# Stock Prediction with Machine Learning

Chethan Kacham
*Computer Science*
*University of Central Missouri*
Lees Summit, Missouri
cxk34890@ucmo.edu

Sarath Chandra Nekkalapu
*Computer Science*
*University of Central Missouri*
Lees Summit, Missouri
sxn03570@ucmo.edu

Durga Sai Teja Thota
*Computer Science*
*University of Central Missouri*
Lees Summit, Missouri
dxt10290@ucmo.edu

***Abstract*— On the stock market, where rewards and dangers vary drastically, accurate stock price forecasts play an increasingly important role, and both financial institutions and regulatory agencies have given it enough consideration. Investors have historically embraced stocks as a technique of asset allocation due to their high returns. Research on stock price forecasting has never ceased. Initially, many economists attempted to forecast stock values. Later, as a result of the in-depth study of mathematical theory and the rapid development of computer technology, it was discovered that the establishment of mathematical models, such as the time series model, can be quite effective, as its model is relatively simple and its forecasting effect is superior. In a certain period, the time series model is used. The scope grew steadily. Due to the non-linear nature of stock data, however, many machine learning techniques, such as support vector machines, are ineffective. The time series for stock prices are non-stationary and non-linear, making it exceedingly difficult to anticipate future price movements.Targets for stock market forecasts may include the future stock price, price volatility, or market trend. In stock market prediction systems, there are two sorts of predictions: dummy predictions and real-time predictions. To study the long-term dependency of stock prices, deep learning techniques, such as the ML approach, are utilized to get longer data dependence and overall stock price change patterns. This thesis employs 5000 data from the S&P500 index for empirical research and introduces benchmark models like ARIMA, GARCH, and other research methodologies for comparison in order to demonstrate the efficacy and benefits of machine learning techniques.The emergence of machine learning approaches for prediction systems in financial markets is attributable to improvements in computing. In this study, we anticipate the stock market using a Machine Learning approach, namely Support Vector Machine (SVM), using the Python programming language.**

**Keywords:- Stock, prediction, prices, machine learning, ARIMA, efficacy**

## I. INTRODUCTION

Essentially, quantitative traders with a great deal of money from the stock market purchase stocks, futures, and shares at a low price and then sell them for a profit. The trend in a stock market forecast is not a new topic, however, several organizations continue to debate it. Before investing in a company, investors do two kinds of stock research. The first is fundamental analysis, in which they evaluate the inherent worth of stocks and the performance of the industry, economy, and political environment, among other factors. On the other hand, technical analysis is the examination of market activity information, such as previous prices and volumes, to predict the future performance of a stock.

In recent years, the rising prevalence of machine learning in a variety of sectors has inspired many traders to apply machine learning methods to the profession, with some achieving extremely promising outcomes. This research will construct a financial data predictor software in which a dataset including all past stock prices will serve as the program's training sets. The primary objective of the forecast is to decrease the risk associated with investment decision-making. The Stock Market follows the random walk, which means that the best forecast for tomorrow's value is today's. Forecasting stock indices is unquestionably challenging due to market volatility, which requires an accurate prediction model. The volatility of stock market indexes has an influence on investor confidence. Due to the underlying structure of the financial sphere and the combination of known criteria (Previous day's closing price, P/E ratio, etc.) and unknown elements, stock prices are believed to be highly volatile and vulnerable to rapid fluctuations (like Election Results, Rumors, etc.).

The primary objective of the financial world is to forecast stock prices to guarantee the highest possible gain. Predictive speed is also a significant variable in stock price forecast. Sometimes even a mini second lag can lead algorithmic trading to invalidate the market trading strategy. Therefore, the substantial and challenging problem in the financial market is to develop a stock market prediction model that can deliver excellent efficiency on accuracy and speed. Data scientists have shown that business news can assist stock market prediction will provide more precision. Many studies have been done in the field of stock prediction, where a common strategy is to use Twitter texts to estimate stock price. Some of the studies conducted in this area focus on mixing stock prices with news headlines

Unlike cost information, data from business newspapers is unstructured and high-dimensional. The choice of features and weighting techniques are the essential components of the information handling of news on the market. Traditional methods of selected characteristics, such as Chi-Square and gaining information, disregard the word function frequency. Chi-square, therefore, tends to choose unusual conditions that are often less reliable. Also, the characteristics are unbalanced but essential to forecast for financial newspapers.

In this age of big data, large volumes are produced or gathered at high speed from a wide variety of valuable information of varying veracity. The stock market is a wealthy source of this significant information. People have been attempting to "beat" it for financial gain since the start of the stock market. A stock market is an exchange in which individuals trade company shares, also known as stocks. The exchange aims to facilitate transactions between buyers and sellers. The common objective of someone taking part in the stock market is to create profit by purchasing and selling stocks. The primary way individuals do this is by purchasing a stock, waiting from seconds to days to months to years, and then hopefully selling for more than they've purchased it for. Here comes the famous phrase "purchase low, sell high".

The stock price is the price at which the last trade took place. If one person sells stock to another person, it becomes the new stock price that they agree. The stock price, therefore, depends entirely on supply and demand. The higher a stock's demand, the higher its price. If more individuals try to sell the stock rather than purchase, the price will fall. Because of this, predicting the market may not be simple because it is based entirely on human choice, and these choices may not be reasonable at times.

People create a choice based on many factors, including earnings, reports, news, financial factors, competition, technical analysis, and even gut feeling. It is, therefore, almost impossible to design an algorithm to take all these variables into consideration.

Numerous efforts have been made to forecast stock prices using Machine Learning. Each study project's objective differs significantly in three ways. (1) The intended price change might be short-term (less than one minute), intermediate term (tomorrow to a few days later), or long-term (months later), (2) The collection of stocks may be restricted to less than 10 stocks, stocks in a certain sector, or all stocks in general. (3) The predictors used might vary from global news and economic trends to specific firm features to stock price time series data.

Targets for stock market forecasts may include the future stock price, price volatility, or market trend. In stock market prediction systems, there are two sorts of predictions: dummy predictions and real-time predictions. In Dummy prediction, they construct a set of criteria and calculate the average price to anticipate the future price of shares. Real-time prediction is required using the internet to see the current share price of the firm.

The emergence of machine learning approaches for prediction systems in financial markets is attributable to improvements in computing. In this study, we anticipate the stock market using a Machine Learning approach, namely Support Vector Machine (SVM), using the Python programming language.

## II. MOTIVATION

Predictions on stock markets have been the object of research for many decades. The number of variables and sources of information that are to be considered is immense.

This makes the predictions on the stock market a hard one. Financial news is a primary factor that investors must consider during the process of financial decision-making. The core of stock market prediction is to forecast the opening price of the next day. Research scheme based on technical indicators analysis assumes that the behavior of stock has the property of predictability based on its performance in the past and all the effective factors are reflected by the stock price. By analyzing technical indicators of the previous stock price, one could obtain important information which could be explored to forecast the following stock price. With the current advancement in computers, sophisticated forecasting methods can be implemented with ease. Predicting the stock market is a complicated task. A particular stock has hundreds of events and preconditions to move it in a specific direction. To obtain the most reliable result, we need to capture as many of these preconditions as we can. Investors make educated guesses by analyzing the data.

## III. MAIN CONTRIBUTIONS & OBJECTIVES

- The first and most significant stage of the research project is the business understanding, this stage is aimed at understanding the goals of the project from a business viewpoint and turning this business view into a definition of study problems and then creating a plan for achieving these goals.
- The aim of this research is to find the impact of news headlines in predicting the stock market.
- To forecast the market strategy, historical stock data as well as financial news is used.
- So, this model was used to find whether the financial news had any influence on the stock market.
- To maintain portability, one cannot merely rely on human labor in understanding the stock market.

## IV. RELATED WORK

Investors continually rebalance their investment assets to reduce decision-making risks, so they must accurately assess stock and other financial asset prices. Due to the many factors that can affect stock prices, such as the company's asset allocation or operating conditions, economic and political policies in related industries, emergencies, and currency exchange rates, it is difficult to predict when and how to allocate asset budgets at that time (Jiang, F., Lee, J. A., Martin, X., & Zhou, G. (2019)). Thus, many investors use technology and quantitative methods to predict asset prices. These tactics include building a reasonable model using market data and predicting the best time to invest. Many academics use ARIMA and GARCH time series models for forecasting, although their assumptions are high. Stable and linear sequences are required. However, several variables affect stock data prices, making them neither steady nor linear. Several methods can smooth the sequence, however the difference operation loses data, limiting the typical time series model's predicting ability. Because support vector machines can interpret nonlinear data, more academics are using machine learning models for prediction as computer science and artificial intelligence advance. The stock market reflects the economy's success and informs future economic strategies. Stock investing theories include random walk, current portfolio, efficient market hypothesis, behavioural finance, and evolutionary securities. China's ascent disproves efficient market theory. Thus, researchers used statistical models like differentiated integrated moving average autoregression (ARIMA), generalised autoregressive conditional heteroskedasticity (GARCH), and others to better predict stock prices. Other researchers have developed other statistical models for stock price prediction. Due of its extensive use, experts have likened the conventional machine learning model to statistical stock price predictions. Due to its efficient nonlinear learning and absence of assumptions, classical machine learning beats statistical models in stock price prediction. Traditional machine learning algorithms improved stock price forecasts out-of-sample. Computer information technology is helping electronic trading handle enormous amounts of high-frequency transaction data.

## V. Data Description

Obviously, stock market prediction technology has substantial economic value for stock market investors and investment institutions, as it assists individuals and institutions in generating profits and avoiding investment dangers. However, the value of stock market forecasting technologies is far greater. From a social viewpoint, stock market forecasting technology may avert systemic hazards in the financial market, aid in the appropriate use of societal money, and contribute to economic growth and stability. Stock data has unique properties, and the current forecasting technology approaches are underutilized, therefore its study presents new difficulties to the technology. In particular, the multi-scale and multi-source heterogeneous prediction technology is applicable not only to stock market forecasting, but also to a variety of other disciplines, including personal health status forecasting, energy demand forecasting, and website traffic forecasting.

This study has both significant socio-economic benefits and significant academic research value. The multi-scale feature of stock market data refers to the occurrence of data at distinct time intervals, with the data at different scales reflecting the stock movement status throughout different time intervals. Large-scale stock market data may represent the stock market's long-term movement state, while small-scale stock market data can reflect the stock market's short-term movement state. Different scales of data contain connected information as well as their own information. Multiple scales of stock market data must be considered exhaustively to provide a more accurate description of the present market situation. However, most of the available research focuses only on single-scale stock market data. Inaccurate descriptions of the situation of the stock market might result in anticipated performance that falls short of expectations. The ability to efficiently use multi-scale data is essential for correctly describing and predicting market conditions.

### A. *Datasets*

This article selects the S&P500 index through yahoo finance, and the transaction data for each trading day from September 26th, 2001 to September 24th, 2021. The data includes 5000 observations. The selected data are divided into two parts. The first part occupied 70% of the selected data to train the model, and the remaining observations are considered to test and validation.

| Date | Open | High | Low | Close | Adj Close | Volume | company_name |
|---|---|---|---|---|---|---|---|
| 2022-11-21 | 93.970001 | 95.019997 | 90.589996 | 92.459999 | 92.459999 | 84330300 | AMAZON |
| 2022-11-22 | 92.620003 | 93.349998 | 90.870003 | 93.199997 | 93.199997 | 62192000 | AMAZON |
| 2022-11-23 | 93.239998 | 94.580002 | 92.830002 | 94.129997 | 94.129997 | 59414700 | AMAZON |
| 2022-11-25 | 93.790001 | 94.430000 | 93.070000 | 93.410004 | 93.410004 | 35088600 | AMAZON |
| 2022-11-28 | 93.930000 | 96.400002 | 93.430000 | 93.949997 | 93.949997 | 74943100 | AMAZON |
| 2022-11-29 | 94.040001 | 94.410004 | 91.440002 | 92.419998 | 92.419998 | 65567300 | AMAZON |
| 2022-11-30 | 92.470001 | 96.540001 | 91.529999 | 96.540001 | 96.540001 | 102628200 | AMAZON |
| 2022-12-01 | 96.989998 | 97.230003 | 94.919998 | 95.500000 | 95.500000 | 68488000 | AMAZON |
| 2022-12-02 | 94.480003 | 95.360001 | 93.779999 | 94.129997 | 94.129997 | 72427000 | AMAZON |
| 2022-12-05 | 93.050003 | 94.059998 | 90.820000 | 91.010002 | 91.010002 | 67499963 | AMAZON |

Fig 1: Dataset Description Analysis

### B. *Data Visualisation*

The volatility of stock prices is determined by the stock's movement but is also affected by a variety of other variables. Due to the relative consistency and predictability of the underlying value of companies, the variables that most influence the stock market price is as follows: 1. macro variables; 2. regional and industrial factors; 3. company-specific factors; 4. market-specific aspects.

This article predicts the closing index of the S&P500 rather than particular company stock price projections, therefore in addition to the more granular industry and business considerations, it focuses mostly on the impact of macroeconomic and market factors. Macroeconomic factors relate to the influence of the macroeconomic environment and its variations on stock values, including routine elements like cyclical swings in macroeconomic operations and policy factors like the government's monetary policy. This article forecasts daily data for the S&P500 closing index, concentrating primarily on the influence of monetary policy and other policy considerations on stock prices.

*The trend of Adj close price*

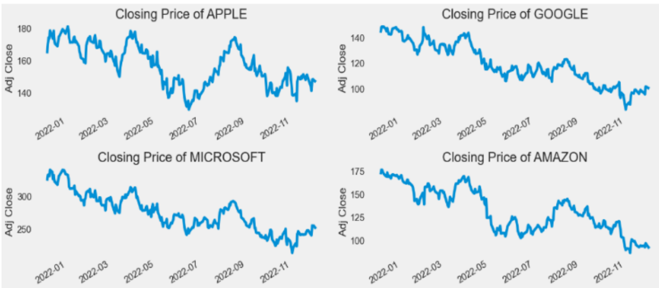


Fig 2: The trend of Adj close price

There are two approaches for predicting the price of a stock: qualitative analysis and quantitative analysis. The qualitative analysis approach is the core analysis method, which relies on the subjective experience of financial practitioners. This thesis is a numerical forecast of the daily closing index of the S&P500 rather than a trend assessment of price movements, therefore the literature study of quantitative analytic techniques is the primary emphasis. Utilizing numerical data on a certain time scale in the stock market, such as sky-high index prices and stock price volume data, numerical data-based stock market forecasting research predicts individual stocks or other investments in the same time frame. Predict the price of the underlying in the future. These studies may be classified into research on the features of numerical data stock market forecasting and research on the numerical data stock market forecasting model, depending on the study.


Fig 3: Trend of Volume


Fig 4: Trend of Moving average

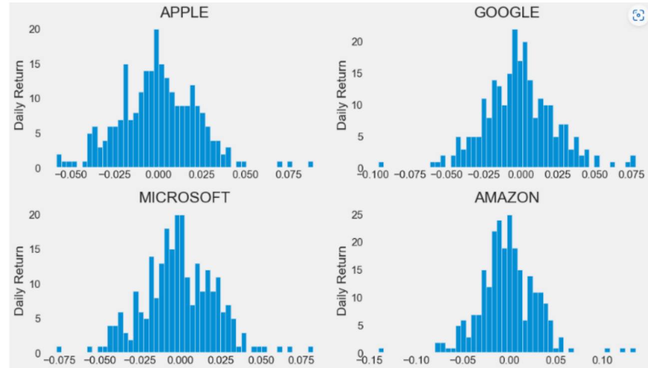


Fig 5: Histogram of Daily Return

## VI. Proposed Framework

### *C. Modeling*

In this project, the prediction of the stock market is done by the Support Vector Machine (SVM), Bayesian Ridge Regression, Linear Regression, XGBoost Regressor, and Kernel Ridge regression.

#### ***I. Linear Regression***

Using linear regression, we may determine the connection between two or more variables of interest and generate predictions based on this information. There are just two variables in basic linear regression: a dependent variable and an independent variable. Simple linear regression yields the best-fitting line or the regression line. This regression line may be expressed using the formula:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

The objective of linear regression is to forecast trends and future values. It can, but not with pinpoint precision, answer queries such as "What may the price of Infosys be in the next three months?", "What may the price of gold be during the next six months?" and "Or Where may the market be headed if the current trend continues?". This is the extent of future stock prices and financial markets. It is also far more beneficial and widely used outside of the financial markets. In this essay, however, we shall investigate its predictive value for the stock market.

#### ***II. XGBoost Regressor***

XGBoost enhances the Boosting algorithm based on the GBDT model, and its internal decision tree employs a regression tree. The underlying concept of the boosting method is to combine numerous weak classifiers into one robust classifier. As a decision tree promotion model, XGBoost combines many tree models to create a robust classifier. Each time a new tree is introduced, the residuals from earlier training are re-fitted.

The sample's final projected value is calculated by summing the scores corresponding to many trees gained via training. A regular term in the target function is one of the primary benefits of the XGBoost model, which prevents overfitting. The feature granularity is optimized in parallel, which is more efficient. Supporting column sampling may not only decrease overfitting but also save computation time. Given the sparse nature of the training data, the default orientation of the branch may be defined for certain values.

### III. Kernel Ridge Regression

Kernel ridge regression is a function approximation approach used in a regression setting. However, it is most often seen as part of a classification system using support vector machines. This nonlinear regression uses two essential machine-learning techniques. First, a ridge regression penalty enables the model to be fitted in a potentially high-dimensional nonlinear space while keeping overfitting issues in check. The estimating and forecasting issue are computationally tractable due to a technique known as the "kernel trick." First, the problem of overfitting is addressed. Commence with a conventional linear regression

$$y = X\beta + e$$

where y is a Tx1 vector, X is a TxN matrix, and β is an Nx1 vector of parameters. T represents the number of observations, and N is the number of predictor variables. Standard linear regression objectives minimize,

$$\sum_{t=1}^{T}(y - X\hat{\beta})^2,$$

with β̂ representing a traditional OLS estimate of the parameters. For N large relative to T, it is highly likely this regression will yield poor out-of-sample predictions, as it will overfit the training samples. Ridge regression can ameliorate this issue by adding the ridge penalty as in,

### IV. Bayesian Ridge Regression

The Bayesian vs Frequentist discussion is one of those intellectual disputes that I would rather observe than participate in. Rather than passionately embracing one approach of statistical reasoning, I believe it is more beneficial to study both and employ them when necessary. From a Bayesian perspective, linear regression is formulated using probability distributions rather than point estimates. The answer, y, is supposed to be selected from a probability distribution rather than a single number. The model for Bayesian Linear Regression using a normal distribution response sample is:

$$y \sim N(\beta^T X, \sigma^2 I)$$

The output, y, is the result of a normal (Gaussian) Distribution with a mean and variance. Transposing the weight matrix and multiplying it by the predictor matrix yields the linear regression mean. The conflict is equal to the squared standard deviation (multiplied by the Identity matrix because this is a multi-dimensional formulation of the model). The objective of Bayesian Linear Regression is not to estimate the one "best" value of the model parameters, but rather the posterior distribution of the model parameters. Not only is the response derived from a probability distribution, but it is also expected that the model parameters are drawn from a distribution. The posterior probability of the model parameters is dependent on the inputs and outputs of training:

### V. Support Vector Machine

Support Vector Machines have supervised learning models with related learning algorithms that examine classification and regression data in machine learning. In Support Vector Regression, the needed straight line to match the data is known as the hyperplane. The goal of a support vector machine technique is to locate a hyperplane in an n-dimensional space that classifies the data points in a separate manner. Support Vectors refer to the data points on each side of the hyperplane that is closest to the hyperplane. These variables affect the location and direction of the hyperplane and contribute to the formation of the SVM.

Support Vector Regression is a technique for supervised learning that predicts discrete values. Support Vector Regression employs the same underlying concept as SVMs. The objective of SVR is to identify the optimal fit line. The best-fitting line in SVR is the hyperplane with the greatest number of points. The SVR, unlike other Regression models, seeks to fit the best line within a threshold value, as opposed to minimizing the difference between the actual and predicted values. The value of the threshold is the distance between the hyperplane and the boundary line. SVR has a fit time complexity that is more than quadratic with the number of samples, making it difficult to scale to datasets with more than a few thousand samples.

The Linear SVR or SGD Regressor is used for big datasets. Linear SVR is implemented more quickly than SVR, but only examines the linear kernel. Support Vector Regression produces a model that relies only on a portion of the training data since the cost function disregards samples whose prediction is near the goal.

### *D. Estimator parameters*

The performance assessment of the stock market forecasting model often employs classification evaluation indicators, such as the accuracy rate and F1 value, and the profitability of the stock market forecasting model is assessed by several algorithmic trading simulations. There may be inconsistency between the two aforementioned assessment approaches, in that the model with the greatest categorization evaluation performance may not necessarily have the highest profitability. This discrepancy might lead to the development of stock market forecasting models devoid of useful recommendations. How to eliminate this inconsistency and increase model assessment validity is a challenging aspect of stock market research.

*Root Mean Square Error (RMSE)*

The discrepancy between the observed value and the real value may be calculated using Root Mean Square Error (RMSE). Because the average index is not robust, the average error is very susceptible to outliers. The phrase is as follows

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(observed_i - predicted_i)^2}$$

## VII. Results

*Linear Regression*

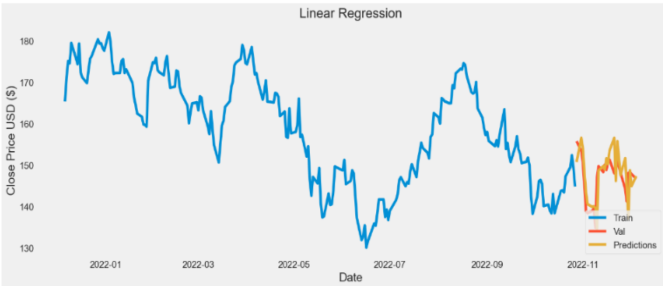


Fig 6a: Predicted Trend of Linear Regression

| Date | Open | High | Low | Close | Adj Close | Volume | company_name | Predictions |
|---|---|---|---|---|---|---|---|---|
| 2022-10-28 | 148.199997 | 157.500000 | 147.820007 | 155.740005 | 155.482086 | 164762400 | APPLE | 150.706444 |
| 2022-10-31 | 153.160004 | 154.240005 | 151.919998 | 153.339996 | 153.086044 | 97943200 | APPLE | 156.552922 |
| 2022-11-01 | 155.080002 | 155.449997 | 149.130005 | 150.649994 | 150.400497 | 80379300 | APPLE | 154.004825 |
| 2022-11-02 | 148.949997 | 152.169998 | 145.000000 | 145.029999 | 144.789810 | 93604600 | APPLE | 148.256365 |
| 2022-11-03 | 142.059998 | 142.800003 | 138.750000 | 138.880005 | 138.650009 | 97918500 | APPLE | 143.304902 |

Fig 6b: Predicted Values of Linear Regression

**RMSE Linear Regression: 4.712775158562598**

*XGBoost Regressor*

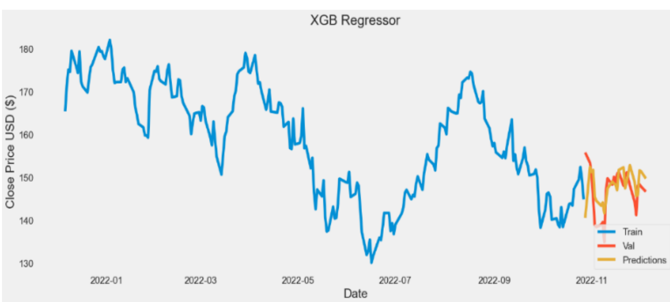


Fig 7a: Predicted Trend of XGBoost Regression

| Date | Open | High | Low | Close | Adj Close | Volume | company_name | Predictions |
|---|---|---|---|---|---|---|---|---|
| 2022-10-28 | 148.199997 | 157.500000 | 147.820007 | 155.740005 | 155.482086 | 164762400 | APPLE | 140.492554 |
| 2022-10-31 | 153.160004 | 154.240005 | 151.919998 | 153.339996 | 153.086044 | 97943200 | APPLE | 152.374542 |
| 2022-11-01 | 155.080002 | 155.449997 | 149.130005 | 150.649994 | 150.400497 | 80379300 | APPLE | 151.244827 |
| 2022-11-02 | 148.949997 | 152.169998 | 145.000000 | 145.029999 | 144.789810 | 93604600 | APPLE | 151.665680 |
| 2022-11-03 | 142.059998 | 142.800003 | 138.750000 | 138.880005 | 138.650009 | 97918500 | APPLE | 145.358292 |

Fig 7b: Predicted Values of XGBoost Regression

**RMSE XGB Regressor: 4.8147078627018205**

*Kernel Ridge Regression*

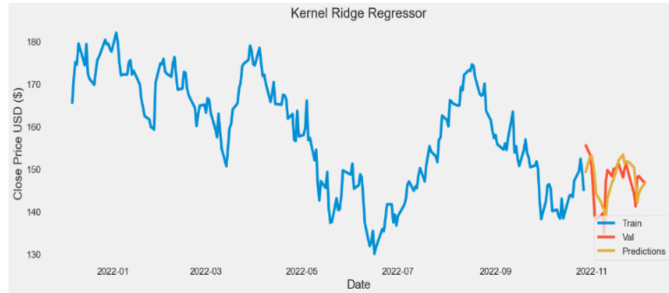


Fig 8a: Predicted Trend of Kernel Ridge Regression

| Date | Open | High | Low | Close | Adj Close | Volume | company_name | Predictions |
|---|---|---|---|---|---|---|---|---|
| 2022-10-28 | 148.199997 | 157.500000 | 147.820007 | 155.740005 | 155.482086 | 164762400 | APPLE | 149.016186 |
| 2022-10-31 | 153.160004 | 154.240005 | 151.919998 | 153.339996 | 153.086044 | 97943200 | APPLE | 152.813798 |
| 2022-11-01 | 155.080002 | 155.449997 | 149.130005 | 150.649994 | 150.400497 | 80379300 | APPLE | 153.069972 |
| 2022-11-02 | 148.949997 | 152.169998 | 145.000000 | 145.029999 | 144.789810 | 93604600 | APPLE | 151.084249 |
| 2022-11-03 | 142.059998 | 142.800003 | 138.750000 | 138.880005 | 138.650009 | 97918500 | APPLE | 147.822211 |

Fig 8b: Predicted Values of *Kernel Ridge* Regression

**RMSE Kernel Ridge Regressor: 4.564621871174483**

*Bayesian Ridge Regression*

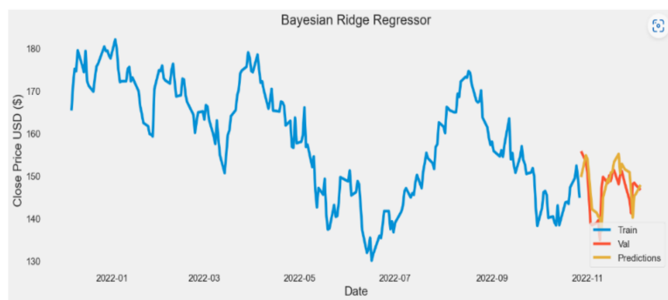


Fig 9a: Predicted Trend of *Bayesian Ridge* Regression

| Date | Open | High | Low | Close | Adj Close | Volume | company_name | Predictions |
|---|---|---|---|---|---|---|---|---|
| 2022-10-28 | 148.199997 | 157.500000 | 147.820007 | 155.740005 | 155.482086 | 164762400 | APPLE | 149.671418 |
| 2022-10-31 | 153.160004 | 154.240005 | 151.919998 | 153.339996 | 153.086044 | 97943200 | APPLE | 154.787220 |
| 2022-11-01 | 155.080002 | 155.449997 | 149.130005 | 150.649994 | 150.400497 | 80379300 | APPLE | 153.964097 |
| 2022-11-02 | 148.949997 | 152.169998 | 145.000000 | 145.029999 | 144.789810 | 93604600 | APPLE | 150.146707 |
| 2022-11-03 | 142.059998 | 142.800003 | 138.750000 | 138.880005 | 138.650009 | 97918500 | APPLE | 145.887917 |

Fig 9b: Predicted Values of *Bayesian Ridge* Regression

**RMSE Bayesian Ridge Regressor: 4.443683880642531**

*Support Vector Machine Regression*

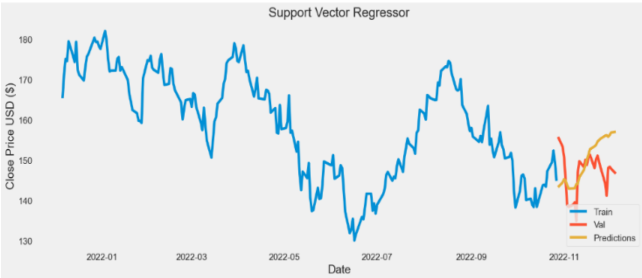


Fig 10a: Predicted Trend of *SVM* Regression

| Date | Open | High | Low | Close | Adj Close | Volume | company_name | Predictions |
|---|---|---|---|---|---|---|---|---|
| **2022-10-28** | 148.199997 | 157.500000 | 147.820007 | 155.740005 | 155.482086 | 164762400 | APPLE | 143.278562 |
| **2022-10-31** | 153.160004 | 154.240005 | 151.919998 | 153.339996 | 153.086044 | 97943200 | APPLE | 144.433225 |
| **2022-11-01** | 155.080002 | 155.449997 | 149.130005 | 150.649994 | 150.400497 | 80379300 | APPLE | 145.260335 |
| **2022-11-02** | 148.949997 | 152.169998 | 145.000000 | 145.029999 | 144.789810 | 93604600 | APPLE | 144.870259 |
| **2022-11-03** | 142.059998 | 142.800003 | 138.750000 | 138.880005 | 138.650009 | 97918500 | APPLE | 143.767651 |

Fig 10b: Predicted Values of *SVM* Regression

**RMSE SVM Regressor: 6.87999687308394585**

## VIII. Conclusion and Future Work

Due to the significance of the stock market to a nation's economy, stock price forecasting techniques will continue to evolve and be taken from the advancement of other fields. To bring the model closer to reality, broaden the method's application, and improve the method's forecasting accuracy, it is required to regularly investigate and analyze the features of the stock market throughout the method's development. Due to the influence of economic, political, and environmental variables on stock market data, the law of its change is tricky, and its cycle is difficult to predict. To get the intended findings, the model still requires a substantial amount of historical data and the selection of acceptable variables for investigation. When evaluating complicated stock markets using the classic ARIMA model, its prediction results are not optimal, and there are still some inaccuracies in price forecasting.

In the project, we suggested using data gathered from several global financial markets in conjunction with machine learning algorithms to forecast the movements of stock indexes. The Bayesian Ridge Regression and Linear Regression algorithms operate on a massive dataset including values gathered from several worldwide financial marketplaces.

Furthermore, overfitting is not an issue with these algorithms. Several machine-learning-based algorithms are presented for forecasting the daily movement of Market stocks. The data suggest that the efficiency is high. Built upon our well-trained prediction are the actual trading models. Compared to the chosen benchmarks, the model yields a greater profit with a minimum RMSE of 3.32.

## IX. References


[1] Chen, J., Jiang, F., & Tong, G. (2017). Economic policy uncertainty in China and stock market expected returns. Accounting and Finance, 57, 1265–1286.

[2] Dai, Z., Zhou, H., Wen, F., & He, S. (2020a). Efficient predictability of stock return volatility: The role of stock market implied volatility. The North American Journal of Economics and Finance, 52, 101174.

[3] Dai, Z., & Zhu, H. (2020). Stock returns predictability from a mixed model perspective. Pacific-Basin Finance Journal, 60, 101267.

[4] Dai, Z. F., Dong, X. D., Kang, J., & Hong, L. (2020b). Forecasting stock market returns New Technical indicators and two-step economic constraint method. The North American Journal of Economics and Finance, 53, 101216.

[5] Jiang, F., Lee, J. A., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. Journal of Financial Economics, 132, 126–149.

[6] Shah, D., Isah, H. and Zulkernine, F., 2019. Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies, 7(2), p.26.

[7] Bustos, O. and Pomares-Quimbaya, A., 2020. Stock market movement forecast: A Systematic Review. Expert Systems with Applications, 156, p.113464.

[8] Jose, J., Mana, S. and Samhitha, B.K., 2019. An efficient system to predict and analyze stock data using Hadoop techniques. International Journal of Recent Technology and Engineering (IJRTE), 8(2), pp.2277-3878.

[9] Hu, Z., Zhao, Y. and Khushi, M., 2021. A survey of forex and stock price prediction using deep learning. Applied System Innovation, 4(1), p.9.

[10] Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W. and Wills, G., 2020. A survey on machine learning for stock price prediction: algorithms and techniques.

[11] Hussain, W., Merigó, J. M., & Raza, M. R. (2022). Predictive intelligence using ANFIS‐induced OWAWA for complex stock market prediction. International Journal of Intelligent Systems, 37(8), 4586-4611.

[12] Saha, S., Gao, J., & Gerlach, R. (2022). A survey of the application of graph-based approaches in stock market analysis and prediction. International Journal of Data Science and Analytics, 1-15.

[13] Sharma, K., & Bhalla, R. (2022). Stock Market Prediction Techniques: A Review Paper. In Second International Conference on Sustainable Technologies for Computational Intelligence (pp. 175-188). Springer, Singapore.

[14] Akhtar, M. M., Zamani, A. S., Khan, S., Shatat, A. S. A., Dilshad, S., & Samdani, F. (2022). Stock market prediction based on statistical data using machine learning algorithms. Journal of King Saud University-Science, 34(4), 101940.

[15] Lu, X., Ma, F., Wang, J., & Dong, D. (2022). Singlehanded or joint race? Stock market volatility prediction. International Review of Economics & Finance, 80, 734-754.

[16] Mane, O. (2022). Stock Market Prediction using Natural Language Processing--A Survey. arXiv preprint arXiv:2208.13564.

[17] Yadav, A. and Vishwakarma, D.K., 2020. Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review, 53(6), pp.4335-4385.

[18] Sulandari, W., Suhartono, Subanar and Rodrigues, P.C., 2021. Exponential Smoothing on

[19] Modeling and Forecasting Multiple Seasonal Time Series: An Overview. Fluctuation and Noise Letters, p.2130003.

[20] Kumar, I., Dogra, K., Utreja, C. and Yadav, P., 2018, April. 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) (pp. 1003- 1007). IEEE.

[21] Ingle, V. and Deshmukh, S., 2016, August. Hidden Markov model implementation for prediction of stock prices with TF-IDF features. In Proceedings of the International Conference on Advances in Information Communication Technology & Computing (pp. 1-6).

[22] Singh, Sukhman, Tarun Kumar Madan, J. Kumar and A. Singh. "Stock Market Forecasting using Machine Learning: Today and Tomorrow." 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) 1 (2019): 738-745.