A Comparative Analysis of Machine Learning Algorithms in Stock Prediction

A. Sravani

Computer Science and Engineering, GITAM Institute of technology, Visakhapatnam, Andhra Pradesh, India
sandavar@gitam.edu

Chintam Anusha

Computer Science and Engineering, Vignan's Institute of Information technology(A), Visakhapatnam, Andhra Pradesh, India
anusha.rohini07@gmail.com

N.V.S. Shankar

Mechanical Engineering, Vignan's Institute of Information technology(A), Visakhapatnam, Andhra Pradesh, India
Shankar.publications@gmail.com

Abstract

In order to earn more money in less time in this pandemic period, the ultimate option is to invest some amount in the stock market. If we invest more then we will have more profit whenever we invest in a good company. In Stock exchange, the goal is to understand the future worth of the economic stock. The recent trend in stock market prediction innovations is making use of machine learning that makes forecasts based up on the worth's of present stock exchange indices by training on their previous values. Our work analyzes machine learning algorithms and also say the best algorithms for predicting stock values. Also comparing results of four algorithms namely Linear Regression, LSTM, k –nearest neighbors, fb-prophet algorithms. Factors considered are open, close, high, date and last. Furthermore, the proposed work examines the use of the prediction system in real-world settings and also problems related to the precision of the overall worth are given, also provides a machine-learning model to forecast the long life of stock in a open market. The effective forecast of the stock will certainly be a excellent possession for stock exchange organizations as well as will certainly provide real-life solutions stock capitalists encounter.

Keywords
Liner regression, LSTM, Stock Market forecast, KNN

1. Introduction
Stock exchange prediction is the act of trying to identify future value of a business stock. The successful forecast of a stocks future rate might yield considerable earnings. The stock exchange is combination of both customers and vendors. The trying to identify the future worth of a stock exchange is stock exchange forecast. The prediction must be expected to be durable, accurate and efficiency [2] The system must be adapted to real world establishing and real-world circumstances. The system is anticipated to take into account of all variables that may affect the supply's value and also efficiency. There are various methods and methods of carrying out the forecast system like artificial intelligence, technological evaluation and time collection aspect structuring. Artificial intelligence includes artificial intelligence which permits the system to discover and boost from previous experiences without being configured time to time. The supply rate movements over a long period of time typically develop a straight contour.[6] Individuals pick to buy those stocks whose costs are expected to increase in the future. As a result of upside and also drawback movements of stock price, usually people are afraid to purchase stocks. Hence, we need to precisely anticipate stock exchange price which can be utilized in the real-world situation. Anticipating methods like direct regression, long short-term memory, fb prophet, k nearest neighbor is made use of in this task.

2. Literature Survey

Among the substantial financial subjects that have fascinated the researcher's focus for years is anticipating the stock returns. Investors in the securities market have been trying to uncover an answer to estimate the stock fads in order to decide the better timing to purchase or market or hold a share. Projecting the stock patterns have actually been done both on qualitative analysis and also quantitative analysis. There are numerous statistical models available for anticipating stock patterns as well as picking a proper model for a specific projecting application relies on the style of the information [5]. Cheng, Li-Chen, Yu-Hsiang Huang, and et. Al, Prediction of stocks is complex due to dynamic, complex, and chaotic environment of the stock market has gained quality recently in neural computational translation, little focus on deep learning models [7]. L. Y. Wei et. Al examines the forecasting performance of stock data obtained from New York Stock Exchange. The empirical results obtained reveal the superiority of neural networks model over ARIMA model. The findings further resolve and clarify contradictory opinions reported in literature over the superiority of neural networks and ARIMA model and vice versa [8]. C. M. Anish and et. Al described the application of conventional time series models, most of these models uses a single variable for forecasting, but there are many noises involutedly in raw data that are caused by changes in weather conditions and environments for daily patient number forecasting [9].

3. Methodology

In this work we advise a prediction variation for the time collection stock market info. This model will automate the treatment of modification of stock price indices based upon technological evaluation along with deals assistance for financial specialists to select the far better timing for acquiring along with offering supplies. Information mining methods are used to create the forecast design as well as additionally Python programs language is used for visualization of end results. It takes previous well worth's right into factor to consider and additionally forecast. By computing the rmse values of exactly how specifically the design forecasts the activity as well as also it is important requirement for fit if the significant purpose of the version is forecast.

Figure 1 represents the proposed model overview comprises of various stages. The first stage is concerned with understanding the objectives of the system. The second stage deals with data collection, data collected from tata global beverages. The third stage deal with data preprocessing that is preparing final data set from the previous stage. The data set is given to Data Handling such that identification, specification evaluation and also projecting are done and also duplicated until an appropriate model is identified for prediction. This is given to supply experts to examine the future stock patterns value. Forecast additionally offers a considerable standard for organizations that have a lasting perception of actions. After this the results are plotted either by line graphs or bar charts. After visualization, discover the correlations to obtain the short-term predictions. Figure 2 depicts the detailed architecture of the system. It involves the following modules.
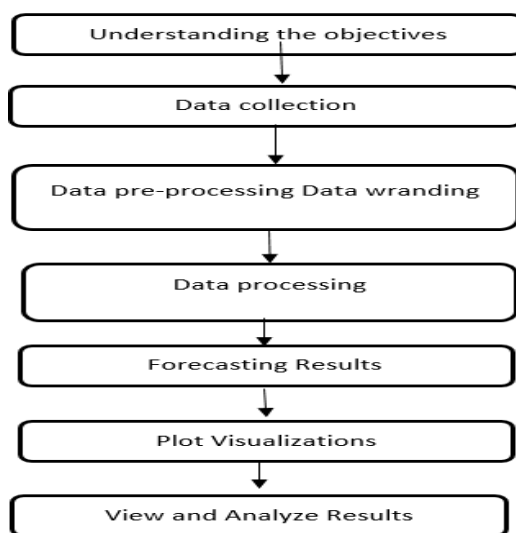

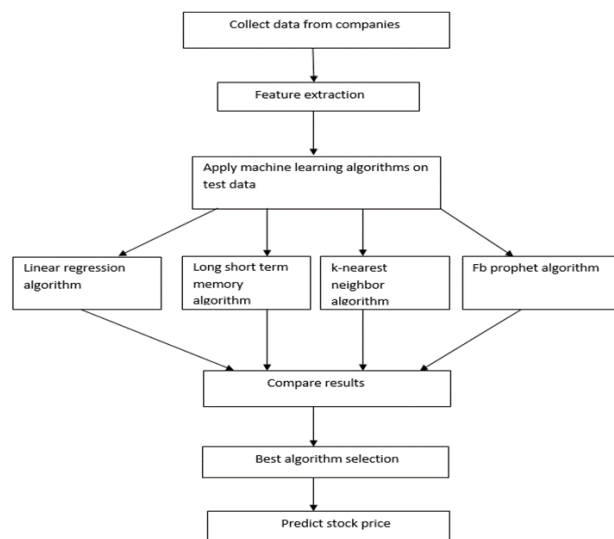Figure 1: Overview of the proposed work

Figure 2: Detailed architecture of the system

3.1 Collected Data

The dataset is collected from Quandl (you can find historic information for various supplies right here) as well as for this work, we have used the information for Tata Global Beverages is represented in Figure 3.

	Date	Open	High	Low	Last	Close	Total Trade Quantity	Turnover (Lacs)
0	2018-10-08	208.00	222.25	206.85	216.00	215.15	4642146.0	10062.83
1	2018-10-05	217.00	218.60	205.90	210.25	209.20	3519515.0	7407.06
2	2018-10-04	223.50	227.80	216.15	217.25	218.20	1728786.0	3815.79
3	2018-10-03	230.00	237.50	225.75	226.45	227.60	1708590.0	3960.27
4	2018-10-01	234.55	234.60	221.05	230.30	230.90	1534749.0	3486.05

Figure 3: Sample Data

3.2 Feature extraction:

It is a process of dimensionality reduction through which a beginning collection of raw information is minimized to even more manageable teams for handling. A normal of these large information collections is a large number of variables that call for a great deal of computing resources to process.

3.3 Training Data & Testing Data:

To start with, we divided the data into 2 parts which is training and also testing. Then we utilize training part for starting evaluation and also define the version.80% data is utilized for training function as well as 20% made use of for testing objective. Training the equipment is like to feeding the information to algorithm to touch up the test data. The training sets are utilized to change and also fit the models. Test information is used is to see how well the maker can predict new services based on training.

3.4 Algorithms used

In this work four different algorithms are used and tested on our dataset. The algorithms are explained below.

3.4.1 Linear Regression Algorithm

The most standard maker finding out algorithm that can be carried out on this data is direct regression. The direct regression version returns a formula that figures out the connection between the independent variables as well as the reliant variable. The mathematical equation is

$$y=a+bx$$

Where y represents approximated value, b stands for Slope of a line, x stands for an independent variable and a stand for obstruct. [2]

3.4.2 K Nearest Neighbor Algorithm

It is supervised algorithm. knn is also known as laziest learning. Based on the independent variables, KNN finds the similarity between new data points and old data points.
The prediction of stock market closing price is computed using KNN as follows:
  a) Determine the number of nearest neighbors, k.
  b) Compute the distance between the training samples and the query record.
  c) Sort all training records according to the distance values.
  d) Use a majority vote for the class labels of k nearest neighbors, and assign it as a prediction value of the query record [3].

3.4.3 FB Prophet Algorithm

Prophet, created as well as originated by Facebook, is a time series projecting collection that needs no information pre-processing as well as is exceptionally basic to carry out. The input for Prophet is an information frame with two columns: day as well as target (ds and y). Prophet attempts to catch the seasonality in the past data and also functions well when the dataset is huge. The Prophet uses a decomposable time collection version with three primary design elements: fad, seasonality, and holidays. They are integrated in the list below formula

$$y(t)=g(t)+s(t)+h(t)+\varepsilon t$$

Here g(t) stands for piecewise straight or logistic growth curve for modelling non-periodic modifications in time collection, s(t) represents periodic modifications (e.g. weekly/yearly seasonality), h(t) stands for impacts of

vacations (individual given) with irregular routines and εt represents mistake term make up any kind of unusual adjustments not suited by the version [4].

3.4.4   Long Short-Term Memory

LSTMs are commonly used for sequence prediction problems and have actually shown to be exceptionally efficient. The factor they function so well is because LSTM is able to save past info that is important, and neglect the info that is not.

LSTM has 3 gates:
   i)   The input gate: The input gateway includes details to the cell state
   ii)   The fail to remember gate: It eliminates the information that is no more needed by the version
   iii) The output entrance: Result Gate at LSTM selects the details to be shown as result [1]

4. Results and Discussions

Figure 4 represents the outcome of linear regression where orange line indicates the actual value and the Green line indicates the predicted value and Figure 5 represents KNN outcome where Blue line indicates the actual value and the Orange line indicates the predicted value. Figure 6 represents the result of Fb Propnet where Orange line indicates the actual value and the Green line indicates the predicted value and Figure 7 represents LSTM outcome where Red line indicates the actual value and the Green line indicates the predicted value
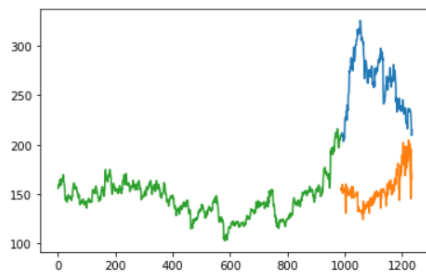

Figure 4: Linear regression outcome
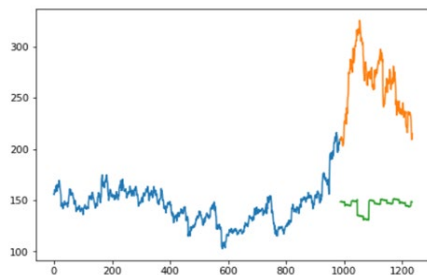

Figure 5: K nearest neighbor outcome
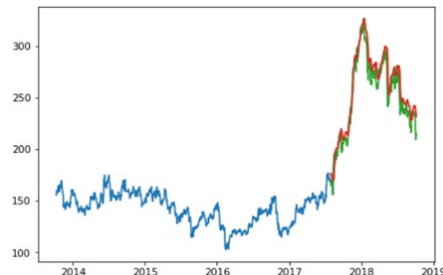

Figure 6: Fb prophet outcome

Figure 7: Long short-term memory outcome

4.1 RMSE scores

The RMSE is the square root of the difference of the residuals. It indicates the outright fit of the design to the data-- just how close the observed data factors are to the design's anticipated values.

RMSE can be taken the standard deviation of the unusual difference, as well as has the beneficial property of remaining in the exact same units as the reaction variable. Reduced worth's of RMSE indicate much better fit. RMSE is an excellent measure of how properly the design forecasts the reaction, and also it is the most crucial requirement for fit if the primary function of the design is prediction.

5. Conclusion

By gauging the RMSE of the various algorithms, we discovered that one of the most suitable formulas for predicting the market cost of a supply based on various information points from the historical data is the LSTM algorithm. The formula will be a terrific possession for brokers as well as capitalists for spending money in the securities market since it is educated on a significant collection of historical information and also has actually been picked after being evaluated on sample information. The job demonstrates the maker learning version to forecast the supply value with more precision as compared to formerly carried out device learning versions.

6. References

Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018). Stock Market analysis Using Machine Learning LSTM algorithm. 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). doi:10.1109/icsccc.2018.8703332

Bhuriya, D., Kaushal, G., Sharma, A., & Singh, U. (2017). Stock market analysis using a linear regression. 2017 International Conference of Electronics, Communication and AerospaceTechnology

Poornima S P,Priyanka C N,Reshma P, Suraj Kr Jaiswal and SurendraBabu K N .Stock market analysis using KNN algorithm. International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249–8958, Volume-8, Issue-5S, May 2019

Indira, Niral, Sowmya, Vijay.stock market analysis using FB prophet algorithm. Charles Davidson College of Engineering, San Jose State University One Washington Square, San Jose, California -95192-0080, USA

Stock Market Analysis: A Review and Taxonomy of Prediction Techniques Dev Shah, Haruna Isah * and Farhana Zulkernine School of Computing, Queen's University, Kingston, ON K7L 2N8, Canada; dshah@cs.queensu.ca (D.S.), International Journal of Financial Studies,2019,7,26

Ashraf S. Hussein,Ibrahim M. Hamed,Mohamed F. Tolba, An Efficient System for Stock Market Prediction, Advances in Intelligent Systems and Computing book series (AISC, volume 323)

Cheng, Li-Chen, Yu-Hsiang Huang, and Mu-En Wu. "Applied attention-based LSTM neural networks in stock prediction." 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018.

L. Y. Wei, "A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting," Applied Soft Computing, vol.42, pp. 368-376, May 2016.

C. M. Anish and B. Majhi, Hybrid nonlinear adaptive scheme for stock market prediction using feedback FLANN and factor analysis, Journal of the Korean Statistical Society, vol. 45, pp.64-76, March 2016

Biography

A. Sravani, working as Assistant Professor in Vignan's Institute of Information Technology (A), Visakhapatnam, India. She has an experience of seven years. Her research interests include Internet of Things, Data Mining and Image Processing. She published six papers in reputed journals out of that four are scopus indexed. Also, she authored a book on image processing.

Ch. Anusha, working as an Assistant Professor in Vignan's Institute of Information Technology (A), Visakhapatnam, India. She has an experience of two years. She has published three paper in reputed international journals. Her research interests include computer networks, mobile ad-hoc networks, machine learning, Data mining.

Dr. N V S Shankar, is working as Associate professor in Department of Mechanical engineering, has published over 20 papers of which 10 are scopus indexed. He has experience of 12 years, both teaching and industry combined.