Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Design and Comparative Analysis of Intelligent Models for Stock Market Forecasting and Management Using Machine Learning Technique

Neha Dewangan[1], Dr. Mayank Singh Parihar[2], Bisahu Ram Sahu[2]

[1]Ph.D. Research Scholar, Department of Computer Science, Dr. C.V. Raman University, Bilaspur, Chhattisgarh, India
[2]Assistant Professor, Department of Computer Science, Dr. C.V. Raman University, Bilaspur, Chhattisgarh, India
Email: nehateekam@gmail.com

At present, 70-80% of the world population is investing in the stock market. This depends on the demand of the investors. There are two types of investors: first, investors who invest less money, and second, those who invest more money in the market. The main objective of the investor is to get more profit with less investment. Currently, the Angel One app, Stable app, Grow the app, and other software have been created to increase investment in the market. This proves how important the stock market is at present. The stock market is not stable; it always fluctuates, and the stock value keeps changing. All the time, it isn't easy to forecast the stock market. Here, we use different types of ML techniques for the best results. In this research paper, we have compared three major machine learning methods: one is SVM, the other is linear regression, and the third is random forest, which falls under supervised machine learning techniques. By comparing all three methods, an attempt has been made to determine which method provides the best forecast value, and the dataset is fetched from Yahoo Finance. The dataset's starting date is January 1, 2015, and the ending date is January 1, 2024. Various performance metrics have been used, such as MAE, MSE, accuracy, R^2 score, etc. After implementing all the methods, we found that linear regression gives the best accuracy of 99.90%.

Keywords: Stock Market, Stock Market Forecasting, Ensemble Methods, Yahoo Finance, Random Forest, Linear Regression, Support Vector Machine, Trading Strategies.

1. Introduction

The stock market is a public market in which companies list their stocks and trade their stocks at an agreed price to raise financial resources. In turn, shareholders receive benefits either in the form of annual profits or bonuses. At the same time, equity holders can also trade shares at an agreed price. If they want to earn money from the difference between the purchase and sale price, they can also trade shares at an agreed price. If an investor wants to trade across the world, there are some world-famous stock markets like the NYSE, NASDAQ, TSE, MSE,

LSE, PSE, SSE, BSE, TSE, etc.

Financial analysts and scholars continue to debate the topic of stock price forecasting. However, in recent years, technological advancements like artificial intelligence and machine learning techniques have made it possible to anticipate stock prices more accurately and effectively. Support vector machines are becoming increasingly common among machine learning algorithms used to forecast stock prices. However, a large portion of research has focused on SVM-based algorithms, ignoring the overfitting of SVM [1] Three indicators are used for the stock market: the first one is the stockbroker, the second one is the stock buyer, and the last one is the stock sale. Support vector machine is a very special type of learning algorithm characterized by capacity control of the decision function, the use of kernel functions, and the sparsity of the solution. The various algorithms used for forecasting can be categorized into linear (AR, MA, ARIMA, ARMA) and nonlinear (ARCH, GARCH) models [2]. Researchers and market analysts are always keen to study and test stock market trends. For this, various statistical techniques such as autoregressive integrated moving average and clustering, etc., are applied. This model provides historical evidence and postulates theory. The research by [3] has conducted extensive research in stock market forecasting. At present, one of the most attractive markets is the financial market. This market has had a significant impact on money-related areas such as employment, business, and GDP. Investors also use key strategies to make decisions regarding the buying and selling of shares in the stock market so that money can be invested with minimum risk and more returns can be obtained.[3] Because stock big data is so widely dispersed, these forecasting techniques, like all others, are unable to provide an accurate prediction of change. The "Efficient market hypothesis" is offered because of the inability to forecast stock prices for many reasons. This theory states that how buyers and sellers respond to the most recent information and the company's future affects the price of securities in the market.[4]

Investors first assess the company's operational approach. By examining historical data, they also attempt to determine which companies may yield profitable stock purchases. Social media and websites that provide financial news can be used for this assessment. However, evaluating such a vast data collection is impossible. For investors to make decisions, an automated decision support system must be developed.

The research explores the application of the SVM network for stock market analysis and forecasting and focuses on methodologies like SVM, linear regression, and random forest techniques.

Researcher [5] study's primary goal is to evaluate pertinent, related studies about the use of machine learning in stock market forecasting. To accomplish this goal, this research employs a methodical examination. The overview of ML in stock market forecasting is the research topic for this systematic investigation. Becoming aware of the developments in machine learning applications for stock market forecasting. Model techniques of ML are commonly employed for casting in the stock market. Here sections discuss the overall procedures that are used in a research paper: section 1 describes the introduction part 2 section shows the related work done by another researcher, section 3 shows some data part, and so on.

2. DISCUSSION ABOUT SOME RELATED WORK

Kuo, R. J. et al. [6] The study describes a hybrid model called MVl-SVM for stock price prediction by integrating multi-view learning with SVM and comparing it with the ARIMA and classic SVM models based on single and multivariate data. In the paper, both market data and the impact of financial media are shown. In order to forecast the Karachi Stock Exchange (KSE) index.

Sharma et al. [7] discussed the different types of approaches and methods available for stock market forecasting, focusing on fundamental and technical analysis for better results. Important factors like the PE ratio, B ratio, PEG ratio, and dividend are evaluated, and the valuation of the stock is also described, including the price-to-book ratio, price-to-earnings ratio, price-to-earnings-growth, and dividend. Singh G. [26] attempts to forecast the stock price using machine learning. Nifty is the listed company in the stock market. In this paper, the researcher has taken the historical data of about 25 years of Nifty 50 from 22 April 1996 to 16 April 2021. A supervised machine-learning technique was implemented. In conclusion, the results of ANN and linear regression obtained equal results; the best result was from stochastic gradient descent.

Bijesh Dhyan et al. [8] have analyzed various analysis techniques used with time series data like ARIMA, model, exponential smoothing, etc. A common technique used for stock price is ARIMA which helps in predicting future returns based on past behavior. In this paper, data from NIFTY 50 has been taken.

Gurjar M et al. [9] shifted the focus of researchers toward non-linear patterns of stock returns. He tried to achieve maximum accuracy by using artificial neural networks, and using the historical data, the ANN model was trained. Stochastic indicators, moving divergence, and RSI are used for some features, and the data set is divided into training and testing sets so that maximum accuracy can be achieved.

Usmani et al.[10] suggested a hybrid machine learning approach based on SVM, Single Layer Perceptron (SLP), Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Autoregressive Integrated Moving Average (ARIMA).

Jiang et al. [11] outlined a novel machine learning-based stock market forecasting method that uses SVM to forecast the next day's stock movement by taking advantage of the temporal correlation between international stock markets and different financial products. To track the real increase in the markets, the same approach is also used with other regression algorithms. Lastly, a basic trading model is created in order to examine how well the suggested prediction algorithm performs in comparison to other standards.

In this paper by Singh et al. [12], all the machine learning techniques have been studied, and the focus was on casting for the stock market using the ANM method; the aim is to provide a review of the applications of ANN in SMF. Introductory material has been provided for SMF to the investors who want to predict the stock price using ANN. How to implement this method of AVN. Its complete procedure has been explained.

3. FEATURE SELECTION

Feature selection is a crucial step that must be completed before implementing machine learning algorithms in data analysis. To lower the dimensionality of the data and exclude characteristics that are irrelevant to the classification job or that can compromise the quality of the information that can be extracted from the data, this feature selection is required. It also saves money and time on computation. The process of finding and acquiring a subset of features from a larger data collection to improve the performance of a classification approach is known as feature selection. All these three feature selection techniques are illustrated by [13] embedding, wrappers, and filters.

1. Filter: The filter methodology, independent of any machine learning algorithm, uses statistical assessment techniques to separate the pertinent characteristics from the data. The main benefit of a filter is its speed in feature selection. However, it is less accurate than other approaches, such as correlation-based feature selection and information-gain mutual information. Using a machine learning algorithm that employs a. search strategy to comb through a space of potential feature subsets,

2. Wrapper: The wrapper method evolutes each subset according to the performance quality of specified techniques.

3. Embedded: The embedded approach, selects features during model training, combining the benefits of filters and wrapper approaches. ML models that may be used for embedded feature selection include random forest, rigid regression, and elastic net. The performance of machine learning can be gently impacted by the feature selection process.

4. FORECASTING METHODS AND DATA TYPES

4.1 TRADITIONAL APPROACHES OF SMF

One of the approaches used for stock market forecasting is the traditional approach which has two parts. The first one is a fundamental analysis and the second is a technical analysis that is illustrated below:

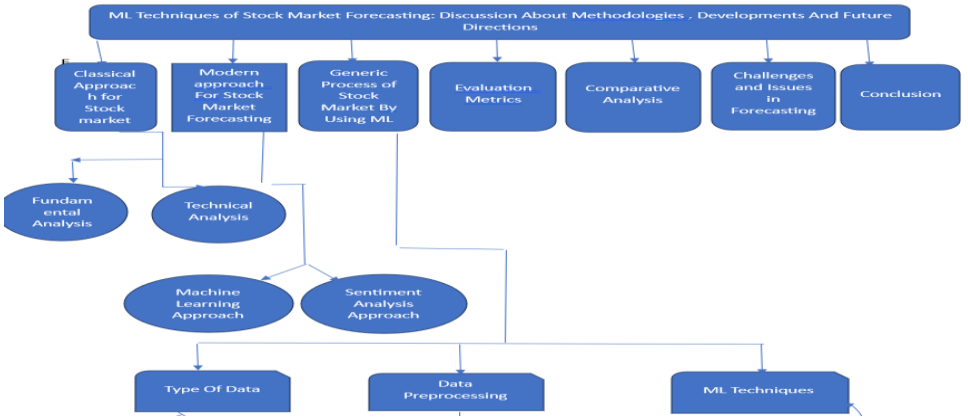

Figure1 : represent all process done in research paper

1. Fundamental Analysis

It is a technique to conceal the stocks' true worth. Comprehending both technical and fundamental analysis. Everything about the company's profile is looked at here, including the analysis of the industry, including market competition, and the assessment of the company's assets, costs, revenue, and future goals. Purchasing shares at a discount and then selling them when the equities reach their real worth, enables investors to profit. These elements aid in determining the company's true average worth from its founding. This is done for long-term analysis, not for short-term estimations [14].

2. Technical Analysis

It is the process of forecasting future market trends by analyzing historical market data, such as volume and cost. Chart patterns and statistical data are the focus of technical analysis. Several mathematical formulas must be used to do the analysis. History repeats itself, according to the research, and the stock market is no exception. To forecast future prices, stock charts, price, volume, and trends are examined. Using the available data to generate forecasts that provide the desired outcome for investors is the goal of TA. Here, we'll talk about some of the key technical indicators that serve as the foundation for future forecasts: volume, the Moving Average Convergence Divergence (MACD), the Simple Moving Average (SMA), the Exponential Moving Average (EMA), and the RSI (Relative Strength Index)[14].

4.2 MODERN APPROACH OF SMF

Like the traditional approach, there is another stock market forecasting approach which is called the modern approach. It includes machine learning methods like SVM, decision tree, linear regression, etc.

The most widely used method for forecasting the future or categorizing data to assist individuals in making the required judgments is learning. To learn from past experiences and evaluate historical data, machine learning algorithms are trained on instances or examples. It is therefore able to recognize patterns to forecast the future as it repeatedly trains over the instances. Here, some of the techniques are explained as follows:

4.2.1 Linear Regression

Linear regression is a technique used to predict the relationship between the dependent and independent variables. The relationship between the two variables is said to be deterministic if one variable can be accurately expressed by the other. A numerical value is predicted by regression. When regression is applied to a dataset, the target values are predetermined. Additionally, by including fresh information, the outcome may be expanded. A pattern may emerge from the relationships that regression creates between predictor and target values. This pattern may be used for different datasets with unknown goal values. As a result, two parts of data are required for regression: one for model definition and the other for model testing. We decide to use linear regression for our analysis in this part. The data is first separated into training and testing sections. After that, we define the model and begin analysis using the training section. Twenty percent of the data is utilized for testing, while eighty percent is used for training [15]. A predictive method for modeling the connection between an independent and dependent variable is regression [16]

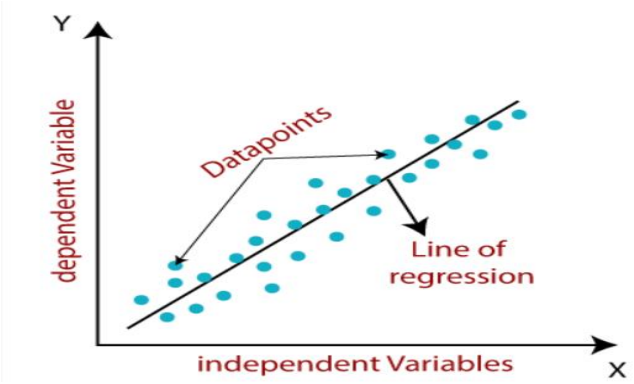

Figure 2: Linear Regression

4.2.2 Support Vector Machine

One of the machine learning algorithms that possesses the desired features, such as the decision function, usage of the kernel method, and also the sparsity of the solution, is known as the Support Vector Machine (SVM) technique.

The Support Vector Regression algorithm, the most widely used supervised type learning algorithm, is designed to solve regression and classification issues. A hyperplane with a maximal margin is found for support vector machine regression, or SVR, so that the greatest amount of data points fall inside those margins. The hyperplane with the greatest number of points in SVR is the best-fit line. Extreme vector points known as support vectors are selected in the SVR method to aid in the creation of a suitable hyperplane. In terms of how it operates, it is comparable to the Support Vector Machine (SVM) algorithm. Working with time series data is one of its uses [17]. SVMs are a set of supervised learning approaches that can be employed for classification and regression problems. The classifier's version is named SVC (support vector classifier). The method's purpose is to find a decision boundary between two classes with vectors. The boundary must be far from any point in the dataset, and support vectors are the sign of observation coordinates with a gap named margin. SVM is a boundary that best separates two classes by employing a line or hyperplane. After studying some research papers, it was found that the size of the training data set should be large enough as per the requirement because SVM is very sensitive to the size of the training data set, due to which the hyperplane is unable to divide the data correctly. Hence, it can be said that while implementing SVM, it is necessary to select the appropriate features. In contrast, the linear regression algorithm does not require a large data set; rather, it gives more preference to high-dimensional features.

The study of research papers shows that when the period of the data set is longer, the accuracy of the forecast increases, whereas when the period of the data set is shorter, the accuracy of the forecast decreases. The reason for this is that the longer the period, the more information we will collect, and our resistance against the noise of forecasting will also be greater.

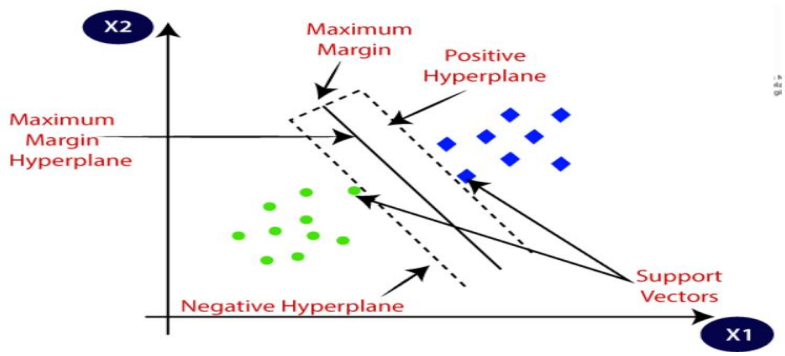

Figure 3: Support Vector Machine

4.2.3 Random Forest

Random forests are used for classification, regression, and creation of decision trees, RFs correct the habit of overfitting decision trees to training sets. The first algorithm for RF, the random subspace method was created by Tin Kam Ho [18]. The stock market is predicted using the random forest method. It provides outstanding accuracy in forecasting since it is regarded as one of the most user-friendly and adaptable machine learning algorithms. Typically, this is applied to categorization jobs. It is quite difficult to make predictions because of the stock market's extreme volatility. We use a random forest classifier, which has the same hyperparameters as a decision tree, to predict the stock market. The decision tool uses a model that resembles a tree. It bases the choice on potential outcomes,

which include factors like utility, resource cost, and event result. In the random forest approach, various observations and characteristics are chosen at random to create many decision trees. The results of these decision trees are then aggregated. The questions on a label or an attribute determine which data is divided into divisions. We used a data set from the financial markets of the previous year that was gathered from a publicly accessible Internet database; 80% of the data was used to train the machine, and the remaining 20% was used for testing. The supervised learning model's fundamental strategy is to identify patterns and correlations in the training set's data and then replicate them for the test set.

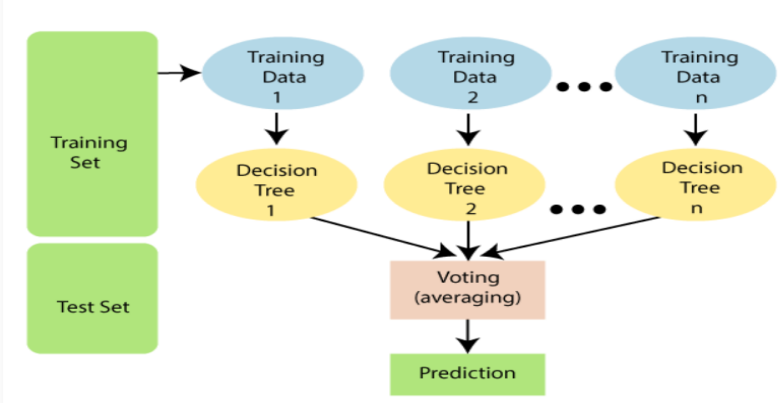

Figure 4: shows the Random Forest

4.3 DATA TYPE

After the discussion of forecasting methods, traditional approach, and modern approach, a discussion of data type is as follows, There are three types of data. First is fundamental data, Second is microeconomic data and the third is raw data [13].

4.3.1 Fundamental data

The fundamental data required for stock trading is provided by historical or fundamental data. The two components of it are the open price and the close price. When the market begins on a trading day, the open price is the initial transaction price; the closing price is the price obtained at the end of an entire day transaction or stock trade. For all past trade data, Yahoo offers daily data [19].

4.3.2 Macroeconomic data

The second group of variables, which are macroeconomic data that significantly affect stock market performance, may provide additional insight into stock price prediction. The US dollar index (USDX), the Consumer Sentiment Index (UMCSENT), the Civilian Unemployment Rate (UNRATE), the Interest Rate (EFFR), and the Cboe Volatility Index (VIX) are the macroeconomic variables that we have chosen. These variables are the representative features that explain the overall status of the economy, according to the model [19].

4.3.3 Raw Data

The first step in creating forecasts is to collect reliable data as the basis. Based on the idea that history repeats itself, the stock market may be impacted by intrinsic past pricing or extrinsic pertinent data sources. The efficient-market theory states that asset prices already take into account all information. The stock market is predicted using a range of extrinsic data sources since many academics disagree with this finding in practice. For instance, market statistics, fundamental data, knowledge graphs, and news are provided, and technical indicators, Wikipedia traffic, Google news counts, and created features are compared.

1. Types of Data: The raw data commonly utilized to predict the stock market is separated into seven groups in this section:

2. Market data: Market data includes all of the trading activity that occurs in a stock market, including open, high, low, and closing prices as well as trade volume. It functions as an input feature (like the previous prices in a look-back window) as well as a prediction target (like the closing price of the next day).

3. Text data: "Text data" is the word used to describe the text that individuals have contributed, for example, through online searches, social media, and news. These data may provide important insights not available in market data, even though they are challenging to get and evaluate as alternative data. A sentiment component (such as positive, neutral, or negative) that may be used for prediction can be obtained by applying sentiment analysis to this text data.

5. WORKFLOW OF RESEARCH

To achieve the goals specified above, the workflow of the research is defined in Figure 1, the generic model of stock forecasting. We will use different financial stock data from specified indices like Apple dataset (APPL Inc.), BSE30, BSE100, etc., collected from the historical data available on the website (www.yahoofinance.com). We will also use other time series data available in different data sources. Then, we will use preprocessing techniques for smoothing data and removing nonlinearity from the financial data set; after that, some intelligent techniques, like different machine learning techniques, will be used for prediction. The performance of these models will be checked by various measures like Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), R^2 score, and accuracy.

Stock market forecasting is the act of making an effort to estimate the future value of a stock or other financial instrument traded on a financial exchange. The computer language used to predict the stock market using machine learning is called Python. In this study, we propose a machine learning system that will be trained on currently available stock data, develop intelligence, and then use the learned information to provide accurate forecasts [20].

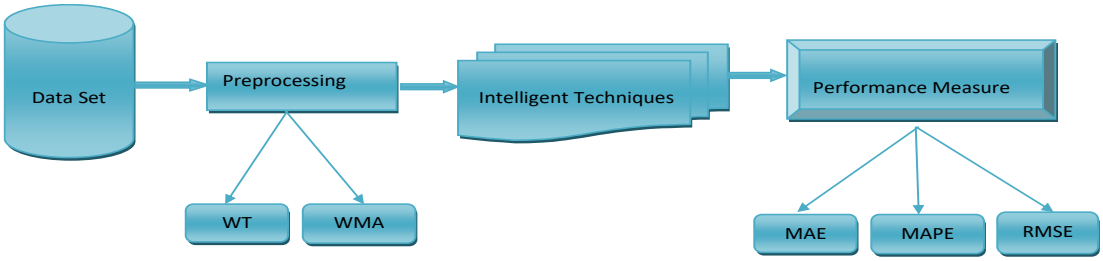

Figure5: Shows the workflow of research.

6. Data collection and preprocessing

First of all, real data is collected from various sources like Yahoo Finance and other websites. Yahoo Finance is based on its price index and its price. It is a group of stock markets and encourages investors to compare the previous value and current value for performance calculation, remove the noise in the collected data, and pre-process the data. Analyst review [21] pre-processed data is useful for stock market forecasting.

Figure 6: Data Preprocessing

Study make use of historical stock price data from Yahoo Finance, which can be accessed using Python's finance package. The information comprises trading volumes and daily adjusted closing prices for the main technology stocks: Apple Inc. (AAPL), from January 1, 2015, to the present day. Yahoo Finance stock information Yahoo Finance provides historical stock price information. Yahoo Finance provides price data for the chosen period for the chosen stock markets in a .csv file format. Date, Open, High, Low, Close, Volume, and Adjusted Close are the seven features of the downloaded data files that, on a given date, display the date of stock trading, the stock's open price, its maximum and lowest trading prices, its closing price, the number of shares traded, and the closing price of a stock when investors receive dividends [22].

1. Import Libraries: Import the necessary libraries like numPy, pandas, Seaborn, Matplotlib, etc. NumPy, pandas, Seaborn, and Matplotlib is standard for data manipulation, visualization, and model building. After collection, all libraries then go to the next step.

2. Imputation of Missing Data: In contrast to other domains, such as sensor data, the issue of missing data is less severe in market data due to its greater reliability and the trading markets' strong support and maintenance of it. The data with a lower sampling frequency, such as market and fundamental data, should be inserted forward by propagating the last valid observation to the next valid one in order to align multiple data types with different sampling frequencies. This will prevent future information from leaking.

3. Denoising: Because stock trading involves a lot of irrational behavior, market data is noisy and can lead to inaccurate predictions and distort the direction of price movement. Stock price time series noise has been removed using the wavelet transform as a signal processing technique.

4. Data preprocessing: Preparing the data for analysis and modeling requires several procedures, including data pretreatment.

5. Filtering (Closing Price): Since the goal of the study is to predict closing stock prices, the dataset is filtered to include only the Close price column.

6. Scaling: To standardize the values between 0 and 1, the closing prices are scaled using the MinMaxScaler from the sklearn library. For all ML models to be trained effectively, this step is essential.

7. Splitting the Dataset into Training and Test Sets: Eighty percent of the dataset is utilized for training, while the remaining twenty percent is used for testing. This division makes it possible to test the model with data that hasn't been seen yet.

8. The Hyperparameter Adjustment: This is used for optimizing the parameters.

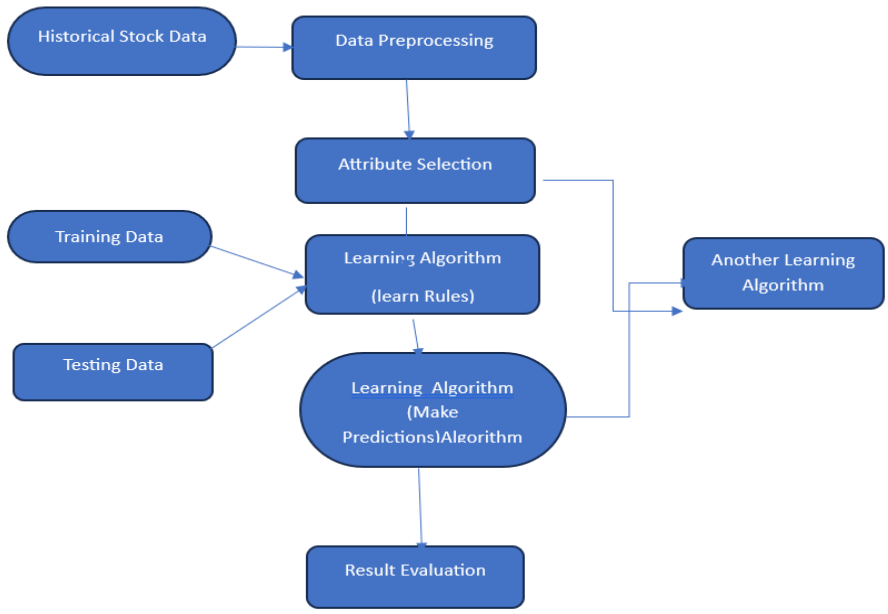

Figure7: About The Learning Environment

Model Evaluation Metrics

The ability of machine learning to improve the stock market or exchange, and forecast predictions is tested using several performance metrics. These performance metrics assess the specific algorithm according to its methodology and dataset. The following are some of the performance metrics that the chosen studies utilized to determine their effectiveness:

a. Accuracy: One measure used to evaluate the model categorization is accuracy. A component of the prediction that our model is accurate is informal accuracy.

b. RMSE: The difference between the expected model values and the retained data is calculated using the root mean square error, or RMSE. The training and assessment database and RMSE are quite similar.

c. MAE: Regression results are calculated using MAE. The sum of the differences between the actual and predicted variables, divided by the total number of data points, is the error prediction in this instance. The calculation of the difference between two continuous variables is known as MAE.

d. MSE: The square average error, or mean squared error (MSE), is a loss function that is used to compute minimal square regression. Additionally, it is calculated by dividing the total number of data points above all data points by the sum of the discrepancies between the predicted and actual variables.

e. MAPE: The most common application of mean absolute percentage error (MAPE) is in stock market forecasting calculations. It is the total of each individual's absolute mistakes divided by the demand. The error average is expressed as a percentage.

Additionally, several participants have made stock market predictions using their database and these performance metrics. Exchange rates in the stock market fluctuate either monthly or annually.

To evaluate the performance of models for stock market forecasting, various metrics are employed [23]. The forecasting model was evaluated using Mean Absolute Error (MAE) and Mean Squared Error (RMSE). The equations for evaluating the two metrics are shown in equations 1 and 2, respectively. The two equations determine the error resulting from the prediction by the given model [24].

[1] MAE = mean (absolute (Predicted Value - Actual Value)) .....(1)

[2] MSE = mean (square (Predicted Value - Actual Value)) .....(2)

In this research, MSE and MAE have been utilized as performance metrics to assess the outcomes. MSE is determined by squaring the difference between the predicted and actual values for each sample point, followed by averaging these squared differences across all sample points. Meanwhile, MAE quantifies the average percentage of absolute errors between the predicted and true values. This is achieved by calculating the absolute difference between the predicted and actual values for each sample point. Additionally, a newly proposed innovative performance measure has been applied [25].

7. EXPERIMENTAL WORK

In this study, one data set Apple (APPL Inc.), and three classifiers Random Forest, Linear Regression, and SVM have been used for SMF because many classification algorithms and data extraction methods exist. It is difficult to compare the classifier built in this work with a similar system in the literature. The classification ranks from highest to lowest accuracy is as follows. It is clear from this result that linear regression produces greater accuracy results. Since SVM has 97.36% and linear regression has 99.88%. All three classifications have the maximum accuracy of 99.90% for Apple Inc. data using the usual data set. The random forest comes in second on the SVM and has the lowest accuracy 97.36%.

Yahoo Finance includes historical stock price information. The Yahoo Finance website provides price information for the specified period for the chosen equity markets in .csv file format. Date, Open, High, Low, Close, Volume, and Adjusted Close are the seven features of the downloaded data files that, on a given date, display the date of stock trading and the stock's opening price. Its maximum and lowest trading prices, its closing price, the number of shares traded, and the closing price of a stock when investors receive dividends [22]. Yahoo Finance's dataset is used to train and test the suggested method. After going through the various models, it is divided into training and testing sets, respectively, and produces the following outcomes:

Price	Adj Close	Close	High	Low	Open	Volume
Ticker	AAPL	AAPL	AAPL	AAPL	AAPL	AAPL
Date						
2015-01-02 00:00:00+00:00	24.373957	27.332500	27.860001	26.837500	27.847500	212818400
2015-01-05 00:00:00+00:00	23.687296	26.562500	27.162500	26.352501	27.072500	257142000
2015-01-06 00:00:00+00:00	23.689537	26.565001	26.857500	26.157499	26.635000	263188400
2015-01-07 00:00:00+00:00	24.021711	26.937500	27.049999	26.674999	26.799999	160423600
2015-01-08 00:00:00+00:00	24.944677	27.972500	28.037500	27.174999	27.307501	237458000
...	...	...	...	...	...	...
2023-12-22 00:00:00+00:00	192.868134	193.600006	195.410004	192.970001	195.179993	37122800
2023-12-26 00:00:00+00:00	192.320221	193.050003	193.889999	192.830002	193.610001	28919300
2023-12-27 00:00:00+00:00	192.419830	193.149994	193.500000	191.089996	192.490005	48087700
2023-12-28 00:00:00+00:00	192.848221	193.580002	194.660004	193.169998	194.139999	34049900
2023-12-29 00:00:00+00:00	191.802185	192.529999	194.399994	191.729996	193.899994	42628800

2264 rows × 6 columns

Table 1: Stock data set taken from yahoo finance of APPL inc. in .csv file.

Results of Regression-Based Models

The plot in the figures shows the outcome of applying the linear regression method to the dataset to forecast changing prices over time. Where the prediction is shown by the green line and the actual data is shown by the blue line.

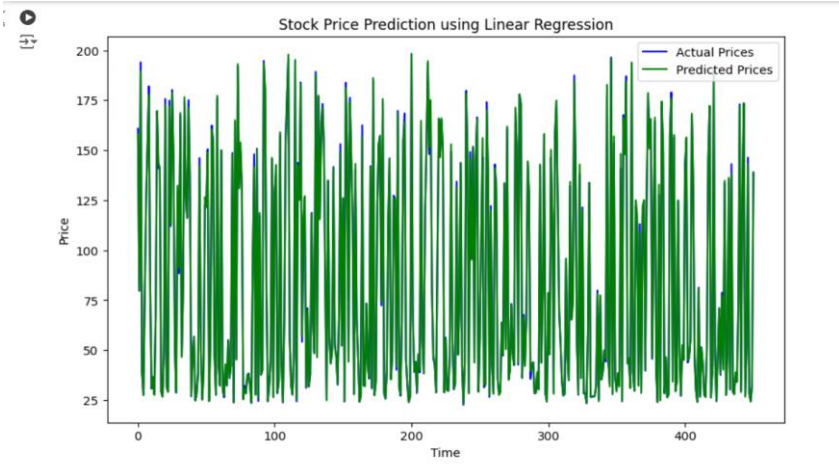

Figure 8: Put price and time on a linear regression plot. (graphical Representation)

Results of Random Forest Models

The plot in the figures shows the outcome of applying the linear regression method to the dataset to forecast changing prices over time. Where the prediction is shown by the green line and the actual data is shown by the blue line.

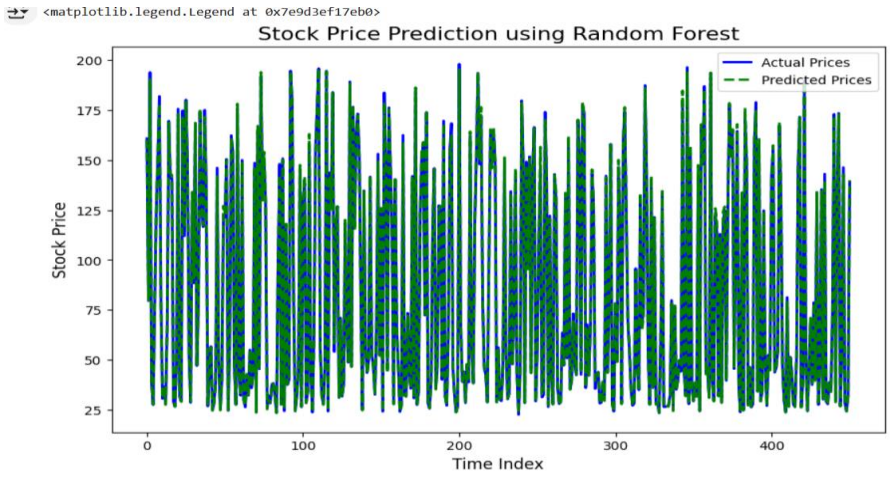

Figure 9: Put price and time on a linear regression plot.

Results of SVM Models

The plot in the figures shows the outcome of applying the linear regression method to the dataset to forecast changing prices over time. Where the prediction is shown by the green line and the actual data is shown by the blue line.

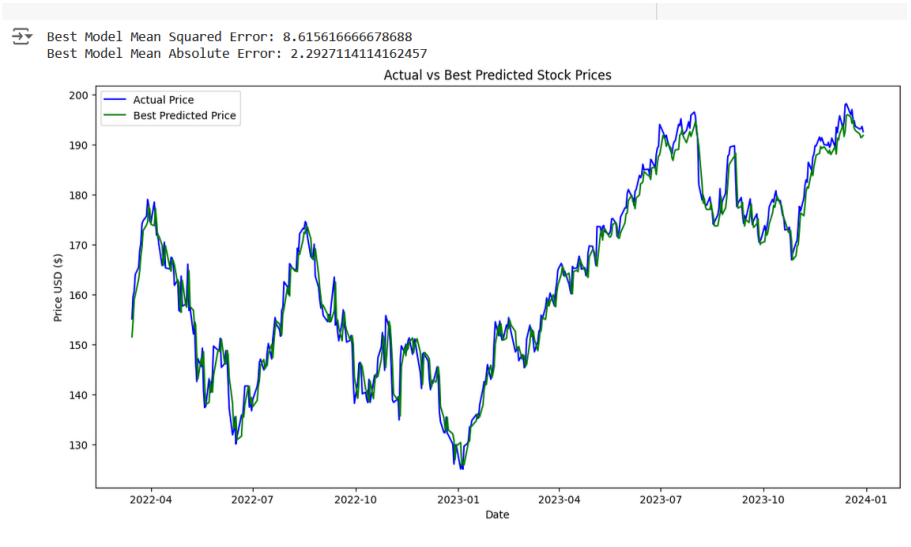

Fig 10: Put price and Date on a linear regression plot.

Performance Evaluation Results

Total No of errors	Linear Regression	Random Forest	SVM
MAE	1.0967751577	1.2025204104	2.3611476380
MSE	3.1619163049	3.7106956165	8.9675521944
R- Square	99.90%	99.98%	97.36%

Table 2: various model performance metrics

8. CONCLUSION

This paper attempts to research the various strategies used in stock market forecasting to determine which of the SVM, RF, and linear regression methods provides the best results and which technique can be used to achieve maximum forecasting accuracy. Its main objective is to study the currently adopted techniques, major methods, performance metrics, and datasets and determine which method will be the best. Here we compare three types of supervised machine learning that used label data and tried to find out which method gives us higher forecasting accuracy. Data was used to train and test the model. We obtained Appl Inc's dataset from Yahoo Finance between January 1, 2015, to January 1, 2024, and used the Python program to do this. We divided the data set into 80% training and 20% testing data. For this performance metrics, MAE, MSE, R-Square, and accuracy were taken for which separate values were obtained from each model, as shown in the Table 2 We are still trying to maximize forecasting accuracy by combining the data from different systems. Yes, this model provides many advantages compared to the selected two benchmarks. Where the accuracy of linear regression is 99.90%, The accuracy of random forest is 99.88% and the accuracy of SVM is 97.36%. Shows that the highest accuracy was obtained in the form of linear regression. It was observed that the performance of the linear regression model was significantly better than the two models implemented. Therefore, further research should consider using different options. Once again, the movement of stock prices is influenced by both past stock data and fundamental data, such as consumer satisfaction with the company's market and online news. Therefore, future studies might look at how user mood and online financial news affect stock price movement prediction.

References

1. Nti, I., Adekoya, A., & Weyori, B. (2020). Efficient stock market prediction using ensemble support vector machine. Open Computer Science, 10(1), 153-163. https://doi.org/10.1515/comp-2020-0199
2. Hiransha, M. E. A. G., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. Procedia computer science, 132, 1351-1362.
3. Kumar, D., Sarangi, P. K., & Verma, R. (2022). A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings, 49, 3187-3191.
4. Aghakhani, K., & Karimi, A. (2016, July). A new approach to predicting stock big data by combining neural networks and the harmony search algorithm. In 5th International Conference on Computer Science, Electrical and Electronics Engineering, Malaysia.
5. Mintarya, L. N., Halim, J. N., Angie, C., Achmad, S., & Kurniawan, A. (2023). Machine learning approaches in stock market prediction: A systematic literature review. Procedia Computer Science, 216, 96-102
6. Kuo, R. J., & Chiu, T. H. (2024). Hybrid of jellyfish and particle swarm optimization algorithm-based support vector machine for stock market trend prediction. Applied Soft Computing, 154, 111394.
7. Panwar, B., Dhuriya, G., Johri, P., Yadav, S. S., & Gaur, N. (2021, March). Stock market prediction using linear regression and SVM. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 629-631). IEEE.
8. Gurjar, M., Naik, P., Mujumdar, G., & Vaidya, T. (2018). Stock market prediction using ANN. International Research Journal of Engineering and Technology, 5(3), 2758-2761.
9. Usmani, M. et al., Stock market prediction using machine learning techniques, 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), IEEE, 2016, 322–327.
10. Vui, C. S., Soon, G. K., On, C. K., Alfred, R., & Anthony, P. (2013, November). A review of stock market prediction with Artificial neural network (ANN). In 2013 IEEE international conference on control system, computing and engineering (pp. 477-482).

11. Aromolaran, O., Aromolaran, D., Isewon, I., & Oyelade, J. (2021). Machine learning approach to gene essentiality prediction: a review. Briefings in Bioinformatics, 22(5), bbab128.
12. Sharma, K., & Bhalla, R. (2022). Stock market prediction techniques: a review paper. In Second International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2021 (pp. 175-188). Springer Singapore
13. Bhuriya, D., Kaushal, G., Sharma, A., & Singh, U. (2017, April). Stock market predication using a linear regression. In 2017 international conference of electronics, communication and aerospace technology (ICECA) (Vol. 2, pp. 510-513). IEEE.
14. Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. Electronics, 10(21), 2717.
15. Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. Decision Analytics Journal, 3, 100071.
16. Subasi, A., Amir, F., Bagedo, K., Shams, A., & Sarirete, A. (2021). Stock market prediction using machine learning. Procedia Computer Science, 194, 173-179.
17. Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM. Machine Learning with Applications, 9, 100320
18. Reddy, V. K. S., & Sai, K. (2018). Stock market prediction using machine learning. International Research Journal of Engineering and Technology (IRJET), 5(10), 1033-1035.
19. Li, Z., Yu, H., Xu, J., Liu, J., & Mo, Y. (2023). Stock market analysis and prediction using LSTM: A case study on technology stocks. Innovations in Applied Engineering and Technology, 1-6.
20. Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. Journal of Ambient Intelligence and Humanized Computing, 1-24.
21. Ajiga, D. I., Adeleye, R. A., Tubokirifuruar, T. S., Bello, B. G., Ndubuisi, N. L., Asuzu, O. F., & Owolabi, O. R. (2024). Machine learning for stock market forecasting: a review of models and accuracy. Finance & Accounting Research Journal, 6(2), 112-124.
22. Joseph, E., Mishra, A., & Rabiu, I. (2019). Forecast on close stock market prediction using support vector machine (SVM). Int J Eng Res Technol (IJERT), 8(02).
23. Zhan, Z., & Kim, S. K. (2024). Versatile time-window sliding machine learning techniques for stock market forecasting. Artificial Intelligence Review, 57(8), 209.
24. Singh, G. (2022). Machine learning models in stock market prediction. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 11(3), 18. Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). https://doi.org/10.35940/ijitee.C9733.0111322