Stock Price Prediction Using Machine Learning

# Ch.Subrahmanya Sresti

*Computer Science and Engineering (AI) Amrita School of Engineering*

# S.Hemanth Sai

*Computer Science and Engineering (AI) Amrita School of Engineering*

# Bhavaharnath Sharma

*Computer Science and Engineering (AI) Amrita School of Engineering*

[bl.en.u4aie22179@bl.students.amrita.edu](mailto:bl.en.u4aie22179@bl.students.amrita.edu) [bl.en.u4aie22171@bl.students.amrita.edu](mailto:bl.en.u4aie22171@bl.students.amrita.edu) [bl.en.u4aie22162@bl.students.amrita.edu](mailto:bl.en.u4aie22162@bl.students.amrita.edu)

***Abstract*—This project uses the machine learning techniques like Simple Linear Regression, Support Vector Regression, Deci- sion Tree Regression, Random Forest Regression to predict the closed price of the stocks based on the open prices of stock recorded in the data set. Using that data as reference, the model estimates the future stock prices to help the user to invest the stocks efficiently. By visualization of actual and predicted process, the efficiency of model can be determined.This project give some insights on the machine learning approach to predict the stock price of the company.**

***Index Terms*—Regression,Simple Linear,Support Vector,Decision Tree,Random Forest**

* 1. INTRODUCTION

This paper explores into the stock market prediction,mainly concentrating on estimating future stocks prices for making decisions.This project aims on using historical data, to predict the variations of stock prices on daily basis. In this project we will use many algorithms like Simple Linear Regression,Support Vector Regression,Decision Tree Regression and Random Forest Regression to provide predictions. For each model there will be visualization between actual and predicted prices, therefore this model can take decisions in stock investments. By testing above machine learning regression algorithms ,we can predict risks and can able to make better decisions.Finally,we can improve how to reduce the risks and plan inverstments will be given in difficult financial situations.

* 1. LITERATURE SURVEY

1. This research paper’s 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. 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.

1. This research paper’s review of literature focus on limita- tions of using machine learning and deep learning algorithms. 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.
2. This paper explores into the stock price prediction using machine learning and deep learning algorithms checking the challange of nonlinearity. 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.

1. This paper explores 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.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.
2. This research paper’s review of literature addresses the critical issue of deadlock avoidance in cloud computing, proposing an algorithm that enhances resource allocation effi- ciency by considering execution time attributes of processes. The existing algorithm focuses on load balancing, while the improved algorithm introduces a Temporary Queue and prior- itizes 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 par- allel processing capabilities when resources permit.
3. This study conducts a reviews on Extreme value the- ory(EVT) in predicting investment risk of covid-19 pan- demic.It reveals research in the domain mainly in multivariate cases and gives the complexities of such analysis.Further research will be done to develop conceptual models sensitive to fluctuations.We use EVT and machine learning algorithms to get good accuracy.
4. In this paper, analysis of stock chart through Machine learning and Deep learning model have been discussed which mainly focuses on finding the hidden patterns. 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.
5. This paper explores the impact of different factors in market trends.It explains the application of statistical tech- niques based on analysis of customer reviews.The paper in- cludes the challenges such as lower classification accuracy and explains the exsisting problems in stock market prediction.
6. This paper 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), Tokyo Stock Exchange(TSE) to know the techeniques’s ef- fectiveness. By using these regression models it produces performance compared to other models,offering important predictions for portopfolio strategies.

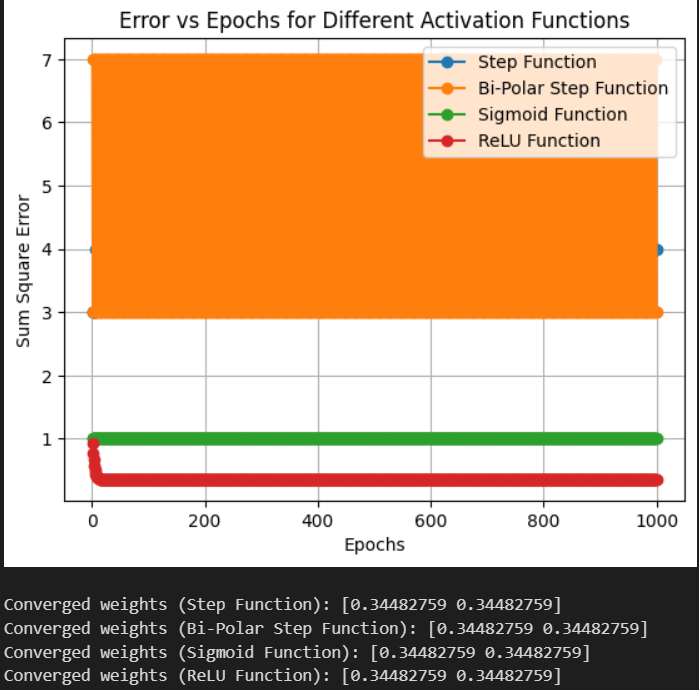
1. This paper focuses on the uses of natural language models such as ChatGPT in finding the returns of stock market with the help of news headlines.The findings focuses on the positive connections between ChatGPT scores and the follow- ing daily stock returns which exceeds the performance of the conventional sentiment analysis. 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.
   1. METHODOLOGY
2. Data Preprocessing(A1) :
   * The data which we used for this project is taken from Kaggle. The data consist of Cat- egorical Variables like Date,Symbol,Series,Prev Close,Open,High,Low,Last,Close etc.
   * Now we are only using ’Open’ as categorical vari- able and ’Close’ as target value.It handles missing values and encodes the data of target variable like ’Open’.After it calculates centroid, spreads and dis- tance between vectors.
3. Data Visualization and Analysis(A2) :
   * Using histogram it visualizes and analyze the dis- tribution of target variable(’Open’) price and calcu- lates the mean and variance of the given dataset.
4. Distance Calculation(A3) :
   * Calculation of Distance is calculated by using Minkowski distance between two selected feature variable for different values of r.
5. Model Training and Evaluation(A4 - A8) :
   * The model is divied into 2 tests. One is ’Training Tests’ and other one is ’Testing Tests’ . It trains by K- nearest neighbours(KNN) classifier and com- putes the accuracy.
   * This model also explores shows the effect or impact on changing the number of neighbours(K).
6. Performance Evaluation(A9) :
   * The code implements the Banker’s algorithm for resource allocation.
   * The code implements the Banker’s algorithm for resource allocation.
   * It checks for a safe sequence of resource allocation and prints the result.
7. Output :
   * Finally after training the model, it calculates the different types of metrics like confusion matrix,

precision, recall and F1 score to know about the models performance.

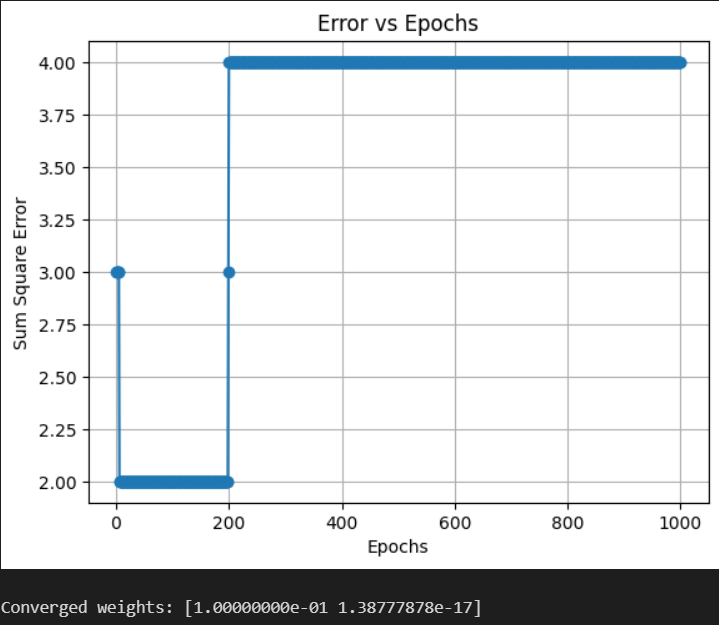
* LAB-5

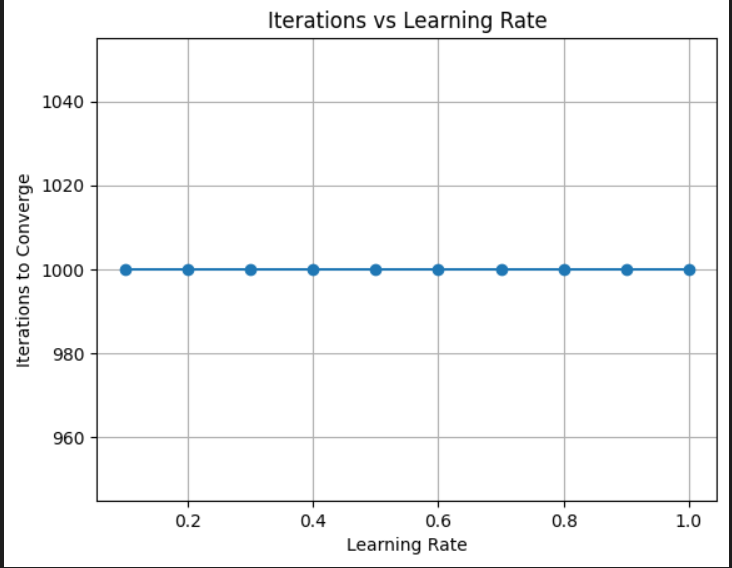
1. A1.Classification Evaluation:
   * We can calculate confusion matrix using predicted and actual values. We will get true positive, true negative, false positive, false negative from confu- sion matrix.
   * It tries to calculates the precision, recall,and F1- score.
   * So based on above metrics, we can decide the model is overfitting or underfitting.
2. Regression Evaluation and Analysis:
   * Explains the Mean Squared Error(MSE), Root Mean Squared Error(RMSE),Mean Absolute Percentage Error(MAPE).
   * It also explains the the performance of regression model by calculating above values.
   * Predict the calculated metrics and analyse the results such as prediciting abilities.
3. A3-A6. Scatter plot and KNN-Classifier:
   * This process explains how 20 data points are defined and it is vizualized using scatter plot.
   * We create test set datasets with values of X and Y.By using KNN classifier we get scatter plot.
4. A7.Hyper Parameter Tuning:
   * We use RandomSearchCV or GridSearchCV for hyperparameter tuning.
   * It analyze the results from the above process.
   * It also tells about the ’K’ value which we selected to tell about the model accuract

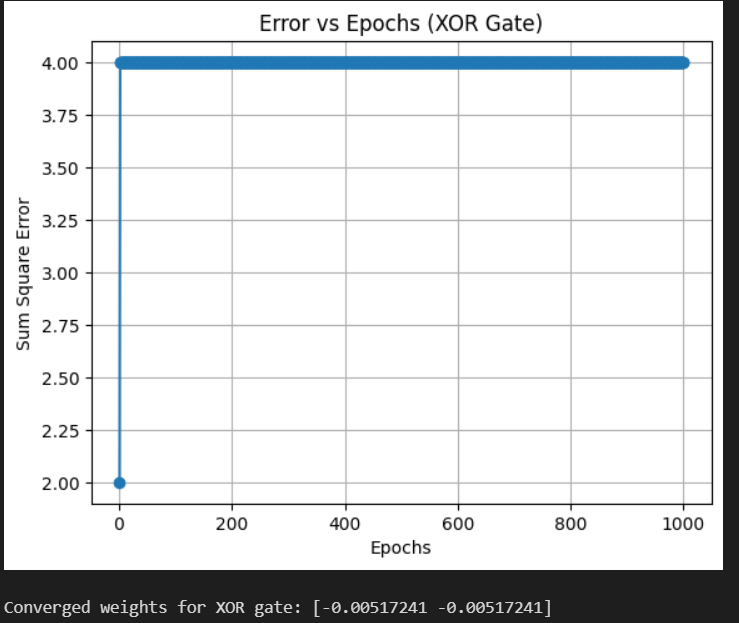
* LAB-6

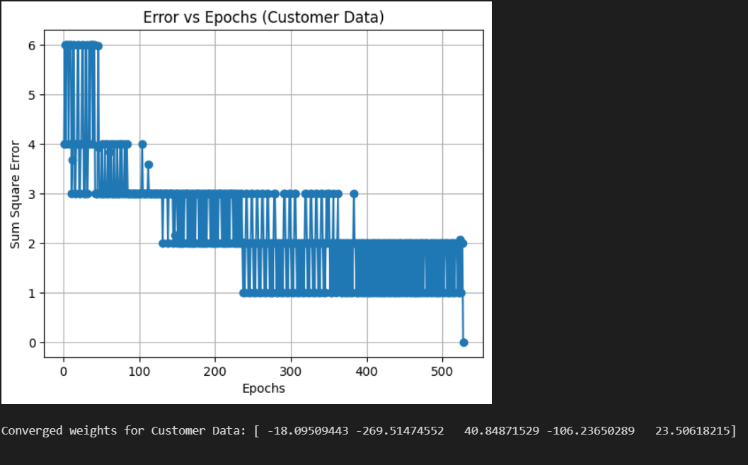
1. Perceptron Learning Algorithm(A1):
   * Perform perceptron algorithm on inputs and outputs of AND gate
   * Assign weights,parameters and learning rates of the algorithm
   * Train perceptron algorithm and print its weights and map epochs vs error.
2. Researching On Activation Functions(A2):
   * For above code, implement diff. activation function such as the sigmoid, step and rectified linear unit.
   * Modify code for above mentioned activation func- tions.
   * Use the step function, the bipolar step function, the sigmoid function, and ReLU.
3. Learning Rates(A3):
   * Initialize different learning rates.
   * Train the same perceptron with different learning rates.
   * Plot and Compare the convergence and learning rate.
4. Training Perceptron to Solve XOR Gate Problem(A4):
   * perform perceptron algorithm on inputs and outputs of XOR gate.
     + Train perceptron algorithm .
     + Plot error vs. epochs to show algorithm’s training.
5. Training the Perceptron on customer data(A5):
   * Label customer data and initialize the learning rate.
   * Implement perceptron with sigmoid function as it’s activation function.
   * Plot error vs. epochs and print convergence weights for customer-data.
6. Comparing perceptron and matrix psuedo inverse(A7):
   * Implement perceptron learning and train it and obtain the learned weights.
   * Use psuedo-inverse method to calculate weights for the AND gate.
   * Calculate the predicted weights using both methods.
   * Compare the performance of perceptron and matrix- psuedo inverse methods.
7. Implementing AND gate using Back-propogation Algo- rithm(A7):
   * We will define input and output pairs for the AND gate.
   * For AND gate: X-AND = np.array([0,0],[0,1],[1,0],[1,1]) and Y-AND = np.array([0,0,0,1])
   * Implement the back propogation algorithm by up- dating the weights and biases until the convergence is done.
   * Stop the learning process if error is less than 0.002.
8. Implementing XOR gate using Back-propogation Algo- rithm(A8):
   * We will define input and output pairs for the AND gate.
   * For AND gate: X-AND = np.array([0,0],[0,1],[1,0],[1,1]) and Y-AND = np.array([0,1,1,0])
   * Implement the back propogation algorithm by up- datinmag the weights and biases until the conver- gence is done.
   * Stop the learning process if error is less than 0.002.
9. Implementing logic gate with 2 output Nodes(A9):
   * Update the neural network
   * We should adjust the target labesls(for AND gate

,[0,1] for output 1 and [1,0] for output 2)

1. Using MLP classifier for AND and XOR gates(A10):
   * Use MLPClassifier() method from scikit-learn.
   * Prepare input-output pairs for these gates.
   * Then we will train the classifier withe parameters like learning rate,activation function.
2. Apply MLPClassifier on project dataset(A11):
   * Prepare the dataset by ensuring labels and features are formatted or not.
   * Load the dataset and preprocess it by cleaning the data.
   * Use MLPClassifier() to train the model.
   * Predict the performance of the model on the dataset.
   1. RESULT 





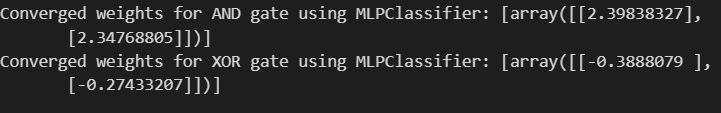












CONCLUSION

In conclusion, 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 wil learn how to manage risks in difficult situations.

REFERENCES

1. *Sharma, A., Bhuriya, D., Singh, U.: Survey of stock mar- ket 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 predic- tion using deep learning for actual investment management.*

*Proceed 2020 Asia Service Sci Software Eng Conf:9–15*

1. *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. Eco- nomic Uncertainty before and during the COVID-19 Pan- demic. Journal of Public Economics 191: 104274 .*
2. *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.*
3. 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
4. Alexander, G.J., Baptista, A.M., 2002. Economic impli- cations of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis. J. Econom. Dynam. Control 26 (7–8), 1159–1193.
5. Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022. “Artificial Intelligence and Jobs: Ev- idence from Online Vacancies.” Journal of Labor Economics 40, no. S1 (April): S293–S340.