



*Dipartimento di Impresa e Management
Cattedra Tecnologie Digitali e
Cambiamenti Organizzativi e Sociali*

Community Sentiment e l'andamento dei Mercati Finanziari

RELATORE
Prof. Stefano Za

CANDIDATO
Stefano Ciccarelli
Matr. 198101

ANNO ACCADEMICO 2017/2018

Community Sentiment e l'andamento dei Mercati Finanziari

0.	<i>Introduzione</i>	4
1.	<i>Capitolo Primo: le Nuove Tecnologie Digitali</i>	5
1.1.	<i>Data Science</i>	5
1.2.	<i>Big Data e IOT</i>	7
1.3.	<i>Machine Learning e Deep Learning</i>	12
1.4.	<i>Esperienza Umana, il fulcro dell'economia</i>	15
2.	<i>Capitolo Secondo: Le Tecnologie ed i Mercati Finanziari</i>	19
2.1.	<i>Le Idee e le Tecnologie Digitali</i>	19
2.2.	<i>Finanza Comportamentale e Social Network: il caso OpenBook</i>	24
2.3.	<i>Fintech e Hybrid Robo Advisoring</i>	32
2.4.	<i>Sentiment Analysis e Mercati Finanziari</i>	44
3.	<i>Capitolo Terzo: Analisi Finanziaria con i Big Data</i>	47
3.1.	<i>Sentiment Analysis: un'Applicazione Pratica in Python</i>	47
3.2.	<i>Analisi Finanziaria tramite il Machine Learning</i>	58
3.3.	<i>Analisi dei Differenziali di Prezzo Sentiment-Based</i>	68
4.	<i>Conclusione</i>	78
5.	<i>Note</i>	80
5.1.	<i>Note Capitolo 1</i>	80
5.2.	<i>Note Capitolo 2</i>	81
6.	<i>Bibliografia</i>	82

Ai miei Genitori

*Un ringraziamento speciale a Gian Luca Comandini per il sapiente lavoro di
divulgazione delle Tecnologie Finanziarie*

0. Introduzione

L'Universo è un sistema continuo, complesso e caotico in costante evoluzione.

Mentre le scienze naturali pongono le proprie fondamenta su basi solide, certe e validate, le scienze sociali devono costantemente adattarsi e combattere contro l'evoluzione dinamica della società, della storia umana e soprattutto contro l'assenza di dati su cui effettuare delle assunzioni o quantomeno delle ipotesi.

Così come Galileo Galilei grazie al telescopio e alla propria resilienza è diventato l'ideatore del "metodo scientifico" e della "fisica moderna", oggi giorno i neo-scienziati economici grazie ai Big Data e alle Tecnologie Digitali, come ad esempio l'Intelligenza Artificiale, possono tracciare l'evoluzione dinamica dei sistemi sociali caotici e complessi, rilevando così pattern e regolarità sottostanti all'Intelligenza Collettiva che prima d'ora sfuggivano alla cognizione umana a causa dei limiti della stessa.

Questo Trend non costituisce solamente un supporto a questa disciplina, ma risulta essere un vero e proprio Game Changer, infatti, la determinazione e diffusione delle Tecnologie Digitali si prospetta essere un effettivo cambio di paradigma collettivo, che può rendere esponenziale tramite la diffusione dei Devices Personali e la progettazione delle Smart Cities il nuovo benessere economico globale.

Per assicurarsi che la crescita ed il progresso siano però inclusivi e diffusi tra la popolazione, è necessario che la consapevolezza di questi cambiamenti si diffonda in maniera trasversale tra chi partecipa al processo economico collettivo, o il rischio dell'avvento di un monopolio cognitivo potrebbe diventare una cruda realtà.

Per queste ragioni, la mia Tesi si incentra sull'analisi delle Tecnologie Digitali principali che porteranno un maggiore impatto a livello economico e su come le decisioni degli operatori economici, anche a livello cognitivo-comportamentale, evolvono impattando sui mercati finanziari, impiegando come campione le tracce lasciate da quest'ultimi online.

1. Capitolo Primo: le Nuove Tecnologie Digitali

1.1. Data Science

L'Economia è una scienza sociale fondante su variabili caotiche.

Spesso queste variabili sulle quali si pone questa nobile scienza sono difficili da identificare, e ancor più da misurare analiticamente nel tempo.

La recente rivoluzione digitale che ha reso democratico l'utilizzo di Smartphone e la ormai prossima generazione di Smart Devices su cui si fonda il nascente ecosistema IOT hanno consentito però un progresso inaspettato in questa disciplina spesso ostica da analizzare.

Infatti, dall'unione dell'intuizione creativa e dell'analisi metodica e scientifica è nato un nuovo universo: la Data Science.

Quest'ultima, spesso identificata più appropriatamente nei suoi "evangelist" quali i Data Scientist, è frutto di un ibrido rivoluzionario che prevede l'applicazione di competenze statistiche e computazionali a questioni spesso distanti da discipline considerate così tecniche.

Il Data Scientist, a discapito del nome è sia un artista, in quanto utilizza la creatività e l'intuizione, che uno scienziato, in quanto applica il metodo scientifico tramite la programmazione e la statistica.

Il proliferare di questa professione non è però affatto casuale, il cui manifestarsi è stata una recente novità congiunturale che è dipesa per lo più da 3 fattori:

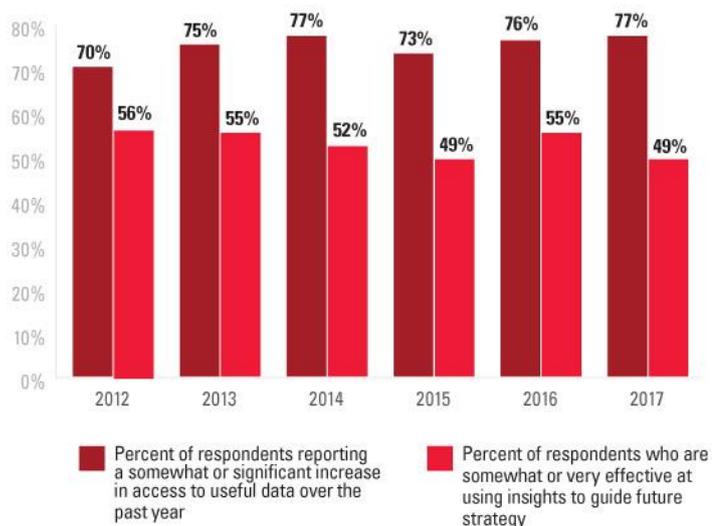
- Aumento esponenziale della quantità di dati disponibili online
- Aumento esponenziale della potenza di calcolo dei Devices
- Necessità di soluzioni eleganti a questioni sempre più complesse per la mente umana

Il coincidere di questi elementi ha portato al successo improvviso di questo settore, al punto da spingere Harvard Business Review a definire nell'ottobre del 2012 il Data

Scientist come “La professione più sexy del XXI secolo” [1], identificando questa figura professionale come una tra le più richieste rispetto alla sua effettiva disponibilità e paragonandola agli analisti quantitativi (“quants”) di Wall Street degli anni 80’ e 90’.

I dati sembrano dunque essere una vera e propria miniera d’oro, ma ad essere precisi, la vera ricchezza non è nei dati in sé, che perlopiù tendono ad essere neutri ai fini del value management, ma nell’abilità di saper trarne conoscenza, ovvero insight utili al processo decisionale aziendale, affinché le strategie aziendali si fondino su solide fondamenta oggettive piuttosto che su assunzioni soggettive prive di validazione o peggio sui biases cognitivi dei singoli manager.

La corsa all’oro deve dunque essere attuata con metodo e razionalità, evitando che le aziende finiscano per voler “incastrare” professionalità e tecnologie in contesti in cui non è necessario, generando inutili costi e lasciandosi trasportare dalla foga e dal hype. È fondamentale dunque che le Aziende, prima di analizzare le tattiche necessarie a rendere operativa una tecnologia all’interno di un contesto organizzativo, analizzino le strategie più efficienti tramite le quali le Tecnologie ed i Sistemi Informativi possano creare valore per l’azienda sul lungo termine.



Fonte: <https://sloanreview.mit.edu/projects/using-analytics-to-improve-customer-engagement/>

1.2. Big Data e IOT

I Big Data sono un fenomeno molto recente, equivalente al processo di accumulo delle informazioni disponibili, che appunto diventano grandi quantità di dati caratterizzati sia dal volume, dalla varietà, dalla veridicità che dalla velocità (le “4V”[2]), e che è esploso grazie alla diffusione esponenziale e democratica di devices intelligenti a basso costo. Tra quest’ultimi è frequente che lo Smartphone ne sia il rappresentante di categoria per eccellenza, anche se spesso in realtà sono innumerevoli i devices diversi da questo che possono essere potenzialmente definiti intelligenti.

Cisco, azienda pionieristica in questo settore, ha infatti coniato inizialmente un termine per definire l’intero ecosistema, attuale e soprattutto futuro, di Devices intelligenti: IOE, l’Internet of Everything, conseguentemente rinominato in IOT, dunque Internet of Things.

Questo ecosistema viene accomunato dalla capacità dei suoi singoli componenti di poter rilevare tramite sensori degli Input dall’ambiente circostante secondo parametri ben precisi, producendo generalmente Dati Strutturati, e dalla capacità di trasmettere quest’ultimi online.

O più precisamente, il Professore Stefano Za definisce questo fenomeno *“Come una rete di oggetti interconnessi tra loro capaci di raccogliere e scambiare informazioni attraverso l’uso di Internet”* [3].

Dunque, con l’aumento della qualità e quantità dei sensori, della potenza di calcolo in loco (Fog Computing) e della banda larga (Cloud Computing), è possibile determinare che il Trend dei Big Data è solo al suo inizio, in quanto nei prossimi anni avremo un’evoluzione di questo sistema che sarà sempre più presente e determinante, in maniera visibile o invisibile, nelle nostre vite.

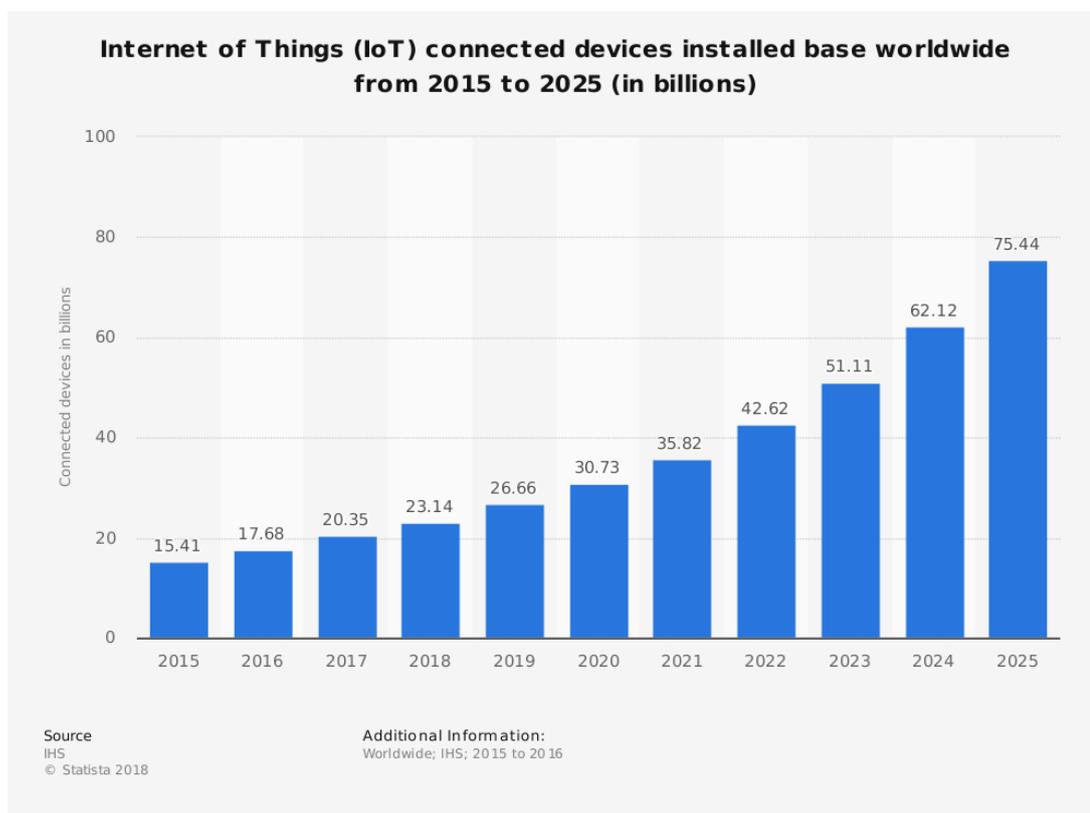


Grafico 1.1

È di fondamentale importanza rilevare che non solo quindi il numero di devices iscrivibili all’ecosistema IOT, stimati pari a 75,44 miliardi di unità nel 2025 (Grafico 1.1), sarà determinante alla creazione di valore aziendale tramite i dati da essi prodotti, ma anche il contemporaneo raffinamento degli algoritmi di Cognitive Computing avrà un ruolo di primaria importanza, in quanto quest’ultimi grazie ai dati ad essi forniti riescono ad emulare su larga scala il processo decisionale umano, effettuando un’analisi top-down, semplificando ed automatizzando alcuni processi aziendali time consuming.

Quest’ultimo fattore non solo consentirà un incremento di produttività per quelle funzioni aziendali routinarie e ripetitive, ma anche per quelle che necessitano di capacità intuitive e creative.

Infatti, il Machine Learning ed il Deep Learning, due sottogruppi del Cognitive Computing, offrono agli Smart Devices la potenzialità di poter apprendere dai dati tramite la rilevazione di pattern al loro interno.

Ciò fornisce, in un contesto organizzativo, non solo la possibilità di ottenere un quadro prospettico più lucido su ciò che è accaduto e sta accadendo in tempo reale all'interno dell'azienda, ma anche la potenzialità di ottenere delle previsioni probabilistiche su scenari futuri, consentendo dunque di poter prendere delle decisioni che consentano di ottimizzare il valore aziendale, minimizzando ed anticipando dunque scenari negativi incerti ed imprevisi.

Di questo fenomeno dunque ne potranno beneficiare tutti gli attori che partecipano al sistema economico globale, in quanto tramite la riduzione dell'incertezza, dunque dei rischi, all'interno dei singoli contesti organizzativi sarà possibile ridurre la variabilità del ciclo economico su scala macroeconomica, a favore della domanda e dell'occupazione globale.

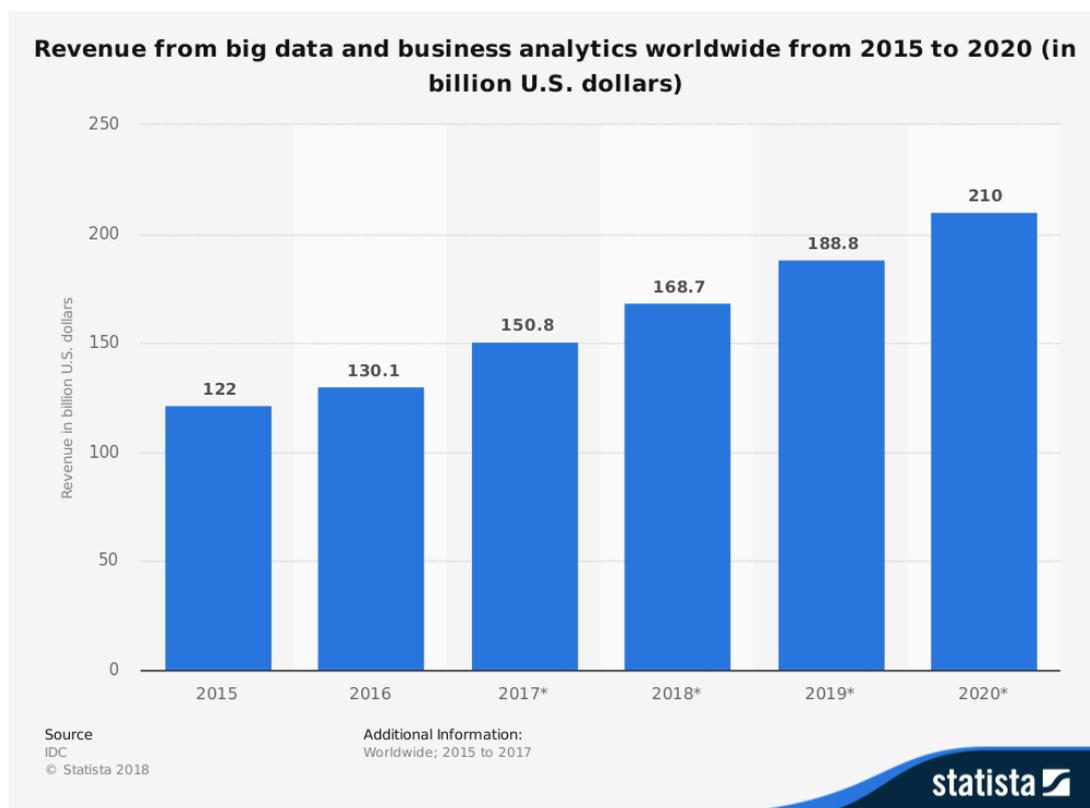


Grafico 1.2

Il Grafico 1.2 mostra, infatti, come dal 2015 al 2020 i ricavi diretti derivanti dai mercati a livello globale che si stanno sviluppando grazie ai Big Data e Business Analytics incrementeranno di circa il 72%, equivalente ad una crescita media annua attesa di circa l'11,45 %, un Trend che dunque apparentemente sembra poter offrire un boost alla redditività finanziaria di qualsiasi Azienda i cui Business sono determinati dai dati e dall'incertezza operativa e finanziaria, tutto questo senza contare che secondo McKinsey in alcuni casi meno del 1% dei dati disponibili viene effettivamente utilizzato, spesso ignorando l'ottimizzazione e la previsione a favore unicamente dell'analisi [4].

Inoltre, sempre secondo gli analisti di McKinsey l'impatto indiretto nella creazione di valore derivante dal IOT sarà in un range stimato tra 3900 \$ miliardi e i 11500 \$ miliardi l'anno nel 2025 analizzando tutti i principali settori dell'economia tramite 150 casi d'uso.

Where is the value potential of the Internet of Things?



Interoperability required to capture 40% of total value



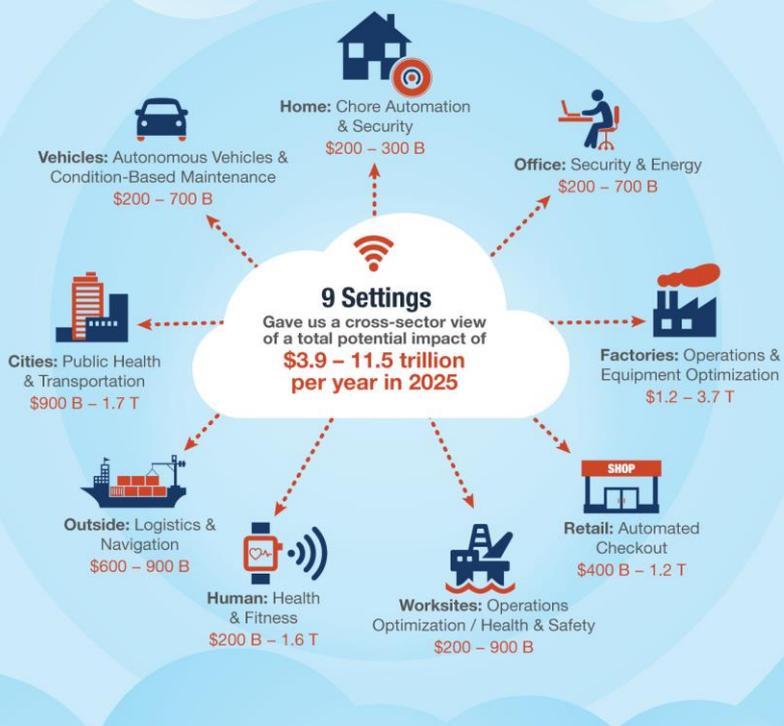
< 1% of data currently used, mostly for alarms or real-time control. More can be used for optimization & prediction.



2X more value from B2B applications than B2C



Developing: 40% / Developed: 60%



Types of Opportunities

Transform business processes
Predictive maintenance, better asset utilization, higher productivity



Enable new business models
E.g., remote monitoring enables anything-as-a-service

1.3. Machine Learning e Deep Learning

I Big Data però non sono composti unicamente da dati strutturati prodotti direttamente dai devices, infatti, questi sono suddivisi tra:

- Dati Strutturati (Organizzati), ossia dati che hanno una struttura e dunque possono essere interpretati dal punto di vista oggettivo e riescono ad essere elaborati da un algoritmo con basse possibilità d'errore.
- Dati non Strutturati (Non Organizzati), ossia dati qualitativi, frequentemente prodotti dall'uomo, come un messaggio di testo o un comando vocale.

Quest'ultimi sono stimati essere tra l'80% ed il 90% dei dati complessivi e sono spesso sottoutilizzati (post Facebook, Twitter, e-mail e documenti vari), infatti, solo in alcuni casi sono applicate delle tecniche di pre-processing per potervi offrire una struttura.

Le principali tecniche di pre-processing impiegate dai Data Scientist sono in grado di ottenere delle rilevazioni strutturate con le seguenti metodologie:

- Enumerazione di vocaboli o intere frasi
- Rilevazione dell'eventuale presenza di caratteri particolari
- Misurazione della lunghezza assoluta e relativa del testo
- Classificazione degli argomenti e del sentiment (positivo / neutro / negativo)

Gli output che ne derivano possono consentire l'elaborazione e la modellizzazione dei dati in maniera efficiente e soprattutto misurabile, anche in caso di media vocali (tramite il Transcript del testo ottenuto dallo Speech Recognition), e video (Image Classification e Object Recognition).

I dati in sé non sono impiegabili ai fini della creazione di valore aziendale, ma è necessario che questi, una volta strutturati e definiti secondo le principali metodologie della Data Science, siano analizzati ed elaborati da degli algoritmi.

Secondo la definizione classica un algoritmo è una procedura che descrive i passaggi logici (ed ottimali) da eseguire per raggiungere una determinata soluzione tramite

determinati input, ma grazie al Machine Learning e al Deep Learning questa concezione degli algoritmi (nella loro accezione più ampia) sta radicalmente cambiando.

Infatti, un algoritmo di Machine Learning, tramite l'apprendimento sulla base degli esempi ad esso forniti (ossia i dati), è in grado di rendere "intelligente" una macchina, che di conseguenza sceglie la soluzione ottimale (generalizzazione estratta dai dati). Quest'ultimo infatti è in grado di rilevare dei pattern e delle regolarità statistiche all'interno di Data Set (insiemi di dati) ad esso forniti ed effettuare di conseguenza delle previsioni o delle classificazioni.

Questi pattern spesso sfuggono, a causa della loro quantità e complessità, alla cognizione umana e dunque l'uomo non sarebbe mai in grado, senza gli algoritmi di Machine Learning appunto, di rilevarli.

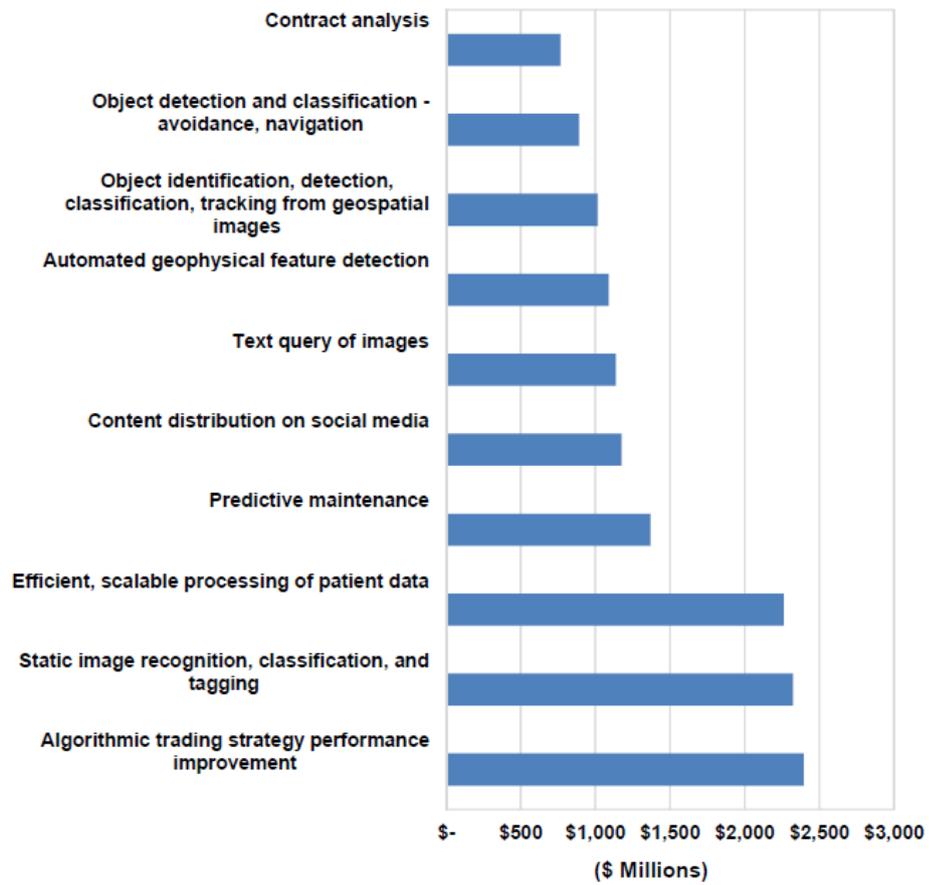
Nel quadro generale delle cose, la velocizzazione dei processi d'analisi insieme all'incremento prospettico qualitativo della stessa offrono un'ottica tramite la quale osservare l'ambiente, inteso su più livelli, dell'Impresa (e non) completamente nuova, che se correttamente applicata e integrata all'interno dei processi aziendali (e trasversalmente ad essi) può consentire l'accelerazione della nuova era industriale, ovvero dell'Industria 4.0 (la quarta rivoluzione industriale [5]).

Come mostra infatti il grafico 1.3, grazie agli sviluppi degli applicativi del Cognitive Computing, dunque dell'Intelligenza Artificiale, si stima incrementeranno i ricavi dei prodotti e servizi basati su IA, dove al primo posto (per ricavi assoluti) ritroviamo i servizi utili all'incremento delle performance nelle strategie di Trading Algoritmico.

Già attualmente gli intermediari finanziari utilizzano l'IA per guidare le proprie strategie di portafoglio, ma in realtà lo switch sarà ancora più drastico, sino ad arrivare, come stimato dagli analisti, al 2025 dove circa metà delle attività finanziarie saranno scambiate da un'IA. [6]

Gli analisti di Tractica stimano inoltre che la fetta più grande di questi ricavi è appartenente al Deep Learning, che si prevede nel 2025 sarà pari a 16 \$ miliardi.

Chart 1.2 Artificial Intelligence Revenue, Top 10 Use Cases, World Markets: 2025



(Source: Tractica)

Grafico 1.3 [6]

1.4. Esperienza Umana, il fulcro dell'economia

L'8 Febbraio 1998 l'allora Ministro dell'Economia Carlo Azeglio Ciampi tramite una votazione telefonica dichiarò che il retro della futura moneta da 1 euro “dovesse recare l'Uomo Vitruviano di Leonardo da Vinci” [7].

Una scelta certamente singolare agli occhi di molti, ma decisamente non casuale: l'intento infatti fu quello di richiamare il Rinascimento Italiano, la cui rivoluzione evolutiva si basava sulla centralità dell'uomo nella misurazione della natura e dei fenomeni circostanti.

Il simbolismo di Ciampi fu esternalizzato anche da egli stesso dove dichiarava che << *questa rappresenta la "moneta al servizio dell'Uomo" invece dell'Uomo al servizio del denaro* >>.

La moneta è dunque principalmente uno strumento di pagamento *per l'uomo*, così come l'economia è, nella sua accezione generale, uno strumento per *ottimizzare* la suddivisione delle risorse *limitate* con il fine di massimizzare il benessere collettivo ed individuale tramite il soddisfacimento dei bisogni *tendenzialmente illimitati*.

La moneta, come strumento di pagamento, unità di misura e traslatore del valore nel tempo e nello spazio, è innanzitutto soggetta operativamente al meccanismo di determinazione dei prezzi, quest'ultimi singolarmente e nella loro eterogeneità possono essere dei veri e propri *vettori di informazioni* che induttivamente possono trasmettere lo stato della domanda o dell'offerta di un determinato bene o servizio.

Dunque, assumendo il principio della *razionalità limitata* ideato da Herbert A. Simon [8], e dell'*euristica* nella sua accezione cognitivo-comportamentale, è possibile determinare come l'esperienza umana definita da Ciampi sia interpretabile in maniera relativa e singolare rispetto ai singoli individui (ossia fondata su informazioni e facoltà cognitive limitate, entrambe emanate secondo una distribuzione casuale, in maniera dinamica e non indipendente), e soprattutto quanto la qualità (ossia il livello di valore percepito individualmente) di quest'ultima sia in realtà una colonna portante

dell'economia stessa in quanto scienza sociale, da intendersi dunque come variabile fondante da analizzare e massimizzare al fine del benessere economico complessivo.

Il Valore Economico, o anche il valore dell'esperienza che ve ne deriva, assumendolo come obiettivo da ottimizzare per l'economia, non è però un fattore od un fenomeno unicamente dipendente dalle risorse limitate.

I prezzi infatti talvolta non riescono a comunicare effettivamente i costi (sociali o ambientali ad esempio) o il valore effettivo di un determinato bene o servizio standardizzato, in quanto la dimensione del *valore esperienziale* è una percezione soggettiva che può prescindere dal *valore economico* di un bene e tendenzialmente risulta difficile da replicare in termini assoluti.

Questo fattore dunque potrebbe paradossalmente finire per influenzare tramite esternalità negative (o positive) il *valore esperienziale* collettivo (inteso come valore percepito derivante dalle risorse circostanti, a prescindere dalla tipologia di bene economico), senza che la moneta (originariamente concepita come "al servizio dell'uomo") sia in grado di misurarne o comunicarne l'andamento al "mercato" o alla collettività di attori economici.

Viceversa, anche il *valore economico* di un bene o servizio può mutare al variare del *valore esperienziale*, come spesso accade in caso di fluttuazioni improvvise della domanda globale o nelle bolle finanziarie.

Il cervello umano infatti è determinante, nei termini strutturali e cognitivi, nella definizione dell'esperienza e del valore esperienziale in funzione di un determinato bene unitario omogeneo, e la sua evoluzione biologica nel tempo, seppur lenta, ha incrementato non solo il valore percepito delle risorse limitate circostanti, ma anche le modalità potenziali percepite secondo le quali è possibile combinarle, ovviamente con lo scopo di migliorare l'esperienza umana stessa tramite la minimizzazione dello sforzo o la massimizzazione di un determinato risultato.

Questo orientamento binomiale non è casuale, infatti rappresenta una condizione intrinseca del cervello che propende per l'attuazione di comportamenti utili ad ottimizzare i risultati in funzione dell'evoluzione dinamica dell'ambiente circostante.

Un'analogia concettuale, derivabile dal concetto di prezzo come "vettore di informazioni", è iscrivibile nella tendenza degli esseri umani, in quanto animali sociali, verso la comunicazione.

Comunichiamo informazioni, costantemente in maniera cosciente o meno, verso gli individui che interagiscono con noi o con le nostre tracce (dati, libri, media, ecc.) al fine di trasmettere lo stato dell'ambiente e della realtà circostanti nel tempo.

Questo è forse il fenomeno più importante, che ha determinato l'evoluzione umana e che soprattutto ha garantito la massimizzazione del valore economico-esperienziale nel tempo.

Un incremento nelle modalità, velocità e profondità dimensionali delle informazioni trasmesse tra gli individui equivale direttamente ed indirettamente ad un incremento esponenziale del valore economico-esperienziale generato, come confermato dai trend esplicitati nei paragrafi precedenti.

Il valore esperienziale è infatti un fattore che può dipendere da più variabili, le cui effettive colonne portanti sono l'apprendimento sociale e l'influenza sociale.

In quanto animali sociali abbiamo, infatti, adottato con il tempo la facoltà di comunicare e recepire informazioni da altri, inizialmente tramite l'utilizzo della parola, poi circa 5200 anni fa tramite la scrittura, che ha esteso il potere dell'influenza e dell'apprendimento oltre i vincoli spaziali e temporali.

Verso il 1440 è sopraggiunta la stampa, a cui si è aggiunta la radio verso i primi del XX secolo ed infine la Televisione tra gli inizi e la metà dello stesso secolo.

Il problema della tipologia di comunicazione derivante da questi strumenti è il meccanismo unilaterale dell'influenza sociale: le idee e le informazioni trasmesse erano

strettamente limitate e frutto concettuale della mente di pochi uomini con un elevato potere.

La vera rivoluzione del valore esperienziale è sopraggiunta infatti con l'avvento di Internet, intorno agli inizi degli anni 90'.

Il World Wide Web ha consentito a tutti gli attori economici di partecipare al processo di influenza ed apprendimento sociale di massa, abbattendo vincoli temporali e spaziali, rendendoli al contempo sia protagonisti che spettatori.

All'evoluzione tecnologica, non è però susseguita l'evoluzione biologica, perlomeno in termini di velocità.

Secondo le neuroscienze sono permasti, infatti, la facoltà di apprendimento sociale ed i processi biologici alla base del processo decisionale umano, apparentemente cristallizzati agli anni dell'invenzione della scrittura. [9]

Ecco che quindi gli uomini tendono a lasciare che il valore esperienziale, per come viene concepito e percepito, venga influenzato dalle opinioni e dalle informazioni che derivano dagli individui circostanti, spesso in ordine di similarità (affinità secondo i gusti personali) o di autorità percepita (in materia ad esempio dello specifico bene). E' stato dimostrato infatti che dinanzi ad un consenso diffuso a livello massivo e multilaterale, solo all'incirca il 30% delle persone ha il "coraggio" di selezionare un'opzione alternativa. [10]

Quest'ultima rivoluzione derivante dalla nuova multilateralità dei dati riguardo l'esperienza umana, che in un certo senso chiude il cerchio aperto con il primo paragrafo, è l'inizio di una nuova fase conoscitiva sia per l'economia in quanto scienza sociale, sia per l'umanità in generale, per il benessere derivante non solo dalla spinta propulsiva verso l'innovazione garantita dalle nuove tecnologie informatiche, ma anche dall'affinamento del valore esperienziale individuale tramite una maggiore personalizzazione ed indicizzazione dinamica dei principali concetti di ottimizzazione economica.

2. Capitolo Secondo: Le Tecnologie ed i Mercati Finanziari

2.1. Le Idee e le Tecnologie Digitali

Sinora abbiamo discusso riguardo il valore esperienziale, tralasciando l'analisi approfondita della sua funzione contestuale e delle variabili causali dalle quali è determinato.

Il valore esperienziale è innanzitutto, come anticipatamente detto, determinato dalle facoltà cognitive umane (limitate e caotiche) e dunque è fondamentale soggetto al processo decisionale [1].

Ciò significa che a livello individuale le decisioni umane non detengono uno stato di ottimo omogeneo nel tempo ed universalmente valido: si tratta di condizioni funzionali e dinamiche che non necessariamente sono fondate sulla razionalità assoluta e sulla perfetta informazione.

Quello che influenza maggiormente ogni singola scelta personale può derivare da dinamiche sociali e contestuali che mutano a seconda delle condizioni delle potenzialità informative e dello stato emotivo individuale.

Adam Smith, infatti, nel classico "Teoria dei sentimenti morali" sosteneva proprio che l'economia, o meglio il motore costituito dalla "mano invisibile" che ne è alla base, fosse guidata non unicamente dalla concorrenza tra gli operatori economici, ma dal tessuto sociale e dalle dinamiche contestuali che ne derivano.

Dunque, non unicamente i beni sono scambiati tra gli attori economici, ma anche idee ed informazioni.

Il processo di trasmissione avviene tramite il meccanismo dell'apprendimento sociale, che a seconda dell'intensità e della frequenza delle relazioni tra gli individui conduce verso un determinato livello di flusso di idee.

Infatti, le nuove informazioni, i nuovi concetti, dunque le nuove idee funzionali, garantiscono non solo un incremento dell'efficienza strategica ed operativa in un

contesto organizzativo, così come studiato dal Knowledge Management, bensì anche una variazione esponenziale dei comportamenti e del grado di innovazione.

Se ne deriva quindi che il valore economico di un determinato bene o servizio è strettamente dipendente dal valore esperienziale che se ne è in grado di trarre (a prescindere dallo scopo utilitaristico o accessorio), e che dunque il valore esperienziale, includendo bias cognitivi e assunzioni irrazionali, è in funzione delle idee che riguardano quei beni stessi, come ad esempio le percezioni emotive o le modalità di utilizzo concepibili.

O più precisamente, il tasso di variazione del valore esperienziale nel tempo è determinato dalla velocità e dalla quantità del flusso delle idee tra gli individui, assumendo come costanti le facoltà cognitive di quest'ultimi nel breve e medio termine.

Un'idea non è unicamente un concetto, ma a livello operativo spesso è proprio iscrivibile come una strategia, da attuare dinanzi a più opzioni possibili, che determina il comportamento finale di un individuo all'interno di un sistema economico.

Come descrive, infatti, Kahneman in "Thinking, Fast and Slow", le idee quando sono compatibili e coerenti con il nostro modo di essere e soprattutto sono influenti, finiscono per risultare in "azioni abitudinarie", impiegate quando necessario dal "pensiero veloce".

La forza che funge da attrito alla loro propagazione è per l'appunto la sua stessa origine: la mente.

I processi biologici che determinano il processo decisionale sono infatti ancorati a logiche e procedure ancestrali, tendenzialmente, infatti, capita agli agenti economici di discostarsi ampiamente dagli agenti razionali concepiti dalla teoria neoclassica.

Può capitare che in merito ad un determinato argomento non si inneschi lo stesso meccanismo di apprendimento sociale che avverrebbe osservando gli individui più competenti, come auspicabile in condizione di ottimo, bensì che vengano ponderate maggiormente le idee provenienti da individui secondo un principio di similarità o di

autorità apparente, altresì definito da Bahador Bahrami come equality bias, non garantendo la diversificazione delle idee e portando all'autoreferenzialità di gruppo e conducendo infine all'errore sistematico.

Questa è una condizione che affligge infatti le moderne democrazie come noi tutti le conosciamo, che tendono dunque verso l'euristica dell'eguaglianza, fattore che minimizzando il costo cognitivo dell'introduzione delle nuove idee, può portare alla generazione di bias cognitivi e dunque errori sistematici diffusi nel processo decisionale collettivo.

Il focolaio è dilagato con le tecnologie di comunicazione di massa, che hanno radicalmente cambiato il processo di diffusione delle idee, poste alle basi delle singole opinioni e dei comportamenti individuali.

La logica fondante del sistema democratico e del libero mercato è infatti proprio il processo decisionale individuale, che su larga scala per scelte condivise tende a distribuirsi in maniera normale attorno ad un determinato valore o una determinata opzione, tendendo dunque, quando popolazioni sufficientemente grandi lo consentono, a minimizzare l'errore.

Viceversa, i sistemi fortemente accentrati e con un modello ad "economia pianificata" si sono dimostrati scarsamente funzionali nel tempo proprio a causa dell'assenza di indipendenza e di una sufficientemente elevata diversificazione all'interno del direttivo, che necessariamente hanno teso dunque verso forti errori decisionali sistematici.

Nel testo "Il Manifesto della Geniocrazia", è citato il concetto di "Social Media Socialism", la cui analisi si fonda sull'evoluzione dei Social Network e del Cognitive Computing, il cui ibrido può, già attualmente, profilare gli interessi, le idee, le abitudini e dunque i comportamenti di ogni singolo individuo, arrivando non solo in ottica predittiva ad anticipare l'evoluzione dei comportamenti dei singoli attori economici, ma anche in quella "correttiva" pianificare tramite incentivi di rete, come descritto dalla "Nudge Theory" (concetto analizzato anche dal premio Nobel per l'economia Cass

Sunstein [2]) nell'ambito del paternalismo libertario, i comportamenti e le idee future degli operatori appartenenti ad un determinato sistema economico.

Questo ne risulta in un ipotetico scenario futuro dove l'errore sistematico presente nel processo decisionale può essere ridotto ulteriormente a beneficio della collettività, ponendo però come fondamentale un adeguata supervisione e regolamentazione riguardo i processi di "back-end" degli ambienti digitali, spesso occultati da multinazionali private, che eludono dunque le nazioni concettualmente sovrane.

Il meccanismo alla base dell'intelligenza collettiva, che è definita dal flusso di idee, è infatti il fenomeno della "saggezza della folla", concetto approfondito dagli studi di James Surowiecki. [3]

Egli infatti analizza come nell'effettuare una stima, un grande gruppo di individui la cui opinione individuale venga sommata con lo scopo di effettuare una media, tende verso la compensazione degli errori di valutazione, ottenendo un valore ottimale e più vicino al valore effettivo rispetto alla stima ottenibile dal singolo individuo.

Questo fenomeno però richiede una condizione imprescindibile: l'indipendenza di ogni singolo individuo nell'effettuare la propria valutazione, infatti se all'interno del sistema d'analisi è presente la possibilità di comunicare, allora le opinioni tenderanno verso una minore indipendenza e di conseguenza verso un maggior errore di valutazione sistematico.

Online diviene sempre più frequente l'interazione sociale, fattore che determina la crescente interdipendenza degli attori economici, con l'incremento dello scambio sia di singole idee che di opinioni più strutturate.

Questo fenomeno se non adeguatamente supervisionato può incrementare la diffusione di idee in maniera omogenea tra la popolazione, fattore che può condurre alla riduzione della diversificazione cognitiva e dell'indipendenza necessari alla "saggezza della folla".

A dimostrazione dell'influenza sociale online e dell'assenza di indipendenza totale nel processo decisionale nel 2014 un gruppo di ricercatori ha dimostrato che su Twitter tendenzialmente quando una persona pubblica un Tweet con un sentiment positivo, precedentemente alla pubblicazione nel feed erano contenuti un quantitativo di tweet positivi almeno il 4% superiore rispetto quelli con un sentiment negativo. [4]

In effetti è stato dimostrato che l'utilizzo di Twitter incrementa dal 65% al 75% lo stato di eccitazione emotiva, in quanto la struttura del Social Network è stata pensata per aggirare i filtri emotivi descritti da Kahneman in "Thinking, Fast and Slow", un fattore che può rendere molto influenzabili gli individui che interagiscono sul Social.

Viceversa, le Tecnologie Digitali e nello specifico i Social Network dovrebbero supportare la riduzione di eventuali errori sistematici a cui la "saggezza della folla" potrebbe essere esposta anche in condizioni di indipendenza, a causa magari di bias sistematici propri del cervello umano, che non consentono anche in caso di distribuzione normale di ottenere un valore dove è minimizzato l'errore.

Di questo fenomeno ne beneficerebbero non solo i sistemi politici, evitando dunque le recenti crisi sistemiche della democrazia moderna, ma anche le organizzazioni e gli operatori economici del libero mercato, tramite la riduzione della probabilità di svilupparsi di bolle finanziarie, limitando dunque gli attacchi speculativi, o l'incertezza derivante da un'improvvisa riduzione della domanda globale.

Per fare ciò, è necessario che il fenomeno del "Social Media Socialism" sia interessato dall'individuazione degli individui con minore indipendenza nel processo di formazione delle opinioni e al contempo dalla profilazione dello status quo del livello di conoscenza individuale.

In questo modo è possibile ottimizzare il feed d'esposizione dei singoli individui in modo tale da incrementare il livello di diversificazione di conoscenze utili ai fini del processo decisionale collettivo e all'aumento dell'indipendenza rispetto gli individui eccessivamente influenti sull'opinione pubblica a beneficio di chi è particolarmente influenzabile.

2.2. Finanza Comportamentale e Social Network: il caso OpenBook

Una volta esplicitato come si forma il valore percepito e come si propaga il flusso delle idee, decisivo ai fini dei processi decisionali del singolo individuo, diviene di fondamentale importanza comprendere come questo finisca per incidere nella creazione di valore aziendale e nel conseguimento dei rendimenti per i singoli investitori.

Un flusso delle idee con un'eccessiva resistenza, causato magari dall'eccessiva distanza tra gli individui di un'organizzazione aziendale o dall'assenza di incentivi strutturali che portino al coinvolgimento e all'interazione, può condurre ad un'accentuata riduzione del valore aziendale creato, oltre che una minore sostenibilità nel tempo derivante da una minore flessibilità aziendale nell'adattarsi all'ambiente esterno.

Questo non vale unicamente per le aziende, ma anche per i singoli individui che si muovono in contesti come i mercati finanziari.

I singoli investitori, così come nelle aziende circoscritte, decidono dove impiegare il proprio surplus di capitale in cambio di un rendimento reale traslato ad un determinato tempo "t" futuro.

Spesso però i rendimenti attesi sono frutto di stime che possono fondarsi su idee irrazionali o bias comportamentali, conducendo a scelte ben distanti dal rendimento ottimo di mercato rispetto al rischio, risultando dunque in una condizione dove il singolo investitore si addossa rischi superiori in cambio di rendimenti inferiori.

Questo comportamento è stato analizzato infatti tramite un esperimento di Alex Pentland, Yaniv Altshuler e Wei Pan, ripreso anche dalla Harvard Business Review [5] e approfondito nel libro "Fisica Sociale" [6].

Nel 2011 sono stati raccolti dalla piattaforma eToro (dedicata alle transazioni finanziarie su base giornaliera come il day trading) operazioni riguardo 1,6 milioni di utenti, per un totale di circa 10 milioni di scambi finanziari.

L'aspetto interessante di questa ricerca ha riguardato una feature particolare di questa piattaforma: il Social Network presente dentro eToro: OpenBook

Tramite questa funzionalità per gli utenti è possibile replicare esattamente le transazioni ed i portafogli di altri utenti, quest'ultimi classificati in base al loro rendimento.

L'incentivo di rete principale nel rendere disponibili le proprie idee di trading al pubblico consiste nel ricevere una piccola remunerazione da parte della piattaforma ogni qual volta qualcuno inizia a replicare la propria strategia.

Questo *nudge* ha consentito il rapido dilagarsi del flusso delle idee e dunque dell'apprendimento sociale.

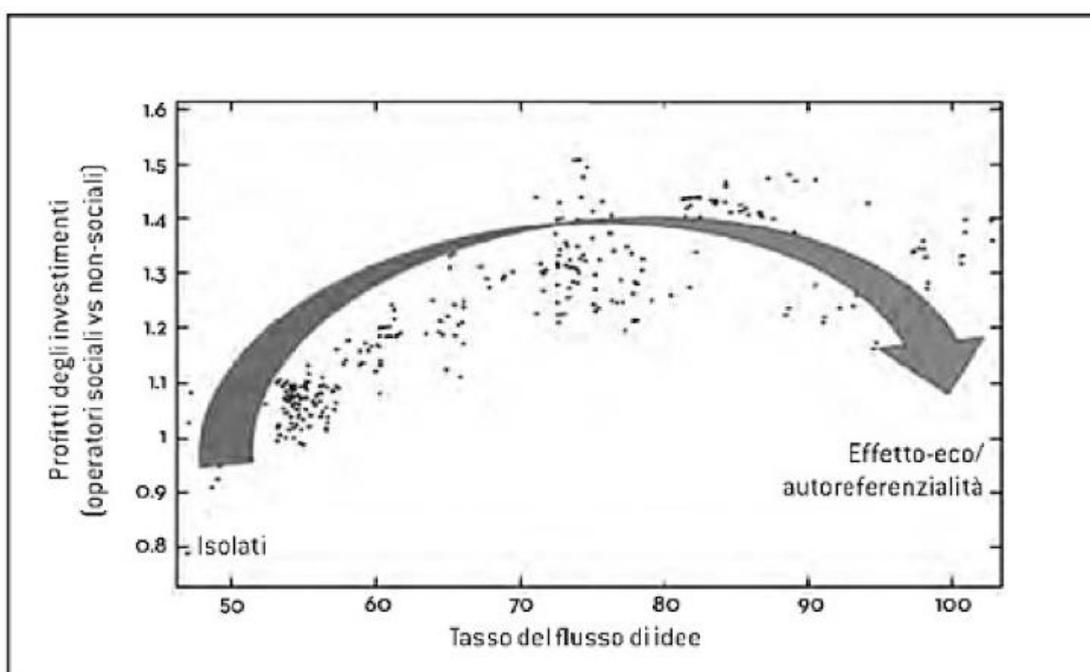


Grafico 2.1 [6]

Il Grafico 2.1 mostra dunque le “Prestazioni medie dei trader eToro in una giornata” [6] e rivela un insight fondamentale: l'apprendimento sociale, se ben ponderato tra la diversificazione collettiva e la razionalità individuale indipendente consente di ottenere rendimenti dagli investimenti superiori di circa il 30% rispetto a quelli ottenuti dagli investitori isolati.

Dunque, come Pentland giustamente fa notare, il fenomeno della “saggezza della folla” va a ritrovarsi nel mezzo tra coloro che fanno eccessivo affidamento sugli altri, ossia

l'effetto gregge che conduce all'autoreferenzialità, e coloro che invece sono completamente isolati dagli altri e che dunque non apprendono dai potenziali flussi di idee.

Le Tecnologie Digitali hanno introdotto per l'appunto la possibilità di decentralizzare la fonte dei flussi di idee, rendendo i nodi non più unici recettori ma anche democratici comunicatori.

Questo fenomeno va però governato, in quanto se da un lato incrementa la qualità del processo decisionale collettivo e dunque individuale tramite l'apprendimento sociale dagli usi di una community online, dall'altro un'eccessiva fiducia dei singoli nodi verso altri nodi più influenti può condurre ad un effetto gregge che tende a vanificare l'efficienza e la minimizzazione dell'errore inizialmente apportata dalla saggezza della folla.

Infatti, quella che inizialmente era stata concepita come una democrazia decentralizzata, sta diventando un'oligarchia dell'attenzione, dove diviene sempre più centrale l'individuo che è in grado di bypassare i filtri cognitivi individuali. E' per questo motivo che diviene fondamentale che la regolamentazioni e la supervisioni del legislatore primario e secondario pongano attenzione all'accenramento dell'influenza cognitiva sui singoli nodi di uno specifico sistema economico, ideando nuovi sistemi di disincentivi di rete alla formazione di "monopoli cognitivi" che possano condurre all'errore sistematico e al bias cognitivo collettivo, sfociando dunque nell'autoreferenzialità e riducendo, anche se in maniera non apparente, le opportunità del libero mercato.

Affinché la domanda globale raggiunga un punto di ottimo, è necessario che non solo le imprese produttrici siano sottoposte a concorrenza per la creazione di valore economico, ma che anche i nodi che tendono verso l'influenza della domanda stessa lo siano, in modo tale che il valore esperienziale sia ottimizzato, affinché quest'ultimo garantisca l'efficacia del processo decisionale individuale a beneficio del benessere collettivo.

I singoli nodi, da identificare con i singoli attori di un sistema economico, non sono infatti in grado di riconoscere lo stato di autoreferenzialità di un determinato flusso di idee.

Tendenzialmente infatti, ogni singolo nodo tende a replicare nella rete una stessa idea tramite definizioni e argomentazioni all'apparenza differenti.

Di conseguenza allo stato cosciente, un'idea può non essere riconosciuta alla sua eventuale replica, conducendo dunque il singolo nodo a intuire che gli altri nodi della rete siano giunti alle proprie stesse idee tramite un processo razionale individuale.

L'incremento di fiducia derivante dalla riprova sociale, dunque di utilizzo effettivo dell'idea nel processo decisionale, è quello che può portare la collettività verso scelte finanziarie ed economiche che conducono al fenomeno della bolla speculativa, ossia al panico borsistico ed al collasso economico e finanziario.

Sempre nello studio di Pentland è stato evidenziato come nel social network di eToro un investitore lettone essendosi classificato come uno tra quelli di maggior successo in termini di rendimento ha di conseguenza iniziato ad ottenere un ampio seguito.

I singoli utenti però non sono stati in grado di riconoscere la struttura organizzativa della rete che si era creata e di conseguenza hanno iniziato a ricadere nel fenomeno dell'autoreferenzialità.

Infatti, ogni singolo nodo credeva di replicare la strategia frutto di più nodi differenti (dunque indipendenti solo all'apparenza) senza però sapere che quegli stessi nodi stavano replicando la strategia originata dal lettone.

Il risultato finale di questa situazione, dove i singoli nodi stavano attuando una strategia di apprendimento sociale diversificato all'apparenza, ha condotto ad un unico replicarsi di un processo decisionale di un singolo nodo originario, dunque soggetto a bias cognitivi ed errori sistematici non diversificati, che ha determinato lo scoppio di una bolla finanziaria sulla piattaforma.

Una simile situazione è spiegata in maniera più approfondita però dallo studio di Niklas Karlsson, George Loewenstein e Canegie Mellon [7]: infatti, l'apprendimento sociale è determinato dalla tipologia di idee e di informazioni che recepiamo.

La mente umana tende a sottostimare l'effetto positivo derivante dalle informazioni classificate con un sentiment negativo e invece a sovrastimare il valore delle informazioni con un sentiment positivo.

La tendenza dunque degli operatori economici verso il "sentirsi emotivamente bene" influenza notevolmente le scelte finanziarie, come ad esempio nel caso degli investimenti, ed è la causa alla base dell'autoreferenzialità e della ricerca della riprova sociale.

L'apprendimento sociale che inizialmente ha determinato ad esempio la propensione verso un determinato investimento possiede un attrito cognitivo inferiore rispetto al momento in cui si ricevono informazioni negative sull'andamento dell'investimento stesso.

Lo studio dei 3 esperti [7], ha monitorato la propensione degli investitori verso il controllo dell'andamento del proprio portafoglio titoli in rapporto all'andamento del mercato, quando era nulla la volontà antecedente di effettuare una transazione.

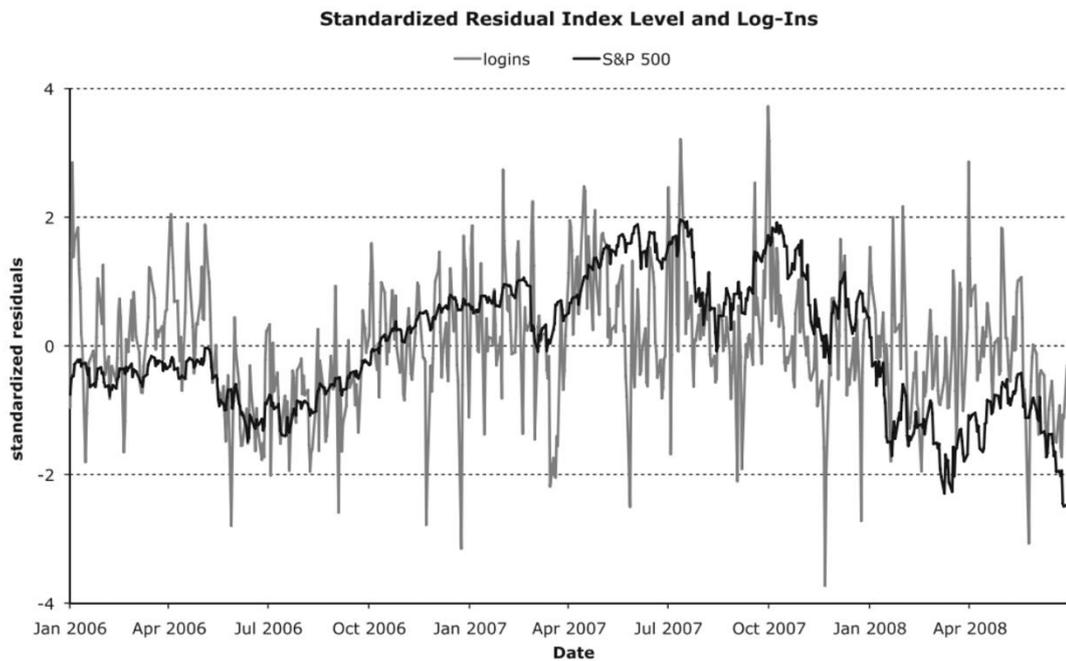


Grafico 2.2 [7]

La linea nera mostra l'andamento del valore dell'indice S&P 500, un indice che rappresenta l'andamento complessivo di 500 titoli considerati tra i più rilevanti scambiati presso il NASDAQ e/o il NYSE.

Quella grigia invece rappresenta il numero di accessi ai conti che avessero come fine unicamente quello di monitorare l'andamento del portafoglio.

Lo studio, anche se riguarda una categoria di investitore differente rispetto al day trader, dimostra un bias comportamentale finanziario comune agli investitori. Infatti, come rilevato dal grafico, ogni qual volta che il mercato sembra avere un trend crescente incrementa il numero di log in, viceversa quando scende il numero di accessi tende a diminuire.

L'insight derivante dallo studio è che la probabilità che un investitore controlli l'andamento effettivo del proprio portafoglio incrementa quando le performance

generali del mercato sono positive, viceversa scende quando le performance del mercato sono negative.

Lo studio ha determinato che dunque la tendenza del cervello umano a rinnegare informazioni negative e ad ignorare i trend avversi che potrebbero provocare negli agenti economici una condizione emotiva-esperienziale negativa avrebbe un impatto fortemente negativo sui risultati finanziari.

Solo verso la fine del 2008, con l'arrivo della crisi finanziaria ormai generalmente evidente, il numero degli accessi si è fatto più frequente, in un momento in cui ormai era troppo tardi per ripararsi dalle perdite.

Forse un atteggiamento più razionale e distaccato da parte degli investitori avrebbe garantito una graduale riduzione delle esposizioni di quelle posizioni sopravvalutate, riducendo notevolmente la possibilità del tracollo improvviso del mercato finanziario, ma finché sarà la mente biologica ad interagire con un mondo nuovo e complesso il numero di errori sistematici e bias cognitivi sarà destinato ad incrementare all'aumentare degli operatori economici umani che da tutto il mondo potranno interagire con i mercati finanziari.

La diversificazione del rischio finanziario richiede dunque anche una diversificazione informativa dei singoli individui.

Affinché l'intelligenza collettiva riduca gli errori sistematici infatti è di fondamentale importanza che l'individuo apporti informazioni differenti alla collettività degli operatori economici, viceversa infatti le fonti di informazione tradizionali sono spesso causa di scelte finanziarie erranee in quanto origine unilaterale di un flusso di idee strettamente limitato e scarsamente diversificato.

Per questa ragione, oltre che a causa dell'inclinazione degli agenti economici verso le informazioni con un sentiment positivo a scapito di quello negativo nel filtro dell'apprendimento sociale applicato al flusso delle idee, è di fondamentale importanza che le tecnologie digitali non solo supportino il processo decisionale dell'individuo ideando un'esposizione strutturale dinamica alle informazioni e alle idee, bensì tramite

gli sviluppi del Cognitive Computing ne ottimizzino le facoltà, in base agli obiettivi ed al profilo individuale.

2.3. Fintech e Hybrid Robo Advising

Le Tecnologie Digitali stanno però introducendo un'opportunità disruptive per il processo decisionale nel settore degli investimenti finanziari: il Robo Advising.

Gli sviluppi di Software ed algoritmi sempre più raffinati insieme al recente impeto del Cognitive Computing, alla cui base vi è il Machine Learning e il Deep Learning, hanno consentito la decentralizzazione e la disintermediazione della consulenza finanziaria nella strutturazione di un portafoglio di investimenti.

All'interno di un intermediario finanziario il processo di profilazione e di consulenza generalmente infatti dipende dagli operatori economici, ossia dal processo decisionale umano.

Questo fattore è molto importante al fine della determinazione dei rendimenti effettivi assestati secondo il rischio del portafoglio finale dell'investitore: infatti i singoli agenti che profilano l'investitore e lo supportano durante il processo di composizione del portafoglio oltre ad essere essi stessi soggetti a bias comportamentali finanziari e a valutazioni erranee, sono tendenzialmente influenzati dall'interesse principale dell'intermediario stesso, che tende a sottoporre con una priorità di esposizione le tipologie di titoli da cui trae maggior convenienza nella transazione.

Questo fenomeno nel processo di formazione collettiva degli investimenti nel mercato finanziario tende a produrre risultati inefficienti secondo la teoria neoclassica finanziaria, non rientrando dunque nella frontiera dei portafogli ideata da Markowitz [8].

Infatti, tralasciando la traslazione del processo decisionale derivante dalla questione analizzata dalla Agency Theory, la condizione di ottimo non si verifica in quanto richiederebbe i seguenti prerequisiti:

- Che gli operatori economici siano perfettamente razionali
- Che il prezzo sui mercati finanziari si determini unicamente in base al valore intrinseco degli asset sottostanti

Queste due assunzioni hanno contribuito all'ideazione dell'Ipotesi dei mercati efficienti e del Capital Asset Pricing Model.

Una condizione strutturale necessaria a seguito della modellizzazione matematica dei processi finanziari derivante da queste 2 ipotesi dovrebbe dunque essere la seguente:

- 1) Il valore di mercato di un titolo tende verso il suo valore fondamentale
- 2) I mercati finanziari reagiscono in maniera rapida alle nuove informazioni
- 3) I prezzi si determinano in maniera casuale in base alla casualità di arrivo delle nuove informazioni
- 4) Nessun investitore può ricevere una remunerazione sempre superiore rispetto a quanto è richiesto in proporzione al rischio sistemico

Il differenziale tra realtà effettiva e la realtà descritta da queste 4 condizioni è evidente dalla presenza in passato di bolle speculative, dove gli operatori economici non sono stati in grado di replicare il valore effettivo del titolo sottostante, spesso per motivi derivanti dai bias comportamentali descritti nei paragrafi precedenti.

Infatti, non sempre è presente l'indipendenza informativa necessaria al manifestarsi della razionalità collettiva derivante dal fenomeno della saggezza della folla, ma anzi talvolta esistono condizioni in cui nodi particolarmente influenti della rete sociale tendono a traslare i bias comportamentali sistemici.

Una condizione simile non si verificherebbe invece in un mercato finanziario in cui gli interessi, ovvero i profili, siano rappresentati da agenti virtuali razionali. Questo è possibile grazie al Robo Advisoring e ai prossimi sviluppi del Cognitive Computing.

Il Robo Advisoring oltre ad avere fee più basse, efficientando i costi dal classico 1,35% ad un range dallo 0,2% al 1%, detiene anche minori barriere all'entrata, che portano l'investimento minimo da 50.000\$ per un financial planning a soli 500\$ [9]. Ma non è qui che si manifesta la vera rivoluzione: infatti il Cognitive Computing applicato ad un Robo Advisor può ridurre i bias comportamentali e il problema del rapporto principal-agent [10], garantendo contemporaneamente una razionale gestione

patrimoniale tramite l'ausilio delle informazioni pressoché istantanee dei Big Data e dell'ecosistema IOT, che altrimenti non sarebbero utilizzabili da un agente economico umano e dunque trasmissibili al mercato sotto forma del processo della determinazione dei prezzi.

Un simile scenario si avvicinerebbe notevolmente alla realtà matematizzata della finanza neoclassica, a discapito della finanza comportamentale.

Infatti, con una perfetta informazione pubblicamente e democraticamente disponibile presso agenti razionali il cui compito è rappresentare il profilo di rischio-rendimento dei singoli investitori, la condizione di un mercato efficiente diviene sempre più probabile.

Un margine di casualità e di errore sarà necessariamente presente in quanto vincolato alla condizione intrinseca dello status quo caotico proprio dell'universo, ma sarà per l'appunto minimizzato dall'assenza di assunzioni irrazionali casuali.

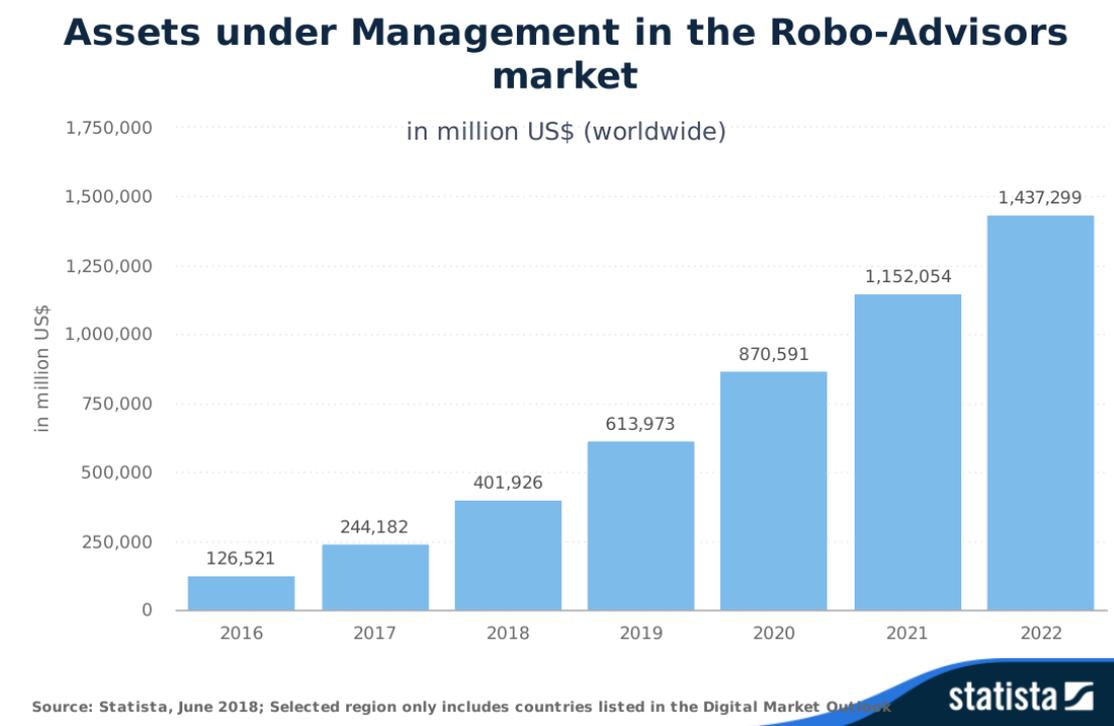


Grafico 2.3

Il Trend e la convenienza della gestione da parte dei Robo Advisor sono confermati dall'adozione da parte del mercato di questa tecnologia, analizzata dal report di Statista nel grafico 2.3.

Il passaggio da una gestione di asset pari a circa 126,5 \$ miliardi nel 2016 ad una pari a circa 1437,3\$ miliardi nel 2022 in tutto il mondo è il segno evidente di una condizione disruptive di questo mercato e dell'esigenza sottostante di razionalizzare sempre più i mercati finanziari in contemporanea al loro efficientamento, tendendo dunque verso la realizzazione della visione neoclassica troppo frequentemente definita come irrealistica.

Contemporaneamente a questo fattore, una volta delegato il processo di formazione di un portafoglio e della gestione finanziaria, il rendimento è vincolato ad un unico comportamento possibile da parte dell'operatore economico umano: fidarsi. Infatti, la mente biologica tende nel flusso delle idee a generare un maggiore attrito all'incrementare della scarsità di fiducia verso un altro operatore economico (umano e non).

Dunque, minimizzati i bias comportamentali derivanti dal processo decisionale della gestione del portafoglio, possono rimanere delle condizioni proprie della complessità, derivante dal differenziale cognitivo tra la società e/o l'intermediario che gestisce i Robo Advisor e l'investitore finale, che possono condurre verso una mancanza di fiducia da fornire al delegato.

Proprio per questa ragione, gli intermediari finanziari stanno esplorando le opportunità derivanti dalla tecnologia Blockchain (Grafico 2.4), che nella sua essenza di tecnologia trustless, consente di validare pubblicamente (e non) in maniera decentralizzata e trasparente le transazioni, portando il rischio derivante dal fidarsi del delegato virtuale ad essere prossimo allo zero.

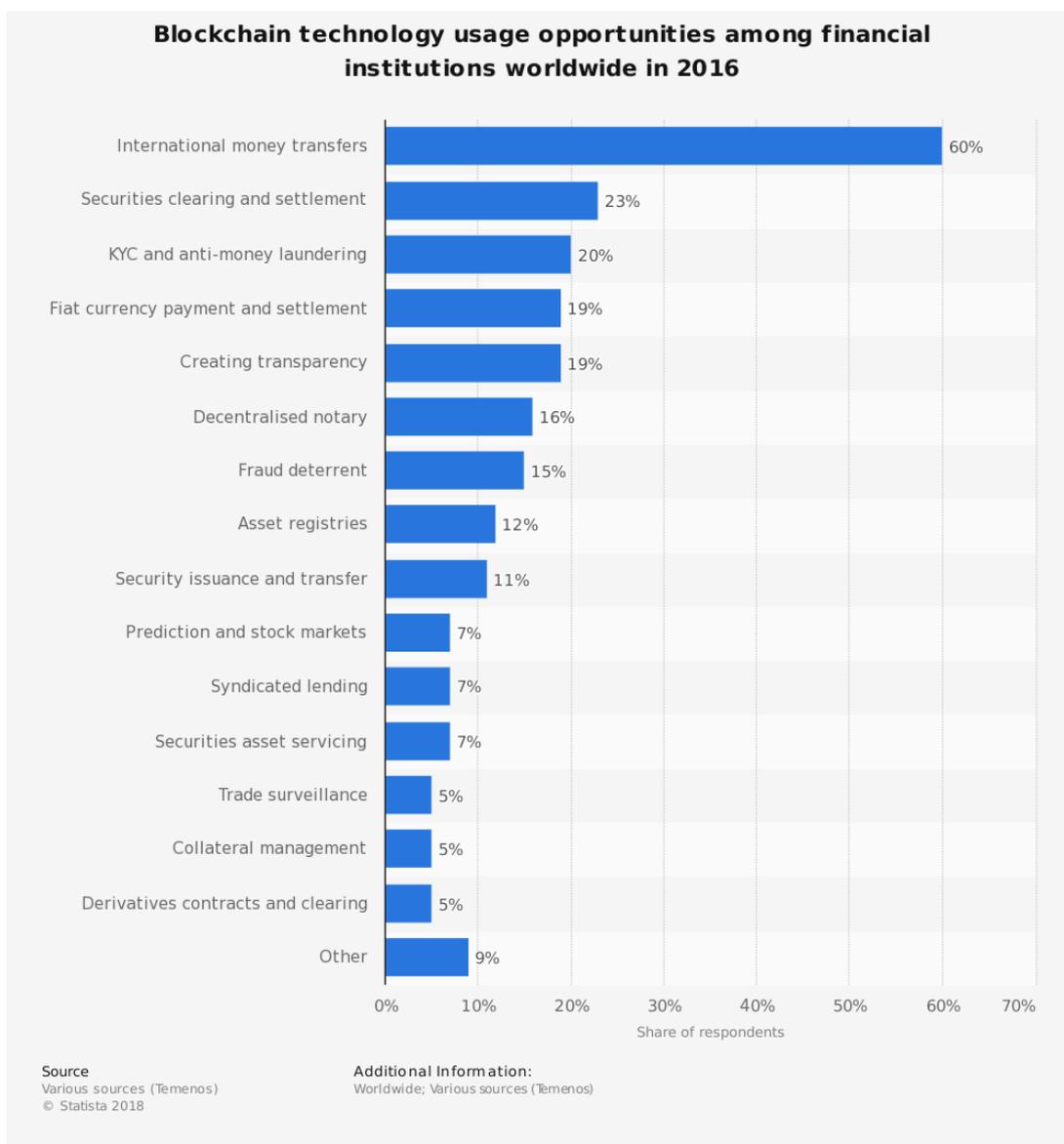


Grafico 2.4

Come dimostrato infatti dal grafico 2.4, il 19% degli intermediari finanziari intervistati al 2016 vuole utilizzare la Blockchain per essere più trasparente verso i propri clienti, mentre il 5% per incrementare la sorveglianza negli scambi (per ridurre le manipolazioni dei mercati, le frodi finanziarie, ecc.).

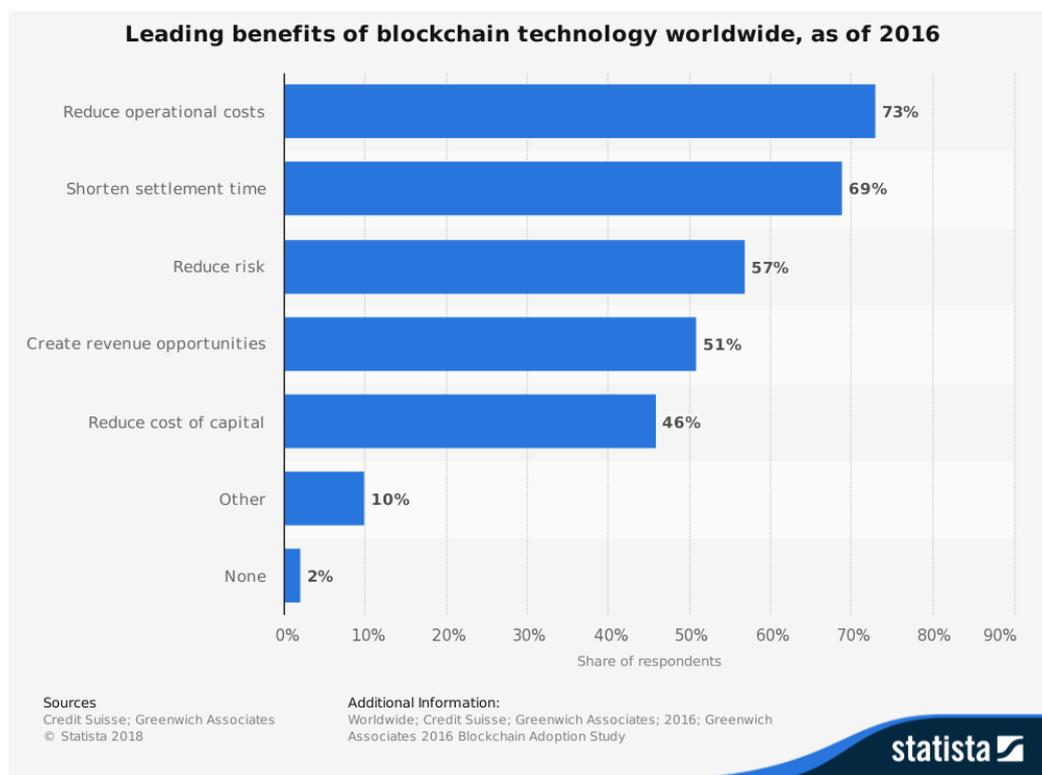


Grafico 2.5

Secondo il grafico 2.5, nel 2016 ben il 57% degli intervistati ha dichiarato di voler utilizzare infatti la Blockchain per ridurre i rischi presenti all'interno del contesto operativo aziendale, spesso derivanti dalla fiducia richiesta dal processo di delega intrinseco alle funzioni aziendali.

L'evoluzione delle tecnologie precedentemente descritte consente un'evoluzione esponenziale, sia in termini di redditività che di modello di business, di tutti i principali settori, che però diversamente da quanto percepito non è un processo immediato ed improvviso, ma un fenomeno in verità graduale che da tempo sta prendendo piede all'interno dei principali istituti finanziari e dei contesti aziendali a livello globale.

Più nello specifico il fenomeno delle tecnologie finanziarie, come il Robo Advisoring, il Robo Trading costituiscono sottogruppi di tecnologie che hanno incominciato a prendere piede già dai primi anni 90'.

Le aree principali del Fintech infatti sono principalmente 3:

- Investimento
- Transazionale
- Finanziamento

La digitalizzazione dei processi sottostanti le relazioni e le interazioni con i clienti da parte degli intermediari finanziari è stata fortemente trainata dalla contemporanea espansione della base utenti potenziale, come dimostrato dal Grafico 2.6.

Solo l'Home Banking ad esempio viene utilizzato mediamente da circa la metà degli Europei, che si contrappone però al 31% italiano (Grafico 2.7), che tralascia intuire un'avversione al cambio di paradigma tipica della struttura sociale del nostro paese.

Il problema sottostante questa condizione che si emana e si manifesta nella numerosità delle filiali presenti sul territorio decade necessariamente nell'impossibilità della riduzione dei costi, che nel bilancio degli intermediari, anche a causa della recente riduzione dei margini da interesse (e dell'eccessiva presenza di Non Performing Loans), finisce per tradursi in una globale riduzione della redditività, incrementando dunque il rischio sistemico.

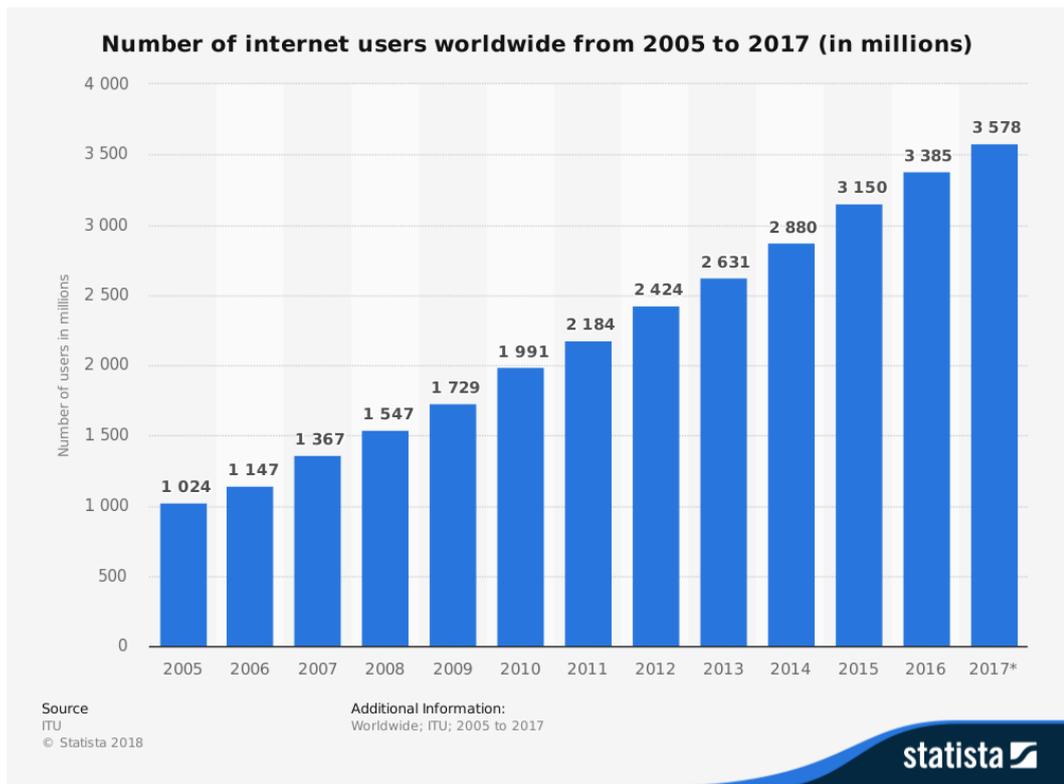


Grafico 2.6

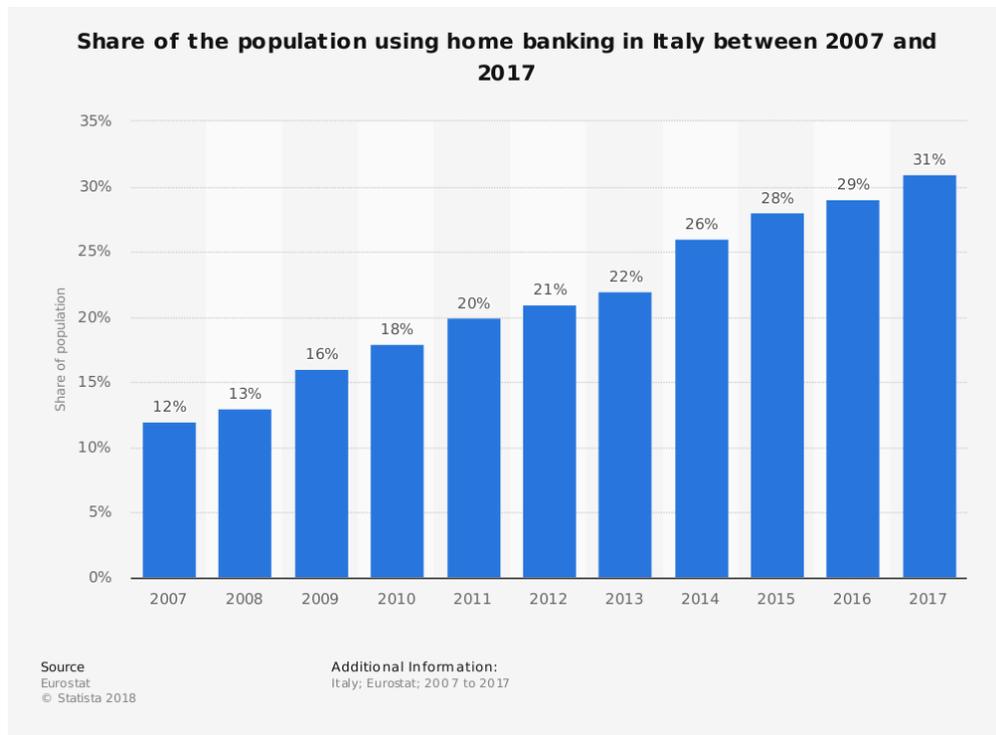


Grafico 2.7

È per questa ragione che il Fintech, così come il Robo Advisoring che ne fa parte, diventa non più un'evoluzione dettata unicamente dal tentativo di innovare i processi, ma una sentita condizione necessaria alla sostenibilità di lungo termine per la Governance degli Intermediari.

L'attrito italiano nella diffusione dei servizi Fintech precedentemente citati è però da identificarsi anche talvolta nello skill mismatch per quanto riguarda le tecnologie finanziarie, infatti, effettuando un benchmarking con le altre nazioni, è possibile rilevare che negli ultimi anni le competenze più richieste oltreoceano per ricoprire posizioni da "Portfolio Manager" o "Risk Manager" sono più informatiche che economiche: da C# a Python, sino ad arrivare a Matlab.

Abilità che come si può dedurre dal Grafico 2.8 risultano in termini assoluti probabilmente più rare da ricercare sul mercato (assumendo in pari quota la dilazione dei Background) per un Intermediario rispetto alla media dei Competitors Europei.

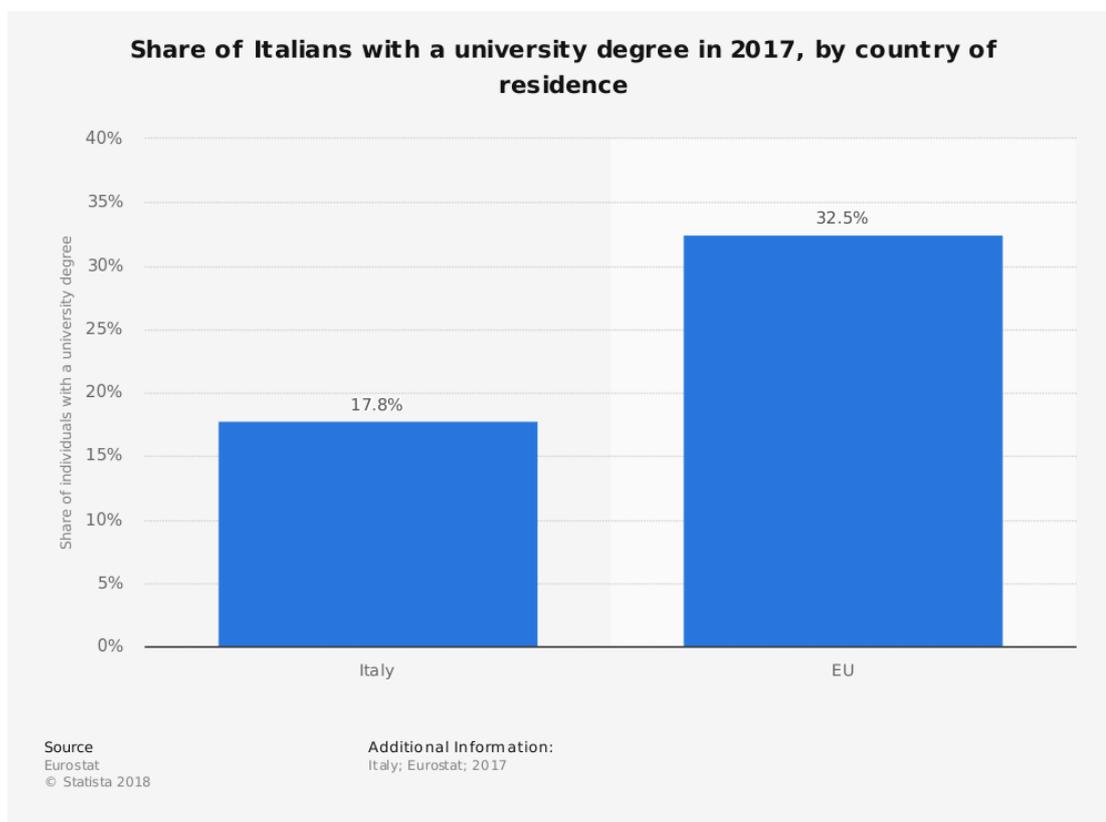


Grafico 2.8

In termini di costi, la transizione non risulterebbe infatti onerosa per quanto riguarda le tecnologie in sé, come ad esempio i Software e gli Hardware necessari alle Tecnologie Finanziarie, ma la gran quota di questi sarebbe in realtà indirizzata sulla formazione e sulla generazione di Know How interno tra le risorse umane.

Infatti, i costi sottostanti un Robo Advisor sono pressoché nulli, eventualmente un Robotrader che opera in intervalli temporali relativamente ristretti o addirittura un HTF potrebbe comportare dei costi elevati ma che per gli intermediari finanziari maggiori non sarebbero necessariamente onerosi se propriamente ammortizzati.

La variabile formazione è molto importante però che venga identificata come chiave, perché non solo è importante nell'ideazione, progettazione e gestione dei servizi digitalizzati che supportano le economie di scala efficientando dunque i costi per gli intermediari, ma sono in realtà di fondamentale importanza nella comunicazione verso il cliente finale.

L'adozione del Robo Advisor è risultata immediata per il mercato americano per via di una questione culturale, dove l'utente americano medio oltre ad essere propenso per l'innovazione ed i cambiamenti tende anche ad avere una curva di apprendimento molto rapida nell'interazione in autonomia con una nuova tecnologia.

Le condizioni intrinseche del nostro mercato però non consentono, anche ipotizzando sullo stesso livello il grado di innovazione effettivo dei servizi, un'adozione altrettanto rapida della tecnologia sul mercato.

Per questa ragione, nel libro "Fintech la finanza digitale" [11], Enrico Malverti descrive il concetto di Hybrid Robo-Advisor, dove i modelli decisionali robotizzati vengono affiancati dal gestore di portafoglio.

Questo fenomeno diviene sempre più rilevante al fine di "umanizzare", tradurre e semplificare al fine di comunicare con l'utente finale.

Come sempre descritto nel testo le effettive modalità operative possono essere 3:

- “*Compartimenti stagni*”: ossia la suddivisione di quali Asset possono essere gestiti dalla macchina e quali dall’uomo
- “*Adattamento manuale del Risk Management*”: in questo caso le Strategie effettive sono intraprese dalla macchina a priori, susseguite da operazioni di contenimento del rischio a posteriori tramite il recepimento umano di variabili non misurabili dalla macchina
- “*Selezione dei segnali coerenti con la propria view*”: In questo caso la macchina funge solo da supporto alle decisioni in quanto si prendono in considerazione unicamente le informazioni “interessanti” per le proprie decisioni.

Da questo fenomeno del “Centauro Digitale” derivano una serie di casistiche che sono analizzate nel contesto normativo della MiFID 2, ovvero dal tentativo di armonizzazione europea che consiste nella tutela del cliente tramite una maggiore trasparenza ed una profilazione dei rischi accettabili a seconda della categoria di cliente, che a sua volta dunque ha diritto a ricevere degli alert a seguito di una ricomposizione del suo portafoglio.

Una maggiore comprensione di questo strumento digitale da parte del cliente finale può essere infatti accompagnata tramite l’ibridazione uomo-macchina, in quanto per fare questo è necessario un approccio comunicativo-recettivo empatico la cui logica decisionale-comportamentale si fonda sul processo “Bottom Up” descritto nel “Pensiero Flessibile” [12], la cui determinazione si discosta dalla logica analitica Booleana del processo “Top Down” attuabile da una macchina.

Questa funzione dell’uomo si rende necessaria al Robo Advisoring dunque finché sarà il processo decisionale umano a gestire i flussi monetari a livello individuale nel libero mercato, in quanto sarà necessario accompagnare ed ottimizzare il processo comunicativo ed interpretativo verso gli operatori economici a seconda della mission strategica di lungo termine aziendale.

Se ne può dunque concludere che fin tanto che l’IA non giunga al punto di singolarità tecnologica [13] e sia completamente delegato dunque il processo alla base dei flussi monetari anche a livello individuale, sarà necessario ibridare ed accompagnare le

evoluzioni tecnologiche finanziarie tramite risorse umane altamente formate e che siano in grado di tradurre il flusso informativo complesso e caotico in maniera semplice agli Shareholders e Stakeholders finali.

Infatti, tramite la scalabilità intrinseca alle Tecnologie Digitali e nello specifico al Cloud Computing, la sfida moderna per l'economia non sarà tanto sfruttare le economie di scala del capitale, ma quelle derivanti dalla conoscenza umana.

2.4. Sentiment Analysis e Mercati Finanziari

Una volta compresa l'importanza dell'ibrido uomo-macchina ai fini dei processi decisionali economici, possiamo analizzare come i primi tentativi di unificare le due competenze siano stati supportati dal recente sviluppo della Data Science, e più nello specifico degli algoritmi di Sentiment Analysis e dell'incremento dei dati disponibili Online.

Infatti, già dall'Ottobre 2010 Johan Bollen, Huina Mao, Xiao-Jun Zeng hanno tentato di impiegare Algoritmi di Sentiment Analysis per processare quasi 10 milioni di Tweet (appartenenti a circa 2,7 milioni di utenti) con l'obiettivo di trovare una correlazione tra il Sentiment del mercato campionato su Twitter con il Dow Jones Industrial Average (DJIA) [14].

Le metodologie impiegate si basarono su 2 diverse tipologie di algoritmo:

- OpinionFinder, che misura se un commento è positivo o negativo
- Google-Profile of Mood States (GPOMS) che stima il Sentiment in ben 6 diverse dimensioni (Calma, Allerta, Sicurezza, Vitalità, Gentilezza, e Felicità)

I risultati per l'epoca furono senza precedenti: l'accuratezza nel predire l'andamento del DJIA tramite l'evoluzione emotiva del mercato campionato tramite Twitter fu pari al 87.6%

Da allora furono numerosi i tentativi di emulare l'impresa compiuta dai 3 Neo Data Scientist, tra le quali il lavoro di Anshul Mittal e Arpit Goel, entrambi professori alla Stanford University, che basandosi sul lavoro precedentemente citato hanno ottenuto tramite l'impiego di "Self Organizing Fuzzing Neural Networks" (SOFNN) ed una metodologia molto simile a quella precedente un'accuratezza pari a circa il 75.56% [15].

I risultati di queste ricerche sono di fondamentale importanza: infatti, essere in grado di velocizzare il trasferimento delle informazioni ai mercati finanziari non solo consente di rendere i mercati efficienti, ma l'analisi sottostante dell'origine e dell'evoluzione del

Sentiment stesso permetterebbe di anticipare ed evitare bolle speculative o manipolazioni esterne del mercato.

Nel testo “The Financial Consequences of Behavioural Biases” [16] è approfondito, infatti, il concetto Bias della Rappresentatività, dove interpretando degli studi del 2003 di Mitchell e Utkus sulla Finanza Comportamentale si giunge alla conclusione che le persone che sono tenute a prendere decisioni difficili tendono a basarsi sulle informazioni facilmente disponibili: un fenomeno che se analizzato nel contesto dei piccoli investitori consente di far comprendere come in caso di situazioni complesse, il presentare informazioni per primi o in maniera semplice consenta di manipolare facilmente il mercato ed il suo Sentiment aggregato.

Una trattazione strutturata del Text Mining per la predizione dei mercati finanziari è contenuta nel Paper di Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah e David Chek Ling Ngo [17] del Giugno 2014.

Più nello specifico viene descritta la nascita di questa disciplina come ibrido tra la linguistica, il Machine Learning e l’Economia Comportamentale ed il suo Focus sull’analisi fondamentale.

Infatti, nel paper essendo descritti i mercati come poco efficienti, non sempre risulta possibile fare affidamento sulla semplice analisi tecnica, ossia dell’individuazione di specifici pattern intra-giornalieri nelle transazioni, in quanto questi ultimi non sempre consentono di comprendere cosa è effettivamente avvenuto.

L’analisi fondamentale su cui ci si concentra prevedendo invece l’interpretazione di informazioni non solamente finanziarie, ma provenienti da più fonti, viene descritta come più promettente, specialmente come ambito di applicazione del “Opinion Mining” (Sentiment Analysis).

L’ambito di applicazione dell’Opinion Mining in questo paper è analizzato in un ecosistema di fonti informative più ampio: dai Blog, alle Newsletter ai giornali sino ad arrivare a Twitter.

La caratteristica in comune di queste fonti di dati è la caratteristica strutturale dei Big Data che generano e diffondono quotidianamente: i dati essendo di tipo non strutturato necessitano dell'elaborazione da parte di algoritmi di Sentiment Analysis per poter garantire una struttura oggettiva al fine del loro impiego in analisi finanziarie.

La fase di strutturazione del dato viene definita come fase di Pre-Processing.

Ad essa sussegue la Feature-Selection che consiste nella scelta della caratteristica fondamentale sulla quale ponderare l'Output del proprio algoritmo sulla base di determinati Input (Feature-Representation).

A seguito della trattazione riguardo le modalità di estrazione e di classificazione delle caratteristiche ottenibili tramite la Sentiment Analysis, una maggiore attenzione è posta agli algoritmi di Machine Learning impiegabili:

- SVM, Support Vector Machine
- Algoritmo di Regressione
- Naïve Bayes
- Alberi Decisionali
- Algoritmi Combinatori
- Esperimenti Multi-Algoritmo

È importante per gli operatori economici comprendere le caratteristiche di ogni singola tipologia di algoritmo, in quanto per un dato livello di Dati disponibili Online, all'incrementare dell'efficienza e della potenza di calcolo degli algoritmi sottostanti si è in grado di efficientare i risultati in maniera scalabile, con impatti a livello Macroeconomico sui mercati economici e finanziari.

Ai fini della trattazione, nel prossimo capitolo ci concentreremo sull'algoritmo di Regressione tramite l'impiego del modulo Python Sci-Kit Learn e del modulo TextBlob per la Sentiment Analysis, al fine di dimostrare come i nuovi modelli di Business di relazionano con i mercati finanziari e come quest'ultimi basano le proprie valutazioni sui primi.

3. Capitolo Terzo: Analisi Finanziaria con i Big Data

3.1. Sentiment Analysis: un'Applicazione Pratica in Python

Una volta determinato i processi alla base della finanza comportamentale, i loro sviluppi affiancati dalle tecnologie digitali, ossia dall'incremento dei flussi di idee e dell'interdipendenza, e infine i loro futuri sviluppi rappresentati da una maggiore razionalizzazione tendente verso lo scenario descritto dalla finanza neoclassica, può essere utile domandarsi come utilizzare al meglio le tracce digitali derivanti dalle interazioni tra i nodi all'interno di un Social Network al fine di campionare ed interpretare lo status quo della percezione prevalente e del flusso delle idee.

È infatti la tipologia di informazione (positiva o negativa) e la sua velocità di diffusione all'interno di una rete sociale, ovvero della sua effettiva propagazione, insieme al livello di fiducia, a determinare il meccanismo di formazione dei prezzi sui mercati finanziari in un'economia prevalentemente determinata da operatori il cui processo decisionale fonda sulla mente biologica.

I recenti sviluppi del Cognitive Computing e la rapida diffusione dei Social Network che coinvolgono a livello globale una fascia sempre più ampia della popolazione possono giungere in soccorso nell'identificazione delle evoluzioni dinamiche dell'intelligenza collettiva online e di come questa effettivamente interagisca sui sistemi economici e finanziari.

Come inizialmente introdotto, spesso gli esseri umani che interagiscono online producono dati non strutturati, come testi, opinioni e messaggi il cui contenuto può variare sia in termini di argomenti che di polarità (positivo, neutro o negativo).

Su questo aspetto interviene la Sentiment Analysis, dove algoritmi di Natural Language Processing rendono possibile estrarre e di conseguenza quantificare informazioni soggettive sullo stato emotivo di un operatore economico che interagisce online.

L'obiettivo di questa tesi è attuare una dimostrazione pratica e sperimentale di come questo sia possibile e come siano applicabili queste tecniche e queste metodologie

introdotte dalle nuove Tecnologie Digitali per fini utili alle scienze economiche e della gestione aziendale.

Per fare questo dunque mi avvarrò dei Big Data disponibili online sui Social Network tramite tecniche di Web Scraping, che strutturerò in DataSet tramite l'impiego di algoritmi di NLP, il tutto utilizzando il linguaggio di programmazione Python.

La Sentiment Analysis generalmente si divide in 3 fasi:

- 1) Raccolta DataSet da Big Data disponibili
- 2) Applicazione degli algoritmi di NLP al DataSet
- 3) Presentazione dei risultati tramite grafici

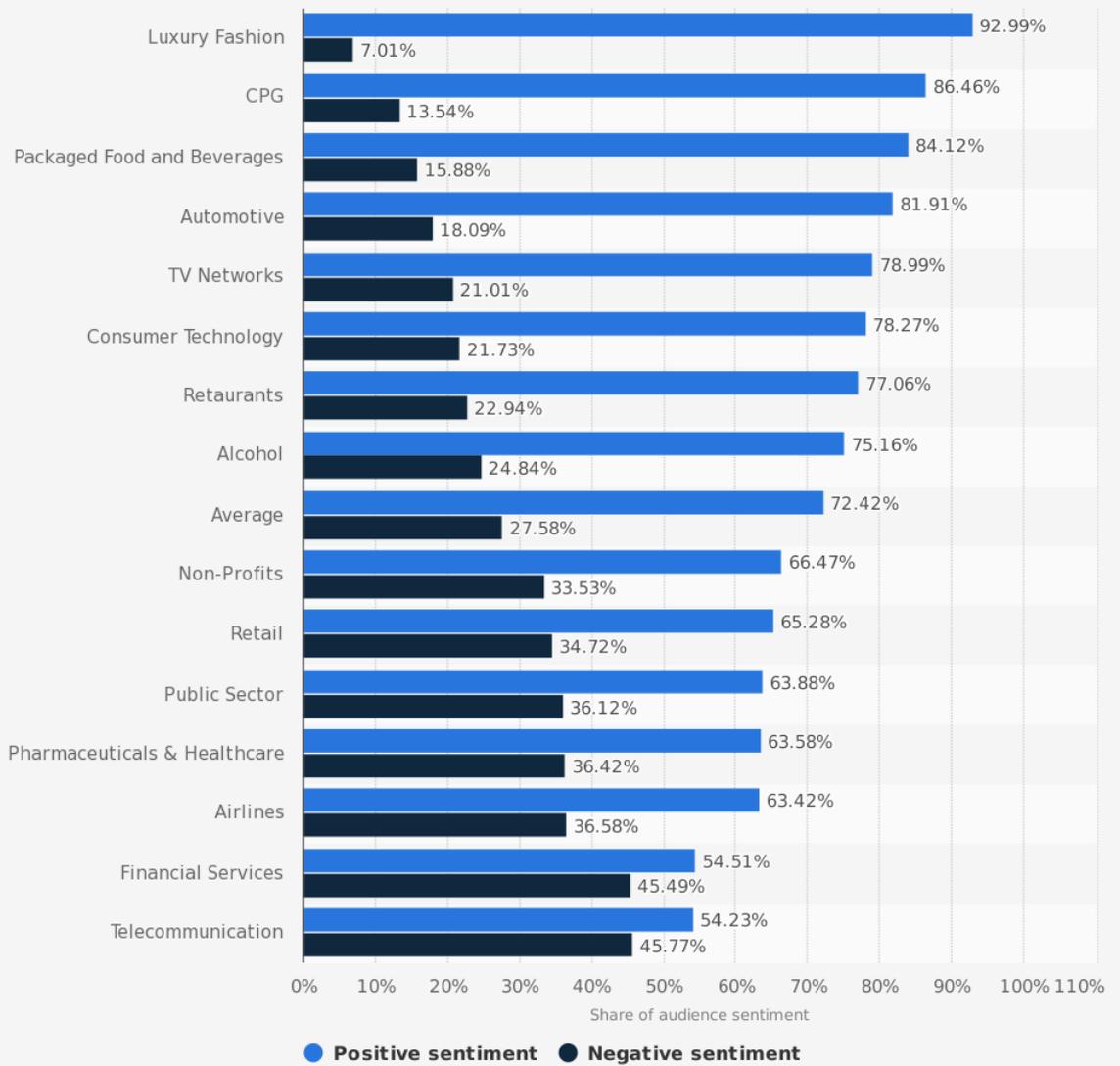
Per la prima fase tra i Social disponibili da dove ottenere il DataSet ho scelto Twitter poiché il mix tra accessibilità, semplicità d'uso e standardizzazione dei contenuti oltre il fatto che il contesto per cui viene impiegato riguarda spesso questioni economico-politiche, che ne fanno l'ideale per questa tipologia di analisi su larga scala per quanto riguarda l'intera popolazione, o anche per finalità campionarie.

Per farsi un'idea, come dimostra il grafico 3.1 i Brands che rientrano nel Luxury ed il Fashion ottengono il consenso più ampio tra la popolazione, mentre al polo opposto le Telecomunicazioni sono la categoria più dibattuta dagli utenti del Social.

Come azienda soggetto del nostro esperimento di Sentiment Analysis sceglieremo Samsung, dove sono stati raccolti le opinioni (Tweet e Re-Tweet) del periodo compreso tra il 18 ed il 28 Agosto 2018 per un totale di circa 14500 osservazioni.

L'accesso ai Big Data di Twitter per questo specifico caso è avvenuto tramite API, delle "chiavi d'accesso" ai Big Data rese disponibili da parte di Twitter agli sviluppatori, di conseguenza ho utilizzato Python ed alcuni suoi moduli per normalizzare i dati e renderli strutturati inserendoli in un Data Set con estensione CSV (Immagine 3.2).

Average audience sentiment of brands on Twitter as of 1st half 2017, by vertical



Source
Brandwatch
© Statista 2018

Additional Information:
Worldwide; H1 2017; 450 respondents

Grafico 3.1

```

import tweepy
import csv
import pandas as pd
import sys

# TWITTER API
consumer_key = '████████████████████████████████████████'
consumer_secret = '████████████████████████████████████████████████████████████████████████████████'
access_token = '████████████████████████████████████████████████████████████████████████████████'
access_token_secret = '████████████████████████████████████████████████████████████████████████████████'

non_bmp_map = dict.fromkeys(range(0x10000, sys.maxunicode + 1), 0xffffd)

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

api = tweepy.API(auth,wait_on_rate_limit=True)

# Creazione del DataSet
DataSet = open('████████████████████████████████████████████████████████████████████████████████\\Samsung.csv', 'a')

#Variabile per l'inserimento dei Big Data all'interno del DataSet
InputDataSet = csv.writer(DataSet)

#Accesso ai Big Data su Twitter tramite le API
for tweet in tweepy.Cursor(api.search,q="#samsung",count=100,
                           lang="en",
                           since="2010-01-01").items():
    #pre-processing e normalizzazione dei Tweet
    message = (tweet.text).translate(non_bmp_map)

    #Inserimento dei Big Data all'interno del DataSet
    InputDataSet.writerow([tweet.created_at, tweet.text.encode('utf-8')])

# Stefano Ciccarelli

```

Ln: 39 Col: 20

Immagine 3.2

Il risultato è dunque un DataSet con all'interno circa 14500 righe e 2 colonne, la prima che contiene informazioni riguardo l'orario di pubblicazione del Tweet e la seconda che invece riguarda il suo contenuto.

Per ottenere il Sentiment è però prima necessario cercare di capire che tipo di dati che ci troviamo davanti tramite un esempio.

```

from textblob import TextBlob
import pandas as pd
import datetime

#Accedo al DataSet appena creato tramite i Big Data
messaggi = pd.read_csv("██████████████████████\Samsung.csv",parse_dates=["Data"])

#Converto la Data in un formato standardizzato rendendola di tipo "datetime"
messaggi["Data"] = pd.to_datetime(messaggi.Data)

#Elimino le righe vuote
messaggi.dropna(inplace = True)

#Salvo in un nuovo DataSet
messaggi.to_csv("██████████████████████\Samsung_DataSet.csv")

#Accedo al nuovo DataSet e trasformo la colonna del DateTime in Indice del DataSet
messaggi = pd.read_csv("██████████████████████\Samsung_DataSet.csv", index_col=0)
messaggi.index = messaggi.Data

#Creo una funzione per ricevere come Output il livello di Sentiment
#dato come Input un messaggio di Testo
def Sentiment(messaggio):
    return TextBlob(messaggio).sentiment.polarity

#Creo una nuova colonna dove è contenuto il livello di Sentiment per ogni messaggio corrispondente
messaggi['Sentiment'] = messaggi["Testo"].apply(Sentiment)

#Salvo la versione definitiva del DataSet con Sentiment
messaggi.to_csv("██████████████████████\Samsung_DataSet_Sentiment.csv")

#Stefano Ciccarelli
|

```

Immagine 3.3

Nell'immagine 3.3 eseguo una prima fase di pulizia del Dataset per lavorare successivamente su di esso in maniera più efficiente, inoltre lo rendo disponibile sin da subito per ricavare informazioni, come mostrato dalle funzioni:

- Describe, che ci restituisce il numero di osservazioni presenti nel DataSet, il numero di osservazioni uniche, l'osservazione più diffusa e dunque la sua frequenza all'interno delle osservazioni totali in caso di osservazioni qualitative, altrimenti alcune informazioni statistiche in caso di osservazioni quantitative.
- Head o Tail, che ci restituiscono rispettivamente le prime 5 le ultime 5 osservazioni del DataSet in modo tale da comprendere sin da subito il tipo di osservazione su cui si andrà a lavorare.

```

>>> messaggi.describe()
count          14413  Testo
unique         7347  Numero di osservazioni uniche
top            b"RT @PoeticCases: We're having another giveaw...  Twitt più frequente (Re-Twitt)
freq          3498  Numero di volte presente il Twitt più frequente (numero
>>> messaggi.tail()
                        di Re-Twitt di un unico Twitt originale)
                        Testo
Data
8/18/2018 18:43 b"RT @PoeticCases: We're having another giveaw...
8/18/2018 18:41 b'MICRO SD CARD SAMSUNG EVO PLUS MB-MC128G 128...
8/18/2018 18:41 b'RT @monsanto_corp: BACK STREAMING LETS GO!!...
8/18/2018 18:41 b'BACK STREAMING LETS GO!! GRIND MODE \nhttps...
8/18/2018 18:40 b'Shoreline Samsung Case #etsy #accessories #c...
>>> #Stefano Ciccarelli

```

Ln: 262 Col: 23

Immagine 3.4 (Senza Sentiment)

```

>>> messaggi.head()
Data \
Data
2018-08-28 18:35:00 2018-08-28 18:35:00
2018-08-28 18:34:00 2018-08-28 18:34:00
2018-08-28 18:31:00 2018-08-28 18:31:00
2018-08-28 18:31:00 2018-08-28 18:31:00
2018-08-28 18:30:00 2018-08-28 18:30:00
Testo \
Data
2018-08-28 18:35:00 b'll Best Accessories To Buy From The Market F...
2018-08-28 18:34:00 b'RT @Androidheadline: Win A #Samsung #Galaxy ...
2018-08-28 18:31:00 b'RT @Androidheadline: Win A #Samsung #Galaxy ...
2018-08-28 18:31:00 b'RT @Framed: Want one? Like + Retweet & F...
2018-08-28 18:30:00 b'Samsung Galaxy Note 9 Screens have a Horribl...
Sentiment
Data
2018-08-28 18:35:00 0.568182
2018-08-28 18:34:00 0.800000
2018-08-28 18:31:00 0.800000
2018-08-28 18:31:00 0.585227
2018-08-28 18:30:00 -1.000000
>>> messaggi.describe()
Sentiment
count 14413.000000
mean 0.187592
std 0.278973
min -1.000000
25% 0.000000
50% 0.101010
75% 0.307143
max 1.000000
>>> #Stefano Ciccarelli

```

Immagine 3.5 (Con Sentiment)

Abbiamo quindi una situazione dove i Tweet unici sono circa la metà rispetto a quelli complessivi, il che implica che avremo in metà del DataSet dei Re-Tweet, dunque non opinioni personali di chi le condivide, ma opinioni altrui in cui ci si rispecchia, tendenti verso il valore del Sentiment Originale.

Nella seconda parte dell'Immagine 3.4 invece possiamo osservare come il DataSet sia suddiviso in 2 colonne: la prima che riporta la data, di pubblicazione o del re-tweet, la seconda invece riporta il Tweet in sé, con dicitura iniziale RT in caso si tratti di un Re-Tweet.

Con la creazione di una funzione (definita "Sentiment"), disponibile nell'Immagine 3.3, utilizziamo un algoritmo di Natural Language Processing che data una stringa di testo come Input, offre il valore del Sentiment corrispondente come Output.

Nell'Immagine 3.5 dunque abbiamo una descrizione del Sentiment, tra cui vi è il valore medio pari a circa 0,188 (su una scala che va da -1, ossia il valore esperienziale peggiore possibile, ad 1, il migliore possibile) ed una deviazione standard pari a circa 0,279.

Sempre nell'Immagine 3.5 ritroviamo che oltre ad aver corrisposto il valore del Sentiment per ogni osservazione disponibile, abbiamo reso la Data come indice del DataSet sotto forma di Datetime (il cui formato standardizzato ci permetterà di applicare algoritmi appositi).

Il nostro obiettivo è però ottenere una serie storica giornaliera dell'andamento del Sentiment in modo tale da poter effettuare poi successivamente una correlazione con le serie storiche delle quotazioni di borsa alla chiusura, come sarà dimostrato nel prossimo paragrafo.

Di conseguenza, come illustrato nell'Immagine 3.6, applico una la funzione "resample('D').mean()" (grazie alla conversione della data in formato Dataframe e soprattutto avendola resa come indice del DataSet) per trasformare la colonna "Sentiment" su base giornaliera, ottenendo come valore la media dei valori intra-giornalieri.

```

from textblob import TextBlob
import pandas as pd
import datetime
import matplotlib.pyplot as plt
|
#Accedo al DataSet appena creato tramite i Big Data
messaggi = pd.read_csv("████████████████████\\Samsung_DataSet_Sentiment.csv", parse_dates = [0], index_col = 0)

#Creo un nuovo DataSet --- Sentiment_Giornalieri
#Ogni Riga al posto di rappresentare un Tweet rappresenta un singolo giorno
#Il valore del Sentiment sarà una media dei Sentiment di ogni giorno
Sentiment_Giornalieri = messaggi[["Sentiment"]].resample("D").mean()

#Salvo il nuovo DataSet
Sentiment_Giornalieri.to_csv("████████████████████/DataSet_sentiment.csv")

#Stefano Ciccarelli

```

Immagine 3.6

Successivamente creo un altro DataSet il cui contenuto è l'Indice, ovvero la data (con i giorni) e il Sentiment, per un totale di 11 osservazioni, Immagine 3.7.

```

>>> Sentiment_Giornalieri
      Sentiment
Data
2018-08-18    0.142893
2018-08-19    0.157955
2018-08-20    0.146676
2018-08-21    0.222794
2018-08-22    0.206512
2018-08-23    0.165144
2018-08-24    0.224115
2018-08-25    0.175228
2018-08-26    0.196319
2018-08-27    0.190473
2018-08-28    0.162767
>>> #Stefano Ciccarelli|

```

Immagine 3.7

```

>>> Sentiment_Giornalieri.Sentiment.plot(kind = "line")
<matplotlib.axes._subplots.AxesSubplot object at 0x000002670FCBE588>
>>> plt.show()
#Stefano Ciccarelli|

```

Immagine 3.8

Nell'Immagine 3.8 sono contenuti i comandi inseriti nell'IDLE di Python che consentono di ottenere come Output un grafico a linee che mostra l'andamento del Sentiment medio su base giornaliera, Immagine 3.9.

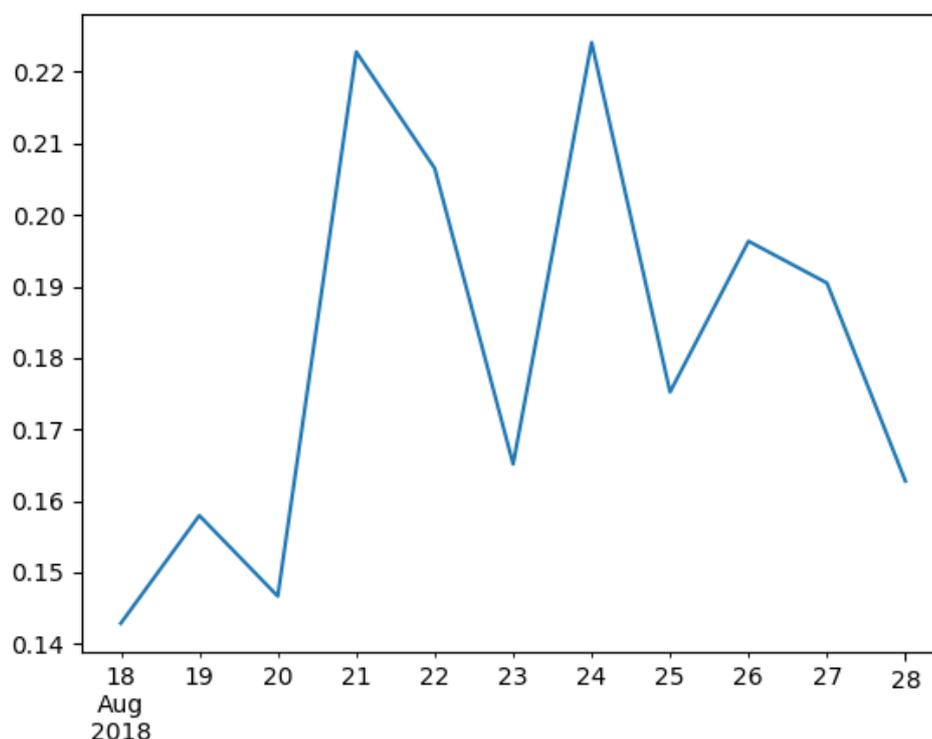


Immagine 3.9

Questo è quindi l'andamento e la variazione del Sentiment medio giornaliero della folla tra il 18 ed il 28 Agosto 2018 rilevabile tramite Twitter, che in un certo senso esprime una percezione media discretamente positiva verso i prodotti Samsung.

In un contesto Aziendale è molto importante valutare la percezione generale d'Impresa tramite la Sentiment Analysis, in quanto l'opinione collettiva non influenza unicamente i Ricavi immediati tramite l'influenza sociale che conduce all'incremento delle vendite, ma agisce anche sulla creazione di valore per gli Shareholders nel lungo termine in quanto è una metrica che sintetizzata può essere utilizzata nel determinare la percezione e la fiducia del mercato nei confronti di una determinata azienda.

Essendo gli algoritmi precedentemente esposti scalabili, e dunque applicabili su una maggiore quantità di dati, ed inoltre i Big Data disponibili online riguardando situazioni anche più di lungo termine, ho deciso di raccogliere i dati riguardo al Sentiment di Samsung sul mercato USA a partire dal 2010, raccogliendo circa 2,5 milioni e mezzo di Tweet ed utilizzando un processo simile a quello precedente per le misurazioni di breve termine.

Il risultato finale è rappresentato nell'Immagine 3.10

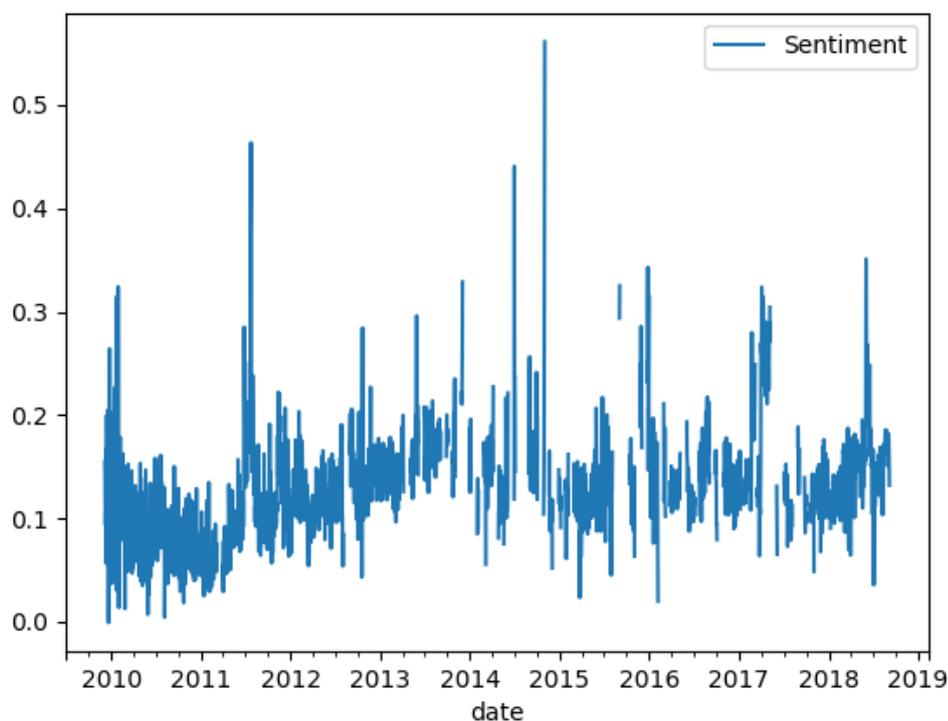


Immagine 3.10

Il risultato raffigura una percezione nettamente positiva, ma è molto importante includere nell'analisi anche l'incremento del flusso delle idee (ovvero delle quantità di idee scambiate e di operatori economici che vi partecipano) al fine di effettuare un'approfondita analisi storica, dunque volendo attuare un'analisi non solo dell'andamento della percezione media quotidiana di mercato, ma anche l'incremento complessivo dell'interazione online come aggregato giornaliero è possibile ottenere un

risultato che sintetizza, tramite la somma dei valori infra-giornalieri, il Sentiment della Community (Immagine 3.11).

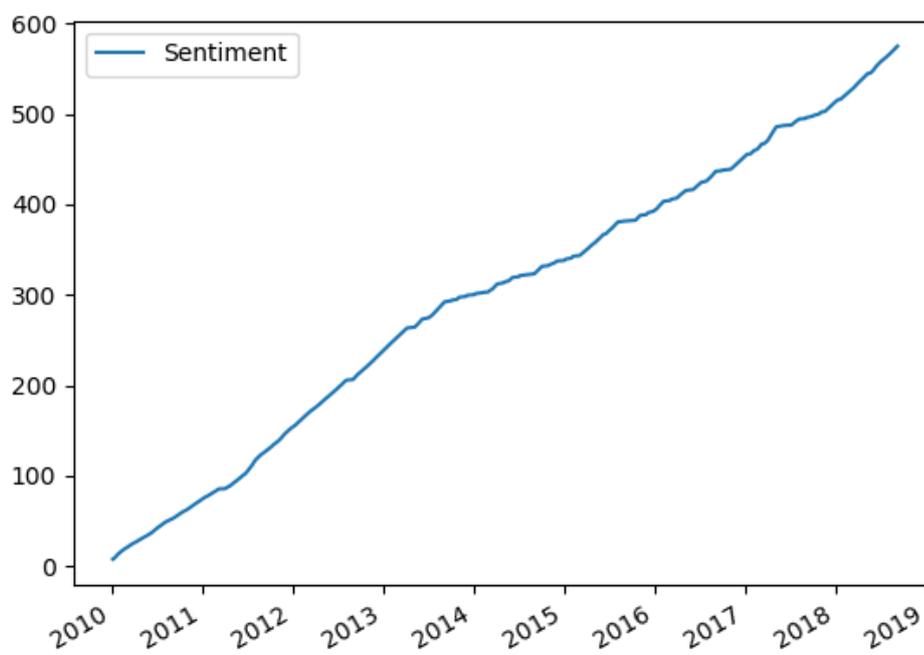


Immagine 3.11

Il significato di questa relazione è che non solo Samsung è riuscita ad incrementare la base utenti, ma nel tempo ha saputo garantirsi un Sentiment Netto positivo nel flusso delle idee ed opinioni riguardo l'azienda, in quanto la diffusione di un'Idea o un'Opinione con un Sentiment negativo influenza negativamente la relazione descritta nel grafico.

3.2. Analisi Finanziaria tramite il Machine Learning

Alcuni dei recenti maggiori scandali manageriali, fondati sul dilemma Agent-Principal, sono stati determinati da due assunti fondamentali:

- Il compito del Manager è quello di massimizzare il valore per gli Shareholders, in modo tale da massimizzare il benessere di tutti gli operatori economici
- Il Valore creato è misurabile grazie ai mercati, considerati efficienti, dunque tramite il prezzo delle azioni sul mercato, o più precisamente in base al loro incremento nel tempo.

Il secondo assunto però decade nel momento in cui il mercato, composto da operatori economici umani, viene influenzato da percezioni, comunicazioni e flussi di idee fondamentalmente biasati o inconsistenti, ossia prive di basi fattuali o percezioni erronee delle tali.

Questo sistema ha generato forti incentivi affinché i Manager finissero per falsificare o diffondere informazioni finanziarie erronee o false, in modo tale da incrementare il valore percepito dell'azienda sul mercato.

Il problema di questo meccanismo è la sua difficoltà di identificazione ed interpretazione che apre nuove sfide alla regolamentazione ed al legislatore.

All'identificazione di queste fattispecie va affiancata dunque una correlazione del Sentiment di mercato, che a seguito dell'atomizzazione delle relative fenomenologie originatrici consente di determinare, valutando la bontà di quest'ultime, se vi è un incremento qualitativo percettivo dell'impresa consistente e assiomatico o se vi è un'effettiva volontà esogena di manipolare la percezione collettiva.

Infatti, all'origine di una variazione del Sentiment di mercato di un'azienda vi è un flusso informativo e di idee che possono dipendere da fonti multilaterali che diffondono notizie utili al processo decisionale economico-finanziario individuale.

Se però alla base di un incremento del Sentiment di mercato vi fosse una serie di notizie e di flussi di idee i cui assiomi sono effettivamente inconsistenti e non verificabili nella

realtà fattuale, ci si ritroverebbe, infatti, davanti ad un tentativo di manipolazione della percezione collettiva di mercato.

Un'analisi finanziaria simile diventa possibile dunque grazie ai nuovi strumenti Digitali che consentono, tramite la corretta supervisione, di incrementare il benessere collettivo di lungo termine riducendo il rischio di moral hazard manageriale, o di speculazione da parte di terzi.

L'analisi finanziaria viene divisa in 3 parti:

- 1) Ottenimento delle serie storiche di mercato riguardo una specifica azienda, ed in particolare i prezzi alla chiusura
- 2) Unione del DataSet delle serie storiche dei prezzi alla chiusura con il DataSet delle serie storiche del Sentiment medio giornaliero campionato da Twitter
- 3) Tracciamento in scala dei DataSet e generazione del grafico

Per effettuare questa analisi è stata scelta come Target la società Tesla, Inc. Il periodo di riferimento è da inizio 2010 sino a Settembre 2018, il numero di osservazioni totali corrisponde a 992618 (circa 1 milione) e il criterio di determinazione è stato l'hashtag "#tesla" a cui è stata applicata la procedura di Sentiment Analysis mostrata nel paragrafo precedente.

L'obiettivo in questo paragrafo non è solo dimostrare la relazione tra Sentiment e Titolo in generale, ma anche come le società innovative ed i nuovi modelli di business si relazionano con i mercati finanziari, e come quest'ultimi determinano il prezzo di mercato dell'azione, che tende a discostarsi da quello determinabile da un'analisi fondamentale derivabile dai Financial Statements.

Iniziamo ottenendo il DataSet riguardo le serie storiche da Yahoo Finance (Immagine 3.11):

```

from pandas_datareader import data as pdr

import pandas as pd

import fix_yahoo_finance as yf

import matplotlib.pyplot as plt

from sklearn.preprocessing import StandardScaler

yf.pdr_override()

#Download delle Serie Storiche
dati = pdr.get_data_yahoo("TSLA", start="2010-01-01", end="2018-09-05")

#Inserimento delle Serie Storiche in un DataSet
Serie_Storiche = pd.DataFrame(dati)

#Salvataggio del DataSet
Serie_Storiche.to_csv("C:\[redacted]//Serie_Sto

#Accedo al DataSet appena creato tramite i Big Data
#Tramite il valore "parse_dates = [0]" trasformo l'indice in un Datetime
Serie = pd.read_csv("C:\[redacted]//Serie_Storic

#Riduco in scala i valori tramite StandardScaler - Modulo Scikit-Learn
Scala = StandardScaler()
Prezzi_Volumi = Serie[["Volume", "Close"]]
Prezzi_Volumi_InScala = Scala.fit_transform(Prezzi_Volumi)

#Stefano Ciccarelli|

```

Immagine 3.11

Una volta ottenuto dunque le Serie Storiche dal 2010 al 2018 delle azioni Tesla è possibile già ottenere alcuni Insights ed informazioni utili, infatti eseguendo i comandi si ottiene, in ottica esplorativa, il grafico che mostra l'andamento del prezzo alla chiusura, mostrato nell'Immagine 3.12

Di conseguenza possiamo anche ottenere, trasformando i valori in scala, il grafico che mostra la relazione tra il prezzo alla chiusura e le quantità acquistate, che definisce la dinamicità, sui mercati finanziari, di quell'azienda come mostrato nell'Immagine 3.13. Questa funzione si rivela un valido strumento che mette in relazione l'andamento del titolo con la sua rispettiva liquidità sul mercato.

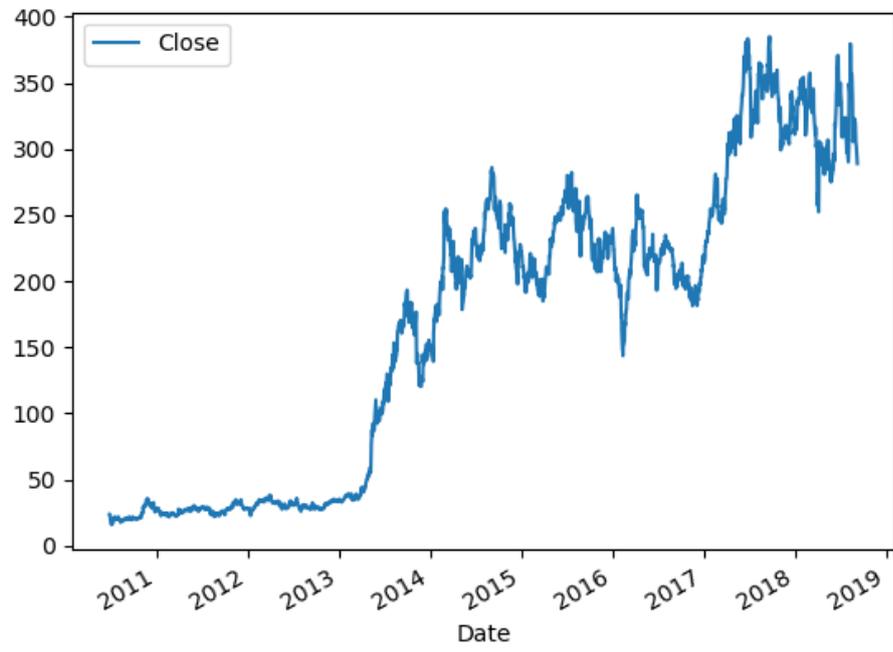


Immagine 3.12

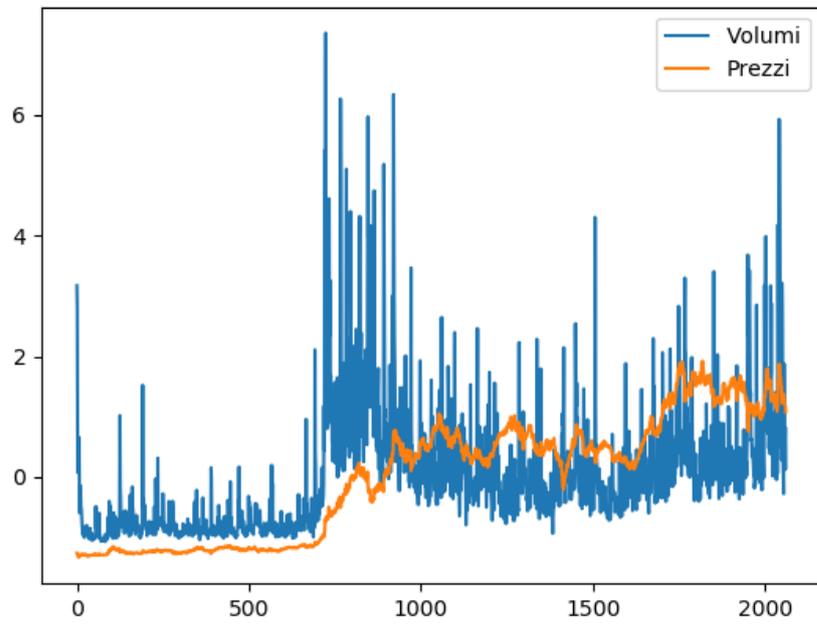


Immagine 3.13

Pertanto, come già anticipato abbiamo anche il DataSet contenente i Twitt riguardo Tesla, a cui applichiamo, come anticipato, su larga scala la procedura descritta nel paragrafo precedente, ottenendo di conseguenza il grafico contenuto nell'Immagine 3.14

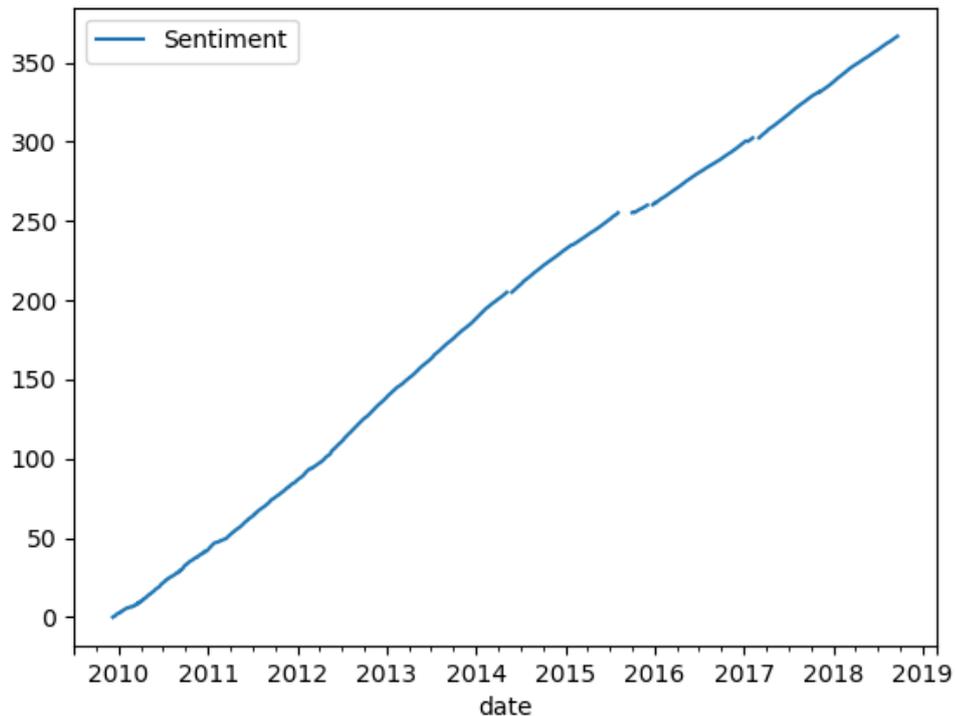


Immagine 3.14

Di conseguenza tramite l'utilizzo del modulo Standard Scaler contenuto in Scikit-Learn (per gli applicativi di Pre-Processing) rendiamo i valori dei prezzi alla chiusura del Titolo Tesla ("Close") in scala rispetto all'aggregato giornaliero del Sentiment della Community Online.

I comandi di Input sono descritti nell'Immagine 3.15:

```

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.preprocessing import StandardScaler

import seaborn as sns

#Accedo al DataSet delle Serie Storiche
DataSetSerie = pd.read_csv("C:\[redacted]//Serie_Storiche

#Accedo al DataSet del Sentiment
DataSetSentiment = pd.read_csv("C:\[redacted]//DataSet_

#Unisco la colonna dei prezzi di chiusura del DataSet delle Serie Storiche
#Insieme alla colonna Sentiment del DataSet del Sentiment
Prezzi_Sentiment = pd.concat([DataSetSerie.Close,DataSetSentiment.Sentiment], axis = 1)

#Creo una nuova colonna dove è contenuto il valore del Sentiment del giorno precedente
Prezzi_Sentiment["Sentiment Ieri"] = Prezzi_Sentiment["Sentiment"].shift(1)

#Rimuovo i valori mancanti derivanti dalle giornate di chiusura della borsa
Prezzi_Sentiment.dropna(inplace = True)

#Salvo il nuovo DataSet Serie_Storiche - Sentiment
Prezzi_Sentiment.to_csv("C:\[redacted]//Tesla_Prezzi_TO

#Riduco in scala i valori tramite StandardScaler - Modulo Scikit-Learn
Scala = StandardScaler()
Prezzi_Sentiment_InScala = Scala.fit_transform(Prezzi_Sentiment)

#Stefano Ciccarelli

```

Immagine 3.15

Tramite i vari comandi contenuti nell'Immagine 3.15, creiamo un'ulteriore colonna il cui contenuto corrisponde al Sentiment del giorno precedente, in modo da poter successivamente ottenere una correlazione che descriva anche come nel breve termine il prezzo di un determinato giorno sia connesso al valore del Sentiment del giorno precedente.

L'Output sarà dunque l'Immagine 3.16:

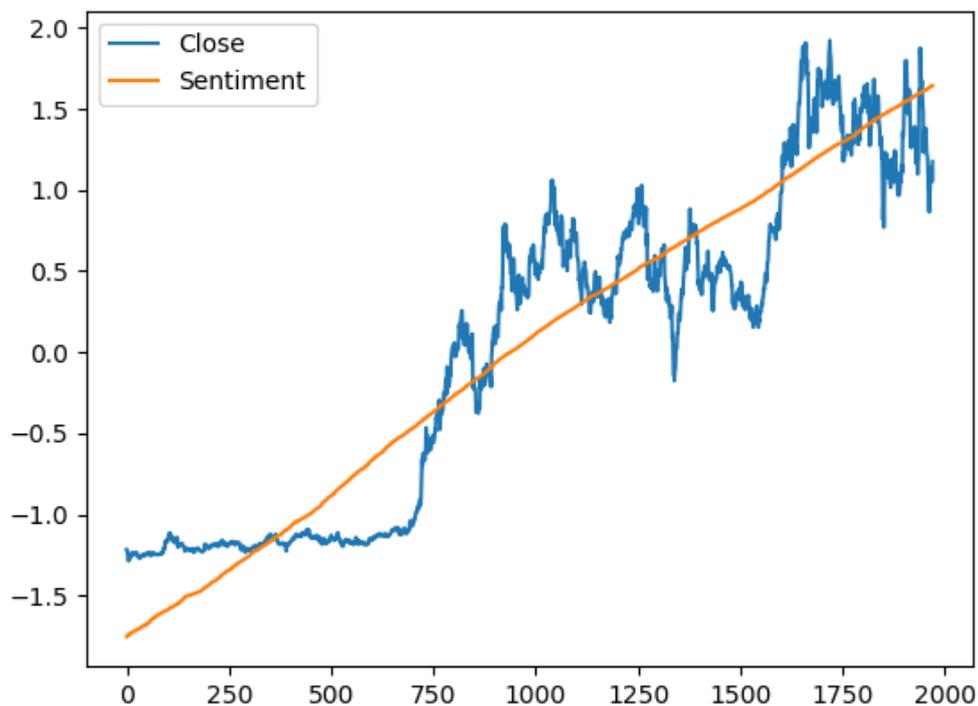


Immagine 3.16

```
>>> Prezzi_Sentiment.corr()
                Close  Sentiment  Sentiment Ieri
Close           1.000000  0.936775    0.936779
Sentiment       0.936775  1.000000    1.000000
Sentiment Ieri  0.936779  1.000000    1.000000
>>> #Stefano Ciccarelli
```

Immagine 3.17

Nella Command Line del IDLE di Python eseguendo il comando “.corr()”, come dimostrato dall’Immagine 3.17, è possibile analizzare la correlazione di lungo termine tra i Prezzi alla chiusura giornalieri (Close) e il valore aggregato del Sentiment, pari a circa il 94%

Inoltre, anche il Sentiment della giornata precedente dimostra avere una correlazione simile rispetto al prezzo di chiusura.

Di conseguenza applichiamo un modello di Regressione Lineare la cui variabile di risposta è il Prezzo alla Chiusura e la variabile sottostante il livello aggregato di Sentiment.

Rappresentiamo il tutto con una Linea di Best Fit nell'Immagine 3.18

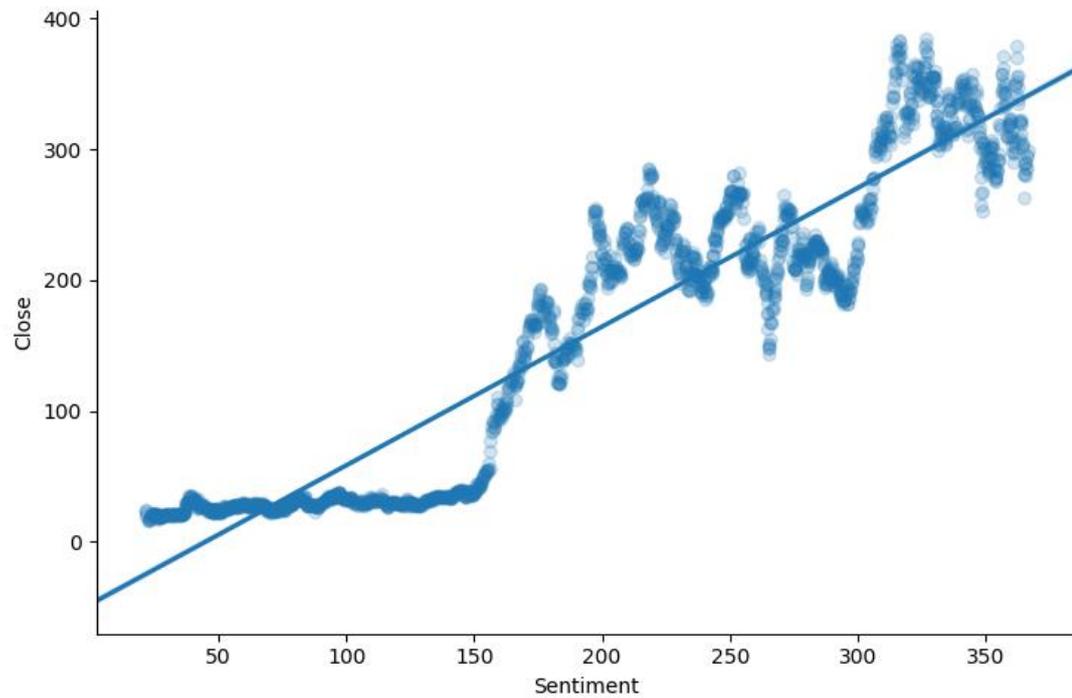


Immagine 3.18

Per trovare l'Intercetta e il coefficiente B del nostro modello di regressione abbiamo dunque bisogno del codice presente nell'Immagine 3.19

L'intercetta, ovvero β_0 , che corrisponde al valore del Prezzo alla Chiusura quando il Sentiment diventa pari a 0 è equivalente a circa $-47,82$

Invece il Coefficiente β_1 è pari a circa 1.06, esso rappresenta il valore di quanto si muove il Prezzo alla Chiusura rispetto ad una variazione unitaria del valore aggregato giornaliero del Sentiment

Provando ad effettuare delle stime otteniamo che se il Sentiment dovesse crollare a 250, ovvero di circa il 32% rispetto all'ultimo valore disponibile allora il prezzo dovrebbe raggiungere il valore di 217,47, un decremento pari a circa il 27%.

Per quanto riguarda invece le stime su un ipotetico incremento del Sentiment verso il 450 di valore aggregato generato dalla Community Online, allora scopriamo che il valore del titolo sul mercato potrebbe salire fino a 429,70, un incremento del 43,7% a fronte di un incremento del 22,8 % del livello del Sentiment.

$$y = \beta_0 + \beta_1 x$$

$$\text{Prezzo alla Chiusura} \cong -47.82 + 1.06 * \text{Sentiment}$$

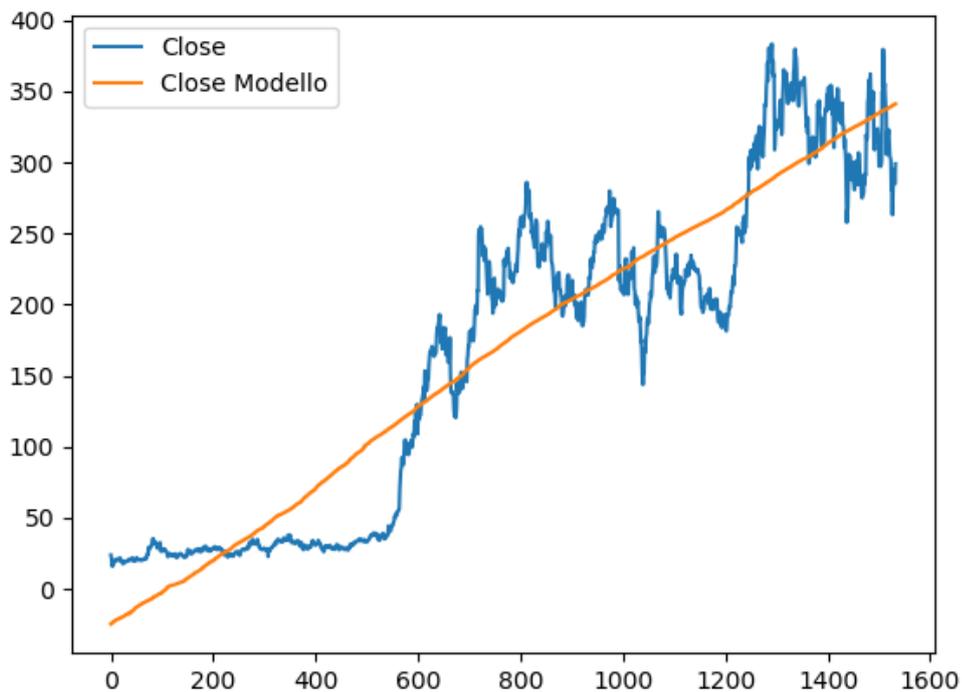


Immagine 3.22

Ricapitolando, abbiamo nel file CSV due serie storiche che rappresentano rispettivamente il valore del Prezzo alla Chiusura ed il valore del Prezzo Atteso sulla base del Sentiment di quel giorno ricavato dal modello di Regressione Lineare.

Creando una terza colonna possiamo rappresentare dunque l'andamento giornaliero del differenziale tra Prezzo Effettivo e Prezzo Atteso, mostrato nell'Immagine 3.23.

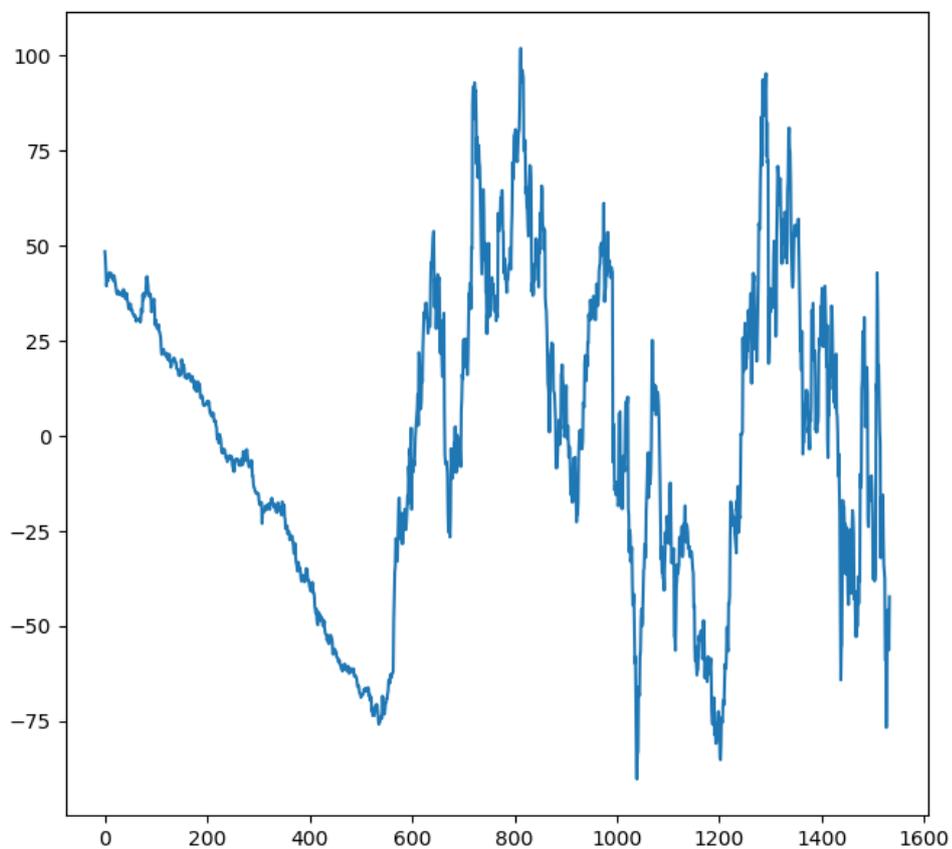


Immagine 3.23

Questa nuova serie storica ha le seguenti statistiche (Immagine 3.24):

```
>>> modello["Differenziale Close"].describe()
count      1533.000000
mean       -0.125566
std        40.441055
min        -90.201496
25%       -29.068674
50%         0.360435
75%        32.484555
max        101.934702
Name: Differenziale Close, dtype: float64
>>> #Stefano Ciccarelli
```

Immagine 3.24

Abbiamo dunque una media prossima al valore 0 ed una deviazione standard pari a circa 40.

È possibile ricavarne dunque che il valore atteso è tendente verso 0, di conseguenza che il prezzo effettivo tende nel tempo verso il prezzo atteso ricavato dal modello di regressione.

Questo tipo di analisi, oltre alla previsione del prezzo atteso alla chiusura sulla base del livello del Sentiment atteso, dimostra anche quando il prezzo effettivo è in surplus rispetto al prezzo derivabile dal reale sentiment di mercato, consentendo dunque di anticipare le possibili oscillazioni di prezzo sul mercato assumendo come stabile l'evoluzione del Sentiment.

```
import pandas as pd
import numpy as np
from scipy import stats

#Accedo al DataSet
modello = pd.read_csv("C:\[redacted]\Tesla_Prezzi_TOTSent

#Lista degli intervalli dei Range
ll = [-110,-100,-90,-80,-70,-60,-50,-40,-30,-20,-10,0,10,20,30,40,50,60,70,80,90,100,110]

#Creo un dizionario per contenere i Range e le corrispondenti frequenze
dl = {}

#Creo le Chiavi nel Dizionario dei Range
for x in ll:
    dl[str(x)+"|" +str(x+10)] = 0

#Trasformo la Serie Storica dei differenziali alla chiusura in una lista
test3 = list(modello["Differenziale Close"])

#Per ogni valore nella lista degli intervalli determino la rispettiva frequenza
for x in ll:
    for i in test3:
        if i > x and i <= int(x + 10):
            dl[str(x)+"|" +str(x+10)] += 1

#Trasformo il Dizionario in un DataSet
test = pd.Series(dl)

#Stefano Ciccarelli
```

Immagine 3.25

Dopo aver trasformato quindi le Serie Storiche dei Differenziali tra Prezzo Effettivo e Prezzo Atteso in una lista, e dopo averla ordinata in ordine crescente, eseguo in Python la lista di comandi contenuti nell'Immagine 3.25.

Creo un dizionario (denominato "d1"), avente come indici (o chiavi) il range dei numeri continui e come valori la frequenza delle osservazioni (quante osservazioni si trovano in quel range).

Di conseguenza avendo ottenuta la serie delle chiavi-valori, trasformo il dizionario in un DataSet, che a sua volta viene rappresentato sotto forma di grafico a barre come mostrato nell'Immagine 3.26.

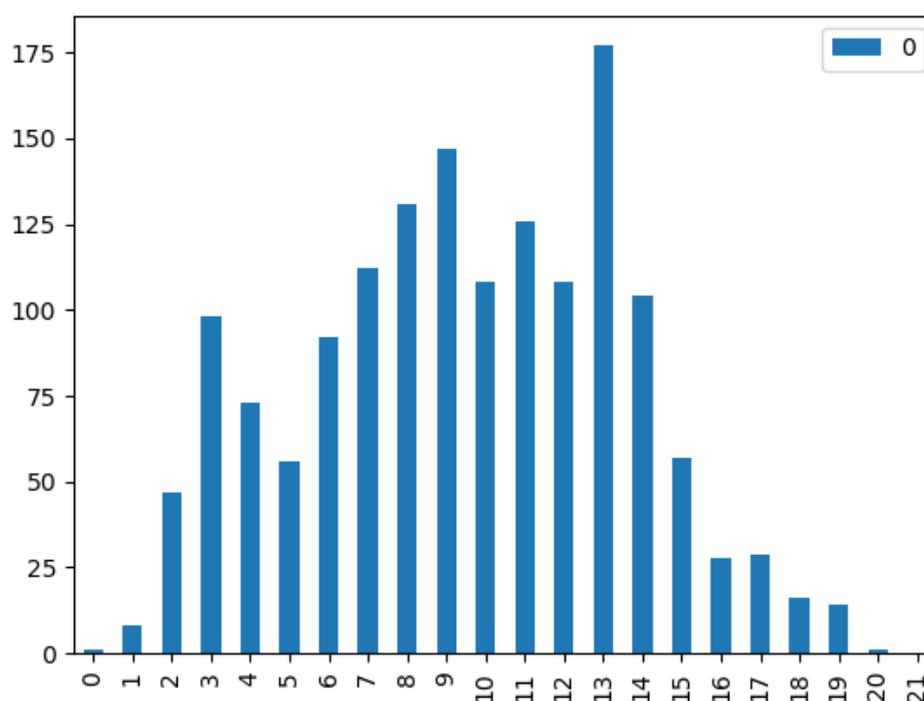


Immagine 3.26

L'indice dei valori corrispondenti (da 0 a 21) è rappresentato nell'Immagine 3.27.

Range Differenziale	Frequenza	Indice
-110 -100	0	0
-100 -90	1	1
-90 -80	8	2
-80 -70	47	3
-70 -60	98	4
-60 -50	73	5
-50 -40	56	6
-40 -30	92	7
-30 -20	112	8
-20 -10	131	9
-10 0	147	10
0 10	108	11
10 20	126	12
20 30	108	13
30 40	177	14
40 50	104	15
50 60	57	16
60 70	28	17
70 80	29	18
80 90	16	19
90 100	14	20
100 110	1	21
110 120	0	22
#Stefano Ciccarelli		

Immagine 3.27

```
>>> modello["Differenziale Close"].describe()
count      1533.000000
mean        -0.125566
std         40.441055
min         -90.201496
25%        -29.068674
50%         0.360435
75%         32.484555
max         101.934702
Name: Differenziale Close, dtype: float64
>>> sigma = modello["Differenziale Close"].std()
>>> mu = modello["Differenziale Close"].mean()
>>> x = np.linspace(mu - 3*sigma, mu + 3*sigma, 100)
>>> plt.plot(x, mlab.normpdf(x, mu, sigma))
[<matplotlib.lines.Line2D object at 0x000001C232303550>]
>>> plt.show()
#Stefano Ciccarelli|
```

Immagine 3.28

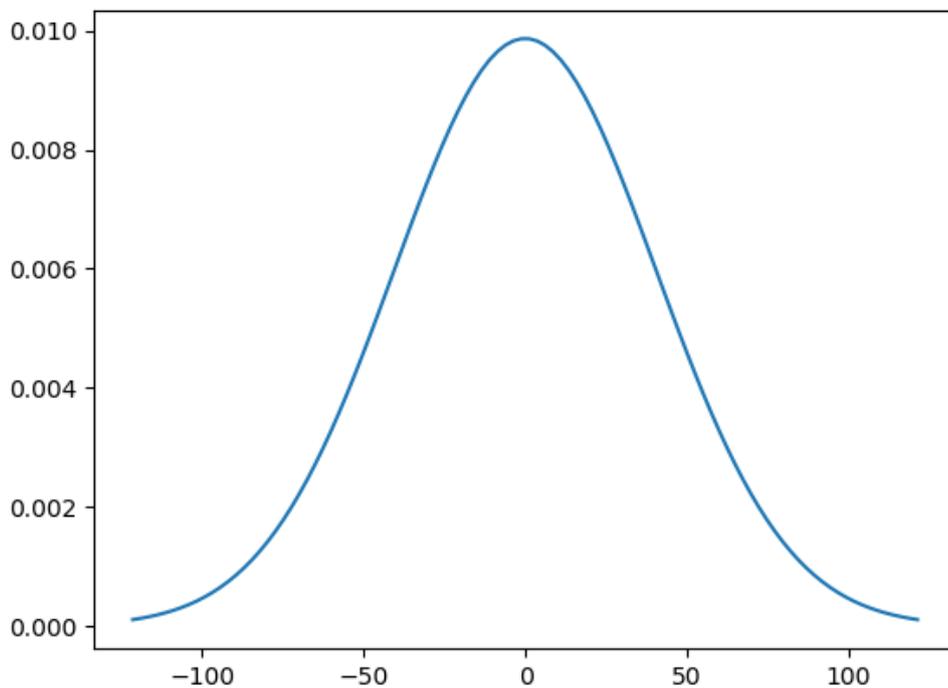


Immagine 3.29

Nell'Immagine 3.28 otteniamo ricaviamo la Deviazione Standard della Serie Storica dei Differenziali ed insieme alla sua media otteniamo il grafico illustrato nell'Immagine 3.29 che rappresenta la Gaussiana (o distribuzione Normale) corrispondente alla distribuzione normale della rispettiva media e deviazione standard.

La forma della distribuzione delle frequenze suddivise in range dei differenziali ci lascia intendere che la distribuzione sia Normale e con media tendente verso 0.

La sovrapposizione dei due grafici è mostrata nell'Immagine 3.30

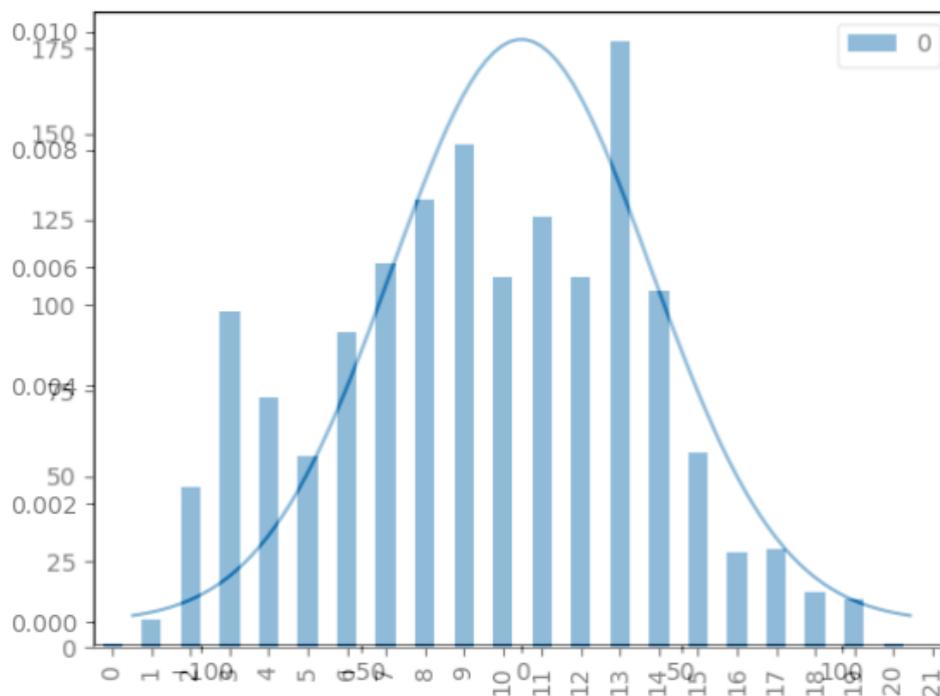


Immagine 3.30

Che la distribuzione delle frequenze sia normale viene di conseguenza confermato tramite un Test di Normalità (Immagine 3.31) che impiega il modulo Stats di Scipy (l'Ipotesi Nulla iniziale è che la distribuzione sia Normale):

```

>>> test1
      Range  Frequenze
0  -110|-100          0
1  -100|-90           1
2   -90|-80           8
3   -80|-70          47
4   -70|-60          98
5   -60|-50          73
6   -50|-40          56
7   -40|-30          92
8   -30|-20         112
9   -20|-10         131
10    -10|0         147
11     0|10         108
12    10|20         126
13    20|30         108
14    30|40         177
15    40|50         104
16    50|60          57
17    60|70          28
18    70|80          29
19    80|90          16
20    90|100         14
21   100|110          1
22   110|120          0
>>> l = list(test1["Frequenze"])
>>> k2, p = stats.normaltest(l)
>>> alpha = 1e-3
>>> if p < alpha:
    print("L'Ipotesi Nulla può essere Rigettata")
else:
    print("L'Ipotesi Nulla non può essere Rigettata")

L'Ipotesi Nulla non può essere Rigettata
>>> #Stefano Ciccarelli

```

Immagine 3.31

Il significato di questa “convergenza”, basata sulla “Legge dei Grandi Numeri” tra il Prezzo Effettivo ed il Prezzo Atteso derivato dal livello di Sentiment Aggregato, è da ritrovarsi nel processo di determinazione del valore attuale dei ricavi futuri dell’azienda stimati sulla base della fiducia e dunque del Sentiment presente sul mercato, che tende ad avere nel breve termine errori casuali strutturali a causa dell’attrito intrinseco tra il

mercato informativo-sociale, ossia l'insieme di tutti gli Stakeholders che sono soggetti al flusso delle idee e delle informazioni, ed il mercato finanziario, ossia l'insieme degli Shareholders che sono soggetti al processo non solamente di recezione ma bensì anche di razionalizzazione ai fini del processo decisionale finanziario, che può essere minimizzato o quantomeno velocizzato grazie al supporto di un sistema di Hybrid-Robo Advisor.

Se ne deriva dunque che il valore attuale sui mercati finanziari di un'azienda come Tesla dipende principalmente dalla fiducia che gli Shareholders e gli Stakeholders ripongono in essa e in particolare dall'evoluzione della percezione collettiva delle funzioni e del benessere economico atteso attribuitogli.

Questo "assunto" è valido intrinsecamente sempre più per i nuovi modelli di Business, che come Tesla impiegano anni prima di distribuire dividendi o addirittura raggiungere il Break Even Point, in quanto in assenza di previsioni e stime finanziarie attendibili la ricchezza dei nuovi modelli di Business si misura effettivamente tramite il Sentiment della Community Online e sull'ecosistema fiduciario che un'Azienda è in grado di costruirsi tramite le relazioni e le interconnessioni dirette garantite dalle Tecnologie Digitali.

4. Conclusione

Partendo da un'analisi delle Tecnologie Digitali, del Comportamento Umano e di come quest'ultimo si manifesta e misura Online in maniera aggregata ho cercato di stabilire un filo conduttore per comprendere come i modelli di Business si stiano spostando sempre verso più un'interdipendenza con l'intelligenza collettiva e come questi siano influenzati da come questa interagisce Online.

La dimostrazione che il Prezzo alla Chiusura di Tesla possa essere approssimato tramite la trasformazione lineare del livello del Sentiment Aggregato impiegando un algoritmo di Machine Learning come il Modello di Regressione è una nuova porta per le valutazioni finanziarie, in quanto il livello di fiducia e la grandezza della Community di un'azienda può avere un impatto notevole e delle implicazioni superiori alle stime e alle serie storiche economico-finanziarie al fine di aggiornare il valore aziendale futuro.

Non necessariamente questo è un fattore negativo, in quanto il Sentiment Aggregato può essere utilizzato come indice di misura del livello di gradimento dell'azienda sul mercato o come stimatore dell'intelligenza collettiva ("saggezza della folla") in caso di presenza di operatori razionali e professionali, ma allo stesso tempo non bisogna mai dimenticare di porre attenzione alle fonti di origine del Sentiment stesso.

Ad esempio, un improvviso incremento del Sentiment a seguito di notizie ed informazioni non verificabili o addirittura false potrebbe essere sintomo di una possibile bolla finanziaria presente sul mercato o di moral hazard manageriale, per questo è necessario accompagnare agli strumenti delle Tecnologie Digitali un ferreo spirito critico ed analitico.

Partendo da strumenti comuni di Opinion Mining e di Sentiment Polarity evidenziati nel capitolo 2.4, mi sono concentrato su un'analisi finanziaria che tendesse verso la ricostruzione del prezzo alla chiusura partendo unicamente dal valore del sentiment aggregato, piuttosto che tentare di predire le evoluzioni del mercato sulla base delle variazioni del sentiment come invece effettuato nei paper citati.

Questo fattore è però anche un limite del modello creato in questa tesi: non tutte le aziende, come ad esempio le unlisted companies, sono necessariamente altamente correlate con il livello del Sentiment, ma anzi si tratta di un'interrelazione tipica dei nuovi modelli di Business che stanno proliferando grazie alle Tecnologie Digitali, il cui Asset principale è costituito dalla Community Online che sono riusciti a coinvolgere.

Anche se i Big Data che ho campionato tramite milioni di osservazioni (applicando il modulo Twint di Python) sono stati rilevati da Twitter, ci sono numerose possibili applicazioni che possono essere introdotte come ad esempio il Web Scraping degli articoli di giornale, dei blog online, ecc.

Questo apre enormi possibilità per raffinare la qualità del modello finale e per la sua possibile integrazione con modelli di Algorithmic Trading e di Robo Advisor.

Infine, il modello finale dei differenziali dimostra come nel lungo termine, avendo una distribuzione Normale, il valore effettivo di mercato tende a corrispondere a quello derivato dalla trasformazione lineare del valore del Sentiment Aggregato, ma affinché si voglia utilizzare questo modello per il fine di ridurre il rischio finanziario specifico è importante tenere in considerazione che è necessario un ottimo modello predittivo del Sentiment, che tenga in considerazione la variabilità del flusso di informazioni (online e offline) verso il mercato e soprattutto del flusso delle idee tra gli agenti economici.

5. Note

5.1. Note Capitolo 1

- 1) <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>
- 2) <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>
- 3) Internet of Things, Stefano Za, Luiss University Press, ISBN 9788861053106
- 4) <https://healthcare.mckinsey.com/unlocking-potential-internet-things>
- 5) M. Hermann, T. Pentek and B. Otto, “Design Principles for Industrie 4.0 Scenarios”, 2016, 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, 2016, pp. 3928-3937.
- 6) <https://www.top500.org/news/market-for-artificial-intelligence-projected-to-hit-36-billion-by-2025/>
- 7) https://it.wikipedia.org/wiki/Carlo_Azeglio_Ciampi
- 8) https://it.wikipedia.org/wiki/Razionalit%C3%A0_limitata
- 9) J. Hawks, How has the Human Brain Evolved? “Scientific American”, 2013, p. 6.
- 10) Micah Edelson, Tali Sharot, R. J. Dolan e Y. Dudai, Following the Crowd: Brain Substrates Of Long-Term Memory Conformity, “Science”, 333, n. 6038, 2011, pp. 108-11.

5.2. Note Capitolo 2

1. <https://it.wikipedia.org/wiki/Decisione>
2. https://en.wikipedia.org/wiki/Nudge_theory
3. https://it.wikipedia.org/wiki/Saggezza_della_folla
4. E. Ferrara e Z. Yang, Measuring Emotional Contagion in Social Media, “PLoS One”, 10, n. 11, 2015, e0142390.
5. <https://hbr.org/2013/11/beyond-the-echo-chamber>
6. Alex Pentland, “Fisica Sociale: Come si propagano le buone idee”, UBE ISBN 9788883502286
7. Karlsson, Niklas, George Loewenstein e Duane Seppi, The Ostrich Effect; Selective Attention to Information, “Journal of Risk and Uncertainty”, 38, n.2, 2009, pp. 95-115.
8. Markowitz, H., 1952, "Portfolio Selection, "Journal of Finance, 7(1), 77-99
9. https://it.wikipedia.org/wiki/Robo_advisor
10. <http://www.smh.com.au/business/banking-and-finance/asics-greg-medcraft-says-robo-advice-can-reduce-fees-and-conflicts-20151105-gkrmxw.html>
11. “Fintech: La Finanza Digitale, Strategie di Investimento con i Roboadvisor”, Enrico Malverti, Davide Bulgarelli, Gabriele Villa, ISBN 9788820384883
12. Il pensiero flessibile, Leonard Mlodinow, ISBN 9788820063993
13. https://it.wikipedia.org/wiki/Singularit%C3%A0_tecnologica
14. <https://arxiv.org/pdf/1010.3003.pdf> , Twitter mood predicts the stock market. Johan Bollen, Huina Mao, Xiao-Jun Zeng
15. <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf> , Stock Prediction Using Twitter Sentiment Analysis, Anshul Mittal, Arpit Goel, Stanford University.
16. The Financial Consequences of Behavioral Biases: An Analysis of Bias in Corporate Finance and Financial Planning, Imad A. Moosa, Vikash Ramiah, ISBN 9783319693880
17. Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah e David Chek Ling Ngo, “Text Mining for Market Prediction: A systematic review”

6. Bibliografia

Testi approfonditi:

- 1) Internet of Things, Stefano Za, Luiss University Press, ISBN 9788861053106
- 2) Alex Pentland, “Fisica Sociale: Come si propagano le buone idee”, UBE ISBN 9788883502286
- 3) “Fintech: La Finanza Digitale, Strategie di Investimento con i Roboadvisor”, Enrico Malverti, Davide Bulgarelli, Gabriele Villa, ISBN 9788820384883
- 4) Il pensiero flessibile, Leonard Mlodinow, ISBN 9788820063993
- 5) The Financial Consequences of Behavioral Biases: An Analysis of Bias in Corporate Finance and Financial Planning, Imad A. Moosa, Vikash Ramiah, ISBN 9783319693880
- 6) La Scienza della Persuasione: il Nostro Potere di Cambiare gli Altri, Tali Sharot, ISBN 9788807091063
- 7) Data Science, Sinan Ozdemir, ISBN 9788850334193
- 8) Hands-On Machine Learning with Scikit-Learn & TensorFlow, Aurélien Géron ISBN 9781491962299
- 9) Nudge, Improving Decision About Health, Wealth, and Happiness, Richard H. Thaler and Cass R. Sunstein ISBN 9780143115267
- 10) L’Algoritmo Definitivo, Pedro Domingos, ISBN 9788833927060
- 11) Intelligenza Artificiale, Jerry Kaplan, ISBN 9788861052673