Università
Ca'Foscari
Venezia

Corso di Laurea magistrale
in Statistica e Sistemi Informativi Aziendali

Tesi di Laurea

# Tecniche di Deep Learning per l'implementazione di un Trading System

**Relatore**
Ch.mo Prof. Claudio Pizzi

**Laureando**
Ivano Boso
Matricola 809757

**Anno Accademico**
2018 / 2019

# Sommario


**1 Introduzione 6**

**2 Analisi Tecnica 8**

2.1 Analisi Tecnica ed Analisi Fondamentale 8

2.2 Analisi grafica 11

2.2.1 Grafici 11

2.2.2 Trendline 16

2.3 Indicatori di tendenza 18

2.3.1 Medie mobili 18

2.3.2 Bande di Bollinger 20

2.3.3 Average Directional Movement Index (ADX) 23

2.3.4 Aroon Indicator 25

2.4 Indicatori di momento 26

2.4.1 Rate of Change (ROC) 26

2.4.2 Moving Average Convergence/Divergence (MACD) 27

2.4.3 Relative Strength Index (RSI) 29

**3 Reti Neurali Artificiali (ANN) 30**

3.1 Introduzione alle Reti Neurali Artificiali 30

3.2 Architetture delle Reti Neurali Artificiali 31

3.3 Neuroni e funzioni di attivazione 32

3.4 L'algoritmo di apprendimento 34

3.5 Funzioni di costo 35

3.6 Delta Rule e Algoritmo di Backward Propagation of Errors 37

3.7 Generalizzazione del modello 39

3.8 Deep Learning 40

3.8.1 Artificial Intelligence, Machine Learning e Deep Learning 40

3.8.2 I Back End per il Deep Learning e la libreria Keras 42

**4 Il Trading System 44**

4.1 Il Dataset 44

4.2 Definizione del modello 48

4.3 Preparazione dei dati 49

4.3.1 Feature Scaling 50

*4.3.2* Valutazione del modello e scomposizione dei dati 51

4.4 Struttura della rete neurale 52

4.5 Iperparametri 55
4.6 Implementazione del Trading System 56
4.7 Analisi di scenario 58
4.7.1 Scenario 1: Fase laterale seguita da mercato bullish 59
4.7.2 Scenario 1 bis: Fase laterale seguita da mercato bullish, e riduzione del numero di variabili esplicative 61
4.7.3 Scenario 2: Fase laterale seguita da mercato bearish 62

**5 Conclusioni 64**
**6 Appendice 66**
6.1 Ulteriori indicatori di analisi tecnica utilizzati nel trading system 66
6.1.1 Know Sure Thing 66
6.1.2 Money Flow Index 67
6.1.3 Parabolic Stop-and-Reverse (Parabolic SAR) 68
6.1.4 On Balance Volume 69
6.1.5 Stochastic Oscillator / Stochastic Momentum Index - Stochastic Fast %K 70
6.1.6 William's %R 71
6.1.7 Triple Smoothed Exponential Oscillator 71

**7 Bibliografia 74**

# 1 Introduzione

La possibilità di prevedere con anticipo i movimenti dei mercati finanziari è un argomento che ricopre una notevole importanza per i diversi attori che operano nelle borse mondiali, siano essi investitori istituzionali oppure privati.

Nel corso degli anni sono state numerose le discipline che hanno affrontato il tema della previsione con approcci di tipo sia quantitativo che qualitativo. Una delle metodologie che tutt'oggi rimane tra le più diffuse tra gli analisti è sicuramente l'**analisi tecnica** che, come verrà spiegato con maggior dettaglio nel capitolo 2, si avvale di numerosi strumenti di tipo grafico e quantitativo per prevedere l'andamento dei prezzi, attraverso lo studio dei dati di mercato disponibili in un determinato momento (principalmente prezzi e volumi passati).

Se da una parte l'accesso alle informazioni è accessibile in larga scala a tutti gli operatori di mercato, bisogna altresì considerare che, per gli investitori non professionali, l'interpretazione dei segnali generati dagli indicatori di analisi tecnica non è di facile lettura, in quanto implica una profonda conoscenza della materia ed un'esperienza pregressa sugli scenari economici in cui gli indicatori devono essere contestualizzati.

Bisogna inoltre tenere conto che tra clientela professionale e retail esistono importanti differenze di accesso ai mercati in termini di masse, strumenti, commissioni e tempo da dedicare al trading. In particolare, con riferimento all'aspetto commissionale, bisogna considerare che l'incidenza delle commissioni bancarie sugli investimenti dei consumatori è notevolmente più elevata rispetto agli investitori istituzionali. Ultima ma non meno importante è la componente tempo: per un investitore privato è sicuramente diversa la

possibilità di seguire i mercati e gli investimenti in maniera continuativa durante le giornate di borsa rispetto ad un professionista del settore.

In questo scenario si è delineata l'idea alla base di questo elaborato, ovvero la costruzione di un **trading system** che tenga conto delle peculiarità di un investitore non professionale, e che possa elaborare i segnali operativi che derivano dall'analisi tecnica, traducendoli in suggerimenti sulle azioni da compiere sul mercato.

Definiti quindi gli indicatori di analisi tecnica come input del modello, si è proceduto a "decodificare" le informazioni in segnali operativi attraverso le **Reti Neurali Artificiali**, un sistema computazionale ispirato ai processi biologici, che si caratterizzano per la capacità di adattarsi a problemi di natura complessa, le cui componenti sono legate da relazioni non lineari che rendono i segnali difficilmente codificabili.

L'implementazione del modello è avvenuta attraverso strumenti informatici propri del **Deep Learning**, una sottocategoria dell'intelligenza artificiale; tale disciplina permette infatti l'utilizzo di reti neurali complesse, e sta attualmente riscuotendo grande successo in una moltitudine di settori che spaziano dalla traduzione automatica dei testi alla guida autonoma delle automobili.

L'elaborato mira dunque a fornire un punto di raccordo tra le tecniche di previsione classiche, basate sull'utilizzo degli indicatori di analisi tecnica, ed i sistemi evoluti di intelligenza artificiale al fine di supportare il processo decisionale dei consumatori nelle operazioni di trading.

# 2 Analisi Tecnica

## 2.1 Analisi Tecnica ed Analisi Fondamentale

Nel 1957, in una delle primissime fonti, l'analisi tecnica viene descritta da Edwards e Magee come "la scienza che riporta, solitamente in forma grafica, l'effettiva cronologia delle contrattazioni per un titolo od un indice, e poi deriva da quella particolare rappresentazione la probabile evoluzione futura" (Edwards & Magee, 1992). L'analisi tecnica può essere pertanto definita come lo studio dell'andamento dei prezzi attraverso strumenti a carattere prevalentemente grafico, allo scopo di prevedere le tendenze di breve periodo. Quello che maggiormente interessa i *chartists*, così vengono definiti in letteratura gli agenti economici che fanno largo uso dell'analisi tecnica, sono infatti le inversioni di tendenza dei prezzi e la capacità di anticiparle utilizzando le evidenze empiriche della storia passata dei prezzi.

Sempre Edwards e Magee descrivono la filosofia dei chartists sintetizzandola in quattro punti (Edwards & Magee, 1992):

1. il valore di mercato di un'attività è determinato unicamente dall'interazione tra offerta e domanda;
2. l'offerta e la domanda sono governate a ciascun istante da centinaia di fattori, alcuni razionali ed altri irrazionali. Le informazioni, le opinioni, gli stati d'animo, le previsioni (accurate o meno) riguardo al futuro si combinano in maniera varia in questa equazione. Nessun individuo può sperare di comprenderle e pesarle tutte, ma il mercato lo fa automaticamente;

3. a parte fluttuazioni di poco conto, i prezzi si muovono lungo tendenze che persistono per una lunghezza apprezzabile di tempo;
4. i cambiamenti nella tendenza che rappresentano un importante cambiamento nell'equilibrio fra offerta e domanda, qualunque ne sia la causa, sono individuabili prima o poi nelle azioni del mercato stesso.

Se da un lato le analogie tra analisi tecnica ed analisi quantitativa sono numerose, da un altro lato bisogna evidenziare una sostanziale differenza in quanto, l'analisi quantitativa mira a riprodurre il comportamento dei prezzi attraverso l'utilizzo di modelli matematici e statistici complessi. Nell'analisi tecnica invece, all'analisi grafica, vengono affiancati degli indicatori di tendenza e di momento a carattere matematico e non probabilistico.

Nonostante questa natura non probabilistica, le combinazioni di indicatori proprie dell'analisi tecnica sono un punto di partenza per la creazione dei segnali di acquisto o di vendita in molti dei trading system, compresi i più evoluti che si basano su moderni sistemi di intelligenza artificiale.

Frequentemente l'analisi tecnica viene contrapposta all'analisi dei fondamentali (o analisi fondamentale), che utilizza invece metodi qualitativi e quantitativi per determinare il valore intrinseco di un'azienda (fair value). Questo valore si riferisce al valore effettivo di un'azione che, rispetto al valore corrente di mercato, potrebbe essere sopravvalutato (sottovalutato) e quindi soggetto ad un'anomala discesa (risalita). Vengono pertanto prese in considerazione specifiche informazioni, come ad esempio i conti economici e i rendiconti finanziari aziendali, e notizie di più ampio raggio di carattere microeconomico e macroeconomico (politica internazionale, stato di salute dell'economia di un intero paese, livello di competitività di mercato, comunicati stampa). L'analisi fondamentale considera pertanto un quadro più esteso che mira a spiegare l'influenza sui mercati finanziari di determinati fattori estrinseci.

Nella realtà analisi tecnica e analisi fondamentale vengono spesso integrate. Attraverso l'analisi dei fondamentali si individuano le aree strategiche sopravvalutate o sottovalutate, e con l'aiuto dell'analisi tecnica vengono individuate le tendenze di breve periodo, siano esse a rialzo (mercato bullish) o a ribasso (mercato bearish).

Un altro approccio interessante dell'analisi tecnica è lo studio sulla natura umana poiché, se da un lato è vero quanto affermato da Edwards e Magee circa i numerosi fattori razionali ed irrazionali che governano il mercato in un determinato istante, da un altro

lato è possibile rilevare una certa coerenza sui comportamenti degli operatori di mercato in presenza di situazioni simili. Pring (1991) afferma che "Technical analysis is based on the assumption that people will continue to make the same mistakes they have made". E sempre nello stesso testo di Pring vengono definite tre macroaree in cui opera l'analisi tecnica:

1. gli indicatori di Sentimento o di Aspettativa. Sono indicatori che registrano i comportamenti dei diversi operatori di mercato come ad esempio gli investitori comuni, gli investitori professionisti o gli OICR[1]. "Just as the pendulum of a clock continually moves from one extreme to another, so the sentiment indexes (which monitor the emotions of the investors) move from one extreme at a bear market bottom to another at a bull market top". Questi indicatori derivano da indagini effettuate su specifici gruppi di investitori, nelle quali vengono rilevate delle analogie di comportamento nei momenti in cui si registrano i cambi di tendenza nel mercato. Anche la Sentiment Analysis ha subito degli importanti sviluppi nell'ultimo decennio grazie alla disponibilità delle informazioni ed alle crescenti capacità di computo degli elaboratori. Permane il fine ultimo ovvero riuscire a prevedere anche la parte irrazionale del comportamento degli operatori al fine di migliorare la previsione dei cambi di tendenza;
2. gli indicatori di flusso di fondi si basano sulle stime della liquidità di mercato e, in funzione di queste, stimano i movimenti potenziali che potrebbero occorrere nel mercato. Viene analizzata la posizione finanziaria dei diversi operatori di mercato con una particolare attenzione nei confronti dei gestori di fondi comuni di investimento. Rapportando i potenziali compratori e venditori pertanto. si possono anticipare le tendenze future circa gli andamenti dei prezzi. Ad esempio, se sono presenti più compratori che venditori, il prezzo dovrà salire;
3. indicatori di struttura del mercato. Analizzano le caratteristiche degli indicatori, in quanto si basano sul presupposto che il comportamento dei prezzi e degli indicatori cominciano a divergere nelle fasi finali di tendenza, e rappresentano dei segnali di deterioramento nelle fasi di crescita, e viceversa di deterioramento nelle fasi di declino. L'approccio tecnico si basa sulla teoria che la psicologia di massa si muova tra i due estremi di panico e fiducia (e relative accezioni) in quanto

[1] Organismi di investimento collettivo del risparmio. Sono organismi con forma giuridica variabile che investono in strumenti finanziari o altre attività somme di denaro raccolte tra il pubblico di risparmiatori.

"riflesso della psicologia di massa" e dove lo scopo, ancora una volta, consiste nel prevedere l'inversione di queste tendenze.

Un altro principio su cui si fonda infatti l'analisi tecnica, è il perdurare della psicologia di massa poiché l'evoluzione delle oscillazioni emotive si completa nel giro di diversi anni. È superfluo dire che l'individuazione precoce di questo cambio di tendenza rappresenta un ulteriore punto di forza in ottica previsiva.

Per misurare gli aspetti emotivi legati alla fiducia che gli investitori ripongono nel mercato vengono utilizzati quattro importanti indicatori: i prezzi, il tempo, i volumi di scambio e l'ampiezza, quest'ultima intesa come la numerosità dei titoli oggetto di studio che sono presenti nel mercato.

## 2.2 Analisi grafica

L'analisi grafica rappresenta una delle colonne portanti dell'analisi tecnica e si basa sulla ricerca di particolari strutture grafiche, spesso chiamate anche figure. Il giudizio dell'analista assume un ruolo dominante in quanto non esiste un metodo rigoroso che determini il posizionamento e l'inclinazione delle rette che formano le figure. Tale attività viene infatti svolta dagli operatori in funzione di una serie di indicazioni che derivano in parte dalla teoria ed in parte dall'esperienza pregressa.

È bene ricordare che la base dell'intera analisi tecnica si basa sull'esame della successione di minimi e massimi relativi della serie storica dei prezzi e, a tal proposito, la frequenza giornaliera dei prezzi di chiusura è quella che meglio si presta all'analisi dei dati; i principi dell'analisi tecnica possono essere applicati anche a rilevazioni intra-giornaliere, che però hanno un'evoluzione meno deterministica e risultano pertanto meno attendibili.

Di seguito riportiamo alcuni tra i principali strumenti dell'analisi tecnica.

### 2.2.1 Grafici

Il grafico rappresenta lo strumento di primaria importanza per l'analista tecnico. Anche in considerazione di quanto già detto circa la capacità di cogliere i segnali dettata dall'esperienza e dalla conoscenza dei mercati, l'impatto visivo dei grafici rappresenta un punto imprescindibile per l'analisi, nonostante la disponibilità di strumenti a carattere maggiormente quantitativo sia progressivamente aumentata durante gli anni.

Generalmente le serie storiche finanziare vengono rappresentate su piani cartesiani che riportano i prezzi di chiusura nell'asse delle ordinate ed il tempo in quello delle ascisse. A

seconda delle rappresentazioni il grafico può incorporare informazioni circa i volumi scambiati, prezzo minimo, prezzo massimo, prezzo di apertura e prezzo di chiusura.

I prezzi di frequente vengono riportati su scala logaritmica, con il vantaggio di rappresentare graficamente con una buona approssimazione gli scostamenti percentuali dei prezzi.

Le tipologie di grafico che vengono utilizzate per condurre le analisi sono numerose. Di seguito vengono riportati i grafici maggiormente utilizzati:

- grafici lineari (close only chart);
- grafici a candele (candlestick chart);
- grafici a barre (bar chart).

I **Grafici Lineari** vengono costruiti unendo i prezzi di chiusura dei diversi periodi delle giornate di mercato. Rappresentano lo strumento più semplice di lettura ma presentano il limite di non fornire altri importanti elementi quali i prezzi di apertura, minimi, massimi ed i volumi scambiati, presenti invece negli altri grafici. Si adatta tuttavia molto bene alla rappresentazione dei dati nel medio e nel lungo periodo, in quanto rende facilmente identificabili le tendenze nei diversi periodi.

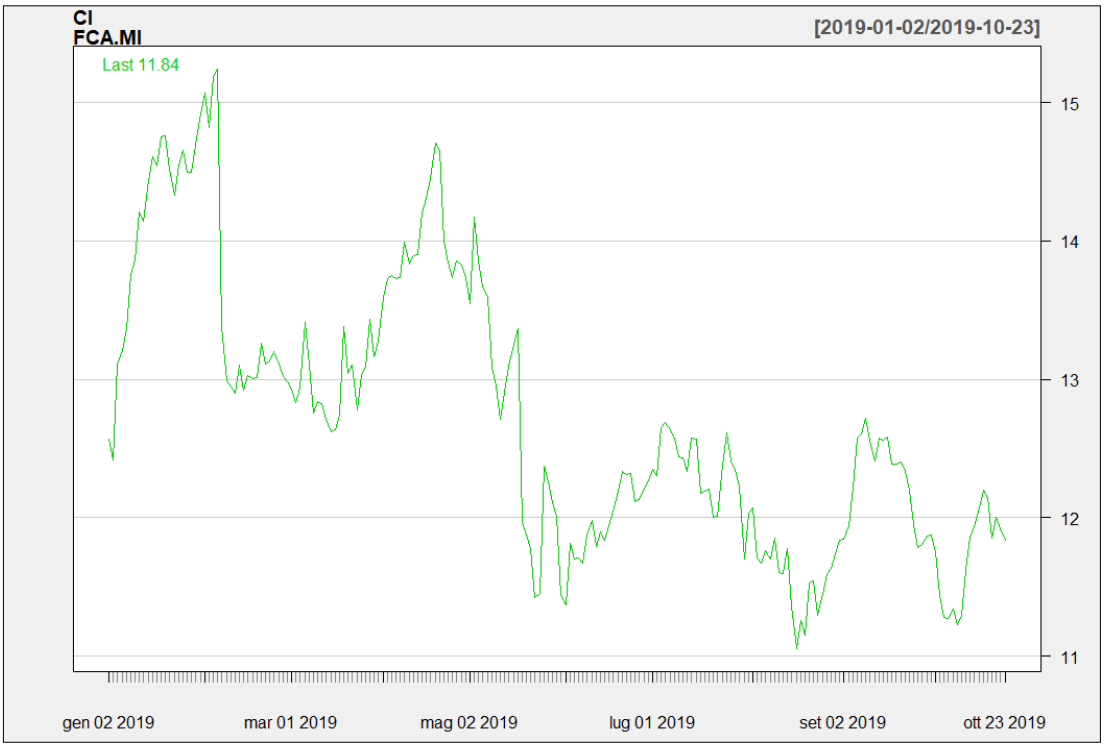


*Figura 2.1: grafico lineare dei prezzi di chiusura sulla borsa di Milano del titolo Fiat.*

I **grafici a candele** ed i **grafici a barre** hanno il vantaggio di rappresentare in maniera completa una situazione di breve periodo dove, per ogni giornata, si possono evincere i prezzi di apertura, di chiusura, massimo, minimo ed il volume scambiato.

L'elemento base del candlestick è appunto la candela, formata da un corpo rettangolare (real body), e da due ombre (shadow) che consistono in due linee verticali al di sotto e al di sopra del corpo. In caso di seduta borsistica positiva, che si realizza quando nella giornata il prezzo di chiusura di un titolo è maggiore del prezzo di apertura, il corpo del rettangolo è solitamente di colore bianco, e gli estremi superiori ed inferiori del rettangolo rappresentano rispettivamente il prezzo di chiusura e di apertura del titolo. Viceversa, in caso di seduta negativa, il corpo del rettangolo è solitamente di colore nero, e la base inferiore indica il prezzo di apertura e quella superiore il prezzo di chiusura. Le ombre raffigurano invece il prezzo massimo (upper shadow) e minimo (lower shadow) di una specifica seduta.

Il grafico a barre ha una rappresentazione analoga a quella del grafico a candele: la rappresentazione avviene mediante delle linee verticali di colori diversi a seconda che le sedute borsistiche siano positive o negative, e dove l'estremo superiore della linea rappresenta il prezzo massimo della seduta, l'estremo inferiore il prezzo minimo, il trattino orizzontale a sinistra il prezzo di apertura, e quello a destra il prezzo di chiusura.

In Figura 2.2 viene riportato un confronto tra la rappresentazione mediante grafico a candele e grafico a barre, della quotazione giornaliera di un titolo in caso di seduta positiva e nel caso di seduta negativa.

Come si può notare la rappresentazione è abbastanza simile, con il vantaggio del grafico a candele di rendere un maggiore impatto visivo delle differenze tra i prezzi di apertura ed i prezzi di chiusura. Inoltre, i grafici a candela, forniscono diverse indicazioni operative in funzione delle figure che si vengono a creare.

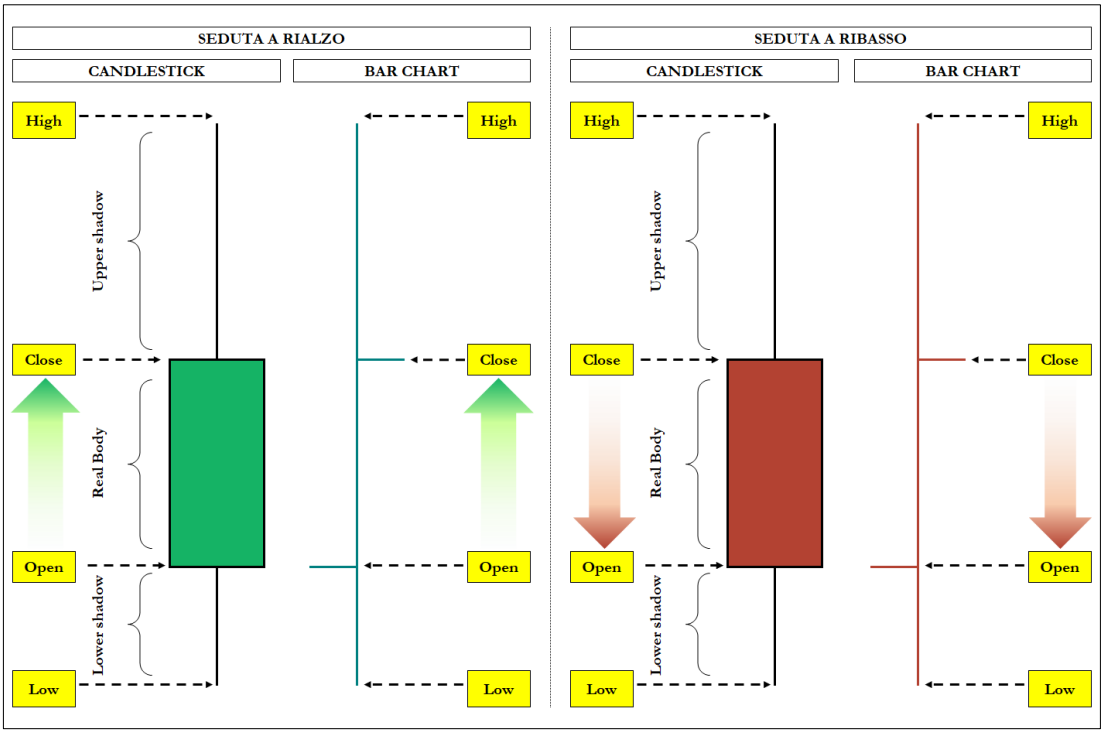


*Figura 2.2: Rappresentazioni grafiche del prezzo giornaliero di un titolo mediante grafico a candele e grafico a barre, nelle ipotesi di seduta a rialzo e di seduta a ribasso.*

In Figura 2.3 vengono riportati il candlestick chart nella parte superiore ed il bar chart nella parte inferiore. Nell'esempio i due grafici rappresentano il medesimo fenomeno: i prezzi del titolo Fiat, quotato nella borsa di Milano, nel periodo tra 30 agosto 2019 ed il 23 ottobre 2019. I grafici sono creati con il software R (2019), dove i colori di default sono il verde, per le sedute borsistiche positive, ed il rosso per quelle negativa. Nella parte inferiore dei grafici sono rappresentati i volumi scambiati nelle singole giornate che mantengono le caratteristiche cromatiche sopra descritte. I grafici sono stati posti a confronto per evidenziare la stretta analogia grafica nella rappresentazione dei dati. In entrambi i casi è bene considerare che i grafici, per quanto siano in grado di fornire una veloce lettura di un titolo, devono essere accompagnati da altri indicatori per fornire informazioni attendibili sui cambi di tendenza.

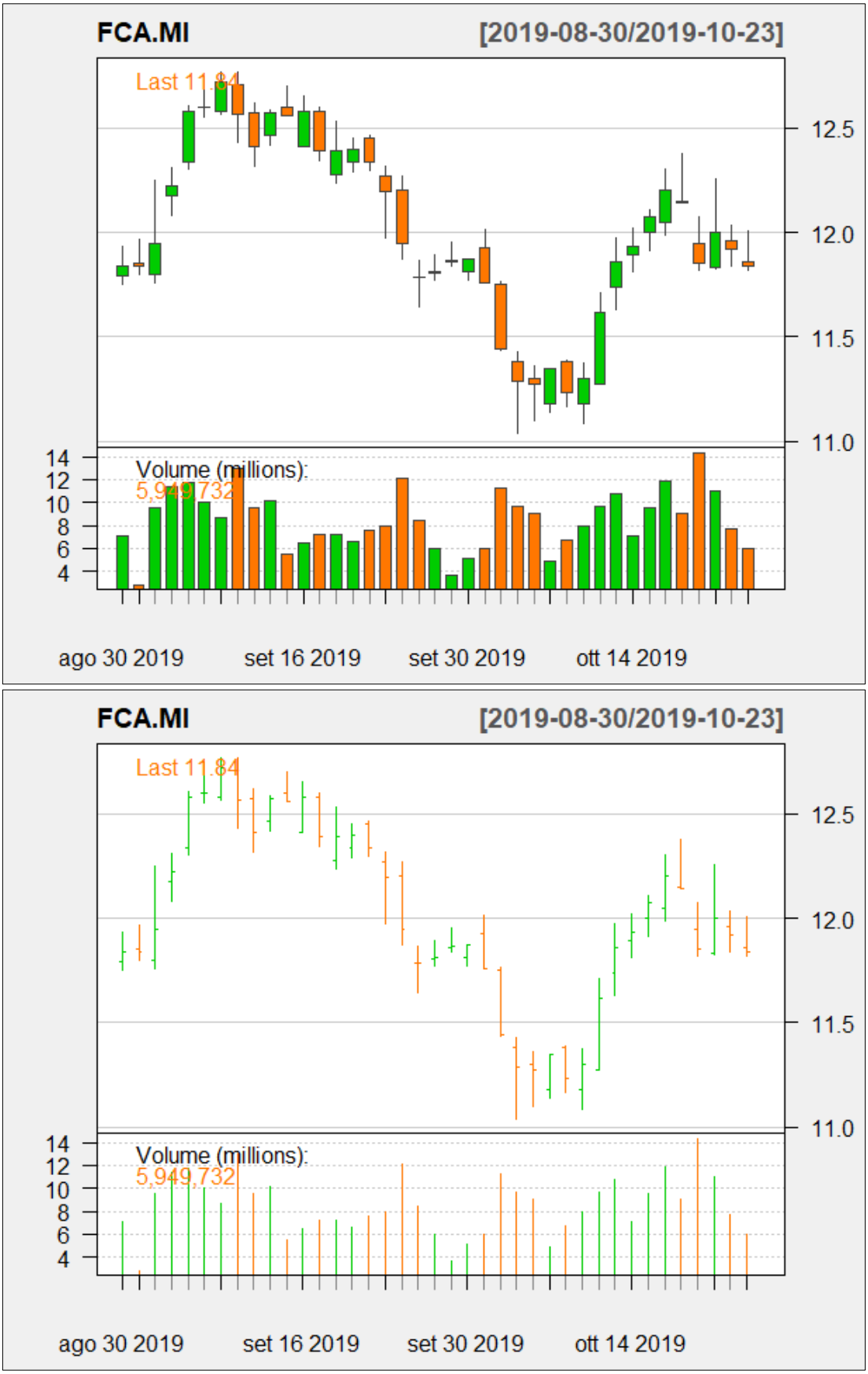


*Figura 2.3: grafico a candele (in alto) e grafico a barre (in basso) dei prezzi e dei volumi del titolo Fiat sulla borsa di Milano.*

### 2.2.2 Trendline

La linea di tendenza, o trendline, è rappresentata da una retta che unisce i punti di minimo locale nelle fasi di crescita, ed i punti di massimo locale nelle fasi di declino.

In funzione della direzione dei prezzi si possono manifestare tre tipi di trend:

- trend rialzista (up-trend): i prezzi presentano un andamento temporale crescente dei massimi e dei minimi;
- trend ribassista (down-trend): i prezzi presentano un andamento temporale decrescente dei massimi e dei minimi.
- trend neutrale (sideways-trend): le quotazioni sono composte da massimi e minimi che mantengono lo stesso livello nel tempo.

I segnali operativi legati alla trendline sono associati al concetto di rottura, ovvero l'attraversamento della linea di tendenza da parte della serie dei prezzi. Infatti, la rottura di una linea ribassista genera un segnale di acquisto mentre, la rottura di un *up-trend* è indice di un segnale di vendita.

La trendline è di veloce comprensione poiché rende di facile lettura la presenza di un particolare andamento in un determinato intervallo di tempo; tuttavia, tra tutti gli strumenti propri dell'analisi tecnica, rappresenta uno degli indicatori in cui l'esperienza dell'operatore gioca un ruolo fondamentale. A differenza di quanto detto in precedenza infatti, le rotture possono anche indicare l'inizio di un'inversione di tendenza oppure la prosecuzione della tendenza in atto ma con un'inclinazione diversa; l'interpretazione su quale dei due fenomeni stia avvenendo viene lasciata all'analista che combina la lettura del grafico ad altri indicatori, alla propria esperienza, coadiuvato a regole deterministiche che possono aiutare nell'identificare false inversioni di tendenza. Per esempio, allo scopo di eludere i segnali deboli, è consigliato prendere in considerazione movimenti di rottura con una distanza minima del 3% tra prezzo e trendline, rafforzati da evidenze di volumi di scambio elevati in corrispondenza della data.

Il concetto di trend è alla base dell'analisi tecnica in quanto uno dei suoi fondamenti è proprio il concetto di *follow trend*, ovvero che i prezzi tendono a mantenere la direzione assunta per un certo periodo di tempo.

Nell'esempio riportato in Figura 2.4 vengono evidenziate due rotture: a rialzo in corrispondenza della freccia rossa, ovvero dell'attraversamento della linea rossa dal basso verso l'alto da parte della serie dei prezzi e, viceversa al ribasso in corrispondenza della

freccia blu. È importante ricordare che sono diverse le caratteristiche che rendono una linea di tendenza più o meno attendibile, prime fra tutte la lunghezza, l'inclinazione ed il numero di volte in cui viene toccata od avvicinata.

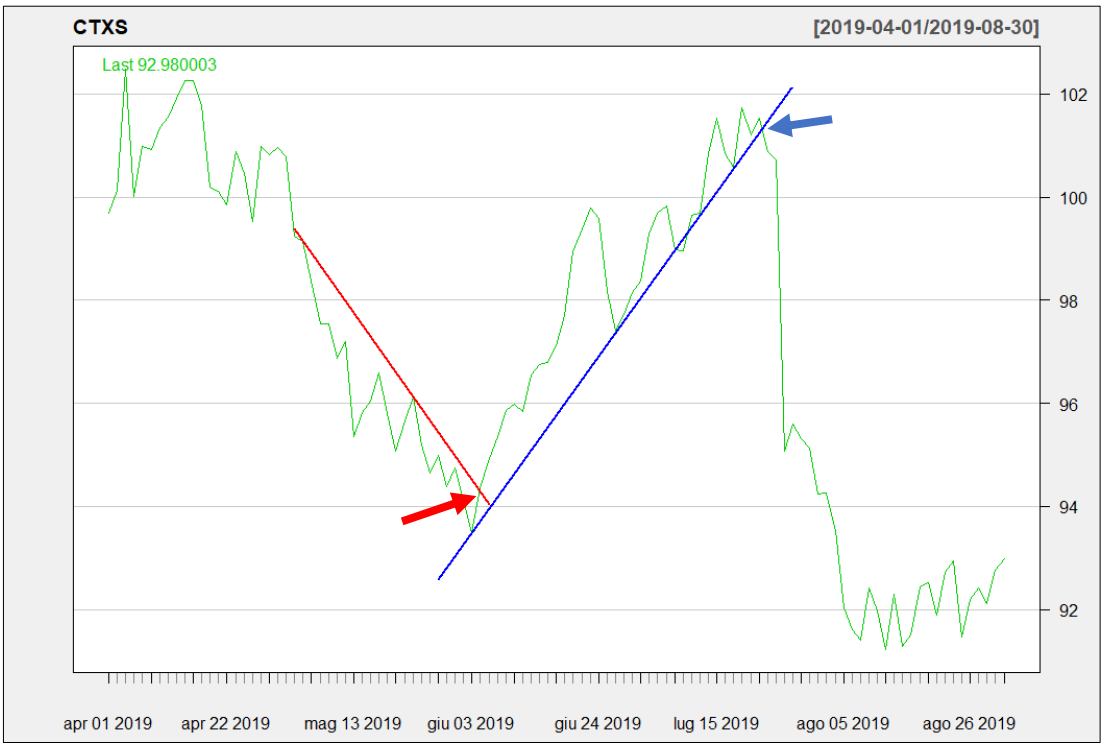


*Figura 2.4: linee di tendenza a ribasso (colore rosso) ed a rialzo (colore blu). Le frecce indicano i punti di rottura in cui si assiste all'inversione del trend.*

La linea di tendenza da origine ad un altro indicatore di fondamentale importanza nell'analisi tecnica, ovvero il canale di tendenza. Il canale di tendenza viene costruito creando inizialmente una retta parallela alla trendline che passi almeno per un punto di massimo relativo nelle fasi bullish, o di minimo relativo nelle fasi bearish.

Con il canale di tendenza vengono definiti altri due importanti indicatori, i supporti (le linee costruite sui minimi) e le resistenze (le linee costruite sui massimi). La rottura da parte della serie dei prezzi può pertanto essere indicativa di un'inversione di tendenza nel caso di un trend rialzista se rompe una linea di supporto, o può indicare un'accelerazione della tendenza in atto se rompe una linea di resistenza (viceversa nel caso di trend ribassista). I supporti e le resistenze sono delle concentrazioni di domanda, o di offerta, capaci di fermare temporaneamente dei movimenti di prezzo. Se il trend viene mantenuto, i prezzi oscillano all'interno del canale, fornendo dei segnali operativi tanto più validi quanto le quotazioni si appoggiano sulle estremità dei supporti e delle resistenze senza oltrepassarli.

## 2.3 Indicatori di tendenza

Gli indicatori di tendenza hanno lo scopo di individuare, come suggerisce il nome stesso, uno specifico trend in atto. Come si vedrà di seguito, l'utilizzo di questi indicatori permette di identificare fasi del mercato *very bearish, bearish, bullish e very bullish*.

### 2.3.1 Medie mobili

Le medie mobili sono uno strumento ampiamente utilizzato, che ha il vantaggio di eliminare, in maniera semplice, le oscillazioni di breve periodo negli andamenti irregolari che si presentano nel caso di serie storiche finanziarie. Diversamente, per ottenere informazioni importanti dal punto di vista statistico, bisognerebbe fare ricorso a polinomi di grado elevato, che presentano una complessità di interpretazione dei risultati decisamente maggiore.

Date le caratteristiche implicite delle serie storiche, le medie mobili utilizzate sono "non centrate", ovvero vengono calcolate a ritroso per ogni istante temporale su un numero prefissato di termini.

Si distinguono tre tipi di medie mobili, che vengono descritte di seguito.

La **Media Mobile Semplice** (Simple Moving Average, SMA), ottenuta calcolando una finestra mobile di *n* quotazioni:

$$SMA(n)_t = \frac{1}{n}\sum_{i=0}^{n-1} P_{t-i} \qquad [2.1]$$

Dove *n* rappresenta il numero di ritardi scelti, e $P_t$ il prezzo del titolo al tempo *t*. Una media mobile è detta più o meno veloce a seconda del numero *n* di ritardi che si va ad impostare. Tanto più grande è *n* tanto più lenta sarà la media mobile; pertanto il cambio di direzione rispetto alla serie originale dei prezzi subirà un ritardo funzionale al parametro *n*.

La **Media Mobile Ponderata** (Weighted Moving Average, WMA), è anch'essa una media degli ultimi *n* prezzi, ma pesati affinché le osservazioni più recenti diano un contributo maggiore alla determinazione della media rispetto ai valori meno recenti. La media mobile ponderata maggiormente utilizzata è quella riportata nella *[2.2]* ponendo $\alpha_n = n$, ovvero,

prendendo in esame una media mobile ponderata a *n* giorni, il valore più recente viene moltiplicato per *n*, quello che lo precede per *n-1*, via via fino al valore meno recente che viene moltiplicato per uno. La serie di questi valori viene così sommata e quindi divisa per la somma dei fattori di ponderazione. In questo modo viene fornita una media in cui il valore più recente pesa *n* volte rispetto al valore meno recente.

$$WMA(n)_t = \frac{\sum_{i=0}^{n-1} P_{t-i} \cdot \alpha_{n-i}}{\sum_{i=0}^{n-1} \alpha_{n-i}} \qquad con\ \alpha_n > \alpha_{n-1} > \cdots > \alpha_1 \qquad [2.2]$$

La **Media Mobile Esponenziale (Exponential Moving Average, EMA)** è la media mobile maggiormente utilizzata dai trader; diversamente dalle due medie mobili appena descritte, la media mobile esponenziale viene ottenuta in maniera ricorsiva, e come per la media mobile ponderata attribuisce una maggiore importanza al dato più recente.

$$EMA_t = (1-\alpha)EMA_{t-1} + \alpha P_t \quad con\ 0 \leq \alpha \leq 1,\ EMA_0 = P_0 \qquad [2.3]$$

Il valore di α viene deciso in maniera funzionale all'importanza che dev'essere data al prezzo più recente; un valore di α prossimo ad uno darà origine ad una serie di medie esponenziali ponderate più fedele alla serie dei prezzi originale, mentre un valore che tende allo zero genererà una serie più "liscia"; in corrispondenza di α=0 si otterrà una linea parallela all'asse delle ascisse in corrispondenza del primo prezzo $P_0$. In questo contesto il parametro α è definito anche parametro di lisciamento (smoothing). Analogamente a quanto avviene per la media mobile semplice, è possibile fissare il parametro α in funzione del numero delle osservazioni *n* che si vuole utilizzare attraverso la formula:

$$\alpha = \frac{2}{n+2} \qquad [2.4]$$

Le medie mobili sono degli strumenti che riflettono la tendenza in atto ed hanno la funzione di anticiparla. In particolar modo i segnali più attendibili arrivano dalle

intersezioni di medie mobili più o meno veloci con la serie dei prezzi. Ad esempio, sono oggetto di particolare attenzione da parte degli analisti, i segnali che derivano dalla costruzione di una media mobile veloce a quattro elementi ($n$ = 4), con una media mobile più lenta a nove elementi ($n$ = 9). Il confronto tra le due medie mobili e la serie dei prezzi determina gli attributi di mercato riportati in Tabella 2.1 (Gallo & Pacini, 2015).

| | |
|---|---|
| mercato bullish | $MA(4)_t > MA(9)_t$ |
| mercato very bullish | $P_t > MA(4)_t > MA(9)_t$<br>$MA(4)_t > MA(4)_{t-1}$<br>$MA(9)_t > MA(9)_{t-1}$ |
| mercato bearish | $MA(4)_t < MA(9)_t$ |
| mercato very bearish | $P_t < MA(4)_t < MA(9)_t$<br>$MA(4)_t < MA(4)_{t-1}$<br>$MA(9)_t < MA(9)_{t-1}$ |

*Tabella 2.1: Le tendenze di mercato individuate dal confronto tra i prezzi e le medie mobili a n=4, ed n =9 elementi.*

I tagli dal basso verso l'alto da parte delle quotazioni di un titolo nei confronti delle medie mobili sono una conferma di un trend rialzista, in quanto evidenziano uno scenario in cui i prezzi si sono mantenuti sempre al di sotto della media degli $n$ giorni precedenti fino al momento dell'intersezione; una volta tagliata la serie è verosimile che i prezzi si mantengano al di sopra per un certo periodo. Questa evidenza può pertanto essere considerata un segnale di acquisto.

È importante ricordare che il significato dei segnali di intersezione è legato anche al numero di ritardi coperti dalla media mobile; maggiore è l'arco temporale $n$ che identifica la media mobile, maggiore è l'attendibilità di un segnale di intersezione.

La reazione delle medie mobili, più lenta rispetto alla serie dei prezzi, le rende allo stesso tempo un'area di resistenza o di supporto. Un declino dei prezzi in una fase di crescita infatti, trova nella media mobile un supporto viceversa, un rialzo in una fase di decrescita, identifica nella media mobile una resistenza.

### 2.3.2 Bande di Bollinger

Le bande di Bollinger rappresentano uno degli strumenti più importanti utilizzato dagli analisti tecnici; sono annoverabili agli indicatori di tendenza e forniscono al contempo importanti indicazioni circa la volatilità del mercato, scopo per il quale sono state create.

Prendono il nome da John Bollinger, un analista finanziario americano, che le ha ideate e formalizzate ispirandosi alla teoria di John Hurst (Bollinger, 2002).

Questo particolare strumento si basa sul concetto di "sacche di negoziazione", altrimenti conosciute come "envelopes"; le evidenze empiriche suggeriscono infatti che il prezzo di un titolo gravita tipicamente all'interno di queste "sacche", composte da due bande di trading e costruite a loro volta attorno ad una banda mediana.

La **banda mediana** corrisponde ad una media mobile semplice tipicamente con *n* posto uguale a 20 elementi.

$$SMA(n)_t = \frac{1}{n}\sum_{i=0}^{n-1} P_{t-i} \qquad [2.5]$$

La **banda inferiore $\boldsymbol{BBL_t}$** e la **banda superiore $\boldsymbol{BBU_t}$** sono costruite, rispettivamente, sottraendo e sommando alla banda mediana, la deviazione standard mobile $\mathrm{MSTD(n)}_t$ premoltiplicata per un parametro *m,* solitamente posto uguale a due, e *n* uguale al numero di ritardi scelti per determinare la banda mediana.

$$BBL_t = SMA(n)_t - m \cdot MSTD(n)_t \qquad [2.6]$$

$$BBU_t = SMA(n)_t + m \cdot MSTD(n)_t \qquad [2.7]$$

dove

$$MSTD(n)_t = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(P_{t-i} - SMA(n)_t)^2} \qquad [2.8]$$

L'ampiezza delle bande varia in funzione della deviazione standard mobile che, per questo indicatore, viene utilizzata come proxy della volatilità del titolo: in fasi di alta volatilità le bande si espandono, al contrario in momenti di bassa volatilità le bande si contraggono (squeeze). Se la serie dei prezzi oscilla all'interno delle due bande, gli incroci con la banda mediana sono ritenuti fisiologici; l'interpretazione di un segnale di cambiamento di tendenza avviene invece se, diversamente, le quotazioni si avvicinano ad una delle due bande, superiore od inferiore. In analogia con quanto già detto circa le aree di supporto e

resistenza infatti, le bande di Bollinger possono essere considerate un canale di tendenza dinamico che fornisce importanti indicazioni operative: se la quotazione si avvicina banda superiore il mercato può essere considerato in una fase di ipercomprato, facendo supporre un'inversione a ribasso della tendenza prossima a manifestarsi. Viceversa, la prossimità delle quotazioni alla banda inferiore indica un'inversione a rialzo di un trend ribassista.

Rispetto all'altezza del canale formato dalle bande di Bollinger, le evidenze empiriche suggeriscono come segnale di inversione, l'intersezione della serie dei prezzi in prossimità del primo o dell'ultimo 30% di ampiezza del canale, come riportato di seguito in Tabella 2.2 (Gallo & Pacini, 2015).

| | |
|---|---|
| mercato bullish | $\frac{P_t - BBL_t}{BBU_t - BBL_t} < 0{,}3$ |
| mercato bearish | $\frac{P_t - BBL_t}{BBU_t - BBL_t} > 0{,}7$ |

*Tabella 2.2: Le tendenze di mercato individuate dalle bande di Bollinger rispetto alla serie dei prezzi.*

È bene sottolineare che un caso particolare si manifesta in presenza di uno squeeze; in questo caso è più probabile una conferma del trend in essere piuttosto che un cambio di tendenza.

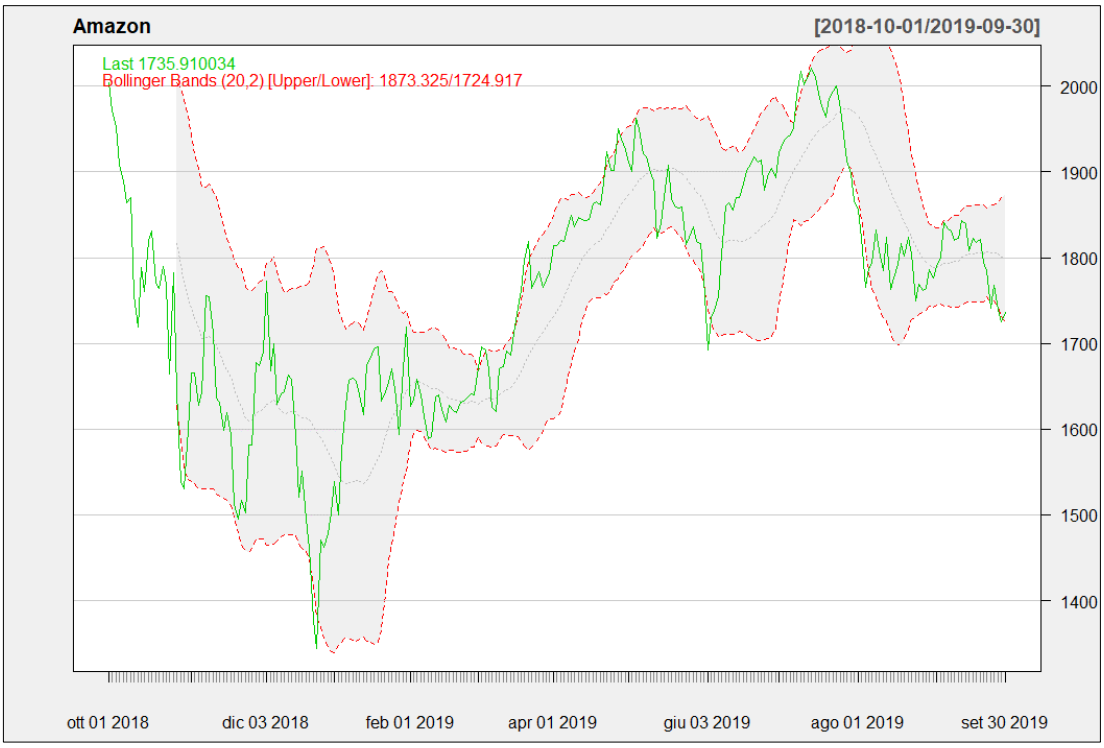


*Figura 2.5: bande di Bollinger su grafico lineare dei prezzi di chiusura del titolo Amazon.*

A conferma dei segnali operativi individuati per mezzo delle bande di Bollinger, la maggior parte degli analisti tecnici utilizza degli ulteriori strumenti, come l'indicatore di forza relativa (RSI) che vedremo di seguito, allo scopo di individuare eventuali divergenze rispetto ai segnali operativi.

### 2.3.3 Average Directional Movement Index (ADX)

L'Average Directional Movement Index è un indicatore di tendenza utilizzato per determinare la forza di un trend, con lo scopo di supportare l'analista nella stima delle diverse fasi della tendenza in atto. Poiché l'indice si riferisce unicamente alla forza del momento del prezzo, a supporto dell'indicatore intervengono altri due indici utilizzati per fornire informazioni sulla direzione del trend: il Plus Directional Indicator (+*DI*) [2.13] ed il Minus Directional Indicator (-*DI*) [2.14].

Definiti quindi $H_t$e $L_t$, rispettivamente il massimo ed il minimo prezzo realizzati in una generica giornata *t*, è necessario preventivamente calcolare il Plus Directional Movement (+*DM*) ed il Minus Directional Movement (-*DM*) come da equazioni [2.9] e [2.10].

$$+DM_t = \begin{cases} (H_t - H_{t-1}) & se\ (H_t - H_{t-1}) > (L_{t-1} - L_t) \wedge (H_t - H_{t-1}) > 0 \\ 0 & altrimenti \end{cases} \quad [2.9]$$

$$-DM_t = \begin{cases} (L_{t-1} - L_t) & se\ (L_{t-1} - L_t) > (H_t - H_{t-1}) \wedge (L_{t-1} - L_t) > 0 \\ 0 & altrimenti \end{cases} \quad [2.10]$$

Per il calcolo dell'ADX sono richiesti inoltre due ulteriori indicatori, il True Range (TR) [2.11] ed una media mobile ad *n* elementi del TR definita Average True Range (ATR) [2.12], con *n* posto solitamente uguale a 14.

$$TR_t = [(H_t - L_t), |H_t - C_{t-1}|, |L_t - C_{t-1}|] \quad [2.11]$$

$$ATR_t = EMA_n(TR_t) \quad [2.12]$$

Dividendo i DM per la ATR, è quindi possibile ottenere i Directional Indicator (DI).

$$+DI_t = \frac{+DM_t}{+ATR_t} \qquad [2.13]$$

$$-DI_t = \frac{-DM_t}{+ATR_t} \qquad [2.14]$$

È quindi possibile calcolare l'Average Directional Movement (ADX) [2.15]

$$ADX_t = \frac{1}{k}\sum_{i=0}^{k-1}\frac{DI_{t-i}^{+} - DI_{t-i}^{-}}{DI_{t-i}^{+} + DI_{t-i}^{-}} \times 100 \qquad [2.15]$$

Oltre all'indicazione di forza della tendenza, i cui livelli sono riassunti nella Tabella 2.3, l'ADX fornisce alcuni importanti segnali operativi definiti da tre regole di seguito elencate:

- *Crossover Rule*: indica di assumere una posizione lunga qualora il *+DI* incroci dal basso verso l'altro il *-DI*. Viceversa, al verificarsi di un incrocio dall'altro verso il basso del *+DI* rispetto al *-DI* il segnale è di assumere una posizione corta.
- *Extreme Point Rule*: individua il momento migliore per eseguire un'operazione di acquisto o di vendita, segnalata a sua volta dalla regola di cui al punto precedente. In caso di posizione lunga la regola prevede di identificare il prezzo massimo della giornata in cui la Crossover Rule ha fornito il segnale di acquisto, e di procedere all'acquisto solamente quando il prezzo supera tale livello. Al contrario, se il +DI scende al di sotto del -DI, il punto estremo è il prezzo minimo del giorno in cui le linee si incontrano, e rappresenta la soglia sotto il quale è consigliato assumere una posizione corta.
- *Turning Point Rule*: in presenza di un trend delineato identifica una possibile inversione di tendenza nel mercato qualora la curva ADX scenda al di sotto degli indicatori DI.

| ADX | Forza della Tendenza |
|---|---|
| 0-25 | Debole |
| 25-50 | Forte |
| 50-75 | Molto Forte |
| 75-100 | Estremamente Forte |

*Tabella 2.3: ADX, classificazione dei livelli di forza della tendenza.*

### 2.3.4 Aroon Indicator

L'indicatore Aroon (Ao) viene utilizzato per identificare dei cambi di tendenza della serie dei prezzi. Viene costruito attraverso il calcolo degli indici AroonUp [2.16] e AroonDown [2.17], che forniscono una misura del tempo impiegato dal prezzo per raggiungere i valori massimi e minimi in un periodo di giorni pari ad $N$ (solitamente con $N$=20).

$$AroonUp = 100 \times \frac{N - n_M}{M} \qquad [2.16]$$

$$AroonDown = 100 \times \frac{N - n_m}{M} \qquad [2.17]$$

con $n_M$ e $n_m$ che rappresentano il numero di giorni impiegati per raggiungere il massimo ed il minimo di periodo.

L'Aroon Indicator [2.18] viene quindi calcolato come differenza tra l'AroonUp e l'AroonDown ed assume valore che oscillano all'interno dell'intervallo (-100, 100).

$$Ao = AroonUp - AroonDown \qquad [2.18]$$

Segnali operativi:

- un valore dell'Ao superiore ai 70 punti è segnale di una forte tendenza rialzista in corso, mentre inferiore a -70 indica un forte trend ribassista;
- il passaggio da valori negativi a positivi dell'Ao è un segnale bullish viceversa, da valori positivi a negativi è un segnale bearish.

## 2.4 Indicatori di momento

A supporto degli indicatori di tendenza si utilizzano gli indicatori di momento, che hanno la caratteristica di fornire indicazioni circa l'attendibilità degli indicatori di tendenza.

### 2.4.1 Rate of Change (ROC)

Il Rate of Change (ROC) è un indicatore della forza del mercato poiché misura il tasso di variazione dei prezzi, tra due istanti di tempo, rispetto ai loro livelli effettivi. Concretamente rappresenta la differenza percentuale di due prezzi in due istanti temporali. In formula:

$$ROC_t(n) = \frac{P_t - P_{t-n}}{P_{t-n}} \approx ln(P_t) - \ln(P_{t-n}) \qquad [2.19]$$

Il ROC oscilla attorno alla linea dello zero (detta anche "linea di equilibrio") a seconda che l'indice si muova in territorio positivo o negativo.

I campi di maggiore utilizzo dell'indice sono legati all'identificazione della spinta direzionale di una determinata tendenza, all'individuazione di situazioni di ipercomprato e ipervenduto, ed alla ricerca delle divergenze.

Nel primo caso, la spinta di una tendenza in crescita può essere confermata da ROC positivi e progressivamente crescenti ovvero, con *k>0,* dev'essere $ROC_{t+k} > ROC_{t+k-1} > ROC_{t+k-2} > \cdots > ROC_t$. Viceversa, una tendenza decrescente viene confermata da ROC negativi e progressivamente decrescenti.

Per determinare le fasi di ipercomprato ed ipervenduto, che rappresentano rispettivamente i segnali di vendita e di acquisto, i lag più utilizzati sono invece i 12 giorni per il breve termine ed i 25 giorni per il medio/lungo termine. Per identificare queste fasi non esistono tuttavia dei livelli standard; questi livelli, solitamente, vengono scelti tracciando due linee sul grafico: la linea che attraversa i due terzi dei minimi rappresenta una soglia di ipervenduto, mentre la linea che attraversa i due terzi dei massimi identifica una soglia di ipercomprato.

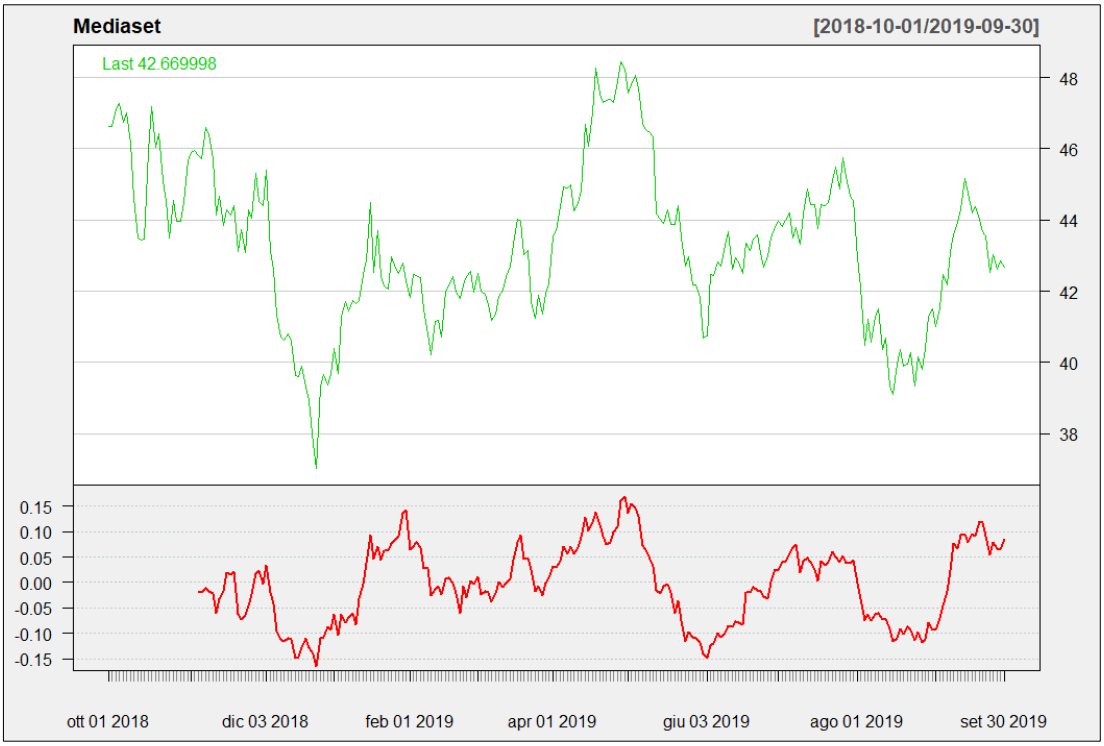


*Figura 2.6: Quotazioni del titolo Mediaset. Nella parte inferiore del grafico è riportata l'oscillazione del ROC con n=25.*

Infine, per determinare delle possibili inversioni di tendenza, si ricorre al confronto tra il ROC ed il movimento del prezzo del titolo; eventuali divergenze tra i due indicatori sono indicative di un imminente cambio di tendenza: si parla di divergenza positiva qualora le quotazioni mostrino un trend decrescente e l'indice ROC sia crescente, viceversa, si parla di divergenza negativa, qualora il ROC sia decrescente e l'andamento dei prezzi sia in ascesa. Il primo caso è un segnale di acquisto mentre il secondo è un segnale di vendita.

### 2.4.2 Moving Average Convergence/Divergence (MACD)

Il Moving Average Convergence/Divergence è un indice sviluppato da Gerald Appel ed è costruito attraverso la combinazione di medie mobili esponenziali.

La rappresentazione grafica dell'indice avviene attraverso due linee, così costruite:

- Differential Line = MACD = $EMA_{12}(P_t) - EMA_{26}(P_t)$
- Signal Line = $EMA_9(MACD)$

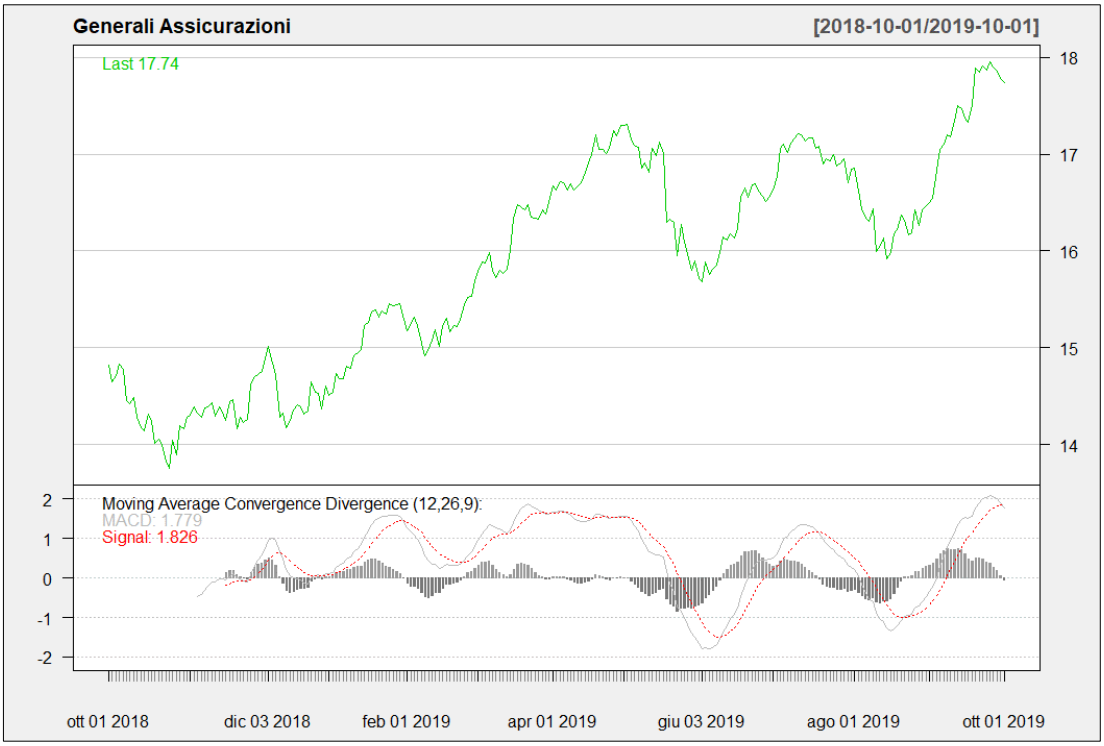


*Figura 2.7: Quotazioni del titolo Generali Assicurazioni. Nella parte inferiore del grafico sono riportate la signal line (colore rosso) e la differential line (colore grigio).*

La differential line è utilizzata per identificare le variazioni di tendenza in atto, che si manifestano attraverso il cambio del segno dell'indice. La media a 12 ritardi infatti, che è più veloce rispetto a quella a 26 ritardi, rileva con anticipo le variazioni di prezzo. Pertanto, un'intersezione della prima ($EMA_{12}$) sulla seconda ($EMA_{26}$) dal basso verso l'altro viene interpretato come segnale di acquisto, mentre un incrocio dall'alto verso il basso identifica un segnale di vendita. La signal line invece svolge una funzione anticipatrice dell'inversione di tendenza da parte del MACD. Si parla di crossover bullish, che rappresenta un segnale di acquisto, se la MACD incrocia la signal line dal basso verso l'alto; si parla invece di crossover bearish se, al contrario, la differential line incrocia la signal line dall'alto verso il basso.

L'attendibilità del segnale è letta anche in funzione di quanto velocemente la media lenta si orienta alla stessa maniera di quella veloce. A questo proposito, un altro elemento di notevole importanza nell'interpretazione dei segnali operativi generati dal MACD, è la convergenza e la divergenza tra la differential e la signal. In questo contesto l'aiuto più importante è rappresentato dall'istogramma riportato in Figura 2.7. L'istogramma infatti, rappresenta la distanza tra le due linee; se il trend a lungo termine è in crescita, allora una divergenza negativa con crossover ribassista viene interpretata come un segnale di

inversione. Diversamente, nella situazione di un trend a ribasso, gli analisti identificano una inversione qualora si manifesti una divergenza positiva con crossover rialzista.

### 2.4.3 Relative Strength Index (RSI)

Il Relative Strength Index (RSI), conosciuto anche come indice di forza relativa, è un oscillatore[2] ideato da John Welles Wilder (Wilder, 1978). È formulato come:

$$RSI_t(n) = 100 \cdot \frac{\sum_{\tau=0}^{n-1} \Delta^+ P_{t-\tau}}{\sum_{\tau=0}^{n-1} |\Delta P_{t-\tau}|} \qquad [2.20]$$

Dove $\Delta$ sono le differenze di prezzo, e $\Delta^+$sono le differenze di prezzo positive. Concretamente, per ogni istante temporale *t*, e dato un numero di lag *n*, all'interno dell'intervallo (*t-n+1, t*) viene rapportata la somma delle variazioni positive di prezzo alla somma, in valore assoluto, di tutte le variazioni.

Presenta pertanto una banda d'oscillazione costante che può presentare valori compresi tra zero, qualora non ci siano variazioni positive all'interno dell'intervallo, e cento, qualora tutte le variazioni abbiano segno positivo. Essendo un indice relativo, permette una comparazione dei valori con alcuni livelli costanti prestabiliti: si identifica una zona di ipervenduto nei casi in cui l'indice scenda sotto una soglia che solitamente viene fissata a venti punti percentuali; al contrario, la soglia per l'ipercomprato, è fissata ad ottanta punti percentuali. A differenza del ROC, il relative strength index risulta meno sensibile a situazioni di particolare volatilità in cui, altri indicatori di momento come il ROC, tendono a creare falsi segnali.

[2] Indici i cui valori sono ricompresi in un range definito di dati.

# 3 Reti Neurali Artificiali (ANN)

## 3.1 Introduzione alle Reti Neurali Artificiali

Le reti neurali artificiali sono dei sistemi di elaborazione dei dati che mirano a replicare in maniera matematica la struttura biologica del sistema nervoso. Tale struttura è costituita da numerosi neuroni, ciascuno interconnesso con migliaia di altri neuroni, e che a loro volta possono ricevere informazioni dall'ambiente esterno o da altre unità interconnesse; si distinguono pertanto le unità di ingresso (input), le unità di uscita (output) e le unità nascoste (hidden). Come il cervello umano combina una serie di input derivanti dagli organi sensoriali, e fornisce degli output che possono spaziare dall'esecuzione di un movimento alla formulazione di un pensiero, così le reti neurali artificiali combinano una serie di input per fornire uno o più output.

Le reti neurali sono diventate degli strumenti molto importanti nella modellazione delle serie storiche sia per la loro capacità di cogliere relazioni non lineari nei dati, sia per rivelarsi particolarmente efficaci nelle elaborazioni in parallelo, ovvero nella capacità di riuscire a trattare molti dati contemporaneamente.

Un concetto spesso accostato alle reti neurali è quello di "black box" poiché, a differenza di un sistema algoritmico dove si può esaminare nel dettaglio il processo che dall'input genera l'output, le ANN sono in grado di generare un risultato valido, o comunque con una alta probabilità di essere accettabile, ma non è possibile definire in maniera analitica il percorso che ha portato alla generazione di tale risultato. A questo proposito è interessante uno studio condotto da Marco Tulio Ribeiro (Ribeiro, Guestrin, & Singh, 2016) dove, in un esperimento di riconoscimento immagini, una ANN multilivello è stata

utilizzata per distinguere due differenti tipologie di immagini, quelle che identificano un lupo e quelle che identificano un husky. Dopo la fase di training della rete, un'immagine veniva identificata correttamente nel 90% dei casi ma non era chiaro il meccanismo alla base di tale risultato. Affinando i test, si è successivamente accertato che la rete neurale trovava gli elementi peculiari nello sfondo dell'immagine in quanto, nella maggior parte delle fotografie utilizzate nella fase di apprendimento, lo sfondo posto dietro al lupo era un paesaggio innevato.

I campi di applicazione delle reti neurali artificiali sono i più svariati e l'esempio appena proposto, che può essere ricondotto al campo della classificazione di oggetti, ne è una conferma. Attraverso l'utilizzo di reti neurali multistrato negli ultimi anni si è così evoluto un approccio denominato "Deep Learning", che ha permesso di sviluppare dei sistemi che vanno dal riconoscimento vocale a complessi software OICR per il riconoscimento della scrittura a mano. Riportando la metodologia ad un campo più ristretto come le aree di applicazione finanziaria, gli ambiti sono comunque numerosi e si spazia, solo per citarne alcuni. dalle strategie di trading alla gestione dei portafogli.

## 3.2 Architetture delle Reti Neurali Artificiali

Le architetture delle reti neurali artificiali ripropongono in maniera semplificata la struttura del cervello umano formata, dai neuroni, preposti all'elaborazione delle informazioni, e dalle sinapsi, che hanno la funzione di trasmettere le informazioni.

Le ANN si distinguono in base al numero di strati utilizzati, e alle modalità di interconnessione tra gli stati ed all'interno di un singolo strato.

Di norma l'organizzazione della ANN è su più strati, distinti in:

- strato di input, i cui neuroni ricevono le informazioni dall'ambiente;
- strati nascosti, se presenti sono composti da neuroni interconnessi ai neuroni di input e di output, o a neuroni appartenenti ad altri strati nascosti;
- strato di output, dove sono presenti uno o più neuroni che forniscono le risposte.

Nell'esempio riportato in Figura 3.1è riportata una rete definita feedforward a più strati. In questa architettura le informazioni dai neuroni di input vengono passate, nell'ordine, ai nodi nascosti e quindi ai nodi di output. Ogni strato ha connessioni entranti dal precedente strato e uscenti in quello successivo, quindi la propagazione del segnale avviene in avanti, in modalità aciclica e senza connessioni trasversali.

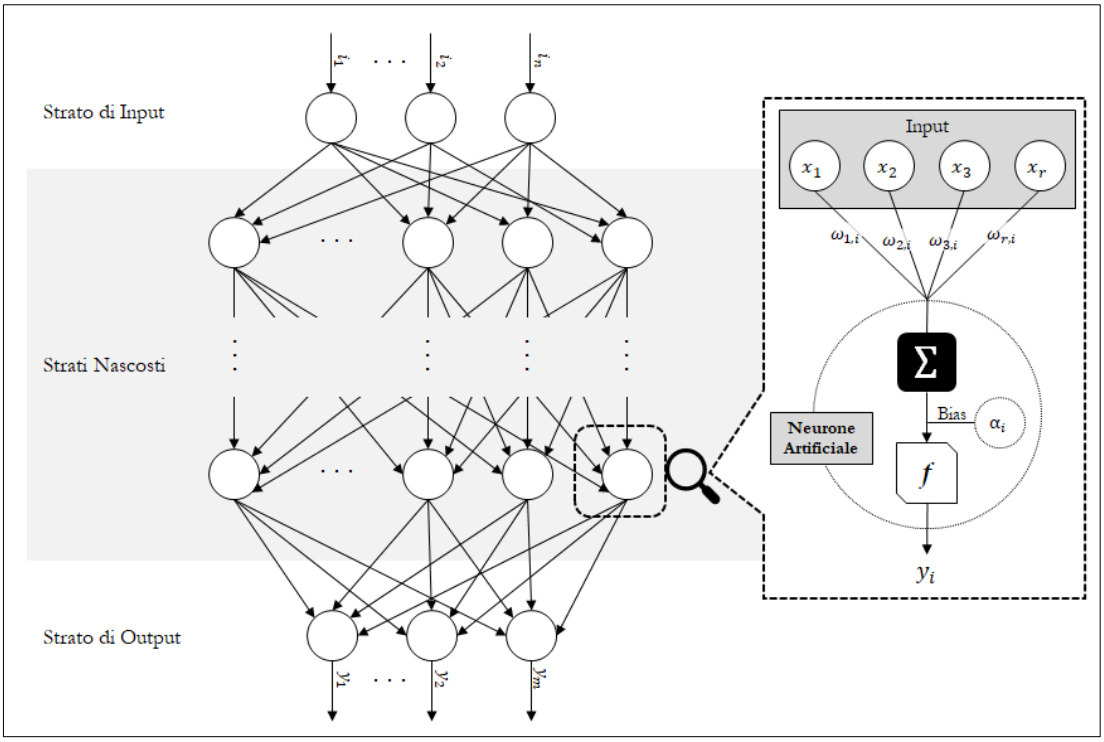


*Figura 3.1: Struttura di una rete neurale artificiale mulistrato.*

L'architettura della ANN influenza la capacità previsiva del processo e dev'essere scelta accuratamente in funzione del trade off tra basso potenziale di apprendimento della rete (se si utilizzano pochi neuroni) e overfitting dei dati (se si utilizzano al contrario molti neuroni).

## 3.3 Neuroni e funzioni di attivazione

Come già anticipato il neurone, che rappresenta l'unità di calcolo fondamentale delle ANN, è formato da un insieme di connessioni (sinapsi); ciascuna sinapsi è caratterizzata da un peso (efficacia sinaptica) che può essere negativo o positivo.

Il neurone artificiale può essere rappresentato matematicamente dalle seguenti equazioni:

$$u_k = \sum_{j=1}^{m} w_{kj} \cdot x_j \qquad [3.1]$$

$$y_k = f(u_k + a_k) \qquad [3.2]$$

dove:

- $x_i$ sono gli input del neurone k;

- $w_i$ sono i pesi sinaptici del neurone k;
- $u_k$ è la combinazione lineare di tutti gli input (pesati) del neurone k;
- $a_k$ è il bias del neurone k;
- $f(\cdot)$ è la funzione di attivazione (altresì detta di trasferimento);
- $y_k$ è l'output generato dal neurone k.

Gli input $x_i$ possono essere gli input reali della ANN, oppure dei valori intermediati provenienti dagli strati nascosti precedenti, e vengono combinati linearmente da un sommatore che li addiziona, rapportandoli ai rispettivi pesi, e li trasferisce ad una funzione di attivazione $f(\cdot)$ che limita l'ampiezza dell'output del neurone. Il neurone artificiale utilizza infatti la combinazione lineare degli input in ingresso $u_k$, come argomento della *funzione di attivazione* $f(\cdot)$, che restituisce a sua volta come output $y_k$ con valori solitamente compresi tra 0 e 1, oppure tra -1 e 1. È inoltre presente un ulteriore elemento, conosciuto come valore soglia, che attiva il neurone qualora sia superato un determinato livello; il valore soglia concretamente è rappresentato da una costante indipendente, detta *bias*, che viene impostata durante il training. Il bias può essere altresì introdotto mediante un ingresso costante ed uguale ad 1, che viene modulato da un peso $w_0$, ed il cui effetto è quello di spostare la soglia di attivazione in funzione dei segnali di input.

La funzione di attivazione è un elemento di fondamentale importanza in quanto rende le ANN estremamente flessibili, aumentandone l'efficacia nella stima di relazioni non lineari e complesse nei dati. Tali funzioni vengono scelte a seconda della struttura utilizzata e, all'interno della stessa ANN, possono essere adottate funzioni di attivazione diverse per ogni strato. Una tra le funzioni maggiormente utilizzate nell'analisi delle serie storiche finanziarie è la funzione logistica, definita come da equazione [3.3]:

$$f_j(z) = \frac{exp(z)}{1 + exp(z)} \qquad [3.3]$$

L'ampio utilizzo di questa funzione deriva dal vantaggio di assumere valori compresi tra zero ed uno, di essere continua e differenziabile.

Specificatamente allo strato di output invece, vengono utilizzate molto spesso le funzioni lineari in quanto evitano che il risultato tenda verso un estremo (minimo o massimo) e rende la ANN interpretabile come una generalizzazione del modello lineare.

## 3.4 L'algoritmo di apprendimento

L'apprendimento è la proprietà che contraddistingue le reti neurali dagli altri modelli, e si concretizza nella capacità di imparare e di migliorare il proprio funzionamento attraverso una serie di aggiustamenti progressivi applicati ai pesi sinaptici.

Esistono diverse modalità di apprendimento che possono definire il percorso che porta alla modellazione di una ANN. Quasi tutte queste regole possono considerarsi delle varianti della regola di Donald Olding Hebb (Hebb, 1949), per la quale si può affermare che se un neurone di entrata ed un neurone di uscita sono fortemente attivi, ovvero sono attivati contemporaneamente per un certo tempo, aumenta la facilità di trasmissione del segnale stesso fra i due neuroni, e pertanto il valore del peso di connessione fra i due neuroni deve essere rafforzato.

Le metodologie di apprendimento che si utilizzano per istruire le reti neurali possono essere raggruppate tre distinti protocolli di apprendimento:

- apprendimento supervisionato (supervised learning);
- apprendimento non supervisionato (unsupervised learning);
- apprendimento per rinforzo (reinforcement learning).

La tipologia da usare è funzionale al campo di applicazione della ANN, ed a seconda che l'architettura sia di tipo feedforward o di tipo feedback.

Nell'apprendimento supervisionato vengono sottoposti alla rete degli input ai quali corrispondono degli output conosciuti. In questo contesto la rete viene addestrata processando anche gli output, individuando le relazioni che li collegano agli input ed impostando i pesi sinaptici con lo scopo di migliorare la risposta del modello.

L'applicazione di tale modello prevede un processo che scompone la serie storica in due sottoinsiemi: il training set utilizzato per addestrare la rete, ed il validation set che ha lo scopo di verificare le prestazioni dell'algoritmo. L'algoritmo viene quindi applicato al training set e procede con una serie di aggiustamenti che vengono effettuati mediante un insieme di regole che mira a minimizzare l'errore medio, ovvero la differenza tra l'output prodotto dalla ANN ed i valori reali registrati nella serie dei dati che forma il training set. Il validation set ha successivamente lo scopo di misurare il grado di generalizzazione del modello costruito sul training set in quanto, se la rete non è in grado di fornire delle risposte adeguate a ad un set di dati esterno a quello utilizzato per l'addestramento, è necessario ripercorrere la fase di training.

L'apprendimento non supervisionato invece, è fondato su algoritmi di addestramento basati sui soli neuroni di input che tentano di individuare dei cluster rappresentativi dei dati di ingresso. Non disponendo di un target dunque, la rete cerca di comprendere alcune proprietà degli ingressi solamente in funzione della loro disposizione. Anche in questo caso la metodologia di apprendimento è dinamica, ma sono i neuroni stessi a modificare i pesi sinaptici, lasciando alla stessa rete il compito di identificare una classificazione logica dei dati in input.

Nel metodo per rinforzo infine, un particolare algoritmo svolge delle determinate azioni sull'ambiente e viene guidato nell'apprendimento dalle retroazioni che l'ambiente ha sull'algoritmo stesso. Le azioni vengono processate attraverso una funzione di rinforzo, che misura il grado di efficacia di un'azione rispetto ad un obiettivo prestabilito. La funzione di rinforzo assegna quindi una ricompensa (valore reale positivo), o una penalizzazione (valore reale negativo), proporzionale all'avvicinamento o all'allontanamento dell'obbiettivo.

Le tre categorie di apprendimento sopra menzionate sono comunque accomunate dalla possibilità di formulare il modello generale come funzione di ottimizzazione dei pesi sinaptici; in particolare nel caso dell'apprendimento supervisionato e con rinforzo, il problema consiste nel minimizzare una funzione di errore data dalla differenza tra i valori stimati e gli output reali.

## 3.5 Funzioni di costo

Nelle reti neurali artificiali gli algoritmi di apprendimento lavorano in maniera iterativa e si interrompono secondo una regola di arresto definita da una funzione di costo (denominata anche cost function o errore di previsione).

Tra gli indicatori dell'errore di previsione utilizzati più frequentemente si trovano:

- l'errore assoluto medio (MAE, mean absolute error), riportato in equazione [3.4] dove $y_t$ rappresenta il valore osservato dell'output e $\hat{y}_t$ il valore stimato. Consiste nella media aritmetica delle differenze degli errori di previsione in valore assoluto;

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t| \qquad [3.4]$$

- l'errore quadratico medio (MSE, mean squared error) [3.5], che consiste nella media aritmetica delle differenze degli errori di previsione al quadrato. È utilizzato più spesso nella forma "rooted", riportata in equazione [3.6], in quanto espresso nelle stesse unità della variabile di interesse (in questo caso di parla di RMSE).

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 \qquad [3.5]$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2} \qquad [3.6]$$

- il mean absolute percentage error (MAPE) è la media aritmetica dei rapporti tra il valore assoluto degli errori di previsione e gli output originari in percentuale;

$$MAPE = \frac{100\%}{T}\sum_{t=1}^{T}\frac{|y_t - \hat{y}_t|}{y_t} \qquad [3.7]$$

- l'Entropia Incrociata, o Cross Entropy, è basata sull'utilizzo dei logaritmi, è sempre non negativa e tende a zero quanto più le risposte tendono al valore originario (Nielsen, 2015).

$$Cross\ Entropy = \frac{1}{T}\sum_{t=1}^{T}[\hat{y}_t \cdot \ln y_t + (1 - \hat{y}_t) \cdot \ln(1 - y_t)] \qquad [3.8]$$

Tutti gli indicatori presentati in questo paragrafo sono accumunati dal misurare gli scostamenti tra l'output stimato e quello originario, e ciascuno presenta dei punti a favore e sfavore a seconda della struttura dei dati che si deve analizzare. Ad esempio, il MAE rispetto all'MSE è preferibile in presenza di numerosi outlier[3] valori il cui valore viene

[3] In statistica viene definito "outlier" un valore molto distante dalle altre osservazioni disponibili. Solitamente si usa definire un intervallo [Q1 – k(Q3-Q1), Q3 + k(Q3-Q1)] dove Q1 e Q3 sono rispettivamente il primo ed il terzo quartile, e k una costante che assume solitamente valore unitario. I valori al di fuori dell'intervallo vengono considerati "outlier".

accentuato dall'elevazione al quadrato (Goodfellow, Bengio, & Courville, 2016, p. 181). Al contrario, in presenza di calcoli matematici complessi nell'applicazione dell'algoritmo di Backpropagation Error, è preferibile l'utilizzo del MSE rispetto al MAE in quanto l'utilizzo del valore assoluto peggiora l'efficacia dell'elaborazione (Chai & Draxler, 2014). Il differenziale sarebbe nullo solo qualora tutti gli elementi coinvolti e precedentemente descritti (input, architettura, funzioni di attivazione…) siano in grado di descrivere completamente il fenomeno oggetto di analisi.

## 3.6 Delta Rule e Algoritmo di Backward Propagation of Errors

Il Backward Propagation of Errors, solitamente abbreviato in Backpropagation Error, è sicuramente l'algoritmo più utilizzato per l'addestramento delle reti neurali artificiali e può essere considerato una generalizzazione della Delta Rule (o Regola Delta), che a sua volta è una regola utilizzata per aggiornare i pesi sinaptici basata sul metodo di Discesa del Gradiente (Gradient Descent), una procedura iterativa che porta ad individuare un punto di minimo globale di una funzione sulla base di informazioni disponibili localmente.

La Regola Delta valuta infatti il modello basandosi su un input il cui corrispondente output esatto sia noto, e corregge ogni peso sinaptico in una quantità proporzionale, e di segno opposto, rispetto al contributo che lo stesso peso concorre a fornire alla funzione di costo precedentemente individuata.

Ipotizzando una struttura di una rete neurale con *t* output, l'errore registrato dalla *t-esima* unità può essere definito come da equazione [3.9], dove $\hat{y}_t$ è la risposta stimata, mentre $y_t$ è la risposta desiderata (il valore realmente assunto dalla serie).

$$\delta_t = \hat{y}_t - y_t \qquad [3.9]$$

Definita quindi una generica funzione di costo $E$ che sia in grado di misurare la distanza tra gli $\hat{y}_t$ e gli $y_t$, l'apprendimento consiste nel minimizzare $E$, modificando i pesi nella direzione opposta al gradiente della funzione stessa, da cui il termine discesa del gradiente, alla ricerca del minimo. Si prenda ad esempio la funzione MSE (mean squared error) [3.10] che è sicuramente tra le più utilizzate nel campo delle reti neurali artificiali; tale funzione risulta sempre non negativa e, come per tutte le funzioni di costo, vale la condizione per

cui $E \approx 0$ se $\hat{y}_t \approx y_t$; pertanto le predizioni generate dal modello di rete neurale risulteranno tanto più accurate quanto maggiore sarà la prossimità di $E$ a 0.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2 \qquad [3.10]$$

Poiché il valore di $\hat{y}_t$dipende direttamente dai pesi sinaptici, per procedere alla minimizzazione di $E$ è necessario modificare i pesi sinaptici $w$ del modello di un certo ammontare $\Delta w$, applicando la cosiddetta regola delta [3.11].

$$\Delta w = -\eta \frac{\partial E}{\partial w_{ij}} \qquad [3.11]$$

Con:

- $\eta$ coefficiente di apprendimento, o learning rate, rappresentato da un numero reale compreso tra zero ed uno, e che determina la velocità di apprendimento del neurone. Rappresentando l'intensità con cui si muove l'apprendimento, si genera un trade off tra apprendimento lento ($\eta$ vicino allo zero) ed apprendimento veloce ($\eta$ vicino ad uno) che però genera la rilevazione di punti sub ottimali
- $\frac{\partial E}{\partial w_{ij}}$ derivata parziale della funzione di costo rispetto all'insieme dei pesi sinaptici, che rappresenta il contributo di ciascun parametro all'errore del modello.

In questo contesto se l'errore E ed i pesi aumentano in maniera direttamente proporzionale, la derivata di E rapportata a quella di $w$ assume un valore positivo e pertanto, in considerazione del segno negativo davanti alla derivata i pesi $w$ vengono diminuiti. Viceversa, se l'errore E ed i pesi $w$ si muovono in direzioni opposte, i pesi sinaptici $w$ vengono aumentati.

La minimizzazione di una funzione di costo attraverso la tecnica descritta nei precedenti paragrafi rappresenta la base per l'implementazione del Backpropagation Error

Algorithm, che deve il proprio nome alla particolare tecnica di propagazione all'indietro dell'errore, dallo strato di output agli strati nascosti via via precedenti.

Si definisce il gradiente della funzione di costo come da equazione [3.12]:

$$\nabla E(\vec{w}) = \left[\frac{\partial E}{\partial w_{ij}}\right]_{ij} \qquad [3.12]$$

che rappresenta il saggio della variazione della funzione di costo $E$, per uno specifico neurone *i* in uno strato *j* della rete, rispetto alle oscillazioni della somma pesata degli input connessi a *i* nel medesimo strato.

Tramite iterazioni successive, denominate epoche, l'algoritmo di Backpropagation Error mira a minimizzare l'errore aggiornando i pesi in funzione del gradiente come da [3.13]:

$$\vec{w} = \vec{w}_c - \eta \nabla E(\vec{w}) \qquad [3.13]$$

con $\vec{w}$ che rappresenta il vettore dei pesi aggiornati, $\vec{w}_c$ identifica il vettore dei pesi casuali[4] che inizializzano la rete e $\eta$ il learning rate definito in precedenza.

## 3.7 Generalizzazione del modello

Definiti tutti i componenti propri di una Rete Neurale Artificiale si passa quindi alla generalizzazione, ovvero a verificare che il modello fornisca risposte plausibili ad input diversi da quelli utilizzati nella fase di addestramento.

Vengono pertanto sottoposti al modello i dati presenti nel validation set, e vengono nuovamente misurati gli scostamenti dalle risposte effettive per verificare la bontà previsionale. È infatti possibile che la rete descriva in maniera adeguata il training set e che, per esempi di test che non sono stati inclusi nel training set, perda di efficacia

[4] I valori iniziali dei pesi vengono generalmente campionati da una distribuzione di media nulla e varianza unitaria.

generando una mappa input/output non corretta, questo avviene quando si è in presenza di un fenomeno denominato overfitting od overtraining.

Tra i metodi più comuni per contrastare l'overfitting spiccano l'early stopping, il weight decay ed il dropout. Il ricorso all'early stopping si concretizza con la valutazione della risposta su un nuovo campione, si parla in questo caso di previsioni out of sample, mentre vengono modificati i pesi in funzione degli input del training set; la funzione di errore dipendente dai nuovi input tende ad incrementare al verificarsi dell'overfitting, ed in questo contesto viene interrotto il training.

Il weight decay invece è un metodo sviluppato da Krogh e Hertz (1991), e consiste in una tecnica di penalizzazione che permette di ridurre la varianza del modello senza ridurre il numero di unità nascoste, un'operazione che a sua volta andrebbe ad inficiare la flessibilità del modello. Nella sostanza viene sommata alla funzione di costo, la norma dei pesi moltiplicata per una costante fissata dal ricercatore. In questo modo la stima del modello può essere effettuata in maniera semplice, aggiungendo un termine nell'equazione di aggiornamento dei parametri, senza costi computazionali incrementali.

Il dropout infine è una tecnica che, ad ogni passo dell'addestramento, "spegne" casualmente un certo numero di neuroni nei livelli nascosti determinando nella sostanza delle reti neurali ridotte rispetto all'architettura impostata. In questa maniera l'output di un neurone nullo, al quale corrisponde un gradiente pari a zero, non viene propagato agli altri neuroni collegati e non contribuisce all'aggiornamento degli altri pesi sinaptici.

## 3.8 Deep Learning

### 3.8.1 Artificial Intelligence, Machine Learning e Deep Learning

All'inizio del capitolo precedente è stato introdotto il termine "deep learning" come un approccio che, attraverso l'utilizzo delle reti neurali, permette di sviluppare dei sistemi evoluti di elaborazione di grandi quantità di dati. Ovviamente il concetto alla base del deep learning è ben più ampio e per essere compreso è necessario correlarlo ad altri due termini diventati di uso comune in questo contesto: il machine learning (o apprendimento automatico) e l'intelligenza artificiale (AI, Artificial Intelligence). I tre elementi possono essere rappresentati secondo lo schema riportato in Figura 3.2.

L'intelligenza artificiale ha visto i suoi primi sviluppi a partire dall'anno 1950 e può essere riassunta con una definizione di Francois Chollet (Chollet, Deep Learning with R, 2018,

p. 4): *"lo sforzo di automatizzare le attività intellettuali normalmente svolte dagli umani"*. Da questa definizione si evince che l'intelligenza artificiale comprende anche tutte quelle attività svolte da programmi che sono stati specificatamente predisposti per determinati compiti; in questo contesto, il termine Machine Learning rappresenta un sotto insieme dell'AI, e identifica un approccio che pone l'enfasi sulla capacità di un programma, in corrispondenza di determinate condizioni ambientali, di auto apprendere e modificare le proprie scelte. In buona sostanza il Machine learning automatizza la costruzione del modello analitico.

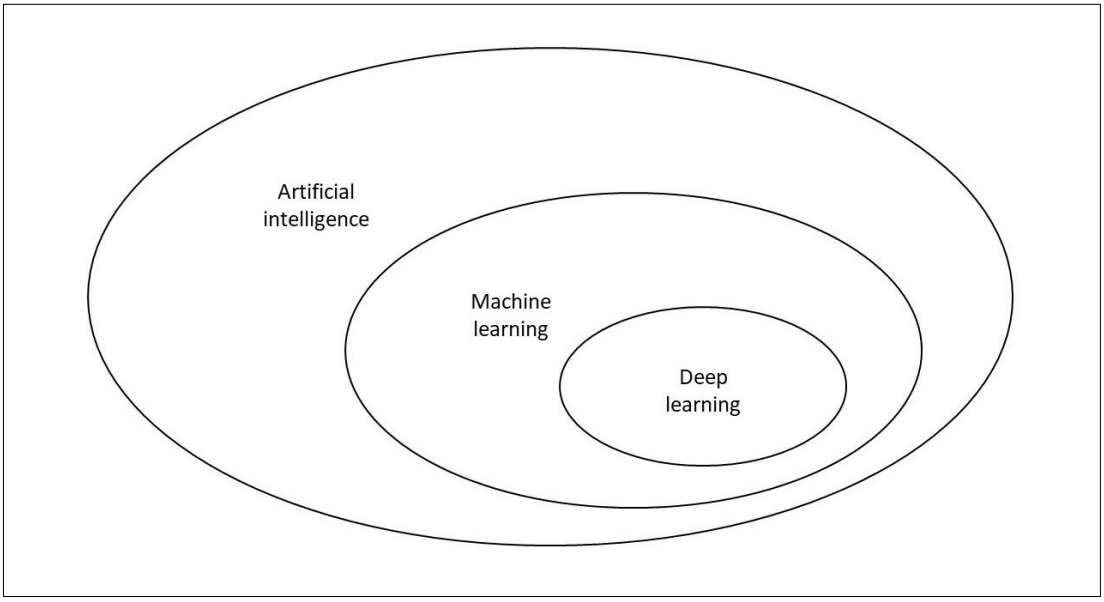


*Figura 3.2: Artificial intelligence, machine learning e deep learning (Chollet, Deep Learning with R, 2018).*

Le reti neurali sono uno dei principali strumenti utilizzati nell'apprendimento automatico in considerazione della loro particolare struttura adattiva ai dati oggetto di analisi.

Il deep learning infine è un sottoinsieme dell'apprendimento automatico il cui approccio è basato su gerarchie di rappresentazione progressiva, rappresentate dai livelli, dove le informazioni di alto livello vengono definite in base alle informazioni di livello più basso. Si concretizza con l'utilizzo modelli neurali con strutture complesse in termini sia di numero di strati nascosti, che di numero di nodi.

È interessante notare il parallelo tra l'evoluzione del deep learning ed un altro termine diventato oramai di uso comune: big data. Quest'ultimo termine, big data, è stato coniato da Doug Laney (Douglas, 2001) per definire le capacità di trattare enormi moli di dati da parte di alcuni particolari algoritmi, in poco tempo e, in termini informatici, con poche risorse. Erroneamente si tende a semplificare il concetto di big data con l'enorme quantità

di dati generati da un numero crescente di piattaforme come, solo per citarne alcune, social media e applicazioni degli smartphone. Ma approfondendone il significato, si può facilmente capire l'importante connessione, e sovrapposizione, che avviene tra le definizioni di big data e deep learning. I big data infatti coinvolgono i campi dell'informatica e della statistica, e richiedono delle tecnologie specifiche non solo nelle fasi di raccolta ed immagazzinamento di grandi volumi di dati, ma soprattutto nell'elaborazione e nella visualizzazione, per sviluppare forme di lettura ed interpretazione complesse.

Per rendere l'idea delle potenzialità legate allo sviluppo del deep learning e dei big data in ambito di analisi delle serie storiche finanziarie, si pensi a questo elaborato: nelle sezioni successive verrà presentato un trading system che si basa sulle informazioni generate da una particolare azione nella Borsa di New York, legate ai prezzi ed ai volumi giornalieri. A queste informazioni potrebbero essere affiancati ad esempio degli indicatori che derivano da un processo dell'elaborazione dei testi e del linguaggio noto come *sentiment analysis*. Questa disciplina permette di interpretare le informazioni generate da fonti come la stampa, il web ed i social media, e di trasformarle in "indicatori di sentimento" che descrivono in maniera numerica quale sia l'opinione in un particolare momento su un determinato argomento, che nel caso in specie potrebbe essere il valore del titolo azionario stesso, il settore a cui appartiene il titolo o le notizie economiche della regione di appartenenza dell'azienda. La sentiment analysis in un trading system permetterebbe pertanto di introdurre una proxy della componente emotiva che, a sua volta, potrebbe rappresentare un elemento importante a spiegare una parte della volatilità del titolo, al fine di prevederne il valore in periodi successivi al momento di analisi.

### 3.8.2 I Back End per il Deep Learning e la libreria Keras

I sistemi di deep learning possono essere implementati grazie all'utilizzo di alcuni particolari back end[5]. In questo elaborato, il trading system che verrà descritto nelle sezioni successive è stato creato con l'utilizzo di Keras (Chollet, 2015), un framework[6]

[5] Software con il quale l'utente interagisce indirettamente, solitamente attraverso l'utilizzo di un'applicazione front-end.

[6] In informatica, un framework è un metalinguaggio, ovvero un linguaggio che deriva da un altro linguaggio, attraverso il quale uno sviluppatore può svolgere un compito in maniera semplificata. Utilizzando un Framework, infatti, lo sviluppatore accede a strutture, comandi e funzionalità nativamente non previste dal linguaggio originale (Bossi, 2019).

che permette l'utilizzo di tre tra i back end più diffusi per la generazione di sistemi di deep learning:

- il TensorFlow back end, sviluppato da Google (https://tensorflow.org);
- il Theano back end, sviluppato dal MILA Lab presso Universitè de Montréal (http://deeplearning.net/software/theano);
- il Microsoft Cognitive Toolkit (CNTK) back end, sviluppato da Microsoft (https://github.com/Microsoft/CNTK).

Il trading system che verrà presentato in questo elaborato è stato sviluppato attraverso RStudio (2019) un ambiente di sviluppo integrato per il software R (2019).

# 4 Il Trading System

Come anticipato, il fine ultimo di questo elaborato è la costruzione di un trading system che combini le informazioni derivanti dalla storia passata di un titolo ed i relativi indicatori di analisi tecnica, alle metodologie di deep learning. A tale scopo verrà presentato un modello che si basa su una rete neurale ad apprendimento supervisionato e, nello specifico, verrà utilizzata una tipologia di rete neurale "fully connected". Tale modello è il più semplice tra le classi presenti nella libreria di Keras e viene applicato in maniera sequenziale; pertanto le connessioni uscenti da ogni layer rappresentano l'input per lo strato successivo, e non sono presenti loop e connessioni trasversali.

## 4.1 Il Dataset

L'analisi è stata condotta sul titolo azionario dell'azienda Citrix Systems Inc., una multinazionale nata nel 1989 a Fort Lauderdale (Florida, USA) il cui core business è basato sul cloud computing e le tecnologie per la virtualizzazione desktop e server. Il titolo è quotato sul Nasdaq di New York dal 1995. A partire dal 1997 inizia la propria espansione grazie all'acquisizione di numerose società appartenenti a diversi settori tecnologici e, nel 2017, comincia una collaborazione con Google per la creazione di un reparto di ricerca per l'intelligenza artificiale. A gennaio 2020 presenta una capitalizzazione maggiore a 16 miliardi di dollari USA, più di 130 milioni di azioni nel mercato e volumi medi giornalieri, nell'ultimo anno, superiori a 300 mila scambi.

| Anno | Mese | Società/Prodotto | Anno | Mese | Società/Prodotto |
|---|---|---|---|---|---|
| 1997 | set | DataPac | 2005 | nov | Teros |
| 1998 | gen | NTrigue | 2006 | mag | Reflectent |
| 1998 | giu | APM | 2006 | ago | Orbital Data |
| 1998 | lug | VDOnet | 2006 | dic | Ardence Inc. |
| 1999 | lug | ViewSoft | 2007 | feb | Aurema |
| 2000 | feb | Innovex Group | 2007 | set | QuickTree |
| 2001 | mar | Sequoia Software | 2007 | ott | XenSource |
| 2003 | dic | Expertcity | 2008 | mag | sepagoProfile |
| 2004 | nov | Net6 | 2008 | nov | Vapps |
| 2005 | giu | Netscaler | 2010 | ago | VMLogix Inc |

*Tabella 4.1: piano delle acquisizioni di Citrix Systems Inc. (fonte: it.wikipedia.org/wiki/Citrix_Systems)*

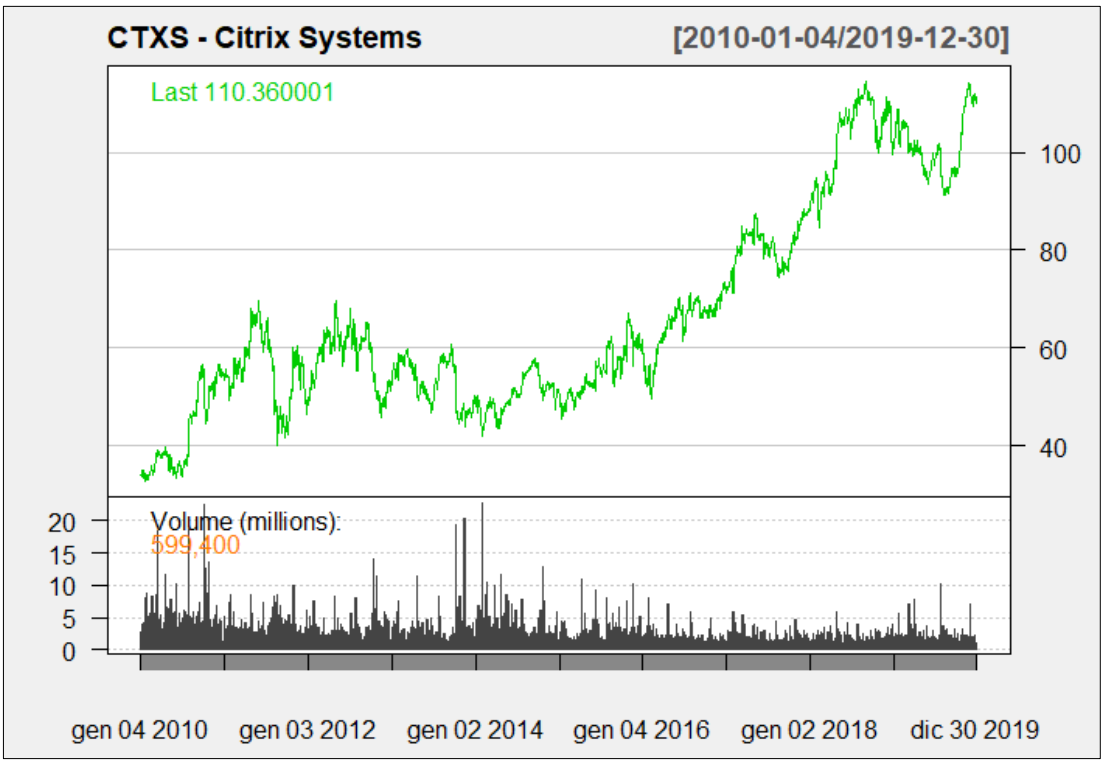


*Figura 4.1: grafico lineare dei prezzi di chiusura del titolo Citrix System.*

Il periodo di analisi è compreso tra le date 01/01/2010 e 31/12/2019.

La prima parte della serie si connota all'interno di un mercato laterale, con una tendenza intermedia soggetta a forti oscillazioni. Nel mese di maggio 2011 ad esempio il titolo quotava 70 USD, per poi precipitare a 40 USD nell'agosto dello stesso, proseguendo poi con oscillazioni meno marcate fino a gennaio 2015. Da quest'ultimo punto in poi si realizza una fase primaria tendenzialmente ascendente che presenta tuttavia delle fasi intermedie laterali, con oscillazioni negative anche superiori al 25% in archi temporali di

quattro mesi. Lo scenario si configura all'interno del piano delle acquisizioni riportato in Tabella 4.1 che giustifica la volatilità del titolo.

In questo contesto l'analisi verrà dapprima svolta costruendo un modello sull'intero dataset, simulando l'ultimo anno di trading con i segnali operativi forniti dalla rete neurale; in seconda battuta verranno presentati degli scenari "ridotti" prendendo in esame un numero di dati minori su cui verrà addestrata la rete in tre fasi distinte: una fase ascendente, una discendente ed una laterale.

Viene quindi scaricata la serie giornaliera dei prezzi di chiusura, di apertura, massimi, minimi ed i volumi scambiati del titolo e, per ogni giornata di borsa, vengono calcolati gli indicatori di analisi tecnica, creando così una matrice a due dimensioni il cui tracciato record è riportato in Tabella 4.2.

| Campo | Descrizione |
|---|---|
| Open | prezzo di apertura della giornata di borsa |
| High | prezzo massimo rilevato nella giornata di borsa |
| Low | prezzo minimo rilevato nella giornata di borsa |
| Close | prezzo di chiusura della giornata di borsa |
| Volume | volumi scambiati nella giornata di borsa |
| Adjusted | prezzo adjusted della giornata di borsa |
| i_ema4 | media mobile esponenziale a 4 elementi calcolata sul prezzo di chiusura |
| i_ema9 | media mobile esponenziale a 9 elementi calcolata sul prezzo di chiusura |
| i_ema200 | media mobile esponenziale a 200 elementi calcolata sul prezzo di chiusura |
| i_bBands_pctB | valore percentuale del prezzo di chiusura della giornata, rispetto all'ampiezza delle bande di Bollinger |
| i_roc | ROC calcolato sul prezzo di chiusura della giornata |
| i_macd_macd | MACD - Differential Line calcolata con medie mobili esponenziali a 12 e 26 elementi |
| i_macd_signal | MACD - Signal Line - media mobile esponenziale a 9 elementi della Differential Line |
| i_adx_DIp | ADX - Positive Direction Index |
| i_adx_DIn | ADX - Negative Direction Index |
| i_adx_DX | ADX - Direction Index |
| i_adx_ADX | ADX - Average Direction Index (trend strength) |
| i_aroon_aroonUp | Aroon up indicator |
| i_aroon_aroonDn | Aroon down indicator |
| i_aroon_oscillator | Aroon oscillator (aroonUp - aroonDn) |
| i_kst_kst | Know Sure Thing |
| i_kst_signal | Know Sure Thing |
| i_mfi | Money Flow Index |
| i_sar | Parabolic Stop-and-Reverse |
| i_obv | On Balance Volume |
| i_stochOSC_fastK | Stochastic Oscillator / Stochastic Momentum Index - Stochastic Fast %K |
| i_stochOSC_fastD | Stochastic Oscillator / Stochastic Momentum Index - Stochastic Fast %D |
| i_stochOSC_slowD | Stochastic Oscillator / Stochastic Momentum Index - Stochastic Slow %D |
| i_stochWPR | William's %R |
| i_trix_TRIX | Triple Smoothed Exponential Oscillator |
| i_trix_signal | Triple Smoothed Exponential Oscillator - Signal |

*Tabella 4.2: Tracciato record della matrice dei prezzi giornalieri del titolo, dei volumi e degli indicatori di analisi tecnica.*

## 4.2 Definizione del modello

Definite le variabili di input del modello è essenziale definire anche l'output, ovvero cosa la rete neurale dovrà prevedere. In funzione dell'output infatti, il problema assume connotazioni diverse che coinvolgono la struttura della rete neurale; parlando di reti neurali supervisionate, di seguito vengono riportate le diverse configurazioni che si possono riscontrare nei modelli.

- Regressione scalare: l'output è un numero, o per l'appunto uno scalare. Nel caso di un trading system, un problema di regressione scalare può essere impostato scegliendo come output il prezzo di chiusura della giornata successiva rispetto al momento di analisi. I prezzi di chiusura di ogni giornata ritardati di un periodo, verranno utilizzati come target per ogni giornata di quotazione, e rappresentano la variabile obiettivo sulla quale sarà addestrata e misurata la rete.
- Regressione vettoriale: vale quanto detto per la regressione scalare, ad eccezione del fatto che l'output in questo caso non è uno scalare, bensì un vettore. Ad esempio, qualora in un'immagine si cerchi il volto di una persona, il vettore che contiene le coordinate che incorniciano il volto è un esempio di output vettoriale.
- Classificazione binaria: l'output è rappresentato da due classi, mutuamente esclusive. È l'esempio riportato nel primo paragrafo del capitolo precedente, che riguarda l'identificazione di un husky o di un lupo in una foto.
- Classificazione multi-classe: vale quanto detto per la classificazione binaria, ma le classi sono più di due. L'output è comunque univoco e identifica una sola classe.
- Classificazione multi-etichetta: a differenza della classificazione multi-classe, è ammesso che un output identifichi più di una classe. Riprendendo l'esempio dell'husky e del lupo, se un'immagine può contenere entrambi gli animali il modello deve essere impostato come un problema di classificazione multi-etichetta affinché vengano rilevati.

Ciò premesso, volendo ricondurre la trattazione alla costruzione di un trading system, un'alternativa concreta a quanto già descritto nel caso di regressione scalare può essere rappresentata dal caso della classificazione multi-classe. Per ogni giornata di borsa infatti, in funzione di quelli che saranno i prezzi attesi per le giornate successive, è possibile ricondurre l'azione da intraprendere ad una (e solo una) delle seguenti tre classi:

- sell (vendita);

- do nothing (non compiere alcuna azione);
- buy (acquisto).

In questo contesto è stata operata la scelta di individuare delle opportunità di operatività in base ad una serie di minimi e di massimi locali, che corrisponderanno ai segnali, rispettivamente, di acquisto e di vendita. Per fare ciò, per ogni osservazione $p_t$ è stata presa una finestra temporale non centrata di ampiezza $m$ definendo un intervallo $(t, t+m)$, e si è verificato se all'interno dell'intervallo il prezzo di chiusura rappresenta un minimo od un massimo per un numero di giorni maggiore a $m$ (Figura 4.2).

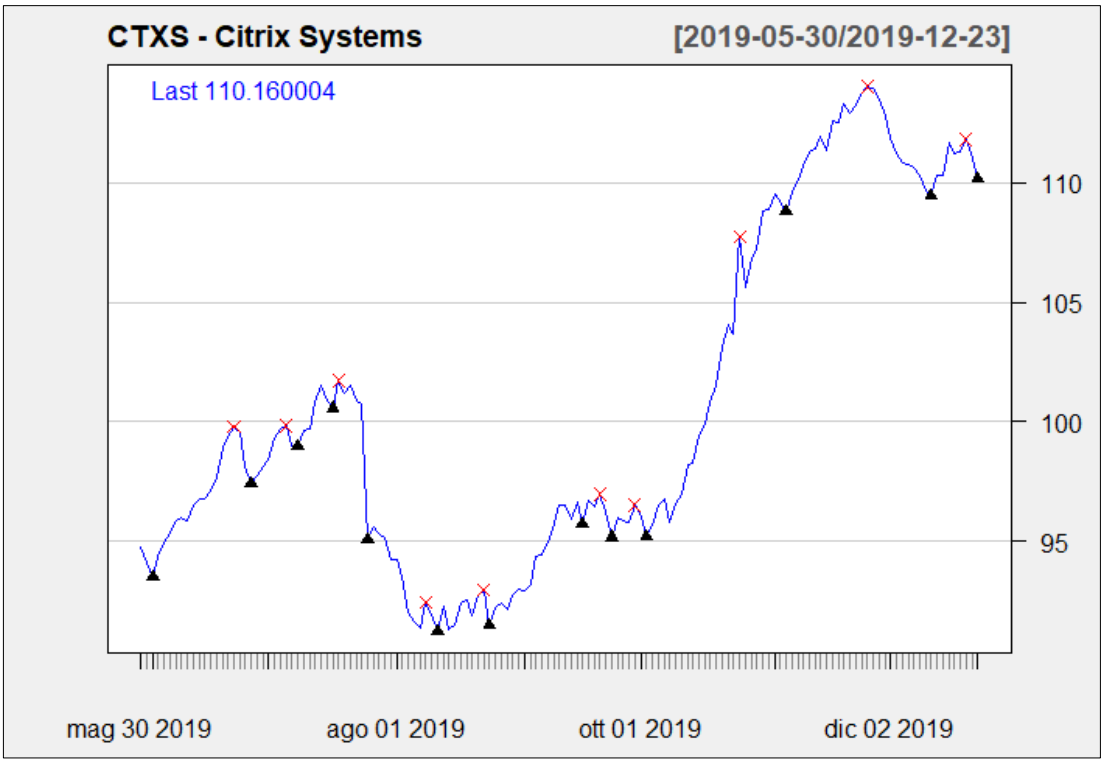


*Figura 4.2: Serie dei segnali di acquisto (triangoli neri) e di vendita (croci rosse). I segnali coincidono con le variabili target che addestreranno il modello*

## 4.3 Preparazione dei dati

La rappresentazione delle tre categorie di output avviene con il metodo denominato one-hot encoding, altrimenti detto categorical encoding. In questa modalità viene definita una variabile binaria per ciascuna delle tre etichette, che potrà assumere valore uno, qualora venga rilevato il segnale relativo all'etichetta, oppure zero in caso contrario. In ogni riga una categoria sarà valorizzata ad uno, e le altre due a zero essendo ammessa l'identificazione di una sola etichetta (ad esempio non avrebbe senso generare un segnale di "sell" ed uno di "buy" contemporaneamente).Rispetto al tracciato record riportato in

Tabella 4.2, alla matrice dei dati vengono quindi aggiunte le tre colonne relative alle etichette "sell", "do_nothing" e "buy" come da esempio riportato in Tabella 4.3.

| *(index)* | sell | do_nothing | buy | Open | | i_trix_signal |
|---|---|---|---|---|---|---|
| 02/11/2010 | 0 | 1 | 0 | 51,30594 | … | 0,18081 |
| 03/11/2010 | 0 | 1 | 0 | 51,95095 | … | 0,16218 |
| 04/11/2010 | 1 | 0 | 0 | 52,64373 | … | 0,14646 |
| 05/11/2010 | 0 | 1 | 0 | 52,80299 | … | 0,13347 |
| 08/11/2010 | 0 | 1 | 0 | 51,75983 | … | 0,12299 |
| 09/11/2010 | 0 | 0 | 1 | 52,00669 | … | 0,11466 |
| 10/11/2010 | 0 | 1 | 0 | 51,51298 | … | 0,10840 |
| 11/11/2010 | 1 | 0 | 0 | 51,91909 | … | 0,10413 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

*Tabella 4.3: One-hot encoding. Categorizzazione delle variabili target "sell", "do_nothing" e "buy" nel trading system.*

### 4.3.1 Feature Scaling

Un problema comune alla maggior parte dei problemi di machine learning consiste nel confrontare dei dati che hanno unità di misura diverse tra loro.

Nel dataset introdotto nei precedenti paragrafi ad esempio, la media mobile esponenziale a 4 elementi spazia in un range che va da 34 a 113, mentre Triple Smoothed Exponential Oscillator oscilla tra -0.83 e 0.72.

Sebbene la rete neurale dovrebbe essere in grado di adattarsi anche a dati così eterogenei, l'apprendimento potrebbe tuttavia rivelarsi più difficoltoso, innescando degli aggiornamenti del gradiente più ampi ed impedendo alla rete di convergere (Chollet, 2018, p. 77 e 92); è importante quindi procedere con il feature scaling, ovvero il ridimensionamento dei dati.

Le modalità per ridimensionare i dati a cui si ricorre solitamente sono due:

- la *standardizzazione*; consiste nel ridimensionare ciascuna feature affinché abbia un valore medio pari a zero e deviazione standard pari ad uno; la formula per la standardizzazione di una generica feature *j* è riportata nell'equazione [4.1]

$$Z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \qquad [4.1]$$

con $x_{ij}$ il valore della singola osservazione, $\mu_j$ il valore medio della feature e $\sigma_j$ la deviazione standard.

- La *normalizzazione*, definita anche *Min-Max Scaling*, riporta invece i dati di ciascuna variabile in un range solitamente ricompreso tra zero ed uno.

$$Z_{ij} = \frac{x_{ij} - min_j}{max_j - min_j} \qquad [4.2]$$

Con $max_j$ e $min_j$ rispettivamente il massimo ed il minimo valore registrato dai valori di una generica feature *j*.

Nel trading system i dati sono stati ridimensionati attraverso la standardizzazione, che ha il vantaggio di gestire in maniera migliore gli outlier.

### *4.3.2* Valutazione del modello e scomposizione dei dati

La valutazione del modello avverrà attraverso una simulazione a passi progressivi, impostando un numero di finestre temporali mobili pari al numero di giorni di trading che si intende simulare. Pertanto, dati:

- *i*, il numero giorni di trading da simulare;
- *N*, il numero di sample disponibili;
- $n = (N - i + 1)$, la dimensione delle *i* finestre temporali.
- *training_ratio*, la percentuale dei dati utilizzati per il training set rispetto agli *n* samples presenti in ciascuna finestra;

verrà testato il modello per un numero di volte pari ad *i*, secondo lo schema rappresentato in Figura 4.3.

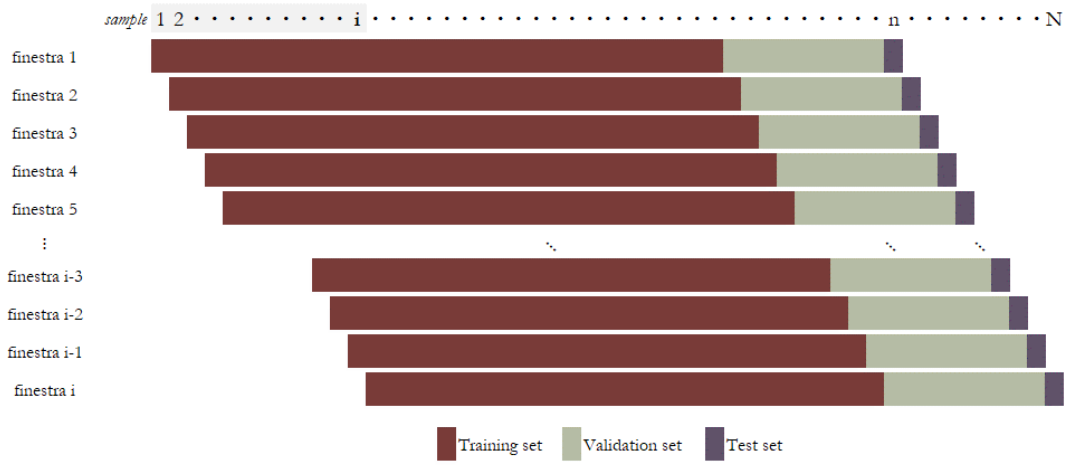


*Figura 4.3: Rappresentazione della scomposizione dei dati per la creazione dei tre set nelle diverse finestre temporali.*

Ad esempio, definito un *training_ratio* pari a 75%, e con riferimento ai primi *n* sample disponibili, per la prima finestra temporale viene utilizzato il sample *n* per testare il modello, il primo 75% dei dati (arrotondato per difetto) costituisce il training set, ed i dati rimanenti vengono impiegati per il validation set (25%). Nella seconda finestra tutte le osservazioni saranno scalate di una posizione e così via fino alla finestra *i*.

Gli *i* test set forniranno quindi un'azione di tipo "sell", o "buy", o "do nothing" per ogni giornata di borsa; sarà quindi possibile valutare l'efficienza del trading system comparando l'insieme di azioni fornite con una strategia di tipo "*buy and hold*", ovvero ipotizzando che il titolo sia acquistato nella giornata in corrispondenza dell'osservazione *n-esima* e venduto nell'ultima giornata di borsa disponibile nella finestra di osservazione (*sample N-esimo*).

## 4.4 Struttura della rete neurale

La definizione della rete neurale è sicuramente la configurazione che ricopre maggiore importanza nel processo. La scelta del numero degli strati (definito anche profondità della rete), nonché la numerosità dei neuroni, incide in maniera importante nella capacità di adattamento del modello e nei tempi di elaborazione. Infatti, come anticipato nel capitolo precedente, un modello complesso tende ad avere buone performance nei dati di training, ma rischia di generalizzare in maniera poco efficace nel momento in cui si trova ad elaborare osservazioni non presenti nella fase di addestramento, viceversa un modello troppo semplice tende a non cogliere le componenti non lineari all'interno dei dati e perdere di efficacia in termini di previsione. La configurazione dell'architettura avviene

pertanto a step progressivi, partendo da reti semplici con un solo hidden layer ed un numero limitato di neuroni, incrementandone progressivamente la complessità.

Si procede quindi con la costruzione di una rete neurale con tre strati totali: un input layer, un hidden layer ed un output layer. Il numero di neuroni da assegnare all'ultimo layer è una scelta obbligata, in quanto bisogna assegnare un neurone a ciascuno dei tre output. Per le dimensioni degli altri layer invece, il numero dei neuroni nascosti deve essere modulato in funzione del numero di indicatori di analisi tecnica che vengono forniti come input. È necessario infatti evitare che gli strati intermedi abbiano una dimensione ridotta rispetto agli input per evitare che si creino dei colli di bottiglia a causa di una rete sottodimensionata. Per questo motivo lo strato intermedio conta inizialmente un numero di neuroni pari al numero delle variabili fornite come input.

Per la scelta del modello è necessario monitorare due particolari fasi: l'*ottimizzazione*, che si riferisce alla capacità del modello di adattarsi ai dati di training, e la fase di *generalizzazione*; che riguarda la risposta del modello ai dati non inclusi nella fase di addestramento.

A tale scopo, è necessario valutare e scegliere accuratamente tre importanti elementi:

- la *funzione di attivazione*, poiché determina la trasformazione dei dati nei passaggi da uno strato a quello successivo;
- la *funzione di costo*, o *loss function*, poiché determina la misura del successo della rete;
- l'*optimizer* (*ottimizzatore*) che determina come i pesi della rete neurale devono essere modificati in fase di addestramento in base alla funzione di costo.

Con riferimento al primo punto, si descrivono di seguito le funzioni di attivazione scelte per il modello:

- per lo strato di input e per quello intermedio si è optato per una funzione denominata Parametric Rectified Linear Unit (PReLU) [4.4] che è una versione modificata della Rectified Linear Unit (RELU) [4.3];

ReLU
$$f(x) = \begin{cases} x & se\ x \geq 0 \\ 0 & se\ x < 0 \end{cases} \qquad [4.3]$$

PReLU
$$f(x_i) = \begin{cases} x_i & se\ x_i \geq 0 \\ \alpha_i x_i & se\ x_i < 0 \end{cases} \qquad [4.4]$$

Dove …

dove $x$ corrisponde al generico output di un nodo della rete. Nell'equazione della PreLU è inoltre presente un parametro $\alpha_i$ che viene aggiornato ad ogni iterazione dalla rete in analogia con quanto avviene per gli altri pesi, e che moltiplica il valore dell'input qualora sia negativo.

- per lo strato di output una funzione di attivazione Softmax [4.5], una generalizzazione su più dimensioni della funzione logistica che in fase di addestramento, ad ogni iterazione, restituisce un vettore di dimensione $k$ (con $k$ pari a tre, il numero degli output), dove l'output *k-esimo* è un numero reale compreso tra 0 e 1. L'output rappresenta pertanto la probabilità che l'osservazione appartenga alla classe *i-esima* e la somma dei tre output ad ogni iterazione è pari ad 1. In fase di previsione l'output verrà ricodificato in un vettore binario nuovamente di dimensione 3, dove l'elemento in corrispondenza del valore massimo sarà impostato a uno, e gli altri due elementi a zero. Per esempio, l'output (0.2, 0.3, 0.5) sarà ricodificato con il vettore (0, 0, 1).

Softmax $$f_k(x) = \frac{e^{x_k}}{\sum_{l=1}^{k} e^{x_l}}$$ [4.5]

In generale la funzione RELU è considerata la funzione di attivazione standard nel deep learning per due fondamentali ragioni: la sua semplicità, che si traduce in tempi computazionali minori, e la miglior risposta al "vanishing gradient problem" (Thomas, 2019). L'utilizzo della funzione di attivazione PReLU rispetto alla ReLU risponde invece all'esigenza di mitigare gli effetti del cosiddetto "dying ReLU problem" in quanto, qualora i valori di una serie di input consecutivi siano negativi, i relativi pesi smettono di essere aggiustati, senza possibilità di essere rettificati nella direzione corretta (Lu, Yeonjong, Yanhui, & George, 2019). La funzione di attivazione ReLU infatti, restituisce uno zero per ogni numero negativo ed il suo gradiente assumerà lo stesso valore. In questo modo, se il neurone assumerà un valore negativo, e finché il gradiente rimarrà pari a zero, il neurone interromperà l'apprendimento poiché la discesa del gradiente non modificherà i pesi. In questo contesto il neurone viene definito "morto".

Per quanto riguarda la Loss function invece, viene utilizzata una funzione denominata *Categorical Crossentropy*. Tale funzione ha lo scopo di misurare la distanza tra due

distribuzioni di probabilità che, nel caso del trading system, sono gli output dell'ultimo strato, corretti dalla funzione *Softmax*, e le variabili target codificate in precedenza.

Relativamente all'ottimizzatore infine, nella libreria Keras ne vengono implementate delle varianti dell'algoritmo di Back-propagation Error, utilizzate in base al problema che si sta trattando. In questo elaborato viene utilizzato l'ottimizzatore "RMSprop" che ha il vantaggio di attenuare le oscillazioni del gradiente in direzione verticale, permettendo di poter aumentare il learning rate per convergere più velocemente al minimo (Rohith, 2018).

È importante evidenziare che, poiché il trading system che si andrà a descrivere sarà composto da tre output, saranno presenti altrettante funzioni di costo; tali funzioni verranno comunque ricondotte ad un unico scalare, ottenuto come media delle tre loss function. L'ottimizzatore processerà pertanto un unico valore continuo ad ogni iterazione.

## 4.5 Iperparametri

Nel paragrafo precedente è stato descritto come il numero di strati nascosti ed il numero di neuroni per ciascuno strato, siano determinanti nelle performance di una rete neurale artificiale. Questi, insieme al learning rate, alle tecniche di dropout, al numero di epoche ed al batch size, sono denominati "iperparametri".

Con riferimento agli ultimi due iperparametri sopra citati, il numero di epoche si configura in considerazione della struttura iterativa delle reti neurali, poiché i dati devono essere processati per un numero di volte sufficiente ad aggiornare i parametri senza incorrere ad overfitting, e rappresenta di fatto il numero di volte in cui l'intero training set viene fornito alla rete.

Infine, per ogni epoca il training ed il validation set vengono forniti in sottoinsiemi di dimensione prefissata, chiamati appunto batch poiché, nella maggior parte dei casi, il numero di osservazioni da campionare è elevato e la memoria di sistema è limitata, nonostante la crescente potenza di calcolo degli elaboratori. La dimensione del batch determina un trade off tra allocamento della memoria di sistema e tempi di apprendimento: quanti più piccoli sono i batch, maggiore è la memoria di sistema disponibile ma maggiore è anche il tempo di calcolo impiegato per ciclare tutti i dati ad ogni epoca.

Rispetto ad un determinato modello di rete neurale artificiale, un iperparametro può essere pertanto definito come una variabile esterna, in quanto il suo valore viene impostato in un momento anteriore a quello dell'apprendimento.

## 4.6 Implementazione del Trading System

Definito il modello di partenza della rete neurale ed il settaggio iniziale degli iperparametri, si procede con l'implementazione del modello.

Le suddivisione del dataset riportata in Tabella 4.4 evidenzia quanto anticipato in precedenza: viene riservato il 75% dei dati per il training set ed il rimanente 25% per il validation set. Come si può notare, rispetto alla serie caricata inizialmente, e relativa al periodo 01/01/2010 – 31/12/2019, vengono eliminate molte delle osservazioni iniziali, e parte delle osservazioni finali, a causa della natura delle feature e delle variabili target. Le prime infatti comprendono medie mobili ed altri indicatori asimmetrici rispetto al momento in cui vengono calcolati, per i quali non possono essere calcolati i valori iniziali; le variabili target invece definiscono i punti di massimo e di minimo in funzione di *m* periodi successivi a quello di osservazione, e non possono pertanto essere calcolate per gli ultimi *m* periodi.

| | size | first date | last date |
|---|---|---|---|
| Train_size | 1825 | 2010-04-26 | 2017-07-24 |
| Validation_size | 608 | 2017-07-25 | 2019-12-20 |
| Total | 2434 | | |

*Tabella 4.4: Struttura iniziale del dataset.*

Per quanto riguarda le variabili target inoltre, è necessario considerare che rispetto alle 2434 osservazioni complessive, i segnali di "do_nothing" sono 1972 e rappresentano l'81% del totale. In termini di accuracy, che rappresenta la percentuale degli output generati dalla rete neurale che coincidono con le variabili obiettivo, è necessario tener conto che se un ipotetico modello restituisse la risposta do_nothing a tutte le osservazioni, raggiungerebbe comunque un'accuratezza dell'81%. Questa percentuale verrà utilizzata come benchmark di riferimento per l'accuratezza complessiva del modello.

Viene quindi addestrata la rete neurale con struttura 24-24-3 per cinquanta epoche, ed i risultati in termini di loss e di accuracy sono rappresentati in Figura 4.4

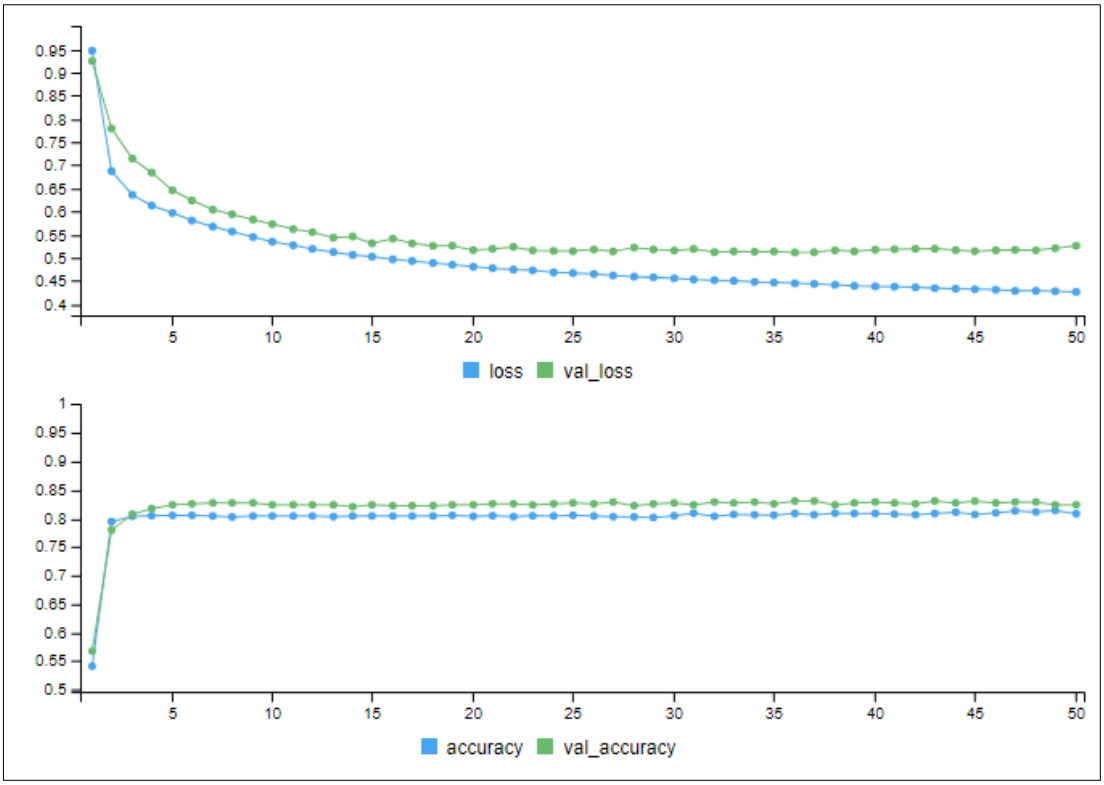


*Figura 4.4: Valori per epoca delle loss function (nella parte superiore) e dell'accuratezza (nella parte inferiore) dei dati di training (in azzurro) e di validation (in verde).*

Si può notare come ad ogni epoca le loss function diminuiscano per effetto dell'ottimizzazione attraverso la discesa del gradiente, che corregge i pesi sinaptici del modello e riduce la distanza tra l'output della rete neurale e le variabili target. Dalla ventesima epoca tuttavia, la loss function misurata sul validation set smette di decrescere e si stabilizza su valori prossimi a 0.55. Al fine di evitare un overtraining, che comporta oltretutto un allungamento dei tempi computazionali, viene definito un criterio di "early stopping" sulla loss function, che arresta l'apprendimento qualora non rilevi diminuzioni di perdita sul validation set per più di due epoche.

Per quanto riguarda l'accuracy invece, si può notare come i valori si approssimino già dalle prime epoche al benchmark definito in precedenza (81%) proprio per l'elevata frequenza di indicazioni "do nothing" rispetto alle frequenze delle classi "sell" e "buy". Nel caso in specie bisogna sottolineare che l'accuracy ha un ruolo importante ma non discriminante nell'identificazione del miglior modello di classificazione, poiché alcuni modelli con accuratezza minore rispetto ad altri, e che quindi rilevano una percentuale minore di segnali corretti, in verità possono rivelarsi più efficienti in quanto riescono a cogliere dei segnali di vendita o acquisto temporalmente vicini ai segnali effettivi (le variabili target) che non vengono però rilevati dall'accuracy. Per questo motivo l'efficacia del modello

verrà effettuata in primis sul rendimento medio, ottenuto dalle simulazioni effettuate sul modello prescelto.

La struttura della rete neurale sopra citata, la cui stampa di sistema è riportata in Figura 4.5, rappresenta la configurazione iniziale della rete neurale artificiale poiché, nelle diverse simulazioni, sono stati testati i modelli modificando progressivamente il numero di neuroni per ogni strato, e quindi il numero di strati al fine di monitorarne l'efficacia in termini di risposta.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 24)                600
_________________________________________________________________
dense_1 (Dense)              (None, 24)                600
_________________________________________________________________
dense_2 (Dense)              (None, 3)                 75
=================================================================
Total params: 1,275
Trainable params: 1,275
Non-trainable params: 0
_________________________________________________________________
```

*Figura 4.5: Keras – stampa del modello di rete neurale iniziale con struttura 24-24-3. Il numero di parametri rappresentano le sinapsi che collegano i neuroni ed include i bias.*

## 4.7 Analisi di scenario

Si procede quindi alla simulazione dei segnali di trading secondo il processo a finestre mobili introdotto nel paragrafo 4.3.2. I test saranno ripetuti per il numero di volte indicati nelle tabelle di sintesi, che a loro volta forniranno il dato medio delle caratteristiche principali: validation loss, accuracy, rendimento e tempo impiegato per ogni ciclo.

La comparazione dei modelli avverrà confrontando il rendimento medio definito dalle reti neurali (Rendimento medio ANN) con il rendimento di una strategia "*buy and hold*" (Rendimento buy-hold). Per completezza verrà anche riportato il miglior rendimento realizzabile (Rendimento best case), che si conseguirebbe qualora l'output di ciascuna finestra mobile coincidesse con la classificazione indicata dalle variabili obiettivo.

Nello specifico, escludendo ipotesi di short selling, ed ipotizzando un portafoglio inizialmente vuoto, verrà simulato un acquisto al primo segnale "buy" generato nel periodo di lunghezza *i* (giorni di trading da simulare). Successivamente, procedendo in ordine cronologico:

- qualora il portafoglio sia non vuoto ed in presenza di un segnale "sell" si simulerà la vendita al prezzo di apertura del giorno successivo;

- qualora il portafoglio sia vuoto ed in presenza di un segnale "buy" si simulerà l'acquisto al prezzo di apertura del giorno successivo, reinvestendo il capitale derivante dalla precedente vendita;
- non saranno eseguite azioni altrimenti.

### 4.7.1 Scenario 1: Fase laterale seguita da mercato bullish

La prima serie di test avviene sull'intervallo temporale ridefinito dopo la standardizzazione (26/04/2010 - 20/12/2019), simulando un'applicazione sugli ultimi 152 periodi.

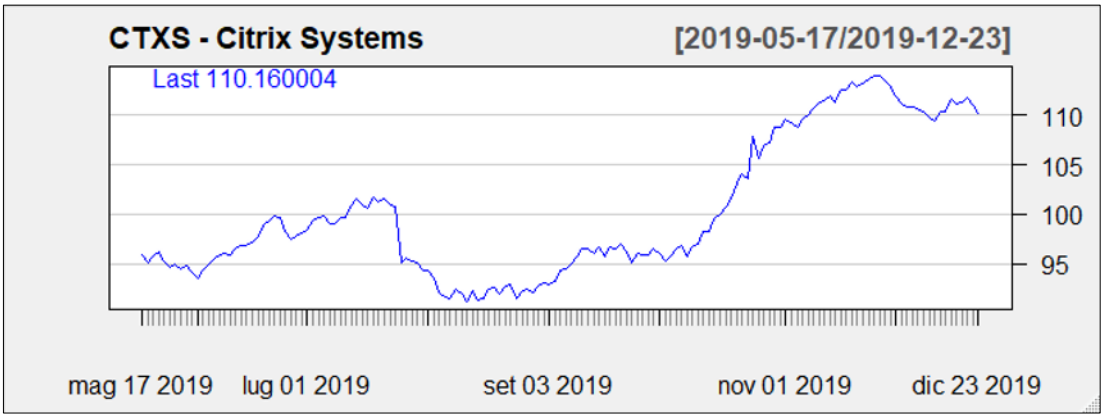


*Figura 4.6: Scenario 1. Andamento dei prezzi di chiusura del titolo Citrix Systems Inc. nel periodo 17/5/19 – 23/12/19.*

Come si può osservare dal grafico riportato in Figura 4.6, il titolo attraversa una fase laterale fino al mese di ottobre quando poi, da una quotazione di circa 95 dollari, raggiunge quota 114 a inizio dicembre per poi subire una flessione che lo porta al prezzo di 110.

| | | | |
|---|---|---|---|
| Training set size | 1588 | Giorni di simulazione | 152 |
| Validation set size | 529 | Rendimento (best case) | 28,7% |
| Data inizio simulazione | 17/05/2019 | Rendimento (buy-hold) | 15,9% |
| Data fine simulazione | 23/12/2019 | Features | 24 |

| Struttura della rete | Numero di simulazioni | Validation Loss Media | Validation Accuracy Media | Rendimento (ANN) | Tempo impiegato (Secondi per test) |
|---|---|---|---|---|---|
| 24/24/3 | 20 | 0,64 | 0,81 | 9,1% | 611 |
| 28/28/3 | 20 | 0,59 | 0,81 | **14,8%** | 614 |
| 30/30/3 | 20 | 0,59 | **0,54** | 10,9% | 658 |
| 28/28/28/3 | 20 | 0,57 | 0,75 | 8,7% | 694 |
| 28/12/6/3 | 20 | 0,63 | 0,69 | 8,1% | 687 |

*Tabella 4.5: Scenario 1. Comparazione delle performance dei modelli testati.*

L'analisi è stata condotta testando i modelli per un numero di volte pari a venti. Le caratteristiche principali delle diverse simulazioni sono sintetizzate in Tabella 4.5, dove si

evince che il modello di rete neurale che, in termini di rendimento, porta al miglior risultato è quello con struttura 28/28/3 (rendimento ANN = 14,8%), nonostante l'accuracy del modello con struttura 30/30/3 risulti nettamente inferiore (0,54), per effetto di quanto anticipato nel paragrafo 4.6.

Definito il modello con la migliore performance sono state operata ulteriori due simulazioni incrementando di una unità il numero di hidden layer, senza tuttavia ottenere miglioramenti.

In questo scenario il rendimento del miglior modello (14,8%) risulta peggiorativo rispetto al rendimento ottenuto con una strategia "*buy and hold*" (15,8%), ma al contempo permette di disporre mediamente di una maggiore liquidità sul titolo.

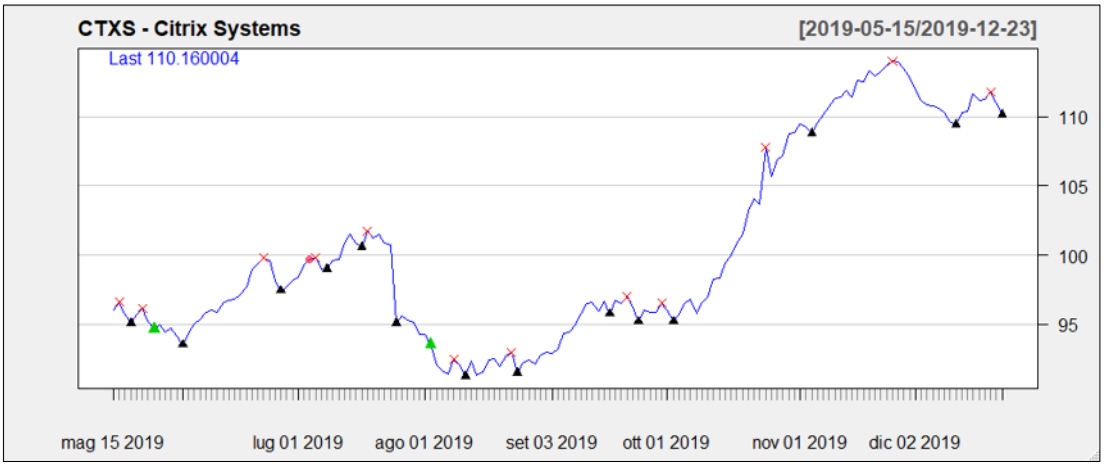


*Figura 4.7: Simulazione ottenuta con il modello a struttura 28/28/3. Rispetto alle variabili target che indicano i segnali di acquisto (triangoli neri) ed i segnali di vendita (croci rosse), vengono aggiunti i segnali restituiti dal modello (acquisto con i triangoli verdi e vendita con il pallino rosso)*

In Figura 4.7 viene riportata una delle venti simulazioni ottenute sul modello con struttura 28/28/3, in cui sono riportati i segnali generati dalla rete neurale: due segnali di acquisto, rappresentati dai triangoli di colore verde, ed un segnale di vendita rappresentato dal pallino di colore rosso in corrispondenza del dato del primo luglio.

La numerosità dei segnali diversi da "do nothing" forniti dalla rete si conferma bassa, se confrontata con il numero di segnali con i quali è stata addestrata la rete in fase di supervisione; questo avviene in coerenza con lo scopo del trading system oggetto di questo elaborato destinato una clientela di tipo retail che, a differenza di un investitore professionale deve tenere conto di due elementi di notevole importanza quali il tempo da dedicare al trading ed le commissioni bancarie. Un numero elevato di segnali di acquisto e di vendita rende necessariamente più dispendioso il processo di trading e rischia di

vedere eroso l'eventuale profitto, o addirittura renderlo negativo, proprio a causa delle commissioni.

### 4.7.2 Scenario 1 bis: Fase laterale seguita da mercato bullish, e riduzione del numero di variabili esplicative

Il numero di input, che nel caso del trading system è rappresentato dagli indicatori di analisi tecnica, incide sicuramente sulla complessità computazionale del modello a rete neurale in quanto comporta un numero di pesi sinaptici da calcolare. In questo contesto si è provato a procedere con l'eliminazione di alcune variabili che presentino, in valore assoluto, una soglia di correlazione superiore a *r* con altre variabili presenti nel dataset.

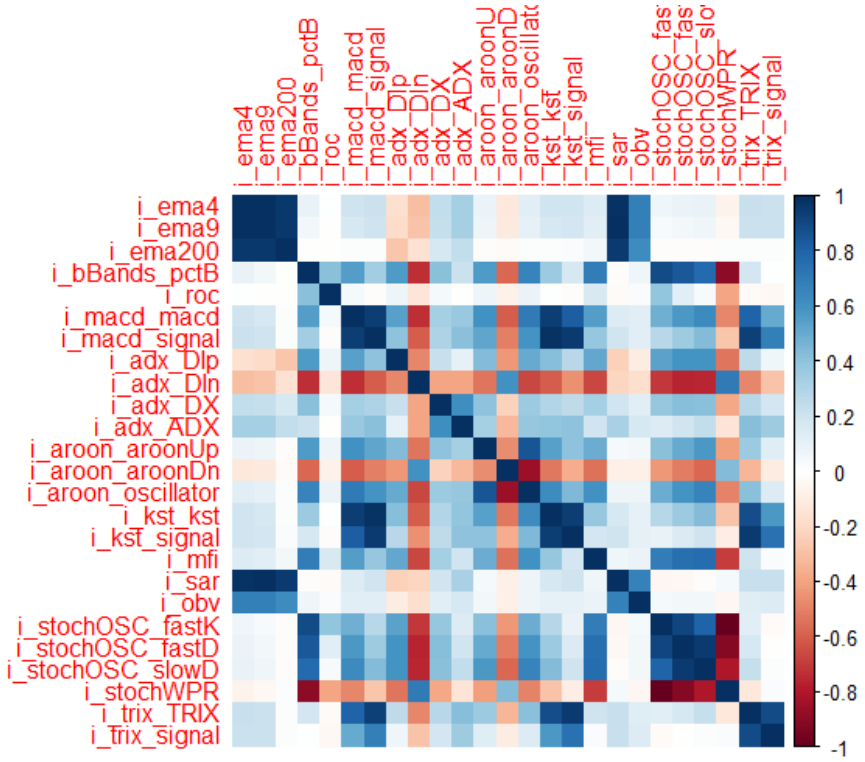


*Figura 4.8: Correlogramma degli indicatori di analisi tecnica.*

Pertanto, posto $r = 0.90$, e scegliendo di mantenere tra le variabili esplicative le medie mobili esponenziali a 4 e 9 elementi indipendente dalla correlazione con altre variabili, il processo porta all'esclusione di 9 indicatori: i_trix_TRIX, i_stochWPR, i_stochOSC_slowD, i_stochOSC_fastD, i_sar, i_kst_signal, i_kst_kst, i_macd_signal, i_ema200.

Si ripercorre quindi il processo descritto nel paragrafo precedente, partendo però da un modello con un numero inferiore di neuroni, in coerenza con il numero di feature che è diminuito per effetto della rimozione delle variabili a più alta correlazione.

| | | | |
|---|---|---|---|
| Training set size | 1588 | Giorni di simulazione | 152 |
| Validation set size | 529 | Rendimento (best case) | 28,7% |
| Data inizio simulazione | 17/05/2019 | Rendimento (buy-hold) | 15,9% |
| Data fine simulazione | 23/12/2019 | Features | 15 |

| Struttura della rete | Numero di simulazioni | Validation Loss Media | Validation Accuracy Media | Rendimento (ANN) | Tempo impiegato (Secondi per test) |
|---|---|---|---|---|---|
| 15/15/3 | 20 | 0,71 | 0,79 | 3,1% | 604 |
| 18/18/3 | 20 | 0,70 | 0,74 | **5,9%** | 607 |
| 20/20/3 | 20 | 0,68 | 0,66 | 4,9% | 637 |
| 28/28/3 | 20 | 0,62 | 0,73 | 3,7% | 642 |

*Tabella 4.6: Scenario 1BIS. Comparazione delle performance dei modelli testati.*

Come si evince dalla Tabella 4.6 il miglior rendimento ottenuto (5,9%) si registra sul modello con struttura 18/18/3; rendimento che tuttavia si distanzia notevolmente da quello ottenuto con il modello 28/28/3 (14,8%) calcolato nel paragrafo precedente e che comprende un maggior numero di feature.

Si conclude quindi che gli indicatori individuati nella fase iniziale, per quanto in alcuni casi presentino un'elevata correlazione con altre variabili descrittive, svolgono un ruolo determinante nel processo di classificazione e devono pertanto essere mantenuti. Ad ulteriore conferma, si osservino anche i tempi computazionali che presentano un miglioramento trascurabile nel modello a minor numero di input.

### 4.7.3 Scenario 2: Fase laterale seguita da mercato bearish

In ultima istanza i modelli oggetto di test nel primo scenario sono stati testati in uno scenario con una fase intermedia discendente successiva ad una fase laterale.

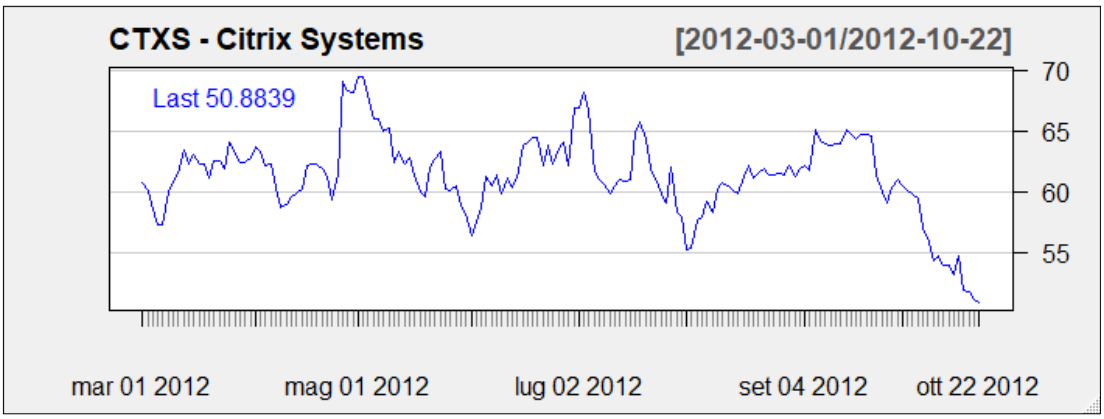


*Figura 4.9: Scenario 2. Andamento dei prezzi di chiusura del titolo Citrix Systems Inc. nel periodo 16/3/12 – 23/10/12.*

Cambiando il periodo di simulazione è stato modificato anche il periodo complessivo di riferimento dei dati che, in questo scenario, è ricompreso all'interno dell'intervallo

01/01/2002-31/10/2012 allo scopo di garantire un numero di dati per l'addestramento coerente con il primo scenario.

| | | | |
|---|---|---|---|
| Training set size | 1871 | Giorni di simulazione | 152 |
| Validation set size | 623 | Rendimento (best case) | 60,6% |
| Data inizio simulazione | 16/03/2012 | Rendimento (buy-hold) | -18,3% |
| Data fine simulazione | 23/10/2012 | Features | 24 |

| Struttura della rete | Numero di simulazioni | Validation Loss Media | Validation Accuracy Media | Rendimento (ANN) | Tempo impiegato (Secondi per test) |
|---|---|---|---|---|---|
| 24/24/3 | 10 | 0,65 | 0,80 | -8,0% | 488 |
| 28/28/3 | 10 | 0,59 | 0,80 | **-5,9%** | 687 |
| 30/30/3 | 10 | 0,58 | 0,56 | -6,8% | 706 |

*Tabella 4.7: Scenario 2. Comparazione delle performance dei modelli testati.*

Come evidenziato in Tabella 4.7 il modello che presenta il miglior rendimento, seppur negativo, è ancora una volta la rete neurale con struttura 28/28/3 (-5.9%).

Il modello risulta tuttavia nettamente migliorativo se confrontato con la strategia "*buy and hold*" (-18,3%) confermando una buona reazione alle fasi negative.

# 5 Conclusioni

In questo elaborato è stato proposto un trading system che, sulla base di 24 indicatori derivanti dall'analisi tecnica, ed attraverso l'implementazione di una rete neurale artificiale complessa, è in grado di fornire dei segnali operativi utili ad un investitore di tipo non istituzionale.

Le tecnologie legate al Deep Learning, hanno fornito un importante supporto per la definizione del modello sia in termini di tempi computazionali che in termini informatici, grazie ad un insieme di strumenti che hanno permesso di eseguire un accurato "tuning" sugli iperparametri che compongono la rete. Quest'ultimo tema ha rappresentato al contempo la maggiore difficoltà incontrata nella definizione del modello a causa della moltitudine di configurazioni ottenibili spostando le "leve" degli iperparametri. Alcune regole generali hanno sicuramente aiutato nel trovare la direzione da percorrere tuttavia, per comprendere il problema, risulta esplicativa una frase di Boris Hanin, matematico docente alla Texas A&M University e Visiting Scientist presso Facebook AI Research: *"la migliore approssimazione a ciò che sappiamo, è che non sappiamo quasi nulla di come funzionano effettivamente le reti neurali e di come sarebbe una teoria davvero profonda*". Lo stato dei fatti ha imposto pertanto di procedere con numerose sessioni di test per giungere ad una soluzione subottimale rispetto a tutte le configurazioni testate. I modelli presentati nel precedente capitolo sono infatti solo alcuni dei numerosi modelli che sono stati oggetto di studio per giungere al risultato finale.

L'utilizzo del Deep Learning ha permesso di gestire una notevole quantità di input, cogliendo per ciascun indicatore delle componenti non lineari che sono risultate

discriminanti per la definizione del modello. È risultato infatti che, l'eventuale esclusione degli indicatori che presentano un'elevata correlazione con altre variabili di input, porta ad una performance peggiore in termini di rendimento senza migliorare in maniera rilevante i tempi di elaborazione.

La ricerca è stata quindi ricondotta ad un modello di classificazione multi-classe e, dalle simulazioni svolte sulle serie storiche osservate, è emersa una buona reazione nelle fasi con tendenza decrescente, dove si registrano delle performance migliori rispetto ad una strategia "*buy and hold*". L'utilizzo del trading system è risultato particolarmente utile anche nelle fasi laterali, permettendo di mantenere in media una maggiore liquidità sull'investimento.

Per la valutazione delle performance sono stati testati i modelli molteplici volte al fine di ricavare il dato medio del rendimento effettivo. L'applicazione di un modello di classificazione sulle serie storiche comporta infatti la necessità di considerare non solo se la rete fornisce un segnale di acquisto (o di vendita) errato, ma anche di capire quanto questa indicazione "cade" temporalmente vicino ad un segnale di acquisto (o di vendita) indicato dalle variabili target. Questo meccanismo non può essere colto dall'indice di accuratezza, che verifica esclusivamente la percentuale di volte in cui l'output del modello corrisponde alla categoria individuata dalla variabile obiettivo.

L'applicazione del modello ha portato alla creazione di un numero di segnali operativi di tipo "buy" o "sell" decisamente inferiore se confrontati con la numerosità dei segnali con i quali è stata addestrata la rete, in coerenza con lo scopo di questo elaborato, ovvero di costruire un trading system che tenga conto delle esigenze di una clientela di tipo retail, dove l'aspetto commissionale è strettamente legato al numero di transazioni ed incide in maniera importante sul rendimento.

# 6 Appendice

## 6.1 Ulteriori indicatori di analisi tecnica utilizzati nel trading system

### 6.1.1 Know Sure Thing

Il Know Sure Thing è un oscillatore preposto all'individuazione di segnali di trading e a confermare tendenze in atto.

Per la costruzione dell'indice è necessario dapprima calcolare le medie mobili di quattro distinti ROC [6.1]; solitamente vengono utilizzati i lag 10, 15, 20 e 30 per costruire i ROC, e medie mobili a 10 elementi per i primi tre ROC ed una media mobile a 15 elementi per il ROC con 30 lag.

Successivamente le 4 medie mobili vengono combinate, attribuendo un peso progressivo, come riportato in [6.2].

Viene quindi calcolata la Signal Line come media mobile a 9 ritardi [6.3].

$$SMA_{m,n} = \frac{1}{m}\sum_{i=0}^{m-1} ROC_{n,t-i} \qquad [6.1]$$

$$KST = SMA_{10,10} \times 1 + SMA_{10,15} \times 2 + SMA_{10,20} \times 3 + SMA_{15,30} \times 4 \qquad [6.2]$$

$$SL_t = \frac{1}{9}\sum_{i=0}^{8} KST_{t-i} \qquad [6.3]$$

**Segnali Operativi**

- mercato bullish: $SL_t$ attraversa dal basso verso l'alto il KST;
- mercato bearish: $SL_t$ attraversa dall'alto verso il basso il KST;
- vendita: KST attraversa dal basso verso l'alto la linea dello zero;
- acquisto: KST attraversa dall'alto verso il basso la linea dello zero.

### 6.1.2 Money Flow Index

Il Money Flow Index è un oscillatore che misura la pressione del flusso di acquisto e di vendita, al fine di determinare le zone di ipervenduto o ipercomprato. È simile Relative Strength Index che però non tiene conto dei volumi di scambio.

Per la costruzione dell'indice è necessario calcolare dapprima il Tipical Price [6.4] ottenuto dal valore medio di prezzo massimo, minimo e di chiusura della giornata.

Si costruisce quindi il Money Flow [6.5] moltiplicando il Tipical Price per i volumi della giornata.

Successivamente, dato un periodo *t* solitamente fissato a 14 giorni, si calcola il Money Flow Ratio [6.6] come rapporto tra le somme dei Money Flow positivi ($MF^{+}_{t-i}$) e Money Flow negativi ($MF^{-}_{t-i}$) di periodo (un Money Flow viene considerato positivo se risulta essere maggiore al valore della giornata precedente, negativo nel caso contrario).

Viene quindi calcolato il Money Flow Index con la formula riportata in [6.7].

$$TP_t = \frac{H_t + L_t + C_t}{3} \qquad [6.4]$$

$$MF_t = TP_t \times V_t \qquad [6.5]$$

$$MFR_t = \frac{\sum_{i=0}^{k-1} MF_{t-i}^{+}}{\sum_{i=0}^{k-1} MF_{t-i}^{-}} \qquad [6.6]$$

$$MFI_t = 100 - \frac{100}{1 + MFR_t} \qquad [6.7]$$

**Segnali Operativi**

- area di ipercomprato: $MFI_t > 80\%$
- area di ipervenduto: $MFI_t < 20\%$

### 6.1.3 Parabolic Stop-and-Reverse (Parabolic SAR)

Il Parabolic Stop-and-Reverse è un indicatore utilizzato per determinare punti di inversione dei prezzi (stop and reverse) in presenza di tendenze ben delineate; risulta poco efficace pertanto in situazioni di mercato laterale.

Il calcolo dell'indice è riportato in equazione [6.8], ed il computo avviene con una formula diversa a seconda che si sia in presenza di una tendenza a rialzo ($C_t \geq C_{t-1}$ con $C_t$ che rappresenta il prezzo di chiusura al tempo $t$) o al ribasso ($C_t < C_{t-1}$). Gli altri elementi dell'equazione sono:

- gli *EP* (extreme points), che corrispondono al prezzo massimo del periodo bullish, qualora sia in atto una tendenza crescente, o ad il prezzo minimo del periodo bearish, qualora sia in atto una tendenza decrescente;
- gli *SP* (significant points), che corrispondono al prezzo massimo di chiusura del periodo bullish, qualora sia in atto una tendenza crescente, o ad il prezzo minimo di chiusura del periodo bearish, qualora sia in atto una tendenza decrescente;
- l'*AF* (fattore di accelerazione), il cui valore standard iniziale è solitamente posto uguale a 0.02 e viene incrementato a *step* progressivi, anch'essi posti usualmente uguali a 0.02, qualora all'interno della sessione venga individuato un nuovo punto estremo. L'incremento avviene fino al raggiungimento del *maximum step* il cui valore standard è uguale a 0.2.

$$SAR_t = \begin{cases} RisingSAR = SAR_{t-1} + AF_{t-1} \times (EP_{t-1} - SAR_{t-1}) \;\; se\; C_t \geq C_{t-1} \\ FallingSAR = SAR_{t-1} - AF_{t-1} \times (EP_{t-1} - SAR_{t-1}) \;\; se\; C_t < C_{t-1} \end{cases} \qquad [6.8]$$

Poiché la formula del *SAR* è iterativa (per il calcolo del *SAR* ad un determinato time frame è necessario conoscere il valore del *SAR* all'istante precedente), il primo *SAR* viene fatto corrispondere al prezzo massimo (in presenza di tendenza crescente) o minimo (in presenza di trend decrescente) della giornata che precede l'apertura della posizione.

**Segnali operativi**

- vendita: *SAR* attraversa dall'alto verso il basso la linea dei prezzi;
- acquisto: *SAR* attraversa dal basso verso l'alto la linea dei prezzi;

### 6.1.4 On Balance Volume

L'On Balance Volume è un indicatore ideato per misurare i flussi dei volumi in entrata e in uscita di uno strumento finanziario. Il calcolo dell'indice avviene come indicato nell'equazione [6.9] dove, al valore dell'indice della giornata precedente ($OBV_{t-1}$) vengono aggiunti o sottratti i volumi di scambio della giornata ($V_t$), a seconda che il prezzo di chiusura risulti maggiore o minore rispetto alla chiusura della giornata precedente.

$$OBV_t = OBV_{t-1} + \begin{cases} V_t & se\;\; C_t > C_{t-1} \\ 0 & se\;\; C_t = C_{t-1} \\ -V_t & se\;\; C_t > C_{t-1} \end{cases} \qquad [6.9]$$

con

- $C_t$ = prezzo di chiusura della *t-esima* giornata;
- $V_t$ = volumi scambiati nella *t-esima* giornata.

**Segnali operativi**

- vendita: *OBV* decrescente con prezzi in aumento
- acquisto: *OBV* crescente con prezzi in diminuzione

### 6.1.5 Stochastic Oscillator / Stochastic Momentum Index - Stochastic Fast %K

Lo Stochastic Oscillator indica la prossimità del prezzo di chiusura al massimo e minimo di giornata (si è osservata una relazione tra trend rialzisti e chiusure prossime ai massimi, e trend ribassisti e chiusure prossime ai minimi).

$$Fast\ Line\ =\ \%K_{t,n} =\ 100\ \times \frac{C_t - m_{t,n}}{M_{t,n} - m_{t,n}} \qquad [6.10]$$

$$Slow\ Line\ =\ \%D_t = SMA_3(\%K_{t,n}) \qquad [6.11]$$

con

- $C_t$ = prezzo di chiusura della *t-esima* giornata;
- $m_{t,n}$ = prezzo minimo di chiusura registrato nell'intervallo (*t-n, t*);
- $M_{t,n}$ = prezzo massimo di chiusura registrato nell'intervallo (*t-n, t*).

I segnali operativi sono generati dalla *Fast Line* e dalla *Slow Line*, le cui formulazioni base sono riportate nelle equazioni [6.10] e [6.11], che vengono poi declinate in tre diverse varianti dello Stochastic Oscillator: Fast, Slow, Full.

$$FastStochasticOscillator_t \qquad \begin{aligned} \%K_{fast}\ &=\ \%K_{t,14} \\ \%D_{fast}\ &=\ SMA_3(\%K_{t,14}) \end{aligned} \qquad [6.12]$$

$$SlowStochasticOscillator_t \qquad \begin{aligned} \%K_{slow}\ &=\ SMA_3(\%K_{fast}) \\ \%D_{slow}\ &=\ SMA_3(\%K_{slow}) \end{aligned} \qquad [6.13]$$

$$FullStochasticOscillator_t \qquad \begin{aligned} \%K_{full}\ &=\ SMA_p(\%K_{t,n}) \\ \%D_{full}\ &=\ SMA_p(\%K_{full}) \end{aligned} \qquad [6.14]$$

Il *Fast Stochastic Oscillator* [6.12] viene generato utilizzando la formulazione base e ponendo *n* uguale a 14.

Lo *Slow Stochastic Oscillator* [6.13] utilizza invece una media mobile a tre ritardi della *Fast Line* ($\%Kf_{ast}$)del Fast Stochastic Oscillator [6.12], per generare la Fast Line $\%Kf_{slow}$.

Il *Full Stochastic Oscillator* [6.14] infine può essere considerato una versione "libera" dell'indicatore, in quanto permette di operare su entrambi i parametri, ovvero i giorni da considerare e i ritardi della media mobile.

Gli indici *%D* e *%K* oscillano all'interno dell'intervallo (0, 100).

**Segnali operativi**

- ipercomprato: $\%D > 75$;
- ipervenduto: $\%D < 25$;
- acquisto: $\%D_{fast}$ attraversa dal basso verso l'alto $\%D_{slow}$;
- vendita: $\%D_{fast}$ attraversa dall'alto verso il basso $\%D_{slow}$.

### 6.1.6 William's %R

È un oscillatore simile allo Stochastic Oscillator che però, in funzione della sua formulazione [6.15], assume valori ricompresi nell'intervallo (-100, 0).

$$WPR_t = -100 \times \frac{M_{t,n} - C_t}{M_{t,n} - m_{t,n}} \qquad [6.15]$$

con

- $C_t$ = prezzo di chiusura della *t-esima* giornata;
- $m_{t,n}$ = prezzo minimo di chiusura registrato nell'intervallo (*t-n, t*);
- $M_{t,n}$ = prezzo massimo di chiusura registrato nell'intervallo (*t-n, t*).

**Segnali operativi**

- ipercomprato: $WPR_t > -20$;
- ipervenduto: $WPR_t < -80$.

### 6.1.7 Triple Smoothed Exponential Oscillator

Il Triple Smoothed Exponential Oscillator (*TRIX*) è un indicatore di momentum, costruito su una tripla media mobile ed utilizzato per individuare inversioni di tendenza.

$$TRIX_t = \frac{EMA3_t - EMA3_{t-1}}{EMA3_{t-1}} \qquad [6.16]$$

con

$$EMA3 = EMA_{SM^3} = EMA_{15}(EMA_{SM^2})$$

$$EMA2 = EMA_{SM^2} = EMA_{15}(EMA_{SM})$$

$$EMA1 = EMA_{SM} = EMA_{15}(C_t)$$

$$C_t = prezzo\ di\ chiusura\ della\ giornata\ t - esima$$

$$Signal\ line = EMA_9(C_t)$$

**Segnali operativi**

- Mercato bullish: *TRIX* attraversa dal basso verso l'alto la *Signal Line*;
- Mercato bearish: *TRIX* attraversa dall'alto verso il basso la *Signal Line*;
- Acquisto: *TRIX* attraversa dal basso verso l'alto la linea dello zero;
- Vendita: *TRIX* attraversa dall'alto verso il basso la linea dello zero.

# 7 Bibliografia


Achelis, S. (.). *Technical Analysis from A to Z*. .: .

Beber, A. (1999). *Introduzione all'analisi tecnica*. Trento: Alea - Centro di Ricerca sui Rischi Finanziari.

Bollinger, J. (2002). *Bollinger on Bollinger Bands*. McGraw Hill Professional.

Bossi, M. (2019, 04 26). *https://mrwebmaster.it/primi-passi/framework_12725.html*.

Cacique da Costa, T. R., Nazàrio, R., Zica Bergo, G., Amorin SObreiro, V., & Kimura, H. (2015). Tradyng System based on the use of technical analysis: A computational experiment. *Journal of Behavioral and Experimental Finance*, 42- 55.

Chai, T., & Draxler, R. R. (2014). *Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature*. Tratto da Geoscentific Model Development: https://doi.org/10.5194/gmd-7-1247-2014

Chollet, F. (2015). *Keras*. Tratto da GitHub repository: https://github.com/fchollet/keras

Chollet, F. (2018). *Deep Learning with R*. New York: Manning Publications Co.

Deboeck, G. J. (1995). *Neural network in finance: A tutorial*. Milano: Università Cattolica del Sacro Cuore.

Douglas, L. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*.

Edwards, R. D., & Magee, J. (1992). *Technical Analysis of Stock Trends.* Boston: J. Magee Institute.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of Variance of United Kingdom Inflation. *Econometrica*, 987-1008.

Fama, E., & Blume, M. (1966). Filter Rules and Stock-Market Trading. *The Journal of Business*, 226-241.

Fister, I., Yang, X.-S., Fister, I., Brest, J., & Fister, D. (2013). A Brief Review of Nature-Inspired Algorithms for Optimization. *ELEKTROTEHNISKI VESTNIK* , 1-7.

Galati, L., & Tumietto, A. (1999). *La previsione nei mercati.* Bancaria Editrice.

Gallo, G. M., & Pacini, B. (2015). *Metodi quantitativi per i mercati finanziari.* Roma: Carrocci Editore S.p.A.

Gately, E. (1995). *Neural Networks for Financial Forecasting.* P.J. Kaufman.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *2015 IEEE International Conference on Computer Vision (ICCV).*

Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory.* New York: Wiley.

Isidore, & Christie. (2018). Fundamental Analysis versus Technical Analysis - a Comparative Review. *International Journal of Recent Scientific Research*, 23009 - 23013.

Kar, A. (2016). Bio inspired computing - A review of algorithms and scope of applications. *Expert System With Applications*, 20-32.

Keras Documentation. (s.d.). *Advanced Activations Layers_ PReLU.* Tratto da Keras Documentation: https://keras.io/layers/advanced-activations/#prelu

Krogh, A., & Hertz, J. A. (1991). NIPS'91 Proceedings of the 4th International Conference on Neural Information Processing Systems. Denver: Morgan Kaufmann Publishers Inc.

Lin, Q. (2018). Technical Analysis and Stock Return Predictability: an Aligned Approach. *Journal of Financial Markets*, 103-123.

Lu, L., Yeonjong, S., Yanhui, S., & George, E. K. (2019). *Dying ReLU and Initialization: Theory and Numerical Examples.*

Michell, T. (1997). *Machine Learning.* McGraw Hill.

Mostafa, M. (2010). Forecasting stock exchange movements using neural networks: Empirical evidence from Kuwait. *Expert Systems with Applications*, 6302-6309.

Nielsen, M. A. (2015). *Neural Networks and Deep Learning.* Determination Press.

Niemira, M. P., & Klein, P. A. (s.d.). *Forecasting Financial and Economic Cycles.*

Pring, M. J. (1991). *Technical Analysis Explained.* New York: McGraw-Hill.

Quarteroni, A., Sacco, R., & Saleri, F. (2014). *Matematica numerica.* Milano: Springer Verlag.

R Core Team - R Foundation for Statistical Computing. (2019). R: A Language and Environment for Statistical Computing,SBN 3-900051-07-0, URL http://www.R-project.org. Vienna, Austria.

Rhea, R. (1932). *The Dow Theory: An Explanation of Its Development and an Attempt to Define Its Usefulness as an Aid in Speculation.* .: Barron.

Ribeiro, M. T., Guestrin, C., & Singh, S. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *KDD2016*. San Francisco.

Rohith, G. (2018, 06 19). *A Look at Gradient Descent and RMSprop Optimizers*. Tratto da https://towardsdatascience.com/a-look-at-gradient-descent-and-rmsprop-optimizers-f77d483ef08b

RStudio Team. (2019). RStudio: Integrated Development for R. Boston, MA.

Thomas, A. (2019). *The vanishing gradient problem and ReLUs – a TensorFlow investigation.* Tratto da adventuresinmachinelearning: https://adventuresinmachinelearning.com/vanishing-gradient-problem-tensorflow/

Wilder, J. W. (1978). *New Concepts in Technical Trading System.* Greensboro: Trend Research.