

POLITECNICO DI TORINO

Facoltà di Ingegneria

Corso di Laurea in Ingegneria Informatica (Computer Engineering)

Tesi di Laurea Magistrale

Analisi di strategie di investimento con tecniche di data mining



Relatori

Prof. Luca Cagliero

Prof. Paolo Garza

Candidato

Aimar Luca

Aprile 2018

Indice

1.	Il mercato finanziario.....	1
1.1.	Le azioni.....	1
1.2.	La borsa e gli indici	2
1.3.	Le obbligazioni.....	4
1.4.	I derivati.....	5
1.4.1.	I futures.....	5
1.4.2.	I certificati.....	6
1.4.3.	I contratti per differenza	6
1.4.4.	Le opzioni.....	6
2.	Data mining	9
2.1.	KDD	9
2.2.	Tecniche di data mining.....	11
3.	La regressione.....	13
3.1.	Regressione Lineare Semplice	13
3.2.	Regressione Lineare Multipla	15
3.3.	Support Vector Machine	16
3.4.	Gradient Boosting Trees	21
4.	Operare in borsa.....	23
4.1.	Posizioni long.....	24
4.2.	Posizioni short	24
4.3.	Stop loss.....	25
4.4.	Stop profit.....	25
4.5.	Take profit	26
4.6.	Orizzonte temporale delle operazioni.....	27
4.6.1.	Scalping.....	27
4.6.2.	Day trading	27
4.6.3.	Swing Trading	27
4.6.4.	Position Trading o Buy & Hold.....	28
5.	Il progetto	29
5.1.	Confronto con lo stato dell'arte	29
5.2.	L'obiettivo della tesi	31

5.3.	Crawling dei dati azionari	32
5.3.1.	Fonti.....	32
5.3.2.	Yahoo! Finance API.....	32
5.3.3.	Google Finance API.....	33
5.4.	Rapidminer Studio	34
5.5.	Gli indicatori	34
5.5.1.	Moving Average (MA).....	35
5.5.2.	Exponential Moving Average (EMA).....	35
5.5.3.	On Balance Volume (OBV)	36
5.5.4.	Average True Range (ATR).....	37
5.5.5.	Relative Strenght Index (RSI)	37
5.5.6.	Money Flow Index (MFI).....	38
5.5.7.	Aroon Oscillator	38
5.5.8.	Commodity Channel Index (CCI):.....	39
5.6.	Preparazione dei dati.....	40
5.6.1.	Windowing.....	40
5.7.	I test considerati	42
5.7.1.	Modello Chiusura, Volume, Indicatore.....	45
5.7.2.	Modello Test 2	46
5.7.3.	Modello Chiusura, Volume, Indicatore_X - Chiusura, Volume, Indicatore_Y.....	47
5.7.4.	Modello Chiusura, Indicatore_X – Chiusura, Indicatore_Y.....	47
5.7.5.	Modello unico.....	47
6.	Analisi dei risultati	49
6.1.	Utilizzo della soglia	49
6.2.	Risultati.....	50
6.2.1.	Modello Chiusura, Volume, Indicatore_X – USA 2015 long	50
6.2.2.	Modello Chiusura, Indicatore_X – Chiusura, Indicatore_Y – USA 2011 short	56
6.2.3.	Modello unico.....	61
6.3.	Analisi dell’utilizzo della soglia	67
6.3.1.	Scelta della soglia percentuale più opportuna	69
6.4.	Analisi dei migliori indicatori nei test ensemble	71
6.5.	Significatività statistica	79
7.	Conclusioni	85
8.	Bibliografia.....	87

1. Il mercato finanziario

Per mercato finanziario si intende il luogo in cui gli strumenti finanziari vengono scambiati. Grazie alla diffusione del sistema telematico di negoziazione non è più necessaria l'esistenza di un luogo fisico in cui poter effettuare gli scambi, quindi con il termine luogo si comprendono, in senso più ampio, anche tutte le transazioni effettuate attraverso i broker online. Questi svolgono un compito di intermediazione tra l'acquirente e il venditore, ed il loro dovere è quello di assicurare al compratore il miglior prezzo disponibile sul mercato del bene finanziario in questione.

Gli strumenti finanziari principali sono:

- le azioni;
- le obbligazioni;
- i derivati:
 - i contratti futures;
 - i certificati;
 - i CFD;
 - le opzioni.

1.1. Le azioni

L'azione è uno strumento finanziario che costituisce una frazione percentuale di una società, la quale può essere costituita in forma di società per azioni, accomandita per azioni o società cooperative per azioni. Questi titoli azionari rappresentano quote di partecipazione nella società in questione, sottoscrivendo le quali l'investitore acquisisce lo status di socio dell'emittente.

Le azioni sono indivisibili, di uguale valore ed attribuiscono ai proprietari uguali diritti. Questi si dividono in amministrativi, come il diritto di voto, ed economici, come il diritto al dividendo in caso di utili. Il luogo dove sono negoziati i titoli azionari è il mercato azionario.

Il mercato azionario si divide poi in altri due segmenti:

- il mercato primario;
- il mercato secondario.

Il mercato primario è dedicato ai momenti in cui una società si quota in borsa. Ciò avviene attraverso quella che è chiamata offerta pubblica iniziale, detta IPO, dall'inglese *initial public offering*, il cui obiettivo è quello di vendere una parte della società a terzi per raccogliere nuovi capitali utili alla crescita dell'azienda per poter così finanziare futuri investimenti.

Esempi di IPO recenti sono stati:

- Facebook, 18 maggio 2012;
- Alibaba Holdings Group, 18 settembre 2014;
- Ferrari, 21 ottobre 2015.

Il mercato secondario è invece il luogo in cui i titoli già circolanti vengono scambiati. Una volta terminato il primo periodo, in cui una società viene negoziata sul mercato primario, l'azione viene trasferita sul secondario.

1.2. La borsa e gli indici

Ogni paese ha almeno una borsa valori dedicata alla compravendita di azioni delle società che si sono volute quotare in tale paese.

Le dieci maggiori borse valori per capitalizzazione di mercato sono le seguenti:

MARKET	PAESE	FINE 2016 (\$TN)
NYSE	USA	19.573
Nasdaq	USA	7.779
Japan Exchange Group	GIAPPONE	5.062
Shanghai SE	CINA	4.104
London SE Group	UK	3.496
Shenzhen SE	CINA	3.217
Euronext	EUROPA	3.493
Hong Kong Exchanges	HONG KONG	3.193
Deutsche Börse	GERMANIA	1.732
TMX Group	CANADA	2.042
		53.691

Alla fine del 2016 le maggiori borse mondiali capitalizzavano quasi cinquantaquattro trilioni di dollari, mentre si stima che ad oggi la capitalizzazione mondiale totale sia di settantasei trilioni di dollari, il massimo mai registrato nella storia del mercato azionario.

La capitalizzazione di una società si ottiene dal prodotto tra il numero totale di azioni in circolazione per il prezzo di ogni azione. Questo valore indica il valore sul mercato dell'azienda.

Per capitalizzazione mondiale si intende quindi la somma di tutte le capitalizzazioni societarie presenti sulle borse del mondo.

Quando azioni di diverse società vengono tracciate insieme si forma quello che viene chiamato un indice. Questo viene creato per poter misurare l'andamento di uno specifico settore di mercato azionario, ad esempio:

- S&P 500: contiene le cinquecento aziende con capitalizzazione più grande della borsa di New York;
- S&P MidCap: contiene solo le aziende la cui capitalizzazione è compresa tra 750 milioni di dollari e 3 miliardi di dollari;
- FTSE MIB: contiene le quaranta aziende italiane, anche aventi sede legale all'estero, con capitalizzazione più grande della borsa di Milano.

L'andamento di un indice è quindi legato all'andamento delle azioni che lo compongono, ma lo si può calcolare in diversi modi:

- Media non pesata: si sommano i rendimenti delle singole azioni e si divide per il numero totale di azioni;
- Media pesata: in base ad un fattore, che spesso è la capitalizzazione, si dà più o meno importanza alle società che compongono l'indice;
- Media pesata con tetto massimo: simile al metodo della media pesata, introduce però un peso massimo che un'azienda può avere nell'indice. In questo modo si evitano situazioni troppo sbilanciate che portano l'indice a seguire l'andamento di poche società importanti.

Si può investire direttamente su un indice e non solo sulle singole azioni delle società. In questo modo si riesce a distribuire implicitamente il rischio su diversi titoli con un solo investimento. Le componenti degli indici variano con il tempo, a causa del cambiamento delle condizioni economiche delle società che lo compongono, e ciò permette di avere nel proprio investimento sempre le migliori aziende sul mercato per un determinato settore.

Entrambi gli indici analizzati usano una media pesata per calcolare il loro andamento. In particolare le dieci componenti più importanti del FTSE MIB sono:

SOCIETA'	PESO (%)
Enel	11.68
Intesa San Paolo	10.99
Unicredit	10.16
Eni	10.11
Generali	5.83
Fiat Chrysler Automobiles NV	4.77
Atlantia	4.1
STMicroelectronics	3.38
Ferrari NV	3.22
Cnh Industrial	3.14
	67.39

Le principali dieci componenti del S&P 500 invece sono:

SOCIETA'	PESO (%)
Apple Inc.	3.84
Microsoft Corp.	2.87
Amazon.com Inc.	2.06
Facebook Inc.	1.87
Berkshire Hathaway Inc.	1.64
JPMorgan Chase & Co.	1.62
Johnson & Johnson	1.62
Exxon Mobil Corporation	1.56
Alphabet Inc. Class C	1.39
Alphabet Inc. Class A	1.38
	19.85

Il 25% delle aziende quindi spiegano il 67.39% dell'andamento dell'indice italiano FTSE MIB, mentre per l'indice americano S&P500 il 2% delle aziende ne determina per il 19.85% l'andamento.

1.3. Le obbligazioni

Le obbligazioni rappresentano un prestito per un periodo di tempo predeterminato, ricevendo in cambio degli interessi e la promessa della restituzione del capitale alla scadenza del contratto.

Possono essere emesse da banche, imprese, stati o da enti sovranazionali.

I titoli obbligazionari possono essere di due tipi:

- a tasso fisso;
- a tasso variabile.

Con le obbligazioni a tasso fisso, alla scadenza del prestito, si ottiene il rimborso del capitale investito più il pagamento del tasso d'interesse, fissato all'emissione del titolo, chiamato cedola. Generalmente le obbligazioni a tasso fisso prevedono il pagamento di cedole semestrali o annuali. Le obbligazioni a tasso variabile hanno invece un rendimento che può cambiare seguendo il tasso d'interesse del mercato, la loro rendita è quindi legata ad esso in ogni momento, e per questo motivo sono considerate più sicure rispetto alle precedenti.

1.4. I derivati

Questa categoria di strumenti finanziari comprende un vasto insieme di prodotti, che hanno però in comune un aspetto. Tutti dipendono da un sottostante, che può essere una materia prima, una valuta, un'azione o anche un tasso di interesse, da cui il loro prezzo dipende.

1.4.1. I futures

I contratti futures sono emessi per obbligare un compratore ed un venditore a scambiarsi un bene, materiale o finanziario, ad un certo prezzo in un preciso momento futuro. Per questo motivo hanno quindi una scadenza, alla scadenza della quale il contratto cessa di esistere. Ogni contratto corrisponde ad una determinata quantità del bene a cui si riferisce. La nascita di questo tipo di contratto è avvenuta nella seconda metà del 1800 per proteggere i coltivatori ed i commercianti da grandi variazioni di prezzo dei beni trattati che avrebbero potuto causare perdite economiche significative. Conoscendo in anticipo quale sarà il prezzo di un bene è possibile continuare a produrlo conoscendo quali saranno i possibili guadagni. Usando i contratti futures non si avranno perdite causate da ribassi del prezzo del bene lavorato ma non si avranno guadagni ulteriori che si sarebbero

potuti ottenere solo dal fatto che la materia prima lavorata ha acquistato valore. A partire dagli anni '70 oltre alle materie prime sono stati sviluppati futures riguardanti indici azionari, valute e titoli di stato. A cause del loro funzionamento è possibile perdere più del capitale a disposizione se non si prendono i necessari provvedimenti. Per i contratti futures il lotto minimo di negoziazione varia a seconda dell'attività finanziaria sulla quale il contratto è scritto.

1.4.2. I certificati

Simili ai futures, ma con alcune differenze. La prima è il fatto che non è possibile perdere più del capitale, la seconda importante differenza riguarda il numero minimo di lotti. In questo tipo di contratto non c'è un minimo, per cui si può negoziare anche un solo certificato alla volta.

1.4.3. I contratti per differenza

Questo genere di contratto è uno strumento finanziario derivato che viene stipulato tra un compratore ed un venditore. I due partecipanti al contratto si scambiano denaro in base alla differenza di prezzo che l'attività sottostante ha registrato dal momento dell'apertura del contratto fino alla sua chiusura. Se il prezzo dell'attività ha avuto un incremento in tale lasso temporale, per il compratore si sarà realizzato un guadagno, mentre per il venditore una perdita. Viceversa, se il prezzo dell'attività ha avuto un decremento, per il compratore si sarà realizzata una perdita, mentre per il venditore un guadagno.

1.4.4. Le opzioni

Le opzioni sono una forma di contratto che permette al possessore di poter acquistare o vendere lo strumento finanziario a cui si riferisce ad una certa data ad un certo prezzo. Esistono due tipi di opzioni: *call* e *put*.

- Opzioni Call

È un contratto tra due entità, in cui il compratore dell'opzione ha il diritto, ma non l'obbligo, di comprare, ad esempio, un titolo azionario al prezzo dichiarato sul contratto entro una certa data ed il venditore ha l'obbligo di vendere l'azione se l'opzione viene esercitata dal compratore.

Se ad esempio una certa azione è scambiata a 50€, il compratore può acquistare l'opzione call se pensa che il prezzo dell'azione dovrebbe aumentare. Un possibile contratto potrebbe fissare un prezzo di acquisto pari a 55€, pagando una certa cifra per siglare il contratto. Se, entro il termine del contratto, l'azione avrà raggiunto un prezzo di 60€ il compratore eserciterà il contratto assicurandosi le azioni al prezzo stabilito di 55€, generando così un guadagno di 5€ per azione meno il costo del contratto.

- Opzioni Put

È un contratto tra due entità, in cui il compratore dell'opzione ha il diritto, ma non l'obbligo, di vendere, ad esempio, un titolo azionario al prezzo dichiarato sul contratto entro una certa data ed il venditore ha l'obbligo di comprare l'azione se l'opzione viene esercitata dal compratore.

Se ad esempio una certa azione è scambiata a 50€, il compratore può vendere l'opzione put se pensa che il prezzo dell'azione dovrebbe scendere. Un possibile contratto potrebbe fissare un prezzo di vendita pari a 45€, pagando una certa cifra per siglare il contratto. Se, entro il termine del contratto, l'azione avrà raggiunto un prezzo di 40€ il compratore eserciterà il contratto assicurandosi le azioni al prezzo stabilito di 45€, generando così un guadagno di 5€ per azione meno il costo del contratto.

I vantaggi dell'utilizzo delle opzioni, rispetto alle tradizionali azioni, sono dovuti alla limitazione delle perdite pari al costo del contratto e al fatto di dover utilizzare un capitale minore rispetto a quello necessario per acquistare un equivalente numero di azioni.

2. Data mining

Con il termine data mining si vogliono includere tutte quelle attività volte ad estrarre conoscenza dai dati. Questo tipo di processo è diventato possibile a partire dagli ultimi anni, periodo in cui la quantità di dati generati ed utilizzabili è cresciuta notevolmente. In questa crescita di raccolta dati sono coinvolti praticamente tutti gli aspetti della vita quotidiana e non, partendo da tutto ciò riguardante gli smartphone, gli acquisti fatti online, fino ad arrivare agli apparecchi del mondo IoT. In tutti i settori citati è fondamentale, per le aziende produttrici, capire come gli strumenti vengono utilizzati al fine di proporre sul mercato opzioni più competitive rispetto ai concorrenti. Proprio questo ultimo passaggio mette in luce l'importanza dell'utilizzo del data mining, perché permette di ricavare informazioni dai dati a disposizione al fine di migliorare un prodotto.

Data l'importanza che i mercati finanziari ricoprono oggi e la loro natura descrivibile con sequenze numeriche, con l'utilizzo dei computer in questo settore è stata possibile la raccolta e la diffusione di una grande quantità di dati. Basti pensare che per ogni azione quotata in borsa e per ogni compravendita di ogni titolo azionario viene generato un record contenente almeno il nome dell'azione, la data e l'ora dell'acquisto o vendita, il prezzo del titolo in quel momento e il numero di azioni scambiate. I dati generati dal mercato azionario vengono analizzati per costruire strategie di investimento profittevoli, e tutti i passi che ne compongono lo sviluppo sono comuni alle tecniche di data mining.

2.1. KDD

Il data mining in realtà è solo parte di un processo più ampio chiamato *knowledge discovery from data*, ossia estrarre conoscenza dai dati. È un concetto generico meglio definito in [1], composto da diversi passaggi, il cui obiettivo è quello di identificare nei dati a disposizione pattern nuovi, utili e chiari.

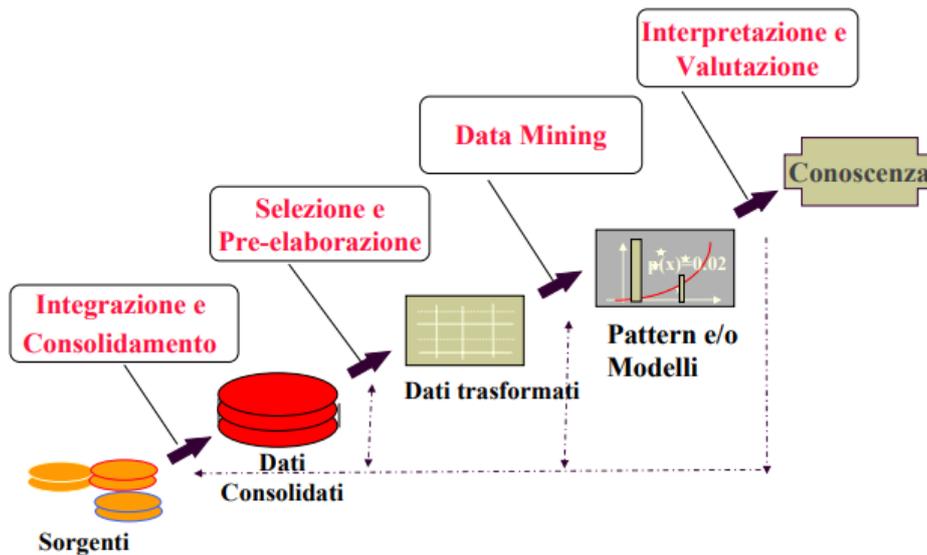


Figura 1 - Fasi del processo di estrazione di conoscenza [2]

Le fasi che compongono il processo sono:

- **Integrazione e consolidamento:** consiste nel raccogliere i dati, anche da diverse fonti, e formare un unico database.
- **Selezione:** fase in cui vengono selezionati, con un certo criterio, i dati che verranno utilizzati successivamente. L'obiettivo è di andare a considerare solo informazioni utili ai fini del processo. Il risultato ottenuto è un insieme di dati chiamati *target data*.
- **Pre-elaborazione:** fase in cui i dati precedentemente ottenuti vengono preparati per essere analizzati. Questo comprende la rimozione di valori anomali, di rumore ed eventualmente la normalizzazione dei dati.
- **Data mining:** è la fase chiave di tutto il processo, consiste nell'utilizzo di particolari tecniche per identificare pattern significativi nei dati.
- **Interpretazione e valutazione:** fase in cui vengono analizzati i pattern messi in evidenza precedentemente, per poter capire quali potrebbero essere utili per i successivi test.

2.2. Tecniche di data mining

Attraverso il data mining si possono risolvere essenzialmente due categorie di problemi: la predizione di un valore e la descrizione di un insieme di dati.

L'obiettivo della predizione è di identificare il possibile valore, non noto precedentemente, di un certo attributo. Nel caso in cui si voglia predire un valore continuo si parla di regressione, mentre se l'obiettivo è individuare a quale categoria appartiene un elemento si parla di classificazione. In questo lavoro di tesi è stata utilizzata la tecnica della regressione, approfondita in seguito. In generale è valido assumere che i modelli predittivi sono costruiti a partire da un set di dati chiamato *training data*, in cui è noto il valore dell'attributo di cui si vorrà fare la predizione nel set di dati usati per il test. Questo tipo di apprendimento è chiamato apprendimento supervisionato.

L'obiettivo della descrizione è invece quello di capire quali sono le principali caratteristiche di un certo set di dati. Il *clustering* è la tecnica più utilizzata per sviluppare questo tipo di analisi, in quanto raggruppa, in base ai valori degli attributi, elementi simili tra loro. Questo tipo di apprendimento è chiamato apprendimento non supervisionato.

3. La regressione

Per regressione si intende una tecnica usata per analizzare una serie di dati al fine di scoprire eventuali relazioni tra una o più variabili indipendenti ed una variabile dipendente continua di cui si vuole fare una previsione.

Per fare ciò si costruisce una funzione di regressione, che è una funzione delle variabili indipendenti.

Nella forma più generale possibile questa prende la seguente forma:

$$Y = f(X_1, X_2, \dots, X_n)$$

dove:

- Y è la variabile dipendente;
- X_i è una variabile indipendente;
- ε è la variabile casuale di errore.

I parametri di tale funzione sono stimati in modo tale che la curva di regressione si avvicini il più possibile all'insieme dei dati, al fine di minimizzare l'errore. Gli algoritmi che permettono di applicare la regressione ad un set di dati sono molti, tra cui quelli illustrati nel seguito, usati per lo svolgimento della tesi.

3.1. Regressione Lineare Semplice

Il modello di regressione lineare è dato dall'equazione:

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

dove:

- i indica l'indice del dato a cui si fa riferimento tra le osservazioni, $i = 1 \dots n$;
- Y_i è la variabile dipendente;
- X_i è la variabile indipendente o regressore;
- $\beta_0 + \beta_1 X$ è la retta di regressione o funzione di regressione della popolazione;
- β_0 è l'intercetta della retta di regressione della popolazione;

- β_1 è il coefficiente angolare della retta di regressione della popolazione. Il segno di β_1 indica se la relazione lineare è positiva o negativa;
- e_i è l'errore casuale in Y corrispondente all' i -esima osservazione.

Al fine di ridurre l'errore e spesso viene usato il metodo dei minimi quadrati, una tecnica di ottimizzazione il cui obiettivo è quello di minimizzare la somma dei quadrati delle distanze tra i dati osservati e quelli della retta che rappresenta la funzione lineare delle variabili indipendenti.

Su un grafico questo si rappresenta come la distanza minima tra la retta interpolante e il dato al quadrato, minore è quindi questo valore e migliore sarà il modello generato.

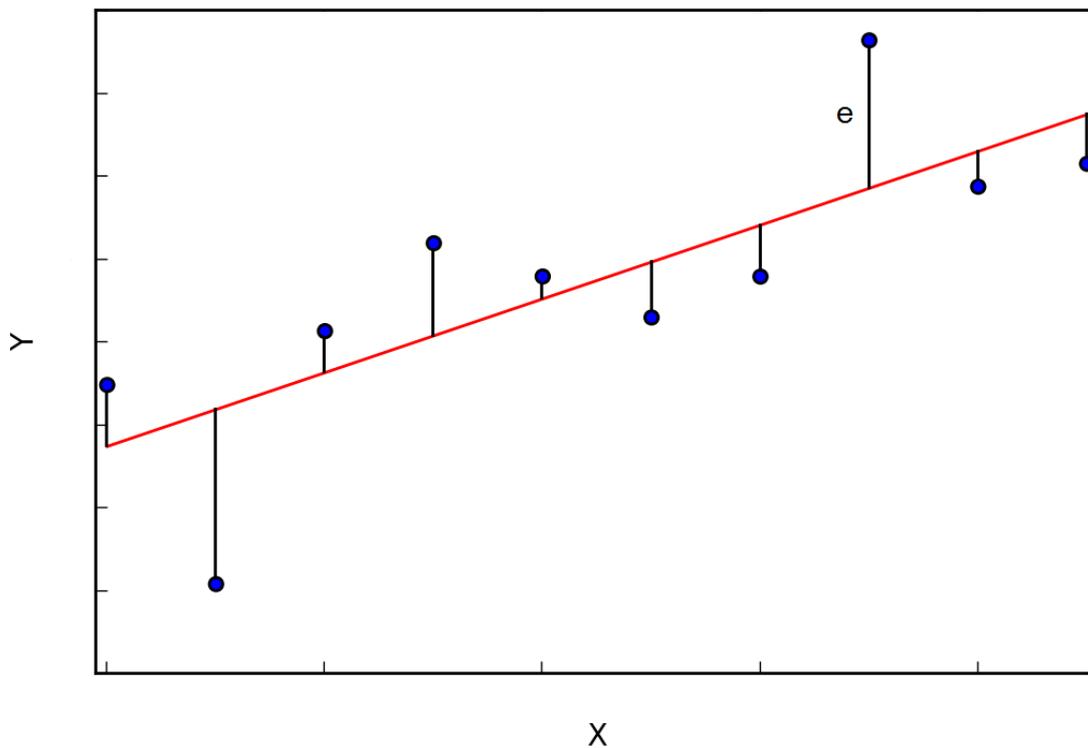


Figura 2 - Visualizzazione dei quadrati minimi

Attraverso questo metodo si vanno a determinare i parametri β_0 e β_1 .

In formule, si vuole minimizzare:

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

dove:

- Y_i è il vero valore di Y per l'osservazione di i ;
- \hat{Y}_i è il valore stimato di Y per l'osservazione di i .

Si tratta quindi di minimizzare il seguente termine:

$$\sum_{i=1}^n [(Y_i - (\beta_0 + \beta_1 X_i + e))]^2$$

risolvendolo rispetto alle due incognite β_0 e β_1 , chiamate coefficienti di regressione.

3.2. Regressione Lineare Multipla

Quando più variabili contribuiscono a spiegare la variabile dipendente Y il modello di regressione lineare semplice si amplia per incorporare le variabili aggiunte.

Questo porta, usando l'algebra lineare, alla seguente formula:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + e_i$$

dove:

- i indica l'indice del dato a cui si fa riferimento tra le osservazioni, $i = 1 \dots n$;
- Y_i è l' i -esimo valore della variabile dipendente;
- $X_{1i}, X_{2i}, \dots, X_{ki}$ sono le i -esime osservazioni di ciascuno dei k regressori;
- $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$ è la retta di regressione o funzione di regressione della popolazione;
- β_0 è l'intercetta della retta di regressione della popolazione;
- β_1 è il coefficiente angolare di X_1 , β_2 è il coefficiente angolare di X_2 , β_k è il coefficiente angolare di X_k ;
- e_i è l'errore casuale in Y corrispondente all' i -esima osservazione.

La formula di cui sopra si può però riscrivere in forma compatta come:

$$Y = X\beta + e$$

dove:

- Y è un vettore i -dimensionale, chiamato *response vector*;
- X è una matrice di dimensione $k * i + 1$ chiamata *design matrix*;
- β è un vettore $(k + 1)$ -dimensionale chiamato *slope vector*;
- e è un vettore i -dimensionale chiamato *error vector*;
- $X\beta$ indica il prodotto tra matrice e vettore.

Anche in questo caso si deve stimare il vettore dei coefficienti β in modo del tutto analogo al metodo dei minimi quadrati usato per la regressione lineare semplice.

Si tratta di minimizzare il seguente termine:

$$\sum_{i=1}^n [(Y_i - (\beta_0 + \beta_{1i}X_{1i} + \beta_{2i}X_{2i} + \dots + \beta_{3i}X_{3i} + e))]^2$$

rispetto a tutti i possibili valori del vettore β .

3.3. Support Vector Machine

Attraverso l'utilizzo del SVM si possono svolgere operazioni sia di classificazione che di regressione, e i due algoritmi hanno caratteristiche comuni.

- Classificazione

Un SVM classificatore ha l'obiettivo di trovare un iperpiano che massimizzi la distanza, chiamata margine, tra le classi. Il termine *support vector* indica proprio il vettore che definisce l'iperpiano divisore.

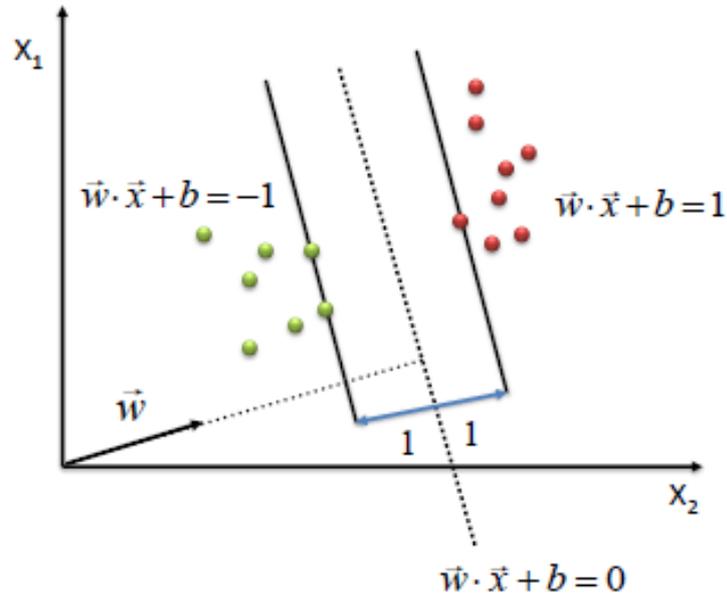


Figura 3 - Esempio di classificazione attraverso SVM [3]

Per trovare l'iperpiano ottimo bisogna massimizzare il margine w e nel caso ideale come nella Figura 3 l'iperpiano trovato separa in modo netto le due classi.

Matematicamente questo si esprime nel seguente modo:

$$\max\left(\frac{2}{\|w\|}\right)$$

Per cui vale:

$$(w x + b) \geq 1, \forall x \in C1$$

$$(w x + b) \leq -1, \forall x \in C2$$

Per trovare i valori di w e b bisogna risolvere la seguente funzione obiettivo:

$$\min\left(\frac{1}{2}\|w\|^2\right)$$

Per cui vale:

$$y_i(w x_i + b) \geq 1, \forall x_i$$

Non sempre però si è nella situazione ideale, per cui ci possono essere dei casi in cui uno o più dati appartenenti ad una certa classe siano disposti nello spazio definito per un'altra classe, producendo così una classificazione errata. In questo caso bisogna introdurre una penalità ξ , per cui se un dato è classificato erroneamente il suo valore aumenta. Il nuovo obiettivo è quindi massimizzare il margine w e minimizzare la penalità ottenuta dalla errata classificazione.

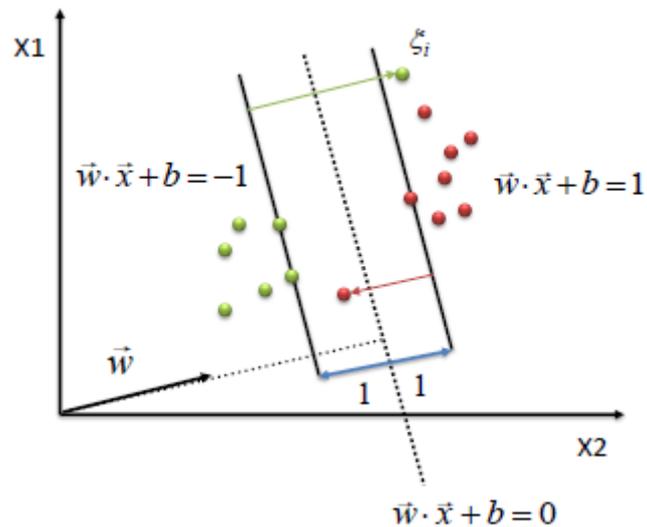


Figura 4 - Esempio di classificazione attraverso SVM con penalità ξ [3]

In formule ciò si esprime come di seguito:

$$\begin{cases} y_i(w x_i + b) \geq 1 - \xi_i, \forall x_i \\ \xi_i \geq 0, \forall i \end{cases}$$

Negli esempi precedenti per separare le due classi veniva usata una linea retta, ma ci sono occasioni in cui si otterrebbe una separazione migliore usando funzioni non lineari. Per fare ciò l'SVM usa una funzione del kernel non lineare per fare un mapping dei dati in uno spazio dimensionale diverso dove è possibile separare le due classi linearmente: questo processo è chiamato *kernel trick*.

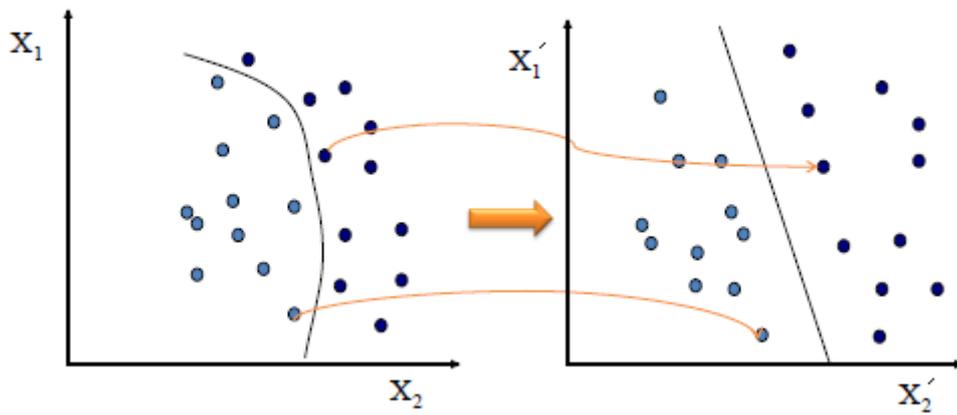


Figura 5 - Esempio di classificazione attraverso SVM con kernel trick [3]

- Regressione

L'SVM può anche essere usato come metodo di regressione, lasciando intatte le caratteristiche tipiche dell'algoritmo. In questo caso si aggiunge una tolleranza ϵ , un valore al di sotto del quale l'errore viene ignorato.

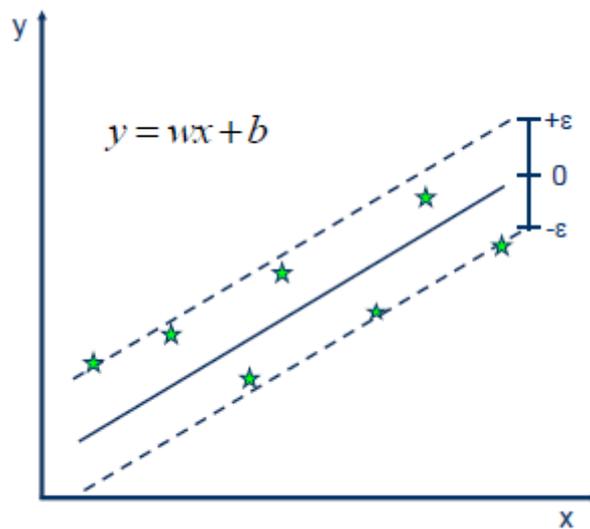


Figura 6 - Esempio di Support Vector Regression con tolleranza ϵ [4]

Formalmente per risolvere il problema si tratta di minimizzare:

$$\min \left(\frac{1}{2} \|w\|^2 \right)$$

Per cui vale:

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases}$$

Si assume quindi che il problema sia risolvibile, che esista quindi una funzione che riesca ad approssimare tutti i punti con al più uno scostamento ε .

Se però la tolleranza ε non viene rispettata come in Figura 7, bisogna introdurre un errore ξ , che ha il compito di aggiungere una penalità rendendo però possibile arrivare ad un risultato.

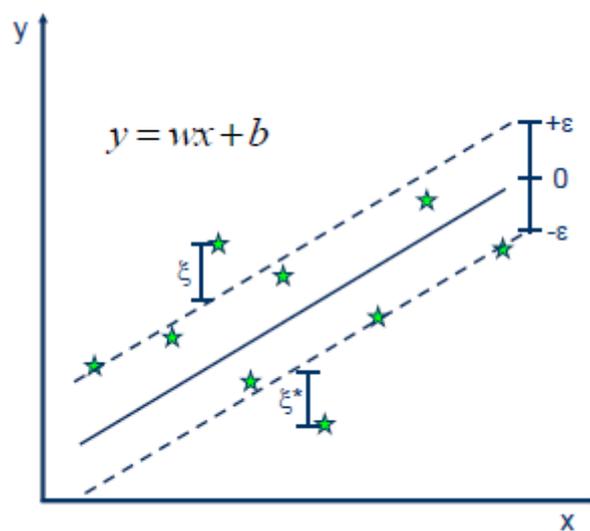


Figura 7 - Esempio di Support Vector Regression con errore ξ [4]

La funzione d'errore ha quindi l'aspetto del grafico:

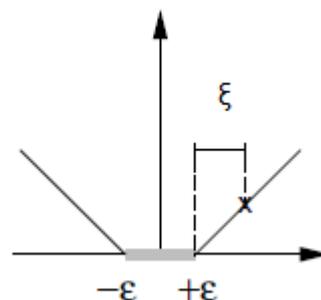


Figura 8 - Grafico della funzione d'errore ξ [5]

Formalmente questa nuova formulazione del problema si esprime nel seguente modo:

$$\min \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right)$$

Per cui vale:

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i \\ \xi_i + \xi_i^* \geq 0 \\ C \geq 0 \end{cases}$$

La variabile C è un parametro fondamentale per modificare il comportamento di un SVM, valido sia per la classificazione che per la regressione, per cui a seconda del valore che assume si creano situazioni differenti. Usando un valore alto si dà molta importanza al fatto che i dati siano classificati correttamente, mentre usando un valore basso ci si concentra sull'aver il margine più ampio possibile.

3.4. Gradient Boosting Trees

È un algoritmo che usa un insieme di alberi decisionali “*deboli*” per creare un predittore più efficace al fine di risolvere problemi di classificazione o di regressione.

L'idea chiave del *gradient boosting trees* è che si creino molti alberi in serie, dove ogni nuovo albero creato cerca di correggere l'errore prodotto dagli alberi precedenti.

I due parametri più importanti sono la profondità degli alberi, che tipicamente è un valore compreso tra uno e cinque, e il *learning rate*, che controlla quanto l'albero può correggere l'errore degli alberi precedenti. Modificando questi valori si va ad influire sulla bontà del modello finale generato; impostando valori troppo alti di *learning rate* si rischia di trovarsi in una situazione di *overfitting*, conviene quindi abbassarlo per generalizzare meglio il problema.

4. Operare in borsa

Investendo in azioni si possono usare due strategie, ossia si può decidere di comprare i titoli di un'azienda pensando che si trovi in un buon periodo e che quindi ipotizzando che i suoi titoli guadagneranno valore, o in alternativa si può anche investire contro una società. Questo significa che si prevede un ribasso delle azioni e quindi si investe in quella direzione. Le due modalità sopra descritte sono chiamate rispettivamente strategia *long* e strategia *short*.

La strategia long è quella più comune, perché semplice da capire ed attuare, ma anche perché il mercato per la maggior parte del tempo ha un andamento al rialzo. Se si considerano i guadagni giornalieri in realtà solo il 53% dei giorni sono in guadagno, mentre nel 47% delle volte si ha una perdita. Se però si allunga la finestra di tempo e si sceglie un anno per calcolare l'andamento del mercato si ricava che è in rialzo per il 73% delle volte. Se poi si analizza un orizzonte temporale ancora più lungo, di dieci anni, si ottiene un risultato ancora più estremo, con il 93% delle volte in cui il mercato ha generato un profitto.

La strategia short si manifesta quindi per periodi più brevi, ma questo non significa che sia meno redditizia. Infatti, storicamente, il mercato ha proposto lunghi periodi di rialzo moderato e brevi periodi di fortissimo ribasso, ed i seguenti sono i casi più eclatanti di recessioni seguendo l'indice S&P500:

- Settembre 1929 – Giugno 1932: -86.1%¹ in trentaquattro mesi;
- Agosto 1987 – Dicembre 1987: -33.5% in tre mesi;
- Ottobre 2007 – Marzo 2009: -56.4% in diciassette mesi.

Un altro aspetto da tenere in considerazione è l'orizzonte temporale degli investimenti fatti, a seconda della cui durata si creano strategie e situazioni diverse. Una singola operazione può durare da minuti ad anni, e ciò comporta grandi differenze a livello di gestione.

4.1. Posizioni *long*

Per posizione long si intende il classico investimento, in cui si dà fiducia ad una società comprandone le azioni, e si attende che il loro valore cresca.

Se si comprano ad esempio cento azioni della società X ad un prezzo di 100€ l'una, per un totale di 10000€, e dopo un mese tali azioni hanno un valore di 104€, significa che l'investimento ha fruttato il 4%, portando il capitale investito alla somma di 10400€.

Se invece dopo un mese il prezzo del titolo fosse stato di 97€ si avrebbe avuto una perdita del 3% del capitale.

Non esiste un guadagno massimo raggiungibile, ma questo è calcolato in base all'aumento del prezzo del titolo azionario. Se l'azione, durante la sua storia, riuscisse a moltiplicare di molte volte il suo valore farebbe generare guadagni importanti ai proprietari, anche molto maggiori rispetto al capitale investito in origine. Le perdite invece sono limitate al totale del capitale investito, questo perché al massimo il valore di un'azione può azzerarsi, perdendo tutto il suo valore.

4.2. Posizioni *short*

Usare la strategia short è meno comune rispetto a quella tradizionale, ed il suo funzionamento è esattamente contrario rispetto alla strategia long.

Riprendendo l'esempio precedente: se si comprano cento azioni della società X ad un prezzo di 100€ l'una, per un totale di 10000€, e dopo un mese tali azioni hanno un valore di 104€, significa che l'investimento ha perso il 4%, portando il capitale investito alla somma di 9600€.

Se invece dopo un mese il prezzo fosse stato di 97€ si avrebbe avuto un guadagno del 3% del capitale.

Mentre nella posizione long si comprano le azioni, nella posizione short il meccanismo è il seguente:

- L'acquirente chiede al suo broker, che può anche essere una banca, di *imprestargli* le azioni della società X al valore di 100€ l'una, che il broker deve avere in posizione long in quel momento;
- Il broker vende le sue azioni sul mercato, e accredita il loro valore sul conto dell'acquirente;

- Dopo un mese l'azione perde valore e si trova a quota 97€;
- L'acquirente vuole chiudere la sua posizione in profitto, chiede al suo broker quindi usare i soldi precedentemente accreditati sul suo conto per comprare delle azioni al prezzo di mercato di 97€;
- Il broker ricompra le azioni e si riporta alla sua situazione iniziale;
- L'acquirente può quindi incassare il guadagno di 3€ per ogni azione trattata dovuto al fatto di aver permesso al broker ricevere dei titoli ad un prezzo minore di quello a cui li avevano in portafoglio.

A differenza della posizione long, durante un'operazione short le perdite sono teoricamente infinite. Ciò è dovuto al fatto che in questo caso, quando l'azione guadagna valore, la posizione short va in perdita e siccome un'azione può aumentare il suo prezzo all'infinito, si può perdere più dell'investimento iniziale.

Si verifica quando, ad esempio, un utente inizia la transazione short a 100€ ad azione, ma questa raggiunge i 250€. Per il trader ci sarà una perdita di 150€, maggiore dell'investimento iniziale.

4.3. Stop loss

Per proteggersi da perdite troppo elevate ed incontrollate si può inserire un limite, che una volta raggiunto, fa scattare un ordine per chiudere l'operazione automaticamente.

Se ad esempio si compra con strategia long un'azione a 100€ e si decide che al più si accetterà una perdita del 4%, si imposta uno stop loss a 96€. Se il prezzo dell'azione toccherà quella quota allora automaticamente il sistema genererà un ordine per chiudere l'operazione, accettando così una perdita limitata, conoscendone già a priori il massimo importo.

4.4. Stop profit

Una variante dello stop loss è lo stop profit, in cui si imposta sempre una soglia percentuale che segue il prezzo nei momenti di rialzo, andando a modificare il valore numerico dello stop loss. Questo tipo di stop loss viene quindi calcolato sul prezzo massimo raggiunto dall'azione. Se come

nell'esempio precedente si compra un'azione a 100€ e si decide che al più si accetterà una perdita mobile del 4%, inizialmente si imposterà uno stop loss al valore di 96€. Se però il titolo salisse fino a 115€, questo valore cambierà diventando 110.4€. Questo metodo permette di assicurare dei profitti anche nel caso in cui dopo un periodo di trend positivo si generi un trend negativo che li avrebbe potuto azzerare.

Impostare il valore di stop loss più corretto per una certa operazione non è immediato e dipende da diversi fattori, tra cui la volatilità del mercato e l'orizzonte temporale dell'investimento.

In momenti di alta volatilità sarebbe opportuno aumentare lo stop loss perché ci potrebbero essere dei falsi trend, in realtà movimenti temporanei del prezzo dell'azione, che attiverebbero prematuramente l'ordine automatico.

L'altro parametro è il l'orizzonte temporale dell'operazione: se si vuole fare un investimento di pochi giorni si imposterà tipicamente uno stop loss inferiore rispetto ad un investimento che potrebbe durare un anno, questo perché con durate maggiori gli *swing* possono essere più importanti ma comunque temporanei.

4.5. Take profit

Discorso opposto allo stop loss è quello del take profit. In questo caso non si imposta un limite alle potenziali perdite ma si imposta un limite ai potenziali profitti.

Quando si raggiunge un certo livello predeterminato di guadagno il sistema invia automaticamente un ordine per chiudere la posizione e ricavare così i profitti imposti precedentemente come obiettivo.

Questo parametro è più delicato da impostare perché taglia i potenziali guadagni, è quindi utile usarlo solo in situazioni in cui non si è sicuri che il trend desiderato possa continuare.

4.6. Orizzonte temporale delle operazioni

4.6.1. Scalping

Si tratta di trading ad altissima frequenza, di durata di qualche minuto al massimo. Con questa strategia si effettuano moltissime operazioni giornaliere, ciascuna con guadagni poco elevati.

In questo tipo di operazioni viene sfruttato ciò che è chiamato *effetto leva*. La leva è uno strumento offerto dai broker per poter effettuare un investimento maggiore rispetto al capitale disponibile. Usare una leva di 10:1 permette di comprare azioni per 1000€ investendone però solo 100€. I restanti 900€ in realtà sono anticipati dal broker. In questo modo le variazioni di prezzo comportano guadagni e perdite molto elevati rispetto al capitale investito, perché sono riferiti al totale mosso con la leva.

4.6.2. Day trading

È una strategia comune, consiste nell'aprire le posizioni al mattino e chiuderle entro la chiusura serale dei mercati. Usando un lasso temporale a livello di ore è difficile intercettare in singole gravi perdite dato che l'azione non avrà grosse variazioni ed in più non si sfrutterà il trend di base che il titolo starà affrontando. Questo però può essere sia un punto di forza che di debolezza di questa strategia perché da una parte è possibile trovare dei movimenti diversi dall'andamento generale, dall'altra parte sfruttando il trend in essere sarebbe più facile investire nella direzione giusta.

4.6.3. Swing Trading

Si tratta di uno stile di trading sul medio-lungo termine, con operazioni che durano diversi giorni o poche settimane. La caratteristica principale di questa strategia è legato all'inseguimento del trend di fondo e questo comporta una durata maggiore dell'investimento.

Questo genere di operazioni nasce dallo studio dei grafici, specialità chiamata analisi tecnica, che favorisce l'identificazione di trend azionari.

4.6.4. Position Trading o Buy & Hold

È uno stile di investimento basato completamente sul lungo periodo, diventa quindi importante l'analisi economica delle aziende interessate attraverso l'analisi economica fondamentale. In questo tipo di studio si vanno ad identificare i punti di forza di una società ed in generale del paese a cui appartiene, per decidere se può essere conveniente, a distanza di mesi o anche anni, investire in determinati titoli azionari.

5. Il progetto

In questo lavoro di tesi sono stati considerati tutti i titoli presenti negli indici azionari del FTSE MIB e dello S&P 500. In questo modo verranno trattati sia il listino italiano che quello statunitense.

Il primo anno preso in considerazione è il 2011, periodo in cui il listino italiano ha subito una perdita del 26.16% mentre il listino americano ha chiuso sostanzialmente in parità ma con trend negativi alternati a trend positivi. Durante questo anno sarà testata la strategia short, per approfittare della discesa delle azioni per generare profitti. Il secondo anno considerato è il 2013, in cui il listino italiano ha guadagnato il 22.5% alternando nella prima metà dell'anno salite a discese del prezzo, mentre il listino americano ha chiuso in positivo del 25.27% con un trend di base costantemente positivo. Per questo secondo anno saranno sfruttati i trend positivi per usare una strategia long.

Successivamente saranno anche invertite le strategie, long per il 2011 e short per il 2013, al fine di considerare il caso in cui si investa in senso opposto al trend di fondo del mercato. L'ultimo anno preso in considerazione è il 2015, in cui dopo alcuni mesi di rialzo o pareggio si è verificato un periodo di ribasso fino a fine anno. Anche in questo caso saranno testate entrambe le strategie long e short.

5.1. Confronto con lo stato dell'arte

In letteratura si trovano studi in cui vengono analizzate diverse strategie di investimento con ragionamenti di fondo diversi. Nella maggior parte dei casi gli strumenti finanziari presi in esame sono le azioni, ma non mancano analisi sull'andamento delle materie prime o di coppie di valute.

La prima grossa distinzione che si può fare riguarda la durata dell'investimento proposto, che varia da alcuni minuti fino a diversi mesi. Ciò che cambia radicalmente sono i dati utilizzati per generare le raccomandazioni. In certe trattazioni vengono analizzati esclusivamente i dati numerici dello strumento finanziario in questione, mentre in altre occasioni si considerano anche i dati economici dell'azienda selezionata o notizie ad essa legate.

Tra la letteratura che si occupa dell'analisi delle strategie a lungo termine l'obiettivo principale è quello di bilanciare al meglio il peso delle azioni contenute nel proprio portafoglio. A tal proposito

in [6] viene usata intelligenza artificiale per analizzare i dati finanziari delle varie aziende prese in esame al fine di massimizzare i guadagni riducendo il rischio. Per risolvere lo stesso problema in [7] e [8] vengono usati algoritmi genetici, e in [9] si aggiungono ancora dei vincoli riguardanti la somma minima e massima da investire in un'azione e la probabilità minima che il portfolio nel suo complesso guadagni almeno un certo valore percentuale.

Se invece si considerano le ricerche riguardanti l'analisi delle strategie a breve termine in [10] gli autori propongono di abbinare l'analisi degli articoli economici riguardanti le società quotate in borsa all'analisi tecnica e di regressione dei loro titoli azionari. Con le informazioni testuali si dovrebbero cogliere particolari maggiori riguardanti l'andamento di un'azione in quanto vengono spiegati anche i motivi di un rialzo o ribasso sul mercato. Discorso analogo quello presentato nel *paper* [11], dove però vengono solo analizzati gli articoli economici usando algoritmi di classificazione opportuni per il *sentiment analysis* generando predizioni sul valore dell'azione di una società.

Affiancato all'analisi delle notizie si inserisce l'analisi dello streaming dei dati ricevuti da Twitter, come approfondito in [12] e [13].

Considerando invece solo i dati riguardanti l'andamento azionario in forma numerica, l'uso comune è di inserire tutti gli elementi a disposizione nei dati di training e poi effettuare il processo di *features selection* come in [14] e [15]. Viene usato l'algoritmo *K-nearest neighbors* per identificare i momenti in cui comprare e vendere un'azione in [16], rinforzato poi da ulteriori tecniche di trading per confermare la predizione del trend filtro basato sul valore dell'indicatore Relative Strength Index e l'utilizzo dello stop loss e del take profit. Anche in questo caso si tratta però di operazione a medio-lungo termine in quanto hanno una media di quaranta operazioni effettuate all'anno. Con lo stesso obiettivo del medio termine si posiziona [17] in cui vengono utilizzati anche altri algoritmi quali il *Support Vector Machine*, *Naive Bayes* e un *Decision Tree*. I dati vengono trattati in modo diverso in [18] in cui viene applicata la *box theory* usando il *Support Vector Machine*. Secondo questa teoria il prezzo di un'azione oscilla tra un minimo ed un massimo in un certo lasso di tempo, ed il pattern che viene creato è contenuto in tale *box*. Nel momento in cui l'azione rompe questo contenitore il prezzo è pronto a creare un nuovo *box* in quella direzione. Tra i modelli predittivi più utilizzati si trovano le reti neurali, modellate in diverso modo come si può leggere in [19], [20], [21].

Un altro approccio è quello presentato in [22], in cui vengono studiate sequenze ricorrenti di eventi. Una sequenza è una lista ordinata di itemset, dove un itemset è composto da diversi item, che in questo caso sono i prezzi delle azioni in un certo istante di tempo, ed indica che dato un gruppo di

top – k azioni nei giorni precedenti, allora sarà probabile avere una certa azione X tra le migliori k nel giorno successivo. Il metodo usato ha poi previsto l'utilizzo di sequenze pesate in base al guadagno giornaliero delle azioni in esso contenute.

5.2. L'obiettivo della tesi

L'obiettivo è quindi quello di proporre un nuovo metodo di investimento nel campo azionario, ma che in linea di principio possa essere applicabile a qualsiasi altro strumento finanziario che si comporta in modo simile alle azioni.

La strategia usata è quella del day trading, ossia comprare e vendere in giornata l'azione scelta. L'aspetto stimolante del scegliere il day trading come lasso temporale dell'investimento è il fatto di dover scegliere tra centinaia di azioni quella che dovrebbe essere la migliore per il giorno successivo, operazione che manualmente sarebbe pressoché impossibile da svolgere quotidianamente.

Il punto che introduce una novità rispetto alla letteratura disponibile è il fatto di indagare il profitto creato da un ensemble di modelli di regressione generati da sottoinsiemi di dati relativi a specifici indicatori statistici, peculiari al mercato di riferimento, e confrontarlo con quello generato da un modello di regressione unico calcolato a partire da un ampio spettro di indicatori statistici. L'ipotesi è che la specificità delle previsioni basate su indicatori statistici peculiari per il mercato preso in esame garantisca maggiore accuratezza, e quindi profitto, rispetto a previsioni basate su correlazioni più generiche e meno focalizzate sul contesto specifico.

Gli altri due punti che distinguono questo lavoro sono il fatto di aver utilizzato algoritmi di regressione, a differenza della maggior parte della letteratura in cui sono stati usati algoritmi di classificazione che permettono di predire solamente se il titolo azionario nel giorno successivo guadagnerà o perderà valore sul mercato, e l'utilizzo del *windowing* per i dati di training.

Il fatto di usare la regressione permette di capire quali dovrebbero essere le migliori azioni per il giorno successivo potendole ordinare secondo il valore percentuale di guadagno predetto, usando invece la classificazione binaria si otterrebbe un risultato difficile da interpretare per ottenere una strategia affidabile da applicare sul mercato reale.

5.3. Crawling dei dati azionari

Il primo passo per poter investigare i dati è quello di ottenere lo storico delle quotazioni delle azioni che si vogliono prendere in considerazione. Una possibilità è quella di scaricare manualmente le quotazioni da siti web specializzati che lo permettono, oppure usare le API messe a disposizione da Yahoo! Finance e Google Finance. In alternativa si può sfruttare la libreria di Python che si appoggia a Google Finance. In tutti i casi si può scegliere l'arco temporale per cui si vuole scaricare la quotazione dell'azione, da pochi giorni fino all'ingresso in borsa, e la granularità dei dati, ossia ogni quanto tempo conoscere i valori del titolo. In questo lavoro, dato che l'orizzonte della predizione è di un giorno, è stato scelto di avere un dato per ogni giorno di mercato.

5.3.1. Fonti

I dati necessari per compiere analisi di tipo economico-finanziario sono facilmente reperibili su vari siti internet. Nella lista dei siti più importanti rientrano i reparti Finance di Yahoo! [23] e Google [24], in aggiunta a *Borsa Italiana* [25], che offre contenuti particolarmente approfonditi per le azioni quotate sul mercato italiano e *The New York Stock Exchange* [26] per le azioni quotate sulla borsa di New York.

I servizi che offrono Yahoo! Finance e Google Finance sono simili: entrambi offrono la possibilità di analizzare il grafico dell'azione, esaminare informazioni riguardanti i dati economici degli ultimi anni e leggere notizie relative al titolo in questione. Inoltre, è possibile scaricare manualmente i dati giornalieri dell'andamento del prezzo. Con i due portali più specializzati si possono vedere in dettaglio tutti gli ordini scambiati nella giornata, con il prezzo a cui si è comprato/venduto il titolo azionario ed il volume dell'operazione.

5.3.2. Yahoo! Finance API

Il servizio Yahoo! Finance offriva la possibilità di scaricare i dati delle quotazioni della maggior parte delle società quotate in borsa. Il 2 novembre 2017 [27] però è stato cessato il supporto a questo

servizio, rendendolo di fatto inaccessibile. Vengono però esposte le caratteristiche di queste API in quanto usate nel lavoro di tesi.

Tramite le API si possono scaricare i dati delle ultime quotazioni in diversi formati: CSV, JSON, XML.

Attraverso questo servizio si possono ricavare i seguenti dati:

- Prezzo attuale di un'azione, valuta monetaria o indice;
- Prezzo di uno strumento finanziario in un certo periodo di tempo;
- Informazioni societarie dell'azienda in questione.

Scaricando i dati storici di un'azione è possibile impostare la granularità dei dati: oraria, giornaliera, settimanale o mensile. Più il tempo tra un dato e l'altro è ridotto e minore sarà la finestra temporale di cui sarà possibile ottenere i dati.

I valori che si possono scaricare sono:

- Prezzo di apertura;
- Prezzo di chiusura;
- Prezzo minimo giornaliero;
- Prezzo massimo giornaliero;
- Volume (numero di azioni scambiate durante la giornata).

In questo lavoro di tesi è stata usata principalmente la possibilità di scaricare i prezzi delle azioni per un certo periodo di tempo. In particolare sono stati utilizzati i dati riguardanti gli anni 2011, 2013 e 2015 per le azioni contenute negli indici S&P 500 e FTSE MIB.

5.3.3. Google Finance API

Google Finance offre una alternativa simile a quella di Yahoo! Finance ma risulta essere ancora funzionante.

In questo caso i dati che si possono ottenere sono riguardanti solo al prezzo e volume, per un periodo di tempo impostabile a piacere, di una certa azione. Anche in questo caso i valori che si possono scaricare sono:

- Prezzo di apertura;
- Prezzo di chiusura;

- Prezzo minimo giornaliero;
- Prezzo massimo giornaliero;
- Volume (numero di azioni scambiate durante la giornata).

5.4. Rapidminer Studio

Rapidminer Studio [28] è un software che permette l'applicazione di machine learning, data mining, text mining, predictive analytics e business analytics scritto in Java. Il progetto è iniziato nel 2001 nell'università di Dortmund, ed ora conta più di trecentomila utenti che usano il software regolarmente in oltre centocinquanta paesi e circa duecento dipendenti. L'aspetto che lo rende competitivo è il fatto di avere una GUI, *Graphic User Interface*, che permette all'utilizzatore di progettare applicazioni senza dover scrivere molte righe di codice. Risulta quindi essere molto utile per iniziare ad ideare un progetto potendo vedere graficamente i blocchi di cui si ha bisogno.

I blocchi che svolgono le operazioni sono chiamati *operatori* [29], i quali ricevono, processano e inviano all'operatore successivo i dati da rielaborare nuovamente.

Gli operatori più importanti utilizzati in questo lavoro di tesi sono stati:

- *Windowing, Series Extension*: permette di generare dati di training per la regressione utilizzando la tecnica del windowing;
- *Support Vector Machine* [30], *Linear Regression* [31], *Gradient Boosted Trees* [32]: utilizzati per la regressione al fine di costruire i modelli che permettano di predire il prossimo valore della serie presentata in input.

5.5. Gli indicatori

Nell'analisi tecnica delle azioni, durante lo studio del grafico dell'andamento del prezzo, vengono usati anche altri parametri chiamati indicatori od oscillatori.

Gli indicatori sono funzioni matematiche usate per integrare informazioni riguardanti i dati del titolo azionario, tra cui il suo prezzo minimo, massimo, di apertura, di chiusura e il volume di azioni scambiate, al solo prezzo a cui è scambiata un'azione in un certo momento.

Questi strumenti vengono calcolati su un determinato periodo di tempo, il che significa che il valore di un indicatore a periodo N viene calcolato sugli ultimi N valori della serie a cui si riferisce.

Per tutti gli indicatori vale l'osservazione che più il periodo N su cui è calcolato è ampio, e più si tratterà di un indicatore che va ad analizzare il comportamento di lungo periodo di un titolo, anche chiamato comportamento di fondo, mentre se il periodo N è breve si analizzano le caratteristiche del breve termine.

5.5.1. Moving Average (MA)

È la media del prezzo di chiusura degli ultimi n giorni, calcolata come:

$$MA = \frac{\sum Chiusura}{n}$$

Dove:

- n indica il numero di giorni su cui è calcolato l'indicatore

Il risultato che si ottiene è una serie di valori che solitamente viene plottata sovrapposta al grafico del prezzo. In questo modo si riesce a visualizzare un andamento più smussato ma comunque significativo del prezzo reale. Infatti in certi periodi in cui il mercato è molto movimentato può essere difficile individuare con chiarezza i momenti in cui l'andamento era in generale positivo o negativo.

Sono state calcolate medie mobili a periodo cinque, dieci, venti e cinquanta.

5.5.2. Exponential Moving Average (EMA)

È una media pesata, creata per dare più importanza agli ultimi valori della serie.

Calcolata come:

$$EMA = (Chiusura - EMA_{precedente}) * M + EMA_{precedente}$$

Dove:

- $M = \frac{2}{n+1}$;
- n indica il numero di giorni su cui è calcolato l'indicatore.

Dando pesi diversi ai valori in base a quando questi si sono manifestati si riesce a creare un indicatore più reattivo ai cambiamenti improvvisi di prezzo. Anche in questo caso viene rappresentato graficamente come sovrapposto al grafico principale del prezzo dell'azione.

Questo indicatore è stato calcolato su un periodo pari a cinque giorni.

5.5.3. On Balance Volume (OBV)

Viene ottenuto sommando i volumi dei giorni che hanno avuto una chiusura positiva rispetto al giorno precedente e sottraendo i volumi dei giorni che hanno invece avuto una chiusura negativa rispetto al giorno precedente.

$$OBV = \sum Vp - \sum Vn$$

Dove:

- Vp indica il volume di un giorno in cui il prezzo di chiusura è maggiore del prezzo di chiusura del giorno precedente;
- Vn indica il volume di un giorno in cui il prezzo di chiusura è minore del prezzo di chiusura del giorno precedente.

In altre parole si misura la pressione di acquisto e di vendita dell'azione, supponendo che pressione positiva porti ad un aumento del prezzo e una pressione negativa porti ad una riduzione del prezzo. Tipicamente si ha pressione positiva in acquisto e negativa in vendita dell'azione. Questo indicatore viene mostrato nei grafici al di sotto del grafico del prezzo del titolo finanziario ed è stato calcolato su un periodo pari a cinque giorni.

5.5.4. Average True Range (ATR)

Definito come il più grande valore tra:

- Differenza tra il massimo ed il minimo dello stesso giorno;
- Differenza tra il massimo di un giorno e la chiusura del giorno precedente;
- Differenza tra il minimo di un giorno e la chiusura del giorno precedente.

È un indicatore della volatilità dell'azione in cui livelli alti o crescenti segnalano la continuazione del trend attuale, mentre livelli bassi o decrescenti indicano che l'azione si sta stabilizzando in quella che è chiamata una fase laterale del prezzo. Anche questo indicatore viene mostrato nei grafici al di sotto del grafico del prezzo del titolo finanziario ed è stato calcolato su un periodo pari a cinque giorni.

5.5.5. Relative Strength Index (RSI)

È un indicatore del momentum di uno strumento finanziario. Per momentum si intende l'accelerazione che la salita o discesa del prezzo registra durante un movimento sul mercato.

Calcolato come:

$$RSI = 100 - \frac{100}{1 + \left(\frac{Gp}{Gn}\right)}$$

Dove:

- Gp indica il numero di giornate in cui il prezzo di chiusura è maggiore del prezzo di chiusura del giorno precedente;
- Gn indica il numero di giornate in cui il prezzo di chiusura è minore del prezzo di chiusura del giorno precedente.

Seguendo il ragionamento dell'ideatore, quando il prezzo di un'azione si muove verso l'alto rapidamente si entra nell'area dell'*ipercomprato*, mentre quando il prezzo si muove verso il basso

rapidamente si entra nell'area dell'*ipervenduto*. In entrambi i casi dovrebbe verificarsi un'inversione del prezzo rispetto al trend attuale. Anche questo indicatore viene mostrato nei grafici al di sotto del grafico del prezzo del titolo finanziario ed è stato calcolato su un periodo pari a cinque giorni.

5.5.6. Money Flow Index (MFI)

Anche in questo caso si tratta di un indicatore legato al momentum, perché usa sia il prezzo che il volume degli scambi di uno strumento finanziario per misurare la pressione di un movimento.

Calcolato come:

$$MFI = 100 - \frac{100}{1 + \frac{\sum MFp}{\sum MFn}}$$

Dove:

- $MFp = \sum V * \frac{Massimo+Minimo+Chiusura}{3}$

Per le giornate in cui il prezzo di chiusura è maggiore del prezzo di chiusura del giorno precedente;

- $MFn = \sum V * \frac{Massimo+Minimo+Chiusura}{3}$

Per le giornate in cui il prezzo di chiusura è minore del prezzo di chiusura del giorno precedente.

Il Money Flow Index è molto simile al Relative Strength Index, in quanto ne è una versione ponderata in base ai volumi scambiati. Per il resto è del tutto analogo all'RSI.

Questo indicatore è stato calcolato su un periodo pari a cinque giorni.

5.5.7. Aroon Oscillator

Si tratta di un indicatore definito *trend-following* in quanto usa delle misure create per determinare la forza di un movimento. È formato da due componenti, di cui si calcola la differenza per trovare quale delle due è la più influente in un certo momento.

Calcolato come:

$$AROON = AROON_{up} - AROON_{down}$$

Dove:

- $AROON_{up} = 100 * \frac{(n - \text{numero giorni dal massimo prezzo di chiusura})}{n}$;
- $AROON_{down} = 100 * \frac{(n - \text{numero giorni dal minimo prezzo di chiusura})}{n}$;
- n indica il numero di giorni su cui è calcolato l'indicatore.

Il valore che viene considerato come riferimento è lo zero, al di sopra del quale si dovrebbe verificare un trend positivo e al di sotto del quale si dovrebbe verificare un trend negativo. Anche questo indicatore viene mostrato nei grafici al di sotto del grafico del prezzo del titolo finanziario ed è stato calcolato su un periodo pari a cinque giorni.

5.5.8. Commodity Channel Index (CCI):

Ultimo indicatore considerato riferito al momentum, viene usato per identificare possibili inversioni di prezzo dovute a periodi prolungati di rialzo o di ribasso che potrebbero evolversi in un cambio di direzione.

Calcolato come

$$CCI = \frac{TP - SMATP}{0.015 * devMedia}$$

Dove:

- Typical Price: $TP = \frac{Massimo + Minimo + Chiusura}{3}$;
- Simple Moving Average Typical Price: $SMATP = \frac{\sum TP}{n}$;
- Deviazione Media: $devMedia = \frac{SMATP - TP}{n}$;
- n indica il numero di giorni su cui è calcolato l'indicatore.

L'idea base di questo indicatore è che nei prezzi delle azioni si possono trovare fasi cicliche, con massimi e minimi che si alternano ad intervalli periodici. Così come nell'indicatore Relative Strength Index anche in questo caso ci sono le fasce di *ipercomprato* ed *ipervenduto*, rispettivamente quando

assume un valore maggiore di cento e minore di meno cento. Superati questi valori si dovrebbe verificare un'inversione di tendenza del valore dell'azione. Anche questo indicatore viene mostrato nei grafici al di sotto del grafico del prezzo del titolo finanziario ed è stato calcolato su un periodo pari a cinque giorni.

5.6. Preparazione dei dati

Il primo passo da effettuare nel processo di estrazione di conoscenza consiste nello scaricare i dati relativi alle singole azioni contenute nei due indici S&P 500 e FTSE MIB per gli anni selezionati, ossia 2011, 2013 e 2015. Si tratta di circa cinquecento titoli per il mercato americano e quaranta per il mercato italiano. Per ogni società quotata è opportuno avere tutta la storia del prezzo del titolo per evitare di avere buchi temporali nei dati che verranno utilizzati come training.

5.6.1. Windowing

Una delle idee base dell'analisi dei mercati è che l'andamento di un titolo possa condizionare o possa essere condizionato da un altro titolo durante lo stesso giorno o durante il giorno precedente, si è utilizzato il windowing, come anticipato in [22], una strategia per costruire i dati per il training del modello che permette di evidenziare tali comportamenti.

Un esempio di ciò si ha quando un titolo A e un titolo B, spesso appartenenti allo stesso settore economico, hanno comportamenti simili come conseguenza, ad esempio, di alcune notizie importanti. Questo è un comportamento che si verifica frequentemente soprattutto tra i titoli del settore energetico, tecnologico o bancario. Un esempio recente di tali fenomeni si è verificato nei giorni immediatamente successivi all'incidente nucleare di Fukushima Dai-ichi verificatosi l'11 marzo 2011, quando anche in Italia i titoli del comparto energetico persero percentuali comprese tra il 2% e il 3%:

- Saipem: -3%;
- Eni: -2%;
- Snam Rete Gas: -2%;

- Enel: -2%.

Si ha un discorso simile quando l'andamento di un titolo A in un certo giorno provoca direttamente un effetto sul titolo B nel giorno successivo. Questo genere di comportamento è invece più difficile da identificare manualmente usando un gran numero di azioni come in questo caso.

Per poter sfruttare il windowing bisogna unire i dati delle azioni appartenenti allo stesso indice in un unico set secondo la data a cui si riferiscono, per poi applicare l'algoritmo. In questo lavoro di tesi si è applicata una finestra con passo due. Questo significa che per ogni azione si considerano i suoi dati per un certo giorno e per il giorno precedente ritrovandosi nel caso in cui per predire il valore di un'azione al tempo t_3 , vengono usati i valori relativi agli istanti t_1 e t_2 nella prima riga dei dati di training. Nella seconda riga invece, il valore da predire diventa quello relativo al tempo t_4 e nei dati di training sono utilizzati i valori agli istanti t_2 e t_3 . Dato che l'indice di Standard & Poor's contiene circa cinquecento titoli in ogni set di dati di training sono presenti fino a circa millecinquecento attributi, mentre per l'indice FTSE MIB se ne formano centoventi.

Un esempio di windowing in RapidMiner è il seguente:

Date	label	a1-1	a1-0	b1-1	b1-0	c1-1	c1-0
2011-02-21	1.129	0.761	0.910	0.744	0.395	0.489	0.288
2011-02-22	1.117	0.910	0	0.395	0.279	0.288	0
2011-02-23	1.126	0	0.947	0.279	0	0	0.044
2011-02-24	1.140	0.947	0.308	0	0.209	0.044	0.291
2011-02-25	1.160	0.308	0.478	0.209	0.535	0.291	0.598
2011-02-28	1.156	0.478	1	0.535	1	0.598	1
2011-03-01	1.149	1	0.460	1	0.907	1	0.770
2011-03-02	1.141	0.460	0.505	0.907	0.744	0.770	0.505

Figura 9 - Esempio di Windowing in RapidMiner Studio

Dove:

- a1, b1, c1: sono tre ipotetiche azioni;
- a1-0, b1-0, c1-0: sono i valori che le tre azioni assumono in un certo giorno;
- a1-1, b1-1, c1-1: sono i valori che le tre azioni avevano nel giorno precedente;
- label: è il valore rispetto al quale si vuole applicare la regressione.

Per quanto riguarda invece il numero di periodi presi in considerazione per effettuare la regressione si è impostato questo valore a dieci giorni. Nella Figura 8 di cui sopra, sono riportati però solo otto giorni: questo perché l'ultimo giorno viene usato per effettuare la predizione per generare le raccomandazioni e il primo giorno della finestra viene usato come giorno precedente al secondo, e quindi non compare espressamente.

5.7. I test considerati

Per poter effettuare le predizioni sui mercati sono state considerate diverse configurazioni, in base agli aspetti più importanti legati al trading di azioni.

Per prima cosa si è osservato un grafico tipo di un titolo azionario, ad esempio:



Figura 10 - Grafico giornaliero, Amazon Inc. [33]

Nel grafico in Figura 10 il prezzo è raffigurato con la tecnica che usa le *candlesticks*, anche chiamate candele giapponesi.

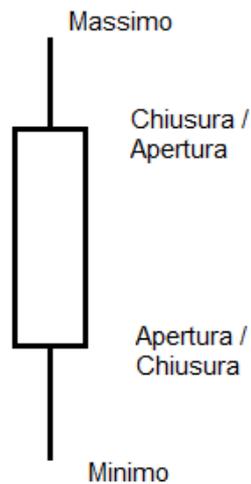


Figura 11 - Esempio di candlestick

In una rappresentazione di questo tipo il prezzo dell'azione, in un certo periodo di tempo, è racchiuso in una superficie rettangolare dove, nei giorni in rialzo il lato superiore rappresenta il prezzo di chiusura ed il lato inferiore rappresenta il prezzo di apertura, viceversa nei giorni di ribasso. Tipicamente viene colorata di verde quando si realizza un profitto e di rosso quando si realizza una perdita.

Dato che un aspetto di questa tesi è anche capire quali indicatori sono i più significativi nella predizione del prezzo dell'azione nel giorno successivo tra i dati forniti da una candlestick si è tenuto solo il prezzo di chiusura.

Un altro elemento fondamentale che si evince dal grafico in Figura 10 è l'importanza del volume, ossia il numero di azioni scambiate in un certo periodo di tempo, che nel caso di cui sopra è di un giorno. È rappresentato sotto forma di rettangoli alla base del grafico ed anche in questo caso si seguono le regole di colore delle candlesticks. Il volume risulta essere importante perché indica la forza di in movimento, se un nuovo trend si verifica in presenza di grandi volumi è più probabile che venga seguito rispetto a casi in cui il movimento non sia appoggiato da grandi investitori.

Per questo motivo si sono considerati casi in cui vengono associati un indicatore alla volta a prezzo di chiusura e volume o solo al prezzo di chiusura, analizzando poi i risultati ottenuti.

Si vogliono considerare due tipi di test, per mettere in evidenza quale configurazione risulta essere la migliore. Da una parte verranno considerati degli ensemble di modelli di regressione generati da sottoinsiemi di dati relativi a specifici indicatori e dall'altra parte sarà considerato un modello di

regressione unico calcolato a partire da un ampio spettro di indicatori, che fotografa quindi una situazione meno specifica del mercato ma più completa.

Una volta generato il modello predittivo relativo ad un certo dataset ed averlo applicato sui dati su cui fare la predizione, si ottengono le raccomandazioni usate per la strategia di trading.

I test considerati comprendono casi in cui vengono usate tutte le raccomandazioni ottenute oppure solo le raccomandazioni in comune tra modelli costruiti a partire da dataset differenti.

In particolare, il processo di intersezione fra le raccomandazioni di due modelli, è formato dai seguenti passi:

- Si costruiscono due dataset formati da alcuni elementi comuni e da un elemento diverso, in Figura 12 questo è raffigurato dai due dataset D1 e D2 che hanno gli attributi X e Y in comune, ma Z e Q differenti;
- Per ogni dataset, ed in questo caso per ogni giornata di trading di cui si vuole predire il valore delle azioni, vengono generati i modelli predittivi ed applicati sui dati di test, ottenendo così per ogni giorno una raccomandazione;
- Per ogni modello predittivo vengono raccolte tutte le raccomandazioni generate, creando i due insiemi di raccomandazioni R1 ed R2, composti da tuple che descrivono per ogni giorno quale azione dovrebbe avere la migliore performance;
- Per ottenere l'intersezione tra i due insiemi di raccomandazioni, si selezionano solamente le azioni che sono state evidenziate dai due diversi modelli predittivi per la stessa giornata di trading.

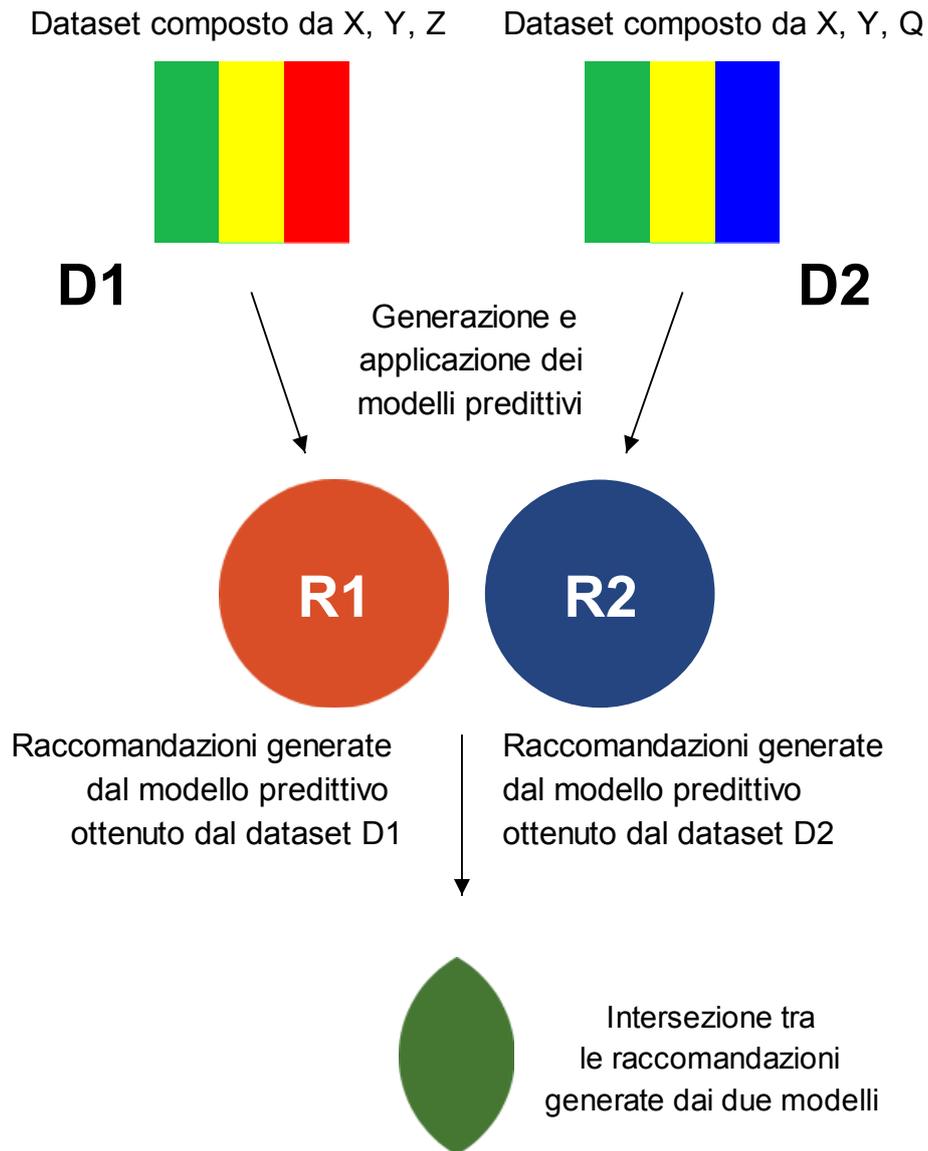


Figura 12 - Descrizione del processo di costruzione delle raccomandazioni

5.7.1. Modello Chiusura, Volume, Indicatore

In questa prima configurazione, per ogni azione viene considerato il set di attributi composto da:

- Prezzo di chiusura giornaliero;
- Volume giornaliero;
- Valore di un indicatore.

Si creano quindi le seguenti combinazioni, di cui si valuteranno in seguito i risultati:

- Chiusura, Volume, MA_5;
- Chiusura, Volume, MA_10;
- Chiusura, Volume, MA_20;
- Chiusura, Volume, MA_50;
- Chiusura, Volume, EMA_5;
- Chiusura, Volume, OBV_5;
- Chiusura, Volume, ATR_5;
- Chiusura, Volume, RSI_5;
- Chiusura, Volume, MFI_5;
- Chiusura, Volume, AROON_5;
- Chiusura, Volume, CCI_5;
- Chiusura, Volume, MOM_5;
- Chiusura, Volume, MACD_5.

5.7.2. Modello Test 2

In questo secondo test si è provato a trovare una configurazione che potesse ottenere risultati buoni cercando di usare modelli generati da dati particolarmente differenti.

Per fare questo si è andata a fare l'intersezione tra le raccomandazioni generate dai dati di training:

- Chiusura, CCI_5;
- Volume, ATR_5.

Dopodiché questo insieme di predizioni è stato ancora ridotto intersecandolo a turno con le raccomandazioni generate da:

- Volume, MA_5;
- Volume, MA_10;
- Volume, MA_20;
- Volume, MA_50;
- Volume, EMA_5;
- Volume, OBV_5;
- Volume, ATR_5;
- Volume, RSI_5;
- Volume, MFI_5;
- Volume, AROON_5;
- Volume, MOM_5;
- Volume, MACD_5.

5.7.3. Modello Chiusura, Volume, Indicatore_X - Chiusura, Volume, Indicatore_Y

In questo test vengono intersecati a due a due gli insiemi delle predizioni ottenute nel modello Chiusura, Volume, Indicatore al fine di identificare raccomandazioni che vengono effettuate da più modelli. Gli indicatori usati per l'intersezione devono essere quindi diversi al fine di effettuare test significativi. In questo caso si creano settantotto set di predizioni, con cardinalità variabile a seconda di quanti elementi comuni i due modelli hanno generato.

5.7.4. Modello Chiusura, Indicatore_X – Chiusura, Indicatore_Y

Questo caso è molto simile al test di intersezione Chiusura, Volume, Indicatore_X - Chiusura, Volume, Indicatore_Y con la differenza che nei dati non viene usato il volume, ma solo il prezzo di chiusura ed un indicatore. Anche in questo caso gli indicatori usati per l'intersezione devono essere quindi diverse al fine di effettuare test significativi. Procedendo in questo modo si creano settantotto set di predizioni, con cardinalità variabile a seconda di quanti elementi comuni i due modelli hanno generato

5.7.5. Modello unico

Siccome l'obiettivo è di confrontare i risultati delle intersezioni delle raccomandazioni di modelli semplici con un singolo modello molto complesso, sono stati preparati i dati per il training utilizzando tutti i dati degli indicatori a disposizione.

Il modello che viene generato usa quindi i dati relativi a:

- Chiusura, Volume, MA_5, MA_10, MA_20, MA_50, EMA_5, OBV_5, ATR_5, RSI_5, MFI_5, AROON_5, CCI_5, MOM_5, MACD_5.

6. Analisi dei risultati

Per ogni modello descritto precedentemente ne è stato calcolato il rendimento usando i seguenti parametri:

- Predizione ad un giorno: si predice il valore che l'azione dovrebbe avere nel giorno successivo all'ultimo giorno presente nei dati di training;
- Stop loss: fissato all'1%, è un valore piuttosto piccolo che lascia poco spazi a bruschi movimenti dell'azione, ma è giustificato dal fatto che la predizione è a breve termine e quindi non si vogliono affrontare grandi perdite giornaliere.

Saranno considerati tutte le combinazioni stato – anno – strategia:

- ITALIA 2011 long;
- ITALIA 2013 long;
- ITALIA 2015 long
- USA 2011 long.
- USA 2013 long;
- USA 2015 long;
- ITALIA 2011 short;
- ITALIA 2013 short;
- ITALIA 2015 short;
- USA 2011 short;
- USA 2013 short;
- USA 2015 short.

Per ognuna di queste combinazioni verranno generate le predizioni usando i tre modelli predittivi di regressione scelti:

- SVM con kernel polinomiale di grado due;
- Regressione Lineare;
- Gradient Boosted Trees.

6.1. Utilizzo della soglia

Siccome la predizione di tipo regressivo genera un valore continuo, si possono fare discriminazioni su quali raccomandazioni considerare. Per questo motivo è stata introdotta una soglia rispetto alla quale si scelgono o scartano le predizioni. I valori considerati per la soglia percentuale sono i numeri

interi compresi tra zero e dieci. Intuitivamente aumentando la soglia il guadagno percentuale per operazione dovrebbe aumentare, in quanto il valore predetto dell'azione è maggiore. In questo caso però il numero di raccomandazioni dovrebbe diminuire, generando così un guadagno totale annuale minore.

6.2. Risultati

Di seguito sono riportati solo alcuni dei risultati ottenuti da tutte le configurazioni usate. Tutti i rendimenti mostrati in questo lavoro di tesi sono da intendere lordi, non sono quindi state applicate le commissioni bancarie che possono essere stimate tra lo 0,16% e lo 0,18%, da sottrarre al guadagno percentuale per operazione.

6.2.1. Modello Chiusura, Volume, Indicatore_X – USA 2015 long

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia allo 0 %

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	250	249	249	109.53	-5.95	38.24	0.438	-0.024	0.154
MA_10	250	249	249	117.84	0.22	28.02	0.471	0.001	0.113
MA_20	250	249	248	124.9	4.92	81.96	0.5	0.02	0.33
MA_50	250	249	248	125.95	30.64	95.06	0.504	0.123	0.383
EMA_5	250	250	248	91.32	-6.92	73.2	0.365	-0.028	0.295
OBV_5	250	248	248	118.85	36.01	63.47	0.475	0.145	0.256
ATR_5	250	250	249	80.76	0.76	46.95	0.323	0.003	0.189
RSI_5	250	248	249	110.3	70.42	61.28	0.441	0.284	0.246
MFI_5	250	248	248	106.31	34.31	70.9	0.425	0.138	0.286
AROON_5	250	250	249	160.12	44.9	63.84	0.64	0.18	0.256
CCI_5	250	249	249	121.79	11.9	126.35	0.487	0.048	0.507
MOM_5	250	248	248	129.03	31.23	24.41	0.516	0.126	0.098
MACD_5	250	249	249	128.38	18.31	50.6	0.514	0.074	0.203

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia all'1%

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	249	239	249	110.53	-10.97	38.24	0.444	-0.046	0.154
MA_10	249	235	249	118.84	-3.83	28.02	0.477	-0.016	0.113
MA_20	250	236	248	124.9	0.62	81.96	0.5	0.003	0.33
MA_50	250	235	248	125.95	37.18	95.06	0.504	0.158	0.383
EMA_5	249	243	248	92.32	-4.82	73.2	0.371	-0.02	0.295
OBV_5	249	248	248	119.85	36.01	63.47	0.481	0.145	0.256
ATR_5	249	244	249	81.76	-2.79	46.95	0.328	-0.011	0.189
RSI_5	250	248	249	110.3	70.42	61.28	0.441	0.284	0.246
MFI_5	249	248	248	107.31	34.31	70.9	0.431	0.138	0.286
AROON_5	249	250	249	161.12	44.9	63.84	0.647	0.18	0.256
CCI_5	249	249	249	119.22	11.9	126.35	0.479	0.048	0.507
MOM_5	249	248	248	126.46	31.23	24.41	0.508	0.126	0.098
MACD_5	250	244	249	128.38	12.25	50.6	0.514	0.05	0.203

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 2 %

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	245	180	244	110.02	-0.99	37.68	0.449	-0.005	0.154
MA_10	243	175	238	120.28	17.7	30.23	0.495	0.101	0.127
MA_20	241	167	234	124.29	9.48	89.1	0.516	0.057	0.381
MA_50	245	171	235	129.51	32.97	98.37	0.529	0.193	0.419
EMA_5	246	188	244	93.78	10.87	74.39	0.381	0.058	0.305
OBV_5	245	229	247	123.53	32.04	64.47	0.504	0.14	0.261
ATR_5	243	203	246	86.76	16.12	49.95	0.357	0.079	0.203
RSI_5	245	231	249	111.4	62.59	61.28	0.455	0.271	0.246
MFI_5	246	223	245	104.98	35.54	73.9	0.427	0.159	0.302
AROON_5	242	234	249	164.75	45.86	63.84	0.681	0.196	0.256
CCI_5	244	238	249	121.25	11.3	126.35	0.497	0.047	0.507
MOM_5	244	233	248	126.13	28.2	24.41	0.517	0.121	0.098
MACD_5	245	207	243	117.62	19.3	54.68	0.48	0.093	0.225

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 3 %

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	233	130	213	111.62	11.14	25.08	0.479	0.086	0.118
MA_10	231	108	194	115.86	33.28	11.34	0.502	0.308	0.058
MA_20	233	106	191	121.58	27.48	63.89	0.522	0.259	0.335
MA_50	232	108	192	123.17	28.14	86.84	0.531	0.261	0.452
EMA_5	236	136	209	93.67	16.04	71.37	0.397	0.118	0.341
OBV_5	240	200	232	122.49	28.3	73.98	0.51	0.142	0.319
ATR_5	234	150	224	88.12	19.45	31.87	0.377	0.13	0.142
RSI_5	241	203	233	110.46	55.47	67.74	0.458	0.273	0.291
MFI_5	242	184	231	100.33	36.98	78.95	0.415	0.201	0.342
AROON_5	238	201	233	159.85	17.58	50.86	0.672	0.087	0.218
CCI_5	242	208	234	116.39	27.93	128.87	0.481	0.134	0.551
MOM_5	241	204	235	125.96	35.51	32.84	0.523	0.174	0.14
MACD_5	239	157	219	117.21	30.73	54.26	0.49	0.196	0.248

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 4 %

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	224	85	170	110.9	18.07	18.73	0.495	0.213	0.11
MA_10	222	79	155	120.47	33.66	20.44	0.543	0.426	0.132
MA_20	221	65	147	126.72	24.58	68.88	0.573	0.378	0.469
MA_50	215	75	136	119.4	29.34	62.65	0.555	0.391	0.461
EMA_5	226	97	176	91.41	25.93	39.88	0.404	0.267	0.227
OBV_5	231	160	199	125.48	18.53	65.39	0.543	0.116	0.329
ATR_5	223	101	195	94.56	18.2	11.64	0.424	0.18	0.06
RSI_5	231	159	211	110.79	40.28	64.05	0.48	0.253	0.304
MFI_5	233	142	210	103.76	18.34	74.4	0.445	0.129	0.354
AROON_5	226	169	214	153.12	23.04	51.82	0.678	0.136	0.242
CCI_5	232	179	212	119.7	8.93	120.84	0.516	0.05	0.57
MOM_5	231	161	215	128.41	19.62	30.45	0.556	0.122	0.142
MACD_5	226	107	191	121.34	44.78	29.57	0.537	0.419	0.155

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 5 %

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	209	58	136	114.83	17.84	11.06	0.549	0.308	0.081
MA_10	206	51	114	118.88	22.2	27.03	0.577	0.435	0.237
MA_20	205	48	108	123.24	30.01	45.78	0.601	0.625	0.424
MA_50	203	46	98	117.95	25.76	60.15	0.581	0.56	0.614
EMA_5	211	69	141	99.31	18.01	29.83	0.471	0.261	0.212
OBV_5	220	119	179	133.14	23.27	62.71	0.605	0.196	0.35
ATR_5	203	76	152	98.26	17.27	11.27	0.484	0.227	0.074
RSI_5	220	113	182	120.0	39.51	54.96	0.545	0.35	0.302
MFI_5	219	107	172	105.47	25.9	66.56	0.482	0.242	0.387
AROON_5	216	140	179	161.67	19.56	51.71	0.748	0.14	0.289
CCI_5	223	139	190	118.52	19.0	88.53	0.531	0.137	0.466
MOM_5	221	121	173	132.31	26.96	0.44	0.599	0.223	0.003
MACD_5	210	79	157	125.29	38.33	18.81	0.597	0.485	0.12

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 6 %

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	190	47	108	122.05	16.66	19.23	0.642	0.355	0.178
MA_10	187	38	88	122.55	21.1	29.16	0.655	0.555	0.331
MA_20	187	37	75	124.56	24.18	31.88	0.666	0.653	0.425
MA_50	183	37	63	123.51	17.05	54.37	0.675	0.461	0.863
EMA_5	191	54	111	103.18	18.55	45.16	0.54	0.344	0.407
OBV_5	201	94	140	135.49	24.31	48.99	0.674	0.259	0.35
ATR_5	184	62	127	101.77	18.68	13.16	0.553	0.301	0.104
RSI_5	202	84	147	121.35	36.87	60.5	0.601	0.439	0.412
MFI_5	203	85	146	107.43	24.17	41.17	0.529	0.284	0.282
AROON_5	200	111	147	166.96	16.94	27.45	0.835	0.153	0.187
CCI_5	205	109	172	121.41	33.41	84.45	0.592	0.306	0.491
MOM_5	201	87	143	136.66	15.35	-2.19	0.68	0.176	-0.015
MACD_5	194	60	126	125.14	34.68	22.13	0.645	0.578	0.176

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 7 %

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	165	39	80	102.81	22.49	20.04	0.623	0.577	0.251
MA_10	165	31	75	124.1	16.18	22.57	0.752	0.522	0.301
MA_20	162	32	59	128.74	17.84	40.75	0.795	0.558	0.691
MA_50	159	26	55	120.69	15.96	60.19	0.759	0.614	1.094
EMA_5	174	41	84	105.62	18.86	35.01	0.607	0.46	0.417
OBV_5	185	72	116	146.6	29.81	46.45	0.792	0.414	0.4
ATR_5	166	47	101	100.2	25.62	3.94	0.604	0.545	0.039
RSI_5	188	69	118	130.46	37.26	25.27	0.694	0.54	0.214
MFI_5	192	62	120	112.06	22.06	45.35	0.584	0.356	0.378
AROON_5	186	91	122	164.87	23.24	34.15	0.886	0.255	0.28
CCI_5	193	89	132	128.38	23.91	97.98	0.665	0.269	0.742
MOM_5	187	70	120	142.11	22.15	12.97	0.76	0.316	0.108
MACD_5	176	49	96	133.96	18.18	13.19	0.761	0.371	0.137

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia all'8 %

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	147	32	73	82.74	20.44	23.85	0.563	0.639	0.327
MA_10	148	28	59	97.18	14.31	23.14	0.657	0.511	0.392
MA_20	150	25	50	95.06	21.36	42.5	0.634	0.854	0.85
MA_50	144	24	48	96.96	15.48	67.19	0.673	0.645	1.4
EMA_5	156	33	75	104.1	19.92	37.96	0.667	0.604	0.506
OBV_5	163	57	101	108.53	29.39	40.94	0.666	0.516	0.405
ATR_5	150	40	82	71.99	29.41	12.47	0.48	0.735	0.152
RSI_5	169	59	103	121.9	40.82	27.6	0.721	0.692	0.268
MFI_5	171	55	99	88.06	17.52	29.73	0.515	0.319	0.3
AROON_5	168	79	101	138.21	23.97	33.43	0.823	0.303	0.331
CCI_5	174	70	114	131.46	24.07	75.66	0.756	0.344	0.664
MOM_5	171	60	102	134.06	24.13	20.64	0.784	0.402	0.202
MACD_5	159	45	80	127.25	15.84	23.22	0.8	0.352	0.29

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 9 %

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	138	27	66	82.92	18.04	30.85	0.601	0.668	0.467
MA_10	133	24	48	99.51	17.49	26.66	0.748	0.729	0.555
MA_20	133	22	45	98.71	21.92	39.53	0.742	0.996	0.878
MA_50	129	20	42	85.85	16.21	54.13	0.665	0.811	1.289
EMA_5	139	30	65	78.04	16.57	40.72	0.561	0.552	0.627
OBV_5	148	48	86	107.5	32.69	44.1	0.726	0.681	0.513
ATR_5	136	36	69	65.54	23.93	16.23	0.482	0.665	0.235
RSI_5	150	47	85	93.39	31.49	25.79	0.623	0.67	0.303
MFI_5	158	45	86	64.07	16.69	36.03	0.406	0.371	0.419
AROON_5	151	66	89	129.01	27.3	32.09	0.854	0.414	0.361
CCI_5	158	59	100	102.32	16.47	67.86	0.648	0.279	0.679
MOM_5	152	46	89	103.5	31.44	20.2	0.681	0.683	0.227
MACD_5	141	36	68	103.08	21.27	29.85	0.731	0.591	0.439

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 10 %

X	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	124	26	53	84.45	19.04	37.15	0.681	0.732	0.701
MA_10	117	22	40	93.38	19.49	9.25	0.798	0.886	0.231
MA_20	116	21	39	89.27	20.49	36.15	0.77	0.976	0.927
MA_50	114	18	34	91.29	18.21	30.84	0.801	1.012	0.907
EMA_5	124	25	56	80.77	20.75	18.07	0.651	0.83	0.323
OBV_5	140	41	78	106.75	28.77	36.11	0.762	0.702	0.463
ATR_5	119	33	58	68.37	21.66	21.15	0.575	0.656	0.365
RSI_5	139	41	75	93.08	29.14	32.24	0.67	0.711	0.43
MFI_5	143	37	77	68.37	22.32	39.3	0.478	0.603	0.51
AROON_5	138	55	78	122.29	25.65	34.56	0.886	0.466	0.443
CCI_5	145	57	81	98.94	11.75	73.22	0.682	0.206	0.904
MOM_5	140	38	74	103.92	30.65	17.12	0.742	0.807	0.231
MACD_5	126	28	63	100.97	19.71	24.99	0.801	0.704	0.397

6.2.2. Modello Chiusura, Indicatore_X – Chiusura, Indicatore_Y – USA 2011 short

Vengono riportati solo gli esiti relativi a parte delle intersezioni considerate.

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia allo 0 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	163	129	116	42.44	93.37	94.04	0.26	0.724	0.811
MA_5	MA_20	153	127	109	79.03	104.27	79.0	0.517	0.821	0.725
MA_5	MA_50	147	107	104	66.99	74.82	61.48	0.456	0.699	0.591
MA_5	EMA_5	207	167	125	83.73	121.41	91.05	0.404	0.727	0.728
MA_5	OBV_5	149	72	94	59.95	62.41	88.92	0.402	0.867	0.946
MA_5	ATR_5	148	141	108	87.96	78.69	83.2	0.594	0.558	0.77
MA_5	RSI_5	144	73	106	71.63	55.61	66.1	0.497	0.762	0.624
MA_5	MFI_5	152	76	101	84.42	61.96	71.38	0.555	0.815	0.707
MA_5	AROON_5	139	47	93	97.86	41.08	73.28	0.704	0.874	0.788
MA_5	CCI_5	134	53	80	105.04	32.39	65.26	0.784	0.611	0.816
MA_5	MOM_5	148	79	99	69.23	58.16	85.37	0.468	0.736	0.862
MA_5	MACD_5	174	147	101	82.0	120.77	58.84	0.471	0.822	0.583

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia all'1 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	163	127	115	42.44	94.77	93.25	0.26	0.746	0.811
MA_5	MA_20	153	125	108	79.03	102.1	78.21	0.517	0.817	0.724
MA_5	MA_50	147	106	104	66.99	75.22	61.48	0.456	0.71	0.591
MA_5	EMA_5	207	164	121	83.73	123.81	92.71	0.404	0.755	0.766
MA_5	OBV_5	149	70	92	59.95	64.41	90.36	0.402	0.92	0.982
MA_5	ATR_5	148	138	107	87.96	81.1	83.65	0.594	0.588	0.782
MA_5	RSI_5	143	70	105	72.63	58.02	66.55	0.508	0.829	0.634
MA_5	MFI_5	152	73	100	84.42	64.36	71.83	0.555	0.882	0.718
MA_5	AROON_5	138	47	93	98.86	41.08	73.28	0.716	0.874	0.788
MA_5	CCI_5	133	52	79	106.04	32.79	65.71	0.797	0.631	0.832
MA_5	MOM_5	147	77	99	70.23	59.57	85.37	0.478	0.774	0.862
MA_5	MACD_5	174	143	100	82.0	121.78	59.84	0.471	0.852	0.598

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 2 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	158	106	104	35.89	84.0	89.26	0.227	0.792	0.858
MA_5	MA_20	151	102	95	81.03	85.78	73.74	0.537	0.841	0.776
MA_5	MA_50	146	96	95	67.99	65.14	59.51	0.466	0.679	0.626
MA_5	EMA_5	207	132	111	83.73	111.57	88.57	0.404	0.845	0.798
MA_5	OBV_5	149	61	86	59.95	68.43	88.01	0.402	1.122	1.023
MA_5	ATR_5	144	121	99	89.14	74.52	80.38	0.619	0.616	0.812
MA_5	RSI_5	143	63	97	72.63	60.98	70.09	0.508	0.968	0.723
MA_5	MFI_5	152	64	91	84.42	64.61	70.89	0.555	1.009	0.779
MA_5	AROON_5	138	38	88	98.86	30.66	75.63	0.716	0.807	0.859
MA_5	CCI_5	133	45	69	106.04	31.81	63.37	0.797	0.707	0.918
MA_5	MOM_5	147	67	94	70.23	65.53	87.72	0.478	0.978	0.933
MA_5	MACD_5	174	113	91	82.0	112.78	59.83	0.471	0.998	0.657

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 3 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	150	73	69	35.74	77.49	67.91	0.238	1.062	0.984
MA_5	MA_20	138	73	66	75.02	74.33	58.12	0.544	1.018	0.881
MA_5	MA_50	136	62	69	74.64	59.08	46.56	0.549	0.953	0.675
MA_5	EMA_5	196	98	89	88.0	90.84	73.18	0.449	0.927	0.822
MA_5	OBV_5	143	47	66	60.61	52.25	59.56	0.424	1.112	0.902
MA_5	ATR_5	136	86	79	93.18	70.74	61.6	0.685	0.823	0.78
MA_5	RSI_5	139	49	77	72.33	45.18	45.74	0.52	0.922	0.594
MA_5	MFI_5	146	48	76	88.08	57.05	46.35	0.603	1.189	0.61
MA_5	AROON_5	134	31	70	97.53	32.13	55.24	0.728	1.036	0.789
MA_5	CCI_5	129	39	57	107.7	29.47	57.93	0.835	0.756	1.016
MA_5	MOM_5	144	54	74	68.93	47.17	72.63	0.479	0.874	0.981
MA_5	MACD_5	168	85	73	76.78	98.32	52.54	0.457	1.157	0.72

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 4 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	131	55	44	44.55	86.56	58.54	0.34	1.574	1.33
MA_5	MA_20	126	56	43	75.11	81.29	57.47	0.596	1.452	1.337
MA_5	MA_50	124	48	49	73.86	62.72	46.44	0.596	1.307	0.948
MA_5	EMA_5	171	72	59	90.27	105.73	52.47	0.528	1.468	0.889
MA_5	OBV_5	128	36	47	59.17	44.97	49.64	0.462	1.249	1.056
MA_5	ATR_5	125	59	48	89.14	60.56	59.79	0.713	1.026	1.246
MA_5	RSI_5	127	33	49	75.0	35.0	30.08	0.591	1.061	0.614
MA_5	MFI_5	130	34	52	84.73	54.36	34.29	0.652	1.599	0.659
MA_5	AROON_5	120	23	50	106.85	33.15	52.93	0.89	1.441	1.059
MA_5	CCI_5	116	32	41	110.06	27.55	52.05	0.949	0.861	1.269
MA_5	MOM_5	129	41	51	75.27	44.27	66.1	0.583	1.08	1.296
MA_5	MACD_5	149	61	47	79.22	103.52	53.5	0.532	1.697	1.138

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 5 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	120	44	33	44.94	69.51	52.44	0.375	1.58	1.589
MA_5	MA_20	109	44	29	49.38	62.82	56.63	0.453	1.428	1.953
MA_5	MA_50	109	36	30	62.53	50.71	48.18	0.574	1.409	1.606
MA_5	EMA_5	152	57	41	63.17	105.17	54.28	0.416	1.845	1.324
MA_5	OBV_5	109	29	30	48.85	45.22	39.75	0.448	1.559	1.325
MA_5	ATR_5	108	44	35	56.67	54.59	40.63	0.525	1.241	1.161
MA_5	RSI_5	108	25	32	63.11	34.22	35.65	0.584	1.369	1.114
MA_5	MFI_5	111	31	33	70.42	55.14	29.6	0.634	1.779	0.897
MA_5	AROON_5	104	15	35	93.59	18.91	51.4	0.9	1.261	1.469
MA_5	CCI_5	96	20	29	95.5	7.62	42.45	0.995	0.381	1.464
MA_5	MOM_5	112	30	34	63.05	37.15	57.51	0.563	1.238	1.691
MA_5	MACD_5	130	53	33	64.45	102.83	47.33	0.496	1.94	1.434

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 6 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	102	41	20	51.41	70.05	55.14	0.504	1.709	2.757
MA_5	MA_20	90	41	20	55.53	54.29	47.4	0.617	1.324	2.37
MA_5	MA_50	92	34	20	67.21	41.19	42.63	0.731	1.211	2.131
MA_5	EMA_5	132	48	30	73.61	85.8	48.07	0.558	1.787	1.602
MA_5	OBV_5	99	25	19	52.67	36.07	34.99	0.532	1.443	1.842
MA_5	ATR_5	91	40	25	54.51	47.07	43.07	0.599	1.177	1.723
MA_5	RSI_5	98	21	21	64.97	35.42	30.89	0.663	1.687	1.471
MA_5	MFI_5	101	26	22	74.86	51.18	22.15	0.741	1.969	1.007
MA_5	AROON_5	92	12	22	100.04	21.91	53.14	1.087	1.826	2.416
MA_5	CCI_5	84	16	20	96.53	-2.42	41.01	1.149	-0.151	2.05
MA_5	MOM_5	102	26	23	67.58	35.39	57.07	0.663	1.361	2.481
MA_5	MACD_5	114	43	25	39.93	73.44	45.89	0.35	1.708	1.836

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 7 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	88	34	16	57.71	61.96	41.3	0.656	1.822	2.582
MA_5	MA_20	74	36	18	59.88	59.29	39.3	0.809	1.647	2.184
MA_5	MA_50	72	29	14	62.51	46.19	23.55	0.868	1.593	1.682
MA_5	EMA_5	110	39	24	61.58	87.75	42.98	0.56	2.25	1.791
MA_5	OBV_5	81	20	16	57.33	39.12	29.25	0.708	1.956	1.828
MA_5	ATR_5	78	26	17	59.69	37.75	26.22	0.765	1.452	1.542
MA_5	RSI_5	80	16	16	52.89	20.45	26.15	0.661	1.278	1.634
MA_5	MFI_5	82	18	20	46.72	33.45	24.15	0.57	1.858	1.207
MA_5	AROON_5	75	7	18	85.75	3.83	41.39	1.143	0.548	2.299
MA_5	CCI_5	71	12	12	75.57	-0.37	32.24	1.064	-0.031	2.687
MA_5	MOM_5	82	21	19	53.38	25.28	47.01	0.651	1.204	2.474
MA_5	MACD_5	91	33	19	48.45	76.39	40.46	0.532	2.315	2.129

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia all'8 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	69	26	14	52.1	46.31	28.26	0.755	1.781	2.019
MA_5	MA_20	61	29	14	58.68	34.11	28.26	0.962	1.176	2.019
MA_5	MA_50	62	25	12	54.46	44.03	10.51	0.878	1.761	0.876
MA_5	EMA_5	92	30	14	49.74	63.77	27.13	0.541	2.126	1.938
MA_5	OBV_5	63	16	13	35.62	26.15	31.25	0.565	1.634	2.404
MA_5	ATR_5	62	21	12	43.87	23.33	28.22	0.708	1.111	2.351
MA_5	RSI_5	60	13	12	36.68	23.45	29.15	0.611	1.804	2.429
MA_5	MFI_5	61	16	11	36.62	25.36	31.15	0.6	1.585	2.832
MA_5	AROON_5	58	6	11	55.96	4.83	46.39	0.965	0.806	4.217
MA_5	CCI_5	55	8	8	48.53	2.83	36.24	0.882	0.354	4.53
MA_5	MOM_5	61	18	12	35.68	28.28	37.96	0.585	1.571	3.164
MA_5	MACD_5	75	25	13	33.58	45.25	31.41	0.448	1.81	2.416

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 9 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	56	23	10	47.56	35.45	10.6	0.849	1.541	1.06
MA_5	MA_20	46	24	10	26.25	19.43	11.51	0.571	0.81	1.151
MA_5	MA_50	52	18	9	48.72	27.08	11.6	0.937	1.505	1.289
MA_5	EMA_5	73	24	12	49.62	42.74	8.38	0.68	1.781	0.698
MA_5	OBV_5	49	14	10	37.86	20.48	12.71	0.773	1.463	1.271
MA_5	ATR_5	50	17	9	24.23	9.16	9.18	0.485	0.539	1.02
MA_5	RSI_5	48	12	9	43.0	24.45	10.61	0.896	2.038	1.178
MA_5	MFI_5	50	15	8	42.95	19.48	12.61	0.859	1.299	1.576
MA_5	AROON_5	47	4	8	59.61	-1.98	14.59	1.268	-0.496	1.824
MA_5	CCI_5	47	8	6	51.85	2.83	17.49	1.103	0.354	2.915
MA_5	MOM_5	50	17	9	41.0	29.28	19.42	0.82	1.723	2.158
MA_5	MACD_5	61	18	9	32.6	16.46	13.87	0.534	0.914	1.541

Risultati ottenuti selezionando il migliore titolo azionario per ogni giornata con soglia al 10 %

X	Y	Numero Operazioni			Totale %			Guadagno % per Operazione		
		SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
MA_5	MA_10	43	16	6	15.28	-0.53	9.14	0.355	-0.033	1.523
MA_5	MA_20	36	19	5	13.79	2.45	7.15	0.383	0.129	1.43
MA_5	MA_50	45	15	5	32.1	14.26	10.14	0.713	0.95	2.027
MA_5	EMA_5	58	21	6	32.62	15.92	9.19	0.562	0.758	1.532
MA_5	OBV_5	41	10	7	24.06	-5.34	1.66	0.587	-0.534	0.238
MA_5	ATR_5	38	15	7	7.76	-4.66	11.18	0.204	-0.311	1.597
MA_5	RSI_5	39	10	8	23.17	17.63	11.61	0.594	1.763	1.451
MA_5	MFI_5	41	12	6	23.06	-7.34	1.66	0.562	-0.611	0.277
MA_5	AROON_5	39	4	7	48.75	-1.98	11.7	1.25	-0.496	1.671
MA_5	CCI_5	36	6	4	36.91	-3.98	2.75	1.025	-0.664	0.688
MA_5	MOM_5	42	12	8	23.06	11.64	20.42	0.549	0.97	2.553
MA_5	MACD_5	52	14	6	14.74	11.64	16.87	0.283	0.831	2.812

6.2.3. Modello unico

- ITALIA 2013 long

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	255	255	255	57.32	31.12	64.42	0.225	0.122	0.253
1	226	116	196	51.49	19.08	68.97	0.228	0.164	0.352
2	175	45	111	44.3	14.96	51.39	0.253	0.333	0.463
3	119	18	47	17.03	15.81	22.96	0.143	0.878	0.488
4	76	10	27	12.67	4.85	17.56	0.167	0.485	0.65
5	47	6	12	5.8	-1.13	10.05	0.123	-0.188	0.837
6	34	4	8	-1.74	0.87	5.26	-0.051	0.218	0.657
7	21	3	5	0.63	1.87	0.45	0.03	0.623	0.09
8	17	3	4	3.71	1.87	1.45	0.219	0.623	0.362
9	13	3	3	5.34	1.87	2.45	0.411	0.623	0.816
10	9	3	3	7.07	1.87	2.45	0.785	0.623	0.816

- USA 2013 long

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	251	252	251	89.43	43.95	56.82	0.356	0.174	0.226
1	248	223	250	87.22	42.87	56.6	0.352	0.192	0.226
2	230	113	218	87.58	6.56	42.36	0.381	0.058	0.194
3	218	59	158	85.81	15.38	24.24	0.394	0.261	0.153
4	202	23	110	92.43	13.15	11.11	0.458	0.572	0.101
5	183	18	80	78.71	7.4	23.35	0.43	0.411	0.292
6	168	15	56	67.2	4.82	23.74	0.4	0.321	0.424
7	141	15	46	45.54	4.82	18.73	0.323	0.321	0.407
8	130	14	40	45.14	5.82	15.12	0.347	0.415	0.378
9	118	13	31	32.89	6.82	11.71	0.279	0.524	0.378
10	102	13	28	17.7	6.82	14.71	0.173	0.524	0.525

- ITALIA 2011 short

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	255	255	255	57.32	31.12	64.42	0.225	0.122	0.253
1	180	142	196	51.3	18.25	45.6	0.285	0.128	0.233
2	110	71	126	57.2	3.69	16.1	0.52	0.052	0.128
3	70	38	63	-3.09	4.78	11.71	-0.044	0.126	0.186
4	47	21	42	-5.31	8.56	-1.97	-0.113	0.408	-0.047
5	35	8	24	0.47	7.42	-1.39	0.013	0.927	-0.058
6	18	3	12	8.82	-1.17	-12.0	0.49	-0.39	-1.0
7	16	2	6	5.25	-0.17	-6.0	0.328	-0.085	-1.0
8	12	1	3	0.78	0.83	-3.0	0.065	0.829	-1.0
9	8	0	3	-5.48	0	-3.0	-0.685	0	-1.0
10	6	0	2	-3.48	0	-2.0	-0.58	0	-1.0

- USA 2011 short

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	250	251	250	196.42	115.19	127.6	0.786	0.459	0.51
1	250	225	245	196.42	112.56	130.26	0.786	0.5	0.532
2	248	162	224	198.42	127.18	134.39	0.8	0.785	0.6
3	235	96	161	202.72	106.13	102.04	0.863	1.106	0.634
4	225	66	106	200.19	83.81	73.18	0.89	1.27	0.69
5	197	49	66	155.79	38.24	57.21	0.791	0.78	0.867
6	178	35	39	132.66	37.31	50.93	0.745	1.066	1.306
7	149	27	23	127.64	29.68	21.75	0.857	1.099	0.946
8	120	21	18	129.76	25.72	9.4	1.081	1.225	0.522
9	105	15	8	112.37	22.24	10.07	1.07	1.483	1.259
10	88	13	8	90.23	6.49	10.07	1.025	0.499	1.259

- USA 2015 long

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	250	249	249	82.85	20.16	62.34	0.331	0.081	0.25
1	250	226	249	82.85	16.48	62.34	0.331	0.073	0.25
2	241	147	233	81.04	18.68	57.63	0.336	0.127	0.247
3	228	83	195	81.89	10.59	49.25	0.359	0.128	0.253
4	219	54	151	80.6	28.5	39.33	0.368	0.528	0.26
5	201	40	101	89.84	33.12	35.44	0.447	0.828	0.351
6	186	34	74	67.57	29.59	42.68	0.363	0.87	0.577
7	158	28	58	57.35	27.61	32.82	0.363	0.986	0.566
8	146	25	49	59.06	29.64	41.82	0.405	1.186	0.853
9	136	24	40	58.46	26.73	19.92	0.43	1.114	0.498
10	128	20	37	60.11	20.37	21.0	0.47	1.019	0.567

- ITALIA 2015 long

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	250	248	250	54.09	53.49	76.64	0.216	0.216	0.307
1	210	130	182	45.43	20.45	63.19	0.216	0.157	0.347
2	151	45	108	36.2	12.68	18.24	0.24	0.282	0.169
3	103	15	56	38.07	4.51	16.09	0.37	0.3	0.287
4	75	8	26	41.8	8.47	4.77	0.557	1.059	0.184
5	60	2	11	37.41	2.35	0.57	0.624	1.174	0.052
6	38	1	5	37.89	-1.0	1.48	0.997	-1.0	0.296
7	30	0	2	36.86	0	-2.0	1.229	0	-1.0
8	23	0	0	32.12	0	0	1.397	0	0
9	19	0	0	23.03	0	0	1.212	0	0
10	13	0	0	19.73	0	0	1.518	0	0

- ITALIA 2013 short

Soglia	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	250	251	251	-0.52	-12.86	-23.68	-0.002	-0.051	-0.094
1	185	131	211	6.11	12.13	-15.48	0.033	0.093	-0.073
2	115	50	140	-4.45	6.76	-11.52	-0.039	0.135	-0.082
3	60	19	61	-5.7	11.45	3.86	-0.095	0.602	0.063
4	34	11	27	-3.47	1.91	-3.13	-0.102	0.174	-0.116
5	20	5	6	-9.1	-2.29	-1.44	-0.455	-0.457	-0.239
6	15	4	4	-4.1	-4.0	0.56	-0.274	-1.0	0.141
7	11	0	3	-4.86	0	1.56	-0.442	0	0.522
8	8	0	2	-3.17	0	0.09	-0.396	0	0.045
9	7	0	1	-2.17	0	-1.0	-0.309	0	-1.0
10	6	0	1	-1.17	0	-1.0	-0.194	0	-1.0

- USA 2013 short

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	252	252	252	3.48	-25.52	-7.32	0.014	-0.101	-0.029
1	252	228	252	3.48	-16.54	-7.32	0.014	-0.073	-0.029
2	252	106	227	3.48	11.04	2.44	0.014	0.104	0.011
3	240	44	147	3.7	-0.0	-7.24	0.015	-0.0	-0.049
4	218	19	64	6.41	-2.42	-5.71	0.029	-0.127	-0.089
5	182	14	29	20.44	1.14	10.65	0.112	0.082	0.367
6	144	10	14	15.46	1.03	5.92	0.107	0.103	0.423
7	121	7	9	11.26	0.11	7.13	0.093	0.015	0.792
8	90	7	7	18.87	0.11	7.99	0.21	0.015	1.141
9	77	7	6	19.83	0.11	8.03	0.258	0.015	1.338
10	60	6	2	21.28	-0.07	-0.4	0.355	-0.012	-0.198

- ITALIA 2011 long

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	254	249	252	43.33	28.75	13.02	0.171	0.115	0.052
1	221	143	204	43.71	10.93	17.65	0.198	0.076	0.087
2	168	84	128	50.4	13.48	9.03	0.3	0.161	0.071
3	136	47	91	41.29	13.97	0.3	0.304	0.297	0.003
4	111	31	67	19.85	11.92	6.19	0.179	0.385	0.092
5	94	15	42	19.03	10.0	4.34	0.202	0.667	0.103
6	80	11	31	13.42	-3.38	7.48	0.168	-0.307	0.241
7	67	8	17	2.54	-0.38	0.08	0.038	-0.048	0.005
8	58	5	14	-18.1	2.62	3.08	-0.312	0.524	0.22
9	49	3	11	-16.1	4.62	3.02	-0.329	1.539	0.274
10	42	2	6	-20.89	-2.0	0.4	-0.497	-1.0	0.067

- USA 2011 long

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	249	248	249	137.53	134.38	172.18	0.552	0.542	0.691
1	249	224	248	137.53	112.47	172.24	0.552	0.502	0.695
2	241	165	232	139.54	103.76	175.68	0.579	0.629	0.757
3	233	103	192	141.59	66.32	169.74	0.608	0.644	0.884
4	215	75	146	149.85	62.43	150.49	0.697	0.832	1.031
5	199	48	113	130.63	52.5	135.03	0.656	1.094	1.195
6	178	34	83	122.53	40.88	114.72	0.688	1.202	1.382
7	163	29	65	122.03	44.24	60.69	0.749	1.526	0.934
8	149	25	47	105.4	36.92	49.69	0.707	1.477	1.057
9	138	22	40	108.88	30.29	41.19	0.789	1.377	1.03
10	127	19	32	116.96	25.73	32.01	0.921	1.354	1.0

- USA 2015 short

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	250	249	247	47.07	55.28	33.53	0.188	0.222	0.136
1	250	232	245	47.07	53.78	32.87	0.188	0.232	0.134
2	249	140	217	44.23	43.27	30.22	0.178	0.309	0.139
3	244	74	148	43.64	1.83	30.89	0.179	0.16	0.209
4	225	42	84	44.5	-1.98	21.86	0.198	-0.047	0.26
5	205	26	47	41.48	-2.42	11.4	0.202	-0.093	0.243
6	188	19	25	36.49	-7.03	5.49	0.194	-0.37	0.22
7	165	13	17	34.65	-1.03	5.17	0.21	-0.079	0.304
8	144	12	11	12.64	-1.55	8.24	0.088	-0.129	0.749
9	124	12	6	8.77	-1.55	4.9	0.071	-0.129	0.817
10	113	11	4	8.09	-0.55	2.05	0.072	-0.05	0.512

- ITALIA 2015 short

Soglia %	Numero Operazioni			Totale %			Guadagno % per Operazione		
	SVM	LIN	GBT	SVM	LIN	GBT	SVM	LIN	GBT
0	253	253	253	8.05	29.62	20.7	0.032	0.117	0.082
1	176	135	198	1.7	21.37	17.34	0.01	0.158	0.088
2	74	37	115	-12.39	19.28	14.09	-0.167	0.521	0.123
3	36	11	47	-1.93	8.1	13.33	-0.054	0.736	0.284
4	14	4	18	3.26	-0.01	6.6	0.233	-0.002	0.367
5	7	4	9	8.57	-0.01	10.1	1.224	-0.002	1.122
6	7	2	1	8.57	-2.0	-1.0	1.224	-1.0	-1.0
7	5	1	0	2.71	-1.0	0	0.543	-1.0	0
8	4	1	0	2.25	-1.0	0	0.562	-1.0	0
9	2	0	0	-2.0	0	0	-1.0	0	0
10	2	0	0	-2.0	0	0	-1.0	0	0

6.3. Analisi dell'utilizzo della soglia

Qualitativamente si osserva che, come previsto, il numero di operazioni consigliate diminuisce all'aumentare della soglia, e sembra che il guadagno percentuale medio per operazione cresca al crescere della soglia.

I seguenti due grafici sono stati ottenuti a partire dall'analisi dei test ensemble. In particolare in Figura 13 gli andamenti sono stati ricavati dalle medie dei parametri guadagno percentuale per operazione e numero di operazioni effettuate di tutte le combinazioni paese – anno – strategia per quanto riguarda i test ensemble.

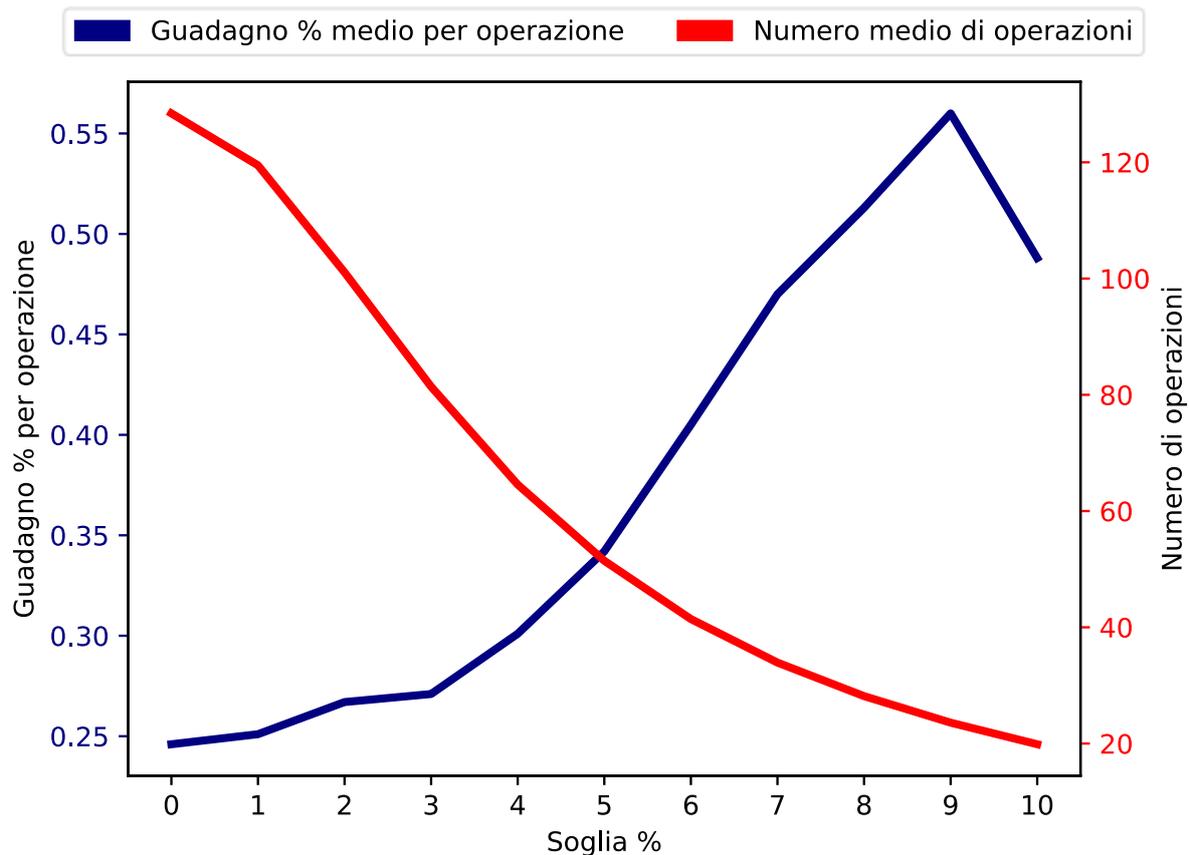


Figura 13 - Andamento del rendimento e del numero di operazioni al variare della soglia

Il numero di operazioni suggerite diminuisce all'aumentare della soglia in modo lineare, ed il guadagno medio per operazione ha effettivamente un andamento crescente all'aumentare della soglia. Sono quindi rispettate le previsioni riguardanti l'utilità di inserire una soglia percentuale nella strategia di investimento per scegliere quali raccomandazioni accettare.

È conveniente studiare anche l'andamento dei risultati ottenuti dal modello predittivo unico, per capire se si comporta nello stesso modo dei test ensemble. Per poterlo verificare si procede graficamente confrontando i rendimenti percentuali medi per operazione ottenuti al variare della soglia percentuale, usata per selezionare le raccomandazioni generate dal sistema.

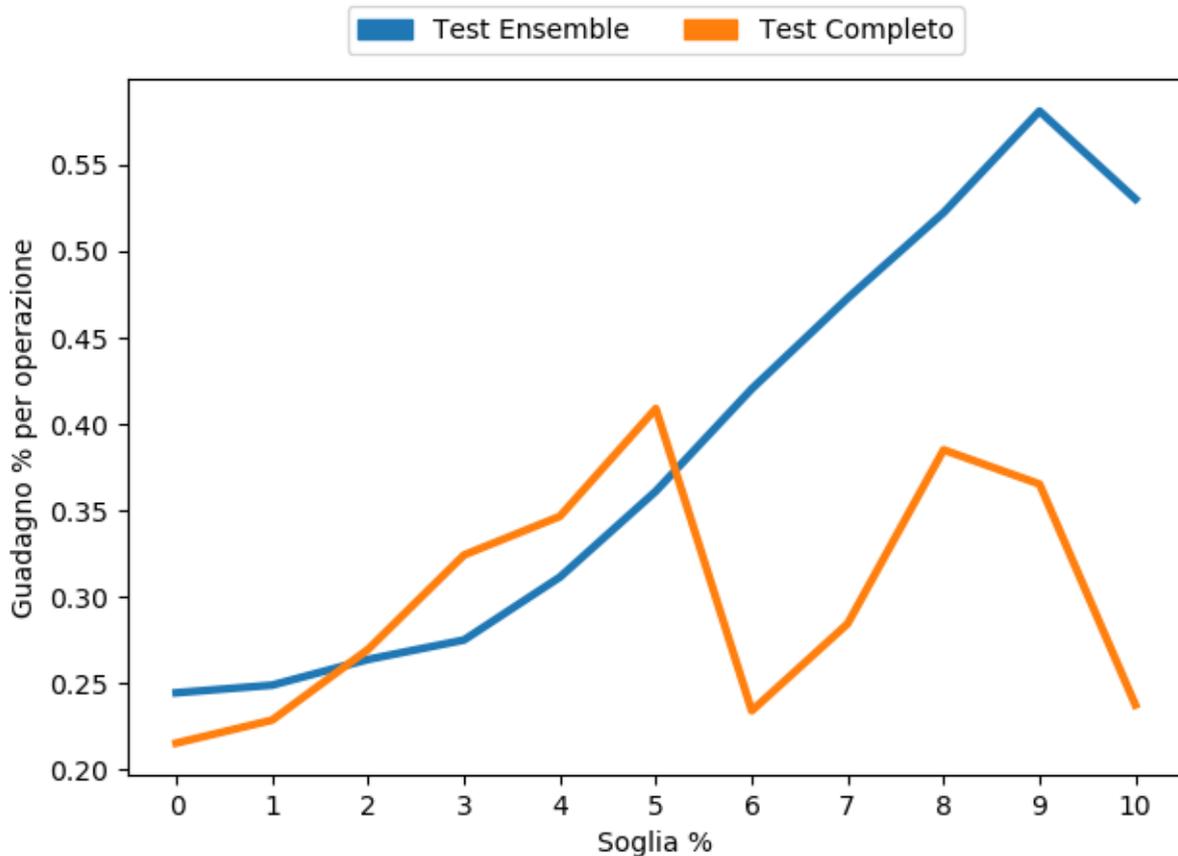


Figura 14 - Confronto dei rendimenti dei test ensemble e del test completo

Anche nel grafico in Figura 14 sono state calcolate le medie del guadagno percentuale per operazione per tutte le combinazioni di paese – anno – strategia e per tutti i test ensemble e test unico.

Si ricava che i due metodi hanno comportamenti diversi: come detto i modelli ensemble migliorano il guadagno percentuale medio per operazione al crescere della soglia percentuale, mentre il modello unico non sembra avere un andamento definito al variare della soglia.

6.3.1. Scelta della soglia percentuale più opportuna

Siccome l'obiettivo di una strategia di trading è quello di massimizzare il guadagno totale, la scelta della soglia percentuale più opportuna deve essere fatta di conseguenza. Conviene quindi calcolare quale, alla fine dell'anno, risulta essere la strategia più conveniente.

Le regole di money management del trading suggeriscono di fissare come dimensione di ogni investimento una percentuale del capitale a disposizione compresa tra il 2% e il 5%.

Per poter definire la scelta di quale sia la migliore soglia percentuale da utilizzare, si considerano i dati ottenuti dall'analisi della stessa nei modelli ensemble, confrontando il caso base in cui si accettano tutte le predizioni positive al caso di utilizzo di una soglia percentuale alta come filtro delle raccomandazioni. Per avere risultati il più generali possibile, il guadagno percentuale per operazione ad una determinata soglia percentuale è stato calcolato facendo la media di tutti i guadagni percentuali per operazione di ogni combinazione stato – anno – strategia. Con la stessa procedura si è determinato il numero di operazioni effettuate ad una certa soglia percentuale. Per calcolare il guadagno totale annuale si è ipotizzato di usare una somma fissa, determinata in base ad una percentuale del capitale iniziale a disposizione, per ogni operazione eseguita. Moltiplicando quindi la somma ottenuta per il guadagno percentuale per operazione e per il numero di operazioni si ottiene il guadagno totale annuale.

Soglia [%]	Capitale a disposizione [€]	Dimensione % investimento	Guadagno % per Operazione	Numero Operazioni	Guadagno Totale Annuale [€]
0	10000	5	0.246	128.42	157.96
7	10000	5	0.47	33.93	79.74
7	10000	20	0.47	33.93	318.94

Tabella 1 - Esempio di calcolo di rendimento

Se si impostasse la stessa percentuale del capitale per definire l'importo dell'investimento ad una soglia percentuale nulla o ad una alta, il guadagno totale sarebbe maggiore per l'investimento fatto con la soglia uguale allo zero per cento. Se invece si sceglie di aumentare la percentuale del capitale per l'investimento nel caso della soglia più alta allora si ottiene una rendita totale maggiore.

Seguendo questi dati si può concludere che se si vuole diversificare maggiormente il rischio conviene investire in tutte le raccomandazioni generate da un modello predittivo con soglia pari allo zero per cento. Se invece si ha confidenza con il sistema e si decide di investire percentuali maggiori del proprio capitale utilizzando una soglia percentuale alta, si possono realizzare profitti più elevati. Usando questa tecnica di confronto, in generale è quindi difficile trarre una unica regola da seguire, ma conviene affidarsi all'esperienza del trader per poter effettuare strategie di investimento efficaci.

Un altro metodo per confrontare strategie di trading con un numero di operazioni molto diverso è quello di confrontarne i rendimenti a parità di investimento totale. Così facendo, è possibile ricavare, a parità di impegno finanziario, quale risulta essere la linea di investimento più profittevole. Con questa modalità di confronto, la strategia con un rendimento percentuale per operazione maggiore sarà la migliore.

Riprendendo l'esempio precedente:

Soglia [%]	Capitale Investito [€]	Guadagno % per Operazione	Numero Operazioni	Guadagno Totale Annuale [€]
0	10000	0.246	128.42	24.6
7	10000	0.47	33.93	47

Dai risultati ricavati si ottiene la conferma che, a parità di capitale investito, è preferibile usare una strategia di investimento con un alto guadagno percentuale per operazione, in modo da generare un guadagno totale annuale maggiore.

6.4. Analisi dei migliori indicatori nei test ensemble

Si è poi analizzato come i modelli ensemble si comportano confrontandoli con il modello unico. Per farlo, si sono paragonati i risultati di rendita, contando quante volte i modelli ensemble hanno generato un guadagno percentuale medio per operazione maggiore rispetto a quello del modello unico.

Per poter ordinare le combinazioni di indicatori in base ad una metrica comune si è usata la seguente funzione:

$$\sum (p_1, p_2, p_3)$$

Dove:

- p_1 : punteggio ottenuto usando il modello SVM polinomiale di grado due;
- p_2 : punteggio ottenuto usando il modello di regressione lineare;
- p_3 : punteggio ottenuto usando il modello Gradient Boosted Trees.

Il risultato ottenuto indica quindi quante volte complessivamente un certo test ensemble ha battuto il test unico usando come parametro di confronto il guadagno percentuale medio per operazione. I risultati, divisi a seconda di quanti indicatori sono chiamati in causa, e comprensivi di ogni combinazione paese – anno – strategia, sono in ordine:

- Modello Chiusura, Volume, Indicatore_X

Indicatore	% meglio del Modello Unico
MA_50	48
MOM_5	47
AROON_5	46
CCI_5	45
MA_20	45
OBV_5	45
RSI_5	44
MA_10	44
MACD_5	41
MA_5	40
MFI_5	40
ATR_5	39
EMA_5	39

Tabella 2 - Confronto Test Chiusura, Volume, Indicatore e Test Unico

- Modello 2

Indicatore	% meglio del Modello Unico
MA_20	49
MA_10	47
MA_50	46
MA_5	44
EMA_5	43
MACD_5	42
AROON_5	40
ATR_5	40
CCI_5	36
RSI_5	36
MFI_5	35
MOM_5	34
OBV_5	33

Tabella 3 - Confronto Test 2 e Test Unico

- Modello Chiusura, Indicatore_X – Modello Chiusura, Indicatore_Y

Indicatore X	Indicatore Y	% meglio del Modello Unico
MA_20	MFI_5	55
MA_20	OBV_5	55
MA_5	AROON_5	55
MA_50	MFI_20	54
MA_50	AROON_5	54
EMA_5	OBV_5	53
MA_50	RSI_5	53
MA_5	OBV_5	53
MA_50	OBV_5	52
RSI_5	AROON_5	52
MA_5	MA_20	52
MA_20	CCI_5	52
MA_20	MACD_5	51

Tabella 4 - Confronto Test ensemble Chiusura, Indicatore_X – Chiusura, Indicatore_Y e Test Unico

- Modello Chiusura, Volume, Indicatore_X – Modello Chiusura, Volume, Indicatore_Y

Indicatore X	Indicatore Y	% meglio del Modello Unico
MA_50	AROON_5	53
MA_50	MFI_5	53
MA_50	MOM_5	52
MA_10	MOM_5	51
MA_10	MFI_5	51
MA_10	OBV_5	50
MA_20	CCI_5	49
MA_20	RSI_5	49
RSI_5	AROON_5	49
MA_20	MFI_5	48
MA_50	OBV_5	48
MA_20	MA_50	48
OBV_5	AROON_5	48

Tabella 5 - Confronto Test ensemble Chiusura, Volume, Indicatore_X – Chiusura, Volume, Indicatore_Y e Test Unico

Dai risultati si può notare che in alcuni casi i risultati dei modelli ensemble risultano avere più frequentemente migliori rispetto al modello unico. In particolare, usando le raccomandazioni generate dal modello che sfrutta solo un indicatore, quindi senza effettuare intersezioni, in nessun caso si riesce ad ottenere una percentuale maggiore del 50% nei confronti del test con tutti i dati.

Se invece si considerano i rendimenti ottenuti dalle raccomandazioni ricavate da intersezioni di modelli, si nota che spesso si riesce a fare meglio del test completo. In *Tabella 4* e *Tabella 5* sono riportati solo alcuni dei settantotto casi creati.

Si può quindi affermare che usando lo stesso modello predittivo è conveniente effettuare un'intersezione dei risultati ottenuti, al fine di aumentare il guadagno percentuale medio per operazione rispetto a quello del modello unico.

Un altro aspetto interessante è che, per quanto riguarda il modello Chiusura, Volume, Indicatore_X – Modello Chiusura, Volume, Indicatore_Y ed il modello Chiusura, Indicatore_X – Modello Chiusura, Indicatore_Y, uno dei due indicatori migliori è sempre una media del prezzo di chiusura. Anche negli altri due modelli esaminati, Chiusura, Volume, Indicatore_X e quello relativo al test 2 il migliore indicatore risulta essere una media del prezzo di chiusura. Questo significa che nel prezzo stesso dell'azione sono contenute molte informazioni riguardanti lo stato dell'azienda e rimane quindi l'elemento cardine di una strategia di investimento.

Si prosegue andando quindi a selezionare la migliore configurazione per ciascuno dei quattro test effettuati, ossia quella che rispetta la funzione obiettivo:

$$\max \sum (p_1, p_2, p_3)$$

I migliori indicatori risultano essere, per ogni modello predittivo:

- Modello Chiusura, Volume, Indicatore_X

Indicatore	% meglio del Modello Unico
MA_50	48

- Modello 2

Indicatore	% meglio del Modello Unico
MA_20	49

- Modello Chiusura, Indicatore_X – Modello Chiusura, Indicatore_Y

Indicatore X	Indicatore Y	% meglio del Modello Unico
MA_5	AROON_5	55

- Modello Chiusura, Volume, Indicatore_X – Modello Chiusura, Volume, Indicatore_Y

Indicatore X	Indicatore Y	% meglio del Modello Unico
MA_50	AROON_5	53

I rendimenti ottenuti dai diversi modelli ensemble sviluppati e dal modello unico sono descritti nei grafici, calcolati per ogni combinazione stato – anno – strategia ed usando un investimento totale uguale per tutte le diverse strategie:

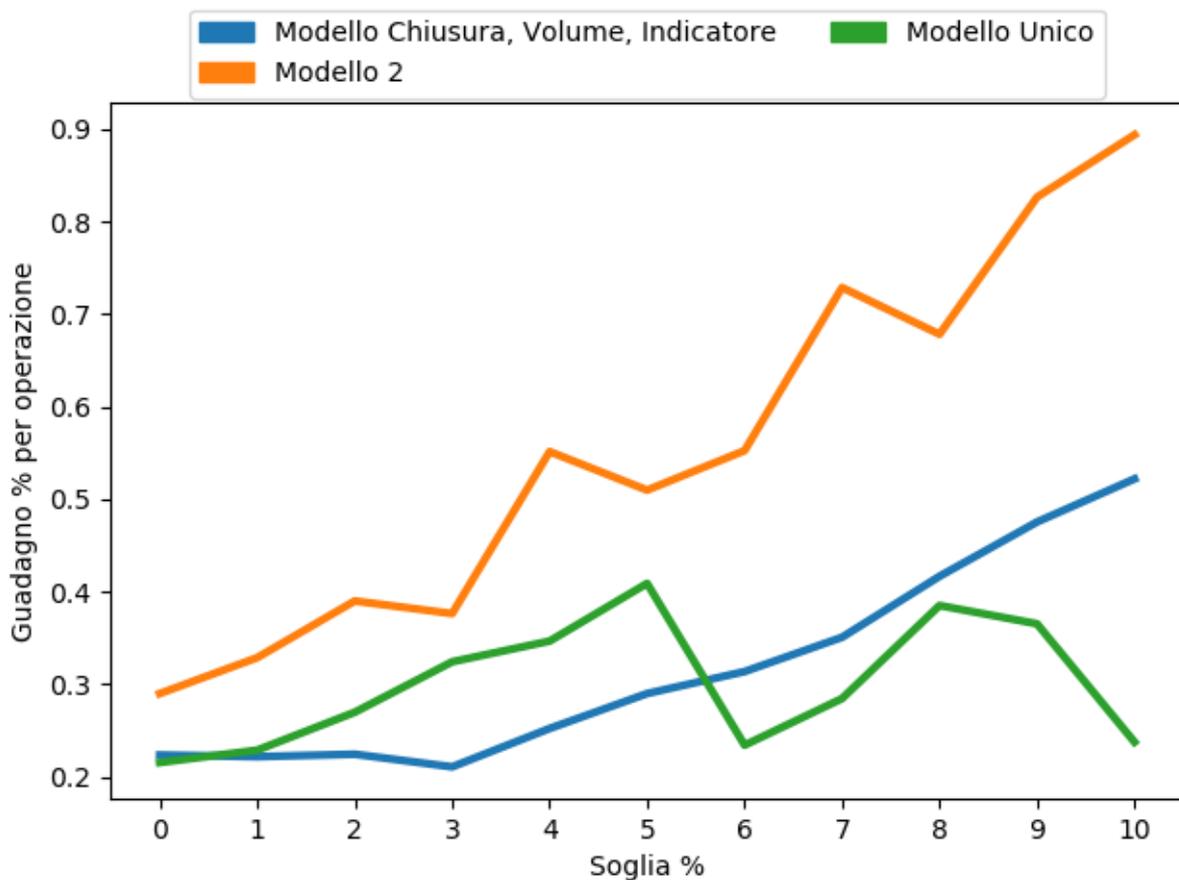


Figura 15 - Grafico del rendimento ottenuti dai test ensemble

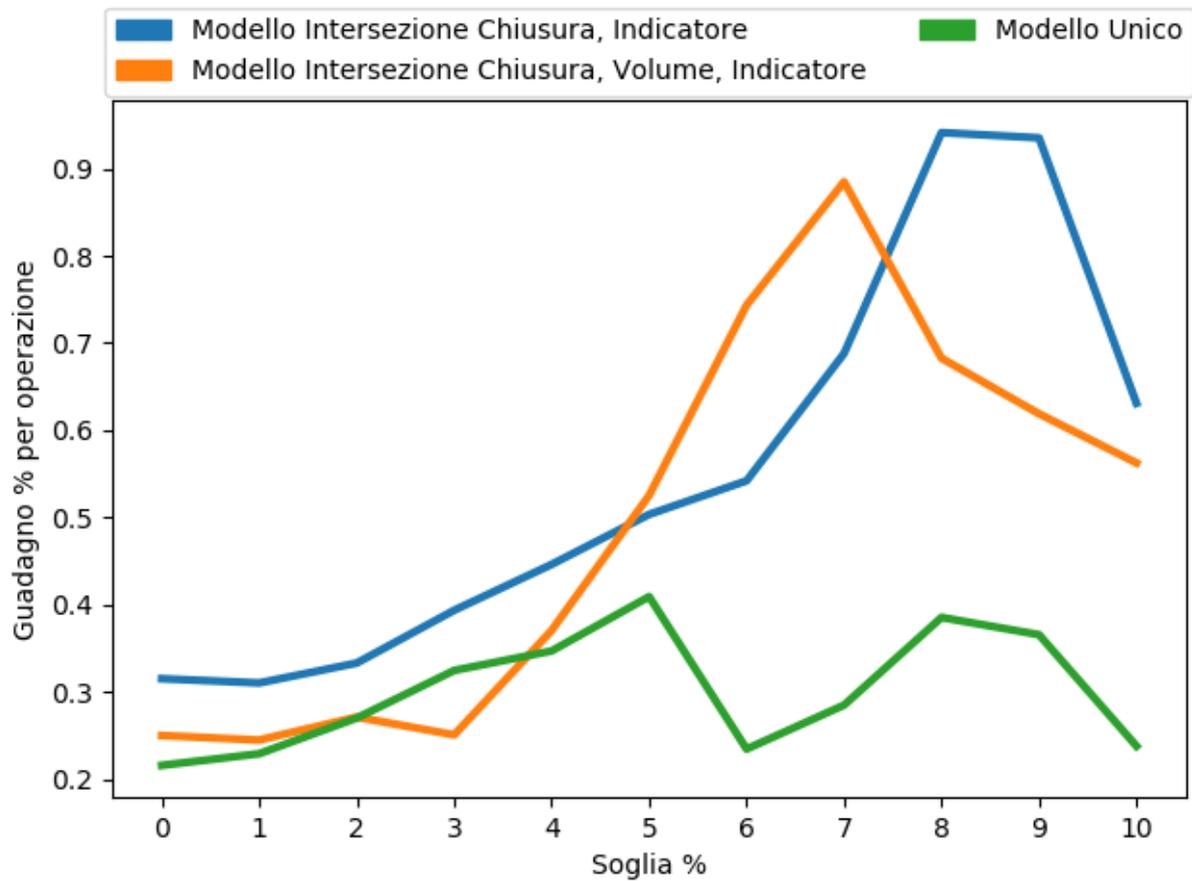


Figura 16 - Grafico del rendimento ottenuti dai test ensemble

Il modello unico risulta essere il peggiore tra i casi proposti. In particolare, i cinque casi analizzati sono sostanzialmente equivalenti quando la soglia è minore del 5%, al di sopra della quale i rendimenti diventano visibilmente diversi. I modelli generati da intersezioni di specifici indicatori riportano risultati molto migliori rispetto al modello unico.

Di seguito sono riportati i dati più completi riguardanti i test migliori, derivati dall'analisi dei risultati divisi per stato o anno. Per ogni soglia percentuale a disposizione è stata evidenziata quale sarebbe stata la strategia migliore secondo il parametro del guadagno percentuale per operazione. Per strategia si intende la scelta dei parametri: tipo di modello predittivo, test usato, stato, anno, strategia di trading.

- Stato USA

Soglia	Numero Operazioni	Totale %	Guadagno % per Operazione	Migliore strategia
0	110	124.49	1.13	GBT CVXCVY USA 2011 long
1	110	124.49	1.13	GBT CVXCVY USA 2011 long
2	58	67.92	1.17	LIN CXCX USA 2011 long
3	47	62.81	1.34	LIN CXCX USA 2011 long
4	68	99.6	1.47	GBT CXCX USA 2011 long
5	54	94.49	1.75	GBT CXCX USA 2011 long
6	22	53.14	2.42	GB CXCX USA 2011 short
7	1	7.82	7.82	LIN 2 USA 2011 short
8	1	7.82	7.82	LIN 2 USA 2011 short
9	1	7.82	7.82	LIN 2 USA 2011 short
10	1	7.82	7.82	LIN 2 USA 2011 short

- Stato ITA

Soglia	Numero Operazioni	Totale %	Guadagno % per Operazione	Migliore strategia
0	8	9.7	1.213	LIN 2 ITA 2015 short
1	5	10.86	2.172	LIN 2 ITA 2015 short
2	3	8.93	2.976	LIN 2 ITA 2015 short
3	2	7.63	3.813	LIN 2 ITA 2015 short
4	1	7.76	7.757	LIN 2 ITA 2015 short
5	1	7.76	7.757	LIN 2 ITA 2015 short
6	1	7.76	7.757	LIN 2 ITA 2015 short
7	1	7.76	7.757	LIN 2 ITA 2015 short
8	1	7.76	7.757	LIN 2 ITA 2015 short
9	1	7.76	7.757	LIN 2 ITA 2015 short
10	1	7.76	7.757	LIN 2 ITA 2015 short

- Anno 2011

Soglia	Numero Operazioni	Totale %	Guadagno % per Operazione	Migliore strategia
0	110	124.49	1.132	TREE CVXCVY USA 2011 long
1	110	124.49	1.132	TREE CVXCVY USA 2011 long
2	58	67.92	1.171	LIN CXCX USA 2011 long
3	3	5.12	1.708	LIN 2 ITA 2011 short
4	1	2.51	2.505	LIN 2 ITA 2011 short
5	1	2.51	2.505	LIN 2 ITA 2011 short
6	1	2.51	2.505	LIN 2 ITA 2011 short
7	1	7.82	7.816	LIN 2 USA 2011 short
8	1	7.82	7.816	LIN 2 USA 2011 short
9	1	7.82	7.816	LIN 2 USA 2011 short
10	1	7.82	7.816	LIN 2 USA 2011 short

- Anno 2013

Soglia	Numero Operazioni	Totale %	Guadagno % per Operazione	Migliore strategia
0	38	28.4	0.747	LIN CXCX ITA 2013 long
1	6	6.45	1.075	LIN 2 ITA 2013 long
2	5	6.15	1.231	LIN 2 ITA 2013 long
3	11	12.08	1.098	LIN CXCX ITA 2013 long
4	10	13.08	1.308	LIN CXCX ITA 2013 long
5	7	12.02	1.717	LIN CXCX ITA 2013 long
6	1	3.83	3.834	SVMpoly CVXCVY ITA 2013 short
7	1	3.83	3.834	SVMpoly CVXCVY ITA 2013 short
8	1	3.83	3.834	SVMpoly CXCX ITA 2013 short
9	1	3.83	3.834	SVMpoly CXCX ITA 2013 short
10	1	4.45	4.448	TREE 2 ITA 2013 long

- Anno 2015

Soglia	Numero Operazioni	Totale %	Guadagno % per Operazione	Migliore strategia
0	8	9.7	1.213	LIN 2 ITA 2015 short
1	5	10.86	2.172	LIN 2 ITA 2015 short
2	3	8.93	2.976	LIN 2 ITA 2015 short
3	2	7.63	3.813	LIN 2 ITA 2015 short
4	1	7.76	7.757	LIN 2 ITA 2015 short
5	1	7.76	7.757	LIN 2 ITA 2015 short
6	1	7.76	7.757	LIN 2 ITA 2015 short
7	1	7.76	7.757	LIN 2 ITA 2015 short
8	1	7.76	7.757	LIN 2 ITA 2015 short
9	1	7.76	7.757	LIN 2 ITA 2015 short
10	1	7.76	7.757	LIN 2 ITA 2015 short

Dai risultati si nota che i migliori test risultano essere quelli che effettuano intersezioni tra le raccomandazioni generate da modelli semplici. In particolare è molto presente il test 2, che per selezionare le azioni su cui investire utilizza due intersezioni. Nei risultati di questo test compaiono poche operazioni, cosa dovuta al fatto di aver effettuato un'ulteriore intersezione, che però generano i migliori guadagni percentuali per operazione.

6.5. Significatività statistica

Per validare i risultati si effettua un test statistico di significatività, che assume vera l'ipotesi zero o anche chiamata ipotesi nulla. Questa consiste nel considerare insiemi di dati come equivalenti rispetto ad un certo parametro, eventuali differenze sono quindi soltanto conseguenza del caso. Il parametro considerato in questo caso, per decidere se accettare o rifiutare l'ipotesi zero, è la posizione media finale dei test presi in analisi calcolata rispetto al guadagno percentuale giornaliero. Il risultato del test di significatività statistica non ha valenza assoluta, ma è una misura di probabilità di accettare o rifiutare l'ipotesi zero. Il livello di significatività indica proprio la probabilità di sbagliare a considerare vera o falsa l'ipotesi di partenza.

Per il test di significatività statistica è stato usato il seguente algoritmo:

- Calcolare il guadagno o la perdita per ogni giorno di trading, per ogni tipo di modello e per ogni combinazione stato – anno – strategia;
- Per ogni giorno calcolare la classifica delle strategie in base al rendimento percentuale giornaliero ottenuto;
- Calcolare la differenza finale tra le medie dei posizionamenti giornalieri e confrontarla con il valore critico VC , il quale stabilisce se questa differenza è statisticamente significativa.

$$VC = q_{\alpha} * \sqrt{\frac{K(K + 1)}{6 * N}}$$

Dove:

- N indica il numero di giorni di trading, circa 250 all'anno;
- K indica il numero di strategie confrontate, sedici in questo caso;
- α indica il livello di significatività adottato, e q_{α} è il suo relativo valore tabellare.

Il valore di α è impostato al 5%. Per selezionare il relativo valore q_{α} , seguendo [34], è stato scelto quello corrispondente al test di Bonferroni – Dunn.

I test considerati sono i seguenti:

- Modello unico;
- Modelli ensemble;
- Test Random, in cui è stata scelta l'azione che nel giorno precedente aveva avuto la migliore performance.

Di seguito sono riportati i risultati delle classifiche ottenuti secondo l'algoritmo esposto in precedenza, calcolato usando tutte le soglie percentuali disponibili.

MODELLO	ITA 2015 LONG	ITA 2015 SHORT	ITA 2013 LONG	ITA 2013 SHORT	ITA 2011 LONG	ITA 2011 SHORT
T3 SVM	6.92	9.58	7.37	9.44	6.90	9.42
CVX SVM	6.44	9.65	7.49	9.69	7.10	9.64
T2 SVM	<u>4.71</u>	3.02	4.54	2.63	4.09	2.84
CXCY SVM	5.12	3.79	5.26	3.50	4.70	3.71
CVXCVY SVM	5.09	3.61	5.30	3.20	4.42	3.40
T3 LIN	7.22	10.06	6.96	9.82	7.10	9.49
CVX LIN	6.78	9.99	6.99	9.86	7.13	9.57
T2 LIN	5.05	<u>2.81</u>	4.58	<u>2.40</u>	<u>4.01</u>	<u>2.57</u>
CXCY LIN	5.7	3.48	4.82	3.18	4.33	3.40
CVXCVY LIN	5.09	3.08	<u>4.51</u>	2.59	4.62	3.37
T3 GBT	6.42	9.77	6.61	9.75	6.46	9.38
CVX BGT	6.42	9.88	6.79	9.51	7.10	9.16
T2 GBT	5.14	3.03	4.68	2.88	4.31	3.17
CXCY GBT	5.37	4.42	5.74	4.62	5.22	4.51
CVXCVY GBT	5.63	3.85	5.31	4.16	4.68	3.97
Random	7.61	7.11	7.83	6.93	7.85	6.70

MODELLO	USA 2015 LONG	USA 2015 SHORT	USA 2013 LONG	USA 2013 SHORT	USA 2011 LONG	USA 2011 SHORT	MEDIA TOTALE
T3 SVM	5.57	9.03	5.44	9.34	6.32	9.26	7.88
CVX SVM	5.61	8.96	5.63	9.12	6.23	9.48	7.92
T2 SVM	4.45	3.96	<u>4.36</u>	4.06	4.65	3.17	3.87
CXCY SVM	4.68	4.24	4.61	4.49	4.96	3.92	4.42
CVXCVY SVM	4.51	4.39	4.70	4.00	4.55	3.61	4.23
T3 LIN	6.78	9.47	7.32	9.69	6.51	9.85	8.36
CVX LIN	6.57	9.39	6.97	9.44	6.05	9.58	8.19
T2 LIN	<u>4.21</u>	<u>3.28</u>	4.88	<u>2.99</u>	<u>4.01</u>	2.76	<u>3.63</u>
CXCY LIN	4.77	3.87	5.19	3.67	4.65	3.36	4.20
CVXCVY LIN	4.78	3.75	5.23	3.42	4.45	3.07	4.00
T3 GBT	6.77	9.14	6.30	9.89	6.25	9.30	8.00
CVX BGT	6.62	9.49	6.17	9.57	6.10	9.58	8.03
T2 GBT	4.43	3.49	4.82	3.26	4.02	<u>2.88</u>	3.84
CXCY GBT	4.78	3.94	5.12	3.86	4.60	3.48	4.64
CVXCVY GBT	4.88	3.82	5.6	3.94	4.30	3.41	4.46
Random	7.66	6.13	8.51	6.86	7.06	5.85	7.18

Nella maggior parte dei casi la differenza delle posizioni finali medie tra i risultati ottenuti dai modelli ensemble effettuati rispetto ai rendimenti del modello unico, a parità di modello predittivo, è maggiore del valore critico VC , per cui l'ipotesi che la differenza di popolazione degli insiemi sia dovuta al caso è rifiutata.

Se i risultati riportati nella colonna della posizione media di tutte le combinazioni stato – anno – strategia vengono ordinati dal migliore al peggiore si ricavano i seguenti dati:

POSIZIONE	POSIZIONE MEDIA	TEST
1	3.63	TEST 2 LIN
2	3.84	TEST 2 GBT
3	3.87	TEST 2 SVM
4	4	TEST CVXCVY LIN
5	4.2	TEST CXCXY LIN
6	4.23	TEST CVXCVY SVM
7	4.42	TEST CXCXY SVM
8	4.46	TEST CVXCVY GBT
9	4.64	TEST CXCXY GBT
10	7.18	RANDOM
11	7.88	TEST COMPLETO SVM
12	7.92	TEST CVX SVM
13	8	TEST COMPLETO TREE
14	8.03	TEST CVX GBT
15	8.19	TEST CVX LIN
16	8.36	TEST COMPLETO LIN

Tabella 3 - Classifica dei test considerati

A questo punto appaiono evidenti le osservazioni finali alle premesse di questo lavoro di tesi.

L'obiettivo principale era di capire quale tra diverse strategie di investimento fosse la migliore. Sono state confrontate quattro tipi di strategie:

- Test Random: non ha un modello predittivo, ma sceglie l'azione che nel giorno precedente aveva avuto il rendimento migliore;
- Modello Unico e modello Chiusura, Volume, Indicatore: hanno modelli predittivi diversi, ma in comune hanno il modo in cui viene effettuata la scelta delle raccomandazioni, perché entrambi non effettuano intersezioni di modelli;

- Modello Chiusura, Volume, Indicatore_X – Chiusura, Volume, Indicatore_Y e modello Chiusura, Indicatore_X – Chiusura, Indicatore_Y: sfruttano dati diversi per generare le predizioni ma entrambi effettuano una intersezione per generare le raccomandazioni;
- Modello 2: per generare le raccomandazioni effettua due intersezioni.

Dalla Tabella 3 appare evidente che più intersezioni sono effettuate tra modelli predittivi semplici per generare le raccomandazioni giornaliere e migliore sarà il rendimento medio per operazione. Infatti il Test 2 si classifica nelle prime tre posizioni con tutti i modelli predittivi usati, dopo si alternano i risultati di test che usano una intersezione tra modelli e per ultimi si posizionano i test che non effettuano intersezioni. In questo ultimo caso i risultati non sembrano essere influenzati dalla complessità e completezza dei dati usati per il training.

7. Conclusioni

In questa tesi sono stati confrontati diversi modelli predittivi al fine di capire quale fosse il migliore, introducendo l'utilizzo della regressione e del *windowing*.

Per poter predire il valore dei titoli azionari contenuti negli indici S&P 500 e FTSE MIB nel giorno successivo ai dati di training, innanzitutto era necessario scaricare i dati relativi all'andamento del prezzo delle singole azioni. Una volta ottenuti questi valori sono stati calcolati molti indicatori statistici, tipici del mercato azionario, che vengono sfruttati dai trader come supporto per eseguire operazioni di acquisto e vendita di titoli finanziari.

I casi analizzati comprendevano modelli ensemble frutto di intersezioni tra modelli specifici per una determinata caratteristica dell'andamento azionario ed un modello complesso, comprensivo di tutti gli indicatori calcolati, che non sfruttava specificità del comportamento dell'azione ma piuttosto considerava una situazione più globale del mercato.

Per effettuare la regressione sono state due i parametri chiave di questo lavoro: la lunghezza della finestra temporale e l'utilizzo del *windowing* con un passo di due giorni. Come finestra è stata scelta una durata di dieci giorni, ossia due settimane lavorative del mercato azionario. Questa scelta è dovuta al fatto che, siccome il valore predetto è quello relativo al giorno successivo all'ultimo presente nella finestra, non avrebbe avuto significato usare sequenze temporali molto più ampie. L'utilizzo del *windowing* con un passo di due giorni permette invece di evidenziare relazioni tra andamenti azionari di due giorni consecutivi. Siccome nei dati di training sono presenti tutte le azioni appartenenti ad un certo indice, eventuali relazioni verranno messe in evidenza anche tra titoli di aziende differenti.

Nell'analisi dei risultati è stata introdotta una soglia percentuale, usata per discriminare quali azioni selezionare a seconda del valore predetto, al cui aumentare tipicamente si nota un aumento del guadagno percentuale.

Confrontando i risultati ottenuti dai diversi modelli creati, si può affermare che i modelli ensemble risultano essere qualitativamente superiori, in quanto generano un guadagno percentuale maggiore. In particolare, il guadagno percentuale giornaliero, e quindi il rendimento, sembra essere correlato al numero di intersezioni effettuate tra modelli. Risulta quindi essere una scelta migliore adottare strategie derivate da raccomandazioni di intersezioni di modelli specifici calcolati su una

selezione di indicatori statistici usati in analisi tecnica piuttosto che seguire un modello generico calcolato su un'ampia collezione di indicatori.

8. Bibliografia

- [1] W. J. Frawley, G. Piatetsky-Shapiro e C. J. Matheus, «Knowledge Discovery in Databases: An Overview,» *AI Magazine*, vol. 13, n. 3, pp. 1-30, 1992.
- [2] L. Pontieri, «ICAR - CNR, Istituto di Calcolo e Reti ad Alte Prestazioni,» 2004. [Online]. Available: <http://staff.icar.cnr.it/pontieri/>.
- [3] S. Sayad, «Support Vector Machine,» [Online]. Available: http://www.saedsayad.com/support_vector_machine.htm.
- [4] S. Saed, «Support Vector Regression,» [Online]. Available: http://www.saedsayad.com/support_vector_machine_reg.htm.
- [5] A. J. Smola e S. Berhard, «A tutorial on support vector regression,» *Statistics and Computing*, n. 14, pp. 199-222, 2004.
- [6] M. E. H. Pedersen, «Artificial Intelligence for Long-Term Investing,» 2016.
- [7] . L. Chang-Chun e L. Yi-Ting, «Genetic algorithms for portfolio selection problems» *European Journal of Operational Research* 185, pp. 393-404, 2008.
- [8] F. L. F. L. D. Q. J.M. Berutich, «Robust technical trading strategies using {GP} for algorithmic portfolio selection,» *Expert Systems with Applications*, n. 46, pp. 307-315, 2016.
- [9] W.-G. Zhang, . W. Chen e Y.-L. Wang, «The Adaptive Genetic Algorithms for Portfolio Selection,» *IJCSNS International Journal of Computer Science and Network Security*, vol. VI, n. 1, pp. 196-200, 2006.
- [10] B. Wuthrich, V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, J. Zhang e W. Lam, «Daily Stock Market Forecast from Textual Web Data,» 1998.
- [11] J. Kalyani, H. N. Bharathi e R. Jyothi, «Stock trend prediction using news».
- [12] R. Pimprikar , S. Ramachandran e . K. Senthilkumar, «Use of machine learning algorithms and Twitter sentiment analysis for stock market prediction,» *International Journal of Pure and Applied Mathematics*, vol. 115, n. 6, pp. 521-526, 2017.
- [13] J. Bollen, H. Mao e X.-J. Zeng, «Twitter mood predicts the stock market,» 2010.

- [14] G. Bonde e R. Khaled, «Extracting the best features for predicting stock prices using machine learning».
- [15] Y. Dai e Y. Zhang, «Machine Learning in Stock Price Trend Forecasting».
- [16] L. A. Teixeira e A. L. Inácio de Oliveira, «A method for automatic stock trading combining technical analysis and nearest neighbor classification,» *Expert Systems with Applications* , vol. 37, p. 6885–6890, 2010.
- [17] R. Dash e P. K. Dash , «A hybrid stock trading framework integrating technical analysis with machine learning techniques,» *The Journal of Finance and Data Science*, vol. 2, pp. 42-57, 2016.
- [18] Q. Wen, Z. Yang, Y. Song e P. Jia, «Automatic stock decision support system based on box theory and SVM algorithm,» *Expert Systems with Applications* , vol. 37, pp. 1015-1022, 2010.
- [19] L. C. Tiong, D. C. Ngo e Y. Lee, «Stock Price Prediction Model using Candlestick Pattern Feature,» *International Journal Of Interactive Digital Media*, vol. 1, n. 3, pp. 58-64, 2013.
- [20] M. Kordos e A. Cwiok, «A New Approach to Neural Network based».
- [21] A. Skabar e I. Cloete, «Neural Networks, Financial Trading and the Efficient Markets Hypothesis».
- [22] E. Baralis, L. Cagliero, T. Cerquitelli, P. Garza e F. Pulvirenti, «Discovering profitable stocks for intraday trading,» *Information Sciences*, n. 405, pp. 91-106, 2017.
- [23] «Yahoo! Finance,» [Online]. Available: <https://finance.yahoo.com>.
- [24] «Google Finance,» [Online]. Available: <https://finance.google.com>.
- [25] «Borsa Italiana,» [Online]. Available: <http://www.borsaitaliana.it>.
- [26] «The New York Stock Exchange,» [Online]. Available: <https://www.nyse.com>.
- [27] «Yahoo! Finance,» 09 11 2017. [Online]. Available: <https://forums.yahoo.net/t5/Known-issues-and-updates/Yahoo-Finance-API-disabled/m-p/392760#M14%C2%A0>.
- [28] «Rapidminer,» [Online]. Available: <https://rapidminer.com/>.
- [29] «Operator Manual - RapidMiner Documentation,» [Online]. Available: <https://docs.rapidminer.com/latest/studio/operators/>.
- [30] «Support Vector Machine - RapidMiner Documentation,» [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/support_vector_machines/support_vector_machine.html.
- [31] «Linear Regression - RapidMiner Documentation,» [Online]. Available:

https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/functions/linear_regression.html.

[32] «Gradient Boosted Trees - RapidMiner Documentation,» [Online]. Available:

https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/gradient_boosted_trees.html.

[33] «AMZN Interactive Chart | Amazon.com, Inc. Stock - Yahoo Finance,» [Online]. Available:

<https://finance.yahoo.com/quote/AMZN/chart?p=AMZN>.

[34] I. Kononenko e M. Kukar, Machine Learning and Data Mining, Woodhead Publishing, 2007.