

Market Price Anticipation using Sentimental Analysis

Asst. Prof Akanksha Bana¹, Belal Ahmad² and Arpit Rai³

¹⁻³ABES Institute of Technology, Ghaziabad, India

Email: akanksha.bana@abesit.edu.in, belal2020csds097@abesit.edu.in, arpit2020csds068@abesit.edu.in

Abstract—This study combines sentiment analysis, autoregressive integrated moving average (ARIMA) models, recurrent neural network (RNN) and long short-term memory (LSTM) to provide a thorough method of stock price prediction. In order to improve the precision and resilience of stock price projections, the study attempts to capitalize on the advantages of each approach. The first step of the investigation is gathering important textual material and historical stock price data from financial news, social media, and other sources. A strong foundation for capturing both short-term dependence and long-term trends in stock prices is provided by the combination of LSTM and ARIMA models. Utilizing natural language processing (NLP) methods, the sentiment analysis component draws insightful conclusions from textual data and incorporates market sentiment as a critical component of the prediction model.

Index Terms— LSTM, ARIMA, Machine Learning, Stock Market Prediction, Financial Time Series Forecasting.

I. INTRODUCTION

The last thirty years have seen significant changes in the financial system due to rising lending rates. Some firms choose to engage in this business despite the high risk involved. Therefore, stock price increase estimate is of interest to both professional and individual investors. Collecting pertinent textual content and historical stock price data from financial news, social media, and other sources is the first phase in the research.[1] The combination of LSTM and ARIMA models provides a solid basis for capturing both long-term trends and short-term dependency in stock prices. The market (LSTM) and Graphical Processing Circuit (GRU) are the two most appropriate RNN structures that are included into the sentiment analysis component, which uses natural language processing. (NLP) techniques to extract meaningful conclusions from textual input. It has been observed that Recurrent Neural Network (RNN) is one of the several varieties of deep learning models for text analysis. Traditional stock methods, such as the ARIMA model, are utilised to build an autoregressive model that forecasts stock price. A decision tree system is suggested by the nonlinear forecasting model, which calls for enormous amounts of financial data based on rough sets. This approach combines the benefits of decision trees with rough sets, but overfitting happens. The extreme noise present in the datasets will disturb the trend. An artificial neural network, which is frequently used in time series prediction, is recommended as a solution since it works with nonlinear data models.[2] In real-world applications, the local optimal problem is solved by ANN and reduced by SVM. The model's generalizability is enhanced using SVM. Random forest gives better results than SVM after optimization of parameters Reinforcement learning, and agent-based learning apart from supervised and unsupervised learning. Q-learning is a model-free learning algorithm were used buying and selling actions can be used with a layered neural network. The stock data trend is identified by the Gaussian process. The programme will acquire knowledge of geometric shapes. Using CNN as a discriminator in

conjunction with LSTM and RNN, a generative adversarial network forecasts the price.

In this research, we apply three stock market analysis techniques: LSTM (Long Short-Term Memory), ARIMA (Autoregressive Integrated Moving Average), and UCM (Unobserved Component Model). Selecting the collection of relevant parameters for producing good forecasts is a complex process. An artificial recurrent neural network architecture with feedback connections used in deep learning is called LSTM. It has the ability to analyse complete data sequences in addition to individual data points [3]. The integrated autoregressive moving average model (ARIMA) is a generalisation that is mostly utilised in time series analysis in statistics. Equally spaced univariate time series data are evaluated and forecasted using the unobserved components model (UCM) approach. The UCMs process is also known as structural models in the time series literature. Recent research has indicated that long short-term memory (LSTM) networks are more predictively accurate.

The goal of stock market prediction is to estimate a company's stock's future worth. Stock prices and NASDAQ data were obtained via the Yahoo Finance API, while NSE data was obtained through the Alpha Vantage API. After being processed before, the data was added to machine learning models. In conclusion, the outcomes for every model demonstrate that, in comparison to alternative approaches, Neural Networks offer a higher degree of accuracy in predicting market directions. In stock market forecasting, vector support machines and case-based consultations are also used. Furthermore, they discovered that gathering event data for a predictive model is crucial to improving forecasting accuracy. For the purpose of making the most accurate short-term forecasts and projections, the web offers the most recent information available about the stock market [4]. Seasonal and non-seasonal ARIMA models are the two types of ARIMA models that can be used for prediction. Because of the characteristics of the stock data in our instance, the off-season ARIMA model was employed.

A multitude of elements, including those related to the economy, the body, psychology, reason, and other significant areas, influence stock values. Despite the difficulty in predicting stock trends, investors continue to find innovative ways to reduce investing risk and raise the likelihood of a profitable outcome. The ARIMA model was used in the proposed system to forecast the stock values of ICICI Bank and Reliance Industries, which were then implemented using a variety of Python tools. Yahoo Finance was used to get historical stock data for ICICI Bank and Reliance Industries. An effective univariate forecasting technique for projecting future values of a time series is the ARIMA model [5]. In light of the necessity of identifying a model to assess stock price fluctuations with adequate information for decision-making. One significant subgroup of prediction functions and algorithms is time-series analysis.

Predicting stock share prices is a very difficult endeavour. The feasibility of this accomplishment has been the subject of investigation for many years, and the results of published studies reveal the startling fact that most projection models are generally unable to produce accurate predictions. Because of this technique's advantage in short-term memory, it is hypothesised that this function would gain more from traditional machine learning approaches than from ones that focus on efficiency [6]. LSTM is the selected algorithm used in this work. Neural networks are really good at many things, and one of those things is predicting future stock prices. In conclusion, a hybrid LSTM and Arima-EMD financial time series forecasting model is proposed [7]. While deep learning has been employed for financial time series prediction by a number of academics, most of them have applied it to classification issues, and comparatively few have applied it to regression challenges. Furthermore, we chose the enhanced hybrid model over a number of widely used deep learning methods, including convolutional neural networks (CNN), deep belief networks, and LSTM.

II. OBJECTIVE

To develop a stock price prediction model using Python Flask that would provide Real-time or near Real-time prediction ensuring that users have access to the most current information regarding stocks.

Basic Machine learning Algorithms were used in general, to give imprecise results, so for the enhancement of the prediction values, the use of sentimental analysis is done to provide desirable accuracy.

Investigate the impact of external factors such as news sentiment, macroeconomic indicators, and global events on real-time predictions. Enhance the model to incorporate relevant external data sources and evaluate their influence on the model's reliability.

III. LITERATURE REVIEW

Predicting the economy is a dangerous endeavour since the stock market is a very intricate mathematical movement system with numerous factors influencing its fluctuation law. The simulation results show that in our suggested model for stock time series prediction, the Long Short-Term Memory Network (LSTMN) is a more

advanced RNN, or sequential network, that enables information to be stored. This model also presents the political and LSTM network models. It can address the RNN's vanishing gradient issue. One of the less well-known but no less effective variations is the Gated Recurrent Unit Network (GRU). Unlike LSTM, it just has three gates and does not maintain an internal cell state. The concealed state of the Gated Recurrent Unit contains the data stored in the Internal Cell State of an LSTM recurrent unit.[1]. This pooled data is received by the next Gated Recurrent Unit, which uses a decision tree method with a lower error rate to analyse the input data decoding and help predict the stock index. When new souls are overwritten or replaced, previous stored memory loses value over time, making RNN incapable of learning and intelligence. As an alternative, use the management reporting to determine if you should buy, sell, or hold onto the stock in its conventional form. in a few different ways.

First, we connect backpropagation and traditional Arima. The primary goal of stock price prediction is to ascertain a company's future stock value. Market conditions and various industries have an ongoing impact on stock prices, which fluctuate constantly. Machine learning methods have difficulties when dealing with high dimensionality data because highly linked dimensions or attributes may affect the model's accuracy. In order to train linear regression methods for future stock price prediction, PCA is employed to minimise dimensionality. R-squared value is used for assessment, and root is the same as square value. Keywords: r square value, root mean square error, linear regression, and principal component analysis [2]. This paper is useful for identifying different optimum principal components that are used when examining the automatic selection of parameters for techniques, such as k for k-NN, kernel parameters in SVM, and the number of PCs. Future prices are significantly impacted by sentiment analysis. As a result, the study can yield a very accurate mixture forecast.

The purpose of this study is to compare the accuracy of the ARIMA, LSTM, and Linear Regression algorithms used to predict stock prices of the Indian NSE and the American NASDAQ. These machine learning techniques are used to both historical stock prices from two years ago and current market prices. Though LSTM and Linear Regression seem more effective than ARIMA in the NSE (Indian Company) stocks [3]. The suggested algorithms perform optimally when used with NSE stocks and NASDAQ stock market data. The primary drawback of stock price prediction is that it is not subject to time-based fluctuations. In addition, the stock price is subject to influence from economic, socio-political, and corporate factors as well as investments, current affairs, and competing company listings [4]. These occurrences are not taken into account by the aforementioned models when forecasting the stock price to predict the stock price more accurately we need to involve these events and also add the annual budget of the company while forecasting using the sentimental analysis Like a new model can be made by training the model with ANN and predicting the stock price with ARIMA while using the sentiments, this can be a future research work for Stock price prediction.

The volatile nature of the Stock Market predicts it is a cumbersome task. Stock prices are influenced by not only economic factors but also various physical, psychological, rational, and other important parameters. In this research work, the prediction of stock prices is conducted utilizing the Auto-Regressive Integrated Moving Average (ARIMA) Model. The volatile nature of stock prices makes them difficult to predict. The specific instances of ICICI Bank and Reliance Industries have been used for verifying the hypothesis [5]. The only drawback of this analysis is that the ARIMA model holds higher accuracy for short-term predictions. Time-series analysis is an important subset of prediction algorithms and functions. It is regarded as an apt tool for predicting the trends in the stock market and logistics Top of Form.

Predicting stock index prices is a challenging endeavour, and share price information is a highly complex, volatile, and dynamic environment due to a multitude of technological and non-technology factors. In addition to deep learning approaches, numerous scholarly investigations from various fields have been conducted to address this issue; among the numerous technologies employed are machine learning techniques. When used to this kind of prediction, several machine learning techniques in this sector were able to yield results that were satisfactory [6]. This study examines the prediction of stock market prices using the Long Short-Term Memory (LSTM) model on historical stock index price data. Indications analysis is also included to enhance the accuracy of the results. The goal of this study was to gather historical price data sets for Agilent Technology's common stock and American Airlines Group's common stock. Several LSTM tests were conducted, and the results were assessed using RMSE and RMSPE values, which ensure improved performance for the LSTM approach employed.

IV. DATA FLOW DIAGRAM

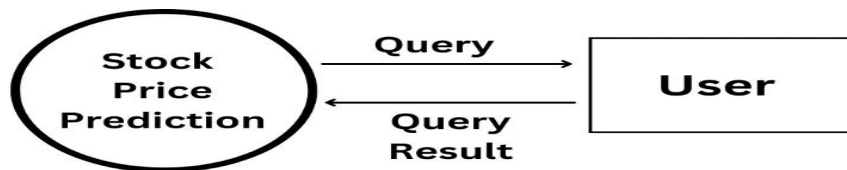


Fig 1 Level 0 dfd

V. METHODOLOGY

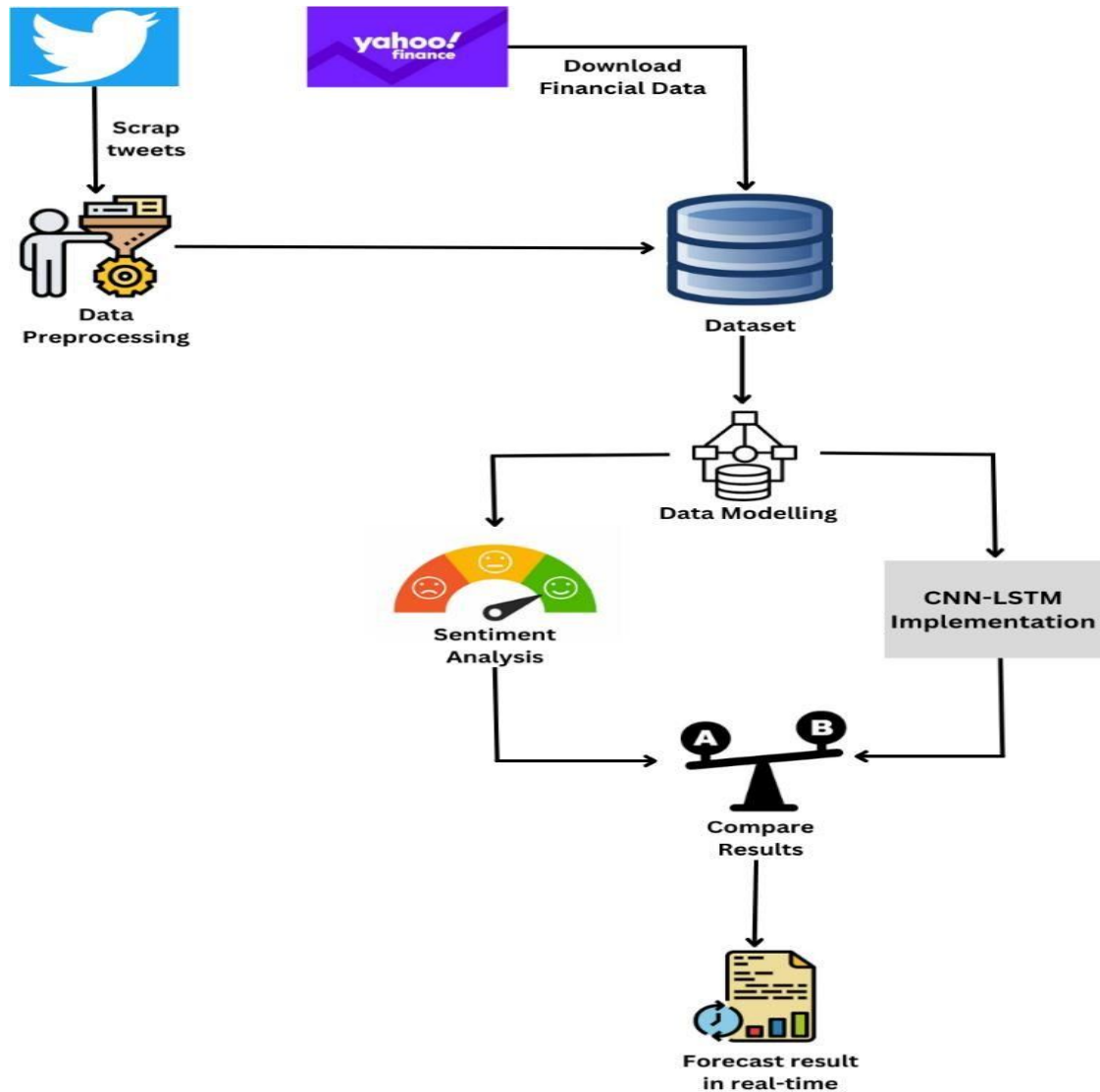


Fig.2 Methodology

A. Data Gathering

Financial Data from Yahoo Finance: This involves accessing historical financial data, such as stock prices, trading volumes, and other relevant metrics, from Yahoo Finance's database. This data serves as the foundation for forecasting future trends.

B. Data Pre-processin

Cleaning and Preparation: This involves transforming raw data into a usable format for analysis. This might include: Addressing missing values by imputation or removal, ensuring consistent formatting

C. Sentimental Analysis

Analysing Twitter Sentiment: This step employs techniques to extract sentiment from the collected tweets. It aims to determine whether the overall sentiment is positive, negative, or neutral towards the financial assets in question. This information can be integrated into the forecasting models.

D. Data Modelling

CNN-LSTM Implementation: This step explores a deep-learning approach that combines CNNs and LSTMs. CNNs excel at extracting features from data, while LSTMs are well-suited for capturing long term dependencies in time series data.

E. Result Comparison

Evaluating Model Performance: The accuracy and effectiveness of the and CNN-LSTM models are compared using appropriate evaluation metrics. This helps determine which model produces more reliable and accurate forecasts.

F. Real-Time Forecasting

Generating Forecasts: The chosen model, whether ARIMA or CNNLSTM, is used to generate real-time forecasts of financial data. These forecasts can inform investment decisions, risk management strategies, and other financial activities.

VI. RESULT

The research project proposes an integrated approach for precision for the prediction of stock prices for various companies whose stock ticker symbol is present on Yahoo Finance. In our web app we have taken real time data present on yahoo finance website of various stock ticker symbols till year 2022 and used LSTM and Sentimental Analysis integrated model for better accuracy in the prediction. In comparison with above researches done on this topic it has been seen that integrated model shows more accuracy.

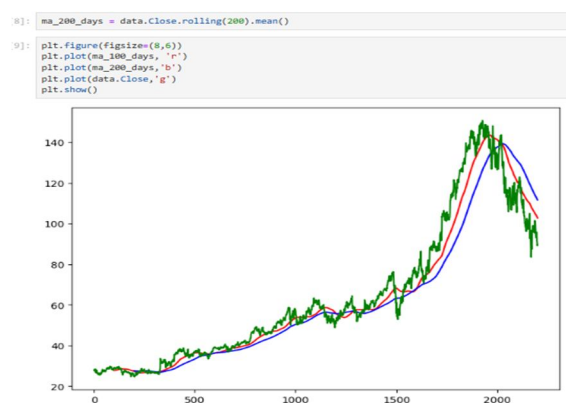


Fig.3 Comparison MA100 vs MA200 vs Price

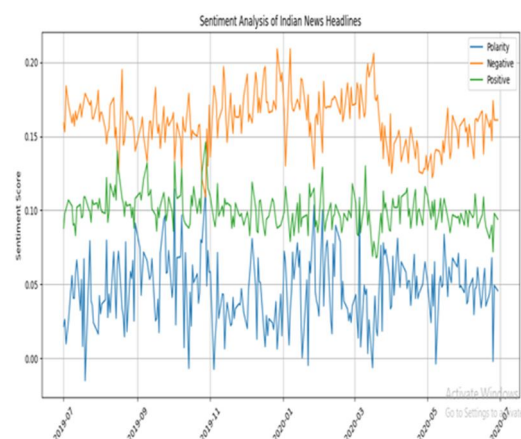


Fig.4 Graph of polarity from sentimental analysis

In the fig 3. We see the comparison of moving averages MA (100 Days) vs MA (200Days)vs Original Price , the technique of moving averages is type of RNN method of machine learning which help us to compare the average of stock data of 100 days than 200 days for better prediction.

In Fig.4. we analyse sentiments from news headlines taken till year 2022 from where we get the polarity graph of positive, negative and polarity lines in which polarity shows that a news is how much positive or negative. for its effect on stock prices.

Fig.5 In this picture we see the Graphical User Interface which consist of a heading showing the name of web app that is "MARKET PRICE ANTICIPATION" below it is the space for writing stock ticker symbol by which real time data of that stock is fetched and prediction is made.

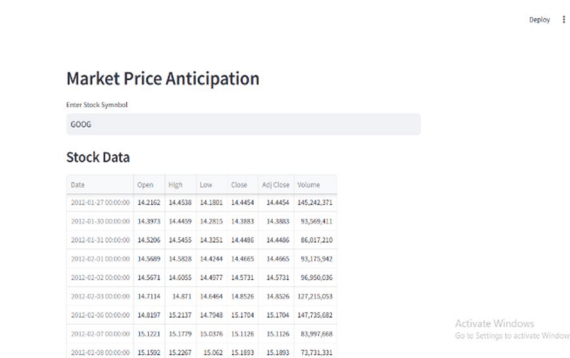


Fig.5 Stock Prediction web app GUI

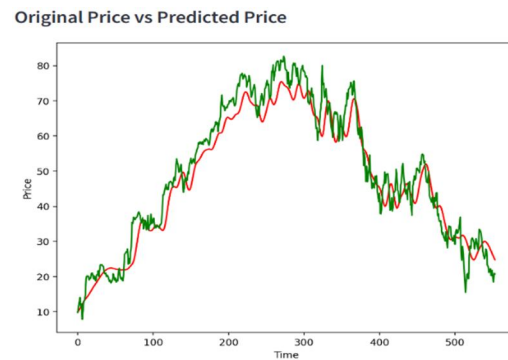


Fig.6 Result showing prediction of google stock

In fig.6 shows the Result of above illustrated model in which there are two lines where green line is the actual price of the stock and the red line is the predicted line of the stock. The gaps in between the green line and red line show the amount of loss made by user in the real investment.

VII. FUTURE SCOPE

Future development of sentiment analysis, LSTM, and ARIMA-integrated stock price prediction models has enormous potential to handle the intricate and dynamic structure of financial markets. It is expected that more advanced LSTM architectures that can recognise complex patterns in stock price data would be produced by ongoing research at the nexus of deep learning and time series forecasting. The integration of sophisticated sentiment analysis methods, utilising advancements in natural language processing, is anticipated to augment models' comprehension of textual information and elevate the precision of forecasts influenced by market sentiment. Further broadening the feature area for prediction models could include the inclusion of macroeconomic indicators and different data sources in the future.

VIII. CONCLUSION

The integration of LSTM, ARIMA, and emotional analysis has resulted in a comprehensive and inventive approach for financial market forecasting, marking the end of this final year research that was centred upon stock price prediction. LSTM has made it easier to represent complex temporal connections. A useful layer of contextual information has been added with the integration of sentiment analysis, which measures market sentiment from textual data. The study has produced a predictive model and given insights into the interpretability and explainability of stock price fluctuations through intensive experimentation and analysis. The efficacious execution of these techniques highlights their possible relevance in actual financial situations.

REFERENCES

- [1] M. Sri and R. U. Kiran, "Stock Price Prediction Using LSTM and GRU", doi: 10.32628/IJSRCSE.
- [2] Prof. Sulochana Sonkamble, Vaibhav Vyas, Prathamesh Shimpi, Aniket Mule, and Mihir Sonawane, "Stock Price Prediction System," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, pp.273–277, Apr. 2023, doi: 10.32628/cseit2390229.
- [3] N. Sharma, P. Pareek, A. Ghosh, and K. Nagarohith, "A Peer Reviewed Open Access International Journal," Journal of Development Economics and Management Research Studies (JDMS), no. 11, pp. 55–66, doi: 10.53422
- [4] K. Sohil Patel, U. Rajesh Prahladka, J. Babulal Patel, and Y. Shelar, "Stock Price Prediction using Machine Learning Algorithms: ARIMA, LSTM & Linear Regression," International Research Journal of Engineering and Technology, 2022
- [5] A. Ganesan and A. Kannan, "Stock Price Prediction using ARIMA Model," International Research Journal of Engineering and Technology, 2021,
- [6] W. Makhoulf, R. Malhotra, and A. Head, "Article ID: IJARET_11_10_160 Cite this Article: Walaa Makhoulf and [12] Dr. Ruchika Malhotra, Stock Market Price Prediction Using LongShort Term Memory (Lstm)," International Journal of Advanced Research in Engineering and Technology (IJARET), vol. 11, no. 10, pp. 1692– 1707, 2020, doi: 10.34218/IJARET.11.10.2020.160
- [7] M. Ali, D. M. Khan, H. M. Alshanbari, and A. A. A. H. El-Bagoury, "Prediction of Complex Stock Market Data Using an Improved Hybrid EMD-LSTM Model," Applied Sciences (Switzerland), vol. 13, no. 3, Feb. 2023, doi: 10.3390/app13031429.

- [8] K. Yadav, M. Yadav, and S. Saini, "Stock values predictions using deep learning based hybrid . models," *CAAI Trans Intell Technol*, vol. 7, no. 1, pp. 107– [15] 116, Mar. 2022, doi: 10.1049/cit2.12052.
- [9] A. K. V Sivanagapavan, A. Divya, B. Nisha, A. Sai, V. Kumar, and K. Bhargavi, "STOCK MARKET PREDICTION," *The International journal of analytical and experimental modal analysis*, vol. 12, no. 5, pp. 1598–1524, May 2020,
- [10] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," *Artif Intell Rev*, vol. 53, no. 4, pp. 3007– 3057, Apr. 2020, doi: 10.1007/s10462-019- 09754