

# Prediction Stock price using Data science technique

Submitted in partial fulfillment of the requirements for the award  
of Bachelor of Engineering Degree in Computer Science and  
Engineering

By

**Jophra Joseph  
(38110218) &  
Ieshwarya.G  
(38110203)**



**DEPARTMENT OF COMPUTER SCIENCE AND  
ENGINEERING**

**SCHOOL OF COMPUTING**

**SATHYABAMA**

**INSTITUTE OF SCIENCE AND TECHNOLOGY  
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade "A" by NAAC**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119**

**MAY - 2022**



**SATHYABAMA**  
**INSTITUTE OF SCIENCE AND TECHNOLOGY**  
**(DEEMED TO BE UNIVERSITY)**  
**Accredited with Grade “A” by NAAC**  
(Established under Section 3 of UGC Act, 1956)  
JEPPIAAR NAGAR, RAJIV GANDHI SALAI  
CHENNAI– 600119  
[www.sathyabama.ac.in](http://www.sathyabama.ac.in)



---

**DEPARTMENT OF COMPUTER SCIENCE AND  
ENGINEERING**

**BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the bonafide work of **Jophra Joseph (38110218)** and **Ieshwarya.G (38110203)** who carried out the project entitled “**Prediction Stock price using Data science technique**” under my supervision from DEC 2021 to MAY 2022.

Internal Guide

**Mrs.M.s.Roobini,M.E., Ph.D.**

Department

Head of the

---

Submitted for Viva voce Examination held on \_\_\_\_\_

**InternalExaminer**

**ExternalExaminer**

## **DECLARATION**

We, **Jophra Joseph and Ieshwarya.G** hereby declare that the project report entitled **Prediction Stock price using Data science technique** done by us under the guidance of **Mrs.M.s.Roobini,M.E., Ph.D.**, is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering.

**DATE:**

**PLACE:**

**SIGNATURE OF THE CANDIDATE**

## ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D, Dean**, School of Computing, **Dr. S. Vigneshwari, M.E., Ph.D.** and **Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department of Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Mrs.M.s.Roobini,M.E., Ph.D.**, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

## **ABSTRACT**

Persuasion of the topic is aimed at predicting the stability of future markets stocks. The movement of future market evolution has been studied by many researchers. Data fluctuates in stocks, so it becomes a source of efficiency. The purpose of this article is to propose a machine-learning-based model for predicting the stock price Index value from comparing supervise classification machine learning algorithms by predicting stock price increases or stable states. Moreover, the aim of this study is to compare and discuss the performance of various machine learning algorithms based on the data from the transport traffic department. For this project, logistic regression, random forest, decision trees, and naive bayes theorem were used to calculate stock price predictions. An evaluation of the proposed machine learning algorithm technique has been made with regard to identifying the confusion matrix, categorizing data based on priority and comparing it with the best levels of precision, recall and the F1 score.

(i)  
**LIST OF FIGURES**

| Figure No. | Figure name                              | Pg.no |
|------------|------------------------------------------|-------|
| 4.1        | OVERALL ARCHITECTURE                     | 8     |
| 4.2        | PROPOSED ARCHITECTURE DIAGRAM            | 8     |
| 4.3        | DROP THE GIVEN DATA                      | 9     |
| 4.4        | IN THE WAKE OF ERASING THE GIVEN DATA    | 10    |
| 4.5        | DATASET                                  | 10    |
| 4.6        | STOCK STABLE GRAPH PREDICTION            | 11    |
| 5.1        | PIE CHART                                | 11    |
| 5.2        | GRAPH OF DISTRIBUTION                    | 12    |
| 5.3        | PIE CHART                                | 12    |
| 5.4        | HISTOGRAM OF ALGORITHM                   | 13    |
| 5.5        | HEATMAP                                  | 13    |
| 5.6        | ACCURACY RESULT OF LOGISTIC REGRESSION   | 14    |
| 5.7        | GRAPH OF LOGISTIC REGRESSION             | 14    |
| 5.8        | CONFUSION MATRIX FOR LOGISTIC REGRESSION | 15    |
| 5.9        | ACCURACY RESULTS OF RANDOM FOREST        | 15    |
| 5.10       | GRAPH OF RANDOM FOREST                   | 16    |
| 5.11       | CONFUSION MATRIX FOR RANDOM FOREST       | 16    |
| 5.12       | ACCURACY OF DECISION TREE                | 17    |
| 5.13       | GRAPH OF DECISION TREE                   | 17    |
| 5.14       | CONFUSION MATRIX FOR DECISION TREE       | 17    |
| 5.15       | ACCURACY OF NAIVE BAYES THEOREM          | 18    |
| 5.16       | GRAPH OF NAIVE BAYES THEOREM             | 18    |
| 5.17       | CONFUSION MATRIX FOR NAIVE BAYES THEOREM | 19    |
| 5.18       | SYSTEM ARCHITECTURE                      | 20    |
| 5.19       | OVERALL CONFUDION MATRIX                 | 21    |
| 5.20       | OVERALL ACCURACY                         | 22    |

(ii)

## LIST OF CONTENTS

| Chapter no. | Title                                          | Page no. |
|-------------|------------------------------------------------|----------|
|             | ABSTRACT                                       | I        |
|             | LIST OF FIGURES                                | ii       |
| 1           | INTRODUCTION                                   | 1        |
| 2           | LITERATURE SURVEY                              | 1        |
| 3           | 3.1 AIM                                        | 5        |
|             | 3.2 SCOPE OF THE PROJECT                       | 5        |
| 4           | 4.1 MATERIALS                                  |          |
|             | 4.1.1 HARDWARE REQUIREMENTS                    | 6        |
|             | 4.1.2 SOFTWARE REQUIREMENTS                    |          |
|             | 4.1.3 NON-FUNCTIONAL REQUIREMENTS              |          |
|             | 4.2 SYSTEM ANALYSIS                            | 7        |
|             | 4.3 SYSTEM METHODOLOGY                         | 7        |
|             | 4.4 EXISTING SYSTEM                            | 8        |
|             | 4.5 FEASIBILITY STUDY                          | 8        |
| 5           | 5.1 EXPLORATION DATA ANALYSIS OF VISUALIZATION | 11       |

|   |                                                                                   |          |
|---|-----------------------------------------------------------------------------------|----------|
|   |                                                                                   |          |
|   | 5.2 ALGORITHM AND TECHNIQUE                                                       | 13       |
|   | 5.3 COMPARING ALGORITHM WITH<br>PREDICTION IN THE FORM OF BEST<br>ACCURACY RESULT | 19<br>20 |
|   | 5.4 SYSTEM ARCHITECTURE                                                           |          |
| 6 | 6.1 CONCLUSION<br>6.2 FUTURE WORK                                                 | 22       |
|   | REFERENCES                                                                        | 23       |
|   | APPENDIX                                                                          |          |
|   | 1. CODE                                                                           | 24       |
|   | 2. OUPUT SCREENSHOTS                                                              | 48       |

## **CHAPTER 1**

### **INTRODUCTION**

The field of data science is an interdisciplinary one involving use of scientific methods, algorithms, processes, and systems to extract knowledge and insights from structured and unstructured data and apply knowledge to a wide range of application domains. A machine learning-based stock price prediction service helps you predict the future value of stocks and other financial assets you trade on an exchange. The entire purpose of predicting stock prices is to make significant profits. It's challenging to predict the performance of the stock market. The Stock Market Prediction serves as a tool for predicting the movement of a financial exchange's stock price in the future. Investors can earn more profit when they accurately predict share price movements. The machine-learning process enables highly automated efficiency in a variety of ways. Predicting future results, making recommendations for future action, and revising decisions according to successful outcomes are all possible with the system.

## **CHAPTER 2**

### **LITERATURE SURVEY**

Artistic examination is a rundown of articles that analyze key components of present-day science or the procedure utilized in the subject. The source is the second wellspring of data that is distributed, and at times the data that is distributed at specific occasions in a specific part.[1]. Lobna Nassar,Ifeanyi Emmanuel Okwuchi, Muhammad Saad in their work the ARIMA model has a high mistake rate, its exhibition is lower than the standard ML. Likewise, Gradient Boosting (GB) is typically the best since you get a little MAPE mistake. At long last, the smooth activity of the DL LSTM is superior to any remaining standard ML FP tests (Watermelon and Bok Choi). This is because of the way that there are not many business sectors for these two FPs that leave us with data near nature during following. The models that work

best as far as absolute worth are a mix of better DL, ATTCNN-LSTM, and ML, and easier DL models as far as value suspicions, particularly later input [2]. Xie Chen, Deepu Rajan, Chai Quek in their article depicts the construction of the Hybridstein-Wiener Hybridstein Hammerstein-Wiener. Dubious obligations and neurological choices dependent on unsure principles set up during preparing. This requires preparing data to have the option to appropriately mirror the conduct of the whole framework. Nonetheless, press assessments might vary in the circumstance of the planned time designation. What's more, preparing data can be gotten internet based when the assessment data is as an adjustment of the type of data trade at times, like financial issues. Notwithstanding powerful correlations with the asymptotic benefits of the Hamerstein-Wiener model for mathematical factors, the faint sensory system isn't viable with the consistent state framework. Model execution is surveyed on two arrangements of information showing monetary expenses. [3]. Rubi Gupta, Min Chen where they investigate stockTwits information and comprehend the effect of stock costs on the securities exchange. To start with, we utilize two kinds of feelings in this action: harassing (great) and labor (terrible). Adding an unbiased viewpoint can decrease commotion and increment working environment legitimacy. Second, our examination is restricted to five organizations. A wide scope of enterprises, or everything StockTwits records, can give definite data, and the key is to utilize stock costs proficiently. They utilize the enthusiastic name given by StockTwits clients since this variant is a genuine learning movement. StockTwits is an extremely new internet-based webpage that is turning out to be progressively famous among clients to analyze discussions and feelings around the stock and monetary business sectors. He gave modest proof that feelings positively affect value suspicions. [4]. Jiannan Chen, Junping Du, Feifei Kou in their paper Climate estimates give the premise to investigating monetary data. At the point when stock costs vacillate throughout a significant stretch of time, stocks have astounding data. Because of the idea of stock data, this report incorporates an incomplete expense-based calculation (STPA). To get data about the program, we take a Bidirectional Gated Recurrent Unit (BGRU) with thought pieces. All estimations depend on a factual device (STPA), which requires computations in each of the three areas. As such, the capacity worth of the

vector structure change is the level, and the BGRU contains the worth of the short and the normal worth change. The STPA, which utilizes a two-way entryway repeater unit to get more cash-flow, tells the best way to construct a material utilizing the STPA technique. The STPA method performs better compared to normal stock change computations in information assortment, demonstrating that the procedure is sufficient. In light of the aftereffects of the review, the STPA procedure works better compared to existing methods in deciding changes in the level of factual information, showing the strength of the proposed technique. [5]. Rahma Firsty Fitriyana, Brady Rikumahu, Andry Alamsyah they offer cost is a significant piece of the organization's security activities and measures up to the assumptions when contrasting the offer cost and the variables influencing it. The issue is that there are innumerable things that can be utilized to uncover stock costs, which permits a monetarily propelled individual to pick things that can be utilized to uncover stock costs. The review remembered an examination of key parts for request to make a critical commitment to the worth of offers without compromising the data and execution of the five associations. In light of the way that there are many elements that influence the worth of the offers, including macroeconomic variables, the review looks at the basic factors that decide the worth of the offers in addition to new resources and incomes to accomplish genuine outcomes, like macroeconomic elements. the strategy can be utilized. eliminated from this review. The following test is to incorporate these factors to decide the effect on stock costs. [6]. Vargas, Carlos E. M. dos Anjos, Gustavo L G. Bicharae Alexandre G. Evsukoff proposed an approach as far as data, this action involves everyday inside and out research on stock costs dependent on financial capital and exceptional divulgence.

Set 1: stochastic (% K), stochastic (% D), stochastic (% K), stochastic (% D), stochastic (% K), uncommon power, speed increase, William (% R), the distinction among 5 and/Distribution (A/D) is variable

Set 2: Ratio Transfer Ratio, Moving Average Ratio, Strength Index, Number, and Bollinger Bands to be changed. Studies have shown that convolutional neural networks (CNNs) are greater at recovering literary data than an organization of recurrent neural networks (RNNs), and that RNNs are better at getting to explicit

articles. There are two methods for imparting data, just as a perplexing perspective on trade along these lines. As indicated by this article: CNN's Mixed Models for Financial Information and Short-Term Capacity for SI-RCNN explicit scales, the LSTM network for explicit scales just, is classified "IRNN". The consequences of each example are utilized as a punishment commitment. Regardless, the gauge is that costs will ascend sooner rather than later. The proposed procedure delineates the vital message of cash in planning results, and the changed arrangement for explicit markers doesn't exist.[7]. Marios Mourelatos, Thomas Amorgianiotis, Christos Alexakos, Spiridon Likothanassis. In their proposed method involved with foreseeing and showing a rundown of monetary forms is troublesome. This issue is extremely challenging because of its temperament. It's unique, power and clamor. The present reality faces many difficulties. Also, they need to retaliate. There are many sorts of illegal tax avoidance in various nations. This is undeniably challenging, with countless data that won't ever be neglected. Here we want current AI arrangements. This article will tell you the best way to exhibit and share Deep Learning in light of monetary use.[8]. C. Narendra Babu and B. Eswara Reddy had implemented most ideal way to join straight and level examples is to give a decent method for pausing. Because of the different expectation programs, one straight or non-direct model is great, and the estimation data is gathered. The engine generally comprises of an autoregressive and a pseudo-neural design. In this article, we will investigate the new outlines; The completed framework will anticipate your experience with the ARIMA-ANN stream diagram. [9] Rudra Kalyan Nayaka, Debahuti Mishraa, Amiya Kumar Rathb In their paper depicts the idea of Support Vector Machine (SVM) true to form in India. The motivation behind this article, for instance, is to acquire inside and out data and other monetary data that remembers two documents for Indian arrangement. The Bombay Stock Exchange (BSE Sensex) and CNX Nifty utilize extraordinary control procedures and hardware. Hindering the trading of data about esteem, startling and refreshed data. Request driven interest from SVM and KNN depends on this information. The given model was recently composed utilizing Error and afterward adjusted later correlation. For instance, Interestingly, FLIT2NS and CEFLANN. [10]. Chung-Ho Su, Ching-Hsue Cheng presented sitting tight at stock costs is a hot issue for contributors,

dealers, and traders. Don't sweat it, elusive. Thusly, the stock cost is given more turnaround time, except for a drawn-out pragmatic technique with few issues. Thusly, this paper shows ANFIS (Adaptive Neuro Fuzzy A Time Series) innovation dependent on Strategic Comparative Analysis.[11].Hari Krishnan Andi.He proposedAn Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model. Using a normalized dataset, the proposed research study aims to improve the accuracy of bitcoin price predictions. The bitcoin market has been exploited for abnormal profits using Machine Learning and Artificial Intelligence (AI).[12].Dr. Jennifer S. Raj.They proposed Evolutionary Computing Techniques for Resolving Load Dispatch Problem.This problem scope aims to meet the demand while meeting the various constraints in the operation, minimizing operational costs.[13].C Anand. He proposed Comparison of Stock Price Prediction Models using Pre-trained Neural Networks A model is trained with closing prices from the National Stock Exchange (NSE) of India.[14].Dr. Akey Sungheetha, Dr. Rajesh Sharma R. In their paper Trans capsule model for sentiment cclassification. Based on the sentiment polarity of the text in the document, aspect-level sentiment classification determines and categorizes the text[15]. Dr. J. Samuel Manoharan Study of Variants of Extreme Learning Machine (ELM) Brands and its Performance Measure on Classification Algorithm.In this research article, we attempt to explain how variants are formed.that ELM can be applied to different applications. By using a neural network with a novel feed-forward algorithm, this procedure can be improved and optimized.

## **CHAPTER 3**

### **3.1 AIM**

This project aims to predict Stock Price using Data science technique.

### **3.2 SCOPE**

Here the scope of the project is that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety,

decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

## **CHAPTER 4**

### **4.1 MATERIALS**

#### **4.1.1 HARDWARE REQUIREMENTS**

- Processor : Pentium IV/III
- Hard disk : minimum 80 GB
- RAM : minimum 2 GB

#### **4.1.2 SOFTWARE REQUIREMENTS**

- Operating System : Windows
- Tool : Anaconda with Jupyter Notebook

#### **4.1.3 Non-Functional Requirements:**

Process of functional steps,

- Problem define
- Preparing data
- Evaluating algorithms
- Improving results
- Prediction the result

#### **4.1.4 Functional requirements:**

The software requirements specification is a technical specification of requirements for the software product. It is the first step in the requirements analysis process. It lists requirements of a particular software system. The following details to follow the special libraries like sk-learn, pandas, numpy, matplotlib and seaborn.

## 4.2 SYSTEM ANALYSIS

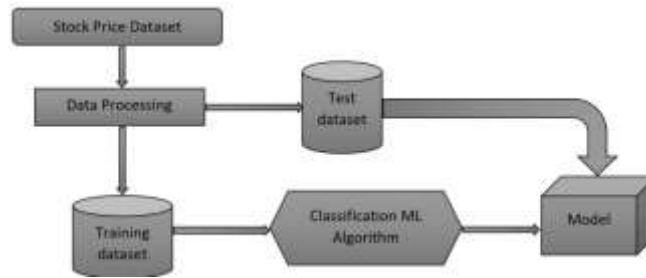
The objective is to foster AI at an anticipated stock cost so the most effective way to contrast controlled calculations is with supplant AI with prescient outcomes. The objective of the task is to incorporate clinic choices with PC based patient documents, which will lessen clinical mistakes, work on understanding security, decrease pointless preparing, and work on tolerant results. The thought is to assemble dependability as an apparatus for data and examination, for instance, by extricating data that can establish an information rich climate to assist with working on the nature of clinical choices.

## 4.3 SYSTEM METHODOLOGY

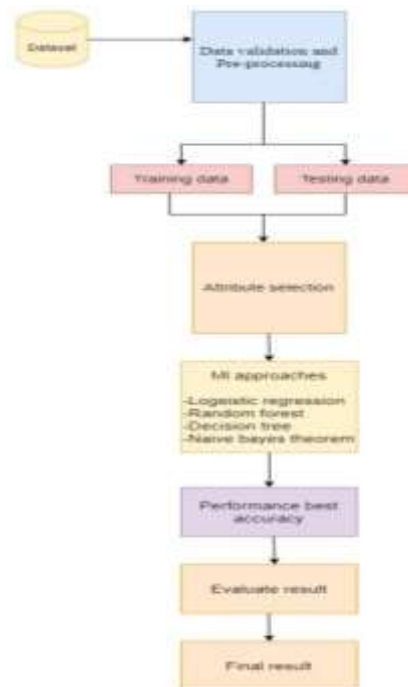
It is vital to think about the presentation of multi-AI calculations, and to examine the exhibition of numerous calculations in Python and scikit-learn. Underscore the various realities that share normal facts, contrasts, and examples. The accompanying segments will tell you the best way to do this utilizing scikit-learn. The way in to a decent correlation of AI is even | Each recipe is assessed similarly at a similar degree of information, and this should be possible by constraining every calculation to test the actual material. Look at the accompanying four distinct calculations The pictorial representation of overall architecture is shown in Fig 4.1 and in Fig 4.2 illustrates the proposed architecture model.

### List of Modules:

- Data Pre-processing
- Data Analysis of Visualization
- Comparing Algorithm with prediction in the form of best accuracy result  
Deployment Using Flask.



*Fig .4.1 Overall Architecture*



*Fig.4.2 Proposed architecture diagram*

#### 4.4 Existing system:

By using prior information about the structure of the model, the Autoregressive integrated moving average (ARIMA) model and the Vector Autoregression (VAR) method are used. Integrating the RNN (LSTM) into the BST method further enhances its performance.

## 4.5 Feasibility study:

**Data Wrangling:** Data Disputes: The data will be remembered for this part of the report and will be handled and refined for examination. Ensure the classes are obviously characterized, advocated, and that there re activities that should be followed.

**Data collection:** This dataset was collected from the Kaggle website. It contains 1989 records of stock price features that have been classified into two groups Increase, Decrease. The information forecast informational index is partitioned into two sections: preparing and experimentation. By and large, a 7: 3 proportion is utilized to limit preparing and testing. During the preparation, trial speculations were made dependent on information models and real outcomes created utilizing Standard Forest, operations, tree approval calculations, and Support Vector Classifier (SVC).

**Pre-processing:** The data got is probably going to be deficient, which could prompt irregularities. For best outcomes, data should be handled to further develop calculation execution. Departure should be eliminated and altered.

### **Data Pre-processing**

The AI machine is utilized to work out the standard blunder of the machine model (ML), which is extremely near the genuine mistake pace of the set-up information. On the off chance that the size of the data is adequately huge to address the populace, a confirmation interaction isn't needed. Assessment is progressively becoming one of the main parts of displaying. Affirmation programming is utilized consistently, yet in addition in example confirmation. This data is utilized by AI staff to all the more likely match hyper scales. The most common way of gathering, breaking down, and settling on content, quality, and information handling sets aside time. Getting data and highlights helps in data recovery; This data will assist you with picking a calculation to use to construct your model. Fig 4.3 depicts the collection of data sets and Fig 4.4 illustrates the pre-processed data

| Instance # | Date | Label     | Topic1 | Topic2                                             | Topic3                                                     | Topic4                             | Topic5                                      | Topic6         | Topic7        | Topic8        | Topic9        | Topic10       | Topic11       |
|------------|------|-----------|--------|----------------------------------------------------|------------------------------------------------------------|------------------------------------|---------------------------------------------|----------------|---------------|---------------|---------------|---------------|---------------|
| 0          | 1988 | 7/1/2018  | 1      | 6,117-year-old woman in Blackie City finally he... | Self-portraits of French artists on pavement (Chicago) ... | The president of France says a ... | Black Man Who Must Use Public Restrooms ... | 500+ Names ... | Blackie ...   | Blackie ...   | Blackie ...   | Blackie ...   | Blackie ...   |
| 1          | 1987 | 6/30/2018 | 1      | James ...                                          | James ...                                                  | James ...                          | James ...                                   | James ...      | James ...     | James ...     | James ...     | James ...     | James ...     |
| 2          | 1988 | 6/28/2018 | 1      | Explosion ...                                      | Explosion ...                                              | Explosion ...                      | Explosion ...                               | Explosion ...  | Explosion ... | Explosion ... | Explosion ... | Explosion ... | Explosion ... |
| 3          | 1988 | 6/28/2018 | 1      | 2,000 ...                                          | 2,000 ...                                                  | 2,000 ...                          | 2,000 ...                                   | 2,000 ...      | 2,000 ...     | 2,000 ...     | 2,000 ...     | 2,000 ...     | 2,000 ...     |

Fig.4.3 Drop the given data

| Instance # | Date | Label     | Topic1 | Topic2                                             | Topic3                                                     | Topic4                             | Topic5                                      | Topic6         | Topic7        | Topic8        | Topic9        | Topic10       | Topic11       |
|------------|------|-----------|--------|----------------------------------------------------|------------------------------------------------------------|------------------------------------|---------------------------------------------|----------------|---------------|---------------|---------------|---------------|---------------|
| 0          | 1988 | 7/1/2018  | 1      | 6,117-year-old woman in Blackie City finally he... | Self-portraits of French artists on pavement (Chicago) ... | The president of France says a ... | Black Man Who Must Use Public Restrooms ... | 500+ Names ... | Blackie ...   | Blackie ...   | Blackie ...   | Blackie ...   | Blackie ...   |
| 1          | 1987 | 6/30/2018 | 1      | James ...                                          | James ...                                                  | James ...                          | James ...                                   | James ...      | James ...     | James ...     | James ...     | James ...     | James ...     |
| 2          | 1988 | 6/28/2018 | 1      | Explosion ...                                      | Explosion ...                                              | Explosion ...                      | Explosion ...                               | Explosion ...  | Explosion ... | Explosion ... | Explosion ... | Explosion ... | Explosion ... |
| 3          | 1988 | 6/28/2018 | 1      | 2,000 ...                                          | 2,000 ...                                                  | 2,000 ...                          | 2,000 ...                                   | 2,000 ...      | 2,000 ...     | 2,000 ...     | 2,000 ...     | 2,000 ...     | 2,000 ...     |

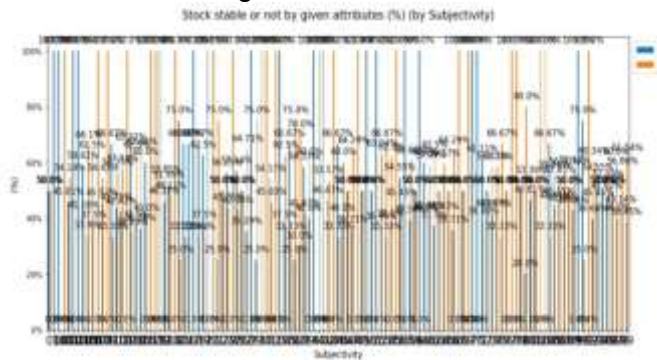
Fig.4.4 In the wake of erasing the given data

## Data Validation/ Cleaning/Preparing Process

Bundles of books are being imported and the data gave is bundled. Distinguish changes dependent on the idea of the media and the sort of media, and assess deficient qualities and values. Information approval depends on an example of information put away later an example practice and is utilized to gauge the showcase abilities to change the example, and use strategies to utilize a greater amount of the approval and estimation data set during the assessment of your example. To analyze the one-variable, the two-variable, and the multi-variable way, the information clearing/arrangement is finished by changing the names given by the segment segments. The most common way of erasing data will fluctuate contingent upon the data gave. In Fig 4.6 describes the datasets and Fig 4.6 presents graph of stock stable prediction.

|       | Unnamed: 0  | Label       | Subjectivity | Objectivity | Positive    | Neutral     | Negative    |
|-------|-------------|-------------|--------------|-------------|-------------|-------------|-------------|
| count | 1986.000000 | 1986.000000 | 1986.000000  | 1986.000000 | 1986.000000 | 1986.000000 | 1986.000000 |
| mean  | 994.843907  | 0.534743    | 56.712522    | 43.287478   | 19.966718   | 43.287478   | 36.745804   |
| std   | 574.301065  | 0.480917    | 21.557927    | 21.557927   | 16.373424   | 21.557927   | 21.109970   |
| min   | 0.000000    | 0.000000    | 0.000000     | 0.000000    | 0.000000    | 0.000000    | 0.000000    |
| 25%   | 496.250000  | 0.000000    | 40.000000    | 20.571429   | 8.333333    | 20.571429   | 21.859903   |
| 50%   | 995.500000  | 1.000000    | 54.545455    | 45.454545   | 18.181818   | 45.454545   | 33.333333   |
| 75%   | 1491.750000 | 1.000000    | 71.428571    | 60.000000   | 28.571429   | 60.000000   | 50.000000   |
| max   | 1986.000000 | 1.000000    | 100.000000   | 100.000000  | 100.000000  | 100.000000  | 100.000000  |

*Fig.4.5 Dataset*



*Fig 4.6 Stock stable graph prediction*

## CHAPTER 5

### 5.1 EXPLORATION DATA ANALYSIS OF VISUALIZATION:

The utilization of data is a vital ability in the field of AI and AI measurements. Arithmetic arrangements with the understanding and computation of numerical data. Data portrayal is a significant device for acquiring data quality. This can be helpful while investigating and exploring set up information to decide patterns, deception, congestion, and substantially more. Addressing data and dissecting research data is an extraordinary field, and it is necessitated that you read a portion of the books that are recorded for more data. Fig 5.1 represents the propagation by variables pie chart.

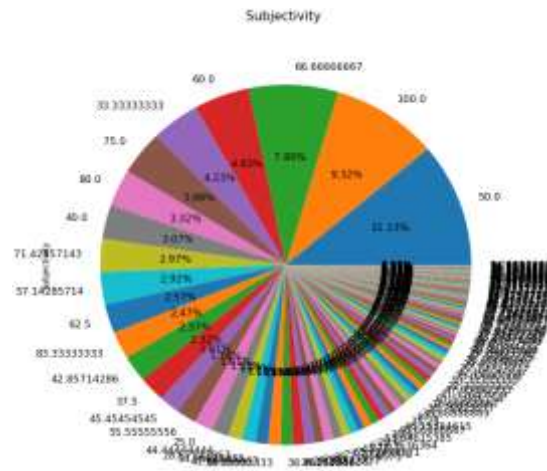


Fig.5.1 Pie chart

It tends to be pointless on the off chance that the data isn't apparent, like drawings or tables. The capacity to rapidly get to test data and different devices and procedures is fundamental in the field of arithmetic and AI. Fig 5.2 and Fig 5.3 represents datasets after visualization .

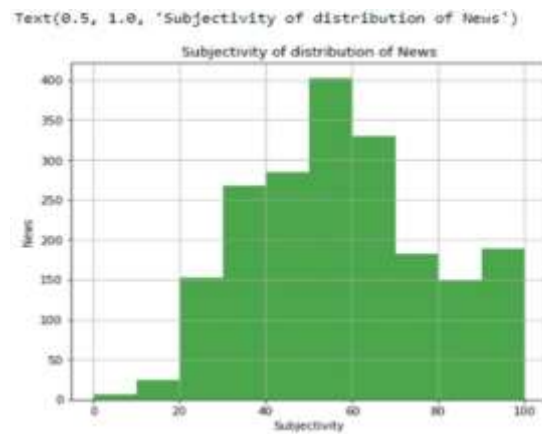
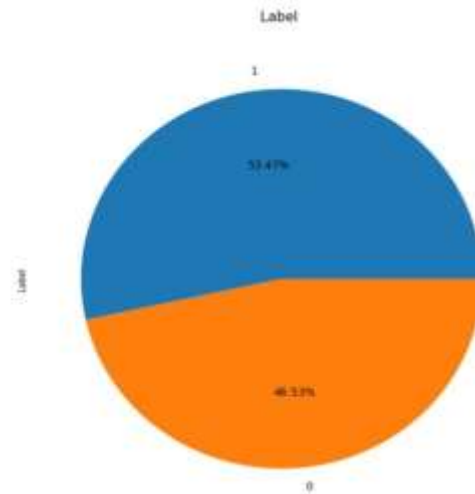
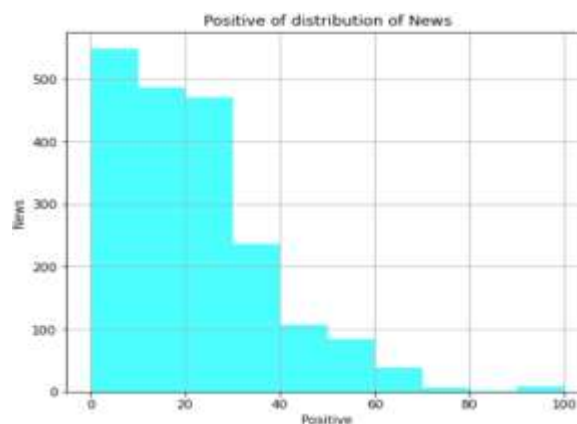


Fig.5.2 graph of distribution



*Fig.5.3 Pie chart*

Altering our information before it is remembered for a pre-handling calculation. Pre-handling information and how to change the fundamental data in the assortment of good data. At the end of the day, data is constantly gotten in better places, gathered badly, making it difficult to examine. To accomplish preferred outcomes over the one utilized in Machine Learning, data should be painstakingly ready. Fig 5.4 depicts the histogram of positive distribution of news and Fig 5.5 represents heatmap plot diagram.



*Fig .5.4 Histogram of algorithm*



Fig.5.5 Heatmap

## 5.2 ALGORITHM AND TECHNIQUE

### Algorithm Explanation:

A rundown is a methodical investigation of mathematical and factual techniques by which a PC program gains from given information and executes examination to arrange new things. The calculation chooses the name to be refreshed with the data dependent on the new model and adjusts the design to the new data that isn't composed later the information investigation.

### Logistic Regression

It is a measurable technique for examining set up information with free factors affecting the outcomes. Two factors are utilized to assess the outcomes (provided that there are two outcomes). The motivation behind review is to decide the proper model to show the connection between a free element (indicator or clarification) and two advantageous elements (variable = reaction or variable reaction). The learning machine is utilized to recognize the chance of a bit-by-bit change in a machine called strategic relapse. An opposite calculated variable is a paired variable comprising of information composed as 1 or 0. (indeed, achievement, and so on). Fig 5.6 provides accuracy results of logistic regression, Fig 5.7 graph of accuracy comparison of stock in logistic regression and Fig 5.8 represents the confusion matrix of logistic regression.

```

Accuracy result of Logistic Regression is: 54
Classification report of Logistic Regression Results:
              precision    recall  f1-score   support

     0           0.00       0.00       0.00         277
     1           0.54       1.00       0.70         319

 accuracy          0.54          0.54          0.54          596
 macro avg         0.27          0.50          0.35          596
 weighted avg      0.29          0.54          0.37          596

Confusion Matrix result of Logistic Regression is:
[[ 0 277]
 [ 0 319]]

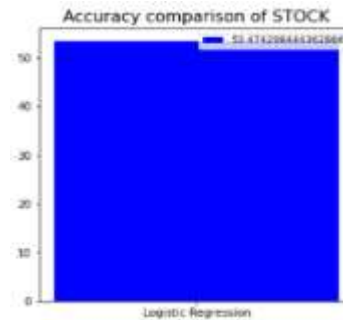
Sensitivity : 0.0
Specificity : 1.0

Cross validation test results of accuracy:
[0.53517588 0.53652393 0.53400504 0.53400504 0.53400504]

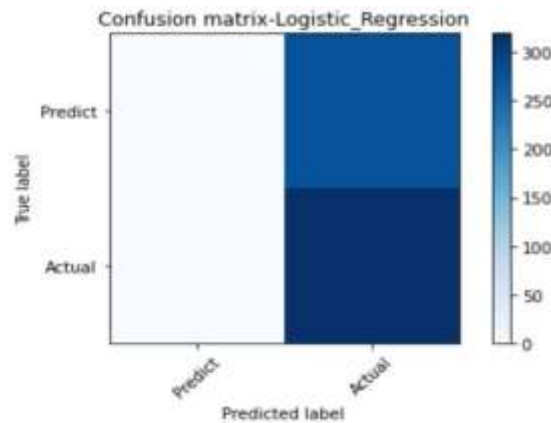
Accuracy result of Logistic Regression is: 53.474298444362866
Cross validation of Logistic Regression is: 0.535234899328859

```

*Fig .5.6 Accuracy result of Logistic regression*



*Fig 5.7 Graph of Logistic regression*



*Fig 5.8 Confusion matrix for Logistic regression*

## Random forest

Typical memory, otherwise called unlawful capacity, is a technique for preparing in stages, withdrawing, and different exercises performed via preparing an enormous number of choice trees and afterward carrying out them. tree free theory (relapse).

Chosen woodlands generally pay for the course of transformation to the development of trees. Typical memory is a kind of preparing dependent on AI. The last worth can be determined when contrasted with the normal worth from all trees in the timberland. All in all, each woodland tree predicts the stage wherein another record will become alright assuming that it is executed. Then, at that point, the new archive will be classified by greater part vote. Fig 5.9 provides accuracy results of random forest , Fig 5.10 graph of accuracy comparison of stock in random forest and Fig 5.11 represents the confusion matrix of random forest.

```

Accuracy result of Random Forest Classifier is: 67

Classification report of Random Forest Results:

              precision    recall  F1-score   support

     0       0.66        0.61        0.63        277
     1       0.68        0.73        0.70        319

 accuracy          0.67          0.67          0.67          596
  macro avg          0.67          0.67          0.67          596
 weighted avg          0.67          0.67          0.67          596

Confusion Matrix result of Random Forest Classifier is:
[[168 109]
 [ 87 232]]

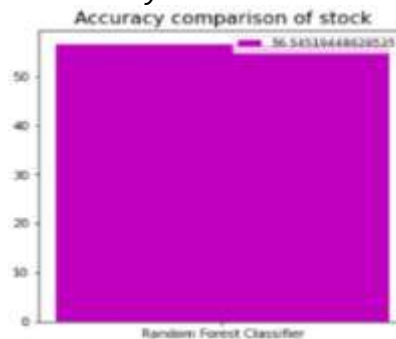
Sensitivity : 0.6064981949458483
Specificity : 0.7272727272727273

Cross validation test results of accuracy:
[0.57788945 0.52896725 0.5768262 0.59949622 0.5448806 ]

Cross validation of Random Forest Classifier is: 56.54519448628535

```

*Fig 5.9 Accuracy results of Random Forest*



*Fig.5.10 Graph of Random Forest*

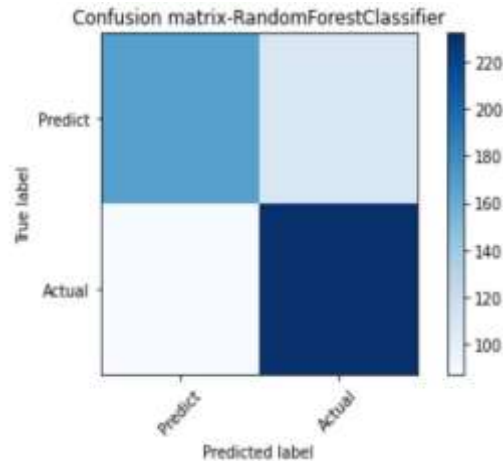


Fig.5.11 Confusion matrix for Random Forest

## Decision tree

The choice tree is a strong and well-known calculation. The choice tree calculation is carried out to control the calculation. It tends to be utilized persistently and in stages. In the first place, how about we survey the entire studio in general. Items should be arranged for data, yet documentation shows that resources keep on being partitioned on different occasions dependent on esteem attributes. At the front or the front, we utilize a mathematical strategy to decide the qualities. The choice tree is not difficult to set up, make different branches and distinguish twisting because of clamor or blockage. Fig 5.12 provides accuracy results of decision tree, Fig 5.13 graph of accuracy comparison of stock in decision tree and Fig 5.14 represents the confusion matrix of decision tree.

```

Accuracy of Decision Tree Classifier 63
Classification report of Decision Tree Results:
              precision    recall  f1-score   support

     0       0.60      0.63      0.61         277
     1       0.66      0.64      0.65         319

 accuracy          0.63
 macro avg         0.63      0.63      0.63         596
 weighted avg         0.63      0.63      0.63         596

Confusion Matrix result of Decision Tree Classifier is:
[[174 103]
 [116 203]]

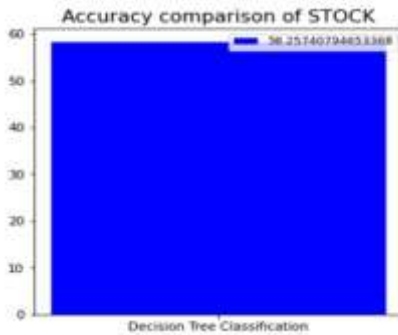
Sensitivity : 0.628158844765343
Specificity : 0.6363636363636364

Cross validation test results of accuracy:
[0.59045226 0.59697733 0.55667506 0.63224181 0.53652393]

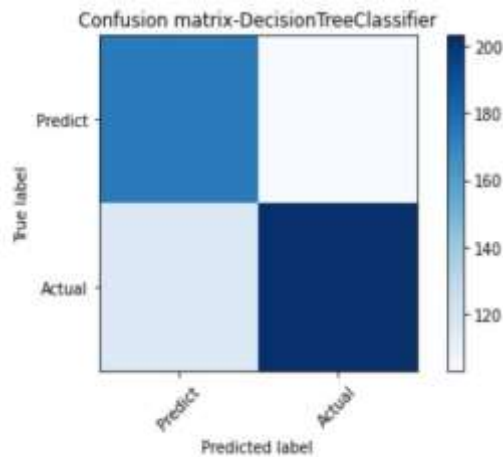
Cross validation of Decision Tree Classifier is: 58.25740794653368

```

Fig.5.12 Accuracy of Decision Tree



*Fig 5.13 Graph of Decision Tree*



*Fig 5.14 Confusion matrix for Decision Tree*

### **Naive Bayes Theorem**

The Naive Bayes calculation is a simple method for anticipating as indicated by the capacities of each component in every class. To feature the issue of theoretical show however much as could reasonably be expected, this is the learning system you should utilize. It is feasible to envision that the potential outcomes of each worth related still up in the air autonomously of any remaining elements. This is really smart, but at the same time it's quick and simple. Prescience, it is feasible to compute the likelihood of a judgment at each stage and afterward pick the most plausible worth. This thought is viewed as basic since it is not difficult to ascertain. Autonomy in the private area is the word for this theory. Fig 5.15 provides accuracy results of Naive Bayes theorem Fig 5.16 graph of accuracy comparison of stock in Naive Bayes theorem and Fig 5.17 represents the confusion matrix of Naive Bayes theorem.

Accuracy result of Naive Bayes Algorithm is 56

Classification report of Naive Bayes Results:

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.58      | 0.15   | 0.24     | 277     |
| 1            | 0.55      | 0.91   | 0.69     | 319     |
| accuracy     |           |        | 0.56     | 596     |
| macro avg    | 0.57      | 0.53   | 0.46     | 596     |
| weighted avg | 0.57      | 0.56   | 0.48     | 596     |

Confusion Matrix result of Naive Bayes is:

```
[[ 42 235]
 [ 30 289]]
```

Sensitivity : 0.15162454873646208

Specificity : 0.9059561128526645

Cross validation test results of accuracy:

```
[0.53266332 0.53904282 0.5465995 0.55415617 0.55415617]
```

Cross validation of Naive Bayes Algorithm is: 54.53235953065073

Fig 5.15 Accuracy of Naive Bayes Theorem

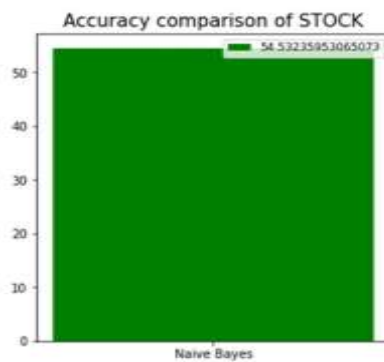


Fig 5.16 Graph of Naive Bayes Theorem

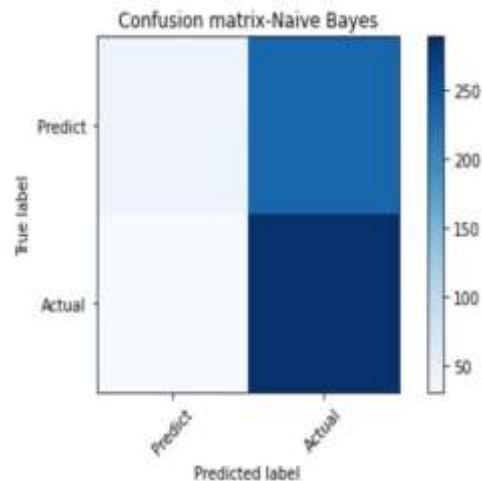


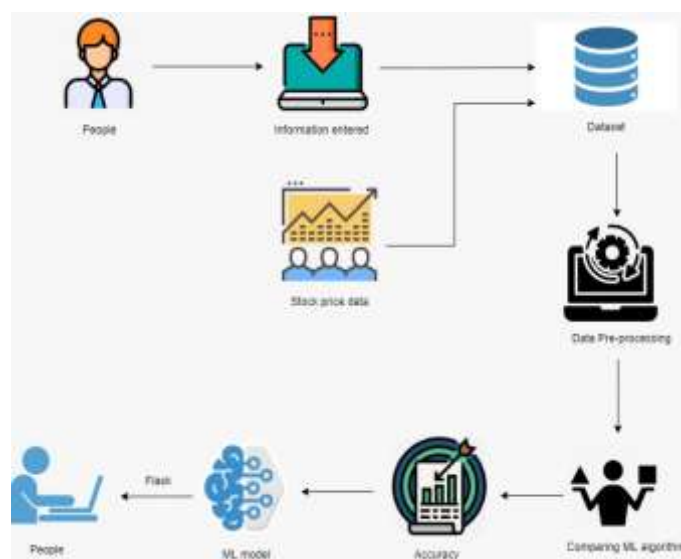
Fig 5.17 Confusion matrix for Naive Bayes Theorem

### 5.3COMPARING ALGORITHM WITH PREDECTION IN THE FORM OF BEST ACCURACY RESULT

Altering our information before it is remembered for a pre-handling calculation. Pre-handling information and how to change the fundamental data in the assortment of good data. As such, data is constantly gotten in better places, gathered badly, making it difficult to examine. To accomplish preferred outcomes over the one utilized in Machine Learning, data should be painstakingly ready. The way in to a decent correlation of calculations is to assess every strategy similarly for a similar data, which should be possible by requiring every calculation to be tried on a similar hardware.

The K-accreditation strategy is utilized to assess every calculation, and in particular, the preparation data is isolated into areas and set with a similar one-of-a-kind seed to assess every calculation similarly. Introduce Scikit-Learn libraries prior to contrasting the calculation and fabricate an AI machine utilizing it.

## 5.4 System architecture:



*Fig 5.18 System Architecture*

The above Fig 5.18 depicts the pictorial representation of the system architecture.

**Prediction Result by Accuracy:**

Line correlations with individual prophets are additionally used to decide esteem with regards to backwardness. The normal worth is endless. The result of the calculation should be executed as a variable. Indeed, it burns-through time and assets, however it is less productive than deals.

**False positives (FP):**

A person who is relied upon to pay has not paid. The normal outcome is indeed, when there is no genuine stage.

**False negatives (FN):**

Some unacceptable individual can manage the cost of it. At the point when the normal order is awful and the genuine characterization is great.

**True positives (TP):**

Default is a term used to portray a neglected individual. It precisely predicts a decent worth, the genuine class esteem is indeed, and shows the worth of the normal arrangement.

**True negatives (TN):**

Level of every single anticipated expense; all in all, how frequently does the model accurately foresee payers and non-payers?

True Positive Rate (TPR) =  $TP / (TP + FN)$

False Positive rate (FPR) =  $FP / (FP + TN)$

Fig 5.19 represents overall confusion matrix for the classifier.

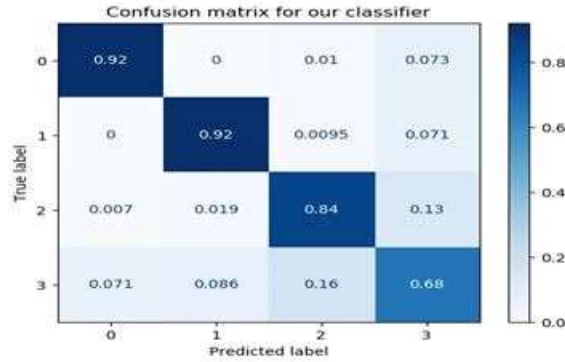


Fig 5.19 Overall Confusion Matrix

**Accuracy:** Percentage of every normal return; at the end of the day, how frequently does the model accurately foresee payers and non-payers?

**Accuracy calculation:**

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

**Precision:** The chance of good expectations will materialize. The choice to foresee beneficial things is a proportion of good expectations and great presumptions. decidedly. We have an exactness of 0.788, which is generally excellent.

$$\text{Accuracy} = \text{TP} / (\text{TP} + \text{FP})$$

**Recall:** This is a level of all around arranged exercises. (The level of individuals who affront them is an example that obviously shows that I disregard them.)

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Recall (Sensation) - Memory is a proportion of consistency and cautious perception in all classifications - yes.

The normal precision and memory score is F1. Consequently, these focuses show the upsides and downsides. Albeit not altogether clear, F1 is a higher priority than unwaveringness, particularly on the off chance that the division of work is lopsided.

At the point when great and bogus have a similar worth, trustworthiness functions admirably. Assuming that the cost of good and awful is altogether different, it's ideal to watch Precision and Remember.

General Formula:

$$\text{F- Measure} = 2\text{TP} / (2\text{TP} + \text{FP} + \text{FN})$$

F1-Score Formula:

$$F1 \text{ Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

Fig 5.20 represent the most accurate result of stock prediction using random forest algorithm.

```
Accuracy result of Random Forest Classifier is: 67
Classification report of Random Forest Results:
      precision    recall  f1-score   support
0               0.66       0.61       0.63        277
1               0.68       0.73       0.70        319

 accuracy: 0.67
 macro avg: 0.67   0.67   0.67   596
weighted avg: 0.67   0.67   0.67   596

Confusion Matrix result of Random Forest Classifier is:
[[168 100]
 [ 87 232]]

Sensitivity : 0.6804981949458883
Specificity : 0.7272727272727273

Cross validation test results of accuracy:
[0.57788945 0.52896725 0.5768262 0.59940622 0.5448886 ]

Cross validation of Random Forest Classifier is: 56.54519448628535
```

*Fig 5.20 Overall Accuracy*

## CHAPTER 6

### 6.1 Conclusion

Our proposal was to consider test data other than the preparation data bundle. The genuine computation of the informational collection of the refreshed framework we needed was 67% utilizing a one-of-a-kind timberland calculation. Information was cleaned and handled, lacks were distinguished, research was led, and a model was created and assessed to figure out which model got the most noteworthy score eventually. Utilize this program to observe stock value estimates.

### 6.2 Future Work

- Share cloud association value figures.
- Working on the execution of man-made consciousness.

## REFERENCES

[1] O. Badawi and M. J. Breslow, "Readmissions and death after icu discharge: development and validation of two predictive models," PloS one, vol. 7, no. 11, p.

e48758, 2012.

[2] Kim, Kyoung-jae, and Ingoo Han. "Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index." *Expert systems with Applications* 19.2 (2000): 125-132.

[3] Lee, J.W., 2001, June. Stock price prediction using reinforcement learning. In *ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings* (Cat. No. 01TH8570) (Vol. 1, pp. 690-695). IEEE.

[4] Ariyo, Adebiyi A., Adewumi O. Adewumi, and Charles K. Ayo. "Stock price prediction using the ARIMA model." *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*. IEEE, 2014.

[5] Manuel R. Vargas, Carlos E.M. dos Anjos, Gustavo L.G. Bichara, Alexandre G. Evsukoff (2018). "Deep Learning for Stock Market Prediction Using Technical Indicators and Financial News Articles". *2018 International Joint Conference on Neural Networks (IJCNN)*

[6] Vivek Rajput, Sarika Bobade . "Stock Market Prediction Using Hybrid Approach". *International Conference on Computing, Communication and Automation (ICCCA2016)*.

[7] C. H. Su and C. H. Cheng, "A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock," *Neurocomputing*, vol. 205, pp. 264-273, September 2016.

[8] 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

[9] Mohammad Obaidur Rahman, Md. Sabir Hossain\*, Ta-Seen Junaid, Md. Shafiul Alam Forhad, Muhammad Kamal Hossen , "Predicting Prices of Stock Market using Gated Recurrent Units (GRUs) Neural Networks", *IJCSNS International Journal of Computer Science and Network Security*, VOL.19 No.1, 2019.

[10] Wenping Zhang, Chunping Li, Yunming Ye, Wenjie Li and Eric W.T. Ngai , "Dynamic Business Network Analysis for Correlated Stock Price Movement Prediction", *IEEE Intelligent Systems* Volume: 30 , Issue: 2 , Mar.-Apr. 2015.

[11] Batra, Rakhi, and Sher Muhammad Daudpota. "Integrating StockTwits with

sentiment analysis for better prediction of stock price movement." 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). IEEE, 2018.

[12] Andi, Hari Krishnan. "An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model." Journal of Soft Computing Paradigm 3, no. 3 (2021): 205-217.

[13]Raj, Dr, and S. Jennifer. "Evolutionary Computing Techniques for Resolving Load Dispatch Problem." Journal of Soft Computing Paradigm 2, no. 1 (2020): 47-56.

[14]Anand, C. "Comparison of Stock Price Prediction Models using Pre-trained Neural Networks." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 3, no. 02 (2021): 122-134.

[15]Sungheetha, Akey, and Rajesh Sharma. "Transcapsule model for sentiment classification." Journal of Artificial Intelligence 2, no. 03 (2020): 163-169.

[16]Manoharan, J. Samuel. "Study of Variants of Extreme Learning Machine (ELM) Brands and its Performance Measure on Classification Algorithm." Journal of Soft Computing Paradigm (JSCP) 3, no. 02 (2021): 83-95.

## **APPENDIX**

### **Module – 1**

#### **Pre-Processing**

```
#import library packages
```

```
import pandas as pd
```

```
import numpy as np
```

```
import warnings
```

```
warnings.filterwarnings("ignore")
```

```
#Load given dataset
```

```
data = pd.read_csv("heart.csv")
```

Before drop the given dataset:

```
data.head()
```

```

#shape
data.shape
After drop the given dataset:
df = data.dropna()
df.head()
#shape
df.shape
#columns
df.columns
#To describe the dataframe
df.describe()
#Checking datatype and information about dataset
df.info()
Checking duplicate values of dataframe

#Checking for duplicate data
df.duplicated()
sum(df.duplicated())
#Checking sum of missing values
df.isnull().sum()
df.age.unique()
df.sex.unique()
df.cp.unique()
df.trestbps.unique()
df.chol.unique()
df.restecg.unique()
df.target.unique()
print("Age of patient ranges :", sorted(df['age'].unique()))
df["target"].value_counts()
df.corr()

```

## Module – 2

### Visualization

```
#import library packages
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import warnings
warnings.filterwarnings('ignore')
data = pd.read_csv("heart.csv")
df = data.dropna()
df.columns
pd.crosstab(df.age,df.target)
pd.crosstab(df.cp,df.target)
#Histogram Plot of Age distribution
df["age"].hist(figsize=(10,8), color="red")
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("No of Patients")
plt.show()
#barplot for ecg and target
fig, ax = plt.subplots(figsize=(15,8))
sns.barplot(x = "restecg", y = "target", ax=ax, data=df)
plt.title("ECG vs TARGET")
#Propagation by variable
def PropByVar(df, variable):
    dataframe_pie = df[variable].value_counts()
    ax = dataframe_pie.plot.pie(figsize=(10,10), autopct='%1.2f%%', fontsize = 12)
    ax.set_title(variable + '\n', fontsize = 15)
    return np.round(dataframe_pie/df.shape[0]*100,2)
```

```

PropByVar(df, 'thal')
fig, ax = plt.subplots(figsize=(15,6))
sns.boxplot(df.age, ax =ax)
plt.title("Age distribution")
plt.show()
sns.pairplot(df)
plt.show()
fig, ax = plt.subplots(figsize=(15,6))
sns.violinplot(y = df['age'], x = df['target'], ax=ax)
plt.title("Patients age and target")
plt.show()
# Heatmap plot diagram
fig, ax = plt.subplots(figsize=(15,10))
sns.heatmap(df.corr(), ax=ax, annot=True)
Splitting Train/Test:
#preprocessing, split test and dataset, split response variable
X = df.drop(labels='target', axis=1)
#Response variable
y = df.loc[:, 'target']
#We'll use a test size of 30%. We also stratify the split on the response variable,
which is very important to do because there are so few fraudulent transactions.
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1,
stratify=y)
print("Number of training dataset: ", len(X_train))
print("Number of test dataset: ", len(X_test))
print("Total number of dataset: ", len(X_train)+len(X_test))
def qul_No_qul_bar_plot(df, bygroup):
    dataframe_by_Group = pd.crosstab(df[bygroup], columns=df["target"], normalize =
'index')
    dataframe_by_Group = np.round((dataframe_by_Group * 100), decimals=2)

```

```

ax = dataframe_by_Group.plot.bar(figsize=(15,7));
vals = ax.get_yticks()
ax.set_yticklabels(['{:3.0f}%'.format(x) for x in vals]);
ax.set_xticklabels(dataframe_by_Group.index,rotation = 0, fontsize = 15);
ax.set_title('Stroke having or not by given attributes (%) (by ' +
dataframe_by_Group.index.name + ')\n', fontsize = 15)
ax.set_xlabel(dataframe_by_Group.index.name, fontsize = 12)
ax.set_ylabel('(%)', fontsize = 12)
ax.legend(loc = 'upper left',bbox_to_anchor=(1.0,1.0), fontsize= 12)
rects = ax.patches

# Add Data Labels

for rect in rects:
    height = rect.get_height()
    ax.text(rect.get_x() + rect.get_width()/2,
            height + 2,
            str(height)+'%',
            ha='center',
            va='bottom',
            fontsize = 12)
return dataframe_by_Group
qul_No_qul_bar_plot(df, 'ca')

```

## Module – 3

### Logistic Regression Algorithm

*#import library packages*

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

```

import seaborn as sns
import numpy as np
import warnings
warnings.filterwarnings('ignore')
#Load given dataset
data = pd.read_csv("heart.csv")
df=data.dropna()
df.columns

#According to the cross-validated MCC scores, the random forest is the best-
performing model, so now let's evaluate its performance on the test set.
from sklearn.metrics import confusion_matrix, classification_report,
matthews_corrcoef, cohen_kappa_score, accuracy_score, average_precision_score,
roc_auc_score
X = df.drop(labels='target', axis=1)
#Response variable
y = df.loc[:, 'target']
#We'll use a test size of 30%. We also stratify the split on the response variable,
which is very important to do because there are so few fraudulent transactions.
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1,
stratify=y)
Logistic Regression :

from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score

logR= LogisticRegression()

logR.fit(X_train,y_train)

predictR = logR.predict(X_test)

```

```

print("")
print('Classification report of Logistic Regression Results:')
print("")
print(classification_report(y_test,predictR))

print("")
cm=confusion_matrix(y_test,predictR)
print('Confusion Matrix result of Logistic Regression is:\n',cm)
print("")
sensitivity = cm[0,0]/(cm[0,0]+cm[0,1])
print('Sensitivity : ', sensitivity )
print("")
specificity = cm[1,1]/(cm[1,0]+cm[1,1])
print('Specificity : ', specificity)
print("")

accuracy = cross_val_score(logR, X, y, scoring='accuracy')
print('Cross validation test results of accuracy:')
print(accuracy)
#get the mean of each fold
print("")
print("Accuracy result of Logistic Regression is:",accuracy.mean() * 100)
LR=accuracy.mean() * 100
def graph():
    import matplotlib.pyplot as plt
    data=[LR]
    alg="Logistic Regression"
    plt.figure(figsize=(5,5))
    b=plt.bar(alg,data,color=("b"))
    plt.title("Accuracy comparison of Stock price",fontsize=15)

```

```

plt.legend(b,data,fontsize=9)
graph()
TP = cm[0][0]
FP = cm[1][0]
FN = cm[1][1]
TN = cm[0][1]
print("True Positive :",TP)
print("True Negative :",TN)
print("False Positive :",FP)
print("False Negative :",FN)
print("")
TPR = TP/(TP+FN)
TNR = TN/(TN+FP)
FPR = FP/(FP+TN)
FNR = FN/(TP+FN)
print("True Positive Rate :",TPR)
print("True Negative Rate :",TNR)
print("False Positive Rate :",FPR)
print("False Negative Rate :",FNR)
print("")
PPV = TP/(TP+FP)
NPV = TN/(TN+FN)
print("Positive Predictive Value :",PPV)
print("Negative predictive value :",NPV)
def plot_confusion_matrix(cm2, title='Confusion matrix-LogisticRegression',
cmap=plt.cm.Blues):
    target_names=['Predict','Actual']
    plt.imshow(cm2, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(target_names))

```

```

plt.xticks(tick_marks, target_names, rotation=45)
plt.yticks(tick_marks, target_names)
plt.tight_layout()
plt.ylabel('True label')
plt.xlabel('Predicted label')

cm2=confusion_matrix(y_test, predictR)
print('Confusion matrix-LogisticRegression:')
print(cm2)
plot_confusion_matrix(cm2)
import joblib
joblib.dump(logR,"model.pkl")
['model.pkl']

```

## Module – 4

### Random Forest Algorithm

```

#import library packages
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import warnings
warnings.filterwarnings('ignore')
#Load given dataset
data = pd.read_csv("heart.csv")
df=data.dropna()
df.columns

#According to the cross-validated MCC scores, the random forest is the best-
performing model, so now let's evaluate its performance on the test set.

```

```

from sklearn.metrics import confusion_matrix, classification_report,
matthews_corrcoef, cohen_kappa_score, accuracy_score, average_precision_score,
roc_auc_score
X = df.drop(labels='target', axis=1)
#Response variable
y = df.loc[:, 'target']
#We'll use a test size of 30%. We also stratify the split on the response variable,
which is very important to do because there are so few fraudulent transactions.
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1,
stratify=y)
RandomForestClassifier:

from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score

rfc = RandomForestClassifier()

rfc.fit(X_train,y_train)

predictR = rfc.predict(X_test)

print("")
print('Classification report of Random Forest Classifier Results:')
print("")
print(classification_report(y_test,predictR))

print("")
cm=confusion_matrix(y_test,predictR)
print('Confusion Matrix result of Random Forest Classifier is:\n',cm)
print("")

```

```

sensitivity = cm[0,0]/(cm[0,0]+cm[0,1])
print('Sensitivity : ', sensitivity )
print("")
specificity = cm[1,1]/(cm[1,0]+cm[1,1])
print('Specificity : ', specificity)
print("")

accuracy = cross_val_score(rfc, X, y, scoring='accuracy')
print('Cross validation test results of accuracy:')
print(accuracy)
#get the mean of each fold
print("")
print("Accuracy result of Random Forest Classifier is:",accuracy.mean() * 100)
LR=accuracy.mean() * 100
def graph():
    import matplotlib.pyplot as plt
    data=[LR]
    alg="Random Fores tClassifier"
    plt.figure(figsize=(5,5))
    b=plt.bar(alg,data,color=("b"))
    plt.title("Accuracy comparison of Stock price",fontsize=15)
    plt.legend(b,data,fontsize=9)
graph()
TP = cm[0][0]
FP = cm[1][0]
FN = cm[1][1]
TN = cm[0][1]
print("True Positive :",TP)
print("True Negative :",TN)
print("False Positive :",FP)
print("False Negative :",FN)

```

```

print("")
TPR = TP/(TP+FN)
TNR = TN/(TN+FP)
FPR = FP/(FP+TN)
FNR = FN/(TP+FN)
print("True Positive Rate :",TPR)
print("True Negative Rate :",TNR)
print("False Positive Rate :",FPR)
print("False Negative Rate :",FNR)
print("")
PPV = TP/(TP+FP)
NPV = TN/(TN+FN)
print("Positive Predictive Value :",PPV)
print("Negative predictive value :",NPV)
def plot_confusion_matrix(cm2, title='Confusion matrix-RandomForestClassifier',
cmap=plt.cm.Blues):
    target_names=['Predict','Actual']
    plt.imshow(cm2, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(target_names))
    plt.xticks(tick_marks, target_names, rotation=45)
    plt.yticks(tick_marks, target_names)
    plt.tight_layout()
    plt.ylabel('True label')
    plt.xlabel('Predicted label')

cm2=confusion_matrix(y_test, predictR)
print('Confusion matrix-RandomForestClassifier:')
print(cm2)
plot_confusion_matrix(cm2)

```

## Module – 5

### Decision Tree Algorithm

```
#import library packages
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import warnings
warnings.filterwarnings('ignore')
#Load given dataset
data = pd.read_csv("heart.csv")
df=data.dropna()
df.columns

#According to the cross-validated MCC scores, the random forest is the best-
performing model, so now let's evaluate its performance on the test set.
from sklearn.metrics import confusion_matrix, classification_report,
matthews_corrcoef, cohen_kappa_score, accuracy_score, average_precision_score,
roc_auc_score
X = df.drop(labels='target', axis=1)
#Response variable
y = df.loc[:, 'target']
#We'll use a test size of 30%. We also stratify the split on the response variable,
which is very important to do because there are so few fraudulent transactions.
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1,
stratify=y)
Decision Tree Classifier:

from sklearn.metrics import accuracy_score, confusion_matrix
```

```

from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import cross_val_score

dtc = DecisionTreeClassifier()

dtc.fit(X_train,y_train)

predictR = dtc.predict(X_test)

print("")
print('Classification report of Decision Tree Classifier Results:')
print("")
print(classification_report(y_test,predictR))

print("")
cm=confusion_matrix(y_test,predictR)
print('Confusion Matrix result of Decision Tree Classifier is:\n',cm)
print("")
sensitivity = cm[0,0]/(cm[0,0]+cm[0,1])
print('Sensitivity : ', sensitivity )
print("")
specificity = cm[1,1]/(cm[1,0]+cm[1,1])
print('Specificity : ', specificity)
print("")

accuracy = cross_val_score(dtc, X, y, scoring='accuracy')
print('Cross validation test results of accuracy:')
print(accuracy)
#get the mean of each fold
print("")
print("Accuracy result of Decision Tree Classifier is:",accuracy.mean() * 100)

```

```

LR=accuracy.mean() * 100
def graph():
    import matplotlib.pyplot as plt
    data=[LR]
    alg="Decision Tree Classifier "
    plt.figure(figsize=(5,5))
    b=plt.bar(alg,data,color=("b"))
    plt.title("Accuracy comparison of Stock price",fontsize=15)
    plt.legend(b,data,fontsize=9)
graph()
TP = cm[0][0]
FP = cm[1][0]
FN = cm[1][1]
TN = cm[0][1]
print("True Positive :",TP)
print("True Negative :",TN)
print("False Positive :",FP)
print("False Negative :",FN)
print("")
TPR = TP/(TP+FN)
TNR = TN/(TN+FP)
FPR = FP/(FP+TN)
FNR = FN/(TP+FN)
print("True Positive Rate :",TPR)
print("True Negative Rate :",TNR)
print("False Positive Rate :",FPR)
print("False Negative Rate :",FNR)
print("")
PPV = TP/(TP+FP)
NPV = TN/(TN+FN)
print("Positive Predictive Value :",PPV)

```

```

print("Negative predictive value :",NPV)
def plot_confusion_matrix(cm2, title='Confusion matrix-DecisionTreeClassifier',
cmap=plt.cm.Blues):
    target_names=['Predict','Actual']
    plt.imshow(cm2, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(target_names))
    plt.xticks(tick_marks, target_names, rotation=45)
    plt.yticks(tick_marks, target_names)
    plt.tight_layout()
    plt.ylabel('True label')
    plt.xlabel('Predicted label')

cm2=confusion_matrix(y_test, predictR)
print('Confusion matrix-DecisionTreeClassifier:')
print(cm2)
plot_confusion_matrix(cm2)

```

## Module – 6

### Naïve Bayes Algorithm

```

#import library packages
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import warnings
warnings.filterwarnings('ignore')
#Load given dataset

```

```

data = pd.read_csv("heart.csv")
df = data.dropna()
df.columns

#According to the cross-validated MCC scores, the random forest is the best-
performing model, so now let's evaluate its performance on the test set.

from sklearn.metrics import confusion_matrix, classification_report,
matthews_corrcoef, cohen_kappa_score, accuracy_score, average_precision_score,
roc_auc_score
X = df.drop(labels='target', axis=1)
#Response variable
y = df.loc[:, 'target']
#We'll use a test size of 30%. We also stratify the split on the response variable,
which is very important to do because there are so few fraudulent transactions.
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1,
stratify=y)
Naive Bayes:

from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import cross_val_score

nb = GaussianNB()

nb.fit(X_train, y_train)

predictR = nb.predict(X_test)

print("")
print('Classification report of Naive Bayes Results:')
print("")
print(classification_report(y_test, predictR))

```

```

print("")
cm=confusion_matrix(y_test,predictR)
print('Confusion Matrix result of Naive Bayes is:\n',cm)
print("")
sensitivity = cm[0,0]/(cm[0,0]+cm[0,1])
print('Sensitivity : ', sensitivity )
print("")
specificity = cm[1,1]/(cm[1,0]+cm[1,1])
print('Specificity : ', specificity)
print("")

accuracy = cross_val_score(nb, X, y, scoring='accuracy')
print('Cross validation test results of accuracy:')
print(accuracy)
#get the mean of each fold
print("")
print("Accuracy result of Naive Bayes is:",accuracy.mean() * 100)
LR=accuracy.mean() * 100
def graph():
    import matplotlib.pyplot as plt
    data=[LR]
    alg="GaussianNB"
    plt.figure(figsize=(5,5))
    b=plt.bar(alg,data,color=("b"))
    plt.title("Accuracy comparison of Stock price",fontsize=15)
    plt.legend(b,data,fontsize=9)
graph()
TP = cm[0][0]
FP = cm[1][0]
FN = cm[1][1]

```

```

TN = cm[0][1]
print("True Positive :",TP)
print("True Negative :",TN)
print("False Positive :",FP)
print("False Negative :",FN)
print("")
TPR = TP/(TP+FN)
TNR = TN/(TN+FP)
FPR = FP/(FP+TN)
FNR = FN/(TP+FN)
print("True Positive Rate :",TPR)
print("True Negative Rate :",TNR)
print("False Positive Rate :",FPR)
print("False Negative Rate :",FNR)
print("")
PPV = TP/(TP+FP)
NPV = TN/(TN+FN)
print("Positive Predictive Value :",PPV)
print("Negative predictive value :",NPV)
def plot_confusion_matrix(cm2, title='Confusion matrix-Naive Bayes',
cmap=plt.cm.Blues):
    target_names=['Predict','Actual']
    plt.imshow(cm2, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(target_names))
    plt.xticks(tick_marks, target_names, rotation=45)
    plt.yticks(tick_marks, target_names)
    plt.tight_layout()
    plt.ylabel('True label')
    plt.xlabel('Predicted label')

```

```

cm2=confusion_matrix(y_test, predictR)
print('Confusion matrix-Naive Bayes:')
print(cm2)
plot_confusion_matrix(cm2)

```

## HTML Code:

```

<!DOCTYPE html>
<html >
<!--From https://codepen.io/frytyler/pen/EGdtg-->
<head>
  <meta charset="UTF-8">
  <title>TITLE</title>
<link rel="stylesheet" href="{{ url_for('static', filename='css/bootstrap.min.css') }}">
  <link      href='https://fonts.googleapis.com/css?family=Pacifico'      rel='stylesheet'
type='text/css'>
<link      href='https://fonts.googleapis.com/css?family=Arimo'      rel='stylesheet'
type='text/css'>
<link      href='https://fonts.googleapis.com/css?family=Hind:300'      rel='stylesheet'
type='text/css'>
<link      href='https://fonts.googleapis.com/css?family=Open+Sans+Condensed:300'
rel='stylesheet' type='text/css'>
<style>
.back{
  background-image: url("{{ url_for('static', filename='image/img.jpg') }}");
  background-repeat: no-repeat;
  background-attachment: fixed;
  background-size: 100% 100%;
}
.white{

```

```

color:white;
}
.space{
margin:10px 30px;
padding:10px 10px;
background: palegreen;
width:500px
}
.gap{
padding:10px 20px;
}
</style>

</head>

<body class="back">
<div>
    <div class="jumbotron">
        <h1 style="text-align:center">STOCK PRICE CLASSIFICATION USING
MACHINE LEARNING TECHNIQUE </h1>

    </div>

    <!-- Main Input For Receiving Query to our ML -->
    <form class="form-group" action="{{ url_for('predict')}}"method="post">

        <div class="row">
            <div class="gap col-md-6 ">
                <label class="white" for="">Open</label>
                <input type="number" class="space form-control" step="0.01"
name="Open" placeholder="Open" required="required" /><br>

```

```
<label class="white" for="">High</label>
<input type="number" class="space form-control" step="0.01"
name="High" placeholder="High" required="required" /><br>
```

```
<label class="white" for="">Low</label>
<input type="number" class="space form-control" step="0.01"
name="Low" placeholder="Low" required="required" /><br>
```

```
<label class="white" for="">Close</label>
<input type="number" class="space form-control" step="0.01"
name="Close" placeholder="Close" required="required" /><br>
```

```
<label class="white" for="">Volume</label>
<input type="number" class="space form-control" step="0.01"
name="Volume" placeholder="Volume" required="required" /><br>
```

```
<label class="white" for="">Adj_Close</label>
<input type="number" class="space form-control" step="0.01"
name="Adj_close" placeholder="Adj_close" required="required" /><br>
```

```
</div>
```

```
<div class="row">
<div class="gap col-md-6 ">
<label class="white" for="">Subjectivity</label>
<input type="number" class="space form-control" step="0.01"
name="Subjectivity" placeholder="Subjectivity" required="required" /><br>
```

```
<label class="white" for="">Objectivity</label>
```

```
        <input type="number" class="space form-control" step="0.01"
name="Objectivity" placeholder="Objectivity" required="required" /><br>
```

```
        <label class="white" for="">Positive</label>
        <input type="number" class="space form-control" step="0.01"
name="Positive" placeholder="Positive" required="required" /><br>
```

```
        <label class="white" for="">Neutral</label>
        <input type="number" class="space form-control" step="0.01"
name="Neutral" placeholder="Neutral" required="required" /><br>
```

```
        <label class="white" for="">Negative</label>
        <input type="number" class="space form-control" step="0.01"
name="Negative" placeholder="Negative" required="required" /><br>
```

```
    </div>
<div style="padding:2% 35%">
    <button type="submit" class="btn btn-success btn-block"
style="width:350px;padding:20px">Predict</button>
</div>
```

```
    </form>
```

```
    <br>
```

```
    <br>
```

```
<div style="background:skyblue;padding:2% 40%">
```

```
    {{ prediction_text }}
```

```
</div>
```

```
</div>
```

```
</body>
</html>
```

### Flask Deploy:

```
import numpy as np
from flask import Flask, request, jsonify, render_template
import pickle
import joblib
app = Flask(__name__)
model = joblib.load('model.pkl')
@app.route('/')
def home():
    return render_template('index.html')
@app.route('/predict',methods=['POST'])
def predict():
    """
    For rendering results on HTML GUI
    """
    int_features = [(x) for x in request.form.values()]
    final_features = [np.array(int_features)]
    print(final_features)
    prediction = model.predict(final_features)
    print(prediction)
    output = prediction[0]
    if output == 1:
        return render_template('index.html', prediction_text='STOCK PRICE NOT
OCCURED')
    else:
        return render_template('index.html', prediction_text='STOCK PRICE
OCCURED')
    print(output)
```

```
if __name__ == "__main__":  
    app.run(host="localhost", port=8000)
```

## Output Screenshots:



