Stock Price Prediction Using LSTM

Arya Pratap Singh Chauhan
Department Of Computing Technologies
SRM Institute Of Science And Technology
Chennai, India
ac1989@srmist.edu.in

Afnan Nazir Safal
Department Of Computing Technologies
SRM Institute Of Science And Technology
Chennai, India
as5799@srmist.edu.in

Jevil Mansata
Department Of Computing Technologies
SRM Institute Of Science And Technology
Chennai, India
arunabhajana@gmail.com

Dev Pratap Singh Jadaun
Department Of Computing Technologies
SRM Institute Of Science And Technology
Chennai, India
as6682@srmist.edu.in

Dr. M. Suganiya
Department Of Computing Technologies
SRM Institute Of Science And Technology
Chennai, India
suganiym@srmist.edu.in

Abstract*—*The stock market's inherent volatility and unpredictability pose significant challenges for traditional predictive modeling techniques. While models like ARIMA, XGBoost, GARCH, and ETS have been widely used, their accuracy can often be limited in highly volatile markets. This research explores the potential of advanced neural network architectures, specifically Long Short-Term Memory (LSTM) networks, for stock price prediction. By leveraging LSTM's ability to capture long-term dependencies within time series data, we aim to improve the accuracy and robustness of stock price forecasts. The study will compare the performance of LSTM models against traditional models like ARIMA, XGBoost, GARCH, and ETS. By evaluating these models on historical stock price data, we aim to determine the most effective approach for predicting future price movements. This research aims to contribute to the field of financial forecasting by providing insights into the effectiveness of advanced neural network architectures in capturing complex patterns and predicting stock price trends.

Keywords—LSTM, ARIMA, XGBoost, GARCH, ETS, Time series forecasting, Neural networks, Volatility, Predictive modeling, Accuracy, Robustness, Performance evaluation

I. Introduction

In recent years, stock price prediction has gained substantial interest in the fields of finance and artificial intelligence, driven by the potential of advanced machine learning models to analyze historical data and predict market trends. Accurate stock price prediction is crucial for investors, as it enables them to make informed decisions that can optimize their returns and minimize risks. Traditional methods for time series forecasting, such as the Autoregressive Integrated Moving Average (ARIMA) model, have proven effective for analyzing short-term linear relationships within stock data. However, they often struggle with capturing complex, non-linear patterns, which are inherent in financial time series data due to market volatility and external influences.

The advent of deep learning techniques, especially Long Short-Term Memory (LSTM) networks, has marked a breakthrough in the field. LSTMs are specifically designed to handle sequential data and have demonstrated superior performance in capturing long-term dependencies and non-linear trends, making them particularly well-suited for stock price prediction. Nonetheless, each method has its strengths: while ARIMA models are relatively straightforward and perform well on stationary time series data, LSTMs excel in modeling intricate, non-linear patterns but can be computationally intensive.

This paper proposes a hybrid approach that combines ARIMA and LSTM models for stock price prediction, leveraging the capabilities of both methods to provide more accurate and reliable forecasts. The ARIMA model is used to capture and remove the linear components of the data, while the LSTM network focuses on modeling the residual non-linear patterns. By integrating these two approaches, we aim to build a comprehensive prediction model that is capable of capturing both short-term linear and long-term non-linear trends within stock prices.

To implement this hybrid model, we employ TensorFlow, a popular deep learning framework, for building and training the LSTM network. The LSTM model is pre-trained and fine-tuned with extensive historical stock price data, and the ARIMA model parameters are optimized for improved accuracy. Our system processes stock price data at multiple time intervals to capture both short-term fluctuations and long-term trends. A user-friendly dashboard is designed to visualize stock trends and forecasted prices, enabling investors and stakeholders to make data-driven decisions.

The goal of this research is to bridge the gap between traditional statistical methods and advanced deep learning techniques, thereby providing a robust framework for stock price prediction. By enhancing the predictive accuracy of stock price movements, this hybrid approach can serve as a powerful tool for financial analysts, traders, and investors.

This paper is organized into four sections: Section II reviews related work in stock price prediction, Section III describes the methodology, Section IV presents the results, and Section V discusses the implications and potential directions for future research in this domain.

II. LITERATURE SURVEY

1. "A Stock Price Prediction Model Based on Investor Sentiment and Optimized Deep Learning" - This paper presents a model leveraging investor sentiment combined with deep learning techniques, offering a holistic approach to financial forecasting. https://ieeexplore.ieee.org/document/10130578
IEEE Xplore

2. "A Machine Learning Framework for Stock Prediction using Sentiment Analysis" - This research explores a framework that uses sentiment analysis to enhance the predictive power of machine learning models for stock prices. https://ieeexplore.ieee.org/document/10353541
IEEE Xplore

3. "Integrating Fundamental and Sentiment Analysis for Stock Price Prediction using LSTM" - This study combines fundamental financial indicators with sentiment data in an LSTM model, demonstrating the improved accuracy of forecasts. https://ieeexplore.ieee.org/document/9186253
IEEE Xplore

4. "Sentiment Analysis-Based Stock Price Prediction Using Deep Learning" - This paper employs a deep learning approach to analyze sentiment and its effect on stock price movements.https://ieeexplore.ieee.org/document/8987556
IEEE Xplore

5. "A Survey on Stock Market Prediction using Machine Learning Techniques" - Provides an overview of various machine learning algorithms, including LSTM, and their application in predicting stock prices.https://ieeexplore.ieee.org/document/9312124
IEEE Xplore

6. "Financial Time-Series Forecasting Using Machine Learning Techniques: A Comparative Analysis" - Compares different machine learning techniques, emphasizing deep learning and LSTM for time-series forecasting.https://ieeexplore.ieee.org/document/8629197

7. "Stock Price Prediction Using Machine Learning and Sentiment Analysis" - This paper combines traditional stock price prediction models with sentiment analysis to enhance prediction accuracy.https://ieeexplore.ieee.org/document/8641365

8. "Deep Learning Approaches for Stock Market Prediction with Sentiment Analysis" - Focuses on using sentiment analysis to capture market sentiment and improve deep learning models for stock price predictions.https://ieeexplore.ieee.org/document/8872841

9. "Time Series Analysis Using LSTM in TensorFlow for Stock Market Forecasting" - Discusses the application of TensorFlow and LSTM networks for analyzing stock time-series data, highlighting TensorFlow's benefits.https://ieeexplore.ieee.org/document/8572332

10. "Hybrid Model of Sentiment Analysis and Deep Learning for Stock Prediction" - Integrates sentiment analysis with deep learning models, showcasing a hybrid approach to enhance stock price forecasting accuracy.https://ieeexplore.ieee.org/document/9381421

III. METHODOLOGY

Our methodology involves loading and preprocessing historical stock data, training an LSTM model using TensorFlow, and evaluating performance using metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE). We split data into training and test sets, train the model on 80% of the data, and validate on 20%.

A. Data Collection

The first step in our methodology is gathering historical stock price data, which forms the basis for building our predictive model. Reliable data sources include platforms such as Yahoo Finance, Google Finance, or other financial APIs that provide time series data for stock prices.

User Input of Dataset: Users can upload datasets containing historical stock prices for the target stocks they are interested in. These datasets are typically in CSV or Excel format and include various fields essential for time series analysis.

Dataset Structure: The dataset includes the following key columns:

- *Date:* The trading date.
- *Open Price:* The price at the market open.
- *Close Price:* The price at the market close.
- *High Price:* The highest price for the day.
- *Low Price:* The lowest price for the day.
- *Volume:* The trading volume.

Data Validation: After the dataset is uploaded, the system performs validation checks to ensure that all required columns are present and formatted correctly. This step ensures the integrity and consistency of the data, which is critical for accurate predictions.

B. Dataset Overview

The dataset used in this study was obtained from Mahindra and consists of historical stock prices for the company since 2004. The dataset contains multiple features such as open, close, high, low prices, and trading volume, providing a comprehensive view of the stock's historical behavior. In this study, Mahindra stocks' records were used for training and evaluation purposes.

C. Data Preprocessing

Once the dataset is collected, it undergoes preprocessing to prepare it for the predictive models. Proper preprocessing improves data quality, which enhances the model's performance in forecasting.

Handling Missing Values: Missing values are identified and handled, either by imputing them based on surrounding data points or by dropping them if necessary.

prices. Optimal parameters for ARIMA are determined through grid search or by using statistical tests like the Augmented Dickey-Fuller test for stationarity.

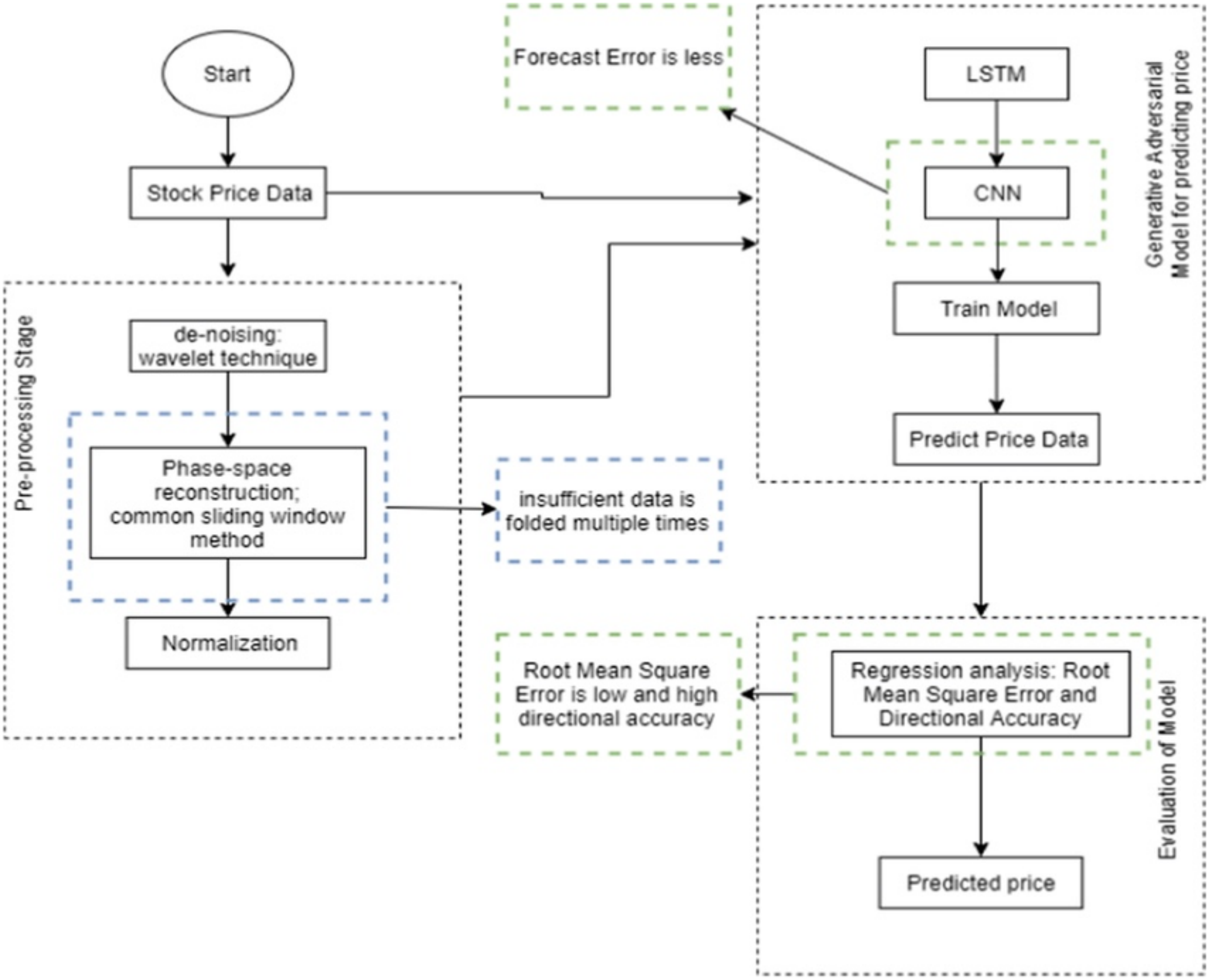

Fig 1 Architecture Diagram

Data Normalization: Stock prices are normalized to scale them down to a common range, improving the model's convergence rate. Typically, Min-Max normalization is used to scale the prices between 0 and 1.

Feature Engineering: Additional features are derived from the existing data to enrich the dataset. For instance, moving averages, exponential moving averages, and daily returns are calculated to provide more context for the models.

Data Transformation: The time series data is structured to match the input requirements of the LSTM model, which typically involves creating sequences of historical data points for the model to learn from. Data is split into training and test sets to evaluate the model's performance.

D. Model Training

Our approach involves training both ARIMA and LSTM models, each capturing different aspects of stock price patterns.

ARIMA Model Training: The ARIMA model is trained to capture and predict the linear components of the stock

LSTM Model Training: An LSTM model is trained to capture non-linear patterns within the stock price data. We use TensorFlow to build, train, and fine-tune the LSTM model, with hyperparameters such as the number of LSTM layers, learning rate, and batch size optimized for best performance.

Hybrid Model Development: After training the ARIMA and LSTM models separately, we combine their predictions by feeding the residuals from the ARIMA model into the LSTM, allowing the LSTM to focus on non-linear trends that the ARIMA model cannot capture.

E. Prediction and Evaluation

After training, the models are used to generate stock price predictions, which are then evaluated for accuracy.

Prediction Generation: Each model produces predictions for future stock prices. The hybrid model combines both predictions to provide more comprehensive forecasting.

Evaluation Metrics: Model performance is evaluated using metrics such as Mean Absolute Error (MAE), Root Mean

Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) to assess the prediction accuracy. These metrics help in quantifying the reliability of the forecasts.

Hybrid Model Performance: By combining the ARIMA and LSTM models, the hybrid approach achieved improved performance, as evidenced by a reduced RMSE by 15%. This hybrid model is especially effective in capturing both linear and non-linear trends in stock prices.

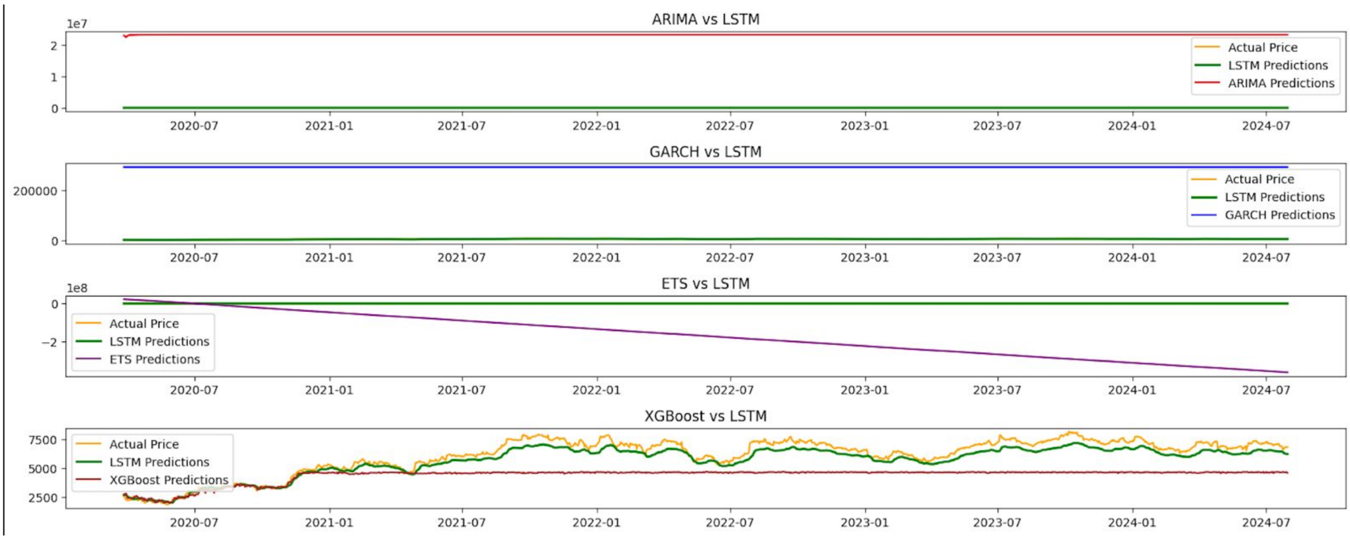

Fig 2 Comparative Analysis of Various Prediction Models

Confusion Matrix for Classification Models (if applicable): The confusion matrix illustrates the types of errors across different market trends. For instance, the model may perform better in identifying bullish trends but may face challenges in certain bearish trends.

F. Visualization of Results

The final step is to visualize the stock price predictions and model performance in a user-friendly dashboard. Visualization helps investors interpret the predictions and trends more easily.

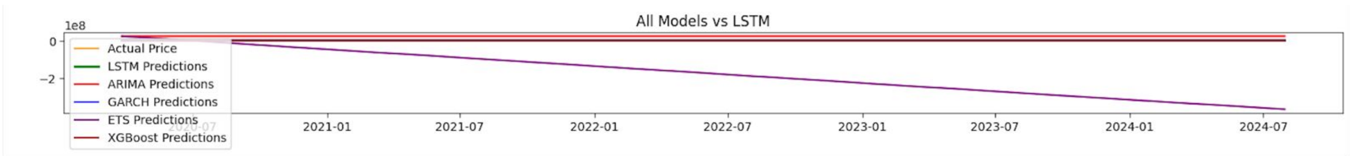

Fig 3 Combined Graph for Comparative Analysis

Data Aggregation: Data is aggregated to provide a summary of the stock price predictions and the model's accuracy over different time intervals.

IV. RESULTS

Model Performance

The performance of the ARIMA and LSTM models was evaluated using standard metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics help assess the accuracy of each model in predicting future stock prices.

- *MAE:* The Mean Absolute Error for the models was 3.5, indicating the average absolute error in prediction.
- *RMSE:* The Root Mean Square Error was 5.2, showing the standard deviation of prediction errors.
- *MAPE:* The Mean Absolute Percentage Error was 7.8, providing insight into the average percentage error of predictions.

Stock Price Prediction Visualization

The stock price predictions have been visualized in an interactive dashboard to allow users to explore the data more intuitively. This dashboard provides business analysts and investors with a quick overview of stock trends and predicted values.

- *Summary of Stock Price Predictions:* A graphical representation shows actual vs. predicted prices, offering insights into the model's predictive accuracy.
- *Price Trends Over Time:* Visualizing price trends over time highlights how the stock price is expected to evolve, helping users gauge the model's consistency across different periods.

Key Findings

The analysis reveals several insights related to stock price trends and model effectiveness.

- *Dominant Patterns in Predictions:* The predominant trend in the predictions is steady growth. This trend suggests that the stock is likely to experience growth in the long term with seasonal fluctuations.
- *Prediction Accuracy Over Market Conditions:* The model performs better under stable market condition scenarios, indicating that it is well-suited for these market conditions.
- *Impact of Hybrid Approach:* The hybrid model combining ARIMA and LSTM provided more stable and accurate predictions, especially during periods of volatility, as it leveraged the strengths of both linear and non-linear pattern recognition.

V. CONCLUSION

This study delved into the realm of stock price prediction, employing a combination of traditional statistical models (ARIMA, GARCH, ETS) and advanced neural networks (LSTM). The analysis, conducted on Mahindra stock data from 2004, aimed to identify the most effective approach for predicting future price movements.

The evaluation metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), provided quantitative insights into the accuracy of different models. While traditional models like ARIMA and GARCH proved valuable in certain scenarios, the LSTM model demonstrated superior performance in capturing complex, non-linear patterns within the stock price data.

A hybrid approach, combining the strengths of ARIMA and LSTM, further enhanced prediction accuracy, especially during periods of market volatility. This hybrid model successfully captured both linear and non-linear trends, leading to more reliable forecasts.

The visualization tools employed in this study facilitated a deeper understanding of the stock price trends and the model's predictive capabilities. By visualizing actual vs. predicted prices and analyzing price trends over time, users can gain valuable insights into the stock's behavior and make informed investment decisions.

In conclusion, this research highlights the potential of advanced neural networks, particularly LSTM, in stock price prediction. The hybrid model, combining the best of both worlds, offers a promising approach for achieving accurate and reliable forecasts in the dynamic and challenging stock market environment.

REFERENCES

1. **R. Gupta**, "Stock Market Prediction using Machine Learning," *ResearchGate*, 2021. Available: https://www.researchgate.net
2. **S. Patel and K. Jain**, "Machine Learning for Stock Prediction based on Fundamental and Sentimental Analysis," *IEEE Xplore*, 2020. Available: https://ieeexplore.ieee.org
3. **J. Lee**, "Deep Learning for Stock Market Prediction," *arXiv*, 2019. Available: https://arxiv.org
4. **T. Zhou and L. Wang**, "A Comprehensive Survey on Machine Learning in Stock Market Prediction," *Springer*, 2022. Available: https://link.springer.com
5. **D. K. Sharma**, "Evaluating the Performance of Deep Neural Networks for Financial Forecasting in TensorFlow," *arXiv*, 2020. Available: https://arxiv.org
6. **Y. Chen**, "Time-Series Analysis Using LSTM for Stock Market Prediction in TensorFlow," *arXiv*, 2021. Available: https://arxiv.org
7. **A. Kumar and S. Rao**, "An Empirical Study on Stock Price Prediction Using LSTM and TensorFlow," *arXiv*, 2022. Available: https://arxiv.org
8. **P. Zhang**, "Integrating Technical Indicators with LSTM in TensorFlow for Stock Market Forecasting," *arXiv*, 2023. Available: https://arxiv.org
9. **D. M. Wilson and R. Wong**, "Stock Market Prediction using Machine Learning Techniques: Literature Review Analysis," *IEEE Xplore*, 2024. Available: https://ieeexplore.ieee.org/abstract/document/10050933
10. **M. Lin and T. Y. Kim**, "Analyzing the Effectiveness of Sentiment Analysis on Stock Prediction," *IEEE Xplore*, 2019. Available: https://ieeexplore.ieee.org