



UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Economiche ed Aziendali - “Marco Fanno”

Corso Di Laurea In Economia

CREDIT SCORING: UTILIZZO DELL'INTELLIGENZA ARTIFICIALE

Relatore:

Preg.Mo Professore Parigi Bruno Maria

Laureando *Alberti Gabriele*

Matricola *1135905*

Anno Accademico
2018-2019

Il candidato, sottoponendo il presente lavoro, dichiara, sotto la propria personale responsabilità, che il lavoro è originale e che non è stato già sottoposto, in tutto o in parte, dal candidato o da altri soggetti, in altre Università italiane o straniere ai fini del conseguimento di un titolo accademico. Il candidato dichiara altresì che tutti i materiali utilizzati ai fini della predisposizione dell'elaborato sono stati opportunamente citati nel testo e riportati nella sezione finale 'Riferimenti bibliografici e sitografici' e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo al documento originale.

SOMMARIO

INTRODUZIONE	1
CAPITOLO 1 : IL CREDIT SCORING	3
1.1 METODI TRADIZIONALI DI CREDIT-ASSESSMENT	3
1.2 APPLICAZIONE DELLE NUOVE TECNOLOGIE AL MERCATO DEL CREDIT SCORING ...	6
1.3 GLI ALGORITMI E IL MACHINE LEARNING.....	7
1.4 L'ESPANSIONE DEL MERCATO DEL CREDITO.....	9
CAPITOLO 2 : BENEFICI E RISCHI COLLEGATI ALL'UTILIZZO DI BIG DATA E ADVANCED ANALYTICS.....	12
2.1 LE NUOVE TIPOLOGIE DI DATI.....	12
2.2 COME L'UTILIZZO DELLE NUOVE FORME DI DATI AUMENTA I BENEFICI DI CONSUMATORI E PRODUTTORI	15
2.3 UNA CONVENIENZA CONDIVISA	17
2.4 CONSEGUENZE ECONOMICHE NEGATIVE	20
2.5 PROBLEMATICHE SOCIALI.....	22
CAPITOLO 3 : ZESTFINANCE INC., COME IL CREDIT SCORING AVANZATO DIVENTA REALTA'	25
3.1 DEFINIZIONE DEL PROBLEMA E SPECIFICAZIONE DELLE VARIABILI OBIETTIVO.....	25
3.2 RACCOLTA E TRASFORMAZIONE DEI DATI.....	26
3.3 SVILUPPO DEL MODELLO FINALE ATTRAVERSO L'ANALISI DEI TRAINING DATA E LA SCELTA DELLE FUNZIONALITA' CHIAVE	28
CONCLUSIONE	31
RIFERIMENTI BIBLIOGRAFICI E SITOGRAFICI.....	34

INTRODUZIONE

In questa prova finale si cercherà di spiegare come negli ultimi anni le nuove tecnologie, artificial intelligence e machine learning, abbiano reso possibile l'implementazione di metodi di valutazione dell'affidabilità creditizia basati su dati dapprima non considerati. Si presenteranno rischi e benefici collegati a questo nuovo paradigma.

I nuovi punteggi di credit scoring diventano delle valutazioni a tutto tondo e come si potrà intuire nel corso della prova, qualsiasi tipologia di dato viene ora presa in considerazione per inferire conclusioni dalle informazioni di una persona. Si dimostrerà, l'alta efficienza e funzionalità di queste nuove scoperte tecnologiche, si discuterà della loro utilità sociale così come delle questioni problematiche che potrebbero minarne l'espansione globale.

E' diventato molto importante a livello mondiale il tema dell'inclusione finanziaria, un argomento che tocca qualunque cittadino più o meno profondamente. Oggi però, grazie alle nuove modalità di valutazione dei rischi collegati al credito, questo problema può essere risolto.

Il contesto che si va a creare grazie all'avvento delle nuove tecnologie è un contesto altamente specializzato, colmo di opportunità di miglioramento e di sviluppo così come di rischi, problematiche sociali e discriminatorie. Il mercato dei dati viene ormai considerato il nuovo petrolio e l'unico modo per non subire conseguenze negative è quello di conoscere a fondo i meccanismi e le conseguenze dei nostri comportamenti in un ambiente altamente interconnesso e controllato, dove qualsiasi azione compiuta online o offline può avere ripercussioni sul nostro futuro e sulle nostre possibilità.

Questi argomenti mi hanno particolarmente colpito per la larga possibilità di applicazione, per l'aumento dismisurato dell'efficienza nei campi in cui vengono utilizzati e per la possibilità di essere sfruttati anche a livello sociale per diminuire le disuguaglianze e le disparità. Sono partito con un pensiero totalmente positivo e a favore di queste tecnologie, tuttavia nel percorso ho potuto constatare che si tratta di processi in via di sviluppo che, se non adeguatamente studiati e capiti, possono portare a conclusioni negative e controproducenti, facendo perdere totalmente la fiducia nel progresso tecnologico.

Si è cercato di dare un'impronta principalmente economica all'elaborato, tralasciando i rischi e le problematiche di natura giuridica per una questione di preferenze personali.

Nel primo capitolo della prova si darà una spiegazione all'evoluzione tecnologica in termini di cambiamento dei sistemi di scoring applicati, continuando con una definizione concettuale

di quali siano i nuovi metodi e di come essi possano essere applicati al mercato del credito. Si conclude la prima parte sottolineando il problema dell'inclusione finanziaria che ha portato all'espansione del P2P lending, analizzando congiuntamente i fattori economici che ne hanno influenzato lo sviluppo.

Il secondo capitolo è concentrato invece sulle nuove tecnologie; si spiegherà il ruolo dei big data nel contesto attuale e si procederà ad analizzare quali possano essere i benefici per i consumatori e per i fornitori. In seguito, si discuterà delle convenienze economiche dei lenders, dei loro vantaggi e obiettivi lucrativi e sociali. Si concluderà il capitolo facendo notare che il nuovo modo di servire il mercato non è esente da critiche e problemi, facendo luce su questioni discriminatorie, problemi di trasparenza e possibili ritorni negativi dell'inclusione finanziaria.

Si concluderà questa prova finale apportando un esempio concreto di quello che è uno dei maggiori rappresentanti dei concetti descritti in precedenza. Si tratta della società ZestFinance, leader nella distribuzione di mutui ai consumatori e pioniere nella costruzione di metodi alternativi di credit scoring. Si cercherà di fare luce su quale sia il processo che porta ad un punteggio di credito e alla conseguente valutazione, apportando un confronto concreto.

CAPITOLO 1 : IL CREDIT SCORING

1.1 METODI TRADIZIONALI DI CREDIT-ASSESSMENT

Un modello di credit scoring è uno strumento che viene tipicamente usato nei processi valutativi e decisionali dove il quesito a cui dare risposta è la concessione o il rigetto di una domanda di prestito. Questo strumento è il risultato di modelli statistici che, basandosi su alcune caratteristiche e informazioni sui richiedenti, permette ai creditori di distinguere tra un candidato buono e uno rischioso, fornendo al soggetto decisore una stima della probabilità di default dello stesso. L'output del modello consiste in un punteggio numerico di tre cifre su una scala che può variare a seconda dell'ente che fornisce il servizio.

Nel corso del tempo i sistemi per la valutazione dell'affidabilità creditizia hanno riflettuto il desiderio e l'obiettivo dei creditori di imporre un tasso di interesse appropriato per il rischio di prestare moneta ad altre persone; essi sono stati fondamentali per determinare il successo finanziario di molti cittadini. Infatti, senza un punteggio favorevole erogato da uno dei maggiori uffici di reporting, i consumatori non potevano permettersi di comprare una casa, di sostenere spese familiari o di mandare i figli al college.

I metodi automatici di valutazione dell'affidabilità al credito sono un'innovazione relativamente recente. Prima degli anni '80 infatti le decisioni di approvazione erano affidate a impiegati e specialisti che valutavano le richieste manualmente su base individuale, applicando tecniche statistiche.

Nel tempo sono stati sviluppati diversi metodi matematici e statistici grazie ai quali è possibile implementare decisioni e valutare l'affidabilità dei richiedenti. Tra i più comuni ci sono: il modello a probabilità lineare, la regressione logistica, il modello probit e il modello dell'analisi discriminante. Le prime tre sono tecniche statistiche standard utilizzate per stimare la probabilità di default sulla base dei dati storici riguardo la performance dei prestiti passati e riguardo le caratteristiche dei richiedenti. Queste tecniche presentano delle differenze: nel modello a probabilità lineare si assume che ci sia una relazione lineare tra la probabilità di default e i fattori analizzati; la regressione logistica invece permette di stimare che la probabilità di un evento (il default) accada o meno, attraverso la previsione di un risultato binario dipendente da un set di variabili indipendenti tra loro, in questo caso la probabilità di default può dipendere dalle caratteristiche dei richiedenti in modi più complicati a seconda dei coefficienti Beta analizzati. Il modello probit è una variante della regressione logistica che invece di utilizzare la cosiddetta trasformazione logistica per determinare la funzione di probabilità, si avvale della funzione di distribuzione cumulativa della

distribuzione normale (Gurný and Gurný, 2013). Per quanto riguarda l'analisi discriminante la peculiarità risiede nel fatto che invece di stimare la probabilità di default di un richiedente, questo metodo divide i richiedenti in classi ad alto e basso rischio di default, analizzandoli sulla base delle classi costruite.

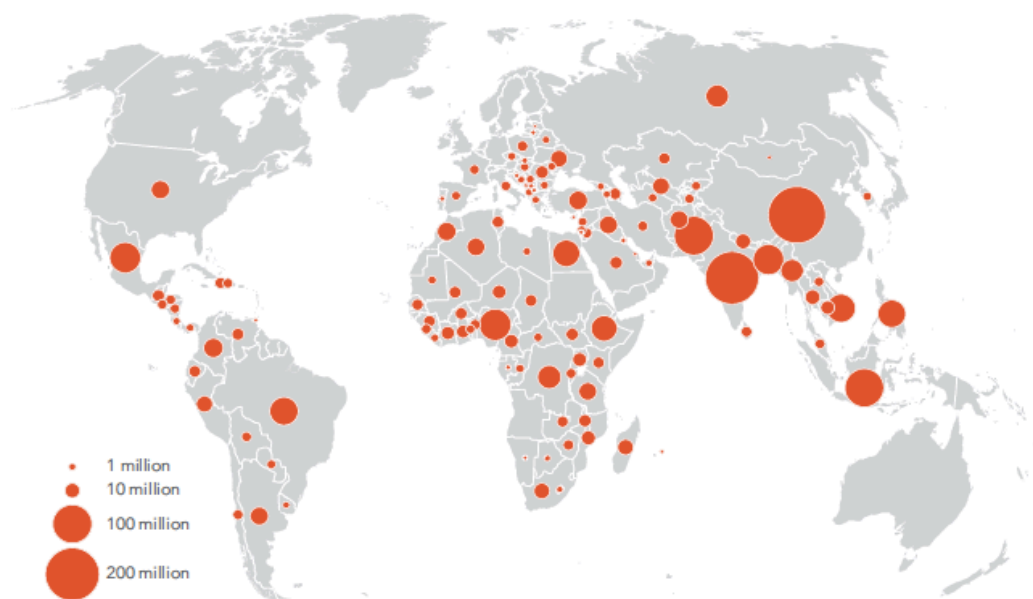
Durante questi processi inizialmente manuali, gli addetti si trovavano ad analizzare grosse quantità di dati e nella maggior parte dei casi le valutazioni potevano soffrire di pregiudizi personali o errori di calcolo.

I metodi automatizzati iniziarono invece ad essere largamente utilizzati solo a partire dagli anni '90 quando gli istituti finanziari constatarono che le nuove metodologie si dimostravano essere un'alternativa migliore, che poteva incrementare l'efficienza e allo stesso tempo eliminare la maggior parte di pregiudizi e discriminazioni legati al pensiero umano. Tuttavia, con il tempo ci si accorse che i miglioramenti apportati non erano all'altezza delle aspettative e che l'utilizzo della tecnologia, sebbene migliorasse la performance rispetto al lavoro umano, penalizzava diversi richiedenti.

Si è costantemente cercato di includere tutti i cittadini nel sistema finanziario ma con scarsi risultati e solo di recente il problema è stato riconosciuto a livello mondiale. Nel 2017 il Global Findex Database, il data set lanciato grazie all'appoggio della fondazione Bill & Melinda Gates, che si propone di studiare il progresso dell'inclusione finanziaria dimostrava (Figura 1) come fossero ancora 1,7 miliardi gli individui considerati non valutabili e quindi esclusi dall'accesso alle tradizionali forme di credito.

Globally, 1.7 billion adults lack an account
Adults without an account, 2017

Figura 1



Source: Global Findex database.

Note: Data are not displayed for economies where the share of adults without an account is 5 percent or less.

Si trattava di individui con una cronologia di credito limitata o assente, persone che non avevano posseduto un account di credito attivo per almeno sei mesi e che, data la limitata quantità di dati considerata dai tradizionali strumenti d'analisi, non venivano nemmeno valutati dalle agenzie di reporting. Questo comportava un'esclusione dal sistema finanziario e da quello bancario.

In principio, il metodo utilizzato da una delle società di analisi dati e servizi finanziari più sviluppate, la Fair Isaac Corporation, meglio conosciuta con l'acronimo FICO, specializzata nei servizi di credit scoring, guardava principalmente alla cronologia dei pagamenti, alla disponibilità finanziaria, alla cronologia creditizia e alla tipologia di strumenti finanziari utilizzati da un determinato consumatore, omettendo lo studio di fattori come la sua storia lavorativa, il suo salario e altre variabili che potevano essere correlate con una valutazione di questo tipo (FICO, 2019).

I dati del 2017 possono essere considerati obsoleti in un'era digitale come quella che stiamo affrontando, tuttavia come riportato nel sondaggio condotto da The Harris Poll nel gennaio del 2019 su un campione di adulti statunitensi, considerati tali da 18 anni in su, emerge che il 40% degli Americani interrogati sostiene di essere escluso dai sistemi finanziari e di essere stato limitato da un mero punteggio creditizio nel momento in cui stipulava una richiesta di fondi in prestito. Fondi che sarebbero stati necessari per raggiungere un obiettivo finanziario chiave nella vita di ciascuno: comprare una macchina, sostenere spese per la famiglia o ristrutturare la casa. I punteggi sono ancora più preoccupanti quando si analizzano i dati riguardanti diversi raggruppamenti: sono infatti il 57% degli ispanici interrogati a ritenere di essere esclusi dal sistema finanziario, 55% invece se si guarda alla popolazione afroamericana e infine ben il 59% dei giovani, dai diciott'anni ai venti tre anni (The Harris Poll, 2019).

Il modello tradizionale di credit scoring, considerato una delle grandi determinanti del successo finanziario statunitense del ventesimo secolo, non è riuscito a stare al passo con il consumatore d'oggi, limitando l'inclusione finanziaria e riducendo le possibilità di crescita e sviluppo economico individuale così come di quello economico. L'aumento delle fonti di dati disponibili per valutare i richiedenti in modo più accurato e più equo non è stato compreso dai tradizionali approcci e si tratta perciò di una mancanza da colmare.

Come si evince dal sondaggio condotto da The Harris Poll, la maggior parte degli americani, il 70%, appoggia nuovi metodi per dimostrare agli istituti finanziari la propria affidabilità per le richieste di credito.

1.2 APPLICAZIONE DELLE NUOVE TECNOLOGIE AL MERCATO DEL CREDIT SCORING

Nel corso degli ultimi anni, anche grazie a innovazioni tecnologiche senza precedenti, si è diffusa un'opinione generalmente riconosciuta riguardo all'incapacità dei metodi tradizionali di intercettare i bisogni dei consumatori che non possiedono una storia creditizia, o che non ne hanno una sufficientemente sviluppata per poter richiedere un prestito ed aver accesso a tariffe ragionevoli ed eque.

Questo ha portato all'emergere di metodi alternativi, tecnologie incentrate sull'utilizzo dei big data che promettono performance migliori per i processi di sottoscrizione dei prestiti e, allo stesso tempo, una diminuzione della disparità sociale.

Sebbene i fattori analizzati tradizionalmente rimangano centrali nel sistema che si va sviluppando, l'industria del credito si prepara ad affrontare un rapido cambio di rotta verso metodi nuovi ed alternativi. Anche le tradizionali agenzie di valutazione e reporting stanno iniziando a sviluppare modelli più sofisticati affidandosi allo studio delle nuove fonti di dati disponibili.

In questo nuovo panorama creditizio la valutazione viene rilasciata dopo aver analizzato nuove tipologie di dati che si ritiene possano avere un valore significativo per spiegare l'affidabilità creditizia di una persona. Si cerca di sviluppare profili universali che integrino informazioni provenienti dalle attività, online e offline, dei consumatori.

Un esempio eclatante è dato dalla più grande agenzia di reporting FICO che, dopo aver testato diversi modelli costruiti per analizzare, tra gli altri, anche dati alternativi, ha presentato il suo modello di scoring alternativo in collaborazione con un'altra grande agenzia di reporting statunitense, Equifax: *FICO Score XD*. Con questo nuovo processo si pone l'obiettivo di soddisfare e colmare la carenza di offerta di servizi finanziari per i cosiddetti thin-file borrowers, coloro che hanno poche garanzie e informazioni da offrire, utilizzando dati provenienti dai conti delle linee telefoniche via cavo e mobili. Mentre alcuni di questi dati potrebbero sembrare logicamente relazionati con l'abilità del consumatore di gestire un prestito, come per esempio la cronologia dei pagamenti delle utenze (Robinson and Yu, 2014) altri tipi di dati, invece, vengono integrati nei modelli sebbene manchi un'intuitiva connessione con la possibile affidabilità del creditore: Berg and al. (2018) dimostrano come anche semplici variabili costruite a partire dalle "digital footprints" dei consumatori hanno un potere predittivo che uguaglia o eccede il potere previsionale dei dati tradizionali; esempi di variabili che prendono in considerazione nel loro documento sono: il tipo di dispositivo dal

quale ci si collega alla rete, il sistema operativo che si utilizza, il servizio di posta elettronica di cui si usufruisce, il canale attraverso il quale si raggiunge una piattaforma online e altre.

1.3 GLI ALGORITMI E IL MACHINE LEARNING

Ciò che rende fattibile lo studio e l'analisi delle nuove tipologie di dati e delle enormi quantità che oggi sono disponibili sul mercato è la nuova tecnologia del Machine Learning applicato ai ben conosciuti algoritmi classici. Come descritto da Hurley e Adebayo (2016, p. 12) quest'ultimi possono essere implementati per risolvere problemi di natura strutturata, detti "well-defined" e non-strutturata, detti "ill-defined". In particolare, nel campo del credit assesment, si utilizzano gli algoritmi per risolvere un problema che è considerato non-strutturato data la alta complessità di calcolo, la presenza di incertezza alla base del processo e all'interno dei dati di input. In parole semplici, un problema non-strutturato può avere diverse risposte corrette, ma alcune di queste risposte possono essere migliori delle altre.

Un banale esempio di problema strutturato può essere il calcolo della circonferenza di un cerchio in base alla misura del raggio: è previsto un unico ed inequivocabile risultato. Per quanto riguarda i problemi detti ill-defined si supponga di voler preparare una pietanza, senza sapere esattamente gli ingredienti e le quantità necessarie; un metodo per raggiungere il risultato finale potrebbe essere quello di preparare diverse combinazioni random di tutti gli ingredienti possibili, scartare gli ingredienti e le combinazioni che risultano errati, migliorare le proporzioni di quelli corretti e ricominciare fino ad arrivare alla pietanza desiderata. Questa analogia culinaria, sebbene non perfetta, offre un'idea di come venga svolto il processo iterativo del machine learning (Hurley and Adebayo, 2016).

La criticità di questi processi è dovuta al fatto che gli algoritmi che stanno alla base dei modelli non sono statici, bensì ci si aspetta che essi cambino in risposta all'immissione di nuovi dati di input negli stessi.

Il machine learning, considerato una sottocategoria dell'artificial intelligence, consiste in un metodo matematico e prevede l'individuazione di una sequenza di azioni per la risoluzione di un problema; una tecnologia che ha la capacità di ottimizzarsi automaticamente attraverso esperienza e reiterazione e costruita con un limitato se non assente intervento umano. L'ampia diffusione riscontrata si deve all'abilità di individuare automaticamente schemi intrinseci e non lineari in un ampio bacino di dati che vengono poi usati per scoprire nuove relazioni e tendenze, per predire risultati futuri e svolgere altre tipologie di decisione in presenza di incertezza. Una volta che il processo di machine learning è completo, i costruttori utilizzano i trend e le intuizioni riscontrate nei dati per sviluppare un modello che possa prevedere

l'output desiderato. Si tratta di un processo prova-errore con l'obiettivo di arrivare ad un risultato previsto; tutto questo è diventato possibile solo di recente grazie al raggiungimento di una potenza di calcolo tale da rendere fattibile l'esecuzione di milioni di iterazioni in un lasso di tempo relativamente breve.

Esistono diversi processi di machine learning (Figura 2) che si differenziano a seconda del modo in cui gli algoritmi apprendono dai dati e a seconda del livello di intervento umano, questi processi sono: “supervised, unsupervised, reinforcement e deep learning” (Schindler, Souza Moraes and al., 2017, p. 9).

Nel primo tipo, viene inserito nell'algoritmo un set di dati detti training data, a una parte dei quali viene assegnata un'etichetta, un risultato. Per esempio, in un data set di transazioni effettuate di una banca, le etichette su alcuni dati vanno ad identificare gli individui fraudolenti e quelli che sono invece in regola; si tratta di dati già verificati a cui è associata una risposta corretta, i quali costituiscono la base dalla quale l'algoritmo deve apprendere. L'algoritmo imparerà una regola generale di classificazione e la userà per classificare i dati restanti e assegnare anche ad essi un'etichetta.

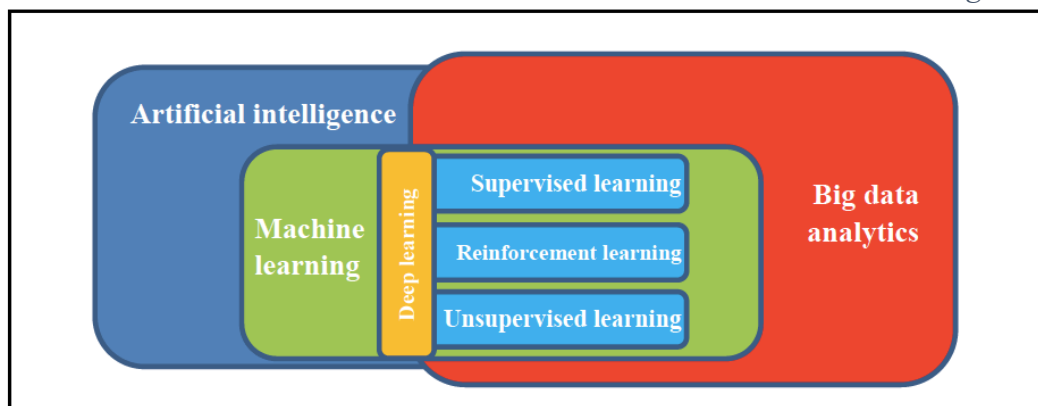
Nel secondo caso, il machine learning unsupervised, si parla di vera e propria intelligenza artificiale. In questo caso il data set non è provvisto di etichette, si chiede all'algoritmo di determinare degli schemi intrinseci nei dati attraverso l'individuazione di aggregati di osservazioni che dipendono da caratteristiche sottostanti simili, i cosiddetti cluster. Per esempio, un algoritmo di questo tipo potrebbe essere impostato per individuare il prezzo di titoli che presentano caratteristiche simili ad altri considerati illiquidi e difficili da prezzare; se l'algoritmo individua un aggregato appropriato per i titoli illiquidi, il prezzo dei titoli presenti nel cluster potrebbe essere usato come benchmark nell'analisi di altre posizioni aperte.

Il reinforcement learning si trova al centro tra i primi due metodi citati; in questo caso si utilizza una serie di dati non etichettati e l'algoritmo sceglie una specifica azione per ciascun dato trattato, ricevendo un feedback, anche umano, che lo addestra. Un esempio può essere il caso dello sviluppo di macchine a guida automatica (Schindler, Souza Moraes and al., 2017).

Infine, il deep learning è una forma di machine learning evoluta che apprende attraverso dei network neurali artificiali che simulano l'attività e le connessioni del cervello umano, permettendo al computer di analizzare i dati attraverso una struttura di pensiero simile a quella umana. Questa tecnologia di ultima generazione non ha bisogno di un programmatore che le indichi cosa fare poiché il concetto alla base è quello di fornirle un'ingente mole di dati

dalla quale estrapolare autonomamente dei pattern, in modo tale da rendere altamente indipendente il processo di learning e quello di previsione (Marr, 2018).

Figura 2



Schindler, Souza Moreas and Al., 2017

1.4 L'ESPANSIONE DEL MERCATO DEL CREDITO

Una varietà di fattori che hanno contribuito generalmente all'incremento dell'uso della FinTech, hanno allo stesso tempo spronato l'adozione dell'intelligenza artificiale e del machine learning nei servizi finanziari. Dalla parte dell'offerta, il mercato del credit scoring ha beneficiato della disponibilità di queste tecnologie in gran parte grazie all'implementazione delle stesse dapprima in altri campi. Questo mercato infatti si è sviluppato grazie alla disponibilità di computer altamente performanti, che hanno portato alla vendita di processori più rapidi, all'abbassamento dei costi degli hardware e al migliore accesso alla potenza dei computer tramite i servizi di clouding. Allo stesso modo si sono diffusi database economici e l'analisi di dati tramite software e algoritmi sofisticati. Mentre i costi delle tecnologie si abbassavano, il volume di dati creati andava aumentando e tutto ciò apre nuovi scenari per l'implementazione di nuove tecniche per la valutazione dell'affidabilità creditizia.

L'ambiente post crisi poi, presentando bassi tassi di interesse e rendimenti dei bond sovrani vicini allo zero, ha portato gli investitori a cercare sponde più interessanti con rendimenti potenzialmente maggiori. Un nuovo target di investimento sono diventati i mercati del P2P lending, collegati a modelli di valutazione creditizia artificiali. I mutui esborsati tramite le nuove piattaforme di P2P offrono tipicamente rendimenti maggiori anche se presentando solitamente rischi più alti (IOSCO, 2017).

Dalla parte della domanda, le istituzioni finanziarie hanno incentivi nell'utilizzo di intelligenza artificiale e machine learning per scopi economici. Le opportunità maggiori si

hanno in termini di riduzione dei costi, miglioramenti produttivi e migliore gestione del rischio; il tutto in vista di una redditività maggiore. In altri casi gli investimenti per implementare le nuove tecnologie vengono sostenuti per stare al passo con la concorrenza e questo spesso garantisce un tasso di innovazione sempre crescente.

Fino a poco tempo fa, i finanziatori e gli istituti di credito dovevano confrontarsi con limiti e vincoli tecnologici che riducevano la possibilità di raccogliere, immagazzinare e analizzare dati. Di recente invece sono proprio gli istituti finanziari che, in un periodo di strette regolamentazioni e tassi di interessi minimi, appoggiano la ricerca di nuove fonti di profitto. Wells Fargo, Bank of America e JPMorgan Chase, già nel 2015, preoccupate dalla mancanza di cittadini valutabili tramite i servizi tradizionali, si dichiararono favorevoli a servire più consumatori ad un livello di rischio accettabile, consumatori da individuare con metodi alternativi. Per questo, anche i tre più grandi istituti bancari d'America, appoggiarono lo sviluppo delle nuove metodologie di FICO per lo scoring dei consumatori, descritte a breve nel seguito del testo (McLannahan, 2015).

Per analizzare fonti di dati riguardo le attività online e offline dei consumatori e trarne conclusioni affidabili, tutta l'industria della valutazione del credito sta convertendo verso l'utilizzo di modelli più complicati e algoritmi più complessi. L'abilità di sfruttare dati aggiuntivi riguardo ai consumatori, permette ai creditori una segmentazione dei clienti più veloce ed economica, raggiungendo allo stesso tempo un numero di utenti maggiore.

Come accennato in precedenza, anche la Fair Isaac Corporation, appurò la presenza di più di 50 milioni di consumatori statunitensi senza un punteggio FICO, per mancanza di informazioni negli archivi. Appoggiata dalle grandi istituzioni bancarie americane, deciso di implementare un metodo per la valutazione di questi consumatori. Il loro approccio consiste nell'utilizzo di dati alternativi, affiancati alla base dalle informazioni classiche già presenti negli archivi. Will Lansing, CEO dell'azienda in questione, nell'Ottobre del 2015 riportò al Financial Times che FICO nelle sue operazioni valutative, incominciava a guardare oltre che alla cronologia dei rimborsi delle carte di credito, anche altre informazioni provenienti da piattaforme come Facebook. Lo stesso Lansing poi disse: "la quantità di volte che una persona dice 'ubriaco' nel proprio profilo Facebook, ha un certo valore per predire se gli individui ripagheranno o meno un debito" (McLannahan, 2015).

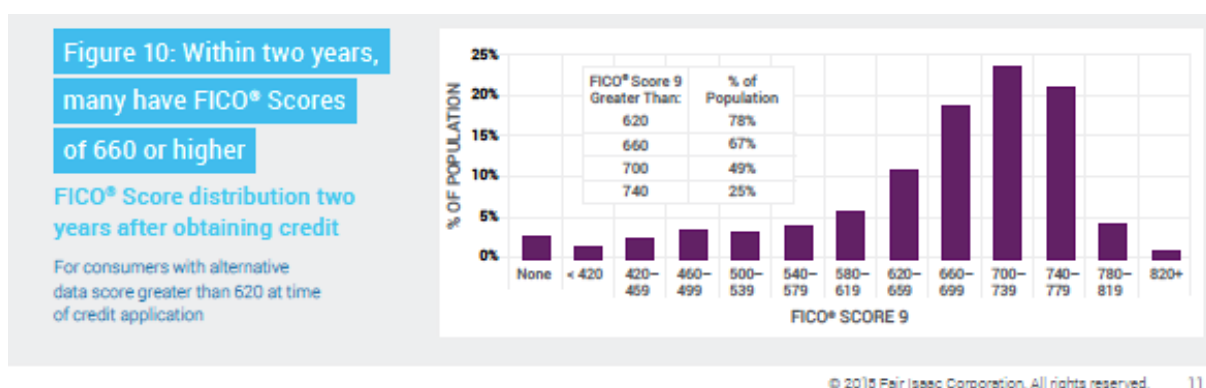
La ricerca spiegata in FICO Insights White Paper No. 90 del 2015, dimostra che con i dati alternativi corretti e accuratamente scelti si è in grado di valutare una grande parte dei consumatori precedentemente considerati "unscorable". Più del 50% di questi individui sono diventati valutabili. Come si legge dal documento, la percentuale di consumatori divenuta

valutabile varia a seconda del segmento che si considera. In particolare, vengono divisi questi consumatori in tre segmenti all'interno dei quali si va a vedere quanti di loro possono essere valutati in seguito all'utilizzo di dati alternativi: il primo segmento racchiude coloro che avevano perso l'accesso al credito e mostra una percentuale di miglioramento del 47%, il secondo coloro i quali fanno richiesta per la prima volta, sono il 76% le persone divenute valutabili e infine l'ultimo mostra le persone inattive, dove il miglioramento è del 43%. Il risultato della ricerca riguardo ai punteggi è incoraggiante; più di un terzo dei richiedenti valutati grazie al nuovo metodo raggiunge un punteggio di 620 o maggiore, tale da garantirgli un accesso al credito. I finanziatori in questo modo possono effettivamente identificare le probabilità di rischio e espandere in maniera sicura l'accesso ai finanziamenti.

Per confermare i risultati della ricerca, FICO è stata in grado di dimostrare che la maggior parte dei consumatori ai quali è stato concesso un finanziamento grazie alla nuova tipologia di valutazione, sono stati capaci di gestire la propria obbligazione responsabilmente nel tempo.

Come mostra un grafico all'interno della ricerca (Figura 3), la maggioranza dei richiedenti che avevano ottenuto un punteggio di 620 o maggiore all'apertura del finanziamento, 24 mesi dopo risultavano avere un punteggio FICO di 620 o maggiore.

Figura 3



Il 67% di queste persone si presentava con un punteggio maggiore di 40 punti, mentre addirittura il 49% raggiungeva un punteggio di 700 o maggiore il quale è indicatore di un richiedente affidabile e rende l'individuo eleggibile per una serie di offerte di credito. Questo per dimostrare come individui in precedenza esclusi dai sistemi finanziari siano riusciti a dimostrarsi adempienti e ad ottenere condizioni di prestito migliore grazie alle analisi più approfondite.

CAPITOLO 2 : BENEFICI E RISCHI COLLEGATI ALL'UTILIZZO DI BIG DATA E ADVANCED ANALYTICS

L'innovazione tecnologica nel settore creditizio è un fenomeno globale in rapida crescita in particolare nell'economie in via di sviluppo e nei mercati emergenti. L'utilizzo di nuove tecnologie sta cambiando il modo attraverso il quale i prodotti e i servizi finanziari vengono sviluppati e consegnati ai consumatori, incrementando in larga scala il numero dei potenziali utilizzatori, semplificando l'accesso da postazioni remote, allargando la copertura a luoghi rurali e riducendo in gran parte i costi di realizzazione e distribuzione dei servizi.

Queste innovazioni sono particolarmente rilevanti perché incoraggiano le economie e i mercati in via di sviluppo ad adottare strategie di inclusione attraverso la finanza digitale, con l'obiettivo di sfruttare le potenziali esternalità positive per la crescita economica e per dare a ciascun cittadino la possibilità di accedere a fonti di finanziamento. Tuttavia, ci possono essere dei riscontri negativi se si tiene conto del problema della mancanza di trasparenza dei sistemi, delle possibili ripercussioni discriminatorie e della privacy personale.

2.1 LE NUOVE TIPOLOGIE DI DATI

Una parte chiave del processo di quest'evoluzione tecnologica è l'utilizzo di nuove tipologie di dati e delle rispettive nuove metodologie per processarli e utilizzarli con lo scopo di fornire servizi finanziari.

Oggi il termine big-data è diventato d'uso comune, indicato per descrivere tutto l'insieme di dati caratterizzati da una crescita esponenziale in termini di volume, varietà e velocità di diffusione (Figura 4). I big-data sono codici informatici soggetti ad analisi avanzate, le cosiddette advanced data analytics, che servono per estrapolare dai singoli numeri un trend da poter applicare alle questioni finanziarie e non solo.

La definizione che si legge nel Joint Committee Discussion Paper (2016), spiega in modo dettagliato le tre V sulle quali si fonda il concetto di big data: per volume si intende l'immenso e crescente ammontare di dati generato dalle nuove forme di mass data; ci sono visioni differenti, tuttavia spesso si parla di almeno 2,5 exabytes generati ogni giorno e si prevede la creazione di più di 100 zetabytes di dati entro il 2020. Un exabyte corrisponde ad un byte elevato alla diciottesima potenza di dieci, mentre per lo zetabyte la potenza viene elevata ben ventuno volte (TechTerms, 2012); la velocità si riferisce al tasso di espansione e crescita delle informazioni mentre la varietà si riferisce alle differenti tipologie di dati che vengono collezionati, analizzati e studiati (EBA, EIOPA and ESMA, 2016)

Figura 4



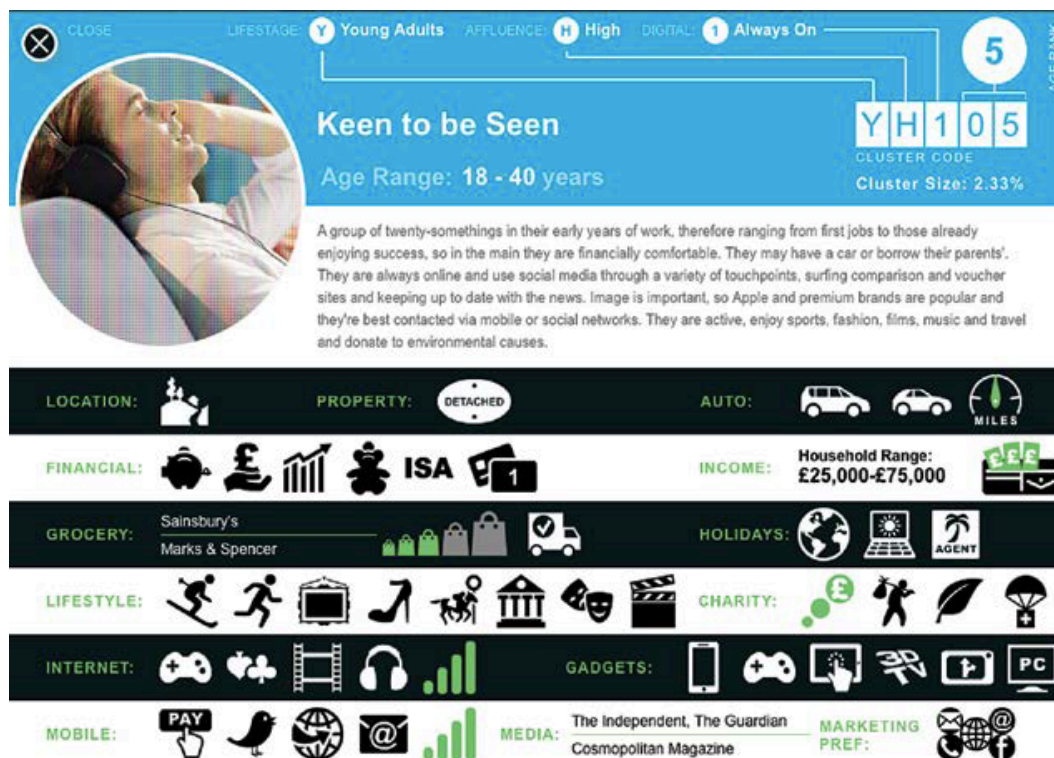
Source: *Big Data, Small Credit* 2016

Fonti di informazioni codificate e nuove forme per processarle sono contributori chiave per supportare l'esplosione nell'accessibilità a servizi finanziari convenienti e costruiti su misura per qualsiasi tipo di consumatore, served, unserved e underserved (Cheney, 2016).

Tradizionalmente, gli istituti finanziari fornitori di servizi di credito, hanno accesso a dati organizzati nei database convenzionali e fruibili direttamente presso i raccoglitori. Si tratta di informazioni quali: data-credit history e punteggi ottenuti in passato, documenti di identità, dati demografici, dati raccolti nei sondaggi pubblici e dati riguardo alle transazioni personali effettuate. Le innovazioni tecnologiche permettono l'utilizzo di altre forme di dati, considerati nuovi per il fatto di non essere mai stati utilizzati prima in questi settori o perché non direttamente correlati con l'utilizzo dei servizi che questi istituti offrono. Si tratta infatti di informazioni estratte dalle attività compiute sui social media e sulle nuove piattaforme di moneta virtuale, informazioni riguardo le ricerche nei siti internet, così come riguardo le abitudini di shopping dei consumatori; cronologie dei pagamenti delle utenze, dati psicometrici, feedback sull'utilizzo di servizi di intrattenimento mobili, dispositivi collegati al cosiddetto internet delle cose, messaggi telefonici, mail, chat online, dati studiati dalle assicurazioni per determinare i rischi e stipulare le polizze. E ancora informazioni relative ai tempi di spedizione e consegna scelti per gli acquisti, recensioni e lamentele dei clienti e altri simili. Esistono veri e propri cataloghi dove le aziende possono trovare diverse configurazioni di dati anonimizzati dei consumatori a seconda di ciò che vogliono conoscere; un esempio è il pacchetto offerto nel catalogo della compagnia statunitense di database marketing Acxiom, il colosso di brokeraggio dei dati si presenta con un database che copre 700 milioni di persone, un portafoglio di 7 000 clienti e un fatturato annuo di 1,1 miliardi di dollari. Chiamato Personicx, il servizio fornisce un sistema di segmentazione e visualizzazione di 47 milioni di consumatori del Regno Unito, che permette agli esperti di marketing di organizzare i propri

clienti in base ai dati demografici, agli stili di vita e alle abitudini di acquisto (Figura 5). E' destabilizzante leggere che il servizio è disponibile a tre livelli: individuale, per nucleo familiare o per codice postale (Acxiom, 2018).

Figura 5



Acxiom, Data Catalogue 2018

Le novità di queste informazioni poggiano sul fatto che alcune di esse possono anche apparire non strutturate, come le ultime citate, definite tali perché non presentano un modello pre-costruito e/o non sono organizzate in una maniera predefinita come lo erano invece i dati tradizionali (Nemschoff, 2014)

Tutte queste nuove tipologie di informazioni, insieme ai database pubblici dei tribunali, alle sentenze di bancarotta e agli account social, possono essere sfruttati e presi in considerazione nel big-data analytics per il credit scoring.

Un altro fattore di fondamentale importanza in questo scenario, che influenza tanto la crescita dei benefici quanto la presenza di rischi, è la rapida espansione dell'utilizzo degli smartphone nel fruire di servizi finanziari. Nel dicembre del 2016, c'erano circa 2 miliardi di telefoni al mondo e ci si aspetta che questo numero raggiunga i 4 miliardi entro il 2020, con la maggior parte della crescita che si presenterà nei mercati emergenti e nelle economie in via di sviluppo (Barnes, 2016). Molte economie stanno riconoscendo il potere di questo strumento come divulgatore di informazioni a basso costo, sfruttabile per espandere l'inclusione finanziaria e offrire prodotti strutturati sulla base delle esigenze del singolo consumatore.

2.2 COME L'UTILIZZO DELLE NUOVE FORME DI DATI AUMENTA I BENEFICI DI CONSUMATORI E PRODUTTORI

Le nuove tipologie d'informazione, in particolare i big-data insieme alle relative modalità d'analisi, garantiscono agli istituti finanziari e non solo di sfruttare nuove opportunità per usare data sets più capienti in modo da combinare i dati tradizionali con informazioni in tempo reale per offrire prodotti studiati su misura.

Queste tecnologie stanno diventando di fondamentale importanza perché grazie ad esse è possibile raggiungere consumatori non ancora valutati con un'ottica di mercato incentrata sul consumatore, creare un punteggio di credito per consumatori sprovvisti di una formale o consistente credit history, identificare le volontà dei singoli clienti e costruire dei profili altamente personalizzati; si può segmentare il mercato in maniera più profonda e dettagliata; valutare e prevenire comportamenti di frode e infine si possono aumentare le performance con cui si rilevano i potenziali rischi di default attraverso determinati modelli di scoring.

In prospettiva di crescita economica, l'adozione dell'artificial intelligence in questo campo ha un enorme potenziale per migliorare l'efficienza del sistema economico e per contribuire alla crescita. Per esempio, grazie all'intelligenza artificiale si assiste ad un aumento della velocità, contemporaneamente ad una riduzione dei costi per i pagamenti e per il regolamento delle transazioni, questo facilita la creazione di dati fondamentali per il successivo studio dell'affidabilità creditizia del consumatore. La tecnologia può poi aiutare ad individuare il momento giusto per l'esecuzione di alcuni scambi, ovvero il momento in cui si è certi che ci sia una controparte con una domanda esattamente corrispondente, questo potrebbe stimolare transazioni nell'economia reale.

Negli ultimi tempi molte imprese, incluse quelle non finanziarie appaiono ampiamente favorevoli a queste nuove tecnologie; inoltre la crescita degli investimenti per la ricerca e lo sviluppo di nuove interfacce possono stimolare l'investimento generale così come lo sviluppo economico. Le innovazioni più tecnologiche potrebbero essere disponibili solo per le grandi aziende che possono permettersi questi investimenti, tuttavia questo sviluppo aprirebbe la strada anche alle realtà più piccole che potrebbero svilupparsi come follower di tecnologie già disponibili.

Le nuove tecnologie, possono essere implementate per individuare rischi eccessivi e transazioni troppo complicate, in questo modo gli istituti finanziari potrebbero usufruirne per costruire strategie di copertura più efficaci per la gestione del rischio. Il fatto che questi

strumenti favoriscano la crescita di nuove piattaforme di credito per connettere direttamente creditori e mutuatari, riduce la dipendenza da prestiti bancari, e di conseguenza la leva finanziaria delle banche tradizionali. Questo potrebbe garantire una struttura di condivisione del rischio più diversificata nel sistema finanziario globale (World Bank Group, 2018).

Un esempio di come il mercato del credit scoring si stia evolvendo è dato dall'idea di Lenddo, azienda fondata ad Hong Kong, premiata al World Economic Forum del 2014 come Tech Pioneers, che iniziò le sue prime operazioni nelle Filippine già nel 2011. Lenddo usa le digital footprints per tracciare e fornire dei record finanziari alle banche e agli istituti finanziari; si avvale di algoritmi basati su non-traditional data e machine learning per dare un punteggio di credito agli individui; come si legge nel sito web, la sua missione iniziale era quella di migliorare le vite dei consumatori della classe media emergente nei paesi in via di sviluppo attraverso la fornitura di micro-prestiti per scopi specifici (Lenddo Website, 2019). Il focus si è evoluto per lavorare con istituzioni finanziarie e altre aziende partner, in modo che gli stessi clienti potessero a loro volta usare i social network e altri dati provenienti da Facebook, LinkedIn, Google, Yahoo e Twitter per provare la loro identità e l'affidabilità creditizia. Nel 2013 in un articolo di The Economist si leggeva: "Forse nessun'azienda è andata tanto avanti quanto ha fatto Lenddo, 'loan-seekers' chiedono agli amici su Facebook di garantire per loro", per determinare se coloro che accettano di fare da garanti sono amici reali piuttosto che semplici contatti nel social network, il software tecnologico di Lenddo controlla la presenza di un linguaggio comune e condiviso o di parole che suggeriscono affinità. Inoltre, il punteggio di credito di chi si dimostra garante, viene danneggiato qualora il garantito non riuscisse a ripagare il debito (ANON., 2013).

Passando a realtà ancora in via di sviluppo, un'azienda keniota che unisce l'obiettivo dell'inclusione finanziaria all'utilizzo delle nuove tecnologie, è M-Shwari, co-fondata dall'operatore di telefonia mobile Safaricom e dalla Banca Commerciale d'Africa. Qui, il processo di approvazione dei prestiti consiste in una prima analisi relativa all'utilizzo del proprio smartphone. Viene condotta con l'obiettivo di fissare un limite di credito iniziale a seconda del profilo di rischio individuato dagli algoritmi. Nel corso del tempo, attraverso i diversi comportamenti di risparmio e rimborso, i limiti di credito del cliente saranno rivisti; nel 2014 la Banca Commerciale d'Africa ha riportato che l'unità M-Shwari ha un valore relativo al rapporto perdita/prestito al di sotto della media di 5% del settore e che tra il 2013 e il 2014 sono stati sborsati circa 289 milioni di dollari in prestiti (Hurley and Adebayo, 2016).

In generale in paesi in via di sviluppo come il Kenya, c'è un'idea comune positiva verso gli operatori di telefonia mobile, verso le nuove forme di valutazione e assegnazione del credito

così come verso le nuove tipologie di dati che possono servire da garanzia: i primi forniscono un'ottima quantità e qualità di dati che permette a start-up, banche e istituti di credito di avere accesso ad un numero elevato di potenziali clienti da servire attraverso piccoli prestiti costruiti su misura, mentre i conseguenti comportamenti di prestito e rimborso diventano, nel lungo periodo, le basi per rivedere i limiti di credito iniziali, i prezzi e i termini delle linee di credito disponibili. Inizialmente il valore dei prestiti è limitato, spesso dai 30\$ in giù e raramente il contratto si protrae per più di un mese, tuttavia l'obiettivo principale nel lungo periodo è quello di costruire una borrowing history che possa determinare un futuro accesso al credito tradizionale (Chen and Faz, 2015).

2.3 UNA CONVENIENZA CONDIVISA

Si è parlato dei potenziali benefici che i big-data, il machine learning e le nuove tecniche di analisi possono apportare ai metodi di scoring e di conseguenza ai consumatori.

Tuttavia, anche qui si parla di una medaglia a due facce, infatti la motivazione di fondo degli istituti finanziari non è solo di natura sociale; si cerca pur sempre un margine di guadagno.

Dalla parte dell'offerta i benefici derivanti dai digital data e dalla loro applicazione sono altrettanti; grazie all'innovazione digitale, istituti finanziari e non possono trovare nuovi potenziali clienti, possono costruire relazioni più profonde e personalizzate con gli stessi e infine gestire il rischio con il supporto di patterns individuati tecnologicamente. Trovare nuovi clienti è importante, specialmente per il problema dell'inclusione finanziaria, tuttavia la raccolta tradizionale dei dati necessari per iniziare a costruire una relazione con il cliente è onerosa, sia per i fornitori che per gli utenti stessi e spesso porta a delle valutazioni di credito imprecise.

Una volta che il cliente è entrato in contatto con l'azienda, lo studio delle nuove tipologie di dati può suggerire indizi su come migliorare la relazione. I vantaggi in questa fase della vendita sono dati dalla capacità degli algoritmi e dell'intelligenza artificiale di individuare correlazioni insite nei dati. Queste vengono poi utilizzate per diminuire i punti di attrito con i clienti o per identificare più velocemente opportunità di "up-sell e cross-sell" (EBA, EIOPA and ESMA, 2016, p. 24) per espandere la relazione, coinvolgere il cliente e migliorare i propri rendimenti.

Per quanto riguarda il caso dei milioni di consumatori unbanked e underbanked, il problema è che effettivamente, in questo tipo di mercato, ci sono state delle barriere all'entrata ed è questo uno dei principali problemi alla base della situazione che si ci si trova ad affrontare

ora. In passato non c'era convenienza a soddisfare i bisogni di certe fasce di consumatori, si presentavano ostacoli legati alla rischiosità e all'incertezza delle scelte. Le banche non erano disposte ad assumersi il rischio di estendere la propria offerta ai cittadini con redditi bassi, senza una storia creditizia o senza la possibilità di dare garanzie; allo stesso tempo i tipi di prodotti offerti risultavano o poco attraenti o comunque troppo onerosi per essere offerti a questi consumatori senza troppi rischi. Infine, i canali di distribuzione non erano adatti e tanto meno non erano stati studiati per garantire efficienza di costo a supporto dell'offerta; i modelli operativi e di consegna dei servizi presentavano costi troppo alti e parallelamente richiedevano un enorme utilizzo di risorse per offrire prodotti a basso costo in quantità tale da essere conveniente.

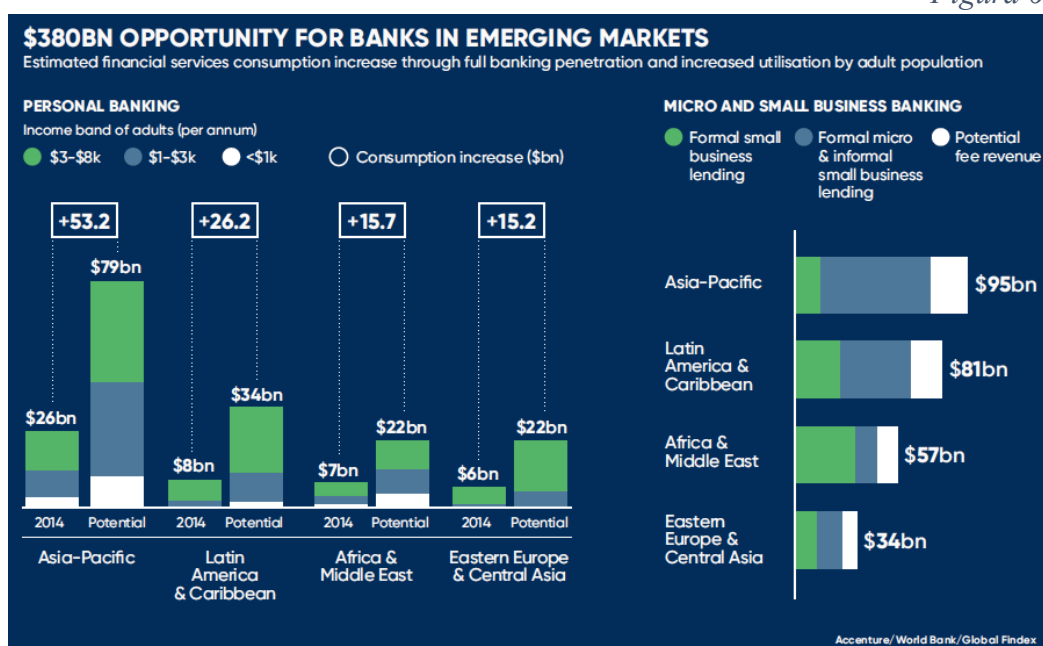
Alla fine del 2014, con l'obiettivo di sviluppare un'analisi quantitativa che spiegasse numericamente quanto la tecnologia poteva tornare utile anche nel prestito bancario, il Consulting Group to Assist the Poor incaricò l'istituto di ricerca McKinsey and Company di esaminare a fondo e analiticamente la questione nel mercato scelto come riferimento: la Tanzania. Questo paese nello stesso anno, si presentava con una popolazione composta tra l'80% e il 90% da individui che non lavoravano in un'istituzione formale e che perciò non ricevevano un salario regolare, al contempo la stessa percentuale non aveva mai avuto una relazione finanziaria con un'istituzione bancaria e infine, essi vivevano guadagnando in media circa 2\$ al giorno. Tuttavia, la rapida espansione dei pagamenti digitali, che aveva raggiunto il 44% degli adulti, veniva considerata come indicatore positivo di crescita e perciò si decise di studiare questo mercato emergente.

Venne costruito un prodotto consistente in un micro-finanziamento di 180\$ per un anno e si studiò l'impatto che i dati alternativi insieme alle nuove tecniche di analisi avanzata potevano avere sui costi di distribuzione del prodotto. In particolare, per questo tipo di prodotto, che prevedeva un tasso di interesse mensile del 6%, il risparmio finale sui costi oscillava tra il 20% e il 30%. I miglioramenti più drastici si ottenevano da un abbattimento del 30-40% nelle spese di marketing per segmentare la popolazione, raggiungendo allo stesso tempo un hit-rate più alto; si riducevano del 30% circa i costi di richiesta del finanziamento attraverso la digitalizzazione e l'automazione dei processi ottenendo contemporaneamente uno snellimento del percorso applicativo; infine si notava un miglioramento nel tasso di perdita che oscillava tra il 25-40%, ottenuto attraverso una migliore segmentazione e una valutazione più accurata del rischio. Il modello iniziale presentava un costo di distribuzione di 66\$; attraverso i dati alternativi il prodotto venne offerto a 38\$ raggiungendo il 25% della popolazione adulta studiata (CGAP, 2014).

Economie consolidate così come paesi emergenti, sono densamente popolate di piccole e micro-imprese che non hanno accesso ai mercati finanziari. Raggiungere, attrarre e servire in maniera adeguata queste piccole realtà insieme ai milioni di cittadini esclusi, porterebbe ad una doppia vittoria per le banche e per gli istituti finanziari: da una parte nuovi introiti, dall'altra dividendi sociali maggiori.

In modo da capire effettivamente l'impatto potenziale per le banche, Moreno, Vazirani e Lillis (2015), di Accenture, hanno modellato i comportamenti finanziari dei consumatori e delle imprese già servite dal sistema, applicandoli ai cittadini e alle piccole imprese escluse (Figura 6). Essi hanno stimato che, insediandosi nei mercati emergenti, il sistema bancario possa generare almeno 380 miliardi di dollari in redditi annuali entro il 2020: per prima cosa, colmando il divario creditizio delle piccole imprese, proponendo tassi medi ai mercati e spese per commissioni costanti per i servizi erogati, scelta che potrebbe generare redditi per 270 miliardi di dollari; in seguito, includendo i cosiddetti non-valutati nel sistema finanziario e infine aumentando i livelli di spesa media per servizi nei paesi con redditi minori per ottenere altri 110 miliardi di dollari di introiti.

Figura 6



Moreno, Vazirani and Lillis, 2015

La maggior parte dei potenziali redditi chiaramente deriva dalla potenziale offerta di prodotti finanziari ai consumatori delle popolazioni dei mercati emergenti che non hanno ancora accesso al credito.

Tuttavia le banche non sempre sono in grado di stare al passo con le nuove tecnologie visti i grossi investimenti in strutture, immobili e gli alti costi fissi che presentano; senza contare la perdita di credibilità che molti istituti finanziari blasonati si troverebbero ad affrontare se

presentassero metodi di valutazione ed analisi basati su dati alternativi, provenienti dai social network e altri di cui, a causa della mancanza di trasparenza del machine learning, non si può dimostrare una chiara correlazione matematica con le variabili di decisione tradizionali.

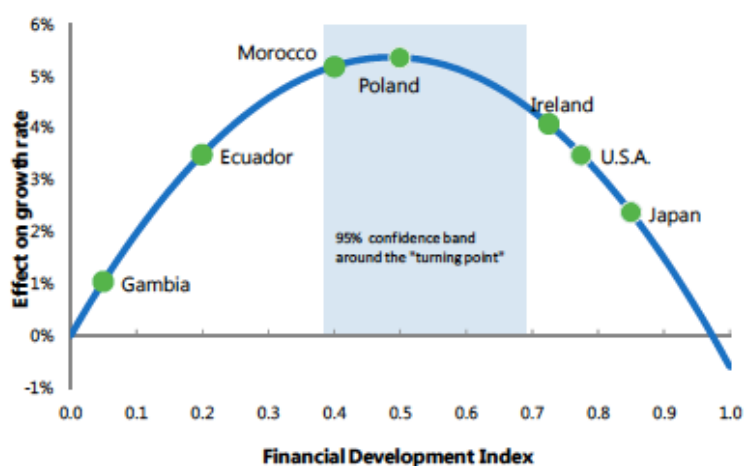
2.4 CONSEGUENZE ECONOMICHE NEGATIVE

Se da un lato, gli aspetti positivi dell'avvento dell'artificial intelligence e del machine learning nel credit scoring sono tanti, dall'altro i problemi che ne derivano sembrano essere trascurati o trattati con secondaria importanza. Diventa cruciale che le società, i consumatori, i legislatori e gli altri stakeholders comprendano e cerchino di limitare i rischi associati a questo sviluppo.

Sorgono problemi che risultano direttamente collegati a quello che potrebbe essere uno degli scopi principali per l'applicazione dell'artificial intelligence in campo finanziario, ovvero per raggiungere l'inclusione finanziaria.

Come mostrato da Sahay, Čihák and al. (2015) sulla base di un campione di 128 paesi nell'arco di 33 anni, dal 1980 al 2013, lo sviluppo e l'inclusione finanziaria aumentano la crescita economica, ma gli effetti si riducono per alti livelli di sviluppo finanziario e all'aumentare dello stesso diventano negativi. In questo contesto l'indice di sviluppo finanziario, FD, riassume la struttura delle istituzioni e dei mercati finanziari in termini di: profondità, intesa come grandezza e liquidità disponibile; accessibilità, ovvero la semplicità con cui individui e società possono accedere ai servizi finanziari e infine l'efficienza, ovvero l'abilità degli istituti di offrire servizi a basso costo con una redditività sufficiente. Il livello di sviluppo finanziario al di sopra del quale gli effetti positivi sulla crescita iniziano a diminuire è dimostrato essere tra 0,4 e 0,7 ad un intervallo di confidenza del 95% (Figura 7).

Figura 7

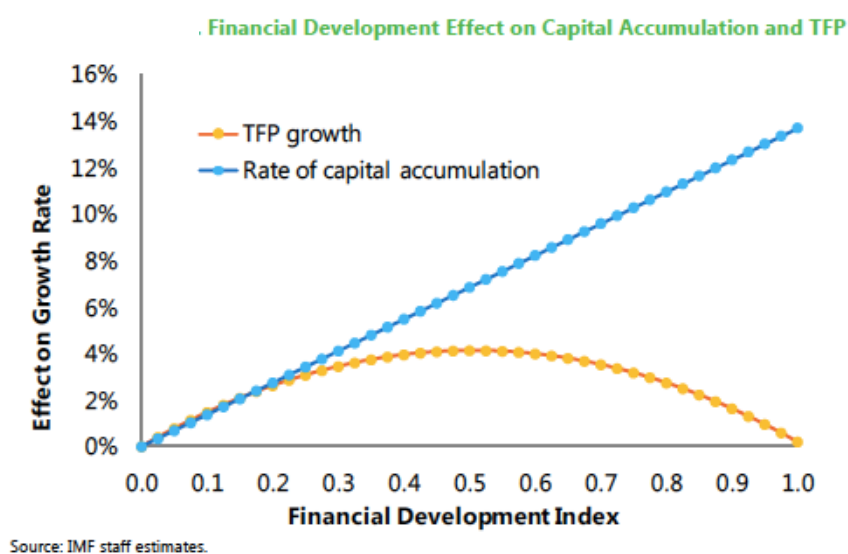


Source: IMF staff estimates.

Concettualmente ci sono diversi canali attraverso i quali un alto livello di sviluppo finanziario potrebbe avere un effetto negativo sulla crescita economica.

Philippon e Reshef (2012) dimostrano che un livello di sviluppo finanziario troppo alto comporta una divergenza di talento e capitale umano dai settori produttivi verso il settore finanziario e questo comporta uno squilibrio del mercato. Guardando a due componenti della crescita economica, l'accumulazione di capitale e la produttività totale dei fattori (TFP), l'evidenza empirica suggerisce che l'effetto di un esagerato tasso di sviluppo finanziario ha impatti negativi sulla crescita della curva di produttività totale dei fattori. Come illustrato nella Figura 8, i risultati indicano che un alto tasso di FD non impedisce l'accumulazione di capitale ma comporta una perdita di efficienza negli investimenti. In altre parole, diverse funzioni del settore finanziario, come la mobilitazione dei risparmi e la facilitazione delle transazioni potrebbero rimanere intaccate ad alti livelli di sviluppo finanziario ma altre funzioni, come l'efficienza nell'allocazione del capitale tra i fattori produttivi, potrebbero risentirne comportando una discesa negativa della curva.

Figura 8



Inoltre, uno dei rischi maggiori è la crescita di piattaforme secondarie e indipendenti dalle tradizionali banche di prestito ai consumatori; l'aumento sproporzionato dell'ambiente competitivo potrebbe erodere la profittabilità delle banche. In particolare, l'avvento di queste nuove piattaforme potrebbe sgretolare alcuni rami tradizionali e erodere le basi di reddito, aumentando la vulnerabilità delle banche alle perdite, indebolendo contemporaneamente il loro accesso a finanziamenti esterni ed interni. Allo stesso tempo, per rimanere competitive queste banche potrebbero vedersi costrette ad abbassare gli standard minimi per l'accettazione di prestiti, o potrebbero rispondere abbattendo i costi e prendendosi carico di rischi operativi e

reputazionali maggiori per stare al passo con i nuovi sistemi automatizzati. L'avvento di queste nuove tecnologie ha apportato un incremento sostanziale nella velocità del tasso di cambiamento a cui le banche devono adeguarsi, costringendo alcune di esse ad assumersi rischi senza troppe analisi approfondite; questo potrebbe ridurre la fiducia verso i tradizionali sistemi bancari (Schindler, Knobloch, and al. 2017).

L'abilità delle istituzioni finanziarie di utilizzare i big data provenienti dalle nuove fonti potrebbe portare ad una maggiore dipendenza da variabili macroeconomiche precedentemente non considerate. Diventa importante la comprensione di settori non finanziari come l'e-commerce, la sharing economy e il FinTech credit, inteso come tutte le attività collegate al credito facilitate dalla presenza di piattaforme elettroniche dove i mutuatari sono direttamente collegate con i prestatori di capitale (Dudley and Knot, 2017). In generale, le maggiori interconnessioni che si andrebbero a creare nel sistema finanziario potrebbero aiutare a condividere meglio il rischio così come potrebbero fungere da capro espiatorio nel caso di shock finanziari nazionali. Allo stesso tempo interconnessioni largamente diffuse andrebbero a aumentare l'impatto di shock estremi in tutto il sistema globale.

Un'altra questione di fondamentale importanza è la cosiddetta "third-party reliance" (Wilkins and al. 2017, p.23). Se un segmento cruciale delle istituzioni finanziarie si affida alle medesime fonti di dati a livello globale, o alle stesse strategie algoritmiche, allora in presenza di condizioni di mercato sfavorevoli, uno shock per tali fonti di dati condivisi o per le strategie adottate, potrebbe colpire tutto il sistema come se fosse una singola unità con una gravità maggiore. Ciò potrebbe verificarsi anche se, in apparenza, il sistema è costituito da decine, centinaia o addirittura migliaia di istituzioni finanziarie giuridicamente indipendenti. Di conseguenza l'adozione collettiva di queste tecnologie ad apprendimento automatico potrebbe introdurre nuovi rischi o comunque interconnettere più facilmente quelli esistenti. (Schindler, Souza Moreas, and al., 2017). Un esempio sono i servizi di cloud computing offerti da un numero limitato di parti e adottati in larga scala, i quali potrebbero avere implicazioni significative qualora si presentassero problemi operativi. Disfunzioni di questi fornitori terzi, magari dovute a difficoltà operative, rappresentano una minaccia e una fonte di rischio sistemico tanto più forte quanto più esse collegano reciprocamente mercati o istituzioni di centrale importanza.

2.5 PROBLEMATICHE SOCIALI

Anche se le nuove tecnologie si propongono di incrementare a dismisura il grado di inclusione finanziaria, incontrando interessi e bisogni di consumatori unbanked, gli algoritmi

potrebbero nascondere al loro interno pregiudizi collegati alla mente umana che li progetta, portando a discriminazioni non desiderate o nel peggiore dei casi ad un rafforzamento degli stessi pregiudizi. A ciò si aggiunge una mancanza di legislazione e proibizioni per l'uso di dati nella sottoscrizione di crediti. I problemi potrebbero essere ancora più gravi qualora gli algoritmi discriminassero contro i consumatori meno disposti a condividere i propri dati online. Questo argomento apre una discussione sociale sull'equità e sul bisogno di trasparenza dei nuovi modelli, dei dati e ancor più delle fonti utilizzate, infine è necessario far chiarezza sul tema dei consensi al prelevamento e all'utilizzo dei dati personali.

Nel dicembre del 2012 Facebook rilevava un brevetto riguardante un metodo per l'autorizzazione e l'autenticazione degli individui basato sul loro network di conoscenti. La maggior parte delle invenzioni presenti nel brevetto discute le azioni giornaliere su come gestire equamente un social network ma, come quarta ed ultima forma di realizzazione descritta nel sommario del testo, il colosso digitale viene definito come prestatore di capitale: il concetto esplicitato in questa sezione del testo prevede che quando un individuo presenta una domanda di credito, il prestatore di capitale esamini il punteggio di credito dei membri del network di cui questo individuo fa parte, poi, se la media del punteggio creditizio del network raggiunge un minimo prestabilito, il prestatore continua il processo di approvazione altrimenti la domanda verrebbe rifiutata. In merito a questo brevetto sono state mosse diverse critiche secondo le quali, se Facebook implementasse un sistema di questo genere, farebbe risorgere le discriminazioni storiche, basando valutazioni finanziarie sul tipo di comunità con cui si è interconnessi ed escludendo a priori alcuni soggetti solo per una questione di amicizie o conoscenze (Hurley and Adebayo, 2016).

Mentre un obiettivo fondamentale del credit scoring moderno è quello di garantire al consumatore una maggiore consapevolezza di come il suo comportamento possa influenzare il punteggio di credito, i metodi alternativi presentano nuove sfide in termini di equità e trasparenza. In generale potrebbe non essere facilmente riscontrabile da consumatori e legislatori, quali specifiche informazioni vengano utilizzate nei metodi alternativi, come queste impattino l'abilità di un consumatore di garantire per un mutuo, come influenzino il prezzo dello stesso o quali comportamenti potrebbero essere attuati per migliorare le proprie condizioni di sottoscrizione e i prezzi pagati. Purtroppo, spesso e volentieri i consumatori non sanno nemmeno quali azioni siano oggetto d'analisi e quali potrebbero influire positivamente o negativamente nel punteggio finale.

In modo simile, le recenti innovazioni potrebbero creare rischi in campi quali la sicurezza, la privacy e la proprietà dei dati sebbene garantiscano un aumento della convenienza ai

consumatori; le piattaforme che si appoggiano all'utilizzo dei big data, potrebbero essere più vulnerabili a certi rischi operativi rispetto alle banche tradizionali; cyber-rischi dovuti alla loro dipendenza dai processi digitali. I recenti scandali di frodi e attacchi cyber hanno illustrato l'importanza di questi temi. Inoltre i metodi di scoring alternativi dipendono molto da terze parti, come fornitori di dati e di sistemi di clouding, rendendoli soggetti a possibili ripercussioni in caso di eventuali malfunzionamenti o interruzioni di tali servizi esternalizzati. Essendo strumenti nuovi e mai testati in un intero ciclo economico, è importante che vengano monitorati i rischi di liquidità e di credito per i mutui sottoscritti tramite le nuove piattaforme FinTech attraverso lo studio di dati alternativi (Brainard, 2016).

CAPITOLO 3 : ZESTFINANCE INC., COME IL CREDIT SCORING AVANZATO DIVENTA REALTA'.

Una delle realtà più prominenti nel mercato del credit scoring alternativo e dell'industria di sottoscrizione al credito è ZestFinance Inc. Fondata nel 2009 da Douglas Merrill, la società vende all'esterno metodi di big data credit scoring ai fornitori di crediti. Allo stesso tempo offre finanziamenti e già nel 2015 erano circa 100 000 i finanziamenti erogati. La società ad oggi è autorizzata ad operare esclusivamente per i consumatori di diversi paesi degli Stati Uniti.

ZestFinance ha un approccio basato sul fatto che ogni dato e ogni informazione è utilizzabile per valutare l'affidabilità creditizia: ZAML, Zest Automated Machine Learning, è il nome dato all'algoritmo che prevede di utilizzare il machine learning per valutare l'affidabilità al credito. Il punto di forza di questa società è il fatto di offrire un metodo che non soffra del problema di trasparenza; i modelli sono costruiti appositamente per garantire una spiegazione dei procedimenti e dei calcoli. La società giustamente non ha fornito informazioni dettagliate riguardo alle sue fonti di dati e agli algoritmi utilizzati, tuttavia i brevetti consultabili pubblicamente forniscono informazioni al riguardo.

Come si evince dal brevetto presentato da ZestFinance, il modello valutativo di proprietà della compagnia statunitense si articola in tre principali passaggi da seguire per costruire un sistema di valutazione: per prima cosa bisogna definire il problema da risolvere, ovvero impostare la propria definizione di persona affidabile al credito e specificare una variabile obiettivo che rappresenti il risultato che il soggetto si propone di prevedere; come seconda cosa deve avvenire la raccolta dei dati e la loro trasformazione in codici utilizzabili; infine si procede con lo sviluppo del modello e con il controllo della sua accuratezza tramite il training. Quest'ultimo è il processo attraverso il quale si allena l'algoritmo in modo che dia risposte corrette a seconda degli input inseriti e a seconda delle relazioni captate, è un aspetto caratterizzante delle nuove tecnologie e in particolare del machine learning.

3.1 DEFINIZIONE DEL PROBLEMA E SPECIFICAZIONE DELLE VARIABILI OBIETTIVO

Il processo per arrivare ad un modello che sia in grado di competere con problemi non strutturati come la valutazione del credito è un percorso complicato che richiede diverse iterazioni, tentativi e fallimenti.

Anche se il primo passaggio può sembrare ovvio, nel caso di problemi non strutturati come la previsione del creditworthiness, dove non c'è una singola risposta corretta, articolare una propria e quantificabile definizione del quesito da risolvere è cruciale (Abbot, 2012).

I data scientist devono tradurre il problema in una domanda a cui può essere associata una risposta espressa in termini formali, termini che un computer dev'essere in grado di interpretare. Perciò viene definito un risultato di interesse, comunemente definito come variabile obiettivo. La valutazione del credito pone molte sfide per questo primo passaggio poiché ci sono opzioni multiple per i data scientist che vogliono dare una definizione codificata di persona affidabile, stabilendo poi una variabile obiettivo forte, che possa venire utilizzata per le previsioni future.

Sfortunatamente vi è scarsa informazione su come le società che utilizzano metodi di scoring alternativo, in particolare ZestFinance, definiscano il concetto di persona affidabile al credito, così come c'è poca luce su quali siano le variabili obiettivo e le etichette di classe utilizzate per suddividere i possibili debitori. Nel testo del brevetto di proprietà ZestFinance infatti non vi è alcuna definizione citata.

3.2 RACCOLTA E TRASFORMAZIONE DEI DATI

Una volta identificata la variabile obiettivo e stabilite le etichette di classe necessarie ad identificare univocamente le classi di richiedenti, i data scientist devono raccogliere i dati degli individui per i quali i diversi risultati corretti sono già conosciuti. Queste informazioni costituiscono il training data che verrà usato per implementare il processo di machine learning che costituirà il modello finale. Il training data infatti, dà la possibilità al computer di capire a quale tipologia di dato è associata una determinata risposta; da qui attraverso il machine learning il computer estrapolerà un pattern per lo studio di dati non verificati. La visione comune suggerisce che maggiori sono i dati disponibili per l'allenamento degli algoritmi, maggiore sarà l'accuratezza del modello predittivo finale (Hurley and Adebayo, 2016).

ZestFinance afferma di raccogliere migliaia di dati per ogni individuo che analizza; questi dati vengono agglomerati in quattro principali categorie. La prima racchiude i dati del richiedente, dati che l'individuo fornisce tramite l'iniziale processo di iscrizione così come altre informazioni, come l'attività di ricerca web, che potrebbe essere prelevata dal dispositivo dell'individuo nello stesso momento in cui lui o lei richiedano un prestito online. Un esempio di informazione prelevata può essere la quantità di tempo spesa dall'individuo per leggere i termini e le condizioni di utilizzo della piattaforma: ciò si ritiene possa determinare il grado di responsabilità di un richiedente.

La seconda categoria, si riferisce ai dati proprietari dell'individuo, ovvero informazioni ottenute tramite database privati, governativi o tramite i data brokers come Acxiom, presentato in precedenza.

Nella terza categoria sono immagazzinati i dati pubblici delle persone, informazioni che ZestFinance ottiene da ricerche automatiche in Internet e da tecniche come il web crawling e il web scraping. La prima tecnica consiste in una tecnologia informatica utilizzata per scandagliare automaticamente le pagine web allo scopo di individuare collegamenti ipertestuali all'interno delle stesse. Attraverso parole chiavi e argomenti trattati, il software cataloga le pagine principali e gli indirizzi URL collegati in modo da offrire poi risultati delle ricerche in modo veloce e più accurato. La seconda tecnica invece scandaglia la rete attraverso un software che emula la navigazione nel web compiuta da utenti in carne ed ossa andando a prelevare determinate informazioni da differenti portali web con diversi scopi: per monitorare l'andamento di una promozione online o per la ricerca di dati sensibili da rivendere ad altri utenti.

Infine, la quarta categoria, raccoglie informazioni riguardo alle attività compiute nei social network, includendo dati aggregati derivanti dai post online del richiedente e da qualsiasi informazione grafica per uno o tutti i membri della rete sociale dell'individuo (Merrill and al., 2014).

Raccolti i dati, i data scientist li utilizzeranno come dati di training, traducendoli in codici decifrabili e processabili da un computer. Una volta che i dati sono stati resi utilizzabili, ZestFinance attraverso il machine learning, processa i dati trasformati per generare una pluralità di classi di scelta indipendenti tra loro, quest'ultime descrivono aspetti specifici di aggregati di richiedenti simili; il brevetto si riferisce a queste classi di scelta utilizzando il termine "metavariabiles" (Merrill and al., 2014, p. 5).

Queste meta-variabili indipendenti posizionano ciascun individuo in una determinata categoria attraverso l'inferenza da uno o più classi di dati trasformati. Per esempio, una meta-variabile può comparare il reddito di un richiedente con il reddito medio di individui della stessa zona geografica e con professioni simili e quindi effettuare un controllo di veridicità costruendo una classe di riferimento: si ritiene che la probabilità di trovare un outlier sia piuttosto bassa e che un richiedente che dichiara un reddito maggiore rispetto a quello di persone residenti nelle stesse zone geografiche con simili professioni, sia un segnale d'allarme.

Tuttavia, in questa fase di costruzione sorge il problema dei dati non neutrali. Se i credit scorers si affidano a metodi di raccolta che non sono in grado di catturare dei campioni

rappresentativi di tutti i gruppi sociali, alcuni di questi potrebbero subire trattamenti sfavorevoli o venire ignorati dai modelli finali dei creditori. Come sostenuto nel 2016 dall'Electronic Privacy Information Center, dato che i credit scorers che sfruttano i big data sono principalmente focalizzati sulla raccolta massiva di informazioni, la qualità totale delle stesse potrebbe subire conseguenze negative.

3.3 SVILUPPO DEL MODELLO FINALE ATTRAVERSO L'ANALISI DEI TRAINING DATA E LA SCELTA DELLE FUNZIONALITA' CHIAVE

A questo punto può iniziare il processo di machine learning. Non tutti gli input inseriti per allenare gli algoritmi si dimostrano rilevanti, infatti molti di questi vengono scartati dopo che il sistema riconosce che gli stessi non influenzano la variabile obiettivo.

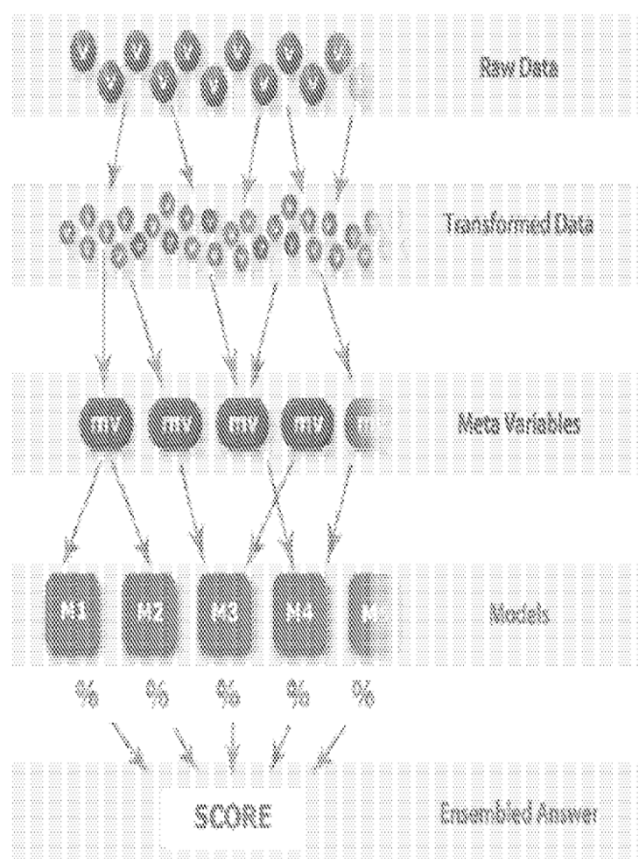
Il processo di machine learning include tipicamente un procedimento di ottimizzazione, attuato con lo scopo di identificare le variabili di input più significative e i relativi pesi appropriati da assegnare a ciascuna di esse.

Per sviluppare il modello finale, i data scientist utilizzano programmi capaci di ripetere diverse iterazioni e analizzare migliaia di combinazioni di data points in modo da identificare i fattori rilevanti che meglio si correlano con la variabile obiettivo, lo score finale; questo processo iterativo prende il nome di "feature selection" (Merrill and al., 2014, p. 8). Nel passaggio finale del modello di ZestFinance (Figura 9), le meta-variabili più rilevanti vengono inserite in più algoritmi statistici e finanziari ognuno con diverse abilità previsionali.

Un algoritmo chiamato Naive Bayes controlla se alcuni tratti individuali, ad esempio per quanto tempo i richiedenti hanno mantenuto un conto bancario attivo, aiutano a prevedere i casi di insolvenza. Un altro, chiamato Random Forest, divide i debitori in gruppi indipendenti tra loro senza caratteristiche preselezionate e incarica il deep learning di trovare dei modelli intrinseci nei dati.

I risultati degli algoritmi sono poi uniti in un punteggio da zero a cento che rappresenta lo score. Non è tuttavia dichiarato il valore dello score al di sopra del quale la richiesta di prestito viene accettata. Da qui fuoriesce la decisione finale di esborso o rifiuto; ogni volta che il modello viene applicato per valutare un nuovo consumatore vengono generati nuovi dati, essi poi vengono reinseriti nel processo di machine learning con lo scopo di sostenere il miglioramento del modello nel corso del tempo. Informazioni considerate irrilevanti in precedenza possono assumere nuovi significati e insegnare al sistema come processarle.

Figura 9



Merrill and Al., 2014 US Patent Appl.

Un richiedente che cambia residenza troppo spesso è considerato instabile, allo stesso tempo se dalle advanced analytics emerge che il richiedente non legge i termini e le condizioni applicate al prestito in questione, egli si dimostra imprudente. Lo stesso Merrill in un'intervista al Washington Post ha dichiarato che un persona che compila i campi del modulo richieste di ZestFinance in lettere maiuscole risulta essere più rischioso rispetto a chi utilizza anche le lettere minuscole; il fondatore, pur dimostrando la sua affermazione, ha poi confessato di non sapere il perché di ciò (Lippert, 2014).

Merrill nella stessa intervista ha confermato che usando il suo modello per distinguere i richiedenti in affidabili o meno, si ha un tasso di default del 15%, la metà della media dei cosiddetti payday loan.

Nell'intervista viene presentato un esempio di come ZestFinance, raccogliendo più di 10 000 dati informatici prende la decisione finale di approvazione del prestito (Figura 10). L'ipotesi si basa sull'analisi della richiesta di un'infermiera a confronto con la richiesta di una guardia carceraria.

Figura 10

<i>Nurse</i>		<i>Prison guard</i>
\$32,000	Reported income	\$65,000
\$800	Rent	\$1,200
2 in 10 years	Number of addresses	7 in 5 years
1	Social Security numbers	4
15 in 45 minutes	Number of ZestFinance Web pages accessed while completing application	3 in 7 minutes
Approved	Loan status	Rejected

Source: Lippert 2014

Una minore spesa per l'affitto spiega un rapporto entrate-uscite maggiore, collegato con una capacità di recupero da un possibile default più alta. Un numero minore di indirizzi di residenza indica una maggiore stabilità, mentre l'abilità del modello di individuare il numero di pagine web lette durante la richiesta di credito serve per distinguere consumatori attenti da consumatori superficiali.

Infondo alla tabella si legge lo stato del prestito: in base alle informazioni studiate la richiesta dell'infermiera viene accettata mentre quella della guardia carceraria verrà rifiutata.

CONCLUSIONE¹

Questa prova finale ha cercato di fare chiarezza su come le nuove frontiere della tecnologia, l'artificial intelligence e il machine learning, stiano subentrando nelle metodologie di credit scoring e di come si affianchino e talora si possano sostituire ai metodi più tradizionali e statici. A tal fine si è condotta una ricerca graduale che dapprima spiegasse questi nuovi concetti spesso fraintesi. Si è poi affiancato lo studio della nuova metodologia utilizzata, il machine learning, descrivendo e spiegando come questa nuova tecnologia applicata al credit scoring possa risolvere il problema dell'inclusione finanziaria. Grazie ai nuovi metodi intelligenti di valutazione dell'affidabilità di credito il panorama globale sta migliorando e come dimostrato nella prova, si stanno ampliando a dismisura le possibilità di piccole imprese e consumatori di beneficiare di sistemi finanziari costruiti su misura così come di tariffe eque e individuali.

E' stato constatato come la questione dei dati e delle 'tracce digitali' sia di fondamentale importanza in questo contesto. Qualsiasi tipologia di azione compiuta offline ed online ha delle conseguenze in termini di informazioni. Senza che ce ne rendiamo conto, il più piccolo movimento viene studiato e può influenzare le nostre opportunità future, le nostre offerte e i consigli che riceviamo sul web. In particolare, è stato spiegato il fenomeno dei big data e di come vengano utilizzati nella valutazione dell'affidabilità al credito. Mi ha colpito particolarmente trovare un sistema ben impostato e allo stesso tempo molto sconosciuto e profondo. Si è parlato di un mercato di compravendita dei dati di cui gran parte delle persone non sono a conoscenza e con questo tema si è aperto il discorso etico e morale. I problemi di privacy, proprietà dei dati e diffusione degli stessi è un tema a parere mio sottovalutato da molti; purtroppo essendo queste tecnologie un fenomeno di recente applicazione, credo che le fonti di informazioni disponibili siano solo la punta dell'iceberg e finché questi sistemi non vengono compresi e studiati a fondo non si può sapere quanto effettivamente possano essere nocivi per la sicurezza mondiale.

In ogni caso i benefici riscontrati finora sono molteplici sia dalla parte dei consumatori che dalla parte dei produttori. Grazie ai nuovi metodi per la valutazione dell'affidabilità al credito è stato dimostrato un incremento in termini di efficienza, una segmentazione del mercato più efficace e un passo avanti per garantire a tutti i cittadini un accesso a fonti di finanziamento esterne. Si è parlato infatti di metodi intelligenti di valutazione che possono colmare il gap per quei consumatori considerati unbanked e underbanked, utilizzando dati alternativi e big data analytics laddove mancano le informazioni tradizionalmente raccolte dalle agenzie di reporting.

Essendo innovazioni recenti, è fondamentale che la novità sia affrontata con cautela e che siano sondati diversi scenari prima di procedere a conclusioni affrettate. Attualmente la direzione comune è quella di soffermarsi principalmente su benefici e profitti apportati da queste nuove possibilità. E' tuttavia importante comprendere che ci sono diversi rischi collegati e che a causa della scarsa informazione i consumatori in generale, non sono nemmeno a conoscenza delle conseguenze delle loro azioni in questo sistema globale così interconnesso. Focalizzarsi solo sulla novità in sé e sulle potenzialità benefiche dei big data nel credit scoring, dando per assodata la loro affidabilità, potrebbe rivelarsi una catastrofe in quanto è in gioco la privacy di miliardi di persone. Inoltre, date le innumerevoli interconnessioni del mondo moderno, rischi e problemi di natura locale potrebbero velocemente diffondersi ed ampliarsi a livello globale.

L'applicazione delle nuove tecnologie con l'obiettivo di ottenere benefici economici e sociali maggiori, risulta essere ancora in fase di sviluppo e come tale presenta i classici problemi di strutturazione e di novità che qualsiasi innovazione comporta. Sarà importante che i regolatori e le autorità preposte siano in grado di affrontare le sfide in modo tale da non rendere l'innovazione una mera questione di profitto. Supportando politiche di collaborazione e trasparenza a favore di un cittadino informato e consapevole, le autorità dovranno essere capaci di punire le scelte mirate al solo rendimento. I responsabili che affronteranno queste problematiche, dovranno essere in grado di creare un ambiente idoneo allo sviluppo di sistemi chiari e dettagliati, modelli che siano in grado di spiegare concettualmente le relazioni matematiche subordinate alle nuove tecnologie, altrimenti si arriverà ad un punto in cui ci saranno computer molto efficienti ma indecifrabili.

RIFERIMENTI BIBLIOGRAFICI E SITOGRAFICI

BIBLIOGRAFIA

- ACXIOM, 2018. *Acxiom Data Catalogue, For Audience Creation And Analytics*, Acxiom, Maggio.
- BERG, T., BURG, V., GOMBOVIĆ, A., MANJU, P., 2019. *On the Rise of the FinTechs - Credit Scoring using Digital Footprints*.
- BRAINARD, L., 2018. *What Are We Learning about Artificial Intelligence in Financial services ?*, Board of Governors of the Federal Reserve System at, *Fintech and the New Financial Landscape*. Philadelphia, Pennsylvania.
- COSTA, A., DEB, A., KUBZANSKY, M., 2016. *Big Data, Small Credit. The Digital Revolution and Its Impact on Emerging Market Consumers*, Omidyar Network, 20 Jun.
- EBA, EIOPA and ESMA, 2016, *Joint Committee Discussion Paper on the use of Big Data by Financial Institutions*.
- GURNÝ, P. AND GURNÝ, M. 2013. *Comparison of Credit Scoring Models on Probability of Default Estimation for Us Banks*, *Prague Economic Paper*.
- HURLEY, M. AND ADEBAYO, J., 2016. *Credit Scoring In the Era of Big Data*, *Yale Journal of Law & Technology*, Vol. 18, article 5.
- INTERNATIONAL ORGANIZATION OF SECURITIES COMMISSIONS (IOSCO), 2017. *Research Report On Financial Technologies (Fintech)*, February.
- LUNT, C., 2012. *Authorization and authentication based on an individual's social network*. U.S. Patent Application No. 9,100,400 filed Aug. 2.
- MERRILL, C. D. AND AL., 2014. *System and method for building and validating a credit scoring function*. U.S. Patent Application No. 14/276,632, May 13.
- MORENO, J. P., VAZIRANI, M., LILLIS, S., 2015. *Billion Reasons to Bank Inclusively*, Accenture Banking.
- ROBINSON AND YU, 2014. *Knowing the Score: New Data, Underwriting, and Marketing in the Consumer Credit Marketplace*, October 29.
- SAHAY, R., ČIHÁK, M. AND AL., 2015. *Rethinking Financial Deepening: Stability and Growth in Emerging Markets*, International Monetary Fund, May.

- SCHINDLER, J., KNOBLOCH, C., AND AL. 2017. *FinTech Credit, Market Structure, business models and financial stability implications*, FSB and CGSF, May 22.
- SCHINDLER, J., SOUZA MOREAS, A. AND AL., 2017. *Artificial Intelligence and Machine Learning in Financial Services - Market Developments and Financial Stability Implications*. The Financial Stability Board.
- SKANTZOS, N., CASTELEIN, N., 2016. *Credit scoring - Case study in data analytics*. Deloitte.
- THE HARRIS POLL, 2019. *A Flawn In The System: Are Credit Scores Undermining The American Dream*.
- WILKINS, C. AND AL. 2017. *Financial Stability Implications from Fintech: Supervisory and Regulatory Issues that Merit Authorities' Attention*, The Financial Stability Board, June 27.
- WORLD BANK GROUP, 2018. *Discussion Note: Financial Consumer Protection and New Forms of Data Processing Beyond Credit Reporting*, November.

SITOGRAFIA

- ABBOT, D. 2012. Why Defining the Target Variable in Predictive Analytics is Critical, Apr. 5. Disponibile su < <http://abbottanalytics.blogspot.com/2012/04/why-defining-target-variable-in.html> > [08/07/2019]
- ABOUT US: LENNDO WEBSITE Disponibile su < <https://www.lenddo.com/about.html> > [07/06/2019]
- ANON., 2013. Credit scores, Stat Oil, Lenders are turning to social media to assess borrowers. *The Economist* [Online], Feb 9, Disponibile su <https://www.economist.com/finance-and-economics/2013/02/09/stat-oil> [12/07/2019]
- BARNES, P., 2016. Smartphones Are Changing Emerging Markets' Business Landscape, *Market Realist* [Online]. Disponibile su <https://marketrealist.com/2016/12/smartphones-changing-ems-business-landscape/> [28/05/2019]
- CHEN, G., FAZ, X. 2015. *The Potential of Digital Data: How Far Can It Advance Financial Inclusion?*. Focus Note 100 Washington, D.C.: CGAP, Gennaio. Disponibile su < https://www.cgap.org/sites/default/files/Focus-Note-The-Potential-of-Digital-Data-Jan-2015_1.pdf > [15/06/2019]

- CHENEY, C., 2016. How Alternative Credit Scoring Is Transforming Lending in the Developing World, *Devex*, [Online] Disponibile su <https://www.devex.com/news/how-alternative-credit-scoring-is-transforming-lending-in-the-developing-world-88487> [15/05/2019]
- CONSULTING GROUP TO ASSIST THE POOR, 2014. *Projecting Impact of Non-Traditional Data and Advanced Analytics on Delivery Costs*. Disponibile su <<https://www.slideshare.net/CGAP/projecting-impact-of-nontraditional-data-and-advanced-analytics-on-delivery-costs>> [15/06/2019]
- CULL, R., TILMAN, E., HOLLE, N., 2014. *Financial Inclusion and Recent Development: Recent Impact Development*, CGAP Focus Note no. 92. Disponibile su <http://documents.worldbank.org/curated/en/269601468153288448/pdf/881690BRI0FN920Box385210B000PUBLIC0.pdf> [01/06/2019]
- ELECTRONIC PRIVACY INFORMATION CENTER, 2016. *Credit Scoring* [Online]. Disponibile su <<https://epic.org/privacy/creditscoring>> [10/07/2019]
- FICO, 2019 *What's not in my FICO scores*, MY FICO [Online]. Disponibile su <https://www.myfico.com/credit-education/credit-scores/whats-not-in-your-credit-score> [29/06/2019]
- LIPPERT, J., 2014. ZestFinance issues small, high-rate loans, uses big data to weed out deadbeats. *The Washington Post* [Online], Disponibile su <https://www.washingtonpost.com/business/zestfinance-issues-small-high-rate-loans-uses-big-data-to-weed-out-deadbeats/2014/10/10/e34986b6-4d71-11e4-aa5e-7153e466a02d_story.html?noredirect=on> [29/08/2019]
- LOHR, S., 2015. Big Data Underwriting for Payday Loans, *New York Times* [Online]. Disponibile su http://bits.blogs.nytimes.com/2015/01/19/big-data-underwriting-for-payday-loans/?_r=0 [01/07/2019]
- MARR, B., 2018. 10 Amazing Examples Of How Deep Learning AI Is Used In Practice?, *Forbes* [Online], Disponibile su <<https://www.forbes.com/sites/bernardmarr/2018/08/20/10-amazing-examples-of-how-deep-learning-ai-is-used-in-practice/#6e158460f98a>> [11/08/2019]
- MCLANNAHAN, B., 2015. Being 'wasted' on Facebook may damage your credit score, *Financial Times* [Online]. Disponibile su <<https://www.ft.com/content/d6daedee-706a-11e5-9b9e-690fdac72044#axzz3pPUmUqGJ>> [01/08/2019]

- NEMSCHOFF, M., 2014. A Quick Guide to Structured and Unstructured Data, *Smart Data Collective* [Online]. Disponibile su <https://www.smartdatacollective.com/quick-guide-structured-and-unstructured-data/> [10/05/2019]
- PHILIPPON, T. AND RESHEF, A., 2012. Wages and Human Capital in the U.S. Finance Industry: 1909–2006. *Quarterly Journal of Economics*. Disponibile su <<https://academic.oup.com/qje/article-abstract/127/4/1551/1843161?redirectedFrom=fulltext>> [29/07/2019]
- RAMIREZ, E., Federal Trade Commission Chairwoman, *Protecting Privacy in the Era of Big Data*, Remarks in the International Conference on Big Data from a Privacy Perspective. Honk Kong June 10, 2015. Disponibile su <https://www.ftc.gov/system/files/documents/public_statements/671661/150610era_bigdata.pdf> [1/08/2019]
- TECHTERMS, 2012. The Tech Terms Computer Dictionary, December 15. Disponibile su <<https://techterms.com/definition/exabyte>> [20/06/2019]
- YSAGUIRRE, D., 2013. Banking the Unbanked: A How-To, *Forbes* [Online], June 14. Disponibile su <<https://www.forbes.com/sites/ashoka/2013/06/14/banking-the-unbanked-a-how-to/#77cdd4ce5727>> [01/05/2019]

ⁱ Lunghezza elaborato: 11041 parole.