# Forecasting Stock Markets Trends using Machine Learning Algorithms

Lahari Kotapati
*Department of CSE,*
*SRM University-AP, India*
*lahari_kotapati@srmap.edu.in*

Reethu Bhargavi Sajjala
*Department of CSE,*
*SRM University-AP, India*
*bhargavi_sajjala@srmap.edu.in*

Sricharan Gudi
*Department of CSE,*
*SRM University-AP, India*
*sricharan_gudi@srmap.edu.in*

Beecha Venkata Naga Hareesh
*Department of CSE,*
*SRM University-AP, India*
*nagahareesh_beecha@srmap.edu.in*

Srilatha Tokala
*ACT Lab, Department of CSE,*
*SRM University-AP, India*
*srilatha_tokala@srmap.edu.in*

Murali Krishna Enduri
*ACT Lab, Department of CSE,*
*SRM University-AP, India*
*muralikrishna.e@srmap.edu.in*

***Abstract*—Prices on the stock market vary frequently as a result of many economic, political, and social reasons. It is a fluid and challenging environment. Investors look for efficient tools to enhance their investing strategy and make well-informed judgements. his study investigates how stock market fluctuations are predicted using machine learning approaches and how well they can identify complex patterns. Various algorithms, including Logistic Regression, K-Nearest Neighbours, Support Vector Machines, Random Forest, Decision Tree, Naive Bayes, Long Short-Term Memory, Multilayer Perceptron, and XG Boost are evaluated and compared based on their accuracy, precision, recall, and F1 score. The Multilayer Perceptron emerges as the most accurate predictor, showcasing its ability to handle complex relationships and learn from historical data. The results of this study provide insightful information for investors who want to use predictions from machine learning in their decision-making.**

**Keywords- political; stock market; investors; machine learning**

## I. Introduction

Stock market predictions are a difficult and crucial task, as it has significant implications for investors, traders, and financial institutions. Accurate predictions can assist traders by arriving at smart choices, optimize their portfolio, and mitigate risks. In 2019, the value of the world's stock markets exceeded $85 trillion (Pound, 2019). Investors used to rely on their personal experiences to understand market trends, but this is no longer possible due to the size of the market and the pace of transaction [1]. Machine learning algorithms have emerged as powerful tools in analysing complex financial data and making predictions. Since it might lead to significant returns, stock forecasting has long been a popular topic. Due to the stock data sequence's resemblance to a random walk, stock market analysis is not a simple task [2]. The current investigation will look at how different machine learning algorithms are used to forecast stock market movements and how efficiently they perform at producing precise projections. Machine learning is commonly used in stock market prediction [3]. ML is beneficial for stock market prediction because it can process large datasets quickly and find hidden patterns that may not be evident to humans.

Traders and investors may use ML algorithms to make better judgments, spot possible trends, and manage risks more successfully, which will improve the stock market's performance. The data is available for download from the website named Kaggle and it gives us information about the historical datasets of a particular company for future predictions. In this study, we explored multiple algorithms for machine learning for stock market prediction [4]. The methods studied include Logistic Regression, KNN, SVM, Random Forest, Decision Tree, Naive Bayes, LSTM, MLP, and XG Boost. By collecting historical data sets of selected companies, we trained and tested each algorithm to evaluate their performance in predicting future stock prices. The objective was to determine which algorithm provides the best accurate forecasts in the dynamic and constantly shifting stock market environment. The study's findings are very important for traders and investors since they can help them make more informed decisions about the stock market.

## II. Related Works

Predicting stock prices is essential for investment strategy, but it is not foolproof due to future uncertainties. Fundamental analysis assesses a stock's value based on internal and external factors, while technical analysis predicts prices from historical data and trends. Various methods, like deep learning, have been proposed for prediction. Deep learning algorithms like ANN, RNN, LSTM, SLSTM, and BLSTM show promise, and combining historical stock data with social media can achieve up to 70% accuracy. Successful predictions support investors in making informed decisions [5]. Numerous Machine Learning techniques have been researched to predict stock values. investigated. The time frames being predicted (short, long, etc.), the stocks being considered (certain industries, all stocks, etc.), and the types of data being used as predictors (global trends, company-specific data, historical stock prices, etc.) vary amongst these initiatives. The study makes a distinction between two prediction techniques: one makes forecasts based on rules (dummy prediction), while the other makes predictions based on real-time data, such as recently updated internet stock prices real-time prediction. Radial Basis Function (RBF) and SVM are two machine learning approaches used in this project. These methods are used with data gathered from various international financial marketplaces. SVM is a good option as a result of its capacity to handle massive datasets, for stock market prediction. and its ability to prevent overfitting. Through SVM and RBF, the project underscores the potential of these techniques for accurate stock market predictions. This study makes a contribution insight at the intersection of machine learning and finance, aiding better-informed investment choices. The work demonstrates the potential for refined stock market predictions, which could lead to improved investment strategies [6].

Researchers collected data, including news and Twitter feeds, to understand if factors like oil rates and foreign exchange influence market performance. They tested four machine learning algorithms - Single Layer Perceptron, MLP, RBF, and SVM- to see which one performed best. According to the findings the MLP was the best in foretelling market performance, especially on new and unseen data. It achieved a 77% accuracy rate. This suggests that machine learning techniques can accurately anticipate stock market movements trends, even with limited data. Interestingly, they found that factors like oil rates were strongly related to market behavior, while foreign exchange had little impact [7]. The main goal was to enhance risk mitigation techniques in investment strategies. Notably, results revealed that the Random Forest and Bagging algorithms outperformed others when using a leaked dataset, demonstrating impressive predictive accuracy. The study's meticulous examination of these algorithms and their exceptional performance highlights the complexity of objectively assessing diverse methods for forecasting stock market trends [8].

LSTM, an advanced version of RNN, is used for long-term data retention, making it more reliable. LSTM provides higher accuracy than regression-based models. Machine learning gets closer to enabling better stock market forecasting [9]. To improve predictions, researchers use LS-SVM and optimize it with PSO. LS-SVM-PSO outperforms other methods, including ANN-BP. Technical indicators like RSI, MFI, EMA, SO, and MACD help traders assess stock conditions. LS-SVM-PSO with technical indicators shows promise for accurate daily stock price prediction [10]. ANN models such feedforward NN, BPNN, and hybrid NN. Even though these models provided outcomes that were satisfactory, ongoing research aims to identify fresher strategies for even greater precision [11].

## III. METHODOLOGY

The stock market generates several structured and unstructured data points. Machine learning algorithms swiftly process complex data to increase reliability. Both supervised and unsupervised machine learning techniques are used in SMP [12]. To improve stock price predictions, machine learning was applied in the analysis. A special LSTM model outperformed traditional strategies, showing that more accurate methods are conceivable [13]. The flowchart's methodical approach ensures a thorough evaluation procedure, enabling us to find insightful information that advances our comprehension of algorithmic performance as shown in Figure 1.

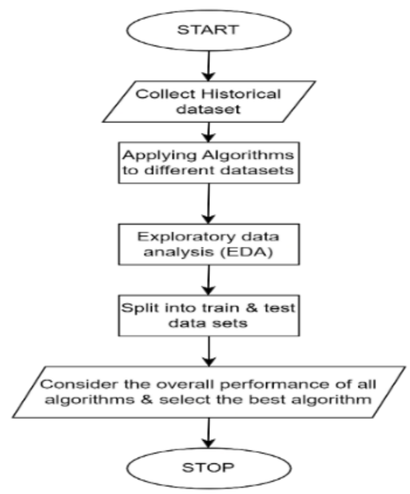


Fig.1 Workflow diagram for the overall procedure.

To recreate real-world circumstances using such historical datasets, increase the validity of our findings. The exploratory data analysis (EDA) step stands as a bridge between data collection and algorithmic evaluation. Through data visualization and statistical techniques, we uncover hidden trends, potential outliers, and data-specific challenges that could impact algorithm performance. Dividing the dataset into sections for training and testing addresses the crucial aspect of generalization. Algorithms that perform exceptionally well during training might exhibit subpar performance on unseen data due to overfitting. By evaluating algorithms on separate test data, we ensure that our analysis accounts for their ability to generalize, offering a more realistic assessment of their practical utility. Accuracy, precision, recall, and F1-score provide a multi-dimensional view of performance, considering factors beyond mere classification correctness.

In this paper we have analysed different machine learning methods employed in stock market predictions. This paper includes techniques such as

### A. Logistic Regression

It involves multiple methods for modeling and evaluating factors, particularly when exploring the connection between variables [14]. By analyzing historical stock market data with consistent company-specific and broader economic variables, logistic regression aims to simulate how input characteristics correspond to the probability of positive or negative stock market outcomes.

### B. K-Nearest Neighbors (KNN)

An indispensable algorithm for predictive analysis, KNN gauges distances between data points in feature space to classify new instances by the majority class among k-nearest neighbors. Utilizing KNN's capacity to capture local patterns and adapt to data distributions, we aim to assess its performance in predicting stock market trends with consistent company-specific and economic factors.

### C. Support Vector Machine (SVM)

SVM is a learning method that is used to construct classifications and regression analysis. SVM is fed a training set of samples labelled with one of two distinct groups [15]. Our predictive modelling approach is centered around SVM. It operates by turning data points into a high-dimensional feature space, then locating the ideal hyperplane to distinguish between favorable and unfavorable stock market outcomes.

### D. Decision Tree

It is a fundamental algorithm that forms the basis for more advanced methods like Random Forest. The process entails generating a tree-like structure with a node for each internal element choice based on a specific feature and a leaf node corresponding to a predicted outcome after recursively dividing the data according to the selected characteristics. Decision Tree seeks to identify key decision points to effectively classify future stock market outcomes.

### E. Random Forest

In contrast to numerous decision trees, it functions by combining separate trees' predictions to provide a strong and accurate final prediction. In our research, we utilize Random Forest to explore how the combination of company-specific and macroeconomic factors can lead to improved stock market predictions.

### F. Naive Bayes

Naive Bayes is a good contender for stock market forecasting because to its effectiveness in handling high-dimensional data and relatively quick training time, which adds important insights to the project's more comprehensive predictive modelling strategy.

### G. Long Short-Term Memory (LSTM)

It plays a pivotal role in exploring sequential patterns and time dependencies within stock market data. Recurrent neural networks (RNNs) such as LSTM are created with the aim of capturing long-term dependencies and retain information over extended time intervals.

### H. Multilayer Perceptron (MLP)

An artificial neural network having numerous layers of linked nodes, also referred to as neurons, is referred to as an MLP. To reduce prediction mistakes and enhance generalization performance, the system adjusts its weights and biases during the training phase. Because it can simulate non-linear connections, MLP is a good choice for reflecting the complexity of financial markets.

### I. XGBoost

Because of its accuracy and speed, XGBoost aims to build a solid and reliable prediction framework. It builds several decision trees successively, with each one correcting the flaws of the one before it. This iterative process enables XGBoost to learn from previous mistakes and continuously improve its predictions by efficiently capturing complicated correlations between the elements.

## IV. Dataset Statistics

It started with the cautious gathering of historical stock price information from a wide range of businesses, including well-known ones like Airtel, IBM, Facebook, and others. The foundation of analysis was laid by the thoughtful structuring of this data, which included elements like date, opening price, closing price, high, low, trading volume, and adjusted closing price. The statistics of the datasets that are taken are described in Table I.

The dataset was rigorously examined for any gaps, mistakes, or missing numbers that would potentially bias the results. The dataset was strategically partitioned into distinct training and testing subsets. This division ensured that 80% of the data was allocated for training the machine learning models, thereby enabling them to grasp complex relationships between attributes like date, prices, and trading volume. The remaining 20% was dedicated to model evaluation, vital for a comprehensive assessment of their predictive capabilities.

| | **Reliance** | **IBM** |
|---|---|---|
| No of rows | 6305 | 5282 |
| No of columns | 7 | 7 |
| No of null values | 127 | 0 |
| Time Period | 01-01-1996 to 01-01-2021 | 23-07-2001 to 20-07-2022 |

Table I: Dataset information

With their training and evaluation complete, the models confidently turned towards the future and the knowledge gained from historical data, they embarked on predicting new, unseen data. The input features extracted from the testing dataset acted as blueprints, guiding the models' predictions for the target variable the elusive closing price. The experience was a reminder of the interdependence of data curation, algorithmic learning, and forecasting skill in the history of finance and data science. From data gathering and cleaning to model training, evaluation, and beyond, the procedures taken constitute a symphony of insights, providing a surface on which the complex design of stock price prediction was gracefully created.

## V. Results

We conducted four distinct measurements: accuracy, precision, recall, and F1-score. These measurements were employed to gauge the accuracy of weather-based predictions, guiding the decision-making process concerning investments in the stock market. This comprehensive approach provides a thorough assessment of the model's effectiveness in determining whether to invest in stocks or not, considering the intricate relationship between weather patterns and market performance.

Let us consider J=True positive, K=True Negative, L=False Positive, M=False Negative

**Accuracy:** Calculates the proportion of times a model was accurate across the entire dataset, which is still accurate if the dataset is class balanced.

$$Accuracy = (J + K)/ (J + K + L + M)$$

**Precision:** a classification model's capacity to focus on relevant data items exclusively. Mathematicians divide the total true positives by the total true positives plus false positives to determine accuracy.

$$Precision = (J)/ (J + L)$$

**Recall:** The capacity of a model to locate all pertinent occurrences in a data source. Before using the recall formula, we need divide the total number of true positives by the sum of true positives and false negatives.

$$Recall = (J)/ (J + M)$$

**F1 score:** A singular metric that fuses both recall and precision by employing the harmonic mean, creating a comprehensive and balanced evaluation of a model's performance in terms of identifying true positives and minimizing false positives.

$$F1\ score= 2\times(precision\times recall)/ (precision + recall)$$

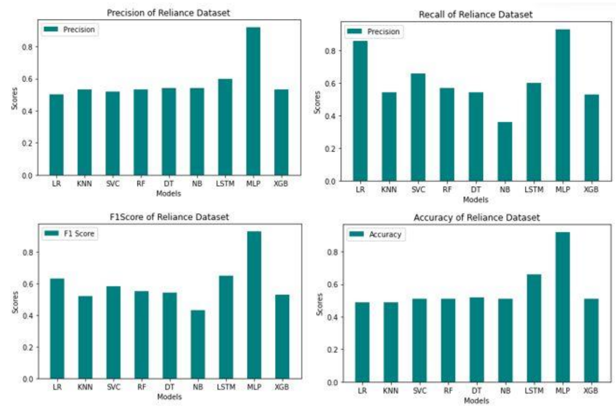


Fig 2 Different performance metrics for Reliance dataset.

Figure 2 shows the performance metrics that are evaluated on different algorithms for predicting the stock market. MLP achieved an outstanding accuracy of 92% when algorithmic accuracy was assessed for the Reliance dataset. The accuracy of the LSTM, which followed closely, was 66%. SVM's accuracy was 51%, while Decision Tree and Naive Bayes' accuracy was also 51%. Similar accuracy results were obtained by XG Boost and Random Forest, which were 51% and 50%, respectively. The accuracy of KNN and logistic regression was slightly lower at 49%. LSTM followed suit with a decent precision of 72%, while the MLP algorithm stood out with the highest score of 93% when we turned our attention to precision. The lowest precision was continuously seen among the four different algorithms logistic regression, KNN, SVM, and Random Forest with all obtaining a precision of 48%. The highest recall score was maintained by MLP at 92%, while Naive Bayes came in second place with a recall score of 67%. SVM, in comparison, had the lowest recall performance, coming in at a mere 34%. LSTM came in second with a creditable score of 65%, followed by MLP with an amazing value of 92% in the F1-Score. On the Reliance dataset, KNN's F1-Score was 40%, the lowest of the classifiers.

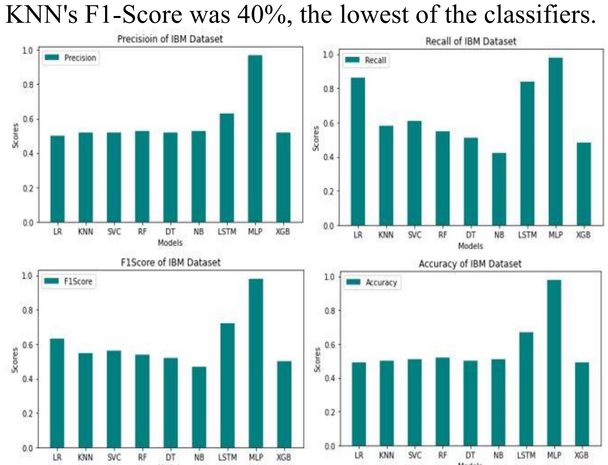


Fig 3 Different performance metrics for IBM dataset.

The algorithms showed varied degrees of accuracy for the IBM dataset as shown in Figure 3. MLP attained a remarkable accuracy of 98%, while LSTM only managed to obtain 67%. Random Forest followed suit with accuracy of 52%, and both Logistic Regression and Random Forest reached accuracy of 52%. Naive Bayes, KNN, and SVM all recorded accuracy levels of 51%, while XG Boost, Decision Tree, and each recorded accuracy levels of 50%. The highest precision, in comparison, was a staggering 97%, and LSTM trailed closely behind with a precision of 63%. At the other end of the scale, XG Boost had the lowest degree of precision (46%). When we turned to recall, MLP scored the highest (97%), closely followed by LSTM (84%). The Nave Bayes model produced the lowest recall score, a meagre 42%. Examining the F1-Score, MLP once more achieved the highest mark with a staggering 98%, closely followed by LSTM with an F1-Score of 72%.

Naive Bayes, on the other hand, had the lowest F1-Score with a respectable performance of 47%. Through meticulous adherence to these systematic steps, the process strives to meticulously select a robust and precise machine learning algorithm capable of effectively predicting stock prices for the enlisted companies. This pursuit leads to invaluable insights, acting as a guiding compass for stakeholders in their intricate investment decisions.

All graphs are plotted with an eye on occurrences marked as '1', indicating a readiness to invest in the stock market, equivalent to a green signal.

## VI. CONCLUSION AND FUTURE WORK

Using machine learning algorithms, stock market predictions is an explosive topic in the business right now. As a result, our research compares nine machine learning algorithms on two distinct stock indices datasets, including IBM and Reliance to aid in risk reduction investment. Furthermore, the results revealed that Multi-Layer Perceptron has the best accuracy and Decision Tree has the lowest accuracy when applied to all datasets, likewise, MLP does not have the issue of overfitting. Bagging with the explored outcomes delivers above satisfactory performance.

Future research on stock market forecasting utilizing machine learning algorithms will boost the model's precision and efficacy. By integrating additional deep learning algorithms with existing MLP, we aim to conduct comprehensive comparisons to determine the best performing approach. The pursuit of improved models will lead to more reliable forecasts, enabling investors to make more informed decisions with less risk.

## REFERENCES


[1] Strader, T. J., Rozycki, J. J., Root, T. H., & Huang, Y. H. J. (2020). Machine learning stock market prediction studies: review and research directions. Journal of International Technology and Information Management, 28(4), 63-83.

[2] Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). A machine learning model for stock market prediction. arXiv preprint arXiv:1402.7351.

[3] Sharma, A., Bhuriya, D., & Singh, U. (2017, April). Survey of stock market prediction using machine learning approach. In *2017 International conference of electronics, communication and aerospace technology (ICECA)* (Vol. 2, pp. 506-509). IEEE.

[4] Umer, M., Awais, M., & Muzammul, M. (2019). Stock market prediction using machine learning (ML) algorithms. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *8*(4), 97-116.

[5] Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., & Wills, G. (2020). A survey on machine learning for stock price prediction: Algorithms and techniques.

[6] Reddy, V. K. S. (2018). Stock market prediction using machine learning. International Research Journal of Engineering and Technology (IRJET), 5(10), 1033-1035.

[7] Usmani, M., Adil, S. H., Raza, K., & Ali, S. S. A. (2016, August). Stock market prediction using machine learning techniques. In 2016 3rd international conference on computer and information sciences (ICCOINS) (pp. 322-327). IEEE.

[8] Subasi, A., Amir, F., Bagedo, K., Shams, A., & Sarirete, A. (2021). Stock market prediction using machine learning. Procedia Computer Science, 194, 173-179.

[9] Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55(1-2), 307-319.

[10] N. Ancona, Classification Properties of Support Vector Machines for Regression, Technical Report, RIIESI/CNRNr. 02/99.

[11] 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.

[12] 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.

[13] Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018, December). Stock market prediction using machine learning. In 2018 first international conference on secure cyber computing and communication (ICSCCC) (pp. 574-576). IEEE.

[14] Sharma, A., Bhuriya, D., & Singh, U. (2017, April). Survey of stock market prediction using machine learning approach. In *2017 International conference of electronics, communication and aerospace technology (ICECA)* (Vol. 2, pp. 506-509). IEEE.

[15] Reddy, V. K. S. (2018). Stock market prediction using machine learning. *International Research Journal of Engineering and Technology (IRJET)*, *5*(10), 1033-1035.