International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05864
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Advanced Sentiment-Based Market Trend Prediction using BERT

Tushar Choudhari..(Department of Data Science
Dr. D. Y. Patil Arts, Commerce and Science College Pimpri)

Purshottam Chaudhari..(Department of Data Science
Dr. D. Y. Patil Arts, Commerce and Science College Pimpri)

----------***----------

Abstract - This study presents an advanced machine learning and natural language processing (NLP) approach for predicting short-term stock market trends using financial news headlines and social media sentiment. Traditional market prediction models rely heavily on historical price data and fail to incorporate textual sentiment effectively. In this research, a fine-tuned BERT (Bidirectional Encoder Representations from Transformers) model is used to analyze financial text and generate sentiment scores. These scores are combined with stock price features such as daily returns and moving averages and fed into a Random Forest classifier to predict whether the market will move upward or downward.
To improve transparency and trust in the system, Explainable AI techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are integrated. SHAP provides global feature importance, while LIME highlights word-level contributions in individual predictions. The model demonstrates strong predictive capability and provides clear interpretability. The system is deployed as an interactive web application using Streamlit.

***Key Words*:** Stock Market Prediction, Sentiment Analysis, BERT, Explainable AI, Random Forest, SHAP, LIME, NLP, Data Science

1.INTRODUCTION

Stock market prediction is a challenging task due to its dynamic and non-linear nature. Traditional models primarily rely on historical stock prices and statistical methods, which fail to capture the influence of external factors such as financial news and public sentiment. With the growth of digital media, large volumes of financial text data are generated daily, providing valuable insights into market behavior.

Recent advancements in Natural Language Processing (NLP), particularly transformer-based models like BERT, enable machines to understand contextual meaning in text. This research focuses on leveraging BERT for sentiment analysis and combining it with machine learning techniques to improve prediction accuracy. Additionally, Explainable AI is incorporated to address the “black box” nature of machine learning models and enhance interpretability.

2.Data Collection and Preprocessing:

The dataset used in this study consists of financial news headlines, social media tweets, and corresponding stock market data. The textual data reflects market sentiment, while stock data includes price-related attributes such as Open, High, Low, and Close values.
Financial text data was collected from publicly available online sources such as financial news platforms and social media discussions related to stock markets. Stock price data was obtained from historical market records.

Model Architecture:

The ERT model was used for sentiment analysis of financial text data Pre-trained "bert-base-uncased" model was fine-tuned using HuggingFace Transformers and PyTorch.The model converts input text into contextual embeddings and predicts sentiment score
Random Forest Classifier was used for market trend prediction Input features include sentiment score and stock-related features such as returns and moving averages.The model consists of multiple decision trees to improve prediction accuracy

Model Evaluation :

The Dataset was split into training and testing sets in the ratio of 80:20 BERT model was fine-tuned for sentiment classification on financial text data Random Forest model

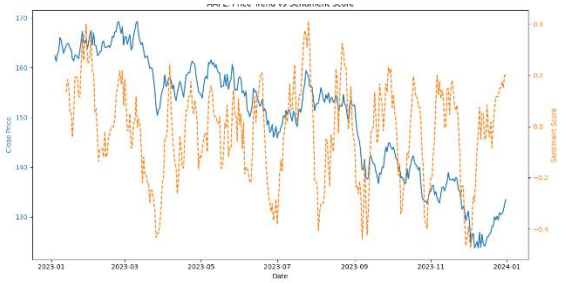
International Scientific Journal of Engineering and Management (ISJEM)
Volume: 05 Issue: 03 | March – 2026
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI:10.55041/ISJEM05864

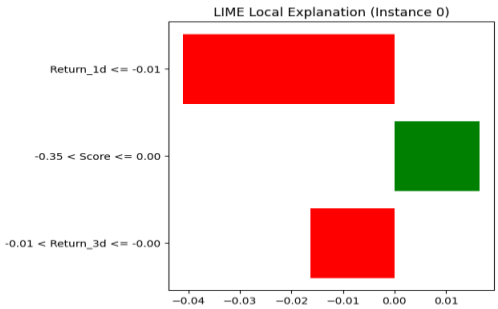
was trained using combined features of sentiment score and stock data Evaluation of the model was performed using standard classification metrics Accuracy was used to measure overall prediction correctness Precision was used to evaluate correctness of positive predictions Recall was used to measure the ability to identify actual positive trends F1-Score was used to balance precision and recall ROC-AUC was used to evaluate model performance across different thresholds The proposed model achieved high accuracy and demonstrated strong predictive performance Results indicate that combining sentiment analysis with stock features improves prediction capability

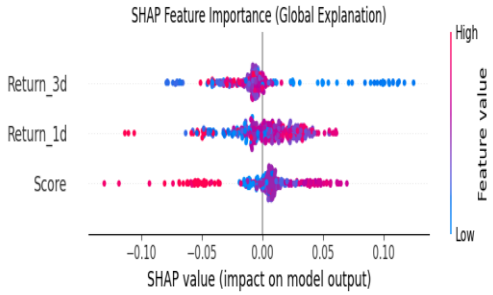
Metric	Random Forest	BERT
Accuracy	0.7934	0.8530
Precision	0.6360	0.8239
Recall	0.8187	0.7593
F1-Score	0.7714	0.8200
ROC-AUC	0.7276	0.9390

Table 1: Model Performance Comparison

Chart







Result and Analysis:

The proposed system achieved high prediction accuracy and demonstrated strong performance in forecasting stock market trends, with the combination of sentiment analysis using BERT and stock data significantly improving results compared to using stock data alone, the evaluation metrics indicate strong classification capability with high accuracy, precision, recall, and F1-score, while the ROC-AUC score shows that the model effectively distinguishes between upward and downward market movements, visualization techniques such as SHAP plots were used to identify important features influencing predictions and LIME explanations highlighted key words in financial text contributing to sentiment classification, the results also show that positive sentiment is generally associated with upward trends and negative sentiment correlates with downward trends, and the integration of Explainable AI improves transparency and helps users understand the decision-making process, overall the system proves to be effective, reliable, and suitable for real-world financial analysis applications

3.CONCLUSION:

This study demonstrates that the proposed BERT-based sentiment analysis combined with a Random Forest model is highly effective for stock market trend prediction, outperforming traditional approaches that rely only on historical stock data, the integration of textual sentiment significantly improves prediction accuracy, recall, F1-score, and ROC-AUC, enabling better identification of market movements, the use of Explainable AI techniques such as SHAP and LIME enhances transparency by providing both global and local interpretability of the model's decisions, making the system more reliable and trustworthy, while minor limitations exist due to dependency on data quality, the overall performance shows strong potential for real-world financial applications, future work can focus on improving model performance using real-time financial data, advanced deep learning models such as LSTM or transformers for time-

series analysis, and deploying the system on cloud platforms for scalability

ACKNOWLEDGEMENT

We would like to express our gratitude to our college, faculty members, and project guide for their continuous support and guidance throughout this research work

REFERENCES

1. Devlin, J. et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

2. Lundberg, S.M., & Lee, S.I. (2017). A Unified Approach to Interpreting Model Predictions (SHAP)

3. Ribeiro, M.T. et al. (2016). Why Should I Trust You? Explaining the Predictions of Any Classifier (LIME)

4. Breiman, L. (2001). Random Forests

5. Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python