OPEN ACCESS

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-5, June 2020

Different Techniques used in Stock Market Prediction

Sai Shourie, Sharadadevi S Kaganurmath

***Abstract**: The stock market has been one of the primary revenue streams for many for years. The stock market is often incalculable and uncertain; therefore predicting the ups and downs of the stock market is an uphill task even for the financial experts, which they been trying to tackle without any little success. But it is now possible to predict stock markets due to rapid improvement in technology which led to better processing speed and more accurate algorithms. It is necessary to forswear the misconception that prediction of stock market is only meant for people who have expertise in finance; hence an application can be developed to guide the user about the tempo of the stock market and risk associated with it.The prediction of prices in stock market is a complicated task, and there are various techniques that are used to solve the problem, this paper investigates some of these techniques and compares the accuracy of each of the methods. Forecasting the time series data is important topic in many economics, statistics, finance and business. Of the many techniques in forecasting time series data such as the Autoregressive, Moving Average, and the Autoregressive Integrated Moving Average, it is the Autoregressive Integrated Moving Average that has higher accuracy and higher precision than other methods. And with recent advancement in computational power of processors and advancement in knowledge of machine learning techniques and deep learning, new algorithms could be made to tackle the problem of predicting the stock market. This paper investigates one of such machine learning algorithms to forecast time series data such as Long Short Term Memory. It is compared with traditional algorithms such as the ARIMA method, to determine how superior the LSTM is compared to the traditional methods for predicting the stock market.*

***Keywords** : Recurrent Neural Networks, Long Short Term Memory, Autoregressive integrated moving average .*

I. INTRODUCTION

The stock market is very unpredictable, and investing in it is involves one of the perilous decisions that can be made by any person, as it can return total loss or very high profit as well, it depends on the investor's acumen in investing stocks that can provide profit. It is imperative to reduce the scope of human error in such decision making situations, so that profit can be increased. The chartist theories suggest there is always some hidden data in historical prices of stock market, and researches have shown that future historical data can predicted by the past information, it supports the theory that historical data have a sound predictive ability. With the application of technologies like machine on a dataset of stock market decreases the risk involved by informing the users about the complexities of the decision he or she is about to make while purchasing stocks. The purpose of this paper is to find out which forecasting method offers better predictions with respect to lower prediction error and higher accuracy.

Revised Manuscript Received on May 15, 2020.
* Correspondence Author

Sai Shourie*, Dept. of Computer Science & Engineering, RV College of Engineering, Bangalore, India, Email: shouriesai7@gmail.com

Sharadadevi S Kaganurmath, Dept. of Computer Science & Engineering, RV College of Engineering, Bangalore, India, Email: sharadadeviks@rvce.edu.in

II. TECHNIQUES TO IMPLEMENT STOCK PREDICTION

A. Recurrent Neural Networks

Recurrent neural networks are one of the effective techniques for replicating a model based on previous information. Long Short-Term memory is a variant of RNN architecture. LSTM includes a unit of computation called as the memory cells; it replaces the artificial neurons which are in the various layers of the network. The network is able to effectively associate memories and input, therefore it can remember and take into consideration the past data and associate it better with the input data dynamically.

Methodology:

- The raw historical data is collected from the dataset, in which 30% is used for testing the model and remaining 70% of data is used for training the model.
- The collected raw data is preprocessed which includes:
  - Data discretization: is data reduction in which continuous data is discretized into intervals without losing considerable amount of information.
  - Data transformation: the data is normalized.
  - Data cleaning: The absent values are filled with a default value.
  - Data integration: After cleaning the dataset, it is divided into testing and training sets, in which 70% is used for training and 30% is used for testing.
- Feature Extraction: The data might contain features that the model doesn't require for predicting the stock value; hence a few features which are required by the model are fed into it.
- Training the model: In this stage, The model is assigned random weights and biases, and based on the error caused by the training data that is fed to the model as input, the weights are adjusted accordingly to the

reduce the error in each iteration. The model that is used consists of a

- Sequential input layer
- Hidden layer having three LSTM layers
- Dense layer with activation
- Dense output layer with linear activation function

- Generation of Output: The target value is compared with the value given by the output layer of LSTM. The error produced which is the difference between the target value and output value is reduced by using back propagation algorithm which modifies the biases and the weights of the network accordingly.

The conversion rate of the algorithm greatly depends on the type of optimizer used, and determines how fast a algorithm can reduce the error to minimum. The algorithm should not lead to a local minimum, it should try to reach global minimum to give best performance and hence there is should an element of randomness included in the algorithm. Therefore taking these requirements into consideration, the chosen algorithm to use is the Adam optimizer. The Adam optimizer incorporates the features of both the RMSprop optimizer and the ADAgrad optimizer optimizer.

B. ARIMA Model

Box-Jenkins methodology is another term used for ARIMA model; it consists of collection of methods for associating, estimating and diagnosing ARIMA models with sequential data. This model is one of the efficient methods in time series forecasting. It consistently gives better performance than other short term complex structural models. ARIMA models are effective enough to get predictions with relatively smaller datasets. In ARIMA model, the output is a linear amalgamation of previous errors and previous values, denoted as

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \quad (1)$$

Where e_t is the error at t and Y_t is the value to be predicted, $\square_j$ and $\square_i$ are the coefficients. q and p are integers that are often referred to as moving average and autoregressive. The steps involved in constructing ARIMA predictive model are model identification, diagnostic checking and parameter estimation. The ARIMA model is the combination of the following methods:

- Auto Regression: In auto-regression, the values of a given time series data are regressed on their own lagged values, which are denoted by the "p" value in the ARIMA model.
- Differencing: This involves converting the non-stationary time series data into stationary time series data; trend is removed by differencing the time series data.
- Moving Average: It represents the number of regressed values of the error term, it is denoted by value 'q'.

Methodology:

- Testing for Stationarity:
  - To fit time series data into the ARIMA model, the prerequisite is that the series should be stationary, which means that the time series data should not contain a trend; therefore it should have constant mean and variance, to increase the accuracy of prediction.
  - To make sure that the time series data does not have a trend, it has to be tested for stationarity. The test to be performed is the Augmented Dickey-Fuller unit root test. For the time series data to be stationary, the value returned from the Augmented Dickey-Fuller unit root test should be less than 0.05. The time series data is considered to be non-stationary if the value returned from the test is greater than 0.05.
  - The differencing method is applied on the time series data to convert it from non-stationary to stationary time series. Calculating the differences between successive numbers of a sequential data is called differencing. The new series that is formed by values from differencing the time series data is tested to determine new statistical properties and other correlations.
  - The first order differences and the second order differences could be derived by consecutively applying the differencing method. It is repeatedly applied until the time series data becomes stationary.
- Identification of q and p
  - After testing for stationarity, the parameters p and q have to be identified, which determines the order of Autoregressive and the order of moving average. And it is identified by the Autocorrelation function and Partial Autocorrelation function.
  - For Autoregressive models, the Autocorrelation function will decrease considerably and the order (p) of the Autoregressive model can be derived by the Partial Autocorrelation function. The order of Autoregressive model is 1, if there is one noticeable spike at lag 1. The order of Autoregressive model is 3, If there are noticeable spikes at lag 1,2,3 on the partial Autocorrelation function.
  - For Moving average models, the Partial Autocorrelation function will decrease considerably. The order of moving average can be derived by the Autocorrelation function. The order of moving average is 1, if there is a considerable spike at lag 1. The order of moving average is 3, if there are considerable spikes at lag 1,2,3.
- Estimation and Forecasting
  - As the parameters (p,d,q) are derived, the ARIMA model is applied on the training data, and the accuracy of ARIMA is determined, and then the trained model is used to predict the outputs of the future values. Then, the error is determined by comparing the predicted values to the forecasted values.

III. RESULT

The dataset containing financial information time series date from 1985 to 2018 is used, in which 70% is training data and 30% is testing data. When tested with both the algorithms, it was observed that the average mean square error produced is 511.481 and 64.213 by ARIMA and LSTM.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The better performance by LSTM is due to iterative optimization used, i.e. obtaining the result multiple times and selecting the optimal one, iteration minimizes errors, it helps transform an under fitted model to optimally fitted model.

LSTM provides better in the case of dealing with large datasets and if required amount of training data is available, but in the case of smaller datasets AMIMA is better, ARIMA requires parameters such as (p, q, d) which are calculated based on the data, while LSTM does not need setting such parameters. But LSTM requires some hyper parameters for tuning it.

IV. CONCLUSIONS

This paper presents a general study and comparison on Stock Market prediction techniques used by various software applications. When the LSTM and ARIMA techniques were implemented and tested, it is concluded that LSTM has better accuracy than ARIMA. This paper advocates the advantages in using deep learning algorithms for financial applications such as stock market prediction. In this paper, forecasting technology such as Long Short-Term Memory unit, ARIMA are used which helps end users, finance experts by giving them information about the stock market.

REFERENCES

1. K. Senthamarai Kannan, P. SailapathiSekar, M.MohamedSathik and P. Arumugam, "Financial stock market forecast using data mining Techniques", 2010, Proceedings of the international multiconference of engineers and computer scientists.
2. JingTao YAO and Chew Lim TAN, "Guidelines for Financial Prediction with Artificial neural networks".
3. Tiffany Hui-Kuangyu and Kun-Huang Huarng, "A Neural network-based fuzzy time series model to improve forecasting", Elsevier, 2010, pp: 3366-3372.
4. Md. Rafiul Hassan and Baikunth Nath, "Stock Market forecasting using Hidden Markov Model: A New Approach", Proceeding of the 2005 5 th international conference on intelligent Systems Design and Application 0-7695-2286-06/05, IEEE 2005.
5. Ching-Hsuecheng, Tai-Liang Chen, Liang-Ying Wei, " A hybrid model based on rough set theory and genetic algorithms for stock price forecasting", 2010, pp. 1610-1629.
6. M.H. Fazel Zarandi, B. Rezaee, I.B. Turksen and E.Neshat, "A type-2 fuzzy rule-based experts system model for stock price analysis", Expert systems with Applications, 2009, pp. 139-154.

AUTHORS PROFILE

Sai Shourie is an Undergraduate Scholar pursuing Computer Science & Engineering in R.V. College of Engineering, Bangalore. He is passionate about Computer Networking and Machine Learning and has a keen interest on pursuing these in his higher studies.. He is currently working under the mentorship of Sharadadevi S Kaganurmath.

Sharadadevi S Kaganurmath is a Professor in R.V. College of Engineering Bangalore. She has over 8 years of experience in teaching. Her primary interests lie in Wireless Sensor Networks. She has 5 publications in International journals and conferences.

Retrieval Number: E9275069520/2020©BEIESP
DOI: 10.35940/ijeat.E9275.069520
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication