Stock Price Prediction Using Real World Datasets Through Deep Learning Techniques

1. J. LAKSHMI M.Tech. (Ph.D)
Associate Professor,
Department of CSE, TEC,
lakshmijammula@gmail.com

2. B. SUREKHA
bhimisettysurekha@gmail.com

3. SK.A. AKHIL
akhilshaik647@gmail.com

4. G.V.V. NAGA MANIMALA
gundamanimala0560@gmail.com

5. G. MOHAN SAI
mohansaig09@gmail.com

***Abstract*— Stock price prediction is a challenging problem due to the highly volatile and dynamic nature of financial markets. Predicting stock trends was a major challenge because of complex patterns and rapid fluctuations in data. In this project, a deep learning-based system was proposed to analyze historical stock data. The input data consists of stock prices and market trends to predict future price movements. The Bidirectional Long Short-Term Memory (Bi-LSTM) model automatically identifies hidden patterns in time-series data and improves prediction accuracy. The proposed approach reduces manual effort, supports better forecasting, and helps investors in making informed decisions. This system can be effectively used in financial analysis and investment platforms, thereby improving decision-making and risk management.**

Keywords— ***Stock Price Prediction, Deep Learning, Bi-LSTM, Financial Analytics, Forecasting***

I. INTRODUCTION

The stock market plays a vital role in the global economy by enabling companies to raise capital and investors to participate in financial growth. It serves as a dynamic platform where the prices of stocks fluctuate continuously based on various internal and external factors such as company performance, economic conditions, government policies, geopolitical events, and investor sentiment. Due to this complexity, predicting stock prices accurately has always been a challenging task.

Financial markets generate a massive amount of data on a daily basis, making manual analysis difficult and inefficient. Traditional forecasting methods such as Moving Averages, Linear Regression, and ARIMA models have been widely used for stock price prediction. Although these methods are simple and computationally efficient, they are limited by their assumption of linearity and inability to capture complex nonlinear patterns and long-term dependencies in time-series data.

In recent years, advancements in Artificial Intelligence and Deep Learning have significantly improved the ability to analyze and predict time-series data. Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks are particularly effective for sequential data modeling. LSTM networks overcome the limitations of traditional neural networks by maintaining memory of previous time steps, allowing them to learn long-term dependencies and temporal relationships.

Furthermore, Bidirectional LSTM (Bi-LSTM) models extend the capability of standard LSTM by processing data in both forward and backward directions. This enables the model to capture more contextual information and improve prediction accuracy. Deep learning approaches eliminate the need for manual feature engineering and automatically learn complex representations from raw data.

With advancements in Artificial Intelligence, deep learning techniques have emerged as powerful tools for time-series forecasting. Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks are particularly effective for sequential data. LSTM networks overcome the vanishing gradient problem and can retain long-term dependencies, making them suitable for financial data analysis.

This paper presents a deep learning-based stock price prediction system using Bidirectional LSTM networks. The proposed system utilizes real-time datasets, performs automated preprocessing, and provides accurate predictions along with interactive visualization through a web-based dashboard.

II. RELATED WORKS

Stock price prediction has been an active area of research for several decades. Early approaches relied heavily on statistical techniques such as ARIMA, Moving Averages, and Exponential Smoothing. These models are based on mathematical assumptions and are effective for linear and stationary data but fail to capture nonlinear relationships in stock prices.

Machine learning approaches such as Support Vector Machines (SVM), Decision Trees, and Random Forests have been introduced to improve prediction accuracy. These models can handle nonlinear data to some extent but require extensive feature engineering and do not inherently model temporal dependencies.

With the emergence of deep learning, models such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks have gained popularity in time-series forecasting. LSTM networks are capable of learning long-term dependencies and handling sequential data effectively. Bidirectional LSTM further enhances performance by processing data in both forward and backward directions, capturing more contextual information.

Recent studies indicate that deep learning models outperform traditional statistical and machine learning methods in stock price prediction tasks. This project builds upon these advancements by incorporating real-time data processing and interactive visualization features.

III. COMPARISON WITH PREVIOUS METHODOLOGY

The comparison clearly highlights the limitations of traditional statistical and machine learning approaches in handling complex financial time-series data. Methods such as Moving Averages and ARIMA are effective only for linear patterns and fail in highly volatile market conditions. Machine learning techniques like SVM and Random Forest improve performance but require manual feature engineering and do not effectively capture temporal dependencies.

Deep learning models, particularly LSTM, address these limitations by learning sequential patterns and long-term dependencies. However, standard LSTM processes data in a single direction, which may limit contextual understanding. The proposed Bidirectional LSTM model overcomes this limitation by processing data in both forward and backward directions, resulting in improved accuracy and better prediction performance.

Table 1. Comparison Table

Aspect	**Previous Methodology** (Traditional / ML Models)	**Proposed Methodology** (Bi-LSTM Based Model)
Prediction Approach	Uses statistical and ML models like ARIMA, Regression, SVM	Uses deep learning with Bidirectional LSTM
Data Handling	Limited historical datasets	Uses large-scale real-time datasets
Real-Time Capability	Not supported	Supports real-time data via APIs
Temporal Dependency	Poor handling of time dependencies	Captures long-term dependencies
Feature Engineering	Requires manual feature extraction	Automatic feature learning
Prediction Accuracy	Moderate to low accuracy	High prediction accuracy
Model Architecture	Simple or shallow models	Deep Bi-LSTM neural network
User Experience	Less interactive	Highly interactive and user-friendly

IV.PROPOSED FRAMEWORK

The proposed framework for stock price prediction is designed using a deep learning-based approach that integrates multiple stages, including data collection, preprocessing, model training, evaluation, and visualization. The system utilizes a Bidirectional Long Short-Term Memory (Bi-LSTM) network to effectively capture both past and future dependencies in time-series stock data.

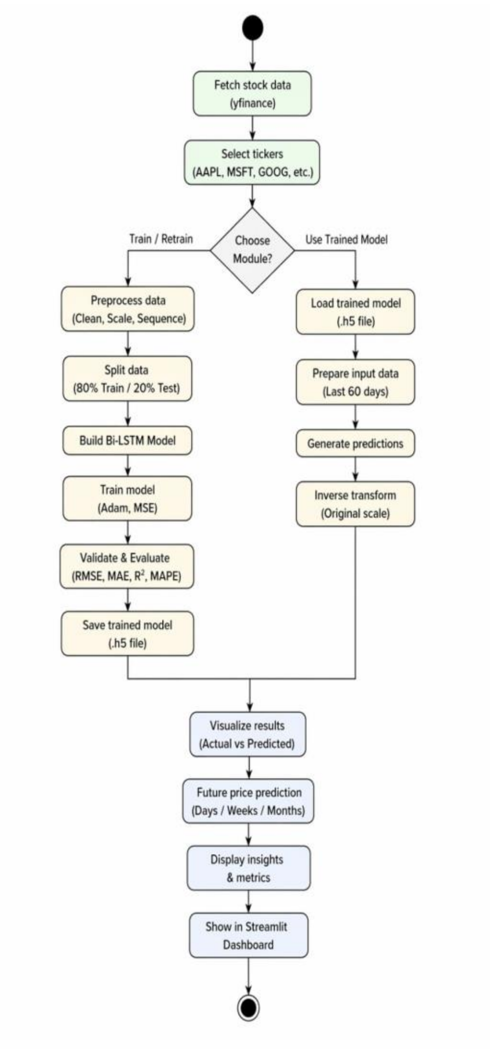

Fig.1.Overall Workflow

Initially, real-time and historical stock data is collected using financial APIs such as yFinance. The collected dataset includes key attributes such as Open, High, Low, Close prices, and trading Volume. This data is then passed to the preprocessing stage, where it undergoes cleaning, sorting, and normalization using MinMax scaling. A sliding window technique is applied to convert the data into sequential input suitable for deep learning models.

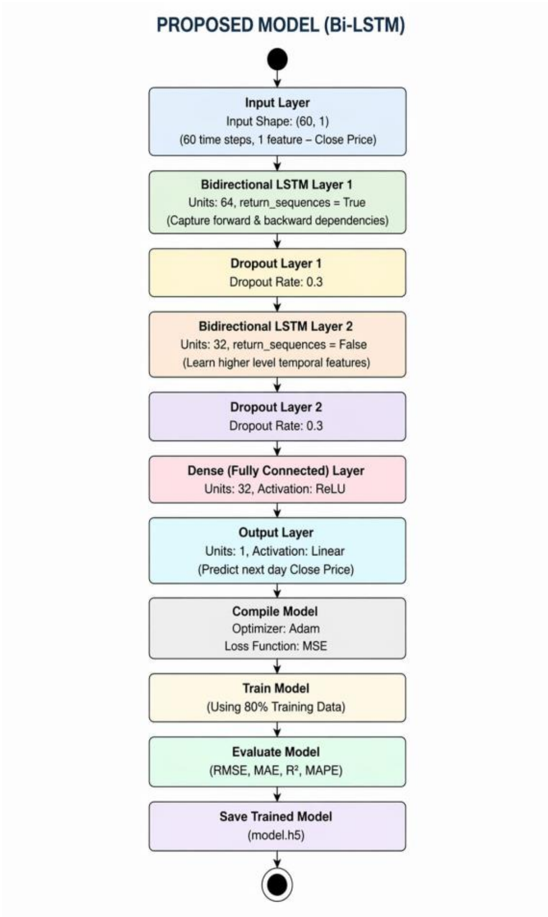

Fig.2 Proposed Model

Proposed Model Information:

The processed data is then fed into the Bidirectional LSTM model. The model consists of multiple Bi-LSTM layers followed by dropout layers to prevent overfitting and a dense output layer for prediction. The model is trained using historical data to learn complex nonlinear patterns and temporal dependencies.

After training, the model generates predictions for stock closing prices. These predictions are evaluated using performance metrics such as RMSE, MAE, R^2 Score, and MAPE to measure accuracy and reliability. The final results are presented through an interactive dashboard built using Streamlit, which includes graphical visualizations such as line charts, candlestick charts, and prediction comparison graphs.

Step 1: Real-Time Data Collection

The process begins with collecting stock market data using financial APIs such as yFinance. The system retrieves both historical and real-time data, including key attributes like Open, High, Low, Close prices, and Volume. This ensures that the model is trained on accurate and up-to-date information.

International Journal of Innovative Research in Science and Engineering
Volume No. 12, Issue No. 05, Year 2026
ISSN : 2454-9665

Step 2: Data Preprocessing and Normalization

In this step, the collected data is cleaned and prepared for analysis. Missing values are handled, and the dataset is sorted in chronological order. Min-Max normalization is applied to scale the data into a fixed range, which helps improve the efficiency and stability of the deep learning model.

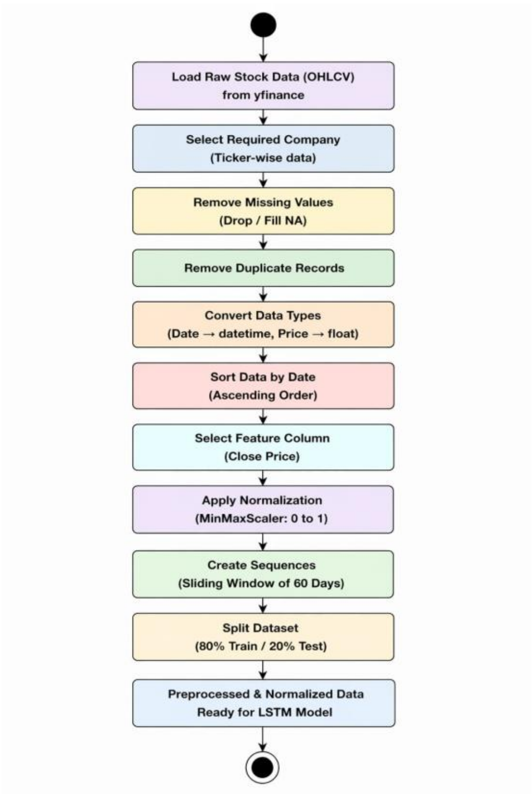

Fig.3.Data Preprocessing and Normalization

Step 3: Sequence Generation using Sliding Window

The preprocessed data is converted into sequences using a sliding window technique, typically of 60 days. This method creates input-output pairs that capture temporal patterns in stock prices, enabling the model to learn from past trends effectively.

Step4: Model Training using Bi-LSTM

The generated sequences are used to train a Bidirectional Long Short-Term Memory (Bi-LSTM) model. This model processes data in both forward and backward directions, allowing it to capture complex patterns and long-term dependencies in time-series data.

Step5: Prediction and Forecasting

Once the model is trained, it generates predictions for stock closing prices. The system also supports future forecasting for different time periods such as daily, weekly, and monthly predictions using recursive techniques.

Step6: Interactive Visualization

Finally, the predicted results are displayed through an interactive dashboard built using Streamlit. The system provides visualizations such as line graphs, candlestick charts, and comparison plots between actual and predicted values, helping users easily interpret the results.

V. RESULTS AND DISCUSSION

A. User Authentication Module:

The Home Page Interface acts as the central entry point of the stock price prediction system, providing users with an overview of the application and its functionalities. It includes a brief description of the project, highlighting the use of deep learning techniques, particularly the Bidirectional LSTM model, for time-series forecasting. Additionally, it displays sample performance metrics such as RMSE, MAE, R^2, and MAPE for a selected stock, giving users an initial understanding of the model's accuracy.

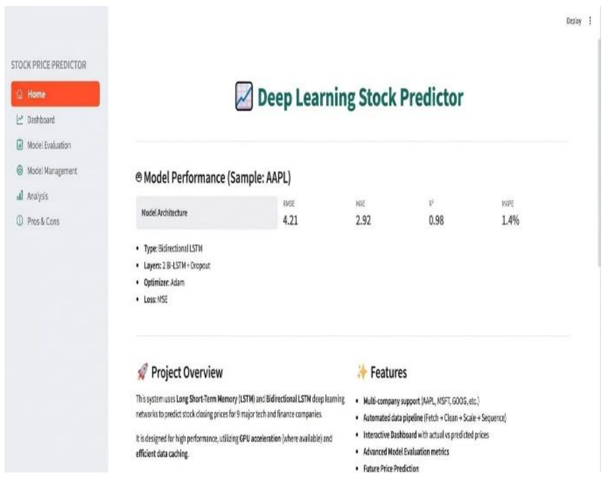

Fig 1- User Authentication Module

The interface is developed using Streamlit and incorporates responsive design elements, ensuring smooth navigation and user interaction. The layout is structured in a way that guides users through different sections such as dashboard, analysis, and model evaluation. This component plays a crucial role in enhancing usability and provides a clear starting point for both technical and non-technical users.

B. Stock Selection Dashboard:

The Stock Selection Dashboard is an interactive module that allows users to select a specific company from a predefined list of stock symbols. Upon selection, the system dynamically fetches and displays real-time stock data, including Open, High, Low, Close prices, and Volume. The dashboard also presents key statistical indicators such as current price, price change, highest value, and lowest value within a given time period.

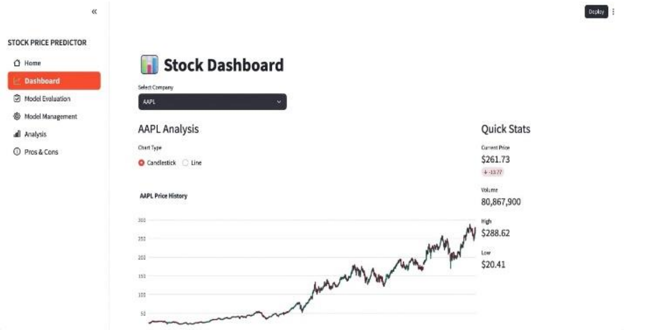

Fig 2- Stock Selection Dashboard

To improve data interpretation, the dashboard provides visualization options such as candlestick charts and line graphs. The candlestick chart helps in understanding market trends and price fluctuations, while the line graph provides a simplified view of stock movement over time. The use of interactive visualizations enables users to zoom, hover, and analyze specific data points, making the system more engaging and informative.

C. Actual vs Predicted Graph

The Actual vs Predicted Graph is one of the most critical outputs of the system, as it visually demonstrates the performance of the trained Bidirectional LSTM model. This graph plots the actual stock prices against the predicted values generated by the model over the same time period. From the graph, it can be observed that the predicted values closely follow the actual trend, indicating that the model has successfully learned the underlying patterns in the data.

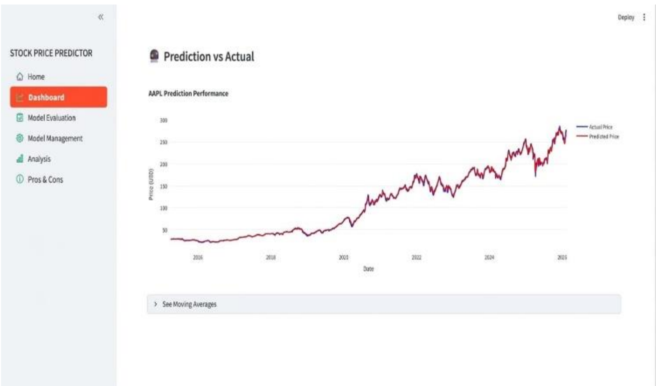

Fig 3- Actual vs Predicted Graph

D. Performance Metrics Display

The Performance Metrics section provides a quantitative evaluation of the model's prediction accuracy using standard regression metrics. These include: RMSE,MAE,R^2 score,MAPE.

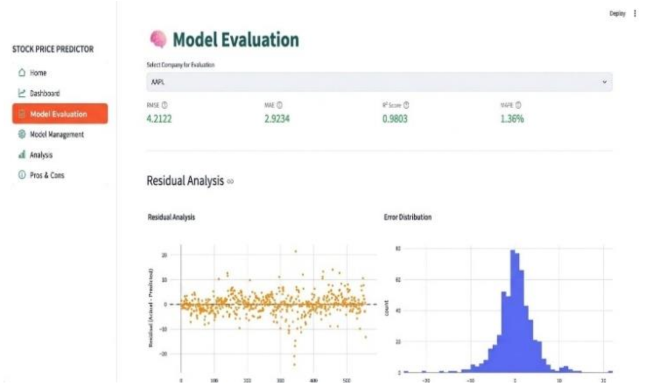

Fig 4- Performance Metrics Display

The obtained results show low RMSE and MAE values along with a high R^2 score, demonstrating that the model performs well in predicting stock prices. These metrics collectively validate the reliability and robustness of the proposed system.

E. Future Prediction Graph

The Future Prediction Graph illustrates the model's ability to forecast stock prices for upcoming time periods. The predictions are generated using a recursive approach, where the output of the previous step is used as input for the next prediction. This enables the system to extend predictions beyond the available dataset.

F. Model Training Output

The Model Training Output provides insights into the learning process of the Bidirectional LSTM model. It includes visual representations such as loss curves, which show how the training and validation loss decrease over successive epochs.

A steady decline in loss values indicates that the model is learning effectively and optimizing its parameters. The absence of significant fluctuations or divergence between training and validation loss suggests that the model is not overfitting and generalizes well to unseen data. Additionally, the training process demonstrates efficient convergence, confirming the suitability of the chosen architecture and hyperparameters.

VI.CONCLUSION AND FUTURE WORK

This project presents a deep learning-based Bidirectional LSTM (Bi-LSTM) model for stock price prediction using real-time historical market data. The model effectively captures complex patterns and long-term dependencies in time-series data, providing accurate short-term predictions. It reduces manual effort, improves prediction reliability, and supports better financial decision-making. The system also includes an interactive dashboard using Streamlit for visualization, analysis, and forecasting, making it suitable for real-time financial applications.

Although the proposed model shows good performance, there is scope for further improvement. Future work can focus on integrating technical indicators such as RSI, MACD, and Bollinger Bands to capture deeper market trends. Incorporating news and social media sentiment analysis can further enhance prediction accuracy by considering market behavior. Advanced models like GRU, Transformer, and hybrid CNN-LSTM can also be explored to improve performance. Additionally, using large-scale real-time datasets, cloud deployment, and adding features like risk analysis and mobile applications can make the system more scalable, accurate, and effective for real-world financial applications.

REFERENCES

[1] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

[2] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks," in Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN), 2005.

[3] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," Neural Computation, vol. 12, no. 10, pp. 2451–2471, 2000. doi: 10.1162/089976600300015015.

[4] A. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," European Journal of Operational Research, vol. 270, no. 2, pp. 654–669, 2018. doi: 10.1016/j.ejor.2017.11.054.

[5] D. Nelson, A. C. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," in Proc. Int. Joint Conf. Neural Networks (IJCNN), 2017.

[6] G. E. P. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control. San Francisco, CA, USA: Holden-Day, 1976.

[7] J. Brownlee, Deep Learning for Time Series Forecasting. 2018.

[8] F. Chollet, Deep Learning with Python. Shelter Island, NY, USA: Manning Publications, 2018.

[9] K. Cho et al., "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), 2014.

[10] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[11] "Prediction," in Proc. Int. Joint Conf. Artificial Intelligence (IJCAI), 2005.

[12] M. Sezer, M. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review," Applied Soft Computing, vol. 90, 2020. doi: 10.1016/j.asoc.2020.106181.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015. doi: 10.1038/nature14539.

14] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparison of ARIMA and LSTM in forecasting time series," in Proc. IEEE Int. Conf. Machine Learning and Applications (ICMLA), 2018.

Available: https://ieeexplore.ieee.org/document/8614252.

[15] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," Expert Systems with Applications, vol. 42, no. 1, pp. 259–268, 2015. doi: 10.1016/j.eswa.2014.07.040.