

LUISS



*Dipartimento di
Impresa e Management*

*Cattedra
Economia dei Mercati e degli
Intermediari Finanziari*

***“MACHINE LEARNING: COME L’INTELLIGENZA ARTIFICIALE STA
RIVOLUZIONANDO LA GESTIONE DEL RISCHIO”***

RELATORE
Prof. Francesco Cerri

CANDIDATO
Giacomo Flores
Matr. 231441

ANNO ACCADEMICO 2020/2021

*Ai miei genitori,
ai miei nonni
e alla mia fidanzata.*

*Grazie per avermi supportato e
sopportato durante tutti questi anni.*

Grazie.

INTRODUZIONE	5
1. MACHINE LEARNING E GESTIONE DEL RISCHIO	6
1.1. MACHINE LEARNING	6
1.1.1. <i>Categorie</i>	7
1.1.2. <i>Metodologie</i>	8
1.1.3. <i>Ambiti di utilizzo</i>	11
1.1.4. <i>Vantaggi e svantaggi</i>	12
1.2. GESTIONE DEL RISCHIO	13
1.2.1. <i>Tipologie di rischio</i>	14
2. IL CONTROLLO DEL RISCHIO FINANZIARIO ATTRAVERSO IL MACHINE LEARNING	17
2.1. EXTREME GRADIENT BOOSTING	19
2.2. IL PROCESSO DI CREDIT SCORING	23
2.2.1. <i>Background</i>	24
2.2.1.1. L'accordo di Basilea sul capitale	28
2.2.2. <i>Credit Scorecards</i>	28
2.2.2.1. Costruzione del set di dati.....	32
2.2.2.2. Modellazione.....	35
3. MACHINE LEARNING APPLICATO AGLI ALTRI TIPI DI RISCHIO	41
3.1. APPLICAZIONE AL RISCHIO DI MERCATO.....	41
3.2. APPLICAZIONE AL RISCHIO OPERATIVO	42
3.3. APPLICAZIONE A REGTECH.....	43

4. LE SFIDE E IL FUTURO DELL'APPRENDIMENTO AUTOMATICO NELLA GESTIONE DEL RISCHIO	45
5. BIBLIOGRAFIA E SITOGRAFIA.....	48

Introduzione

Successivamente alla crisi del 2008, la gestione del rischio nelle società finanziarie ha acquisito un'importanza sempre maggiore. Una notevole ricerca sia nel mondo accademico che in quello aziendale, ha portato un'evoluzione dei metodi di identificazione del rischio con una crescente influenza del machine learning, e delle varie metodologie che lo caratterizzano, alcune di queste sono già state implementate e molte altre sono in fase di esplorazione.

Mckinsey & Co ha sottolineato come le funzioni di rischio nelle società finanziarie, entro il 2025, dovranno essere fundamentalmente diverse da quelle attuali. L'ampliamento e l'approfondimento delle normative, l'evoluzione delle aspettative dei clienti e l'evoluzione dei tipi di rischio, dovrebbero guidare un cambiamento interno della gestione del rischio.

Nuovi prodotti, servizi e tecniche di gestione sono già possibili attraverso l'applicazione di tecnologie in evoluzione ed analisi avanzate. L'apprendimento automatico può consentire la costruzione di modelli di rischio sempre più accurati identificando schemi complessi e non lineari all'interno di grandi set di dati. La potenza predittiva di questi modelli può crescere con ogni bit di informazione aggiunto, migliorando così la affidabilità nel tempo.

L'intelligenza artificiale (IA), e le tecniche di machine learning che ne costituiscono il nucleo, stanno quindi trasformando e rivoluzionando il modo in cui ci avviciniamo alla gestione del rischio. Tutto ciò che ha a che fare con la comprensione e il controllo del rischio è a portata di mano attraverso la crescita di soluzioni basate sull'IA: dal decidere quanto una banca dovrebbe prestare ad un cliente, al fornire segnali di avvertimento, agli operatori del mercato finanziario, sul rischio di una posizione, all'individuare i casi di frode a causa di clienti o di informazioni privilegiate.

In questa tesi, dopo un'introduzione teorica, andrò ad analizzare queste tematiche, concentrandomi sulla gestione del rischio di credito e in particolare il credit scoring.

1. Machine Learning e Gestione del Rischio

1.1. Machine Learning

A livello puramente teorico il machine learning (ML), in italiano “apprendimento automatico”, è una categoria di intelligenza artificiale che permette ai computer di pensare e imparare senza l'intervento umano.

Il termine è stato coniato da Arthur Samuel nel 1959, che ha definito ML come un campo di studio che fornisce capacità di apprendimento ai computer senza essere esplicitamente programmati. Più recentemente, Tom Mitchell ha dato una definizione che si è dimostrata più utile per l'impostazione ingegneristica: "Si dice che un programma informatico apprenda dall'esperienza E rispetto a qualche compito T e a qualche misura di performance P , se le sue prestazioni su T , misurate da P , migliorano con l'esperienza E "

L'apprendimento automatico è un campo multidisciplinare con una vasta gamma di campi di ricerca che rafforzano la sua esistenza. La simulazione dei modelli di ML è significativamente correlata alla Statistica computazionale il cui obiettivo principale è quello di concentrarsi sulla realizzazione di previsioni tramite computer e anche alla “Mathematical Optimization” che mette in relazione modelli, applicazioni e framework con il campo della statistica.

I problemi del mondo reale sono altamente complessi, il che li rende candidati eccellenti per l'applicazione di ML. L'apprendimento automatico può essere infatti applicato a vari settori dell'informatica per progettare e programmare algoritmi espliciti con risultati ad alte prestazioni, ad esempio filtraggio di spam via email, rilevamento di frodi sui social network, stock trading online, rilevamento di volti e forme, diagnosi medica, previsione del traffico, riconoscimento del carattere e raccomandazione del prodotto, e tanti altri.

1.1.1. Categorie

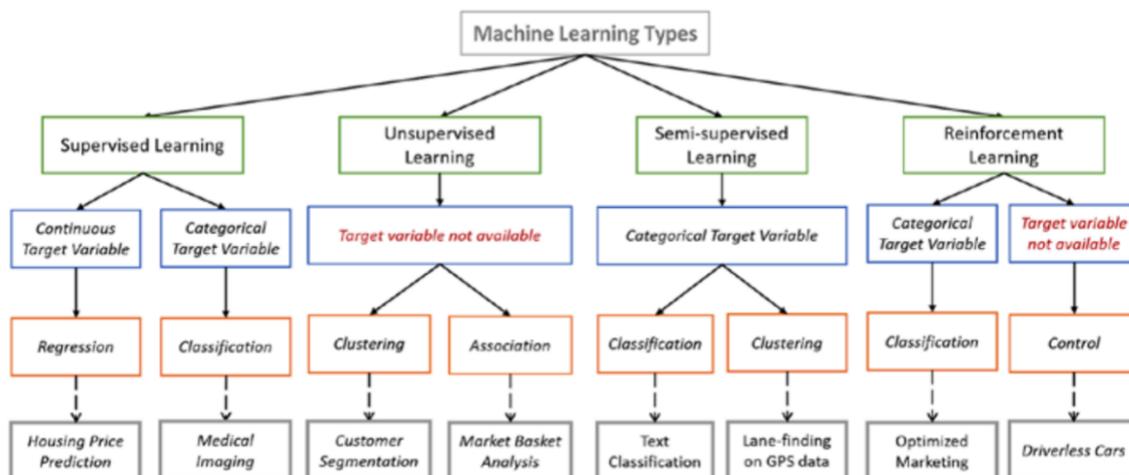
Prima di passare alla risoluzione dei problemi, questi devono essere classificati così che possa essere applicato ad esso l'algoritmo di ML più appropriato.

Le varie categorie sono spiegate di seguito:

- **Problemi di classificazione:** un problema in cui l'output può essere solo uno, scelto tra un numero fisso di classi di output note a priori come Yes/No, True/False. A seconda del numero di classi di output, il problema può essere un problema di classificazione binario o multi-classe.
- **Anomaly Detection Problem:** problemi che analizzano un certo modello e rilevandone cambiamenti o anomalie. Per esempio, le aziende della carta di credito usano gli algoritmi di rilevazione delle anomalie per individuare le transazioni insolite. Tali problemi riguardano la ricerca dei valori anomali.
- **Regression Problem:** hanno il compito di stabilire una relazione tra due o più variabili. Fornito a un modello di regressione un valore x , questo restituirà il corrispondente valore y generato dall'elaborazione di x . Questi sono di solito utilizzati per problemi che si occupano di questioni come, quanto o quanti'.
- **Clustering Problem:** rientra nella categoria degli algoritmi di apprendimento non supervisionati. Il clustering è l'attività di raggruppamento di istanze in gruppi o cluster sulla base di caratteristiche comuni; questo può essere definito come un insieme di tecniche per l'analisi multivariata, cioè quelle parte della statistica in cui l'oggetto dell'analisi è formato da almeno due componenti aventi l'obiettivo di selezionare e raggruppare oggetti omogenei in un dataset.
- **Reinforcement Problem:** gli algoritmi di rinforzo sono utilizzati quando una decisione deve essere presa in base alle esperienze passate e attraverso l'interazione con l'ambiente esterno. Fornisce un modo per programmare gli agenti utilizzando il concetto di ricompense e sanzioni senza specificare come il compito debba essere realizzato.

1.1.2. Metodologie

A seconda di come un algoritmo viene “addestrato” e sulla base della disponibilità dei dati durante la formazione, i paradigmi di apprendimento automatico possono essere classificati in più categorie. Queste includono:



Apprendimento supervisionato

Nell'ambito dell'apprendimento sotto supervisione, una serie di esempi o moduli di formazione sono forniti con i risultati corretti e, sulla base di questi insiemi di formazione, l'algoritmo impara a rispondere in modo più accurato, confrontando i suoi risultati con quelli che sono dati come input. L'apprendimento supervisionato è anche conosciuto come apprendimento tramite esempi o apprendimento da esemplari.

L'obiettivo principale di questa categoria di algoritmi è quello di prevedere a partire dai feature vectors, cioè la descrizione del fenomeno, quale sia l'etichetta corrispondente: infatti, una volta che il modello è addestrato sui dati etichettati, in base alla qualità e alla quantità dei dati, dovrebbe "imparare" le caratteristiche necessarie per definire le interazioni che caratterizzano una specifica classe o popolazione.

L'apprendimento supervisionato trova applicazioni nella previsione basate su dati storici. I compiti di apprendimento supervisionati possono essere ulteriormente classificati come compiti

di classificazione e regressione. In caso di classificazione, le etichette di output sono discrete mentre sono continue in caso di regressione.

In finanza tale metodo può trovare la sua applicazione, per esempio, nel caso in cui si hanno dei rendimenti (etichette) e i rispettivi indicatori (feature vectors) necessari per cercare di prevedere, dato un certo scenario potenziale di mercato, i risultati attesi o per cercare di predire se un portafoglio avrà un esito positivo o negativo.

Apprendimento non supervisionato

L'approccio di apprendimento senza supervisione consiste nel riconoscere modelli di dati esistenti per ricavarne regole. Questa tecnica è appropriata in una situazione in cui le categorie di dati sono sconosciute e quindi non etichettate. L'apprendimento non supervisionato è considerato un approccio statistico per l'apprendimento e si riferisce quindi al problema di trovare una struttura nascosta in dati non marcati.

Nel clustering è possibile cercare la presenza di schemi comuni per stabilire le classi di appartenenza dei dati: ad esempio, data la reazione ad uno shock esogeno di mercato, possiamo differenziare diverse classi di stock?

In questo caso, i dati possono dare maggiori approfondimenti rispetto ad un rating qualitativo ed è inoltre possibile osservare che il mercato può essere soggetto a pregiudizi emotivi, che non tengono necessariamente conto del valore fondamentale delle aziende, a differenza delle agenzie di rating.

Così, nel processo del clustering, è possibile differenziare e prendere in considerazione se un titolo appartiene a una specifica classe di attività o no - a seconda del mercato - utilizzando i dati ex post.

Nel caso in cui si volessero dedurre informazioni rilevanti sulle prestazioni complessive di una serie di aziende, ma si è in presenza di troppi parametri e indici, si può passare alla riduzione della dimensionalità. È possibile fare ciò utilizzando l'analisi dei fattori, al fine di comprendere e analizzare le relazioni tra le diverse caratteristiche oppure l'analisi dei componenti principali per costruire un feature vector a bassa dimensione, con variabili interne non correlate, che sintetizzano il fenomeno sottostante partendo da un elevato numero di variabili caratterizzate da multicollinearità.

Apprendimento Semi-Supervisionato

L'apprendimento semi-supervisionato è applicato a problemi di apprendimento supervisionati che hanno relativamente poche osservazioni etichettate e un numero maggiore di osservazioni senza etichetta che hanno la stessa distribuzione e fanno parte della popolazione di quelle etichettate; può dunque essere visto come un quarto tipo di apprendimento automatico.

In questo caso, al contrario dell'apprendimento non supervisionato, tra tutti i dati presenti nel training set, solo pochi di essi sono stati etichettati.

Generalmente gli algoritmi sono un mix tra approccio supervisionato e non supervisionato, per esempio si hanno pochi bilanci e poche informazioni relative al tipo di industria, si può cercare di classificare tutti i bilanci senza un'industria assegnata per identificare la tipologia alla quale appartengono queste aziende.

Apprendimento per rinforzo

L'obiettivo di questo tipo di apprendimento è quello di costruire un sistema (agente) che attraverso le interazioni con l'ambiente migliori le proprie performance. Per poter migliorare le funzionalità del sistema vengono introdotti dei "rinforzi", cioè dei "segnali di ricompensa". Questo rinforzo non è dato dalle label (etichette), né dai valori corretti di verità, ma è invece una misurazione sulla qualità delle azioni intraprese dal sistema. Per questo motivo, non può dunque essere assimilato ad un apprendimento supervisionato.

L'ambiente può essere analizzato, e alla fine modificato, dall'agente che prende "scelte" secondo una politica e che di conseguenza riceverà una ricompensa specifica, oppure una punizione.

L'obiettivo principale dell'agente è quello di apprendere una politica ottimale a lungo termine dopo una serie di tentativi ed errori eseguiti, al fine di massimizzare la ricompensa prevista.

Ad esempio, un algoritmo di trading tramite apprendimento di rinforzo, che fornisce liquidità al mercato, può imparare dalle serie temporali dei prezzi come massimizzare il rendimento atteso, così come i ricavi, per ridurre il mispricing del bene nel tempo.

1.1.3. Ambiti di utilizzo

Attualmente l'apprendimento automatico ha una vastissima gamma di applicazioni: i sistemi di trading possono essere calibrati per identificare nuove opportunità di investimento; le piattaforme di marketing ed e-commerce possono essere ottimizzate per fornire consigli accurati e personalizzati ai propri utenti in base alla cronologia degli utenti o alle transazioni precedenti; le istituzioni di prestito possono incorporare il ML per predire i crediti inesigibili e per sviluppare un modello di rischio di accreditamento; le banche possono creare strumenti di rilevamento delle frodi.

L'incorporazione dell'apprendimento automatico nell'era del digital-savvy è infinita, dal momento che i commerci ed i governi diventano sempre più informati sulle opportunità che i big data rappresentano a loro favore.

Oltre a quelle già elencate in precedenza, tra le applicazioni più comuni del machine learning troviamo:

- Gestione delle relazioni con i clienti: il software di CRM può usare i modelli di apprendimento automatico per analizzare l'e-mail e segnalare al team di vendita quelle di maggior importanza e, in caso di sistemi più avanzati, possono anche raccomandare risposte potenzialmente efficaci.
- Business intelligence: BI e società di consulenza utilizzano il ML per identificare dati potenzialmente importanti, modelli che si ripetono e anomalie.
- Sistemi di informazione delle risorse umane: i sistemi HRIS possono utilizzare modelli di apprendimento automatico per filtrare attraverso le domande di impiego e identificare i migliori candidati per una posizione aperta.
- Auto a guida autonoma: gli algoritmi di apprendimento automatico possono anche rendere possibile per un'auto semi-autonoma il riconoscimento di un oggetto parzialmente visibile.
- Assistenti virtuali: tipicamente combinano modelli di apprendimento automatico supervisionati e non supervisionati per interpretare il linguaggio e il contesto in modo tale da fornire la risposta più adatta.

1.1.4. Vantaggi e svantaggi

Il machine learning aiuta le imprese a comprendere i bisogni dei loro clienti a un livello più profondo. Raccogliendo i dati e correlandoli con i comportamenti nel lungo periodo, gli algoritmi di apprendimento automatico possono infatti aiutare i team ad adattare le iniziative di sviluppo e marketing alle richieste dei clienti.

Diverse aziende utilizzano il ML come driver primario nei loro modelli di business. Degli esempi possono essere rappresentati da Uber e Google. Mentre il primo utilizza algoritmi per abbinare i driver con i piloti, il secondo utilizza l'apprendimento automatico per scegliere la pubblicità da proporre agli utenti.

Tuttavia, l'apprendimento automatico porta con sé anche degli svantaggi. In primo luogo, può essere un processo costoso, tali progetti richiedono spesso infatti delle infrastrutture di software molto sofisticate.

Un altro problema è quello della polarizzazione. Gli algoritmi formati su set di dati che escludono determinate popolazioni o contengono errori possono portare a modelli inaccurati del mondo che, nella migliore delle ipotesi, falliscono o, nella peggiore, risultano essere discriminatori. Quando un'impresa basa i processi di core business su modelli polarizzati rischia di incorrere in danni normativi e reputazionali.

Tradizionalmente, gli operatori di investimento nel mercato azionario - come ricercatori finanziari, analisti e singoli investitori - devono orientarsi in una molteplicità di informazioni provenienti da diverse aziende di tutto il mondo, al fine di prendere decisioni di investimento redditizie. Tuttavia, alcune informazioni pertinenti possono non essere ampiamente pubblicizzate dai media e possono essere conosciute solo dai pochi che hanno il vantaggio di essere dipendenti della società o residenti nel paese da cui l'informazione deriva. È inoltre evidente che si hanno in circolazione molte più informazioni rispetto a quelle che un essere umano è in grado di raccogliere ed elaborare correttamente. È proprio qui quindi che l'apprendimento automatico entra in gioco.

Ad esempio, un'impresa di gestione patrimoniale può impiegare il machine learning nella sua analisi degli investimenti e nella sua area di ricerca. Ammettiamo che l'asset manager investa solo nelle azioni minerarie, il modello integrato nel sistema analizza il web e raccoglie tutti i

tipi di notizie da imprese, industrie, città e paesi, rilevanti per l'attività dell'impresa andando quindi a costruire i vari set di dati.

1.2. Gestione del rischio

Nel mondo finanziario, la gestione del rischio è il processo di identificazione, analisi e accettazione o attenuazione dell'incertezza nelle decisioni di investimento. In sostanza, la gestione del rischio si verifica quando un investitore o un gestore di fondi analizza e tenta di quantificare le perdite potenziali di un investimento, per poi selezionare l'azione appropriata, dati gli obiettivi di investimento del fondo e la tolleranza al rischio.

Il rischio è inseparabile dal ricavo, ogni investimento comporta un certo grado di rischio, che è considerato vicino a zero nel caso di un titolo risk free o molto elevato per qualcosa come i mercati emergenti azionari o immobiliari in mercati altamente inflazionistici. Il rischio è quantificabile sia in termini assoluti che relativi. Una solida comprensione del rischio nelle sue diverse forme può aiutare gli investitori a comprendere meglio le opportunità, i compromessi e i costi legati ai diversi approcci di investimento.

L'obiettivo della gestione del rischio aziendale è quello di creare un quadro di riferimento che permetta alle aziende di gestire il rischio e l'incertezza, le quali sono presenti in quasi tutte le attività finanziarie ed economiche delle imprese. Il processo di identificazione, valutazione e gestione dei rischi fa parte dello sviluppo strategico delle aziende. Questo deve essere progettato e pianificato al massimo livello, vale a dire dal consiglio di amministrazione. Un approccio integrato di gestione del rischio deve valutare, controllare e monitorare tutti i rischi a cui l'azienda è esposta.

In generale, un rischio puro è una combinazione della probabilità o della frequenza di un evento e delle sue conseguenze, che di solito è negativa, può essere misurato dalla volatilità dei risultati, ma sono spesso necessari momenti più alti della distribuzione. L'incertezza è meno precisa perché la probabilità di un evento incerto è spesso sconosciuta, così come la sua conseguenza. In questo caso, ci riferiamo ad attività precauzionali, piuttosto che preventive, per proteggere contro l'incertezza. Infine, il rischio finanziario consiste nell'intraprendere attività opportunistiche relative a rischi futuri che possono generare risultati positivi o negativi.

La gestione del rischio societario è definita come un insieme di attività finanziarie o operative che massimizzano il valore di una società o di un portafoglio riducendo i costi associati alla

volatilità dei flussi di cassa. Le principali attività di gestione del rischio sono la diversificazione e la copertura del rischio, che si analizzano utilizzando vari strumenti, tra cui derivati e prodotti strutturati, assicurazione di mercato, autoassicurazione e autoprotezione. I costi principali che le aziende cercano di minimizzare sono: costi di difficoltà finanziarie, premio di rischio per i partner (stakeholder), imposte sul reddito attese, e il finanziamento degli investimenti. Il comportamento dei manager nei confronti del rischio (propensione al rischio e avversione al rischio) e la corporate governance influiscono anche sulla scelta delle attività di gestione.

Una gestione inadeguata del rischio può avere gravi conseguenze per le imprese, gli individui e l'economia. Un esempio è indubbiamente rappresentato dal crollo dei mutui subprime nel 2007, che ha contribuito a innescare la recessione. Questo crollo è derivato da cattive decisioni di gestione del rischio: creditori che hanno esteso i mutui a individui con scarso credito; società di investimento che hanno acquistato, confezionato, e rivenduto tali mutui; fondi che hanno investito eccessivamente nei titoli garantiti da ipoteche riconfezionati, ma ancora rischiosi (MBS).

1.2.1. Tipologie di rischio

Rischio di credito

Il rischio di credito è la possibilità di una perdita derivante da un evento creditizio, come il deterioramento delle condizioni finanziarie di un mutuatario, che fa perdere valore o perdere valore a un'attività (comprese le operazioni fuori bilancio).

La gestione del rischio di credito ha lo scopo di mantenere l'esposizione al rischio di credito a un livello ammissibile rispetto al capitale, di mantenere la solidità delle attività e di garantire rendimenti commisurati al rischio.

Senza una gestione efficace del rischio di credito, l'impatto delle perdite corrispondenti sulle operazioni può essere schiacciante.

Può anche fare riferimento al rischio di credito della società con i fornitori. Un'impresa assume un rischio finanziario quando fornisce finanziamenti di acquisti ai propri clienti, a fronte della possibilità che un cliente possa non pagare.

Una società deve gestire i propri obblighi di credito assicurandosi di avere sempre un flusso di cassa sufficiente per pagare i propri conti da pagare le fatture in modo tempestivo. In caso contrario, i fornitori possono interrompere l'erogazione del credito alla società o addirittura interrompere del tutto l'attività con la società.

Rischio di mercato

Il rischio di mercato è il rischio di mutazioni delle condizioni del mercato in cui un'impresa è in concorrenza. Un esempio di rischio di mercato è dovuto alla crescente tendenza dei consumatori a fare acquisti online, il che ha posto le imprese al dettaglio tradizionali davanti a notevoli sfide.

Le aziende che sono state in grado di apportare gli adeguamenti necessari per servire un pubblico di shopping online hanno prosperato e visto una crescita sostanziale delle entrate, mentre le aziende che sono state lente ad adattarsi o hanno fatto scelte sbagliate nella loro reazione al mercato in evoluzione hanno subito ingenti danni economici.

Questo esempio riguarda anche un altro elemento del rischio di mercato, il rischio di essere superato dai concorrenti. In un mercato globale sempre più competitivo e spesso con margini di profitto ristretti, le aziende con maggior successo finanziario sono solite offrire una proposta di valore unica, che le distingue dalla folla e dà loro una solida identità di mercato.

Rischio di liquidità

Il rischio di liquidità comprende la liquidità delle attività e il rischio di liquidità del finanziamento operativo. Per liquidità delle attività si intende la facilità relativa con cui un'impresa può convertire le proprie attività in liquidità in caso di necessità improvvisa e sostanziale di ulteriori flussi di cassa. La liquidità del finanziamento operativo è un riferimento al flusso di cassa giornaliero.

I cali generali o stagionali delle entrate possono presentare un rischio sostanziale se la società si trova improvvisamente senza sufficiente denaro a disposizione per pagare le spese necessarie al funzionamento dell'impresa. Questo è il motivo per cui la gestione dei flussi di cassa è fondamentale per il successo del business e perché analisti e investitori guardano a metriche come il flusso di cassa libero quando si valutano le aziende come un investimento azionario.

Operativo

I rischi operativi si riferiscono ai vari rischi che possono derivare dall'attività ordinaria di un'impresa. La categoria di rischio operativo comprende azioni legali, rischio di frode, problemi di personale e rischio del modello di business, che è il rischio che i modelli di marketing e di piani di crescita di un'azienda possano rivelarsi imprecisi o inadeguati.

2. Il controllo del rischio finanziario attraverso il machine learning

Negli ultimi anni, la quantità di dati raccolti dalle istituzioni finanziarie (IF) è aumentata in modo significativo, sia a causa della ingente quantità di informazioni richieste dal legislatore, sia a causa della digitalizzazione dei servizi, la quale contribuisce a creare una molteplicità di dati di consumo ad alta frequenza e non strutturati. Di conseguenza, le IF hanno un chiaro bisogno di strumenti analitici più potenti, in grado di far fronte a queste grandi quantità di dati di tutti i tipi e formati, e che allo stesso tempo riescano a mantenere o migliorare l'accuratezza dell'analisi.

Dopo la crisi finanziaria del 2008-2009, sono stati introdotti molti nuovi regolamenti e misure di vigilanza che hanno richiesto agli intermediari finanziari di riportare dati più dettagliati e più frequenti su più aspetti dei loro modelli di business e bilanci. Molto importanti sono gli stress test, che si basano su tutti i tipi di dati aziendali, compresi i dati del bilancio e gli aspetti qualitativi del business model.

Questi processi creano grandi quantità di dati che devono essere ben definiti e strutturati, aggregati e consegnati in tempo. Le autorità di regolamentazione hanno pertanto introdotto numerose iniziative per migliorare la qualità dei dati e la capacità degli istituti finanziari di fornirli. I principi del Comitato di Basilea per l'aggregazione dei dati di rischio (Basilea 239) fissano gli standard per le G-SIBS per migliorare i sistemi informatici e le strutture di rendicontazione. L'IFRS 9 mira a migliorare la qualità dei dati di vigilanza.

Oltre ai dati di segnalazione, le IF sono sempre più efficienti nel fornire grandi quantità di dati di bassa qualità. Questi includono output da applicazioni di consumo e altre interazioni digitali con i clienti, come metadati da sistemi di pagamento, e fonti di dati esterni, come i feed dei social media, che possono essere estratti per valutare lo sviluppo del mercato. Questo tipo di dati è tipicamente chiamato "big data".

Avendo praticamente ogni aspetto del business model delle IF regolamentato e sorvegliato con metriche di rischio dettagliate, la gestione di una banca, assicuratore, o asset manager sta diventando sempre più una questione di ottimizzazione entro queste centinaia di vincoli. Per competere in modo efficace hanno quindi bisogno di trovare il modello ottimale, anche estraendo i dati dei consumatori per informazioni dettagliate sulle preferenze e il comportamento dei clienti.

L'ampia gamma di approcci di machine learning è in grado di fornire questa potenza analitica in diversi contesti grazie alla sua capacità di analizzare set di dati estremamente grandi e in modo minuzioso. Per l'estrazione di dati di alta qualità, vengono tipicamente applicate tecniche di machine learning più convenzionali. Per estrarre invece i "big data" - ad alta frequenza, ma bassa qualità - sono più utilizzati il deep learning e le tecniche di rete neurale.

Il rischio di credito è la perdita economica derivante dal mancato adempimento dei propri obblighi contrattuali da parte di una controparte (ad esempio, pagamento tempestivo di interessi o di capitale) o dall'aumento del rischio di inadempimento durante la durata dell'operazione. Tradizionalmente, le società finanziarie hanno applicato le classiche regressioni lineari, logit e probit al rischio di credito del modello. Tuttavia, vi è ora un maggiore interesse da parte degli istituti ad utilizzare l'IA e le tecniche di apprendimento automatico per migliorare le pratiche di gestione del rischio di credito, in parte a causa di prove di incompletezza nelle tecnologie tradizionali. L'evidenza è che le capacità di gestione del rischio di credito possono essere significativamente migliorate sfruttando l'IA e le tecniche di apprendimento automatico grazie alla sua capacità di comprensione di dati non strutturati.

L'uso di IA e le tecniche di apprendimento automatico per modellare il rischio di credito non è un fenomeno nuovo, ma è ad oggi in crescente sviluppo. Nel 1994, Altman & Co. effettuarono una prima analisi comparativa tra i metodi statistici tradizionali di previsione fallimentare e un algoritmo di rete neurale alternativo, conclusero così che un approccio combinato delle due migliorò significativamente l'accuratezza.

In particolare, la crescente complessità della valutazione del rischio di credito ha aperto la porta al machine learning. Ciò è evidente nel crescente mercato dei credit default swap (CDS) dove vi sono molti elementi incerti che determinano sia la probabilità di un evento di default (evento creditizio) sia la stima del costo del default in caso di default. Un'importante scoperta venne resa pubblica nel 2016 da Son et al., utilizzando CDS giornalieri di diverse durate e di gruppi di valutazione diversi da gennaio 2001 a febbraio 2014 riuscirono infatti a dimostrare che i modelli di apprendimento automatico non parametrici che coinvolgono il deep learning danno risultati migliori dei modelli di riferimento tradizionali, in termini di accuratezza delle previsioni e nella proposta di misure pratiche di copertura.

I settori dei prestiti ai consumatori e delle PMI implicano grandi quantità di dati potenziali e fanno sempre più affidamento sull'apprendimento automatico per prendere decisioni migliori

in materia di prestiti. Vi è un sostanziale sostegno empirico per l'efficacia dell'apprendimento automatico. Nell'ambito del credito al consumo, Khandani et al. (2010) propongono una tecnica di apprendimento automatico basata su alberi decisionali e SVM che, se testata su dati reali di prestito, porta a risparmi fino al 25%.

2.1. Extreme gradient boosting

Il rischio di credito può essere misurato con modelli di Machine Learning (ML), in grado di estrarre relazioni non lineari tra le informazioni finanziarie contenute nei bilanci. In una scienza dei dati standard, i modelli sono scelti per ottimizzare l'accuratezza predittiva; mentre nei settori altamente regolamentati, come la finanza o la medicina, i modelli dovrebbero essere scelti in base alla precisione di bilanciamento con la capacità di spiegare. Miglioriamo la scelta selezionando i modelli in base alla loro accuratezza predittiva, e impiegando a posteriori un algoritmo che raggiunge l'esplorabilità. Questo non limita la scelta dei modelli più performanti.

Per esempio, consideriamo, senza perdita di generalità, il modello Extreme Gradient Boost, uno degli algoritmi di apprendimento automatico più popolari e veloci.

Extreme Gradient Boosting (Xgboost) è un modello supervisionato basato sulla combinazione di modelli ad albero con Gradient Boosting, una tecnica di ottimizzazione in grado di supportare diverse attività di apprendimento, come la classificazione e la previsione. Un modello ad albero è un modello di classificazione supervisionato che cerca la partizione delle variabili esplicative che meglio classificano una variabile di risposta (supervisore).

In pratica, un algoritmo di classificazione ad albero viene applicato a diversi “campioni di addestramento” del set di dati. In ogni iterazione, un campione di osservazioni è scelto tra i dati disponibili, utilizzando pesi di campionamento che cambiano nel tempo e ponderando maggiormente le osservazioni con la correlazione minore. Una volta che viene selezionata una “sequenza di alberi” adatta e che sono state svolte le classificazioni, viene attribuito un voto a maggioranza ponderata.

Confronto dei modelli di apprendimento

Una volta scelto un modello predefinito di stima della probabilità, esso deve essere misurato in termini di precisione predittiva, rispetto ad altri modelli, in modo da selezionare il migliore. L'approccio più comune per misurare l'accuratezza predittiva del credito dei modelli di punteggio è quello di dividere in modo casuale i dati disponibili in due parti: un "treno" e un insieme di "test"; il modello viene costruito utilizzando i dati del treno e vengono confrontate predizioni che il modello ottiene sull'insieme di prova, \hat{Y}_n , con i valori reali di \hat{Y} . Per ottenere \hat{Y}_n la probabilità di default stimata è arrotondata in un "default" o "non default", a seconda che una soglia sia passata o meno. Per una determinata soglia T , si può quindi contare la frequenza delle quattro possibili uscite, vale a dire:

1. Falsi Positivi (FP): le aziende previste di default, che non lo fanno;
2. Veri Positivi (TP): le aziende previste di default, che lo fanno;
3. Falsi Negativi (FN): le aziende previsto di non default, che fanno;
4. Veri Negativi (TN): le aziende hanno previsto di non default, che non lo fanno.

Il tasso di errori di classificazione di un modello può essere calcolato come segue:

$$\frac{FP + FN}{TP + TN + FP + FN}$$

Questo caratterizza la proporzione di previsioni errate tra il numero totale di casi.

Il tasso di errore di classificazione dipende dalla soglia scelta e non è, pertanto, uno strumento generalmente concordato di accuratezza predittiva. Una pratica comune è quella di utilizzare la curva Receiver Operating Characteristics (ROC), che traccia il tasso di falso positivo (FPR) sull'asse Y contro il tasso di vero positivo (TPR) sull'asse X, per un intervallo di valori soglia (solitamente valori percentili). FPR e TPR sono quindi calcolati come segue:

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

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

La curva ROC ideale coincide con l'asse Y, una situazione che non può essere realisticamente raggiunta. Il modello migliore sarà quello più vicino ad essa. La curva ROC è di solito riassunta con l'Area sotto il valore della curva ROC (AUROC), un numero compreso tra 0 e 1. Maggiore è l'AUROC, migliore è il modello.

Spiegare le previsioni dei modelli

Le informazioni contenute nelle variabili esplorative vengono utilizzate per localizzare e raggruppare la posizione di ogni individuo (azienda) nel campione, queste, insieme alle probabilità di default previste, permettono una spiegazione molto approfondita del determinante della solvibilità di ogni individuo. In questo specifico contesto, le informazioni sui vari elementi esplicativi sono desunte dai rendiconti finanziari delle società mutuarie, raccolti in un vettore x_n , che rappresenta la composizione finanziaria del bilancio dell'istituzione n .

Per prima cosa viene calcolato il valore Shapley¹ associato ad ogni azienda, in modo da poter utilizzare uno strumento agnostico in grado di interpretare in modo tecnologicamente neutro l'output di un modello di apprendimento automatico altamente accurato. I valori di Shapley di un modello possono essere utilizzati come strumento per trasferire inferenze predittive in uno spazio lineare in modo da potervi applicare una grande varietà di metodi statistici multivariati.

L'approccio Shapley viene sviluppato utilizzando il lavoro di struttura computazionale SHAP, che permette di calcolare i valori Shapley esprimendo previsioni come combinazioni lineari di variabili binarie che descrivono se ogni singola variabile è inclusa o meno nel modello.

Più formalmente, il modello di spiegazione $g(x^1)$ per la predizione $f(x)$ è costruito attraverso un metodo di attribuzione di caratteristica additiva, che scompone la predizione in una funzione lineare delle variabili binarie $z^1 \in \{0, 1\}^M$ e le quantità $\phi_i \in \mathbb{R}$:

$$g(z^1) = \phi_0 + \sum_{i=1}^M \phi_i z_i^1.$$

¹ Concetto di soluzione utilizzato per assegnare una ricompensa ad ogni giocatore presente in una coalizione, in funzione del contributo marginale che apporta ad essa

In altri termini, $g^1(z^1) \approx f(H_x(z^1))$ è un'approssimazione locale delle previsioni dove la funzione locale $H_x(x^1) = x$ mappa le variabili semplificate x^1 in x , $z \approx x$ e M è il numero delle variabili di input selezionate.

Infatti, Lundberg, S. and S.-I. Lee (2017), dimostrano che l'unico metodo di attribuzione di elementi aggiuntivi che soddisfa le proprietà di accuratezza locale, precisione e coerenza è ottenuto attribuendo a ciascuna caratteristica x^1_i un effetto ϕ_i chiamato valore di Shapley, definito come

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} [f_x(z') - f_x(z' \setminus i)]$$

dove f è il modello addestrato, x il vettore di input (caratteristiche), x^1 il vettore delle caratteristiche di input selezionate M . La quantità $f_x(z^1) - f_x(z/i)$ è il contributo di una variabile i ed esprime, per ogni singola previsione, la deviazione dei valori di Shapley dalla loro media.

In altre parole, un valore Shapley rappresenta una quantità unica in grado di costruire un modello esplicativo che localmente approssima linearmente il modello originale, per un input specifico x , (precisione locale), con la proprietà che, ogni volta che una feature è localmente zero, il valore Shapley è zero (missingness) e se in un secondo modello il contributo di una caratteristica è più alto, così sarà il suo valore Shapley (consistency).

Una volta calcolati i valori di Shapley, proponiamo di utilizzare reti di somiglianza, definendo una metrica che fornisce la distanza relativa tra le società, applicando la distanza euclidea tra ciascuna coppia (x_i, x_j) di vettori preconfigurati aziendali.

Si ricava quindi, utilizzando il metodo della rete di correlazione, la rappresentazione Minimal Spanning Tree (MST) delle società. L'MST è un albero senza cicli di una rete complessa, che unisce coppie di vertici con la "distanza" totale minima.

La scelta è motivata dalla considerazione che, per rappresentare tutte le correlazioni di coppia tra N aziende in un grafico, abbiamo bisogno di $N * (N - 1)/2$ spigoli, un numero che cresce rapidamente, rendendo il grafico corrispondente non comprensibile. Il Minimal Spanning Tree semplifica il grafico in un albero di $N - 1$ bordi, che prende $N - 1$ passi per essere completato.

Ad ogni passo, si unisce alle due aziende più vicine, in termini di distanza euclidea tra le variabili esplicative corrispondenti.

In questo contesto di valori Shapley, la somiglianza dei contributi variabili è espressa come matrice simmetrica di dimensione $n \times n$, dove n è il numero di punti di dati nel set di dati (treno). Ogni voce della matrice misura quanto siano simili o distanti due punti dati in termini di contribuzione. La rappresentazione MST associa ad ogni punto il suo vicino più vicino. Per generare l'MST viene usato l'algoritmo EMST Dual-Tree Boruvka e la sua implementazione nel pacchetto R "emstreeR".

La stessa matrice può anche essere usata, in una seconda fase, per un'ulteriore fusione dei nodi, attraverso l'analisi dei cluster. Questo passaggio aggiuntivo può rivelare segmentazioni di punti dati con contributi variabili molto simili, corrispondenti a simili decisioni di credit scoring

2.2. Il processo di credit scoring

L'obiettivo della classificazione è quello di assegnare le corrette etichette di classe ai nuovi elementi di un insieme di dati. Per le istituzioni finanziarie, i sistemi di classificazione, sotto forma di credit scoring, sono utilizzati quotidianamente per valutare i rischi di credito associati ai prestiti. Nel credit scoring, l'affidabilità creditizia di un cliente descrive la sua capacità e disponibilità a rimborsare un'obbligazione finanziaria. Il rischio di credito è definito come il rischio di perdita derivante da qualsiasi cambiamento reale o percepito nella solvibilità di un cliente.

In questa tesi sarà usato il termine credit scoring per descrivere l'insieme dei modelli di decisione e delle tecniche usate dai prestatori per classificare e valutare il rischio di credito presentato dai vari clienti.

L'obiettivo del credit scoring è quello di assegnare sia clienti esistenti che potenziali ad uno dei due gruppi: buono o cattivo. Si ritiene probabile che un membro del gruppo buono rimborsi il proprio obbligo finanziario mentre per un membro del gruppo cattivo è considerato più probabile un inadempimento. In generale, i modelli di credit scoring sono classificati in due diversi tipi, application scoring e behavioural scoring. L'applicazione di punteggio tenta di prevedere il rischio di default di un cliente al momento di una domanda di credito, sulla base di informazioni come i dati demografici del richiedente e le operazioni precedenti. Il punteggio

comportamentale valuta il rischio dei clienti esistenti in base alle loro recenti transazioni contabili.

2.2.1. Background

Nel retail banking, prima dell'uso di sistemi automatizzati di credit scoring, il rischio di credito di un richiedente veniva valutato in modo soggettivo sulla base dell'esperienza dei sottoscrittori. In genere, le informazioni sul cliente erano ottenute attraverso rapporti tra esso e il personale della filiale che frequentava. Il prestito è infatti stato per un lungo periodo un processo di giudizio in cui un dipendente (di solito il direttore della banca) valutava le domande sulla base di criteri noti come le 5C:

1. carattere - il richiedente o uno dei suoi familiari è noto all'organizzazione?
2. Capitale: quanto ha già depositato il richiedente e qual è l'importo del prestito richiesto?;
3. Garanzie (collateral): quali garanzie offre il richiedente?;
4. Capacità: qual è la capacità di rimborso del richiedente?;
5. Condizione: quali sono attualmente le condizioni generali dell'economia?

Ovviamente, tale processo decisionale di concessione del credito presentava una serie di carenze, in particolare per quanto riguarda la coerenza e l'affidabilità. Hand (2001) ha elencato le seguenti carenze chiave:

1. Tali decisioni vengono senza dubbio influenzate dai cambiamenti quotidiani nell'umore dell'operatore bancario;
2. Le decisioni non sono sempre replicabili in quanto i diversi manager non prendono le stesse decisioni;
3. Non vi è una formalizzazione universale del processo di decisione, che spesso rendea difficile l'insegnamento;
4. L'approccio basato sul giudizio umano può trattare solo un numero limitato di domande, con conseguente perdita di entrate.

I prestiti ai consumatori, tuttavia, sono aumentati notevolmente nella seconda metà del XX secolo:



America (USA) tra il 1960 e i primi sei mesi del 2021.

Con l'introduzione e la successiva popolarità delle carte di credito (emesse prima negli Stati Uniti nel 1958 e poi nel Regno Unito nel 1966), la domanda dei consumatori ha reso necessario lo sviluppo e la crescita di metodi oggettivi in grado di automatizzare la decisione di prestito. Durante questo periodo, a partire dagli Stati Uniti, anche la domanda di prodotti ipotecari aumentò a causa e grazie alle “politiche di homeownership” promosse dai governi americani dopo la Grande Depressione degli anni '30, come, ad esempio, la formazione della Federal National Mortgage Association (o Fannie Mae) come parte delle politiche del "New Deal" del Presidente F.D. Roosevelt

Insieme all'aumento della domanda pubblica di credito, le modifiche normative contribuirono anche a far avanzare la causa del credit scoring. Nell'interesse dell'equità e dell'uguaglianza, il Congresso degli Stati Uniti chiese che il processo decisionale in materia di concessione del credito fosse reso trasparente. L'Equal Credit Opportunity Act (ECOA), emanato per la prima volta nel 1974 negli Stati Uniti, e le successive modifiche hanno contribuito a rafforzare il requisito per il rating del credito. La ECOA vieta ai creditori di discriminare un richiedente sulla base di alcuni dettagli individuali proibiti (ad es. razza, colore o religione).

Per primi vennero proposti dei sistemi di credit scoring sviluppati statisticamente al fine di rispettare i requisiti normativi e il regolamento B1 della Federal Reserve statunitense, attuatrice dell'ECOA, il quale stabilisce che tali sistemi di rating del credito devono essere, tra l'altro:

1. "Costruiti sulla base di dati derivati da un confronto empirico di gruppi di campioni o della popolazione di richiedenti meritevoli di credito e non meritevoli di credito che hanno presentato domanda di credito in un lasso di tempo ragionevole";
2. "Sviluppati e convalidati utilizzando principi e metodologie statistiche accettate".

In tal modo, anche se i creditori sono costretti a rinunciare a parte della loro discrezionalità, i sistemi di credit scoring offrono ai prestatori una soluzione trasparente che soddisfa i requisiti della ECOA, in quanto fornisce delle motivazioni chiare ai clienti a cui non viene accettato il prestito.

I computer hanno poi fornito i mezzi necessari per attuare tali procedure in modo automatizzato. Rispetto ai sistemi di valutazione, ciò ha fatto sì che le banche al dettaglio riportassero riduzioni sostanziali del costo delle valutazioni del credito e delle perdite sui prestiti causate dai default dei clienti. Negli anni '80 con miglioramenti nella potenza computazionale (ad es. costo, velocità e capacità di archiviazione) le banche al dettaglio iniziarono a utilizzare metodi statistici per monitorare, rilevare, prevedere e comprendere molti aspetti del comportamento dei clienti. Gradualmente questo ha portato allo sviluppo di tecniche di stima per, tra le altre cose:

1. il rischio di default: la misurazione del rischio di default di un cliente su un prodotto particolare (product default scoring) o per qualsiasi prodotto (customer default scoring);
2. rivelazione delle frodi: tecniche in grado di rilevare le frodi nel più breve tempo possibile;
3. risposta a campagne pubblicitarie: engagement dei clienti
4. fidelizzazione del cliente
5. attrito: clienti insoddisfatti del servizio offerto
6. utilizzo del prodotto
7. profit scoring: tecniche per misurare la redditività di un cliente su un singolo prodotto (product profit scoring) e su tutti i prodotti (customer profit scoring).

Tipicamente, un sistema di credit scoring è implementato utilizzando una carta di credito. La Scorecard assegna punti, automaticamente, alle principali caratteristiche del cliente e agli aspetti della transazione al fine di ricavare un valore numerico che rappresenti il rischio che un cliente, rispetto ad altri, non adempirà ai propri obblighi finanziari. I sistemi di credit scoring non sono perfetti e possono solo stimare il rischio di credito in base alle performance passate infatti, ogni anno, a causa del fallimento dei sistemi di credit scoring nell'identificare gli individui inadempienti, una parte significativa di questi non viene pagata. Spesso, la causa di questo può essere attribuita a circostanze imprevedute come: frode, divorzio, ingenuità finanziaria e debito attraverso la perdita di reddito.



Il tasso di insolvenza (ad es. clienti non in pari con i loro rimborsi dei prestiti) per i prestiti residenziali negli Stati Uniti.

Di conseguenza, vi è un notevole interesse a migliorare i sistemi di valutazione del credito per discernere tra clienti redditizi e non redditizi sulla base del loro futuro comportamento di rimborso. Tra professionisti e ricercatori, è ampiamente riconosciuto che anche un piccolo miglioramento nella valutazione del rischio di credito dei clienti (come sta avvenendo dall'ultima recessione) può comportare notevoli risparmi finanziari.

2.2.1.1. L'accordo di Basilea sul capitale

Insieme con il risparmio finanziario offerto dal credit scoring ci sono anche questioni di regolamentazione a cui aderire. In alcuni paesi si ha una banca centrale che è responsabile della supervisione bancaria; altri paesi invece hanno organi di vigilanza bancaria separati e talvolta multipli. La Banca dei regolamenti internazionali è un'organizzazione internazionale incaricata di promuovere la cooperazione monetaria e finanziaria internazionale tra banche centrali. Il Comitato di Basilea per la vigilanza bancaria (BCBS) è un sottocomitato della BRI incaricato della responsabilità di elaborare orientamenti e raccomandazioni sulla regolamentazione bancaria attraverso l'evoluzione del l'accordo di Basilea sul capitale , il BCBS specifica uno standard internazionale per le banche da utilizzare nel calcolo del capitale necessario per compensare le perdite potenziali derivanti da rischi finanziari e operativi (ad es. l'importo delle liquidità e delle attività liquide che le banche devono accantonare per coprire le perdite impreviste).

2.2.2. Credit Scorecards

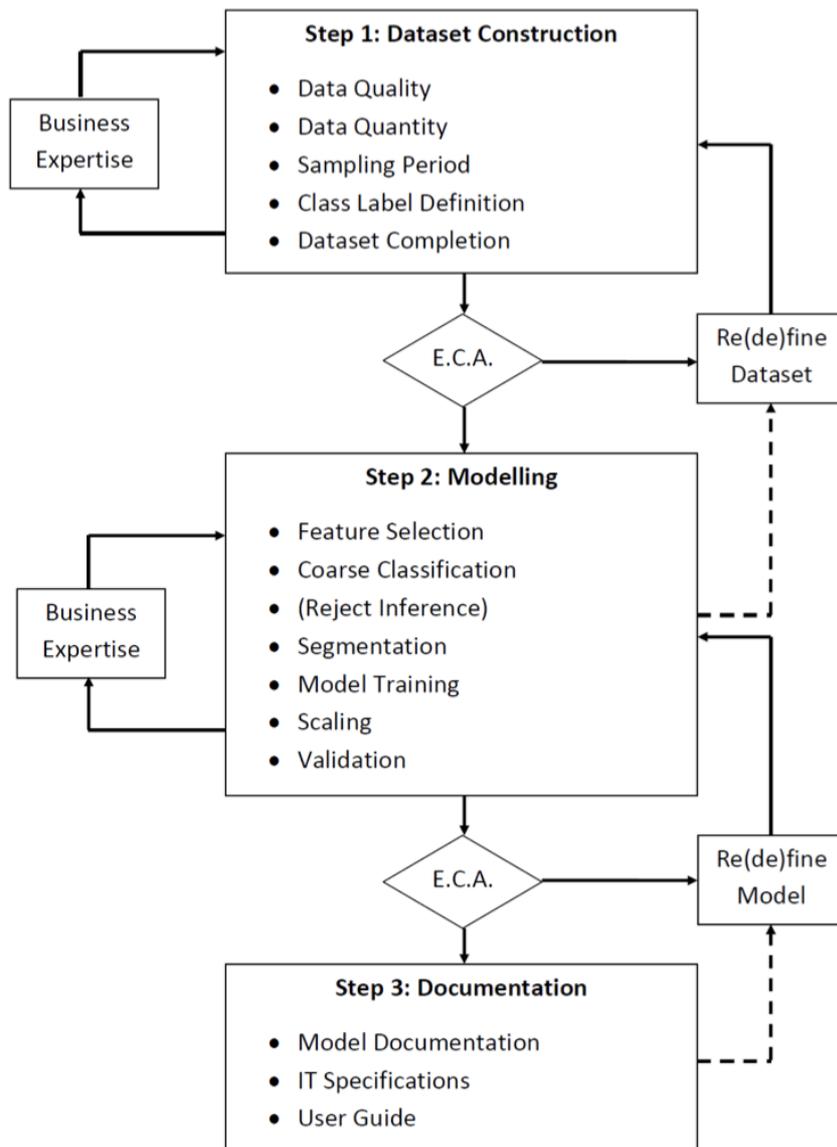
Nella sua forma più semplice, una scorecard (SC) consiste in un gruppo di caratteristiche determinate statisticamente per cercare di determinare la solvibilità di un cliente. Lo scopo delle SC è quello di consentire alle banche di utilizzare un formato strutturato, trasparente e di facile interpretazione con cui valutare l'affidabilità creditizia dei clienti. Nella tabella è riportato un esempio di scheda segnapunti dell'applicazione. La tabella è composta dalle “features” e dagli “attributes” del cliente. La prima colonna descrive una caratteristica particolare del mutuatario o del prestito e può essere selezionata da una qualsiasi delle fonti di dati a disposizione del mutuante. Le caratteristiche considerate, con mezzi statistici o in altro modo, vengono utilizzate per cercare di “predire” lo stato buono/cattivo dei clienti e sono incluse nella scheda di valutazione. Tipicamente, le caratteristiche vengono poi suddivise da un gruppo di uno o più “attributes”, caratterizzati da un insieme di valori mutuamente esclusivi o un intervallo di numeri non sovrapposti al quale sono assegnati un certo numero di punti che contribuiscono a un punteggio di credito complessivo. I punti assegnati ad un attributo si basano sull'analisi dei dati storici, che coinvolge vari fattori come la forza predittiva dell'elemento, la correlazione tra le caratteristiche, nonché considerazioni operative. Maggiore è il punteggio, minore è il rischio di inadempimento di un'obbligazione finanziaria.

La tabella include un esempio che mostra come sarebbe calcolato il punteggio di credito per un richiedente X. Sulla base dei dati in possesso della banca viene assegnato un punteggio di credito di 355 punti da un massimo di 443 della quale l'istituto dovrà decidere se e a quali condizioni approvare il prestito. La selezione di un punteggio di cut-off appropriato è una decisione strategica per la gestione che comporta un trade-off tra: il rischio atteso (i.e. tasso previsto “cattivo”) e il rendimento (profitto).

<i>Feature</i>	<i>Attribute</i>	<i>Points</i>	<i>Attribute value for applicant X</i>	<i>Points for applicant X</i>
Age	< 25	69		
	25 - 29	77		
	30 - 34	84	34	84
	35 - 41	93		
	42 - 50	104		
	50+	110		
Bank Customer	Yes	29	Yes	29
	No	20		
Credit limit on credit card	Blank	60		
	< 2,000	55		
	2,000 - 3,750	59	3,500	59
	3,751 - 6,000	64		
	6,001 - 10,000	71		
	> 10,000	74		
Years at current job	< 1	20		
	1 - 3	24		
	4 - 6	29	4	29
	7+	36		
Accommodation Status	Own	42		
	Rent	28	Rent	28
	Parents	32		
	Other	34		
Self-employed	Yes	25		
	No	41	No	41
Gross Monthly Income	< 2,500	71		
	2,500 - 3,150	79		
	3,151 - 3,850	85	3,750	85
	3,851 - 4,350	92		
	4,351 - 5,100	103		
	> 5100	111		
Score				355

Tabella: Scheda di valutazione della domanda con un punteggio di credito per il richiedente X

Anche se esiste una varietà di modi attraverso la quale può essere sviluppata una SC, il processo standard di sviluppo di queste consiste in una serie di fasi generiche come: costruzione di set di dati, modellizzazione e documentazione. Un modello di processo di sviluppo di SC è mostrato nella figura.



I principali compiti svolti nello sviluppo delle applicazioni e delle schede di valutazione comportamentale sono evidenziati per ciascuna delle tre fasi. Durante le fasi di costruzione e modellazione del set di dati, gli esperti statistici si consultano spesso con esperti aziendali per valutare le somiglianze e le differenze tra i risultati empirici e le conoscenze ed esperienze aziendali. Alla fine di ogni fase, è richiesta l'approvazione del comitato di esperti (ECA) prima di procedere alla fase successiva il quale è composto da sviluppatori di schede di valutazione

ed esperti aziendali che si riuniscono per esaminare lo stato di avanzamento del progetto di sviluppo e determinare se il lavoro precedente richiede affinamento.

Lo sviluppo di scorecard è un processo iterativo e la SC risultante deve soddisfare una serie di criteri di performance fondamentali, tra cui:

1. Stabilità: Gli attributi della SC devono essere stimati da un set di dati dimensionati che coprono un adeguato periodo storico;
2. Discriminatoria: Distinzione tra il “buono” e il “cattivo”;
3. Interpretabile - L'output della SC dovrebbe essere comprensibile e spiegabile ai non esperti;
4. Non eccessivamente complessa: La SC non dovrebbe ne fare troppo affidamento su una singola caratteristica ne su una quantità eccessiva;
5. Prudenziale: Basilea II richiede una stima, tra le altre cose, della probabilità di inadempimento di un portafoglio di prestiti al fine di calcolare l'importo del capitale da accantonare per coprire le perdite. Gli organismi di regolamentazione suggeriscono di sopravvalutare questa misura per coprire le recessioni dell'economia;
6. Robusta: La SC non deve presentare volatilità non necessaria durante il ciclo economico.

Non esiste un quadro quantitativo definitivo per garantire il rispetto dei criteri, poiché il processo decisionale di ciascuna organizzazione finanziaria varia a causa di una serie di fattori quali le risorse disponibili e la cultura contemporanea/regionale. Sebbene misure di regolamentazione come gli Accordi di Basilea forniscano orientamenti, in ultima analisi, è necessaria una fusione di competenze statistiche, legali, informatiche, commerciali e di pianificazione.

Come sottolineato in precedenza, lo sviluppo della scheda di valutazione si basa sul completamento di ciascuna delle tre fasi descritte nella figura. Di particolare interesse per questa ricerca sono la costruzione di dataset e le fasi di modellazione, mentre a fase di documentazione non sarà ulteriormente studiata in questo lavoro poiché meno collegata all'utilizzo del machine learning. Ciò però non significa che essa sia irrilevante in quanto è importante registrare i dettagli delle informazioni utilizzate, delle ipotesi formulate e delle specifiche utilizzate per l'attuazione.

2.2.2.1. Costruzione del set di dati

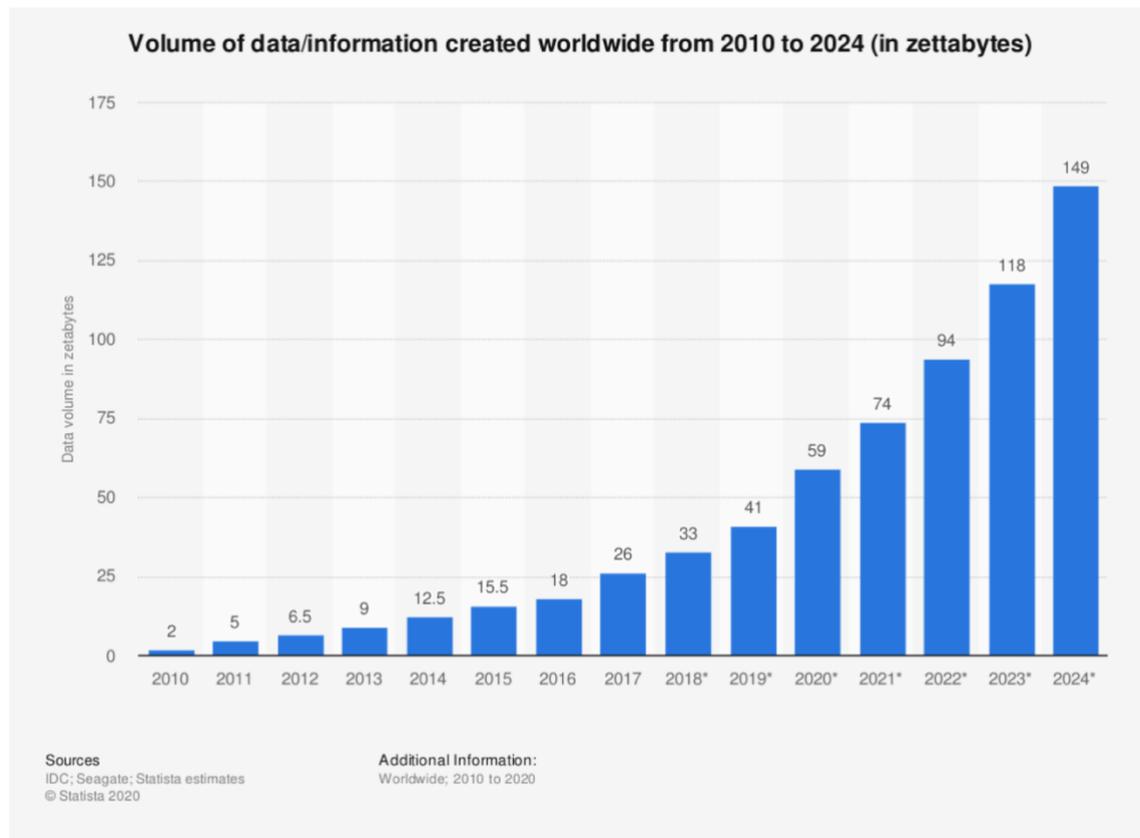
Qualità dei dati

Il successo di una scorecard nel distinguere tra domande “buone” e “cattive” dipende in gran parte dai dati utilizzati durante la fase di sviluppo della SC. La prima questione da affrontare riguarda la qualità dei dati disponibili, la quale è definita da molte caratteristiche, tra cui accuratezza, completezza e coerenza. La prima si riferisce al grado di precisione delle misurazioni di un elemento rispetto al suo valore reale. Esempi comuni tra le cause tipiche di scarsa accuratezza dei dati sono: errori di input dell'utente ed errori nel software. La completezza dei dati si riferisce alla misura in cui mancano dei valori. La coerenza dei dati si riferisce a situazioni in cui sono utilizzate più fonti e, per mancanza di standardizzazione, due o più voci di dati possono entrare in conflitto tra loro. Anche se ci sono vari metodi utilizzati per gestire dati mancanti o incompleti, l'approccio più semplice è quello di rimuovere le voci interessate dal set di dati.

Quantità di dati

Per garantire la costruzione di una scheda di valutazione di alta qualità e robusta, è necessaria una quantità sufficiente di dati dei clienti. Durante questa fase vengono individuate le fonti di dati e vengono stabilite le linee guida su come reperire i dati. La quantità di dati necessari dipende dall'obiettivo della scorecard e dalle proprietà dei dati stessi. Tradizionalmente, nel punteggio delle applicazioni, gli esperti del settore raccomandavano l'uso di 1.500 istanze di ogni classe e, in caso di esito negativo, l'aggiunta di ulteriori 1500

Anderson (2007) attribuisce questi numeri agli anni '60, un'epoca in cui il potere computazionale era limitato e la raccolta di dati era più costosa. Oggi, queste quantità sono ancora ampiamente utilizzate e per molti professionisti la validità delle raccomandazioni si basa sulla consapevolezza che la composizione dei set di dati di credit scoring è omogenea tra prestatori e regioni.



Periodo di campionamento

Come accennato in precedenza, le SC sono costruite utilizzando dati storici. Sebbene i risultati passati non garantiscano i risultati futuri, nel credit scoring la storia è un indicatore abbastanza affidabile. Per generare i set di dati di addestramento, viene effettuata un'istantanea di ogni cliente in due diversi momenti. Il primo all'inizio del prestito quando vengono registrate le caratteristiche dei clienti. Il secondo qualche tempo dopo al punto di osservazione predefinito e a quel punto il cliente è classificato come “buono” o “cattivo”. Il periodo di tempo tra le due istantanee è comunemente indicato come la finestra dei risultati, durante questa fase viene specificata la dimensione della finestra dei risultati, in base agli obiettivi aziendali della scheda di valutazione. Ad esempio, una finestra di risultato breve (ad esempio sei mesi) può essere utilizzata quando l'obiettivo è rilevare i default al primo pagamento mancato senza tener conto della possibilità che i mutuatari possano recuperare. Al fine di limitare le possibilità di classificare in modo sbagliato un cliente e per evitare di sottostimare i creditori, l'istituto deve individuare una finestra di risultato della durata adeguata poiché, se questa risultasse troppo lunga, è possibile che si verificano differenze tra il campione utilizzato durante lo sviluppo

della scheda di valutazione e i campioni futuri non ancora visti. Tali differenze possono derivare da cambiamenti nelle condizioni macroeconomiche, nella strategia aziendale e nelle circostanze personali. Se, invece, la finestra dei risultati fosse troppo breve, informazioni preziose potrebbero essere perse, per esempio, alcuni eventi predefiniti potrebbero non essersi ancora verificati. In genere, per i mutui, la finestra di risultato è identificata tracciando il tasso di default cumulativo mensile il quale è calcolato come il numero totale di mutuatari diviso per il numero totale di inadempimenti. Un plateau nel tasso di default cumulativo mensile indica la scadenza del campione di dati. Tuttavia, non è raro selezionare una finestra di risultato più corta di questo periodo, a condizione che il potere discriminatorio e l'intuitività della SC o non sia influenzato.

Definizione etichetta della classe

Il modo in cui un conto di prestito è definito come “cattivo” dipende dagli obiettivi del sistema di punteggio e dalla visione dell'istituto finanziario di successo o fallimento. In genere, nel credit scoring, un conto di prestito è etichettato come cattivo quando si è verificato un default. Secondo la definizione di Basilea II (paragrafo 452), si ritiene che l'inadempimento abbia avuto luogo quando sono soddisfatti uno dei seguenti criteri o entrambi:

1. il mutuatario è in ritardo nel pagamento di oltre 90 giorni per una qualsiasi obbligazione creditizia rilevante nei confronti del mutuante.
2. il prestatore ritiene che il mutuatario non sia in grado di rimborsare integralmente i propri obblighi di credito nei confronti del prestatore, senza che quest'ultimo ricorra ad azioni quali il realizzo di garanzie (se detenute), ad es. riacquisto di immobili in caso di default del prestito

Le istituzioni finanziarie possono scegliere tra un approccio detto dello “stato attuale”, che classifica un cliente come buono o cattivo in base allo stato del conto alla fine della finestra dei risultati, e un approccio detto dello “stato peggiore”, che classifica un cliente come buono o cattivo in base allo stato del suo account durante la finestra dei risultati.

Completamento del set di dati

La fase finale della costruzione del dataset prevede la suddivisione dei dati in due parti: il campione di allenamento e il campione di prova. Il primo viene utilizzato per costruire la scheda di valutazione mentre il secondo studia la performance delle SC. Ci sono vari modi per dividere i set di dati di allenamento e prova. Di norma, se sono disponibili dati sufficienti, i costruttori

di schede di valutazione optano per l'approccio basato sulla resistenza con una ripartizione 70:30 tra le dimensioni dei due campioni. Invece, nell'approccio di hold-out, una parte del campione di addestramento, chiamato campione di convalida, viene accantonato per regolare i parametri del modello della scorecard

Se i dati disponibili sono insufficienti, possono essere utilizzati approcci statistici standard quali la cross-validation e il bootstrapping per stimare i parametri del modello senza perdere troppe informazioni.

2.2.2.2. Modellazione

Dopo aver generato i set di dati di addestramento e di prova, la scheda di valutazione può essere sviluppata nella fase di modellazione. Durante questa fase si devono attraversare diversi stadi, ognuno delle quali è descritta qui di seguito. Gran parte della discussione che segue si basa sull'ipotesi che la regressione logistica sia forse l'algoritmo più comunemente utilizzato nell'industria del credito al consumo.

Selezione dei “features”

La selezione dei “features” è il processo di scelta di un sottoinsieme tra le caratteristiche disponibili da utilizzare in una scorecard eliminando quelle che sono ridondanti o non completamente affidabili.

Nel credit scoring di solito vi è un'ampia gamma di “features” utilizzabili derivanti dalla varietà di fonti utilizzate per registrare e una scorecard ben strutturata ne utilizza tipicamente tra le 10 e le 20. Sono presenti una serie di validi motivi per effettuare la selezione delle caratteristiche durante la costruzione della SC. In primo luogo, da un punto di vista pratico, per ridurre i costi è importante eliminare il maggior numero possibile di informazioni irrilevanti e ridondanti. In caso contrario, il personale viene retribuito per analizzare e comprendere funzionalità aggiuntive. In secondo luogo, l'identificazione delle caratteristiche predittive aiuta a fornire una visione più chiara e una migliore comprensione della scheda di valutazione. Infine, applicando il principio del rasoio di Occam, una semplice scorecard con un'accuratezza predittiva ottimale è preferibile ad una scorecard più complessa che include molte funzionalità non necessarie. La “maledizione della dimensionalità” è il termine usato quando troppe caratteristiche irrilevanti e ridondanti e poche istanze descrivono la popolazione target; ciò può risultare in sovra-adattamento, per cui il modello indotto classifica accuratamente le circostanze nel campione di

allenamento, comprese quelle rumorose, ma si comporta male se applicato ad un campione precedentemente invisibile.

Classificazione “grossolana”

Dopo che il numero di funzioni è stato ridotto a un livello gestibile, il passo successivo è quello di trasformare i dati in una forma appropriata per il processo di modellazione della scorecard. Il metodo più comunemente usato nel credit scoring è la classificazione “grossolana”. In questa fase gli attributi delle caratteristiche “categoriche” e “ordinali” sono spesso aggregati in un numero minore di categorie, ciò consente di trattare ciascuna categoria di ciascuna caratteristica come una caratteristica fittizia avente un proprio peso nel modello di regressione logistica standard del settore. Al contrario, con una caratteristica continua, si stima un singolo coefficiente di regressione che potrebbe non catturare adeguatamente la relazione non lineare dell'elemento con l'etichetta della classe.

La classificazione grossolana aumenta la robustezza di una SC riducendo la possibilità di sovrapposizioni e creando categorie con un numero sufficiente di osservazioni “buone” e “cattive”. Un altro vantaggio si verifica quando alcune caratteristiche presentano una relazione non monotona con la probabilità di default. Ad esempio, considerare la funzione "tempo di vita con i genitori", negli Stati Uniti, le statistiche suggeriscono che i mutuatari di età superiore ai 30 anni che vivono con i loro genitori sono gli elementi con un maggiore rischio di default, in quanto la norma per lasciare casa è circa a quell'età. Tuttavia, se il mutuatario ha già lasciato la casa di famiglia, può indicare una mancanza di risparmio dovuta al pagamento dell'affitto e di altre bollette domestiche e quindi può aumentare il rischio di default. La classificazione grossolana è usata per accomodare tali relazioni non lineari generando diverse categorie separate, ciascuna delle quali ha relativo peso nel modello standard di regressione logistica. Un ulteriore vantaggio della classificazione grossolana è la sua capacità di incorporare i valori mancanti utilizzando una categoria separata. Analogamente, le instabilità causate da valori anomali e valori estremi possono essere affrontate, in parte, aggregando tali valori in una categoria a sé stante. La classificazione grossolana è anche indicata come “binning”, raggruppamento e discretizzazione.

Respingere l'inferenza

Una volta che viene costruita la scorecard, il valore del risultato (es. buono o cattivo) è disponibile solo per i clienti a cui viene effettivamente concesso il credito, mentre, per i clienti a cui questo viene rifiutato, vengono conservati solamente i valori caratteristici, e non i risultati finali. Si tratta di una forma di “pregiudizio di campione”, spesso indicata come “pregiudizio di rifiuto”, in cui il database dei clienti della banca non è rappresentativo della popolazione nella sua interezza. Molto importante è cercare di risolvere questo pregiudizio stimando come i richiedenti respinti si sarebbero comportati se fossero stati accettati. Utilizzando tecniche di inferenza di rifiuto, gli istituti di credito tentano di migliorare la discriminazione della SC e di fornire una stima accurata delle prestazioni della SC sulla popolazione effettiva a cui sarà applicata, piuttosto che a solo la popolazione accettata.

L'approccio più semplice per affrontare il “pregiudizio di rifiuto” è quello di ottenere le caratteristiche del cliente e il risultato per l'intera popolazione richiedente, concedendo credito a ogni richiedente durante un certo periodo di tempo. Tradizionalmente, i dettaglianti e le società di vendita per corrispondenza hanno utilizzato questo approccio, tuttavia, per molte banche questo risulta essere finanziariamente insostenibile date le probabili perdite.

Diversi approcci di inferenza di rifiuto sono stati sviluppati per evitare questo pregiudizio. Un modo semplice è quello di designare semplicemente ogni candidato respinto come “cattivo”; ovviamente, un inconveniente di questo approccio è che rafforza il pregiudizio delle decisioni precedenti, dato che alcuni gruppi di clienti potrebbero essere etichettati come non affidabili senza la possibilità di smentire questa ipotesi. Due dei più comuni approcci all'inferenza di rifiuto sono l'estrapolazione e l'accrescimento e ci sono diverse varianti per ogni approccio. L'estrapolazione è un metodo relativamente semplice che stima un modello preliminare utilizzando solo i candidati accettati. Successivamente, questo modello è utilizzato per estrapolare la probabilità di inadempimento per i richiedenti respinti, che viene utilizzato per imputare una classificazione buono-cattivo ai richiedenti respinti sulla base di una probabilità di cut-off. Infine, si stima che un nuovo modello utilizzi sia i richiedenti respinti che quelli accettati. Un altro approccio è quello dell'aumento (noto anche come ri-ponderazione); si stima in questo caso un modello utilizzando i richiedenti accettati, ma ogni richiedente è ponderato dall'inverso della probabilità di essere accettato. Per calcolare questa probabilità inversa si stima un secondo modello utilizzando sia i richiedenti accettati che quelli respinti e questo modello cerca di prevedere quali candidati saranno accettati. Sulla base di tale approccio, al fine di

simulare la presenza di applicazioni scartate, un peso sproporzionatamente più elevato è dato ai clienti più marginali.

Segmentazione

Nella modellazione scorecard è importante decidere se segmentare o meno la popolazione e quindi costruire diverse SC per ogni segmento. La segmentazione viene eseguita dividendo la popolazione in diversi gruppi e costruendo SC separate per ogni gruppo. Nel marketing la segmentazione viene regolarmente utilizzata per raggruppare i clienti in gruppi omogenei sulla base dei loro modelli di ricerca e delle informazioni demografiche, come, tra gli altri, il reddito e l'età. Nel credit scoring, lo scopo della segmentazione è quello di migliorare il livello di discriminazione della SC e consentire una maggiore flessibilità al prestatore per quanto riguarda le configurazioni dei prodotti come il tasso di interesse, la struttura di rimborso e altri requisiti di questo tipo. La costruzione e la manutenzione di schede di valutazione aggiuntive comporta un grande lavoro e richiede un'attenta considerazione per limitare il numero di segmenti. Le tre considerazioni che influenzano la decisione di segmentare i dati sono operative, statistiche e strategiche.

Model Training

Con la creazione e l'elaborazione dei set di dati di training e testing può iniziare la formazione del modello predittivo. È prassi standard nel settore di utilizzare la regressione logistica in questa fase dello sviluppo della scorecard. Come evidenziato in precedenza, la regressione logistica è forse l'algoritmo più comunemente utilizzato nell'industria del credito al consumo ma ovviamente può essere presa in considerazione anche qualsiasi tecnica di modellizzazione predittiva. Un modello predittivo viene addestrato utilizzando il set di dati di addestramento e il set di dati di prova viene utilizzato per valutare l'accuratezza del modello. Nonostante sia importante che il modello separi il “bene” e il “male”, è anche necessario considerare come il modello si adatta ai dati al fine di evitare problemi come sovra-montaggio. A seconda delle prestazioni del modello predittivo, può essere infatti necessario rivedere i set di dati rivedendo le fasi precedenti a partire da quella della costruzione del set di dati. Le prestazioni sono spesso misurate utilizzando il coefficiente di Gini e la statistica di Kolmogorov-Smirnov (KS)

Scaling

Lo scorecard scaling è usato per trasformare l'output di un classificatore predittivo in un punteggio che rappresenta un particolare rapporto buono-cattivo. Il ridimensionamento non influisce sulla forza predittiva della scorecard, ma è invece un esercizio di forma eseguito, principalmente, per migliorare la facilità di comprensione e interposizione di una SC agli utenti non esperti. Un sondaggio condotto da Thomas nel 2001 ha identificato un certo numero di proprietà desiderabili SC tra cui:

1. il punteggio totale è positivo;
2. i punti per ogni attributo sono positivi;
3. la presenza di punteggi di riferimento;
4. le differenze tra i punteggi hanno un valore costante in tutta la SC

Convalida

Prima dell'attuazione, la carta di credito deve essere valutata per coerenza e accuratezza. Questo processo è denominato convalida ed è di solito intrapreso da una parte indipendente dal processo di sviluppo ed eseguito utilizzando dati che non sono stati utilizzati nella fase di modellazione. La convalida è un'attività in corso, quindi, viene eseguita non solo dopo lo sviluppo della scheda di valutazione, ma anche su base periodica, soprattutto quando si sono verificati cambiamenti strutturali significativi della scheda di valutazione. Un requisito di Basilea II è, infatti, un processo adeguato al fine di convalidare le stime della probabilità di default (PD), e questo coinvolge il calcolo del potere discriminatorio e la calibrazione.

Il potere discriminatorio si riferisce al grado di separazione delle distribuzioni di punteggi tra “bene” e “male”. Per misurare la potenza di un modello sono normalmente utilizzate una “matrice di confusione” e le misure delle prestazioni che ne derivano. Ulteriori strumenti comunemente usati per valutare i modelli di credit scoring includono curve di potenza (ad es. curva ROC, statistica Kolmogorov-Smirnov) e grafici Cumulative Accuracy Profile (CAP) che illustrano graficamente la potenza di vari modelli su un set di dati.

La calibrazione si riferisce all'accuratezza delle PD, secondo Basilea II, i prestatori devono calibrare la PD di ciascuna classe di rating per garantire che i tassi di default effettivi rientrino nell'intervallo previsto. L'obiettivo della calibrazione della PD è determinare se l'entità della

differenza tra le PD stimate e i tassi di default osservati sia significativa. Tale procedura è comunemente indicata come backtesting.

Conclusione

Il corretto funzionamento e il perfezionamento dei sistemi di valutazione del credito è un argomento ovvio di interesse per le banche, i clienti e le autorità di regolamentazione - come hanno dimostrato i recenti eventi dell'economia mondiale. Ad esempio, ai sensi dell'accordo sul capitale di Basilea, le banche sono ora tenute a fornire alle autorità di regolamentazione competenti una stima accurata della probabilità di inadempimento (PD) dei clienti come parte della stima dei requisiti patrimoniali minimi delle banche.

Nonostante gli ingenti investimenti fatti e l'impegno di istituzioni e imprese, come abbiamo visto in questo capitolo, lo sviluppo delle Scorecard è ancora un processo difficile e molto dettagliato che richiede attenzione a variate sfaccettature. Le tendenze demografiche e gli eventi economici possono creare scenari per i quali le tecniche standard accettate di sviluppo della scheda di valutazione risultano essere inadeguate. Una tale sfida si presenta quando troppo pochi default si verificano nella popolazione campione, presentando difficoltà nella costruzione di una scorecard robusta e affidabile. Un'altra sfida si verifica nel punteggio comportamentale quando i costruttori di SC devono decidere come definire un default e su quale periodo di tempo (i.e. il numero di mesi) basarsi per prevedere il comportamento del cliente. Infine, per molti accademici ottenere dati effettivi di credit scoring è un compito difficile a causa delle leggi sulla privacy dei dati e delle sensibilità commerciali. Utilizzando dati artificiali gli studiosi possono superare queste restrizioni e creare condizioni specifiche in cui indagare problemi specifici ed è questa la soluzione su cui si sta maggiormente concentrando in questo periodo.

3. Machine learning applicato agli altri tipi di rischio

3.1. Applicazione al rischio di mercato

Il rischio di mercato è il rischio derivante dall'investimento, dalla negoziazione e generalmente dall'esposizione ai mercati finanziari. Kumar (2018) fornisce una panoramica strutturale di come l'apprendimento automatico può aiutare nella gestione del rischio di mercato, facendo notare i benefici in ogni fase, dalla preparazione dei dati, alla modellazione, allo stress test, e fornendo un percorso di verifica per la spiegazione del modello.

La negoziazione sui mercati finanziari comporta intrinsecamente il rischio che il modello utilizzato per la negoziazione sia falso, incompleto o non sia più valido. Questo settore è generalmente noto come modello di gestione del rischio. Il machine learning è particolarmente adatto a testare i modelli di mercato per determinare il rischio involontario o emergente nel comportamento di trading. Woodall (2017) descrive una varietà di casi di uso corrente di apprendimento automatico per la convalida dei modelli, compresa la società d'investimento francese Nataxis. Questa, al momento della stesura della relazione, effettuava più di tre milioni di simulazioni a notte, utilizzando l'apprendimento non supervisionato per stabilire nuovi modelli di connessione tra gli attivi e per indagare ulteriormente su eventuali simulazioni emerse dal test che mostravano un andamento irregolare rispetto alle stime medie. Un'applicazione corrente interessante del modello di gestione del rischio è la ditta yields.io che fornisce il monitoraggio in tempo reale del modello, la prova del modello per le deviazioni e la convalida del modello, tutti guidati da IA e dalle tecniche di apprendimento automatico.

Un'altra area di particolare interesse nella categoria del rischio di mercato, importante per le grandi imprese commerciali, è la comprensione dell'impatto della loro contrattazione sui prezzi. Day (2017) esplora il modo in cui le grandi società commerciali utilizzano l'IA, e in particolare le tecniche di clustering, per evitare i costi derivanti dalle ricerche di scambi da effettuare nei mercati illiquidi. Con riferimento a Capital Fund Management, uno dei più grandi fondi hedge in Francia con undici miliardi di dollari in gestione, egli sostiene che fino a due terzi del loro profitto derivante dal commercio possono essere persi a causa dei costi di impatto del mercato. Le tecniche di apprendimento automatico aiutano in modo significativo le imprese che si trovano in questa situazione, identificando le connessioni tra le attività che non sono facilmente osservabili e quindi permettono di entrare nelle posizioni desiderate attraverso una serie di

attività correlate, piuttosto che prendere una grande posizione in una singola attività. L'analisi dei cluster aiuta in particolare a questo proposito, così come i modelli di deep learning.

3.2. Applicazione al rischio operativo

La gestione del rischio operativo implica che l'impresa cerchi di individuare il rischio di perdite finanziarie dirette o indirette derivanti da una serie di disaggregazioni operative potenziali. Questi rischi possono essere interni alle istituzioni (ad esempio, processi interni inadeguati o falliti, persone e sistemi), o da eventi esterni (ad esempio, frodi, sistemi informatici vulnerabili, mancato controllo, errore operativo, procedura trascurata o calamità naturali). Dato l'aumento di quantità, varietà e complessità delle esposizioni al rischio operativo, in particolare per le società finanziarie, si è aperta una strada verso l'IA e soluzioni basate sull'apprendimento automatico.

L'IA può assistere gli istituti in varie fasi del processo di gestione del rischio, che vanno dall'identificazione dell'esposizione al rischio, alla misurazione, alla stima e alla valutazione dei suoi effetti; può anche aiutare ad optare per un'adeguata strategia di attenuazione dei rischi e a trovare strumenti che possano facilitare il cambiamento o il rischio di negoziazione. Così, l'uso di tecniche di IA per la gestione del rischio operativo, che ha iniziato con il tentativo di prevenire le perdite esterne come le frodi con carta di credito, si sta ora espandendo in nuovi settori che comportano l'analisi di ampie raccolte di documenti e l'esecuzione di processi ripetitivi, nonché l'individuazione del riciclaggio di denaro che richiede l'analisi di grandi serie di dati.

L'individuazione della frode finanziaria è un altro caso di uso di gestione del rischio comunemente citato per l'apprendimento automatico e l'IA. Le banche cercano infatti di controllare le frodi finanziarie attraverso la valutazione dei modi migliori per proteggere i loro sistemi, i loro dati, e, in ultima istanza, i loro clienti. La capacità dell'IA di introdurre una migliore automazione dei processi può accelerare il ritmo delle attività di routine, minimizzare l'errore umano, elaborare dati non strutturati per schermare contenuti rilevanti o notizie negative e determinare la connessione degli individui per valutare i clienti e le reti a rischio. Questa stessa analisi di rete può essere utilizzata anche per monitorare dipendenti e operatori.

Le tecniche di raggruppamento e di classificazione possono essere utilizzate per stabilire profili di operatori basati sul comportamento, in cui le combinazioni di dati commerciali, registrazioni

di comunicazioni elettroniche e vocali consentono alle banche di osservare i modelli emergenti di comportamento, prevedere i rischi latenti e individuare i collegamenti tra i dipendenti. Esso consente inoltre alle banche di generare e dare la priorità agli allarmi sulla base dei tipi di attività sospette e del livello di rischio. Ngai et al. (2011) forniscono un'eccellente panoramica delle principali tecniche di IA utilizzate nel rilevamento delle frodi finanziarie, e notano le principali tecniche applicate come alberi di decisione e reti neurali.

Come applicazione pratica, cinque delle più grandi banche nordiche si sono unite per creare una struttura comune contro il riciclaggio di denaro, nota come Nordic KYC Utility. Un'infrastruttura basata sull'IA aiuterà a rispettare le normative e i requisiti relativi alle normative KYC (Know Your Customer) ed evitare l'imposizione di multe da parte dei regolatori. Allo stesso modo, HSBC sta introducendo la tecnologia AI, sviluppata dalla società di analisi dei dati Quantexa, per monitorare i loro processi antiriciclaggio. Ci sono sforzi concreti anche nella prevenzione delle frodi. Ad esempio, una joint venture della Royal Bank of Scotland e della Vocalink nel Regno Unito sta creando un sistema di apprendimento automatico per scansionare le transazioni di piccoli e grandi clienti commerciali per identificare e aggirare fatture false e potenziali casi di frode. Uno studio di Colladon e Remondi (2017) utilizzando dati reali di 33.000 transazioni di una società di factoring italiana mostra l'efficacia di tale analisi nel rilevamento delle frodi.

3.3. Applicazione a Regtech

Il rispetto delle norme sulla gestione dei rischi è una funzione fondamentale per le imprese finanziarie, in particolare dopo la crisi finanziaria. Mentre i professionisti della gestione del rischio cercano spesso di tracciare una linea di demarcazione tra ciò che fanno e la necessità, spesso burocratica, del rispetto delle normative, i due aspetti sono inequivocabilmente legati in quanto entrambi riguardano i sistemi globali di gestione del rischio. In tal senso, la conformità è forse meglio legata alla gestione del rischio delle imprese, anche se riguarda specificamente ciascuna delle funzioni di rischio del rischio di credito, del mercato e del rischio operativo.

Regtech, un'area analoga alla Fintech, si è così sviluppata per assistere le imprese nel l'aumento delle richieste di conformità. In questo settore l'IA e l'apprendimento automatico stanno iniziando a svolgere un ruolo significativo guidato dal volume di dati che devono essere valutati e dalla natura non convenzionale di questi dati. Il vantaggio chiave del machine learning in un puro senso Regtech è la capacità di un monitoraggio continuo delle attività aziendali. Arner et

al. (2016) hanno notato questa capacità per approfondimenti in tempo reale, sono riusciti quindi ad evitare le violazioni di conformità, invece che affrontare le conseguenze delle violazioni dopo che queste si sono verificate. Altri vantaggi sono la possibilità di liberare capitale regolamentare grazie ad un miglior monitoraggio, come l'automazione, apportando così una riduzione al budget di circa 70 miliardi di dollari che le principali istituzioni finanziarie spendono ogni anno per la conformità.

Un attore chiave in questo campo è IBM, che dopo l'acquisto di Promontory (una startup Regtech con 600 dipendenti) offre una gamma di soluzioni AI-driven per ridurre i costi Regtech, dimostrando l'interesse diffuso del settore non solo per le startup. Ad esempio, l'analisi della conversazione vocale in tempo reale per essere utilizzato per garantire la conformità. Si tratta di tradurre le conversazioni vocali in testo e poi classificare questo testo utilizzando l'elaborazione del linguaggio naturale in categorie che identificano la potenziale non conformità. Altre applicazioni del machine learning includono la lettura automatica e l'interpretazione delle implicazioni della documentazione normativa, sempre utilizzando il trattamento del linguaggio naturale.

4. Le sfide e il futuro dell'apprendimento automatico nella gestione del rischio

Ci sono alcuni problemi pratici significativi che devono ancora essere affrontati prima che l'IA e le tecniche di apprendimento automatico per la gestione del rischio possano rivendicare il loro pieno potenziale. La più importante di queste è la disponibilità di dati adeguati. Nonostante i software di apprendimento automatico per Python e R possano facilmente leggere tutti i tipi di dati da Excel a SQL e possano eseguire l'elaborazione del linguaggio naturale e delle immagini, la velocità con cui le soluzioni di apprendimento automatico sono state proposte è stata maggiore dell'abilità delle aziende di organizzare adeguatamente i dati interni a cui hanno accesso. I dati sono infatti spesso tenuti in sezioni separate tra i vari dipartimenti e in alcuni casi con problemi di normative interne che ne limitano la condivisione. Dati importanti potrebbero inoltre non essere registrati come dati, ma piuttosto conservati come conoscenza informale dell'impresa.

Un'altra problematica è la disponibilità di personale qualificato capace di mettere in pratica queste nuove tecniche. Un'indagine del 2017, svolta tra le prime 1000 società negli Stati Uniti sull'implementazione dell'IA nelle loro aziende, ha scoperto che la loro più grande preoccupazione era la disponibilità e la capacità del personale di capire e lavorare con queste nuove soluzioni. Sebbene negli ultimi anni siano stati indubbiamente fatti dei passi in avanti, ancora oggi questo problema non è stato del tutto superato. La formazione di personale qualificato richiede infatti tempo; alcune aziende stanno cercando di velocizzare questo processo, ad esempio Goldman Sachs, una tra le società finanziari più grandi e importanti del mondo, ha costruito un campus in India con spazio per 7000 lavoratori, dove poter trovare personale con queste competenze è decisamente più accessibile.

Vi sono inoltre problemi pratici riguardo all'accuratezza dei sistemi di ML. La gamma di approcci di apprendimento automatico disponibili è in rapida crescita, ma questa evoluzione, benché positiva, è anche dovuta alle evidenti limitazioni dei metodi precedenti e alla necessità di superare tali limitazioni. Ciò indica che le aziende non possono semplicemente applicare una soluzione definitiva alla gestione del rischio, poiché questo è un processo in continua mutazione che richiede una costante valutazione e aggiornamento.

Quando si parla di IA, che “lavora” con parziale o piena automazione dal processo dalla raccolta dei dati alla decisione dell'azione da compiere, la supervisione umana è spesso necessaria. Il

caso di Knight Capital serve ad illustrare questo rischio: grazie al loro processo automatizzato di stock trading sono infatti riusciti a rilevare una perdita di \$440 milioni nel giro di soli 45 minuti.

Un'altra problematica a cui l'IA deve far fronte riguarda la trasparenza e l'etica delle decisioni prese. La prima rappresenta un problema per i sempre più popolari metodi di deep learning, i quali lavorano attraverso parametri non controllabili dall'operatore. Un sistema "black box" di questo tipo non favorisce un'efficace sorveglianza del rischio e può causare problemi di conformità normativa, in particolare per quanto riguarda la dimostrazione della validità del modello. Vi inoltre sono questioni etiche più ampie: la discriminazione contro la razza nelle decisioni di prestito è ampiamente vietata, come lo sono le discriminazioni basate sul genere e la fede religiosa. Normalmente queste restrizioni sono incorporate come regole "rigide" nelle tecniche di AI e di apprendimento automatico riguardanti il rischio di credito e le decisioni di prestito. Tuttavia, con il deep learning diventa molto più difficile assicurarsi che il modello non prenda inavvertitamente decisioni basate su basi eticamente sbagliate.

Ovviamente il ruolo futuro per l'IA e l'apprendimento automatico nella gestione del rischio ha anche tante prospettive positive. Tra queste, la più ovvia è che la natura dispendiosa e costosa della gestione del rischio diminuirà significativamente. Per esempio, BBVA, la seconda banca più grande in Spagna, sta investendo molto in soluzioni di IA al fine di sfruttare la loro capacità di automatizzare le attività ripetitive e di organizzare, recuperare e raggruppare dati non convenzionali, come i documenti.

L'IA fornirà inoltre informazioni sempre più accurate e in tempo reale su tutti i tipi di rischi assunti dall'impresa. Man mano che l'organizzazione dei dati diventerà più orientata verso l'uso dell'IA, la consulenza in tempo reale sarà una presenza fondamentale.

Il passo successivo alla conoscenza dei rischi in tempo reale è la comunicazione preventiva; questo è infatti al momento il fine ultimo delle ricerche che gli esperti stanno svolgendo in questi anni. Si sta quindi cercando di sviluppare un sistema di gestione del rischio guidato dall'IA che sia in grado di riconoscere con precisione e in anticipo i rischi aziendali, siano essi di mercato, operativi o di credito. Le tecniche di apprendimento automatico offrono questa capacità che le tecniche statistiche tradizionali non potrebbero invece mai sperare di raggiungere. Pensando al futuro non vi è infatti alcun ostacolo tecnologico ad un vero e proprio sistema di gestione del rischio dell'IA, che intervenga automaticamente per prevenire rischi

ingiustificati, per scongiurare immediatamente esposizioni pericolose e per adeguare dinamicamente la propensione al rischio dell'impresa. Anche se ciò presenterà sicuramente i suoi rischi, che a loro volta dovranno essere gestiti, i progetti di gestione del rischio saranno felicemente impiegati (anche se in un ambiente in rapida evoluzione) per il prossimo futuro.

5. Bibliografia e sitografia

- Martin Leo, 2019 “Machine Learning in Banking Risk Management: A Literature Review”.
- Theo Lynn, John G. Mooney, Pierangelo Rosati, Mark Cummins, 2019 “Disrupting Finance; FinTech and Strategy in the 21st Century”.
- Mashrur, W. Luo, N. A. Zaidi and A. Robles-Kelly, 2020 "Machine Learning for Financial Risk Management: A Survey".
- T. N. Pandey, A. K. Jagadev, S. K. Mohapatra and S. Dehuri, "Credit risk analysis using machine learning classifiers," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS).
- Mingrui Chen, Yann Dautais, LiGuo Huan, Jidong Ge “Data driven credit risk management process: a machine learning approach” ICSSP 2017: Proceedings of the 2017 International Conference on Software and System Process
- Van Liebergen, Bart, 2017. "Machine learning: A revolution in risk management and compliance?," Journal of Financial Transformation, Capco Institute, vol. 45, pages 60-67
- Bart van Liebergen, 2017 “Machine Learning: A Revolution in Risk Management and Compliance?”
- Anil Savio Kavuri & Alistair Milne, 2019 "FinTech and the future of financial services: What are the research gaps?," CAMA Working Papers 2019-18, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.
- Stijn Claessens & Jon Frost & Grant Turner & Feng Zhu, 2018."Fintech credit markets around the world: size, drivers and policy issues," BIS Quarterly Review, Bank for International Settlements.
- Addo, Peter M., Dominique Guegan, and Bertrand Hassani. 2018. "Credit Risk Analysis Using Machine and Deep Learning Models".
- Shai Shalev-Shwartz, Shai Ben-David 204 “Understanding Machine Learning: From Theory to Algorithms”.
- Jafar Alzubi, 2018 “Machine Learning from Theory to Algorithms: An Overview”
- Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, 2018 “Foundations of Machine Learning, second edition”, The MIT press.
- Naeem Siddiqi, 2018 “Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards, 2nd Edition”

- Anderson, 2007 “Thematic Content Analysis: Descriptive Presentation of Qualitative Data”
- Moosa, Imad A., 2007 “Operational Risk: A Survey. Financial Markets, Institutions & Instruments”