iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69326

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Comparative Study of Different Machine Learning Methodology Used for Stock Price Prediction

Prof. Junaid Mandviwala[1], Shivam Samir Shere[2], Krishnakant Narsing Choudhary[3], Sohail Sabir Ansari[4], Saif Azhar Ansari[5]
Department of Artificial Intelligence & Data Science, Rizvi College of Engineering, Mumbai, Maharashtra, India

Abstract: Stock price prediction has been a topic of interest for investors and researchers for a long time because of the monetary returns that can be achieved using correct forecasts. Over the last few years, machine learning (ML) approaches have become increasingly popular for modelling non-linear and complex financial time series data. This article offers a comparative study of three highly acclaimed machine learning approaches Long-Short-Term-Memory (LSTM) networks, Support Vector Machines (SVM), and Random Forests (RF) for forecasting stock prices.

LSTM, an RNN variant, is designed to recognize long-term dependence in sequential information and thus becomes particularly appropriate to model time series trends. SVM, a versatile supervised learning approach, is characterized by its capacity to handle high-dimensional spaces with ease and express non-linear dependencies via kernel maps. Random Forest, an approach to ensemble-based learning, predicts multiple decision tree outputs to circumvent overfitting and provide better generalizability.

Historical stock prices are employed to train and test each model in this research. Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) Score are used to measure prediction accuracy. The findings show that all three models excel in various contexts, but LSTM performs better in identifying temporal patterns, SVM is good with smaller datasets and evident margins, and Random Forest provides robustness and interpretability.

This comparative study sheds light on the strengths and weaknesses of each method, helping practitioners and researchers choose the most suitable model for stock market forecasting applications.

Keywords: Stock Price Prediction, Machine Learning, LSTM, Support Vector Machine (SVM), Random Forest, Time Series Forecasting, Financial Data, Model Comparison, Deep Learning, Predictive Analytics

I. INTRODUCTION

The stock market is a dynamic and intricate financial system that is critical to the world economy. Stock price prediction has always been a difficult task because of the high volatility, non-linearity, and impact of various external factors like market trends, economic indicators, and investor sentiment. Correct stock price prediction can yield great benefits for investors, traders, and financial analysts by facilitating better decision-making and risk management.

Over the past few years, Machine Learning (ML) has evolved as an efficient tool for modelling and forecasting stock market behaviour. ML algorithms are capable of detecting subtle patterns in historical data and making data-driven forecasts with greater accuracy than conventional statistical techniques. Three methodologies among different ML techniques have demonstrated favorable results in time series prediction and financial forecasting tasks: Long Short-Term Memory (LSTM) networks, Support Vector Machines (SVM), and Random Forest (RF).

LSTM is a Recurrent Neural Network (RNN) specifically tailored to learn long-term dependencies in sequential data. Its memory cell structure makes it especially appropriate for modelling time-dependent data like stock prices.

SVM is a supervised learning algorithm that performs well in high-dimensional spaces and does well for regression tasks using kernel-based learning.

Random Forest, a decision tree-based ensemble technique, offers strength, deals efficiently with non-linearity, and suppresses overfitting through averaging the responses of several trees.

This project will compare these three methods based on historical stock price data. The models are compared using standard performance measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) Score. By comparing these methods, the research will determine their strengths, weaknesses, and applicability to stock price prediction tasks.

The results of this project will advance the knowledge of machine learning usage in financial markets and offer insightful information for researchers, data scientists, and market participants who are interested in predictive modelling.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

II. LITERATURE REVIEW

1) *A Random Forest Based Ensemble Learning Model for Stock Price Prediction Computers & Operations Research*

The complexity and volatility of the stock market have fueled a great deal of research into predictive models, with machine learning methods such as Long-Short-Term Memory (LSTM) networks, Support Vector Machines (SVM), and Random Forests (RF) proving highly promising. LSTMs are particularly good at modelling long-term dependencies and extracting hidden patterns in sequential data, although they need large amounts of data and computational power. SVMs, which can deal with both linear and nonlinear relationships using kernel functions, are efficient for financial prediction but hyperparameter-sensitive and less scalable for big data. RFs, being robust and interpretable, can efficiently deal with high-dimensional data and offer feature importance insights, reducing overfitting and capturing complex relationships between predictors.

2) *Deep Learning for Stock Price Prediction. European Journal of Operational Research*

The financial market's complex behavior has fueled multidimensional studies in stock price forecasting with techniques such as Long Short-Term Memory (LSTM) networks, Support Vector Machines (SVM), and Random Forests (RF). LSTMs, a form of Recurrent Neural Network presented in 1997, are capable of identifying long-term dependencies and nonlinearities in time series but need huge datasets and extensive computational power. SVMs, established in 1995, are excellent for managing high-dimensional data and nonlinear connections through kernel functions, although they are computationally cumbersome and susceptible to parameter choices. RF, an ensemble learning approach brought about in 2001, combines decision trees to provide stable predictions and information about feature significance but can be heavy in computational resources while scaling up with numerous trees. Both techniques possess both weaknesses and strengths, and they are all beneficial for various financial forecasting purposes.

3) *Long-Short-Term-Memory Algorithm Based Prediction of Stock Market Exchange*

Prediction of stock prices has been a complex but inevitable work for researchers, analysts, and investors since the financial markets have been highly unpredictable and dynamic. Effective forecasting models are imperative for decision-making guidance, risk mitigation, and profitability maximization. This study focuses on utilizing cutting-edge machine learning methods, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Artificial Neural Networks (ANNs), to develop a strong and stable platform for forecasting stock prices. Although classical RNNs are poor at handling long-term dependencies because of vanishing gradients, LSTM networks overcome these shortcomings using memory cells to capture and make use of temporal patterns efficiently. Also, ANNs perform well in acquiring nonlinear relationships between stock data and, when coupled with sequential models, further increase predictive accuracy. Through combining LSTM, ANN, and Sequential models, this work establishes a scalable and adaptable system that is suited to the complexities of stock price prediction. The modular structure of the Sequential model provides for specialized architectures, supporting the easy integration of these robust algorithms. Combined, these models represent a reliable process for examining past information and forecasting future stock patterns, providing valuable information to investors for sound decision-making. This integrated strategy not only improves the precision and dependability of stock price forecasts but also opens the door for more streamlined investment procedures, meeting the demands of a fast-changing financial world.

III. PROPOSED SYSTEM

This research focuses on addressing the challenge of accurately forecasting stock prices, a task that is critical for investors and financial professionals in managing risks and optimizing investment decisions. The proposed solution is a unified system employing advanced machine learning methodologies Long Short-Term Memory (LSTM) networks, Support Vector Machines (SVM), and Random Forest (RF) to enhance prediction accuracy. The core of the system lies in its structured approach to building and training predictive models for stock price forecasting, leveraging the strengths of each method. LSTMs are designed to handle sequential data and long-term dependencies, making them particularly effective for capturing trends in time-series data. SVMs are robust in modeling both linear and nonlinear relationships in high-dimensional datasets, while RF excels in managing complex interactions among features and providing interpretability through feature importance. The procedure is to split the data into train and test sets, followed by feature scaling in order to scale the data and make it compatible with models. Separate models for LSTM, SVM, and RF are developed after that, of which LSTM necessitates the preparation of sequential train data, while SVM and RF are trained using suitable hyperparameters. Models are trained from processed data and saved for use, with provision for reusability and scalability. Predictions are done based on test data, and results are inverse-scaled if required and then checked against real values using metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Actual vs. predicted stock prices are finally visualized to check the performance of the model and check for trends.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

By integrating LSTM, SVM, and RF into a cohesive system, the proposed approach capitalizes on the unique strengths of each algorithm to deliver robust, scalable, and accurate stock price predictions. This unified methodology offers a significant advancement in financial forecasting, empowering investors and analysts to make informed decisions in dynamic market environments.

IV. SYSTEM DESIGN AND IMPLEMENTATION

A. System Design Algorithm

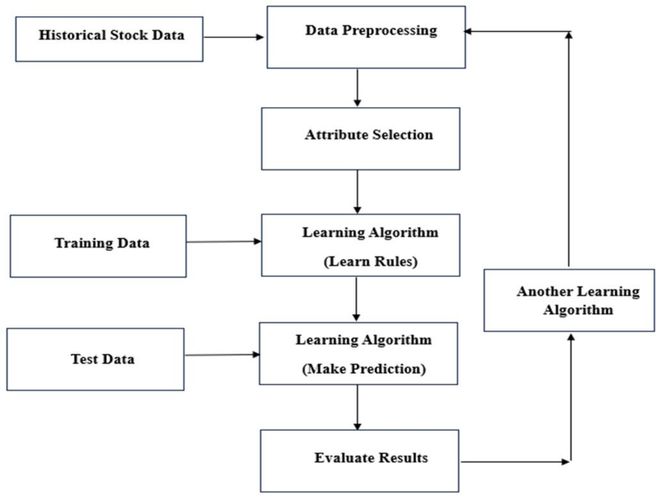

Fig. 4.1. System Design Algorithm

B. Model Implementations

1) Long-Short-Term-Memory (LSTM) Model

LSTM is a specific variant of RNN that can learn from sequential data. In particular, LSTM was designed to solve the problem of long dependencies within the sequence. It differs from traditional RNNs, which are very susceptible to vanishing gradients during training in comparison to other methods. For this reason, the LSTM architecture was designed using memory cells and gating mechanisms (input, forget, and output gates) in order to hold and update information through time.

The Implementation of LSTM Model:

- Data Preparation and Preprocessing: Historical stock data is collected, cleaned, formatted, and explored for trends. The data is split into training and testing sets and normalized using Min-Max Scaler. Training data is reshaped into a 3D format suitable for LSTM.
- Model Development and Training: An LSTM model is built with input, forget, and output gates. The model is trained on the prepared training data and saved for reuse. Testing data is prepared similarly, enabling consistent predictions.
- Prediction, Evaluation, and Visualization: The trained model predicts stock prices, and predictions are inverse-scaled to their original range. Model performance is evaluated using MSE and RMSE, and actual vs predicted prices are plotted to visualize trends and accuracy.

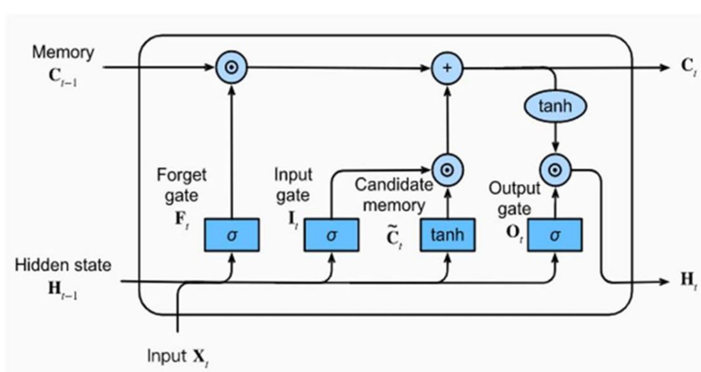

Fig.4.2.1. LSTM Architecture

2) Support Vector Machine (SVM) Model

SVM is a type of supervised learning that may be used for classification as well as regression problems. And in the case of stock price prediction, a variation known as Support Vector Regression (SVR) is employed, which attempts to identify the best hyperplane in a way that will minimize the error with some margin of tolerance like epsilon at the same time.

The Implementation of SVM Model:

- Data Preparation and Preprocessing: Collect historical data, clean and preprocess it into a structured format, explore patterns through visualization, split it into training and testing datasets, and scale the features for SVM compatibility.
- Model Development and Training: Build an SVM model with suitable kernel settings, train it on the processed training data, and save the trained model for reuse. Prepare the testing data similarly to ensure consistency during prediction.
- Prediction, Evaluation, and Visualization: Use the trained SVM model to make predictions, evaluate its performance with metrics like accuracy and error rates, and plot actual vs predicted values to assess model accuracy and trends.

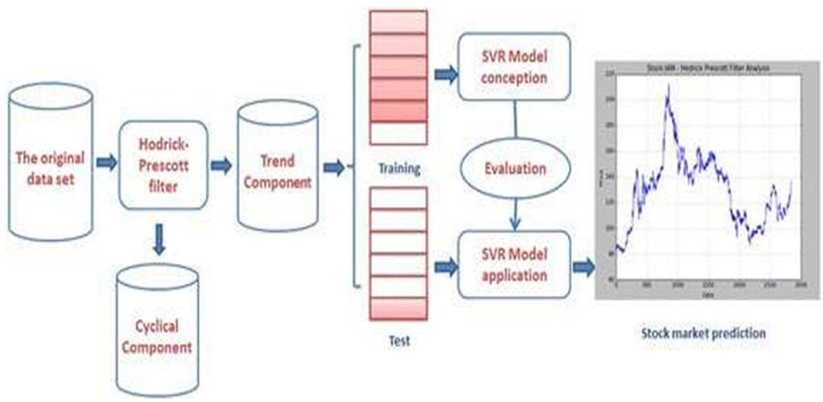

Fig.4.2.2. SVM Architecture

3) Random Forest (RF) Model

Random Forests is a type of ensemble learning technique that, during training, builds up many decision trees and then uses these to classify or regress. It thus proves very effective in handling large data sets, containing thousands of features in each.

The Implementation of RF Model:

- Data Preparation and Preprocessing: Collect and preprocess historical data by cleaning, exploring patterns through visualization, splitting it into training and testing datasets, and scaling the features as necessary for compatibility with the Random Forest model.
- Model Development and Training: Build a Random Forest model with optimized parameters, train it using the prepared training data, and save the trained model for future use. Ensure the testing data is formatted consistently for accurate predictions.
- Prediction, Evaluation, and Visualization: Use the trained model to predict outcomes, evaluate its performance with metrics like accuracy and error rates, and visualize actual vs predicted results to assess model effectiveness and identify trends.

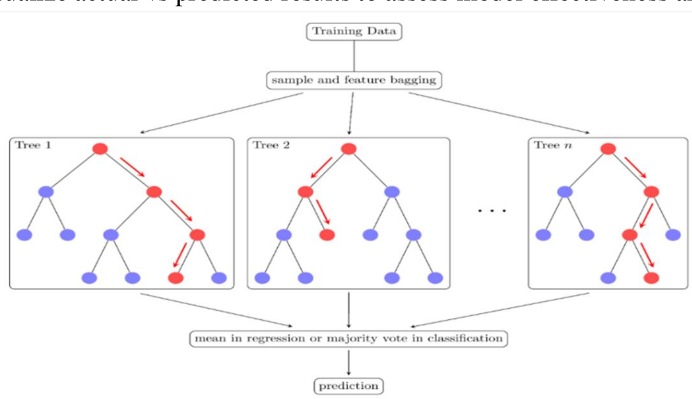

Fig.4.2.3. RF Architecture

4) *Technology Stack:*
- Programming Language: Python 3.11.0
- Libraries: NumPy, Pandas, Scikit-learn, TensorFlow/Keras, Matplotlib, Seaborn,yfinance,etc.
- Data Sources: Historical stock price data from Yahoo Finance.
- Code Editor / IDE's: Jupyter Notebook, Visual Studio.
- Web Framework: Streamlit (for interactive UI)

5) *System Development:*

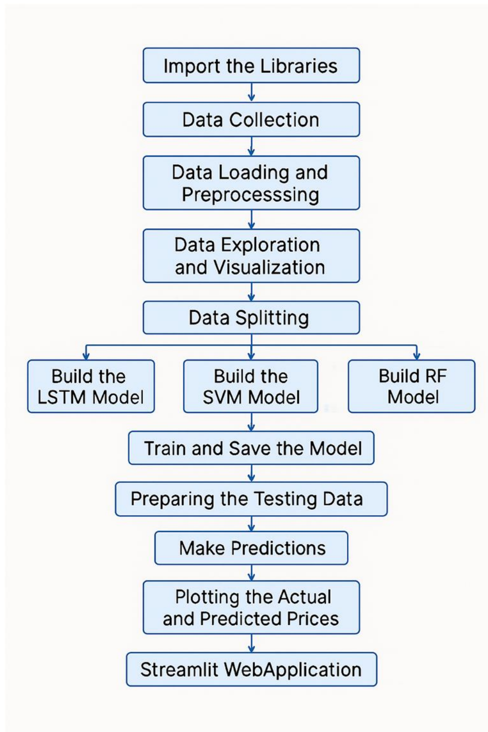

Fig. 4.2.5. System Development Overview

V. RESULTS AND DISCUSSION

A. *Qualitative Results*

Qualitative analysis involved examining specific interactions of different machine learning models to assess functional correctness and prediction accuracy. Key observations include their performance trends and user interpretability in stock price prediction.

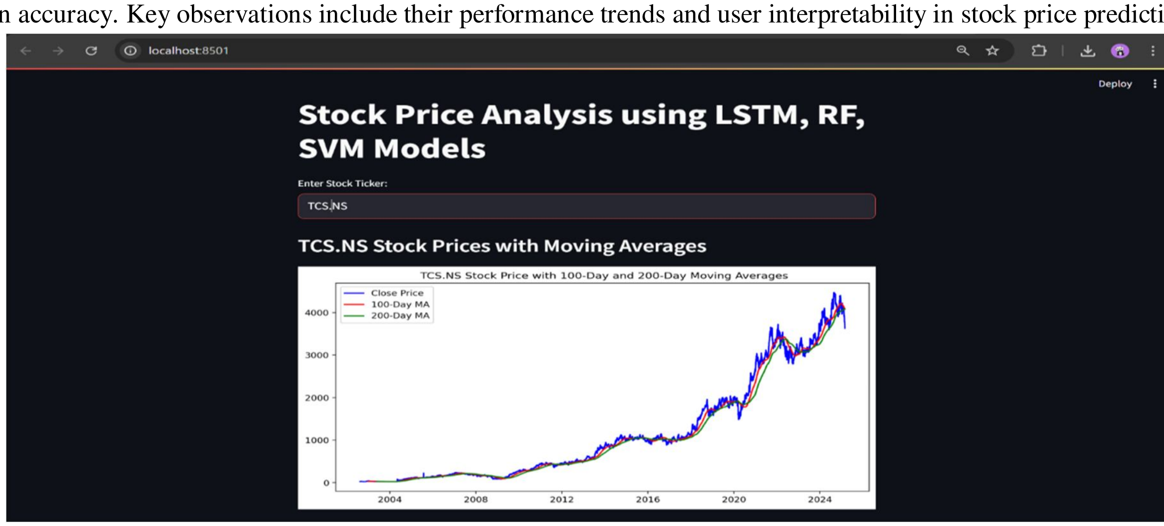

Fig.5.1.1. Stock Price With 100 days & 200 days MA

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

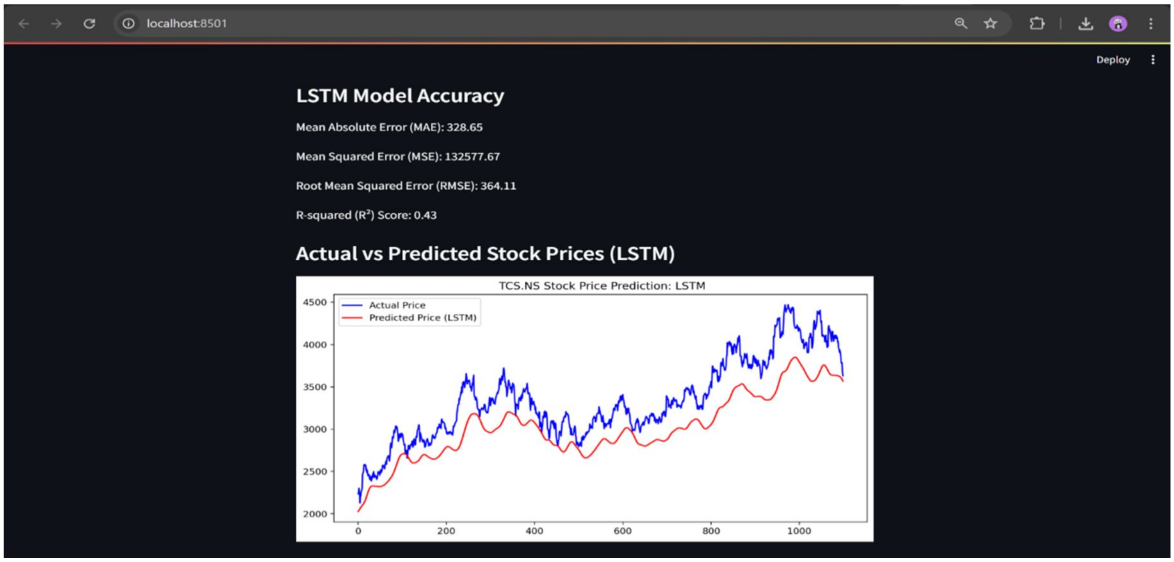

Fig.5.1.2. LSTM Model Accuracy Metric & Graph

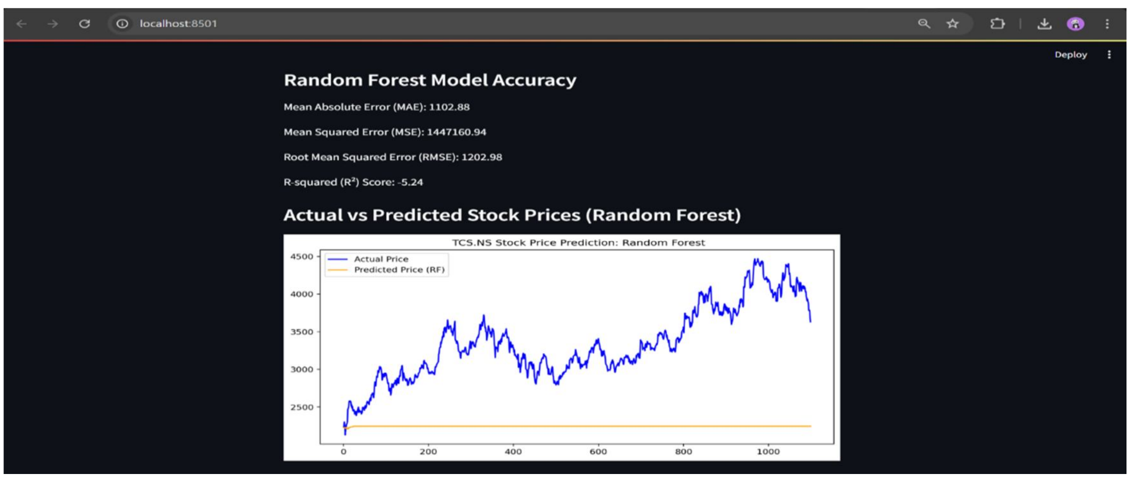

Fig.5.1.3. RF Model Accuracy Metric & Graph

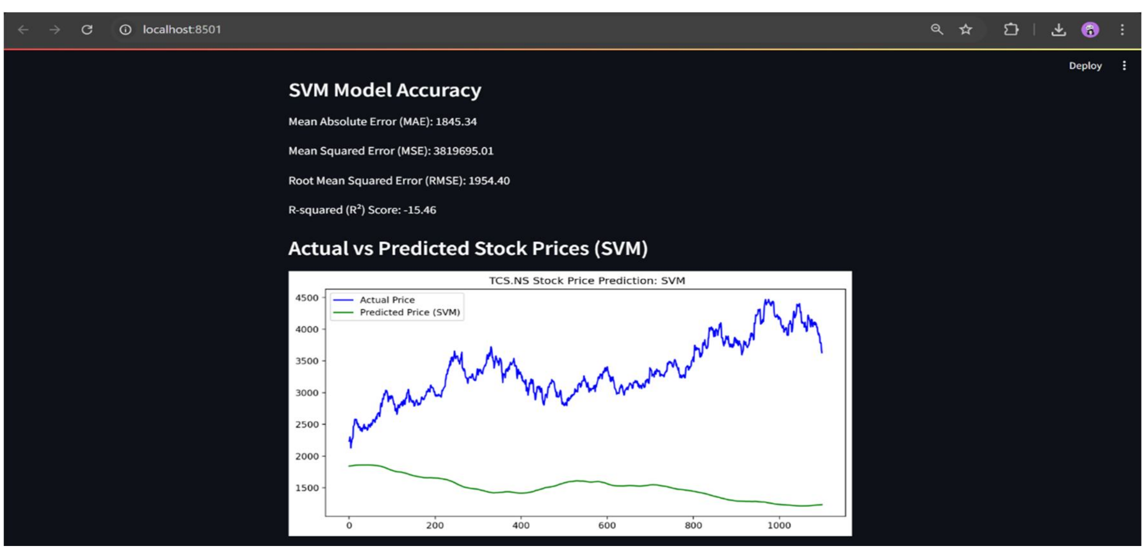

Fig.5.1.4. SVM Model Accuracy Metric & Graph

B. Discussion: Comparative Analysis of LSTM, RF, and SVM Models

Aspect	LSTM	Random Forest	SVM
1. Mean Absolute Error (MAE)	**328.65** (Lowest, better prediction accuracy)	**1,102.88** (Higher, less accurate)	**1,845.34** (Highest, least accurate)
2. Mean Squared Error (MSE)	**132,577.67** (Lowest, better accuracy in capturing variance)	**1,447,160.94** (Moderate, less accurate)	**3,819,695.01** (Highest, poor performance)
3. Root Mean Squared Error (RMSE)	**364.11** (Lowest, better performance)	**1,202.98** (Moderate)	**1,954.40** (Highest, poor performance)
4. R² Score	**0.43** (Positive, captures data variance reasonably well)	**-5.24** (Negative, poor model fit to data)	**-15.46** (Highly negative, poorest model fit)
5. Prediction Accuracy	**Best:** Closest to actual stock prices	**Poor:** Predictions deviate significantly	**Poorest:** Predictions completely fail to match trends
6. Time Complexity	**Moderate:** Sequential nature increases training time	**Low:** Faster for smaller datasets but slower for large trees	**High:** Solving quadratic optimization is computationally intensive
7. Space Complexity	**Moderate:** Requires storing weights for recurrent connections	**High:** Needs space for storing multiple decision trees	**Moderate:** Depends on the number of support vectors
8. Performance in Stock Price Trends	**Best:** Captures trends well	**Poor:** Struggles with complex trends	**Poorest:** Fails to follow stock trends

Fig.5.2. Comparative Study of LSTM, Radom Forest, SVM Model Results

VI. CONCLUSION

This project aimed at using machine learning algorithms to forecast stock market trends based on global financial information. We investigated three leading models—Long Short-Term Memory (LSTM), Support Vector Machines (SVM), and Random Forests (RF) to make daily market forecasts. The outcome proved the potential of these models, as their forecasts outperformed traditional benchmarks, with promise for real-world trading strategies to be more profitable.

LSTM had the highest accuracy in recognizing time-series patterns, SVM worked best with smaller, high-dimensional datasets, and Random Forest delivered consistent, interpretable results. Increasing dataset size, trying other algorithms, and applying different evaluation measures might lead to greater accuracy and reliability.

VII. FUTURE SCOPE

The present research lays a strong groundwork for stock price forecasting through machine learning and deep learning models. Nevertheless, the highly dynamic financial market provides room for future research in the following areas:

1) Integration of Multi-Source Data: Subsequent models can gain from bringing in alternate sources of data like news sentiment, social media trends, and economic indicators. This can enhance model accuracy by taking into account market psychology and macroeconomic changes.
2) Real-Time Forecasting System: Building a pipeline of real-time prediction with ongoing learning may increase the practicality of the model in portfolio management and high-frequency trading.

3) Cross-Market Analysis: The methods can be extended and tried across various stock markets (e.g., NYSE, NASDAQ, BSE, NSE) to test the generalizability and strength of each model.
4) Deployment at Scale: The Streamlit web application can be enhanced and deployed further with cloud platforms (AWS, GCP, Azure) to support larger user bases and make model inference scalable.

VIII. ACKNOWLEDGEMENT

We extend our heartfelt gratitude to Prof. Junaid Mandviwala, Department of Artificial Intelligence and Data Science, Rizvi College. His unwavering support, insightful guidance, and continuous motivation have fostered a positive research atmosphere, significantly contributing to the successful completion of this project and inspiring us towards academic excellence and publication. We also appreciate the Department of Artificial Intelligence and Data Science, Rizvi College of Engineering, Mumbai, for affording the appropriate infrastructure and resources to conduct this research. Special thanks to all the contributors who helped with testing and gave useful feedback during development.

REFERENCES

[1] Tsai, K., & Wang, J. (2009). External technology sourcing and innovation performance in LMT sectors. Research Policy, 38(3), 518–526. https://www.sciencedirect.com/science/article/pii/S0048733308002266
[2] Han, K., & Kim, J. (2000). Genetic quantum algorithm and its application to combinatorial optimization problem. Evolutionary Computation. https://www.researchgate.net/publication/2573070
[3] Usmani, M., Adil, S. H., Raza, K., & Ali, S. S. A. (2016). Stock market prediction using machine learning techniques. 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), 1354–1360. https://www.researchgate.net/publication/367044423
[4] Liu, S., Liao, G., & Ding, Y. (2018). Stock transaction prediction modelling and analysis based on LSTM. 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2787–2790. https://ieeexplore.ieee.org/document/8398027
[5] Hu, Z., Zhu, J., & Tse, K. (2013). Stocks market prediction using support vector machine. 6th International Conference on Information Management, Innovation Management and Industrial Engineering. https://www.irjet.net/archives/V5/i10/IRJET-V5I10193.pdf
[6] Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55, 307–319. https://www.sciencedirect.com/science/article/pii/S0925231203003722
[7] Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day. https://archive.org/details/timeseriesanalys0000boxg_p2r1
[8] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://link.springer.com/article/10.1023/A%3A1010933404324
[9] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning https://link.springer.com/article/10.1007/BF00994018
[10] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research https://www.sciencedirect.com/science/article/pii/S0377221717310652
[11] Streamlit, "Streamlit: The fastest way to build data apps," [Online]. https://streamlit.io/.
[12] Yahoo Finance. https://finance.yahoo.com/
[13] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://direct.mit.edu/neco/article/9/8/1735/6109/Long-Short-Term-Memory