

|  |  |
| --- | --- |
| Stock prediction  [machinelearning::linearregression&  long shorttermmemory | BY::  Devaki.S |

# StockMarketPrediction

***Abstract: The Stock Market is a challenging forum for investment andrequiresimmensebrainstormingbeforeoneshall put their hard earned money to work. This project aims at processing large volumes of data and running comprehensive regression algorithms on the dataset; that will predict the future value of a stock using the regression model with the highest accuracy.Thepurposeofthispaperistoanalyzetheshortcomings ofthecurrentsystemandbuilding atime-seriesmodelthat would mitigatemost ofthembyimplementingmore efficientalgorithms. Usingthismodel,anyonecanmonitorthepreferredstockthatthey wanttoinvestin;andmaximizeprofitbypurchasingvolumeatthe lowest price and liquidating the stock whenit’s at its highest.***

***Keywords: Stock Market, Forecast, Regression, Time-Series Prediction.***

1. **INTRODUCTION**

Stocksformthe corner-stone of any business portfolio andmaybepurchasedprivately, orfrompublicforums.Any such transactionmust conformtolegalnormsthathavebeen established by the government; in order to prevent illegal practices. *“A stock (alsoknownas"shares"or"equity")isa typeof securitythatsignifiesproportionateownership inthe issuing corporation. This entitles the stockholder to that proportion of the corporation's assets and earnings.”* Historically,stockshavesurvivedtherelentlesswrathof time,surpassingallitspredecessors. Stockscan beboughtat the stock exchange or from many online stock brokers.

Significantprofitcanbeyielded fromthesuccessful prediction ofa stocksfuture. *“Stock market predictionisthe act of trying to determine the future value of a company stockorotherfinancialinstrument tradedonanexchange.”* The efficient-market hypothesis indicates that the prices of stock reflect all information, available currently and hence, any changes in price that are not based on information that recentlycameintolight,areconsequentlyunpredictable. The peoplewhoopposethishypothesistherefore,possessmyriad methods andtechnologies which calculatedlyallowthem to gain information about the future price of stock. Stock market price data is tediously voluminous and extremely volatile.Stockmarket trading is an extremely complicated and ever-changingsystemwhere peoplewilleithergain a fortuneor lose their entirelife savings.[11] In this work, an attempt to build a Time-Series prediction model, to predict stock prices.[11] This papertakes the current stock values from the data sets gathered. The data gathered is modelled intovarious subpartsor data sets which is usedtotrain and testthealgorithm.Weuseregression modelsinpython or R tomodelthedata.Werunacomprehensivesearchalgorithm onthe datasets andcreate asummary tablebasedonthe output. We plot the values on a chart and apply regression andclusteringtechniquestofindouttheincreaseordecrease

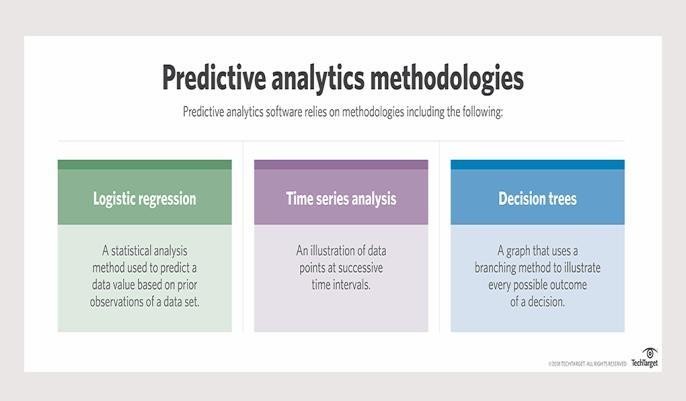
inpriceofthatstock.

Based on the calculation, we extrapolate the current stock prices togeneratea predictionafter a given time. Themodel is developed using supervised machine learningalgorithms. The output will be in graphical form and will change with change in dataset. We expect up to66% in-sample accuracy and 35% out-of-sample accuracy usingsupervised machine learningalgorithmson predictionmodel.Thiswillenablethe user to take better decisions while investing.

1. **LITERATURESURVEY**

The Stock Market of a country is considered an accurate reflection of its economic prowess. Stock Market prices are ever-changing as they are prominently affected by the ebb and flow of finances throughout various economic domains. Prices of stock are invariably dependent on the demand and supply curve, that is a fundamental rule of economics. An increase in demand of a particular stock, will increase its pricewhereasadecreaseinpopularityofanother, willreduce its price. Even though this fluctuation in stock prices in necessary to reap profits from investment that have been made,itisofparamountimportancethatwepredictthefuture valueofa stocksothatlosson investmentsismitigated.This reviewisdonein order topredictthe pricesofstock, soasto make more informed investments.

Recently occurring trends in a market is studied and different typesofmachinelearningclassifiersandregression techniques are applied on them. Various approaches[Fig. 1] and results from past studies are weighted based on various parametrics;[Table 1] and then is displayed in a graphical format. The survey brings to light, different conventional approachesto stock market prediction, thathasalreadybeen built. In addition to that, it discusses recent application of machine learning techniques along with strengths and weaknesses of each technique for effective prediction of stock prices, in the future.[12]

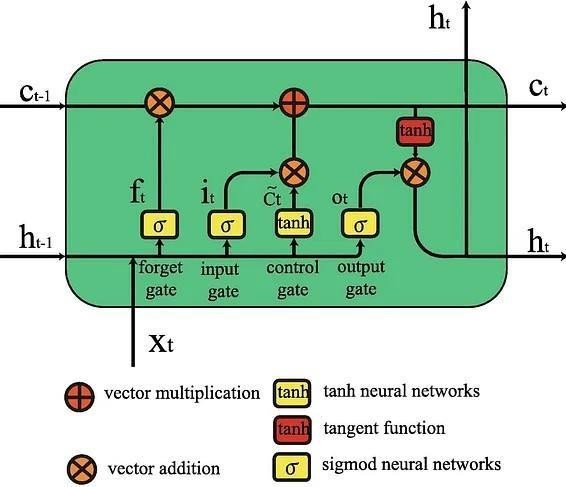


**Fig.1:PredictiveAnalysisMethodologies[13]**

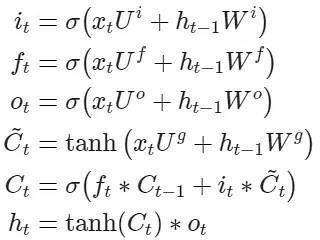
# LSTM(LONGSHORTTERMMEMORY)

UnderstandingLongShortTermMemoryNetworkforStockPricePrediction

LSTMisaRecurrentNeuralNetworkthatworksondatasequences, learningtoretainonly relevantinformationfroma timewindow. Newinformationthenetworklearns is addedto a“memory”thatgetsupdatedwitheachtimestepbasedonhowsignificantthenewsample seems to the model. Over the years, LSTM has revolutionized speech and handwriting recognition, languageunderstanding, forecasting, andseveralother applicationsthathave become the new normal today.

A standard LSTM cell comprises of three gates: the input, output, and forget gate. These gates learn their weights and determinehowmuchofthecurrent data sampleshouldberememberedandhowmuchofthepast learnedcontent should beforgotten.

*As seen in the equations below, i, f, and o represent the three gates: input, forget, and output. C is the cellstatethatpreservesthelearned data,whichisgivenasoutputh.All of this is computed for each timestamp t, considering the learned data from timestamp (t-1).*



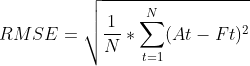
*The forget gate decides what informationandhowmuchof itcan be erased from the current cell state, while the input gate decides what will be added to the current cell state. The output gate, used in the final equation, controls the magnitude of output computed by the first two gates.*

*So, asopposed tostandard feed-forward neural nets, LSTMs have the potential to remember or erase portionsofthepastdatawindowsactively.Itsfeature of reading and trainingon windows (or timesteps) of data makes its training unique*

## Evaluating Prediction PerformanceforStockPrice Prediction

Before putting the algorithms into practice, let’s clarify the metric to measure the performance of our models. Stock price prediction being a fundamental regression problem, we can use RMSE(RootMeanSquaredError)orMAPE(Mean AbsolutePercentageError)tomeasurehowclose or far off our price predictions are from the real world.

Lookingclosely at the formulaof RMSE, we can see how we will be able to consider the difference(orerror)betweentheactual(At)and predicted (Ft) price values for all N timestamps and get an absolute measure of error.



On the other hand, MAPE looks at the error concerning the true value – it will measure relatively how far off the predicted values are fromthetruthinsteadofconsideringtheactual difference. This is a good measure to keep the

errorrangesincheck if wedeal with too large or small values. Forinstance,RMSEforvaluesin the range of 10e6 might blow out of proportion, whereas MAPE will keep error in a fixed range

MAPE

Stock Market Prediction using MachineLearning

ProjectCode

First,wewillimplementasimple LSTM network using [Keras](https://www.projectpro.io/article/keras-for-deep-learning/830)in Python. Let’s take a look at the Stock Prediction using Machine Learning dataset. We can work onactualstockdatafrommajor public companies such as Facebook, Microsoft, or Apple bysimplydownloadingthedata from [finance.yahoo.com.](https://finance.yahoo.com/)

## Downloading the StockPricesDataset for Project on Machine Learning Price Prediction of

**Stocks**

Go to [finance.yahoo.com/](https://finance.yahoo.com/)and searchthecompanywhosedata youwanttoseekforstockprice prediction. For our example, we will look at the Netflix (NFLX) stock over 3 years.

Going

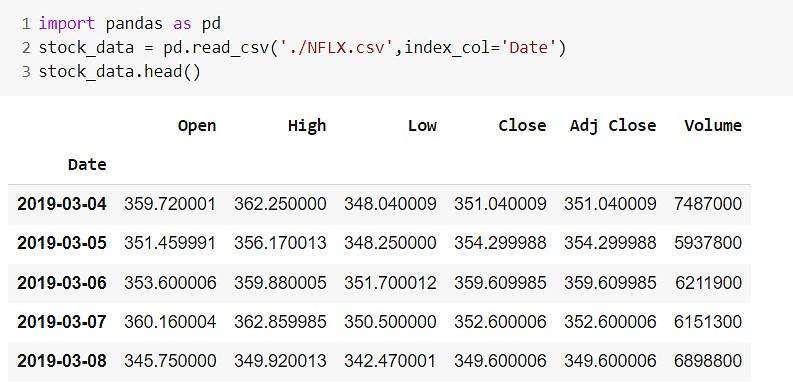
to [finance.yahoo.com/quote/NFLX/history?p=NFLX](https://finance.yahoo.com/quote/NFLX/history?p=NFLX)in the “Historical Data” section, we see the stock data listed each day. We can filter out the timeforwhichwewishtoanalyzeanddownload the CSV file using the download button on the right.



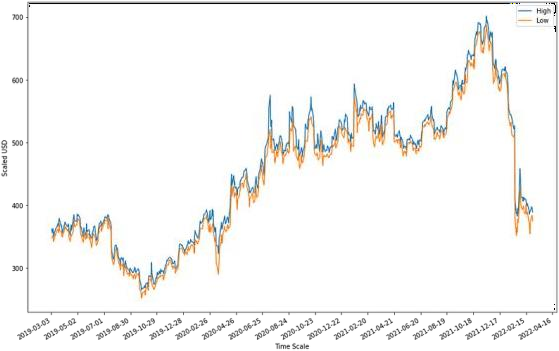
ThedownloadCSVfilewillcontainthedatafor Open, High, Low, Close, Adj Close, Volume for each date, as shown in the image above.

## LoadingtheStockPricesDataset

Load the CSV file as a DataFrame using [Pandas.](https://www.projectpro.io/data-science-in-python-tutorial/python-pandas-tutorial-for-beginners) Since the data is indexed by date (each row represents data from a different date), we can also index our DataFrame by the date column. We have taken the data from March 2019 to March2022.Thiswillalsochallengeourmodelto work with the unpredictable changes caused by the COVID-19 pandemic.



Plotting the High and Low pointsofNetflixstockover3years,wesee the below graph.



As noticeable, around March 2020,weseeasuddendropin theprice,afterwhichitreports steady growth until recently.

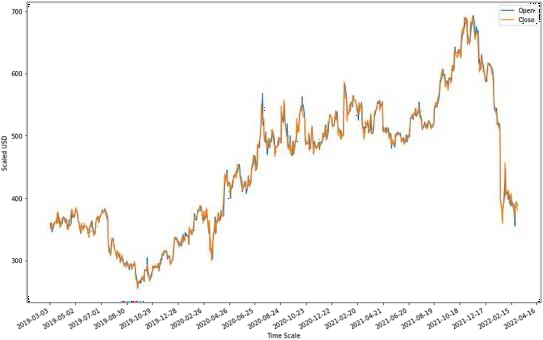
It will be challenging for a model in the stock prediction using machine learning project to correctly estimate the rapid changes that we can see in March2020andFebruary2022. Wewillfocusonevaluatingthe model performance in predicting the more recent values after training it on the past data.

Similarly, plottingthe Open and Closevalueofthestockforeach day gives equivalent observations.

The code for plotting these graphs is as shown below. We use matplotlib to plot the DataFrame columns directly against the Date index column. To make things flexible while plotting against dates, lines 6-8 convert our date strings into datetime format and plot them cleanly and legibly. The interval parameter in line 7 defines theintervalindays betweeneachtickonthedate axis.



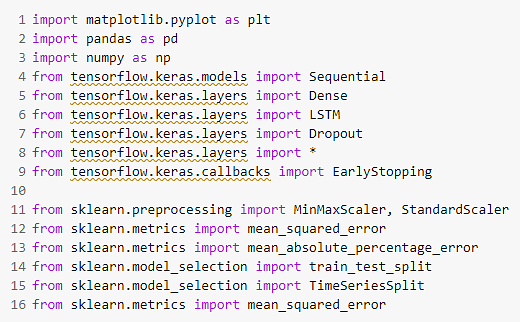
WewillusetheOpen,High,andLowcolumns to predict the Closing value of the Netflix stock for the next day.



## Importing the LibrariesforStock Price Prediction

**Project**

We will be building our [LSTMmodels](https://www.projectpro.io/article/lstm-model/832)using Tensorflow Keras and preprocessing our stock prediction machine learning data using scikit-learn. These imports are used in different steps of the entire process, but it is good to club these statementstogether.Whenever we wish to import something new, just add the statement arbitrarily to the below group.



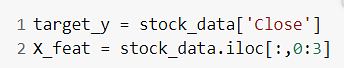
## DataPreprocessing for Stock Market Prediction using

**MachineLearning**

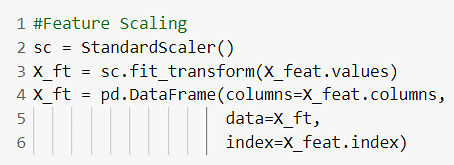
As with any other machine learning model, it is always goodtonormalizeorrescalethe data within a fixed range when dealing with real data.

Thiswillavoidfeatureswithlargernumericvalues to unjustly interfere and bias the model and help achieve rapid convergence in the machine learning stock prediction project.

First,wedefinethefeaturesandthetargetas discussed above.



Next,weuseaStandardScalertorescaleour values between -1 and 1.



Scikit-learn also provides a popularMinMaxScaler preprocessing module. However, considering the context, stock prices might max out or minimise on different days, and using those values to influence others might not be great. The change in values from using either of thesemethodswouldnotbemuch,sowestickto StandardScaler.

Wehave757 datasamplesinthe dataset.

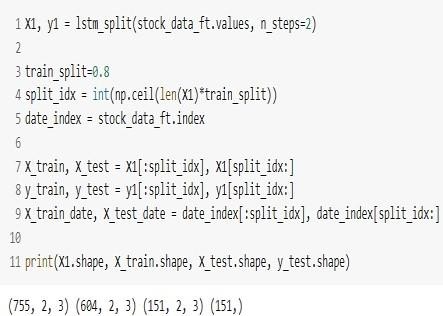
So,thenextstepwouldbetosplititintotraining andtestingsets.Asexplainedabove,thetraining of an LSTM model requires a window or a timestep of data in each training step. For instance, the LSTM will take 10 data samples to predict the 10th one by weighing the first nine input samples in one step. So, we need a different approach than the train\_test\_split provided by scikit-learn.

Let’s define a splitting function called lstm\_split() which will make windows of size “n\_steps” starting from the first sample of data and ending at n\_steps’th sample (if n\_steps=10, then the 10th sample) from the end. We understand the latter part because, for each time step, LSTM will take n\_steps-1 samplesfortrainingandpredict the last sample. Loss calculation is done based on the error in thisprediction.Soifn\_steps=10, you cannot use the last 9 samples to predict anything because the “10th” data point for the current step does not exist in the dataset.

The function below takes the entiredataandcreateswindows ofsizen\_stepsstartingfromthe beginning. The target y will contain the target value correspondingtothen\_steps’th index. So if n\_steps is 10, the first element in X will have features from 10 data samples, and y will contain the target of the 10th data sample.

## TrainandTestSetsforStock Price Prediction Project

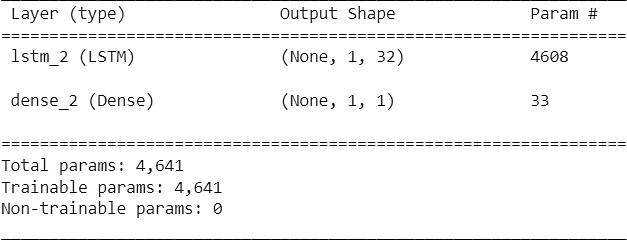
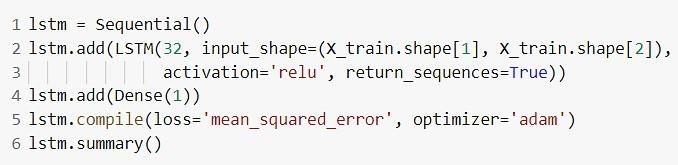
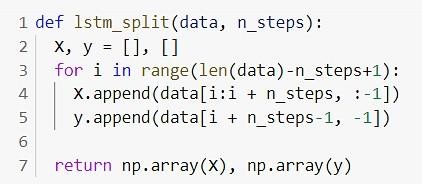
We split our data into training and testing sets. Shufflingisnotpermittedintime-seriesdatasets. Inthebeginning,wetaketwostepsworthofpast datatopredictthecurrentvalue.Thus,themodel will look at yesterday’s and today’s values to predict today’s closing price.



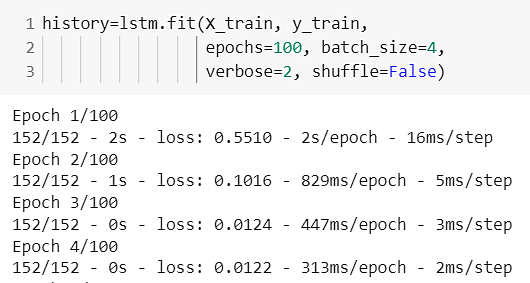
Note above that the size of X1 is n\_steps less thanthatoftheoriginaldataset.Asweexplained above, you cannot use the last two samples of the original set during training or prediction as wedonothavetheircorrespondinggroundtruth values.

## StockPredictionMachine

**LearningProject-Buildingthe LSTM model**

We will use the Sequential and LSTM modules providedbyTensorflowKerastobuildasimple, single-unit LSTM model.

Now we canfit thissimple modeltothetrainingdata.



Given the simplicity of the model and the data, we note thatthelossreductionstagnates after only 20 epochs. You can observe this by plotting the traininglossagainstthenumber of epochs, and LSTM does not learn much after 10-20 epochs.

