



Stock Price Prediction Using Linear Regression, LSTM and Decision Tree

Sarthak Singh, Shaurya Rehan and Vimal Kumar

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

April 18, 2022

Stock Price Prediction Using Linear Regression, LSTM and Decision Tree

Sarthak Singh; Shaurya Rehan ; Vimal Kumar

Guide:- Mrs. Babli Baliyan

Department of Computer Science and Engineering

ABES Engineering College

Abstract: In the cash world stock trading is maybe the principle development. Protections trade assumption is a showing of endeavoring to conclude the future worth of a stock other financial instrument traded on a money related exchange. This paper explains the gauge of a stock using Machine Learning. The particular and head or the time series assessment is used by a huge piece of the stockbrokers while making the stock assumptions. The programming language is used to anticipate the protections trade using AI is Python. In this paper we propose a Machine Learning (ML) approach that will be ready from the available stocks data and gain knowledge and a short time later includes the acquired data for an exact figure. In this setting this study uses an AI strategy called Linear Regression, LSTM and Decision Tree to anticipate stock expenses for the immense and little capitalizations and in the three special business areas, using costs with both step by step and approved frequencies.

Keywords: *Machine Learning Algorithms, Stock Market, Linear Regression, LSTM, Decision Tree .*

1. INTRODUCTION

Fundamentally, quantitative brokers with huge amount of cash from securities exchanges purchase stocks subordinates what's more, values at an unobtrusive expense and later on selling them at exorbitant cost. The pattern in a securities exchange forecast is certifiably not something else but rather this issue is proceeded to be discussed by various affiliations. There are two sorts to dissect stocks which financial backers perform prior to placing assets into a stock, first is the head examination, in this investigation financial backers take a gander at the inherent worth of stocks, and execution of the business, economy, world of politics and so on to conclude that whether or not to contribute. Then again, the specialized examination it is a development of stocks by the method for concentrating on the measurements created by market movement, for example, past costs and volumes.

In the new years, expanding unmistakable quality of AI in different ventures have edified various vendors to apply AI methods to the field, and some of them have made exceptionally reassuring results. This paper will foster a monetary information indicator program in which there will be a dataset putting away all chronicled stock costs and information will be treated as preparing sets for the program. The fundamental justification behind the estimate is to diminish weakness connected with venture independent direction.

Securities exchange follows the irregular walk, which suggests that the best forecast you can have about the upcoming worth is the current worth. Evidently, the expecting stock records is extremely challenging a direct result of the market unpredictability that needs precise estimate model. The securities exchange records are exceptionally fluctuating and it impacts the financial backer's conviction. Stock costs are considered an incredibly novel and helpless against quick changes because of stowed away nature of the money related space and somewhat considering the mix of a known limits (Previous day's end esteem, P/E extent, etc) and the dark elements (like Election Results, Rumors and so on) value. There has been various endeavors to anticipate stock cost with Machine Learning. The point of convergence of every investigation projects moves an extraordinary arrangement in three ways. (1) The zeroing in on change can be close term (under a second), present second (tomorrow to a few days sometime later), and a long stretch (months afterward), (2) The game plan of stocks can be in confined to under 10 explicit stock, to stocks explicitly industry, to generally all stocks. (3) The indicators utilized can go from a overall news and economy design, to explicit characteristics of the association, to just time series data of the stock cost.

The conceivable protections trade estimate target can be the future stock expense or the shakiness of the expenses or market pattern. In the forecast there are two sorts like faker and an ongoing expectation which is utilized in securities exchange expectation framework. In Dummy forecast they have characterize some arrangement of rules and anticipate the future expense of offers by working out the ordinary expense. In the constant expectation mandatory utilized web and saw current cost of portions of the organization.

Computational advances have provoked show of AI systems

for the judicious structures in money related business areas. In this paper we are using a Machine Learning method i.e., Support Vector Machine (SVM) to predict the protections trade and we are including Python language for programming.

2. METHODOLOGY

In this project we are using three different machine learning approaches Linear regression, LSTM (Long Short-Term Memory) and Decision Tree Algorithm.

2.1 Linear Regression

Linear regression is used in business, science, and pretty much whatever other field where expectations and anticipating are pertinent. It recognizes the connections between a reliant variable and at least one free factor. Straightforward direct relapse is characterized by utilizing a component to anticipate a result. That is what the future holds. Financial exchange estimating is an appealing use of straight relapse. Current AI bundles like scikit-learn for executing these examinations conceivable in a couple of lines of code.

2.2 LSTM (Long Short Term Memory)

Long transient memory (LSTM) is a model that builds the memory of repetitive brain organizations. Intermittent brain networks hold transient memory in that they permit prior deciding data to be utilized in the current brain organizations. For guaranteed assignments, the prior information is utilized. We may not have a rundown of all of the previous data for the brain hub. In RNNs, LSTMs are broadly utilized in Neural organizations. Their viability ought to be executed to different arrangement demonstrating issues in numerous application spaces like video, NLP, geospatial, and time-series.

One of the principle issues with RNN is the disappearing slope issue, and it arises because of the rehashed utilization of similar boundaries, in RNN blocks, at each progression. We should attempt to utilize various boundaries to beat this issue at each time step.

We attempt to track down an equilibrium in such a circumstance. We bring novel boundaries at each progression while summing up factor length successions and keeping the general measure of learnable boundaries consistent. We present gated RNN cells like LSTM and GRU.

Gated cells hold interior factors, which are Gates. This worth of each door at each time step relies upon the data around then advance, including early states. The worth of the door then becomes increased by the various factors important to impact them. Time-series information is a progression of information values assembled over the long haul breaks, permitting us to follow contrasts over the long run. Time-series information can follow progress over milliseconds, days, and years.

Early, our point of view of time-series information implied more static; the regular highs and lows under temperature, the opening and shutting measure of the financial exchange. Presently we will go to the coding part. We will carry out LSTM on the stocks dataset.

2.3 Decision Tree Algorithm

Choice Tree estimation has a spot with the gathering of controlled learning computations. Not at all like other oversaw learning estimations, the decision tree computation can be used for dealing with backslide and gathering issues also.

The goal of using a Decision Tree is to make a readiness model that can use to predict the class or worth of the objective variable by acquiring direct decision principles concluded from before data(training data).

In Decision Trees, for predicting a class name for a record we start from the foundation of the tree. We dissect the potential gains of the root quality with the record's trademark. In light of connection, we follow the branch contrasting with that value and jump to the accompanying center point.

3. Data Set Collection and Preprocessing

Data Set Collection and Preprocessing is an exceptionally fundamental module that for the most part manages the assortment of right datasets. The dataset that will be utilized in market expectation should be sifted in view of various viewpoints.

Information preprocessing is a piece of information mining, which includes changing crude information into a more cognizant structure.

The information preprocessing includes looking at for missing qualities, searching for all out values, parting the dataset into preparing and test information for additional handling. We have used TESLA dataset taken from Kaggle which consists of 2416 instances and 7 attributes.

Date	Open	High	Low	Close	Adj Close	Volume
#####	19	25	17.54	23.89	23.89	18766300
#####	25.79	30.42	23.3	23.83	23.83	17187100
7/1/2010	25	25.92	20.27	21.96	21.96	8218800
7/2/2010	23	23.1	18.71	19.2	19.2	5139800
7/6/2010	20	20	15.83	16.11	16.11	6866900
7/7/2010	16.4	16.63	14.98	15.8	15.8	6921700
7/8/2010	16.14	17.52	15.57	17.46	17.46	7711400
7/9/2010	17.58	17.9	16.55	17.4	17.4	4050600
#####	17.95	18.07	17	17.05	17.05	2202500
#####	17.39	18.64	16.9	18.14	18.14	2680100
#####	17.94	20.15	17.76	19.84	19.84	4195200
#####	19.94	21.5	19	19.89	19.89	3739800
#####	20.7	21.3	20.05	20.64	20.64	2621300
#####	21.37	22.25	20.92	21.91	21.91	2486500
#####	21.85	21.85	20.05	20.3	20.3	1825300
#####	20.66	20.9	19.5	20.22	20.22	1252500
#####	20.5	21.25	20.37	21	21	957800
#####	21.19	21.56	21.06	21.29	21.29	653600
#####	21.5	21.5	20.3	20.95	20.95	922200
#####	20.91	21.18	20.26	20.55	20.55	619700
#####	20.55	20.9	20.51	20.72	20.72	467200
#####	20.77	20.88	20	20.35	20.35	616000

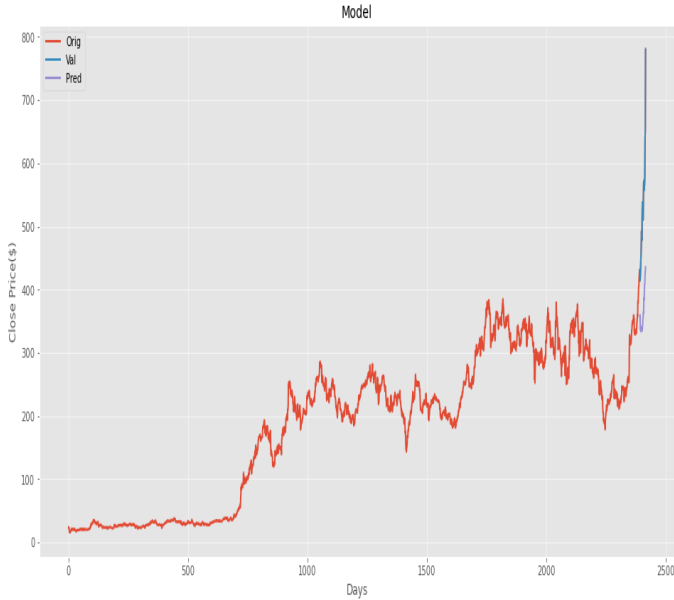
The preparation sets are utilized to tune and fit the models. It is separated into the apportion of 80:20, where 80% is for preparing set and rest 20% is utilized for the testing stage.

4. Result

4.1 Using Linear Regression

Linear Regression can be utilized to track down a connection between at least two factors of interest and permits us to make forecasts once these connections are found.

Results on TESLA Dataset



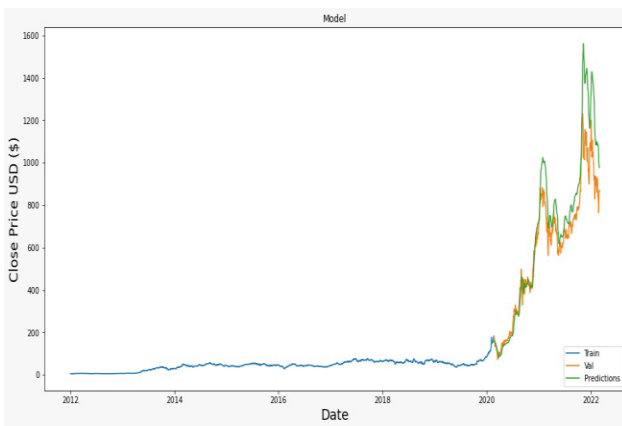
This algorithm gives an accuracy of **76%** approx..

4.2 Using LSTM

LSTM networks are an augmentation of repetitive brain organizations (RNNs) for the most part acquainted with handle circumstances where RNNs fizzle.

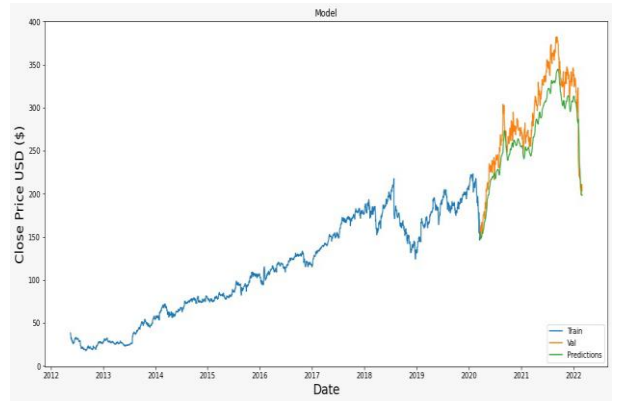
The three pieces of LSTM cell are known as doors. The initial segment is called neglect entryway, the subsequent part is known as the info door and the last one is known as the result door.

Results on TESLA Dataset



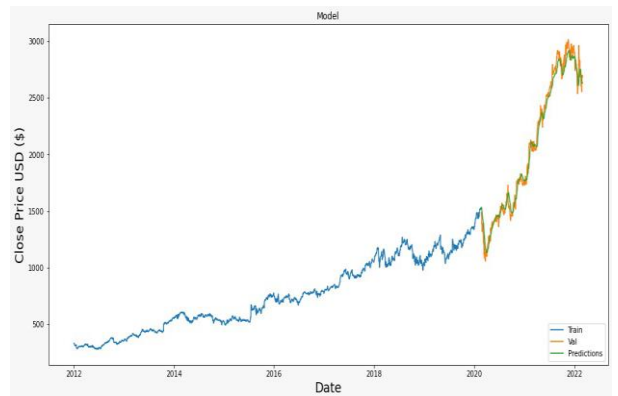
This algorithm has a root mean square error of **22.3**.

Results on Facebook Dataset



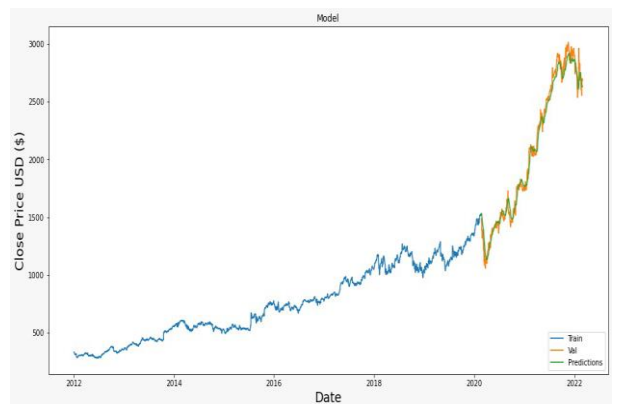
This algorithm has a root mean square error of **8.4**.

Results on Google Dataset



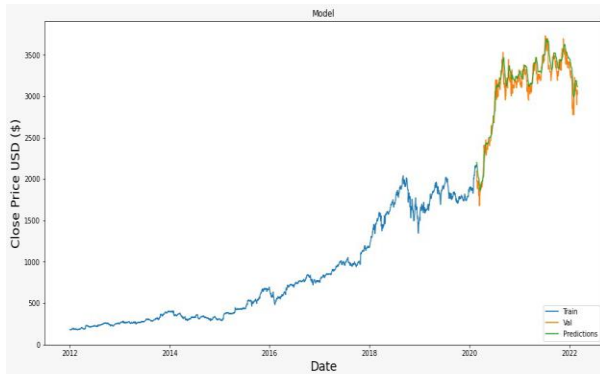
This algorithm has a root mean square error of **4.2**.

Results on Apple Dataset



This algorithm has a root mean square error of **7.7**.

Results on Amazon Dataset



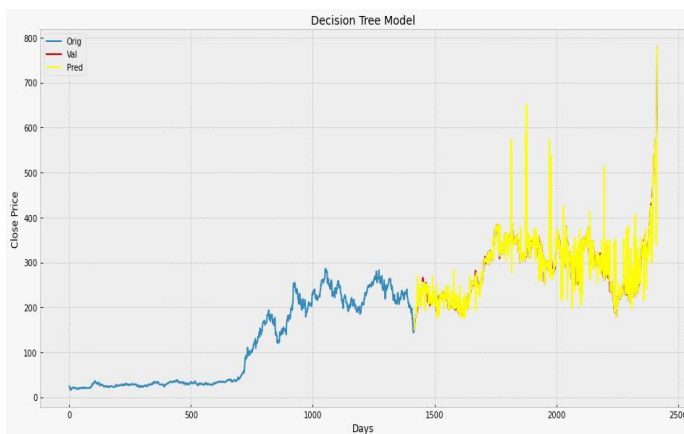
This algorithm has a root mean square error of **19.2**.

4.3 Results using Decision Tree

Decision Tree is a directed AI procedure that can be utilized both for grouping as well as relapse issues, but for the most part it is liked for arrangement issues.

It is a tree organized classifier, where interior hubs address the elements of a dataset, branches address the choice guidelines and each leaf hub addresses the result.

Results on TESLA Dataset



This algorithm gives test score accuracy of **28%** and train score accuracy of **96%**.

5. Conclusion

As our objective was to predict the future value of a stock we were able to achieve our goal using all the three techniques we have discussed i.e., Linear Regression, LSTM and Decision Tree, if we compare the algorithms and their results then the best one with the highest precision is LSTM (Long Short Term Memory).

- At first, we used the same TESLA dataset and the deflection from the real value was 22.3.

- Then we used the same algorithm on AMAZON, FACEBOOK, APPLE and GOOGLE dataset and the deflection from the real values were 19.2, 8.4, 7.7 and 4.2.

6. Future Scope

In future we have decided to create an application where the LSTM model will be integrated, and it will provide the next day stock value of every registered stock in that application.

7. References

- [1] Shah, Isah and Zulkernine;2019. Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies, 7(2), pg 26
- [2] Hu, Zhao and Khushi;2021. A survey of forex and stock price prediction using deep learning. Applied System Innovation, 4(1), pg 9.
- [3] Obthong, M.Tantisantiwong, N.Jeamwatthanachai and Wills, G., 2020. A survey on machine learning for stock price prediction: algorithms and techniques.
- [4] A stock price prediction method based on deep learning technology ;Xuan Ji, Jiachen Wang, Zhijun Yan March;2021.