevoli nei confronti di determinate categorie di individui, identificate tramite l'utilizzo di particolari attributi sensibili, comportando il rischio di esclusione finanziaria. La discriminazione si può presentare in maniera diretta o indiretta. Nel primo caso si fa riferimento all'utilizzo di variabili che etichettano dei soggetti, individui o imprese che siano, come vulnerabili. Nel secondo caso invece non vengono utilizzate in maniera esplicita tali variabili, bensì parametri che siano correlati con l'appartenenza ad un gruppo vulnerabile. In questi casi, come la Banca d'Italia suggerisce nell'Occasional Paper n. 721, è d'aiuto ricorrere alla definizione di fairness, che consente di valutare la presenza di eventuali meccanismi di discriminazione. L'adozione di una definizione appropriata potrebbe dipendere soprattutto dall'atteggiamento dell'analista, rigido o permissivo. Sta di fatto che la definizione maggiormente adottata sembra essere quella di eliminare tali variabili sensibili dal modello. Come definito da uno studio dell'International Monetary Fund, si potrebbe assistere ad una sorta di falsificazione degli indicatori. Nel momento in cui i soggetti richiedenti comprendono quali siano gli indicatori utilizzati per il calcolo del proprio scoring, sarebbero motivati, fin dove sia possibile, a modificare artificiosamente i valori. Queste distorsioni però potrebbero risultare comuni anche agli algoritmi tradizionali utilizzati, in passato e tutt'ora. La questione tende ad aggravarsi proprio nel momento in cui vengono implementati nella valutazione del merito di credito, le tecnologie ML. Queste in quanto tali, e come definito nei precedenti capitoli,

sono tecnologie che apprendono e si addestrano tramite il loro stesso utilizzo. La presenza di queste distorsioni potrebbe generare un meccanismo in cui la distorsione venga confermata e addirittura rinforzata. Si pensi ad esempio ad un modello non adeguatamente specificato che porti alla sistematica negazione di finanziamenti nei confronti di determinate categorie di soggetti o imprese. Si costituirebbe un bias storico che si proietterà sui campioni estratti, rafforzando la discriminazione del modello stesso. L'utilizzo dei Big Data cerca di limitare questa situazione sgradevole in virtù dell'accesso ad informazioni altamente ricche e dettagliate. Di contro però l'avvento dei Big Data porta con sé un fattore discriminante dovuto dal fatto che soltanto alcune generazioni posseggono una propria identità digitale da cui estrapolare dati, mentre altre verrebbero lasciate fuori perché scarsamente rappresentate all'interno del mondo digitale o del tutto assenti. In questi casi ricercare l'equità dei modelli stesse potrebbe non risolvere il problema della discriminazione.

4.4 Una questione etica tra privacy e spiegabilità

L'utilizzo di tecniche di ML, come già ripetutamente evidenziato, porta ad avere un'accuratezza predittiva superiore. A ciò però si contrappone una sorta di oscurità di funzionamento, tale da denotarle come delle black box. Alla battaglia per garantire trasparenza dei modelli utilizzati, da parte delle autorità di vigilanza,

dovrà affiancarsi la “sfida” degli istituti di credito contro le già definite normative di tutela della privacy, vista la possibilità di usufruire anche di dati sensibili.

Uno studio dell’università di Malta (Demajo, Vella e Dingli) ha cercato di proporre un modello che potesse garantire spiegabilità dei risultati, come le autorità chiedono, senza compromettere l’accuratezza predittiva. Il modello in questione incorpora l’algoritmo XGBoost, il quale si basa su dei campioni che rappresentano l’esatto comportamento del sistema che si vuol simulare. Le questioni che hanno mosso la volontà di ricercare nuovi modelli per il credit scoring, come quello appena citato, sono da ricondurre nella necessità di comprendere ciò che accade all’interno degli algoritmi, in maniera tale da poter giustificare i risultati raggiunti, sia per l’appunto alle autorità che ai clienti. Lo studio si è prefissato di creare un modello che possa dare motivazioni e giustificazioni differenti in base ai diversi soggetti interessati al mercato del credito. In particolare, nel credit scoring possono essere identificati tre soggetti richiedenti di giustificazioni in relazione dei risultati in output:

- gli operatori interni all’istituzione che siano responsabili della concessione di prestiti;
- i clienti richiedenti nell’eventuale negazione del credito;
- i manager dell’istituto o le autorità di vigilanza con lo sguardo rivolto ai punti deboli dei modelli.

Ognuno di questi, si attende delle spiegazioni differenti. Per quel che riguarda il personale interno alla banca, si farà riferimento a spiegazioni locali che permettono di giustificare la singola richiesta di prestito. In questo modo il funzionario avrebbe la possibilità di confrontare e soprattutto verificare la validità del modello esaminando richieste di prestito precedenti dallo stesso esito e di soggetti con caratteristiche simili. Questo aiuterebbe ad individuare la causa del rifiuto di una specifica domanda di prestito rispetto a quelle accettate. È una tecnica che potrebbe supportare tale questione è quella della permutazione dei regressori all'interno degli algoritmi. Un regressore sarà importante se al variare dei suoi valori, l'errore predittivo aumenta in maniera notevole.

I mancati mutuatari preferiranno invece delle spiegazioni su come e perché sia stato negato il prestito. In questo caso le giustificazioni vengono solitamente fornite sotto forma di livelli di rilevanza delle caratteristiche utilizzate all'interno dei modelli.

Infine, per quanto riguarda le autorità e i manager bancari, si ha un interesse ad ottenere una spiegabilità dei modelli in termini globali. Questi come già accennato, pongono l'attenzione sull'accuratezza e il corretto funzionamento dell'intero modello, piuttosto che delle singole previsioni. Solo attraverso una comprensione globale, diventa possibile interpretare i risultati in maniera tale da individuare eventuali lacune ed apportare i conseguenti miglioramenti.

Tale modello non farebbe che portare vantaggi a tutti i soggetti interessati, visto che lo studio condotto dalla Banca d'Italia evidenzia come le tecniche di spiegabilità

vengono utilizzate prevalentemente per comprendere le caratteristiche del dataset utilizzato per addestrare il modello e la logica dietro le sue decisioni. Si nota per cui un vuoto nei confronti della clientela esterna, nonostante le normative del GDPR impongano che su sua richiesta il cliente debba ricevere le informazioni in merito alla logica utilizzata dal modello (diritto di spiegazione). A questo si lega fortemente anche la questione sulla privacy, poiché tali algoritmi potrebbero includere anche l'utilizzo di dati personali sensibili, con il rischio di arrivare ad escludere soggetti sulla base di motivazioni poco etiche. Il problema potrebbe affievolirsi, come suggerisce il Fondo Internazionale Monetario, escludendo tali variabili potenzialmente discriminatorie, e lasciando alle autorità di vigilanza il compito di monitorare le variabili in input dei modelli.

4.5 Quadro normativo e visione delle istituzioni internazionali

L'intelligenza artificiale nonostante sia ormai presente da molto, come visto nei capitoli precedenti, solo in tempi più recenti ha registrato un'imponente diffusione anche tra il pubblico privato. Basti pensare al fenomeno ChatGPT di OpenAI che ha completamente stravolto l'atteggiamento delle persone nei confronti di queste nuove macchine. Tutto ciò però porta benefici alla società solo se accompagnato da regole e limiti che ne garantiscono un progresso stabile e genuino. Prima che la situazione possa sfuggire di mano, negli ultimi anni sempre più istituzioni nazionali e internazionali hanno definito la loro visione a riguardo. Tra le istituzioni di

maggior rilievo, la Commissione Europea ha presentato la bozza del “Regolamento sull’approccio europeo per l’intelligenza artificiale”. Il quadro presentato si pone in primis di assicurare l’allineamento dei sistemi IA alla normativa sui diritti fondamentali, rendere l’Unione Europea un’area in cui prosperi lo sviluppo di tali sistemi assicurando certezza per gli investimenti su di essi, e prevenire la frammentazione del mercato garantendo i cittadini UE da sistemi illeciti e inaffidabili. La normativa definisce poi il concetto di sistemi ad alto rischio, che richiedono tutele e presidi maggiori rispetto alle altre, ciò denota la chiara adozione di un approccio basato sul rischio da parte dell’Europa. La misurazione del merito creditizio rientra in questa categoria visto il rischio che tecniche AI utilizzate in quest’ambito, come evidenziato in precedenza, potrebbero provocare dinamiche discriminatorie. In tal senso sono stati identificati dalla Commissione alcuni obblighi a cui i fornitori/produttori di queste tecnologie devono sottostare, e senza ombra di dubbio, l’argomento interessa sia agli intermediari che decidessero di appoggiarsi a soggetti esterni competenti in materia, sia per quelli che opterebbero per uno sviluppo interno dei modelli. Tra gli obblighi se ne possono ricordare alcuni come l’utilizzo di dataset di addestramento pertinenti, rappresentativi, completi e soprattutto dotati di adeguate proprietà statistiche in riferimento ai soggetti a cui è destinato; si richiede un adeguato livello di accuratezza, robustezza e sicurezza (con particolare attenzione ai dati personali), oltre ad essere richiesta anche la presenza del controllo umano che sia in grado di individuare eventuali distorsioni e

intervenire al riguardo. L'attenzione del sistema finanziario deve però porsi anche nei confronti di quelle che la normativa definisce pratiche vietate, poiché tra di esse, ci sono anche i sistemi che attribuiscono dei punteggi sociali in base al comportamento delle persone, e nella definizione dello score creditizio di persone fisiche tale divieto non può passare inosservato.

Nell'agosto del 2020, anche il Hong Kong Monetary Authority ha pubblicato un report, sviluppando delle linee guida per gli istituti bancari che avessero l'intenzione di implementare le tecnologie IA. L'attenzione è stata posta sulla definizione di principi di alto livello per guidare le banche nell'adozione di questi sistemi. L'approccio dell'HKMA si basa su due principi: la tecnologia neutrale e basata sul rischio. La neutralità tecnologica ha come obiettivo quello di creare un ambiente favorevole allo sviluppo da parte delle banche di nuove tecnologie che migliorino la loro efficienza operativa. L'approccio basato sul rischio, che riprende la proposta europea, implica una supervisione più attenta sulle istituzioni che adottino sistemi più complessi tali da avere un impatto maggiore sui clienti, rispetto a quelle banche che utilizzano sistemi più semplici. Linee guida queste che, mirano a rafforzare la governance in tre diverse aree: model risk management, consumer protection e cyber security. L'HKMA in riferimento ai principi di alto livello, identifica tre aspetti chiave: governance, progettazione e sviluppo e, monitoraggio e manutenzione continui. Con governance si vuol sottolineare la responsabilità della direzione della banca sui risultati dei modelli IA utilizzati, si rende perciò

necessario stabilire un quadro di governance che possa supervisionare i modelli e mitigare i potenziali rischi. Per quanto riguarda la progettazione e lo sviluppo, la direzione deve ugualmente definire procedure che possano garantire la qualità dei dati, dovrebbero per cui esserci le competenze sufficienti. Nel caso in cui la banca richieda a terze parti la progettazione del modello, dovrà condurre revisioni nei confronti dei fornitori, e avere un'adeguata comprensione del funzionamento del modello, poiché rimane sempre in capo alla banca l'obbligo di spiegare i risultati del modello alle parti interessate. Ciò che invece rientra nell'area del monitoraggio e manutenzione, sono revisioni periodiche dei modelli sul rispetto dei requisiti di privacy e protezione dei dati nonché l'implementazione di una cyber security in grado di respingere attacchi hacker (rischio operativo).

In linea con la visione europea anche Hong Kong obbliga le banche a dichiarare, prima di fornire servizi finanziari, che il servizio in questione, in caso, sia basato su tecnologie di IA. In aggiunta a ciò, l'HKMA suggerisce anche alle banche, per migliorare la protezione di dati che con l'utilizzo di queste nuove tecnologie aumentano a dismisura, di raccogliere e archiviare solo la quantità minima di dati per la quantità minima di tempo (minimizzazione di dati).

Tra le diverse posizioni di istituzioni internazionali, spicca sicuramente anche quella del Monetary Authority of Singapore che nel 2018 presentò quattro principi per l'utilizzo dell'intelligenza artificiale, denotati FEAT (fairness, ethics,

accountability e trasparenza). Nel 2019 la banca centrale olandese aggiunse al FEAT due ulteriori principi, trasformandolo in SAFEST (soundness and skills):

- Soundness, le applicazioni IA devono essere accurate e prevedere meccanismi che attivino la revisione quando il modello produce risultati inaspettati;
- Accountability, la direzione ha sempre la responsabilità sui rischi legati all'IA, la fornitura dei modelli da terzi parti non può mai essere una scusa;
- Fairness, si deve evitare l'utilizzo di dati sensibili in input, i quali potrebbero provocare discriminazioni ingiuste;
- Ethics, si suggerisce di prendere in considerazione regole come la privacy dei dati;
- Skills, è necessario che la banca adegui il suo personale interno attraverso l'assunzione di soggetti competenti sull'IA;
- Trasparenza, è necessario documentare i punti deboli dei modelli e sviluppare strumenti che facilitino la spiegabilità degli stessi.

5 CONCLUSIONI

Le questioni analizzate permettono di concludere che l'intelligenza artificiale, nel contesto storico in cui ci troviamo sta trovando carburante senza limiti per poter esprimere tutte le sue potenzialità. Nel settore finanziario, su cui il lavoro si concentra, l'enorme mole di dati permette di aumentare l'accuratezza delle analisi, motivo per cui la stragrande maggioranza degli attori utilizzano servizi IA, e coloro che non rientrano in questo gruppo hanno già in programma di implementarla all'interno dei rispettivi business. Diventa per cui, l'intelligenza artificiale, protagonista dei futuri sviluppi nella misurazione del merito di credito dei soggetti, permettendo alle banche e alle altre istituzioni finanziarie, di erogare credito "sicuro". Questo potrebbe rivelarsi un vantaggio non solo per gli intermediari, ma anche per le maggiori istituzioni nazionali e internazionali che abbiano l'obiettivo della stabilità finanziaria del sistema. Tuttavia, è necessario tenere presente dei rischi che tali tecnologie incorporano. Riprendendo il lavoro si sono evidenziati soprattutto problemi etici riguardanti la privacy e la spiegabilità dei modelli, oltre al pericolo di esclusione finanziaria dovuto a discriminazioni ingiuste e la possibilità di una specificazione errata del modello. Dalle numerose fonti consultate, è evidente che le attenzioni maggiori delle autorità di regolamentazione, si riversano sulla questione della protezione dei dati e sull'apertura dei modelli IA, descritti come delle black-box. Diventano a questo punto necessari, interventi normativi, che possano garantire un genuino avanzamento dello stato tecnologico,

in maniera tale che tutti gli attori interessati (banche, autorità di vigilanza, clienti e terzi fornitori), possano usufruire dei benefici senza dover correre il rischio di vedersi negare il credito ingiustificatamente, oltre che non ottenere giustificazioni al riguardo, ma soprattutto senza veder violata la propria privacy. Le prime proposte di regolamentazione si dirigono, a supporto di ciò, verso, da una parte un approccio basato sul rischio, e dall'altra un approccio antropocentrico, che faccia in modo che la dignità umana non sia mai lesa, e che queste macchine siano soltanto di supporto e mai in sostituzione all'uomo.

SITOGRAFIA

<https://schiaffonati.faculty.polimi.it/pubblicazioni/H1.pdf>

<https://www.intelligenzaartificiale.it>

https://www.bancaditalia.it/compiti/vigilanza/normativa/consultazioni/2006/basilea2/Doc_Cons_IRB.pdf

<https://arxiv.org/pdf/2012.03749.pdf>

<https://www.agendadigitale.eu/cultura-digitale/ai-e-finanza-tra-discriminazioni-e-imprecisioni-rischi-e-contromisure/>

<https://www.aof.org.hk/docs/default-source/hkimr/applied-research-report/airep.pdf>

<https://www.bankofengland.co.uk/research/fintech/ai-public-private-forum>

<https://www.mise.gov.it/index.php/it/strategia-intelligenza-artificiale/contesto>

BIBLIOGRAFIA

A. TURING, *Computer Machinery and intelligence*, “Mind 49”, 1950, pp. 433-460.

BONACCORSI DI PATTI E., CALABRESI F., DE VARTI B., FEDERICO F., ... & RINNA G. (2022). *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano*. Bank of Italy Occasional Paper, (721).

BAZARBASH M. (2019). *FinTech in Financial Inclusion Machine Learning Applications in Assessing Credit Risk*. IMF Working Paper (109).

CASCARINO G., MOSCATELLI M., & PARLAPIANO F. (2022). *Explainable Artificial Intelligence: interpreting default forecasting model based on Machine Learning*. Bank of Italy Occasional Paper (674).

