# Forecasting Sensex: A Comparative Analysis of Predictive Models

Shadab Qureshi, Tanmay Agarwal, Yashwardhan Chouksey

Department of Computer Engineering, STME , NMIMS, Indore

*Abstract—*

In this paper we check out the evolving terrain of stock markets via discussing the Sensex, India's iconic fairness index. The take a look at amalgamates critical sides of stock market evaluation: stock visualization and charge prediction. It seeks to augment our understanding approximately the intricacies of the inventory market and facilitates better-informed choice making on this context. Our research begins with an exploration of inventory visualization modes. Our intention is to offer insight into the intricacies of Sensex styles and traits via the usage of increase ddata visualization equipment and strategies. We analyse and visualize the beyond stock statistics to offer our readers a extra transparent view of the marketplace behaviour. Our 2d section delves into price prediction. We use system gaining knowledge of algorithms and statistical models to create extra correct predictions of Sensex price movements. Our goal has been to use past data, market indicators and other macroeconomic elements to assess potential future paths for buyers.Our look at utilises a large facts-set over multiple years that is consultant of varied monetary and geopolitical conditions.

KEYWORDS: Sensex, Machine learning, Deep learning, Artificial intelligence, Python, SQl

## I. INTRODUCTION

In the ever-evolving global of economic markets, the want for accurate and sturdy forecasts is critical to knowledgeable choice making. Sense[1] index is an critical indicator of monetary developments and displays the pinnacle groups at the National Stock Exchange of India. This study is dedicated to a complete evaluation of historical rate information of Tata Motors[2]and HDF[3]. Our fundamental goal is to create reliable merchandise and price estimates for those corporations, as a consequence offering a higher information of stakeholders within the economic marketplace. To attain this, we use three unique forecasting tools: Auto ARIMAX[4], Facebook Prophet[5] and LightGBM[6]. These fashions all have particular features and skills, making them appropriate for a extensive variety of time predictions. The motives behind this take a look at are many: First of all, Tata Motors and HDFC are essential. Share fees are stricken by many factors which include international change, enterprise trends and organization development. Accurate predictions help investors and market members determine whether or not to buy, promote or maintain those shares. Secondly, Sensex, which is partly stimulated with the aid of Tata Motors and HDFC, affords

benchmark information for stocks. Serve as a reference point for the overall market. Predicting the share prices of these companies can provide insight into underlying business trends and overall business performance.Thirdly, linear regression[7], a trader can identify key price points—entry price, stop-loss price, and exit prices. A stock's price and time period determine the system parameters for linear regression, making the method universally applicable.

The three models chosen for this study (Auto ARIMAX, Facebook Prophet and LightGBM) represent various methods for time estimation. Auto ARIMAX uses the autoregressive integrated moving average method of exogenous variables. Facebook Prophet specializes in capturing seasonal and holiday effects in real-time data. LightGBM is a versatile gradient boosting framework known for its efficiency and adaptability. By applying this model to Tata Motors and HDFC stock price data, we aim to evaluate their performance in the context of the Sensex performance index and their suitability for predicting stock prices.

Financial analysis aims to optimize strategies in the dynamic Indian economy. The results of this research can inform investment decisions and strategies and provide people with important tools to navigate the challenging financial environment.

II. LITERATURE SURVEY

In this examine, the focal point is on sentiment analysis inside the context of the Indian inventory marketplace. The research explores various sentiment category methodologies, which include device learning and lexicon-primarily based processes. The proposed machine makes use of Python programming to fetch stay records from on-line news portals and perform sentiment analysis on datasets. The findings suggest a giant improvement in predicting inventory market conditions. By using sentiment evaluation, buyers can make more knowledgeable selections when it comes to stock marketplace investments, ultimately contributing to a extra balanced and economically solid percentage market[8].

This survey delves into the dynamic intersection of finance and deep gaining knowledge of by using summarizing the modern day improvements in stock marketplace prediction strategies. Focusing at the length from 2017 to 2019, it covers the whole predictive workflow, from information series and processing to version development and assessment. Emphasizing the need for reproducibility and sharing of code, the survey offers a complete review of the evolving landscape of deep learning models for inventory market forecasting. Its findings provide valuable insights that expand beyond stock prediction, reaping benefits other financial time series forecasting domain names like alternate charges and cryptocurrency expenses[9].

This look at addresses the tricky assignment of predicting stock marketplace fluctuations by way of comparing diverse predictor models over different time horizons, starting from 1 to 30 days in advance. The studies spans a decade of data from the Iranian inventory market, encompassing 4 distinct marketplace agencies. By employing six tree-based models (Decision Tree, Bagging,

Random Forest, Adaboost, Gradient Boosting, and XGBoost) and three neural network-based totally algorithms (ANN, RNN, and LSTM), the examine rigorously examines prediction overall performance. The results screen the dynamic nature of inventory values and the importance of version parameter choice in accomplishing accurate predictions, shedding light on both profitable opportunities and risk reduction for traders inside the Iranian inventory market[10].

This study, investigates the application of deep getting to know strategies, in particular LSTM neural networks with embedded layers and automatic encoders, for stock marketplace prediction, focusing at the Shanghai A-percentage composite index and specific stocks. The findings display promising predictive accuracy, outperforming stochastic techniques, and highlighting the capability of these models for economic analysis. Additionally, this study underscores the significance of interdisciplinary tactics for advanced stock market forecasting. Future paintings may involve incorporating text data and expanding to different international markets, aiming for even higher accuracy in predictions[11]

This paper explores the software of Artificial Neural Networks (ANNs) for stock market prediction, highlighting the efficacy of ANNs in dealing with complicated, nonlinear relationships in monetary facts. The look at showcases the superior overall performance of ANNs when in comparison to different models and algorithms, such as genetic algorithms and linear regression analysis, in predicting numerous stock market parameters. These findings emphasize the capability of ANNs as a treasured tool for stock market forecasting and endorse future possibilities for his or her utility in distinct domain names and comparative studies with alternative models.[12]

This paper gives a complete review of the software of Artificial Neural Networks (ANN) in inventory marketplace prediction. ANNs have established their effectiveness in forecasting stock marketplace trends, capitalizing on their ability to examine, generalize, and adapt to non-linear facts styles. Researchers have constantly sought to decorate prediction accuracy thru hybrid models and by using thinking about additional outside elements. The dynamic and non-linear nature of the inventory marketplace, encouraged by using diverse outside elements, gives ongoing challenges and opportunities for improving prediction accuracy using ANN-primarily based procedures.[13]

This paper delves into the world of inventory market fee prediction, a multidisciplinary discipline at the intersection of pc technology, economics, finance, and management. It highlights the complexity of stock statistics, often unsure and incomplete, making accurate predictions hard yet potentially rewarding. The have a look at explores the efficient market speculation (EMH) and emphasizes the usage of machine gaining knowledge of methods, particularly Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees. The researchers advise that those strategies, whilst carried out to numerous records resources and sentiment analysis, can beautify inventory marketplace prediction accuracy. The findings demonstrate promising

consequences, showcasing the ability for progressed prediction fashions via leveraging advanced generation and artificial intelligence.[14]

This paper explores the fascinating realm of stock market prediction, a hard undertaking that has garnered sizable interest from the AI network. With the evolution of synthetic neural networks (ANNs), traders keep high hopes for unraveling the intricacies of financial markets. ANNs, known for their pattern popularity and machine getting to know capabilities, have emerge as a common records mining approach across various domain names, together with economics, commercial enterprise, and science. What units ANNs aside is their potential to discover nonlinear relationships within complex records, making them advanced equipment for stock market prediction in comparison to standard fashions. By efficaciously approximating complicated capabilities, ANNs play a pivotal position in modeling the relationships between input and output. The paper also delves into stock marketplace basics and the performance of the inventory marketplace in keeping with the efficient market hypothesis (EMH). Overall, ANNs provide promise in enhancing stock marketplace prediction, with their adaptability to noisy information and capacity to tackle difficult nonlinear demanding situations, setting them aside as a precious tool in this area.[15]

This paper delves into the fascinating realm of inventory marketplace prediction with the aid of presenting a genetic set of rules (GA) optimized hybrid system. This system, combining selection timber and help vector machines (SVM), ambitions to are expecting one-day beforehand trends in inventory markets. It extracts and selects relevant capabilities from ancient information using a decision tree and employs SVM for trend prediction. The hybrid gadget in addition optimizes parameters using GA to beautify prediction accuracy. The study validates the gadget's performance the use of BSE-SENSEX records and compares it with artificial neural community (ANN) and Naïve Bayes-primarily based systems. Impressively, the hybrid machine outperforms the others, supplying capacity benefits in stock marketplace fashion prediction and trading profits.[16]

The international of financial markets is a dynamic and complicated realm that extensively affects diverse components of our lives. This paper delves into the multifaceted area of inventory marketplace analysis and prediction. It highlights the diverse variety of techniques and processes used to decipher the intricacies of monetary markets, encompassing statistical methods, sample reputation, gadget learning, and sentiment analysis. The study affords a taxonomy of recent advancements in these regions, losing mild at the successes and challenges they pose. It emphasizes the significance of sturdy stay trying out, algorithmic trading, and sentiment evaluation while spotting the ongoing evolution of economic markets and the need for progressive research to cope with emerging demanding situations and beautify marketplace analysis in an increasingly interconnected worldwide panorama.[17]

This observe focuses on leveraging Artificial Neural Networks (ANNs) to are expecting future percentage costs and assess the growth of organizations across diverse sectors. The objective is to

decide the choicest time span for forecasting percentage costs, thinking about exclusive sectors. Using Long Short-Term Memory (LSTM) networks, the take a look at makes predictions for 3-month, 6-month, 1-year, and three-12 months time frames. Results suggest that errors tiers lower notably with longer test periods, suggesting the effectiveness of LSTM-based totally models for long-term ancient records. The studies concludes that corporations inside the identical region showcase comparable dependencies and growth charges, facilitating marketplace analysis and enhancing prediction accuracy via incorporating extra relevant parameters.[18]

This take a look at introduces a novel approach, known as Adaptive PSO-LSTM, which combines adaptive particle swarm optimization (PSO) with deep gaining knowledge of for inventory fee time series forecasting. The objective is to optimize the weights and biases of lengthy quick-time period memory (LSTM) networks and completely related layers (FCL) the use of PSO, thereby improving the forecasting accuracy. The technique is carried out to a few stock marketplace indices over numerous time horizons, and the outcomes reveal the superior overall performance of Adaptive PSO-LSTM as compared to different strategies, along with traditional LSTM, PSO-LSTM, and GA-LSTM. The studies indicates the ability for further exploration of nature-inspired and evolutionary techniques in monetary time collection forecasting.[19]

In this examine, the assignment of predicting stock market developments is addressed the use of deep getting to know fashions, such as a Long Short-Term Memory (LSTM) recurrent neural network and a Deep Neural Network (DNN). The Efficient Market Hypothesis has lengthy advised that predicting inventory markets is quite tough because of the multitude of variables and human sentiments at play. The observe evaluated the fashions' predictive skills the usage of metrics which includes Root Mean Square Error (RMSE), forecast bias, and directional accuracy (DA). The results confirmed that both the LSTM and DNN models finished properly in each day predictions, with the LSTM excelling in making weekly predictions. These models exhibit promise for figuring out developments and making longer-term predictions in risky inventory marketplace datasets, specifically while extra attributes are included and overfitting is averted.[20]

## III. METHODOLOGY

### A. Data Collection

Our research is based on historical price and volume data of 50 stocks comprising the Sensex index of the National Stock Exchange(NSE)[21]of India. The dataset contains daily data where price and volume for each stock are separated into separate .csv files. We also have a metadata file containing macro information about the product. The data range is from January 1, 2000 to April 30, 2021.

### B. Data Pre-Processing

Ensuring the quality of data and validity of analysis had previously become difficult. Preliminary data is the foundation of all data research and forms the basis for meaningful insights. Using the Python programming language and the pandas library, we performed several important steps, including loading data from CSV files, adjusting the required data, understanding the properties of the data, and performing quality data checks to identify and correct gaps or inconsistencies. This data processing ensures that the data set is ready for subsequent analysis in financial analysis and business forecasting.

C. Model Selection

Selection of the forecast model is an important decision that affects the quality and reliability of the forecast. In our studies, we selected a aggregate of three one of a kind models to estimate market value: Auto ARIMAX, Facebook Prophet and LightGBM. Each version is selected based totally on its specific strengths and features. - Auto ARIMAX: Known for its properly consequences within the evaluation length, this model is right for capturing the effects of external have an impact on on inventory charges as it is able to comply with special parameters. Automated pattern choice technique ensures the first-class selection. - Facebook Prophet: Selected for its information in working with seasons and vacations in economic facts and its strengths in statistics control.The intuitive interface simplifies the process and adapts well to various data complexities and external influences. LightGBM: Required for large files and complex environments due to its higher speed and efficiency. Although primarily used for machine learning, LightGBM's flexibility makes it useful for real-time prediction.

Our model selection is based on this model's collective ability to identify competitor costs including seasonality, other factors, missing data, and the study was good. We aim to provide a comprehensive and robust analysis of financial data by leveraging the strengths of each model.

D. Model training

Each model completes the training process according to its own criteria.

- Automatic ARIMAX: A careful training process involves automatic decision making, including autoregressive (AR) and moving average (MA) orders from grid search and model fit evaluation. Exogenous variables are included to capture the impact of other factors on stock prices.

- Facebook Prophet: The model's user-friendly interface makes it easy to train and see seasonal patterns and holiday effects. Robustness to missing data allows consistent training that adapts to complex data and outliers.

- LightGBM: This gradient boosting machine learning framework is efficiently trained to optimize hyperparameters and fine-tune feature engineering. Its performance in big data processing and improved forecasting capabilities with transformation for regression tasks and time series forecasting.

-Linear regression : Linear regression is the analysis of separate variables to outline a single dating and is a useful degree for technical and quantitative analysis in monetary markets. Plotting stock charges alongside a ordinary distribution—bell curve—can permit buyers to look when a stock is overbought or oversold.

By tailoring the training process to each model's unique strengths and needs, we ensure our predictive models are successful in capturing the nuances and trends of Tata Motors and HDFC pricing information. This training model provides a basis for subsequent evaluation and validation of our predictions.

IV. RESULTS

A.Model Performance

In our research, we applied three different forecasting models—Auto ARIMAX, Facebook Prophet, and LightGBM—on our dataset. The performance of each model was evaluate using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics.
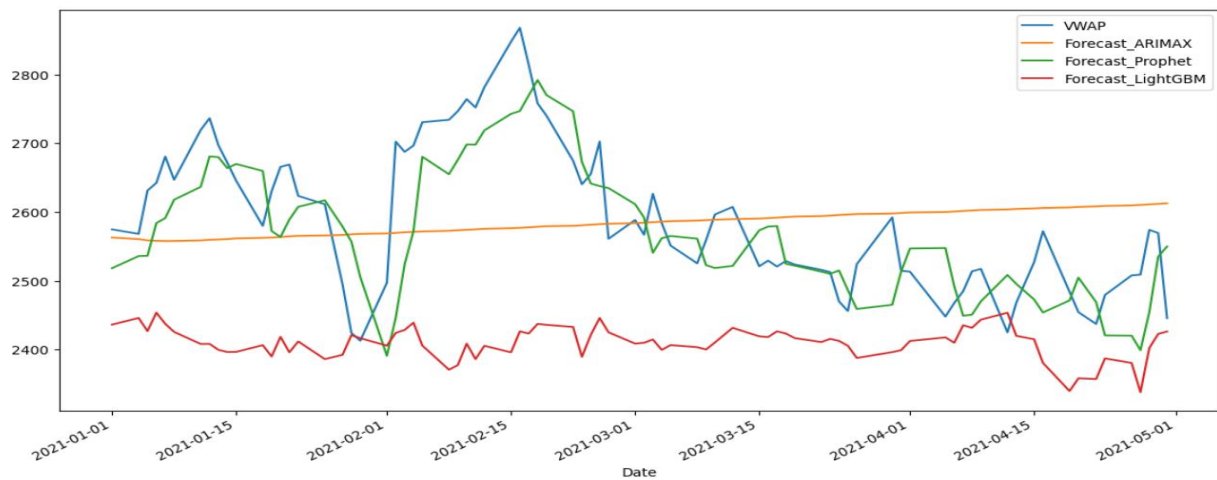
B.Interpretation of Result

In conclusion, based on the provided RMSE and MAE values:

| Company Name | HDFC | | Tata Motor | |
| --- | --- | --- | --- | --- |
| Model /value | RMSE | MAE | RMSE | MAE |
| *Auto ARIMAX* | 118.404 | 101.198 | 116.49 | 109.84 |

| | | | | |
|---|---|---|---|---|
| *Facebook Prophet* | 74.168 | 60.779 | 33.915 | 39.076 |
| *LightGBM* | 212.53 | 184.25 | 33.915 | 17.307 |
| Linear Regression | 0.999 | 3.753 | 0.956 | 3.254 |

*Table 1:table show the value of RMSE and MAE of the above model*



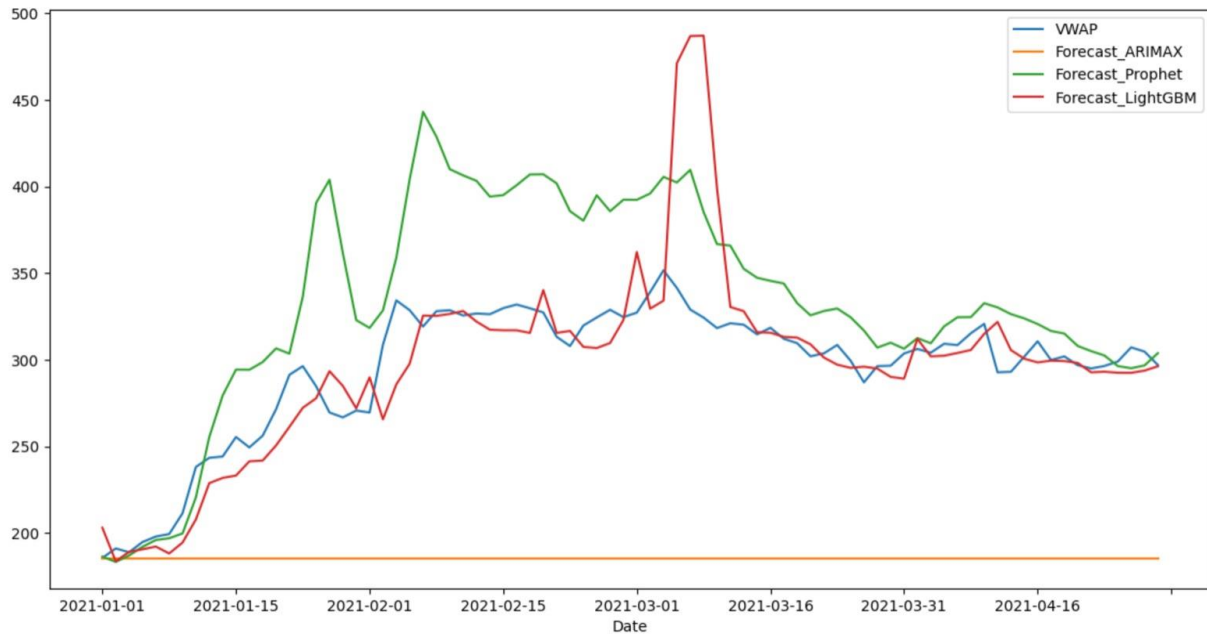*Figure 1:Predicted graph of HDFC contain ARIMAX ,Prophet and LightGBM Forecast model with VWAP graph line*

*Figure 2:predicted graph of Tata Motor contain ARIMAX ,Prophet and LightGBM Forecast model with VWAP graph line*
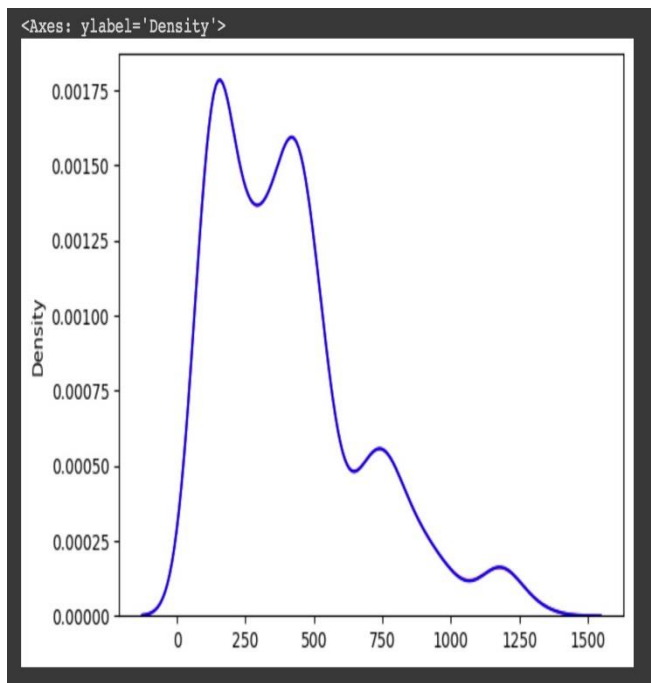


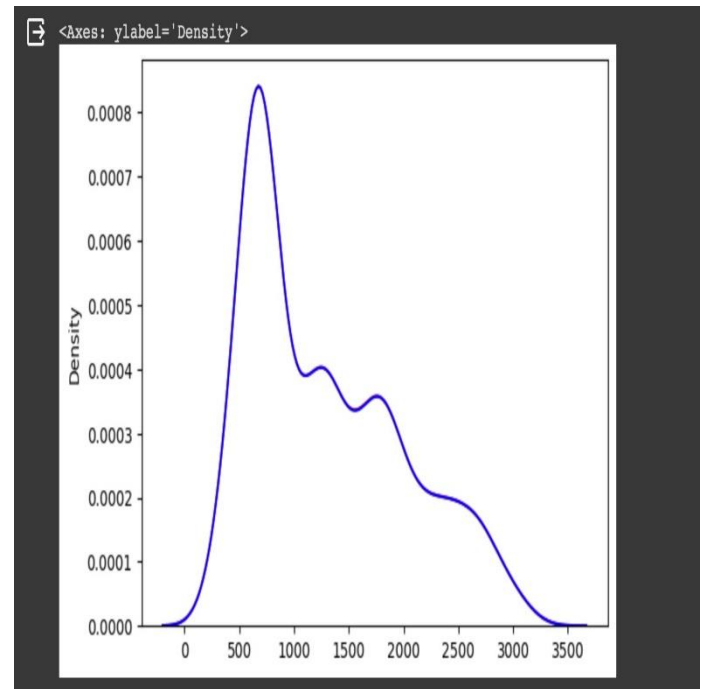*Figure 4: predicted graph of Tata Motor contain linear regression*



Figure 3predicted graph of HDFC contain linear regression

## V. CONCLUSION

In summary, considering Tata Motors' RMSE and MAE values and HDFC data, we can make the following important observations:

In these two cases, Facebook Prophet consistently proves to be the most accurate forecasting model. Its excellent ability to capture seasonal and holiday effects makes it reliable in analyzing financial data over time.

LightGBM outperforms Auto ARIMAX, maintaining good performance on both datasets. Although it cannot provide the same accuracy as Facebook Prophet, its results and flexibility make it a useful tool for financial transactions.

Linear regression device studying method is best for predictive evaluation of statistical records. The authors located that the linear regression model is perfect for predicting stock expenses.

In Tata Motors and HDFC datasets, Auto ARIMAX always shows the highest error, indicating low prediction accuracy. This suggests that it may not be the most appropriate model for financial forecasting in certain situations.

It is important to emphasize that other factors besides RMSE and MAE should be considered in the selection of the best prediction model in any exercise. Factors such as model complexity, interpretability, computational efficiency, and data specificity should be considered. Selecting the most appropriate model should meet the specific needs and objectives of the current study or application to ensure it works well for the purpose.

REFERENCES

[1] Das, N., & Pattanayak, J. K. (2009). Analysis of the fundamental factors affecting the market price of shares constituting the Indian index: a study of SENSEX. *International Journal of Accounting and Finance*, *1*(3), 323-355.

[2] https://www.kaggle.com/datasets/meetnagadia/hdfc-stock-price-20002021

[3] https://www.kaggle.com/datasets/manishkr1754/tata-motors-employee-reviews-dataset

[4] Shilpa, G. N., & Sheshadri, G. S. (2019). ARIMAX model for short-term electrical load forecasting. *Int. J. Recent Technol. Eng.(IJRTE)*, *8*(4).

[5] Daraghmeh, M., Agarwal, A., Manzano, R., & Zaman, M. (2021, June). Time series forecasting using Facebook Prophet for cloud resource management. In *2021 IEEE International Conference on Communications Workshops (ICC Workshops)* (pp. 1-6). IEEE.

[6] Chen, C., Zhang, Q., Ma, Q., & Yu, B. (2019). LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. *Chemometrics and Intelligent Laboratory Systems*, *191*, 54-64.

[7] Yao, W., & Li, L. (2014). A new regression model: modal linear regression. *Scandinavian Journal of Statistics*, *41*(3), 656-671.

[8] Bhardwaj, A., Narayan, Y., & Dutta, M. (2015). Sentiment analysis for Indian stock market prediction using Sensex and Nifty. *Procedia computer science*, *70*, 85-91.

[9] Agrawal, J. G., Chourasia, V., & Mittra, A. (2013). State-of-the-art in stock prediction techniques. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *2*(4), 1360-1366.

[10] Jorgenson, D. W., Weitzman, M. L., ZXhang, Y. X., Haxo, Y. M., & Mat, Y. X. (2023). Can Neural Networks Predict Stock Market?. *AC Investment Research Journal*, *220*(44).

[11] Patel, R., Choudhary, V., Saxena, D., & Singh, A. K. (2021, June). Review of stock prediction using machine learning techniques. In *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 840-846). IEEE.

[12] Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies*, *7*(2), 26.

[13] 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.

[14] Dase, R. K., & Pawar, D. D. (2010). Application of Artificial Neural Network for stock market predictions: A review of literature. *International Journal of Machine Intelligence*, *2*(2), 14-17.

[15] Nayak, A., Pai, M. M., & Pai, R. M. (2016). Prediction models for Indian stock market. *Procedia Computer Science*, *89*, 441-449.

[16] Soni, S. (2011). Applications of ANNs in stock market prediction: a survey. *International Journal of Computer Science & Engineering Technology*, *2*(3), 71-83.

[17] Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. *Electronics*, *10*(21), 2717.

[18] Bharathi, S., & Geetha, A. (2017). Sentiment analysis for effective stock market prediction. *International Journal of Intelligent Engineering and Systems*, *10*(3), 146-154.

[19] Yoo, P. D., Kim, M. H., & Jan, T. (2005, November). Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)* (Vol. 2, pp. 835-841). IEEE.

[20] Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015, June). Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*.

[21] Mittal, A., & Goel, A. (2012). Stock prediction using twitter sentiment analysis. *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)*, *15*, 2352.