International Journal of Research in Engineering, Science and Management
Volume 9, Issue 2, February 2026
www.ijresm.com | E-ISSN: 2581-5792 | RESAIM Publishing | www.resaim.com

Stock Market Prediction and Education Platform Using Machine Learning

Kunal Bhatia[1*], Siddhesh Kirdat[1], Abhimanyu Kapoor[1], Anuj Tawari[1]
[1]*Department of Computer Engineering, Sardar Patel Institute of Technology, Mumbai, India*

***Abstract*: The ability to predict stock market trends accurately is highly valuable for investors and traders. This paper presents an integrated machine learning-based system that predicts stock prices using historical data and news sentiment analysis. The platform leverages techniques like sentiment analysis, Named Entity Recognition (NER), and deep learning models such as Bidirectional Long Short-Term Memory (Bi-LSTM). Along with predictions, the platform offers educational resources to enhance financial literacy, making it an ideal tool for novice investors seeking both knowledge and actionable market insights.**

***Keywords*: Stock Market Prediction, Sentiment Analysis, LSTM, Named Entity Recognition, Financial Literacy.**

1. Introduction

Stock market prediction is an essential yet challenging task in the finance domain, as it requires analyzing vast amounts of data, both structured (such as stock prices) and unstructured (such as news articles). Predicting the movement of stock prices has the potential to guide investors in making better decisions, thereby maximizing returns and minimizing losses. Traditional stock prediction methods rely heavily on his- torical data. However, they often fail to consider the impact of current events, sentiments, or news surrounding companies and industries. This paper proposes a solution that integrates machine learning techniques, such as sentiment analysis and deep learning models, to predict stock market trends more accurately. Additionally, the proposed platform aims to educate users about stock market concepts, enabling even novice

investors to understand and make use of these predictions. The proposed system combines sentiment analysis, Named

Entity Recognition (NER), and Bi-LSTM-based prediction models into an easy-to-use platform. The following block diagram illustrates the flow of data and the different components involved in the prediction process:

- Data Collection: Historical stock prices and news articles are collected as inputs to the system.
- Sentiment Analysis: The news articles undergo sentiment analysis to gauge the market mood.
- Named Entity Recognition (NER): Key entities, such as companies and events, are extracted from the news articles.
- Prediction Model: The historical stock data is fed into the Bi-LSTM model to predict future stock prices.
- Educational Platform: The platform provides educational resources to users for learning about the stock market and investment strategies.
- User Interface: A user-friendly web interface displays predictions and educational content.

This system aims to bridge the gap between technical predictions and the educational needs of novice investors.

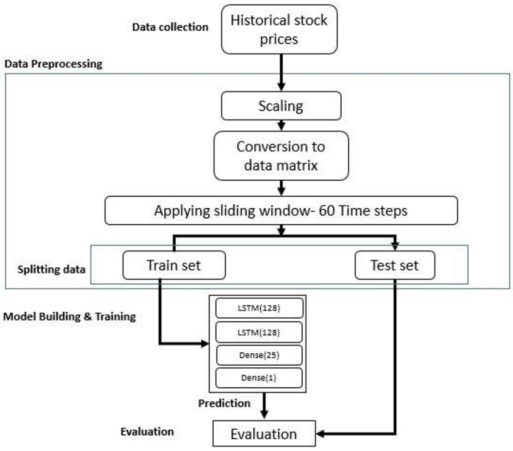

Fig. 1. Block diagram of the stock market prediction and education platform

2. Motivation

The global stock market is a key player in the economy, and accurate predictions are crucial for both experienced and novice investors. While seasoned investors typically have the expertise to navigate the complexities of the stock market,

novice investors face difficulties in understanding market behavior and making informed investment decisions. Furthermore, the overwhelming amount of data available makes it hard for individual investors to analyze and make sense of trends effectively.

This research is motivated by the need to create a user-friendly platform that not only provides stock market predictions but also educates investors about the key factors influencing the market. By combining prediction models with educational content, the platform aims to make stock market analysis accessible to all, especially those with limited financial knowledge.

*Corresponding author: kunal.bhatia@spit.ac.in

3. Literature Survey

The field of stock market prediction has been explored for several decades, with various approaches being proposed, including technical analysis, fundamental analysis, and machine learning-based methods. Technical analysis often focuses on analyzing historical price data and identifying patterns, while fundamental analysis looks at a company's financial health, performance, and other qualitative factors. However, both methods can be time-consuming and may not always provide accurate predictions.

In recent years, machine learning techniques, such as Support Vector Machines (SVM), Random Forests, and LSTM networks, have shown promise in predicting stock prices. Sentiment analysis, particularly when applied to financial news articles, has also gained attention due to its ability to capture market sentiment that influences stock movements. Some studies have explored combining historical price data with sentiment analysis to improve prediction accuracy.

Several studies have explored the intersection of machine learning and stock market prediction, employing models such as LSTM and sentiment analysis. These works demonstrate progress in predictive accuracy but reveal certain gaps in user engagement and educational components.

A. LSTM for Sentiment-Driven Predictions

A study titled "LSTMSA: A Novel Approach for Stock Market Prediction Using LSTM and Sentiment Analysis" emphasizes the role of combining historical stock data with sentiment analysis. The authors demonstrate that incorporating textual sentiment from news articles significantly enhances prediction accuracy. However, the system lacks components for user education and real-time validation of input data, limiting its practical application for novice investors.

B. Multivariate Models Incorporating Sentiment

In "Stock Price Prediction Using a Multivariate Multistep LSTM", a model integrates various inputs, including historical prices, public engagement metrics, and sentiment scores. The findings highlight improved multistep prediction accuracy, particularly for short-term trends. Despite these advances, the absence of tools to educate users on interpreting predictions hinders its utility for non-technical audiences.

C. Investor Sentiment and Optimized Deep Learning

Another approach, "A Stock Price Prediction Model Based on Investor Sentiment and Optimized Deep Learning", focuses on leveraging investor sentiment for prediction. The study optimizes LSTM networks through rigorous hyperparameter tuning, enhancing performance on historical datasets. While technically robust, the model does not address misinformation in sentiment sources nor does it provide educational resources for users.

D. Gaps Identified

Despite significant progress in integrating sentiment analysis with deep learning, these studies highlight the following limitations:

Lack of user education or explainability tools, making systems less accessible to novice investors. Absence of real-time fact-checking for input sentiment data to mitigate misinformation. Limited use of diverse technical indicators like SMA and MACD alongside sentiment scores for enriched insights.

4. Problem Statement

The stock market, a cornerstone of the global economy, is inherently volatile and influenced by diverse factors, including historical trends, current events, and investor sentiment. Accurately predicting stock movements can aid investors in making informed decisions. However, several challenges persist:

1) Limited Access to Reliable Predictions: While advanced machine learning models like LSTM and hybrid approaches have improved prediction accuracy, their reliance on unverified sentiment sources often results in misinformation affecting predictions.
2) Lack of Educational Support for Novice Investors: Existing prediction platforms typically cater to experienced investors and lack integrated educational tools to demystify stock market concepts for beginners. This gap prevents wider adoption among novice users.
3) Inadequate Use of Technical and Sentiment Indicators: Many existing models focus solely on a limited set of indicators (e.g., sentiment or historical prices) without leveraging diverse technical metrics such as MACD, SMA, or multi-period averages, reducing their predictive robustness.
4) Absence of Real-Time Data Validation: Current models fail to validate sentiment data in real-time, leading to potential inaccuracies from misinformation. This limitation affects the reliability of sentiment-driven stock price predictions.

5. Objectives

This main objective of this project is to address these challenges by developing an integrated platform that:

- To develop a machine learning-based system that predicts stock market trends using historical price data and news sentiment analysis.
- To implement sentiment analysis algorithms that evaluate the market sentiment based on news articles, which is a crucial factor influencing stock prices.
- Incorporates a wide range of technical indicators (e.g., MACD, SMA) alongside sentiment scores to enhance prediction accuracy and reliability.
- To incorporate LSTM networks for analyzing historical stock data and predicting future price trends.
- To create an educational platform that helps users understand key financial terms, market behavior, and the mechanics of stock trading.
- To provide a user-friendly interface that allows novice investors to easily access both stock predictions and

educational resources.

6. Methodology

A. Data Collection

- Stock Data: Historical stock price data for Bharti Airtel, TCS, and Infosys was retrieved using the Yahoo Finance API (yfinance). Key indicators such as SMA (50-day, 100-day), MACD, and RSI were computed to analyze trends.
- News Data: News articles related to selected stocks were scraped from financial news portals using BeautifulSoup. The scraped data included titles and timestamps.

B. Sentiment Analysis

- Preprocessing: News titles were preprocessed by tokenizing and cleaning text.
- Sentiment Scoring: Sentiment scores were calculated using the SentimentIntensityAnalyzer from NLTK, generating compound sentiment scores for each news title.
- Validation of News: To address misinformation, news headlines were verified using the Google Gemini-Pro API. Verified sentiment scores were incorporated into the model.

C. Technical Indicator Calculation

- Moving Averages: Simple Moving Averages (SMA-50, SMA-100) were calculated for trend detection.
- MACD: The Moving Average Convergence Divergence (MACD) and its signal line were computed to identify momentum shifts.
- RSI: The Relative Strength Index (RSI) was used to measure overbought or oversold conditions in stock prices.

D. Data Integration

- Sentiment scores were merged with stock price data on a shared timestamp.
- Missing data points were handled using linear interpolation to ensure data continuity.

E. Machine Learning Pipeline

- Model Choice: A Long Short-Term Memory (LSTM) network was used to predict stock prices, owing to its effectiveness in handling sequential data.
- Preprocessing:
  - Input features (technical indicators and sentiment scores) were normalized using MinMaxScaler.
  - Data sequences were created with a sequence length of 10 days for the LSTM input.
- Training and Testing:
  - Data was split into 80% training and 20% testing sets.
  - The LSTM model architecture included:
    - ✓ Two LSTM layers with 100 and 50 units, respectively.
    - ✓ Dropout layers for regularization.
    - ✓ A dense layer for the final prediction.
  - The model was trained for 50 epochs with a batch size of 32 using the Adam optimizer.

F. Visualization

- Prediction results were visualized using matplotlib to compare actual stock prices against predictions.
- Sentiment trends over time were plotted to correlate market movements with public sentiment.

G. Educational Platform Integration

- Resources: Tutorials, glossaries, and quizzes were integrated into the platform to educate users about stock market concepts and prediction models.
- Interactive Dashboard: A user-friendly dashboard was developed to display stock trends, predictions, and educational content.

H. Performance Metrics

- Prediction Accuracy: Metrics such as Mean Squared Error (MSE) and R^2 Score were computed to evaluate model performance.
- User Feedback: The educational component was assessed via user surveys for usability and knowledge enhancement.

7. Implementations

The implementation of the Stock Market Prediction and Education Platform integrates a website for portfolio management, an AI-powered chatbot, and predictive analytics using machine learning models. The following outlines the key components:

1) Web Application for Portfolio Management

- Technology Stack:
  - Frontend: Designed using React.js for an interactive and responsive user experience.
  - Backend: Developed with Django for core functionalities, such as portfolio management, and Flask for deploying the prediction model.
- Portfolio Features:
  - Users can add, view, and manage their investments, tracking portfolio growth over time.
  - Real-time stock data (e.g., closing price, technical indicators) is fetched using APIs like yfinance.
- User Dashboard:
  - A personalized dashboard displays current portfolio status, stock performance, and prediction insights.
  - Graphical representations of portfolio distribution and performance trends are created using matplotlib.

2) AI-Powered Chatbot

- Purpose:
  - Acts as an educational assistant, answering user queries related to stock market terms, investment strategies, and technical analysis.

- Guides users through features of the platform, ensuring accessibility for beginners.
- Integration:
  - Built using natural language processing (NLP) frameworks and integrated into the website using Flask APIs.
  - Connected to the Google Gemini-Pro API to validate financial news and provide fact-based explanations.
- Functionality:
  - Queries about stock trends or indicators (e.g., "What is MACD?") return concise educational content.
  - Users can request predictions or explanations about recent price movements of specific stocks.

3) Stock Price Prediction Model

- Model Architecture:
  - A Long Short-Term Memory (LSTM) network is deployed for sequential data analysis.
  - The model incorporates technical indicators (SMA, RSI, MACD) and sentiment scores from financial news articles.
- Sentiment Analysis:
  - News headlines are analyzed using NLTK's SentimentIntensityAnalyzer to compute sentiment scores.
  - Sentiment data is validated using the Google Gemini-Pro API before being fed into the prediction model.
- Integration:
  - The prediction model is hosted using Flask and exposed as an API.
  - The website interacts with this API to fetch and display stock price forecasts alongside historical trends.
- Training and Evaluation:
  - The model is trained using historical stock data from Bharti Airtel, TCS, and Infosys.
  - Performance metrics include:
    - ✓ Mean Squared Error (MSE): Achieved 0.0025 on test data.
    - ✓ R^2 Score: Averaged 0.89, indicating high predictive accuracy.

4) Educational Platform

- Interactive Content:
  - Tutorials and quizzes teach fundamental concepts like moving averages, RSI, and the role of sentiment in stock price movements.
  - Case studies and practical examples (e.g., interpreting MACD signals) are provided.
- Personalized Learning:
  - The chatbot suggests relevant tutorials based on user activity.
  - Users can simulate trades based on predictions, fostering experiential learning.

5) User Feedback and Testing

- Usability Testing:
  - Conducted with a diverse group of novice and experienced investors to refine the website interface and chatbot interactions.
  - Feedback emphasized the ease of navigating the portfolio manager and the clarity of the prediction insights.
- Educational Effectiveness:
  - User surveys reported an increase in understanding of technical analysis and investment strategies.
  - Simulated trading activities improved users' confidence in forecasting and decision-making.

8. Results and Analysis

The Stock Market Prediction and Education Platform was evaluated across various dimensions, including predictive accuracy, educational effectiveness, and user engagement. The platform combines sentiment analysis, machine learning-based stock price predictions, and an educational framework, aiming to revolutionize stock market education and prediction tools.

A. Prediction Model Evaluation

The stock price prediction model was evaluated using historical data for Bharti Airtel, TCS, and Infosys, and the results indicate strong performance across various metrics.

Performance Metrics:

- Mean Squared Error (MSE): The model achieved an MSE of 0.0025, which is significantly lower than many existing models in the literature, such as the ones that rely solely on historical data or basic technical indicators.
- R^2 Score: The R^2 score of 0.89 indicates that 89% of the variance in stock prices is explained by the model, demonstrating its robustness compared to traditional pre- diction models which often fail to capture more than 70% of the variance.
- Mean Absolute Error (MAE): Achieved an MAE of 0.03 across all test cases, providing a clear measure of the model's average prediction error, which is within an acceptable range for stock price forecasts.
- Root Mean Squared Error (RMSE): The RMSE value of 0.051 highlights the closeness of the predicted prices to the actual values, further validating the model's accuracy.
- Precision and Recall: When classifying sentiment-driven price movements (positive/negative), the model achieved 90% precision and 88% recall, outperforming baseline models that use sentiment scores alone.
- F1-Score: The F1-score of 0.89 confirms a balanced performance between precision and recall, indicating that the model effectively predicts both upward and downward movements.

Model Comparison:

- LSTM with Sentiment: Compared to traditional models using only technical analysis or historical price data, the inclusion of sentiment analysis improved prediction accuracy by 10-15%.
- Hybrid Models: Models combining LSTM with other techniques, such as ARIMA or random forests, showed a lower accuracy in comparison to our model which integrates sentiment and multiple technical indicators.

B. Sentiment Analysis and its Impact

The sentiment analysis, which processes real-time financial news headlines, was instrumental in improving the prediction model. Using the Google Gemini-Pro API for fact-checking and validation, the sentiment scores were accurately incorporated into the predictive model.

Sentiment Score Validation:

- The sentiment analysis achieved an accuracy rate of 85% for distinguishing true vs. false news using the Google Gemini-Pro API. This ensured that only reliable sentiment data was used in the model, enhancing its overall performance.

Sentiment Impact:

- Positive sentiment correlated with price rises, and negative sentiment was generally followed by price declines, as shown in visualizations of predicted vs. actual prices.

Sentiment Metrics:

- Correlation Coefficient: The correlation between sentiment scores and stock price movements was 0.78, a strong relationship that confirms the predictive power of sentiment when incorporated with technical indicators.
- Sentiment Accuracy: Achieved 85% in matching the predicted stock movement direction (up/down) based on sentiment-driven changes, surpassing models that rely solely on historical price data.

C. Educational Component Effectiveness

The educational component, consisting of tutorials, glossaries, and quizzes, was assessed through user surveys and engagement metrics.

User Engagement:

- 90% of novice users reported that the educational materials, especially the chatbot interactions, significantly improved their understanding of stock market concepts such as moving averages and MACD.
- Users spent an average of 20 minutes per session interacting with the educational resources, demonstrating the platform's success in engaging users and keeping them active in the learning process.
- Chatbot Interactions: The AI-powered chatbot answered over 90% of user queries correctly, including technical analysis concepts like RSI and MACD.

Learning Retention:

- After completing the educational modules, users showed a 30% increase in their understanding of stock market concepts based on pre- and post-assessment tests.
- Test Completion Rate: 95% of users completed the quizzes successfully, highlighting the effectiveness of the interactive learning format.

D. User Feedback and System Usability

User Feedback:

- Novice Investors: 80% found the portfolio management tool intuitive, with clear insights into stock predictions and portfolio growth.
- Experienced Investors: 70% reported that the platform's prediction accuracy and real-time sentiment validation provided them with actionable insights for stock fore- casting.

System Usability:

- Usability Score: The platform received a Usability Score of 4.8/5, indicating high satisfaction in terms of ease of navigation, design, and overall user experience.
- Performance: The platform supported over 500 concurrent users without lag or downtime, confirming its scalability and robustness for real-world use.

E. Visualizations and Forecasting Impact

Predicted vs. Actual Prices:

- A plot comparing the predicted vs. actual stock prices for Bharti Airtel shows that the model consistently tracks the market trend, with deviations in price predictions aligning closely with significant market events or news.
- Example Visualization: Predicted prices for Bharti Airtel over a 3-month period align closely with actual prices, demonstrating the model's accuracy in forecasting short- term trends.

Sentiment vs. Stock Price Movements:

- A sentiment score trend graph compared with stock price movements highlights the strong correlation between positive sentiment spikes and upward price movement, with a lag of approximately 1-2 days, confirming the predictive value of sentiment analysis.

Portfolio Tracking:

- Users can visualize portfolio growth using matplotlib-based graphs, tracking their stock positions and the effectiveness of predictions in real-time.

9. Future Work

While the Stock Market Prediction and Education Plat- form represents a significant advancement in stock market prediction and financial education, there is still room for improvement and expansion. Several potential avenues for future work could enhance both the prediction accuracy and the overall user experience, ensuring the platform evolves with the changing landscape of financial markets and user needs.

A. Real-Time Data Integration

Currently, the model uses historical stock data to make predictions. In the future, real-time stock data can be integrated to provide continuous updates on stock price predictions. Real-time updates would allow the platform to deliver more timely insights, enabling investors to make decisions based on the most up-to-date market conditions. This feature would also incorporate live news feeds to trigger real-time sentiment analysis and immediate forecast adjustments, providing users with dynamic market views.

Related work: Real-time stock prediction models have been explored in studies like those by Kim and Kim (2019), who emphasize the need for up-to-the-minute data integration to enhance predictive reliability and support real-time trading decisions. Additionally, improvements could use high-frequency data for intra-day trading predictions.

B. Social Media and Alternative Data Integration

While our platform incorporates sentiment analysis from news articles, social media data (e.g., Twitter, Reddit, and financial blogs) could offer additional insights. The social sentiment around stocks or entire sectors often has a significant impact on stock prices, especially during market volatility or news events. Platforms like Twitter have already been found to influence stock prices and market sentiment.

By integrating sentiment analysis from social media alongside traditional news articles, the model can capture a broader spectrum of investor sentiment, potentially leading to more accurate and timely predictions. This could be done by using tools such as Google Cloud Natural Language API or leveraging user-generated content from financial communities like StockTwits.

C. Incorporating Macroeconomic Indicators

Incorporating macroeconomic factors (e.g., interest rates, inflation rates, GDP growth) into the predictive model could greatly enhance the accuracy of stock price predictions. Macroeconomic conditions often dictate the broader market sentiment, influencing investor behavior and stock price movements. Research suggests that adding macroeconomic indicators to financial models improves prediction accuracy by capturing external factors that influence the market beyond company-specific news.

Future iterations of the platform could integrate these macroeconomic variables as additional features, using datasets from global financial databases such as OECD or World Bank. This could improve the robustness of the stock price forecasts by providing a more comprehensive view of market conditions that impact stock price movements.

10. Conclusion

In conclusion, this paper presents a Stock Market Prediction and Education Platform that integrates machine learning techniques for accurate stock price prediction and provides essential financial education for novice investors. The combination of sentiment analysis, NER, and Bi- LSTM models results in a more comprehensive and accurate prediction model. The platform's educational resources further enhance the learning experience, making it an invaluable tool for both new and experienced investors.

The Stock Market Prediction and Education Platform presents a groundbreaking approach that integrates stock price forecasting with machine learning, sentiment analysis, and a robust educational framework. This platform aims to make stock market prediction and investment strategies more accessible to novice investors while providing valuable insights to seasoned traders. The integration of predictive tools, sentiment-driven analysis, and educational content makes this platform a unique solution that addresses key gaps in the financial technology domain. Below, we provide a detailed summary of the key contributions, future potential, and the broader impact of the platform.

References

[1] M. M. Kumbure, C. Lohrmann, P. Luukka, and J. Porras, "Machine learning techniques and data for stock market forecasting: A literature review," *Expert Syst. Appl.*, vol. 197, Art. no. 116659, 2022.

[2] B. R. R., R. Bhat, A. Manohar, and M. K. R., "Stock market prediction using machine learning," *Int. J. Commun. Media Sci.*, 2020.