

Prediction of Stock Price Using Data Science Techniques

K.M.Jaashika¹, Mr.A.Mathan Gopi²

¹Dept of Computer Science

² Assistant professor, Dept of Computer Science

^{1,2} SRI MUTHUKUMARAN INSTITUTE OF TECHNOLOGY, India

Abstract- stock price prediction has become an important research area. The aim is to predict machine learning based techniques for stock price prediction results in best accuracy. The analysis of dataset by supervised machine learning technique(SMLT) to capture several information's like, variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments and analyze the data validation, data cleaning/preparing and data visualization will be done on the entire given dataset. To propose a machine learning-based method to accurately predict the stock price Index value by prediction results in the form of stock price increase or stable state best accuracy from comparing supervise classification machine learning algorithms. Dataset with evaluation classification report, identify the confusion matrix and to categorizing data from priority and the result shows that the effectiveness of the proposed machine learning algorithm technique can be compared with best accuracy with precision, Recall and F1 Score

I. INTRODUCTION

Generally, predicting how the stock market will perform is one of the most difficult things to do. It can be described as one of the most critical process to predict that. This is a very complex task and has uncertainties. To prevent this problem in One of the most interesting (or perhaps most profitable) time series data using machine learning techniques. To propose a machine learning-based method to accurately predict the stock price Index value by prediction results in the form of stock price increase or stable state best accuracy from comparing supervise classification machine learning algorithms.

II. PROPOSED SYSTEM

➤ Exploratory Data Analysis of stock Prediction

Multiple datasets from different sources would be combined to form a generalized dataset, and then different machine learning algorithms would be applied to extract patterns and to obtain results with maximum accuracy.

Data Wrangling

In this section of the report will load in the data, check for cleanliness, and then trim and clean given dataset for analysis. Make sure that the document steps carefully and justify for cleaning decisions.

Data collection

The data set collected for predicting given data is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using machine learning algorithms are applied on the Training set and based on the test result accuracy, Test set prediction is done.

III. LIST OF MODULES

- Data Pre-processing
- Data Analysis of Visualization
- Comparing Algorithm with prediction in the form of best accuracy result
- Deployment Using Flask

- Data Pre-processing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model

- Data Analysis of Visualization

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data

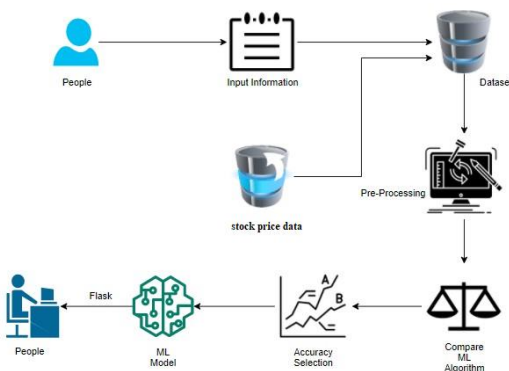
- **Comparing Algorithm with prediction in the form of best accuracy result**

Dataset with evaluation classification report, identify the confusion matrix and to categorizing data from priority and the result shows that the effectiveness of the proposed machine learning algorithm technique can be compared with best accuracy with precision, Recall and F1 Score

• Deployment Using Flask

Based upon the best accuracy of the data set we can have a decision whether the stock Price where going to increase or decrease.

IV. BLOCK DIAGRAM



V. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score will be find out. This application can help to find the Prediction of Stock price.

VI. FUTURE WORK

- Stock price prediction to connect with cloud.
- To optimize the work to implement in Artificial Intelligence environment.

REFERENCES

- [1] Jun Lee, S and Siau, K. (2001), "A review of data mining techniques", *Industrial Management & Data System*.
- [2] DallasThornton, Roland M.Mueller, PaulusSchoutsen, Josvan Hillegersberg, "Predicting Healthcare Fraud in Medicaid: A Multidimensional Data Model and Analysis Techniques for Fraud Detection"
- [3] Hossein Joudaki, Arash Rashidian, Behrouz Minaei-Bidgoli, Mahmood Mahmoodi, Bijan Geraili, Mahdi Nasiri & Mohammad Arab, "Using Data Mining to Detect Health Care Fraud and Abuse:"
- [4] Pedro A. Ortega, Cristian J. Figueroa, Cristian J. Figueroa, "A Medical Claim Fraud/Abuse Detection System based on Data Mining:"
- [5] Leonard Wafula Wakoli, Abkul Orto and Stephen Mageto, "Application of The K-Means Clustering Algorithm In Medical Claims Fraud/ Abuse Detection",
- [6] Mayank Garg, Akshit Monga, Priyank Bjatt, Anuja Arora, "Android App Behaviour Classification Using Topic Modelling Techniques and Outlier detection using AppPermissions",
- [7] El Bachir Belhadji, Georges Dionne and Faouzi Tarkhani, "A Model for the Detection of InsuranceFraud",
- [8] A. Fiat, and M. Naor, "Broadcast Encryption," in *Annual International Cryptology Conference (CRYPTO)*, 1993, pp. 480–491.
- [9] L. Zhang, C. Hu, Q. Wu, J. Domingo-Ferrer, and B. Qin, "PrivacyPreserving Vehicular Communication Authentication with Hierarchical Aggregation and Fast Response," *IEEE Transactions on Computers*, vol. 65, no. 8, pp. 2562–2574, 2016.
- [10] L. Zhang, Q. Wu, J. Domingo-Ferrer, B. Qin, and C. Hu, "Distributed Aggregate Privacy-Preserving Authentication in VANETs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 516–526, 2017.
- [11] J. Liu, J. Li, L. Zhang, F. Dai, Y. Zhang, X. Meng, and J. Shen, "Secure Intelligent Traffic Light Control Using Fog Computing," *Future Generation Computer Systems*, vol. 78, part 2, pp.
- [12] M. Burmester, and Y. G. Desmedt, "A Secure and Efficient Conference Key Distribution System," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, 1995, pp. 275–286.
- [13] S. Jiang, "Group key agreement with local connectivity", *IEEE Transactions on Dependable and Secure Computing*, vol. 13, no. 3, pp.326–339, 2016.
- [14] Q. Wu, B. Qin, L. Zhang, J. Domingo-Ferrer, and O. Farras, "Bridg- ´ ing Broadcast Encryption and Group Key Agreement," in *Annual International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT)*, 2011, pp. 143–160.
- [15] Q. Wu, B. Qin, L. Zhang, J. Domingo-Ferrer, O. Farras, and J. A. ´ Manjon, "Contributory Broadcast Encryption with Efficient En- ´ cryptation and Short Ciphertexts," *IEEE Transactions on Computers*, vol. 65, no. 2, pp. 466–479, 2016.