

(Stock Price Prediction using Machine Learning Algorithms)

Project Report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology .

in

Computer Science and and Engineering /Information Technology

By

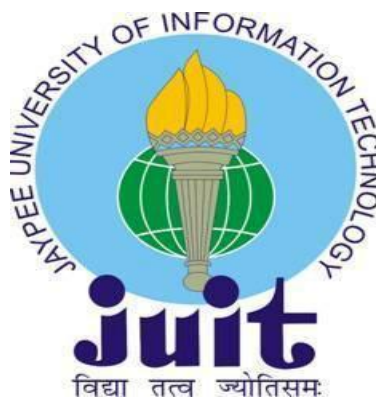
Anand Aman(171275)

Pratimaan Priyam(171240)

Under the Supervision of

(Dr. Hemraj Saini)

To



Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology Waknaghat , Solan-173234,

Himachal Pradesh

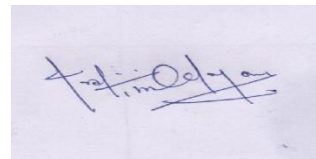
Candidate's Declaration

We hereby declare that the work presented in this report entitled "**Stock Price Prediction using Machine Learning Algorithms**", in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science Engineering /Information Technology** submitted in the department of Computer Science and Engineering and Information Technology, Jaypee University of Information Technology Waknaghat is an authentic record of my own book carried out over a period from January,2021 to May,2021 under the supervision of **Dr.Hemraj Saini , Associate Professor , Computer Science Engineering and Information Technology, Jaypee University of Information Technology , Solan.**

Name of Student: Anand Aman

Name of Student: Pratimaan Priyam

Anand Aman



Roll No. : 171275

Roll No. : 171240

Course: B.Tech C.S.E

Course: B.Tech C.S.E

Year: 4th

Year: 4th

Semester: 8th

Semester: 8th

This is to certify that the above statement made by the candidate is true to the best of my knowledge .

Supervisor Name: Dr. Hemraj Saini

Designation: Associate Professor

Department Name: Computer Science Engineering and Information Technology

Dated: 16th May , 2021

Acknowledgement

It gives us a great sense of pleasure the report of the B.Tech Major Project undertaken during B.Tech 4th year . This project is itself is an acknowledgement to the inspiration , drive and technical assistance contributed to it by many individuals . This project would never have seen the light of the day without the help and guidance that we have received .

Our heartiest thanks to Dr. Samir Dev Gupta and Head of Dept., Department of C.S.E & I.T for providing us with an encouraging platform to develop this project , which thus helped us in shaping our abilities towards a constructive goal .

We owe special debt. of gratitude to Dr. Hemraj Saini, Associate Professor Department of C.S.E and I.T, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and preservice have been a constant source of inspiration for us. He has showered us with all his extensively experienced ideas and insightful comments at virtually all stages of the project & has also taught us about the latest industry-oriented technologies.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind guidance and cooperation during the development of our project. Last but not the least , we acknowledge our friends for their contribution in the completion of the project .

Pratimaan Priyam

Anand Aman

TABLE OF CONTENTS

Title.....	(i)
Certificate.....	(ii)
Acknowledgement.....	(iii)
List of Tables.....	(v)
Abstract.....	(vi)
Chapter 1 INTRODUCTION.....	1
Chapter 2 MAJOR PROJECT SDLC.....	13
Chapter 3 IMPLEMENTATION OF THE MAJOR PROJECT.....	20
Chapter 4 RESULTS.....	28
REFERENCES.....	32

List of Abbreviations:

IDE- Integrated Development Environment

JSON- Java Script Object Notation

OTP- One Time Password

XML- Extensible Markup Language

UI- User Interface

API- Application Programming Interface

REST- Restfull API

GPS- Global Positioning System

DOM- Document Object Model

SDK- Software Development Kit

HTR – Handwritten Text Recognition

ABSTRACT

The project aims to provide retail investors to navigate through the stock market. This is achieved through the use of several machine learning algorithm. A deeper understanding of Random Forest by taking OHLC (open-high-low-close) data for Apple,Google stock for 5year period. This will keep unrelated discussion at a minimum and try to focus on everything that is related to our understanding of Random Forest for this post. It would be helpful if you are familiar with concepts such as mean squared error and R-Squared statistics but it's not necessary to understand this Concept.

1.1 Introduction

Stock price Prediction is an exercise for trying to predict the future value of a company stock and other instrument which will be traded financially on an average. If our system will predict the stock's future price successfully then it may yield a very significant profit.

The efficient market hypothesis suggest that stock price is the function of company's information and the expectations. The information which has currently revealed about the company is also reflected immediately in the company's stock price. It implies that all the public information about the company, which contains the price history, will already reflected in the current stocks price. Similarly, changing in the stock price will reflect the new information's release, market change and random movement around the value that reflects the existing information set. In 1973, Burton Malkiel claimed that stock price the stock price cannot be predicted accurately by looking on the price history. In return, Malkiel argued that the stock price square measure represented best by a applied mathematics method called "random walk" which suggests that the deviation of every day is random and unpredictable. It LED Malkiel to a conclusion that paying money services to the individuals to plug prediction really hurts, instead of helped, web portfolio come back. Whereas, the economical hypothesis of market finds a favor among money lecturers, it's juncture to the instances within which the expertise of the particular market differs from the prediction of unpredictability the hypothesis implies. There square measure an oversized variety of industries that has grownup up round the implication proposition that there square measure some analyst WHO predict it higher than the others, which might be not possible beneath the economical market hypothesis that if the stock worth prediction firms didn't provide one thing that it's client believes to be important .In 1984, one of the most successful and famous investor, Warren Buffet, rebutted the efficient hypothesis during his speech at Columbia University.

1.1.1 Introduction to Machine Learning

Machine Learning is the technique to enable machines to learn the thing which is not programmed explicitly.

This involve as much maths as much it involve CS.

This involve computers to get train using given data sets, and to use this trainings in predicting the property of the new data.

Example: we could train computers by feeding 1000 images of cat and 1000 images more which is not of a cat , and tell each time to the computers that an image is cat or non cat.

And then if we will show to the computer a new image , then by the given training, the computer should must be able to tell that whether the new image is cat or not.

The Process of training and predicting involve the use of specialized algorithm. We give the training data to a algorithm, and then the algorithm use the training data to tell prediction on this new test data.

These are the various machine learning algorithms:

Man-made intelligence computations are secluded into basically two orders as overseen or independent. Using named advisers for anticipate future outcomes, AI models can be used where past learning is used for new data. Beginning with the train and test dataset, the controlled learning makes a prohibitive ability to make creation regards desires. After adequate getting ready from the dataset the structure will give wants any especially new information gave . The estimation can similarly make assessment its yield with special one , arranged yield gave and find goofs to reveal extra enhancements in the model in like way to give a prevalent.

Some machine learning methods

AI calculations are isolated into essentially two classifications as managed or unaided. Utilizing named guides to foresee future results, AI models can be utilized where past learning is utilized for new information. Starting with the train and test

dataset, the administered learning makes a restrictive capacity to make creation esteems expectations. After sufficient preparing from the dataset the framework will give desires any exceptionally new info gave . The calculation can likewise make examination its yield with unique one , planned yield gave and discover blunders to roll out additional improvements in the model in like manner to give a superior.

1.1.2 Introduction to Random Forest

Random forest is that the primarily a collection of call tree.

Random Forest area unit made exploitation the technique referred to as material that work just like the illustration below.

For each call tree , take a random subsets of coaching dataset and match the choice Tree on it dataset. Repeat it as Mach time because the numbers of tree given for the forest.

Now, every of the model are a few things prophetic on a set of dataset. Since we've every which way chosen the set of the information, error on every of these model mustn't be correlate and it's vital.

If we have a tendency to take into account the common of those totally different model, we have a tendency to area unit merging the informations that is contained altogether of the model that represent the total dataset.

The main plan regarding the random forest is to supply multiple call tree's from the random subsets of dataset , since the set selected is totally random therefore we have a tendency to cowl the whole dataset .

These set area unit accustomed multiple prediction.

Long Short Term Memory :

Long immediate memory (LSTM) networks ar a sort of continual neural network capable of learning order dependence in sequence prediction issues. This is a behavior needed in advanced drawback domains like MT, speech recognition, and more. LSTMs ar a fancy space of deep learning. It ar usually exhausting to urge your hands around what LSTMs ar, and also the means terms like two-way and sequence-to-sequence relate to the arena . In this post, you will get insight into LSTMs mistreatment the words of analysis scientists that developed the ways and applied them to new and vital issues. There ar few that ar higher at clearly and exactly articulating each the promise of LSTMs and also the means they work than the specialists that developed them. We will explore key queries inside the sector of LSTMs mistreatment quotes from the specialists, and if you're interested, you will be able to dive into the primary papers from that the quotes were taken .

One of the appeals of RNNs is that the concept that they could be able to connect previous info to this task, like victimisation previous video frames would possibly inform the understanding of this frame. If RNNs may do that, they'd be extraordinarily helpful. however will they? It depends.

Sometimes, we have a tendency to solely ought to look into recent info to perform this task. for instance, contemplate a language model making an attempt to predict

consequent word supported the previous ones. If we have a tendency to try to predict the last word in “the clouds area unit within the sky,” we have a tendency to don't want from now on context – it's pretty obvious consequent word goes to be sky. In such cases, wherever the gap between the relevant info and also the place that it's required is tiny, RNNs will learn to use the past info.

1.2 Problem Statement

Apple or Tesla has continuously been on high|the highest} and top since terribly early 2013 that happened to be short once the emotional of their 1st automobile. There square measure a various quantity of factors that affects the tesla company.

There exist many reasonably issue that have an effect on the value over another factors. ought to|we ought to always|we must always} confirm these issue and therefore the feature and that we should try and predict the stock value on a given day and that we should compare it with the output that has been recorded truly .The problem of over fitting that occurred was conjointly a retardant ,which should be overcome.

These factors low, high, open, close, date, etc.

The Retail capitalist spent plenty of your time to find the investment chance. made investors will get the skilled money consulting service, the everyday retail investors cannot take this advantage and for them the prices square measure preventive .

Therefore, the retail investors can got to fathom the market by themselves and create the educated call by their own. These makes the investment plenty nerve-wracking within the fashionable society. And sadly, the human square measure irrational within the nature.

So, while not data-driven models, quantitative, call can get swayed by psychological feature bias or personal feeling, leading to the extra losses. And though the investors square measure cautious enough, most of them don't have that abundant ability to method an enormous quantity of information that is needed to form the nice judgment. The Institutional capitalist depends on the delicate model that is supported by technology to avoid the traps, however the retail capitalist don't have the access of such style of technology and that they typically realize themselves insulation behind the market.

So, this final year project targets to function a possible tool to seek out out the hidden patterns within the trend of stock value.

1.3 Objectives

To develop and style the model to predict the stock worth of company on a given day and to ascertain if it truly matches the first output or not and analyse manually that by what p.c its differs from the first.

The main goal of this Project is to assist retail capitalist that use the machine learning to assist them in navigating the quick ever-changing stock markets. The project targets to introduce & democratize the present machine learning technology for the retail capitalist.

We will perform associate degree experiment with few of the options in our dataset in predicting the adjusted terms of the corporate Tesla stock on a given day.

Those predictions don't seem to be essentially smart however these square measure solely used for the illustrative purpose, Since No predictions square measure 100% true. So, the bound and also the boundary of stock worth are shown as an example the commerce vary the capitalist would be viewing.

1.4 Methodology

Machine Learning may be a methodology of machines wherever we have a tendency to may train the pc machines to find out the items from numerous dataset. And by victimisation numerous machine learning algorithmic rule and by researching the bunch of the value functions the concept of machine learning algorithmic rule might be use within the prognostic model usually employing a relationship between those factors. With the advancement of deep learning ideas within the project that we've enforced. bound tangled points for sensible|the great|the nice} accuracy and good performance of the model. This project can use the fashionable science for the integral operate.

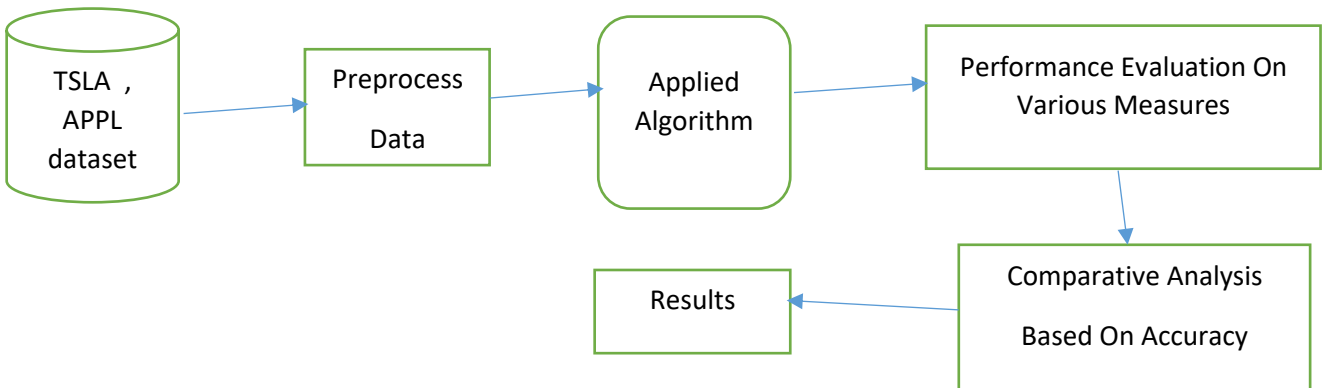


Fig. 1. Proposed Model Diagram

1.4 Organization

The model that we have implemented is Random forest classifier.

All coding has been done in Jupyter Notebook.

We have extensively used the pandas, numpy, matplotlib and sklearn packages.

The dataset contains 7 different features with 2416 instances .

Features are numerical.

At last, a best model which could classify unbalanced/imbalanced multiclass dataset is obtained.

1.5 Tools to be used

Jupyter Notebook

Pandas

Numpy

Matplotlib

Scikit Learn

MSExcel

Chapter-2 LITERATURE SURVEY

I continually accustomed assume that why some corporations increase at a rising like Facebook and Google whereas some corporations like yahoo and Tata Nano don't grows therewith pace in spite of a decent market strategy.

After learning few factors one thing that clicked in my mind is that the success of any company heavily depends upon it's stock worth.

That's why i made a decision to develop a system will|which may|which might} predict the output properly and if i will be {able to|i'll} be able to do thus then it can yield in an exceedingly nice success for a great deal of corporations.

From the analysis paper "Machine Learning accessible price Trend Forecasting" written by Y. Dai and Y.Zhang in Stanford, they used selections like letter relation, PX_volume, PX_EBITDA., 10-day volatility, 50-day moving average, etc. to predict the next-day stock price associate degreed an extended stock price .

The machine learning algorithms used inside the analysis ar activity Regression, man of science Discriminant Analysis, Quadratic Discriminant Analysis, and SVM.

The accuracy relation is created public as a results of the vary of days that the model properly classified the testing info over the whole vary of testing days. With the short term model predicting later day stock price, it's extremely low accuracy, the Quadratic Discriminant Analysis is that the best among all models, it scored a fifty eight.2% accuracy.

With the long run model predicting later n days stock costs, the longer the timeframe, the higher within the accuracy for SVM. With a time window of forty four days, the SVM model's accuracy reached seventy nine.3%. except for that, it utterly was found that by increasing the amount of selections, the accuracy enlarged. once all of the sixteen choices were used, the accuracy of the model reached seventy 9, whereas it fell to sixty fourth once alone eight choices were used, and fifty fifth if only one feature was used.Our project will investigate however the timeframe would have an effect on the accuracy of worth predictions of various models.

As models ought to reach an exact threshold to own significance for the users to figure as a reference, it's essential for North American country to optimize our model to work out what the best parameters and model structure ar for our stock worth prediction purpose.

The analysis paper “Predicting stock and stock indicant movement exploitation Trend settled knowledge Preparation and machine learning techniques” written by J. Patel, S. Shah, P. Thakkar, and K. Kotecha for the “Expert Systems with Applications” international journal incontestable how to use trend settled knowledge to predict stock worth movement .

They conducted experiments in exploitation ten technical indicators’ signals as inputs, then they use prediction models to predict whether or not the stock can go up or down within the returning ten days, Technical analysis indicators embody SMA, EMA, Momentum, random SK, random SK, MACD, RSI, etc.

The prediction models they need used embody ANN, SVM, Random Forest, and Naive theorem models. The model outputs “up” or “down” movement signals.

Experiments have shown random forest scored the best performance with eighty three.56% accuracy with their inputs

Researches have shown that large-scale evolution will auto-generate neural network model architectures and hyperparameters with performance comparable progressive human-designed models.

In a analysis in 2017, a large-scale evolution for locating image classification neural networks was run.

It started with a large population of randomised straightforward 1-layer models, then slowly evolved the population by removing a poor model and generating a replacement model by mutating some parameters of a decent model in every iteration.

After many hours of running the rule with vast computing power, most models within the population achieved progressive results on CIFAR datasets.

In every iteration, solely a straightforward mutation that modified one parameter was applied, that allowed looking in an exceedingly giant search area.

The paper showed the of likelihood finding sensible by models victimization various machine power to exchange human-machine learning specialists and has set the inspiration of democratizing machine learning with AutoML.

The project’s problem is getting ready to predict the stock price for succeeding day. “next day” is chosen as a resultof the timeframe as short term.price movements tend to bank extra on trend momentum and.value pattern, whereas long-standing

time price movements depend on the fundamentals of a stock (e.g. company management capabilities, revenue model, market demand, economics factors, etc.).

RMSE and completely scores different square measure made public to provide extra in-depth insights on a model certainty performance and finance-domain-based comparisons between models for investors

For the analysis side, the system is meant to be as durable as potential facilitate to model testing. each model are going be to made public by a attempt of model decisions and input decisions, specifying the model configurations and so the inputs it takes.

This accelerates the tactic of testing out fully completely different model and/or input configuration mixtures.

Two key options area unit chosen because the input. initial may be a fixed-length list of some raw historical knowledge like stock worth and daily proportion modification.

The mounted length chosen specifies the length of the period to appear back from these days once predicting future stock costs.

pertaining to the principle of technical analysis, because the stock worth reflects all relevant data, a technical analyst would concentrate on the mercantilism pattern of the stock instead of the economic fundamentals and company fundamentals.

Therefore, by obtaining a amount of historical stock costs because the input for the coaching model, it may well be a chunk of helpful data to find the mercantilism patterns and thence predicting the trend of future stock costs.

Given a group lookback amount, it's assumed that the worth movement patterns that area unit prognostic would occur within the such period.

The feature of second input is arithmetic moving averages. one among the plain approaches for retail investors the trend being spotted of the market is thru moving averages. With the sturdy system style, completely different amount of moving averages may well be used because the input into the model for stock worth

prediction, for instance, a group of twenty, 30, 300, 600 days moving averages, that area unit usually employed by investors.

For 1-day prediction, there'll be one output unit that is that the stock worth within the following day.

The predicted stock worth of can then be the input of prediction being succeeded, to predict the stock worth within the second day, the method repeats till all ten predictions area unit generated.

2.1 Methodology – Design

After relevant stock knowledge square measure retrieved from the third-party knowledge supplier through the cloud, the backend pre-processes the info and builds the models.

After that, predictions square measure created victimisation appropriate model and algorithmic rule like here we've got used random forest algorithmic rule that uses call trees for prediction.

Long remembering (LSTM) networks square measure a form of repeated neural network capable of learning order dependence in sequence prediction issues.

This is a behavior needed in advanced drawback domains like MT , speech recognition, and more.

LSTMs square measure a fancy space of deep learning. It square measure usually exhausting to urge your hands around what LSTMs square measure, and also the method terms like duplex and sequence-to-sequence relate to the world .

In this post, you will get insight into LSTMs victimisation the words of analysis scientists that developed the strategies and applied them to new and necessary issues.

There square measure few that square measure higher at clearly and exactly articulating each the promise of LSTMs and also the method they work than the consultants that developed them.

We will explore key queries inside the sphere of LSTMs victimisation quotes from the consultants, and if you're interested, you will be able to dive into the primary papers from that the quotes were taken .

Support Vector Machines :

The arrangement insinuates a planned AI technique are may be utilized for together backslide and portrayal plans. Nevertheless, this is by and large applied to gathering estimations. For Support vector, our point is too arranging every datum spot with coordinates various dimensional spaces in which alludes an amount ascribes in the data any place each belonging has a spot with a specific encourage. After this, we will likely complete get-together through glancing through a hyperplane which disconnects the two classes uncommonly fine.

Reinforce vectors imply information which is close the hyperplane and impact the circumstance similarly as a preface to the hyperplane. Utilizing the directions, I can development the limits inside the classifiers. Annihilating assistance directions will change the spot on the hyperspace.

2.2 Decision Trees

As the name suggests, decisions are made by splitting data into groups that are homogenous within themselves but heterogenous with others.

A decision tree is more of a real tree kind of chart with hubs speaking to where we pick a quality and pose an inquiry; edges speak to the appropriate responses the to the inquiry; and the bottom

most without child nodes speak to the genuine yield or class mark. They are utilized in non-direct basic leadership with basic straight choice surface.

Decision trees organize the models by organizing them from the roots to a certain leaf center down the tree, with both the leaf hub offering the model a location. -- hub in the tree is conducted as an application for some property, and each edge falling out of that hub is compared to one of the most common responses to it. This process is calling itself again and again in nature and is rehabilitated for each stable subtree.

How about we represent this with assistance of a model. We should expect we need to take on badminton on a specific day — state Saturday — in what capacity will you choose whether should to play or not. Suppose you go and verify if it's sweltering or cold, check the speed of the breeze and mugginess, how the climate is, for example is it bright, shady, or stormy. You consider every one of these components to choose in the event that you need to play or not.

The baseline methodology that is being in the decision trees is known as the algorithm ID3 (by Quinlan). The ID3 method constructs trees which take decision of it own by using a greedy, from top-down approach. Shortly, the algorithm levels have been:-Pick the best attribute for the NODE — Assign A as the decision attribute (test case). - For every value of node , make a new further of the NODE. – align the order of the training data to the appropriate further node leaf. - If given cases are accurately predicted, then don't do forward and else loop over the new further nodes.

The next major questions, now, is how to select the best characteristic. For ID3, we consider the best attributes in term of which attributes have the most knowledge benefit, a metric which communicates how nicely an attribute divides the data into classes based on prediction.

A split is predicated on some criteria (score) that achieves the most effective attainable division of information in an exceedingly method that optimizes for that criteria.

we tend to decision this criteria, data gain that essentially tells U.S. what proportion our criteria improves as we tend to carry on creating those splits.

The next major questions, now, is how to select the best characteristic. For ID3, we

consider the best attributes in term of which attributes have the most knowledge benefit, a metric which communicates how nicely an attribute divides the data into classes based on predcition.

This is one in all my favorite rule and that i use it quite often. it's a sort of supervised learning rule that's principally used for classification issues.

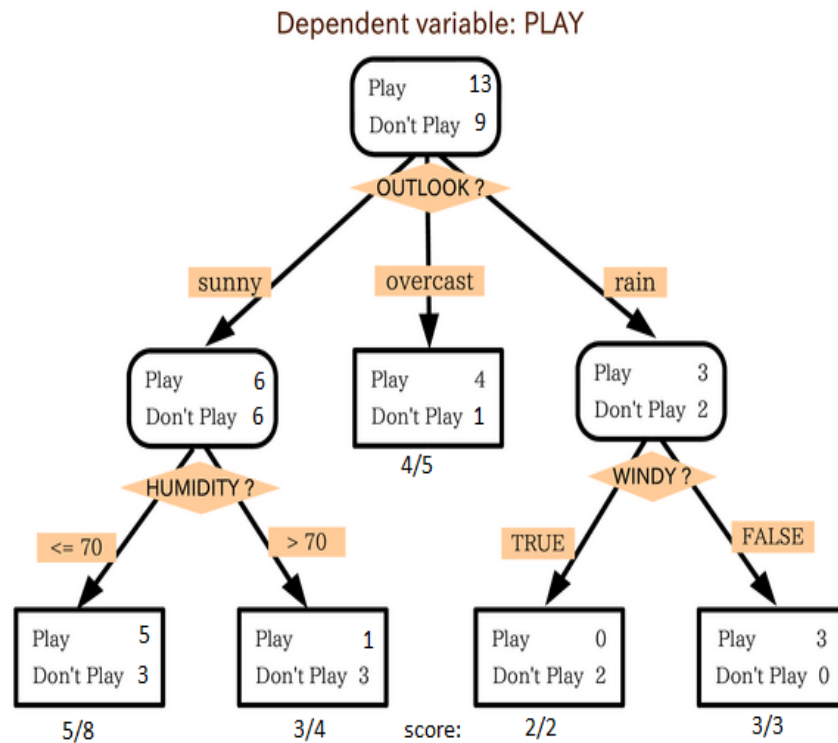


Fig 2 Decision Tree

Above image shows however population is assessed into four totally different teams supported multiple attributes to spot 'if they're going to play or not'

The best thanks to perceive however call tree works, is to play Jezzball – a classic game from Microsoft (image above). basically, you've got space|a space} with moving walls and you would like to make walls specified most area gets cleared off while not the balls.

So, on every occasion you split the area with a wall, you're attempting to make a pair of totally different populations with within the same space. call trees add terribly similar fashion by dividing a population in as totally different teams as doable .

Chapter-3 SYSTEM DEVELOPMENT

In this Chapter, we used random forest machine learning algorithm on our dataset to predict stock prices .

The main reason behind using Random forest algorithm in this project is that it works based upon decision tree which chooses some of the best attributes that mostly effects the problem. Hence it may predict the best output.

3.1 Random Forest Classifier

Random Forest Classifier is supervised machine learning algorithms, it creates multiple call trees and eventually merges them along to present the ultimate prediction. Multiple call trees ar supported varied conditional statements that helps in nice accuracy. because of such nice variety of random call trees, we are able to avoid over fitting .

Random forest ar associate ensemble learning methodology for classification, regression and alternative tasks that operate by constructing a large number of call trees at coaching time and outputting the category that's the mode of the categories (classification) or mean prediction (regression) of the individual trees. Random call forests correct for call trees' habit of overfitting to their coaching set.

The first algorithmic program for random call forests was created by Tin Kam metallic element exploitation the random topological space methodology, which, in Ho's formulation, may be a thanks to implement the "stochastic discrimination" approach to classification projected by Eugene Kleinberg.

An extension of the algorithmic program was developed by Leo Breiman and Adele bargainer, WHO registered "Random Forests" as a trademark . The extension

combines Breiman's "bagging" plan and random choice of options, introduced 1st by metallic element and later severally by Amit and Geman so as to construct a set of call trees with controlled variance.

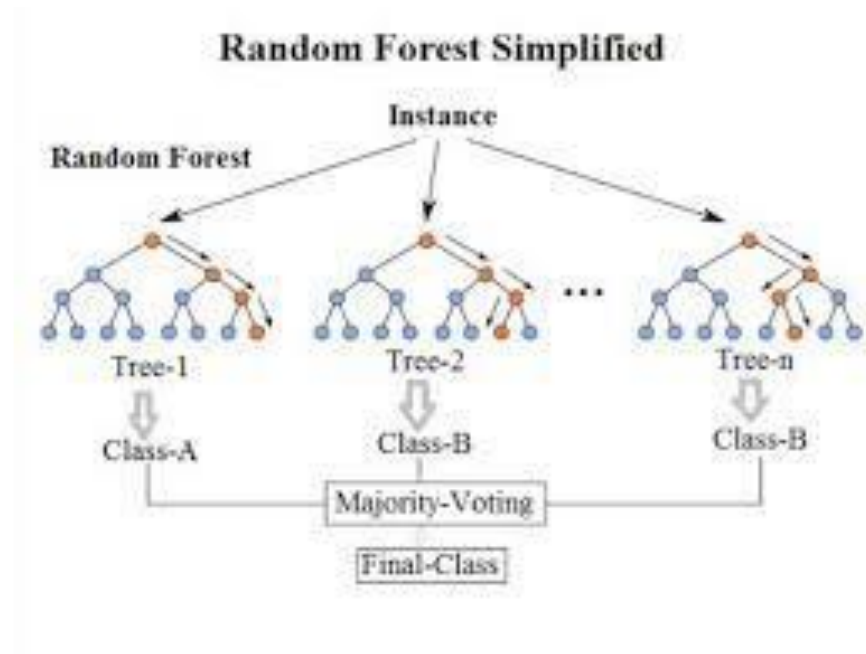


Figure 4-Classification in Random Forest

3.2 Task Performed

3.2.1 Dataset Description

We collected tesla dataset from Kaggle and performed computer simulation on one dataset. Dataset is a Apple and Tesla Price Prediction dataset. Dataset contained 7 different features with 2416 instances. The features describe date, open , high, low, close , adj close, volume. The dataset contains features expressed on nominal, ordinal, or interval scales.

3.3 Random Forest

According to Breiman's approach, every call tree within the forest is made by 1st choosing willy-nilly, at every node, a tiny low cluster of input coordinates (also referred to as options or variables hereafter) to separate on and, secondly, by scheming the simplest split supported these options within the coaching set. The tree is full-grown victimisation CART methodology to most size, while not pruning. This randomisation theme is mixed with fabric to resample, with replacement, the coaching information set anytime a brand new individual tree is made.

Random forest is also a supervised classification machine learning algorithmic program that uses ensemble technique. Simply put, a random forest is made from various call trees and helps to tackle the matter of overfitting in call trees. These call trees square measure at random made by choosing random options from the given dataset.

Random forest arrives at a alternative or prediction supported the utmost range of votes received from the selection trees. the result that is received , for a most range of times through the assorted call trees is taken into consideration as a result of the ultimate outcome by the random forest.

LSTM :

1.Import the required libraries

2. we fetched datasets from kaggle therefore we'd like to line the beginning and finish dates and pass these parameters to the operate for winning the info.

3. we visualised the fetched information within the higher than step. For simplicity, solely the day-wise closing costs are unreal.
4. There are twelve columns within the fetched information. Several of the columns don't seem to be of our interest therefore solely important columns are hand-picked to make the most dataset.
5. we have a tendency to preprocess the info order in to organize it for the LSTM model. the info fetched in the first step is employed for coaching purpose solely. For testing purpose, totally different information are going to be fetched later.
6. we have a tendency to add additional LSTM layers and adjusted the so as dropout to enhance the the model's accuracy.
8. Now, our model is trained and wishes to be tested on the testing information. For this purpose, fetch the new information for a special amount. Preprocessing steps are similar as we've got finished coaching information.

3.4 Feature Importance

Out of all the 7 features Date, open, close, high, low, adj close and volume which are going to predict stock price. These all features play an important role in prediction of price

3.5 Test Plan

The check set up for this prognosticative model stock worth has been taken from actual information of tesla stock worth dataset from kaggle. we've developed numerous ways or approach towards developing a systemized and synchronised method of victimisation the data for the aim of our model. the data specifically targets the crucial points needed for our analysis. Moreover, the check set up is in accordance with our model and should be useful for any enhancements or developments in our model or models related to this kind of analysis.

3.6. Data Set

For the purpose of this study, Tesla dataset is considered as it is the best dataset for the present study. It contains 2416 records. Every record has 7 attributes out of those one attribute is object variable. All 7 attributes contains only numeric data and are not null. Each record contains information about single day.

Date:

Open:

High:

Low:

Close:

Adj close:

Volume:

Outcome:

3.7 Data Preprocessing

Data Preprocessing is one of the most necessary options needed for the coaching of the model. Since all the columns or rows will not be helpful for the model or information | the information } set that is offered is not among the shape throughout that it square measure usually used for the coaching of the machine altogether these cases data preprocessing could be a crucial issue that determines the healthy begin of the models.

Data preprocessing is also a method that is used to point out knowledge to helpful format. It square measure usually through filling empty values, changing categorical variables to binary format through cryptography.

```

from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler(feature_range=(0,1))
df1=scaler.fit_transform(np.array(df1).reshape(-1,1))

```

Features	values
date	2416 non-null object
pen	2416 non-null float64
high	2416 non-null float64
low	2416 non-null float64
close	2416 non-null float64
Adj close	2416 non-null float64
volume	2416 non-null int64

Table 3.7 features and their values

3.8 Algorithm

```
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size = 0.3)
```

```
Model=RandomForestRegressor(n_estimators=40,bootstrap=True,min_samples_leaf=25)
```

```
model.fit(X_train, y_train)
```

```
y_pred=model.predict(X_test)
```

```
conf = model.score(X_test, y_test)
```

```
print(conf)
model.fit(X,y)
predictions = model.predict(X_Check)
print(predictions)
```

Chapter-4 PERFORMANCE ANALYSIS

After an excellent and warm learning expertise and surfing careful observations of our model, considering a lot of totally different approach for a specific result, we will finally measure the key results and explore the correct outcomes. The Machine Learning rule taken into consideration well-tried to be very helpful for the gathering of right information for validation.

For the machine learning model we've created. we'll calculate totally different hyper parameters and check all for the validity of the model. Accuracy MSE and RMSE square measure one in all the foremost vital performance measures. These is calculated simply.

Some operations and data manipulations performed are

```
df.shape
```

```
output- (2416,7)
```

#plotting the data

```
plt.figure(figsize=(16,8))  
plt.title('Close Price History')  
plt.plot(df['Close'], color='red')  
plt.xlabel('Date', fontsize=18)  
plt.ylabel('Close Price USD', fontsize = 18)  
plt.show()
```

```
output-
```

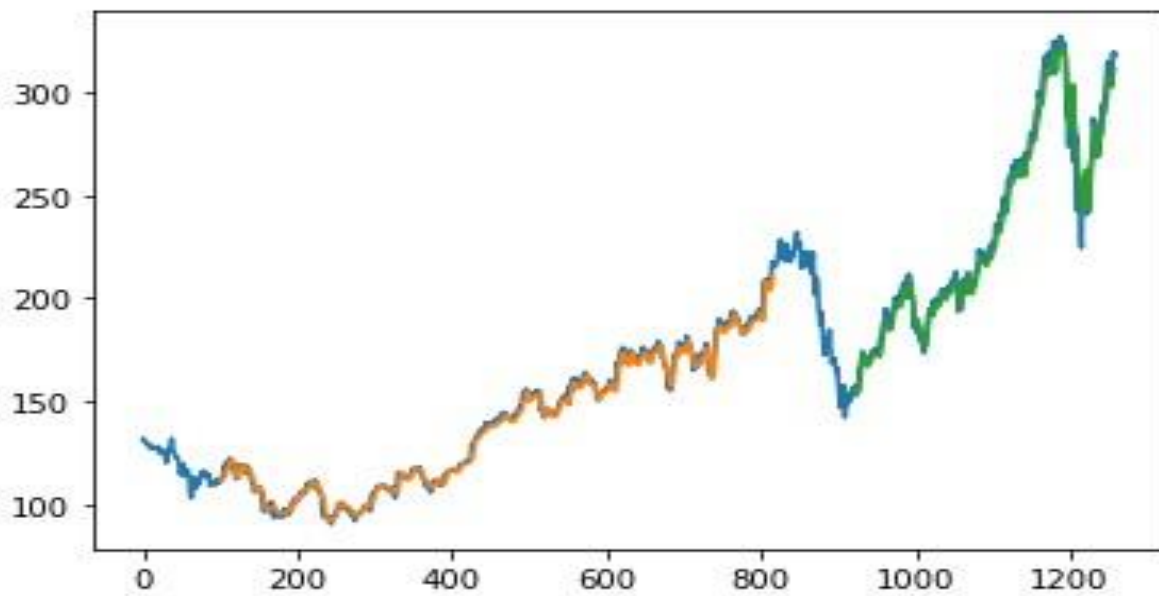



Fig 4.2 Graph

4.1 If test size =0.2

X_train.shape

Output-

(1884, 4)

```
#Calculate the confidence value by applying the model to testing data
```

```
conf = model.score(X_test, y_test)
print(conf)
```

output-

0.9316009542537967

#Plot the results

```
result.plot(figsize=(20,10), linewidth=1.5)
plt.legend(loc=2, prop={'size':20})
plt.xlabel('Date')
plt.ylabel('Price')
```

output-

```
Text(0, 0.5, 'Price')
```



Graph showing forecasted value of price with respect to date

4.2 If test size =0.3

```
X_train.shape
```

```
Output – (1670, 4)
```

#Calculate the confidence value by applying the model to testing data

```
conf = model.score(X_test, y_test)  
print(conf)
```

```
output-
```

```
0.9042539650830713
```

#Plot the results

```
result.plot(figsize=(20,10), linewidth=1.5)  
plt.legend(loc=2, prop={'size':20})  
plt.xlabel('Date')  
plt.ylabel('Price')
```

```
output-
```

```
Text(0, 0.5, 'Price')
```



Plot of forecasted price vs date

4.2 Results

Above figures represent different performance values of our model algorithm calculated on various measures like confidence value is having a great difference in both the cases. From above analysis it is analyzed that random forest is best suited for this prediction showing the maximum accuracy.

As we used a pre-modified and neat and clean dataset that was already feature-engineered so we don't need to bother about any discussion around that. And although, we selected and worked on training random forest algorithm a pretty decent model, so we got better results. We can even do better by using more features and hyper parameters. A decent dataset and more organized features will give best results.

5.1 Conclusions

One of the necessary real-world issues is that the prediction of stock costs for any company like Tesla and exploitation these results for the expansion of company and conjointly to be employed by the retail investors available market. during this study, systematic efforts square measure created in coming up with a system which ends within the prediction of tesla stock costs. throughout this work, one machine learning formula was studied and evaluated on varied measures. Experiments square measure performed on TSLA Stock value info. Experimental results confirm the adequacy of the designed system with considerable accuracy exploitation random forest formula. In future, the designed system with the used machine learning formula are often wont to predict stock costs for alternative corporations. The work are often extended and improved for the automation exchange and stock value analysis together with another machine learning algorithms.

5.2 Future Scopes

In future, we tend to aim to think about a lot of options like firms future plans and checking their dependency on share capital. we tend to conjointly aim to extend samples to urge a lot of correct results for giving correct steering to retail investors . The a lot of information we tend to regain square measure the samples for cross validation for the cases run on the model.

This will be a particularly necessary tool within the futures to come back and can be thought-about because the biggest inventions so as to unravel exchange issues. the speed of individuals with exchange investment can step by step rise and therefore this may be a profit to the grouping.

Investors and prediction manufacturers will use it for his or her own validation, for practice etc. folks having a doubt or simply wished to own a traditional investment will take its thought into the account.

Hence by fashionable technologies and advancement in future articles this downside would have an improved resolution and timely use to the current system are helpful.

Therefore with facilitate of machine learning formula this model is capable of finding a timely resolution within the mean course of your time. what is more for future add this field there square measure several observations that may be taken {into account|under thought|into consideration} for helpful consideration.

Apart from that this model will be deployed as associate degree internet application or robot application which might be offered without charge and may be tested by real users. These results will facilitate America rising the model if we tend to shift towards reinforcement learning.

In today's time wherever milliliter is growing at higher rate, With correct steering and victimization of Principal part Analysis, This model accuracy will be reached to a mark of ninetieth.

REFERENCES

1. https://en.wikipedia.org/wiki/Random_forest
2. https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.html
3. <https://github.com/prnvg/Stock-Price-Prediction>
4. https://www.cse.ust.hk/~rossiter/fyp/103_RO4_Final_201819.pdf
5. <https://analyticsindiamag.com/hands-on-guide-to-lstm-recurrent-neural-network-for-stock-market-prediction/>
6. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
7. <https://lazyprogrammer.me/data-science-interview-questions-random-walk-hypothesis-and-stock-price-prediction/>
8. www.cse.ust.hk
9. www.irjet.net
10. machinelearningmastery.com
11. K. VijayaKumar, B. Lavanya, I. Nirmala, S. Sofia Caroline. "Random Forest Algorithm for the Prediction of Diabetes" , 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019
12. Deepti Sisodia, Dilip Singh Sisodia. "Prediction of Diabetes using Classification Algorithms" , Procedia Computer Science, 2018
- 13.