

Machine Learning Tools to Assist Financial Marketing

Chandan G Raj¹, Dhanush N V², Yedhunandan Kishore³, Surya Bhushan Ojha⁴, Dr. Dattatreya P Mankame⁵

^{1, 2, 3, 4, 5}Dept of Computer Science

^{1, 2, 3, 4, 5}Atria Institute of Technology

Abstract- *Technical and quantitative analysis in financial marketing make use of mathematical and statistical tools to help investors decide on the optimum moment to initiate and close orders. While these traditional approaches have served their purpose to some extent, new techniques arising from the field of computational intelligence such as machine learning have emerged to analyze financial information. In this paper, the aim is to predict the future value of the financial stocks of a company. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. Machine learning itself employs different models to make prediction easier and authentic. The paper focuses on the use of Regression based Machine learning to predict stock values. Factors considered are open, close, low, high and volume. 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, Linear Regression, Dataset, Stock Market.

I. INTRODUCTION

Finance is one of the main pillars on which an economy grows. It plays a very important role in the fast economic growth of a country. If the Financial market rises, then countries economic growth would be high. If Financial market falls, then countries economic growth would be down. The Financial markets have evolved drastically with time. In the past, predictions were simply based on available data and qualitative and quantitative assessment done by financial analysts. Today, however, with the markets being unstable and the rise of electronic trading platforms, predictions have to range from a few seconds ahead to days or months with the data that is available. Having a proper understanding of how the financial market works certifies greater financial returns. The financial 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. Small ownerships, brokerage corporations, banking sector, all depend on this very body to make revenue and divide risks. The basic idea is quite simple, companies will list their shares

as small commodities called Stocks. They do this so that they can raise money for the firm. A company lists its stock at a price called the IPO or initial public offering. This is the offer price at which the company sells the stock and raises money. After which these stocks are the property of the owner and he may sell them at any price to a buyer at an Exchange such as BSE (Bombay Stock Exchange). Traders and buyers continue selling these shares at their own price but the company only gets to keep the money made during the IPO. The continuous hoping of shares from one party to another in order to make more profits, results in an increase of price of the particular share after every profitable transaction. However, if the company issues more stocks at a lower IPO, then the market price for exchange goes down and the traders suffer a loss. This exact phenomenon is the reason why people fear in investing in stock markets. A Stock market is basically nonlinear in nature and the research on stock market is one of the most important issues in recent years. People invest in stock market based on some prediction. For predicting the stock market prices people search methods and tools which will increase their profits, while minimizing their risks. A correct prediction of stocks can lead to huge profits for the seller and the broker. The attempt to determine the future value of the stock market is known as a stock market prediction. The dependency of a stock price on various factors is unstable and it is never possible to predict the exact price of the stock. But with the utilization of technology like Machine Learning, it is possible to develop an algorithm to minimize this window of error and get the most accurate output possible. Machine Learning is a mathematical technique of finding patterns in the data provided and applying what is learnt on a new set of data and making predictions. It is an art of making the machine learn new patterns without explicitly programming the machine to do so. The basic process in Machine Learning from data gathering to obtaining predictions can be summarized as follows: initially, the user analyses the data and selects the input format which is preferred. Then, the user has to decide which machine learning algorithm will best suit the use case. Then the user makes an analytical model based on the chosen mathematical algorithm. Later, the user trains the model that he prepared using the input data chosen. Finally, the trained model is tested on other data for finding patterns, making predictions and to generate scores. The vital part of machine learning is the dataset used. The dataset should be as concrete as possible because a little

change in the data can perpetuate massive changes in the outcome. In this project, supervised machine learning is employed on a dataset obtained from Yahoo Finance. This dataset comprises of following five variables: open, close, low, high and volume. The dataset is pre-processed and tuned up for real analysis. The model is then tested on the test data. Regression model is engaged for this conjecture separately. Finally, the graphs for the fluctuation of prices with the dates and are plotted.

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.

II. PROBLEM DEFINITION

Prediction of the financial market or stocks 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 hyperparameters. 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 information about Stock market prediction mechanisms currently being used.

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 regression 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 of fit" 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 intelligence techniques to predict the prices of the stock is an increasing 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 be n 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 increased.

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 depend 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. PROPOSED SYSTEM

In this proposed system, we focus on predicting the stock values using a machine learning algorithm called Linear Regression. In this model, 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 into 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

Stock market prediction seems a complex problem because there are many factors that have yet to be addressed and it doesn't seem statistical at first. But by proper use of machine learning techniques, one can relate previous data to the current data and train the machine to learn from it and make appropriate assumptions.

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 that 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.

Regression classifier

In general, the Regression based Model is used for predicting continuous values through some given independent values. Regression uses a given linear function for predicting continuous values:

$$V = a + bK + \text{error} \quad (1)$$

Where, V is a continuous value; K represents known independent values; and, a, b are coefficients.

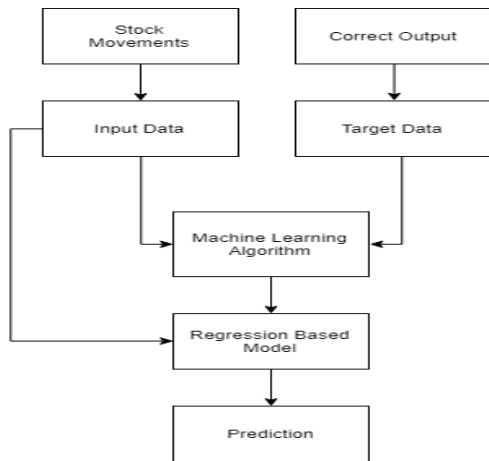


Fig. 1 Flow Chart for Regression Based Model

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. The R-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.

The functioning of the system mentioned in this paper is given below:

1. The raw data used is from Yahoo finance.
2. The attributes used for feature extraction are ‘date’ and ‘closing price’ of a stock.
3. Features used to predict the momentum of stock price of particular company are ‘stock momentum’, ‘index volatility’, ‘sector momentum’. These features are scaled.
4. The dataset is split into training and test data set.
5. The training dataset is used for model training and test dataset is used for prediction. The significance of a feature is determined using the R^2 values.
6. The values of the test data are predicted and the results are evaluated. The result is given on the basis of accuracy, confusion matrix and time required for the model used.

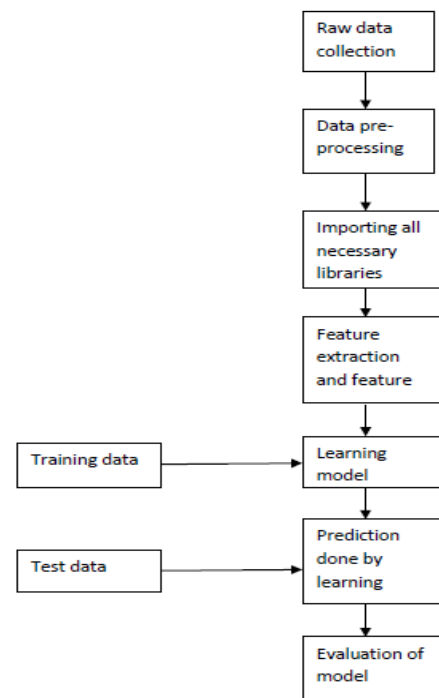


Fig. 2. Flowchart of application

VI. EXPERIMENTAL RESULTS

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 trained model:

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.

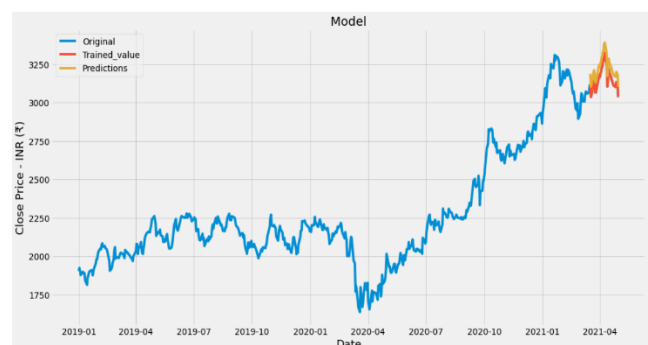


Fig. Plot between Price and Date Using Regression

VII. CONCLUSION

The aim of our paper is to help the investors for investing money in the stock market. The prediction plays a very important role in stock market business which is very

complicated and challenging process due to dynamic nature of the stock market.

In general, stock market data(s) are time-variant and are nonlinear in pattern, predicting the future price of stock is a challenging task. These types of Predictions will provide the users with good information about the current running status of stock price. With some historical data, you can use forecasting tools to predict into the future a specific metric. This paper was an attempt to determine the future prices of the stocks of a company with greater accuracy and reliability using machine learning tools. The primary contribution of the researchers being the application of the novel Linear Regression Model as a means of determining the stock prices. 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. FUTURE ENHANCEMENT

In the future, the accuracy of the stock market prediction system can be further improved by utilizing a much bigger dataset than the one being utilized currently. Furthermore, other emerging models of Machine Learning could also be studied to check for the accuracy rate resulted by them. Sentiment analysis through Machine Learning on how news affects the stock prices of a company is also a very promising area. Other deep learning based models can also be used for prediction purposes.

Various neural network techniques can be applied to the processed data to make the machine more powerful. Advanced neural network techniques can be applied to the processed dataset. The neural network techniques make use of time series to accurately predict the stock values. Related tweets from twitter can also be considered while predicting the future stock prices.

The methods mentioned in the paper can be applied to real time data to get real time predictions of the stock value. Thus it can be deployed in real world application providing a more reliable way of understanding stock price changes. Accurate graphs can be plotted for a particular company to understand the patterns of stock values and thus making it easy to understand specific patterns.

REFERENCES

- [1] Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, Lokesh Chouhan, "Stock Market Prediction Using Machine Learning", ICSCCC 2018.
- [2] Srinath Ravikumar, Prasad Saraf, "Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms", IEEE 2020
- [3] Prakhar Vats, Prof. Krishna Samdani, "Study on Machine Learning Techniques In Financial Markets", IEEE 2019.
- [4] Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
- [5] Kunal Pahwa, Neha Agarwal, "Stock Market Analysis using Supervised Machine Learning", International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 2019.
- [6] D. Bhuriya, G. Kaushal, A. Sharma and U. Singh, "Stock Market Prediction Using A Linear Regression," in International Conference on Electronics, Communication and Aerospace Technology, 2017.
- [7] Vivek Kanade, Bhausahab Devikar, Sayali Phadatare, Pranali Munde, Shubhangi Sonone. "Stock Market Prediction: Using Historical Data Analysis", IJARCSSE2017.
- [8] A. Sharma, D. Bhuriya and U. Singh, "Survey of Stock Market Prediction Using Machine Learning Approach," in International Conference on Electronics, Communication and Aerospace Technology, 2017.
- [9] Pei-Yuan Zhou, Keith C.C. Chan, Member, IEEE, and Carol Xiaojuan Ou, "Corporate Communication Network and Stock Price Movements: Insights From Data Mining", IEEE 2018.
- [10] Abhijoy Sarkar, Abhaya Kumar Sahoo, Sourav Sah, Chittaranjan Pradhan. "LSTMSA: A Novel Approach for Stock Market Prediction Using LSTM and Sentiment Analysis", IEEE 2020.
- [11] Agarwal (July 14, 2017). "Introduction to the Stock Market".
- [12] Kannan, K. Senthamarai, et al. "Financial stock market forecast using data mining techniques." Proceedings of the International Multiconference of Engineers and computer scientists. Vol. 1.2010.