e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

STOCK PREDICTION USING SENTIMENT ANALYSIS ON TWITTER USING LSTM

T. Padmaja*1, Gaddipati Rakesh*2, Venishetty Sriram*3, V. Naveen Kumar*4

*1Assistant Professor Department Of Computer Science And Engineering Malla Reddy College Of Engineering & Technology Hyderabad, India.

*2,3,4Final Year Student Department Of Computer Science And Engineering Malla Reddy College Of Engineering & Technology Hyderabad, India.

DOI : https://www.doi.org/10.56726/IRJMETS50000

ABSTRACT

It is challenging to predict stock fluctuations in the financial markets due to the wide range of unanticipated factors that they contain. This research explores how sentiment analysis and Long Short-Term Memory (LSTM) networks can be combined with Twitter data to enhance stock market predictions. Our method scrapes historical stock data from Yahoo Finance and real-time sentiment data from Twitter using a variety of tools, including Pandas, NumPy, TensorFlow, and Keras. To ensure the accuracy of future predictions, data must be carefully cleaned and transformed during preparation. Features are chosen and the model is trained with the help of LSTM, a deep learning architecture that is well-known for its proficiency with sequence prediction issues. The project also emphasizes how important it is to evaluate model performance, learn from previous studies, and use automated techniques, like Auto Arima, for baseline comparisons. The findings show that Twitter emotions have a positive effect on forecast accuracy, which indicates the potential of proactive decision-making in the stock market. As we work through the challenges of this project, we visualize a time when sentiment-driven insights will greatly improve financial forecasting, resulting in better market outcomes and decision-making.

I. INTRODUCTION

The complex nature and volatility of financial markets have long fascinated investors and researchers alike, which resulted in an endless search for trustworthy forecasting models. To tackle the challenging task of stock market forecasting, our research combines sentiment analysis on Twitter data with advanced deep learning techniques. Including sentiment from Twitter aims to capture the general sentiment of the market in a time of information flow and social media's widespread influence—a factor that often goes unnoticed in traditional financial analyses.

Knowing that stock markets are fundamentally unpredictable and that a variety of outside factors, including political developments and opinions shared on social media, can significantly affect stock prices served as a boost for this study. Recognizing the limitations of conventional models, we investigate the application of powerful recurrent neural networks (RNNs) called Long Short-Term Memory (LSTM) networks to capture intricate patterns in time-series data. Because of its ability to retain contextual information for extended periods of time, LSTM is a viable choice for modelling movements in the stock market.

We carry out our research in multiple stages, beginning with data collection and concluding with model implementation. We use Snscrape to extract sentiment from Twitter in real time and Yahoo Finance to get historical stock data over an extended period of time. Preprocessing of both datasets has been done in-depth to ensure the accuracy of any subsequent studies. Feature selection becomes a crucial step in increasing the prediction power of our models. Relevant variables can be found here.

We still use many different tools and libraries, including Pandas, NumPy, Transformers, TensorFlow, and Keras, because our process is still very much reliant on technology. Together, these technologies enable seamless sentiment analysis, deep learning architecture deployment, and data processing.

This introduction establishes the framework for a detailed analysis of our project and emphasizes the need for innovative solutions in a financial environment that is continuously shifting. Our goal is to significantly advance

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

the quickly evolving field of stock market prediction by offering insights that go beyond traditional methods and consider the dynamic structure of modern financial ecosystems.

II. LITERATURE REVIEW

2.1 Stock Market Prediction Models: A Historical Overview

The pursuit of predicting stock market movements has been ongoing, evolving with diverse models and methodologies. Traditional techniques like time series analysis and statistical models, exemplified by Autoregressive Integrated Moving Average (ARIMA) models, have long been foundational. Mandelbrot's work (1963) on fractal patterns in financial markets introduced complexity to market movements.

Machine learning's rise marked a paradigm shift. Bao et al. (2017) explored Support Vector Machines (SVM) and Random Forests. Machine learning, however, evolves, and novel approaches are continually explored. In this context, our research aligns with the trajectory of leveraging both traditional and contemporary models.

2.2 Sentiment Analysis in Financial Markets

The advent of social media, especially Twitter, has transformed financial analysis. Zhang et al. (2011) and Tumarkin and Whitelaw (2001) correlated news sentiments with stock price movements. Recent studies, like Das and Chen (2007), emphasize the impact of sentiments from social media on market volatility and trading behaviour.

From rule-based methods to machine learning, sentiment analysis models have evolved. Natural Language Processing (NLP), represented by models like VADER (Hutto and Gilbert, 2014), enables nuanced sentiment analysis. Bollen et al.'s (2011) Financial Sentiment Lexicon demonstrates specialized lexicons' importance.

2.3 Deep Learning in Stock Market Prediction

Deep learning, especially Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, has redefined predictive modeling. Brownlee's (2018) work on using LSTMs for time series forecasting exemplifies their success. Studies by Ding et al. (2015) and Zheng et al. (2014) showcase the efficacy of LSTMs in forecasting stock prices.

Our research aligns with this evolution, aiming to harness the power of LSTMs for their ability to capture temporal dependencies, a crucial aspect of financial markets.

2.4 Integration of Sentiment Analysis and Deep Learning

The integration of sentiment analysis and deep learning has shown promise in enhancing predictive accuracy. Ding et al. (2014) incorporate sentiment scores from financial news into an LSTM-based model, showcasing improved performance. Mao et al. (2019) emphasizes the significance of sentiment-aware features in enhancing predictive models. The studies you provided further enrich the understanding of sentiment analysis and machine learning techniques in stock market prediction. While each study adds valuable insights, our research aims to contribute by integrating sentiment analysis from Twitter with LSTM-based architectures, providing a holistic approach to stock market prediction.

III. METHODOLOGY

1. Data Collection

The study uses two primary data sources: Yahoo Finance's stock market data and Twitter data obtained through the Snscrape programme. Stock market statistics include key components such as Open, High, Low, Close, Adj Close, and Volume.

2. Data Integration

A single dataset is produced by combining stock and Twitter data based on matching timestamps. This integrated dataset is used to build time series modelling that comes after.

3. Time Series Modeling

1. ARIMA Model

The Auto ARIMA algorithm is employed for automated model selection in the ARIMA (Autoregressive Integrated

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

Moving Average) time series model. The model is configured to predict Open and Close prices, mapping features accordingly.

2. LSTM-CNN Model

The LSTM-CNN model is designed with a 1D-CNN architecture followed by BiCudaLSTM layers. Reshaped time series data with target values and features are used to train this model. Predicting Open and Adjusted Close prices is the main goal of the training process.

4. Forecasting

Future values of the Open and Close prices are predicted using both the LSTM-CNN and ARIMA models. This phase involves predicting stock market trends based on the trained models.

5. Performance Evaluation

A range of metrics, including Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), are used to evaluate the performance of the ARIMA and LSTM-CNN models. This comparative analysis sheds light on each model's relative efficacy.

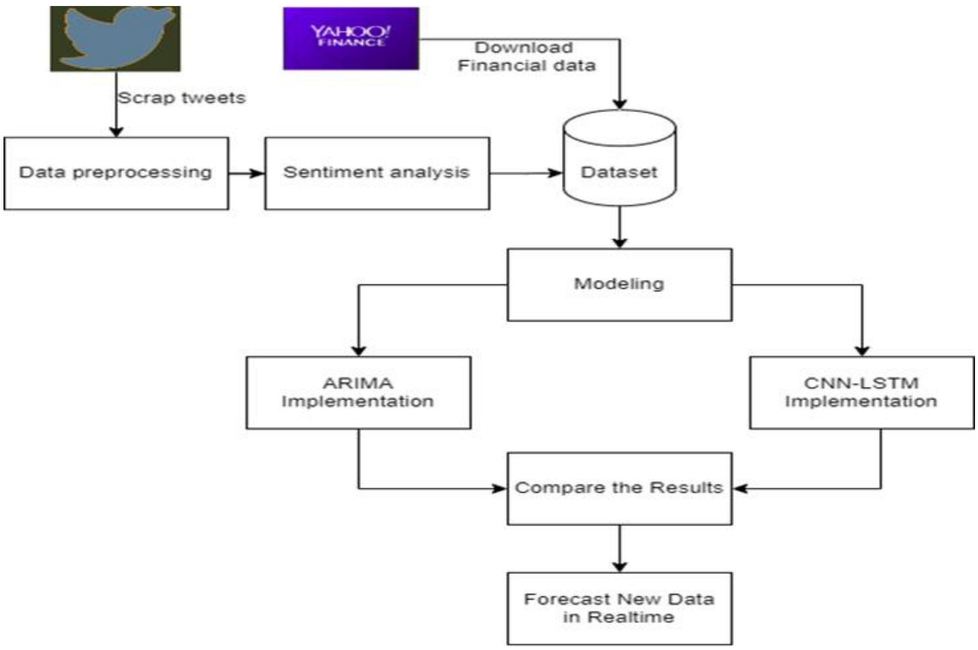

Fig 1: System Architecture

IV. RESULT AND ANALYSIS

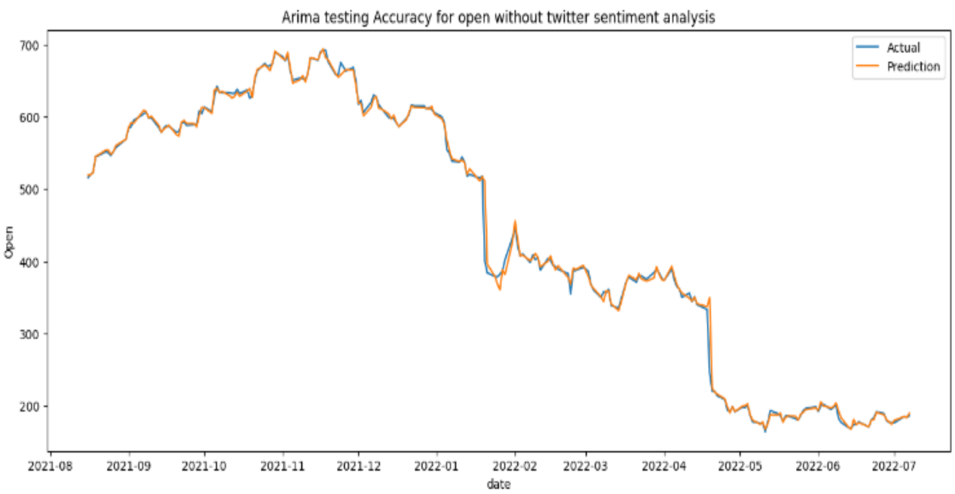

Fig 2: ARIMA Accuracy without Twitter Sentiment

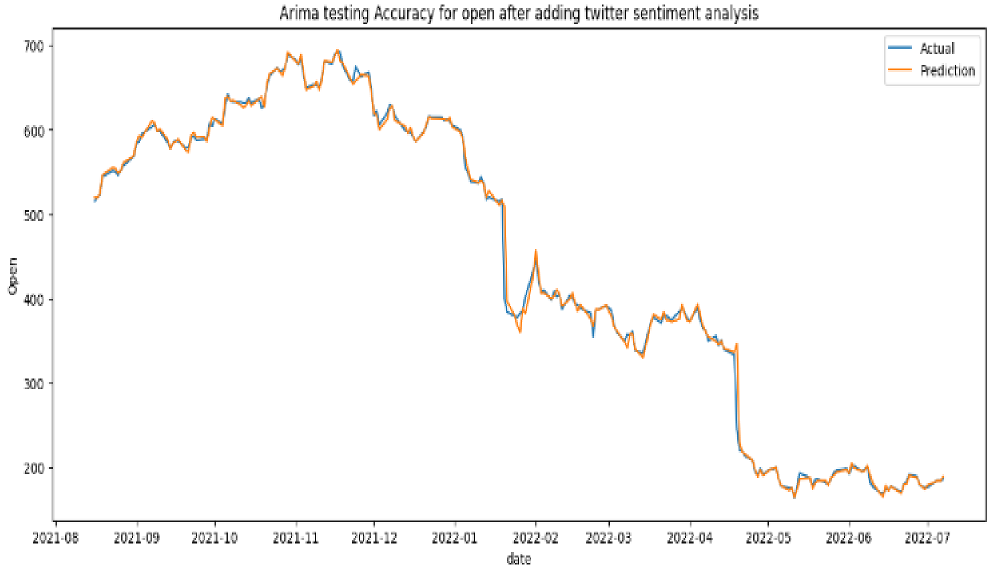

Fig 3: ARIMA Accuracy with Twitter Sentiment

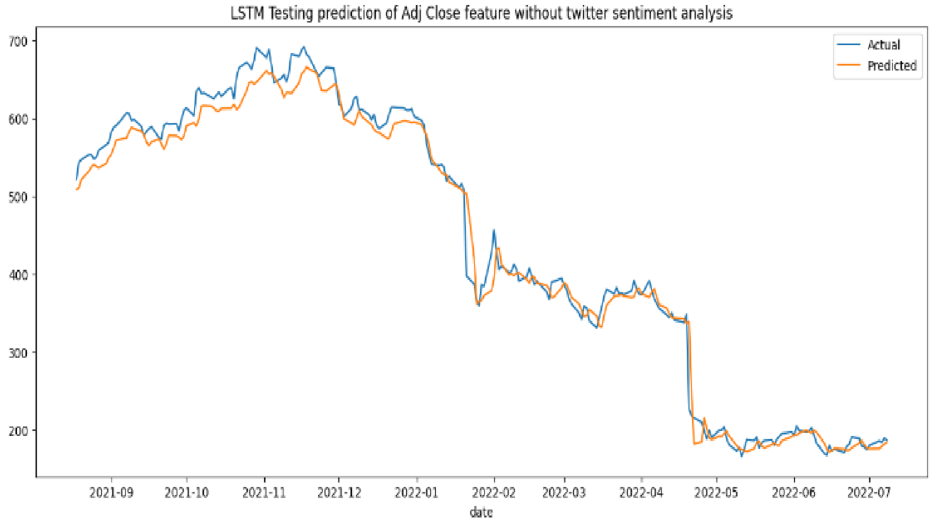

Fig 4: LSTM Adj Close Accuracy without Twitter Sentiment

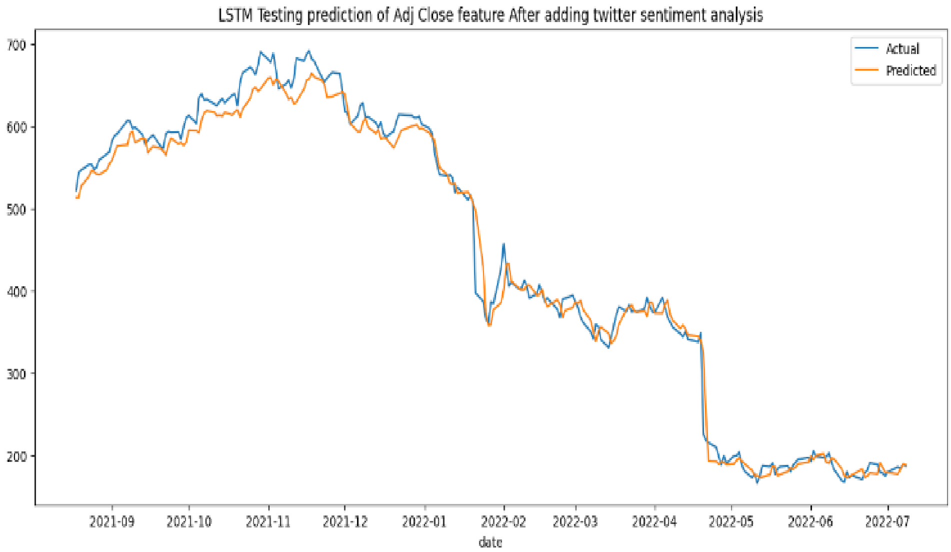

Fig 5: LSTM Adj Close Accuracy with Twitter Sentiment

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

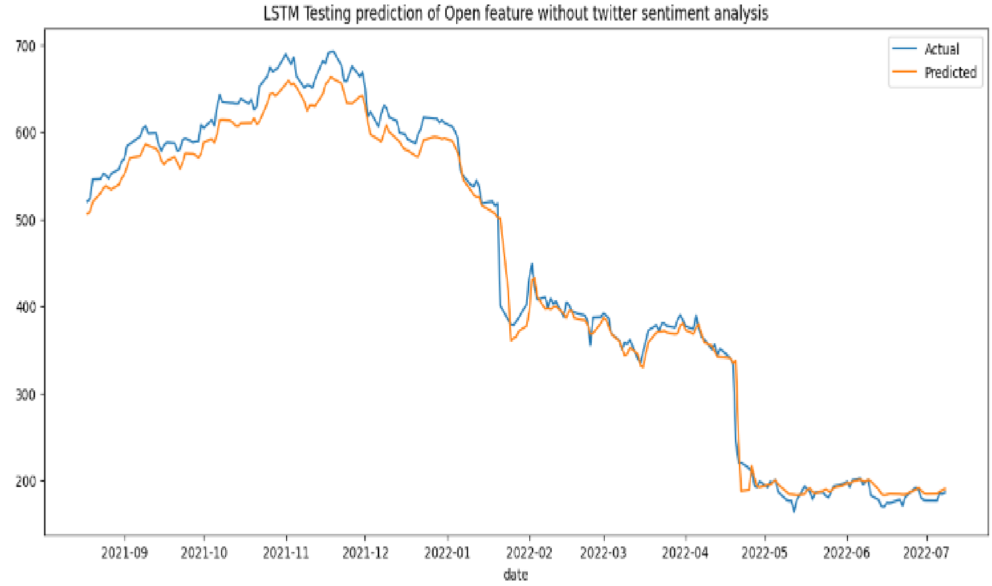

Fig 6: LSTM Open Feature without Twitter Sentiment

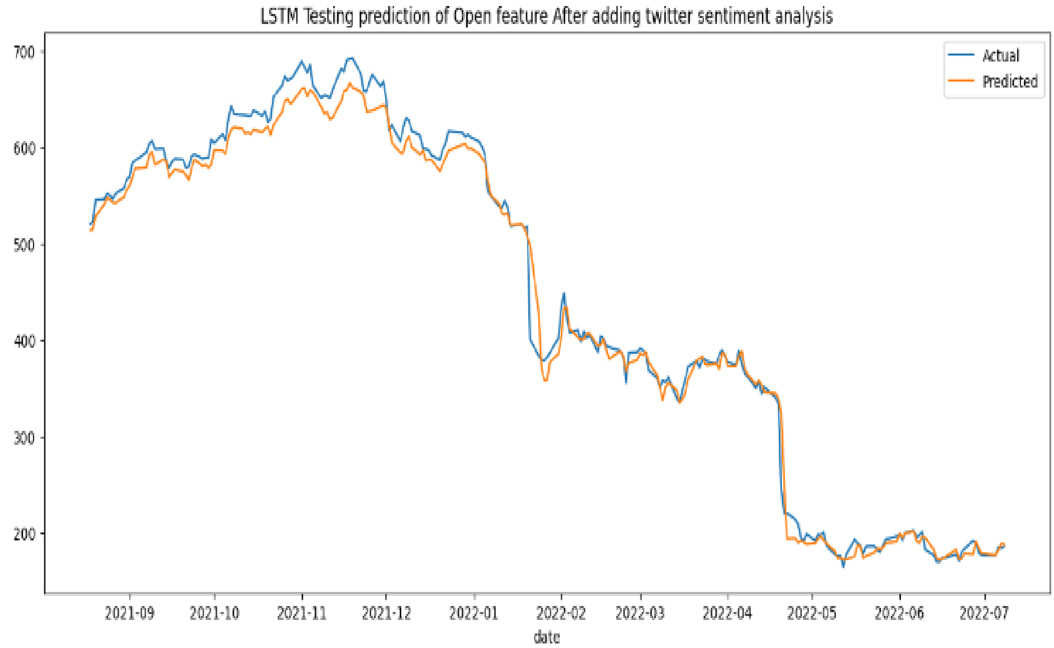

Fig 7: LSTM Open Feature with Twitter Sentiment

V. CONCLUSION

In conclusion, this study uses sentiment analysis from Twitter data to dive into the complex field of stock market prediction. A comprehensive method for predicting stock prices is provided by the combination of both modern deep learning techniques—embodied in the LSTM-CNN model—and conventional time series modelling, represented by the ARIMA model.

Using the Auto ARIMA algorithm simplifies the process of choosing an ARIMA model and makes it more flexible in accordance with the properties of the dataset. Convolutional and long short-term memory layers are used by the LSTM-CNN model to effectively capture temporal patterns and feature representations from the integrated data.

The forecasting stage shows the accuracy with which the models can predict future stock prices, giving investors and other financial decision-makers important information. Metrics for performance evaluation, such

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), make it easier to compare models and highlight their advantages and disadvantages.

A comprehensive understanding of the variables influencing stock market trends is made possible by the addition of sentiment analysis from Twitter data, which provides an additional layer of contextual information. The findings give market participants a more nuanced understanding of the possible influence of social media sentiment on stock prices.

The research findings add to the ongoing discussion about the incorporation of alternative data sources, like sentiment from social media, in stock market forecasting as financial markets continue to change. The knowledge acquired opens up new avenues for investigation and improvement of prediction models, resulting in a more comprehensive comprehension of the complex relationship between sentiment on social media and financial markets. The ultimate goal of this research is to equip investors with better tools so they can make wise decisions in the dynamically shifting world of the stock market.

VI. REFERENCES

[1] Mehar Vijha, Deeksha Chandola ,Vinay Anand Tikkiwal, Arun Kumar, (ICCIDS2019), Stock Closing Price Prediction using Machine Learning Techniques.

[2] Sandipan Biswas, Shivnath Ghosh, Sandip Roy, Rajesh Bose, Sanjay Soni, 2023, A Study of Stock Market Prediction through Sentiment Analysis.

[3] Ghaith Abdulsattar , Jabbar Alkubaisi, Siti Sakira Kamaruddin & Husniza Husni, 2018, Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers.

[4] Zhaoyu Li , December 2014 , Naïve Bayes Algorithm for Twitter Sentiment Analysis And its implementation in Map Reduce.

[5] Jyoti Gautam, Mihir Atrey, Nitima Malsa, Abhishek Balyan, Rabindra Nath Shaw, and Ankush Ghosh, April 2021, Twitter Data Sentiment Analysis Using Naive Bayes Classifier and Generation of Heat Map for Analyzing Intensity Geographically.

[6] Padmanayana, Varsha, Bhavya K, : 22 July 2021, Stock Market Prediction Using Twitter Sentiment Analysis.

[7] Samsir, Deci Irmayani, Firman Edi, Junaidi Mustapa Harahap, Jupriaman Rizki Kurniawan Rangkuti, Basyarul Ulya, Ronal Watrianthos, March 2020, Naives Bayes Algorithm for Twitter Sentiment Analysis.

[8] Shen Rong, Zhang Bao-wen, (2018), The Research of Regression Model in Machine Learning Field.