International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Comparative Analysis of Financial Forecasting Techniques in Machine Learning for Netflix Stock Price Prediction

S. Habeeb Mohamed Sathak Amina[1], K Annsheela[2], A H Sumaiya Jailani[3]

[1,2]Assistant Professors, Department of Computer Science and Research Centre, Thassim Beevi Abdul Kader College For Women
[3]MCA Student, Thassim Beevi Abdul Kader College For Women

Abstract
In the stock market, predicting the price of stocks is still a hot research issue. One of the most challenging tasks in the field of computation is prediction of the stock market. The prediction depends on a number of variables, including physiological vs. Physical elements, rational vs. irrational investor behavior, investor attitude, market rumors, etc. All these factors work together to make stock values unpredictable and very challenging to predict accurately. Supervised Machine Learning (SML) algorithms is used to predict future value by using past stock price data. ML approaches have the ability to reveal trends and knowledge that hadn't previously noticed, and they can be used to predict that are incredibly accurate.
The primary objective of this project is to predict stock prices of Netflix for analyzing profit on day closing. Data was gathered between the years of 2002 and 2022, and separated into two parts: training set and testing set. Only the testing portion is to be used for the final forecast. Four machine learning algorithms such as linear regression, decision tree, MLP, and KNN are used to forecast the close value of the Netflix stock price. This project's output adds to the body of knowledge on predicting stock prices and offers investors some new information for accurate prediction of stock prices.

Keyword: Netflix, Stock, Prediction, Supervised Machine Learning, Linear Regression

I. INTRODUCTION

One of the most difficult and challenging is stock market analysis and forecasting.. This is due to a number of factors, including price fluctuations and a wide range of other dependent and independent variables that affect the market value of a particular stock. Any stock market expert will find it very challenging to accurately predict the market's growth and drop in considering all of these factors. The most recent developments in market research and stock market prediction, however, we have started to incorporate such methods in analyzing stock market data since the introduction of machine learning and its powerful algorithms. In conclusion, several firms use machine learning algorithms extensively for stock market forecasting. This article will take over a simple method of stock evaluation and prediction.

II. LITERATURE REVIEW

CHAU Tsun Man , et. al [4], assumed retail investors with a third-party investment mobile application to navigate through the stock market. The project provides a framework for making machine learning technology accessible to the general population. It lays the path for future development of Auto ML in the financial context as well as for the extension and testing of new models.

Author Shikha Verma et. al [3] predicts the stock market as profit in form of a Chart using Python. The attributes include in the model are Date, Open, Close, High, Low, Volume and Adj Close. The profit of various years are calculated and shown in the form of graphs. The results suggest that NFLX functions well.

Y.Puja et. al [3] assumed that, in Stock Market Prediction, the point is to foresee the future worth of the monetary loads of an organization. This research paper focuses on the application of regression and LSTM-based machine learning to forecast stock features, including open, close, low, high, and volume.

Shakir Khan et. al [2] applied and compared auto ARIMA (Auto Regressive Integrated Moving Average model). Using historical Netflix stock data spanning five years, two ARIMA models are customised to produce an accurate stock forecasting model. The ARIMA model had the most accurate findings across the three models when calculating the MAPE and holdout testing, indicating its potential for precise stock forecasting.

Israt Jahan et. al [1] assumed Stock market is generally very unpredictable in nature. There are numerous variables that could affect a stock's price, including market trend, supply and demand, the state of the world economy, public opinion, sensitive financial information, earnings announcements, historical pricing, and many more. These aspects help to explain why accurate prediction is difficult. However, by analysing massive data and creating an accurate prediction model with the aid of cutting-edge technologies like data mining and machine learning, we can reduce the chance of human mistake. In this study, a supervised machine learning system is used to forecast closing prices for certain stocks using sample data. On the time-series data of the stocks, a Recurrent Neural Network (RNN) method is applied in particular. Cross-checking is done between the anticipated closing prices and the actual closing price. Finally, it is proposed that this model be applied to forecast other financial products that are volatile.

III. METHODOLOGY

The following methods were processed to analyze the data

- Data Pre-processing
- Data Visualization
- Algorithm 1:Linear Regression
- Algorithm 2:Decision Tree
- Algorithm 3:Multi-Layer Perceptron(MLP)
- Algorithm 4: K-Nearest Neighbors(KNN)
- Deployment

A Data Pre-Processing

Data preprocessing is a necessary first step before using any machine learning technology , as the algorithms learn from the data and the learning outcome for problem solving significantly rely on the appropriate data required to solve a specific problem, Before using data, preparation is necessary. The idea of transforming unclean data into clean data is known as data preparation. Before the technique is applied, the dataset is preprocessed to check for missing values, noisy data, and other anomalies. The data format

needs to be appropriate for ML.

```
In [3]: data=pd.read_csv('NFLX.csv')

In [4]: data
```

Out[4]:

	Date	Open	High	Low	Close	Adj Close	Volume
0	2002-05-24	1.214286	1.225000	1.197143	1.210000	1.210000	11104800
1	2002-05-28	1.213571	1.232143	1.157143	1.157143	1.157143	6609400
2	2002-05-29	1.164286	1.164286	1.085714	1.103571	1.103571	6757800
3	2002-05-30	1.107857	1.107857	1.071429	1.071429	1.071429	10154200
4	2002-05-31	1.078571	1.078571	1.071429	1.076429	1.076429	8464400
...	...	...	...	...	...	...	...
5111	2022-09-13	226.500000	227.690002	217.720001	218.130005	218.130005	8000100
5112	2022-09-14	219.820007	228.000000	215.800003	224.119995	224.119995	8230300
5113	2022-09-15	230.470001	242.289993	229.100006	235.380005	235.380005	19454100
5114	2022-09-16	231.220001	241.360001	227.830002	240.130005	240.130005	18871600
5115	2022-09-19	245.050003	245.250000	235.750000	243.630005	243.630005	17895500

5116 rows × 7 columns

Fig. 1. Data Collection from 2002 to 2022

B Data Visualisation

Data visualisation is the process of displaying data using popular images like infographics, charts, and even animations. These understandable informational visual representations communicate complex data relationships and data-driven insights. It's vital to keep in mind that data visualisation is not only used by data teams but may also be used for a variety of other purposes. In addition, management uses it to communicate organisational hierarchy and structure, and data scientists and analysts use it to identify and explain patterns and trends.

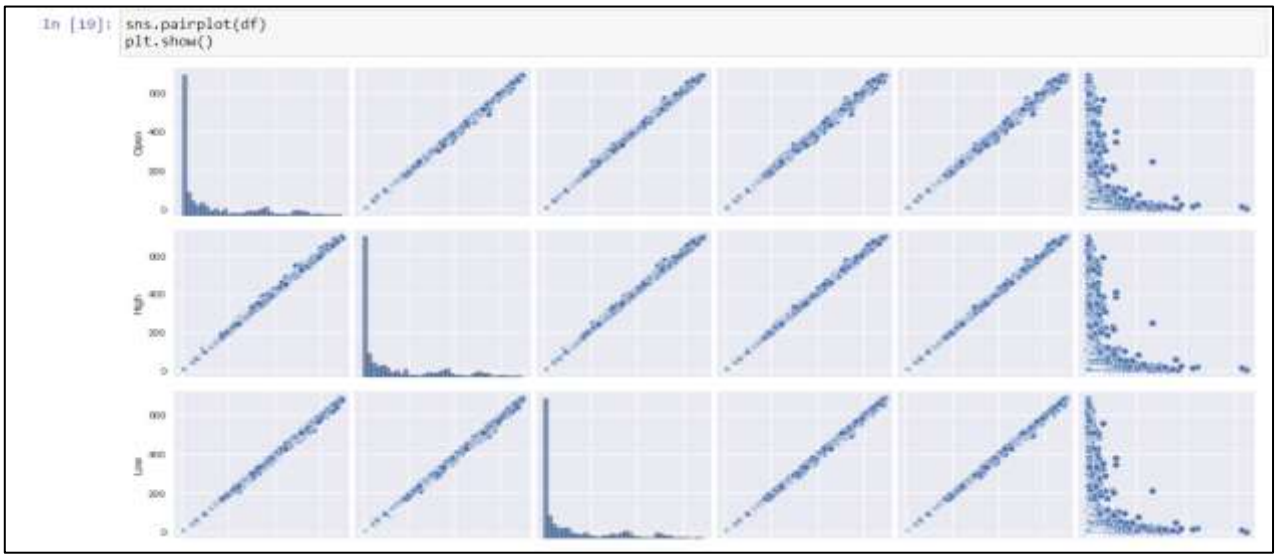

Fig. 2. Plotting of Data

Fig. 3. Visualization of Data

C Algorithm1: Linear Regression

An algorithm called linear regression uses a linear relationship between an independent variable and a dependent variable to forecast the result of upcoming events. It is a statistical technique used in forecast analysis in data science and machine learning. The predictor or explanatory variable that does not change as a result of the shift. The response or outcome variable under study or analysis is the dependent variable, and the regression model forecasts its value. Thus, linear regression is a supervised learning algorithm that predicts continuous or numeric variable by simulating a mathematical connection between the variables. This analytical approach has benefits. When there are at least two factors present in the data, as is the case in stock market forecasting, managing a portfolio, conducting scientific research, etc. The linear regression model is shown as a curved linear path.

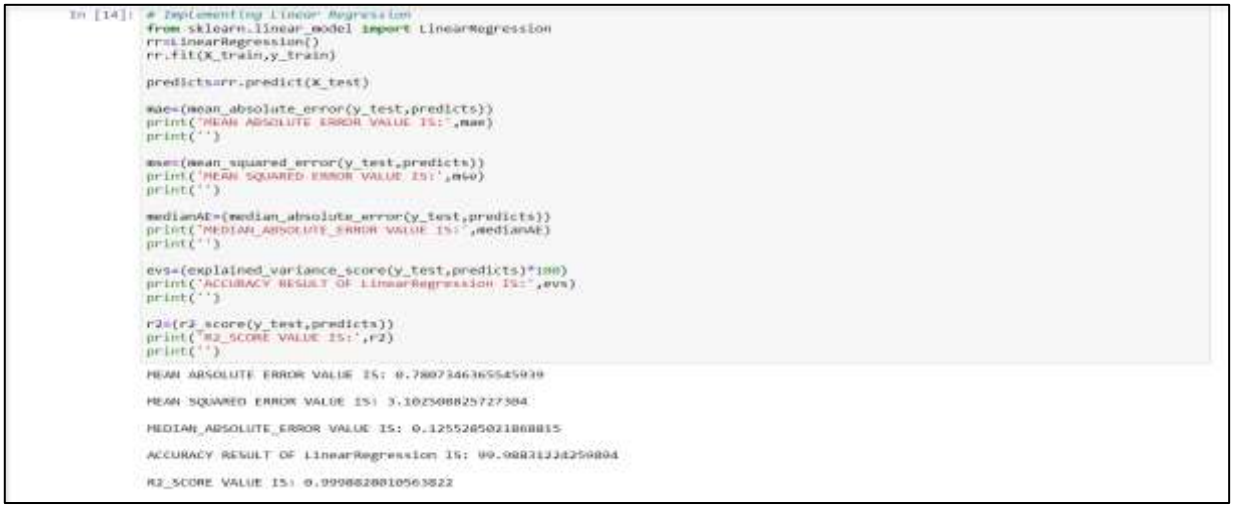

Fig. 4. Implementation of Linear Regression

D Algorithm 2: Decision Tree

A supervised learning method called the decision tree can be applied to categorization and regression issues. It is a tree-structured regressor, where interior nodes stand in for a dataset's features, branches for the decision-making process, and each leaf node for the result. The Decision Node and Leaf Node are the two elements in a decision tree. While Leaf nodes are the results of decisions and do not have any additional branches, Decision nodes are used to make decisions and have numerous branches. The given dataset's

features are used to conduct the test or make the decisions. It is a graphical representation for obtaining all feasible answers to a decision or issue based on predetermined conditions. Because it is comparable to a decision tree, It begins with the root node and extends on additional branches to create a structure akin to a tree. The CART algorithm, which means for Classification and Regression Tree algorithm, is used to construct a tree. A decision tree merely poses a query and divides the tree into subtrees according to the response (Yes/No).

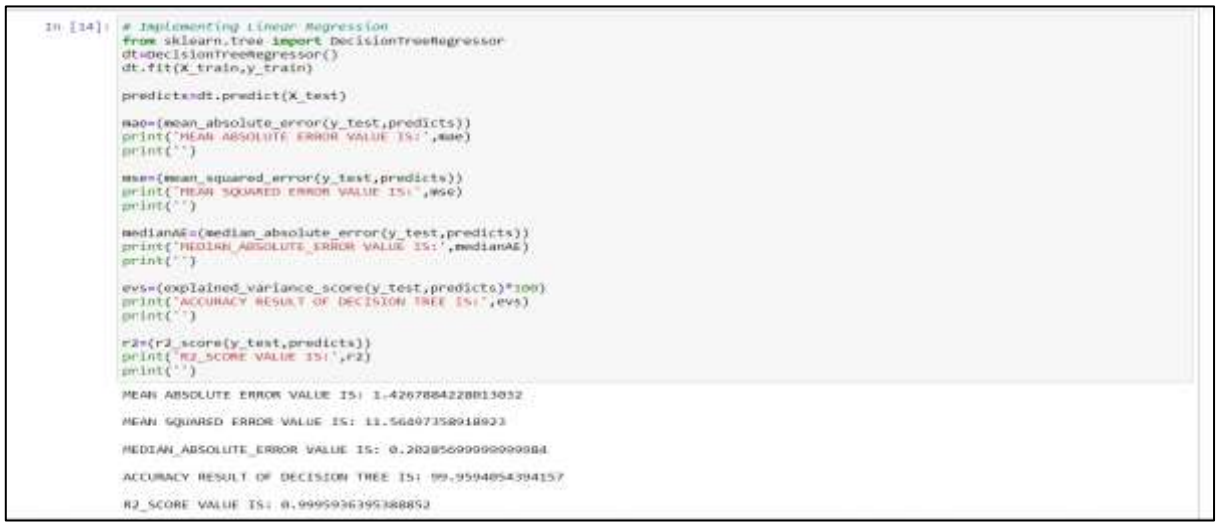

Fig. 5. Implementation of Decision Tree

E Algorithm 3:Multi-Layer Perceptron (MLP)

A supervised learning method that learns a function is the multi-layer perceptron (MLP). The number of dimensions for input and output are equal after training on a dataset. A non-linear function approximate for either classification or regression can be learned from a set of features. There may be one or more non-linear layers, known as hidden layers, between the input layer and the output layer, which distinguishes it from logistic regression. A single hidden layer MLP with scalar output is depicted in Fig. 6. Type MLP Regressor uses a multi-layer perceptron (MLP) that may be thought of as employing the identity function as an activation function and trains using backpropagation with no activation function in the output layer.

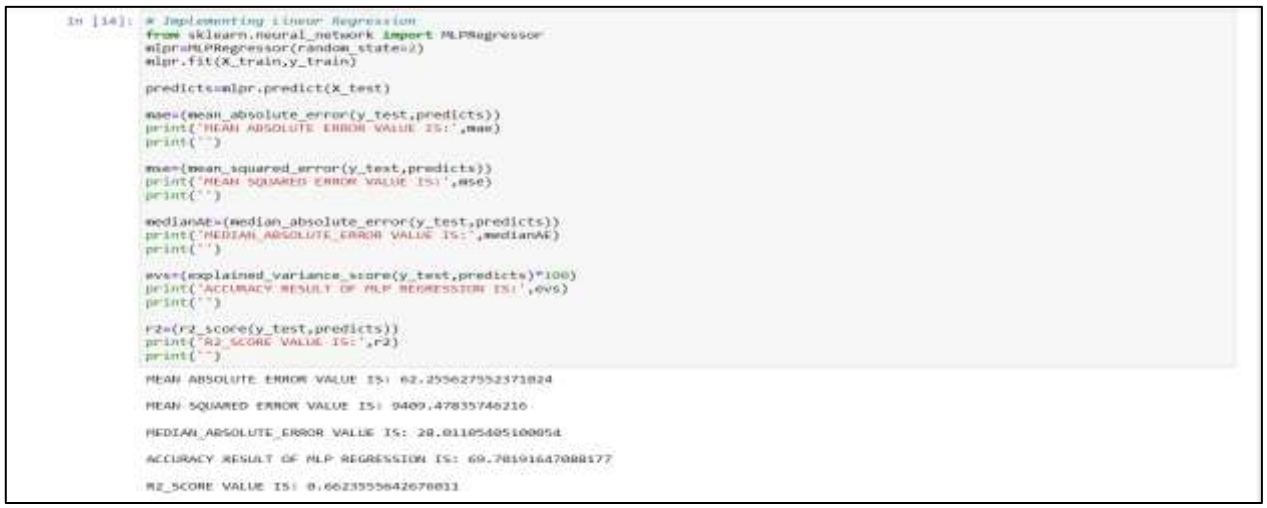

Fig. 6. Implementation of MLP

F Algorithm 4: K-Nearest Neighbors (KNN)

The k-nearest neighbors algorithm also referred to as KNN or k-NN, is a supervised learning classifier that employs proximity to make classifications or projections about the grouping of a single data point. It is employed in a variety of contexts, including handwriting recognition, image recognition, and video identification. KNN can achieve high accuracy in a broad range of prediction-type problems, and it is most helpful when labelled data is either prohibitively expensive or impossible to obtain. KNN is a non-parametric, lazy learning method. In order to forecast the classification of the new sample point, it utilizes data from multiple classes. KNN is non-parametric because the model is distributed from the data and no assumptions are made about the data being examined.

Fig. 7. Implementation of KNN

G Deployment

a) Django (Web FrameWork)

Python-based Django is a micro web platform. Its status as a micro-framework is a result of the fact that it does not need specific tools or modules. It lacks any components where pre-existing third-party libraries already provide common tasks, such as a database abstraction layer, form validation, or other components. Django, however, allows extensions that can add application features just like they were built into the core of Django.

b) Deploying the model to predict the output

In this module the trained machine learning model is converted into pickle data format file (.pkl file) which is then deployed for providing better user interface and predicting the output of Netflix Stock price prediction

IV RESULTS AND DISCUSSION

Table 1. Experimental Result

Name	Mean Absolute	Mean Squared	Median Absolute	R2 Score Value	Accuracy Result
Linear Regression	**0.7873**	**3.1025**	**0.1255**	**0.9999**	**99.9883**
Decision Tree	1.4267	11.5649	0.2028	0.9995	99.9594

MLP	62.25556	9409.47	28.0110	0.6623	69.7019
KNN	113.844	25802.012	75.5479	0.0692	6.9432

This section shows the predicted results obtained by using four algorithms. By testing these algorithms, the best accuracy is received from **Linear Regression Algorithm.** A System was constructed and data was collected from the website for analysis. This system predicted Netflix stock price from dataset with multiples attribute values. The user should enter current stock price value given in the template and the system will predict closing end value.

V. CONCLUSION

Data was gathered between 2002 and 2022 and split into Training and Test sets. ML prediction models were used to predict the close value of Netflix stock price. The linear regression model was the most reliable for predicting the long-term price of Netflix shares. Future enhancements include deploying the project in the cloud and applying sentimental analysis to predict future price movement.

REFERENCES

1. Y.Puja, Vivekparghania, Dinesh Bhawnani "Stock Market Prediction Using Machine Learning & Python", International Research Journal of Modernization In Engineering Technologies And Science, 2022
2. Shakir Khan , Hela Alghulaiakh "ARIMA Model For Accurate Time Series Stocks Forecasting", International Research Journal Of Modernization In Engineering Technology And Science(IJACSA) International Journal Of Advanced Computer Science And Applications, Vol. 11, No. 7, 2020
3. S. Singh, P. Gupta, "Comparative Study ID3, CART And C4.5 Decision Tree Algorithm: A Survey", International Journal Of Advanced Information Science And Technology (IJAIST) ISSN: 2319:2682 Vol.27, No.27, 2014.
4. S Habeeb Mohamed Sathak Amina, "Predictive Analytics using Machine Learning Techniques in Real Time Applications", International Journal of Scientific Research in Engineering and Management (IJSREM), ISSN: 2582-3930, Volume: 06 Issue: 07 | July – 2022
5. Kanade, P. A. (2020), "Machine Learning Model For Stock Market Prediction", International Journal For Research In Applied Science And Engineering Technology, 8(6), 209-216.Doi:10.22214/Ijraset.2020.6030.
6. Neha, & Rohan Lal. (2019), "Prediction Analysis of Stock Price Using Machine Learning" October 02, 2020. International Journal Of Innovations & Research Analysis (IJIRA) 39 ISSN: 2583-0295, Impact Factor: 5.449, Volume 03, No. 01(I), January- March, 2023, Pp 39-44.
7. I. K. Nti, A. Felix Adekoya, · Benjamin, and A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," Artif. Intell. Rev., vol. 53, pp. 3007–3057, 123AD, doi: 10.1007/s10462-019-09754-z
8. CHAU Tsun Man, SUEN Heung Ping, TO Cheuk Lam, WONG Cheuk Kin "Stock Price Prediction App Using Machine Learning Models Optimized By Evolution", The Hongkong University Of Science And Technology,2019
9. Junaid Maqbool, Preeti Aggarwal, Ravreet Kaur, Ajay Mittal, Ishfaq Ali Ganaie, "Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach", International Conference on Machine Learning and Data Engineering, 1877-0509 © 2023 The Authors. Published by Elsevier B.V.

10. Yash Mehta, Atharva Malhar, Dr. Radha Shankarmani, “Stock Price Prediction using Machine Learning and Sentiment Analysis”, 2021 2nd International Conference for Emerging Technology (INCET) Belgaum, India. May 21-23, 2021
11. S. K. Khatri and A. Srivastava, “Using sentimental analysis in prediction of stock market investment,” 2016 5th Int. Conf. Reliab. Infocom Technol. Optim. ICRITO 2016 Trends Futur. Dir., pp. 566–569, 2016, doi: 10.1109/ICRITO.2016.7785019
12. Shikhaverma , Arti , Shiwani "Prediction Of Netflix Stock Prices Using Machine Learning", in National Seminar On Emerging Trends Of Technologies In Current Era,2022
13. Israt Jahan "Stock Price Prediction Using Recurrent Neural Networks", Midwest Instruction And Computing Symposium 2018
14. H. Hauska, P. H. Swain, “The Decision Tree Classifier: Design And Potential”, 1975.
15. Y. Song, Y. Lu, “Decision Tree Methods: Applications For Classification And Prediction”, Shanghai Arch Psychiatry, Apr 25; 27(2): 130–135. 2015.
16. Paraskevas Koukaras, Christina Nousi, Christos Tjortjis, “Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning”, Telecom 2022, 3, 358-378. https://doi.org/10.3390/telecom3020019
17. Rousidis, D.; Koukaras, P.; Tjortjis, C. Social media prediction: A literature review. Multimed. Tools Appl. 2020, 79, 6279–6311.
18. Gurjar, M.; Naik, P.; Mujumdar, G.; Vaidya, T. Stock market prediction using ANN. Int. Res. J. Eng. Technol. 2018, 5, 2758–2761.
19. Yixin Guo, “Stock Price Prediction Using Machine Learning”, Södertörn University, School of Social Science Master, Economics Spring 2022
20. E. Chong, C. Han, and F. C. Park, “Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies,” Expert Syst. Appl., vol. 83, no. September, pp. 187–205, 2017, doi: 10.1016/j.eswa.2017.04.030
21. P. Chakraborty, U. S. Pria, M. R. A. H. Rony, and M. A. Majumdar, “Predicting stock movement using sentiment analysis of Twitter feed,” 2017 6th Int. Conf. Informatics, Electron. Vis. 2017 7th Int. Symp. Comput. Med. Heal. Technol. ICIEV-ISCMHT 2017, vol. 2018-Janua, pp. 1–6, 2018, doi: 10.1109/ICIEV.2017.8338584.