Hybrid Deep Learning Framework for Stock Market Price Prediction Using News Sentiment and Time-Series Analysis

Indraja S*, Nidhi K S *, Suraj S G*, Vishwas D G*, Sayyed Johar**

*Students, VII Sem, Dept. of AIML, Jawaharlal Nehru New College of Engineering, Shivamogga
**Assistant Professor, Dept. of AIML, Jawaharlal Nehru New College of Engineering, Shivamogga

Abstract—**Stock market price prediction is a challenging problem due to the highly dynamic nature of financial markets, which are influenced by both historical price movements and external factors such as public sentiment and real-world events. Traditional time-series forecasting models rely primarily on numerical data and often fail to capture the impact of news-driven market behavior. To address this limitation, this paper proposes a hybrid stock market prediction framework that integrates deep learning–based time-series modeling with news sentiment analysis. The proposed system employs a Long Short-Term Memory (LSTM) network trained on historical stock prices and derived technical indicators to model temporal dependencies in market trends. In parallel, financial news articles are processed using Natural Language Processing (NLP) techniques to extract sentiment scores that reflect market optimism or pessimism. These sentiment features are fused with numerical stock data to generate refined price predictions.**

A web-based dashboard is developed to visualize live stock prices, candlestick charts, sentiment trends, and predicted price movements in real time, enabling intuitive analysis for investors and traders. Experimental evaluation demonstrates that the hybrid LSTM-sentiment model achieves improved prediction accuracy compared to conventional time-series forecasting approaches. The results highlight the effectiveness of combining quantitative market data with qualitative news sentiment, providing a robust decision-support tool for intelligent stock market analysis.

Keywords— **Stock Market Prediction; News Sentiment Analysis; Deep Learning; Long Short-Term Memory (LSTM); Natural Language Processing; Financial Time-Series Forecasting; Technical Indicators; Decision Support System**

I. INTRODUCTION

Stock market prediction is a prominent field of research due to the complex, volatile, and highly dynamic nature of financial markets. Stock movement is influenced not only by numerical indicators but also by psychological and socio-economic factors communicated through news. Therefore, analyzing both historical price patterns and real-time market sentiment is crucial to forecast stock prices accurately. With advancements in Artificial Intelligence, Natural Language Processing (NLP), and Deep Learning, hybrid models have emerged that combine numerical and textual information for improved forecasting accuracy. This project focuses on building such a hybrid model.

The stock market plays a critical role in the global economy and directly influences investment decisions, corporate planning, and financial growth. Due to its volatile nature, even small changes in sentiment can trigger large fluctuations in stock prices. Therefore, timely and accurate prediction of stock behaviour is essential for minimizing risks and maximizing returns. This work becomes important for the following reasons:

- News drives market movement: A large portion of stock price fluctuations are driven by reactions to news such as quarterly earnings, government regulations, mergers and acquisitions, global crises, and industry announcements. Incorporating news sentiment provides a more realistic picture of price behaviour.
- Bridges the gap between numerical and behavioural analysis: Traditional forecasting models rely only on historical price data and cannot detect the psychological effect of market sentiment. Combining LSTM prediction with sentiment analysis improves reliability.
- Supports informed decision-making: The system provides investors and analysts with a real-time dashboard that helps in

evaluating market conditions, risk levels, and profit opportunities before taking decisions.

🞢 Useful for both beginners and professional traders: Even users with limited financial knowledge can interpret prediction and sentiment insights to understand whether the market is bullish, bearish, or neutral.

🞢 Applicable across global financial markets: Since the model is independent of a specific stock exchange, it can be applied to any company or index worldwide, making it scalable for real-world applications.

Thus, the paper holds significant importance as it not only forecasts market price but also captures market psychology, enabling more dependable and intelligent trading decisions. Stock market prediction is a complex and challenging task due to the highly dynamic nature of financial markets and the influence of numerous external factors. Conventional stock price prediction approaches primarily depend on historical price trends and technical indicators, often overlooking the impact of real-time information such as financial news, economic events, and market sentiment. As a result, these methods may fail to capture sudden market movements driven by investor perception and external developments. The problem addressed in this project is the lack of an integrated predictive system that effectively combines quantitative historical stock data with qualitative information derived from financial news. To overcome this limitation, the proposed work leverages the use of Natural Language Processing (NLP) techniques to analyze unstructured news articles and extract sentiment and contextual insights that influence stock price movements. By integrating news-based sentiment analysis with historical stock data, the system aims to enhance prediction accuracy and provide a more comprehensive understanding of market behavior.

II. RELATED WORK

Stock market prediction has long been an active research area due to the complex, nonlinear, and highly volatile nature of financial markets. Traditional time-series forecasting models primarily rely on historical price data and technical indicators; however, these approaches often fail to capture the influence of external factors such as public sentiment, economic events, and financial news. As a result,

recent research has increasingly focused on integrating textual information from news sources with numerical stock data to enhance prediction accuracy.

Early studies explored market prediction using daily news headlines combined with time-series analysis. These works employed n-gram models, linguistic feature engineering, and sentiment polarity extraction to generate additional features such as subjectivity, objectivity, and sentiment scores. By integrating these features with Long Short-Term Memory (LSTM) models trained on Dow Jones Index data, researchers demonstrated marginal improvements over baseline time-series models. However, results indicated that achieving high accuracy on headline-based datasets is challenging due to limited data size, lack of news categorization, and simplistic sentiment modeling, leading to weak overall predictive performance.

Subsequent research investigated the integration of sentiment scores from financial news with machine learning regression models. Studies using Multi-Layer Perceptron (MLP) regression combined with sentiment scores derived from tools such as VADER, TextBlob, and Flair evaluated the influence of each sentiment analyzer on stock price prediction. Experimental results showed that deep semantic sentiment models like Flair outperformed lexicon-based approaches and achieved trend prediction accuracy of up to 75% over extended horizons. Nevertheless, these approaches were limited by short prediction windows, overfitting risks, and simplistic labeling of news relevance, which reduced robustness in real-world market conditions.

Several works extended sentiment-integrated prediction models into full-fledged systems by incorporating real-time data acquisition and visualization. End-to-end web-based applications combining LSTM models, live stock price feeds, and news scraping demonstrated improved short-term predictive accuracy, with errors typically ranging between 0.1% and 3%. While such systems highlighted the practical applicability of hybrid prediction models, their reliance on static sentiment indicators, limited data sources, and short-term forecasting constrained long-term generalization and adaptability.

Parallel research streams focused on purely numerical machine learning approaches using

historical market data. Studies employing Support Vector Machines (SVM) and Random Forest (RF) regressors on NIFTY-50 datasets showed strong predictive performance, with RF models achieving lower RMSE and higher R² scores compared to SVM. Despite their effectiveness, these models lacked integration with real-time news or sentiment data, limiting their ability to respond to sudden market shifts driven by external events. Additionally, basic feature engineering and the absence of hyperparameter optimization reduced their extensibility.

Deep learning–centric studies further evaluated multiple regression models, including KNN, Linear Regression, Decision Trees, Support Vector Regression, and LSTM, across a wide range of Indian stocks. Results consistently indicated that LSTM models outperform traditional algorithms in terms of RMSE, R², and SMAPE, confirming their superiority in capturing temporal dependencies in financial data. However, these studies primarily relied on historical price information and did not incorporate fundamental or sentiment-based features, restricting contextual awareness.

More comprehensive approaches integrated macroeconomic indicators such as gold prices, crude oil prices, exchange rates, and government securities alongside news sentiment and historical stock data. By comparing LSTM and Random Forest regressors,

these studies demonstrated that sentiment analysis can enhance prediction accuracy for certain stocks, although performance varied across sectors. High computational cost, noise in news data, fixed content windows, and lack of pretrained language models were identified as major limitations affecting scalability and accuracy.

Overall, existing literature indicates that hybrid models combining deep learning–based time-series forecasting with news sentiment analysis offer measurable improvements over standalone numerical models. LSTM consistently emerges as the most effective architecture for temporal modeling, while advanced sentiment extraction methods contribute contextual awareness. However, challenges such as limited datasets, weak event modeling, static sentiment analysis, and lack of company-specific relevance remain unresolved. These gaps motivate the development of an intelligent hybrid prediction framework that tightly integrates LSTM-based numerical modeling with robust NLP-driven sentiment analysis and real-time visualization to support more accurate and actionable stock market forecasting.

III. SYSTEM ARCHITECTURE

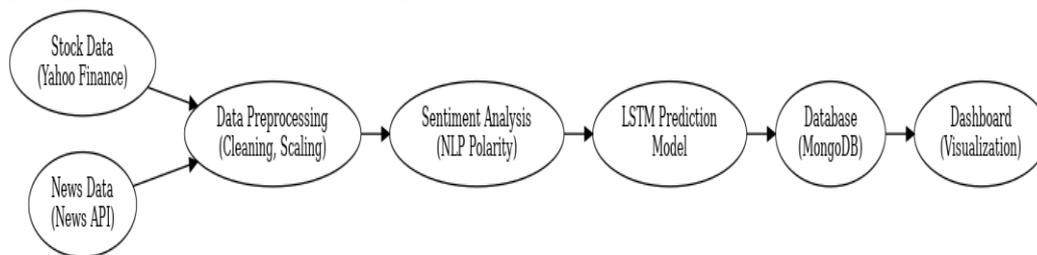The modular architecture of the proposed project is illustrated in Figure 1.

Figure 1: System Architecture

Different modules involved are:
- Data Collection Module
  This module fetches historical and live stock market data such as opening price, closing price, high, low, and volume from Yahoo Finance. Simultaneously, it retrieves real-time financial news headlines related to the selected stock from news APIs. This ensures that both numerical trends and qualitative information are captured.
- Data Preprocessing Module
  The collected stock data is cleaned, normalized, and transformed into a time-series format suitable for LSTM input. News headlines

undergo text preprocessing steps such as tokenization, removal of stop words, and elimination of irrelevant characters. This module ensures data consistency and quality before analysis.
- Sentiment Analysis Module
  The sentiment analysis module applies NLP techniques to determine the polarity of financial news. Each headline is classified as positive, negative, or neutral, and corresponding sentiment scores are generated. These scores numerically represent market sentiment and are later merged with stock price data.

➕ Prediction Module

The prediction module is the core of the system. It employs an LSTM-based deep learning model to learn historical stock price patterns. Sentiment scores are injected as additional features to enhance prediction accuracy during news-driven market movements.

➕ Database Management Module

All processed data, sentiment scores, and prediction results are stored in a structured database. This enables efficient data retrieval, historical analysis, and visualization.

➕ Visualization and Dashboard Module

This module displays real-time stock trends, sentiment indicators, and predicted prices using interactive charts and graphs. It serves as the user interface for monitoring stock behavior and prediction outcomes.

All modules work together to collect market data, analyze financial news, predict stock prices and display reliable insights for investors and users.

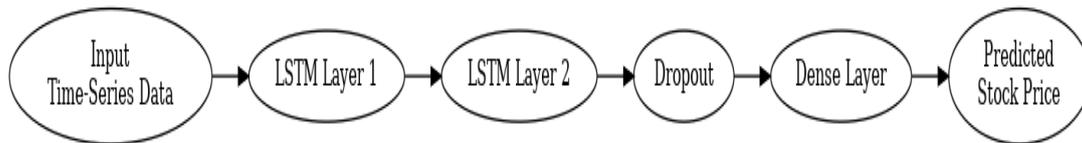The architecture of the LSTM model used in the project is shown in Figure 2.

Figure 2: LSTM Architecture

The architecture begins by feeding the time-series closing price dataset into stacked LSTM layers that analyze long-term dependencies in stock movement. The output from the LSTM network then flows to dense layers which refine the extracted features into a single predicted closing price. Activation functions and dropout layers are used to avoid overfitting, while the Mean Squared Error (MSE) loss function optimizes prediction learning. The final output node provides the predicted stock price for the next day.

The architecture of the sentiment-processing model is given in Figure 3.
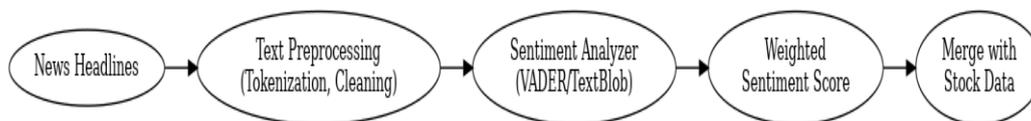

Figure 3: Sentiment Analysis Architecture

The sentiment analysis pipeline begins with raw financial news headlines. These headlines are first cleaned and tokenized to remove irrelevant information. A polarity detection algorithm then assigns sentiment values representing positive, negative, and neutral emotions. Weighted sentiment scores are calculated to represent overall market mood for a given trading day. These numerical sentiment values are synchronized with stock price data and appended as additional features. This architecture ensures that psychological and emotional aspects of the market are effectively captured and utilized by the prediction model.

IV. IMPLEMENTATION

Algorithm: News-Driven Stock Price Prediction System

Input:
- Stock ticker symbol
- Date range
- News sources

Output:
- Predicted stock price
- Evaluation metrics (RMSE, MAPE)
- Visual analytics

BEGIN

1. INITIALIZE system environment
   LOAD required libraries for data processing, NLP, ML, and visualization

2. ACCEPT user input
   INPUT stock_ticker
   INPUT start_date, end_date

// -------------------- DATA ACQUISITION ------------------------

3. CALL fetch_stock_data(stock_ticker, start_date, end_date)
   a. Retrieve Open, High, Low, Close, Volume data from Yahoo Finance
   b. Store stock data in dataframe STOCK_DF

4. CALL fetch_news_data(stock_ticker, start_date, end_date)
   a. Scrape financial news headlines from trusted sources
   b. Store headlines with timestamps in NEWS_DF

// -------------------- DATA PREPROCESSING --------------------

5. CALL preprocess_stock_data(STOCK_DF)
   a. Handle missing values
   b. Normalize numerical features
   c. Create lag features if required

6. CALL preprocess_news_data(NEWS_DF)
   a. Remove stop words and punctuation
   b. Convert text to lowercase
   c. Tokenize headlines

// -------------------- SENTIMENT ANALYSIS --------------------

7. FOR each headline in NEWS_DF DO
   a. CALL sentiment_analyzer(headline)
   b. Compute sentiment score (positive, negative, neutral, compound)
   c. Store sentiment score with corresponding date
  END FOR

8. CALL aggregate_daily_sentiment(NEWS_DF)
   a. Combine multiple headlines per day
   b. Generate daily sentiment index

// -------------------- FEATURE ENGINEERING --------------------

9. CALL merge_features(STOCK_DF, SENTIMENT_DF)
   a. Align stock prices and sentiment using date
   b. Add macroeconomic indicators (optional)
   c. Create final feature matrix FEATURES_DF

// -------------------- TRAIN–TEST SPLIT --------------------

10. SPLIT FEATURES_DF into:
   TRAIN_SET (80%)
   TEST_SET (20%)

// -------------------- MODEL BUILDING --------------------

11. INITIALIZE LSTM model
   a. Define input shape
   b. Add LSTM layers
   c. Add Dense output layer

12. TRAIN LSTM model using TRAIN_SET
   a. Optimize loss function (MSE)
   b. Use backpropagation through time

13. INITIALIZE Random Forest Regressor (optional)
14. TRAIN Random Forest using TRAIN_SET sentiment features

// -------------------- PREDICTION --------------------
15. CALL predict_stock_price(LSTM_MODEL, TEST_SET)
   a. Generate predicted prices
   b. Store predictions

// -------------------- MODEL EVALUATION --------------------
16. CALL evaluate_model(PREDICTIONS, ACTUAL_VALUES)
   a. Compute RMSE
   b. Compute MAPE
// -------------------- VISUALIZATION --------------------
17. CALL plot_results()
   a. Plot actual vs predicted stock prices
   b. Plot sentiment trend vs price movement
   c. Generate performance graphs
// -------------------- OUTPUT --------------------
18. DISPLAY predicted prices and evaluation metrics
19. DISPLAY visualization dashboards
END

V. RESULTS AND DISCUSSIONS

Figure 4 illustrates an AAPL Intelligent Dashboard that integrates real-time market data, sentiment analysis, and LSTM-based forecasting into a single visual interface. The top summary panel presents key indicators: the current live price, the short-term next-minute trend, the LSTM-predicted closing price, and aggregated news sentiment. This layout allows users to quickly assess both quantitative signals (prices and predictions) and qualitative signals (news sentiment), helping bridge traditional technical analysis with AI-

driven insights. The sentiment score and label (e.g., Positive) indicate how recent news may be influencing market psychology, complementing the numerical forecasts.

The main chart visualizes historical intraday price movement alongside the LSTM forecast trajectory. The dashed trend line and highlighted prediction point indicate the model's expectation of near-future price direction based on learned temporal patterns. By overlaying predictions on actual price action, the dashboard makes it easy to compare model behavior with real market movement. Overall, the dashboard demonstrates how deep learning (LSTM) combined with sentiment signals can support informed, data-driven decision making by offering an interpretable, real-time view of expected price trends rather than relying solely on past prices or static indicators.
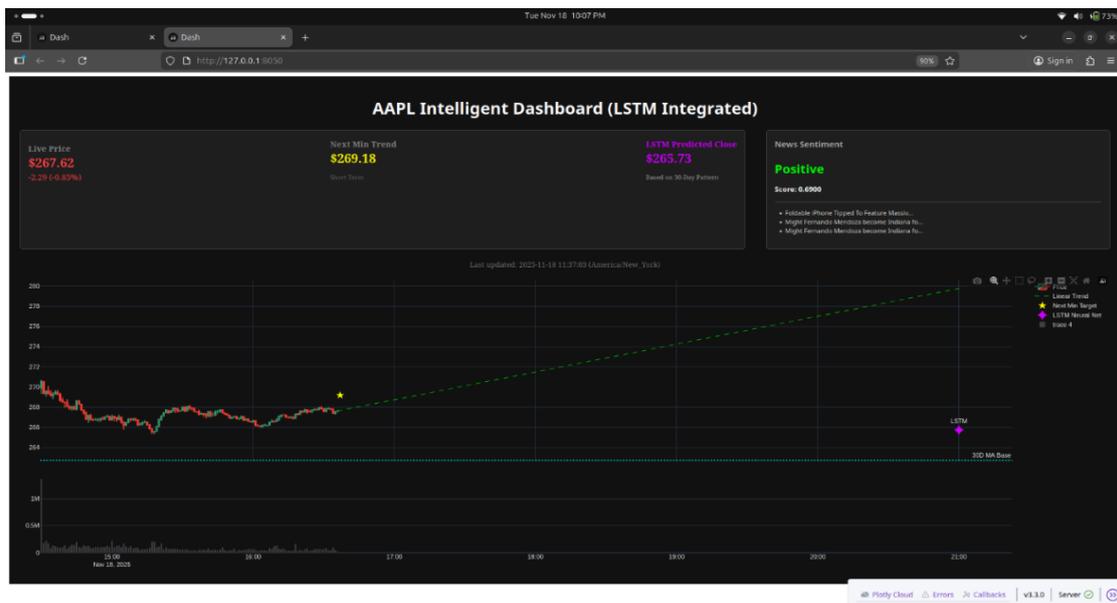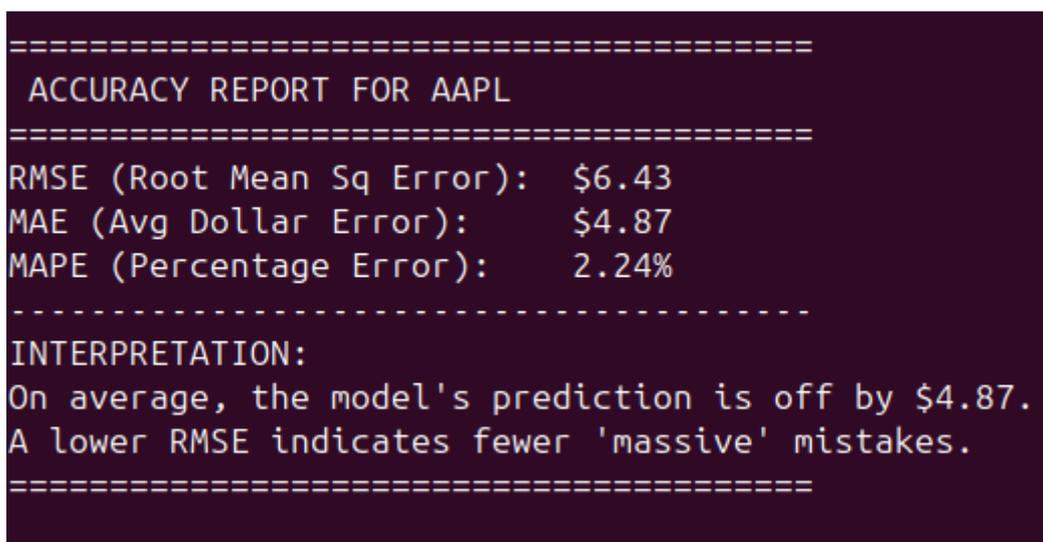

Figure 4: Dashboard Results

Figure 5 Accuracy

Figure 5 shows an accuracy report for AAPL that summarizes how close your model's predicted prices are to the actual market prices, using three standard regression error metrics:

RMSE (Root Mean Square Error) = $6.43

- 🔱 RMSE measures the *typical size of prediction error*, but it penalizes large errors more heavily because the errors are squared before averaging.
- 🔱 Interpretation: Most predictions are reasonably close, but if the model makes a few big mistakes, RMSE will rise quickly.
- 🔱 In your case, $6.43 means the model's errors (in dollars) are usually around that scale, with extra weight given to larger misses. This is why RMSE is often used to judge whether the model avoids "massive" wrong predictions.

MAE (Mean Absolute Error) = $4.87

- 🔱 MAE measures the average absolute difference between predicted and actual prices.
- 🔱 It is more straightforward than RMSE because it treats all errors equally (no squaring).
- 🔱 Interpretation: On average, your predicted AAPL price is off by about $4.87 per prediction. This matches the report line: *"On average, the model's prediction is off by $4.87."*

MAPE (Mean Absolute Percentage Error) = 2.24%

- 🔱 MAPE expresses the average error as a percentage of the actual price, making it easier to compare performance across stocks with different price ranges.
- 🔱 Interpretation: A typical prediction deviates by about 2.24% from the true price.
- 🔱 Example: If AAPL is around $250, then 2.24% ≈ $5.60, which is close in scale to your MAE (~$4.87). That consistency is a good sign.

Observations

- 🔱 MAE ($4.87) tells you the *average miss* in dollars (easy to explain in viva).
- 🔱 RMSE ($6.43) being higher than MAE is normal and indicates there are some larger errors (outliers), but not extremely huge ones.
- 🔱 MAPE (2.24%) suggests the model is reasonably accurate in relative terms (low percentage deviation), especially for short-term forecasting.
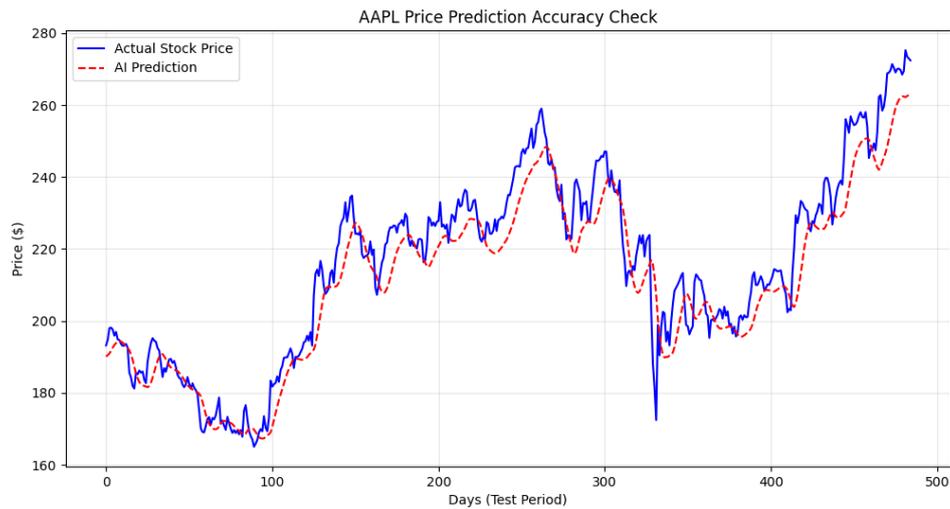

Figure 6 Accuracy graph

Figure 6 illustrates the AAPL stock price prediction accuracy by comparing the actual market prices (solid blue line) with the AI-based predicted prices (red dashed line) over the test period. The horizontal axis represents time in days, while the vertical axis shows the stock price in US dollars. The close alignment between the two curves across most of the timeline indicates that the model has effectively learned the underlying price trends and short-term fluctuations from historical data combined with news-based features.

A key observation is that the AI prediction smoothly follows the overall upward and downward movements of the actual stock price, capturing major trend reversals and sustained growth phases. Minor deviations occur during sudden price drops or sharp spikes, which are typical in stock markets due to unexpected news or high volatility. However, even in these cases, the predicted curve quickly converges back toward the actual price trajectory. This behavior demonstrates the robustness of the LSTM-based model in handling time-series data and confirms its

ability to provide reliable short-term stock price forecasts rather than exact point-to-point matching.

Overall, the graph visually validates the quantitative performance metrics (low RMSE, MAE, and MAPE) reported earlier. It shows that the model is well-suited for trend prediction and decision support, helping investors understand market direction and sentiment-driven movements, while acknowledging that exact prediction of sudden market shocks remains inherently challenging.

VI. CONCLUSION AND FUTURE SCOPE

This work combines historical market data with financial news sentiment, the system bridges the gap between numerical market trends and psychological market reactions. Designed for investors and market researchers, the system enhances stock prediction by providing real-time insights into price fluctuations along with the sentiment impact of news. Through its interactive dashboard, users can easily observe market trends, sentiment levels and the predicted closing price, enabling more informed and confident investment decisions.

This work also highlights the importance of market sentiment in financial forecasting and demonstrates how sentiment-driven predictive models can outperform traditional price-only models, especially during volatile market conditions. The system lays a solid foundation for intelligent and sentiment-aware stock market analytics, contributing to smarter, more adaptive investment decision-making.

The future of the stock price prediction system lies in expanding its capabilities to handle multiple stocks and multiple financial news sources simultaneously, further enhancing prediction reliability. Integration of multilingual sentiment analysis would allow the system to process international market news, improving global applicability. Additionally, incorporating advanced deep learning architectures such as Transformer-based models could strengthen forecasting accuracy during extreme market volatility.

BIBLIOGRAPHY

[1] Ali Hassanzadeh et al., Stock Market Prediction using Daily News Headlines, 2020.

[2] Junaid Ahmed et al., Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach, 2023.

[3] Rajkar, A., Kumaria, A., Raut, A., & Kulkarni, N. (2021). Stock Market Price Prediction and Analysis. *International Journal of Engineering Research & Technology (IJERT)*, *10*(06), 47-51.

[4] Eswar Reddy et al., Prediction of Financial Stock Market Based on Machine Learning Technique, 2023.

[5] Narayana Rao et al., Stock Price Prediction using Sentiment Analysis and Deep Learning for Indian Markets, 2022.

[6] Harmanjeet Singh et al., Analysis of Stock Market Prediction Models Using Deep Learning, 2021.

[7] Deepak Kumar et al., A Systematic Review of Stock Market Prediction using Machine Learning and Statistical Techniques, 2021.

[8] Malti Bansal et al., Stock Market Prediction with High Accuracy using Machine Learning Techniques, 2022.

[9] X. Pang et al., An Innovative Neural Network Approach for Stock Market Prediction, 2018.

[10] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, Financial Time Series Forecasting with Deep Learning: A Systematic Literature Review, 2020.

[11] P. C. Tetlock, Giving Content to Investor Sentiment: The Role of Media in the Stock Market, 2007.

[12] J. Bollen, H. Mao, and X. Zeng, Twitter Mood Predicts the Stock Market, 2011.

[13] X. Ding et al., Deep Learning for Event-Driven Stock Prediction, 2015.