ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

PREDICTING STOCK MARKET TRENDS USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS VIA CONTINUOUS AND BINARY DATA

Lingamgunta Maneesha
Email: lingamguntamaneesha@gmail.com
M. Tech, Department of Computer Science and Engineering.
Annamacharya Institute of Technology and Science, Hyderabad, Telangana, India.

Ramesh Babu Varugu
Email: rameshvarugu82@gmail.com
Assistant Professor & HOD, Department of CSE.
Annamacharya Institute of Technology and Science, Hyderabad, Telangana, India.

Abstract - The nature of stock market movement has always been ambiguous for investors because of various influential factors. This study aims to significantly reduce the risk of trend prediction with machine learning and deep learning algorithms. Four stock market groups, namely diversified financials, petroleum, non-metallic minerals and basic metals from Tehran stock exchange, are chosen for experimental evaluations. This study compares nine machine learning models (Decision Tree, Random Forest, Adaptive Boosting (Adaboost), eXtreme Gradient Boosting (XGBoost), Support Vector Classifier (SVC), Naïve Bayes, K-Nearest Neighbors (KNN), Logistic Regression and Artificial Neural Network (ANN)) and two powerful deep learning methods (Recurrent Neural Network (RNN) and Long short-term memory (LSTM). Ten technical indicators from ten years of historical data are our input values, and two ways are supposed for employing them. Firstly calculating the indicators by stock trading values as continuous data, and secondly converting indicators to binary data before using. Each prediction model is evaluated by three metrics based on the input ways. The evaluation results indicate that for the continuous data, RNN and LSTM outperform other prediction models with a considerable difference. Also, results show that in the binary data evaluation, those deep learning methods are the best; however, the difference becomes less because of the noticeable improvement of models' performance in the second way.

Keywords – Stock market, trends prediction, classi cation, machine learning, deep learning.

I. INTRODUCTION

The task of stock prediction has always been a challenging problem for statistics experts and finance. The main reason behind this prediction is buying stocks that are likely to increase in price and then selling stocks that are probably to fall. Generally, there are two ways for stock market prediction. Fundamental analysis is one of them and relies on a company's technique and fundamental information like market position, expenses and annual growth rates. The second one is the technical analysis method, which concentrates on previous stock prices and values. This analysis uses historical charts and patterns to predict future prices. Stock markets were normally predicted by financial experts in the past time. However, data scientists have started solving prediction problems with the progress of learning techniques.

ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

Also, computer scientists have begun using machine learning methods to improve the performance of prediction models and enhance the accuracy of predictions. Employing deep learning was the next phase in improving prediction models with better performance. Stock market prediction is full of challenges, and data scientists usually confront some problems when they try to develop a predictive model. Complexity and nonlinearity are two main challenges caused by the instability of stock market and the correlation between investment psychology and market behavior. It is clear that there are always unpredictable factors such as the public image of companies or political situation of countries, which affect stock markets trend.

Therefore, if the data gained from stock values are efficiently preprocessed and suitable algorithms are employed, the trend of stock values and index can be predicted. In stock market prediction systems, machine learning and deep learning approaches can help investors and traders through their decisions. These methods intend to automatically recognize and learn patterns among big amounts of information. The algorithms can be effectively self-learning, and can tackle the predicting task of price punctuations in order to improve trading strategies.

Since recent years, many methods have been improved to predict stock market trends. The implementation of a model combination with Genetic Algorithms (GA), Artificial Neural Networks and Hidden Markov Model (HMM) was proposed by Hassan et al. the purpose was transforming the daily stock values to independent groups of prices as inputs to HMM. The predictability of financial trend with SVM model by evaluating the weekly trend of NIKKEI 225 index was investigated by Huang et al. A comparison between SVM, Linear Discriminant method, Elman Back propagation Neural Networks and Quadratic Discriminant method was their goal. The results indicated that SVM was the best classifier method. New financial prediction algorithm based on SVM ensemble was proposed by Sun et al.

The method for choosing SVM ensemble's base classifiers was proposed by deeming both diversity analysis and individual prediction. Final results showed that SVM ensemble was importantly better than individual SVM for classification. Ten data mining methods were employed by Ou et al. to predict value trends of Hang index from Hong Kong market. The methods involved Tree based classification, K-nearest neighbor, Bayesian classification, SVM and neural network. Results indicated that the SVM outperformed other predictive models. The value punctuations by a developed Legendre neural network was forecasted by Liu et al. By assuming investors' positions and their decisions by analyzing the prior data on the values. Indeed, they examined a random function (time strength) in the prediction model. Araujo et al. proposed the morphological rank linear forecasting approach to compare its results with time-delay added evolutionary forecasting approach and multilayer perceptron networks. From the above research background, it is clear that each of the algorithms can effectively solve stock prediction problems. However, it is vital to notice that there are specific limitations for each of them. The prediction results not only are affected by the representation of the input data but also depend on the prediction method. Moreover, using only prominent features and identifying them as input

ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

data instead of all features can noticeably develop the accuracy of the prediction models.

II. EXISTING SYSTEM

In light of employing bagging and majority vote methods, Tsai et al. used two different kinds of ensemble classifiers, such as heterogeneous and homogeneous methods. They also consider macroeconomic features and financial ratios from Taiwan stock market to examine the performance of models. The results demonstrated that with respect to the investment returns and prediction accuracy, ensemble classifiers were superior to single classifiers.

Ballings et al. compared the performance of AdaBoost, Random Forest and kernel factory versus single models involving SVM, KNN, Logistic Regression and ANN. They predict European company's prices for one year ahead. The final results showed that Random Forest outperformed among all models.

III. PROPOSED SYSTEM

In this research, we concentrate on comparing prediction performance of different machine learning models and deep learning methods to predict stock market movement. Many technical indicators are utilized as inputs to our models. Our study includes two different approaches for inputs, continuous data and binary data, to investigate the effect of preprocessing; the former uses stock trading data (open, close, high and low values) while the latter employs preprocessing step to convert continuous data to binary one. Each technical indicator has its specific possibility of up or down movement based on market inherent properties.

The performance of the mentioned models is compared for the both approaches with classification metrics, and the best tuning parameter for each is reported. All experimental tests are done with past years of historical data of four stock market groups that are totally crucial for investors.

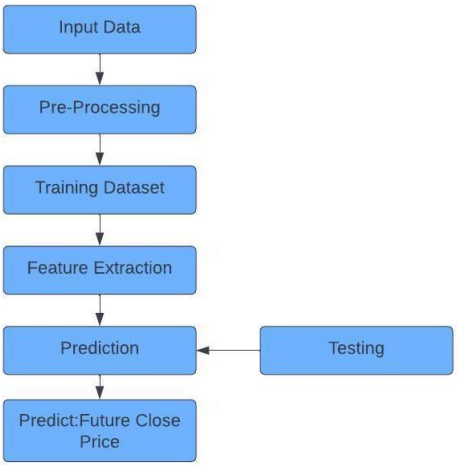
IV. METHODOLOGY:

To predict stock movements, they propose a machine learning approach that consists of five main steps, as depicted in the below diagram.

- In step I, they scrape two sets of stocks data from online resources. These datasets include: (a) Open, High, Low, Close and Volume (OHLCV) price data on Amazon, Apple, Microsoft, IBM, Facebook, and Tesla stocks from Yahoo finance (b) Public tweets about these companies from Twitter. Both of these datasets span from 2008 to 2018.
- In step II they perform data pre-processing followed by extraction of various informative features from tweets using NLP techniques.
- In Step III they fit machine learning models, and evaluate model accuracies. If the accuracy is less than a certain threshold T, Step IV is triggered.
- In step IV, we perform feature selection and transformation and then refit the machine learning models with improved set of features. Models accuracies are then compared with those in previous step.
- Finally, in step V we further try to improve classification performance using regularized model stacking. Latent Dirichlet Allocation (LDA) is used for topic modelling to validate the legitimacy of tweets corpus while Granger Causality analysis is

ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

used to check whether there is a statistically significant causality between sentiments driven from tweets and the stock returns.



V. REQUIREMENT ANALYSIS

IMPLEMENTATION:

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Analyze and Prediction
- Accuracy on test set
- Saving the Trained Model

MODULES DESCSRIPTION:

Data Collection:

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get; the better our model will perform.

There are several techniques to collect the data, like web scraping, manual interventions and etc.

The dataset taken from the

Link:https://www1.nseindia.com/products/content/equities/indices/historical_index_data.htm

Dataset:

The dataset consists of 100252 individual data. There are 8 columns in the dataset but we taken only two country data set one India and other totally which are described below.

Date - specifies trading date
Open - opening price
High - maximum price during the day
Low - minimum price during the day
Close - close price adjusted for splits
Adj Close - adjusted close price adjusted for both dividends and splits.
Volume - the number of shares that changed hands during a given day

Data Preparation:

We will transform the data by getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain. Next we drop or remove all columns except for the columns that we want to retain. Finally we drop or remove the rows that have missing values from the data set.

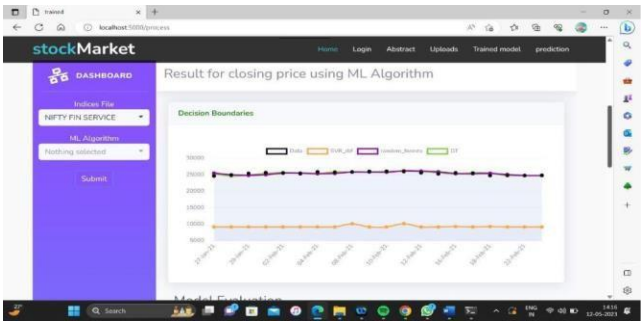
VI. SYSTEM TESTING:

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product. It is the process of exercising

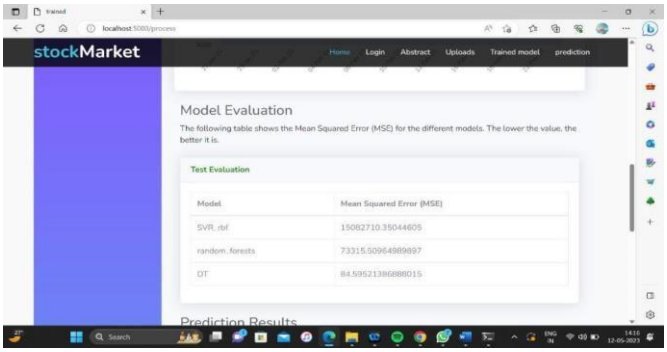
ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

software with the intent of ensuring that the software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

VII. RESULTS



Shows the comparison graph of the algorithms SVR_rbf, Random Forest, Decision tree.



Model Evaluation of the used algorithms is displayed which depicts the best value among the three algorithms.

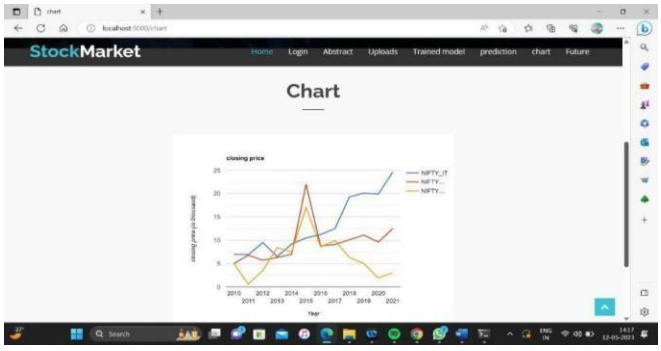
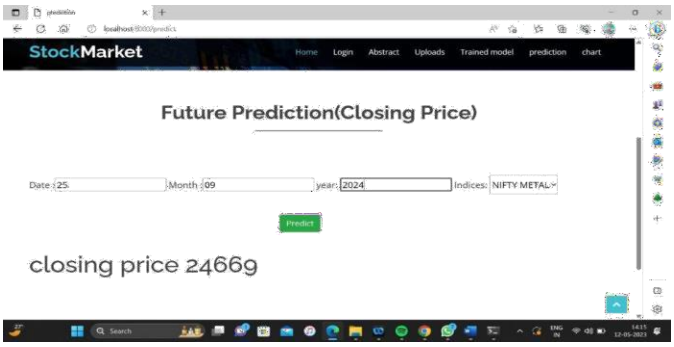


Chart description of the three factors, IT, Metel, Finances, Closing Price from the year



Future prediction of a given date is given with respect to the indices

VIII. CONCLUSION

The purpose of this study was the prediction task of stock market movement by machine learning and deep learning algorithms. Four stock market groups, namely diversified financials, petroleum, non-metallic minerals and basic metals, from Tehran stock exchange were chosen, and the dataset was based on ten years of historical records with ten technical features. Also, nine machine learning models (Decision Tree, Random Forest, Adaboost, XGBoost, SVC, Naïve Bayes, KNN, Logistic Regression and ANN) and two deep learning methods (RNN and LSTM) were employed as predictors. We supposed two approaches for input values to models, continuous data and binary data, and we employed three classification metrics for evaluations. Our experimental works showed that there was a significant improvement in the performance of models when they use binary data instead of continuous one. Indeed, deep learning algorithms (RNN and LSTM) were our superior models in both approaches.

IX. FUTURE SCOPE

Overall, it is obvious that the whole of algorithms predict well as they are trained with continuous values (up to 67%), but the models' performance is remarkably improved when they are trained with binary data (up to 83%). The

ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

result behind this improvement is interpreted as follows: an extra layer is employed in the second approach, and the duty of the layer is comparing each current continuous value (at time t) with previous value (at time t-1). So the future up or down trend is identified and when binary data is given as the input values to the predictors, we enter data with a recognized trend based on each feature's property. This critical layer is able to convert non-stationary values in the first approach to trend deterministic values in the second one, and algorithms must find the correlation between input trends and output movement as an easier prediction task. Despite careful tries to find valuable researches on the similar stock market, no important paper could be reported and compared; therefore, this deficiency is presented as a novelty of this study, which will be a baseline for future articles.

X. REFERENCES

[1] S. Bouktif, A. Fiaz, and M. Awad, "Augmented textual features based stock market prediction," *IEEE Access*, vol. 8, pp. 40269–40282, 2020.

[2] M. Khedmati, F. Seifi, and M. J. Azizi, "Time series forecasting of bitcoin price based on autoregressive integrated moving average and machine learning approaches," Int. J. Eng., vol. 33, no. 7, pp. 1293–1303, 2020.

[3] B. Lutz, N. Pröllochs, and D. Neumann, "Predicting sentence-level polarity labels of financial new using abnormal stock returns," Expert Syst. Appl., vol. 148, Jun. 2020, Art. no. 113223.

[4] N. Seong and K. Nam, "Predicting stock movements based on financial news with segmentation," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 113988.

[5] Pushpendra Singh Sisodia and J. Gordon, Anish Gupta , Yogesh Gupta "Stock Market Analysis and prediction for Nifty 50 using LSTM Deep Learning Approach ," (ICIPTM)., vol. 112, pp. 88–97, Aug. 2018.

[6] D. C. Furlaneto, L. S. Oliveira, D. Menotti, and G. D. C. Cavalcanti, "Bias effect on predicting market trends with EMD," *Expert Syst. Appl.*, vol. 82, pp. 19–26, Oct. 2017.

[7] S.M. Ajitha, S.Karthick, Mu.Karthikeyan, N.T.Naveen Balaji " Amazon Stock Price Prediction By Using Smlt." IRJET e-ISSN: 2395-0056 Vol. 10, Apr 2023

[8] T. Swathi, N. Kasiviswanath & A. Ananda Rao, "An optimal deep learning-based LSTM for stock price prediction using twitter sentiment analysis." IEEE Trans. Ind. Informat., vol. 17, no. 1, pp. 720–727, Jan. 2021.

[9] X. Wu, Q. Ye, H. Hong, and Y. Li, "Stock selection model based on machine learning with wisdom of experts and crowds," IEEE Intell. Syst., vol. 35, no. 2, pp. 54–64, Mar. 2020.

[10] R. Ren, D. D. Wu, and T. Liu, "Forecasting stock market movement direction using sentiment analysis and support vector machine," IEEE Syst. J., vol. 13, no. 1, pp. 760–770, Mar. 2018.

[11] T. Fischer and C. Krauss, "Deep learning with long short-term memory 600 networks for financial market predictions," Eur. J. Oper. Res., vol. 270, 601 no. 2, pp. 654–669, Dec. 2017.