

POLITECNICO DI TORINO

Collegio di Ingegneria Matematica

Corso di Laurea in Ingegneria Matematica

Tesi di Laurea Magistrale

**Studio di sistemi per il trading
intraday basato su tecniche di
classificazione e pattern
recognition**



Relatori:

prof. Luca Cagliero

prof. Paolo Garza

Candidato:

Erika PASTORELLI

ANNO ACCADEMICO 2017-2018

Ringraziamenti

Un sentito ringraziamento va ai miei relatori, i professori L. Cagliero e P. Garza, i quali hanno saputo indirizzarmi egregiamente lungo questo lavoro di tesi, sempre con grande gentilezza, comprensione e disponibilità nell'ascoltare e chiarire eventuali dubbi.

Ringrazio i miei genitori, forse sarà un po' scontato, senza cui non avrei avuto la possibilità di perseguire gli studi universitari lontano da casa; li ringrazio per avermi supportato sia finanziariamente che psicologicamente, per tutte le telefonate quotidiane, ma soprattutto per non essersi mai arresi con me e con il mio carattere turbolento.

Ringrazio mio fratello Mirko per aver tracciato per primo il cammino che ho potuto percorrere, per i pranzi domenicali insieme, e per tutti quei saggi consigli che solo un fratello maggiore può donare.

Ringrazio il Collegio Renato Einaudi per avermi accolto negli ultimi sei anni come una seconda casa e per avermi regalato la possibilità di incontrare alcune tra le persone più importanti della mia vita, tra cui il ragazzo che da più di un anno sopporta paranoie, sbalzi d'umore, e condivide con me sacrifici e successi.

Sia l'esperienza collegiale, che quella universitaria, mi hanno concesso l'opportunità di incrociare il cammino di moltissime persone, ognuna delle quali mi ha lasciato qualcosa di sé e della sua storia. A colleghi, amici e gruppi di lavoro, dico grazie a tutti coloro che mi sono stati vicini, debellando ogni momento di solitudine con un abbraccio o un sorriso.

Grazie agli "amici di giù", che mi accolgono ogni qualvolta torno nella mia prima casa, come se nulla fosse mai cambiato.

A me stessa auguro semplicemente di trovare la strada giusta, seppur impervia mi renderà felice.

Indice

Ringraziamenti	II
1 Introduzione	1
2 Strategie di trading e money management	4
2.1 Strategie di trading	4
2.2 Metodologie di analisi	6
2.3 I fondamenti dell'analisi tecnica	8
2.3.1 Analisi grafica	9
2.3.2 Esempi di pattern	11
2.4 Elenco dei pattern considerati	14
2.5 Il money management	15
2.5.1 Come massimizzare i profitti	16
3 Strategie di trading data-driven basate su tecniche di classificazione	18
3.1 Stato dell'arte	19
4 Il sistema di trading utilizzato	26
5 Strategie di integrazione di tecniche di pattern recognition per l'identificazione di figure di analisi tecnica	31
6 Gli esperimenti	34
6.1 Il mercato	34
6.2 I risultati	39
6.2.1 Confronto tra le strategie di pattern recognition	40
6.2.2 Strategia 3: confronto tra i classificatori	41
6.2.3 Confronto tra le tipologie di filtro	42
6.2.4 Simulazione di una strategia di money management	44

7 Conclusioni e sviluppi futuri	46
7.1 Conclusioni	46
7.2 Sviluppi futuri	47
Bibliografia	63

Capitolo 1

Introduzione

Da molti anni ormai, lo sviluppo della "Tecnologia dell'informazione" (in inglese *Information Technology-IT*), ha reso possibile la memorizzazione di una grande mole di dati di ogni genere, la cui eterogeneità e complessità talvolta, ne rende impossibile una comprensione adeguata. Da qui nasce l'esigenza di applicare strumenti che rendano immediati lo studio e la successiva estrazione di informazioni potenzialmente utili; uno tra tutti il *data mining*.

Ambito di forte interesse è quello finanziario la cui forte digitalizzazione quotidianamente permette, a chiunque possieda una connessione Internet, di potere accedere alle notizie e ai dati di mercato in tempo reale. In tale contesto prende piede la tendenza a voler sfruttare tali informazioni per un proprio tornaconto personale, in una delle pratiche più diffuse di compravendita di asset finanziari, il **trading online**.

Qui il data mining, e in generale tutti i principali strumenti di apprendimento automatico, giocano un ruolo fondamentale; sempre più studi infatti se ne servono per poter prevedere l'andamento dei prezzi, al fine di rendere profittevole il proprio investimento.

Purtroppo, essendo il trading online un'attività facilmente accessibile, si è resa congeniale l'idea che anche gli utenti non professionisti, i principianti nel settore, possano ottenere cospicui guadagni senza sforzi e pericoli.

In realtà il trading è un'operazione piuttosto rischiosa: ecco perché è necessario conoscere adeguatamente il mercato cui si fa riferimento (il quale dovrà essere liquido e volatile), gli strumenti da poter sfruttare nella propria strategia, e quali sono i rischi connessi. Il rischio non potrà mai essere eliminato del tutto, ma potrà essere minimizzato mediante un adeguato programma di *money management*.

Da tali considerazioni nasce l'esigenza di costruire un modello di predizione sufficientemente completo che, prendendo coscienza del rischio, seppur in minima parte, si serva al suo interno di una delle strategie messe a disposizione per i trader, l'Analisi Tecnica (in dettaglio i *candlestick giapponesi*), disciplina che permette la previsione

del movimento del mercato mediante lo studio sistematico di grafici costruiti sull'andamento passato.

Il lavoro di tesi si è incentrato sulla progettazione di un sistema di trading in grado di generare automaticamente segnali di trading intraday sul mercato azionario (ad es. compra l'azione X se il suo prezzo salirà nelle prossime 24 ore). In base all'andamento del mercato, i segnali possono essere di tipo long (scommetto al rialzo sul prezzo di una azione) o short (scommetto al ribasso).

Il sistema proposto si basa sull'applicazione di tecniche di classificazione che apprendono dai dati storici pattern ricorrenti utili a predire il range di variazione futuro di un'azione (ad es., si predice per l'azione X una variazione percentuale giornaliera del prezzo maggiore dell'1%). In base alle predizioni effettuate dal classificatore si individua un sottoinsieme di azioni potenzialmente profittevoli nel breve periodo e si considera la confidenza delle predizioni generate dai classificatori per individuare i segnali di trading da suggerire al trader.

Il lavoro di tesi estende un precedente lavoro di ricerca finalizzato ad individuare gli algoritmi di classificazione maggiormente accurati nella predizione delle variazioni giornaliere di prezzo delle azioni. Nello specifico, per poter analizzare i prezzi, sono state analizzate le performance di quattro diversi classificatori, valutando le prestazioni in termini di profitto e tempo impiegato; a questi si è reputato opportuno aggiungere un altro classificatore bayesiano, più accurato del Naive Bayes, l'AODE. Le variabili considerate per predire il prezzo di chiusura dell'azione, includono prezzo di apertura, prezzo di chiusura, massimi e minimi giornalieri. Inoltre in questo lavoro si è integrato uno step di pattern recognition applicato ai segnali storici, utile a riconoscere i pattern grafici (rialzisti o ribassisti) dell'analisi tecnica, comunemente usati nel trading azionario.

Il contributo che la seguente tesi apporta rispetto al lavoro di ricerca precedente, afferisce principalmente ai miglioramenti che tali pattern grafici possono apportare al modello costruito in precedenza. In particolare, si cerca di rispondere alle seguenti domande:

- (i) Come integrare i pattern grafici all'interno del trading system basato su algoritmi di classificazione;
- (ii) Qual è l'impatto dell'integrazione del pattern recognition sulle prestazioni del trading system;
- (iii) In quali istanti temporali passati è opportuno analizzare la presenza di pattern grafici;
- (iv) Come incide l'affidabilità dei pattern grafici sulle performance del trading system.

Per rispondere alle domande sopra riportate, si sono effettuati vari esperimenti volti ad analizzare diversi scenari considerando tre misure di affidabilità dei pattern grafici, due istanti temporali, e invertendo tra loro gli step di pattern recognition e di

classificazione. Al fine di poter estendere il precedente lavoro esplorativo di ricerca, oltre ad aggiungere informazioni grafiche aggiuntive, si è tenuto conto anche di un primo miglioramento riguardo alla gestione del capitale d'investimento sviluppando una strategia di money management basata sulle considerazioni nate lungo lo svolgimento del lavoro stesso.

I risultati sperimentali hanno evidenziato che l'analisi basata su pattern grafici di analisi tecnica migliora le prestazioni del sistema, nello specifico la presenza dei pattern grafici è da ricercarsi nei giorni che precedono nell'immediato quello di predizione, e deve limitarsi ad un sottoinsieme di pattern ad alta affidabilità. Inoltre, l'integrazione di tali pattern risulta maggiormente efficace se utilizzata a posteriori rispetto alla classificazione.

Per quanto riguarda gli esperimenti, al fine di esplorare tre andamenti di mercato eterogenei, questi sono stati eseguiti su tre anni differenti: 2011 (trend ribassista), 2013 (trend rialzista), 2015 (trend misto); inoltre sono stati analizzati due indici (FTSE MIB e S&P500), ipotizzando per ciascun anno due posizioni di investimento (long e short).

La trattazione si articola in 7 capitoli:

Il *capitolo 2*, che segue questa introduzione, tratta in linea generale i vari possibili stili di trading con cui è possibile operare e soprattutto delle analisi a supporto dei sistemi di trading automatici. In particolare si fa riferimento all'Analisi Tecnica e al Money Management.

Il *capitolo 3* è interamente dedicato allo stato dell'arte per ciò che riguarda il data mining applicato al trading, dal 1988 ad oggi si riassumono i principali studi presenti in letteratura.

Il *capitolo 4* descrive il sistema di trading utilizzato come base per il seguente lavoro di tesi.

Il *capitolo 5* si focalizza sul principale contributo apportato dalla tesi al lavoro preesistente, ovvero sull'inserimento di pattern grafici dell'analisi tecnica, e sugli esperimenti effettuati in merito.

Il *capitolo 6* descrive il contesto dell'analisi, descrivendo i mercati e gli anni relativi agli esperimenti effettuati; di seguito vengono discussi i risultati ottenuti.

Il *capitolo 7* conclude la trattazione riepilogando i risultati ottenuti e suggerendo eventuali sviluppi futuri.

Capitolo 2

Strategie di trading e money management

La parola *trading*, derivante dall'inglese "to trade" (commerciare), significa tecnicamente commercio, negoziazione. In realtà con tale termine si indica un'attività di compravendita di azioni (in generale afferisce all'intero mondo delle Borse), volta all'ottenimento di un profitto; più precisamente si può parlare di speculazione sui mercati finanziari, non con accezione negativa, ma riferita al significato più alto di "trarre profitto mediante operazioni finanziarie e commerciali".

Grazie allo sviluppo di Internet nel 1999, il trading "classico" operato fisicamente in borsa, si è trasformato sempre più in trading *online*, ovvero la compravendita di asset per via telematica, per la quale occorre disporre semplicemente di una connessione internet e di un dispositivo da cui collegarsi. Tale attività si svolge principalmente su piattaforme online gestite da broker finanziari regolamentati CONSOB¹. Dalla figura 2.1, risulta come tale attività sia di interesse in tutto il mondo, coinvolgendo circa 9.6 milioni di persone (dato aggiornato al 2017).

2.1 Strategie di trading

A seconda dell'orizzonte temporale su cui opera il trader, si distinguono varie strategie di trading ([42],[43]) aventi punti di forza e di debolezza differenti in quanto il profitto e i rischi che ne derivano sono di diversa origine. Obiettivo comunque, resta tuttavia quello di sfruttare per quanto possibile i movimenti dei prezzi del mercato, così da poter ricavare un profitto più o meno alto, a breve o lungo termine.

¹CONSOB: Commissione nazionale per le società e la Borsa; istituita nel 1974 è un'autorità indipendente che si occupa di tutelare gli investitori e garantire la trasparenza del mercato.

²Immagine tratta da: <https://brokernotes.co/forex-traders-map/>.

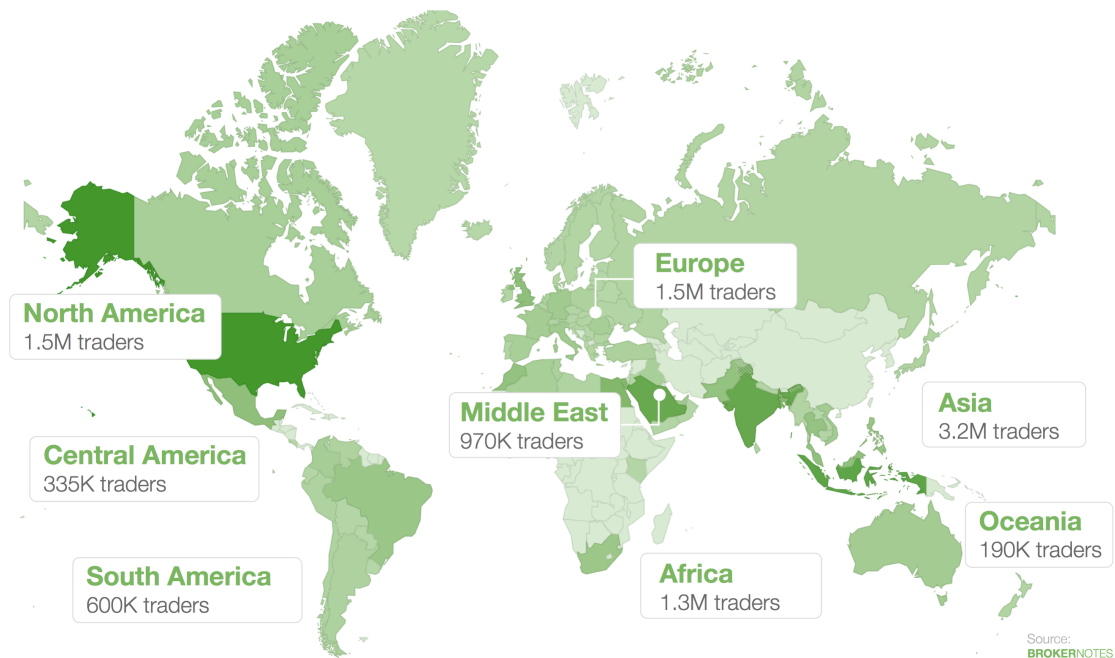


Figura 2.1: Numero di utenti che operano trading online, diviso per continente (2017).²

Buy-and-hold. La strategia *buy-and-hold*, letteralmente "Acquista e mantieni", è probabilmente quella che più di tutti si discosta dall'idea di trading attivo. In questo caso infatti, si aprono delle posizioni e le si mantengono per un lungo periodo, che può variare da poche settimane a più mesi. Coloro che operano secondo questa linea vengono detti "cassettisti", proprio per la loro tendenza ad acquisire una posizione e metterla nel cassetto.

L'obiettivo è quello di puntare sui trend maggiori del mercato, al rialzo o al ribasso, ignorando le piccole oscillazioni di breve e medio periodo. Nel caso non vi siano profonde inversioni di tendenza, una simile strategia può rivelarsi profittevole, d'altro canto, se vi è una svolta al ribasso e ci si dimentica di chiudere le posizioni aperte molto tempo prima, si potrà pregiudicare il guadagno ottenuto. Il rischio principale diviene quello di affossarsi nella staticità, non solo allontanando del tutto il divertimento che può derivare dal fare trading, ma soprattutto non riuscendo a sfruttare al meglio l'andamento del mercato.

Day trading. Con il termine *day trading*, o più comunemente detto *intraday*, si fa riferimento alla tipologia di trading più diffusa, tant'è vero che molto spesso lo si utilizza come sinonimo del trading stesso.

Come dice la parola stessa si riferisce ad un commercio condotto in un'unica giornata, tutte le operazioni infatti vengono chiuse ogni sera e riaperte la mattina seguente.

Data la sua struttura, tra i punti di forza risiede un'alta avversione alle perdite ingenti, di contro però non permette di godere appieno dei trend di fondo, in quanto le posizioni vengono chiuse anche quando potrebbero continuare a generare profitto.

Swing. La via intermedia tra il trading intraday e il buy-and-hold viene rappresentata da un ulteriore stile di trading detto *swing*. Il tempo di apertura di una posizione è variabile, supera il giorno e generalmente non va oltre una settimana, di conseguenza può permettere al trader di beneficiare di vari aspetti che né il day trading che il buy-and-hold sfruttano al meglio. Si seguono le oscillazioni del mercato, cercando di aprire alla nascita di un trend al rialzo, per poi vendere appena il prezzo subisce un'inversione al ribasso (se la posizione è short la strategia sarà opposta); per permettere un risultato ottimale non è sufficiente seguire l'andamento dei prezzi, ma occorre dotarsi di altri strumenti, quali quelli tipici di analisi fondamentale e tecnica, in modo da predire il momento esatto in cui il mercato cambierà direzione.

Scalping. Ultima frontiera del trading, infine, è quella rappresentata dallo *scalping*, termine col quale si identifica una serie di operazioni aperte e chiuse in brevissimo tempo, a distanza di minuti o anche di pochissimi secondi. E' quindi uno stile di trading da definirsi "estremo", forse il più grezzo fra tutti, in quanto non necessita nessuna conoscenza di base relativa al commercio in Borsa; avere dei nervi ben saldi, quindi, è forse l'unica prerogativa di cui necessita. Del tutto nulla è la tendenza a voler analizzare e sfruttare a proprio favore la tendenza di mercato; il susseguirsi di posizioni intervallate a limitata distanza l'una dall'altra, punta solo a ottenere piccoli profitti limitando i rischi il più possibile, ecco perché l'inserimento di uno stop loss in questo caso assume un'importanza significativa.

Ovviamente chi opera secondo stili differenti cercherà dei mercati più o meno tranquilli su cui operare; se un mercato piatto può essere avverso per lo swing che sfrutta le oscillazioni, non lo è per lo scalper, che cercherà un mercato tranquillo su cui poter proporre continuamente le stesse operazioni ottenendo di volta in volta piccoli profitti.

2.2 Metodologie di analisi

Dall'analisi delle strategie di trading, si evince come talvolta non sia il prezzo in sé l'oggetto di studio di un trader, bensì il timing dell'investimento, ovvero decidere quando entrare o uscire dalla borsa.

Per tale motivo, dopo aver definito lo stile che si vuole seguire, il più adatto alle proprie esigenze in fatto di tempo, profitti attesi e rischi da correre, si deve decidere su quali strumenti, tra quelli tipicamente messi a disposizione dei trader, si vuole puntare.

In finanza, quando si parla di **Analisi Tecnica**, ci si riferisce allo studio puro dei prezzi storici del mercato finanziario, al fine di riuscire a prevederne le tendenze future; in tale analisi ci si può avvalere principalmente dell'uso sistematico di grafici tipici (*pattern*) o di indicatori numerici (tra i principali Media Mobile³, RSI, Momentum, ADX, Bande di Bollinger, [44]).

Essa è solo uno dei tre principali strumenti finalizzati allo studio degli investimenti. Secondo la **Random Walk Theory** è impossibile effettuare una predizione sui prezzi di mercato, in quanto essi fluttuano in modo casuale e imprevedibile, indipendentemente dal loro valore passato, sotto l'ipotesi di efficienza dei mercati finanziari.

L'**Analisi Fondamentale** mira a determinare il prezzo di un titolo, focalizzandosi sulle cause, ovvero le forze economiche della società cui si fa riferimento. In realtà, l'approccio tecnico include quello fondamentale, in quanto i dati fondamentali si riflettono intrinsecamente nei prezzi di mercato; il viceversa non si verifica, ecco perché tale analisi assume cotanta valenza.

L'approccio migliore resta, tuttavia, quello in cui le due analisi di integrano, ma sviluppandosi su diversi orizzonti temporali; l'impronta tecnica potrebbe essere più utile, ad esempio, nel caso in cui si verificano delle oscillazioni repentine poiché si ha bisogno di tempo per studiarne le cause, quindi a breve termine essa potrà fornire più informazioni di quella fondamentale.

L'analisi tecnica è probabilmente la strategia più utilizzata, purtroppo per questo motivo talvolta è soggetta ad "emulazioni" da parte di trader esperti che cercano di ricreare appositamente delle figure tipiche nei prezzi, in modo da ingannare gli altri utenti ([45]). Per questo motivo occorre avere una conoscenza trasversale di tutti gli strumenti a nostra disposizione, ultimi fra tutti l'**Analisi Ciclica**, considerata forse una delle ultime frontiere del trading. Quest'ultima non parte dalla volontà di predire un movimento futuro per anticiparlo, ma da quella di capire in che fase si è in quel preciso momento, se in una fase rialzista o ribassista del trend di mercato. Essa si basa sulla consapevolezza che i prezzi si muovono in *cicli*, ovvero in un susseguirsi di minimi e massimi fino al minimo successivo: individuare in che momento del ciclo

³Media Mobile: Corrisponde alla media dei prezzi in un arco temporale definito, mobile perché viene ricalcolata in modo dinamico ogni qualvolta ci si sposta al timeframe successivo.

RSI: Relative Strength Index; è un indicatore compreso tra 0 e 100 che indica la correlazione tra giorni "positivi" e giorni "negativi".

Momentum: Differenza percentuale tra il prezzo di chiusura della candela corrente e di quello di una delle candele precedenti.

ADX: Average Directional Index; è un indicatore volto a quantificare la forza e la robustezza del trend in atto.

Bande di Bollinger: costituiscono un indicatore di volatilità del mercato, in quanto misurano la deviazione standard del prezzo.

(o di un eventuale sottociclo) si è, potrà agevolare il trader nell'investimento. Uno dei vantaggi è che essa non è manipolabile.

In questo lavoro di tesi ci concentriamo sull'utilizzo dei pattern grafici dell'analisi tecnica all'interno del sistema di trading, per questo motivo sarà discussa in modo più approfondito nel paragrafo seguente.

Successivamente si accennerà brevemente al *money management*, sistema di supporto per il trader per ridurre il rischio e gestire al meglio le proprie risorse.

2.3 I fondamenti dell'analisi tecnica

L'Analisi Tecnica trae origine dalle teorie di Charles Dow, da lui elaborate a cavallo tra XIX e XX secolo, pubblicate in più articoli sul Wall Street Journal. Il suo lavoro venne applicato agli indici del mercato azionario industriale e ferroviario che egli aveva creato; tuttavia la maggior parte dei concetti sono estendibili a più indici; le stesse basi dell'analisi tecnica, applicate in principio solo al mercato azionario, si sono estese in seguito a quello valutario e delle materie prime, fino agli indici di borsa.

La previsione futura dei prezzi, resa possibile dallo studio di dati storici, si fonda su tre premesse fondamentali:

1. **Il mercato "sconta" tutto.** Tale teoria è alla base di tutte le previsioni economiche fondamentali. Il presupposto di base è che i prezzi di borsa incorporino/riflettano tutti i fattori economici, monetari, politici e psicologici che ne hanno determinato l'andamento. In tale principio risiede anche il concetto di efficienza dei mercati.
2. **I prezzi si muovono dentro un "trend".** Tale assunzione afferma che il mercato segue dei "trend", destinati a persistere finché non mostreranno segni di esaurimento o inversione. L'analista tecnico dovrà identificare la tendenza primaria del mercato attuale così da poter assumere delle posizioni coerenti con la sua direzione, fino al verificarsi di un successivo cambiamento di tendenza. Nel mercato si sviluppano tre tipi di trend simultaneamente, i quali differiscono l'un l'altro in quanto a durata:
 - *Trend primario o di lungo termine.* Tale movimento può durare da meno di un anno a più anni. In generale lo si definisce bullish se è orientato al rialzo dei prezzi ("toro"), o bearish ("orso") nel caso contrario;
 - *Trend secondario o di medio termine.* Esso si sviluppa all'interno del precedente e può durare da circa tre settimane a vari mesi;
 - *Trend minore o di breve termine.* La durata varia da poche ore ad un massimo di tre settimane.

3. **La storia si ripete.** L'assunto che sta alla base di tale asserzione è che, poiché gli "attori" in gioco sono sempre gli stessi, ovvero essere umani il cui comportamento tende a ripetersi nel tempo con i medesimi errori, anche il mercato, al verificarsi di certe condizioni, si evolverà di conseguenza.

Tali presupposti mostrano come lo studio di movimenti passati dei prezzi, e l'individuazione di pattern ricorrenti, sia tutto ciò di cui necessita l'analisi tecnica per poter formulare previsioni fondate sotto il profilo statistico.

E' bene sottolineare come i grafici, accompagnati da indicatori tecnici, di cui tale analisi si serve, non determinano un movimento del mercato, ne riflettono semplicemente la psicologia ribassista o rialzista; la teoria dell'autoalimentazione, secondo cui gli analisi tecnici condizionerebbero il mercato stesso tramite vendite e acquisti, è quindi da biasimare. ⁴

2.3.1 Analisi grafica

L'analisi tecnica è supportata dall'utilizzo di grafici. I vantaggi consistono principalmente nell'adattabilità a diverse dimensioni temporali, e nell'opportunità di fare previsioni circa l'andamento del mercato, qualora i meccanismi che lo regolano siano stati compresi in modo efficace, e si riescano a identificare adeguatamente formazioni di prezzi ricorrenti.

Tali grafici, per poter essere costruiti, necessitano delle medesime informazioni relative all'unità di tempo considerata. Ipotizzando di voler costruire un grafico giornaliero, dovremmo disporre di *prezzo massimo*, *minimo*, di *apertura* e *chiusura* della giornata. In alcuni casi vi si affiancano anche *volume*, ovvero l'ammontare degli scambi in un dato mercato e in un determinato giorno, e l'*open interest*, ossi il numero totale di operazioni long o short in quella giornata.

I grafici possono essere di più tipi:

- Grafico point&figure (P&F);
- Grafico a linee;
- Grafico a barre;
- Grafico a candele giapponesi (*candlestick*).

Il grafico *point&figure* rivela solamente le tendenze, non mostra infatti alcun potere previsionale. La variabile tempo non è tenuta in considerazione, si rappresentano esclusivamente i prezzi e i loro movimenti mediante una successione di colonne di croci (X), indicanti i prezzi al rialzo e cerchietti (O) per quelli al ribasso.

⁴Per approfondimenti sull'Analisi Tecnica si consiglia la consultazione del libro *Analisi Tecnica dei Mercati Finanziari* di M.J.Pring [19].

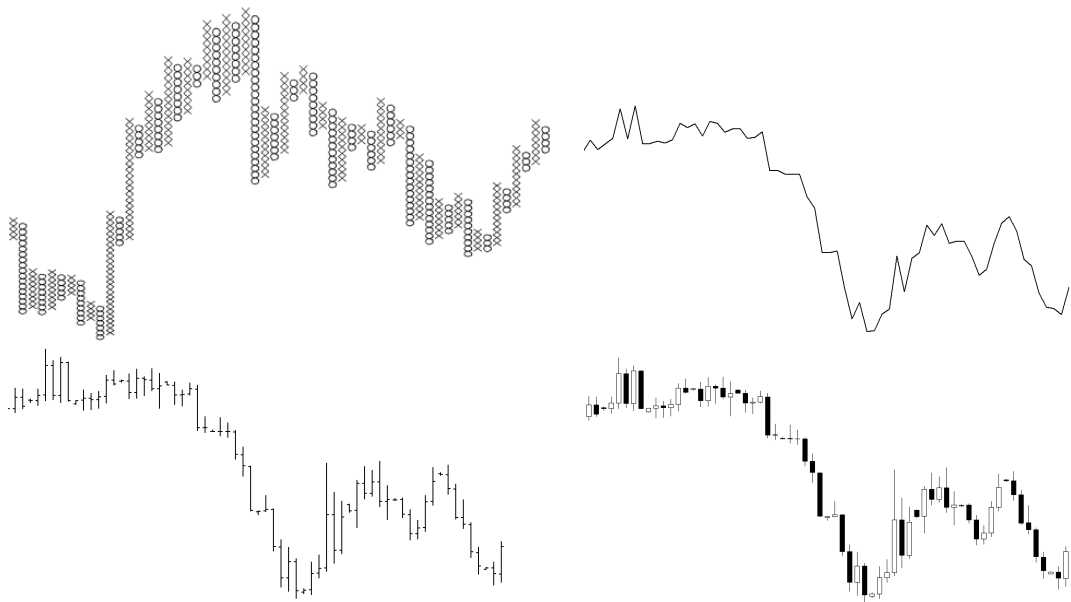


Figura 2.2: Partendo dall'alto a sinistra: P&F, a Linee, a Barre e Candlestick.⁵

Il grafico a *linee* è il più semplice: seguendo l'asse temporale, si costruisce unendo tramite linee tutti i prezzi di chiusura di giornata. In tal modo è possibile studiare in modo approssimativo l'andamento del mercato, distinguendo tra rialzista, ribassista o laterale.

Il grafico a *barre*, il più utilizzato, coniuga le informazioni circa tempo e prezzo e necessita di tutte le informazioni relative a quest'ultimo, ovvero valore massimo, minimo, di apertura e chiusura del time frame che si sta considerando.

La sua costruzione è in realtà semplice: fissata l'unità di tempo, in corrispondenza si disegna una barra verticale, la cui ampiezza è pari alla differenza tra prezzo massimo e minimo; altre due tacchette orizzontali, a sinistra e destra di quella principale, indicano rispettivamente il prezzo di apertura e chiusura.

Infine, al contrario dei precedenti, può contenere ulteriori informazioni come volumi e open interests, rappresentati mediante ulteriori barre alla base del grafico, ciascuna in corrispondenza del giorno cui fa riferimento.

I grafici a *candele giapponesi*, o semplicemente *candlestick*, hanno origine in Giappone (furono utilizzati per la prima volta nel XVIII secolo per prevedere l'andamento dei prezzi del riso) e solo recentemente il loro utilizzo si è diffuso in Occidente. Essi, in realtà, rappresentano una versione più intuitiva di quelli a barre, in quanto le informazioni racchiuse sono le stesse.

⁵Immagini tratte da <https://www.tradingblog.it> e <https://www.BlogFinanza.com>.

Mentre nel grafico precedente, abbiamo semplicemente delle barre unidimensionali, qui la barra principale viene sostituita da un rettangolo colorato detto **corpo centrale**, mentre delle lineette, dette **ombre**, definiscono la distanza dal corpo centrale ai valori di minimo e massimo del singolo time frame (2.3).

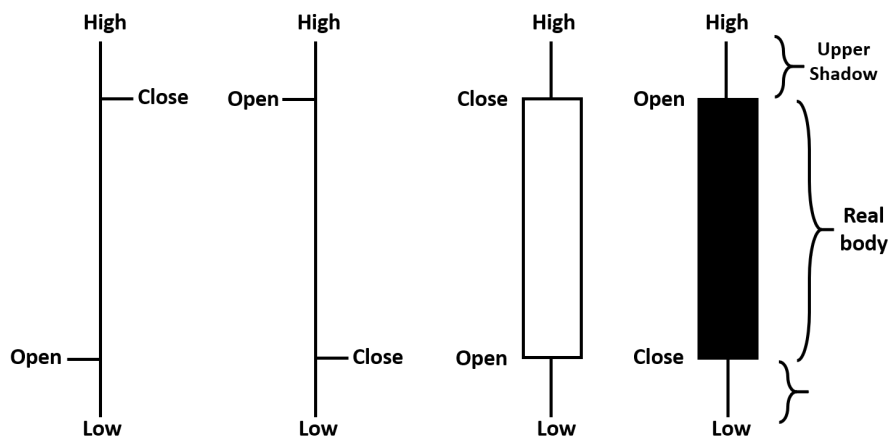


Figura 2.3: Esempio di barra e candela a confronto.

Ciò che li rende più intuitivi è il colore delle barre centrali, le quali possono essere bianche (o verdi), se il prezzo di chiusura è maggiore di quello di apertura, o nere (o rosse) se accade il viceversa.

In questo modo si può avere un'immediata analisi del movimento dei prezzi, e si ha la possibilità di riconoscere dei pattern significativi, in grado di suggerire probabili trend e/o inversioni di prezzo.

Di seguito si discuteranno i pattern più conosciuti.

2.3.2 Esempi di pattern

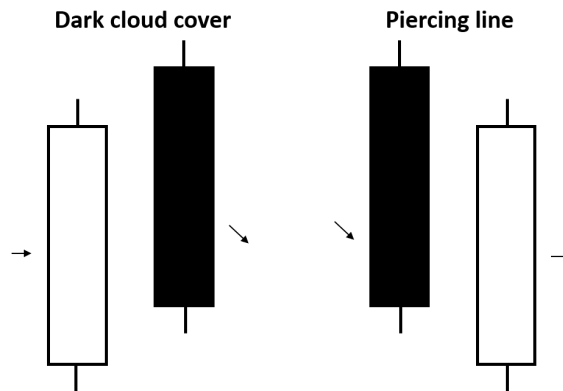
Se si considerano delle successioni di candele, combinate in modo differente tra loro, si possono individuare dei veri e propri pattern; essi possono essere costruiti mediante l'ausilio di una sola fino ad un massimo di cinque candele.

La loro utilità risiede essenzialmente nel fatto che gli uomini, comportandosi sempre nel medesimo modo, genereranno tendenzialmente le stesse reazioni nel mercato. Nonostante la maggior parte delle possibili formazioni simboleggi un'inversione della direzione dei prezzi, altri ne evidenziano la continuazione.

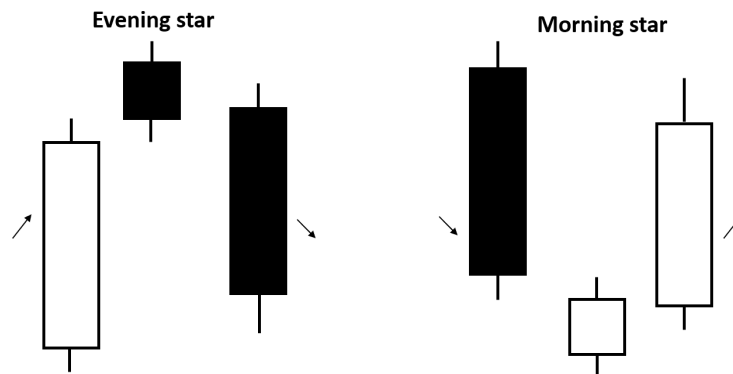
Modelli di inversione

Dark cloud cover e Piercing line. Tali pattern costituiscono l'uno l'opposto dell'altro. Composti da sole due candele, il primo, detto *dark cloud cover*, mostra

una tendenza ribassista: nonostante la prima candela sia di tipo *long white line*, simbolo di un forte gap rialzista (corpo centrale lungo e bianco, ombre corte), e il giorno seguente il prezzo di apertura sia maggiore della chiusura precedente, un evidente ribasso viene rappresentato tramite una *long black line*. Viceversa accade nel caso del *piercing line*, segno di un cambiamento di trend rialzista.



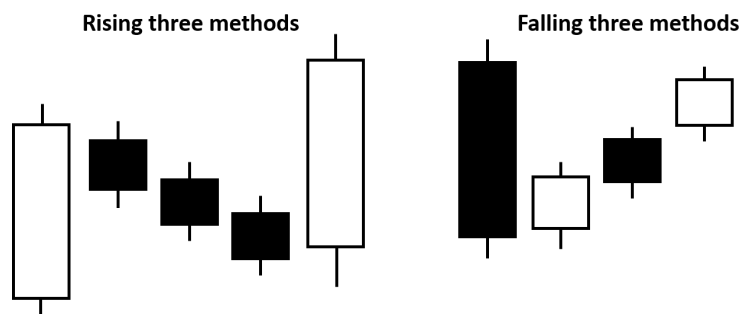
Evening star e Morning star. Così come nel caso precedente, tali configurazioni indicano trend d'inversione opposti. Come suggeriscono gli stessi nomi, *evening star* induce un movimento al ribasso, viceversa il *morning star*. In dettaglio essi si compongono di tre candele. Prendendo in considerazione il primo dei due pattern, una lunga candela bianca ed una nera saranno intervallate da una terza di corpo poco esteso, la quale giace al di sopra della prima e dà luogo ad un modello a stella (*star*).



Modelli di continuazione

Rising e falling three methods. Vi sono modelli che confermano un trend già in atto, per questo sono detti di continuazione. Il rising e falling three methods sono

opposti; dal nome si evince come il primo indichi una tendenza rialzista, mentre il secondo di tipo ribassista. Entrambi sono costituiti da cinque candlestick in successione. Nel caso del rising three methods, si ha una lunga candela bianca seguita da tre candele nere aventi il corpo molto ridotto (in realtà è sufficiente averne almeno due su tre), le quali in successione si dispongono lungo una linea al ribasso, ma i cui valori massimi e minimi restano confinati tra quelli della prima candela, l'ultima, la quinta, è di nuovo lunga e bianca.



In questo lavoro di tesi si sono considerati 61 dei pattern messi a disposizione dalla libreria *TA-Lib*, nata come un hobby nel 1999 per mano di Mario Fortier. Oggi viene utilizzata in innumerevoli applicazioni, sia open-source, che private e commerciali. Essa include circa 200 indicatori, e permette la ricognizione di pattern grafici.

Infine, poichè non tutte le configurazioni si dimostrano performanti allo stesso modo, anche perché la veridicità della loro previsione dipende da vari fattori (tra cui la bravura dell'analista), si è presa in considerazione per ciascuna, una misura di *affidabilità* da 1 a 3, indicate rispettivamente come *bassa*, *media* e *alta*, in modo da "stimare" la probabilità di successo.

Nella tabella seguente (sezione 2.4) vengono riportati tutti i pattern in ordine alfabetico, tra parentesi il numero di candele di cui ciascuna configurazione è costituita. Risultano suddivise in base alla misura detta sopra.⁶

⁶Le misure di affidabilità sono state ricavate da una libreria all'interno di *investing.com* [6].

2.4 Elenco dei pattern considerati

Pattern di affidabilità bassa

Belt Hold (2)
 Doji (1)
 Gap Side White Three Crows (3)
 Hammer (1)
 Hanging Man (1)
 Harami (2)
 Harami Cross (2)
 High Wave (1)
 Inverted Hammer (1)
 Long Legend Doji (1)
 Rickshaw Man (1)
 Separating Lines (2)
 Shooting Star (1)
 ShortLine (1)
 SpinningTop (1)
 Takuri (1)
 Three Line Strike (3)
 Thrusting (2)

Pattern di affidabilità media

Advance Block (3)
 BreakAway (5)
 Doji Star (2)
 Dragonfly Doji (1)
 Engulfing (2)
 Gravestone Doji (1)
 Hikkake (3)
 Hikkake Mod (3)
 Homing Pigeon (2)
 Ladder Bottom (5)
 Long Line (1)
 Matching Low (2)
 On Neck (2)

Piercing Line (2)
 StalledPattern (3)
 Stick Sandwich (3)
 Tasuki Gap (3)
 Three River (3)
 Three Stars in the South (3)
 Three White Soldiers (3)
 Tri-Star (3)
 Two Crows (3)
 Upside Gap Three Methods (3)

Pattern di affidabilità alta

Abandoned Baby (3)
 Closing Marubozu (1)
 ConcealBabysWall (4)
 Counter Attack (2)
 Dark Cloud Cover (2)
 Evening Doji Star (3)
 Evening Star (3)
 Falling Three Methods (3)
 Identical Three Crows (3)
 In Neck (2)
 Kicking By Length (2)
 Kicking (2)
 Maruzobu (1)
 Mat Hold (5)
 Morning Doji Star (3)
 Morning Star (3)
 Rising Three Methods (3)
 Three Black Crows (3)
 Three Inside Up (3)
 Three Outside Up (3)
 Upside Gap Two Crows (3)

2.5 Il money management

Relativamente a denaro e prudenza, Mark Twain recitava: "Ci sono due casi in cui l'uomo non deve speculare in borsa: quando non ha soldi e quando i soldi ce li ha". Da ciò si può evincere la filosofia alla base del *Money Management*, termine col quale non si indica altro che una strategia di "gestione del denaro".

Ben diverso dal concetto di *risk management*, col quale spesso si confonde, sebbene esso stesso comprenda in parte una componente di gestione del rischio.

Quest'ultimo, infatti, parlando di trading, si preoccupa principalmente del rischio legato che il mercato assume, in particolare di quanto possa essere incerta un'operazione, della sua evoluzione, e di fattori pratici, quali il settaggio migliore di parametri quali lo stop loss.

Il Money Management, al contrario, si compone principalmente di un secondo ambito, detto *position sizing* (dimensionamento della posizione operativa), ovvero della corretta ripartizione del capitale tra gli svariati asset del portafoglio.

Seppur potrebbe sembrare un concetto banale, in realtà esso incorpora varie nozioni e necessita di una buona esperienza, affinché si riesca a massimizzare il profitto, riducendo il più possibile la perdita potenziale. Ricordiamo che il capitale è una quantità limitata e che operare in Borsa significa essere soggetti ad una certa quantità di incertezza.

Per poter impiegare le proprie risorse in modo ottimale, un buon trader dovrà essere in grado di ideare una buona strategia di investimento; essa in genere si fonda sui seguenti criteri:

- Conoscenza approfondita del mercato in cui si intende investire;
- Stima del massimo rischio per il portafoglio; più nel dettaglio la percentuale del capitale iniziale perduta lungo l'operazione di trading (*drawdown*). Di solito esso viene calcolato come differenza tra un picco e una depressione relativi al capitale azionario.
- Definizione degli ordini di stop loss e take profit; ovvero il proprio profilo di rischio/rendimento, tali parametri definiranno quanto si potrà procedere prima si chiuda la posizione;
- Diversificazione del portafoglio al fine di limitare il numero di azioni strettamente correlate tra loro;
- Investimento di massimo il 5% su una singola azione del portafoglio.

Esempio pratico

L'esempio sottostante è tratto dalla pagina web [10].

Trade N°	Totale Conto	2% di rischio	Trade N°	Totale Conto	10% di rischio
1	€20.000	€400	1	€20.000	€2.000
2	€19.600	€392	2	€18.000	€1.800
3	€19.208	€384	3	€16.200	€1.620
4	€18.824	€376	4	€14.580	€1.458
5	€18.447	€369	5	€13.122	€1.312
6	€18.078	€362	6	€11.810	€1.181
7	€17.717	€354	7	€10.629	€1.063
8	€17.363	€347	8	€9.566	€957
9	€17.015	€340	9	€8.609	€861
10	€16.675	€333	10	€7.748	€775

Tabella 2.1: Esempio di strategie a confronto: rischio del 2% e del 10% per ciascun trade.

Analizzando i dati riportati in tabella 2.1, si evince come un esempio semplice come quello riportato, sià in grado di illustrare una grande differenza tra le due percentuali di rischio: 2% e 10% su ogni singola operazione; si ipotizza un capitale iniziale di €20.000.

Tale dislivello lo si nota soprattutto nel caso in cui si registri una sequenza di posizionamenti negativi in cui si perde per 10 posizioni di seguito. Nel primo caso (quello più conservativo), il capitale rimanente ammonterebbe a €16.675, mentre nel secondo a soli €7.748, perdendo in tal modo circa il 61% della quota iniziale.

Tale esempio evidenzia come una buona gestione del denaro, ed una buona impostazione dei parametri che la regolano, determinano un notevole squilibrio tra situazioni più o meno favorevoli; è importante cercare in modo di restare sempre con un capitale che ci permetta di "restare in gioco".

2.5.1 Come massimizzare i profitti

Come già detto in precedenza, riuscire a stabilire correttamente le percentuali di capitale da investire per ciascun asset, è alla base della strategia in questione.

Fra le varie metodologie utili ad ottimizzare tale ripartizione, le più discusse sono due:

- La formula di Kelly;
- La formula di Larry Williams.

le quali saranno spiegate brevemente di seguito.

Formula di Kelly

La formula di Kelly, dovuta a John Larry Kelly (1923-1965), veniva inizialmente impiegata nel gioco d'azzardo, al fine di massimizzare il profitto tratto dalle scommesse su lungo periodo. Solo successivamente ha trovato riscontro nel mercato finanziario, nel quale ha assunto la funzione di calcolo della frazione di capitale da impiegare in ciascuna operazione (F).

Essa si basa su due componenti: W e R ed assume la forma seguente:

$$F = W - \frac{1 - W}{R}$$

W esprime la probabilità di "vincita", ovvero che una operazione si generi un risultato positivo; è pari al rapporto tra il numero delle operazioni che si sono tradotte in un guadagno ed il numero totale delle operazioni. R è invece la divisione tra valore medio dei trade chiusi in profitto, e quello dei trade chiusi in perdita.

Ne risulta un numero inferiore ad 1, indice del fatto che sia una percentuale. Il difetto di tale formula consta della eccessiva semplicità, essa infatti trascura fattori importanti, quali la volatilità del mercato e il drawdown del sistema.

Formula di Larry Williams

La formula di Larry Williams è un'evoluzione della precedente in quanto permette di stabilire una percentuale F, tenendo conto del rischio che accada il contrario di ciò che si spera:

$$F = \frac{\text{Capitale} \times \text{Rischio}\%}{\text{drawdown massimo}}$$

Essa dipende anche dal drawdown, già definito nel primo paragrafo. Il fattore di tolleranza al rischio, essendo soggettivo, renderà tale equazione aleatoria, portando a risultati sempre diversi e di conseguenza poco attendibili.

Capitolo 3

Strategie di trading data-driven basate su tecniche di classificazione

Oggigiorno, il concetto di trading fa riferimento al concetto più ristretto di *trading online*, in cui tale commercio avviene sulla rete, grazie a dei broker che operano su piattaforme online. Qui gli investitori, maggiormente esperti o non, si servono di sistemi di trading automatizzati, ovvero condotti in automatico tramite computer grazie ai quali diviene possibile monitorare in tempo reale il mercato e agire di conseguenza in modo istantaneo investendo sugli asset più convenienti.

In tale sistema si stabiliscono delle regole di azione a priori, le quali saranno poi eseguite senza l'intervento manuale del trader. Un simile strumento può portare con sé vari vantaggi, tra i quali la possibilità di mettere in disparte le emozioni, rendere il profitto costante e omogeneo nel tempo limitando i rischi e di conseguenza le perdite, diversificare il proprio portafoglio, e soprattutto di reperire le informazioni più disparate da vari contesti, e sfruttarle a proprio favore. Di contro tale mezzo ha bisogno di monitoraggio costante e può essere fallace, qualora la connessione internet venga persa ad esempio, e un ordine comandato dal sistema, non riesca a raggiungere in tempo il mercato.

Alla base della costruzione di tali sistemi, ogni giorno sempre più sofisticati, vi è la capacità di memorizzare una grande mole di dati eterogenei ed analizzarli al fine di scovare informazioni potenzialmente utili. In ciò può essere riassunto il concetto di *data mining*, ovvero quel processo che, integrando concetti provenienti dal machine learning, dalla statistica e dalle basi di dati, permette l'estrazione di informazioni non note da un volume ampio di dati, rivelando in essi anche la presenza di relazioni complesse.

Il data mining risponde quindi all'esigenza di sfruttare tale ammontare di informazione a proprio beneficio, e trova applicazione in innumerevoli campi, tra cui quello

finanziario di nostro interesse.

Nel paragrafo sottostante si presentano alcuni degli studi condotti in tale ambito, così da analizzare il contributo che ciascuno ha apportato alla ricerca, permettendo la delineazione di sistemi di trading sempre più efficaci e profittevoli.

Nello specifico si ci riferisce per lo più al *trading intraday*, termine con il quale si indica l'insieme di operazioni condotte sul breve periodo (lungo l'arco di una stessa giornata), per cui si possono esigere strategie di investimento diverse da quelle adatte a lungo termine.

3.1 Stato dell'arte

Grazie al costante sviluppo della tecnologia e con l'avvento dell' "era dei dati", anche il settore finanziario, assimilando quotidianamente un'ingente quantità di informazioni provenienti dal mercato, ha mostrato la necessità di applicare i potenti tools di data mining per gestirle e sfruttarle a supporto delle decisioni.

Una delle applicazioni di maggior interesse di tale strumento, consiste nella predizione del prezzo di un'azione o semplicemente del trend del mercato, al ribasso o al rialzo, così da agevolare investitori più o meno esperti, nelle proprie strategie di trading. In realtà, secondo l'ipotesi dei mercati efficienti (EMH), descritta per la prima volta da Eugene Fama nell'articolo *Efficient Capital Markets: A Review of Theory and Empirical Work* (1970), è impossibile "battere il mercato", in quanto i prezzi dei titoli riflettono nell'immediato tutte le informazioni disponibili, impedendo così qualsiasi forma di arbitraggio. Diviene così impossibile qualsiasi tipo di predizione, in quanto i prezzi si muovono in modo randomico; l'unica tecnica d'investimento utile secondo tale ipotesi sarà quella passiva del *buy-and-hold*, che come già accennato si limita all'acquisizione di posizioni al rialzo e le mantiene per un lungo periodo.

Quello che gli studi recenti in tale ambito cercano di dimostrare, è la non credibilità dell'EMH e la possibilità di costruire modelli di training in grado di superare i limiti del settore finanziario, ottenendo delle performance superiori a quello del *buy-and-hold*, presa in molti casi come riferimento da battere.

Primi contributi. L'interesse nella predizione dei prezzi di titoli risale agli anni '80; già nel 1988 White ([41]) cercava di predire i prezzi giornalieri di IBM mediante ausilio di reti neurali, avendo già compreso la necessità di applicare tecniche in grado di cogliere la non linearità intrinseca del movimento dei prezzi. Fin da subito quest'ultima, insieme alla naturale complessità dei trend di mercato, diviene la caratteristica principale da tener in considerazione quando si sviluppa una strategia di trading.

I maggior contributi alla ricerca in tale ramo si sviluppano a partire dagli anni 2000. Già nel 2002 Leight Leigh, Purvis e Ragusa ([39]) percepiscono la necessità di costruire un sistema "eclettico", quasi anticonvenzionale per quei tempi, introducendo componenti provenienti sia dal machine learning (reti neurali - NN e algoritmi genetici - GA) che dall'analisi tecnica (template matching del pattern "*bull-flag*") di cui si riconosce per la prima volta la potenzialità. Tale metodo di template matching viene analizzato anche da Wang e Chan nel 2008 ([29]), i quali si concentrano sul training del mercato statunitense (indice NASDAQ¹); essi si discostano dal *bull-flag* pattern, oggetto di studio di molti (T.L. Chen, F.Y. Chen nel 2016 [19]), per focalizzarsi su due nuovi chart patterns: *rounding top* e *saucer*. L'obiettivo è quello di promuovere un nuovo metodo di template matching mediante l'ausilio di una matrice di facile costruzione, adattabile a più configurazioni in modo tale da essere utilizzato in sistemi reali di trading.

Nel 2010 anche Parracho, Neves e Horta([28]) propongono un sistema di trading automatico, capace di generare segnali di tipo buy/sell, basandosi sulla presenza o meno di pattern grafici al rialzo. Il contributo maggiore risiede nel determinare una tecnica di pattern-matching semplice in grado di inglobare diversi template. Nel dettaglio si effettua una moltiplicazione tra due matrici 10x10, ottenendo un valore di "fit" del matching; i parametri fondanti del metodo, quali l'ampiezza della finestra da considerare nel match, e i valori *Fit-Buy* e *Fit-Sell* (che regolano i segnali di acquisto o vendita), vengono ottimizzati mediante GA. L'esperimento, condotto su 13 anni (1998-2010) e su i tre indici Dow Jones², S&P500, NYSE³, sembra garantire prestazioni migliori del buy-and-hold in tutti i casi.

Modelli Ensemble. Un approccio rivoluzionario entra in gioco grazie a Kim, Min ed Han nel 2006, i quali, per ovviare alla complessità del problema di predizione in ambito finanziario, propongono un modello *Ensemble* di apprendimento in cui si combinano insieme più classificatori soggetti allo stesso dataset, in modo da acquisire i pregi di ciascuno rafforzando il sistema globale e riducendo la varianza dell'errore di predizione. L'indice di riferimento è il KOSPI⁴, analizzato su 11 anni (1990-2001), di cui 5 di training, 3 di test e 3 di validazione. Viene proposto un metodo di combinazione basato su GA di una rete neurale e due modelli basati su

¹NASDAQ: National Association of Securities Dealers Automated Quotation, è un indice della borsa americana comprendente i principali titoli tecnologici tra cui le compagnie informatiche come Microsoft, Cisco Systems, IBM, Apple, Google, Yahoo, Facebook e Amazon.

²Dow Jones: Dow Jones Industrial Average è il più noto indice azionario della borsa di New York; esso ingloba i 30 titoli principali di Wall Street.

³NYSE: New York Stock Exchange.

⁴KOSPI, Korea Composite Stock Price Index, è l'indice di Borsa della Corea del Sud.

opinioni di due utenti più e meno esperti del settore. Gli esperimenti mostrano come la NN sia più performante degli altri in circostanze più regolari (quando i profitti predetti sono compresi tra -3% e +3%), il viceversa vale per grandi cambiamenti di prezzo.

I modelli Ensemble vengono riproposti anche in studi successivi. Tra tutti Chang, Liu, Fan, Lin e Lai ([30], 2009) sviluppano un sistema per generare segnali "buy/sell" su breve e lungo periodo, in grado di riconoscere i punti di svolta del trend mediante una fase di preprocessing PLR (Piecewise linear representation), da innescare in input ad un complesso di classificatori formato da due diverse tipologie di reti neurali (BPN e ENN). Nel 2011 Tsai, Lin, Yen e Chen ([35]), ampliano il modello precedente prendendo in considerazione più casistiche: vengono confrontati due modelli di combinazione *major voting* e *bagging*, e considerati come classificatori rete neurale, albero di decisione (DT) e regressione logistica (LR), aggregati in modo omogeneo o eterogeneo. Gli esperimenti, condotti sul dataset di azioni dell'industria elettronica del Taiwan Economic Journal (TEJ) (anni 2002-2006), mostrano una maggior accuratezza di predizione nei modelli multipli eterogenei combinati con *major voting*.

Studi basati su reti neurali. In letteratura tuttavia, sono molteplici i lavori concentrati sull'analisi di modelli che si appoggiano ad un unico classificatore; tra questi oggetto di studio più frequente resta la rete neurale, in quanto non necessita di ipotesi a priori sul dataset, e riesce a catturare le relazioni fra le variabili durante la costruzione del modello senza vincolo di linearità. Tra tutti Mostafa ([25], 2010) propone due modelli per predire la variazione del prezzo di chiusura del KSE (Kuwait Stock Exchange), analizzando i dati dal 2001 al 2003: rete neurale di tipo multilayer perceptron (MLP) e general regression (GRNN). Nel 2011 Kara, Boyacioglu e Baykan ([26]) confronta due algoritmi di classificazione: rete neurale artificiale (ANN) ed SVM, mostrando come il primo raggiunga in media delle performance maggiori rispetto al secondo. Il mercato di riferimento è quello dell'ISE (Istanbul Stock Exchange, periodo 1997-2007). Qiu, Song e Akagi ([27], 2016) infine applicano il modello ANN per predire l'indice giapponese Nikkei 225⁵. Il loro maggior contributo consiste nel confrontare più strategie di ottimizzazione affiancate all'algoritmo di backpropagation (BP), al fine di superare i limiti di quello classico (che si appoggia generalmente alla discesa del gradiente), ovvero una lenta convergenza che rischia di cadere in un minimo locale; per tale motivo vengono esplorate due alternative basate su GA e SA⁶, le quali portano ad un minore MSE e ad una maggior

⁵Nikkei 225: indice contenente i 225 titoli delle maggiori 225 compagnie quotate al Tokio Stock Exchange (TSE).

⁶GA - Genetic algorithm, SA - Simulated Annealing

accuratezza nella predizione.

Anche Kwon e Moon ([38], 2007), ripropongono il GA per ottimizzare la rete neurale, testata su 36 compagnie del NASDAQ e NYSE per 13 anni, dal 1992 al 2004. Nel 2011 Wang, Wang, Zhang e Guo ([33]) migliorano il classico modello di NN con BP, eliminando il "rumore" dai dati originali mediante l'uso della trasformata *Wavelet*, mostrando come, fissati parametri specifici, tal strategia sia superiore in termini di MSE, RMSE e MAPE⁷.

Studi basati su altri classificatori. In realtà le tecniche di machine learning applicate in ambito finanziario sono molteplici e non tutte si fondano esclusivamente sulle reti neurali (vedi Zbikowski [23]). Nel 2010 ad esempio Texeira e de Oliveira ([24]), confrontano la classica strategia buy-and-hold ad un nuovo modello per il trading automatico, che combini gli indicatori tecnici propri dell'analisi tecnica (tra cui stop loss, stop gain, media mobile - MA, RSI, etc.), con l'algoritmo k-NN (k-Nearest Neighbor). Seppur quest'ultimo mostri risultati peggiori rispetto alla NN, la sua struttura gli consente di essere più veloce e semplice, permettendo esclusivamente di analizzare l'impatto di tali indicatori sul risultato finale; il modello, applicato su 15 azioni del l'indice Bovespa (São Paul Stock Exchange), si mostra più performante del buy-and-hold in 12 casi su 15, determinando su 8 un profitto maggiore del 30%.

Anche Patel, Shah, Thakkar e Kotecha ([22] 2015), basano la loro analisi sugli indicatori tecnici, proponendo un modello predittivo non esclusivamente intraday, cercando di stimare i prezzi delle azioni con 1, 10, 15 e 30 giorni di anticipo. Esso viene testato su due indici indiani: Nifty 50 e BSE⁸, e sugli anni dal 2002 al 2013. Lo studio si fonda un approccio a due fasi, in cui la prima applica l'algoritmo SVR (Support Vector Regression), seguito nella seconda fase nuovamente da SVR, da un ANN, o da una Random Forest (RF), corrispondente ad un ensemble di n alberi di regressione, la cui predizione finale altro non è che la media delle singole predizioni. Più in dettaglio, il primo SVR determina i parametri che saranno dati in ingresso al secondo classificatore. Dall'esperimento l'approccio a due fasi sembra essere più performante, specialmente quello fondato su rete neurale.

Il filo conduttore che unisce tutti i lavori citati, così come altri presenti in letteratura, è la staticità dei sistemi di trading costruiti, ovvero l'essere stati designati per uno specifico mercato, in un determinato periodo storico, senza possibilità di adattarsi

⁷MSE - Mean squared error, RMSE - Root mean square error, MAPE - Mean absolute percentage error.

⁸Nifty 50: National Stock Exchange of India, esso rappresenta 50 azioni appartenenti a 12 settori indiani, equamente pesati.

BSE: S&P Bombay Stock Exchange Sensex, comprende 30 aziende simbolo dell'economia indiana.

in modo intelligente a diverse condizioni. Da ciò nasce l'esigenza di sperimentare un modello di supporto alle decisioni che si adatti a più indici mantenendo delle buone prestazioni.

Nel 2015, Wang, Wang, Zhao e Tan([34]), per ovviare a tale staticità, propongono un metodo qualitativo basato sul concetto di "grafo dinamico bayesiano", la cui struttura dovrebbe catturare e rappresentare le relazioni che intercorrono in un determinato periodo temporale fra tutti i fattori economici considerati in ingresso al sistema stesso. Si dimostra, solo a livello empirico, che la topologia di tale grafo (nodo = feature, arco = correlazione fra nodi), è rappresentativa di un trend di mercato: se essa cambia, il trend ha subito una svolta. Seppur tale tecnica possieda dei limiti, tra cui l'impossibilità nel distinguere tra un'inversione al rialzo o al ribasso, e si ponga solo come un aiuto teorico per studi successivi, è un primo passo per trovare una strategia di trading che sia realmente applicabile on line, in quanto è sufficiente monitorare la topologia di tale struttura nel tempo.

Chiang, Enke, Wu e Wang ([20], 2016) propongono un altro tipo di approccio adattivo basato su ANN, ottimizzato in fase di BP grazie alla Particle Swarm Optimization (PSO). Tale sistema si ripropone di predire la direzione dei prezzi (output compreso tra 0.5 e 1 indica un rialzo, viceversa se compreso tra 0 e 0.5), servendosi in input di 12 variabili indipendenti, tra cui il prezzo di chiusura, il volume, e altri indicatori tecnici quali MA, MPP⁹, PPO¹⁰. Grazie ad una misura di DCP (Direction Correctness Percentage) si stabilisce la validità del modello; ogni qualvolta viene considerato un indice differente in ingresso, si testano più parametri dell'algoritmo affinché si riescano a trovare quelli più performanti, in questo modo è possibile adattarsi a più input differenti. Tuttavia, tale strategia è stata testata solo sui due indici S&P500 e NASDAQ nel periodo 2008-2010, mostrando ad ogni modo una miglior accuratezza di predizione rispetto al buy-and-hold.

Studi recenti. Negli ultimissimi anni la ricerca in tale ambito ha ampliato sicuramente i propri orizzonti, introducendo tecniche che prendano in considerazione aspetti tralasciati in studi antecedenti.

Baralis, Cagliero, Cerquitelli, Garza e Pulvirenti ([18], 2017), ad esempio, propongono un sistema di trading automatico che tenga conto delle influenze che due azioni x ed y , potrebbero generare sul prezzo di z . Il modello, volto alla predizione delle variazioni di prezzo, si basa essenzialmente su due tecniche di data mining: regressione e individuazione di sequenze temporali (sequence mining), avente il fine di

⁹MPP: Moving Price level Percentage; esprime la relazione fra supporto e resistenza in un trend.

¹⁰PPO: Percentage Price Oscillator; indica la relazione fra due medie mobili esponenziali a breve e lungo termine

identificare tra i dati storici, dei pattern che si propongono frequentemente. Più in dettaglio si vogliono scoprire le sequenze più usuali di azioni, così da poter affermare che se un'azione appartiene al set dei k asset più profittevoli in un istante temporale, è più probabile che faccia parte dello stesso insieme all'istante successivo (in una strategia *intraday*, come in questo caso, si considera si procede di giorno in giorno). Poiché è importante considerare in questo tipo di analisi il profitto generato da un'azione, lo si inserisce all'interno dell'algoritmo, andando a "pesare" le sequenze in base al profitto medio delle azioni che ve ne fanno parte. Per quanto riguarda la regressione, gli algoritmi considerati sono tre: SVM, regressione lineare e albero di decisione. Lo studio, condotto per tre indici differenti: NASDAQ-100¹¹, Dow Jones e FTSE MIB, fa riferimento ad investimenti sia di tipo long che short focalizzandosi, per quanto riguarda il training, su due anni in cui entrambi i mercati (statunitense e italiano) hanno mostrato tendenze opposte: il 2011, anno delle crisi dei mercati italiani e spagnoli, ed il 2013, anno di forte crescita per quello US. Gli esperimenti hanno mostrato una preferenza per il sistema di *weighted sequence mining*, sia per quanto concerne i profitti medi giornalieri ottenuti, sia in relazione alla buona interpretabilità del modello.

Chen e Hao, ([31], 2017), prendono come riferimento il mercato cinese (Shanghai e Shenzhen Stock Exchange Indices), durante il periodo 2008-2014 durante il quale esso ha vissuto una crescita economica stabile, e ne predicono i prezzi sia su breve (1-5gg), medio (10-15gg) e lungo termine (20-30gg). Il framework da essi proposto, si serve di una struttura ibrida tra SVM e KNN, utilizzati nella fase di train e di test rispettivamente; la maggior novità consiste nel pesare le n features in ingresso, mediante l'utilizzo di una matrice $n \times n$, avente sulla diagonale una misura di *information gain* di ciascuna feature. In ingresso vengono considerati, oltre ai prezzi di ingresso, uscita, minimo e massimo di ciascun asset in un determinato giorno, il volume di trade e 9 indicatori tecnici; l'etichetta di classe sarà la variazione del prezzo di chiusura, normalizzata a +1 e -1 a seconda che sia positiva o negativa. In particolare, il modello si serve del FWSVM (FW = Feature Weighted), per determinare la label nei dati di train per utilizzare poi il FWKNN nei dati di test, a cui verrà assegnata l'etichetta di classe in base a quella dei k vicini. Dal confronto di tale tecnica sia con il semplice SVM che con l'ibrido SVM-KNN non pesato, essa risulta la più performante, specialmente sul breve periodo.

Ultimi contributi alla letteratura in ambito trading automatico provengono da Geva e Zahavi ([21]) e da Naranjo, Arroyo e Santos ([37]).

Il maggior apporto dei primi consiste nell'introdurre nel modello le notizie testuali

¹¹NASDAQ-100: Contiene le 100 maggiori compagnie dell'indice NASDAQ appartenenti al settore biotecnologico e tecnologico.

del mondo finanziario provenienti da più fonti (riviste, web etc.). Essendo preponderante il problema legato alla non linearità della relazione tra le informazioni in ingresso, viene prediletta la rete neurale come tecnica di data mining, nonostante vengano testati anche albero di decisione e regressione logistica, risultati meno profittevoli. L'esperimento consiste nell'affiancare agli usuali dati di mercato in ingresso, le informazioni testuali sotto cinque diverse rappresentazioni, aggiunte una alla volta al dataset così da analizzare l'efficacia che ognuna di queste apporta al modello originale. Si considerano in successione: 1 - conteggio delle parole, 2 - occorrenze di termini reputati più rilevanti, 3 - classificazione delle notizie in 16 diverse categorie di business, 4 - punteggio compreso tra -1 e +1 del sentimento negativo o positivo che il testo trasmette, 5 - punteggio al punto 4 calibrato manualmente secondo l'influenza che si pensa possa avere sul profitto finale. Considerato l'indice S&P500 (anni 2006 e 2007), fra tutte la struttura più performante risulta quella contenente le rappresentazioni 1, 3 e 5.

Naranjo et al. infine, sviluppano un approccio innovativo volto all'identificazione dei pattern grafici dell'analisi tecnica (*candlestick patterns*), individuando come oggetti di studio tra i più significativi l'Hammer, il Kicking Bullish, e il Piercing Line. La reale novità consiste tuttavia nel modo in cui essi vengono determinati, mediante l'ausilio della logica *fuzzy* (vedi anche [36]), la quale introduce nel modello dell'incertezza nel sistema di decisione di trading, sia in input attraverso l'inserimento di regole del tipo "se-allora" ("se x vera, allora y..."), che in output attribuendo al pattern un grado di affidabilità. Grazie a tale strategia e alle misure di incertezza, è possibile suggerire agli investitori non solo quando ma soprattutto quanto investire. In particolare, sono state sviluppate sui mercati NASDAQ-100 e Euro Stoxx (anni 2009-2013) due diverse tecniche basate su tale logica, la prima meno conservativa in quanto investe l'intero capitale a disposizione, la seconda più oculata. Dal confronto delle due con la classica buy-and-hold, queste sono risultate meno performanti in termini di profitti ottenuti ma meno rischiose, dato evincibile dal minor drawdown. Alla luce di quanto riportato, il lavoro di tesi qui riportato, si trova in linea con le ultime analisi affrontate.

Capitolo 4

Il sistema di trading utilizzato

In questo capitolo verrà descritto per intero il sistema di trading utilizzato per lo svolgimento del lavoro di tesi.

Dopo aver accennato brevemente a software e hardware utilizzati, si riassumeranno le fasi tipiche del KDD (*Knowledge Discovery in Databases*) del sistema, il processo che porta alla conoscenza dei dati e all'estrazione delle informazioni, di cui il data mining, sebbene in molti casi venga scambiato come pseudonimo dell'intero processo, in realtà ne costituisce solo una fase intermedia (figura 4.1).

Software e Hardware utilizzati

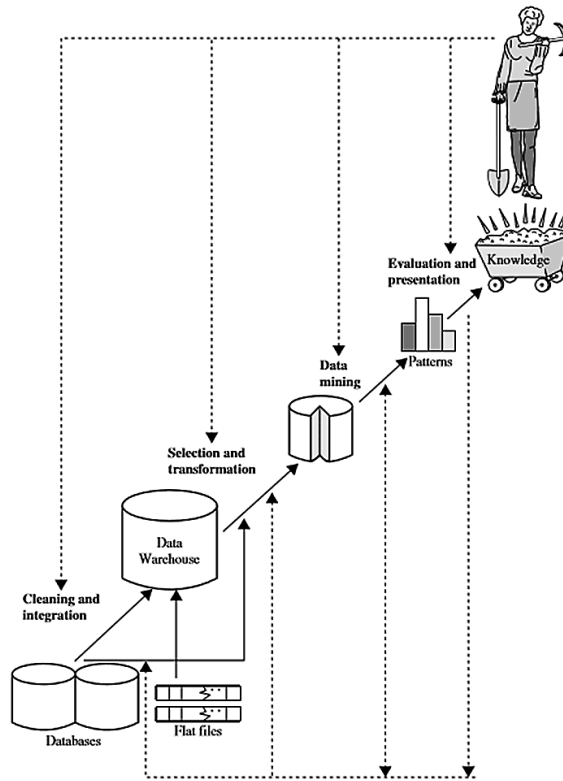
Il sistema è stato sviluppato integralmente mediante l'utilizzo del Software Rapidminer (versione 5.3.015) e di un tool sviluppato nel linguaggio di programmazione C#.

Grazie al primo è stata possibile la creazione del modello di classificazione; esso infatti, grazie ad un'interfaccia grafica, senza bisogno di scrivere del codice, è in grado di sviluppare vari processi tipici del machine learning e del text mining. In esso è stato possibile aggiungere dei plugin utili all'analisi, nello specifico ci si è serviti dell'ulteriore libreria Weka¹.

Il secondo ha permesso invece il download, la fase di pre-processing dei dati da dare poi in input al classificatore, e l'elaborazione finale di quest'ultimi.

Per quanto riguarda l'Hardware, tutti gli esperimenti sono stati effettuati su un computer di processore Intel i5-4210U e RAM 8 GB 2,7 GHz, sistema operativo Windows 10 Home.

¹Weka: Waikato Environment for Knowledge Analysis; estensione di Rapidminer contenente algoritmi e schemi aggiuntivi del machine learning appartenenti all'omonimo Software.

Figura 4.1: Rappresentazione del processo KDD.²

Acquisizione e trasformazione dei dati

L'intero processo ha inizio con il download dei dati utili mediante API di *Yahoo! Finance*; dopo aver definito la data di inizio e di fine della fascia temporale considerata e gli attributi relativi ai prezzi delle azioni, si avrà un dataset in cui ad una data specifica vengono associati i prezzi di apertura, chiusura, minimo e massimo di giornata delle azioni appartenenti al mercato di riferimento. In caso alcune informazioni non fossero disponibili vengono assegnati dei valori null gestite poi nella fase successiva.

Poiché, una volta generato tale file, questo resta invariato per tutte le simulazioni relative alla stessa fascia temporale e allo stesso mercato, il download viene effettuato una volta sola, permettendo così di accelerare gli step successivi riutilizzando le informazioni presenti in memoria.

Tutte le variabili caratterizzanti gli esperimenti vengono settate all'inizio in un file

²Immagine tratta da: *What is data mining?* URL: <https://busy.org/@machinelearning/what-is-data-mining>.

di configurazione chiamato *config.txt*, nel quale vengono definiti anche i path dei file e dei processi coinvolti nel sistema.

In particolare, fissato un giorno in cui si voglia predire il portafoglio di azioni da comprare/vendere, l'intero processo andrà a coinvolgere tre file; qualora si voglia lanciare l'intero processo ciclicamente per più giorni consecutivi, questi saranno aggiornati ad ogni iterazione:

- *prezzi.csv*: file contenente i prezzi di tutti gli asset riferiti ad una determinata finestra di training;
- *list.csv*: file contenente la lista degli asset su cui agirà il classificatore per poter generare la predizione;
- *portfolio.csv*: file contenente il portafoglio delle azioni selezionate come le migliori da commerciare nel giorno considerato.

La seconda fase riguarda sia il *pre-processing* dei dati, in cui questi vengono puliti da eventuali valori nulli e outliers che possono invalidare l'intero processo, che la selezione e la trasformazione degli stessi.

Scelto il giorno di predizione, avendo deciso di utilizzare una finestra di training lunga 10 giorni (variabile *windowTrain*), verranno selezionati tutti i prezzi di chiusura relativi ai dieci giorni precedenti il *giornoDaPredire*: questi andranno a comporre il file *prezzi.csv*. Il file *list.csv* conterrà invece la lista delle azioni che non presentano valori nulli o outliers nel dataset originale; questo può essere modificato prima o dopo il processo di classificazione, nel caso in cui venga applicato o meno lo step di *pattern recognition* (descritto in modo più approfondito nei capitoli 5 e 6), avente l'obiettivo di selezionare tra tutte le azioni, quelle in possesso di determinate caratteristiche.

Durante questa cernita di titoli, a seconda del tipo di filtraggio effettuato, si potranno o selezionare degli asset da non mantenere nell'elenco generale, ed in quel caso andranno a riempire una lista chiamata *stocksToRemove*, per poi essere sottratte dall'insieme originario, oppure verranno scelti quelli da preservare ed inseriti nella lista *stocksToMaintain*, che rimpiazzerà di fatto il file *list.csv*.

Nel dettaglio tale selezione viene effettuata per una azione alla volta, esaminando in relazione ad essa, uno per uno tutti i 61 pattern dell'analisi tecnica considerati. Tale operazione è resa possibile grazie alla libreria open source TA-Lib ([5]), i cui metodi finalizzati allo riconoscimento di un pattern specifico, ricevono in ingresso i prezzi di apertura, chiusura, minimo e massimo relativi all'intervallo di tempo considerato (i 10 gg della finestra di training), e restituiscono in output un vettore *taOut*, costituito dai valori -100, 0 o 100, in corrispondenza del giorno conclusivo di un pattern, slittato di un valore pari alla variabile *outBeginIdx*: in particolare -100 nel caso in cui la figura in questione sia di tipo *bearish* (al ribasso), 100 nel caso opposto *bullish*.

Applicazione del processo di classificazione

Dopo la fase di pre-processing effettuata tramite il tool di C#, avendo generato correttamente i file di input del classificatore, si può eseguire il processo di Rapidminer via linea di comanda ed ottenere così la predizione.

Il processo di Rapidminer, prima ancora di applicare il classificatore scelto, si preoccupa, attraverso una fase di *windowing* sui dati, di creare per ciascuna azione un ulteriore attributo pari al prezzo di chiusura del giorno precedente rispetto a quello che si considera in ciascuna riga del dataset, relativa ad uno dei 10 giorni di training.

Per poter dar vita al processo predittivo è necessario creare per ciascuno di essi, il *label*, ovvero l'etichetta di classe che deve essere predetta. In questo caso il sistema si occupa di generare dei segnali, ovvero di accettare o meno un asset all'interno del portafoglio se il prezzo di chiusura subirà una variazione maggiore dell'1% (in caso di strategia long), o inferiore al -1% (se short); ecco perché l'etichetta di classe, dopo aver discretizzato la variazione percentuale del prezzo in $]-\infty, -1\%],]-1\%, +1\%[, [1\%, +\infty[$, assumerà rispettivamente i valori "-", "=" e "+". Scelto il classificatore (per ciascuno di essi è necessario un differente processo di Rapidminer), questo riceverà in ingresso l'input modificato come sopra; potrà così generare il modello ed effettuare una predizione per ciascuna delle azioni presente in *list.csv*.

I parametri di ciascun classificatore sono quelli assegnati di default dal Software (vedi Tabelle 4.1). In dettaglio ne sono stati presi in considerazione cinque differenti: Naive Bayes, Decision Tree, SVM, Neural Network e AODE.

La predizione viene effettuata applicando il modello al giorno di predizione che non è stato etichettato alla fase precedente. A seconda dei tre valori di confidenza, relativi ciascuno ad un valore di etichetta differente (-, = e +), e a seconda della strategia di investimento scelta (long o short), verranno prese in considerazione nel portafoglio finale, solo le azioni avente come massimo valore di confidenza quello relativo alla classe "+" se la posizione è di tipo long, o "-", se di tipo short.

Elaborazione dei risultati

Dopo aver effettuato la predizione per tutti gli asset, il portafoglio generato per un determinato giorno di predizione, andrà ad aggiornare il file *portafoglio.csv*, fino a completamento del ciclo su tutti i giorni della finestra scelta.

Il tool in C#, potrà così calcolare la variazione percentuale del prezzo di chiusura per ciascuna delle azioni in portafoglio, ed valutare nel calcolo l'inserimento dello stop loss nel caso in cui tale opzione fosse stata abilitata all'inizio nel file di configurazione.

Neural Network		SVM	
Parametro	Valore	Parametro	Valore
training cicle	500	svm type	C-SVC
learning rate	0.3	kernel type	Poly
momentum	0.2	degree	3
error epsilon	$1.0 \cdot 10^{-5}$	gamma	0.0
		coef	0.0
		C	0.0
		epsilon	0.001
		calculate confidence	disabilitato
(a)		(b)	
Decision Tree		AODE	
Parametro	Valore	Parametro	Valore
criterion	gain_ratio	frequency limit	1.0
minimal size for split	4	m-estimate	disabilitato
minimal leaf size	2	weight for m-estimate	1.0
minimal gain	0.1		
maximal depth	20		
confidence	0.25		
(c)		(d)	
Naive Bayes			
Parametro	Valore		
laplace correction	abilitato		
(e)			

Tabella 4.1: Configurazioni di tutti i classificatori.

Capitolo 5

Strategie di integrazione di tecniche di pattern recognition per l'identificazione di figure di analisi tecnica

Nel mondo del trading azionario l'analisi tecnica, grazie ai suoi innumerevoli punti di forza, è oggi uno degli strumenti più diffusi; non solo è di facile applicazione necessitando di pochi dati (reperibili con facilità) per essere messa in atto, ma soprattutto può essere utilizzata su più archi temporali e adattata a diversi strumenti finanziari (obbligazioni, azioni etc.).

Per tale motivo, in questo lavoro di tesi, si è deciso di affiancare al processo di classificazione originale un sistema di *pattern recognition* applicato ai segnali storici, che identificasse e analizzasse i pattern grafici (rialzisti o ribassisti) dell'analisi tecnica; in particolare si prediligono i *candlestick giapponesi*. Quest'ultimi, essendo tra tutti i più intuitivi, permettono di comprendere nell'immediato quale tendenza stanno assumendo i prezzi, e sono in grado di anticipare i cambiamenti del trend, così da favorire i trader nell'individuare il momento più adatto per entrare o uscire dal mercato.

La realizzazione di tale sistema di riconoscimento si attua mediante un filtro il quale si occupa di selezionare l'insieme di azioni su cui applicare il metodo di predizione, seguendo diversi criteri elencati di seguito (per maggiori dettagli vedi Capitolo 4). Su una delle caratteristiche principali delle configurazioni, ovvero che esse predicano un andamento del mercato al rialzo o al ribasso, e sulla possibilità di scegliere tra le due posizioni di investimento (short e long), si fondano le due tipologie di filtro considerate:

- **Pruning di azioni associate ad un pattern in controtendenza.** In tal caso, vi è l'identificazione dei pattern contrari alla direzione dell'investimento prescelto. E' sufficiente che almeno un pattern discorde sia presente all'interno dell'intera finestra di training, affinché l'azione che vi è associata, venga inserita nella lista (*stocksToRemove*) di quelle da non considerare.
- **Scelta di azioni associate ad un pattern concorde.** Contrariamente alla situazione precedente, qui vi è l'individuazione dei pattern concordi alla direzione di investimento; se almeno un pattern di questo tipo è presente, l'asset considerato verrà aggiunto alla lista (*stocksToMaintain*) di quelle su cui continuare l'analisi.

Per semplicità, di seguito si farà riferimento ad esse rispettivamente con "**filtro negativo**" e "**filtro positivo**".

Quest'ultimo è quello su cui ci si è focalizzati maggiormente, ponendo attenzione anche ad altri fattori. Nel caso negativo si è andati alla ricerca di una configurazione senza badare al giorno in cui questa si è verificata; viceversa, in questo caso, si è cercato di capire se la presenza del pattern in giorni differenti influenzasse più o meno positivamente il profitto.

Considerata la finestra di training di 10 giorni, si sono considerate due situazioni possibili: nella prima andiamo alla ricerca di un pattern filtrato positivamente in tutta la finestra; nella seconda ci si preoccupa di riscontrare una specifica configurazione negli ultimi giorni a ridosso di quello in cui si effettua la predizione.

Inoltre, si è tenuto conto anche della misura di *affidabilità* del pattern, già descritta al Capitolo 2. Essa indica la probabilità di successo, ossia di predire correttamente se vi sarà un rialzo o un ribasso del prezzo, e in quanto tale potrebbe influire sul risultato.

Facendo variare queste due variabili: giorno in cui si verifica il pattern e la relativa affidabilità (bassa, media o alta), si sono analizzate in totale sei strategie differenti, che per comodità abbiamo distinto numerandole da 1 a 6. Ricapitolando:

- **Strategia 1:** ricerca i pattern all'ultimo giorno della finestra di training, indipendentemente dall'affidabilità;
- **Strategia 2:** ricerca, tra quelli di affidabilità media o alta, i pattern all'ultimo giorno della finestra di training;
- **Strategia 3:** ricerca, tra quelli di affidabilità alta, i pattern all'ultimo giorno della finestra di training;
- **Strategia 4:** ricerca i pattern in qualsiasi giorno della finestra di training, indipendentemente dall'affidabilità;

- **Strategia 5:** ricerca, tra quelli di affidabilità media o alta, i pattern in qualsiasi giorno della finestra di training;
- **Strategia 6:** ricerca, tra quelli di affidabilità alta, i pattern in qualsiasi giorno della finestra di training.

Gli esperimenti basati sul confronto delle diverse strategie, mirano a comprendere come integrare i pattern grafici all'interno del sistema di trading basato su algoritmi di classificazione, in quali istanti temporali è opportuno analizzare la presenza di pattern grafici e come incide l'affidabilità di questi sulle performance del trading system. I risultati vengono riportati in coda al capitolo seguente.

Capitolo 6

Gli esperimenti

L'obiettivo principale di tale lavoro di tesi, essendo questo la continuazione di uno studio precedente, è stato quello di modificare e/o ampliare il modello di trading già sviluppato, inserendo all'interno elementi nuovi, volti ad estendere il numero di casistiche considerate, e stabilire così quali fossero le strategie più profittevoli.

Ciò che ha portato allo studio di scenari differenti è stata la variazione di più fattori: il mercato considerato (italiano e statunitense), il tipo di classificatore utilizzato nel modello predittivo, la posizione d'investimento scelta (long o short), e l'anno in cui la validità del modello è stata verificata (2011, 2013, 2015); a questi se ne aggiunge un ultimo, a cui si deve il maggiore contributo in questa tesi. Primaria importanza assume, infatti, l'inserimento di un *filtro* (vedi Capitolo 5) di analisi tecnica, il quale ci permette di cernere il set di asset su cui applicare il classificatore, mediante il riconoscimento di vari pattern (elencati nel Capitolo 2).

Altri esperimenti hanno portato all'utilizzo di tale filtro non solo a monte del classificatore, ma anche a posteriori, stabilendo quali, tra gli asset selezionati dal classificatore, dovessero restare all'interno del portafoglio. In un ultimo caso, esso è stato applicato in purezza, senza classificare del tutto.

Sia il contesto di analisi che i risultati degli esperimenti verranno riportati qui di seguito.

6.1 Il mercato

Per poter effettuare degli esperimenti il più vario possibile, si è deciso di operare su due tipi di mercato differente e su diversi anni, in modo da poter analizzare situazioni aventi caratteristiche molto diverse tra loro. Ci si è focalizzati sull'indice italiano FTSE MIB e su quello americano S&P500.

Con indice si intende la media ponderata dei titoli che ne fanno parte, di conseguenza il suo andamento sarà il risultato della media degli andamenti dei titoli azionari che

lo compongono, ognuno dei quali contribuisce in modo proporzionale alla propria capitalizzazione di mercato (le aziende più "grandi" hanno un peso maggiore).

Dell'indice FTSE MIB, acronimo di *Financial Times Stock Exchange Milano Indice di Borsa*, fanno parte le azioni delle 40 maggiori aziende italiane, anche se aventi sede legale all'estero.

Nato il 1 ottobre del 2007 dalla fusione tra Gruppo Borsa Italiana e il Gruppo London Stock Exchange, e operativo dal 1 giugno 2009, rappresenta oggi il maggior indice quotato sull' MTA (Mercato Telematico Azionario) della Borsa Italiana, fig.6.1. Esso forma l'indice aggregato FTSE All Share, insieme ad altri due indici italiani: FTSE Mid Cap e FTSE Micro Cap.



Figura 6.1: La Borsa Italiana.



Figura 6.2: Il New York Stock Exchange. Foto tratta da: Spencer Platt/Getty Images.

L'indice americano S&P500 (abbreviato S&P), fondato nel 1957 dalla società Standard&Poor's, è costituito dalle maggiori 500 aziende americane, quotate all'American Stock Exchange (Amex), al New York Stock Exchange (NYSE, fig. 6.2), e al Nasdaq. Esso, gestito da S& P Dow Jones Indices, ha ormai assunto un'importanza superiore allo stesso Dow Jones, rappresentato dai 30 principali titoli americani.

Così come in altri indici, il peso di ciascun asset viene attribuito in modo proporzionale alla propria capitalizzazione di mercato; ciò che lo rende più solido rispetto ad altri, sono sia i criteri con cui vengono selezionate le aziende appartenenti, che il contributo di ciascun titolo, calcolato tenendo in considerazione solo le azioni del commercio pubblico.

Il National Bureau of Economic Research lo reputa uno dei migliori indicatori del ciclo economico statunitense. Per quanto riguarda i periodi scelti per la validazione del modello, la scelta di tre anni differenti (2011, 2013 e 2015), nasce dall'esigenza di voler analizzare tre andamenti di mercato eterogenei.

Risulta evidente dalle figure 6.3 e 6.6, come il 2011 abbia rivelato una tendenza ribassista per entrambi i mercati, mentre il 2013 ha riportato un trend al rialzo 6.4 e 6.7; nel 2015 è difficile individuare con certezza un trend primario in crescita o in calo (fig. 6.5 e 6.8).

Ultima possibile variazione riguarda la posizione d'investimento scelta per un determinato anno, essa può essere di tipo long o short.

Quando si parla di strategie di trading, tali termini in realtà altro non sono che sinonimi di "acquistare" e "vendere". Erroneamente si crede che in Borsa, si possa realizzare un guadagno acquistando dei titoli, per poi rivenderli ad un prezzo più alto. In verità la strategia appena descritta, detta long (si dice "andare lunghi" in questo caso), non è l'unica possibile.



Figura 6.3: Anno 2011: Andamento ribassista dell'indice FTSE MIB.¹



Figura 6.4: Anno 2013: Andamento rialzista dell'indice FTSE MIB.



Figura 6.5: Anno 2015: Andamento misto dell'indice FTSE MIB.

E' possibile infatti ricavare profitto dalla compravendita di titoli tramite meccanismo di *short selling* (vendita allo scoperto), grazie al quale si vendono dei titoli che non si possiedono realmente ma che ci sono stati dati in prestito ad esempio da un istituto bancario in base alla quotazione di mercato valida al momento del prestito, per poi riacquistarli in un secondo momento e pagare il debito dovuto, se il prezzo sarà sceso, si genererà un profitto dato dalla differenza di prezzo di vendita e riacquisto del titolo stesso.

Questo meccanismo, conosciuto come "andare corto", risulta efficace solo se si punta al ribasso del prezzo del titolo, viceversa infatti, qualora il prezzo dovesse aumentare

¹Immagini tratte da: *FTSEMIB INDEX*, URL:<https://it.tradingview.com/chart/w8tmJ5F6/>

al momento si estinguere il debito, si avrà una perdita.

Andare long o short quindi, significa rispettivamente, scommettere al rialzo o al ribasso di un titolo azionario, in quanto nel primo caso si otterrà un guadagno rivendendo ciò che si possiede ad un prezzo maggiore rispetto al quale lo si è acquistato, nel secondo solo se si verifica una situazione opposta.

Nel nostro caso sono state testate entrambe le ipotesi su tutti gli scenari, diversificando sia il mercato che l'anno. Ci si aspetta che una posizione di tipo short determini un profitto maggiore nel 2011 in cui si è riscontrato un trend ribassista, viceversa per il 2013, in cui si ipotizza un maggior rendimento generato da un investimento di tipo long; per il 2015 non si possono effettuare previsioni certe.



Figura 6.6: Anno 2011: Andamento ribassista dell'indice S&P500.²



Figura 6.7: Anno 2013: Andamento rialzista dell'indice S&P500.



Figura 6.8: Anno 2015: Andamento misto dell'indice S&P500.

6.2 I risultati

In questa sezione si riportano i risultati relativi a tutti gli esperimenti effettuati. Si evidenzia come, per ciascuno di questi considereremo entrambi i mercati, FTSEMIB e S&P500, tutti gli anni, ed entrambe le posizioni di investimento.

Non tutti i classificatori saranno presenti nella trattazione, in quanto i lunghi tempi di elaborazione in certi casi hanno reso quasi impossibile l'avanzamento delle varie simulazioni. Come compromesso tra tempo impiegato e prestazione ottenuta, in alcuni casi ci si è focalizzati solo sull'impiego del classificatore Naive Bayes.

Le diverse prestazioni che l'utilizzo di uno o di un altro classificatore comportano, esulano infatti dallo scopo di questa tesi, in cui ci si è voluti soffermare sul ruolo svolto dal filtro di analisi tecnica.

I valori presenti in tabella esprimono la media giornaliera dei profitti su un intero anno. Nello specifico, fissata la finestra di predizione pari ad un anno specifico, per ogni azione predetta dal classificatore in un determinato giorno, il profitto è stato calcolato in modo differente in base alla posizione di investimento. Se si sceglie di andare long si otterrà un guadagno qualora il prezzo salga, viceversa si registrerà una perdita. Contrariamente al caso short in cui si avrà un profitto nel caso in cui il prezzo cali. Per cui, fissato il prezzo di chiusura di un'azione P_t , si avrà:

$$\text{Profitto}_{long} = \frac{P_t - P_{t-1}}{P_t}$$

$$\text{Profitto}_{short} = \frac{P_{t-1} - P_t}{P_t}$$

²Immagini tratte da: *S&P500*, URL:<https://it.tradingview.com/chart/w8tmJ5F6/>

Infine, per semplicità, si è tenuto conto di uno stop loss fisso all'1% e dei costi di transazione pari allo 0.15% per poter rendere le simulazioni più veritiere. In Appendice vengono riportati per completezza i valori di tutte le simulazioni effettuate, in cui tali variabili vengono fissate una alla volta.

6.2.1 Confronto tra le strategie di pattern recognition

Per tutta la prima fase di test si è tenuto fisso il classificatore Naive Bayes per entrambi i mercati, in quanto dal lavoro di tesi precedente esso era risultato il miglior compromesso tra profitto e tempo di elaborazione.

Come già affermato in precedenza, si è cercato, fin da subito, di capire come l'affiancamento di un sistema di pattern recognition potesse influire sul rendimento finale. Avendo deciso di focalizzarsi fin da subito sul ruolo attivo dell'analisi tecnica, le prime simulazioni si sono concentrate sulla ricerca di pattern concordi alla direzione d'investimento prescelta. Di tali pattern si sono studiate due caratteristiche in dettaglio, la misura di affidabilità e il giorno in cui questi si presentavano nella finestra di training.

Dalla combinazione di tali variabili si sono confrontate le sei diverse strategie, descritte al Capitolo 5. Come possiamo notare dai risultati in tabella 6.1 per quanto riguarda l'indice FTSE MIB e in tabella 6.2 per il S&P500, l'andamento risulta essere coerente su entrambi i mercati.

Strategia	Anno 2011		Anno 2013		Anno 2015	
	Long	Short	Long	Short	Long	Short
1	1,450	1,621	1,630	1,353	1,537	1,475
2	1,791	1,768	1,763	1,455	1,734	1,463
3	2,428	2,185	1,917	1,778	1,975	1,543
4	1,057	1,363	1,372	1,060	1,000	1,324
5	1,097	1,134	1,356	1,055	1,287	1,264
6	1,598	1,451	1,546	1,248	1,569	1,324

Tabella 6.1: FTSE MIB - Confronto strategie applicate al filtro in positivo - Classificatore: Naive Bayes.

Fissando la misura di affidabilità, si andranno a confrontare rispettivamente le strategie 1 e 4, 2 e 5, 3 e 6; ciò implica ricercare i pattern in tutta la finestra di training

Strategia	Anno 2011		Anno 2013		Anno 2015	
	Long	Short	Long	Short	Long	Short
1	1,279	1,689	1,589	1,314	1,331	1,442
2	1,335	1,935	1,512	1,365	1,470	1,483
3	1,726	2,583	1,695	1,579	1,580	1,673
4	1,026	1,363	1,372	1,044	1,149	1,144
5	1,005	1,311	1,248	0,988	1,169	1,107
6	1,217	1,866	1,398	1,132	1,402	1,243

Tabella 6.2: S&P500 - Confronto strategie applicate al filtro in positivo - Classificatore: Naive Bayes.

o solo nel giorno antecedente la predizione. Il giorno determina una netta differenza, le strategie 1, 2 e 3 infatti, mostrano un rendimento medio superiore alle ultime tre. Se invece si confrontano le misure di affidabilità, si nota innanzitutto come queste impattino maggiormente nelle prime tre strategie, in quanto qui si ha una maggior variazione di profitto tra una strategia e l'altra e come, considerando via via pattern sempre più "sicuri", si ottenga un miglior rendimento. Eccezione va fatta per i casi 4 e 5, l'andamento crescente del profitto viene ribaltato in quasi la totalità dei casi. Tra tutte, la strategia numero 3 si rivela nettamente più performante delle altre, determinando profitti notevoli, superando in molti casi il 2%.

Gli esperimenti successivi, per tale motivo, sono stati condotti tenendo fissa tale modalità di selezione dei pattern, al fine di poter mettere a confronto altre variabili.

6.2.2 Strategia 3: confronto tra i classificatori

Stabilite le modalità con cui filtrare il set di azioni da passare in input al classificatore, l'attenzione si è spostata sul ruolo di quest'ultimo.

Per quanto riguarda il mercato italiano, il test è stato effettuato su tutti i cinque diversi processi di classificazione, in quello statunitense, a causa delle grosse dimensioni del dataset iniziali, si è preferito tralasciare le reti neurali a causa degli eccessivi tempi di elaborazione.

Dai risultati in tabella 6.3 e 6.4, si evince come NaiveBayes ed SVM siano i classificatori più promettenti su entrambi i mercati, senza che uno riesca a prevaricare nettamente sull'altro.

Per quanto riguarda l'AODE, a differenza degli altri, i cui risultati restano coerenti

Classificatore	Anno 2011		Anno 2013		Anno 2015	
	Long	Short	Long	Short	Long	Short
Naive Bayes	2,428	2,185	1,917	1,778	1,975	1,543
Decision Tree	1,932	2,041	1,915	1,425	1,868	1,340
SVM	2,387	2,437	2,028	1,766	1,705	1,655
Neural Network	1,516	1,647	1,655	1,600	1,824	1,367
W-AODE	1,870	1,177	1,905	1,404	1,846	1,492

Tabella 6.3: FTSE MIB - Strategia 3 - Confronto tra classificatori.

in tutti gli scenari, questo porta a risultati molto variabili (anche negativi).

Classificatore	Anno 2011		Anno 2013		Anno 2015	
	Long	Short	Long	Short	Long	Short
Naive Bayes	1,726	2,583	1,695	1,579	1,729	1,673
Decision Tree	1,706	2,002	1,521	1,507	1,508	1,405
SVM	1,804	2,624	1,726	1,530	1,721	1,516
W-AODE	1,679	4,984	1,861	0,443	1,674	-1,117

Tabella 6.4: S&P500 - Strategia 3 - Confronto tra classificatori.

6.2.3 Confronto tra le tipologie di filtro

Come ultima fase di test riguardante il ruolo che assume il sistema di pattern recognition affiancato al modello predittivo originale, ci si è posta la questione circa il rapporto tra il classificatore stesso ed il filtro, ovvero se fosse più performante classificare prima sul dataset originale, e poi filtrare secondo i criteri esposti nella strategia 3 tra le azioni predette dal classificatore stesso, o se fosse più redditizio prima ridurre il set di asset per poi dare in input al classificatore le azioni sopravvissute alla selezione.

Come classificatore, per coerenza negli esperimenti, è rimasto fisso Naive Bayes. Dall'analisi, come si evince nelle tabelle 6.5 e 6.6, per entrambi i mercati vi è il

medesimo risultato: applicare il classificatore prima del filtro determina un profitto superiore, sebbene la differenza non sia netta, e in alcuni casi pressoché nulla; possiamo quindi concludere che l'integrazione del filtro a monte o a posteriori è indifferente.

Da tali valori si evince come il ruolo della selezione sia determinante; per questo motivo successivamente ci si è chiesti se, alla fine, si potesse evitare di classificare ed applicare direttamente il filtraggio al set di asset in ingresso al programma.

In realtà, i numeri smentiscono del tutto tale supposizione; i profitti infatti, senza classificare, subiscono un ribasso decisivo.

Modello predittivo	Anno 2011		Anno 2013		Anno 2015	
	Long	Short	Long	Short	Long	Short
Classifico e filtro	2,421	2,189	1,947	1,841	2,006	1,516
Filtro e classifico	2,428	2,185	1,884	1,778	1,975	1,543
Solo filtro	1,940	1,890	1,717	1,452	1,851	1,405

Tabella 6.5: FTSE MIB - Strategia 3 - Confronto tra metodi di modello predittivo.

Modello predittivo	Anno 2011		Anno 2013		Anno 2015	
	Long	Short	Long	Short	Long	Short
Classifico e filtro	1,728	2,587	1,695	1,581	1,728	1,668
Filtro e classifico	1,726	2,583	1,695	1,579	1,729	1,673
Solo filtro	1,479	1,893	1,154	1,083	1,202	1,163

Tabella 6.6: S&P500 - Strategia 3 - Confronto tra metodi di modello predittivo.

Avendo decretato che l'approccio migliore consiste nell'applicare in cascata prima il classificatore e poi il filtro, per completezza, lo stesso approccio è stato applicato al modello con filtro in negativo, lasciando nel portafoglio di selezione, tutte le azioni per cui non fosse presente almeno un pattern contrario alla posizione di investimento. Dai risultati in tabella 6.7 e 6.8, si evince come tale strategia sia, confrontata alle ultime tre, la più fallimentare. Ciò avvalorava maggiormente l'importanza del sistema di pattern recognition, il quale però, senza l'aiuto del classificatore, risulta meno efficace.

Classificatore	Anno 2011		Anno 2013		Anno 2015	
	Long	Short	Long	Short	Long	Short
Negativo	1,013	1,252	1,369	1,146	1,251	1,401
Positivo	2,421	2,189	1,947	1,841	2,006	1,516

Tabella 6.7: FTSE MIB - Strategia 3 - Naive Bayes - Confronto tra filtro in positivo e in negativo, applicato in successione al classificatore.

Classificatore	Anno 2011		Anno 2013		Anno 2015	
	Long	Short	Long	Short	Long	Short
Negativo	1,067	1,502	1,511	1,145	1,242	1,304
Positivo	1,728	2,587	1,695	1,581	1,728	1,668

Tabella 6.8: S&P500 - Strategia 3 - Naive Bayes - Confronto tra filtro in positivo e in negativo, applicato in successione al classificatore.

6.2.4 Simulazione di una strategia di money management

Come riportato al Capitolo 2.5, essendo il capitale una quantità limitata, una buona gestione delle proprie risorse diviene un tassello fondamentale in una strategia di trading.

A tal fine, come ultimo test, si è tentato di simulare in modo piuttosto semplice una politica di money management, utilizzando quanto appreso grazie agli esperimenti precedenti in quanto alla politica di pattern recognition.

Ipotizzando di aver fissato un budget iniziale da investire, obiettivo dell'esperimento è stato, infatti, il confronto fra due diversi investimenti, i quali coinvolgono due scenari già ampiamente descritti. Nel primo si considera come portafoglio, quello generato da un modello predittivo basato sulla strategia n.2, ovvero quella in cui si classifica a partire da asset i cui pattern fossero mediamente o altamente affidabili; nel secondo ci si basa esclusivamente sulla strategia n.3, tenendo in considerazione solo azioni affiancate da pattern con alta affidabilità.

La misura di affidabilità diviene quindi, il criterio per decidere su quali asset venga investire maggiormente e su quali meno.

Nello specifico, nel primo investimento si immagina di investire in misura "doppia" su quegli asset supportati da pattern più affidabili. Seguono i profitti medi calcolati in entrambi gli scenari:

$$P_1 = \frac{1 \times \sum_{media} P + 2 \times \sum_{alta} P}{1 \times \#media + 2 \times \#alta}$$

$$P_2 = \frac{\sum_{alta} P}{\#alta}$$

Per semplicità si considera il caso in cui stop loss e costi di transizione vengono trascurati (posti = 0%).

	Anno 2011		Anno 2013		Anno 2015	
	Long	Short	Long	Short	Long	Short
Strategia 2	2,918	3,351	2,362	2,359	2,191	2,307
Strategia 3	3,700	4,008	2,608	2,746	2,628	2,609
Money Management	3,194	3,592	2,449	2,507	2,355	2,419

Tabella 6.9: FTSE MIB

Come si evince dalle tabelle 6.9 e 6.10, in ambedue i mercati, e in tutti gli anni, la politica di money management adoperata non risulta profittevole quanto la strategia n.3, conviene quindi basarsi solo su pattern altamente affidabili.

	Anno 2011		Anno 2013		Anno 2015	
	Long	Short	Long	Short	Long	Short
Strategia 2	2,167	2,532	1,861	1,840	1,991	2,034
Strategia 3	2,929	3,458	2,133	2,197	2,386	2,478
Money Management	2,435	2,872	1,964	1,967	2,134	2,198

Tabella 6.10: S&P500

Capitolo 7

Conclusioni e sviluppi futuri

7.1 Conclusioni

Grazie all'avvento di Internet, una pratica più di nicchia come quella del trading oggi, grazie alla presenza di piattaforme di trading online, è divenuta sempre più accessibile per gli utenti di tutto il mondo. Questa è una delle tante cause che ha permesso l'interesse del data mining nel settore finanziario, così da poter predire il prezzo di un titolo e di conseguenza di guadagnare dal suo acquisto/vendita.

L'obiettivo di tale tesi è stato quello di integrare ad un sistema automatico di trading già implementato, un sistema di *pattern recognition* basato sull'identificazione dei *candlestick giapponesi* propri dell'analisi tecnica. L'interesse nasce dal fatto che quest'ultima è frequentemente utilizzata nel mondo del trading azionario, in quanto si basa esclusivamente sull'analisi storica dell'andamento dei prezzi, ed è per questo di facile applicazione. Purtroppo, la maggior parte degli studi precedenti si è servita di tale risorsa solo sotto il profilo statistico, adoperando gli indicatori tipici di tale analisi per arricchire il set di attributi su cui verrà addestrato il classificatore.

L'apporto di tale tesi consiste quindi nel distaccarsi dagli studi più comuni in letteratura, esplorando le potenzialità dell'analisi tecnica sotto il profilo grafico. Si è cercato di comprendere come integrarla al meglio all'interno di un sistema precedente basato su algoritmi di classificazione, focalizzandosi sull'impatto che tale inserimento potesse riportare sui risultati, e su quali caratteristiche dovessero possedere i pattern presi in considerazione negli esperimenti, a livello di attendibilità e di istante temporale in cui questi si presentano.

Per rispondere a tali quesiti, si è implementato un sistema di filtraggio, volto a selezionare il set di azioni su cui applicare successivamente il classificatore.

Fissata una strategia di investimento, long o short, e una finestra di training di 10 giorni, si sono effettuati vari esperimenti volti o ad eliminare quelle azioni per cui esistesse almeno un pattern contrario alla strategia di investimento, o a mantenere

quelle per cui fosse presente almeno un pattern concorde a tale strategia. Nel secondo caso si sono analizzati vari scenari, in cui è cambiata sia la misura di affidabilità dei pattern ricercati (variando tra "qualsiasi", "solo media e alta", "solo alta"), che l'istante temporale in cui questi si rivelassero all'interno della finestra (considerando come casi possibili "qualsiasi giorno", "solo nell'ultimo giorno").

Infine, scelto lo scenario più promettente tra i precedenti, si sono confrontate le performance di vari classificatori, chiedendosi infine dove l'inserimento di tale filtro potesse generare un profitto maggiore, se a monte del classificatore o a posteriori.

I risultati sperimentali hanno evidenziato che:

- In generale l'integrazione di uno step di pattern recognition basato sull'analisi tecnica, migliora le prestazioni del sistema;
- La scelta dei pattern grafici usati per pilotare il processo deve limitarsi ad un sottoinsieme di pattern che siano concordi alla direzione di investimento, che abbiano un'affidabilità molto alta, ed è da ricercarsi nei giorni che precedono nell'immediato quello di predizione;
- L'integrazione a monte o a posteriori rispetto alla classificazione non determina sostanziali differenze.

Da un ultimo esperimento, avente come obiettivo la validazione di una politica molto semplice di money management, si sono confrontate due strategie di investimento differenti ipotizzando di possedere in entrambe la stessa quota di capitale iniziale e di attribuire un peso differente in denaro ad azioni legate a pattern altamente affidabili rispetto a quelli mediamente affidabili. In realtà tale strategia è risultata meno conveniente rispetto ad investire tutto su asset connesse alle configurazioni più sicure.

In conclusione, si può affermare come, almeno a livello teorico, il lavoro effettuato abbia portato notevoli miglioramenti al sistema di trading precedente, validando l'importanza che l'analisi dei dati storici ha nei confronti delle predizioni future anche a livello grafico.

7.2 Sviluppi futuri

Il lavoro svolto, sebbene abbia preso in considerazione vari scenari, può essere sicuramente migliorato ed ampliato.

Innanzitutto, un possibile esperimento potrebbe riferirsi alla validazione delle performance di tutti i classificatori considerati, introducendone anche di nuovi tra quelli utilizzati frequentemente in letteratura. Al fine di ottimizzare la fase predittiva, andrebbero considerate diverse configurazioni dei parametri di ciascuno, attribuendo valori diversi da quelli di default.

Si dovrebbe anche prendere in considerazione l'idea di incrementare o diminuire l'ampiezza della finestra di training, al fine di stabilire con più esattezza quanti giorni sono necessari per predire il segnale con massima accuratezza.

Per quanto riguarda l'analisi tecnica, se ne potrebbe sfruttare integralmente la potenzialità, valutando l'ipotesi di inserire all'interno del modello altre features provenienti dallo stesso settore. Oltre ai pattern grafici infatti, il classificatore potrebbe basare la propria fase di apprendimento non solo sui prezzi storici, ma anche sugli indicatori di impronta statistica già considerati in molti studi antecedenti a questo. Una simile mutazione potrebbe generare un modello ancora più accurato e profittevole.

Una grande innovazione potrebbe essere infine, quella di creare un sistema di trading che rispecchi il più possibile il reale mondo del trading. Oltre al concetto di stop loss, volto a chiudere una posizione prima che questa generi una perdita eccessiva, sarebbe interessante considerare all'interno del modello anche quello del *take profit* che, viceversa, chiude un'operazione in guadagno; è grazie ad entrambi che la redditività della propria strategia di trading può essere stabilizzata e massimizzata.

Concludendo, si potrebbe simulare il sistema fissando all'inizio sia un budget d'investimento che una politica di money management; stop loss e take profit ([46]) potrebbero essere adottati in modo dinamico (nel cui caso vengono detti *trailing stop*), ovvero cambiandone il valore continuamente in relazione ai guadagni o alle perdite conseguiti fino a quel momento ed in base alla variazione dei prezzi.

Appendice

In appendice si riportano tutti i risultati inerenti agli esperimenti discussi al Capitolo 6.2. Oltre ai confronti già evidenziati precedentemente, è possibile notare come cambino i profitti qualora si inserisca o meno lo stop loss (fissato all'1%), e qualora si considerino o meno i costi di transazione della singola operazione (pari allo 0.15%). Si evidenzia come l'inserimento di uno stop loss nella maggior parte dei casi renda il trading meno profittevole; il rischio complessivo ne trae però giovamento, in quanto la strategia risulta più conservativa.

Strategia	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
1	Long	2,457	2,307	1,600	1,450
	Short	2,924	2,774	1,771	1,621
2	Long	2,918	2,768	1,941	1,791
	Short	3,351	3,201	1,918	1,768
3	Long	3,700	3,550	2,578	2,428
	Short	4,008	3,858	2,335	2,185
4	Long	1,840	1,690	1,207	1,057
	Short	2,502	2,352	1,513	1,363
5	Long	1,891	1,741	1,247	1,097
	Short	2,085	1,935	1,284	1,134
6	Long	2,678	2,528	1,748	1,598
	Short	2,815	2,665	1,601	1,451

Tabella 7.1: Confronto strategie - FTSE MIB - Anno 2011

Strategia	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
1	Long	2,133	1,983	1,780	1,630
	Short	2,089	1,939	1,503	1,353
2	Long	2,362	2,212	1,913	1,763
	Short	2,359	2,209	1,605	1,455
3	Long	2,608	2,458	2,067	1,917
	Short	2,746	2,596	1,928	1,778
4	Long	1,817	1,667	1,522	1,372
	Short	1,672	1,522	1,210	1,060
5	Long	1,840	1,690	1,506	1,356
	Short	1,699	1,549	1,205	1,055
6	Long	2,092	1,942	1,696	1,546
	Short	1,926	1,776	1,398	1,248

Tabella 7.2: Confronto strategie - FTSE MIB - Anno 2013

Strategia	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
1	Long	1,834	1,684	1,687	1,537
	Short	2,155	2,005	1,625	1,475
2	Long	2,191	2,041	1,884	1,734
	Short	2,307	2,157	1,613	1,463
3	Long	2,628	2,478	2,125	1,975
	Short	2,609	2,459	1,693	1,543
4	Long	1,480	1,330	1,406	1,000
	Short	0,848	0,698	1,474	1,324
5	Long	1,585	1,435	1,437	1,287
	Short	0,889	0,739	1,414	1,264
6	Long	1,996	1,846	1,719	1,569
	Short	1,674	1,524	1,474	1,324

Tabella 7.3: Confronto strategie - FTSE MIB - Anno 2015

Strategia	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
1	Long	1,896	1,759	1,427	1,279
	Short	2,252	2,107	1,837	1,689
2	Long	2,167	2,029	1,484	1,335
	Short	2,532	2,383	2,085	1,935
3	Long	2,929	2,780	1,876	1,726
	Short	3,458	3,309	2,733	2,583
4	Long	1,484	1,338	1,175	1,026
	Short	1,636	1,491	1,453	1,304
5	Long	1,505	1,358	1,154	1,005
	Short	1,618	1,470	1,460	1,311
6	Long	1,993	1,845	1,367	1,217
	Short	2,343	2,196	2,014	1,866

Tabella 7.4: Confronto strategie - S&P500 - Anno 2011

Strategia	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
1	Long	1,888	1,743	1,737	1,589
	Short	1,731	1,585	1,461	1,314
2	Long	1,861	1,714	1,661	1,512
	Short	1,840	1,690	1,515	1,365
3	Long	2,133	1,983	1,845	1,695
	Short	2,197	2,047	1,729	1,579
4	Long	1,645	1,499	1,520	1,372
	Short	1,350	1,204	1,191	1,044
5	Long	1,540	1,392	1,397	1,248
	Short	1,304	1,157	1,137	0,988
6	Long	1,751	1,603	1,547	1,398
	Short	1,525	1,377	1,282	1,132

Tabella 7.5: Confronto strategie - S&P500 - Anno 2013

Strategia	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
1	Long	1,769	1,623	1,479	1,331
	Short	1,935	1,788	1,591	1,442
2	Long	1,991	1,843	1,619	1,470
	Short	2,034	1,885	1,633	1,483
3	Long	2,386	2,236	1,730	1,580
	Short	2,478	2,329	1,823	1,673
4	Long	1,522	1,376	1,298	1,149
	Short	1,520	1,373	1,293	1,144
5	Long	1,559	1,412	1,318	1,169
	Short	1,458	1,310	1,256	1,107
6	Long	1,918	1,770	1,552	1,402
	Short	1,788	1,640	1,393	1,243

Tabella 7.6: Confronto strategie - S&P500 - Anno 2015

Classificatore	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
Naive Bayes	Long	3,700	3,550	2,578	2,428
	Short	4,008	3,858	2,335	2,185
Decision Tree	Long	3,557	3,407	2,082	1,932
	Short	3,381	3,231	2,191	2,041
SVM	Long	3,907	3,757	2,537	2,387
	Short	4,461	4,311	2,587	2,437
Neural Network	Long	2,846	2,696	1,666	1,516
	Short	2,803	2,653	1,797	1,647
W-AODE	Long	3,458	3,308	2,020	1,870
	Short	3,898	3,748	1,327	1,177

Tabella 7.7: Strategia 3 - Confronto Classificatori - FTSE MIB - Anno 2011

Classificatore	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
Naive Bayes	Long	2,608	2,458	2,067	1,917
	Short	2,746	2,596	1,928	1,778
Decision Tree	Long	2,708	2,558	2,065	1,915
	Short	2,266	2,116	1,575	1,425
SVM	Long	2,769	2,619	2,178	2,028
	Short	2,915	2,765	1,916	1,766
Neural Network	Long	2,429	2,279	1,805	1,655
	Short	2,212	2,062	1,750	1,600
W-AODE	Long	2,730	2,580	2,055	1,905
	Short	2,555	2,405	1,554	1,404

Tabella 7.8: Strategia 3 - Confronto Classificatori - FTSE MIB - Anno 2013

Classificatore	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
Naive Bayes	Long	2,628	2,478	2,125	1,975
	Short	2,609	2,459	1,693	1,543
Decision Tree	Long	2,528	2,378	2,018	1,868
	Short	2,319	2,171	1,490	1,340
SVM	Long	2,472	2,322	1,855	1,705
	Short	2,746	2,596	1,805	1,655
Neural Network	Long	2,538	2,388	1,974	1,824
	Short	2,216	2,066	1,517	1,367
W-AODE	Long	2,500	2,350	1,996	1,846
	Short	2,819	2,669	1,642	1,492

Tabella 7.9: Strategia 3 - Confronto Classificatori - FTSE MIB - Anno 2015

Classificatore	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
Naive Bayes	Long	2,628	2,478	2,125	1,975
	Short	2,609	2,459	1,693	1,543
Decision Tree	Long	2,528	2,378	2,018	1,868
	Short	2,319	2,171	1,490	1,340
SVM	Long	2,472	2,322	1,855	1,705
	Short	2,746	2,596	1,805	1,655
Neural Network	Long	2,538	2,388	1,974	1,824
	Short	2,216	2,066	1,517	1,367
W-AODE	Long	2,500	2,350	1,996	1,846
	Short	2,819	2,669	1,642	1,492

Tabella 7.10: Strategia 3 - Confronto Classificatori - FTSE MIB - Anno 2015

Anno	Classificatore	Posizione	Stop loss = 0%		Stop loss = 1%	
			Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
2011	W-AODE 3	Long	3,441	3,291	1,709	1,664
		Short	3,891	3,741	1,687	1,537
	W-AODE 5	Long	3,458	3,308	2,020	1,870
		Short	3,898	3,748	1,327	1,177
2013	W-AODE 3	Long	2,699	2,549	1,928	1,778
		Short	2,493	2,343	1,353	1,203
	W-AODE 5	Long	2,730	2,580	2,055	1,905
		Short	2,555	2,405	1,554	1,404
2015	W-AODE 3	Long	2,532	2,382	1,943	1,793
		Short	2,876	2,726	1,632	1,482
	W-AODE 5	Long	2,500	2,350	1,996	1,846
		Short	2,819	2,669	1,642	1,492

Tabella 7.11: Strategia 3 - Confronto tipi di discretizzazione Waode - FTSE MIB

Classificatore	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
Naive Bayes	Long	2,929	2,780	1,876	1,726
	Short	3,458	3,309	2,733	2,583
Decision Tree	Long	2,861	2,712	1,856	1,706
	Short	2,782	2,632	2,152	2,002
SVM	Long	2,954	2,806	1,954	1,804
	Short	3,587	3,437	2,774	2,624
W-AODE	Long	2,732	2,584	1,829	1,679
	Short	5,280	5,130	5,134	4,984

Tabella 7.12: Strategia 3 - Confronto Classificatori - S&P500 - Anno 2011

Classificatore	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
Naive Bayes	Long	2,133	1,983	1,845	1,695
	Short	2,197	2,047	1,729	1,579
Decision Tree	Long	1,919	1,769	1,671	1,521
	Short	1,919	1,769	1,657	1,507
SVM	Long	2,237	2,087	1,876	1,726
	Short	2,241	2,091	1,680	1,530
W-AODE	Long	2,285	2,135	2,011	1,861
	Short	1,420	1,270	0,593	0,443

Tabella 7.13: Strategia 3 - Confronto Classificatori - S&P500 - Anno 2013

Classificatore	Posizione	Stop loss = 0%		Stop loss = 1%	
		Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
Naive Bayes	Long	2,386	2,236	1,879	1,729
	Short	2,478	2,329	1,823	1,673
Decision Tree	Long	2,136	1,987	1,658	1,508
	Short	2,127	1,979	1,555	1,405
SVM	Long	2,436	2,289	1,871	1,721
	Short	2,459	2,310	1,666	1,516
W-AODE	Long	2,275	2,126	1,824	1,674
	Short	2,297	2,147	-0,967	-1,117

Tabella 7.14: Strategia 3 - Confronto Classificatori - S&P500 - Anno 2015

Anno	Filtro	Posizione	Stop loss = 0%		Stop loss = 1%	
			Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
2011	Negativo	Long	2,068	1,918	1,163	1,013
		Short	2,398	2,248	1,402	1,252
	Positivo	Long	3,710	3,560	2,571	2,421
		Short	4,030	3,880	2,339	2,189
2013	Negativo	Long	1,894	1,744	1,519	1,369
		Short	1,981	1,831	1,296	1,146
	Positivo	Long	2,641	2,491	2,101	1,947
		Short	2,780	2,630	1,991	1,841
2015	Negativo	Long	1,431	1,281	1,401	1,251
		Short	2,001	1,851	1,551	1,401
	Positivo	Long	2,654	2,504	2,156	2,006
		Short	2,591	2,441	1,666	1,516

Tabella 7.15: FTSE MIB - Strategia 3 - Confronto tra tipi di filtraggio, applicati in seguito a classificazione.

Anno	Filtro	Posizione	Stop loss = 0%		Stop loss = 1%	
			Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
2011	Negativo	Long	1,738	1,592	1,216	1,067
		Short	2,068	1,918	1,652	1,502
	Positivo	Long	2,917	2,767	1,878	1,728
		Short	3,455	3,305	2,737	2,587
2013	Negativo	Long	1,824	1,677	1,659	1,511
		Short	1,555	1,405	1,295	1,145
	Positivo	Long	2,133	1,983	1,845	1,695
		Short	2,195	2,045	1,731	1,581
2015	Negativo	Long	1,698	1,551	1,390	1,242
		Short	1,783	1,634	1,453	1,304
	Positivo	Long	2,380	2,230	1,878	1,728
		Short	2,455	2,305	1,818	1,668

Tabella 7.16: S&P500 - Strategia 3 - Confronto tra tipi di filtraggio, applicati in seguito a classificazione.

Anno	Modello predittivo	Posizione	Stop loss = 0%		Stop loss = 1%	
			Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
2011	Classifico e filtro	Long	3,710	3,560	2,571	2,421
		Short	4,030	3,880	2,339	2,189
	Filtro e classifico	Long	3,700	3,550	2,578	2,428
		Short	4,008	3,858	2,335	2,185
	Solo filtro	Long	3,141	2,991	2,090	1,940
		Short	3,270	3,120	2,040	1,890
2013	Classifico e filtro	Long	2,641	2,491	2,101	1,951
		Short	2,780	2,630	1,991	1,841
	Filtro e classifico	Long	2,608	2,458	2,067	1,884
		Short	2,746	2,596	1,928	1,778
	Solo filtro	Long	2,325	2,175	1,867	1,717
		Short	2,215	2,065	1,602	1,452
2015	Classifico e filtro	Long	2,654	2,504	2,156	2,006
		Short	2,591	2,441	1,666	1,516
	Filtro e classifico	Long	2,816	2,666	2,125	1,975
		Short	2,609	2,459	1,693	1,543
	Solo filtro	Long	2,467	2,317	2,001	1,851
		Short	2,272	2,122	1,555	1,405

Tabella 7.17: FTSE MIB - Strategia 3 - Tipi di modelli: cambia il ruolo del filtro in positivo.

Anno	Modello predittivo	Posizione	Stop loss = 0%		Stop loss = 1%	
			Ct = 0%	Ct = 0.15%	Ct = 0%	Ct = 0.15%
2011	Classifico e filtro	Long	2,917	2,767	1,878	1,728
		Short	3,455	3,305	2,737	2,587
	Filtro e classifico	Long	2,929	2,780	1,876	1,726
		Short	3,458	3,309	2,733	2,583
	Solo filtro	Long	2,397	2,250	1,629	1,479
		Short	2,582	2,434	2,043	1,893
2013	Classifico e filtro	Long	2,133	1,983	1,845	1,695
		Short	2,195	2,045	1,731	1,581
	Filtro e classifico	Long	2,133	1,983	1,845	1,695
		Short	2,197	2,047	1,729	1,579
	Solo filtro	Long	1,541	1,392	1,304	1,154
		Short	1,511	1,361	1,233	1,083
2015	Classifico e filtro	Long	2,380	2,230	1,878	1,728
		Short	2,455	2,305	1,818	1,668
	Filtro e classifico	Long	2,386	2,236	1,879	1,729
		Short	2,478	2,329	1,823	1,673
	Solo filtro	Long	1,735	1,587	1,352	1,202
		Short	1,718	1,570	1,313	1,163

Tabella 7.18: S&P500 - Strategia 3 - Tipi di modelli: cambia il ruolo del filtro in positivo.

Bibliografia

- [1] *I tre pilastri dell'analisi tecnica*, URL: <https://www.soldionline.it/guide/analisi-tecnica/i-tre-pilastri-dell-analisi-tecnica>.
- [2] John J. Murphy, *Analisi tecnica dei mercati finanziari*, Ulrico Hoepli Editore, 2002.
- [3] Martin J. Pring, *Analisi tecnica dei mercati finanziari*, McGraw-Hill, 1995.
- [4] *Trading forex*, URL: <https://www.blogfinanza.com/forex-online>.
- [5] *TA-Lib :Technical Analysis Library*, URL: <http://ta-lib.org/>.
- [6] *Pattern grafici a candela*, URL: <https://it.investing.com/technical/candlestick-patterns>.
- [7] *Il significato di trading*, URL: <https://www.mercati24.com/significato-di-trading>.
- [8] *Formazione di trading online*, URL: http://www.traderpedia.it/wiki/index.php/Position_sizing
- [9] *Formazione di trading*, URL: <https://www.avatrade.it/formazione>.
- [10] *Cos'è il money management, guida alla gestione del rischio*, URL: <https://www.money.it/Cos-e-il-Money-Management-Guida>.
- [11] P. P. Tan, M. Steinbach e V. V. Kumar. *Introduction to Data Mining*, Pearson Addison-Wesley, 2006.
- [12] *FTSE MIB*, URL: https://it.wikipedia.org/wiki/FTSE_MIB.
- [13] *Cos'è il FTSE MIB*, URL: <http://www.dirittodicritica.com/2011/12/09/ftse-mib-definizione-significato-economia-31893/>.
- [14] *S&P500*, URL: https://it.wikipedia.org/wiki/S%26P_500.
- [15] *Posizioni long e short*, URL: <http://www.valoreazioni.com/approfondimenti-finanziari/posizioni-long-e-short.html>.

- [16] *Automated Trading Systems: The Pros and Cons*, URL: <https://www.investopedia.com/articles/trading/11/automated-trading-systems.asp>
- [17] J. Han, M. Kamber, J. Pei, *Data Mining, concepts and techniques*, Morgan kaufmann, 2012.
- [18] E. Baralis, L. Cagliero, T. Cerquitelli, P. Garza, F. Pulvirenti, "Discovering profitable stocks for intraday trading", *Inf. Sci.* 405 (2017) 91-106.
- [19] T.I. Chen, F.y. Chen, "An intelligent pattern recognition model for supporting investment decisions in stock market", *Inf. Sci.* 346 (2016) 261-274.
- [20] W.C. Chiang, D. Enke, T. Wu, R. Wang, "An adaptive stock index trading decision support system", *Expert Syst. Appl.* 59 (2016) 195-207.
- [21] T. Geva, J. Zahavi, "Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news", *Decis. Support Syst.* 57 (2014) 212-223.
- [22] J. Patel, S. Shah, P. Thakkar, K. Kotecha, "Predicting stock market index using fusion of machine learning techniques", *Expert Syst. Appl.* 42 (2015) 2162-2172.
- [23] K. Zbikowski, "Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy", *Expert Syst. Appl.* 42 (2015) 1797-1805 .
- [24] L.A. Teixeira, A.L.I. De Oliveira, "A method for automatic stock trading combining technical analysis and nearest neighbor classification", *Expert Syst. Appl.* 37 (2010) 6885-6890.
- [25] M.M. Mostafa, "Forecasting stock exchange movements using neural networks: empirical evidence from Kuwait", *Expert Syst. Appl.* 37 (2010) 6302-6309.
- [26] Y. Kara, M.A. Boyacioglu, O.K. Bayka, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the istanbul stock exchange", *Expert Syst. Appl.* 38 (2011) 5311-5319.
- [27] M. Qiu, Yu. Song, F. Akagi, "Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market", *Chaos, Sol. and Frac.* 85 (2016) 1-7.
- [28] P. Parracho, R. Neves, N. Horta, "Trading in Financial Markets using Pattern Recognition Optimized by Genetic Algorithms", *GECCO '10*, 2105-2106.

- [29] J.L. Wanga, S.H. Chan, "Trading rule discovery in the US stock market: An empirical study", *Expert Syst. Appl.* 36 (2009) 5450-5455.
- [30] P.C. Chang, C.H. Liu, C.Y. Fan, J.L. Lin, C.M. Lai, "Ensemble of Neural Networks for Stock Trading Decision Making", *LNAI* 5755 (2009) 1-10.
- [31] Y. Chen, Y.Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction", *Expert Syst. Appl.* 80 (2017) 340-355.
- [32] M.J. Kim, S.H. Min, I. Han, "An evolutionary approach to the combination of multiple classifiers to predict a stock price index", *Expert Syst. Appl.* 31 (2006) 241-247.
- [33] Z. Wang, J.J. Wang, Z.G. Zhang, S.P. Guo, "Forecasting stock indices with back propagation neural network", *Expert Syst. Appl.* 38 (2011) 14346-14355.
- [34] L. Wang, Z. Wang, S. Zhao, S. Tan, "Stock market trend prediction using dynamical Bayesian factor graph", *Expert Syst. Appl.* 42 (2015) 6267-6275.
- [35] C.F. Tsai, Y.C. Lin, D.C. Yen, Y.M. Chen, "Predicting stock returns by classifier ensembles", *Apply Soft Comp.* 11 (2011) 2452-2459.
- [36] M. Pulido, P. Melin, O. Castillo, "Particle swarm optimization of ensemble neural networks with fuzzy aggregation for time series prediction of the mexican stock exchange", *Inf. Sci.* 280 (2014) 188-204.
- [37] R. Naranjo, J. Arroyo, M. Santos, "Fuzzy modeling of stock trading with fuzzy candlesticks", *Expert Syst. Appl.* 93 (2018) 15-27.
- [38] Y.K. Kwon, B.R. Moon, "A Hybrid Neurogenetic Approach for Stock Forecasting", *IEEE Trans. on NNs* 18 (2007) 851-864.
- [39] W. Leigh, R. Purvis, J.M. Ragusa, "Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support", *Decis. Support Syst.* 32 (2002) 361-377.
- [40] H. Wang, A.S. Weigend, "Data Mining for Financial Decision Making", *Decis. Support Syst.* 32 (2002) 417-418.
- [41] H. White, "Economic prediction using neural networks: A case of IBM daily stock returns", *IEEE Inter. Conf. on NNs* 2 (1988) 451-458.
- [42] *4 strategie di trading attivo*, URL: <https://it.insider.pro/investment/2017-04-14/4-strategie-di-trading-attivo/>.

- [43] *Strategie di trading*, URL:<https://www.money.it/+Strategie-di-Trading-+>.
- [44] *I 5 migliori indicatori di analisi tecnica per il trading*, URL:<https://www.money.it/I-5-migliori-indicatori-trading-analisi-tecnica>.
- [45] *Come sfruttare l'analisi ciclica*, URL: <https://www.giacomobruno.it/come-sfruttare-l%E2%80%99analisi-ciclica/>.
- [46] *Stop-loss e Take-Profit*, URL: <https://www.soldionline.it/guide/analisi-tecnica/stop-loss-e-take-profit>.