Stocks Market Analysis and Prediction using Gradient Boosting Neural Network

Atul Kumar Tripathi[1], Prof. Satyarth Tiwari[2]
M. Tech. Scholar, Department of Electronics and Communication, Bhabha Engineering Research Institute, Bhopal[1]
Guide, Department of Electronics and Communication, Bhabha Engineering Research Institute, Bhopal[2]

Abstract: - The financial sector is a very complex, dynamic, and non-linear efficient structure, making the forecasting of stock market progress a demanding task. In the last ten years, numerous surveys have been conducted on mining financial time series data using data mining techniques and traditional statistical methods. By identifying technical indications from prior prices and volumes and feeding them into a prediction model, technical analysis can obtain pertinent information about stocks. Finding highly informative financial indicators requires in-depth understanding of the subject, which isn't always available. A learning model's effectiveness is also significantly impacted by the data representation technique. This research seeks to design a reliable and effective data normalization approach and feature extraction method while taking into account the aforementioned challenges. The stock data is initially acquired, and it is then transformed into a synthetic format for additional processing. To create the preprocessed dataset, preprocessing is conducted to the synthesized stock data. Following the use of the feature selection approach to choose the crucial characteristics, classification is performed using a gradient boosting neural network technique, and the model's performance is assessed in terms of loss maximization absolute error (MAE) and root mean square error (RMSE).

Keywords: - MAE, RMSE, NN, Gradient Boosting

I. INTRODUCTION

Trading profitably in stocks consistently is a challenging task. It has been the belief that becoming a successful trader requires expertise gained from years of experience and a capability to spot the underlying trends from stock price movements. This makes it exceedingly difficult for a layman to trade profitably in stocks as he/she neither has the skills nor the experience. Recommender systems that can assume the role of the „expert" and recommend when to buy/sell stocks can potentially help a layman generate returns through stock trading which are higher than that obtained through other forms of investments such as fixed deposits, bonds etc. The incorporation of expert knowledge into the stock trading recommender systems is one of the major issues that need to be addressed while designing such systems. An attempt has been made in this study to explore novel approaches to integrating expert knowledge into a stock trading recommender system and address some of the issues involved in designing such systems [1, 2].

From the survey of literature on the subject over the past century starting with, considering over a hundred different articles, it is observed that stock trading recommender systems can be broadly classified into three categories. First category of recommender systems rely on fundamental analysis, ie. on utilization of macroeconomic data such as the money supply and industrial production etc. and stock specific data such as the P/E ratio, P/D ratio, dividend yields etc. in order to identify undervalued and overvalued stocks. Indicators from National Bureau of Economic Research (NBER) were also employed for the purpose by earlier systems, as seen [3, 4].

However, the efficacy of such systems is significantly dependent on the knowledge of the system designer. It is also observed that these techniques are suitable for trading in stocks over a long time horizon of several months or years and hence might not be suitable for investors interested in short-term profits. The second category of stock recommender systems employs technical analysis to identify underlying trends in financial data and/or forecast the future values [5].

These identified trends and the forecasts obtained are then used to formulate stock trading rules. The advantage such systems offer is that trades can be carried out over a much shorter span, of the order of a week or even on a daily basis. However, large numbers of technical indicators exist. Selecting the optimal set of technical indicators and identification of the optimal technical indicator parameters under the given market conditions is a challenging task [6].

Typically, traders relying on technical analysis for trading tend to select from their experience, a combination of technical indicators under the given market conditions using which they obtain the trading rules. The parameter selection of each technical indicator has also been traditionally, subjective in nature. Third category of recommender systems is more recent in its origin and employs soft computing techniques to „learn" patterns from the historical stock price data and use this „knowledge" to forecast future prices or offer trading recommendations. It is observed from the literature survey that soft computing based techniques tend to outperform the traditional techniques [7, 8].

II. STOCK MARKET PREDICTION

Stock market prediction is a significant task for the financial decision-making process and investment. Even though stock price prediction is a key problem in the financial world, it contributes to the growth of efficient methods for stock exchange transactions. Generally, stock markets are in the form of non-stationary, non-linear and uncertain even so financial experts recognized it is complex to produce precise predictions. Stock market prediction is a challenging job due to its high dynamic and unstable. Stock market prediction plans to compute the future value of a company stock trade on exchange as well as consistent prediction of future stock prices obtains high profits to investors. Various researches applied numerical data and news for the prediction of the stock market. Commonly, based on the number of information sources, the stock market prediction technique is experimented on selecting the numerical data by analysing the news data [9, 10].

In basic, forecasting behaviours are separated into three levels, such as short, medium and long. Furthermore, stock market movements are influenced by various macroeconomical aspects, like bank exchange rate, commodity price index, investors' expectations, bank rate, general economic conditions, investor's psychology, firms' policies, institutional investors' choices, political events and so on [11]. Additionally, stock value indices are computed using higher market capitalization stocks, whereas several technical parameters are also employed to obtain statistical information about stock price values [12]. In the stock market, there are two assumptions for predicting stock price value. The first one is EMH stating at any time, stock price completely confines all identified information about stock where all identified information's are utilized through market participants and also random price variations obtains new random information's.

Therefore, stock prices execute a random walk, that is every future price does not follow any patterns or trends. This assumption deduces fluctuations, so incomplete or delayed information controls the stock market prices. In addition, an exterior incident influences successive stock market prices, although the precise prediction of a stock price is complex. From the prediction perception, it can be categorized into two types, namely stock price trend and stock price forecast. The stock price trend is also named as classification, and stock price forecast is also termed as regression [14]. Basically, the time duration for stock price trend prediction is highly related with previously selected features [13].

The prediction of stock market future price is very significant for investors, because of the identification of suitable movement of stock price decreases the risk of future trend calculation. The industry, economy and other correlated features are considered to compute the intrinsic value of a company, which helps to forecast stock prices from fundamental analysis method. Stock market decision-making technique is a very complex and significant job because of unstable and complex nature of the stock market. It is necessary to discover a huge quantity of valuable information created through the stock market. In addition, every investor has an imminent requirement for identifying future behaviours of stock prices [15].

Although, it helps the investors to achieve the best profit by identifying the best moment to sell or buy stocks. Normally, trading in stock market can be performed electronically or physically. The investor becomes the owner or partnership of a particular company, while an investor obtains a particular company share. Furthermore, financial data of the stock market is very complex in nature, so for predicting stock market behaviour is also complex. The stock market prediction helps the investors to take investment decisions by offering strong insights regarding stock market behaviour for reducing investment risks [16, 17].

III. PROPOSED METHODOLOGY

Supervised machine learning classifiers can be categorized into multiple types. These types include naïve Bayes, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA), generalized linear models, stochastic gradient descent, support vector machine (SVM), linear support vector classifier (Linear SVC) decision trees, neural network models, nearest neighbours and ensemble methods. The ensemble methods combine weak learners to create strong learners. The objective of these predictive models is to improve the overall accuracy rate. This can be achieved using two strategies. One of the strategies is the use of feature engineering, and the other strategy is the use of boosting algorithms. Boosting algorithms concentrate on those training observations which end up having misclassifications. There are five vastly used boosting methods, which include AdaBoost, CatBoost, LightGBM, XGBoost and gradient boosting.

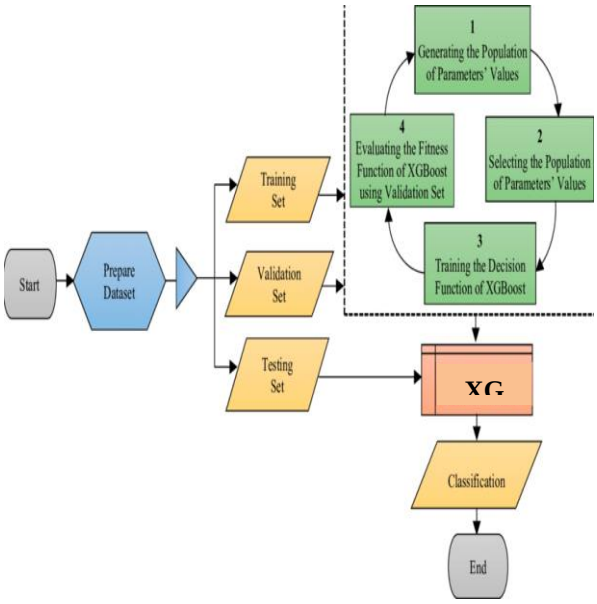

Fig. 1: Flow chart of Proposed Algorithm

Dataset was divided into two datasets (70%/30%, training/testing) to avoid any bias in training and testing. Of the data, 70% was used to train the ML model, and the remaining 30% was used for testing the performance of the proposed activity classification system. The expressions to calculate precision and recall are provided in Equations (2) and (3).

Precision provides a measure of how accurate your model is in predicting the actual positives out of the total positives predicted by your system. Recall provides the number of actual positives captured by our model by classifying these as true positive. F-measure can provide a balance between precision and recall, and it is preferred over accuracy where data is unbalanced.

Algorithm steps:

Input: D= {$(x1,)$, $(x2,)$,…, (xN,yN)}, $L(y,O(x))$

Where: $(y,(x))$ is the approximate loss function.

Begin

Initialize: $(x) = \frac{argmin}{w}\sum_{i=1}^{n} L(yi, w)$

$for\ m=1:M$

$$r_{im} = -\frac{\partial L(yi, O(xi))}{\partial O(xi)}$$

Train weak learner $C_m(x)$ on training data

Calculate $w: w_m = \arg\min \sum_{i=1}^{N} L(yi, O_{m-1}(xi) + wC_m(xi))$

Update : $O_m(x) = O_{m-1}(x) + wC_m(x)$

End for

End

Output: $O_m(x)$

IV. SIMULATION RESULTS

1. Import important libraries
2. Download and access historical dataset
https://finance.yahoo.com/quote/USO/history/
3. Show data in pandas data frame

	index	Date	Open	High	Low	Close	Adj Close	Volume	close_10	close_50	Daily Return
0	0	2006-04-10	546.000000	548.000000	541.359985	544.159973	544.159973	484738	NaN	NaN	NaN
1	1	2006-04-11	546.559998	547.119995	538.400024	545.599976	545.599976	162138	NaN	NaN	0.002646
2	2	2006-04-12	545.760010	550.479980	542.479980	542.719971	542.719971	156038	NaN	NaN	-0.005279
3	3	2006-04-13	540.000000	551.919983	539.200012	550.559998	550.559998	70088	NaN	NaN	0.014446
4	4	2006-04-17	553.599976	559.200012	549.440002	558.320007	558.320007	114713	NaN	NaN	0.014095

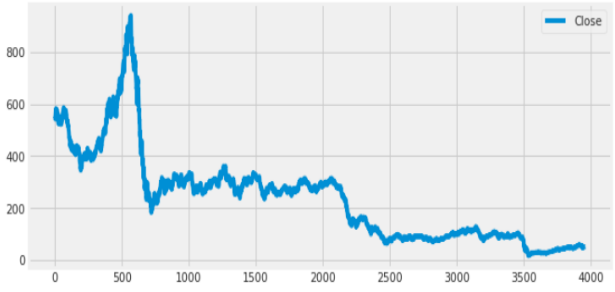

Fig. 2: Closing Price of Perform Exploratory Data Analysis

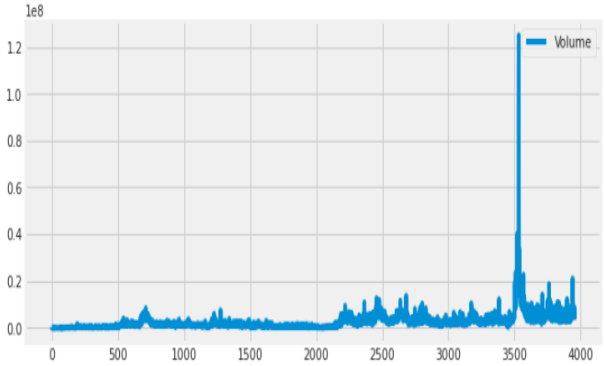

Fig. 3: Volume of Perform Exploratory Data Analysis

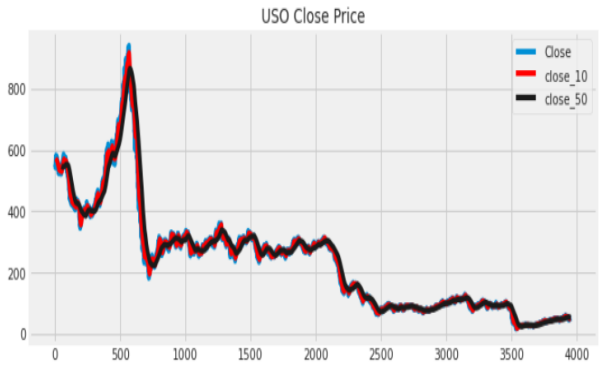

Fig. 4: USO Closing Price of Perform Exploratory Data Analysis

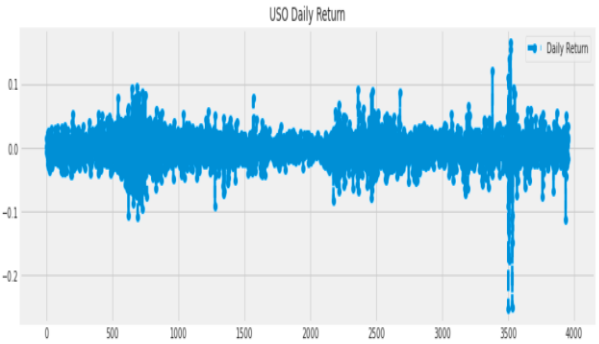

Fig. 5: USO Daily Return of Perform Exploratory Data Analysis

Fig. 6: USO Daily Closing of Perform Exploratory Data Analysis

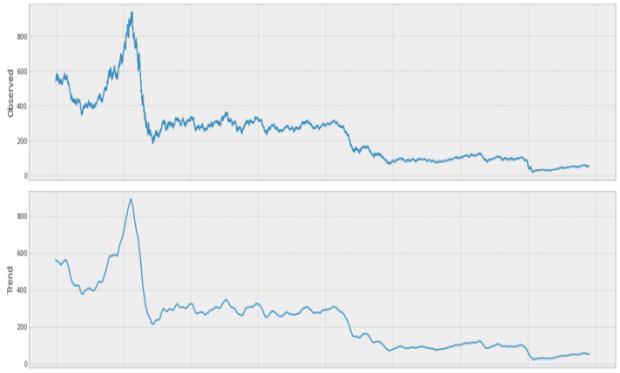

Fig. 7: Observation of Perform Exploratory Data Analysis

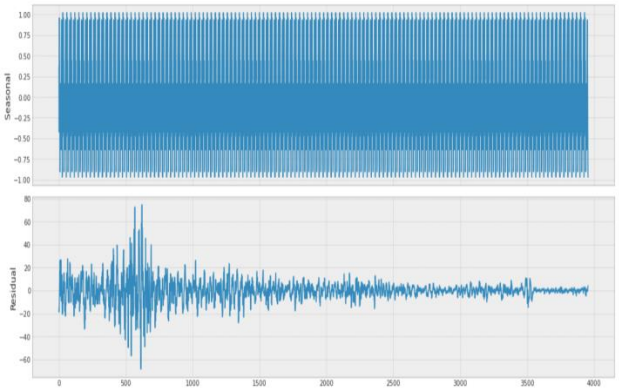

Fig. 8: Trend of Perform Exploratory Data Analysis

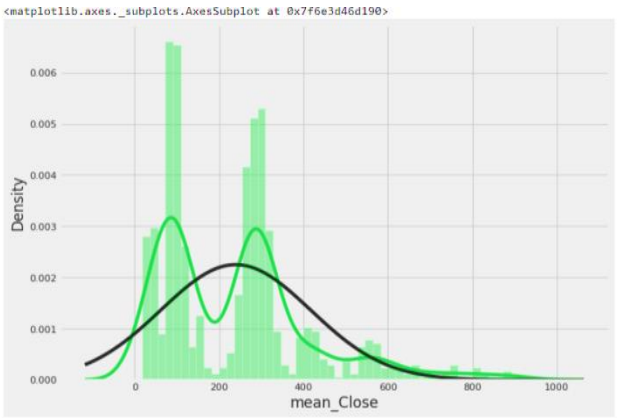

Fig. 9: Density of Perform Exploratory Data Analysis

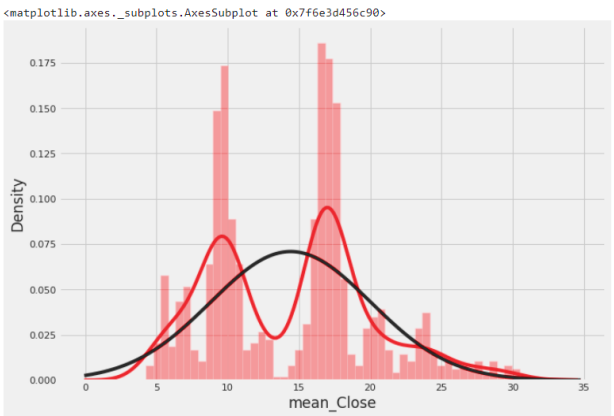

Fig. 10: Mean Deviation of Perform Exploratory Data Analysis

V. CONCLUSION

Since the financial sector is very complex, dynamic, and non-linear; anticipating stock market progress is viewed as a difficult task. Over the past decade, multi-fold surveys have been executed on mining of financial time series details, along with data mining procedures and conventional statistical methodologies. Relevant information about stocks is retrieved in technical analysis by determining technical indicators from previous prices as well as volumes, and fed into a prediction model. Identifying high-information-value financial indicators necessitates in-depth knowledge of the topic, which is not always open. Furthermore, the data representation method has a significant influence on the efficiency of a learning model. Thus, the number and type of input features affect the stock prediction.

REFERENCES

[1] Shravan Raviraj, Manohara Pai M M. and Krithika M Pai, “Share price prediction of Indian Stock Markets using time series data - A Deep Learning Approach”, IEEE Mysore Sub Section International Conference (MysuruCon), IEEE 2021.

[2] J. J. Duarte S. M. Gonzalez and J. C. Cruz "Predicting stock price falls using news data: Evidence from the brazilian market" Computational Economics vol. 57 no. 1 pp. 311-340 2021.

[3] R. Kumar P. Kumar and Y. Kumar "Multi-step time series analysis and foreasting strategy using Arima and evolutionary algorithms" International Journal of Information Technology pp. 1-15 2021.

[4] P. Ghosh A. Neufeld and J. K. Sahoo "Forecasting directional movements of513 stock prices for intraday trading using lstm and random forests" Finance514Research Letters pp. 102280 2021.

[5] Gee-Kok Tong et al. "Construction of Optimal Stock Market Portfolios Using Outlier Detection Algorithm" International Conference on Soft Computing in Data Science 2021.

[6] C. Anand "Comparison of Stock Price Prediction Models using Pre-trained Neural Networks" Journal of Ubiquitous Computing and Communication Technologies (UCCT) vol. 3 no. 02 pp. 122-134 2021.

[7] Hari Krishnan Andi "An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model" Journal of Soft Computing Paradigm vol. 3 no. 3 pp. 205-217 2021.

[8] V. Bhardwaj V. Kukreja and A. Singh "Usage of Prosody Modification and Acoustic Adaptation for Robust Automatic Speech Recognition (ASR) System" Rev. d'Intelligence Artif vol. 35 no. 3 pp. 235-242 2021.

[9] H. Kaur V. Bhardwaj and V. Kadyan "Punjabi Children Speech Recognition System Under Mismatch Conditions Using Discriminative Techniques" Innovations in Computer Science and Engineering: Proceedings of 8th ICICSE pp. 195-202 2021.

[10] P. Jiao A. Veiga and A. Walther "Social media news media and the stock525market" Journal of Economic Behavior & Organization vol. 176 pp. 63-90 2020.

[11] X. Li P. Wu and W. Wang "Incorporating stock prices and news sentiments516for stock market prediction: A

case of hong kong" Information Processing517& Management vol. 57 no. 5 pp. 102212 2020.
[12] A. Moghar and M. Hamiche "Stock market prediction using lstm recurrent neural network" Procedia Computer Science vol. 170 pp. 1168-1173 2020.
[13] P. Jiao A. Veiga and A. Walther "Social media news media and the stock498market" Journal of Economic Behavior & Organization vol. 176 pp. 63-90 2020.
[14] Onur Can Sert et al. "Analysis and prediction in sparse and high dimensional text data: The case of Dow Jones stock market" Physica A: Statistical Mechanics and its Applications vol. 545 pp. 123752 2020.
[15] Xinvi Guo and Jinfeng Li "A novel twitter sentiment analysis model with baseline correlation for financial market prediction with improved efficiency" 2019 Sixth International Conference on Social Networks Analysis Management and Security (SNAMS) 2019.
[16] E. P. TorresP M. Hernandez-Alvarez E. A. Torres Hernandez and S. G. Yoo "Stock Market Data Prediction Using Machine Learning Techniques" Information Technology and Systems pp. 539-547 2019.
[17] Siddique S. Mohanty and D. Panda "A hybrid forecasting model for prediction of stock value of tata steel using support vector regression and particle479swarm optimization" International Journal of Pure and Applied Mathematicsics vol. 119 no. 14 pp. 1719-1727 2018.
[18] Ruchira Ray Prakhar Khandelwal and B. Baranidharan "A survey on stock market prediction using artificial intelligence techniques" 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT) 2018.
[19] M Hiransha E.A Gopalkrishnan Vijay Krishna Menon and K.P Soman "NSE Stock Market Prediction Using Deep-Learning Models" International Conference on Computational Intelligence and Data Science (ICCIDS 2018)" Procedia Computer Science vol. 132 pp. 1351-1362 2018.