5th International Conference on Electronics and Sustainable Communication Systems (ICESC 2024)
IEEE XPlore Part Number: CFP24V66-ART; ISBN: 979-8-3503-7994-5

A Comparison of Stock Price Prediction using Machine Learning Techniques

Mohammed Ali Shaik
School of computer science & Artificial Intelligence
SR university
Warangal, Telangana State, India
niharali@gmail.com

Nallani Lakshmi Sri
School of computer science & Artificial Intelligence
SR university
Warangal, Telangana State, India
nallanilakshmisri1208@gmail.com

2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC) | 979-8-3503-7994-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICESC60852.2024.10689767

***Abstract*- Technological advancements have revolutionized stock market forecasting, with machine learning methods proving more accurate than traditional statistical approaches. Comparing various models, this study found that algorithms like Random Forest results in high performance in predicting Tesla's stock closing prices. These methods mitigate risks associated with investing while maximizing dividends, which are crucial for effective resource allocation and macroeconomic expansion in a country's financial market.**

Keywords— Stock Markets Forecasting, Machine Learning, Traditional statistical Approaches, closing prices, Investing.

I. INTRODUCTION

Stock markets, comprising stockbrokers and dealers, facilitate the trading of stock shares, enhancing liquidity and investor appeal. However, stock market investment carries inherent risks due to the potential for rapid value fluctuations, making stock prices forecasting a challenging task. Statistical methodologies, particularly linear, often fail to account for sudden spikes or drops in stock prices, proving insufficiently accurate for the volatile and unpredictable nature of stock data. Linear Models consist of Regression, SVM, Random Forest, LSTM, and ARIMA models.

II. LITERATURE REVIEW

As per reference [1] the article attempts to use machine learning like k-nearest Neighbors algorithm and time series to analyze Tesla's stock price. Kayako. The research work aims at undertaking the study of such machine learning techniques like Linear Regression [11], Polynomial Regression [12], XGBoost [13], ARIMA [14], Prophet [15], and LSTM [16] to understand and make predictions about Tesla's stock value, where LSTM is least error-prone [17]. Evaluation ways such as Mean square error [18], Root Mean Square Error [19], Mean Absolute Error [20], and Accuracy of trend prediction are introduced to assess the prediction accuracy of each model. Linear regression, polynomial regression [21], and XGBoost [22] are among the models used in the analysis. Tesla stock data from 2016 to 2021 sourced from Kaggle is utilized in the study.

In [2], a comparison study was undertaken on classic stats methods and AI sorcery for foretelling stock costs, uncovering that neural web model are the peak precision. The exponential Smoothing method uses a smoothing constant to maximize prediction accuracy from the last prediction. The Weighted Moving Average (WMA) [23] method uses weights with previous values to predict future values, with different WMA units showing varying prediction errors.

As per reference [3], researchers analyzed Tesla stock prices using machine learning algorithms. Python was chosen as the language for the project, utilizing linear regression for stock price forecasting. The document includes figures illustrating data processing, linear regression, and data splitting.

In [4] the performance of four models in predicting Tesla stock prices is compared. Linear regression stands out with the highest R-squared value of 0.85 for Tesla stock from 2020 to 2022 Super vector regression (SVR) [24] is a part of Support vector Machine (SVM) and is popular for its wide range of applications in regression estimation

In [4] the application of machine learning models, including Kalman filtering and LSTM architectures, for stock price prediction over a 10-year range are discussed. It explores the performance of different algorithms in predicting future stock prices, with a focus on both low-volatility (e.g., Microsoft) and high-volatility (e.g., Tesla) stocks [25]. The study uses historical stock prices of individual technology sector stocks (e.g., MSFT, AAPL, TSLA) and market indices (e.g., NASDAQ, S&P 500) to train and test the machine learning models [26].

As per reference [5] the market prediction by Machine Learning algorithms is crucial for investors to estimate future stock values accurately. Factors like demand and supply, corporate results, and popularity influence stock values, making accurate predictions essential for decision making. Linear regression is commonly used in stock market forecasting, while deep learning-based non-linear regression models outperform linear regression in capturing stock market data changes.

As per reference [6] the variants like Vanilla, Stacked, Bidirectional, and CNN LSTM are compared in predicting Tesla's stock price Stacked LSTM, with multiple hidden layers, shows promising results in stock prediction. LSTM, a type of RNN, retains prior-state information and is crucial for learning long-term dependencies in stock market forecasting

Reference [7] focuses on predicting Tesla's stock prices using deep learning models like MLP and LSTM based on data from investing.com and Yahoo Finance. The models are evaluated using metrics like MAE, MSE, and RMSE to assess the accuracy of the predictions. The study aims to forecast future Tesla stock prices by analyzing historical data from 2015 to 2021, emphasizing the importance of technical indicators and neural networks for accurate predictions.

In [8] researchers have analyzed Tesla's stock investment potential based on historical data, social media sentiment, and investor behavior. Tesla is seen positively on social media and in old-fashioned news, route to political backing and absence of challengers. It recommends using text analytics and AI for a deeper look at social media data about Tesla's stocks. According to the Twitter study, there aren't many negative tweets about Tesla, but a lot of people seem thrilled and proud to invest in the company. [9] focuses on the prediction of Tesla Inc. sock value and analyzed the correlation between the tweets posted by Elon Musk and Tesla's stock price. Machine Learning methods, especially sentiment analysis on social networks, are implemented to build predication models.

As per reference [10], twitter data analysis was used to predict intraday stock return for Tesla, Inc. by utilizing sentiment extraction methods like VADER and SVM. Data cleaning procedures were implemented to adjust for time zone and calendar differences between Twitter and Tesla stock datasets.

III. Methodology

In an experiment utilizing Tesla stock data from June 29, 2010, to July 12, 2022, data was collected and processed from Kaggle, involving the removal of null values and normalization for model compatibility. MATLAB was employed to implement and compare various models, with the performance comparison being a key objective. The dataset includes 6 columns and 3031 rows, with 2121 data points allocated for training and 909 for testing. The training data spans from June 29, 2010, to September 7, 2018, while the test data extends from July 10, 2010, to July 12, 2022. Data preparation involved removing extraneous information and scaling variable ranges to ensure a fair comparison across methodologies. The experiment utilized data from Kaggle covering June 29, 2010, to July 12, 2022, focusing on Tesla stock information. Comprehensive pre-processing steps were undertaken to handle null values and normalize the data. Various machine learning and statistical models, including Linear Regression, Support Vector Machine (SVM), Random Forest, Long Short-Term Memory (LSTM), and AutoRegressive Integrated Moving Average (ARIMA), were applied to predict stock rates. The dataset was divided into 2121 data points for training and 909 for testing, ensuring balanced evaluation of different model performances. Evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were used to assess model accuracy. The purpose of the experiment is to provide updated recommendations for further research and project expansion in stock price prediction, based on insights gained from analyzing Tesla stock data.

IV. Experimental Results

In this Section, we explain the details of experimental and dataset results, then the results of our experiments.

A. Experimental setup:

For this project, the dataset is obtained from Kaggle. This dataset contains 2814 rows of data and 7 columns (features) that we could focus on to build our prediction model i.e., I have used 7 attributes to predict the rise in stock of Tesla.

TABLE I. Comparision Table of Reseafch

Date	Open	High	Low	Close	Adj Close	Volume
########	3.8	5	3.508	4.778	4.778	9383150
########	5.158	6.084	4.66	4.766	4.766	8593550
########	5	5.184	4.054	4.392	4.392	4109400
########	4.6	4.62	3.742	3.84	3.84	2569900
########	4	4	3.166	3.222	3.222	3433450
########	3.28	3.326	2.996	3.16	3.16	3460850
########	3.228	3.504	3.114	3.492	3.492	3855700
########	3.516	3.58	3.31	3.48	3.48	2025300
########	3.59	3.614	3.4	3.41	3.41	1101250
########	3.478	3.728	3.38	3.628	3.628	1340050
########	3.588	4.03	3.552	3.968	3.968	2097600
########	3.988	4.3	3.8	3.978	3.978	1869900
########	4.14	4.26	4.01	4.128	4.128	1310650
########	4.274	4.45	4.184	4.382	4.382	1243250
########	4.37	4.37	4.01	4.06	4.06	912650
########	4.132	4.18	3.9	4.044	4.044	626250
########	4.1	4.25	4.074	4.2	4.2	478900
########	4.238	4.312	4.212	4.258	4.258	326800
########	4.3	4.3	4.06	4.19	4.19	461100

Table 1 comprises of following attributes:

- Date: This column features the date of observation for each data point. They represented the day when the stock market started and as well as the time records were recorded.
- Open: The "Open" price is the "TSLA" stock opening price on a trading day. That is the amount of money for which a security was offered on the market recently.
- High: The "High", which stands for the highest price that TSLA shares managed to reach in the current trading day.
- Low: The "low" denominated 142.59 is the lowest price where TSLA stock moved during trading today.
- Close: The "Close" price is noted as TSLA's closing stock price at the end of the trading day. It is the latest price at which an order occurred, and it represents the last transaction that took place before the market closed.
- Adj Close: The "Adjusted Close" price is the same as the close or closing price but will be adjusted when there is a change between the stock's value and any certain factors such as dividends and stock splits. It is

one of the most effective historical performances analysis in today's times.

- Volume: The "Volume" column indicates with which the given day's activity on the TSLA stock was measured. It denotes the aggregate percentage of shares open during the day's trading session

These parameters are essential for conducting technical analysis, building predictive models, and making investment decisions. For stock prediction, analysts and data scientists often use historical data like this to develop models that attempt to forecast future stock prices or to identify patterns and trends in the market.

B. Data Cleaning:

Data cleaning consists of removing the incorrect, incomplete, inaccurate, irrelevant, or missing data from the data then modifying or replacing them according to the needs of the data! Data cleaning being one of the first elements of data science, is considered fundamental by the data science experts. Data is the most valuable thing for data analysis and machine learning. In computing and Business, Data is the key aspect of all areas. As we deal with today's world of data, we are likely to encounter missing, inaccurate, or erroneous values. If da information is corrupted, the process may be disrupted or the results become unreliable. My dataset has no string which means that I do not need to clean the data.

C. Data Visualization:

Data visualization helps to depict data clearly in a pictorial or graphical representation. Data visualization tools helps to understand various themes like showing trends, the patterns and detection of outliers in data. The data visualization tools and technologies will be playing key role when it comes to analyzing big amounts of data and making evidence-driven decisions. Visualization usage as a tool to cognize fought throughout centuries. More often than not, the type of data visualization used are charts, tables, graphs, maps, dashboards, and so on. However, the figures of our data set is shown in plots from each feature to price as shown in figures 1 to 5.

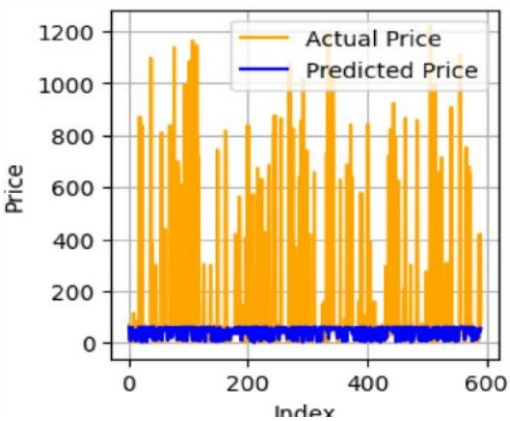

Fig 1. Open price vs high price through (Linear Regression)

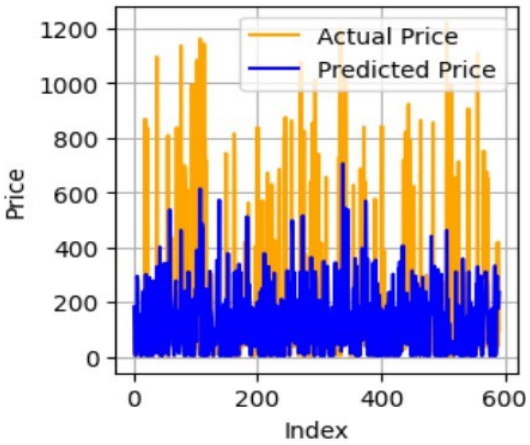

Fig 2. Open price vs low price (RIDGE REGRESSION)

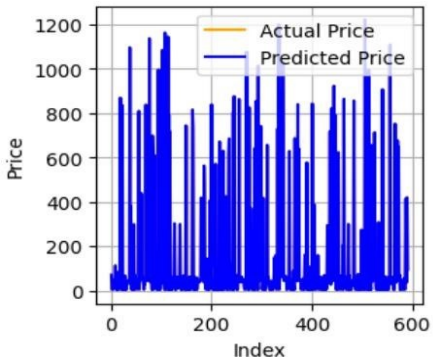

Fig 3. Open price vs low price (SVM)

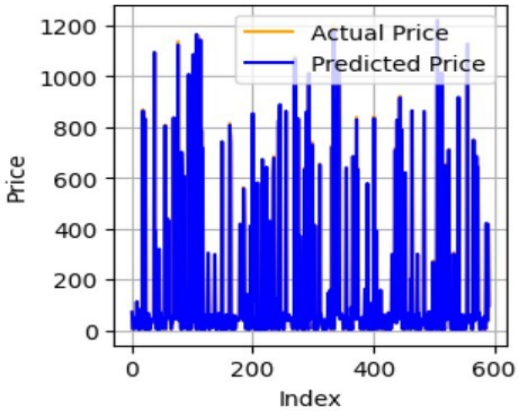

Fig 4. Open price vs Adj Close price (KNN)

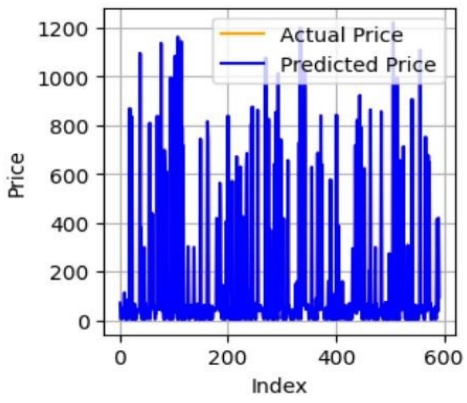

Fig 5. Open price vs volume (Lasso REGRESSION)

V. CONCLUSION

In conclusion, the integration of machine learning methods in stock market forecasting represents a notable increase in accuracy and reliability compared to traditional statistical approaches. This study highlights the superior performance of algorithms such as Random Forest in predicting Tesla's stock closing prices. By enhancing the precision of predictions, these advanced techniques help to mitigate investment risks and maximize dividends. Consequently, this technological progression plays a vital role in efficient resource allocation and contributes positively to macroeconomic growth within a country's financial market.

REFERENCES

[1] M. Ananthi and K. Vijayakumar, "Stock market analysis using candlestick regression and market trend prediction (CKRM)", Journal of Ambient Intelligent and Humanized Computing, vol. 12, pp. 4819-4826, 2021.

[2] Mohammed Ali Shaik and Dhanraj Verma, (2022), "Prediction of Heart Disease using Swarm Intelligence based Machine Learning Algorithms", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020025-1–020025-9; https://doi.org/10.1063/5.0081719, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020025-1 to 020025-9

[3] J Zhao, N Sun and W Cheng, Logistics forum based prediction on stock index using intelligent data analysis and processing of online web posts, 2019, ISSN 1868-5145.

[4] M. A. Shaik, M. Y. Sree, S. S. Vyshnavi, T. Ganesh, D. Sushmitha and N. Shreya, “Fake News Detection using NLP”, 2023 International Conference on Inovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 399-405, doi: 10.1109/ICIDCA56705.2023.10100305.

[5] Z. Wang, S. Ho and Z. Lin, "Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment", 2018 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 1375-1380, 2018.

[6] M. A. Shaik, R. Sreeja, S. Zainab, P. S. Sowmya, T. Akshay and S. Sindhu, “Improving Accuracy of Heart Disease Prediction through Machine Learning Algorithms”, 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 41-46, doi: 10.1109/ICIDCA56705.2023.10100244.

[7] K Vijayakumar and C Arun, "Continuous security assessment of cloud based applications using distributed hashing algorithm in SDLC", Cluster Computing, 2019.

[8] Mohammed Ali Shaik, M. Varshith, S. SriVyshnavi, N. Sanjana and R. Sujith, “Laptop Price Prediction using Machine Learning Algorithms”, 2022 International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS), Nagpur, India, 2022, pp. 226-231, doi: 10.1109/ICETEMS56252.2022.10093357.

[9] Z Peng et al., "Stock analysis and prediction using big data analytics", 2019 international conference on intelligent transportation big data & smart City (ICITBS), pp. 309-312, 2019.

[10] M. A. Shaik, S. k. Koppula, M. Rafiuddin and B. S. Preethi, (2022), "COVID-19 Detector Using Deep Learning”, International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 443-449, doi: 10.1109/ICAAIC53929.2022.9792694.

[11] K Vijayakumar and C Arun, "Automated risk identification using NLP in cloud based development environments", Journal of Ambient Intelligentand Humanized Computing, 2017.

[12] Mohammed Ali Shaik, Geetha Manoharan, B Prashanth, NuneAkhil, Anumandla Akash and Thudi Raja Shekhar Reddy, (2022), "Prediction of Crop Yield using Machine Learning", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020072-1–020072-8; https://doi.org/10.1063/5.0081726, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020072-1 to 020072-8

[13] Farhan Ahmed et al., "Assessing the Impact of COVID-19 Pandemic on the Stock and Commodity Markets Performance and Sustainability: A Comparative Analysis of South Asian Countries", Sustainability, vol. 13.10, pp. 5669, 2021.

[14] Mohammed Ali Shaik and Dhanraj Verma, (2020), Deep learning time series to forecast COVID-19 active cases in INDIA: A comparative study, 2020 IOP Conf. Ser.:Mater.Sci.Eng. 981 022041, doi.org/10.1088/1757-899X/981/2/022041

[15] Ishita Parmar et al., "Stock market prediction using machine learning", 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018.

[16] Mohammed Ali Shaik, “A Survey on Text Classification methods through Machine Learning Methods”, International Journal of Control and Automation (IJCA), ISSN:2005-4297,Volume12,Issue-6 (2019), Pp.390-396.

[17] Praveen, P.; Shaik, M.A.; Kumar, T.S.; Choudhury, T. Smart Farming: Securing Farmers Using Block Chain Technology and IOT. In Blockchain Applications in IoT Ecosystem; Springer: Cham, Switzerland, 2021; pp. 225–238.

[18] V. Reddy and Kranthi Sai, "Stock market prediction using machine learning", International Research Journal of Engineering and Technology (IRJET), vol. 5.10, pp. 1033-1035, 2018.

[19] Mohammed Ali Shaik, “Time Series Forecasting using Vector quantization”, International Journal of Advanced Science and Technology (IJAST), ISSN:2005-4238,Volume-29,Issue-4 (2020), Pp.169-175.

[20] Aryendra Singh, Priyanshi Gupta and Narina Thakur, "An Empirical Research and Comprehensive Analysis of Stock Market Prediction using Machine Learning and Deep Learning techniques", IOP Conference Series: Materials Science and Engineering, vol. 1022, no. 1, 2021.

[21] Mohammed Ali Shaik, Dhanraj Verma, (2021), Agent-MB-DivClues: Multi Agent Mean based Divisive Clustering, Ilkogretim Online - Elementary Education, Vol 20(5), pp. 5597-5603, doi:10.17051/ilkonline.2021.05.629

[22] Mojtaba Nabipour et al., "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis", IEEE Access, vol. 8, pp. 150199-150212, 2020.

[23] Mohammed Ali Shaik and Dhanraj Verma, (2020), Enhanced ANN training model to smooth and time series forecast, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022038, doi.org/10.1088/1757-899X/981/2/022038

[24] Wasiat Khan et al., "Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis", Soft Computing, vol. 24.15, pp. 11019-11043, 2020.

[25] Mohammed Ali Shaik, Dhanraj Verma, P Praveen, K Ranganath and Bonthala Prabhanjan Yadav, (2020), RNN based prediction of spatiotemporal data mining, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022027, doi.org/10.1088/1757-899X/981/2/022027

[26] Kunal Pahwa and Neha Agarwal, "Stock market analysis using supervised machine learning", 2019 International Conference on Machine Learning Big Data Cloud and Parallel Computing (COMITCon), 2019.