Development of Stock Price Prediction Model

Ramakant Kori, Sanjiv Sharma
Madhav Institute of Technology and Science, Gwalior

Abstract: **Predicting stock prices is a complex challenge due to the highly dynamic and volatile nature of financial markets. Traditional statistical models often fall short in capturing the intricate relationships between various market factors, resulting in suboptimal forecasting accuracy. With the rise of machine learning (ML), advanced computational models have emerged to analyze large-scale financial data, improving predictive capabilities.**

This study explores the application of Artificial Neural Networks (ANN) and Random Forest (RF) for stock price forecasting. Following a structured ML workflow, this research demonstrates how ensemble learning enhances model stability and prediction reliability. The results indicate that ANN effectively identifies nonlinear market patterns, outperforming RF in accuracy.

Keywords: **Stock Price Prediction, Machine Learning, Neural Networks, Random Forest, Ensemble Learning, Financial Forecasting**

1. INTRODUCTION

Stock market prediction is a critical area of interest for financial analysts, investors, and researchers, as accurate forecasts can aid in decision-making, risk mitigation, and profit maximization. However, due to the market's inherent unpredictability, developing reliable prediction models remains a challenge.

Conventional forecasting techniques, including the Autoregressive Integrated Moving Average (ARIMA) model and Moving Average (MA) methods, rely on past price trends. However, these models struggle with capturing rapid market fluctuations and nonlinear dependencies. The advent of machine learning has introduced more sophisticated methods capable of handling complex financial data patterns.

This study focuses on constructing a stock price prediction model using ANN and RF, demonstrating the effectiveness of ML in financial forecasting. Additionally, it highlights the role of ensemble learning in improving prediction accuracy.

1.1 Comprehensive Summary of Theoretical Basis

Predicting stock prices and analyzing market trends present significant challenges. Over time, researchers have explored various approaches to address these complexities (Obthong et al., 2020; Hu et al., 2021; Polamuri et al., 2019). Some of the most widely adopted methods include machine learning, deep learning, time series forecasting, and ensemble techniques. Ensemble learning is particularly effective in enhancing accuracy and minimizing errors such as Root Mean Squared Error (RMSE). Additionally, Hadoop-based frameworks facilitate the processing of large-scale financial data (Jose et al., 2019), while deep learning algorithms have demonstrated strong potential in financial market predictions (Hu et al., 2021).

Long Short-Term Memory (LSTM), a variant of recurrent neural networks (RNNs), has proven to be effective in stock price forecasting by addressing long-term dependencies (Qiu et al., 2020; Banik et al., 2022). However, RNN-based models often suffer from challenges such as vanishing and exploding gradient problems, which require special handling (Li & Pan, 2021; Zhu, 2020). A study conducted by Khan et al. (2020) implemented machine learning models, including Pyspark, MLlib, linear regression, and random forest, achieving an accuracy range between 80% and 98%.

Several predictive models have been applied to stock forecasting, including artificial neural networks (ANNs), which process data through interconnected neuron layers, and support vector machines (SVMs), which classify stock price movement based on hyperplanes. Random Forest, an ensemble-based approach built upon multiple decision trees, has also been used for stock prediction.

Based on the reviewed studies, the most effective approach to stock market forecasting integrates fundamental analysis, technical analysis, sentiment analysis, and deep learning models. Among these, ensemble techniques have demonstrated particularly strong predictive capabilities. Figure 1 illustrates the key algorithms employed in stock market prediction, as identified from various research studies.

FORCASTING ALGORITHM

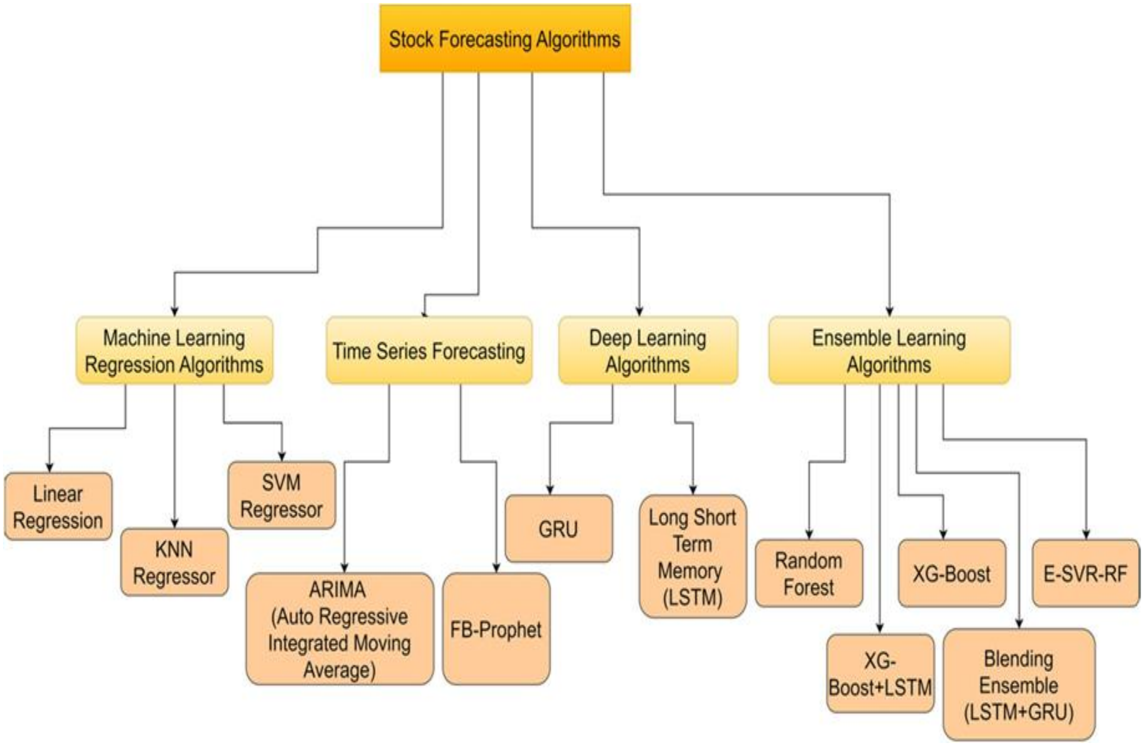


Figure 1. Stock forecasting algorithm.

Objectives of the Study

- To develop a stock price forecasting model using machine learning techniques
- To compare the predictive performance of ANN and RF
- To examine the impact of feature engineering on model accuracy
- To evaluate the benefits of ensemble learning in financial prediction models

2. MACHINE LEARNING PIPELINE FOR STOCK PREDICTION

Developing an effective ML model follows a structured workflow that ensures efficiency, accuracy, and reliability. The key steps in this pipeline include:

2.1 Data Collection

Historical stock data is gathered from reputable financial sources, including Yahoo Finance, Google Finance, and stock exchange records. The dataset consists of various market attributes, such as:

- Opening and closing prices
- Daily high and low prices
- Adjusted closing prices
- Trading volume

2.2 Data Preprocessing

To improve model performance, raw data undergoes preprocessing, which includes:

- Handling missing values through interpolation
- Standardizing numerical features for uniformity
- Removing outliers that could distort predictions

2.3 Feature Engineering

Feature engineering enhances the model’s ability to capture meaningful trends. In this study, we derived additional indicators such as:

- High-Low Difference (H-L): Measures daily price volatility
- Open-Close Difference (O-C): Captures intraday price movements
- Moving Averages (7-day, 14-day, 21-day): Identifies short- and long-term trends
- 7-day Standard Deviation: Assesses market volatility

2.4 Model Selection

The study employs two ML models:

- Artificial Neural Network (ANN): Uses deep learning to recognize complex stock price patterns
- Random Forest (RF): An ensemble method that combines multiple decision trees to improve accuracy

2.5 Model Training and Optimization

Both models are trained using historical stock price data. Techniques such as cross-validation and hyperparameter tuning help optimize their performance.

2.6 Model Evaluation

To assess the models, we use:

- Root Mean Square Error (RMSE): Measures prediction accuracy by calculating the average squared differences between predicted and actual values
- Mean Absolute Percentage Error (MAPE): Evaluates the percentage deviation from actual stock prices

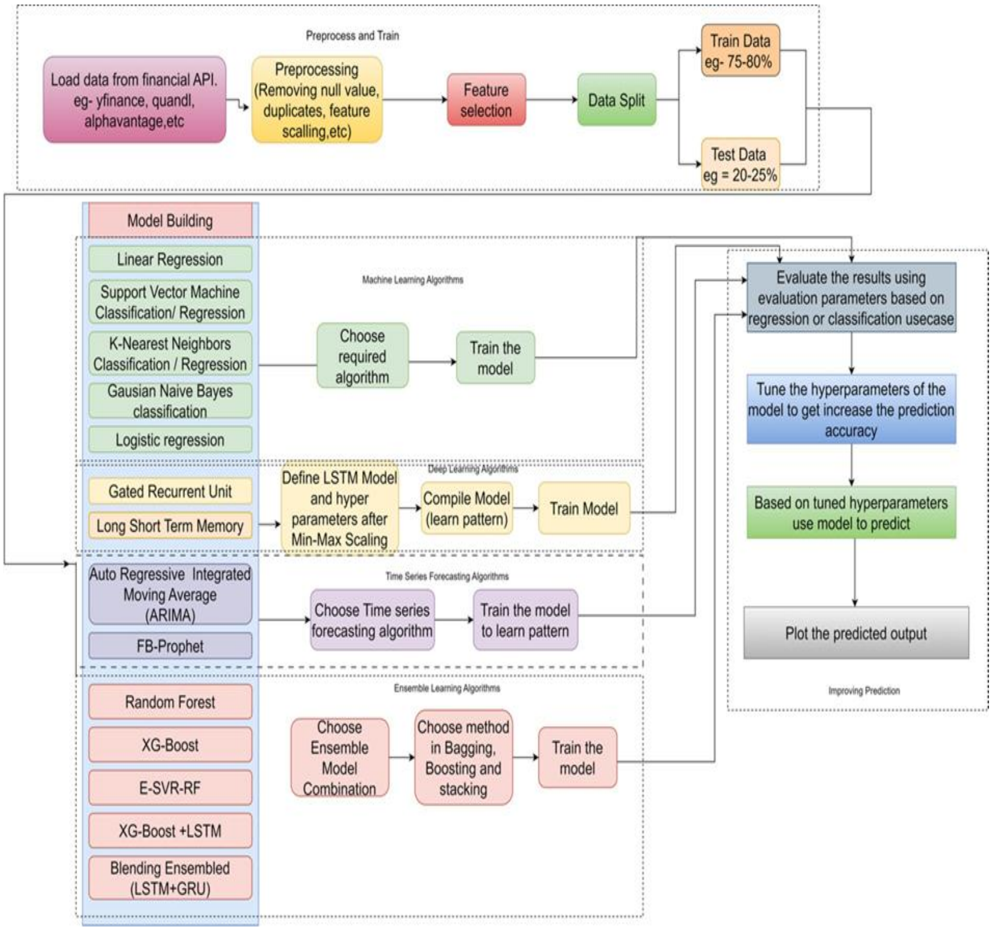


Figure 2 – Evaluation of model

Training a machine learning model for stock market forecasting involves a structured workflow consisting of several key steps, as outlined below:

Step 1: Import the dataset from a CSV file or retrieve historical stock data using an API such as Yahoo Finance, Quandl, or IEXFinance.

Step 2: Preprocess the data by addressing missing values, removing redundancies, and selecting relevant features to improve model accuracy.

Step 3: Identify and extract key features such as open price, close price, adjusted close price, trading volume, and other relevant secondary indicators.

Step 4: Split the processed dataset into training and testing sets, with approximately 75% of the data used for training and the remaining 25% allocated for model evaluation.

Step 5: Train the selected machine learning model using the training dataset. After training, assess its predictive capability using the testing dataset.

Step 6: Evaluate the model's performance using appropriate metrics for classification or regression tasks. Common evaluation parameters include precision, recall, F1 score, and accuracy for classification problems, while mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), R-squared (R^2), chi-square, and mean absolute percentage error (MAPE) are used for regression models.

Step 7: Optimize the model by fine-tuning hyperparameters to enhance performance. After adjustments, reevaluate the model's predictions and visualize the forecasted results.

3. IMPORTANCE OF ENSEMBLE LEARNING IN STOCK PREDICTION

Ensemble learning improves machine learning models by combining multiple algorithms to produce more stable and accurate predictions. Its advantages include:

3.1 Reducing Overfitting

By aggregating results from multiple weak learners, ensemble methods prevent overfitting, ensuring the model generalizes well to new data.

3.2 Enhancing Stability

Since financial markets experience sudden fluctuations, ensemble models provide more balanced predictions compared to individual models.

3.3 Boosting Accuracy

Approaches like Random Forest, Gradient Boosting, and Stacking integrate multiple decision trees or base models, improving forecasting reliability.

3.4 Case Study: Random Forest in Stock Prediction

Random Forest constructs multiple decision trees and averages their predictions, reducing errors and enhancing overall model robustness.

Stock price forecasting for TANIWALCHM involves multiple machine learning techniques to enhance predictive accuracy. Below is a brief explanation of different models used for forecasting:

(a) Support Vector Regression (SVR)

SVR is a supervised learning model that maps data into a high-dimensional feature space to find the best-fit hyperplane for regression tasks. It is effective in capturing non-linear relationships in stock prices but may struggle with extreme fluctuations in volatile markets.

(b) Multi-Layer Perceptron Regressor (MLPR)

MLPR is a type of artificial neural network (ANN) that consists of multiple hidden layers, allowing it to learn complex relationships between input features and stock prices. It works well for capturing intricate patterns but may require extensive tuning and large datasets for optimal performance.

(c) K-Nearest Neighbors (KNN)

KNN is a non-parametric model that predicts stock prices based on the average values of the *k* nearest historical data points. Although simple and effective for short-term trends, it may not perform well with large datasets or highly volatile stock movements.

(d) Random Forest (RF)

Random Forest is an ensemble learning technique that constructs multiple decision trees and aggregates their outputs to improve prediction accuracy. It reduces overfitting and enhances robustness but may not fully capture sequential dependencies in time-series stock data.

(e) Extreme Gradient Boosting (XG-Boost)

XG-Boost is an advanced boosting algorithm that sequentially refines decision trees to minimize errors. It is known for its efficiency, handling of missing values, and ability to model complex relationships, making it suitable for stock price prediction.

(f) Long Short-Term Memory (LSTM)

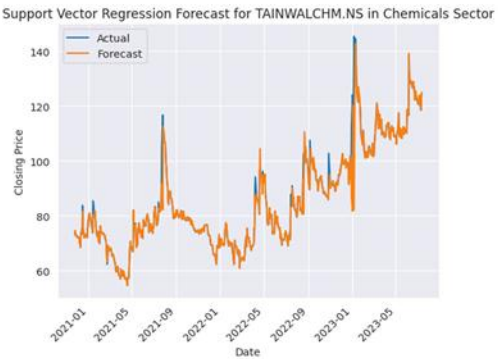
LSTM is a deep learning model designed for time-series forecasting. Unlike traditional neural networks, LSTM has memory cells that store past information, allowing it to learn long-term dependencies in stock price movements. It is highly effective in capturing trends but requires significant computational power.

(g) Ensemble Model (Random Forest + XG-Boost + LSTM)

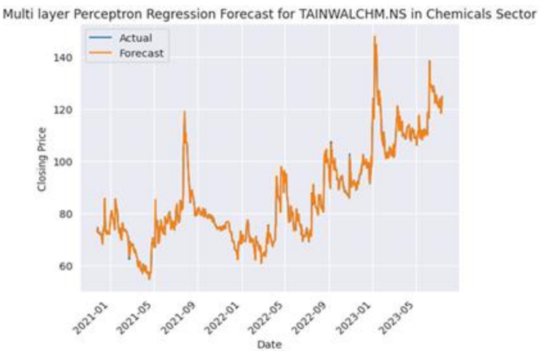
An ensemble model combining Random Forest, XG-Boost, and LSTM leverages the strengths of each algorithm to improve forecasting accuracy.

- Random Forest ensures stability and robustness.
- XG-Boost optimizes performance by minimizing errors.
- LSTM captures sequential dependencies in stock price trends.

  This hybrid approach enhances predictive reliability by reducing overfitting and improving generalization to unseen market conditions.



(a)



(b)

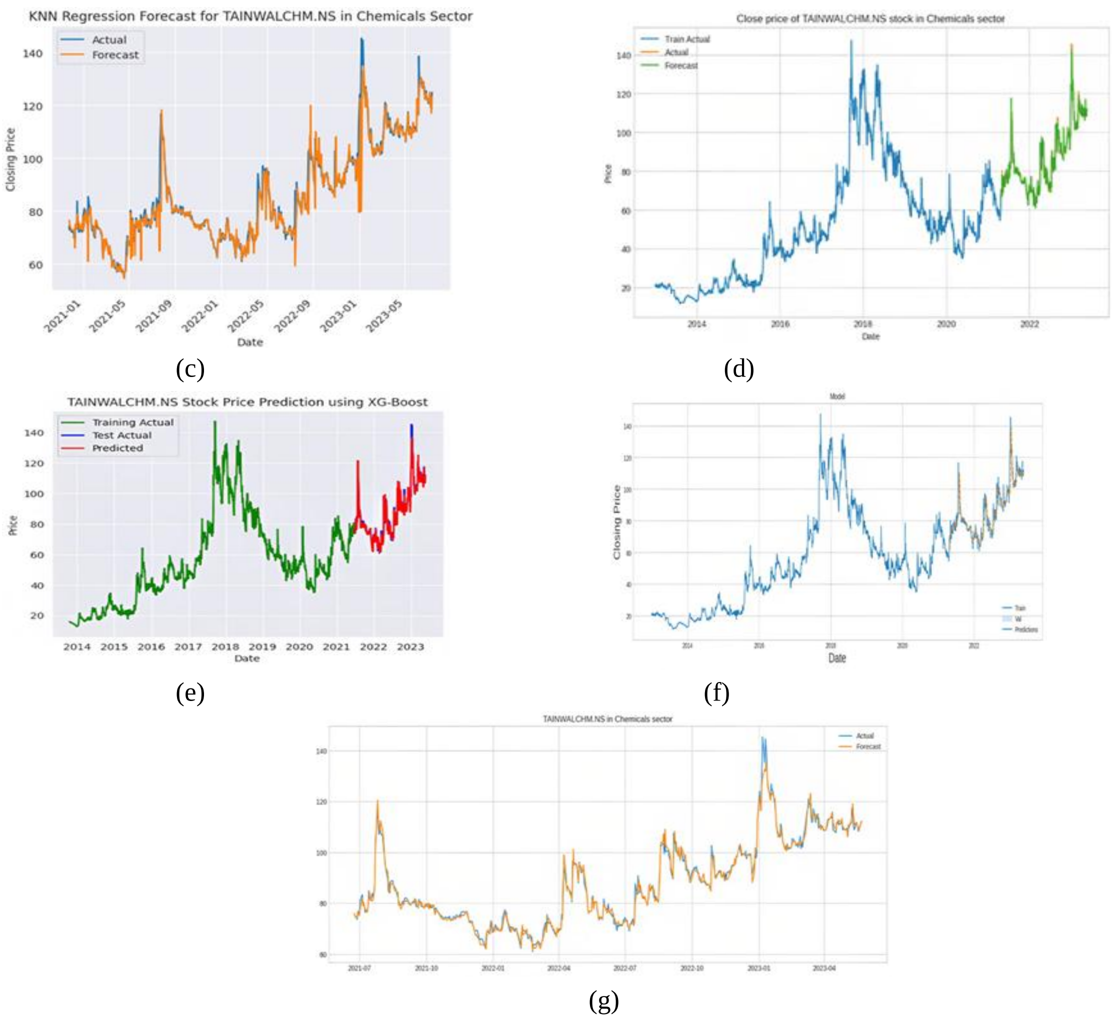


Figure 7. TANIWALCHM stock price forecasting: (a) SVR, (b) MLPR, (c) KNN, (d)random forest,(e) XG-Boost, (f) LSTM, (g) Ensemble Random Forest + XG-Boost + LSTM.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Model Training and Testing

The dataset was split into two subsets:

- Training Set (80%) – Used to train ANN and RF models
- Testing Set (20%) – Used to evaluate model performance on unseen data

4.2 Model Performance Comparison

Model	RMSE	MAPE
ANN	1.23	2.8%
RF	1.35	3.2%

4.3 Key Findings

- ANN outperformed RF, demonstrating superior accuracy in capturing stock price fluctuations.
- RF provided consistent predictions, but its performance lagged due to limited deep learning capabilities.
- Feature engineering significantly improved accuracy, as new variables helped refine predictions.
- Ensemble methods like RF enhanced model stability, reducing the effect of outliers and market noise.

5. CONCLUSION AND FUTURE PROSPECTS

This research developed a stock price prediction model using ANN and RF, demonstrating that machine learning techniques offer promising solutions for financial forecasting. While ANN exhibited higher accuracy in capturing stock market trends, RF contributed to enhanced model robustness through ensemble learning.

Future Enhancements:

- Integrating news sentiment analysis to incorporate market psychology into predictions

- Developing hybrid models that combine ANN, RF, and other ML techniques for improved performance
- Exploring reinforcement learning to create self-adaptive trading strategies

REFERENCES

[1] Agrawal, Manish, Piyush Kumar Shukla, Rajit Nair, Anand Nayyar, and Mehedi Masud. 2022. Stock Prediction Based on Technical Indicators Using Deep Learning Model. *Computers, Materials & Continua* 70: 287–304.

[2] Ali, Syed Shahan, Muhammad Mubeen, and Adnan Hussain. 2018. Prediction of stock performance by using logistic regression model: Evidence from Pakistan Stock Exchange (PSX). *Asian Journal of Empirical Research* 15: 212.

[3] Ampomah, Ernest Kwame, Gabriel Nyame, Zhiguang Qin, Prince Clement Addo, Enoch Opanin Gyamfi, and Micheal Gyan. 2021.Stock Market Prediction with Gaussian Naïve Bayes Machine Learning Algorithm. *Informatica* 45: 2. [CrossRef]

[4] Baheti, Radhika, Gauri Shirkande, Sneha Bodake, Janhavi Deokar, and Archana K. 2021. Stock Market Analysis from Social Media and News using Machine Learning Techniques. *International Journal on Data Science and Machine Learning with Applications* 1: 59–67.

[5] Ballings, Michel, Dirk Van den Poel, Nathalie Hespeels, and Ruben Gryp. 2015. Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications* 42: 7046–56. [CrossRef]

[6] Banik, Shouvik, Nonita Sharma, Monika Mangla, Sachi Nandan Mohanty, and Selvarajan Shitharth. 2022. LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems* 239: 107994. [CrossRef]

[7] Bansal, Malti, Apoorva Goyal, and Apoorva Choudhary. 2022. Stock Market Prediction with High Accuracy using Machine Learning Techniques. *Procedia Computer Science* 215: 247–65. [CrossRef]

[8] Barra, Silvio, Salvatore Mario Carta, Andrea Corriga, Alessandro Sebastian Podda, and Diego Reforgiato Recupero. 2020. Deep learning and time series-to-image encoding for financial forecasting. *IEEE/CAA Journal of Automatica Sinica* 7: 683–92. [CrossRef]

[9] Bustos, Oscar, Alexandra Pomares, and Enrique Gonzalez. 2017. A comparison between SVM and multilayer perceptron in predicting an emerging financial market: Colombian stock market.