IJMSERH

International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

Volume 14, Issue 2, April-June 2026

Impact Factor: 9.274

International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402026

Stock Market Prediction using Python

Mr. M.Suresh Kannan[1], Mrs. D. Saroj Priyadharsini[2]

PG Scholar, Department of Master of Computer Applications, R.V.S College of Engineering, Dindigul, Tamil Nadu, India[1]

Assistant Professor, Department of Master of Computer Applications, R.V.S College of Engineering, Dindigul, Tamil Nadu, India[2]

ABSTRACT: Stock market prediction has been an area of interest for investors as well as researchers for many years due to its volatile, complex and regularly changing in nature, making it difficult to make reliable predictions. Proposed system provides an approach towards prediction of stock market trends using machine learning models like Random Forest model and Support Vector Machine. The Random Forest model is an ensemble learning method that has been an exceedingly successful model for classification and regression. Support vector machine is a machine learning model for classification. However, this model is mostly used for classification. These techniques are used to forecast whether the price of a stock in the future will be higher than its price on a given day, based on historical data while providing an in-depth understanding of the models being used.

I. INTRODUCTION

In statistics and machine learning the problem that involves deciding which category a new observation belongs to among a set of categories using a dataset for training which contains observations whose category membership is mentioned. In machine learning, classification is regarded as an example of supervised learning in which learning takes place using a training dataset containing observations with correctly identified classes provided in advance.

The trend in stock market prediction is not a new thing yet this issue is kept being discussed by the various organization. Being able to predict accurately the future financial outcome is equivalent to earning big money. Proposed work aims at analyzing this problem in an academic way which provides a different way of prediction on the market trend. Proposed work will develop a financial data predictor program there will be dataset storing all historical stock prices and data will be treated as training sets for the program. To predict the stock prices in NSE, this project aims at developing a program, which serves best solution for accurate predicted stock result.

Random Forest model is an ensemble learning method that creates multiple decision trees based on random subsets of data during the training and outputs the mode of classes that have resulted from these decision trees using an input for classification. Random Forest Model involves a general technique called Bootstrap Aggregation which leads to better model performance by decreasing the variance. Support Vector Machine is a machine learning algorithm used for classification and regression. Support Vector Machine is predominantly used for classification problems, it is a maximum margin classification algorithm. In Support Vector Machine each data item is plotted on an n-dimensional space, where n is equal to the number of features and then an attempt is made to classify these points by finding the most suitable plane that differentiates these points better than all the other planes.

II. EXISTING SYSTEM

Sentiment Analysis has been used in multiple areas such as blogging websites, review websites, online retail etc. Sentiment analysis has been a very important social media analytics tool and has been effectively used by E-commerce websites like Snapdeal and Flipkart to filter irrelevant reviews. Sentiment analysis at present plays a vital role in customer service, management of brand reputation and business intelligence. It has also been influential in politics by helping political strategists determine the public opinion on the internet. It played a crucial role in Barack Obama's campaign in 2011 where sentiment analysis was used to predict the responses to campaign messages

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402026

DISADVANTAGES:

- Depends mainly on sentiment analysis
- Low prediction accuracy
- Cannot analyze complex stock patterns
- Not suitable for daily trading
- Affected by sudden market changes
- Difficult to process huge datasets
- Less reliable forecasting results

III. PROPOSED SYSTEM

Proposed system studies stock market prediction using machine learning techniques for evaluating data concerned to the stock market using a novel method for prediction of stock prices to minimize the risk of investment in a stock market. Proposed work uses machine learning algorithm namely Random Forest model and Support Vector Machine. Random Forest model is an ensemble learning method that creates multiple decision trees based on random subsets of data during the training and outputs the mode of classes that have resulted from these decision trees using an input for classification. NSE stock is considered for stock prediction, the dataset collected from 2000 to till date is used.

ADVANTAGES:

- Uses Machine Learning techniques
- Higher prediction accuracy
- Efficient data analysis
- Better investment decision support
- Reduces investment risk
- Faster prediction process
- Handles large datasets effectively
- Reliable forecasting performance

IV. SYSTEM ARCHITECTURE

The system architecture for Stock Price Prediction using Machine Learning Approaches consists of several important modules that work together to predict future stock market trends accurately. Initially, historical stock market data is collected from NSE datasets containing attributes such as open price, close price, high price, low price, and trading volume. The collected data is then passed to the data preprocessing module, where missing values are removed and the dataset is normalized and transformed into a suitable format for analysis. Feature engineering techniques are applied to extract useful information from the dataset, improving the efficiency of the prediction models.

After preprocessing, the processed data is provided to machine learning models such as Random Forest and Support Vector Machine for training and prediction. The training and testing module divides the dataset into training and testing data to evaluate model **performance** effectively. The prediction module generates stock market movement predictions based on learned historical patterns, while the performance evaluation module measures the accuracy and efficiency of the models using evaluation metrics. Finally, the predicted results are displayed to the user through the user interface, helping investors make better financial and investment decisions with reduced risk.

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402026

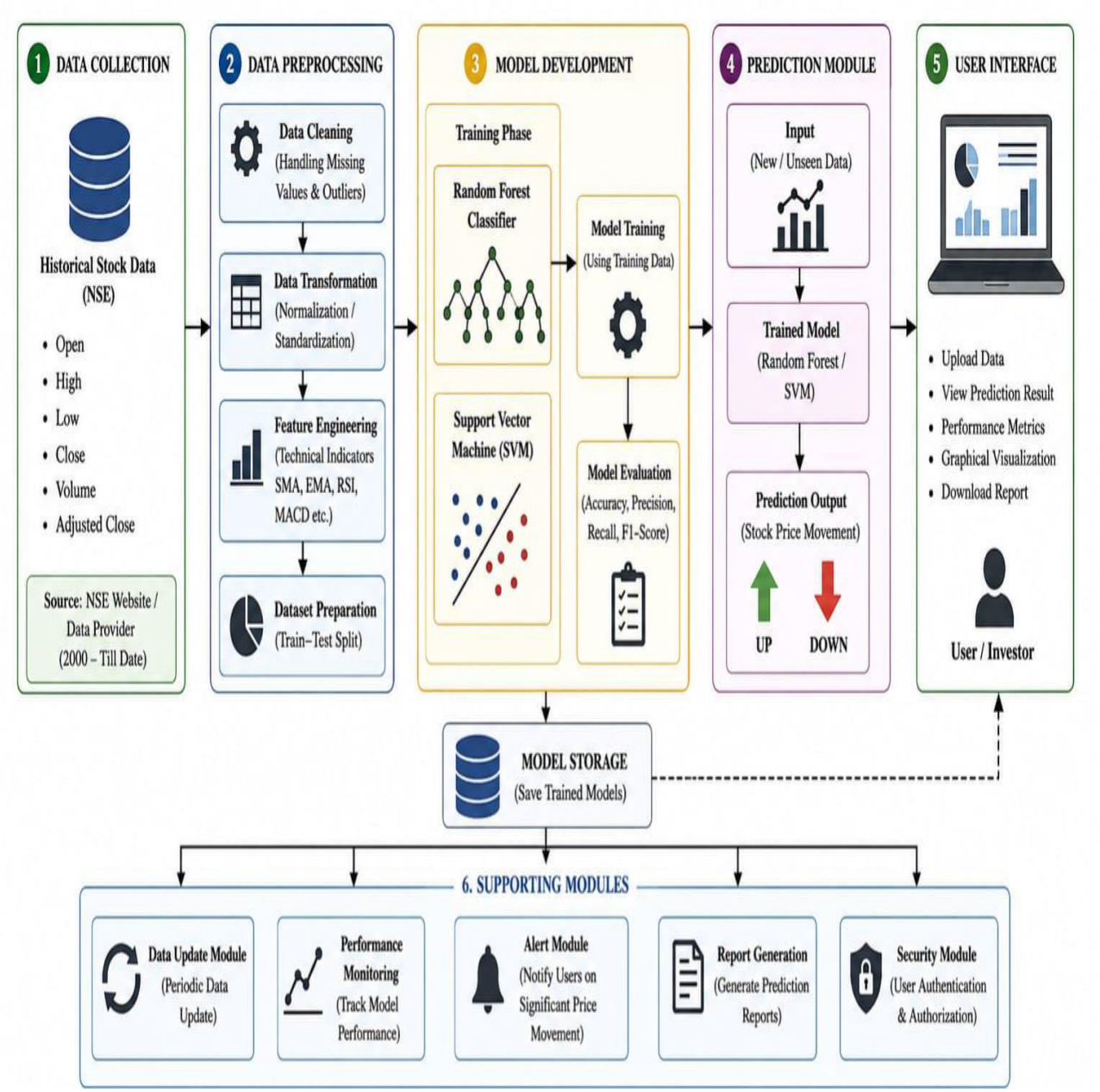

Fig NO 4.1

MODULES

1. Data Collection Module
2. Data Pre processing Module
3. Random Forest Prediction Module
4. Support Vector Machine Module
5. Forecasting and Result Visualization Module

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402026

V. MODULE DESCRIPTIONS

1. **Data Collection Module**
   The Data Collection Module is one of the most important parts of the stock market prediction system. This module is responsible for collecting both live and historical stock market data from sources such as NSE and Yahoo Finance. It gathers important stock details including Open price, High price, Low price, Close price, Volume, VWAP, and previous closing price. The collected data is stored in CSV file format so that it can be easily used for further processing and prediction. The module continuously maintains stock market data from the year 2000 to the present date, which helps in analyzing long-term stock market trends. By collecting accurate and reliable stock information, this module provides the foundation for the entire prediction process. The quality of prediction mainly depends on the quality and accuracy of the data collected through this module.

2. **Data Pre processing Module**
   The Data Preprocessing Module is responsible for preparing the collected stock market dataset for machine learning operations. Since raw stock market data may contain missing values, unwanted information, duplicate entries, or noise, this module performs several preprocessing operations to clean and organize the dataset. It removes invalid or empty values and converts the data into a proper structured format suitable for prediction algorithms. The module also separates the dataset into training and testing data for efficient model learning and evaluation. Data validation techniques are used to ensure the correctness and consistency of the dataset before it is provided to machine learning models. Proper preprocessing improves the performance, accuracy, and reliability of the prediction system. This module plays a major role in reducing prediction errors and increasing the efficiency of stock market forecasting.

3. **Random Forest Prediction Module**
   The Random Forest Prediction Module uses the Random Forest machine learning algorithm to predict stock market trends and future stock prices. Random Forest is an ensemble learning technique that creates multiple decision trees using random subsets of the training dataset. Each decision tree produces a prediction result, and the final output is generated based on majority voting among all trees. This module helps in improving prediction accuracy and reducing variance compared to single decision tree models. It is capable of handling both classification and regression problems effectively. The Random Forest algorithm is highly efficient in analyzing large amounts of stock market data and identifying hidden patterns within the dataset. This module helps investors make better investment decisions by providing accurate stock price predictions based on historical market behavior.

4. **Support Vector Machine Module**
   The Support Vector Machine Module uses the Support Vector Machine (SVM) algorithm for stock market classification and prediction. In this module, stock market data points are plotted in an n-dimensional space where the algorithm identifies the best possible hyperplane that separates the data into different classes. The hyperplane with the maximum margin between classes is selected to improve classification accuracy. SVM is mainly used for classification problems, but it can also support regression analysis. This module is highly effective in handling complex and non-linear stock market data. It helps in analyzing stock price movements and predicting future trends with improved accuracy. By using mathematical optimization techniques, the SVM module increases the reliability and efficiency of the stock market prediction system.

5. **Forecasting and Result Visualization Module**
   The Forecasting and Result Visualization Module is responsible for generating future stock price forecasts and displaying the prediction results in an understandable format. This module presents stock market analysis using graphical representations such as line charts, stock trend graphs, and candlestick charts. It also provides comparison results and accuracy evaluation of machine learning models such as Random Forest and Support Vector Machine. A user-friendly graphical interface is developed to allow users to enter stock symbols, perform predictions, and view forecasting results easily. The module helps investors and users understand stock market behavior clearly through visual analysis. By displaying prediction accuracy and forecasting outcomes, this module supports better decision-making and improves user interaction with the system.

International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402026

VI. RESULTS

1. DASHBOARD

Fig No.6.1

2. PREDICTION PAGE

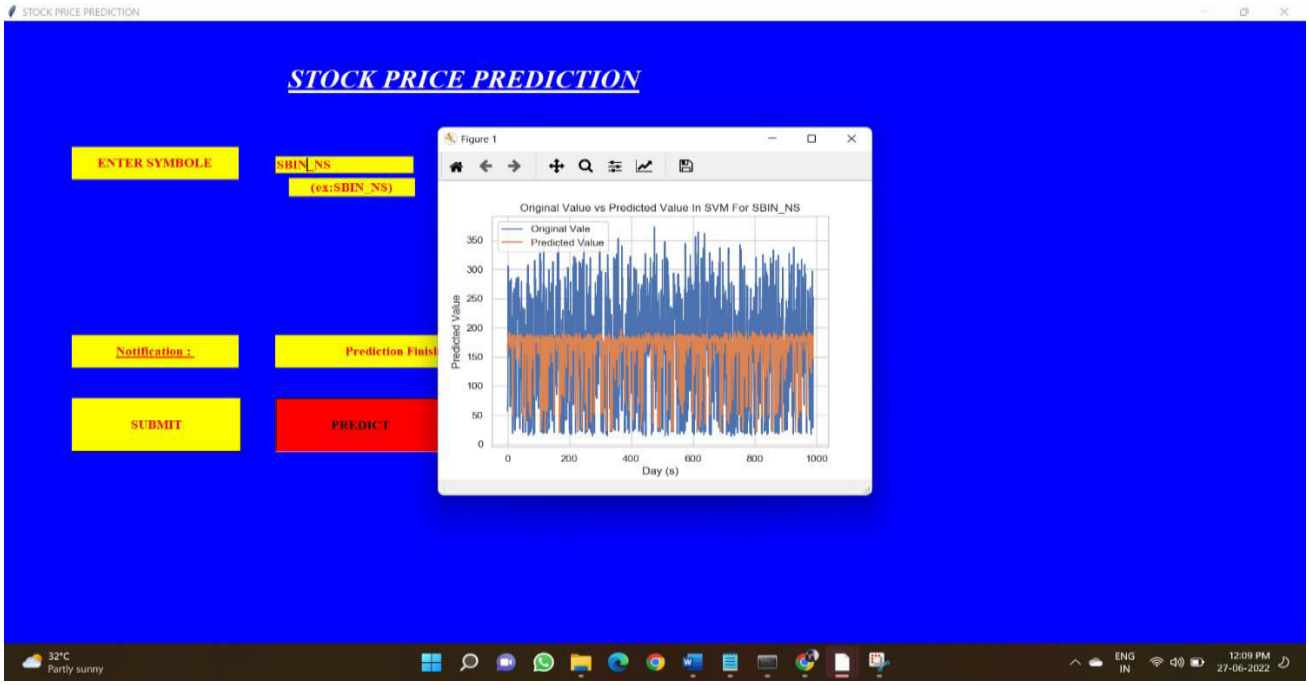

Fig No 6.2

International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402026

3. TRAINING PROCESS PAGE

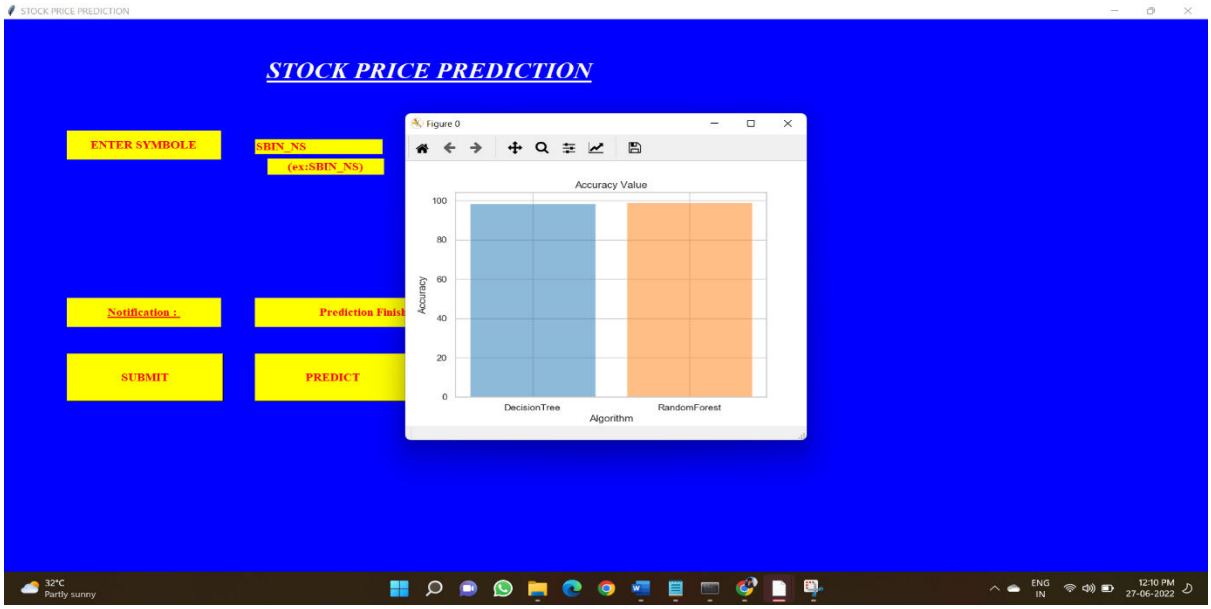

Fig No 6.3

4. FORECAST PAGE

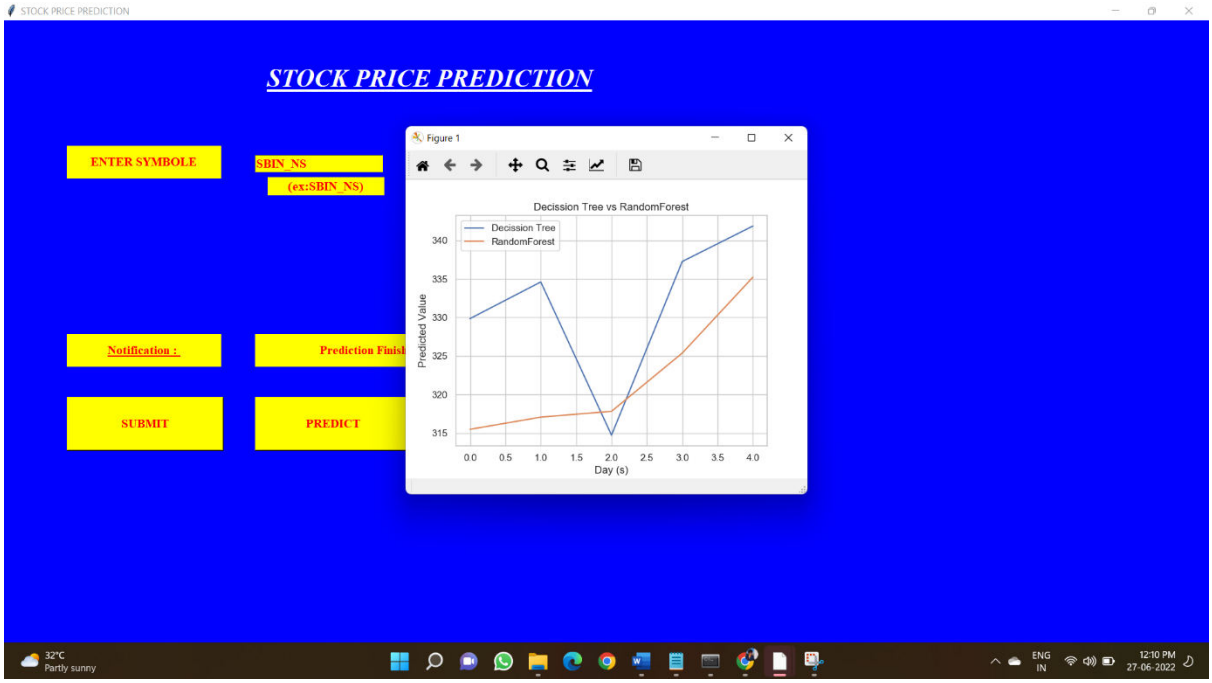

Fig No 6.4

VII. CONCLUSION

The “Stock Market Prediction Using Python” project was developed to analyse and predict stock market trends using machine learning techniques. The system mainly focuses on reducing the risk faced by investors while making investment decisions in the stock market. By using historical stock market data and machine learning algorithms such as Random Forest and Support Vector Machine, the project provides an effective method for predicting future stock

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402026

prices and market trends. The system collects stock data from reliable sources, Pre processes the dataset, and applies advanced prediction techniques to improve forecasting accuracy.

The Random Forest algorithm helps in improving prediction accuracy by creating multiple decision trees and combining their outputs, while the Support Vector Machine algorithm classifies stock market data by identifying the best hyperplane for prediction. These machine learning techniques make the system more reliable and efficient compared to traditional prediction methods. The forecasting and visualization module also helps users understand stock market behaviour through graphs, charts, and prediction outputs. The graphical interface makes the system easy to use and provides clear results for analysis.

The system provides a strong foundation for future improvements and advanced research in stock market analysis. In the future, additional machine learning and deep learning techniques can be integrated to further improve prediction accuracy and system performance.

VIII. FUTURE ENHANCEMENT

Deep Learning techniques such as LSTM and Neural Networks can be integrated to improve prediction accuracy further. Real-time stock market data can be incorporated for live prediction and analysis. Additional technical indicators and financial parameters can be included for better forecasting performance. The system can be enhanced with a user-friendly web or mobile application for investors. Hybrid machine learning models can be developed to achieve more reliable results. News analysis and social media sentiment can be combined with machine learning models for advanced prediction. Cloud-based deployment can be implemented for scalability and faster processing.

REFERENCES

[1] Machine Learning, McGraw-Hill Education, 1997.
[2] Leo Breiman, “Random Forests,” Machine Learning Journal, Vol. 45, No. 1, pp. 5–32, 2001.
[3] Corinna Cortes and Vladimir Vapnik, “Support-Vector Networks,” Machine Learning Journal, 1995.
[4] National Stock Exchange of India historical stock market dataset.
[5] Machine Learning techniques for stock market prediction research papers and journals.
[6] Pattern Recognition and Machine Learning, Springer Publication, 2006.
[7] Trevor Hastie, Robert Tibshirani and Jerome Friedman, “The Elements of Statistical Learning,” Springer, 2009.
[8] Python documentation and related machine learning libraries.
[9] Data Mining concepts and predictive analytics techniques for financial forecasting.
[10] Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O’Reilly Media, 2019.

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

International Journal of Multidisciplinary and Scientific Emerging Research (IJMSERH)

Impact Factor: 9.274

editor@ijmserh.com www.ijmserh.com