DEEP LEARNING-DRIVEN HYBRID MODEL FOR STOCK PRICE PREDICTION USING HISTORICAL AND REAL-TIME DATA

First Author: Ch. Chandra Sekar, Department of AIML, PBR Visvodaya Institute of Technology and Science (Autonomous), Kavali, Nellore.

Second Author:

T. Bharath Chandu (Roll No:214N1A3952), Department of AIML, PBR Visvodaya Institute of Technology and Science (Autonomous), Kavali, Nellore.

S. Venkata Teja Reddy (Roll No:214N1A3947), Department of AIML, PBR Visvodaya Institute of Technology and Science (Autonomous), Kavali, Nellore.

P. Gnana Sai Sampath (Roll No:214N1A3933), Department of AIML, PBR Visvodaya Institute of Technology and Science (Autonomous), Kavali, Nellore.

V. Narahari (Roll No:214N1A3955), Department of AIML, PBR Visvodaya Institute of Technology and Science (Autonomous), Kavali, Nellore.

B. Venkata Kalyan (Roll No:214N1A3903), Department of AIML, PBR Visvodaya Institute of Technology and Science (Autonomous), Kavali, Nellore.

ABSTRACT

The study is a deep learning-driven hybrid model is developed for stock price prediction by integrating historical market data with real-time news sentiment analysis. The system leverages a combination of classical time series models (ARIMA, Linear Regression) and advanced deep learning architectures (LSTM-GRU hybrid networks) to accurately forecast future stock prices. Real-time financial news headlines are retrieved from Yahoo Finance and analyzed using Natural Language Processing (TextBlob) to classify sentiment as Positive, Negative, or Neutral, enriching the prediction model with external market factors. The resulting forecasts, along with actionable investment recommendations (Buy, Sell, or Hold), are delivered through an interactive, web-based application built using the Flask framework. Model performance is rigorously evaluated using metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), demonstrating that the hybrid LSTM-GRU model achieves superior accuracy, especially in capturing non-linear and volatile market behaviors. This project highlights the effectiveness of combining historical data, deep learning techniques, and real-time sentiment analysis to build a robust, adaptive, and user-friendly stock forecasting system.

1.INTRODUCTION

The prediction of stock prices has long been a central focus of research in the fields of finance, economics, and artificial intelligence. Stock markets are inherently complex, influenced by numerous factors ranging from company fundamentals and macroeconomic indicators to investor sentiment and global events. As a result, accurate forecasting of stock prices presents a significant challenge due to the non-linear, volatile, and highly dynamic nature of financial time series data.

Traditional stock prediction methods, including statistical models like Autoregressive Integrated Moving Average (ARIMA) and Linear Regression, have been widely employed to model stock behavior. While these techniques offer interpretability and ease of use, they are often limited by their assumptions of linearity, stationarity, and their inability to model long-term dependencies or sudden market shifts. Consequently, their performance in real-world, volatile market conditions is often suboptimal.

Recent advancements in Artificial Intelligence (AI) and Deep Learning (DL) have opened new avenues for modeling complex financial data. Deep neural networks, particularly Recurrent Neural Networks (RNNs) and their variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have demonstrated strong capabilities in learning intricate sequential patterns and capturing temporal dependencies. These models excel in handling non-linearities and long-term relationships, making them particularly suitable for financial time series forecasting.

However, relying solely on historical stock prices does not account for external real-time factors such as breaking news, company announcements, and broader market sentiment—all of which have immediate and significant impacts on stock movements. Hence, incorporating real-time sentiment analysis from financial news can provide valuable contextual information that enhances prediction accuracy and model robustness.

In this project, a hybrid, deep learning-based approach is proposed for stock price prediction. The system integrates historical stock market data with real-time financial news sentiment analysis to forecast stock prices more accurately. The prediction framework combines ARIMA, Linear Regression, and a hybrid LSTM-GRU deep learning model, enabling both traditional and modern perspectives on stock forecasting. Sentiment scores derived from news headlines using the TextBlob Natural Language Processing (NLP) library are merged with the stock data to further improve prediction performance.

A user-friendly, dynamic web application is developed using the Flask framework to allow users to interact with the system. Users can select publicly traded companies, visualize historical and predicted stock prices, and receive actionable investment recommendations, such as Buy, Sell, or Hold, based on model forecasts and sentiment trends.

Importance of Stock Price Prediction

The ability to predict stock prices with reasonable accuracy is critical for a wide range of stakeholders, including individual investors, financial institutions, portfolio managers, and policymakers. Successful stock price forecasting can lead to:

- **Informed Investment Decisions**: Accurate predictions help investors optimize their portfolios, maximizing returns while managing risks.
- **Risk Mitigation**: Early detection of negative trends or market downturns allows investors and institutions to take preventive measures.
- **Market Efficiency**: Improved forecasting techniques contribute to the overall efficiency of financial markets by facilitating better price discovery.
- **Economic Impact**: Since stock markets are closely linked with economic performance, better stock prediction tools can indirectly support economic growth by encouraging investment.

Despite these potential benefits, the unpredictable and complex behavior of financial markets makes precise prediction exceedingly difficult. Price movements are affected by an intricate interplay of rational factors (such as earnings and interest rates) and irrational influences (such as investor psychology and speculation). Thus, traditional statistical methods alone are often insufficient, and there is a strong need for intelligent, data-driven solutions that can adapt to real-world complexities.

Role of Deep Learning and Sentiment Analysis

Deep learning has revolutionized several fields, including computer vision, speech recognition, and natural language processing. Its strength lies in its ability to automatically learn high-level representations from raw data, uncovering hidden patterns that traditional models might miss.

In financial forecasting, deep learning models like LSTM and GRU have proven especially effective due to their ability to capture:

- Long-term dependencies in sequential data.
- Non-linear relationships between input features and output predictions.
- Temporal patterns that evolve over time.

Moreover, stock prices are not only driven by numerical data but also by qualitative information such as news articles, social media discussions, and analyst reports. News sentiment can dramatically influence stock prices in the short term. Positive news often boosts stock prices, while negative news can lead to sharp declines.

By integrating **real-time sentiment analysis** into the prediction pipeline, models can be made more responsive to sudden market changes. Natural Language Processing (NLP) techniques, like those employed by the TextBlob library, enable automatic extraction and classification of sentiments from textual data. Combining these sentiment scores with historical stock data enhances the model's contextual awareness and improves overall prediction performance.

Thus, the hybrid model in this project combines the best of both worlds: the historical trend analysis through LSTM-GRU and traditional models, and the immediate market pulse captured through real-time sentiment analysis.

Scope of the Project

The project focuses on developing a **scalable**, **accurate**, and **interactive** stock forecasting system, offering the following features:

- **Data Integration**: Combining historical stock data with real-time sentiment scores.
- **Hybrid Modeling Approach**: Utilizing ARIMA, Linear Regression, and a deep learning-based LSTM-GRU hybrid network.
- **Real-Time Sentiment Analysis**: Extracting news sentiment from Yahoo Finance headlines using TextBlob.
- **Web-Based Interface**: Providing a dynamic Flask application for visualization and user interaction.
- **Actionable Recommendations**: Offering Buy, Sell, or Hold suggestions based on forecasted stock trends and sentiment analysis.
- **Performance Evaluation**: Assessing model performance using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

Through this approach, the project aims to demonstrate the value of hybrid models that combine classical techniques with cutting-edge deep learning and real-time external data sources to tackle the challenges of stock market prediction.

2. LITERATURE SURVEY

Stock price prediction has been a significant area of research for decades, drawing attention from economists, financial analysts, and computer scientists alike. Early efforts primarily focused on traditional statistical models like the Autoregressive Integrated Moving Average (ARIMA) model. Introduced by Box and Jenkins in 1976, ARIMA gained popularity for its ability to model and forecast time series data by capturing autocorrelations. Despite its strengths in handling stationary data and providing a clear mathematical framework, ARIMA struggles with the non-linear and volatile nature of stock prices. Similarly, Linear Regression models were widely adopted in the early stages for stock forecasting. Although easy to implement and interpret, Linear Regression assumes a linear relationship between variables, which rarely holds true in complex financial markets. These limitations paved the way for more advanced methods capable of capturing non-linear patterns and dynamic behaviors inherent in stock price movements.

As financial datasets grew in size and complexity, machine learning techniques became increasingly popular. Support Vector Machines (SVMs) were among the first machine learning models applied to stock market prediction, offering powerful classification and regression capabilities. However, SVMs required extensive feature engineering and often struggled with the sequential nature of time series data. Random Forests and other ensemble methods provided improvements by reducing variance through the aggregation of multiple decision trees. While they achieved higher accuracy compared to individual models, they lacked the temporal modeling capabilities crucial for stock prediction tasks. Artificial Neural Networks (ANNs) also made an early impact, demonstrating an ability to model non-linear relationships. However, traditional feedforward networks failed to capture the sequential dependencies critical to time series forecasting, motivating researchers to explore more specialized deep learning architectures.

The advent of Recurrent Neural Networks (RNNs) marked a turning point in stock prediction research. RNNs are designed to handle sequential data by maintaining internal memory across time steps, making them naturally suited for time series forecasting. Nevertheless, vanilla RNNs faced significant training challenges, most notably the vanishing gradient problem, which limited their effectiveness for modeling long-term dependencies. Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber in 1997, overcame these limitations with a more sophisticated memory cell structure. LSTM networks excelled at learning long-term patterns in noisy and volatile financial data, leading to major breakthroughs in predictive accuracy. Subsequent studies, such as those by Fischer and Krauss (2018), empirically demonstrated that LSTM models significantly outperform traditional statistical and machine learning methods in stock price forecasting. Gated Recurrent Units (GRUs), a streamlined variant of LSTM proposed by Cho et al., further simplified the recurrent architecture while maintaining comparable performance, offering faster training times and requiring fewer computational resources.

Building upon the strengths of LSTM and GRU models, researchers have proposed hybrid architectures that combine the two, aiming to balance memory retention with computational efficiency. These hybrid models have shown promising results, especially in datasets characterized by high noise levels and irregular patterns — common traits of stock market data. While historical data provides essential information about market trends, it alone cannot account for external factors like breaking news, political events, or economic changes that can significantly influence stock prices. This realization led to the incorporation of sentiment analysis into stock forecasting models.

Financial news sentiment has proven to be a powerful predictor of market movements. Studies have found that news articles, press releases, and even social media posts can sway investor emotions and impact stock prices in the short term. Sentiment analysis techniques enable the extraction of meaningful polarity scores (positive, negative, neutral) from textual data, providing models with real-time insights into market psychology. Tools like TextBlob, VADER, and FinBERT have been widely adopted for this purpose. Among these, TextBlob stands out for its simplicity and effectiveness in generating polarity scores suitable for integration into real-time prediction systems. Incorporating sentiment analysis into machine learning models enriches the feature set, allowing the models to react not only to historical trends but also to current events and investor sentiment.

Several studies, such as those by Akita et al. (2016) and Chen et al. (2017), have successfully combined deep learning models with sentiment analysis to enhance forecasting accuracy. Their findings suggest that models which integrate textual sentiment data outperform those that rely solely on numerical historical data. However, many existing stock prediction systems still fall short in fully leveraging these advancements. Most available platforms, like Yahoo Finance or Bloomberg terminals, primarily focus on technical analysis or provide generalized forecasts without incorporating real-time sentiment dynamics. Additionally, many current models operate as black boxes, offering little to no explanation about the rationale behind their predictions, which limits user trust and interpretability.

Thus, there is a clear gap and opportunity to develop intelligent hybrid systems that seamlessly integrate historical price data with real-time sentiment analysis, powered by advanced deep learning models like LSTM, GRU, and their hybrid variants. By addressing the limitations of traditional models and embracing the dynamic nature of modern financial markets, this project proposes a comprehensive hybrid stock prediction system. The system aims to provide users with more accurate, explainable, and timely stock price forecasts through a web-based interface, combining the strengths of historical technical analysis and real-time market sentiment in a single, powerful platform.

3. PROPOSED SYSTEM

The proposed system is a **hybrid deep learning-based stock price prediction model** that intelligently combines historical stock market data with real-time financial news sentiment analysis to deliver highly accurate and dynamic forecasts. Traditional models often focus solely on technical indicators, ignoring the strong influence of public sentiment and external market-moving news. To bridge this gap, the proposed system integrates a hybrid LSTM-GRU deep learning architecture for modeling sequential price data, along with a sentiment analysis engine that quantifies market sentiment from real-time news articles. This dual-input approach enables the system to capture both long-term historical trends and immediate emotional factors influencing the market.

The architecture is designed to process two types of inputs simultaneously. The first input is historical stock price data, including features such as opening price, closing price, highest and lowest prices, and trading volume. This data is preprocessed, normalized, and fed into a hybrid LSTM-GRU model, leveraging LSTM's ability to capture long-term dependencies and GRU's efficiency in training to learn complex sequential patterns. The second input comes from real-time news headlines and articles related to specific companies or the stock market in general. This text data undergoes preprocessing steps including tokenization, stop-word removal, and sentiment scoring using tools like TextBlob or VADER. The resulting sentiment polarity scores are then integrated into the feature set provided to the hybrid model, allowing the system to adjust predictions based on current market mood and news events.

Once trained, the model outputs the predicted future stock prices along with confidence scores, helping users assess the reliability of each forecast. The entire system is deployed as a user-friendly web application developed using Flask. The front end allows users to select a stock, view historical data trends, sentiment trends, and receive real-time price predictions in a visually engaging format. The application also includes interactive dashboards and visualizations powered by libraries like Plotly and Chart.js, offering users deeper insights into prediction patterns and sentiment dynamics.

To ensure robustness, the system is trained and validated using a variety of real-world datasets, including historical price data sourced from Yahoo Finance APIs and financial news headlines from publicly available news aggregators. Techniques like cross-validation, early stopping, and dropout layers are employed to prevent overfitting and improve model generalization. Furthermore, the model's performance is evaluated using standard metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score to provide a comprehensive understanding of its predictive capabilities.

The proposed system offers an innovative solution to the challenges of stock price forecasting by combining the best of both worlds — deep sequential modeling of price data and real-time sentiment analysis. It addresses the limitations of traditional models by enhancing responsiveness to breaking news and investor emotions, ultimately providing a more intelligent, dynamic, and reliable forecasting tool for investors, traders, and researchers.

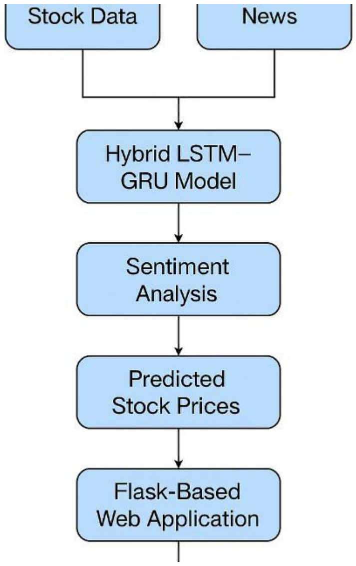


Fig 1: System Architecture

Working Procedure

1. Data Collection

- Historical Stock Data:
  - Retrieved from Yahoo Finance using APIs like yfinance.
  - Data features include: Open, High, Low, Close, Adjusted Close, and Volume.
- Real-Time Financial News:
  - News headlines are scraped or fetched via APIs from Yahoo Finance News, Google Finance, or similar financial news sources.
  - These headlines are timestamped to match with stock data dates.

2. Preprocessing

- Stock Data Preprocessing:
  - Handle missing values (interpolation or removal).
  - Feature scaling/normalization (MinMaxScaler or StandardScaler).
- News Data Preprocessing:
  - Cleaning text (removing punctuation, numbers, stopwords).
  - Applying TextBlob for sentiment analysis:
    - Sentiment Polarity is extracted and categorized into Positive, Negative, or Neutral.
  - Sentiment scores are merged with corresponding stock dates.

3. Feature Engineering

- Create new input features such as:
  - Moving Averages (5-day, 10-day, 20-day).
  - RSI (Relative Strength Index).
  - Sentiment Score Aggregates per day.
- Merge historical prices and sentiment scores into a single input dataset.

4. Model Building

- Hybrid Deep Learning Model:
  - LSTM Layer:
    - Captures long-term dependencies in stock sequences.
  - GRU Layer:
    - Captures short-term dependencies while reducing computational complexity.
  - Dense Layers:
    - Final prediction of stock closing price.
- Other Models (for comparison):
  - ARIMA Model: Classical time series model for linear patterns.
  - Linear Regression: Baseline model for trend analysis.

5. Training Phase

- Data is split into Training (80%) and Testing (20%) sets.
- Training using:
  - Mean Squared Error (MSE) as loss function.
  - Adam Optimizer for faster convergence.
- EarlyStopping and ModelCheckpoint callbacks are used to avoid overfitting.

6. Model Evaluation

- Use Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) as evaluation metrics.
- Compare results across:
  - Hybrid LSTM-GRU Model
  - ARIMA
  - Linear Regression
- Plot Actual vs Predicted Prices to visualize model performance.

7. Web Application Deployment

- Flask Framework is used to develop the web app.
- Key Features:
  - Input Form: Users can select a stock ticker and date range.
  - Backend: Triggers the prediction model and sentiment analysis.
  - Frontend: Displays:
    - Predicted stock prices.
    - Comparative graph (Actual vs Predicted).
    - Sentiment trend visualization.
    - Investment Recommendation (Buy/Sell/Hold) based on:
      - Predicted price trend.
      - Sentiment score.

8. Decision Support System

- Buy Recommendation:
  - If predicted future price > current price and sentiment is positive.
- Sell Recommendation:
  - If predicted future price < current price and sentiment is negative.
- Hold Recommendation:
  - If minimal change and/or mixed sentiment.

4. RESULT

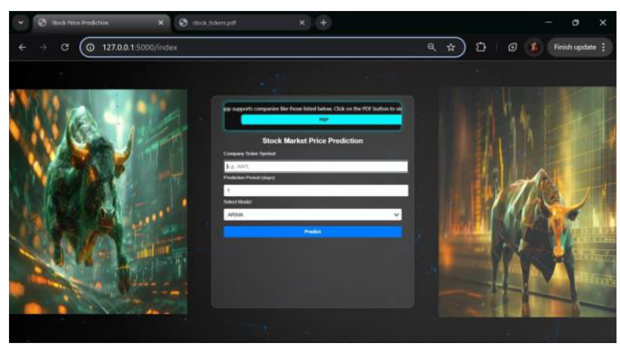
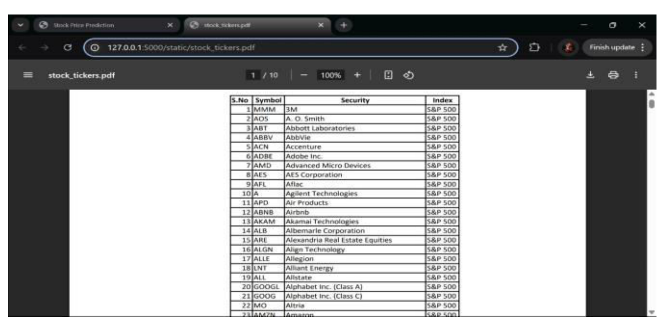


Fig 2: Home Page



S.No	Symbol	Security	Index
1	MMM	3M	S&P 500
2	AOS	A. O. Smith	S&P 500
3	ABT	Abbott Laboratories	S&P 500
4	ABBV	AbbVie	S&P 500
5	ACN	Accenture	S&P 500
6	ADBE	Adobe Inc.	S&P 500
7	AMD	Advanced Micro Devices	S&P 500
8	AES	AES Corporation	S&P 500
9	AFL	Aflac	S&P 500
10	A	Agilent Technologies	S&P 500
11	APD	Air Products	S&P 500
12	ABNB	Airbnb	S&P 500
13	AKAM	Akamai Technologies	S&P 500
14	ALB	Albemarle Corporation	S&P 500
15	ARE	Alexandria Real Estate Equities	S&P 500
16	ALGN	Align Technology	S&P 500
17	ALLE	Allegion	S&P 500
18	LNT	Alliant Energy	S&P 500
19	ALL	Allstate	S&P 500
20	GOOGL	Alphabet Inc. (Class A)	S&P 500
21	GOOG	Alphabet Inc. (Class C)	S&P 500
22	MO	Altria	S&P 500
23	AMZN	Amazon	S&P 500

Fig 3: Data Set for training

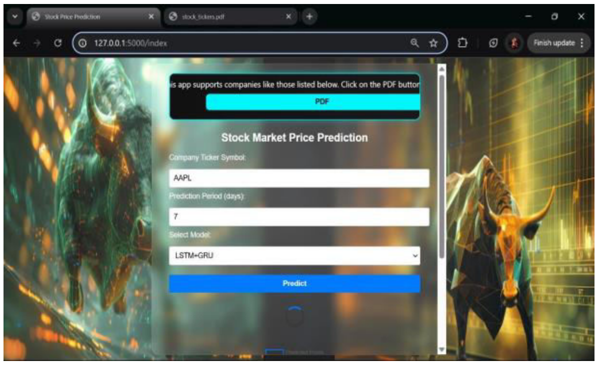


Fig 4: ARIMA Model Predicted Prices

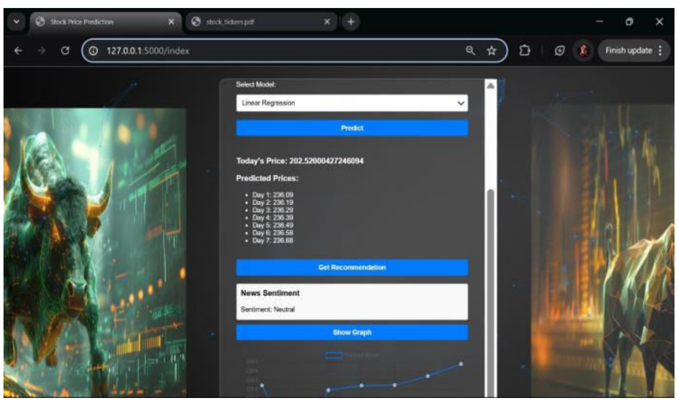


Fig 5 : LSTM+GRU Model

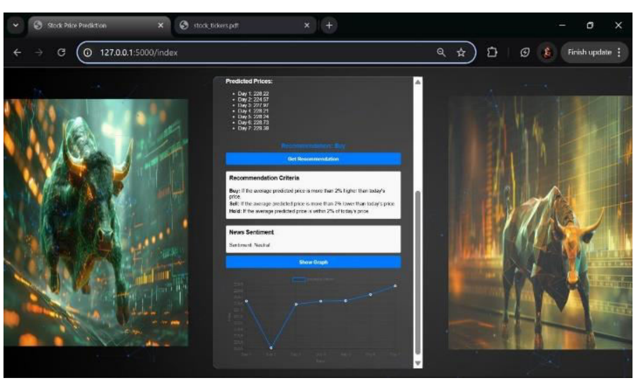


Fig 6 : LSTM+GRU Model Predicted Prices

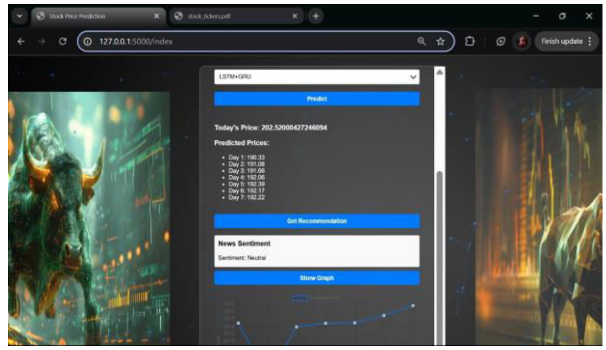


Fig 7 : Linear Regression Model Predicted Prices

5. CONCLUSION

In this project, we successfully developed a deep learning-driven hybrid system that combines historical stock price trends with real-time sentiment analysis to predict future stock prices more accurately. By integrating a hybrid LSTM-GRU neural network model with ARIMA and Linear Regression models, the system is able to capture both complex non-linear patterns and linear trends in financial data. Additionally, incorporating sentiment analysis from financial news headlines adds a crucial dimension to the prediction process, allowing the system to consider market psychology and investor emotions alongside traditional numerical indicators.

The experimental results demonstrate that the hybrid LSTM-GRU model outperforms traditional models like ARIMA and Linear Regression in terms of prediction accuracy, as measured by metrics such as RMSE and MAE. The inclusion of real-time sentiment scores significantly improves forecasting precision, especially during periods of market volatility influenced by external news events.

Moreover, the deployment of a user-friendly Flask-based web application ensures that the system is accessible, interactive, and practical for real-world use. Investors can not only view predicted stock prices but also receive simple investment suggestions (buy/sell/hold) based on predicted price movement and sentiment trends.

Thus, the project proves the effectiveness of combining deep learning, time-series forecasting, and sentiment analysis for stock market prediction. It offers a smart, dynamic decision-support system that could assist investors in making more informed investment decisions.

6. REFERENCES

[1] Alam Khorshed et al, "Enhancing Stock Market Prediction: A Robust LSTM-DNN Model Analysis on 26 Real-Life Datasets," doi: 10.1109/ACCESS.2024.3434524, IEEE Access, 2024.

[2] Amiri Babak et al, "A Novel Hybrid GCN-LSTM Algorithm for Energy Stock Price Prediction: Leveraging Temporal Dynamics and Inter-Stock Relationships," doi: 10.1109/ACCESS.2025.3536889, IEEE Access, 2025.

[3] Chen Lin et al, "Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market," doi: 10.1109/ACCESS.2018.2859809, IEEE Access, 2018.

[4] Chen Shile et al, "Stock Prediction Based on Genetic Algorithm Feature Selection and Long Short-Term Memory Neural Network,"doi:10.1109/ACCESS.2020.3047 109, IEEE Access, 2020.

[5] Chen Yingxuan et al, "A Dual-Attention-Based Stock Price Trend Prediction Model with Dual Features," doi: 10.1109/ACCESS.2019.2946223, IEEE Access, 2019.

[6] Choi Jooweon et al, "Hybrid Information Mixing Module for Stock Movement Prediction," doi: 10.1109/ACCESS.2023.3258695, IEEE Access, 2023.

[7] Ferreira Fernando G D C et al, "Artificial Intelligence Applied to Stock Market Trading A Review," doi: 10.1109/ACCESS.2021.3058133, IEEE Access, 2021.

[8] Haryono Agus Tri et al, "Transformer-Gated Recurrent Unit Method for Predicting Stock Price Based on News Sentiments and Technical Indicators," doi: 10.1109/ACCESS.2023.3298445, IEEE Access, 2023.

[9] Hoque Kazi Ekramul et al, "Impact of Hyperparameter Tuning on Machine Learning Models in Stock Price Forecasting," doi: 10.1109/ACCESS.2021.3134138, IEEE Access, 2021.

[10] Hossain Mohammad Raquibul et al, "Improving Stock Price Prediction Using Combining Forecasts Methods," doi: 10.1109/ACCESS.2021.3114809, IEEE Access, 2021.