Stock Prices Prediction: Survey and Research Direction

Vedant D. Shirodkar[1] Nasir J. Kadri[2] Vikas V. Yadav[3]
[1,2,3]Department of Information Technology
[1,2,3]Padmabhushan Vasantdada Patil Pratishthan College of Engineering (PVPPCOE), Mumbai, India

Abstract—Stock prices prediction is the act of determining the future prices of stock based on the current scenario of the market. Stock Market is highly volatile in nature as many factors are involved in the movement of stock price it is difficult, tedious and a challenging task to forecast exact prices of the stock. Several attempts have been made to predict stock prices. Research is been carried out with end ultimate goal to predict the stock prices more accurately. For an algorithm There are several sorts of research done as such far with the end goal to predict the stock market to complete the depicted measurements. While designing an algorithm for stock prices forecasting one should keep in mind that the algorithm should have accuracy, adaptability to learn new trends and patterns and the prediction should be carried out in less operational time. In this survey paper our aim is to analyse few well known stocks prediction algorithms. Our motive for this literature survey is to detect directions for future machine learning stock market prediction research based upon a review of current literature.

Keywords: Stock Prediction, Survey paper, SVM, HMM, Sentimental Analysis, LSTM, ARIMA, Fuzzy System

I. Introduction

Stock Market Prediction aims to predicting the future value of stocks of companies or another financial instrument. Stock prediction is a challenging problem in the field of finance but due to high financial earning it has fascinated business and academic side. Precise prediction of future prices of stock could produce significant profits; while inaccuracy in prediction may lead to huge monetary losses. For many developing countries stock market acts as a backbone for the economy.

Over years, several methods are been implemented to forecast the future prices. This paper includes some of the well- known algorithms which proved successful for predicting the stock effectively and correspondingly few remarkable paper summaries based on these algorithms with their limitations. This paper includes Support Vector Machine (SVM), Hidden Markov Model (HMM), Sentimental Model Prediction, Fuzzy System based Model and its hybrid versions, Autoregressive integrated moving average model (ARIMA) and Long Short-Term Memory (LSTM) model and its different versions. The conclusion section provides suggestions related to the overall study objective that focuses on identifying directions for future research and recommendations for improving study generalizability.

II. Literature Survey along with Prediction Algorithms

A. Support Vector Machine (SVM)

SVM is a supervised machine learning technique for classification and regression analysis. SVM offers an alternative method to make stock predictions through example categorization. SVM plots the features and tries to maximize the gap between the two categories (rise and fall) of stock price direction. New examples are classified based they belong to. Schumaker and Chen [1] used an SVM in combination with textual examination, inspecting the impact of news articles on stock prices. They created a predictive machine learning method for financial news article analysis using various discrete textual representations: Named Entity, Bag of Words, Phrases with nouns. They investigated a huge set of financial news articles and stock quotes covering stocks listed on the S&P500 during a five-week period. They estimated a discrete stock price twenty minutes after a news article was released. SVMs are sensitive towards outliers. Parameter selection also plays significant role in predictions.

B. Hidden Markov Model (HMM)

An HMM is a statistical model used for machine learning with hidden states, as opposed to a standard Markov chain where all states are disclosed. Md. Rafiul Hassan and Baikunthu Nath in [2] used HMM methodology to estimate stock price for interdepended markets. Because of HMM's proven appropriateness for exhibiting dynamic system it was used for recognizing pattern and for classification of problems. The author summarized the advantage of the HMM was strong statistical foundation. It's able to handle new data robustly and computationally efficient to develop and evaluate similar patterns. The author decides to develop hybrid system using AI paradigms with HMM improve the accuracy and efficiency of forecast the stock market. HMM requires huge volume of data to get well trained and also consumes more computational time to find the different sequence of hidden states that performs the best. Also, the accuracy is reliant on the data fed.

C. Sentimental Model

Sentimental method of stock prediction is used to find the direction of stock prices. This helps to forecast the stock prices and get the hidden indicators from the market which helps in drawing insights on real time basis. This is done with the help of information gathered from twitter data, reviews or some other primary source of information. Sentimental Analysis is applied on the data to extract customer sentiments of the stock market and thus based on the overall calculation predictions are drawn. In the paper The Effects of Twitter Sentiment on Stock Price Returns [3] by G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar and I. Mozetič proved that there exists a correlation between the stock prices and the stocks news data on Twitter. D. Shah, H. Isah, F. Zulkernine [4] in the paper predicting the Effects of News Sentiments on the Stock Market demonstrated a dictionary-based method to predict the sentiment of the news articles. Dictionary Based approach make use of predefined dictionary of words where each word is bracketed with a definite sentiment polarity strength. Sentiment of people such as happy, sad or depressed can be

found out by comparing word against lexicons from dictionaries. Sentimental Analysis gives a base for predicting the stocks in real time based on the market and could be able to track the movement of stock prices in future beforehand. This method proves to be good when coupled with some other pre-existing techniques. Sentimental Analysis when combined with data mining techniques helps in finding an unknown pattern from the data collected over news articles and which have the ability to predict the stock probably. It utilizes updated news to measure its effect on the stock in real time. The data source here plays an important role in the overall performance of the model.

D. Fuzzy System based Model

A Fuzzy c is a clustering technique where an object can belong to more than one cluster with different probabilities. A. Sheta [5] has used Takagi Sugeno (TS) technique to construct fuzzy models for two nonlinear processes, which estimated prediction of the next week S&P 500 for stock ICECA 2017. The development process of the TS fuzzy model is carried out by determining the Membership Functions (MF) in the rule antecedents with the help of model's input data and then estimating the consequence parameters. Least-square estimation was used to estimate these parameters. The end results found were propitious. Fazel Zarandiet al. [6] have built a type-2 fuzzy rule based expert system for stock price analysis. Interval type-2 fuzzy logic system allowed to model rule uncertainties and every membership value of an element was interval itself. The proposed type-2 fuzzy model applied the technical and fundamental indexes as the input variables. Fuzzy models are best suited for searching spherical shaped clusters but is very much sensitive towards the noise. The models do perform well on smaller datasets but losses its efficiency as the size of the data increases. An attempt was made of developing a Hybrid Fuzzy Model by Robert K. Lai et al. [7] which established a financial time series-forecasting by evolving and clustering fuzzy decision tree for stocks in Taiwan Stock Exchange Corporation (TSEC). The forecasting model incorporated a data clustering technique, a fuzzy decision tree (FDT), and genetic algorithms (GA) to build a decision-making system grounded on past data and technical indexes. The past data can be dissected into k sub-clusters by adopting K-means algorithm. GA was then applied to evolve the number of fuzzy terms for each input index in FDT so the forecasting accuracy of the model can be further improved.

E. ARIMA (Autoregressive Integrated Moving Average Model)

ARIMA model is fits to historical data to understand the data in better way. This helps in predicting the future points of the plotted graph and thus provides the prediction. It predicts stock using time series data as a parameter. A. A. Adebiyi., A. O. Adewumi, C. K. Ayo [8] and Hyndman and Athanasopoulos [9] developed an ARIMA model which works well for linear time series data. The linearity could be check by correlation graph. For short-term forecasting it provides sturdy and efficient prediction compared to other models. The predictions are made by the lag which looks for similar patterns in the history and predicts the future prices accordingly. ARIMA model doesn't perform well on non-linear time series. The model doesn't work for every stock as not every stock possesses linearity. The forecast result is entirely dependent on the past results.

F. Long Short-Term Memory (LSTM)

LSTM uses latch system to store the previous history patterns. This is done with the help of Recurrent Neural Networks (RNN). RNN provides a great way to predict all types of time series forecasting data as it stores the previous data back to the nodes using the self-loop. Adil Moghar and Mhamed Hamiche developed a model which predicts the stock prices using LSTM. The accuracy of this LSTM model is pretty high irrespective of the stock. S. Selvin, R. Vinayakumar, E.A. Gopalakrishnan, V. K. Menon and K. Soman [10] developed a hybrid model from LSTM using LSTM, ANN and CNN and used sliding window mechanism which serves as a lag or the future frame of scope of prediction. S. Kim, and M. Kang [11] predicted stock prices using Attention LSTM. Attention LSTM does require continuous reinforcement or backpropagation training to be effective in prediction. LSTM model is capable of analysing and draw the interconnection of patterns existing in data through a self-learning process. The lag in the LSTM serves as a memory which helps in remembering details for long time. K. A. Althelaya, El-Alfy, and S. Mohammed [12] used bidirectional LSTM for shorthand long-term stock market prediction. The main limitation of this method is that it requires lots of data for training the model.

III. CONCLUSION

The main objective of this paper was to examine different existing algorithm for stock prediction and provide direction for the future machine learning algorithms. The survey paper concludes that profit in the stock market may be maximize or minimize based on the various factors which were discussed in the previous section. An ideal algorithm should provide result more precisely, with less computational power and in short period of time. Fuzzy System and ANN proves to be more promising algorithm when compared to the available results but compared to Fuzzy system, ANN identifies new pattern more effectively but still as ANN only deals with historical data which is good enough to provide good results but not sufficient enough to catch the hidden indicators which silently affects the stock prices. These indicators also play a significant role in the prediction and needs to be addressed. News articles may fill this gap and helps one to identify the market sentiment of the particular stock. Many studies have confirmed that the sentiment (mood) analysis has a high influence on the future prices. Thus, a fusion of technical and fundamental analyses can yield a highly efficient prediction.

REFERENCES

[1] Schumaker, R. P., & Chen, H. (2009), "Textual analysis of stock market prediction using breaking financial news", ACM Transactions on Information Systems (TOIS).

[2] Md. Rafiul Hassan and Baikunth Nath, "Stock Market forecasting using Hidden Markov Model: A New Approach", Proceeding of the 2005 5th international conference on intelligent Systems Design and Application 0-7695-2286-06/05, IEEE 2005.
[3] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar and I. Mozetič, "The Effects of Twitter Sentiment on Stock Price Returns", 10.1371/ journal.pone.0138441.
[4] Dev Shah, Haruna Isah and Farhana Zulkernine, "Predicting the Effects of News Sentiments on the Stock Market", 2018 IEEE International Conference on Big Data (Big Data)
[5] A. Sheta, "Software Effort Estimation and Stock Market Prediction Using Takagi-Sugeno Fuzzy Models", InProceedings of the IEEE International Conference on Fuzzy Systems, pp.171-178, Vancouver, BC, 2006.
[6] M.H. FazelZarandi, B. Rezaee, I.B. Turksen and E. Neshat, "A type-2 fuzzy rule-based expert system model for stock price analysis", Expert Systems with Applications, Vol.36, No.1, pp. 139-154, January 2009.
[7] Robert K. Lai, Chin-Yuan Fan, Wei-Hsiu Huang and Pei-Chann Chang, "Evolving and clustering fuzzy decision tree for financial Time series data forecasting", An International Journal of Expert Systems with Applications, Vol.36, No.2, pp. 3761-3773, March 2009.
[8] Ayodele A. Adebiyi., Aderemi O. Adewumi, Charles K. Ayo, "Stock Price Prediction Using the ARIMA Model", 16th International Conference on Computer Modelling and Simulation.
[9] Hyndman, R. J. & Athanasopoulos, G., "Forecasting: principles and practice", 2014 University of Western Australia.
[10] Selvin, S., Vinayakumar, R., Gopalakrishnan, E., Menon, V. K., and Soman, K. (2017), "Stock price prediction using LSTM, RNN and CNN-sliding window model", in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE.
[11] [11] S. Kim and M. Kang, "Financial series prediction using Attention LSTM".
[12] [12] K. A. Althelaya, El-Alfy, and S. Mohammed, "Evaluation of bidirectional LSTM for shortand long-term stock market prediction", in 9th International Conference on Information and Communication Systems (ICICS).

