A Comparative Study of Dynamic Algorithms for Prediction of Indian Stock Markets

Prof. Vaishali Khairnar[1], Ganesh Mestry[2]

[1]Professor, Department of Information Technology, Terna Engineering College, Maharashtra, India
[2]Research Scholar, Terna Engineering College, Maharashtra, India

---

Abstract – *Stock market prediction is a crucial and challenging task due to its nonlinear, evolutionary, complex, and dynamic nature. Exploration on the financial exchange has been a significant issue for analysts lately. Foreseeing the securities exchange pattern precisely will limit the danger and bring a greatest measure of benefit for every one of the partners. Any little changes in the framework can deliver compound mistakes in anticipating what's to be conducted in the framework. Hence, predicting the stock market trend accurately will minimize the risk and bring maximum amount of profit for all the stakeholders. During the last several years, a lot of studies have been done to predict stock market. Also in the course of the most recent couple of years, many machine learning calculations have been utilized trying to gauge stock costs. The forecast of stock market assists us with putting admirably in the financial exchange. The securities exchange is anticipated based on essential investigation and specialized examination. There are unique expectations techniques like KNN, SVM, Artificial Neural Organization, datamining; and Fuzzy framework are present to foresee the trading of stocks. Here we are breaking down these techniques to discover the proficient securities exchange expectation model and to anticipate the stock development with higher exactness. This study will help the researchers and specialists in choosing calculations that can be valuable for a foreseeing the stock market. An overview of different calculations and its boundaries for stock exchange forecast is introduced in this paper. This paper assesses the viability of a kind of Repetitive Neural Network known as Long Short Term Memory (LSTM), SVM, and ARIMA model to execute examinations.*

Key Words: ARIMA, Forecast analysis, Long Short Term Memory, Stock Exchange, SVM, Time Series Analysis.

I. INTRODUCTION

The Stock market assumes a crucial part in the country's monetary development just as the individual economy generally. Figuring out the correct opportunity to purchase and sell the offers is subject to anticipating the patterns in the stock market. The strategy for most precise expectation is to gain from past cases and plan a model to do this by utilizing customary and AI calculations [1]. The Stock market pattern shifts because of a few factors like political, financial matters, climate, society, and so forth [2][3]. There are two kinds of stock examination. One is Fundamental Analysis, which requires investigation of the organization's rudiments, for example, asset report, costs and incomes, yearly returns, organization's profile, and position, and so on. The other one is a Technical Analysis, which manages contemplating the insights created by market exercises like verifiable information, past cost, and volumes [4]. There are two fundamental speculations utilized in customary methodology for securities exchange forecast to be specific Efficient Market Hypothesis (EMH) presented by Fama in 1964 which expresses that stock value future is unusual dependent on the recorded information [5] and Random Walk Hypothesis (RWH) which expresses that stock's future cost is autonomous of its set of experiences. Fundamental analysis is tedious and time consuming and is required while making long term investments. Technical analysis is preferred due to its ability to forecast for short term profits and help the traders in making an informed decision before actually investing. In the past few years many machine learning methods including artificial neural network based, evolutionary and optimization based techniques and various hybrid models have been developed in an attempt to forecast stock market movement. The use of neural network based models has been many popular with highest majority of models using lagged index data of some form. The objective of this work is to study the suitability of a recurrent neural network called Long Short Term Memory (LSTM), SVM and ARIMA model in making predictions about end of day closing price of a stock by using technical analysis.

The rest of this paper is organized as follows. Section II provides a description about the used materials and methods followed by the literature survey. Section III describes the working of the proposed methodology which includes data collection, pre-processing and other implementation details. Section IV contains the comparative analysis of simulations and Section V concludes the paper with findings indicative of the results.

II. MATERIALS AND METHODS

The stock market is a center or hub which allows traders to trade equities, shares, debentures and securities. The transactions can be done both in a physical manner as well as virtual (online) manner. The stock exchange is an integral part of a nation's economy because it is through this place that new investments come into a company. The company can grow, expand and generate revenue which in turn benefits the nation as a whole. It also signals whether there is an investment-friendly atmosphere prevalent or a sluggish bear-like situation in the country. Forecasting the stock market is difficult, challenging and essential. An assumption needs to be made based on which stock market prediction can be done. That is the data that is available in public domain has some relationship capabilities amongst themselves which will help in predicting the future trends of the stock market. Individual investors always search for mechanisms of profit-making in the stock market which will enable them to make low-risk investments and get maximum return of investment. This is the biggest motivation behind this kind of research work which deals with developing stock market predictive models. Figure1 depicts a typical Google stock price prediction graph.

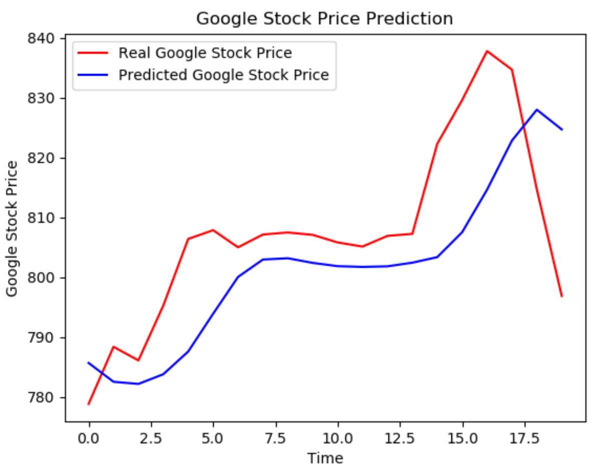

Figure1: Google stock price prediction

A. Related Work

This section discusses the similar work for stock market prediction, implemented by various authors using different algorithms. There are many datasets that can be used for data related to stocks. Cao and Tay [6] used the S&P 500 dataset. Time series forecasting is an essential area in which variables past observations are collected and analyzed for development of the model, to describe the underlying relationships, so that the model can be used to predict the future. ARIMA is one of the widely used time series models due to its statistical properties, and it can represent various types of time series like Pure Autoregressive (AR), Pure Moving Average (MA) and combined AR and MA (ARMA) series. Amin Hedayati, Moein [7] in their study investigated the ANNs ability to forecast the stock market. The daily NASDAQ Dataset was used for investigation. Two types of input dataset was used one was four prior days and another was nine prior days. The Paper discussed the various Training and Transfer functions and 16 different network architectures using the same were built and the two datasets were tested. The author concluded that there is no difference between the prediction ability of the four and nine prior working days as input parameters.Osman and Mustafa [8] used the same dataset to predict the stock market price using Particle Swarm Optimization (PSO) and Least Square Support Vector Machines (LS-SVM). Traditional SVM was reformulated, and the LS-SVM algorithm was presented. LS-SVM comprised of least square function with equality constraints, leading to a linear system which meets the KKT conditions to get an optimal solution. The methodology to accurately select the LS-SVM must be robust against noise and will not need prior user knowledge about the free parameter influence.

B. Types of Algorithms

- **LSTM:**

The neural network used in this paper was Long Short Term Memory (LSTM) which is a type of Recurrent Neural Network (RNN) [9]. A RNN is a type of neural network that has the form of a simple neural network structure which is repeated in loops allowing information to travel from one state to another while also making the network structure less complex. However, one of the biggest drawbacks of RNNs is that in a time-series forecasting problem where data will have long-term dependencies the RNN does handle the long-term dependencies well as the gap between the information under consideration and the instance where it is needed to becomes very large, making the network incapable of connecting information. LSTMs are a kind of RNN which is capable of learning long-term dependencies. In standard RNNs, the repeating module consists of a very basic structure, such as a single tanh layer. LSTMs also have this looping structure, but the repeating module instead of having a single neural network layer, has four collaborating in a special way. Figure2 illustrates an LSTM cell structure. A common architecture of LSTM is comprised an input gate, an output gate, a forget gate and a memory cell. The hidden layers are briefly described below:

- *Cell:* it stores a value (or state), for either a defined period of time. Removal and addition of information from a cell is regulated by gates.
- *Input Gate:* controls the degree to which a new value flows into the cell. Input activation function is tanh.
- *Forget Gate:* controls the degree to which a value remains in the cell. Has a sigmoid activation.
- *Output Gate:* output gate controls the degree to which the value in the cell is used to calculate the output of the LSTM unit.

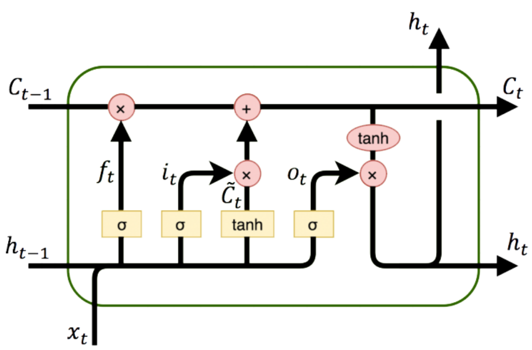

Figure2: LSTM cell structure

- **ARIMA Time Series Model**

The ARIMA (Auto-Regressive Integrated Moving Average) Model is generally used for analysis of time series data and prediction future values of the same time series. Essentially, Time Series is a sequence of numerical data obtained at regular time intervals (frequency) over a period of time. In other words it is a sequence of data a variable assumes over a period of time at fixed intervals. In a univariate time series there is a single variable, whose data changes over regular time intervals. The principal components of a Time Series are –

- Trend,
- Seasonality,
- Cyclicity
- Randomness/Error.

Trend is a general direction in which the series is moving. Trend is the increase or decrease in the series over a period of time; it persists over a long period of time. It is either an upward trend or downward trend. Seasonality is a regular pattern of up and down fluctuations. It generally happens over a short span of time, i.e. within a year. Cyclicity is medium-term variation caused by circumstances which repeat in irregular intervals. Its frequency is longer than seasonal (more than a year). Randomness / Error refer to variations which occur due to unpredictable patterns and do not repeat in any particular pattern. After the 1st three components are taken out of the series we are left with the randomness. A time series can be decomposed into its components – trend, seasonality and randomness. The decomposition can be of two types: Additive decomposition and Multiplicative decomposition. Figure3 represents a standard architecture of the time series model.

- **SUPPORT VECTOR MACHINES**

Support vector machine with adaptive parameters [10] deal with the application of financial time series forecasting. It used the five real futures contracts collated from the Chicago Mercantile Market as datasets. It increases the feasibility by comparing the result from SVM with the multilayer Back-Propagation (BP) neural network and Radial Basis Function (RBF). The adaptive Parameter is proposed by incorporating the financial datasets into the SVM. Though the prediction is satisfactory, it needs several improvements.

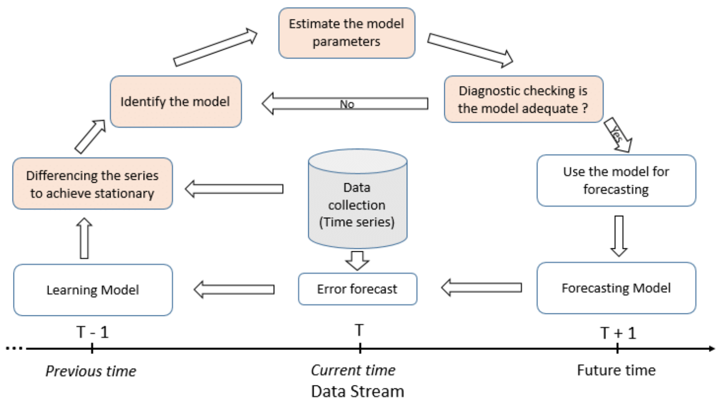

Figure3: ARIMA Architecture

III. PROPOSED METHODOLOGY

- ***Data Collection and pre-processing:***

Historical data containing prices for Apple Inc. (AAPL/NASDAQ) was obtained from Quandl. Quandl API was used in the code and free historical data for AAPL ticker was obtained from the WIKI Prices database. The database contains all stock data including open, high, close, adjacent close, etc. but in our case only 'close' value was chosen as a feature to train the model. The data contained historical data of closing prices for Apple stocks ranging from 01 Jan 2013 to 01 May 2018. Data was scaled into a range of (0,1) and (1,1) in case of every output layer activation by Min-Max scaling before being input to the LSTM. Scaling was done to make the training faster and also facilitate the use of three popular output layer activations namely linear, sigmoid and tanh. Two ranges of scaling were considered to measure the performance of model in both cases. Data was divided into training and test sets by a ratio of 60:40. An option of using a cross‑validation set to measure the error of the back‑propagation algorithm after every iteration was provided by taking a 10% split off the training set. The input i.e. training set was reshaped into a format of [samples, time steps, features]. Figure4 represents the comparison between actual and predicted prices of Quandl.

Figure4: Actual prices Vs predicted prices of Quandl

- ***Training detail:***

The model is composed of a single input layer which takes in the input scaled historical data in three dimensional array form, followed by a dense layer which is a regularly connected Neural network layer. Following the dense layer, a dropout layer [7] with a fraction rate of 0.25 is added to avoid overfitting of data in the model by ignoring randomly selected neurons during training. Dropout of 0.25 implies that 1 out of every 4 inputs will be randomly ignored after each update cycle. Finally, a single output layer was added to the network to make predictions for the closing value of stock data. The network was trained by stochastic gradient descent using the Adaptive moment estimation (Adam) optimizer with a learning rate = 0.001, decay rate for 1st order moment estimates 1 = 0.9 and exponential decay for 2nd order moment estimates 2 = 0.999. The loss function in this case was Mean absolute error (MAE).

- ***Stopping Criterion:***

The rate of convergence for the backpropagation can be controlled by the learning rate. A larger value of would ensure faster convergence, however it may cause the algorithm to oscillate around the minima, whereas a smaller value of would cause the convergence to be very slow. We needed to have some stopping criterion for the algorithm as well, to ensure that it does not run for more epochs than required. For this experiment, we used a two-fold stopping criterion.

SVM, a supervised machine learning algorithm, can be used for both regression and classification problems. This algorithm uses a kernel trick technique that transforms the data and then finds the optimal boundary between outputs. Moreover, SVM shows that it can perform well on non-linear dataset problems, based on the kernel we choose in training SVM model. SVM have been widely used for stock market prediction. In our SVM model, we have tried different kernel algorithms tuning parameters for each model: Linear, Polynomial and RBF. We have trained and tested this model on our training and testing datasets generated. The output is the binary value, 0 when yesterday close price goes down with respect to today close price and 1 when the price goes up. We used scikit-learn library to build this model and we have trained the model and applied GridSearchCV to choose the best parameters to fit our model.

SVR is the same as SVM; however it is used for regression instead of classification. It uses same terms and functionalities as SVM to predict continuous value. In this model, we follow the same process of SVM except for the output, which is not a class, rather end-of-day price.

Prediction models incorporate multiple factors news, social media factor, historical price to increase the accuracy of the result, it built on the principle of technical theories two model was built here as part of business work. Supervised machine learning model is used to build the models. Here the history prices are combined with sentiments. The output of the model is based on the correctness of the input.

SSA and SVM model is used the Shanghai Stock Exchange (SSE) composite Index. It decompose the original stock index into the trend terms, the market fluctuation terms, the noise terms and time series with different economic features. The index value are passed as input to the SVM for prediction and SSA can better grasp the features of original index series than the EEMD, while the SSA-SVM combination have better predictive effect than that of the EEMD-SVM combination prediction.

IV. CONCLUSION

This paper aims to study the stock market prediction using multiple Traditional, Machine learning, and Deep learning algorithms. Along with the algorithms, the survey has focused on various datasets used for stock market prediction, features of these datasets selected as input parameters and the evaluation metrics used for comparing the results of predictions.

Sr. No	Dataset Used	Algorithms used	Evaluation Metrics Used	Features
1	S&P 500	MLP-BP SVM	NMSE MAE DS CP CD	Days, Relative Frequency
2	S&P 500	NN-BP LS-SVM LS-SVM PSO	MSE	Days, Closing Price
3	The Wolf's sunspot data the Canadian lynx data British Pound/US dollar exchange rate data	ARIMA ANN Hybrid	MSE MAD	Days , Price
4	NASDAQ DJIA STI	Component- based three- layer feed forward neural network	One-step Sign prediction rate MSE	Trading Volume, Price
5	NASDAQ	MLP DAN2 GARCH	MSE MAD	Day, Price
6	Karachi, London and New York stock exchange	KNN SVM Naïve Bayes	MAE RMSE Accuracy	Date, Open, Low, High, Close, Volume, Trend, Sentiment & Future trend value
7	DJIA	RNN-LSTM CNN	Mean Accuracy	Price and Date , News
8	NSE,NYSE	MLP RNN LSTM CNN	MAPE	Closing Price , Days
9	BSE	PSO LSM	MAPE	Closing Values, Days

Table1: Comparative Analysis of Algorithms used

V. REFERENCES

[1]. Lijuan Cao and Francis E.H. Tay, "Financial Forecasting Using Support Vector Machines", 2001 Springer-Verlag London Limited.

[2]. Jinqi Tang, Xiong Chen, "Stock Market Prediction Based on Historic Prices and News Titles", ACM ISBN 978-1-4503-6432- 4/18/05.

[3]. Billah, M., and Waheed, S. 2017. "Stock market prediction using an improved training algorithm of neural network", International Conference on Electrical, Computer and Telecommunication Engineering. pp. 1-4. IEEE.

[4]. Nischal Puri, Avinash Agarwal, Prakash Prasad, "A Survey on Machine Learning Approach for Stock Market Prediction", Helix Vol. 8(5): 3705- 3709.

[5]. Chang Sim Vui, Gan Kim Soon, Chin Kim On, and Rayner Alfred, "A Review of Stock Market Prediction with Artificial Neural Network (ANN)", 2013 IEEE International Conference on Control System, Computing and Engineering, 29 Nov. - 1 Dec. 2013, Penang, Malaysia.

[6]. Lijuan Cao and Francis E.H. Tay, "Financial Forecasting Using Support Vector Machines", 2001 Springer-Verlag London Limited.

[7]. Amin Hedayati Moghaddam, Moein Hedayati Moghaddam, Morteza Esfandyari ," Stock Market Index Prediction Using Artificial Neural Network",2077-1886© 2016 Universidad ESAN, Elsevier.

[8]. Osman Hegazy , Omar S. Soliman , and Mustafa Abdul Salam, "A Machine Learning Model for Stock Market Prediction", International Journal of Computer Science and Telecommunications volume 4, Issue 12, December 2013.

[9]. S.Hochreiter and J. Schmidhuber, "Long short-term memory", Neural Computation, Volume 9, Issue 8, 1997, p.1735-1780.

[10]. Cao, L. J., & Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. IEEE Transactions on neural networks, 14(6), 1506-1518.