StockPrice Forecasting: Stock Price Forecasting Using Machine and Deep learning

Diaa Salama AbdElminaam[1],Donia Mostafa[2],
Ahmed AbdelRahman Gaber[3], Ahmed Mohamed Adel [4],Adham Said Ahmed [5] , Bassel AbdelRahim [6]
Faculty of Computer Science
Misr International University, Cairo, Egypt
diaa.salama[7], donia.mostafa[8],
ahmed2103540[9], adham2105856[10], ahmed2100403[11], bassel2107433[12]{@miuegypt.edu.eg}

***Abstract*—The use of machine learning and deep learning in stock market forecasting is a recent trend that generates predictions of current stock prices by training on historical data. We present a comparative analysis of five different models for stock price forecasting using machine learning and deep learning techniques. The models evaluated include Ridge Regression, Linear Regression using Weighted Moving Average (WMA), ARIMA, Gradient Boosting, and Extreme Gradient Boosting (XGB). Performance was assessed using R2 score, MSE, and MAE. The results indicate that the Ridge Regression model outperforms the others, achieving the highest R2-Score (accuracy) in the three datasets with average 99%, thereby demonstrating superior accuracy in predicting stock prices. This study underscores the importance of selecting an appropriate model for accurate stock price forecasting in financial applications. We highlight the effectiveness of the Ridge Regression approach in minimizing prediction errors, making it the most reliable model among those tested. The Linear Regression using Weighted Moving Average model also showed promising results with a reasonable error margin, while the ARIMA, XGBoosting, and Gradient Boosting models exhibited higher error rates.**

I. Introduction

Stock price forecasting plays a crucial role in the financial world shaping investment choices and economic strategies. Can you imagine the impact of accurate predictions on financial gains? It's huge! The rise of machine learning (ML) and deep learning (DL) has revolutionized stock price forecasting thanks to their ability to learn from historical data and make precise predictions. This paper delves into harnessing ML and DL techniques to boost accuracy in predicting stock prices.

Unlike traditional methods like technical analysis and time series models, which struggle with complex financial data patterns, machine learning is adept at uncovering hidden insights for more reliable forecasts. How cool is that? This study aims to scrutinize four models - Random Forest, Weighted Moving Average (WMA), ARIMA, and Gradient Boosting - to determine which one reigns supreme in stock price prediction.

The ultimate aim here is finding the most trustworthy model for predicting stock prices amidst market trends, economic indicators, and investor behavior swaying them. While ARIMA falls short with non-linear dependencies, machine learning models like Random Forest and Gradient Boosting hold promise with their pattern recognition prowess. Let's not forget about the underdog WMA model!

We're diving into Mean Squared Error (MSE) and Mean Absolute Error (MAE) as performance markers for each model tested in this study. Surprisingly enough (or maybe not so much), the WMA model steals the show with its minimal MSE of 0.0108 and MAE of 0.065 - talk about pinpoint precision in forecasting! Random Forest also steps up with decent results while ARIMA and Gradient Boosting stumble a bit with higher error rates.

Machine learning is all about teaching systems to learn independently based on data inputs - sounds pretty sci-fi, right? By training algorithms on datasets through supervised or unsupervised approaches or even reinforcement learning techniques is how these mystical patterns are unveiled in data streams.

Supervised learning vs unsupervised... it's almost like choosing between teachers guiding students or letting them explore freely! The former labels data for training models while the latter uncovers hidden patterns without labels. Smells like success when reinforcement learning enters the chat; using environments to enhance performance over time sounds just like real-life growth!

II. Related Work

Stock price forecasting is a rabbit hole many have dived, unearthing various findings along the way. Countless minds have delved into this realm, producing a plethora of results that span the spectrum of satisfaction. Our journey led us to sift through numerous research papers, extracting nuggets of wisdom to enhance our own exploration. Each paper we encountered will be duly noted in the references section, guiding our course with their insights and analyses.

In[1]The author focused on forecasting using machine learning techniques, The paper shows the huge importance of stock trading and the huge complexity that comes with trying to forecast it due to the influence that unknown factors like political and natural events have on the market.The author used SVM(support vector machine) model with a Radial Basis Function(RBF) kernel. The study used the SVM approach and historical stock data to forecast the prices for large and small companies across different markets. The methodology involves using features such as stock price volatility, momentum of the stock price and sector volatility. The results shows that the model was highly effective at forecasting the stock prices and achieved high accuracy. The author ends with how SVM models are effective for financial market forcasting.

In[2]The author suggested a hybrid approach to forecasting stock prices by using historical market data with real time news sentiment analysis. The author use LSTM(long short term memory) for the forecasting of stock prices based on past data patterns and markets indicators, He also implemented a sentiment analysis of financial news using a dictionary-based approach to understand thee sentiment around companies, Then the scores of the two models are combined to give an accurate forecast. The results shows that using the LSTM and .sentiment analysis together improves the accuracy of the forecasting offering a more safe and reliable recommendation.

In[3] The collected data from various sources including Yahoo Finance and NSE-India to create models to help forecast stock prices, extracting twelve features such as Moving Averages, RSI, and Volatility. the author used train and split technique an splitting it in to a 70 : 30 ration. The models that the author used in his research are as following SVM, Random Forest, KNN, and Softmax. The results were as the following the random forest outperformed all the other models. The study also found that lowering the number of features leads to decrease in the models accuracy.

In[4]The author highlighted the significance of meging Genetic Algorithms and Support Vector Machines (SVM) in a hybrid system for forecasting stock prices, emphasizing the improved performance of this approach compared to stand alone SVM systems. Additionally, they mark the importance of considering the correlation between stock prices of different companies in predicting stock price movements. They also discuss the classification of stock market direction as a two-class classification problem, with class values indicating either a rise or fall in stock prices. Furthermore, they focus on the selection and utilization of technical indicators as crucial input features for enhancing the accuracy of stock market predictions. These findings contribute to the existing body of knowledge in the field of stock market forecasting and machine learning applications.

In[5]a study conducted focusing on the use of deep learning algorithms for stock price forecasting, specifically targeting the close price. The main goal of their research was to test the prediction power of deep learning methods in comparison with other machine learning models such as SVR, ANN, and RF. Their findings shows that LSTM networks, outperformed the machine learning models in terms of accuracy. This study emphasizes the potential of deep learning to handle the nonlinear and dynamic nature of financial markets, giving significant improvements over conventional methods. Moreover, the research recommends further exploration into hybrid models to enhance predictive performance. The work by Nikou et al. contributes to the ongoing advancements in financial forecasting by demonstrating the practical benefits and remarkable performance of deep learning methodologies in the context of stock market predictions.

In[6]the author delves into the application of machine learning algorithms, especially Logistic Regression and SVM, for predicting stock movements in the Chinese stock market. Wang discusses the conclusiveness of these models in generating returns that surpass the HS300 index, highlighting the importance of quantitative investment strategies for investors to maximize earnings. The author compares different machine learning models used in stock market forecasting. Wang's research provides statistical analysis of China stock market data, details the implementation and evaluation of Logistic Regression and SVM models, and presents results showing the superior performance of the SVM model in predicting stock movements. The study emphasizes the significance of selecting appropriate machine learning models for stock selection and investment strategies, with the SVM model demonstrating promising results in achieving higher returns and lower drawdown rates compared to traditional market indices.

In[7]The aim is to investigate the use of machine learning algorithms for stock market forecasting, highlighting the importance of integrating sentiment analysis and news events with historical data to enhance prediction accuracy. The study discusses various machine learning techniques, including SVM, Random Forest, BDT, LSTM, and hybrid methods, for their effectiveness in stock price prediction. The results indicate that fusion algorithms, such as SVR-ANN, offer more accurate predictions compared to individual models.

In[8]the author underscores the importance of forecasting stock prices for informed decision-making, given the stock market's close ties to economic growth and substantial investments. They conducted a comparative analysis of forecasting algorithms, ranging from traditional ML methods to advanced Deep Learning and Neural Network models, as well as Sentiment analysis, Time series analysis, and Graph-Based algorithms. Future research opportunities include combining sentiment analysis with historical stock data for improved predictions, developing more effective stock recommendation systems, leveraging deep learning for enhanced feature extraction.. Based on the metrics and results provided by them, the Convolutional Neural Network(CNN) model in the Deep Learning and Neural Networks category demonstrated strong performance with an RMSE of 0.0087, MAE of 61, and MAPE of 2.16%. This indicates that the CNN model performed well in predicting stock prices based on the evaluation metrics provided.

In[9]This paper focuses on the development of hybrid model for stock price forecasting. They used ANN and decision tress to help increase prediction accuracy. It also aims for Recognizing the strengths and weaknesses of hybrid models.The study uses a dataset from Taiwan's electronic industry, incorporating both fundamental and technical analysis indicators. The author used PCA (Principal Component Analysis) in the data preprocessing after that he preformed the splitting of data by using train and split. The results were as the following the model with used ANN and DT achieved a 77% prediction accuracy outperforming individual ANN and DT model.

In[10] the author studies the forecasting of stock market closing prices using machine learning methods, like ANN, involves historical data from companies. The ANN can make accurate predictions. This model includes input, hidden, and output layers, with variables like H-L, O-C, 7 DAYS MA, 14 DAYS MA, 21 DAYS MA, 7 DAYS STD DEV, and Volume. The study emphasizes the need for efficient models to handle large and complex stock market datasets, showing how machine learning techniques can significantly improve prediction accuracy by 60-86 % compared to traditional methods.

In[11] the author focused on forecasting of the stock market by the use of machine learning algorithms. The paper shows the importance of stock market prediction due to its complexity influenced by various factors like political and natural events. The authors used support vector machines (SVM) to predict next-day stock trends by incorporating temporal correlations among global stock markets. The methodology involved using data from global stock indices, commodity prices, and foreign currencies. The study concludes that using global financial data enhances prediction accuracy and suggests SVM models are effective for financial market forecasting. The simulation results told us that the proposed model outperformed benchmark models, achieving an average profit of $814.6 over five 50-day periods, translating to an annual return rate of about 30 %. The authors suggest further exploration of creative methods and modifications to maximize profits even in bullish markets.

In[12]This study presents a novel approach to predicting stock market movements by integrating machine learning models with news sentiment analysis. The research focuses on utilizing historical stock market data and news sentiment collected over ten years for various stocks like AAPL, GOOGL, AMZN, and FB. The proposed AI framework incorporates Deep Neural Networks (DNN), Support Vector Regression (SVR), and Support Vector Machines (SVM) for prediction, achieving an impressive accuracy of 82.91 %. By combining technical features and sentiment analysis, the models outperform traditional methods and demonstrate the potential for profitable short-term trading strategies. The research highlights the importance of structured event features and the ability of DNN to learn hidden relationships, showcasing improved performance in daily stock price prediction compared to SVM and other models.

In[13]The paper focuses on utilizing machine learning algorithms to predict stock market trends by analyzing social media and financial news data. The primary goal is to enhance prediction accuracy by incorporating sentiment analysis and data mining techniques. The study aims to extract valuable insights from textual data available on social media platforms and news sources. By exploring the relationship between sentiment expressed in tweets and stock market features such as opening, closing, and trading volume, the researchers were able to predict key aspects of China's stock market with improved performance compared to baseline models. The study also emphasizes the impact of events posted on social media platforms on stock market returns, providing meaningful labels for event classification. By building efficient classifiers to judge sentiments of tweets, the researchers aimed to develop strategies for more effective trading.

In[14]The paper proposes some new stock price forecasting methods that integrates traditional financial features with media features using deep learning methods. The method involves training text feature vectors with Doc2Vec, reducing their dimension with stacked auto-encoder. The study compares the proposed method, named Doc-W-LSTM, with benchmark models like ARIMA, RNN, and LSTM. Results show that the Doc-W-LSTM model outperforms the baseline models in terms of MAE, RMSE, and R power 2 values. The goal of the study is to demonstrate that incorporating text features from social media can enhance stock price prediction accuracy, and the proposed method achieves this by effectively combining financial and text features for prediction.

In[15]The author examines the good are machine learning models in forecasting financial distress in companies, using a dataset containing various financial indicators. The study evaluates the performance of three models: Random Forest, Gradient Boosting, and SVM. The results show that the Gradient Boosting model surpasses the others in predictive accuracy, achieving an F1-score of 0.85. Additionally, the study aims to add to the existing literature by highlighting the potential of machine learning techniques in improving financial risk assessment and management. The findings indicate that these models are promising tools for identifying companies at risk of financial distress.

In[16] the objective of the paper was to develop models capable of providing insights into the general tendencies of the stock market during the covid-19 pandemic. The algorithms they used were LSTM and ARIMA to analyze and predict stock market values. They both performed reasonably accurate results, but LSTM outperformed the ARIMA model.

In[17]the author aims to predict stock market performance by applying different models such as: Linear regression, Logistic regression, Artificial Neural Networks using Bak Propagation. Artificial Neural Networks outperformed the other models in predicting stock prices as its results were more accurate, as the Back Propagation algorithm enables the prediction of future stock prices, especially for volatile stocks.

In[18] the target of the research is to use hybrid machine learning in predicting the stock prices. The models used in this study were DNN, and traditional Artificial Neural Networks, whereas the Deep Neural Networks were evaluated with various hidden layers ranging from 12 to 1000, and they were both applied to a PCA transformed data set and untransformed data set. The results showed that the Deep Neural Networks with an optimal number of hidden layers on the PCA transformed data set achieved the highest accuracy.

In[19] the author aims to compare the accuracy achieved in forecasting stock market prices between forecasting models and advanced machine learning techniques. The models they used were ARIMA, and MACD, Support Vector Machine, Random Forest, Gradient Boosting, Artificial Neural Networks. The results were that the advanced machine learning techniques out performed the traditional time series algorithms in the price forecasting.

In[20] The goal is to forecast stock market prices by thoroughly examining and evaluating the performance of various deep learning and machine learning models. The models tested included RandomForest, SVM, Multi-Layer perceptron regression, Knn, XGBoost, LSTM, and an ensemble model that combined LSTM, Random Forest, and XGBoost. The ensemble model demonstrated the best performance, achieving the highest R2 score and the lowest RMSE. The study also highlighted the critical role of hyper-parameter tuning in enhancing model performance.

III. Proposed Methodology

A lot of algorithms were used, and a research was done on each one of the models before training the model and using the models on the datasets. The following diagram represents the steps we went through to get the results.

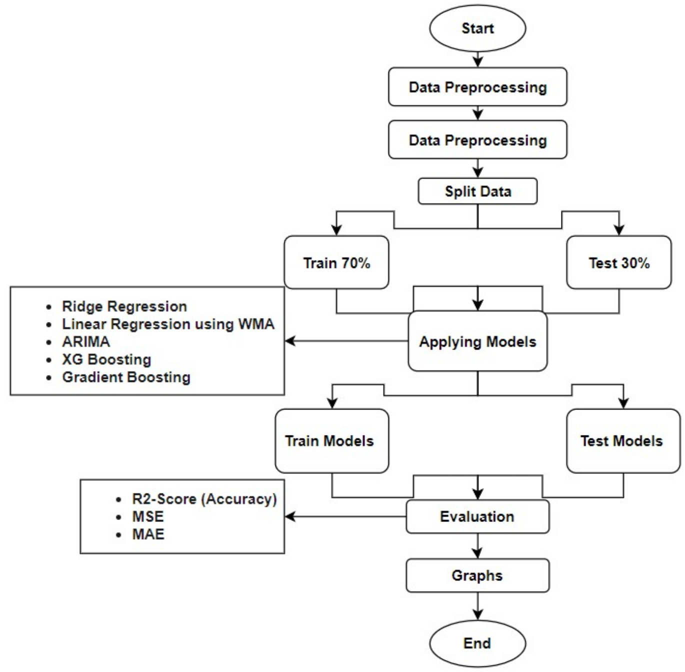

Fig. 1. Stock Prediction process

A. Datasets Descriptions

The data used for this study comprises historical stock price data for three prominent companies: Netflix, Inc. (NFLX), Google (GOOGL), and Yahoo! Inc. (YHOO). These datasets provide comprehensive records of the daily trading activity of each company's stock over a specified period, facilitating detailed analysis of market dynamics and investor behavior. Structured in a tabular format, each row represents a trading day, while columns capture various facets of the stock performance.

B. Key Attributes

The "Date" attribute serves as the primary temporal reference for the data, indicating the day on which the trading occurred. "Open," "High," "Low," and "Close" are a representation of the opening, highest, lowest, and closing prices of the stock, during the trading day. These numerical attributes provide insights into the price fluctuations and trading range of the stock.

The "Adj Close" attribute, present in the Netflix and Yahoo datasets, denotes the adjusted closing price of the stock. While absent in the Google dataset, the inclusion of this attribute enhances the completeness and analytical capabilities of the Netflix and Yahoo datasets.

The "Volume" attribute represents the total trading volume of the stock for the day, reflecting the level of market activity and liquidity.

This comprehensive overview highlights the key attributes present in the stock datasets for Netflix, Google, and Yahoo. Let me know if you need any further details or adjustments!

TABLE I
FEATURES OF THE NETFLIX STOCK DATASET

Feature	Type	Values
Date	Date	Formatted as MM/DD/YYYY
Open	Numerical	From 233.919998 to 692.3499768
High	Numerical	From 250.649994 to 700.98999
Low	Numerical	From 231.229996 to 686.090027
Close	Numerical	From 233.880005 to 691.690002
Adj Close	Numerical	From 233.880005 to 691.690002
Volume	Numerical	From 1144000 to 58904300 shares

TABLE II
FEATURES OF THE GOOGLE STOCK DATASET

Feature	Type	Values
Open	Numerical	From 279.12 to 816.68
High	Numerical	From 281.21 to 816.68
Low	Numerical	From 277.22 to 805.14
Close	Numerical	From 491.2 to 1216.83
Adj Close	Numerical	From 7900 to 24977900

TABLE III
FEATURES OF THE YAHOO STOCK DATASET

Feature	Type	Values
Open	Numerical	From 1833.400024 to 3612.090088
High	Numerical	From 3645.98999 to 3645.98999
Low	Numerical	From 3600.159912 to 3600.159912
Close	Numerical	From 3626.909912 to 3626.909912
Adj Close	Numerical	From 3626.909912 to 3626.909912
Volume	Numerical	From 9044690000 to 9044690000

The data preprocessing steps include data cleaning, which ensured all values are valid and consistent, handling any missing or erroneous entries. Normalization was performed to adjust the scale of numerical features to improve model performance and convergence speed. Feature engineering involved creating additional features that might enhance the predictive power of the models, such as moving averages, volatility measures, and lagged variables.

These data values are instrumental in training and evaluating various machine and deep learning models for stock price forecasting. By leveraging historical price data, the models can learn patterns and trends, enabling them to predict future stock prices with varying degrees of accuracy.

The data preprocessing steps include:

1. **Data Cleaning**: Ensured all values are valid and consistent, handling any missing or erroneous entries.
2. **Normalization**: Adjusted the scale of numerical features to improve model performance and convergence speed.

C. Used Algorithms

The datasets we used were passed into 5 different Machine and deep Learning models which were Ridge Regression, Gradient Boosting, XGBoost, Linear regression using WMA(weighted moving average), and Arima . For each one of the models there were outputs generated, and they were: Accuracy, MSE(mean square error), and MAE(mean absent error). After that the results were plotted and compared. The results, plots, and the results can be found later in the paper.

1) **Gradient Boosting:**
Gradient boosting is a really powerful ensemble in machine learning. Unlike traditional models that operate independently, boosting merges the predictions of more than one weak learner to form one single, more accurate learner. This model is renowned for its prediction speed and accuracy, especially with large and complex datasets. The core principle of boosting involves initially building a model on the training dataset, followed by constructing a second model to correct the errors of the first model.

$$F_0(x) = \arg\min_{\rho} \sum_{i=1}^{N} L\left(y_t, \rho\right) \quad (1)$$

$$\tilde{y}_t = -\left[\frac{\partial L\left(y_i F\left(x_i\right)\right)}{\partial F\left(x_i\right)}\right]_{F(x)=F_{n-1}(x)}, i = 1, N \quad (2)$$

$$a_m = \arg\min_{a,\beta} \sum_{i=1}^{N} \left[\tilde{y}_l - \beta h\left(x_i : a\right)\right]^2 \quad (3)$$

$$\rho_m = \arg\min_{\alpha,\beta} \sum_{t=1}^{N} L\left(y_t, F_{m-1}\left(x_t\right) + \rho h\left(x_t; a_m\right)\right) \quad (4)$$

$$F_m(x) = F_{m-1}(x) + \rho_m h\left(x, a_m\right) \quad (5)$$

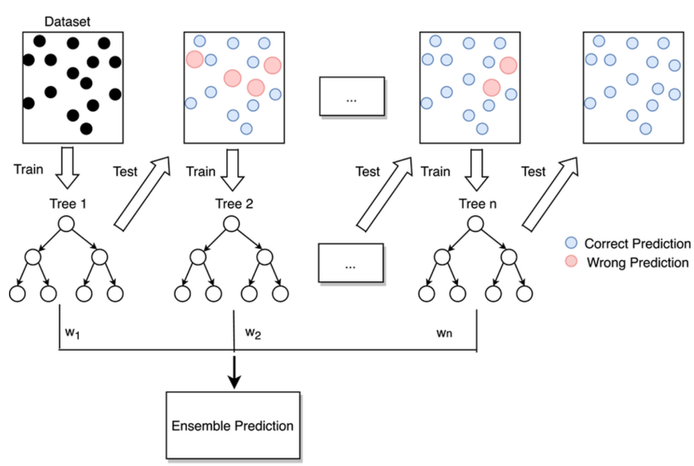

Fig. 2. Illustration of Gradient Boosting

2) **Ridge Regression:**
Ridge regression, also called L2 Regularization, is a type of regularization used in many regression models. Regularization helps reduce errors caused by overfitting on the training data. Ridge Regression specifically addresses multicollinearity in regression analysis, which is beneficial when features have high weights. By penalizing these features, Ridge Regression can produce better results. .

$$\frac{1}{n} \sum_{i=1}^{n} \left(h_\theta(x)^i - y^i\right)^2 + \lambda(\text{ slope })^2 \tag{6}$$

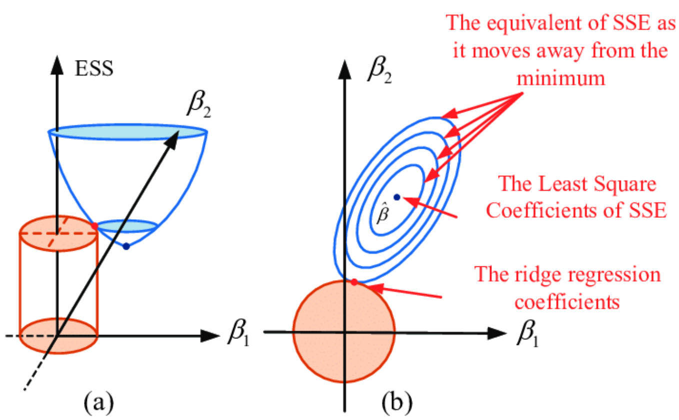

Fig. 3. Illustration of Ridge Regression

3) **XGBoost:**
The Very Acute Enhancing Boosting (XGBoost) mode is a strong tool for supervised learning jobs such as regression, classifications, and ranks. Its major highlights involve building an ensemble of decisions trees one by one to fixing errors, utilizing regularization to avoid overfitting, efficiently handling missing data, and trimming trees for more excellent performance. Xtreme Boost could also favor parallel computing for faster training and is optimized to handle sparse data efficiently. This model has been extensively used in various sectors like finances and bioinformatics because of its capability to capture complex patterns and provide exact outcomes. In general, XGBoost merges the advantages of gradient boosting with sophisticated techniques, creating it a versatile and effective model for predictive analytics.

$$\operatorname{Obj}(\theta) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega\left(f_k\right) \tag{7}$$

Where:
- $l\left(y_i, \hat{y}_i\right)$ is the loss function that measures the difference between the actual value y_i and the predicted value $\hat{y}_i$.
- $\Omega\left(f_k\right)$ is the regularization term for the k-th tree, which helps to control the model complexity.

4) **Arima:**
The Autoregressive integrated moving average(ARIMA) models a famous tool for the forecast of series of time-related information. There are three main pieces: autoregressive shuffle, differencing, and moving average. The autoregressive section examines previous values to predict the current one. The involved part is about making the data stationary by removing trends. The moving average portion looks at the distinctions between observations and the errors between previous time periods. An ARIMA model is written as ARIMA(p,d,q), where p, d, and q specifies what number of lag observations, differing times, and the moving average window size, respectively. By recognizing these values utilizing techniques like the Autocorrelation action(ACF) and Partial Autocorrelation operation(PACF) plots, an ARIMA model can predict upcoming data points accurately. This model is useful in many areas like finance, economics, and environmental science to support decision-making processes.
The arima model equations it is were the combination of the arima component happen :

$$\begin{aligned} y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \\ + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \end{aligned} \tag{8}$$

Where:
- y_t is the value of the time series at time t,
- c is a constant,
- $\phi_1, \phi_2, \ldots, \phi_p$ are the coefficients of the autoregressive,
- $\theta_1, \theta_2, \ldots, \theta_q$ are the coefficients of the moving average,
- ϵ_t is the error at time t,
- $\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-q}$ are the error from previous time periods.
[21]

5) **Linear Regression Using Weighted Moving Average (WMA):**
Linear regression is a strong statistical method used for forecasting values based on historical data. It creates the relation between a target and one or more independent variables by use of a linear equation to observed data. And WMA is a purely statistical method. This process involves assigning varying weights to data points within a moving average calculation. This method is utilized to smooth a realized time series and identify short-term, long-term trends, and seasonal components effectively. By applying a **k-day moving average**, where each data point is multiplied by a specific weight before being averaged, the weighted moving average process offers a more nuanced approach to capturing different patterns

and trends in the data.[22] The flexibility of adjusting weights based on the significance of each data point allows for a more precise analysis of the data, making it a valuable tool in forecasting and trend analysis within time series modeling.

$$Ft = \frac{\sum(weight\,for\,period\,n) \times (demand\,in\,period\,n)}{\sum weights} \quad (9)$$

D. *Performance Metrics*

R2 score also known as coeffecient of determination, is used to evaluate The R2 score, also known as the coefficient of determination, is used to evaluate regression models by measuring the level of variation between the independent variables (features) and the dependent variable (target). Its range is from 0 to 1. Mean Squared Error (MSE) is a metric used to evaluate a regression model's performance by measuring the average of the squares of the errors, where the error is the difference between predicted and actual values. MSE is sensitive to outliers because it squares the errors, giving more weight to larger errors. Mean Absolute Error (MAE) is another metric for evaluating regression models, measuring the average of the absolute differences between predicted and actual values. Unlike MSE, MAE is less sensitive to outliers because it doesn't square the errors. Root Mean Squared Error (RMSE) is the square root of MSE and is measured in the same units as the target variable.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (10)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of data points.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \quad (11)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (12)$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \quad (13)$$

IV. Results and Analysis

The results taken from Gradient Boosting, Arima, Linear Regression, Ridge regression, XGBoost are as follow.

The results below are from the Netflix Stock data set.

TABLE IV
OUTPUTS OF MODELS WITH 70:30 SPLIT OF

Models	MAE	MSE	R2 Score
Arima	8.5319	13.4396	0.9485
Gradient Boosting	25.55	1793.55	0.635
Linear regression	0.60647	0.01077	0.9807
Ridge regression	1.838	6.6757	0.998
XGBoost	14.675	247.03	0.8456

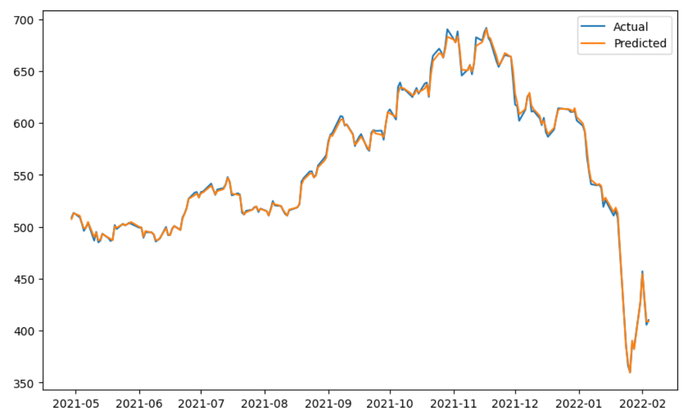

Fig. 4. First dataset the actual data vs the predicted values with Ridge regression

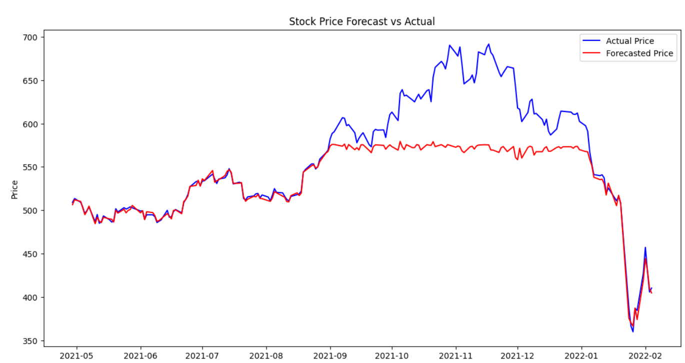

Fig. 5. First dataset the actual data vs the predicted values chart with Gradient Boosting

Fig. 6. First dataset the actual data vs the predicted values chart with Arima

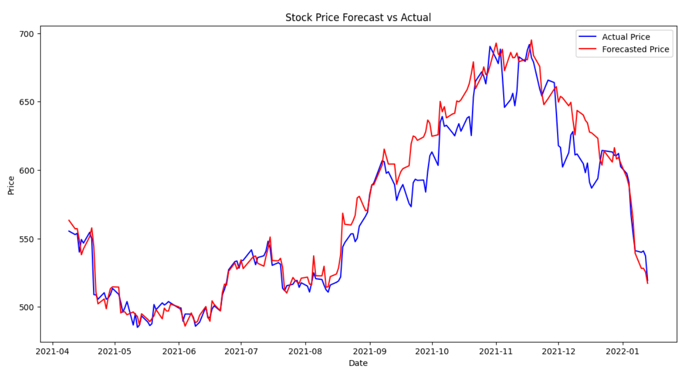

Fig. 7. First dataset the actual data vs the predicted values chart with XGBoost

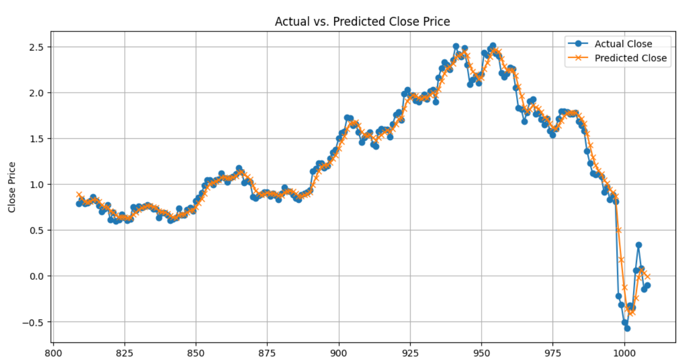

Fig. 8. First dataset the actual data vs the predicted values chart with Linear regression

Linear regression and Ridge regression are the best algorithms, Having an accuracy of 0.9807 and 0.998 with Ridge Regression being better in terms of accuracy. Arima was a took third place with and accuracy of 0.9485. Gradient Boosting was the worst one of all the models with an accuracy of 0.635, and XGBoost had an accuracy of 0.8456.

The results below are from the Google stock data set.

TABLE V
OUTPUTS OF MODELS WITH 70:30 SPLIT OF

Models	MAE	MSE	R2 Score
Arima	6.69	9.2733	0.977
Gradient Boosting	3.073	4.28	0.9937
Linear regression	0.60647	0.01077	0.9397
Ridge regression	2.2458	8.652	0.997
XGBoost	4.152	30.41	0.9896

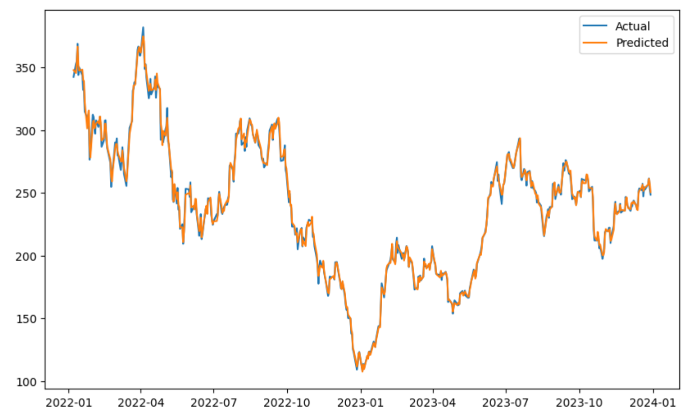

Fig. 9. Second dataset the actual data vs the predicted chart with Ridge regression

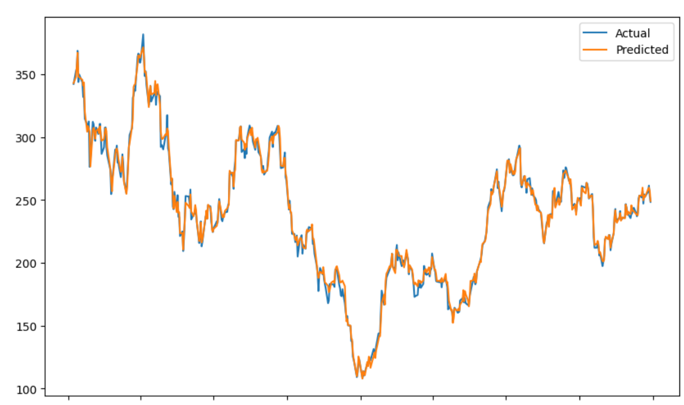

Fig. 10. Second dataset the actual data vs the predicted chart with Gradient Boosting

Fig. 11. Second dataset the actual data vs the predicted chart with Arima

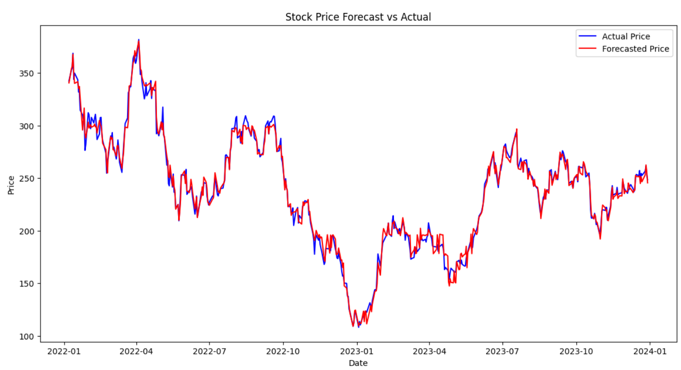

Fig. 12. Second dataset the actual data vs the predicted chart with XGBoost

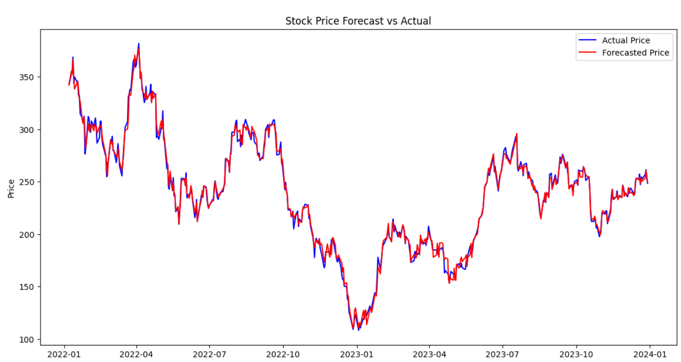

Fig. 13. Second dataset the actual data vs the predicted chart with Linear regression

The second dataset had an overall higher accuracy than the first one. Ridge regression and Gradient Descent are sharing the same accuracy which is 0.99 but the Ridge is a little higher with 0.997 . XGBoost had a much higher accuracy than in the first data set with 0.9896, while Linear regression came last with 0.939.

The results below are from the Yahoo stock data set.

TABLE VI
OUTPUTS OF MODELS WITH 70:30 SPLIT OF

Models	MAE	MSE	R2 Score
Arima	22.129	41.106	0.971
Gradient Boosting	132.91	33848.79	0.5501
Linear regression	0.00856	0.0571	0.980
Ridge regression	5.8272	61.7875	0.999
XGBoost	134.71	33851.26	0.55

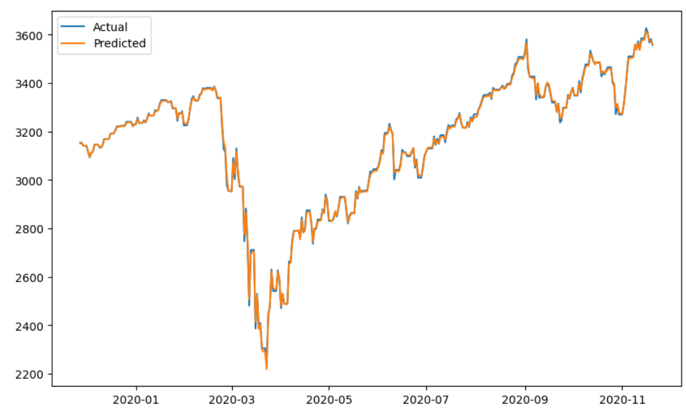

Fig. 14. Third dataset the actual data vs the predicted chart with Ridge regression

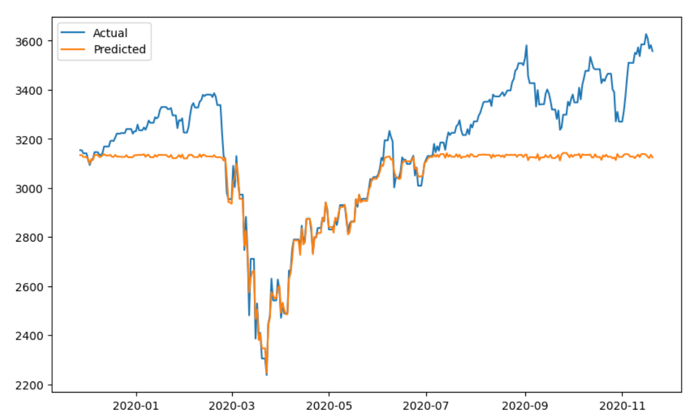

Fig. 15. Third dataset the actual data vs the predicted chart with Gradient Boosting

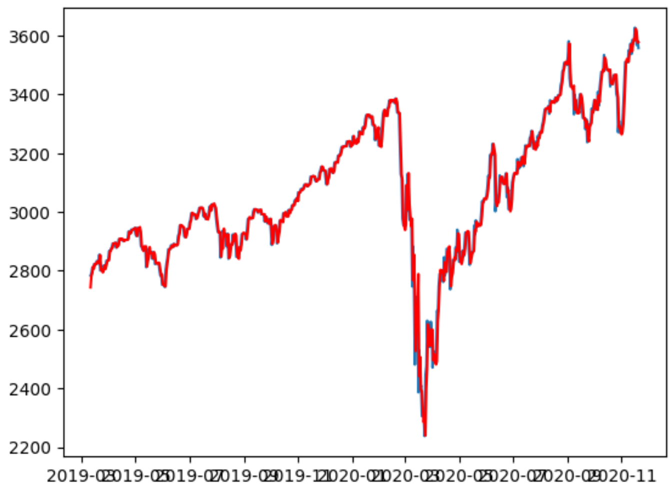

Fig. 16. Third dataset the actual data vs the predicted chart with Arima

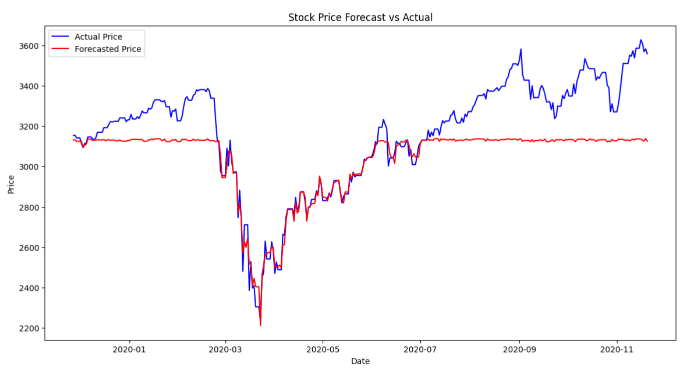

Fig. 17. Third dataset the actual data vs the predicted chart with XGBoost

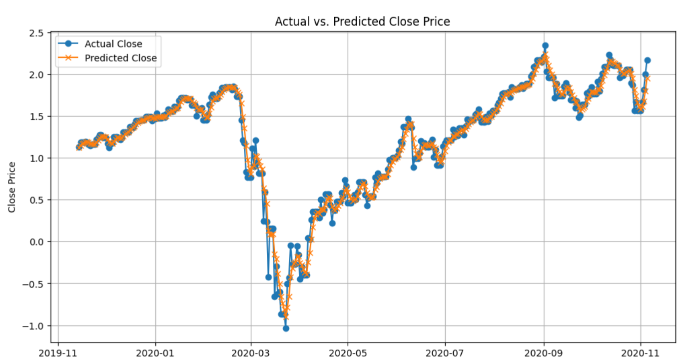

Fig. 18. Third dataset the actual data vs the predicted chart with Linear regression

Unsurprisingly, Ridge regression is the one the best accuracy scoring a 0.999. Results are dowen for almost all the models, All the models had a much higher MSE. XGBoost was in last place, scoring a 0.55 in terms of accuracy.

V. CONCLUSION

We conclude that the application of machine learning algorithms such as Linear regression with Weighted Moving Average, Gradient Boosting, ARIMA, Ridge Regression, and XGBoost has demonstrated promising results in forecasting stock prices. In particular Ridge Regression and ARIMA which remained consistent in their high performances throughout all of our data sets. These algorithms have shown their potential in capturing patterns and trends in historical stock price data, enabling accurate predictions of future stock prices. As the field of machine learning continues to evolve and improve, these algorithms offer valuable tools for analysts and investors to make informed decisions in the dynamic world of stock trading.

VI. ACKNOWLEDGMENT

We would like to express how grateful we are to Dr. Diaa Salama for being an amazing educator, and for wisely guiding and inspiring us to be better learners and high performing academics. And of course we can't forget to mention our heartfelt gratitude and honor to be part of the Misr International University student community, and for the amazing opportunity, learning environment, and beneficial resources provided for us throughout the years.

REFERENCES

[1] V. K. S. Reddy and K. Sai, "Stock market prediction using machine learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 10, pp. 1033–1035, 2018.

[2] S. Sarode, H. G. Tolani, P. Kak, and C. Lifna, "Stock price prediction using machine learning techniques," in *2019 international conference on intelligent sustainable systems (ICISS)*. IEEE, 2019, pp. 177–181.

[3] I. Kumar, K. Dogra, C. Utreja, and P. Yadav, "A comparative study of supervised machine learning algorithms for stock market trend prediction," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, 2018, pp. 1003–1007.

[4] R. Choudhry and K. Garg, "A hybrid machine learning system for stock market forecasting," *International Journal of Computer and Information Engineering*, vol. 2, no. 3, pp. 689–692, 2008.

[5] M. Nikou, G. Mansourfar, and J. Bagherzadeh, "Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms," *Intelligent Systems in Accounting, Finance and Management*, vol. 26, no. 4, pp. 164–174, 2019.

[6] H. Wang, "Stock price prediction based on machine learning approaches," in *Proceedings of the 3rd International Conference on Data Science and Information Technology*, 2020, pp. 1–5.

[7] S. Singh, T. K. Madan, J. Kumar, and A. K. Singh, "Stock market forecasting using machine learning: Today

and tomorrow," in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, vol. 1. IEEE, 2019, pp. 738–745.

[8] P. Soni, Y. Tewari, and D. Krishnan, "Machine learning approaches in stock price prediction: A systematic review," in *Journal of Physics: Conference Series*, vol. 2161, no. 1. IOP Publishing, 2022, p. 012065.

[9] C. F. Tsai and S. P. Wang, "Stock price forecasting by hybrid machine learning techniques," in *Proceedings of the international multiconference of engineers and computer scientists*, vol. 1, no. 755, 2009, p. 60.

[10] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "Stock closing price prediction using machine learning techniques," *Procedia computer science*, vol. 167, pp. 599–606, 2020.

[11] S. Shen, H. Jiang, and T. Zhang, "Stock market forecasting using machine learning algorithms," *Department of Electrical Engineering, Stanford University, Stanford, CA*, pp. 1–5, 2012.

[12] W. E.-H. M. Mokalled and M. Jaber, "Automated stock price prediction using machine learning," in *Proceedings of the Second Financial Narrative Processing Workshop (FNP 2019)*, 2019, pp. 16–24.

[13] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media, news," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–24, 2022.

[14] X. Ji, J. Wang, and Z. Yan, "A stock price prediction method based on deep learning technology," *International Journal of Crowd Science*, vol. 5, no. 1, pp. 55–72, 2021.

[15] M. M. Akhtar, A. S. Zamani, S. Khan, A. S. A. Shatat, S. Dilshad, and F. Samdani, "Stock market prediction based on statistical data using machine learning algorithms," *Journal of King Saud University-Science*, vol. 34, no. 4, p. 101940, 2022.

[16] U. R. Saxena, P. Sharma, and G. Gupta, "Comprehensive study of machine learning algorithms for stock market prediction during covid-19," *Journal of Computers, Mechanical and Management*, vol. 2, no. 1, pp. 01–07, 2023.

[17] S. Agrawal, D. Thakkar, D. Soni, K. Bhimani, and C. Patel, "Stock market prediction using machine learning techniques," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 5, no. 2, pp. 1099–1103, 2019.

[18] X. Zhong and D. Enke, "Predicting the daily return direction of the stock market using hybrid machine learning algorithms," *Financial innovation*, vol. 5, no. 1, pp. 1–20, 2019.

[19] D. I. Ajiga, R. A. Adeleye, T. S. Tubokirifuruar, B. G. Bello, N. L. Ndubuisi, O. F. Asuzu, and O. R. Owolabi, "Machine learning for stock market forecasting: a review of models and accuracy," *Finance & Accounting Research Journal*, vol. 6, no. 2, pp. 112–124, 2024.

[20] G. Sonkavde, D. S. Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, and S. K. Bhat, "Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications," *International Journal of Financial Studies*, vol. 11, no. 3, p. 94, 2023.

[21] S. L. Ho and M. Xie, "The use of arima models for reliability forecasting and analysis," *Computers & industrial engineering*, vol. 35, no. 1-2, pp. 213–216, 1998.

[22] S. H. Shih and C. P. Tsokos, "A weighted moving average process for forecasting," *Journal of Modern Applied Statistical Methods*, vol. 7, pp. 187–197, 2008.