

Machine Learning per la previsione di titoli azionari a supporto di strategie di investimento

Relatore: *Antonio Candeliere*

Co-relatore: *Silvio Bencini*

Relazione della prova finale di:

Giuseppe Sabino

matricola 852287

Anno Accademico 2021-2022

Indice

1	Introduzione	3
2	Dati esperimento	4
2.1	Dow Jones Industrial Average	4
2.2	Statistiche descrittive	5
2.3	Analisi di previsione delle serie temporali	8
2.3.1	Divisione del dataset	9
2.4	Modello media-varianza	9
3	Machine Learning	10
3.1	Arima	11
3.2	Support Vector Regressor	14
3.3	Linear Regressor	16
3.4	KNN Regressor	17
3.5	Decision Tree Regressor	18
3.6	Random Forest Regressor	20
4	Deep Learning	22
4.1	Artificial Neural Network (ANN)	23
4.2	Gated Recurrent Unit (GRU)	24
4.3	Long Short Term Memory (LSTM)	26
5	Risultati Esperimento	27
5.1	Previsione sui componenti	27
5.2	Frontiera efficiente	31
6	Conclusione	35
7	Bibliografia	36

Abstract L'intelligenza artificiale si sta sviluppando sempre di più in tutti i settori del mondo moderno. Grazie al Machine Learning molte analisi di previsioni risultano più semplici e più accurate. In questo esperimento saranno usati 6 algoritmi di Machine Learning (Arima, SVR, Regressione Lineare, KNN Regressor, Decision Tree Regressor, Random Forest Regressor) e 3 di Deep Learning (ANN, GRU, LSTM) per la previsione di titoli azionari, confrontandoli tra loro e dimostrando che le reti neurali, in questo caso ANN, sono le più idonee per compiti complessi e che richiedono molto dati. Tutto ciò per supportare una strategia di investimento, che si dimostrerà vantaggiosa, ma non all'altezza delle previsioni

1 Introduzione

La previsione delle azioni è diventata negli ultimi anni una sfida sempre più interessante e presa in considerazione soprattutto da parte degli esperti, questo perchè, prevedere l'andamento di un'azione, può portare vantaggi significativi all'investitore. L'intelligenza artificiale ha reso possibile affrontare questa sfida nel migliore dei modi, infatti tramite algoritmi di Machine Learning e Deep Learning è possibile prevedere con più accuratezza l'andamento delle azioni.

Ovviamente per quanto efficaci essi siano, è impossibile prevedere con esattezza il prezzo dello stock. Si cerca quindi tramite la validazione di ottenere dei parametri che meglio possano stimare l'andamento generale.

Questo argomento è stato preso in esame da diversi studi [1] [2], alcuni di essi utilizzano algoritmi di classificazione, prevedendo dunque se il prezzo salirà/scenderà [6]. L'obiettivo di questa tesi consiste nell'utilizzare algoritmi di regressione per prevedere il prezzo di un titolo azionario nell'arco di 3 mesi (17\03\2022 - 10\06\2022). Per fare questo saranno confrontati 6 algoritmi di Machine Learning (Arima, SVR, Regressione Lineare, KNN Regressor, Decision Tree Regressor, Random Forest Regressor) e 3 di Deep Learning (ANN, GRU, LSTM). Per paragonare i vari modelli verranno effettuate delle previsioni sul DJIA (Dow Jones Industrial Average), indice americano di importanza mondiale, andando ad utilizzare in seguito il modello che meglio performa in merito al valore di RMSE. Con quest'ultimo si effettueranno delle previsioni sui 30 titoli che compongono l'indice e verranno selezionati coloro che hanno un errore di predizione minore. Saranno poi usati quelli meno correlati tra loro per costruire un portafoglio di investimento. Per semplificare, non sono stati presi in considerazione i costi di transazione.

Nel capitolo 2 vengono introdotti i dati utilizzati, illustrando cos'è il DJIA e come si comporta nel periodo di interesse. I capitoli 3 e 4 riportano rispettivamente semplici nozioni di teoria di Machine Learning e Deep Learning, mostrando la fase di validazione dei parametri e la previsione ottenuta da ciascun modello. Nel capitolo 5 vengono riassunti i risultati finali riguardanti le predizioni delle componenti dell'indice e la composizione del portafoglio. Il capitolo 6 presenta le principali conclusioni

2 Dati esperimento

2.1 Dow Jones Industrial Average

Il Dow Jones Industrial Average ¹, è un indice di Borsa utilizzato per misurare l'andamento dei primi 30 titoli (Tab 1) del NYSE (New York Stock Exchange) e rappresenta un buon indicatore della prestazione del mercato statunitense e mondiale. In generale un indice è una funzione matematica che, nel contesto dei mercati, misura la variazione dei prezzi delle azioni e viene considerato un riferimento dell'andamento dei valori di mercato e dello stato di salute dell'economia. Si tratta di un indice pesato rispetto al prezzo, ciò significa che viene considerato esclusivamente il prezzo di ogni azione (non è dunque necessario il numero di azioni presenti).

In questa tipologia di indice una variazione significativa del prezzo di un singolo componente può influenzare pesantemente il suo valore, indipendentemente dalle dimensioni dell'azienda. Il DJIA non è legato ad alcun settore specifico, i suoi titoli infatti appartengono a diversi settori produttivi. Viene calcolato sommando il prezzo di tutti i titoli che lo compongono e dividendo tale somma per un divisore fisso; Per calcolare il suo valore si usa la seguente formula :

$$DJIA = (\sum_{i=1}^N p_i) / d \quad (1)$$

dove :

N = 30 è il numero di azioni che compongono il Dow Jones

p corrisponde al prezzo della rispettiva azione

d è il Divisore Dow, una costante del valore di circa 0,1321.

¹<https://www.borsaitaliana.it/borsa/glossario/indice-dow-jones-industrial-average.html>

Tabella 1: Componenti DJIA

Nome	Ticker	Nome	Ticker
3M	MMM	Alphabet	GOOG
Amazon	AMZN	Amgen	AMGN
Apple	AAPL	Boeing	BA
Caterpillar	CAT	Chevron	CVX
Cisco	CSCO	Coca-Cola	KO
Disney	DIS	Facebook	META
Goldman Sachs	GS	Home Depot	HD
Honeywell	HON	IBM	IBM
Intel	INTC	Johnson e Johnson	JNJ
JPMorgan	JPM	McDonalds	MCD
Microsoft	MSFT	Merck e CO	MRK
Nike	NKE	Procter e Gamble	PG
Salesforce	CRM	Tesla	TSLA
United Health	UNH	Verizon	VZ
Visa	V	Wal-Mart	WMT

2.2 Statistiche descrittive

Sono stati importati i dati pubblici disponibili su Yahoo! Finance dell'indice DJIA utilizzando come periodo di riferimento 10 anni di dati storici a partire dal 30/11/2012 fino al 10/06/2022

	High	Low	Open	Close	Volume	Adj Close
Date						
2012-11-30	13053.740234	12988.679688	13022.049805	13025.580078	171070000	13025.580078
2012-12-03	13087.320312	12959.419922	13027.730469	12965.599609	112900000	12965.599609
2012-12-04	13022.509766	12940.070312	12966.450195	12951.780273	125290000	12951.780273
2012-12-05	13089.110352	12923.440430	12948.959961	13034.490234	161230000	13034.490234
2012-12-06	13076.879883	13007.839844	13026.190430	13074.040039	116850000	13074.040039
...
2022-06-06	33235.371094	32819.500000	33032.039062	32915.781250	253010000	32915.781250
2022-06-07	33207.449219	32641.849609	32783.031250	33180.140625	270750000	33180.140625
2022-06-08	33156.500000	32824.371094	33087.070312	32910.898438	270470000	32910.898438
2022-06-09	32956.730469	32267.779297	32828.621094	32272.789062	289710000	32272.789062
2022-06-10	32053.519531	31387.839844	32053.519531	31392.789062	362300000	31392.789062

Figura 1: OHLC data

In Figura 1 sono rappresentati gli OHLC data. Nelle varie colonne sono indicati rispettivamente il prezzo più alto di giornata, il più basso, il prezzo di apertura e di chiusura (Open, High, Low, Close). E' presente inoltre il volume, ovvero quante azioni sono state scambiate durante l'arco della giornata e

l'adjusted close che corrisponde invece al prezzo di chiusura una volta considerati i dividendi. Per gli scopi di questo esperimento sarà utilizzato il prezzo di chiusura.

Avendo a disposizione questi dati si può costruire un grafico che mostra l'andamento dell'indice in questi 10 anni.

In Figura 2 è molto evidente un trend di crescita continua fino a febbraio 2020, momento in cui tutto il mercato ha avuto un calo drastico dovuto allo scoppio della pandemia da Sars-Covid-19 la quale ha portato il valore dell'indice da circa 29000 di metà febbraio a circa 17000 di metà marzo. Nel giro di un mese quindi, il suo valore è sceso di 12 mila dollari e sono stati necessari 9 mesi per ritornare al valore pre-pandemia. Un altro periodo da analizzare corrisponde agli ultimi mesi del 2021 quando l'indice toccò quota 35 mila salvo poi riscendere sotto ai 30 mila all'inizio del 2022.

Quest'ultima analisi dimostra quanto il mercato sia difficile da prevedere, soprattutto se influenzato da fattori esterni e imprevedibili.

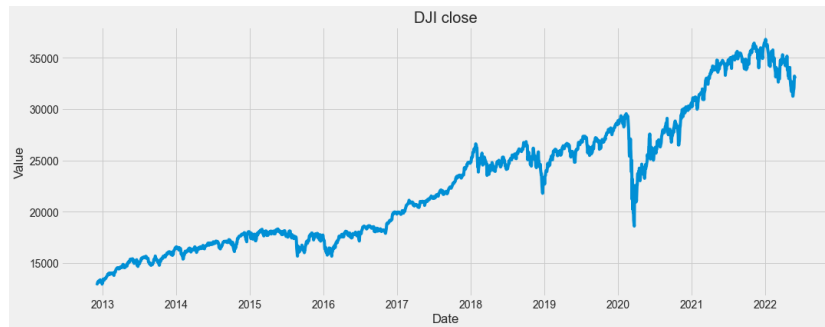


Figura 2: Andamento DJIA

Per le prossime statistiche saranno utilizzati i ritorni semplici netti, ovvero, la percentuale di guadagno, in questo caso giornaliera, dell'indice :

$$(p_t - p_{t-1}) / (p_{t-1}) \quad (2)$$

Prendendo in considerazione la Figura 3 vengono analizzati i tre grafici che la compongono

- Dal primo grafico si nota che la maggior parte dei rendimenti giornalieri si concentra vicino allo 0 e i valori sono principalmente compresi tra $\pm 4\%$.
- Il secondo grafico è un qqplot ovvero la rappresentazione grafica dei quantili di una distribuzione. Visualizza un grafico dei quantili dei dati rispetto ai valori quantili teorici di una distribuzione normale. Se la distribuzione in oggetto è normale, il grafico dei dati appare lineare. Per poter dire che appartiene a una distribuzione normale è sufficiente quindi che i punti siano vicini alla linea presente nel grafico.

I rendimenti semplici seguono dunque una distribuzione normale, ma presenta dei valori estremi che si distaccano da essa, gli outliers, ovvero dei valori anomali, chiaramente distanti dalle altre osservazioni.

- Nel terzo grafico, il boxplot, quanto detto prima è ancora più evidente. Il boxplot permette di rappresentare sullo stesso grafico cinque misure di posizione:

1. valore minimo
2. primo quartile
3. mediana
4. terzo quartile
5. valore massimo.

Il box contiene il 50% centrale delle osservazioni effettuate, quelle comprese tra il primo ed il terzo quartile (scarto interquartile). La linea verde all'interno della scatola invece rappresenta la mediana, il valore che si trova esattamente nel mezzo della distribuzione. I due segmenti che partono dalla scatola e si prolungano verso l'alto e verso il basso sono detti "baffi". Il valore massimo è rappresentato dal baffo superiore, di conseguenza, il valore minimo è rappresentato dal baffo inferiore. Infine i punti isolati indicano gli outliers. La mediana dei valori è 0 e si notano numerosi outliers. I più significativi superano i "-0.10" e "0.10" e vengono associati al periodo del covid

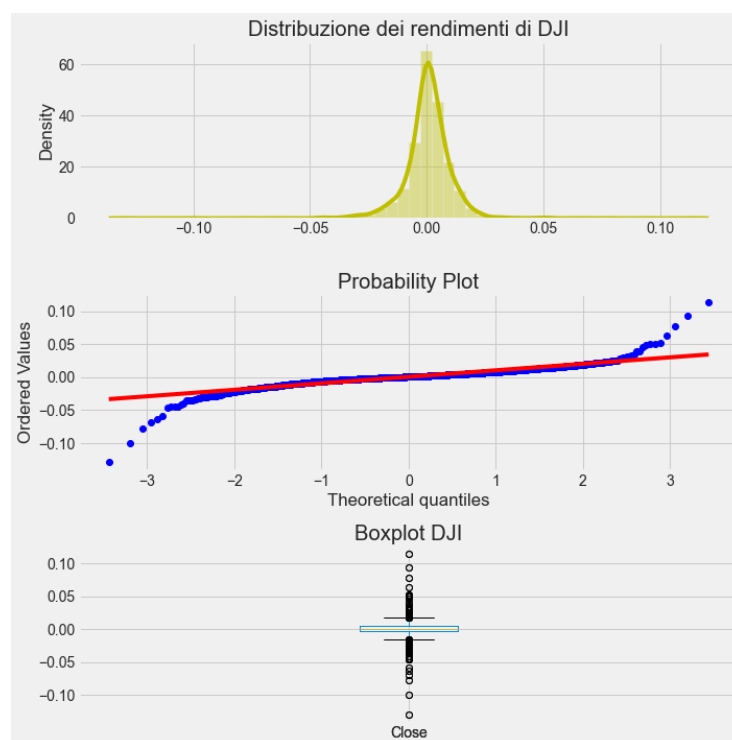


Figura 3: Grafici statistici

Le statistiche descrittive di più valore, che meglio spiegano il comportamento dell' indice sono :

- **Media**: il valore medio dei rendimenti nel periodo osservato. Rappresenta dunque il valore del rendimento che ci si può aspettare da un investimento.
- **Varianza**: fornisce una misura della variabilità dei valori assunti, nello specifico, la misura di quanto essi si discostino quadraticamente rispettivamente dalla media
- **Deviazione standard**: Evidenzia quanto i valori nel set di dati differiscono dalla media, un valore basso della deviazione standard indica che i punti sono vicini ad essa
- **Curtosi**: è un allontanamento dalla normalità distributiva. Il valore dell'indice corrispondente alla distribuzione normale (gaussiana) è 0. Un valore minore di 0 indica una distribuzione platicurtica (appiattimento), mentre un valore maggiore di 0 indica una distribuzione leptocurtica (allungamento).
- **Asimmetria**: una distribuzione è asimmetrica quando la media non coincide con il massimo centrale della curva; la distribuzione normale è simmetrica, ma solitamente i rendimenti sono asimmetrici verso sinistra.

Tabella 2: Statistiche Descrittive

Statistica	Valore
Media	0.00041
Varianza	0.00012
Deviazione Standard	0.01087
Curtosi	23.61782
Asimmetria	-0.64507

Il rendimento giornaliero medio è dello 0.043% e ha una deviazione standard del 1,083%. Nel periodo osservato la media dei rendimenti è positiva, mentre il valore di asimmetria risulta negativo. Ciò significa che mediamente i rendimenti negativi sono numerosi e che in caso di picchi negativi eccessivi è sempre stato in grado di recuperare. Si nota che il valore di curtosi è molto distante da quello di una distribuzione normale, ciò comporta una presenza di outliers significativa.

2.3 Analisi di previsione delle serie temporali

Per serie temporale si intende una lista di date a cui viene associato un valore, è dunque una sequenza presa in punti successivi equidistanti nel tempo [3]. In questo caso corrisponde al prezzo di chiusura giornaliero dell'indice DJIA nell'arco dei 10 anni selezionati. La sua analisi consiste nel assimilare gli eventi passati, nel tentativo di prevedere cosa accadrà in futuro.

Per fare ciò è utile comprendere il trend della serie, ovvero la tendenza di fondo che caratterizza il suo andamento in un periodo di tempo determinato.

Parte di questo lavoro viene svolto dall'intelligenza artificiale e più nello specifico dal Machine Learning e dal Deep Learning.

2.3.1 Divisione del dataset

Un modello di previsione viene solitamente diviso in tre set diversi [4]. Un set di allenamento che deve essere abbastanza grande, in modo da permettere al modello di imparare la relazione presente tra la variabile X e la variabile Y. Un set di validazione in cui vengono validati i parametri in maniera da selezionare i più idonei. Un set di test ovvero una fase finale in cui viene testato il modello. In questo esperimento il dataset iniziale è stato suddiviso nel seguente modo :

- **Train set** : 30\11\2012 - 14\07\2021
- **Validation set** : 15\07\2021 - 16\03\2022
- **Test set** : 17\03\2022 - 10\06\2022

ciò significa che il test set è composto dagli ultimi 3 mesi storici dell'intero dataset.

In figura 4 si evidenziano le tre porzioni suddivise

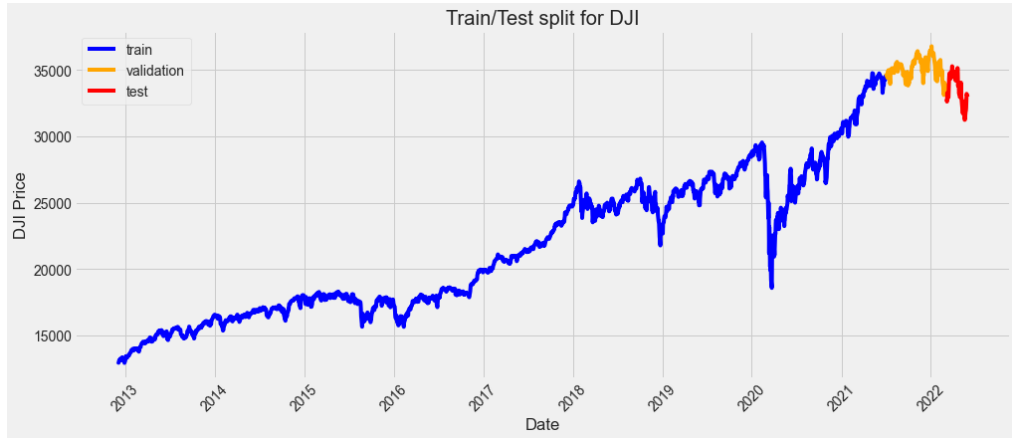


Figura 4: Train/Test set

2.4 Modello media-varianza

Markowitz ha presentato nel 1952 le basi per la selezione del portafoglio , ovvero, trovare una combinazione di strumenti finanziari che produca il miglior ritorno possibile al minor rischio. L'idea fondamentale di questo modello è quella di minimizzare il rischio, in quanto visto come un aspetto negativo. Per fare ciò viene sostenuta l'idea della diversificazione degli asset, in quanto asset scorellati permettono di avere una maggiore diversificazione, e un maggior equilibrio qualora alcuni subissero una variazione negativa. Quindi dal punto di vista pratico viene fissato un determinato valore di ritorno r^* e l'obiettivo è di trovare la soluzione migliore per arrivare a r^* con il minimo rischio.

Ritorno atteso al tempo t del portafoglio w :

$$E(R_t^w) = \sum_{i=1}^N w_i E(R_{i,t}) \quad (3)$$

Varianza del portafoglio w :

$$Var(R_t^w) = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij} \quad (4)$$

Questo si riconduce a un problema di ottimizzazione che si risolve mediante tecnica di programmazione quadratica.

$$\min \quad w' C w \quad (5)$$

$$\text{subject_to:} \quad w' u = r^* \quad (6)$$

$$\text{and} \quad \sum_{i=1}^N w_i = 1 \quad (7)$$

Quindi riassumendo il problema della selezione del portafoglio di Markowitz: trovo i pesi $w=(w_1,..w_N)$ t.c per un dato tasso di rendimento atteso r^* , il ritorno atteso del portafoglio determinato da w è r^* mentre la sua varianza è minima.

Questo problema viene poi rappresentato graficamente nel piano media-varianza: i punti ammissibili vengono chiamati frontiera efficiente e sono i punti favorevoli all'investimento. Un punto interessante è il MVP (Minimum Variance Portfolio) e rappresenta il punto in cui il rischio è minimo ma il ritorno è comunque favorevole.²

3 Machine Learning

Con Machine Learning si intende un sottoinsieme dell'Intelligenza Artificiale (la capacità delle macchine di svolgere compiti dell'intelligenza umana quali pianificazione, riconoscimento di pattern e previsione) e consiste nell' apprendimento di una regola dai dati in maniera automatizzata in modo da migliorare le performance delle predizioni.³

Esistono molti tipi di ML, ma i più rinomati sono sicuramente ML supervisionato e non supervisionato.

Per quanto riguarda il primo si ha una variabile target, quindi un valore specifico che si vorrà conoscere tramite previsioni. In questo modello si agisce direttamente sull'algoritmo insegnandogli i risultati da generare(fase di training) e simulando la previsione vera e propria in una fase di testing. Si divide in due classi :

Classificazione La variabile target ha un valore prefissato, esempio Yes or No.

Es:L'azione X al tempo t avrà un valore maggiore rispetto a t_{-1} ? Le uniche possibilità di risposta saranno Y/N

Regressione La variabile target ha un valore numerico

Es:Quanto varrà l'azione X tra 10 giorni?

²Materiale del corso di Business Intelligence per i Servizi Finanziari, Professor Antonio Candelieri, Anno 2021/2022

³<https://www.oracle.com/it/data-science/machine-learning/what-is-machine-learning/>

Un approccio supervisionato di tipo regressivo è quanto utilizzato nell'esperimento.

Secondo quanto riportato dall'Università della California Berkeley ⁴ "il sistema di apprendimento di un algoritmo di ML si suddivide in tre parti : processo decisionale, funzione di errore, processo di ottimizzazione"

- Processo decisionale : Una serie di step in cui l'algoritmo produrrà una stima su un pattern nei dati
- Funzione di errore: Utilizzata per valutare l'ipotesi confrontandola con esempi noti in modo tale da avere un confronto per valutare l'accuratezza del modello
- Processo di ottimizzazione: se il modello non è ottimale l'algoritmo aggiorna i pesi per ridurre la differenza tra l'esempio noto e la stima del modello per ridurre l'errore. Questa fase viene ripetuta finché non si raggiunge un certo valore di accuratezza.

3.1 Arima

E' un modello che utilizza serie temporali e analisi statistiche per interpretare i dati ed effettuarne una predizione.

Per capire al meglio Arima è necessario introdurre 3 concetti :

Integrated Arima può essere applicato esclusivamente su serie stazionarie, è quindi compito della differenziazione rimuovere la tendenza di una serie per renderla tale

AutoRegressive Un modello AR usa una combinazione lineare di valori passati per effettuare il forecasting

$$Y_t = c + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \epsilon \quad (8)$$

Con: p: ordine autoregressivo; c: costante; ϵ : rumore; θ : peso

Moving Average la previsione o il risultato del modello dipende linearmente dai valori passati. Inoltre, gli errori nella previsione sono funzioni lineari di errori passati.

$$Y_t = Z_t + \lambda_1 Z_{t-1} + \lambda_2 Z_{t-2} + \dots + \lambda_q Z_{t-q} \quad (9)$$

Con: Z: rumore; λ : peso; q: ordine termine moving average

Le equazioni (3) e (4) fanno riferimento a ⁵

Il primo passaggio per costruire un modello arima è fare in modo che la serie sia stazionaria, ovvero che non cambi le sue caratteristiche (Media, Varianza) nel tempo. Il valore di 'd' è il minimo numero di differenziazioni necessarie per rendere una serie stazionaria. 'p' è l'ordine del processo autoregressivo AR e si riferisce al numero di lag da utilizzare come predittori. 'q' è l'ordine del processo

⁴<https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/>

⁵<https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/>

Moving Average MA e si riferisce al numero di errori di previsione ritardati che dovrebbero essere presenti nel modello ARIMA.

Viene utilizzato il comando "ndiffs" che esegue un test di stazionarietà per diversi livelli di d per stimare il numero di differenziazioni necessarie per rendere stazionaria una determinata serie temporale, in questo caso $d=1$

Il prossimo step è quello di identificare se il modello necessita di termini AR, utilizzando il grafico delle autocorrelazioni parziali (PACF) che corrisponde alla funzione di autocorrelazione parziale utilizzata per trovare correlazioni tra la serie e il suo ritardo, dopo aver escluso i contributi dai ritardi intermedi.

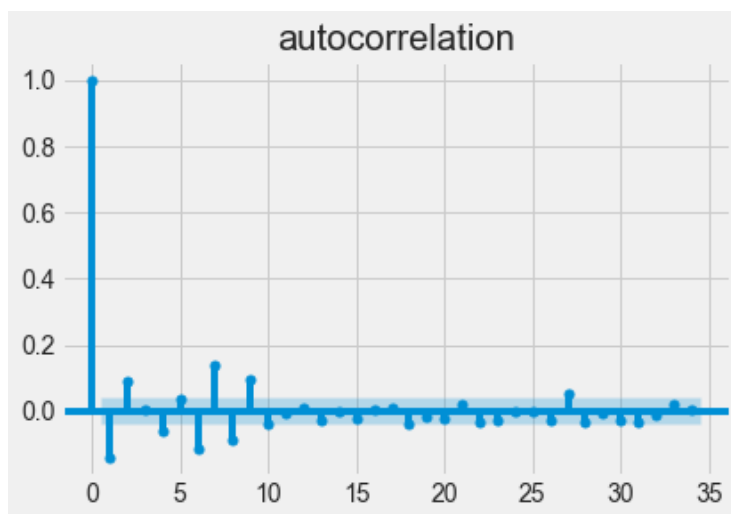


Figura 5: PACF

Viene selezionato il lag 2 in quanto supera leggermente il limite di significatività indicato dalla regione blu.

Quanto fatto ora verrà ripetuto per trovare l'ordine del termine MA, utilizzando però il grafico delle autocorrelazioni (ACF) che descrive quanto il valore attuale della serie sia correlato ai suoi valori passati, considerando tendenza, stagionalità e residui

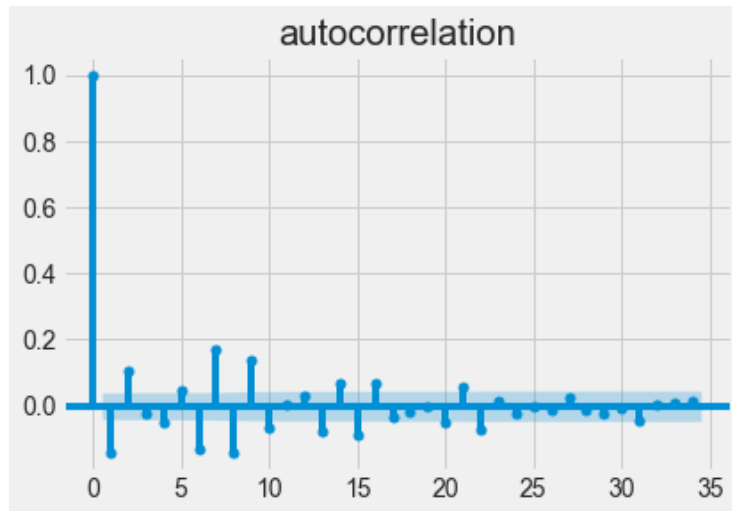


Figura 6: ACF

Per la stessa motivazione precedente, viene utilizzato $q=2$.

Quindi con un modello Arima di ordine $(2,1,2)$ si ottiene la seguente previsione:

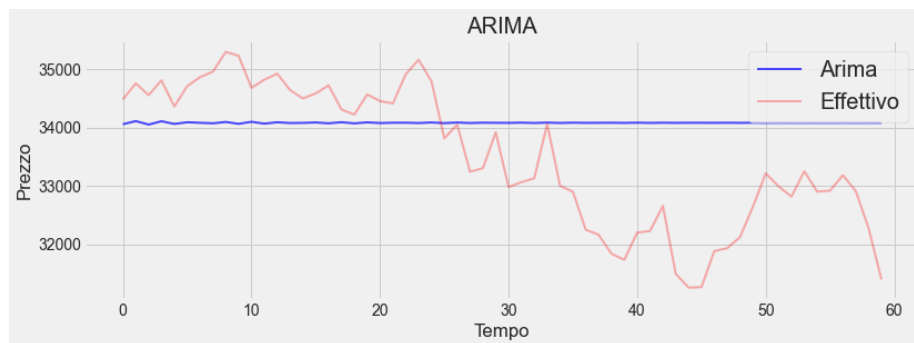


Figura 7: ARIMA

Tabella 3: Forecast Accuracy ARIMA

MAPE	ME	MAE	MPE	RMSE
0.033	569.118	1092.908	0.018	1310.699

E' molto evidente l'appiattimento della previsione, questo perchè per evitare problemi di previsioni maggiori, Arima traccia una linea corrispondente alla media dei valori. Viene riportato in Figura 8 la previsione del modello Arima con un livello di confidenza del 95%, ovvero si può dire quasi certamente che il titolo rimanga all'interno dell'imbuto.

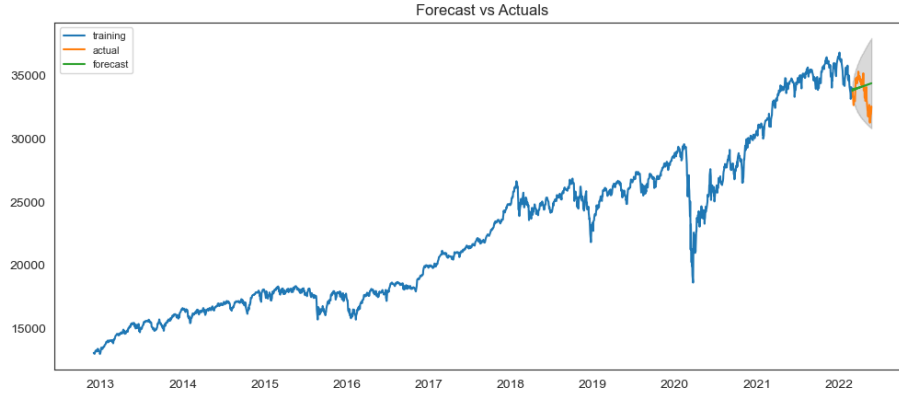


Figura 8: Intervallo di confidenza

In Tabella 3 sono presenti 5 misure di errore :

$$MeanAbsolutPercentageError : \frac{1}{n} \left(\sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \right) \quad (10)$$

$$MeanError : \frac{1}{n} \sum_{t=1}^n (A_t - F_t) \quad (11)$$

$$MeanAbsolutError : \frac{1}{n} \sum_{t=1}^n |(A_t - F_t)| \quad (12)$$

$$MeanPercentageError : \frac{1}{n} \left(\sum_{t=1}^n \frac{A_t - F_t}{A_t} \right) \quad (13)$$

$$RootMeanSquareError : \frac{1}{n} \sum_{t=1}^n \sqrt{(A_t - F_t)^2} \quad (14)$$

Dove A_t rappresenta il valore attuale al tempo t e F_t rappresenta il valore predetto al tempo t

3.2 Support Vector Regressor

Support Vector Regression è, come si può intuire dal nome, un algoritmo di regressione usato per predire variabili di ordine continuo ed è un'estensione del più generico algoritmo SVM, Support Vector Machine, illustrato da Vapnik e Cortes [5]

L'idea del SVR è di adattare l'errore all'interno di una soglia, e quindi di approssimare il miglior valore entro un margine ϵ .

Concetti fondamentali⁶:

- Iperpiano: E' una linea di separazione tra due classi di dati di una dimensione superiore rispetto alla dimensione effettiva e aiuta con la predizione della variabile target.

⁶<https://www.educba.com/support-vector-regression/>

- Kernel: La regressione viene eseguita a una dimensione superiore. é necessario dunque mappare i punti nella sua dimensione superiore. Questo è il compito del kernel.
- Boundary Lines: Rappresentano due linee a distanza ϵ dall'iperpiano
- Support vector: I vettori di supporto sono osservazioni che si trovano sul margine che circonda l'iperpiano. Poiché il margine definisce la distanza minima che le osservazioni dovrebbero avere dal piano, le osservazioni che si trovano su di esso influiscono sull'orientamento e sulla posizione dell'iperpiano.

Gli SVR possono utilizzare una regressione lineare o non lineare utilizzando un kernel che mappa gli input ad una dimensione superiore.

In questo esperimento oltre alla funzione lineare sono state utilizzate tre tipologie di kernel differenti [6]:

$$RBF : K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (15)$$

$$Polinomiale : K(x_i, x_j) = (x_i \cdot x_j + 1)^d \quad (16)$$

$$Sigmoidale : K(x_i, x_j) = \tanh(\alpha x_i^T x_j + \delta) \quad (17)$$

dove γ è la costante della base radiale;

d il grado di funzione polinomiale;

α la pendenza;

δ la costante intercettata.

Per C , parametro di regolarizzazione, che assume i valori 1,10,100 per le 4 tipologie di kernel elencate sopra, è stato selezionato il modello che riporta un valore di RMSE minore.

Tabella 4: Confidence Score

Kernel	C	RMSE
Rbf	1	14112.550
linear	1	1136.0570
poly	1	7706.632
sigmoid	1	14436.655
Rbf	10	11744.324
linear	10	11744.324
poly	10	7669.887
sigmoid	10	14436.655
Rbf	100	4252.16
linear	100	142531.376
poly	100	7584.393
sigmoid	100	14436.655

Quindi al fine di migliorare le prestazioni del modello è stato utilizzato un kernel lineare con $c=1$ ottenendo i seguenti risultati

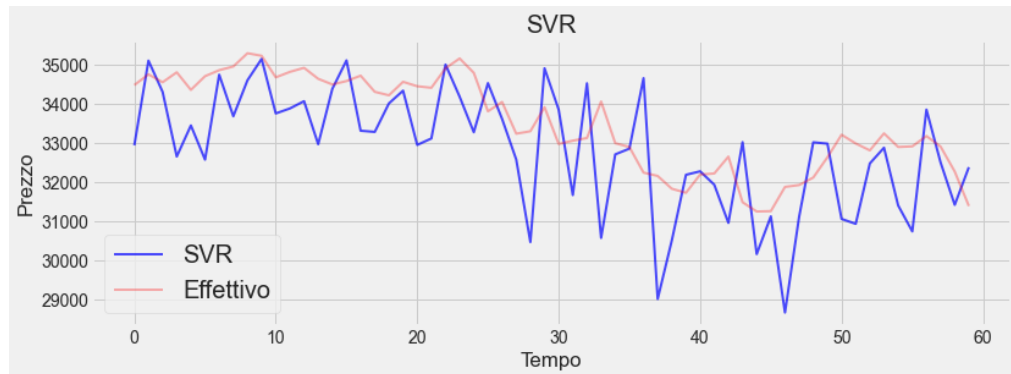


Figura 9: SVR

Tabella 5: Forecast Accuracy SVR

MAPE	ME	MAE	MPE	RMSE
0.032	-666.985	1079.024	-0.019	1365.212

3.3 Linear Regressor

La Regressione lineare⁷ è un modello statistico che studia la relazione tra due variabili, e cerca di prevedere il valore della variabile dipendente (Y), basandosi sul valore della variabile indipendente (X):

$$Y = \beta_0 + \beta_1 X \quad (18)$$

β_0 è chiamato "intercetta" e rappresenta il valore della variabile Y quando X=0

β_1 è chiamato coefficiente angolare e rappresenta la pendenza della retta.
al variare di β_1 si ha che :

- $\beta_1 > 0$, allora Y è direttamente proporzionale a X
- $\beta_1 < 0$, allora Y è inversamente proporzionale a X
- $\beta_1 = 0$, allora Y e X non hanno legami

Tabella 6: Forecast Accuracy Regressione Lineare

MAPE	ME	MAE	MPE	RMSE
0.010	51.904	349.246	0.0016	441.253

⁷<https://www.webtutordimatematica.it/materie/statistica-e-probabilita/modelli-di-regressione/regressione-lineare-semplificata/calcolo-parametri-retta-regressione>

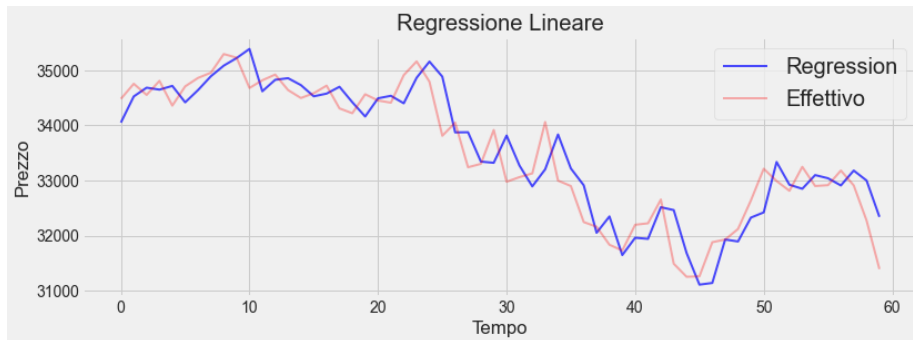


Figura 10: Regressione Lineare

Osservando attentamente si può notare come la previsione sia la serie temporale shiftata di un giorno, quindi con un ritardo, a cui viene aggiunta una soglia di errore. Nonostante i numerosi tentativi non è stato possibile rimuovere il lag dalla previsione, di conseguenza la regressione lineare non sarà presa in considerazione al fine di questo esperimento

3.4 KNN Regressor

KNN è uno degli algoritmi maggiormente utilizzati in tutti gli ambiti per via della sua estrema semplicità [7, 8] e della sua affidabilità [9]

KNN Regressor è un algoritmo utilizzato nel riconoscimento di pattern basandosi sulle caratteristiche dei k oggetti più vicini a quello considerato. L'output è il valore della proprietà per l'oggetto e corrisponde alla media dei valori dei k oggetti più vicini.

Come funziona:

1. Viene calcolata la distanza tra il nuovo punto e ogni punto del training test
2. Vengono selezionati i k punti più vicini, basati sulla distanza calcolata precedentemente.
3. Viene calcolata la media dei valori di questi punti come previsione finale del nuovo punto.

Per scegliere il valore di k ideale, sono stati riportati i valori di RMSE per l'algoritmo con un k che varia da 1 a 50 selezionando il valore di k con RMSE minimo

Prendere un basso valore di k , ad esempio $k=1$ porterebbe ad un overfitting del modello, mentre come si può vedere dalla Figura 11, prendere un valore di k troppo alto riduce le prestazioni.

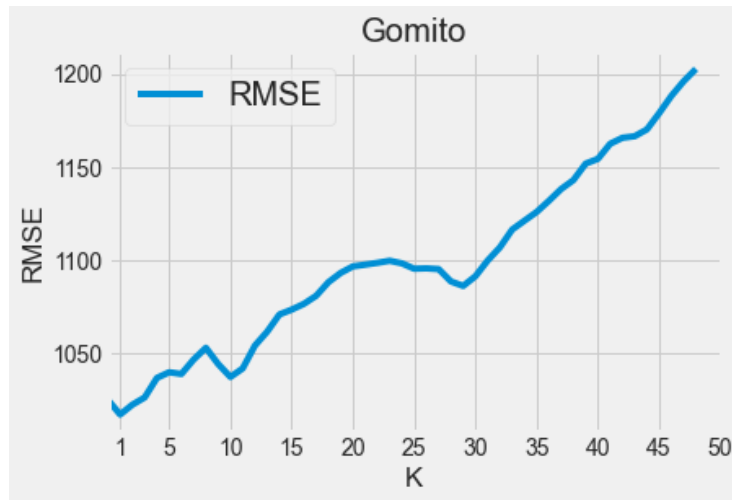


Figura 11: Gomito KNN

Scegliendo quindi $k=10$, si ottengono i seguenti risultati:

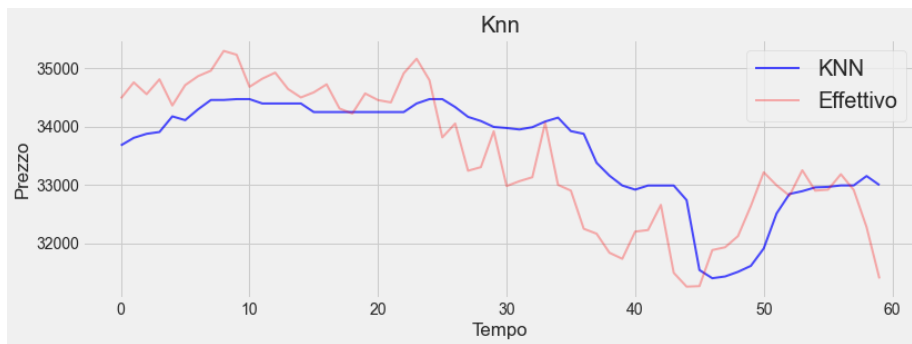


Figura 12: KNN

Tabella 7: Forecast Accuracy KNN

MAPE	ME	MAE	MPE	RMSE
0.0188	71.211	624.413	0.002	760.852

3.5 Decision Tree Regressor

Un albero decisionale è un algoritmo dall'elevata affidabilità [10] che pone diverse domande, ognuna delle quali restringe i possibili valori fino a quando non si è abbastanza sicuri per effettuare una previsione.

Per ogni risposta ci sono rami separati. Indipendentemente dalle risposte alle domande, alla fine si raggiunge una previsione (nodo foglia).

Un albero decisionale impara a mappare i dati agli output nella fase di training. Il modello apprende tutte le relazioni tra i dati e la variabile target.

Dopo la fase di training viene prodotto un albero, calcolando le domande migliori e il loro ordine da porre al fine di effettuare le stime più accurate possibili⁸.

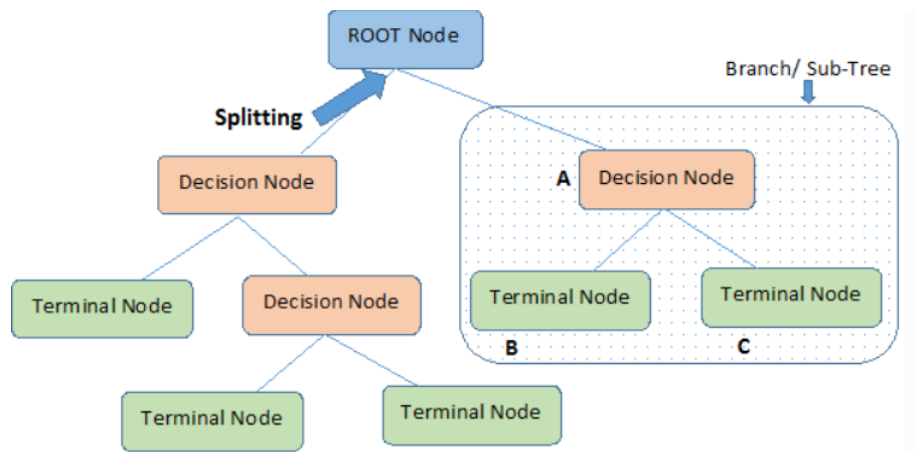


Figura 13: Decision Tree Regressor

Facendo riferimento alla Figura 15

- Root node : Rappresenta l'intero campione che sarà diviso in due o più insiemi omogenei.
- Decision node : Quando un sottonodo si divide in ulteriori sottonodi.
- Terminal node: Nodi che non si dividono, rappresentano i nodi pozzo.
- Branch/sotto-albero: una sottosezione dell'intero albero.

Come per Knn, sono stati riportati i valori di RMSE per l'algoritmo con "max depth"(valore che indica la profondità massima dell'albero) che varia da 1 a 50 e si seleziona il valore più vicino alla punta del gomito che si è creata

⁸<https://gdcoder.com/decision-tree-regressor-explained-in-depth/>

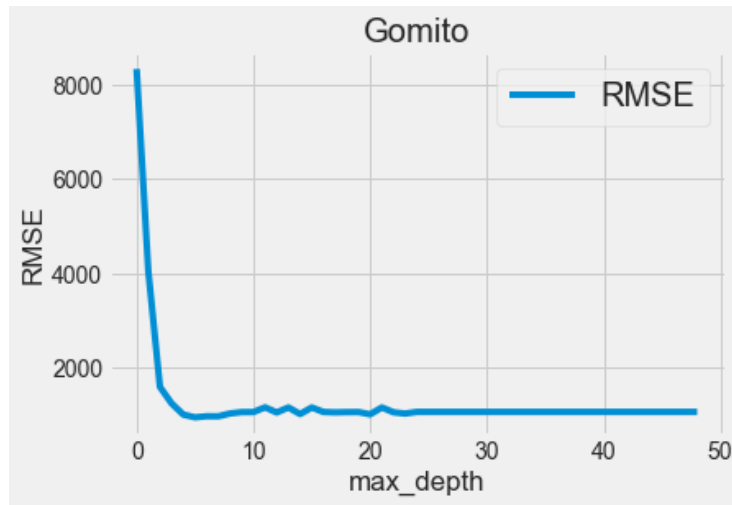


Figura 14: Gomito DTR

Utilizzando 6 come valore di profondità massima avremo la seguente previsione

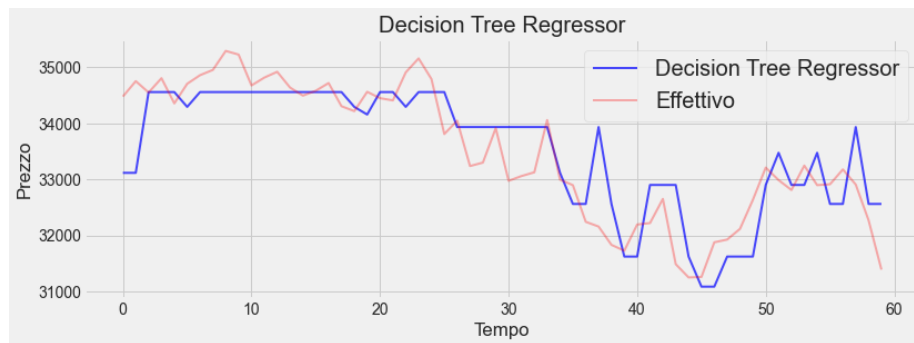


Figura 15: DTR

Tabella 8: Forecast Accuracy DTR

MAPE	ME	MAE	MPE	RMSE
0.0146	10.865	485.979	0.0006	631.053

3.6 Random Forest Regressor

Random Forest è un algoritmo di apprendimento supervisionato. L'idea su cui si basa è quella di combinare più decision tree per determinare l'output finale. Quindi vengono utilizzati più alberi decisionali, unendo le loro previsioni per ottenere una previsione più accurata[10]. Per fare ciò utilizza la tecnica dell'ensemble learning[11], il processo di utilizzare più modelli alla volta o un modello

più volte tenendo presente la media dei risultati di ogni modello, per costruire diversi alberi decisionali in punti dati casuali⁹.

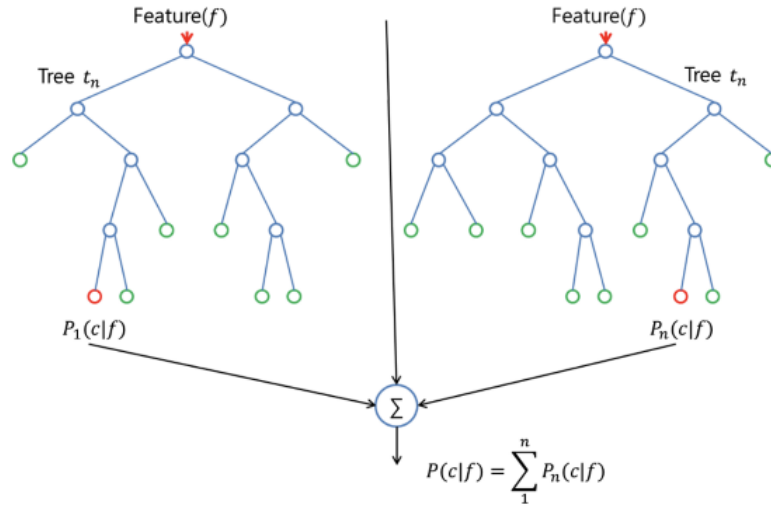


Figura 16: Random Forest Regressor

Per selezionare i parametri ideali, sono stati iterati 10 valori per il numero di alberi all'interno della foresta ($n_estimators$) e 10 valori per la profondità massima (max_depth): [10,20,30,40,50,60,70,80,90,100] valido per entrambi.

Sono stati utilizzati i parametri che danno come risultato un RMSE minore

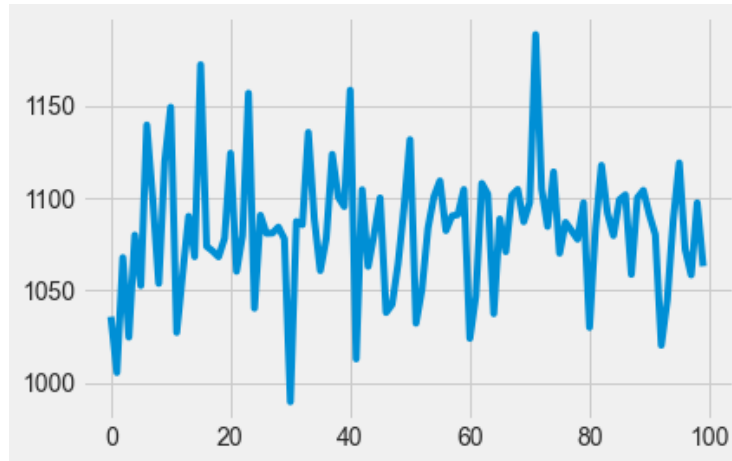


Figura 17: Random Forest Regressor Parameters

$$max_depth : 40, n_estimators : 10 \quad (19)$$

⁹<https://www.aionlinecourse.com/tutorial/machine-learning/random-forest-regression>

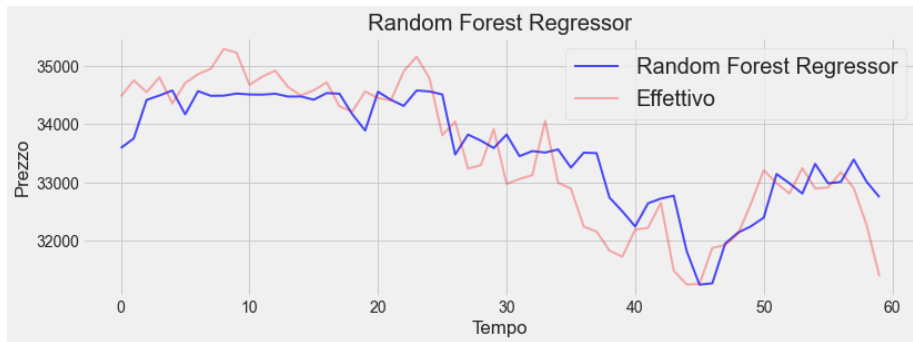


Figura 18: RFR

Tabella 9: Forecast Accuracy RFR

MAPE	ME	MAE	MPE	RMSE
0.0137	42.038	458.755	0.0016	574.703

4 Deep Learning

Il Deep Learning, o apprendimento profondo, è una sottocategoria del Machine Learning e comprende quegli algoritmi che simulano il comportamento del cervello umano utilizzando reti neurali che ne replicano il processo di apprendimento. Il sistema è addestrato attraverso set di dati non strutturati affinché la macchina assimili le informazioni e impari autonomamente.

Quindi grazie al Deep Learning si riuscirà a classificare autonomamente i dati trovando quelli più rilevanti per la soluzione del problema originario, migliorando ulteriormente il livello di prestazione tramite l'apprendimento continuo ¹⁰.

Rete neurale Con il termine rete neurale si intende una rete formata da neuroni cioè gruppi di unità che simulano il comportamento dei neuroni del cervello umano, interconnessi tra loro da collegamenti chiamati "sinapsi" in maniera tale da variare la loro configurazione in base ai dati passati come input.

Generalmente consiste in una divisione in tre strati:

- Strato di input
- Strati/o nascosti/o
- Strato di output

In un rete neurale con un unico strato nascosto, il numero di nodi nello strato di input dipende dalla grandezza dei dati. Tutti i nodi di layer diversi sono collegati tra loro a cascata, quindi l'output di un layer precedente diventa l'input del layer successivo. Il collegamento tra questi due

¹⁰<https://www.ai4business.it/intelligenza-artificiale/deep-learning/deep-learning-cose/>

nodi ha un coefficiente di peso e il suo continuo aggiornamento rappresenta il processo di learning che porterà ad avere un modello ottimale.

Lo strato nascosto applica una funzione di tangente iperbolica sulla somma di questi pesi, funzione di attivazione, che genera valori con tassi di errore minimo tra test e training set. Questi valori rappresentano i dati dello strato di output.

Nel caso in cui non rappresentassero un output ottimale, viene effettuato un processo di back propagation, che consiste in un collegamento tra output layer e hidden layer in cui vengono generati segnali contenenti i migliori pesi possibili per il numero di epochs desiderate.

Questo processo viene ripetuto numerose volte affinché il modello migliori le sue predizioni e diminuisca il tasso di errore [12]

4.1 Artificial Neural Network (ANN)

Per una descrizione completa si fa riferimento al paragrafo precedente. La Figura 19 tratta da [6] illustra una rete artificiale con uno strato di input, uno di output e due strati nascosti.

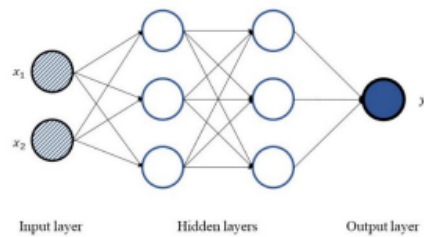


Figura 19: Artificial neural network

la relazione tra input, nodi e bias può essere dimostrata dall'equazione 20

$$Z = f\left(\sum_{i=1}^n x_i w_i + b\right) \quad (20)$$

dove:

- n rappresenta il numero di input;
- x_i sono gli input;
- w_i sono i pesi;
- f è la funzione di attivazione;
- b è il bias;
- z è l'output finale.

Uno dei vantaggi delle reti neurali artificiali, riguardante il loro utilizzo per la predizione dei prezzi delle azioni, è quello di imparare e generalizzare da dati non lineari. Può identificare nuovi campioni di test anche se non sono stati utilizzati durante l'allenamento. Inoltre è in grado di adattarsi al modello e alla relazione dei dati con conseguente miglioramento nelle predizioni e possibilità di lavorare su dataset più complicati. Quindi le ANN sono un modello pratico ed efficiente per la predizione dei prezzi degli stock [13]

Per questo esperimento è stata utilizzata una ANN con due strati nascosti aventi 128 unità ciascuno. Per selezionare il numero di epochs e la batch size ideali è stata utilizzata una funzione che itera diversi valori e calcola il MAPE, così da selezionare la coppia di parametri che produce un errore minore :

$$batch_size : 15, epochs : 10 \quad (21)$$

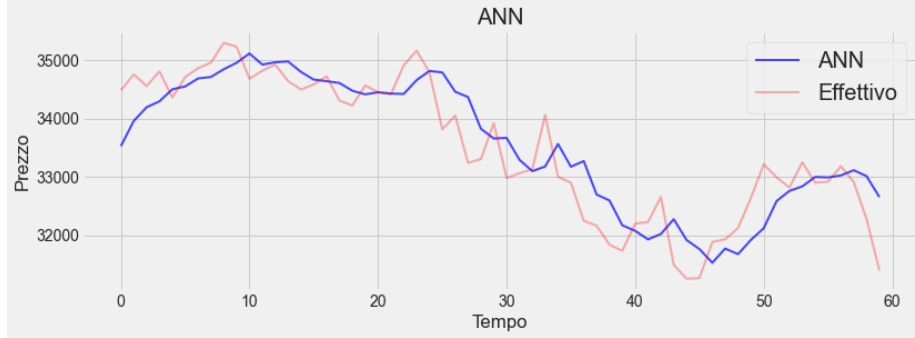


Figura 20: ANN

Tabella 10: Forecast Accuracy ANN

MAPE	ME	MAE	MPE	RMSE
0.0126	44.338	418.259	0.0015	522.930

4.2 Gated Recurrent Unit(GRU)

Prima di entrare nello specifico di una rete GRU, si dà una breve introduzione di una Recurrent Neural Network

RNN sono delle reti neurali in cui l'informazione non naviga in unico verso, ma può dunque tornare al livello precedente creando dei loop o si può avere un collegamento sullo stesso livello. Con questo concetto di ricorrenza, l'output di un neurone può influenzare se stesso in un momento successivo e viene quindi introdotta l'idea di memoria di una rete.

Gated recurrent unit è quindi una rete neurale ricorrente ed è stata proposta per la prima volta da Cho et al. [14] nel 2014, per far fronte al problema della scomparsa del gradiente cambiando il modo di calcolare le unità nascoste.

Una cella GRU è composta da due gate, un update gate(z_t) e un reset gate(r_t). Il primo si occupa di decidere quanti input (x_t) e output precedenti(h_{t-1}) devono essere passati alla cella successiva, il secondo determina quante delle passate informazioni debbano essere cancellate. In questo modo ci si assicura che solo le informazioni rilevanti vengano passate.

Se tutti i valori dei reset gates vengono settati a 1 e tutti i valori degli update gates vengono settati a 0 si ottiene una generica rete neurale[15]

Le seguenti formule spiegano come lavorano i diversi gate:

- Update gate :

$$z_t = \sigma(W_z * [h_{t-1}, x_t]) \quad (22)$$

- Reset gate :

$$r_t = \sigma(W_r * [h_{t-1}, x_t]) \quad (23)$$

In Figura 21 viene illustrata la struttura di un'unità GRU[16]

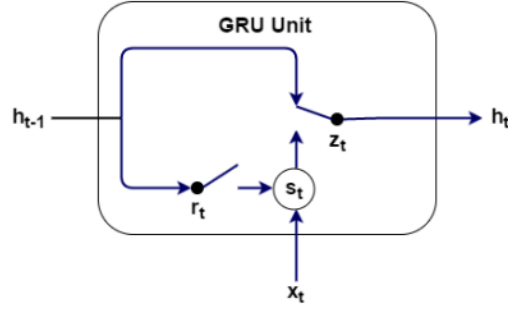


Figura 21: Struttura di un unità GRU

Così come svolto precedentemente vengono utilizzati due strati nascosti da 128 unità ciascuno e riutilizzando la stessa funzione si ottengono i seguenti parametri

$$batch_size : 50, epochs : 10 \quad (24)$$

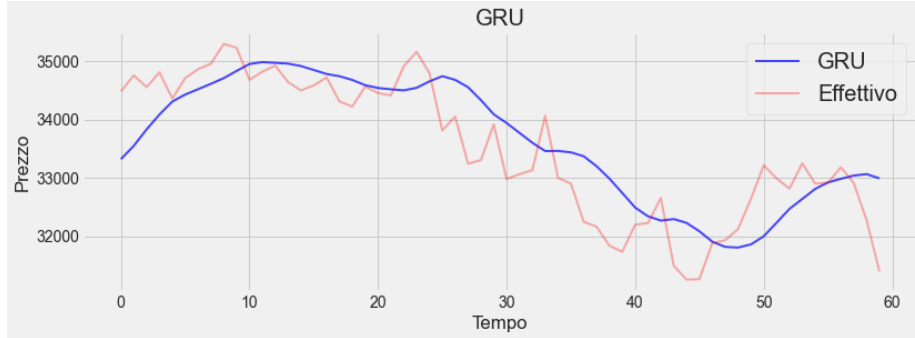


Figura 22: GRU

Tabella 11: Forecast Accuracy GRU

MAPE	ME	MAE	MPE	RMSE
0.016	121.412	535.683	0.003	663.344

4.3 Long Short Term Memory(LSTM)

Long short term memory è anch'essa una rete neurale ricorrente, molto simile alla GRU, ideata per la prima volta da Hochreiter e Schmidhuber nel 1997 [17].

Per risolvere il problema della memoria a breve termine vengono utilizzati diversi tipi di gate in modo tale da consentire solo agli input rilevanti di modificare le informazioni esistenti.

In questa tipologia di rete, sono presenti 3 gate [12]:

- Forget gate: utilizzato per decidere quali informazioni sono rilevanti e quali no; da in output un valore compreso tra 0 e 1 dove "1" conserva interamente l'informazione e "0" la ignora completamente
- Input gate: seleziona quali nuovi dati saranno immagazzinati nella cella quantificando quanto essi siano importanti
- Output gate: decide quale sarà l'output di ciascuna cella basandosi sullo stato di essa e sui dati filtrati aggiunti.

Di seguito viene mostrata una cella LSTM¹¹ :

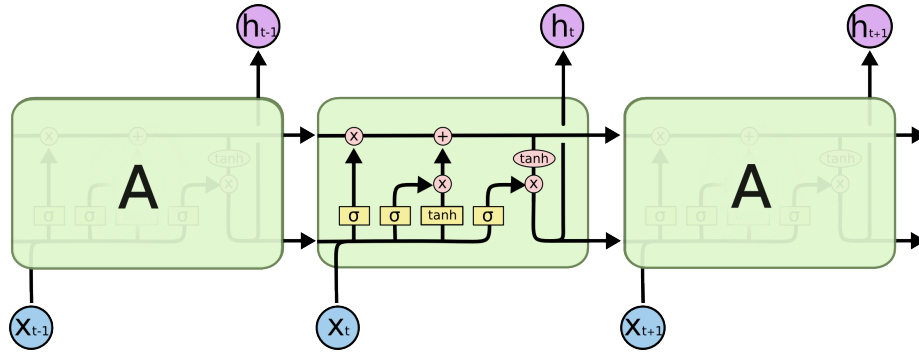


Figura 23: Cella LSTM

Anche in questo caso sono stati utilizzati due strati nascosti da 128 unità ciascuno utilizzando i seguenti parametri :

$$batch_size : 20, epochs : 10 \quad (25)$$

Tabella 12: Forecast Accuracy LSTM

MAPE	ME	MAE	MPE	RMSE
0.020	-86.071	678.710	-0.002	810.103

¹¹<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

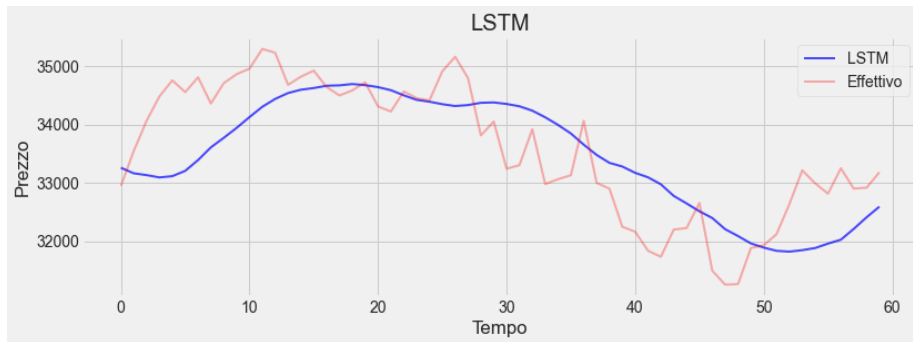


Figura 24: LSTM

5 Risultati Esperimento

5.1 Previsione sui componenti

Facendo riferimento alle Tabelle precedenti, sarà estrapolato il valore di RMSE per ogni modello, così da fare un confronto e stabilire il migliore.

Tabella 13: Forecast accuracy RMSE

Modello	RMSE
Arima	1310.699
SVR	1365.212
LR	441.253
KNN	760.852
DTR	631.053
RFR	574.703
ANN	522.930
GRU	663.344
LSTM	810.103

Avendo a disposizione i valori di RMSE, che può essere considerata come una sorta di distanza normalizzata tra le predizioni effettuate e i valori osservati, il modello che performa meglio in questo esperimento è ANN.

Di conseguenza viene usata una rete neurale artificiale per la previsione dei 30 titoli che compongono l'indice DJIA (Tab 1) e per ognuno di essi viene calcolato il MAPE, che indica il valore assoluto della percentuale di errore media, mostrato in Figura 25. Viene inoltre presa in considerazione la deviazione standard dei titoli per verificare se è correlata alla precisione della previsione. Per fare ciò sono stati presi i rendimenti semplici giornalieri dei titoli.

Tabella 14: MAPE componenti DJIA

Nome	MAPE	STD	Nome	MAPE	STD
3M	1.648	0.0136	Alphabet	10.522	0.0164
Amazon	20.203	0.0198	Amgen	3.858	0.016
Apple	8.491	0.0181	Boeing	19.352	0.0242
Caterpillar	4.155	0.0176	Chevron	4.014	0.018
Cisco	9.356	0.016	Coca-Cola	3.327	0.0114
Disney	13.599	0.016	Facebook	7.781	0.0229
Goldman Sachs	4.6635	0.0176	Home Depot	5.092	0.0147
Honeywell	2.048	0.0141	IBM	4.031	0.0148
Intel	6.691	0.019	Johnson e Johnson	2.172	0.0112
JPmorgan	6.166	0.0168	MC'donalds	3.729	0.0126
Microsoft	7.280	0.0166	Merck e CO	6.098	0.0134
Nike	8.364	0.0169	Protector e Gamble	4.177	0.0115
Salesforce	11.962	0.0214	Tesla	19.054	0.0357
United Health	4.059	0.0159	Verizon	5.064	0.0113
Visa	6.019	0.0158	Wal-Mart	9.587	0.0128

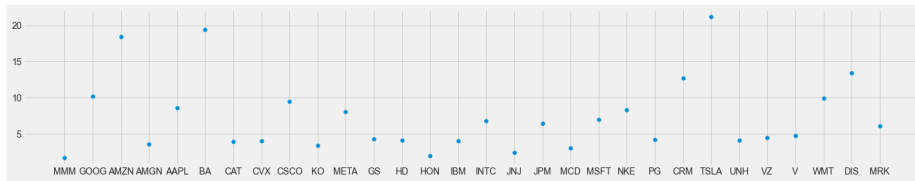


Figura 25: MAPE

Nella maggior parte dei casi, titoli con una deviazione standard più alta hanno un errore percentuale maggiore (spiccano Tesla, Amazon e Boeing) e viceversa, titoli con una deviazione standard medio/bassa presentano delle predizioni più accurate. Ciò era abbastanza intuibile, poiché, come detto prima, la STD rappresenta quanto i valori nel set di dati differiscono dalla media, ed è facile capire come più i valori siano variabili e si allontanino dalla media, più sia difficile avere una predizione accurata.

Lo scopo è dunque quello di utilizzare i titoli con un errore percentuale minore, così che le predizioni effettuate si distacchino il meno possibile dal valore effettivo rendendo più efficiente la strategia di mercato. Saranno dunque presi in considerazione i 10 titoli con un MAPE minore (MMM, AMGN, CAT, CVX, KO, HON, IBM, JNJ, MCD, UNH.) e sarà effettuata un'analisi di correlazione così da selezionare i titoli meno correlati tra loro:

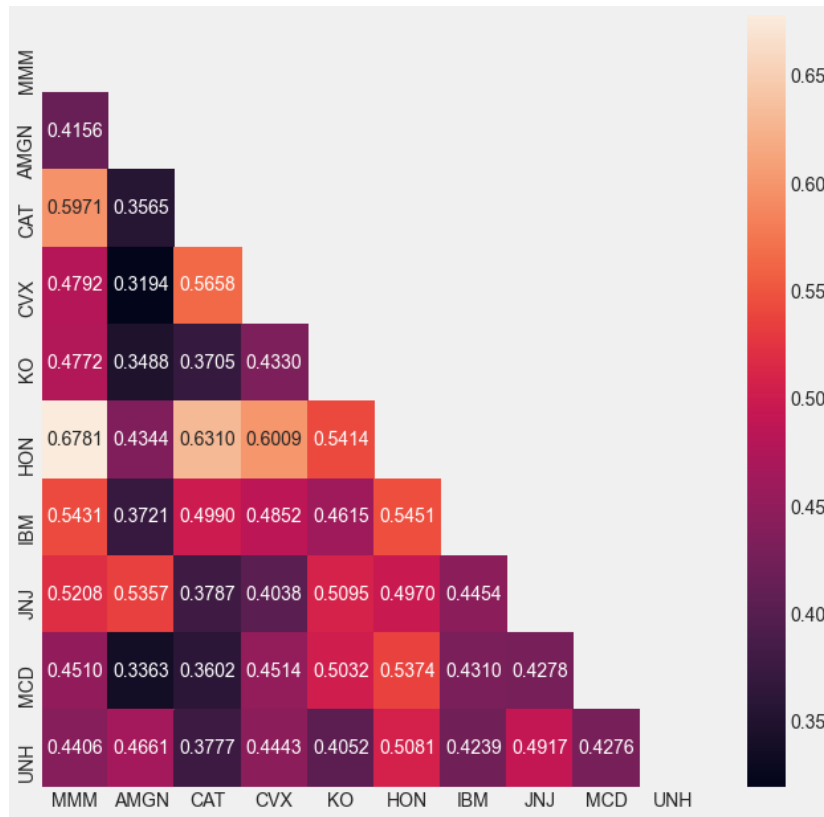


Figura 26: Correlazione tra i titoli

I titoli che saranno presi in considerazione sono quelli meno correlati tra di loro, quindi nel grafico quelli rappresentati da un colore più scuro. Ciò che si nota è che non sono presenti titoli fortemente correlati tra loro, ovvero con un indice di correlazione maggiore di 0.7. Il più alto indice di correlazione si vede infatti tra Honeywell e 3M e corrisponde al 67,8%, mentre il titolo che è generalmente meno correlato con gli altri è Amgen.

Pensando di selezionare 5 di questi 10 titoli per costruire il portafoglio, un buon compromesso per ridurre la correlazione potrebbe essere :

-Amgen

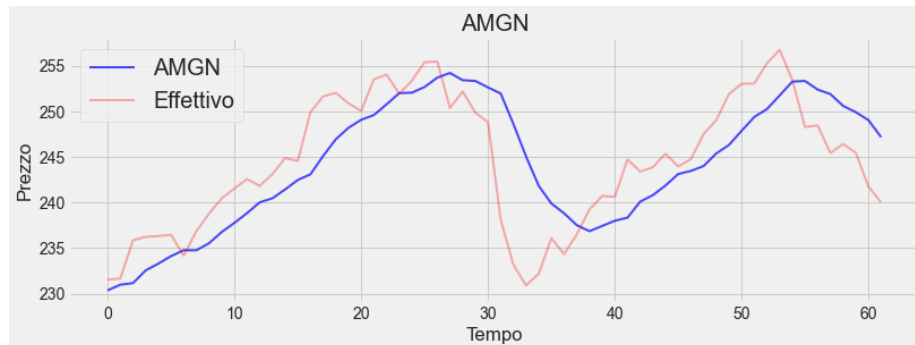


Figura 27: AMGN

-Caterpillar

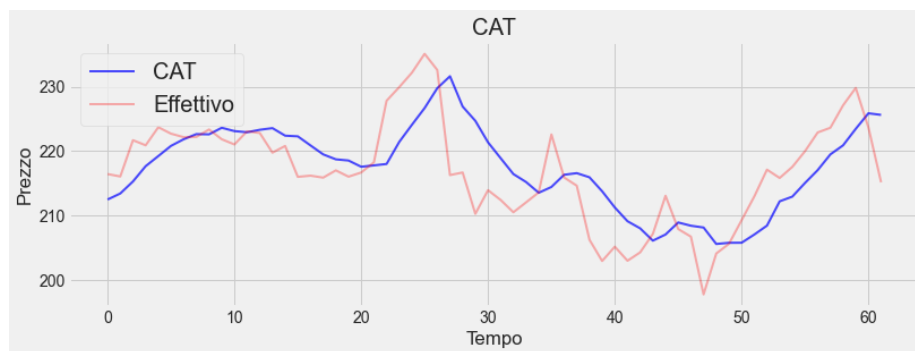


Figura 28: CAT

-Coca-Cola

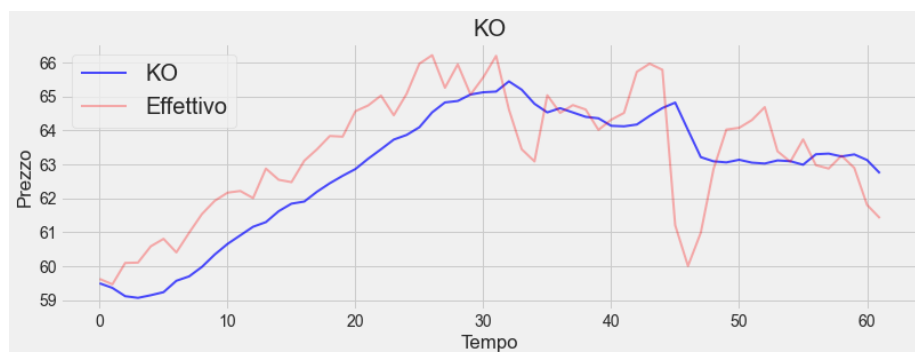


Figura 29: KO

-IBM

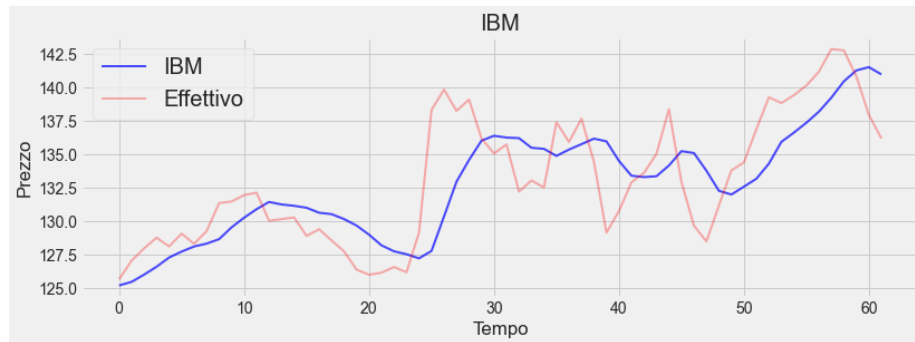


Figura 30: IBM

-United Health

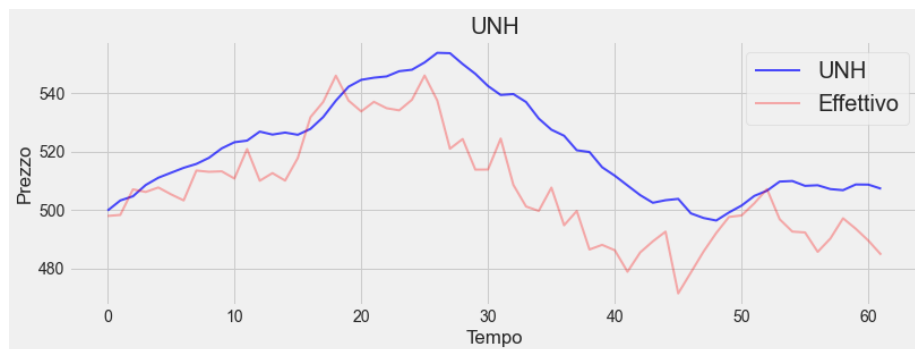


Figura 31: UNH

5.2 Frontiera efficiente

Saranno creati due portafogli e di essi verrà rappresentata la frontiera efficiente. Nel primo portafoglio sono usate le previsioni della rete neurale artificiale e nel secondo sono usati gli ultimi 60 giorni dei dati effettivi così da avere un confronto tra i due.

Per entrambi sono stati simulati 10^6 portafogli, utilizzando 100 di essi per la creazione della frontiera e per entrambi sono stati evidenziati il MVP e il portafoglio con il massimo sharpe ratio.

Per quanto riguarda i dati reali si ha la seguente frontiera efficiente

Si dà inoltre una rappresentazione dei due portafogli elencati in precedenza

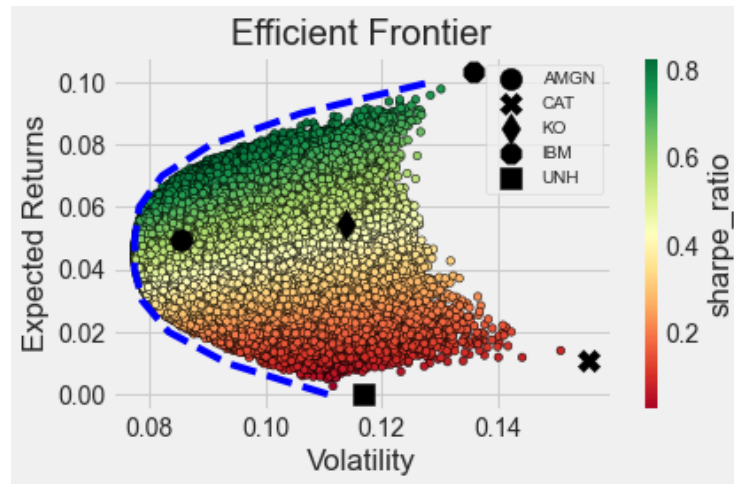


Figura 32: Frontiera efficiente dati effettivi

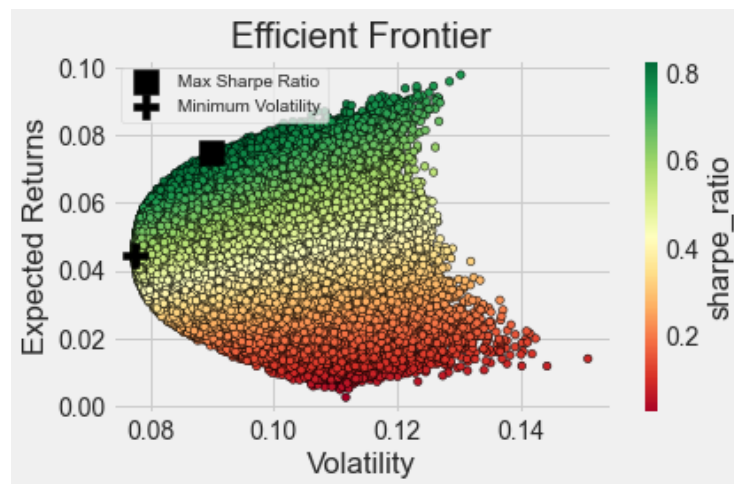


Figura 33: MVP e portafoglio con massimo sharpe ratio con dati effettivi

Tabella 15: Portafoglio dati effettivi

Portafoglio	Ritorno	Volatilità	Sharpe Ratio
Massimo Sharpe Ratio	7.49%	8.99%	83.35%
MVP	4.46%	7.75%	57.63%

Sono inoltre stati calcolati i pesi che ciascun titolo assume all'interno del portafoglio.

Tabella 16: Pesi Titoli Massimo Sharpe Ratio dati effettivi

AMGN	CAT	KO	IBM	UNH
47.39%	0.52%	5.05%	47.03%	0.01%

Tabella 17: Pesi Titoli MVP dati effettivi

AMGN	CAT	KO	IBM	UNH
65.24%	3.82%	15.18%	3.44%	12.32%

Si può notare dalla Figura 32 che AMGN sia il titolo meno volatile, di conseguenza il 65% del MVP è composto da questo titolo mentre IBM e CAT che sono i titoli più volatili compongono circa il 3%. Contrariamente nel Maximum sharpe ratio portfolio IBM compone circa la metà di esso, mentre UNH che ha un ritorno molto vicino allo 0% ha una presenza all'interno del portafoglio dello 0.01%.

Per quanto riguarda le previsioni si ha la seguente frontiera efficiente

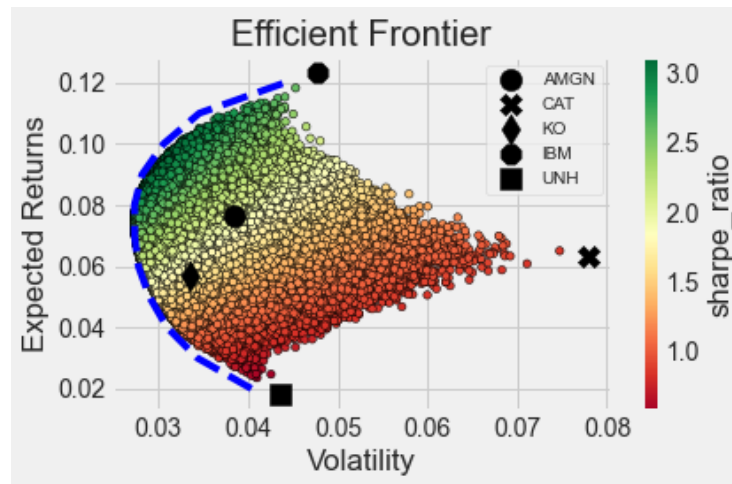


Figura 34: Frontiera efficiente dati predetti

Si da inoltre una rappresentazione dei due portafogli elencati in precedenza

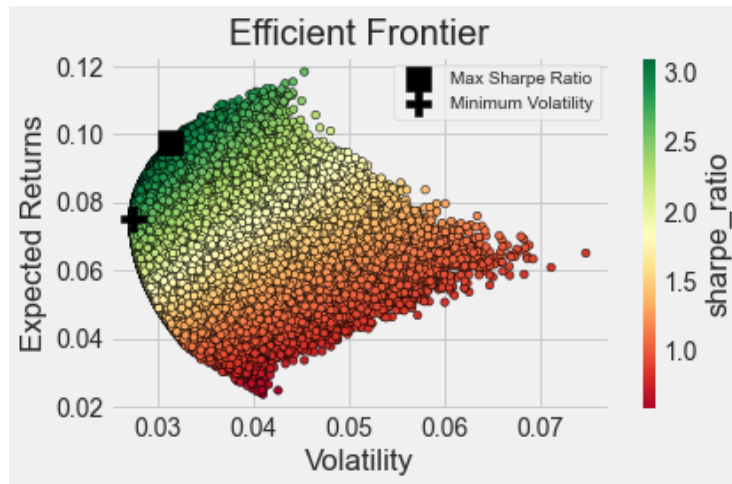


Figura 35: MVP e portafoglio con massimo sharpe ratio con dati predetti

Tabella 18: Portafoglio dati predetti

Portafoglio	Ritorno	Volatilità	Sharpe Ratio
Massimo Sharpe Ratio	9.74%	3.12%	312.04%
MVP	7.52%	2.73%	275.37%

Si nota che lo sharpe ratio è molto più elevato rispetto a prima, poichè la volatilità assume valori minori. Questo perchè è difficile prevedere cambiamenti repentini, ma le previsioni assumono degli andamenti più costanti.

Sono inoltre stati calcolati i pesi che ciascun titolo assume all'interno del portafoglio.

Tabella 19: Pesi Titoli Massimo Sharpe Ratio dati predetti

AMGN	CAT	KO	IBM	UNH
35.81%	0.61%	18.96%	44.31%	0.31%

Tabella 20: Pesi Titoli MVP dati predetti

AMGN	CAT	KO	IBM	UNH
37.09%	2.59%	42.90%	16.81%	0.61%

In questo caso è KO a rappresentare il titolo di dominanza all'interno dell MVP poichè rappresenta appunto il titolo meno volatile. IBM che ha una volatilità circa 3 volte minore rispetto ai dati effettivi, assume una maggior impor-

tanza a discapito di AMGN. Anche in questo caso CAT rappresenta il titolo più volatile e viene di conseguenza preso poco in considerazione.

Vengono infine paragonate le due frontiere efficienti.

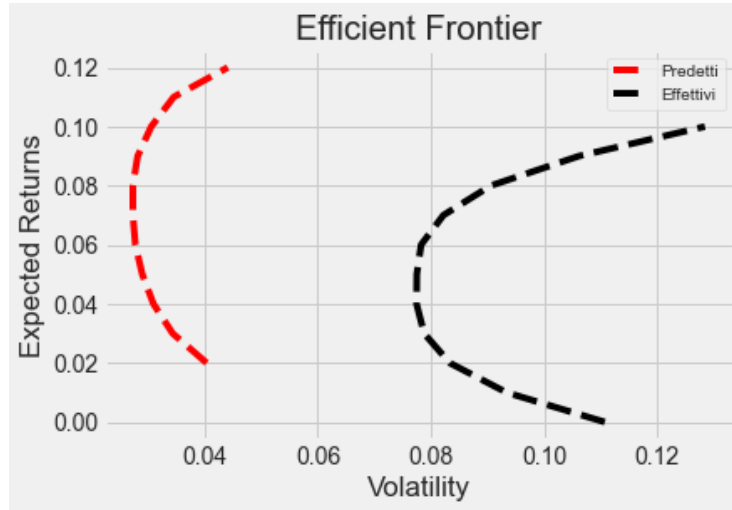


Figura 36: Frontiere efficienti a confronto

Come spiegato in precedenza, la frontiera efficiente dei dati predetti ha una volatilità notevolmente minore, e risulta essere leggermente ottimistica per quanto riguarda il ritorno minimo e il ritorno massimo. Un investimento sul MVP avrebbe portato a un ritorno di circa la metà di quello previsto, mentre un investimento sul portafoglio con massimo sharpe ratio avrebbe portato un ritorno leggermente inferiore. Anche se i dati effettivi non hanno rispettato a pieno le previsioni, si sarebbe ottenuto un guadagno.

6 Conclusione

In questo esperimento è stato utilizzato l'indice DJIA per confrontare diversi modelli di Machine Learning e Deep Learning tra loro per decretare il migliore utilizzando 5 metriche di misura. Per quanto riguarda gli algoritmi regressivi Random Forest Regression ha presentato un valore di RMSE minore, probabilmente dovuto alla tecnica dell'Ensemble Learning che permette la combinazione di più alberi decisionali per migliorare le previsioni. ANN ha avuto le prestazioni migliori di conseguenza è stato selezionato per la previsione dei titoli componenti l'indice DJIA. Di questi titoli sono stati utilizzati i 10 con un MAPE minore così da rappresentare una previsione più affidabile e di essi sono stati scartati i titoli maggiormente correlati tra loro per evitare rischi, seguendo il modello media-varianza per la selezione del portafoglio. Inserendo i seguenti 5 titoli all'interno del portafoglio : AMGN, CAT, KO, IBM, UNH, si otterrà un guadagno investendo nel MVP e nel portafoglio con massimo sharpe ratio, anche se non rispetta a pieno le previsioni effettuate, in quanto investendo nel MVP si avrebbe un ritorno positivo, ma della metà rispetto a quanto previsto. Bisognerebbe inoltre considerare i costi di transazione presenti nel mondo degli

investimenti che avrebbero portato a un guadagno inferiore. A conclusione di ciò si può affermare come utilizzare algoritmi di intelligenza artificiale possa aiutare nel mondo degli investimenti, ma anche che essi non possano garantire l'accuratezza totale della previsione in quanto essa risulta impossibile.

7 Bibliografia

Riferimenti bibliografici

- [1] Kranthi Sai "*Stock Market Prediction Using Machine Learning*"
- [2] Rohit Choudhry, and Kumkum Garg "*A Hybrid Machine Learning System for Stock Market Forecasting* "
- [3] Vaishnavi Gururaj,Shriya V.R.and Dr. Ashwini "*Stock Market Prediction using Linear Regression and Support Vector Machines*", 2019
- [4] Terence Shin "*Several Model Validation Techniques in Python A comprehensive guide to four popular cross validation techniques*", 2021
- [5] Cortes Corinna and Vapnik Vladimir, "*Support-vector networks*," 1995.
- [6] Mojtaba Nabipour, Pooyan Nayyeri,Hamed Jabani,Shabab S. and Amir Mosavi "*Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data;a Comparative Analysis*", 2020
- [7] Tao Ban, Ruibin Zhang, Shaoning Pang, Abdolhossein Sarrafzadeh and Daisuke Inoue "*Referential kNN Regression for Financial Time Series Forecasting*",2013
- [8] Brandsma, T. and Buishand, T. A.: "*Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling*", 1998.
- [9] Andrada-Fèlix, J.,Fernadez-Rodriguez, F., Garcia-Artiles, M.-D., Sosvilla-Rivero,S.: "*An empirical evaluation of non-linear trading rules. Studies in Nonlinear Dynamics & Econometrics*" , 2003.
- [10] S. Ravikumar and P. Saraf, "*Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms*,"2020
- [11] Martin Sewell "*Ensemble learning*", 2011
- [12] Adil Moghar, Mhamed Hamiche "*Stock Market Prediction Using LSTM Recurrent Neural Network*",2020
- [13] Erkam Guresen, Gulgun Kayakutlu , Tugrul U. Daim "*Using artificial neural network models in stock market index prediction*", 2011
- [14] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre et al. "*Learning phrase representations using RNN encoder-decoder for statistical machine translation*",2014.

- [15] Ya Gao, Rong Wang and Enmin Zhou *"Stock Prediction Based on Optimized LSTM and GRU Models"*,2021
- [16] Weiling Chen, Yan Zhang, Chai Kiat Yeo, Chiew Tong Lau, Bu Sung Lee *"Stock Market Prediction Using Neural Networks through News on Online Social Networks"*
- [17] Sepp Hochreiter, Jurgen Schmidhuber *"Long short-term memory"*, 1997

Ringraziamenti

Vorrei dedicare questo piccolo spazio alle persone che con il loro supporto mi hanno aiutato in questi anni di università. Prima di tutti Giulia, perchè pur non sapendo come accendere un computer si è interessata a tutto ciò che ho fatto, assicurandosi che fossi a conoscenza della materia, prima di ciascun esame, e contribuendo in maniera attiva alla preparazione dei progetti. Un Grazie ai miei genitori che mi hanno sostenuto in tutto e per tutto in questo percorso dandomi il loro appoggio costante. Ancor di più un Grazie a mia madre che ha sopportato le giornate più stressanti prima degli esami e che non me lo ha mai fatto pesare. Un ringraziamento al mio relatore Antonio Candelieri che mi ha dato la possibilità di approfondire argomenti avvincenti e utili per il futuro, offrendo piena disponibilità per tutta la durata della tesi. Grazie a tutti i compagni incontrati in università, su tutti Panda, Fabio e Pope con il quale è iniziata un amicizia che ci ha accompagnati in questo viaggio dal primo giorno. Un Grazie soprattutto a Gabriele che è sempre stato il mio appoggio fedele nella preparazione di un esame e la prima persona con la quale gioire al suo superamento. Un Grazie ai miei nonni che, anche se lontani, hanno costantemente fatto sentire la loro presenza e le loro lacrime di felicità. Infine Ringrazio i Boyz, Guzzo e Edo per avermi regalato le migliori avventure e tanti momenti di spensieratezza. In generale un grazie a tutte le persone che mi sono state vicine nei momenti di difficoltà e vi assicuro che ce ne sono stati tanti.