International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 11 | November - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Stock Market Prediction

Pranjal Jain Namisha Patidar Nitin Singh Nandan Singh

Computer Science , AITR , Indore, India

Abstract

Stock market price data is generated in huge volume and it changes every second. Stock market is a complex and challenging system where people will either gain money or lose their entire life savings. In the finance world stock trading is one of the most important activities. In Stock Market Prediction, 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 and LSTM based Machine learning to predict stock values. Factors considered are open; close, low, high and volume.

INTRODUCTION In the recent years, increasing prominence of machine learning in various industries has enlightened many traders to apply machine learning techniques to the field, and some of them have produced quite promising results. Stock market prediction and analysis are some of the most difficult jobs to complete. This paper will develop a financial data predictor program in which there will be a dataset storing all historical stock prices and data will be treated as training sets for the program. The main purpose of the prediction is to reduce uncertainty associated to investment decision making. Stock price prediction is very important as it is used by most of the business people as well as common people. People will either gain money or lose their entire life savings in stock market activity. It is a chaos system. Building accurate model is difficult as variation in price depends on multiple factors such as news, social media data, and fundamentals, production of the company, government bonds, historical price and country's economics The probable stock market prediction target can be the future stock price or the volatility of the prices or market trend. In the prediction there are two types like dummy and a real time prediction which is used in stock market prediction system. In Dummy prediction they have define some set of rules and predict the future price of shares by calculating the average price. In the real time prediction compulsory used internet and saw current price of shares of the company. Computational advances have led to introduction of machine learning techniques for the predictive systems in financial markets. In this paper we are using a Machine Learning technique i.e., Support Vector Machine (SVM) in order to predict the stock market and we are using Python language for programming. There are two common methods to predict the stock market prices2. One among that is chartist or technical theories and the second one is fundamental or intrinsic value analysis. Proposed method is built on the principle of technical theories. Basic assumption of this theory is history tends to repeat itself. Prediction model can be applied on thehistorical data to get future trend.

As researchers have discussed in S. J. Grasmere and R. J. Shiller3 and L. Andrew and M. A. Craig4, as and when new information comes in the market stock market value varies. LSTM, short for Long Short-term Memory, is an extremely powerful algorithm for time series. It can capture historical trend patterns, and predict future values with high accuracy. With multiple factors involved in predicting stock prices, it is challenging to predict stock prices with high accuracy, and this is where machine learning plays a vital role.

2. Literature Review

In this section, we discuss related works. We reviewed the related work in two different domains technical and financial. Kim and

in built a model as a combination of artificial neural networks (ANN) and genetic algorithms (GAs) with discretization of features for predicting stock price index. The data used in their study include the technical indicators as well as the direction of change in the daily Korea stock price index (KOSPI). Idrees et al. published a time series-based prediction approach for the volatility of the stock market. ARIMA is not a new approach in the time series prediction research domain. Their work is more focusing on the feature engineering side. The only weakness of their proposed solution is that the authors did not perform any customization on the existing ARIMA model, which might limit the system performance to be improved. Nekoeiqachkanloo et al. in proposed a system with two different approaches for stock investment. The strengths of their proposed solution are obvious. First, it is a comprehensive system that consists of data pre-processing and two different algorithms to suggest the best investment portions. Second, the system also embedded with a forecasting component, which also retains the features of the time series. Last but not least, their input features are a mix of fundamental features and technical indices that aim to fill in the gap between the financial domain and technical domain.

3. Research Methodology

In this section, we discuss related works. We reviewed the related work in two different domains technical and financial. in built a model as a combination of artificial neural networks (ANN) and genetic algorithms (GAs) with discretization of features for predicting stock price index. The data used in their study include the technical indicators as well as the direction of change in the daily Korea stock

price index (KOSPI). Idrees et al. published a time series-based prediction approach for the volatility of the stock market. ARIMA is not a new approach in the time series prediction research domain. Their work is more focusing on the feature engineering side. The only weakness of their proposed solution is that the authors did not perform any customization on the existing ARIMA model, which might limit the system performance to be improved.

Nekoeiqachkanloo et al. in proposed a system with two different approaches for stock investment. The strengths of their proposed solution are obvious. . However, their work has a weakness in the evaluation part. Instead of evaluating the proposed system on a large dataset, they chose 25 well- known stocks. There is a high possibility that the well- known stocks might potentially share some common hidden features.Atsalakis and Valavanis in proposed a solution of a neuro-fuzzy system, which is composed of controller named as Adaptive Neuro Fuzzy Inference System (ANFIS), to achieve short-term stock price trend prediction. The noticeable strength of this work is the evaluation part. Not only did they compare their proposed system with the popular data models, but also compared with investment strategies. While the weakness that we found from their proposed solution is that their solution architecture is lack of optimization part, which might limit their model performance.

IMPLEMENTATION:

LSTM is a Recurrent Neural Network that works on data sequences, learning to retain only relevant information from a time window. New information the network learns is added to a "memory" that gets updated with each time step based on how significant the new sample seems to the model. Over the years, LSTM has revolutionized speech and handwriting recognition, language understanding, forecasting, and several other applications that have become the new normal today.

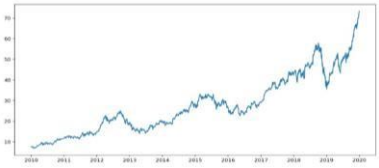
A standard LSTM cell comprises of three gates: the input, output, and forget gate. These gates learn their weights and determine how much of the current data sample should be remembered and how much of the past learned content should be forgotten. This simple structure is an

Fig.2: Stock Trend Prediction

improvement over previous and RNN model. As seen in the equations below, i, f, and o represent the three gates: input, forget, and output. C is the cell state that preserves the learned data, which is given as output h. All of this is computed for each timestamp t,considering the learned data from timestamp (t-1).

$$
\begin{aligned}
i_t &= \sigma\left(x_t U^i + h_{t-1} W^i\right) \\
f_t &= \sigma\left(x_t U^f + h_{t-1} W^f\right) \\
o_t &= \sigma\left(x_t U^o + h_{t-1} W^o\right) \\
\tilde{C}_t &= \tanh\left(x_t U^g + h_{t-1} W^g\right) \\
C_t &= \sigma\left(f_t * C_{t-1} + i_t * \tilde{C}_t\right) \\
h_t &= \tanh(C_t) * o_t
\end{aligned}
$$

The forget gate decides what information and how much of it can be erased from the current cell state, while the input gate decides what will be added to the current cell state. The output gate, used in the final equation, controls the magnitude of output computed by the first two gates. So, as opposed to standard feed-forward neural nets, LSTMs have the potential to remember or erase portions of the past data windows actively. Its feature of reading and training on windows (or time steps) of data makes its training unique.



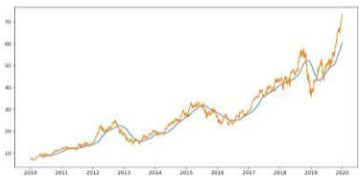


Fig.1: Price VS Time Chart

4. Data Analysis

The various modules of the project would be divided into the segments as described.

I. Data Collection

Data collection is a very basic module and the initial step towards the project. It generally deals with the collection of the right dataset. The dataset that is to be used in the market prediction has to be used to be filtered based on various aspects. Data collection also complements the dataset by adding more data that are external. Our data mainly consists of the previous year stock prices. Initially, we will be analyzing the Kaggle dataset and according to the accuracy, we will be using the model with the data to analyze the predictions accurately.

Data is gathered using following methods.

- News information is collected from 2 different websites using crawler.
- Tweets are collected using twitter API using python language.

II. Pre Processing

Data preprocessing is a part of data mining, which involves transforming raw data into a more coherent format. Raw data is usually inconsistent or incomplete and usually contains many errors. The data pre-processing involves checking out for missing values, looking for categorical values, splitting the data-set into training and test set and finally do a feature scaling to limit the range of variables so that they can be compared on common environs.object using its string representation 55 since the most significant parts of the note can be recreated using the string notation of the pitch. We 56 tokenize those string outputs to feed it into the network. For each example, we use a sequence of 57 the 100 preceding notes in order to predict the next note. We continue this "sliding window" process 58 until we have seen all notes in the file. Our model is fed these inputs of "window," note pairs. These 59 encodings allow us to easily decode the output generated by the network into the correct notes. We 60 write this processed data to our data folder and load them at time of training.

III. Training the Machine

Training the machine is similar to feeding the data to the algorithm to touch up the test data. The Training sets are used to tune and fit the

models. The test sets are untouched, as a model should not be judged based on unseen data. The training of the model includes cross-validation where we get a well-grounded approximate performance of the model using the training data. Tuning models are meant to specifically tune the hyper parameters like the a cross-validated score, for individual sets of hyper parameters. Then, we select the best hyper parameters. The idea behind the training of the model is that we use some initial values with the dataset and then optimize the parameters which we want to in the model. This is kept on repetition until we get the optimal values. Thus, we take the predictions from the trained model on the inputs from the test dataset. Hence, it is divided in the ratio of 80:20 where 80% is for the training set and the rest 20% for a testing set of the data.

IV. Data Scoring

The process of applying a predictive model to a set of data is referred to as scoring the data. The technique used to process the dataset is the Random Forest Algorithm. Random forest involves an ensemble method, which is usually used, for classification as well as regression. Based on the learning models, we achieve interesting results. The last module thus describes how the result of the model can help to predict the probability of a stock to rise and sink based on certain parameters. It also shows the vulnerabilities of a particular stock or entity. The user authentication system control is implemented to make sure that only the authorized entities are accessing the results.

5. Results

However, with the introduction of Machine Learning and its strong algorithms, the most recent market research and Stock Market Prediction advancements have begun to include such approaches in analyzing stock market data. The Opening Value of the stock, the Highest and Lowest values of that stock on the same days, as well as the Closing Value at the end of the day, are all indicated for each date. Furthermore, the total volume of the stocks in the market is provided. With this information, it is up to the job of a Machine Learning Data Scientist to look at the data and develop different algorithms that may help in finding appropriate stocks values. 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 the more will be the accuracy. The algorithms can also be applied for analyzing the contents of publiccomments 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.

6. Conclusion

The techniques that have been utilized in this paper: LSTM and Regression, on the finance dataset. Both the techniques have shown an improvement in theaccuracy of predictions, thereby yielding po sitive results. Use of recently introduced machine learning techniques in the prediction of stocks have yielded promising results and thereby marked the use of them in profitable exchange schemes. It has led to the conclusion that it is possible to predict the stock market with more accuracy and efficiency using machine learning techniques. In the future, the stock market prediction system can be further improved by utilizing a much bigger dataset than the one beingutilized currently. This would help to increase the accuracy of our prediction models. Furthermore,other models of Machine Learning could also be studied to check for the accuracy rate resulted by them.

.

7. References

[1] M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.

[2] K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017, pp. 1-1.

[3] H. Gunduz, Z. Cataltepe and Y. Yaslan, "Stock market direction prediction "

[4] M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," 2016 2nd International Conference on Electrical,

[5] H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5.

[6] K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT- 2010, Chennai, 2010, pp. 59-63.

[7] S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modeling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790.

[8] T. Gao, Y. Chai and Y. Liu, "Applying long short term momory neural networks for predicting stock closing price," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 575-578.

[9] K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," 2018 9th International Conference on Information and Communication Systems

[10] https://www.researchgate.net/publication/354820289_STOCK_MARKET_PREDICTION_SYSTEM_USING_MACHINE_LEARNING_APPROACH