Stock Market Prediction Using Machine Learning

Achyut R Kulkarni[1], Deekshith Lobo[2], Manjunath Ambiger[3], Dimple Halgeri[4]
[1, 2, 3, 4] Dept of Computer Science
[1, 2, 3, 4] Atria Institute of Technology

Abstract- *The main objective of this paper is to find the best model to predict the value of the stock market. During the process of considering various techniques and variables that must be taken into account, we found out that techniques like linear regression, support vector machine, Long short term Memory (LSTM). In, this paper we are going to present and review a more feasible method to predict the stock movement with higher accuracy. The first thing we have taken into account is the dataset of the stock market prices from previous year. The dataset was pre-processed and tuned up for real analysis. Secondly, after pre- processing the data, we will review the use of linear regression, support vector machine and LSTM on the dataset and the outcomes it generates. In addition, the proposed paper examines the use of the prediction system in real-world settings and issues associated with the accuracy of the overall values given. The paper also presents a machine-learning model to predict the longevity of stock in a competitive market. The successful prediction of the stock will be a great asset for the stock market institutions and will provide real-life solutions to the problems that stock investors face.*

Keywords- Machine Learning, Data Pre-processing, Data Mining, Dataset, Stock, Stock Market.

I. INTRODUCTION

The stock market is basically an aggregation of various buyers and sellers of stock. A stock (also known as shares more commonly) in general represents ownership claims on business by a particular individual or a group of people. The attempt to determine the future value of the stock market is known as a stock market prediction. The prediction is expected to be robust, accurate and efficient. The system must work according to the real-life scenarios and should be well suited to real-world settings. The system is also expected to take into account all the variables that might affect the stock's value and performance. There are various methods and ways of implementing the prediction system like Fundamental Analysis, Technical Analysis, Machine Learning, Market Mimicry, and Time series aspect structuring. With the advancement of the digital era, the prediction has moved up into the technological realm. The most prominent and promising technique involves the use of Artificial Neural Networks, Recurrent Neural Networks, that is basically the implementation of machine learning. Machine learning involves artificial intelligence which empowers the system to learn and improve from past experiences without being programmed time and again. Traditional methods of prediction in machine learning use algorithms like Backward Propagation, also known as Backpropagation errors. Lately, many researchers are using more of ensemble learning techniques. It would use low price and time lags to predict future highs while another network would use lagged highs to predict future highs. These predictions were used to form stock prices. Stock market price prediction for short time windows appears to be a random process.

The stock price movement over a long period of time usually stock market refrain people from investing in stocks. Thus, there is a need to accurately predict the stock market which can be used in a real-life scenario. The methods used to predict the stock market includes a time series forecasting along with technical analysis, machine learning modeling and predicting the variable stock market. The datasets of the stock market prediction model include details like the closing price opening price, the data and various other variables that are needed to predict the object variable which is the price in a given day. The previous model used traditional methods of prediction like multivariate analysis with a prediction time series model. Stock market prediction out performs when it is treated as a regression problem but performs well when treated as a classification. The aim is to design a model that gains from the market information utilizing machine learning strategies and gauge the future patterns in stock value development. The Support Vector Machine(SVM) can be used for both classification and regression. It has been observed that SVMs are more used in classification-based problem like ours. The SVM technique, we plot every single data component as a point in n- dimensional space (where n is the number of features of the dataset available) with the value of feature being the value of a particular coordinate and, hence classification is performed by finding the hyperplane that differentiates the two classes explicitly.

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points (such as images), but also entire

sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition, speech recognition and anomaly detection in network traffic or IDS's (intrusion detection systems). A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information in to and out of the cell. In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models.

Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to bean affine function of those values; less commonly, the conditional median or some other quantile is used.

II. PROBLEMDEFINITION

Stock market prediction is basically defined as trying to determine the stock value and offer a robust idea for the people to know and predict the market and the stock prices. It is generally presented using the quarterly financial ratio using the dataset. Thus, relying on a single dataset may not be sufficient for the prediction and can give a result which is inaccurate. Hence, we are contemplating towards the study of machine learning with various datasets integration to predict the market and the stock trends.

The problem with estimating the stock price will remain a problem if a better stock market prediction algorithm is not proposed. Predicting how the stock market will perform is quite difficult. The movement in the stock market is usually determined by the sentiments of thousands of investors. Stock market prediction, calls for an ability to predict the effect of recent events on the investors. These events can be political events like a statement by a political leader, a piece of news on scam etc. It can also be an international event like sharp movements in currencies and commodity etc. All these events affect the corporate earnings, which in turn affects the sentiment of investors. It is beyond the scope of almost all

investors to correctly and consistently predict these hyper parameters.

All these factors make stock price prediction very difficult. Once the right data is collected, it then can be used to train a machine and to generate a predictive result.

III. LITERATURE SURVEY

During a literature survey, we collected some of the informationaboutStockmarketpredictionmechanismscurrentlybeingused.

1. Survey of Stock Market Prediction Using Machine Learning Approach

The stock market prediction has become an increasingly important issue in the present time. One of the methods employed is technical analysis, but such methods do not always yield accurate results. So it is important to develop methods for a more accurate prediction. Generally, investments are made using predictions that are obtained from the stock price after considering all the factors that might affect it. The technique that was employed in this instance was a regression. Since financial stock marks generate enormous amounts of data at any given time a great volume of data needs to undergo analysis before a prediction can be made. Each of the techniques listed under regression has its own advantages and limitations over its other counterparts. One of the noteworthy techniques that were mentioned was linear regression. The way linear regressn models work is that they are often fitted using the least squares approach, but they may alternatively be also be fitted in other ways, such as by diminishing the "lack off it" in some other norm, or by diminishing a handicapped version of the least squares loss function. Conversely, the least squares approach can be utilized to fit nonlinear models.

2. Impact of Financial Ratios and Technical Analysis on Stock Price Prediction Using Random Forests

The use of machine learning and artificial intelligencetechniquestopredictthepricesofthestockisanincreasing trend. More and more researchers invest their time every day in coming up with ways to arrive at techniques that can further improve the accuracy of the stock prediction model.

Due to the vast number of options available, there can ben number of ways on how to predict the price of the stock, but all methods don't work the same way. The output varies for each technique even if the same data set is being applied. In the cited paper the stock price prediction has been

carried out by using the random forest algorithm is being used to predict the price of the stock using financial ratios form the previous quarter. This is just one way of looking at the problem by approaching it using a predictive model, using the random forest to predict the future price of the stock from historical data. However, there are always other factors that influence the price of the stock, such as sentiments of the investor, public opinion about the company, news from various outlets, and even events that cause the entire stock market to fluctuate. By using the financial ratio along with a model that can effectively analyze sentiments the accuracy of the stock price prediction model can be in creased.

3. Stock Market Prediction via Multi- Source Multiple Instance Learning

Accurately predicting the stock market is a challenging task, but the modern web has proved to be a very useful tool in making this task easier. Due to the interconnected format of data, it is easy to extract certain sentiments thus making it easier to establish relationships between various variable and roughly scope out a pattern of investment. Investment pattern from various firms show sign of similarity, and the key to successfully predicting the stock market is to exploit these same consistencies between the data sets. The way stock market information can be predicted successfully is by using more than just technical historical data, and using other methods like the use of sentiment analyzer to derive an important connection between people's emotions and how they are influenced by investment in specific stocks. One more important segment of the prediction process was the extraction of important events from web news to see how it affected stock prices.

4. Stock Market Prediction: Using Long Short Term Memory

LSTM is the advanced version of Recurrent-Neural-Networks (RNN) where the information belonging to previous state persists. These are different from RNNs as they involve long term dependencies and RNNs works on finding the relationship between the recent and the current information. This indicates that the interval of information is relatively smaller than that to LSTM. The main purpose behind using this model in stock market prediction is that the predictions depends on large amounts of data and are generally dependent on the long-term history of the market. So LSTM regulates error by giving an aid to the RNNs through retaining information for older stages making the prediction more accurate. Thus, proving itself as much more reliable compared to other methods.

IV. PROPOSEDSYSTEM

In this proposed system, we focus on predicting the stock values using machine learning algorithms like LSTM, Linear Regression and Support Vector Machines. We proposed the system "Stock market price prediction" we have predicted the stock market price using the SVM. In this proposed system, we were able to train the machine from the various data points from the past to make a future prediction. We took data from the previous year stocks to train the model. We majorly used two machine-learning libraries to solve the problem. The first one was numpy, which was used to clean and manipulate the data, and getting it in to a form ready for analysis. The other was scikit, which was used for real analysis and prediction. The data set we used was from the previous years stock markets collected from the public database available online, 80 % of data was used to train the machine and the rest 20 % to test the data. The basic approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data. We used the python pandas library for data processing which combined different datasets into a data frame. The tuned up data frame allowed us to prepare the data for feature extraction. The data frame features were date and the closing price for a particular day. We used all these features to train the machine on random forest model and predicted the object variable, which is the price for a given day. We also quantified the accuracy by using the predictions for the test set and the actual values.

V. METHODOLOGIES

1. Classification

Classification is an instance of supervised learning where a set is analyzed and categorized based on a common attribute. From the values or the data are given, classification draws some conclusion from the observed value. If more than one input is given then classification will try to predict one or more outcomes for the same. A few classifiers that are used here for the stock market prediction includes the LSTM, Linear Regression, SVM classifier.

SVM classifier

SVM classifier is a type of discriminative classifier. The SVM uses supervised learning i.e. a labeled training data. The output are hyperplanes which categorizes the new dataset. They are supervised learning models that uses associated learning algorithm for classification and as well as regression. Parameters The tuning parameters of SVM classifier are

kernel parameter, gamma parameter and regularization parameter.

- Kernels can be categorized as linear and polynomial kernels calculates the prediction line. In linear kernels prediction for a new input is calculated by the dot product between the input and the support vector.
- C parameter is known as the regularization parameter; it determines whether the accuracy of model is increases or decreases. The default value of c=10.Lower regularization value leads to misclassification.
- Gamma parameter measures the influence of a single training on the model. Low values signify far from the plausible margin and high values signifies closeness from the plausible margin.

Regression classifier

In general, the Regression based Model is used for predicting continuous values through some given independent values. The paper utilizes the gradient descent linear regression algorithm for predicting correct values by minimizing the error function. Linear Regression as governed by the above equation is performed on the data and then the relevant predictions are made. The factors considered for the regression were low, open, high, close and volume. TheR-square confidence test was used to determine the confidence score and the predictions were plotted to show the results of the stock market prices vs time.

LSTM classifier

LSTM is the advanced version of Recurrent-Neural-Networks (RNN) where the information belonging to previous state persists. These are different from RNNs as they involve long term dependencies and RNNs works on finding the relationship between the recent and the current information. This indicates that the interval of information is relatively smaller than that to LSTM. The main purpose behind using this model in stock market prediction is that the predictions depends on large amounts of data and are generally dependent on the long term history of the market.

VI. EXPERIMENTALRESULTS

The proposed system is trained and tested over the dataset taken from Yahoo Finance. It is split into training and testing sets respectively and yields the following results upon passing through the different models:

A. Regression Based Model Results

The plot in figure is the result of application of linear regression algorithm on the dataset to predict varying prices with respect to the time.
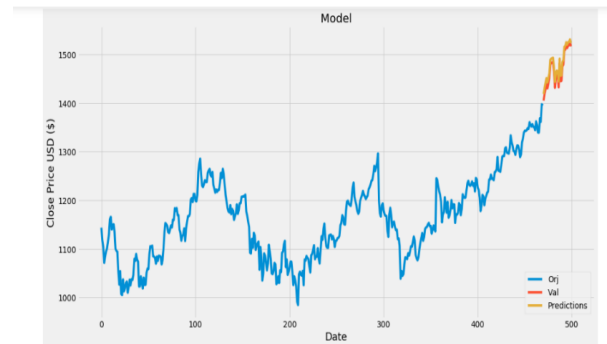

Fig. Plot between Price and Date Using Regression

B. LSTM Based Model Results

The prediction is shown by red line and the actual trend is shown by blue. The proximity of these two lines tells, how efficient the LSTM based model is. The prediction approximates real trend when a considerable amount of time has passed. The more the system is trained and the greater the size of the dataset utilized the greater the accuracy which will be attained. The LSTM Model offered more accuracy than the Regression based Model.
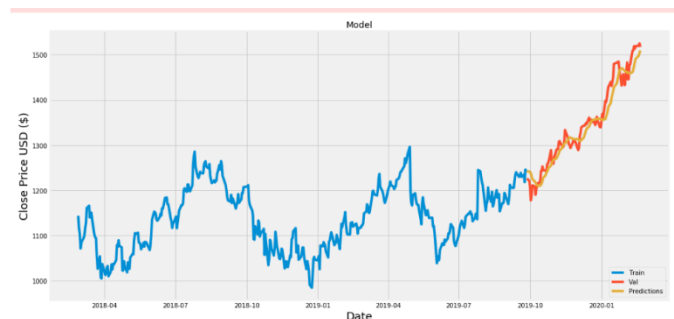
C. SVM Based Model Results

There are seven columns or seven attributes that describe the rise and fall in stock prices. Some of these attributes are (1) HIGH, which describes the highest value the stock had in previous year.(2) LOW, is quite the contrary to HIGH and resembles the lowest value the stock had in previous year (3) OPEN is the value of the stock at the very beginning of the trading day, and (4) CLOSE stands for the price at which the stock is valued before the trading day closes. There are other attributes such as VOLUME and ADJCLOSE.

Using the sklearn libraries we import SVC classifier and fit it with the training data. After training the model with

the data and running the test data through the trained model the confusion matrix obtained is shown below.
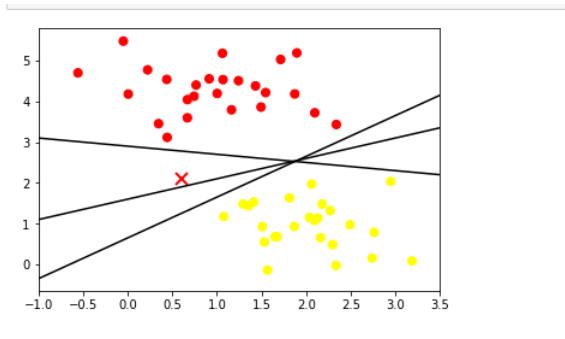

Fig. SVM Prediction

VII. CONCLUSION

By measuring the accuracy of the different algorithms, we found that the most suitable algorithm for predictingthemarketpriceofastockbasedonvariousdatapointsfro mthehistorical data is the LSTM. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

VIII. FUTUREENHANCEMENT

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporation's performance structure as a whole.

REFERENCES

[1] Ashish Sharma, Dinesh Bhuriya, Upendra Singh.
[2] "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
[3] Loke.K.S. "Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests", IEEE,2017.
[4] Xi Zhang1, SiyuQu1, JieyunHuang1, Binxing Fang1, Philip Yu2, "Stock Market Prediction via Multi-Source Multiple Instance Learning." IEEE 2018.
[5] Vivek Kanade, Bhausaheb Devikar, Sayali Phadatare, Pranali Munde, Shubhangi Sonone. "Stock Market Prediction: Using Historical Data Analysis", IJARCSSE2017.
[6] SachinSampatPatil, Prof. Kailash Patidar, Asst. Prof. Megha Jain, "A Survey on Stock Market Prediction Using SVM", IJCTET2016.
[7] https://www.cs.princeton.edu/sites/default/files/uploads/S aahil_magde. pdf
[8] Hakob GRIGORYAN, "A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA)", DSJ 2016.
[9] Raut Sushrut Deepak, ShindeIshaUday, Dr. D. Malathi, "Machine Learning Approach In Stock Market Prediction", IJPAM2017.
[10] Pei-Yuan Zhou , Keith C.C. Chan, *Member, IEEE,* and Carol XiaojuanOu, "Corporate Communication Network and Stock Price Movements: Insights From Data Mining", IEEE2018.