Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1, Paper 2, January 2021

STOCK PRICE TREND FORECASTING USING MACHINE LEARNING

Vaishnavi S[1], Rokitha K[2], Swathi D[3], Moushmi S[4]
[1]Assistant Professor, RMK college of Engineering and Technology
[234] Department of Computer Science and Engineering, RMK College of Engineering and Technology
[1] vaishnavicse@rmkcet.ac.in , [2] rokithakaruppiah@gmail.com, [3]swathidevarajan7177@gmail.com, [4]scmmoushmi003@gmail.com

Abstract— *Nowadays, 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. Hence, 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-variant analysis, bi-variant and multi-variant 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 supervised classification machine learning algorithms. Additionally, to compare and discuss the performance of various machine learning algorithms from the given transport traffic department dataset with evaluation of GUI based user interface stock price prediction by attributes, 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.*
Keywords — Stock prices, stock market, machine learning, regression, classification, prediction.

I. INTRODUCTION

In this modern world, Stock market prediction is the method of determining the future value of a stock or other financial instrument traded on an exchange. A misconception is also associated with people that buying and selling of the stocks/shares in the market is an act of gambling. This misconception can be changed and bringing awareness among people for this. Over the past few years, 90 percent of the data in the world has been created as a result of the creation of 2.5 quintillion bytes of data on a daily basis. A very large amount of data is generated by financial market. It's very difficult for a trader to recognize a pattern and then devise an optimal strategy for making decisions. Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the Prediction – physical factors vs physiological, rational and irrational behavior, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy. Machine Learning can be used as a game changer in predicting the values of stock prices. Machine learning techniques have the potential to unearth patterns and insights we didn't see before, and these can be used to make unerringly accurate predictions. The machine learning is growing at a phenomenal pace in today's world.

The purpose of this study is to find out what variables form the main factors–called the principal component–in determining stock prices, which hopefully can be a reference for further research; especially regarding the factors which affect stock prices in companies in the consumer goods industry sector, by using Principal Component Analysis method which could reduce the number of determinants

of stock prices and form a new component. This research can be a reference for investors in predicting the benefits derived from their investments in the form of shares in manufacturing companies in the consumer goods industry sector in Indonesia, using the principal components, obtained using the Principal Component Analysis method, as the main determinant of stock prices.

II. LITERATURESURVEY

Lobna Nassar[1] proposed the ARIMA model has the highest mean absolute percentage error (MAPE) hence the lowest performance compared to the conventional ML models. In addition, among the conventional the Gradient Boosting (GB) is the best due to having the least MAPE error. Finally, the performance of LSTM simple DL model is higher than all of the tested conventional ML models for two FP (Watermelon and Bok Choy). This is due to having less markets for these two FP which leaves us with data that is closer in nature to time series. It is also found that the best performing model according to the aggregated measure is the compound DL model, ATTCNN-LSTM, which outperforms the ML and simple DL models in accuracy of price prediction especially after adding the attention.

Deepu Rajan[2] A deep hybrid fuzzy neural Hammerstein-Wiener model (FNHW), is proposed in this paper. The implication and inference of a neuro-fuzzy is based on the fuzzy rule base that has been formed during training. It requires the training data to be able to adequately represent entire system behaviors. However, the test data may vary with distribution shift in time series domain. Furthers, the training data may be derived from steadystate while the test data which is in the form of dynamically changing represented by drastic data shift under certain scenario such as financial crisis. the soundness of rule base inference from neuro-fuzzy system on the steady-state data is achieved as well as inheriting the good approximation accuracy and excellent asymptotic tracking advantages of Hammerstein-Wiener model on the dynamically changing data. The effectiveness of proposed model is evaluated on two financial stock price prediction datasets. A deep hybrid fuzzy neural Hammerstein-Wiener network for stock price prediction by combining the benefits of neural fuzzy system and Hammerstein-Wiener model to handle steady-state and dynamically changing data correspondingly. Asymptotic tracking ability of Hammerstein-Wiener model to handle dynamically changing data. We performed the experiments on two different financial stock price prediction datasets and showed that the prediction performance has been significantly improved using our model when compared to other state-of-art neuro-fuzzy systems.

Rubi Gupta[3], Analysis on Stock Twits data and to understand the impact of sentiments on stock price movements. They plan to further improve the work in the following areas. First, in this work, we use two types of sentiments: bullish (positive) and bearish (negative). Adding neutral sentiment might reduce noise and potentially enhance accuracy of the work. Second, our analysis is limited to five companies. An expansion to broader set of companies or all Stock Twits data might yield more insights into the data, leading to more effective application in stock price prediction. They are use the optional sentiment labels provided by Stock Twits users as the ground truth data for model training. Sentiment information is used in addition to the past stock time series data to improve the accuracy of stock price movement prediction. The effectiveness of the proposed work on stock price prediction is demonstrated through experiments on five companies. Stock Twits is a relatively new micro blogging website, which is becoming increasingly popular for users to share their discussions and sentiments about stocks and financial markets. Provided a reasonable evidence that sentiments data has a positive impact on the accuracy of stock price change prediction.

Jiannan Chen[4], Stock trend prediction has always been the focus of research in the field of financial big data. Stock data is complex nonlinear data, while stock price is changing over time. Based on the characteristics of stock data, this paper proposes a financial big data Stock Trend Prediction Algorithm based on attention mechanism (STPA). We adopt Bidirectional Gated Recurrent Unit (BGRU) and attention mechanism to capture the long-term dependence of data on time. Reduction algorithm based on the attention mechanism (STPA) proposed the entire algorithm is divided into three layers. That is, the stock price change trend vector representation layer, the BGRU feature extraction layer, and the stock price change trend prediction attention mechanism layer. STPA method to predict the change trend of financial big data stocks. STPA uses the Bidirectional Gated Recurrent Unit model and introduces attention mechanism technology. STPA method performs better than the current mainstream algorithms in predicting stock changes in the financial stock price data set, which demonstrates the effectiveness of the proposed method. From the experimental results, STPA method performs better than the current mainstream algorithms in predicting stock changes in the financial stock price data set, which demonstrates the effectiveness of the proposed method.

Rahma Firsty Fitriyana[5], stock price is the important factor in achieving the profit in stock investment, and the prediction is usually done by relating the price of a stock to factors that influence it. The problem is, there are a large number of variables that can be used to predict the stock prices so it is difficult for a potential investor to choose which variables should be used in predicting the stock prices. This research used the Principal Component Analysis as the dimension reduction method to form major components that influence the stock prices without losing the information and uses data from five companies. Analysis method can be used to find the main determinants of stock prices by adding new variables to get more accurate results such as macroeconomic factors and adding other financial ratios, because there are many variables affecting stock prices, including macroeconomic factors that did not included in this research. Next research could include those factors to see their impact on stock prices.

III. EXISTING SYSTEM

The existing system uses Fuzzy rough theory that describe real-world situations in a mathematically effective and interpretable way and evolutionary neural networks can be utilized to solve complex problems.

DRAWBACKS:

Their training process takes long time stamp. Then hard to modification for training network. It can't thereby better determine the regularity of stock price prediction data and achieve more accurate prediction results.

IV. PROPOSED SYSTEM

The proposed system implements machine learning approach and the result is shown in the form of user interface of GUI application. Multiple datasets from different sources are combined to form a generalized dataset, and then different machine learning algorithms are applied to extract patterns and to obtain results with maximum accuracy.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1, Paper 2, January 2021

ADVANTAGE:

Accuracy can be predicted with minimum timestamp.Since it is GUI representation, People without mere knowledge over the field can even understand.

V. IMPLEMENTATION

Data Validation:

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters. The validation set is used to evaluate a given model, but this is for frequent evaluation.

Exploration data analysis of visualization:

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more

visceral and stakeholders than measures of association or significance. Data visualization and exploratory data analysis are whole fields themselves and it will recommend a deeper dive into some the books mentioned at the end. Sometimes data does not make sense until it can look at in a visual form, such as with charts and plots. How to chart time series data with line plots and categorical quantities with bar charts. How to summarize data distributions with histograms and box plots. How to summarize the relationship between variables with scatter plots.

Accuracy results of logistic regression and decision tree algorithm :

In the next section you will discover exactly how you can do that in Python with scikit-learn. The key to a fair comparison of machine learning algorithms is ensuring that each algorithm is evaluated in the same way on the same data and it can achieve this by forcing each algorithm to be evaluated on a consistent test harness.

In the example below 2 different algorithms are compared:

Logistic Regression

Decision tree

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1, Paper 2, January 2021

Training the Dataset:

The first line imports iris data set which is already predefined in sklearn module and raw data set is basically a table which contains information about various varieties.For example, to import any algorithm and train_test_split class from sklearn and numpy module for use in this program.To encapsulate load_data() method in data_dataset variable. Further divide the dataset into training data and test data using train_test_split method. The X prefix in variable denotes the feature values and y prefix denotes target values.

Testing the Dataset:

Now, the dimensions of new features in a numpy array called 'n' and it want to predict the species of this features and to do using the predict method which takes this array as input and spits out predicted target value as output.So, the predicted target value comes out to be 0. Finally to find the test score which is the ratio of no. of predictions found correct and total predictions made and finding accuracy score method which basically compares the actual values of the test set with the predicted values.

Accuracy results of Random Forest and SVM algorithms:

In this module we are going to compare the accuracy of Random forest and SVM. The accuracy results of each algorithm

UI based prediction results of stock will rise or not:

Tkinter is a python library for developing GUI (Graphical User Interfaces). We use the tkinter library for creating an application of UI (User Interface), to create windows and all other graphical user interface .The inputs will be given and the output will be predicted whether the stock will rise or not.

V. SYSTEM DESIGN

Software Requirements:

Operating System	: Windows
Tool	: Anaconda with Jupyter Notebook

Hardware requirements:

Processor	: Pentium IV/III
Hard disk	: minimum 80 GB
RAM	: minimum 2 GB

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1, Paper 2, January 2021

VI. SYSTEM ARCHITECTURE

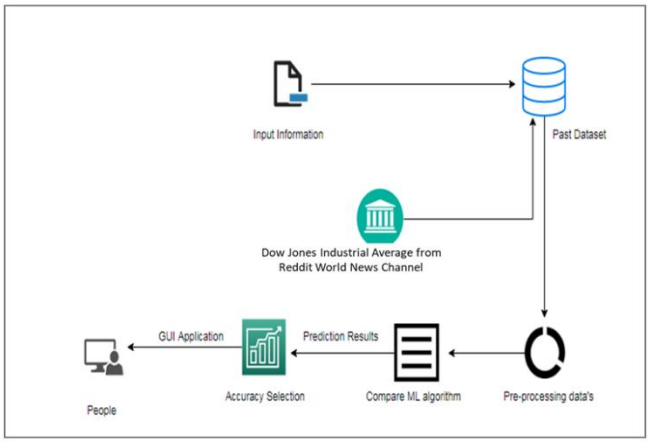

Figure 1

In the above figure 1 the collected dataset is given as input information. It undergoes a pre-processing step where the data value is converted into known value that can be fetched to machine. The four algorithms are trained to the machine. These four algorithms produce results based upon the input values. Based on accuracy rate these algorithms predict the output.

VII. RESULT AND ANALYSIS

The accuracy result was predicted using graphical representation.

LOGISTIC REGRESSION

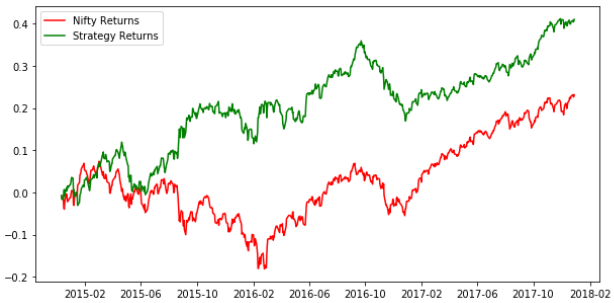

Figure 2

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1, Paper 2, January 2021

The following figure shows the above-defined variable pred, which is a real number, and its conversion between 0 and 1,which represents probality, using the preceding transformation.

RANDOM FOREST

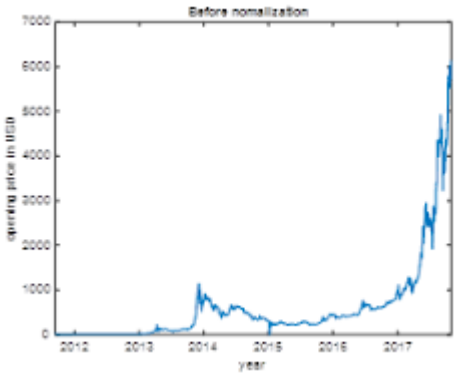

Figure 3

In the above diagram, we can observe that each decision tree has voted or predicted a specific class. The final output is selected by random forest will be class N as it has majority votes or is the predicted output by two or four decision tree.

DECISION TREE

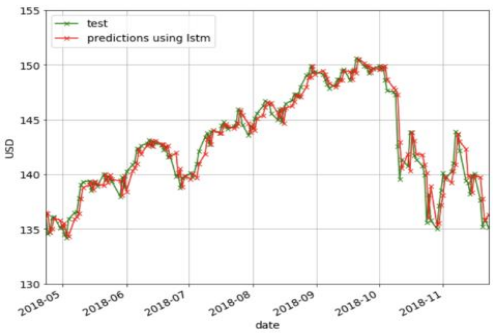

Figure 4

The entire dataset is navigated from the root node of the decision tree down to the leaf, according to set criteria and then the graph is predicted.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1, Paper 2, January 2021

SVM

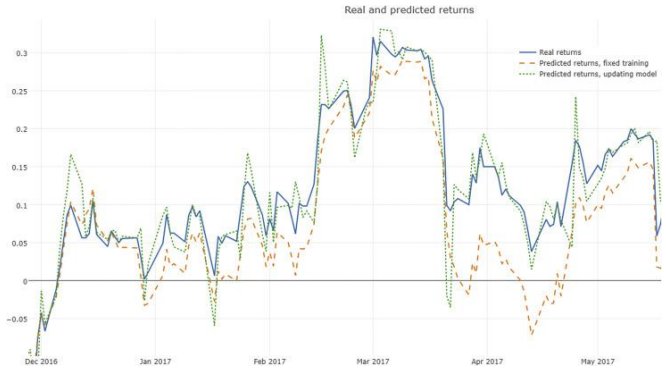

Figure 5

The optimal hyper plane in the middle and then two dotted lines as our boundary lines that go through the closest data point in each class.

VIII. CONCLUSION

In this paper, we study the use of decision trees and support vector machines to predict stock movement direction. Of both these algorithms, we saw that Logistic Regression gave us better results. It is a promising type of tool for stock forecasting. It is superior to the other individual classification methods in forecasting daily movement direction. This is a clear message for stockl forecasters and traders, which can lead to a capital gain. However, each method has its own strengths and weaknesses. In this model, the principal components identified by the LR are used along with internal and external financial factors for forecasting. We also observed that the choice of the indicator function can dramatically improve/reduce the accuracy of the prediction system. Also a particular Machine Learning Algorithm might be better suited to a particular type of stock, say Technology Stocks, whereas the same algorithm might give lower accuracies while predicting some other types of Stocks.

IX. FUTURE WORK

India meteorological department wants to automate the detecting the air quality is good or not from eligibility process (real time).To automate this process by show the prediction result in web application or desktop application. To optimize the work to implement in Artificial Intelligence environment.

REFERENCES

[1] Integrated Long-term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market", DOI 10.1109/ACCESS.2020.2969293, IEEE Access.

[2] "A Deep Hybrid Fuzzy Neural Hammerstein-Wiener Network for Stock Price Prediction ", Xie Chen, Deepu Rajan,

Chai Quek School of Computer Science and Engineering Nanyang Technological University 50 Nanyang Avenue, Singapore.

[3] [3]"Deep Learning Based Approach for Fresh Produce Market Price Prediction", Electrical and Computer Engineering Department University of Waterloo Ontario, Canada

[4] [4]" Prediction of Stock Prices using Machine Learning (Regression Classification) Algorithms ", 2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020

[5] [5]"CUDA parallel computing framework for stock market prediction using K-means clustering", Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC 2020) IEEE Xplore Part Number: CFP20V90-ART; ISBN: 978-1-7281-5461-9

[6] [6]"Principal Component Analysis to Determine Main Factors Stock Price of Consumer Goods Industry", 2020 International Conference on Data Science

[7] [7]"Prediction of Financial Big Data Stock Trends Based on Attention Mechanism", 2020 IEEE International Conference on Knowledge Graph (ICKG)