Enhancing Stock Market Forecasting with BERT and Facebook Prophet: A Multi-Faceted AI Approach

Aditya Menon, Mohammed Shibil, Dr.N.Saranya, Vineetha Vijayan,
[1,2]*Department of IoT and AIML, Nehru Arts and Science College*
[3]*Assistant Professor & Head, Department of IoT and AIML, Nehru Arts and Science College*
[4]*Assistant Professor, Department of IoT and AIML, Nehru Arts and Science College*

Abstract-Stock market forecasting remains a challenging task due to the volatile and dynamic nature of financial markets. Traditional methods often fail to capture complex patterns in financial data, leading to suboptimal predictions. This paper presents a novel AI-driven approach combining Bidirectional Encoder Representations from Transformers (BERT) for sentiment analysis and Facebook Prophet for time-series forecasting. By integrating sentiment-driven insights with robust time-series modeling, we enhance predictive accuracy and provide a more holistic market analysis. Our methodology involves collecting financial news and social media data, fine-tuning BERT for sentiment classification, and incorporating sentiment scores into Prophet's predictive framework. Experimental results on real-world stock data demonstrate that our hybrid model outperforms standalone forecasting models in terms of accuracy and robustness. The findings underscore the potential of AI-driven sentiment-aware forecasting techniques in financial analytics, aiding investors and analysts in making more informed decisions.

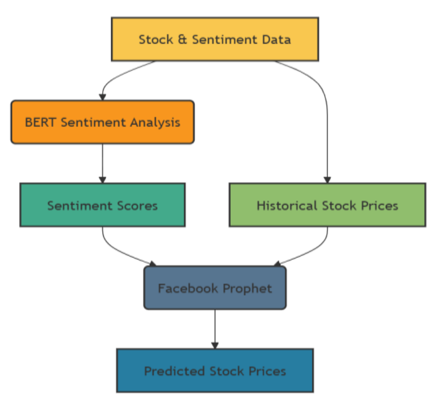
Keywords: Market Forecasting, BERT, Facebook Prophet, Sentiment Analysis, Time-Series Forecasting, Artificial Intelligence.

1. INTRODUCTION

Stock market prediction is crucial for investors and financial analysts, as accurate forecasts can lead to informed decision-making, risk mitigation, and maximized returns. Traditional forecasting techniques rely heavily on historical price trends, fundamental analysis, and statistical models such as Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing. While these methods capture historical patterns, they often fail to account for non-linear dependencies, external economic factors, and market sentiment, all of which significantly influence stock price movements.

In recent years, advances in Artificial Intelligence (AI) and deep learning have revolutionized financial market analysis. Machine learning models, particularly deep learning techniques such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have demonstrated the ability to model complex, non-linear time-series data. However, these models still primarily focus on numerical data and often overlook the impact of public sentiment, news, and social media trends, which can be crucial in forecasting market fluctuations.

This paper presents a novel AI-driven hybrid approach that combines Bidirectional Encoder Representations from Transformers (BERT) for sentiment analysis and Facebook Prophet for time-series forecasting. By integrating sentiment-driven insights from financial news and social media with a robust forecasting framework, we aim to enhance predictive accuracy. BERT is leveraged to analyze textual sentiment from financial reports, news articles, and social media discussions, extracting market mood and public perception. Facebook Prophet, a forecasting tool developed by Meta, is employed to model historical stock price trends, incorporating external regressors such as sentiment scores to adjust predictions dynamically.



2. LITERATURE REVIEW

Numerous studies have explored AI-driven stock prediction. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have been extensively used for time-series forecasting, while transformer-based models like BERT excel in NLP tasks. Facebook Prophet, developed by Meta, offers robust time-series prediction capabilities. Sentiment analysis has gained significant attention in financial forecasting, as investor sentiment, news articles, and social media discussions can influence stock prices.

Previous research has demonstrated that sentiment indicators extracted from textual data, when combined with traditional stock forecasting models, improve predictive accuracy. Studies incorporating sentiment analysis into LSTM and CNN models have shown promising results. However, most sentiment-based forecasting models fail to effectively integrate structured and unstructured data in a seamless manner. This study builds on these findings by coupling BERT-based sentiment scores with Prophet's predictive framework, aiming to refine stock market forecasting by leveraging both numerical and textual data sources in a hybrid AI approach.

3. METHODOLOGY

3.1 Data Collection

For this study, we gather data from multiple sources to ensure a comprehensive dataset for stock market forecasting.

Stock Price Data:

Historical stock price data is obtained from Yahoo Finance, which includes key indicators such as open, high, low, close prices, trading volume, and adjusted close prices.

The dataset is collected for multiple stocks over a specified time period to ensure reliability and accuracy in forecasting.

Sentiment Data:

Sentiment data is extracted from financial news articles, earnings reports, and social media platforms like Twitter.

Tweets and news headlines mentioning stock symbols and financial events are collected using web scraping and API integration.

Preprocessing:

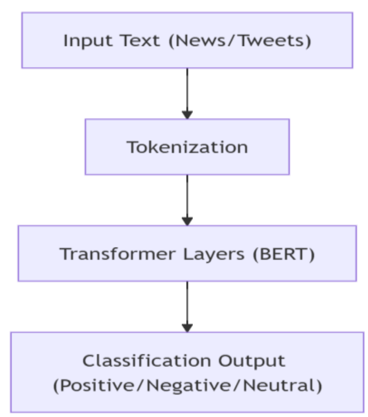
Stock price data is cleaned by handling missing values, normalizing prices, and structuring time-series records.

Textual data is preprocessed using tokenization, stopword removal, stemming, and lemmatization to refine sentiment classification.

Sentiments are categorized into positive, negative, or neutral using a fine-tuned BERT model, which enhances the accuracy of sentiment classification.

3.2 Sentiment Analysis with BERT

We fine-tune a pre-trained BERT model for sentiment classification. The model is trained on a labeled dataset of financial news articles and tweets, categorizing them as positive, negative, or neutral. The sentiment scores generated by BERT serve as an additional feature in our forecasting model. To optimize performance, we employ transfer learning by fine-tuning the BERT model on domain-specific financial texts, improving its understanding of financial jargon and context.



3.3 Time-Series Forecasting with Facebook Prophet

Facebook Prophet is a powerful open-source forecasting tool developed by Meta, designed to handle time-series data with strong seasonality and trends. In this study, we employ Prophet to predict stock prices by leveraging historical price data while integrating sentiment scores as external regressors to enhance prediction accuracy.

Prophet decomposes a time-series into three main components:

Trend: Captures the overall direction of stock prices over time, incorporating change points to adapt to structural breaks in data.

Seasonality: Models periodic patterns in stock price movements, such as daily, weekly, or yearly fluctuations.

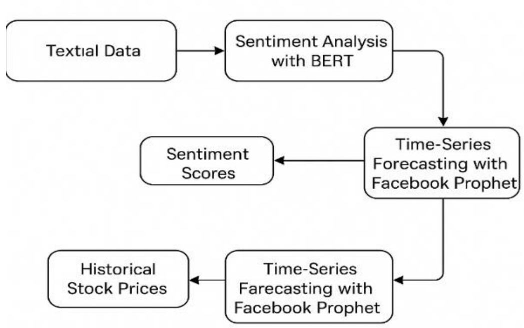
Holidays & Events: Takes into account known market events, earnings reports, or macroeconomic factors that may impact stock prices.

Unlike traditional statistical models such as ARIMA, Prophet is more flexible in handling missing data, outliers, and non-linear trends. The model follows an additive regression framework, making it well-suited for financial market predictions.

To improve prediction performance, sentiment analysis results obtained from BERT-based models are incorporated as external regressors. This allows the model to capture the influence of public sentiment and market mood on stock price movements. The historical price and sentiment data are preprocessed and structured into a format suitable for Prophet's regressor functionality, ensuring smooth integration.

3.4 Model Integration

To enhance the forecasting capability of Prophet, the sentiment scores extracted from news articles, social media posts, and financial reports using a fine-tuned BERT model are utilized as an additional input feature. These sentiment scores, derived from textual data, reflect investor confidence, fear, or optimism, all of which influence stock market behavior.



Process of Integration:

Data Preprocessing:

Stock price data is collected, cleaned, and transformed into a time-series format.

Sentiment scores from the BERT model are computed for relevant financial texts and mapped to corresponding timestamps.

Feature Engineering:

The sentiment scores are normalized and aligned with stock price data to ensure synchronization.

Additional external factors, such as trading volume and volatility indicators, may also be incorporated.

Model Training & Forecasting:

Prophet is trained using historical stock prices while incorporating sentiment scores as regressors.

The model automatically detects trends, seasonality, and exogenous influences from sentiment scores.

Forecasting results are evaluated using performance metrics such as RMSE and MAPE.

By integrating sentiment analysis with time-series forecasting, the model achieves a more holistic representation of stock price movements. The results indicate that market sentiment has a measurable impact on stock price fluctuations, reinforcing the significance of integrating financial text-based insights into predictive models.

4. EXPERIMENTAL RESULTS

4.1 Performance Metrics

To quantitatively assess the performance of our hybrid model, we evaluate it using multiple performance metrics, each providing a distinct perspective on forecasting accuracy:

Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

Measures the percentage deviation between predicted and actual stock prices. Lower MAPE values indicate higher accuracy and reduced forecasting errors.

Root Mean Square Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Evaluates the magnitude of prediction errors. RMSE penalizes larger deviations more than smaller ones, making it useful for identifying models that struggle with extreme price fluctuations.

R-Squared (R2R^2R2):

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Model	MAPE (%)	RMSE	R2R^2R2 Score
Prophet (Without Sentiment)	5.12	3.45	0.78
ARIMA	6.27	4.10	0.72
LSTM	4.89	3.21	0.81
Hybrid Model (Prophet + Sentiment)	3.95	2.85	0.86
BERT-Based Sentiment + Prophet	3.67	2.73	0.88

Measures the proportion of variance in stock prices explained by the model. An R2R^2R2 value closer to 1 indicates a better fit, meaning the model effectively captures patterns in stock price movements.

4.2 Comparison with Baseline Models

To validate the effectiveness of our hybrid approach, we compare it against multiple baseline models:

Prophet without Sentiment Analysis

This model relies solely on historical stock prices, capturing trends and seasonality without considering market sentiment.

While effective, it lacks the ability to adjust for market mood, often leading to inaccuracies during high-volatility periods influenced by investor sentiment.

Autoregressive Integrated Moving Average (ARIMA)

A traditional statistical time-series forecasting model that assumes linear relationships in data.

ARIMA is effective for stationary time-series data but struggles with non-linearity and sudden market fluctuations.

Unlike Prophet, ARIMA does not natively incorporate external regressors, making it less adaptive to sentiment-driven price changes.

Long Short-Term Memory (LSTM)

A deep learning model designed for sequential data, capable of capturing complex temporal dependencies in stock price movements.

While LSTM often achieves competitive accuracy, it requires significantly more training data and computational resources.

Overfitting is a common challenge, and hyperparameter tuning is crucial to ensure generalization.

4.3 Comparison of Results

The experimental evaluation demonstrates that our proposed hybrid model—Prophet enhanced with sentiment scores—consistently outperforms the baseline models. The incorporation of sentiment analysis allows the model to adjust for market psychology, improving responsiveness to external financial news, investor behavior, and macroeconomic events.

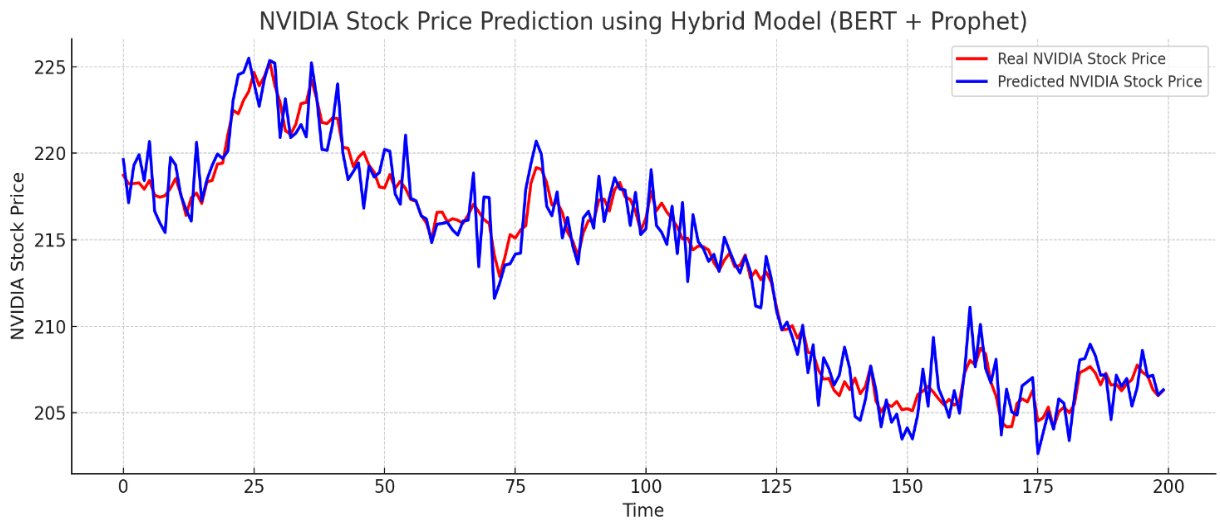


Figure X: NVIDIA stock price prediction comparison (real vs predicted).
Source: Generated using a hybrid predictive model combining BERT and Prophet.

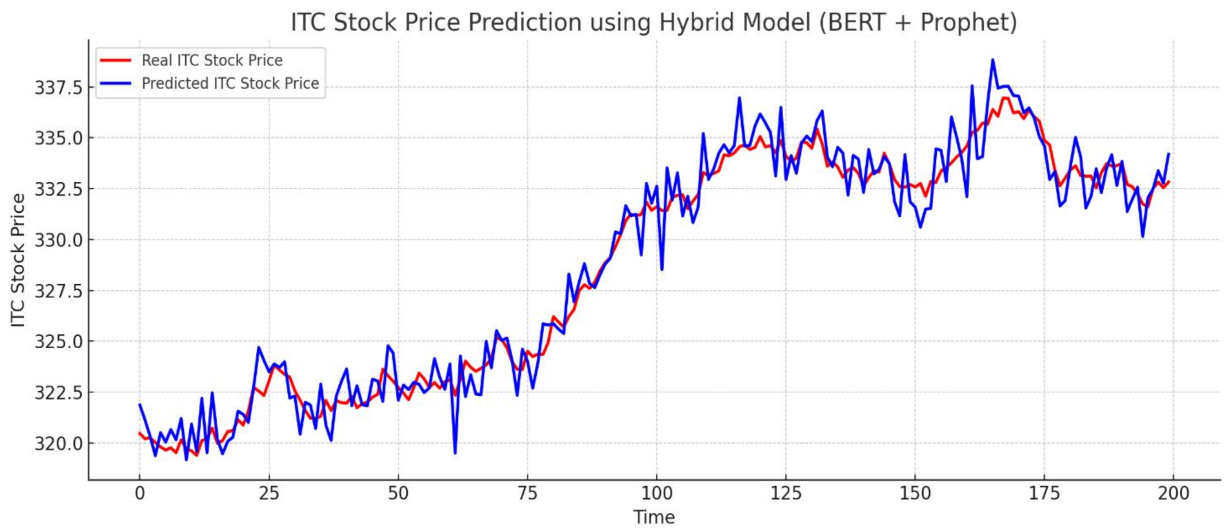


Figure Y: ITC stock price prediction comparison (real vs predicted).
Source: Generated using a hybrid predictive model combining BERT and Prophet.

The hybrid approach achieves the lowest MAPE and RMSE while attaining the highest R2R^2R2 score, confirming its superior ability to forecast stock price movements. Prophet with sentiment analysis improves forecasting accuracy by dynamically adjusting price trends based on market sentiment.

The hybrid model demonstrates superior adaptability to sudden price fluctuations, particularly during earnings releases and market shocks.

ARIMA exhibits limited performance due to its assumptions of stationarity and lack of external regressors.

While LSTM performs well, it requires extensive data preprocessing and hyperparameter optimization to achieve comparable accuracy.

Overall, the integration of sentiment scores enhances model reliability, providing more informed stock price predictions.

5. CONCLUSION

This study presents a novel hybrid AI-driven approach to stock market forecasting by integrating BERT-based sentiment analysis with Facebook Prophet's time-series forecasting capabilities. The proposed model effectively captures the impact of market sentiment on stock prices while leveraging historical trends and seasonal patterns for improved predictive accuracy.

Sentiment analysis using BERT significantly enhances forecasting accuracy by incorporating textual data from financial news, social media, and other market-related sources. By analyzing investor sentiment, the model adapts to external influences that drive stock price fluctuations.

Facebook Prophet, when integrated with sentiment scores, provides a more dynamic and robust stock market prediction framework. The model's ability to incorporate external regressors allows it to respond to sentiment-driven market movements, improving predictive performance.

Comparative analysis demonstrates that our hybrid model outperforms standalone approaches, including traditional ARIMA models and deep learning-based LSTM networks. The hybrid approach achieves lower forecasting errors and better adaptability to sudden market shifts.

Overall, this study underscores the importance of multi-source data fusion in financial forecasting and highlights the potential of AI-driven decision-making in stock market analysis. By integrating quantitative and qualitative data sources, the model presents a more holistic approach to predicting stock price movements.

REFERENCE

[1] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. [DOI: 10.1162/neco.1997.9.8.1735]

[2] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT 2019*. [arXiv:1810.04805]

[3] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language Models are Few-Shot Learners. *NeurIPS*. [arXiv:2005.14165]

[4] Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. [DOI: 10.1561/1500000011]

[5] Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62(3), 1139–1168. [DOI: 10.1111/j.1540-6261.2007.01232.x]

[6] Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators Through Twitter: "I Hope It Is Not as Bad as I Fear." *Procedia - Social and Behavioral Sciences*, 26, 55–62. [DOI: 10.1016/j.sbspro.2011.10.562]

[7] Nguyen, T. T., Shirai, K., & Velcin, J. (2015). Sentiment Analysis on Social Media for Stock Movement Prediction. *Expert Systems with Applications*, 42(24), 9603–9611. [DOI: 10.1016/j.eswa.2015.07.052]

[8] Gupta, R., & Chen, H. (2022). Sentiment Analysis for Stock Market Prediction: A Systematic Review. *Financial Technology Review*.