Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2024)
IEEE Xplore Part Number: CFP24OSV-ART; ISBN: 979-8-3503-7642-5

2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) | 979-8-3503-7642-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/I-SMAC61858.2024.10714754

Analyzing Market Factors for Stock Price Prediction using Deep Learning Techniques

*Suresh K
Department of Computer Science
Christ University
Bangalore-560029, India
suresh.kalaimani@christuniversity.in

Aritra Sarkar
Department of Computer Science
Christ University
Bangalore-560029, India
aritra.sarkar@mca.christuniversity.in

Anjali Rai
Department of Computer Science
Christ University
Bangalore-560029, India
anjali.rai@mca.christuniversity.in

Apoorva Vasishtha
Department of Computer Science
Christ University
Bangalore-560029, India
apoorva.vasishtha@mca.christuniversity.in

Tyrell Fernandes
Department of Computer Science
Christ University
Bangalore-560029, India
tyrell.fernandes@mca.christuniversity.in

Cecil Donald A
Department of Computer Science
Christ University
Bangalore-560029, India
cecil.donald@christuniversity.in

***Abstract*—This paper presents a comprehensive study on stock price predictions by integrating market factors and sentiment analysis of news headlines. The research is divided into two modules, each employing distinct methodologies to enhance the accuracy of stock price forecasts. In the first module, market factors are investigated using three advanced algorithms: Long Short-Term Memory (LSTM), Gradient Boosting Decision Trees (GBDT), and Facebook Prophet (FBPROPHET). These algorithms are evaluated based on metric scores such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The analysis focuses on predicting high and low values of market prices for the period from January to June 2021. The comparative assessment of these algorithms provides insights into their effectiveness in capturing market trends and making precise predictions. In the second module, the paper explores the impact of news headlines on stock prices by extracting sentiment using three distinct algorithms: lexical-based analysis, Naive Bayes, and FinBERT. The sentiment analysis aims to gauge the market sentiment reflected in news articles and assess its influence on stock price movements. Prediction accuracy is calculated for each algorithm, highlighting their strengths in capturing sentiment patterns.**

Keywords—Stock price predictions, LSTM, GBDT, FBPROPHET, MSE, MAE, RMSE, Lexical- Based, Naive Bayes, FinBERT, Financial markets

I. INTRODUCTION (*HEADING 1*)

In finance and investment, the complexities of the stock market is paramount. Our "Market factors and stock price predictions: A comprehensive study" project into this domain by focusing on two critical elements: market sentiment and historical data. The market sentiment reflects investors' emotions and psychological outlook, influencing stock prices and market dynamics. Analyzing sentiment through sources like news, social media, and sentiment analysis tools offers insights into market confidence and risk appetite. Historical data, on the other hand, provides a wealth of information about past stock market performance. Using data analysis and machine learning, we aim to uncover patterns and correlations within this data, enabling us to predict future stock market values. Our project integrates market sentiment and historical data analysis to create a predictive model for stock market performance. This tool empowers investors and analysts with valuable insights for navigating the financial aspects. Join us as we explore the interplay of these factors and predict stock market values.

II. LITERATURE REVIEW

A. A Review of stock price prediction on machine learning techniques

Several existing systems and platforms operate within the same domain as our proposed system. These systems technology, data analysis, and financial expertise provide insights into stock market behavior. Here are a few examples:

Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: This paper proposes a ma- chine learning model that predicts future stock prices by combining historical stock price data with financial news. To comprehend the effects of economic news on stock price and each sentiment scoring algorithm, the study employed three different algorithms to calculate various sentiment scores and combine them in different ways.[4]

Stock price prediction with optimized deep LSTM net- work with artificial rabbit optimization algorithm: The paper presents an optimized deep LSTM network with the ARO model for stock price prediction using the DJIA index dataset. Using MSE, MAE, MAPE, and R2 evaluation criteria, three distinct LSTM models, one artificial neural network model, and an LSTM optimized by a Genetic Algorithm model were compared. LSTM- ARO outperforms the other models.[2]

Forecasting and optimization stock predictions: Varying asset profile, time window, and hyperparameter factors: Their goal is to review the literature to find thorough studies that tackle these issues and improve the state-of- the-art by incorporating new elements like training data ratios, multi-time windows, training batch sizes, stopping criteria, and financial, technical indicators[5]

Machine learning approaches in stock market prediction: This paper reviews 30 studies on machine learning ap- proaches for stock market prediction, including neural networks and support vector machines. The systematic literature review reveals that neural networks are the most commonly used model for stock market prediction, outperforming other models such as support vector ma- chines.[3]

Clustering-based return prediction model for stock pre-selection in portfolio optimization using PSO- CNN+MVF: This paper proposes a hybrid method for stock pre-selection that combines a convolutional neural network with optimized hyperparameters by particle swarm optimization and a mean-variance with forecasting model for portfolio optimization.[6]

These existing systems and platforms serve various segments of the financial industry, from individual investors and traders to financial professionals and institutions. Our proposed method can build upon the strengths of these systems while introducing innovative approaches and insights to contribute to the stock market analysis and prediction field.

B. Challenges in Current Stock Market Analysis Systems

While valuable, existing systems for stock market analysis and prediction come with several limitations our project could address. Here are some standard rules of these systems:

Many existing systems heav- ily rely on historical data for analysis. This reliance can be a limitation when market conditions change rapidly, as historical data may not adequately capture emerging trends or sudden market shocks. Some systems provide delayed data updates, which can be problematic for traders and investors who require real-time information to make timely decisions in a fast-paced market.

Premium financial platforms, like Bloomberg and Thomson Reuters Eikon, can be prohibitively expensive for individual investors and small firms, limiting access to valuable tools and data. Many existing systems are complex and re- quire a steep learning curve, making them less accessible to investors and more minor market participants. Some quantitative trading platforms use proprietary algorithms and models that may lack transparency, making how predictions are generated challenging.

Sentiment analysis can be influenced by noise and irrelevant information, leading to less accurate sentiment indicators. Machine learning-based prediction models can sometimes suffer from overfitting, performing well on historical data but failing to generalize to new market conditions. Stock market sentiment can be manipulated by coordinated efforts, such as social media campaigns or rumor spreading, which can distort senti- ment analysis results.

Using sentiment analysis from social media and news sources may raise privacy and ethical concerns, particularly when analyzing user- generated content without explicit consent.

Our proposed model can aim to address some of these limitations by offering a more transparent, real-time, customizable, and cost-effective solution that strengths existing systems while mitigating their shortcomings. Focusing on data quality, robustness, and adaptability to changing market conditions can also include stock market analysis and prediction.

III. PROPOSED METHODOLOGY

Three advanced algorithms, LSTM, GBDT, and FBPROPHET, are utilized to forecast stock prices based on market factors. These algorithms are trained on historical market data spanning from January to June 2021, and their predictive performance is evaluated using metrics such as MSE, MAE, and RMSE. The aim is to accurately predict high and low market values, providing robust market dynamics.

Long Short Term Memory (LSTM) is a Recurrent Neural Network (RNN) that can capture sequential dependencies in time series data and is well-suited for stock price prediction. Gradient Boosting Decision Trees (GBDT) is a type of ensemble learning method, which means it combines the predictions of multiple base models, in this case, decision trees, to make more accurate predictions FaceBook Prophet (FBPROPHET) is a powerful time series forecasting algorithm that follows the Bayesian approach and can capture complex patterns in the data, such as seasonality, trends, and the effect of holidays

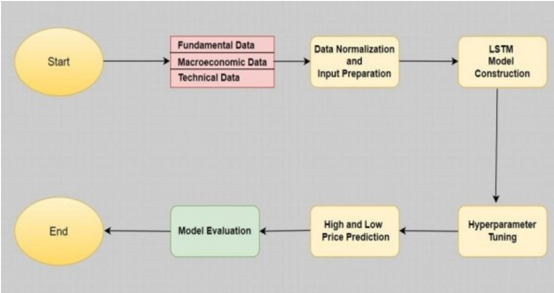


Fig.1. LSTM Model

Model Building involves creating and training a predictive algorithm using data. It includes selecting an algorithm, preparing the data, training the model, and evaluating its performance to make accurate predictions or provide insights. The Prediction Making model is trained and evaluated, it can predict future stock prices. These predictions can inform investment decisions and trading strategies.

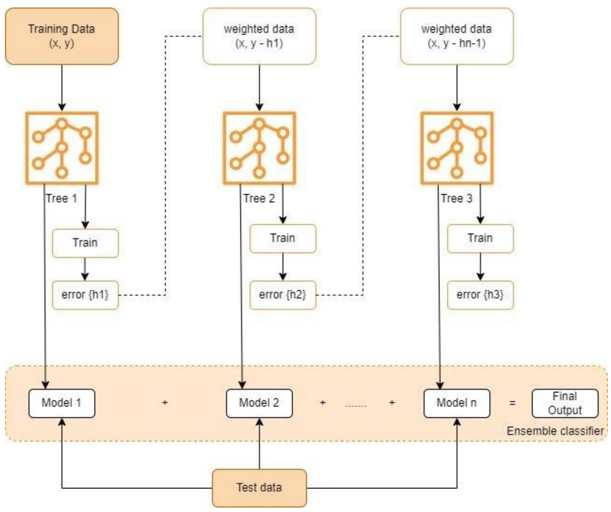


Fig.2. GBDT Model

The Proposed Model focuses on sentiment analysis of news headlines using three distinct algorithms: lexical- based analysis, Naive Bayes, and FinBERT. These models are trained to extract sentiment from news articles, and their prediction accuracy is assessed. The goal is to understand the impact of news sentiment on stock prices, providing additional insights into the factors influencing market movements.

Lexicon-based Approach to sentiment analysis, the process involves setting up a database, loading word lists, reading and preprocessing headlines, calculating sentiment scores for each word, and evaluating these scores to determine the overall sentiment conveyed in the textual data.

FinBERT Algorithm entails importing libraries and loading data, tokenization, model loading, inference, and postprocessing. The results are organized into a data frame, and additional steps involve determining the dominant sentiment and calculating accuracy for comprehensive sentiment analysis.

Naive Bayes Algorithm: In machine learning, they are a family of simple 'probabilistic classifiers' based on applying Bayes' theorem with strong (naive) in- dependence assumptions between the features. Bayes' Theorem is expressed as:

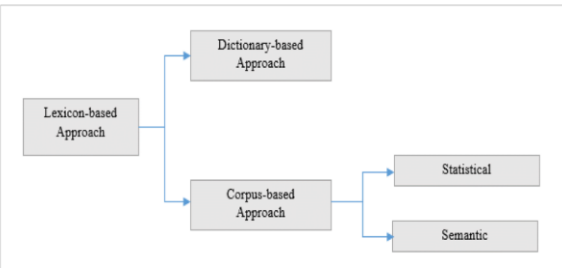
$$P(A B) = \frac{P(B|A) \times P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as:

$$Posterior = \frac{prior \times likelihood}{evidence}$$

A. Prediction using Market Sentiment Analysis

This process entails constructing and training a predictive algorithm using datasets. It encompasses choosing an appropriate algorithm, preparing the data, training the model, and assessing its performance to ensure accurate predictions or uncover valuable insights.



.Fig 3 Lexicon Model

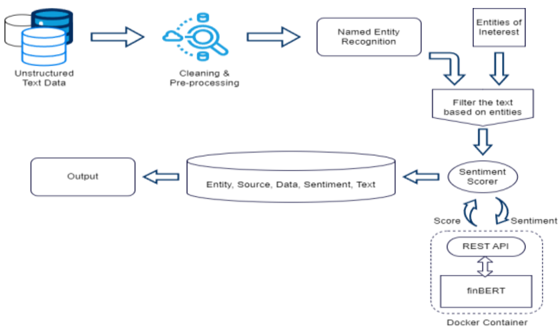
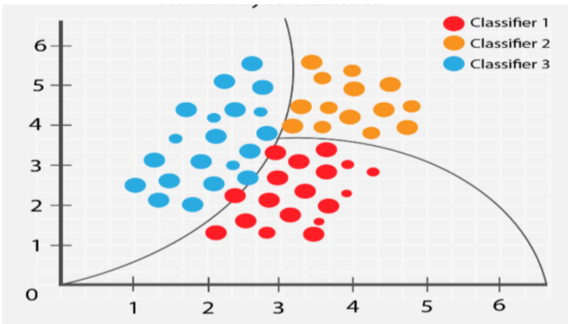


Fig.4. Finbert Model



.Fig 5 Naves Bayes Model

In figure 3, LSTM is designed to handle sequential data, making it particularly useful for language tasks, as it can learn dependencies over long sequences of text. When incorporating a lexicon into an LSTM model. The lexicon can provide additional information about the words in the sequence, such as sentiment or domain-specific tags. These features are then added to the LSTM's inputs.

In figure-4, FinBERT is a pre-trained language model for financial sentiment analysis. It is based on BERT (Bidirectional Encoder Representations from Transformers) but fine-tuned specifically for finance. Unlike general-purpose BERT models, FinBERT is trained on a vast corpus of financial documents, including financial reports, news articles, and earnings call transcripts. This makes it particularly effective for understanding and generating text in the financial domain.

In figure-5, a probabilistic machine learning algorithm based on Bayes' Theorem, with a strong assumption of independence between the features. Although it's primarily used for classification tasks like text classification and spam filtering, it can also be applied to stock prediction in certain cases, especially for classification-based stock movement prediction.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

In module one, time-series plots visualize the actual vs. predicted stock prices for LSTM, GBDT, and FBPROPHET. Comparative bar charts display algorithm performance met- rics (MSE, MAE, RMSE). For the historical data, we visualize the plots of the Actual Values and the Predicted Values of the High and Low His- torical Data from January 2021 to June 2021 in the range of 120-140.

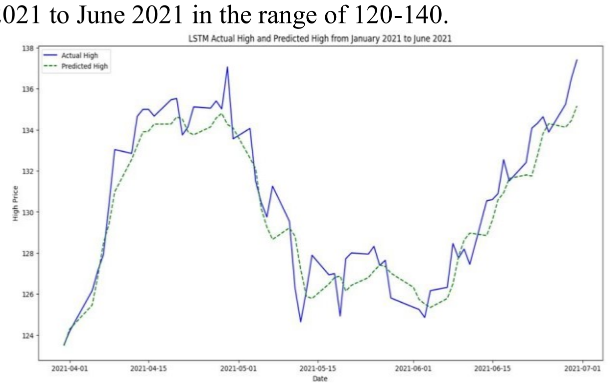


Fig.6. GBDT High

Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2024)
IEEE Xplore Part Number: CFP24OSV-ART; ISBN: 979-8-3503-7642-5

In figure-6, The model uses decision trees in a boosting framework, where trees are built sequentially, and each new tree corrects errors made by the previous one. Grid search is employed to optimize key parameters such as the number of trees, learning rate, and maximum depth. The model's performance is analyzed in volatile market conditions (e.g., during financial crises) to evaluate its robustness.

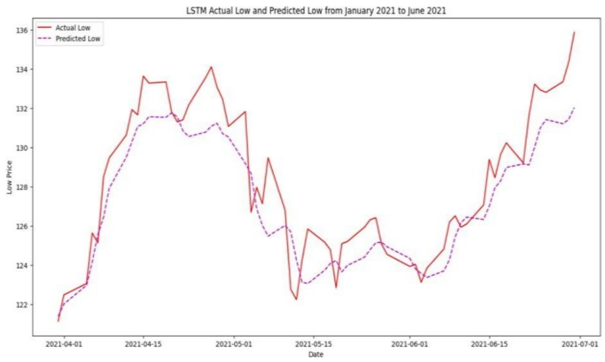


Fig.7. GBDT Low

In figure-7, GBDT can be computationally intensive, so for large datasets, consider using efficient implementations like XGBoost, LightGBM, or CatBoost. Using GBDT for low-stock prediction enables businesses to better manage their inventory, avoid stockouts, and ensure optimal stock levels based on data-driven forecasts.

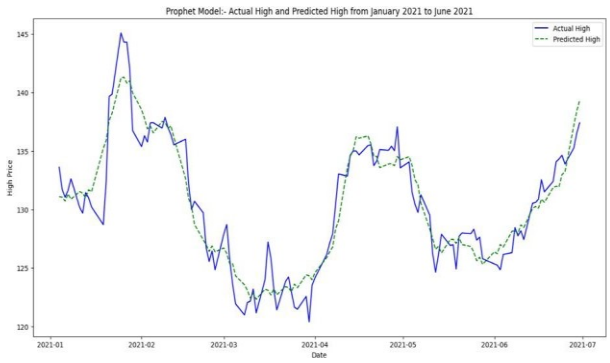


Fig.8. FBPROPHET High

FBProphet, now called Prophet (after its rebranding by Meta), is an open-source tool for forecasting time series data. Developed by Facebook (now Meta), it is particularly well-suited for time series data that display seasonal trends, outliers, or irregular intervals of observation. Its use in stock prediction has gained attention due to its simplicity and effectiveness in handling large, complex datasets.

Fig.9. FBPROPHET Low

For each metric, lower values indicate better performance, and the following results were obtained:-

TABLE I
LSTM ERROR METRICS

Metric	High	Low
Mean Squared Error (MSE)	2.0752	3.3931
Mean Absolute Error (MAE)	1.5148	1.9048
Root Mean Squared Error (RMSE)	1.9405	2.3420

TABLE II
GBDT ERROR METRICS

Metric	High	Low
Mean Squared Error (MSE)	0.6060	0.6491
Mean Absolute Error (MAE)	0.5876	0.6344
Root Mean Squared Error (RMSE)	0.7784	0.8056

TABLE III
FBPROPHET ERROR METRICS

Metric	High	Low
Mean Squared Error (MSE)	3.0311	3.4830
Mean Absolute Error (MAE)	1.3711	1.4007
Root Mean Squared Error (RMSE)	1.7410	1.8662

Finally, we compared all three models using their accuracy scores to determine which algorithm performed best and was the most reliable.

TABLE IV
MODEL PERFORMANCE METRICS

Models	MSE	MAE	RMSE
LSTM High	2.0752	1.5148	1.9405
LSTM Low	3.3931	1.9049	2.3420
GBDT High	0.6060	0.5876	0.7784
GBDT Low	0.6491	0.6344	0.8056
PROPHET High	3.0311	1.3711	1.7410
PROPHET Low	3.4830	1.4007	1.8662

V. CONCLUSION

In the first module, where market factors were analyzed using LSTM, GBDT, and FBPROPHET algorithms, the com- parison of actual and predicted values for high and low prices illustrated the efficacy of each model. The findings indicate that GBDT demonstrated the lowest error, followed by LSTM and FBPROPHET. This suggests that Gradient Boosting Decision Trees is particularly adept at capturing complex market patterns, making it a promising choice for stock price predictions. In the second module, focusing on sentiment analysis of news headlines, the three algorithms, Lexical-based analysis, Naive Bayes, and FinBERT, were evaluated for their accu- racy in predicting market sentiment. The results indicate that Naive Bayes outperformed the other algorithms, followed by FinBERT and lexical-based analysis. Overall, the combined findings from both modules underscore the importance of a multifaceted approach to stock price predictions. Integrating market factors and sentiment analysis enhances the robustness of predictive models, providing stakeholders with more of market behavior. As the financial evolves, further research and refinement of these methodologies will be crucial for adapting to changing market conditions and improving the accuracy of stock price forecasts.

REFERENCES

[1] JinShan Yang, ChenYue Zhao, HaoTong Yu, HeYang Chen, Use GBDT to Predict the Stock Market, Procedia Computer Science, Volume 174, 2020, Pages 161-171, ISSN 1877-0509,https://doi.org/10.1016/j.procs.2020.06.071.

[2] Burak Gu¨lmez, Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm, Expert Systems with Applications, Volume 227, 2023, 120346, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2023.120346.

[3] Latrisha N. Mintarya, Jeta N.M. Halim, Callista Angie, Said Achmad, Aditya Kurniawan, Machine learning approaches in stock market prediction: A systematic literature review, Procedia Computer Science, Volume 216, 2023, Pages 96- 102, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.12.115.

[4] Junaid Maqbool, Preeti Aggarwal, Ravreet Kaur, Ajay Mittal, Ish- faq Ali Ganaie, Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Ap- proach, Procedia Computer Science, Volume 218, 2023, Pages 1067- 1078, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2023.01.086.

[5] Anusha, M., K. Suresh, and M. Chandana. "Earlier Prediction on the heart disease based on supervised machine learning techniques." 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021.

[6] Suresh, K., and V. Pattabiraman. "Developing a customer model for targeted marketing using association graph mining." International Journal of Recent Technology and Engineering 8.2 (2019): 292-296.

[7] Chaher Alzaman, Forecasting and optimization stock predictions: Varying asset profile, time window, and hyperparameter factors, Systems and Soft Computing, Volume 5,2023,200052, ISSN 2772-9419, https://doi.org/10.1016/j.sasc.2023.200052.

[8] YL, Manoj Kumar, and K. Suresh. "Stock Split Analysis and Market Value Predictions by using Enhanced Long Short-Term Memory." 2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT). IEEE, 2022.

[9] Suresh, K., and O. Praveen. "Extracting of patterns using mining methods over damped window." 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2020.

[10] Mahdi Ashrafzadeh, Hasan Mehtari Taheri, Mahmoud Gharehgozlou, Sarfaraz Hashemkhani Zolfani, Clustering-based return prediction model for stock pre-selection in portfolio optimization using PSO-CNN+MVF, Journal of King Saud University - Computer and Information Sciences, Volume 35, Issue 9, 2023,101737, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2023.101737.