VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi - 590018

A PROJECT REPORT ON

"STOCK MARKET PREDICTION"

Submitted in the partial fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING

In

COMPUTER SCIENCE & ENGINEERING

Submitted By

PRAJWAL V BAGARE (1RR18CS100)

SAGAR JAGANNATH SUTAR (1RR18CS116)

SUDHARSHANA VP (1RR18CS142)

SUNEERAJ K S (1RR18CS144)

Under the Guidance of

SRINIVASA R

ASSOCIATE PROFESSOR

Dept of CSE, RRCE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

RAJARAJESWARI COLLEGE OF ENGINEERING

MYSORE ROAD, BANGALORE-560074

(An ISO 9001:2008 Certified Institute)

(2021-22)

RAJARAJESWARI COLLEGE OF ENGINEERING
MYSORE ROAD, BANGALORE-560074
(An ISO 9001:2008 Certified Institute)
(Affiliated to Visvesvaraya Technological University, Belagavi)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

Certified that project work phase-1 entitled

"STOCK MARKET PREDICTION"
Carried out by
PRAJWAL V BAGARE (1RR18CS100)

SAGAR JAGANNATH SUTAR (1RR18CS116)

SUDHARSHANA VP (1RR18CS142)

SUNEERAJ K S (1RR18CS144)

The students of **"RajaRajeswari College of Engineering"** in partial fulfillment for the award of **Bachelor of Engineering in Computer Science & Engineering** of the Visvesvaraya Technological University, Belagavi during the year **2021–2022.** It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The Project report has been approved as it satisfies the academic requirements in respect of project work phase-1 prescribed for the Seventh semester.

…………….................   …………………………   ……………………………
Signature of guide   **Signature of HOD**   **Signature of Principal**

[Mr. Srinivasa R]   **[Dr. S. Usha]**   **[Dr. T. Chandrashekar]**
 Associate Professor   Prof. & HOD   Prof. & Principal
Dept.., of CSE   Dept.., of CSE   RRCE, Bengaluru
RRCE,   Dean(Research)
Bengaluru   RRCE,
                           Bengaluru

External Viva-Voce

Examiners: *Signature*

1. _____

2. _____

ABSTRACT

The stock market is an intriguing assiduity to study. There are colorful variations present in it. Numerous experts have been studying and probing on the colorful trends that the stock market goes through.

One of the major studies has been the attempt to prognosticate the stock prices of colorful companies grounded on literal data Vaticination of stock prices will greatly help people to understand where and how to invest so that the threat of losing plutocrat is minimized.

The proposed system works in two styles – Regression and Classification. In Regression, the system predicts the ending price of stock of a company. In Classification, the system predicts whether the ending price of stock will increase or drop the coming day.

We also use Sentiment Analysis as a supporting factor to determine request sentiment of a particular stock as user sentiment plays a huge part in determining stock prices.

ACKNOWLEDGEMENT

PRAJWAL V BAGARE (IRRI8CS100)

SAGAR JAGANNATH SUTAR (1RR18CS116)

SUDHARSHANA V P (1RR18CS142)

SUNEERAJ K S (1RR18CS144)

CONTENTS

List of Figures

CHAPTER 1

INTRODUCTION

1.1 EXISTING SYSTEM

Existing Systems use only one data source hence the data is extremely biased When Traditional Classifiers are used, Stock Prices seem to be unpredictable. Existing systems do not focus on external events in the environment like news or social media. Other existing systems are manual or analog in nature and are highly speculative, there's no proper way of measuring its accuracy or precision.

Existing systems do not perform well when there is a change in operating environment.

1.2 PROPOSED SYSTEM

We propose to build a standalone application to predict a stock price using more than one method of prediction from various data sources.

In our proposed system we have two main components for supporting buyer decision.

1. Price Prediction using Regression and Classification.
2. Sentiment Analysis to determine market sentiment of a particular stock.

In Price Prediction we use various algorithms to determine the price, out of these various algorithms we perform model evaluation to select the best model that's fitting our data.

The various algorithms are:

1. Multivariate Linear Regression
2. Artificial Neural Networks using Backpropagation
3. Random Forest
4. Sentiment Analysis

1.3 MOTIVATION

Businesses in most cases run over purchaser's pleasure, purchaser opinions approximately their products. Shifts in sentiment on social media were proven to correlate with shifts in inventory markets. Identifying purchaser grievances thereby resolving them results in purchaser pleasure in addition to trustworthiness of an organization. Hence there's a need of an unbiased computerized machine to categorize purchaser opinions regarding any

problem. In today's surroundings in which we're justifiably laid low with statistics overload (even though this doesn't imply higher or deeper insights), businesses may have mountains of purchaser remarks collected; however, for mere humans, it's nevertheless not possible to examine it manually with none type of mistakes or bias. Oftentimes, businesses with the quality intentions locate themselves in an insights vacuum. You recognize you want insights to tell your decision making and you understand that you're missing them, however, don't recognize how quality to get them. Sentiment evaluation offers a few solutions into what the maximum vital problems are, from the attitude of customers, at least. Because sentiment evaluation may be computerized, selections may be made primarily based totally on a tremendous quantity of statistics as opposed to simple intuition.

1.4   OBJECTIVES OF THE WORK

Our main objective is to help support buyer decisions in Intraday trading of stocks in the Indian Stock Market. We aim to achieve our main objective by trying to predict stock prices and determining user sentiment.

We use two methods to predict stock prices, they are:

1. Using Regression and Neural Networks to foresee the closing a stock's price for the next day
2. Using Classification to determine whether the stock price will rise, fall or stay approximately the same for the next day

We also use Sentiment Analysis as a supporting factor as user sentiment plays a huge role in determining stock prices.

CHAPTER 2

LITERATURE SURVEY

A literature survey is an assignment of a previous task accomplished by certain writers, as well as the collection of information or data from research articles published in journals, in order to further our purpose. It's a technique that allows us to generate fresh notions and ideas.

There has been a lot of material written on the same task in the past; some of these articles are used to develop the project's principles.

2.1: TITLE: Survey of Stock Market Prediction Using Machine Learning Approach

Author: Ashish Sharma, Dinesh Bhuriya, Upendra Singh

Publisher: IEEE

Published in: International conference of Electronics, Communication and Aerospace Technology (ICECA) 2017

In this project they have they employed an efficient regression model to estimate stock market price using stock market data in their research. In the stock market, which is a complicated and difficult process, prediction plays a critical role. They did this by modelling and analysing several factors using regression analysis.

Merits and Demerits

This study provided a theoretical approach for predicting stock prices using a basic regression model and was developed. This study described regression models and how their applications can help with price forecasting. This study gave us an idea of how to provide some forecasts by choosing the appropriate factors that affect stock prices as variables.

2.2: TITLE: Analysis and Forecast of Nepalese Stock Prices Using Various Machine Learning Algorithms

Authors: Pushkar Khanal, Shree Raj Shakya

Published: Minutes of the 2016 IOE Graduate Meeting

Publisher: IEEE

In this project, we used machine learning algorithms to predict patterns and analyse stock price trends. Increased accuracy, reliability, and changeability have also increased our reliance on smart trading systems to help us predict and analyse stock prices in a variety of situations.

Forecasting methods like information gain centre, application technology group, classic duration analysis steps, and deep information technology are classified into four categories based on the type of data and tools each method employs to anticipate the market.

When compared to most other machine learning algorithms and traditional engineering methodologies, a support vector machine algorithm combined with boosting produced the most accurate results. However, even in linear regression, the mean square error is reduced.

2.3: TITLE: Multi-kernel learning to predict the direction of stock prices

Authors: Amit Kumar Shirohi, Pradeep Kumar Mahat, Ph.D. Vahida Atter

Publisher: IEEE

Published in: International Conference on Advances in Engineering & Technology Research

For stock price forecasting, this project used various kernel learnings. Its goal is to design a kernel model that integrates fixed-based kernels in a linear fashion. The selection of the proper pre-crafted and prepared kernel is critical to the success of an SVM. They used MKL to combine various kernel types.

The pre-processing component gathers and processes raw market data, then extracts technical characteristics or indications based on past stock prices and trade volumes, before normalising the entire function set.

On the normalised dataset, the forecasting component first generates different basis kernels (RBF and polykernel), then combines these base kernels in MKL, and then rises on the trading day after the previous trading day. Alternatively, the day you set the criterion for predicting the movement of daily stock price trends such as downtrends, or the day you set the criteria for predicting the movement of daily stock price trends such as uptrends. The Prediction Accuracy component analyses the prediction accuracy to compare the proposed and baseline approaches.

This endeavour needed the employment of the correct collection of functions in order to forecast the most precise numbers. In this paper, the two-stage concept was completely described. The first stage guided the user through the process of picking features such as open and closing prices, while the second stage built a new kernel to hold the received data. This gave me a firm grasp on how to apply the appropriate function to the model.

2.4: TITLE: Neural network by predicting stock market data

Authors: Rohit Verma, P Kumar Chore, Upendra Singh

Published in: International Conferences on Electronics, Communications and Aerospace Technologies

Publisher: IEEE

This article described an artificial neural network application for the task of stock index forecasting. They explained the theory behind ANN and the neural network model and its salient features.

In both cases, the outcomes were accurate. Unless there is a huge abrupt difference in the actual data, such as near the right edge, where it becomes impossible to anticipate the changes effectively, the predictions are accurate. On the other hand, this backs up the idea that the stock market is unpredictably volatile. For a single hidden layer, the minimal error in test and training data was as low as 3.5 percent.

As a result, we can see that neural networks are a powerful tool for forecasting stock markets, and that they can be applied to real-world datasets like the Nifty. It was discovered as a result of this experiment that deep learning models can greatly enhance accuracy. A theoretical strategy for predicting stock values using artificial neural networks has been developed by Rohit Verma and colleagues. The data was gathered from the Nifty Stock Index dataset and was based on numbers from the previous few days. The accuracy of this study was 96 percent.

CHAPTER 3

SYSTEM REQUIREMENTS AND SPECIFICATION

3.1 SYSTEM ANALYSIS

It is a method of enquiring and understanding an existing problems system, defining requirements and analyzing the best possible solution is the best way to increase the efficiency of the existing system. It is also considered to be a great problem-solving methodology that will break the existing system into different component modules and evaluate each module to determine the efficiency of the system individually, by doing module wise analysis we will be able to individually increase the speed of each module.

FEASIBILITY STUDY

A feasibility study is a preliminary examination of potential user information and an assessment of the system's resource requirements, benefits, costs, and viability. To include and run the program, a feasibility analysis takes into account a number of constraints. At this point, the resources required to implement computing equipment, people, and expenditures are estimated. The standard resources are contrasted, and a framework cost-cutting advantage analysis is conducted. The process of feasibility analysis includes analyzing the problem and gathering all relevant venture data. The primary purpose of the feasibility study is to establish whether the arrangement is economically feasible.

There are several elements to examine, including specialized achievability, hierarchical achievability, and scheduling practicality. The feasibility analysis examines three areas:

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- OPERATIONAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is being conducted to see whether the system would be helpful to the company. The company's ability to invest in research and development for the system is limited. It is required to demonstrate that the costs are justified. As a result, the complicated system was completed on time and under budget, thanks to the availability of cutting-edge technology.

TECHNICAL FEASIBILITY

This is owing to the lack of a comprehensive framework plan at the time, making factors like execution, expenses (according to the type of innovation to be provided), and so on difficult to obtain. When doing a technical examination, consider the following points: comprehend the many innovations involved with the proposed framework. Before we begin the project, we need be very specific about the technology that will be used to support the new framework. Is there a requirement for innovation? Our "Tweezer" programming is possible since all the necessary elements are readily available.

OPERATIONAL FEASIBILITY

The goal of the study is to establish the system's level of acceptance amid users. This covers the concept of teaching the end-user how to use the technology. The user must not be swayed by the system, but rather accept it as a requirement. The acceptability of stoners is totally dependent on the methods used to educate and familiarize the user with the system. As the system's last user, his confidence level must be boosted so that he is also capable of providing constructive feedback.

3.2 FUNCTIONAL REQUIREMENTS

The software's functionality is addressed via functional requirements from an engineering standpoint. It improves and describes the component flow as well as the structural flow of the same. The functional statement is concerned with categorizing raw datasets and learning from the same dataset. After that, the datasets are grouped into clusters, and their impairment is assessed for efficiency. The data is cleansed once the dataset is cleaned, and the machine learns and identifies the pattern set for the same, going through various iterations and producing output.

3.3 NON-FUNCTIONAL REQUIREMENTS

It is concerned with non-functional external factors. It's a tool for analyzing data. The operations are judged for their performance under the same conditions. Because stock is feasible and always changing, these additional effects and needs enable it to obtain the most recent updates and integrate them in a single step, allowing specialists to work on and resolve any bugs or draughts that may arise.

The efficiency and hit gain ratio are the non-functional parameters that must be met. The code's usability for increased effectiveness and to implement and search for the security console. The system is dependable, and its performance is maintained thanks to the system's integration and portability.

3.4 TOOLS AND TECHNOLOGIES REQUIRED

Hardware Requirements

Intel i5 or higher processor is required.

A minimum of 225MB of RAM is required.

Hard Disk: You'll need at least 2 GB of hard disc space.

Input Devices: The input devices are the keyboard and mouse.

The monitor or laptop screen is employed as the output device.

Software Requirements

1. IDE
    a. PyCharm for Application Development
    b. Jupyter Notebook for Testing and Prototyping
2. Python Packages
    a. Keras
    b. Scikit-Learn
    c. NLPK

CHAPTER 4

SYSTEM DESIGN

System design is mechanism or the flow of outlining the structural design, sections, for a system to meet its specific requirements, it must have interfaces and data. The system design is critical to the creation of the outline and the development of the product. The primary goal of this study is to offer the system's needed functionality to the end user. User must obtain the predicted output in accurate manner and time-efficient manner.
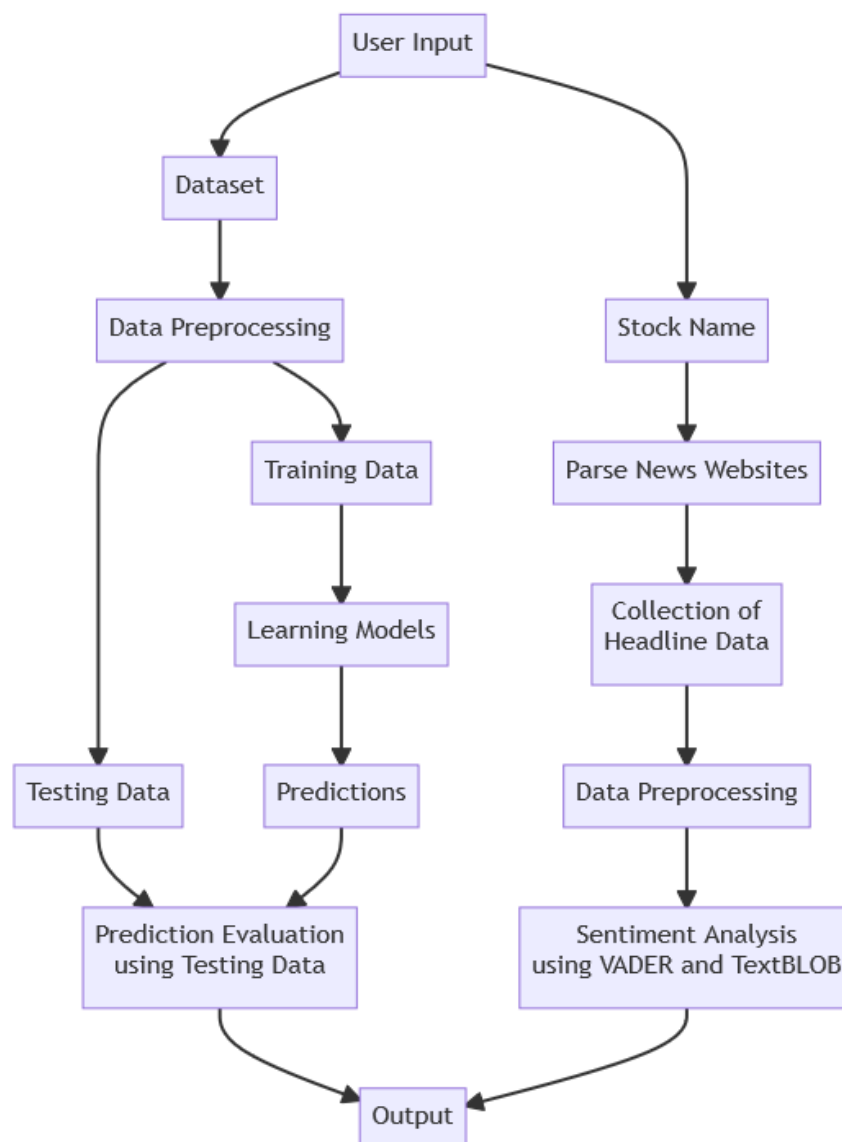
4.1 SYSTEM ARCHITECTURE

Figure-4.1: Overall System Architecture for Price Prediction and Sentiment Analysis

4.2 INPUT/OUTPUT DESIGN

We've developed a GUI for our application, we used the Python Tkinter to build the GUI.

Tkinter is available on Windows and Linux/Unix.

There are 2 main components in our GUI.

1. Data Input

a.   Stock Name

b.   Dataset Input (CSV File)

2. Output

The output section is divided into two categories

a.   Prediction Output

      i.   Price Prediction

     ii.   Price Classification

b.   Sentiment Analysis Output

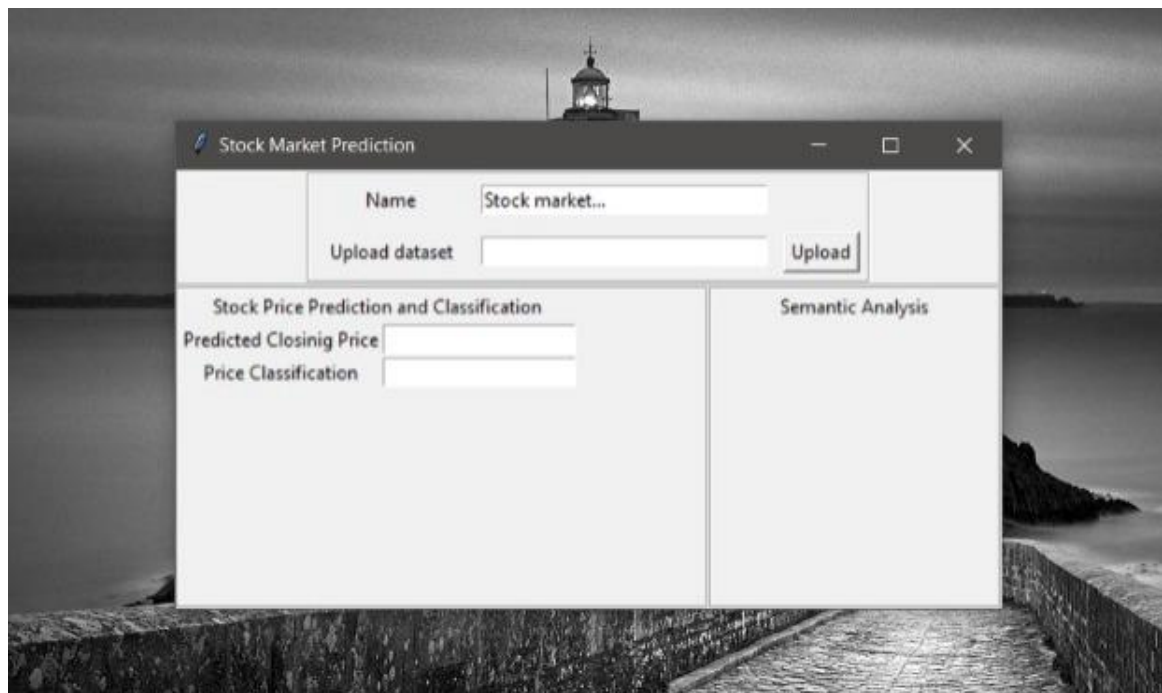The Image below shows a rough version of the GUI and its structure

Figure-4.2: User Interface prototype

4.3 OBJECT ORIENTED DESIGN

4.3.1 SEQUENCE DIAGRAM

Software developers and business analysts use these diagrams to understand the requirements for a new system or to describe an existing process.

Other names for sequence diagrams are event diagrams and event scenarios.

Sequence diagrams can be used as a guidance for companies and other organizations.

• To depict the intricacies of a UML use case, utilize a flowchart.

• Create a logic model for a complex procedure, function, or process.

• Look at how different products and components work together to execute a task.

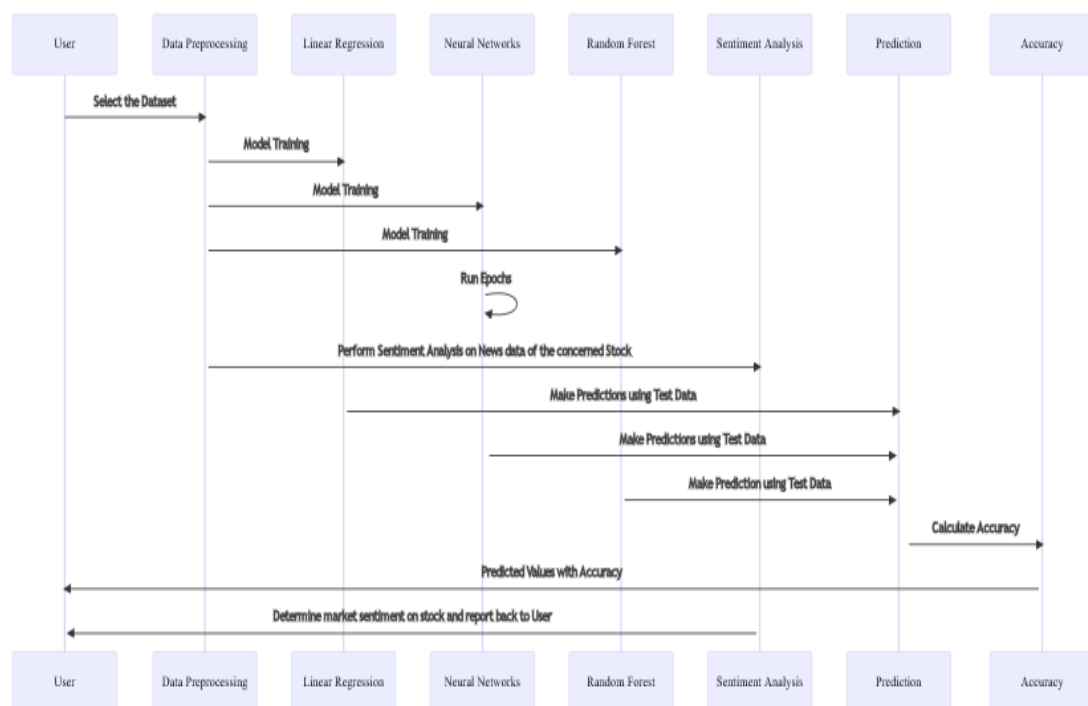• Create a detailed plan and comprehend how a current or future scenario function.

Figure-4.3: Sequence Diagram

4.3.2 ACTIVITY DIAGRAM

A behavioral diagram that explains how a system functions is known as an activity diagram. The control flow from a start point to an end point is depicted in an activity diagram, as well as the numerous paths that can be taken within the activity.
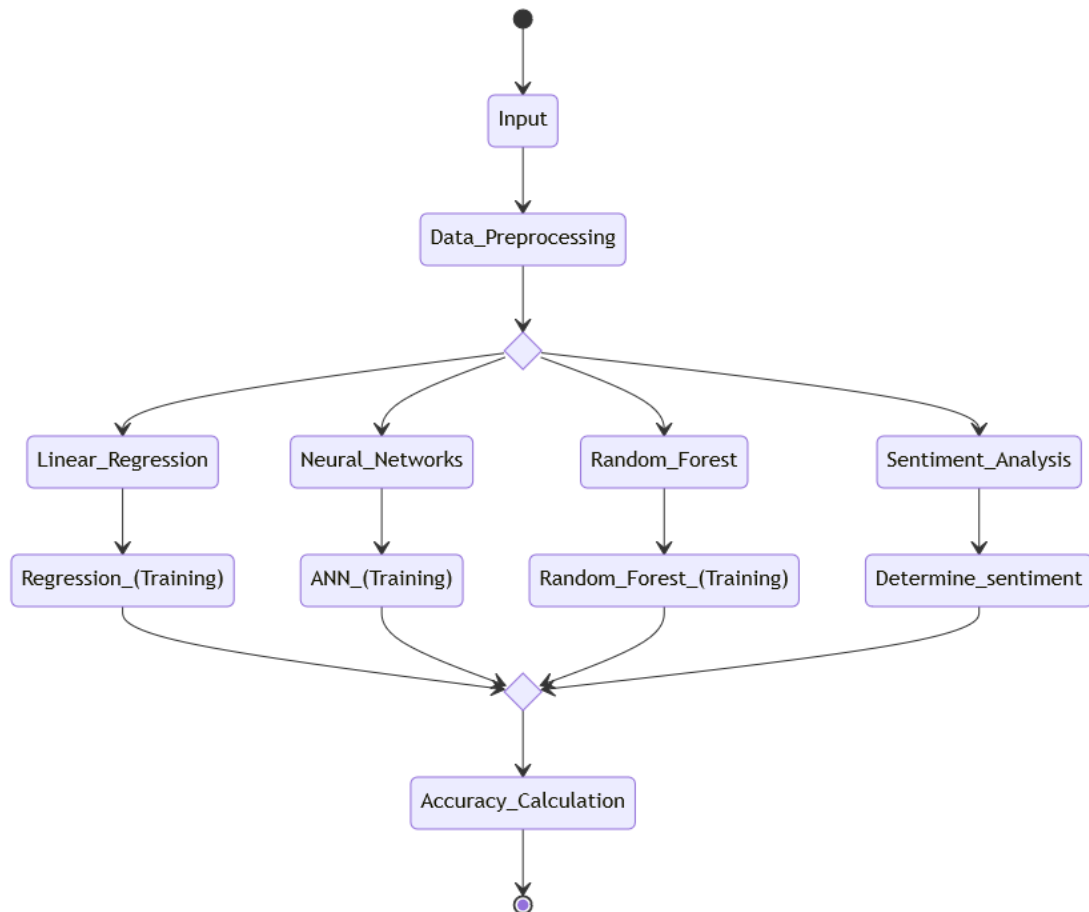

Figure-4.4: Behavioral Diagram

4.3.3 DATAFLOW DIAGRAM

A data flow diagram depicts the process elements and is a graphical representation of data "flow" across an information system.

A DFD is often used as a first stage to offer a high-level overview of the system without getting into too much detail, and it can subsequently be modified.

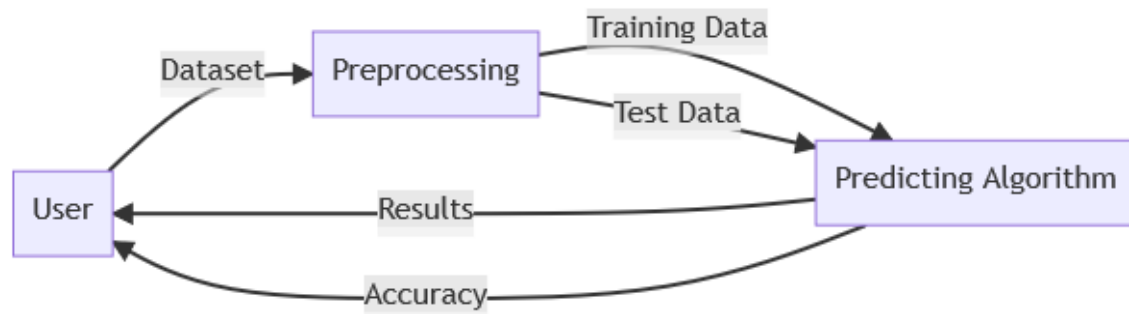Data processing can also be visualized using DFDs.

Figure-4.5: Dataflow diagram

4.4 ALGORITHM

4.4.1 MULTIVARIATE LINEAR REGRESSION

The generalization of univariate linear regression is multivariate linear regression.

There are multiple independent variables, x1,x2,xn, and a dependent variable, as the name implies.

x(i) is the ith training example's input (features).

m denotes the number of training scenarios.

n denotes the number of characteristics.

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 \ldots \ldots \theta_n x_n$

where $\theta_0, \theta_1, \theta_2, \theta_3, \theta_4 \ldots \ldots \theta_n$ are weights

Gradient descent for multiple variables: -

- Our cost function is

$$J(\theta_0, \theta_1, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Repeat $\{$

$$\theta_0 := \theta_0 - \alpha \underbrace{\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})}_{\frac{\partial}{\partial \theta_0} J(\theta)}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})x^{(i)}$$

(simultaneously update θ_0, θ_1) $\}$

Normal Equation for multiple variables

One method of minimizing J is to use gradient descent.

By directly obtaining its derivatives with respect to the j's and setting them to zero, we will reduce J using the "Normal Equation" technique.

This eliminates the need to iterate in order to find the best theta.

The normal equation formula is as follows: $\theta = (X^TX)^{-1}X^Ty$

4.4.2 RANDOM FOREST

The Random Forest procedure functions by instructing numerous unstable classification trees with the help of a secure amount of unintentionally chosen qualities. Later captivating the approach of separate class to construct a robust classifier. Once the training set for the present tree is sketched by sampling with a substitute, almost one-third of the instances are ignored from the sample. The missing information is swapped by proximities, locating outliers, and producing illuminating low-dimensional interpretations of the data. Nevertheless, as this mode picks a controlled number of qualities in each iteration, the implementation of random trees is quicker. The random-trees technique initiates a component of unpredictability into the prototype. Instead of viewing for the elite quality though splitting a node, it examines for the finest items amongst an unplanned subset of qualities. This development commonly produces in an enhanced model.
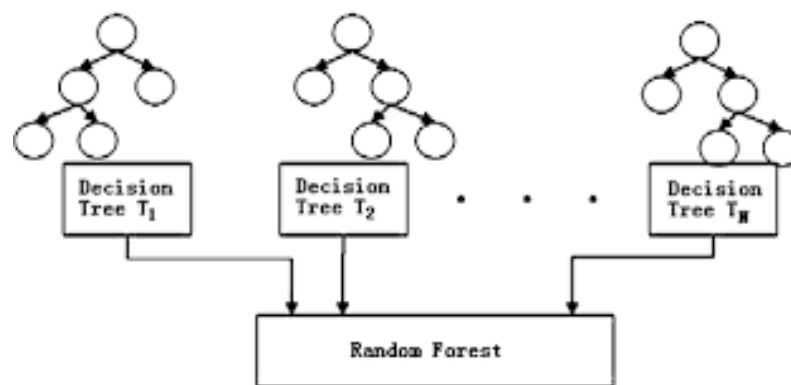

Figure-4.6: Random Forest

4.4.3 Neural Networks using Backpropagation algorithm: -

A neural network is a collection of interconnected I/O units with each connection given a weight based on its attributes. It makes it easier to build prediction models from vast datasets. Neural community training relies heavily on backpropagation. It's a method for fine-tuning the weights of a neural network in the first epoch using the error rate from the previous epoch (i.e., generation). You may reduce errors and increase the model's generalization by fine-tuning the weights, making it more predictable.

Backpropagation is a well-known synthetic neural network training method. In shortened form, it stands for "backward spread of mistakes." This method makes estimating the gradient of a loss characteristic while taking all the variables into account easier.

What is the Backpropagation Algorithm and How Does It Work?

The chain rule is used in the back propagation strategy in neural networks to compute the gradient of the loss characteristic for a single weight.

Unlike a native direct calculation, it efficiently computes one layer at a time.

It works out the gradient.

It expands the computation scope of the delta rule.

Take a look at the diagram below to see an example of a back propagation neural network:
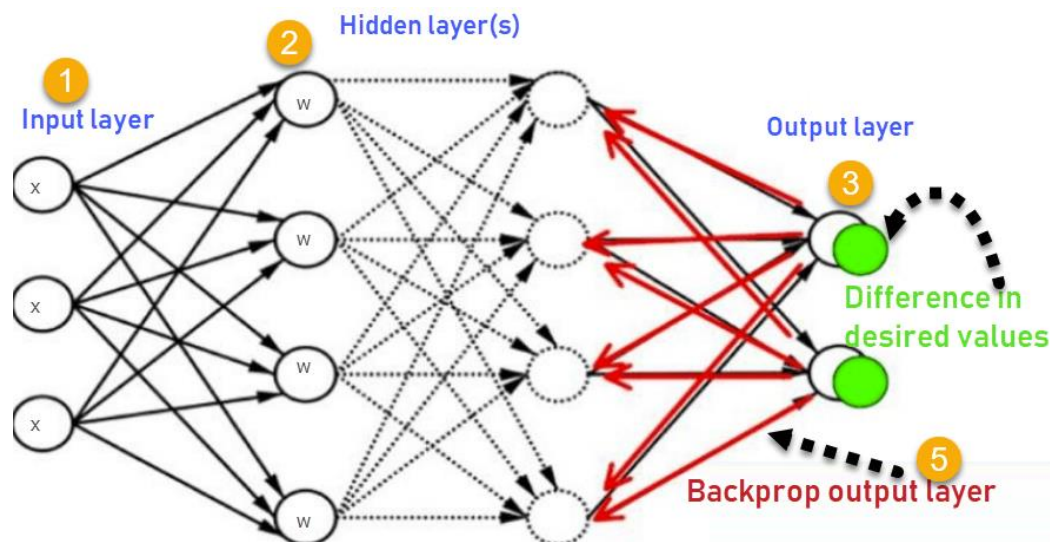

Figure-4.7: Neural Network Structure with Backpropagation

1. A pre-determined path is followed by X inputs.

2. Weights in real life the letters W are used to imitate the input. Usually, the weights are picked at random.

3. Calculate the output for each neuron from the input layer to the hidden layers to the output layer.

4. Calculate how much error there is in the outputs.

REFERENCES

[1] TrevirNath "How Big Data Has changed finance "
https://www.investopedia.com/articles/active-trading/040915/how-big-data-has-changed-finance.asp.

[2] Saheli Roy Choudhury."Machines will soon will be able to learnwithout being programmed" https://www.cnbc.com/2018/04/17/machine-learning-investing-in-ainext- big-thing.html

[3] Yahoo finance data -https://in.finance.yahoo.com/

[4] Rohit Verma Astral institute Technical and Research, Indore (M.P.) , PROF. Pkumar Choure (CSE) Astral institute Technical and Research, Indore (M.P.)"Neural Networks through Stock Market Data Prediction " International Conference on Electronics,Communication and AerospaceTechnology ICECA 2017

[5] Pushkar Khanal ,Shree Raj Shakya "Analysis and Prediction of Stock Prices of Nepal using different Machine Learning Algorithms" Department of Mechanical Engineering, Pulchowk campus, Institute of Engineering, Tribhuvan University, Nepal

[6] Yaojun Wang, Yaoqing Wang "Using Social Media Mining Technology to Assist in Price Prediction of Stock Market" .

[7] Ashish sharma , dinesh bhuriya , upendra singh "Survey of Stock Market Prediction Using Machine Learning Approach "International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.

[8] Amit Kumar Sirohi, Pradeep Kumar Mahato, Dr. Vahida Attar "Multiple Kernel Learning for Stock Price Direction Prediction " IEEE International Conference on Advances in Engineering & Technology Research (ICAETR - 2014),August 01-02, 2014, Dr. Virendra Swarup Group of Institutions, Unnao, India

[9] Book-Elements of Artificial Neural Network,Kishan Mehrotra,Vhilukuri K K Mohan,Sanjay Ranka.