SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

THE COMPARATIVE EFFECTIVENESS OF MACHINE LEARNING MODELS FOR STOCK PRICE PREDICTION

Yamuna Chandrashekhar Mane[1], Shraddha Sanjay Chougule[2]
[1]*M.Sc Data Science-II*
[2]*M.Sc Data Science-II*

ABSTRACT
Predicting stock prices is a key challenge in finance, influencing decisions made by investors and traders. Traditional forecasting methods often struggle to capture the complex and unpredictable nature of stock market movements. This study explores how machine learning models—Random Forest (RF), Support Vector Machines (SVM), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU)—perform in stock price prediction. The models were evaluated based on accuracy, interpretability, and computational efficiency.

Using Tata Motors' stock price data from 2014 to 2024 (sourced from Yahoo Finance), we applied feature engineering techniques such as moving averages, Relative Strength Index (RSI), and Bollinger Bands to improve predictions. Data preprocessing, including handling missing values and normalization, was carried out to enhance model performance.

Our findings show that Random Forest delivered the best results, with the lowest RMSE (0.23), making it the most effective model for stock price forecasting. SVM performed moderately well but fell short in accuracy. LSTM and GRU had higher error rates, suggesting they require extensive fine-tuning and larger datasets for optimal performance.

Overall, Random Forest proved to be the most reliable and interpretable model for stock prediction. Future research can explore hybrid approaches and integrate sentiment analysis from financial news and social media to further improve prediction accuracy.
KEYWORDS: *Stock price prediction, Machine learning, Random Forest, LSTM, SVM, GRU, Financial forecasting.*

INTRODUCTION

In recent years, data science and machine learning have significantly reshaped the landscape of stock market analysis, offering new approaches to understanding complex market behaviour. The stock market, inherently volatile and influenced by a multitude of factors, generates massive amounts of data every day, from real time price changes and trading volumes to corporate announcements and economic indicators. Leveraging this data effectively has become essential for investors seeking a competitive advantage, and machine learning provides powerful tools to discern patterns and make informed predictions.

This research investigates the application of machine learning models, specifically Random Forest, Support Vector Machines (SVM), and neural networks such as LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Units), to predict stock price movements. These algorithms are increasingly recognized for their potential to capture the non-linear, high-dimensional relationships that characterize financial data. Traditional approaches, like technical and fundamental analysis, often fall short in managing the intricacies of modern financial data. Machine learning models, however, can process vast amounts of information, capturing subtle patterns and trends that may not be immediately apparent through conventional methods.

Among these techniques, Random Forest, a robust ensemble learning model, has shown promise in stock market prediction due to its ability to handle large datasets, reduce overfitting, and generate stable predictions by averaging multiple decision trees. This study will compare Random Forest's effectiveness with that of SVM and neural networks, examining their accuracy, precision, and other performance metrics in forecasting stock prices.

Furthermore, by leveraging historical price data and key market indicators, the study aims to develop predictive models that not only outperform traditional methods but also provide a data-driven foundation for investment decisions. In addition to machine learning, this research incorporates statistical analysis techniques to improve feature selection and data preprocessing. Methods such as trend analysis and volatility measures can help refine model inputs, guiding the selection of relevant features that enhance model interpretability and performance. By combining machine learning with statistical analysis, this research aims to provide a comprehensive, multifaceted approach to stock price prediction, allowing for better model accuracy and more nuanced insights into market dynamics.

OBJECTIVES

Objective 1: Evaluate the Predictive Performance of Multiple Machine Learning Models for Stock Price Prediction

- This objective involves implementing and comparing several machine learning algorithms, namely Random Forest, Support Vector Machine (SVM), and Neural Networks (LSTM/GRU). The goal is to assess each model's predictive accuracy and ability to generalize on historical stock price data, focusing on performance indicators such as accuracy, precision, recall, and F1 score. This comparison will provide insights into the strengths and weaknesses of each algorithm in the context of stock price prediction

Objective 2: Analyze the Impact of Feature Engineering and Preprocessing on Model Accuracy

- This objective aims to study how different data preprocessing and feature engineering techniques affect the performance of machine learning models in stock prediction. This includes handling missing values, normalizing data, and selecting the most relevant features based on exploratory analysis. By systematically testing these preprocessing techniques, the research will identify optimal methods for preparing stock data to improve model accuracy and robustness.

Objective 3: Assess the Suitability of Machine Learning Models for Real Time Stock Price Prediction

- This objective evaluates each model's performance in terms of computational efficiency and real-time application potential, particularly for use in fast-paced financial markets. The aim is to determine whether these models can handle real-time data and provide predictions within a reasonable timeframe, balancing between accuracy and latency for timely decision-making.

Objective 4: Identify Model Limitations and Suggest Future Improvements for Stock Market Analysis

- This objective involves analyzing the strengths and limitations of each machine learning model when applied to stock price prediction. It includes discussing the challenges encountered, such as handling noisy or highly volatile data, and proposing future enhancements, including hyperparameter tuning and the integration of additional data sources like technical indicators or sentiment analysis, to improve predictive capabilities.

STATEMENT OF PROBLEM

The stock market is a dynamic and complex environment, heavily influenced by numerous factors, including economic indicators, company performance, investor sentiment, and global events. Predicting stock price movements within this high-dimensional and volatile domain has long been a challenge. While machine learning has emerged as a promising tool for improving stock price prediction, the application of these techniques presents challenges in achieving a balance between predictive accuracy, model interpretability, and practical applicability, especially in real-time trading environments where decision speed is crucial.

Recent advances in machine learning models, including Random Forests, Support Vector Machines (SVM), and neural networks such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have shown potential for high accuracy in stock market prediction. However, complex models like LSTM and GRU often function as "black boxes," limiting transparency and interpretability. This lack of interpretability can be problematic for traders, investors, and analysts, who require clear insights into how and why predictions are made to make informed decisions in high-stakes financial markets.

Among machine learning techniques, decision tree-based models, particularly Random Forests, offer a valuable trade-off between predictive performance and interpretability. Random Forests aggregate multiple decision trees, reducing overfitting and enhancing robustness while providing a clearer reasoning path for each prediction. However, while decision trees are widely used in general machine learning tasks, their performance in the stock market context, especially when combined with statistical techniques, remains underexplored.

Statistical methods, traditionally employed in financial analysis, provide valuable insights into market patterns and relationships through techniques such as time series analysis, correlation measures, and volatility assessments. These methods are transparent and widely accepted within finance, yet they may lack predictive power when used alone, particularly in high-dimensional, non-linear environments like the stock market. This research presents an opportunity to explore the integration of statistical techniques with machine learning models like Random Forests to enhance prediction accuracy while retaining interpretability.

Thus, this research aims to address the gap between highly accurate but complex models and more interpretable but potentially less powerful ones. The study will evaluate the performance of decision tree-based models, specifically Random Forest, and assess the impact of integrating statistical techniques for feature selection and preprocessing. This research also compares these models to other machine learning approaches, examining their predictive accuracy, interpretability, and practical usability in stock market prediction. By investigating the combined performance of Random Forest and statistical methods, this study seeks to determine if a balance between accuracy, usability, and interpretability can be achieved for real-time stock price prediction. The findings will offer

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

valuable insights into the role of decision tree-based models and statistical methods in financial market analysis, potentially benefiting researchers, traders, and investors who require reliable, interpretable models for informed decision-making in dynamic financial markets.

HYPOTHESIS

The study is based on the following hypotheses:

- **H1:** Machine learning models outperform traditional statistical methods in predicting stock prices.
- **H2:** Random Forest provides better interpretability compared to deep learning models while maintaining competitive accuracy.
- **H3:** Data preprocessing techniques significantly improve the performance of stock price prediction models.
- **H4:** Neural Networks (LSTM, GRU) excel in capturing sequential dependencies in stock price data, but their accuracy depends on hyperparameter tuning and dataset quality**.**

METHODOLOGY

Data Collection

The dataset used in this study consists of historical stock prices of Tata Motors, obtained from Yahoo Finance. The dataset includes key financial metrics such as:

- Date: Trading day
- Open: Stock's opening price
- High/Low: Daily price range
- Close: Closing price
- Adjusted Close: Adjusted closing price after corporate actions
- Volume: Number of shares traded

Data Preprocessing

- **Handling Missing Values:** Missing data is filled using interpolation methods to maintain dataset integrity.
- **Feature Engineering:** Moving averages, volatility measures, and sentiment analysis are incorporated.
- **Data Normalization:** Min-Max scaling is applied to improve model convergence.
- **Train-Test Split:** 80% of data is used for training and 20% for testing.

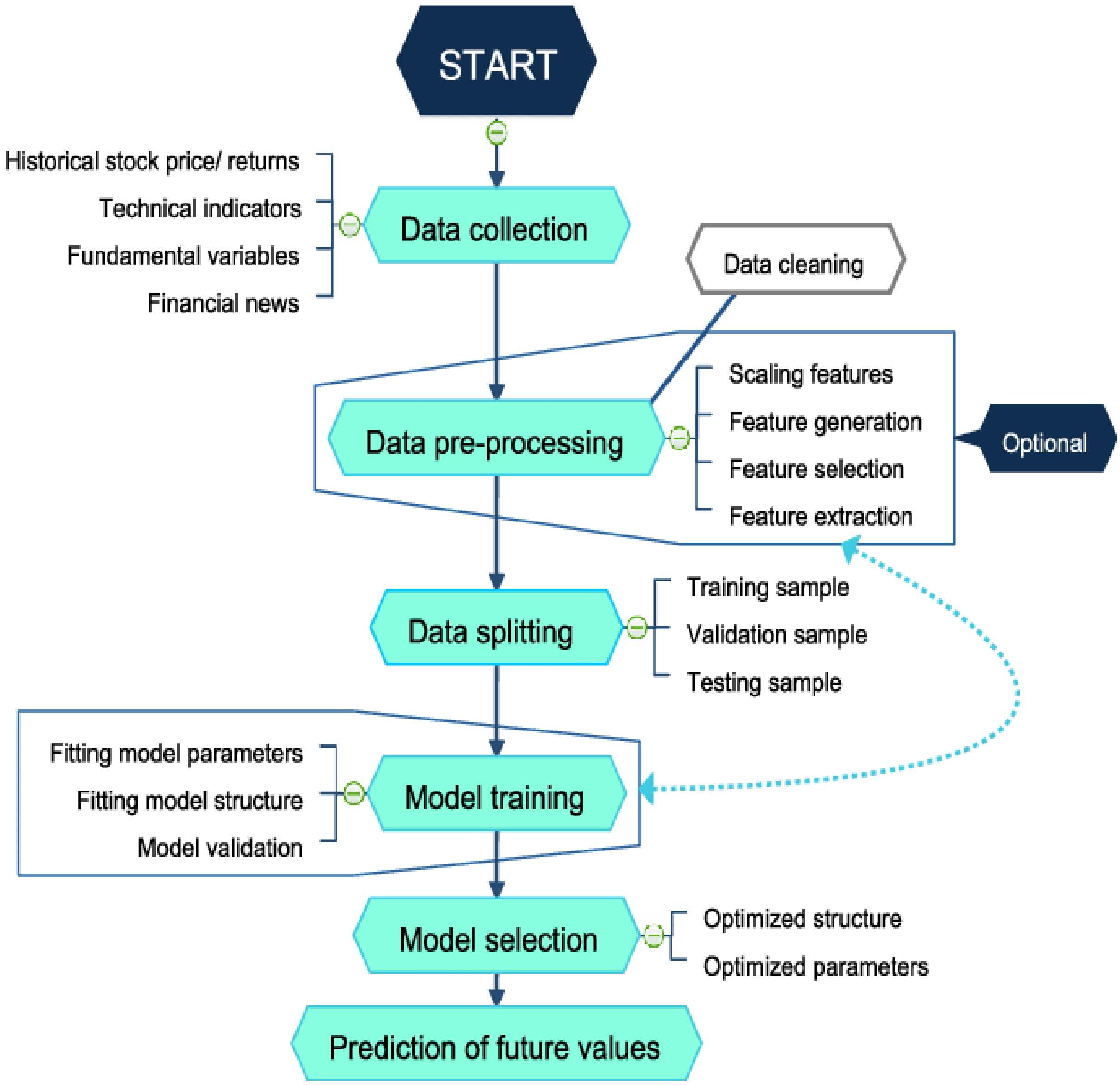
Machine Learning Models Implemented

- **Random Forest:** An ensemble model that averages multiple decision trees to reduce overfitting.
- **Support Vector Machine (SVM):** A supervised learning model using hyperplanes to classify stock price trends.
- **LSTM (Long Short-Term Memory):** A deep learning model specialized in capturing temporal dependencies in financial data.
- **GRU (Gated Recurrent Unit):** Similar to LSTM but computationally efficient, used for sequential forecasting.

Model Development

- **Random Forest:** A decision tree-based ensemble method, initialized with 100 trees.
- **SVM:** Uses an RBF kernel to capture non-linear relationships.
- **LSTM & GRU:** Deep learning models designed for sequential time-series prediction.
- **Evaluation Metric:** Root Mean Squared Error (RMSE).
  Each model is trained and evaluated using *Root Mean Squared Error (RMSE)* and other performance metrics.

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal



Data Analysis
The models are evaluated based on RMSE, with lower values indicating better accuracy.

RESULTS
- **Random Forest:** RMSE = 0.23 (highest accuracy)
- **SVM:** RMSE = 0.37
- **LSTM:** RMSE = 40.84
- **GRU:** RMSE = 48.73

FINDINGS
- Random Forest outperformed deep learning models, likely due to its ability to handle structured, non-sequential data.
- LSTM and GRU struggled with noise and required additional tuning.
- SVM performed reasonably well but lacked the flexibility of ensemble models.

SUGGESTIONS
1. Enhancing Deep Learning Performance: Incorporating sentiment analysis and macroeconomic indicators may improve LSTM and GRU accuracy.
2. Hybrid Models: Combining Random Forest with LSTM could leverage the strengths of both approaches.

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

3.Risk Management Strategies: Models should include volatility predictions to guide investment decisions.
4.Live Market Testing: Future studies should evaluate models under real-time trading conditions.
5.Incorporating Alternative Data: Social media sentiment, news analysis, and global economic trends can refine predictive models.

CONCLUSION

This research demonstrates that machine learning models can significantly improve stock price prediction accuracy. Among the models tested, Random Forest outperformed SVM and deep learning approaches, highlighting its robustness in financial forecasting. While LSTM and GRU offer potential advantages in time-series forecasting, their current performance requires further optimization. Future work should focus on hybrid modeling techniques and real-time deployment to enhance predictive performance.

REFERENCES

1. *Zhang, Y., et al. (2021). "Predicting Stock Prices: A Machine Learning Approach."*
2. *Fischer, T., et al. (2018). "A Deep Learning Framework for Financial Time Series Using LSTM and GRU."*
3. *Ansari, M. A. B., et al. (2020). "Utilizing Machine Learning for Stock Market Prediction: A Comprehensive Review."*
4. *Choudhury, A., et al. (2021). "Stock Price Prediction Using Machine Learning: A Survey."*
5. *Singh, R., et al. (2021). "Machine Learning in Stock Market Prediction: A Review."*

This paper provides a structured approach to stock market prediction, helping both researchers and investors make data-driven decisions based on machine learning insights.