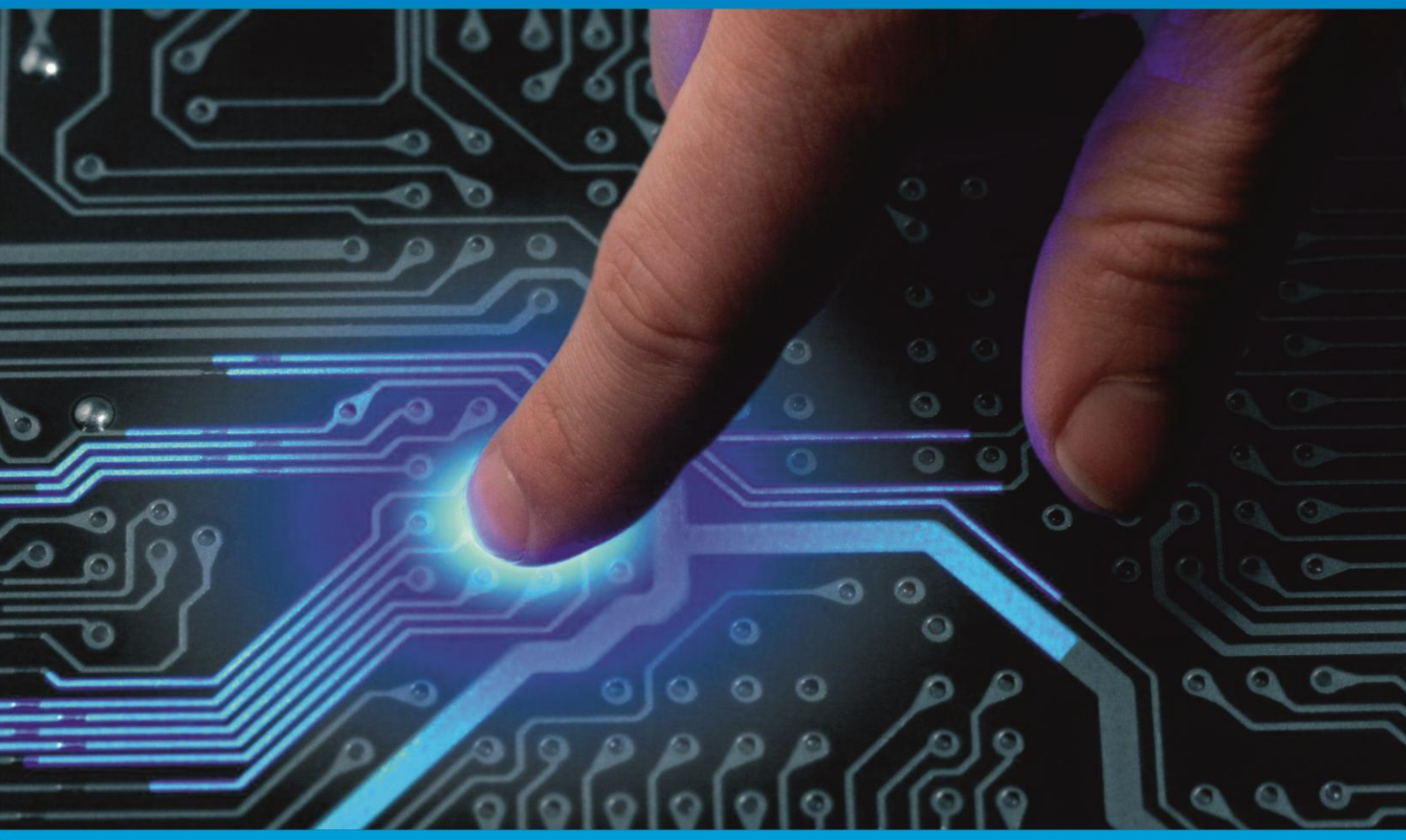
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006102 |

Stock Market Prediction Using Machine Learning

Anish S Madan[1], Krithik M[1], Neeraj Kumar S[1], Vijayakumara YM[2]

UG Student, Dept. of Information Science and Engineering, SMVIT – Bengaluru, Karnataka, India

Assistant Professor, Dept. of Information Science and Engineering, SMVIT – Bengaluru, Karnataka, India

ABSTRACT: The thing of Stock Market Prediction is to read the unborn value of a company's fiscal stocks. Machine literacy, which produces vaticinations grounded on the values of current stock request indicators by training on their previous values, is a recent trend in stock market vaticination technologies. Machine learning makes use of a variety of models to produce accurate prognostications. The exploration focuses on how to prognosticate stock values using Linear Regression, Polynomial Regression, Decision Trees, Sentiment Analysis, and LSTM grounded Machine Learning. Open, near, low, high, and volume are all factors to consider.

KEYWORDS: Prediction, training, Machinelearning, Analyzing,Visualize.

I. INTRODUCTION

A stock market is a public request where you can buy and vend shares for intimately listed companies. The stocks, also known as equities, represent power in the company. A successful stock cast can affect in large earnings for both the dealer and the broker. It's constantly stated that vaticination is chaotic rather than arbitrary, inferring that it can be prognosticated by precisely probing the history of the applicable stock request. Machine literacy is a good fashion to express these types of operations. It forecasts a request value that's near to the palpable worth, perfecting delicacy.

Because of its effective and accurate measures, the operation of machine learning to the field of stock vaticination has piqued the interest of numerous experimenters. Others differ and those with this standpoint retain myriad styles and technologies which purportedly allow them to gain unborn price information. The effective request thesis posits that stock prices are a function of information and rational prospects, and that recently revealed information about a company's prospects is nearly incontinently reflected in the current stock price. This would indicate that all intimately known information about a company, which obviously includes its price history, would formerly be reflected in the current price of the stock. Consequently, changes in the stock price reflect release of new information, changes in the request generally, or arbitrary movements around the value that reflects the being information set. Burton Malkiel, in his influential 1973 work A Random Walk Down Wall Street, claimed that stock prices could thus not be directly prognosticated by looking at price history. The dataset used in machine literacy is pivotal.

Because indeed minor changes in the data might affect in significant changes in the results, the dataset should be as precise as possible. On a dataset taken from Yahoo Finance, supervised machine literacy is used in this study. The following five variables make up this dataset open, close, low, high, and volume. Different shot prices for the stock at different times with nearly direct names are known as open, close, low, and high.

II. SYSTEM ARCHITECTURE

The proposed system will have an input from the dataset which will be uprooted point wise and classified underneath. The bracket fashion used is supervised and the colourful ways of machine position algorithms are enforced on the same.

Training Dataset are created for training the machine and the test cases are deduced and enforced to carry out the visualization and the conniving. The results generated are passed and are imaged in graphic form.

The compliances in the training set form the experience that the algorithm uses to learn. In supervised literacy problems, each observation consists of an observed affair variable and one or further observed input variables.

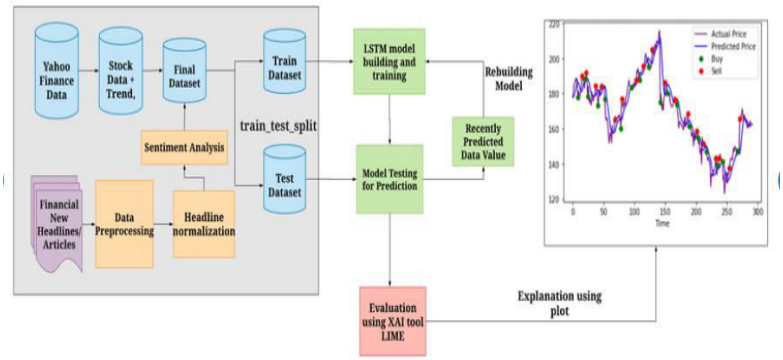
The test set is a set of compliances used to estimate the performance of the model using some performance metric. It's important that no compliances from the training set are included in the testset.However, it'll be delicate to assess whether the algorithm has learned to generalize from the training set or has simply learned it, If the test set does contain exemplifications from the training set.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006102 |

A program that generalizes well will be suitable to effectively perform a task with new data. In discrepancy, a program that memorizes the training data by learning an exorbitantly complex model could prognosticate the values of the response variable for the training set directly, but will fail to prognosticate the value of the response variable for new exemplifications. learning the training set is called over-fitting. A program that memorizes its compliances may not perform its task well, as it could study relations and structures that are noise or coexistence. Balancing memorization and conception, or over-fitting and under- fitting, is a problem common to numerous ML algorithms. Regularization may be applied to numerous models to reduce over-fitting.



III.METHODOLOGY

Stock market soothsaying appears to be a delicate subject to break since there are multitudinous variables that have yet to be considered, and it doesn't appear to be statistical at first. still, with the proper use of machine literacy ways, it's possible to link literal data to current data and train the system to learn from it and make suitable hypotheticals.

The data for the analysis was attained from Yahoo Finance. The needful stock prices and other important parameters were recorded in about 9 lakh records in the dataset. For each day of the time, the data reflected stock prices at specific time ages. It was divided into several divisions, including date, symbol, open, close, low, high, and volume. Only one company's data was taken into consideration for simulation and analysis.

All of the data was stored in a csv train, which was read and turned into a data frame using Python's Pandas package. The data for one specific company was taken from this by sorting the data by the symbol field. Following that, the data was normalized using Python's sklearn module, and the data was divided into training and testing sets. The test set was kept to be 20% of the total dataset.

Although there are many other models for machine learning, this research focused on two of the most essential ones and used them to make predictions.

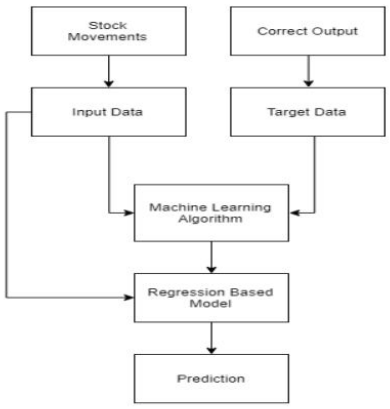


Fig. 1 Flow Chart for Regression Based Model

The enhanced version of Recurrent-Neural Networks (RNN) is LSTM, which keeps information from prior states. These differ from RNNs in that they entail long-term dependencies, whereas RNNs focus on determining the relationship between recent and present data. This shows that the information interval is smaller than the LSTM interval.

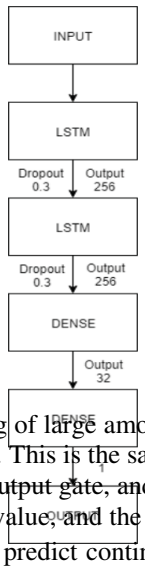
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006102 |

The key reason for utilizing this model in stock market prediction is that the predictions are based on a significant quantity of data and are usually based on the market's long-term history. As a result, LSTM reduces error by assistingRNNs by preserving knowledge from earlier stages, making prediction more accurate. As a result, it has proven to be far more trustworthy than previous methods.

Long Short Term Memory (LSTM) Network Based Model



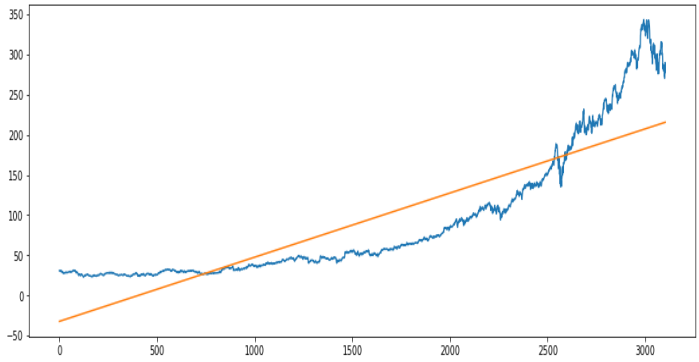
Because the stock market involves the processing of large amounts of data, gradients in relation to the weight matrix may become quite small, lowering the learning rate. This is the same as the Vanishing Gradient problem. This is not the case with LSTM. A remembering cell, input gate, output gate, and forget gate make up the LSTM.
For long-term propagation, the cell remembers the value, and the gates govern it.
In general, the Regression-based Model is used to predict continuous values from a set of independent variables. For predicting continuous values, regression employs a given linear function:

V = a + bK + error

IV. RESULTS

A. Linear Regression Model

The plot shown below is a result of passing the dataset through a linear regression algorithm



We get an accuracy of 74.95% with this algorithm.

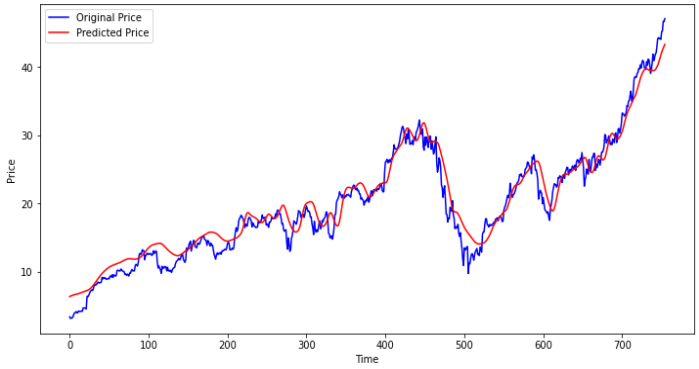
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006102 |

B. LSTM Model

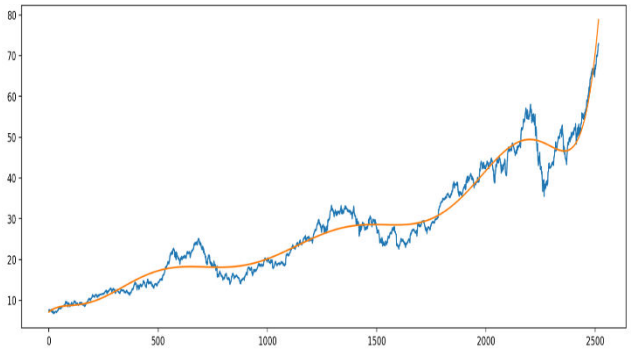
The plot shown below is a result of passing the dataset through a linear regression algorithm



We get an accuracy of 98.32% with the LSTM algorithm.

C. Polynomial Regression Model

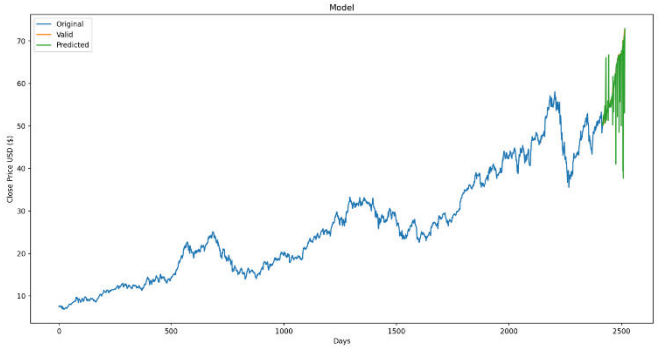
The plot shown below is a result of passing the dataset through a polynomial regression algorithm



We get an accuracy of 95.82% with this algorithm.

D. Decision tree Model

The plot shown below is a result of passing the dataset through a decision tree algorithm



We get an accuracy of 96.73% with this algorithm

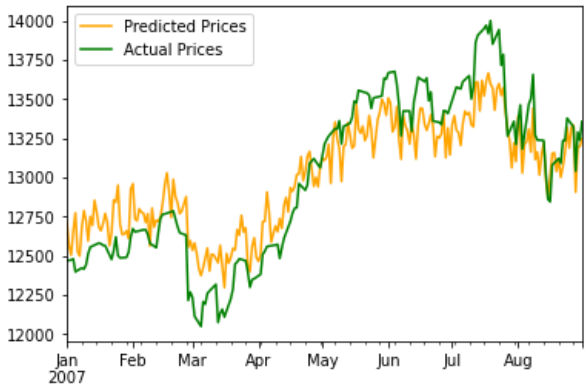
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006102 |

E. Sentiment Analysis

The plot shown below is a result of passing the dataset through a decision tree algorithm



We get an accuracy of 83.42% with this algorithm

From the above graphs, plots and accuracies, we can infer that the LSTM algorithm is the most accurate algorithm to predict stock prices, followed by Decision tree algorithm, followed by Polynomial regression. The least accurate algorithm is the Linear regression algorithm.

REFERENCES

1. Stock Market Prediction Using Machine Learning
   Authors : Ishita Parmar; Navanshu Agarwal; Sheirsh Saxena
2. Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm
   Authors : Lei Zhao; Lin Wang
3. Stock market prediction using Hidden Markov Model
   Authors : Poonam Somani; Shreyas Talele; Suraj Sawant
4. Stock Market Prediction based on Social Sentiments using Machine Learning
   Authors : Tejas Mankar; Tushar Hotchandani

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details