Stock Price Prediction Using Machine Learning

Ch.Subrahmanya Sresti
Computer Science and Engineering (AI)
Amrita School of Engineering
bl.en.u4aie22179@bl.students.amrita.edu

S.Hemanth Sai
Computer Science and Engineering (AI)
Amrita School of Engineering
bl.en.u4aie22171@bl.students.amrita.edu

V.Bhavaharnath Sharma
Computer Science and Engineering (AI)
Amrita School of Engineering
bl.en.u4aie22162@bl.students.amrita.edu

***Abstract*—This project uses machine learning techniques like Simple Linear regression, Support Vector regression, Decision tree regression , Random forest regression, to predict the closing prices and volume of the stocks based on recorded open prices and other features of the dataset. By utilizing this data as reference, the model give the predicted estimations of the future stock prices , helping the users in making more knowledgeable decisions. The efficiency of the model is evaluated through visualizing the comparision of predicted outcomes of data and actual values, there by measuring the performance. This project provides greater understanding of using machine learning techniques for stock price prediction, leading to a better understanding of predictive approachs in finance sector.**

***Index Terms*—Stock Price Prediction, Machine Learning, Regression Models, Support Vector regression, Decision tree regression , Random forest regression Machine Learning Techniques .**

I. INTRODUCTION

Predicting stock prices has long been a pursuit of investors and researchers alike, driven by the tantalizing prospect of financial gain amidst the inherent risks of the stock market. This project delves into the realm of stock price prediction leveraging machine learning techniques, specifically Simple Linear Regression, Support Vector Regression, Decision Tree Regression, and Random Forest Regression. By analyzing historical data of stock open prices, the project aims to forecast future stock prices, thereby offering insights to assist investors in making informed decisions. The stock market is notoriously volatile, influenced by a myriad of factors ranging from economic indicators to geopolitical events and investor sentiment.

Traditional methods of analysis often struggle to capture the complexity and dynamism of market trends. However, machine learning algorithms present a promising avenue for predictive modeling due to their ability to discern patterns and relationships within vast datasets. The primary objective of this project is to develop and evaluate the efficacy of machine learning models in predicting stock prices based on open price data. By training these models on historical stock price data and testing them on unseen data, we seek to ascertain their accuracy and reliability in forecasting future price movements. Through this analysis, investors can gain valuable insights into potential market trends, enabling them to optimize their investment strategies and mitigate risks.

Central to this project is the exploration of various machine learning techniques and their applicability to stock price prediction. Each technique offers unique advantages and limitations, necessitating a comprehensive evaluation to determine the most effective approach. By comparing the performance of different algorithms, we aim to identify the optimal model for predicting stock prices with a high degree of accuracy. Moreover, this project emphasizes the importance of visualization in understanding and evaluating the predictive models. Visual representations of actual versus predicted stock price movements provide intuitive insights into the effectiveness of the models, facilitating decision-making for investors. Through visualizations, users can discern patterns, trends, and potential anomalies in the predicted data, thereby enhancing their confidence in the predictive capabilities of the models.

II. LITERATURE SURVEY

Sharma A. et.al. explain's the deep research using machine learning and deep learning.Researches have searched Artificial Neural Networks(ANN) and Convolutional Neural Networks(CNN) with ANN facing difficulties of overfitting[1]. The CNN method gives output of grayscale 2-D histograms ,giving accuracy by reducing training time and data needs. But both methods exhibit accuracy in stock market prediction.

R. Seethalakshmi et. al. explain about research paper's review of literature focus on limitations of using machine learning and deep learning algorithms[2]. This paper discusses how to face challenges like handling errors or complexities in stock data.It also handles missed data handling including optimization techniques.The review gives the MLS LSTM model within the research.

Basak S. et. al. researched how to explore into the stock price prediction using machine learning and deep learning algorithms checking the challange of nonlinearity[3]. It shows the up-down minor stock price fluctuations and it proposes the NPMM-N period Min-Max labelling method.It develops atrading system using XGBoost and finds its performance against the other labelling methods.It describes the importance of instance selection and suggest future research directions.

Ampomah EK et. al. explains the 32 research works from the year 2011 tp 2022 based on the feature selection and the extraction techniques in stock market prediction, finding correlation criteria RF,PCA and AE as the key methods[4].RF and SVM are the primary Machine learning approches.The study explains the significance of feature analysis in improving accuracy prediction and utilize reduction techniques to

enhance the performance of learning models for stock market predictions.

Abe M et. al. reviews the literature addresses the critical issue of deadlock avoidance in cloud computing, proposing an algorithm that enhances resource allocation efficiency by considering execution time attributes of processe[5]s. The existing algorithm focuses on load balancing, while the improved algorithm introduces a Temporary Queue and prioritizes shorter execution times for faster resource availability. The study emphasizes the importance of response time in cloud computing and suggests future work involving additional process attributes like request importance and exploring parallel processing capabilities when resources permit.

Altig et. al. conducts a reviews on Extreme value theory(EVT) in predicting investment risk of covid-19 pandemic.It reveals research in the domain mainly in multivariate cases and gives the complexities of such analysis[6] .Further research will be done to develop conceptual models sensitive to fluctuations.We use EVT and machine learning algorithms to get good accuracy.

M. Nabipour et. al. explains the analysis of stock chart through Machine learning and Deep learning model have been discussed which mainly focuses on finding the hidden patterns[7]. The speciifying parameters affect stock exchange movements and effect on the stock value volataility.The study tells that a vital role played by the means of the data mining being used in selection of stable stocks for the investors.It also proposes the scope of further research and improvements.

G.A. A.Jabbar Alkubaisi et. al. explores the impact of different factors in market trends.It explains the application of statistical techniques based on analysis of customer reviews[8].The paper includes the challenges such as lower classification accuracy and explains the exsisting problems in stock market prediction.

Alexander et. al. introduces a good approach by combining ML algorithms for stock prediction and mean VaR model for portofolio selection.Random Forest,XGBoost,AdaBoost ,SVR,KNN and ANN are used to predict stock values. It uses monthly datasets from the Bombay Stock Exchange(BSE)[9], Tokyo Stock Exchange(TSE) to know the techeniques's effectiveness. By using these regression models it produces performance compared to other models,offering important predictions for portopfolio strategies.

Acemoglu et. al. focuses on the uses of natural language models such as ChatGPT in finding the returns of stock market with the help of news headlines.The findings focuses on the positive connections between ChatGPT scores and the following daily stock returns which exceeds the performance of the conventional sentiment analysis[10]. The experiment shows the better capability of complex language models like ChatGPT-4 predicting compared to basic models like GPT-1,GPT-2,and BERT.

III. METHODOLOGY

Data Collection and Preprocessing:

Prepare the dataset by doing necessary operations like data cleaning,encoding variables and feature scaling. Preprocessing involved meticoulous data cleaning like normalizing the columns and handling missing values.

Model selection:

There are many types of classifiers to compare with perceptron and MLP networks. So we can choose like Support Vector Machine(SVM),Decision Tree,Random Forest,AdBoost,XGBoost and Naive-Bayes.

Hyper parameter and Tuning:

SO after doing each classifier,define the hyperparameter grid using RandomizedSearchCV().We will get a grid which include a range of values for hyperparameters which impact our model's performance. We can hypertune parameters like hidden layers,number of neurons per layer.

Cross Validation:

Consider K-fold cross validation like k=5 or k=10 to test the performance of the model.RandomizedSearchCV() will split the data into 2 parts like training and testing for every fold,fitting into the training set and select the best hyperparameterson validation performance.

Evaluation of model:

We can evaluate the model by considering the evaluation metrics.We can consider like precision,recall,F1-score, confusion matrix and area under the ROC curve.

Tabulation of results:

Tabulate the results for every classifier using the above metrics for classification.So that we can compare the model's performance and it will help to identify the best-performing model.

IV. RESULT

```
Best hyperparameters for Perceptron: {'max_iter': 1000, 'alpha': 0.1}
Best hyperparameters for MLP: {'hidden_layer_sizes': (50,), 'alpha': 0.0001}
Perceptron Metrics:
Accuracy: 0.509915014164306
Precision: 0.509915014164306
Recall: 1.0
F1 Score: 0.6754221388367729

MLP Metrics:
Accuracy: 0.509915014164306
Precision: 0.509915014164306
Recall: 1.0
F1 Score: 0.6754221388367729
```

Fig. 1. Evaluation Metrics

	Classifier	Accuracy	Precision	Recall	F1 Score
1	SVM	0.484419	0.491228	0.311111	0.380952
2	Decision Tree	0.543909	0.550802	0.572222	0.561308
3	Random Forest	0.538244	0.544974	0.572222	0.558266
4	AdaBoost	0.509915	0.532710	0.316667	0.39723
5	Naive-Bayes	0.464589	0.467626	0.361111	0.407524

TABLE I
EVALUATION METRICS FOR DIFFERENT CLASSIFIERS

CONCLUSION

Based on the results we obtained on various metrics, regreession is best method for stock market prediction. The higher values of precison , recall and F1-score suggests that the classifier is good in classifying the different classes.The models seems to be generalized to unseen data.Therefore model can classify and predict the values in complex situations to capture the patterns. This project approaches at different ways to predict trends using machine learning.We want to understand how to use techniques to predict risks and make decisions on investment.Ultimately we will learn how to manage risks in difficult situations.

REFERENCES

[1] *Sharma A., Bhuriya, D., Singh, U.: Survey of stock market prediction using machine learning approach. In: (ICECA). (2017). .*

[2] *] R. Seethalakshmi, Analysis of stock market predictor variables using linear regression, Int. J. Pure Appl. Math. 119 (15) (2018) 369–378.*

[3] *Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. North American Journal of Economics and Finance, 47, 552–567*

[4] *Ampomah EK, Qin Z, Nyame G (2020) Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement. Information 11:332*

[5] *Abe M, Nakagawa K Cross-sectional stock price prediction using deep learning for actual investment management. Proceed 2020 Asia Service Sci Software Eng Conf:9–15*

[6] *Altig, Dave, Scott Baker, Jose Maria Barrero, Nicholas Bloom, Philip Bunn, Scarlet Chen, Steven J. Davis, Julia Leather, Brent Meyer, Emil Mihaylov, and et al. 2020. Economic Uncertainty before and during the COVID-19 Pandemic. Journal of Public Economics 191: 104274 .*

[7] *M. Nabipour, P. Nayyeri, H. Jabani, S. S, and A. Mosavi, "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, doi: 10.1109/access.2020.3015966.*

[8] *G.A. A.Jabbar Alkubaisi, S.S. Kamaruddin, H. Husni, Jabbar Alkubaisi, Siti Sakira Kamaruddin, Husniza Husni, Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers, Computer and Information Science 11 (1) (2018) 52*

[9] *Alexander, G.J., Baptista, A.M., 2002. Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis. J. Econom. Dynam. Control 26 (7–8), 1159–1193.*

[10] *Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022. "Artificial Intelligence and Jobs: Evidence from Online Vacancies." Journal of Labor Economics 40, no. S1 (April): S293–S340.*