**PROFESSIONAL TRAINING REPORT**

**At**

**Sathyabama Institute of Science and Technology**

# (DEEMED TO BE UNIVERSITY)

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

By

**CH GOPI KRISHNA**

**(Reg.no. 40110233)**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SCHOOL OF COMPUTING**

**SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY**

**JEPPIAAR NAGAR,RAJIV GANDHI SALAI,**

**CHENNAI-600119, TAMILNADU.**

**NOV 2022**

SATHYABAMA

## INSTITUTE OF SCIENCE AND TECHNOLOGY

## (DEEMED TO BE UNIVERSITY)

**Accredited with Grade “A” by NAAC**

**JEPPIAAR NAGAR,RAJIV GANDHI SALAI, CHENNAI - 600119**

**[www.sathyabama.ac.in](http://www.sathyabama.ac.in/)**

## SCHOOL OFCOMPUTING

**BONAFIDE CERTIFICATE**

This is to certify that this Professional Training Report is the bonafide work

Of **CHALLAGUNDLA GOPIKRISHNA (Reg.no.40110233)** who carried out the Project entitled **“ANALYSIS OF RAINFALL PREDICTION”** under our supervision from **December** 2021 to April 2022.

**Internal Guide**

**Dr.Sankari.M, M.E., Ph.D.,**

**Head of the Department**

**Dr.L.LAKSHMANAN, M.E., Ph.D.,**

**submitted for Viva voce Examination held on**

**Internal Examine External Examiner**

# DECLARATION

**I, CHALLAGUNDLA GOPIKRISHNA (Reg. No. 40110233)** hereby

declare that the Professional Training Report on **“ANALYSIS OF RAINFALL PREDICTION”** done under the guidance of **Dr.SANKARI,M.E., Ph.D.,** at **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

**DATE:**

**PLACE: CHENNAI 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 of the Department, Department of Computer Science and Engineering** for providing us the necessary support and details at the right time during the progressive reviews.

I convey my thanks to **Dr. S. VIGNESHWARI, M.E., Ph.D.,and Dr.L.LAKSHMANAN, M.E., Ph.D.,Heads of the Department, Department of Computer Science and Engineering** for providing us the 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 **Dr.SANKARI.M, M.E., Ph.D.,** for his valuable guidance, suggestions

ons 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**

This study seeks a distinctive and efficient machine learning system for the pediction of rainfall. The study experimented with different parameters of the rainfall from INDIAN CLIMATE in order to assess the efficiency and durability of the model. The Linear Regression,Random Forest and neural networks model is focused on this study. The learning of data is completed using hybrid and backpropagation network algorithm. The rainfall parameters in this study are collected, trained and tested to achieve the sustainable results through ANFIS and ANN models. The monthly rainfall predictions obtained after training and testing are then compared with actual data to ensure the accuracy of the model. The results of this study outline that the model is successful in predicting the monthly rainfall data with the particular parameters. The training and testing of data Linear Regression,random forest model helped in not only minimizing the errors up to RMSE of 0.055, 0.035 and 0.025, but also maximizing the reliability and durability of the predicted data. The results of the study highlight that the ANFIS model is most suitable among the artificial networks for the rainfall prediction. The outcome data with ANFIS system presented maximum accuracy with minimum error through the comparison between the actual data and predicted outcome data.

# TABLE OF CONTENTS

**CHAPTER TITLE PAGE NUMBER**

**NO**

|  |  |  |
| --- | --- | --- |
|  | **ABSTRACT** | **v** |
|  | **LIST OF FIGURES** | **ix** |
|  | **LIST OF ABBREVATIONS** | **x** |
| **1.** | **CHAPTER1-INTRODUCTION** | **01** |
|  | 1.1.Aim of study | 04 |
|  | 1.2 .Significance of study | 04 |
|  | 1.3.Limitations of study | 05 |
|  | 1.4.Problem Statement | 05 |
|  | 1.5 Methdology | 05 |
|  | 1.6.The Study Region and Data | 05 |
|  | 1.7.Overview Of Study | 06 |
| **2.** | **CHAPTER 2 -LITERATURE REVIEW** | **07** |
| **3.** | **CHAPTER 3 -RAINFALL** | 11 |
|  | 3.1 Introduction To Rainfall | 11 |
|  | 3.2 Types of Rain | 13 |
|  | 3.3 Measurments Of Rainfall | 16 |
|  |  |  |
| **4.** | **CHAPTER 4-MACHINE LEARNING TECHNOLOGIES** | **21** |
|  | 4.1.Introduction To Machine Learning | 21 |
|  | 4.2.Linear Regression | 23 |
|  | 4.3.LASSO Regressor | 25 |
|  | 4.4.MAE | 28 |
|  | 4.5 MSE | 28 |
|  | 4.6 RMSE | 29 |
| **5.** | **CHAPTER 5-SIMULATION** | **30** |
|  | 5.1.Data Processing | 30 |
|  | 5.2.Data Pre-Processing | 31 |
|  | 5.3 Data Visualization | 33 |
|  | 5.4.Flow Chart For Rainfall Prediction | 34 |
|  | 5.5.Selection Of Input and Output Data | 34 |
|  | 5.6 Feature Extraction | 35 |
|  | 5.7.Training Testing and validation | 36 |
|  | 5.8.ANN | 36 |
| **6** | **CHAPTER 6-DISCUSSIONS AND RESULTS** | **37** |
|  | 6.1.Modelling | 37 |
|  | 6.2Comparing Results | 38 |
| **7** | **CHAPTER 7-ANALYSIS** | **41** |
|  | 7.1 Introduction | 41 |
|  | 7.2.Existing system | 41 |
|  | 7.3.Proposed System | 41 |
|  | 7.4 Software Requirement specifications | 49 |
| **8** | **CHAPTER 8-DESIGN** | **51** |
|  | 8.1.Class Diagram | 52 |
|  | 8.2.Use Case Diagram | 53 |
|  | 8.3 Sequence Diagram | 54 |
|  | 8.4 Collaboration Diagram | 55 |
|  | 8.5.Component Diagram | 55 |
|  | 8.6.Deployment Diagram | 56 |
|  | 8.7 Activity Diagram | 57 |
|  | **8.8.Data Flow Diagram** | **58** |
| **9** | **CHAPTER 9-SCREENSHOTS** | **60** |
|  | 9.1 ML CODE SCREENSHOTS | 60 |
|  | 9.2 GUI Screen Code ScreenShots | 62 |
|  | 9.3 GUI Screens PICS | 64 |
|  | **CONCLUSIONS…….** | **65** |
|  | **REFERENCES….** | **66** |
|  | **APPENDIX** | **67** |

# LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| FIG.NO | FIG NAME | PAGE NO |
| 3.1 | Conventional precipitation | 13 |
| 3.2 | Orographic Rainfall | 14 |
| 3.3 | Cyclonic or frontal rainfall | 15 |
| 4.1 | Neuron Schema | 22 |
| 4.2 | Structure of ANN | 22 |
| 4.3 | Linear regression | 24 |
| 4.4 | Hypothesis of linear regression | 24 |
| 4.5 | Cost Function j | 25 |
| 4.6 | Lasso regression | 26 |
| 5.1 | Flow Chart | 34 |
| 6.1 | Modelling | 37 |
| 6.2 | Linear Regression | 38 |
| 6.3 | Lasso Regression | 38 |
| 6.4 | Random Forest Model | 40 |
| 7.1 | STLC | 42 |
| 8.1 | Class Diagram | 52 |
| 8.2 | Use Case Diagram | 53 |
| 8.3 | Sequence Diagram | 54 |
| 8.4 | Collaboration Diagram | 55 |

**LIST OF ABBREVIATIONS**

# ABBREVIATION EXPLAINATIONS

AI ARTIFICAL INTILLIGENCE

ANN ARTIFICAL NEURAL NETWORK

MAE MEAN ABSOLUTE ERROR

MSE MEAN SQUARE ERROR

RMSE ROOT MEAN SQUARE ERROR

**CHAPTER1**

**Introduction**

Rainfall play important role in forming of fauna and flora of natural life.It is not just significant for the human beings but also for animals, plants and all living things. It plays a significant role in agriculture and farming and undoubtedly; water is one of the most natural resources on earth. The changing climatic conditions and the increasing greenhouse emissions have made it difficult for the human beings and the planet earth to experience the necessary amount of rainfall that is required to satisfy the human needs and its uninterrupted use in everyday life. Therefore, it has become significant to analyze the changing patterns of the rainfall and try to predict the rain not just for the human needs but also to predict for natural disasters that could cause by the unexpected heavy rainfalls. To be more specific and aware of the devastating climatic changing and stay updated; predicting rainfall has been the focus of computer scientist and engineers.

This study is focusing on predicting rainfall using Linear Regression ,Random Forest and Artificial Neural Network. The rainfall prediction will not just assist in analyzing the changing patterns of rainfall but it will also help in organizing the precautionary measures in case of disaster and its management. The rainfall prediction would also assist in planning the policies and strategies to deal with the increasing global issue of ozone depletion. The changing patterns of rainfall are associated much with the global warming; that is increasing of the earth’s temperature due to increased Chlorofluorocarbons emitting from the refrigerators, air conditioners, deodorants and printers etc. that are the significant part of everyday life.

The increasing temperature is actually affecting the climate considerably . Similarly, the rainfall prediction and weather updates not only help in managing the macro level problems like flood and agricultural issues because of poor or extreme rainfall.The rainfall prediction could also contribute to the well-being and comfort of the people by keeping them informed by tracking the rainfall patterns and predicting the rainfall by Linear Regression,Lasso Regressor,Random Forest and Artificial Neural Network. The rainfall predictions help the people to deal with hot and humid weather. The technological development in the modern world has expanded the space for innovation and revolution. Although the issues concerned are probably associated with these technological advancements but one needs to consider the range of possibilities and opportunities that this technological evolution has opened to the human beings.

In addition, the inappropriate or poor rainfall prediction is also one of the reasons that are problematic in the water reserve management. The precise and correct rainfall prediction can not only contribute to the effective and efficient utilization of this natural resource but it can also help in managing the projects and plans for power generation. For this purpose, it is very important to design and operate on a system that would assist in accurate prediction and easy access to the users.

The rainfall prediction will also integrate adaptive Linear Regression,Lasso Regressor,Random Forest with ANN for an increased accuracy and enhanced quality of the predicted output. To analyze the performance of these algorithms; co-relation coefficient will be a key indicator in this study. ANN is the most competent and effective tool for prediction of rainfall that actually contributes to the most accurate forecasting.The Linear Regression,Lasso Regressor,Random Forest is also one of the effective algorithms used for data analysis for the classification. It assigns categories and allocates cases to similar groups/categories. So, each time a data is analyzed; it assigns that data to the most suitable or most similar category it belongs to. This helps in making the regression and allows the user to make a prediction for the similar sets of data or information received each time.

However, rainfall prediction with ANN using backpropagation and hidden layer approach integrated with Linear regression is intended to produce precise and more accurate forecasts. The predictions could be utilized for a maximum range of purposes and thus can play a vital role in minimizing the issues associated with water reserves, agricultural problems with changing climatic conditions and flood management. The appropriate utility and implication of the estimated outcomes could also support the policy and development of strategies about resource management and control with a variety of techniques and approaches that will actually impact the human life in many ways.

* 1. **Aim Of Study**

The aim of the study is the prediction of the rainfall using historical monthly data based on artificial intelligence methodologies such as Neuro-Fuzzy and artificial neural network. The extraction procedures/algorithms will produce the output by classification of the data according to the categories using Neuro-Fuzzy. The similar data will be grouped for the accurate and precise information that will predict rainfall more correctly and with perfect figures. The accurate and exact predictions will help in developing the more appropriate strategies for agriculture and water reserves and will also be informed about the flood to implement precautionary measures. The data for the rainfall prediction is collected from Metrology Department of Erbil, Nicosia and Famagusta. This is the monthly data with all parameters of rainfall including wind speed, direction, air pressure, humidity and temperature. The aim of the proposed study is too effective and efficient in predicting the rainfall with accuracy and precision.

* 1. **Significance Of Study**

Rainfall prediction is significant not only on the micro but also on the macro level. The study is of significance with respect to its vital contribution in the field of agriculture, water reserve management, flood prediction and management with an intention to ease the people by keeping them updated with the weather and rainfall prediction. It is also important to be utilized by the agricultural industries for keeping their crops safe and ensure the production of seasonal fruits and vegetables by updated rainfall prediction. The study will also be significant for the flood management authorities as more precise and accurate prediction for heavy monsoon rains will keep the authorities alert and focused for an upcoming event that of which the destruction could be minimized by taking precautionary measures. The rainfall prediction will impressively help in dealing with the increasing issue of water resource management; as water is a scarce resource and it needs to get saved for the benefit of human beings themselves. Also, it will help the people to manage and plan their social activities accordingly.

* 1. **Limitations Of Study**

1. The data sample is limited to monthly statistics only and does not provide the daily output predictions.

b) The climatic change and the global warming effect may impact the accuracy of the expected output

c) The locations for the data processing used in this study are geographically different and distanced that could also impact the correlation efficient that will measure the performance of the ANFIS and ANN in this research.

d) The system discussed in this particular study will operate with Matlab software

* 1. **Problem Statement**

The accurate and precise rainfall prediction is still lacking which could assist in diverse fields like agriculture, water reservation and flood prediction. The issue is to formulate the calculations for the rainfall prediction that would be based on the previous findings and similarities and will give the output predictions that are reliable and appropriate. The imprecise and inaccurate predictions are not only the waste of time but also the loss of resources and lead to inefficient management of crisis like poor agriculture, poor water reserves and poor management of floods. Therefore, the need is not to formulate only the rainfall predicting system but also a system that is more accurate and precise as compared to the existing rainfall predictors.

* 1. **Methdology**

Spiral model of programming by creating a V0 and test it for feedback from the test sample. It will retrieve the possible alterations to create the next Version of the algorithm. The test will be by stimulating the neural network to retrieve results by archiving.

# The Study region and data

The metrological data including humidity, air pressure, wind speed, wind direction and temperature will be analyzed for three cities.INDIA and it has tall mountains and experiences heavy rain every year. Although, it has very nominal humidity; the data collected for INDIA has some appreciated predicted outcomes.

**1.7 Overview of the Study**

The thesis is designed as follows:

**Chapter 1** is an introduction to the topic of the thesis. Chapter 1 outlines the overview of the study, discusses the aims and significance of this study.

**Chapter 2** reviewed the existing literature and highlighted the previous research on the proposed thesis.

**Chapter 3** is highlighting the general overview to the rainfall prediction and presents the explanation of the rainfall prediction in the field of agriculture, water reservation and flood prediction. It also grants information for the methods and approaches for the accurate prediction in depth.

**Chapter 4** is focused on the explanation of the artificial neural system with the ANFIS and several modelling techniques like RANDOM FOREST algorithms are discussed in details.

**Chapter 5** is an explanation of the stimulation of the data and presents the pre-processing and correlation between the input and output. The application of ANN and ANFIS is also discussed in detail.

**Chapter 6** is the demonstration of results and discussions of the study. It also discusses root- mean-square error to evaluate the best findings for analyzing the accurate rainfall prediction.

**CHAPTER 2**

**LITERATURE REVIEW**

Rainfall prediction is not an easy job especially when expecting the accurate and precise digits for predicting the rain. The rainfall prediction is commonly used to protect the agriculture and production of seasonal fruits and vegetables and to sustain their production and quality in relation to the amount of rain required by them .The rainfall prediction uses several networks and algorithms and obtains the data to be given to the agriculture and production departments. The rainfall prediction is necessary and mandatory especially in the areas where there is heavy rainfall and it’s more often expected. There are huge economies like those of Asia like India and China that that earn a large proportion of their revenue from agriculture and for these economies; rainfall prediction is actually very important .

The rainfall forecasting is prevailing as a popular research in the scientific areas in the modern world of technology and innovation; as it has a huge impact on just the human life but the economies and the living beings as a whole. Rainfall prediction with several Neural Networks has been analyzed previously and the researchers are still trying hard to achieve the more perfect and accurate results in the field of rainfall prediction . The prediction of seasonal rainfall on monthly basis by using the surface data to form annual prediction is also essential for the agricultural activities and therefore the production and supervision of the agriculture and crops. It could be done by recognizing the variations in the supply of moisture in the air. The case of African region illustrates that how this succeeded and how West Africa advantaged from the rainfall prediction in managing their agricultural activities .

Similarly, the short-term streamflow forecasting for the rainfall is also reliable and bias-free. But they are not much effective in predicting the flood and post-processing of rainfall prediction. An approach called raw numerical weather prediction (NWP) was introduced in 2013, where the approach focused on the Bayesian joint probability model to formulate prediction data. The approach formed forecast possibility distributions for each location and it had prediction time for it; collaborative forecasts correlated with space and time was produced in the Southern part of India. This approach focused on Schake shuffle to produce the forecast by the forecast possibility distributions.

Furthermore, the short-term streamflow forecasting could also be used through the artificial neural networks as researched by Zealand, Burn and Simonovic in 1999. The study conducted outlined that ANNs ability to forecast for short-term stream flow and outlined some of the issues that the approach encountered with ANN. Although, ANNs with short-term stream flow can calculate and present complex and nonlinear relationship between input and output with an ability to outline the interface effect as well but has issues in processing some input data with certain type and number. The ANNs also encountered difficulty with dimensions of the hidden layers. This research outcome was represented by the data of Winnipeg River system in Ontario, Canada using the quarter monthly data. The outcomes of the study were encouraging with AANs performed quite well for the four prediction lead-times. The RMSE for the test data of 8 years outlined variation from 5cms to 12.1cms in a forecast from four-time step to two-time step ahead respectively.

Also, the recent decade highlighted the significance of artificial intelligence and it has gained attention in water resource management and engineering as well. ANNs, ANFIS and GP are the driving simulations of AI and they are advantaged over other systems and approaches because of being more reliable and competitive. The adaptive neuro-fuzzy inference system (ANFIS) for time series and ANN for predicting streamflow in Apalachicola River, the United States with that of other neural network techniques like hybrid ; when compared to wavelet-gene expression’ programming approach outlined the following results; ARMA model predicting accurate results for 1 day ahead time whereas, ANFIS forecasted the results for 2 days ahead time. The results from AI using ANFIS were more accurate and could predict 2 days ahead of time data rather than GEP and ANN . But for the 3 days forward data; ANN performed better than other models. For the monthly data; ANN, ANFIS and GEP outperformed as compared to ARMA models in the first part of the study.

Water as is one of the most useful resources of the earth. There is no human and living thing on earth that can survive without water. As, this precious resource is running out because of the increasing temperature of the earth and the unexpected and unappreciated climatic conditions due to global warming.In addition, the comparison among different neural models revealed that Non-linear autoregressive exogenous networks (NARX) and back propagation neural BPN) performed better than distributed time delay neural network (DTDNN) cascade-forward back propagation neural network (CBPN) in outlining more accurate and precise results for rainfall prediction. In comparison, statistical forecasting methodology can also be used for the rainfall prediction that outlines by using two different approaches like traditional linear regression and polynomial-based nonparametric; where nonparametric method outlined more competing results. Both the approaches could predict the 1-3 monthly rainfall forecasting data that could actually impact water resource planning and controlling . The periodic and episodic rainfall data for the south-west peninsula of England has also exposed that atmospheric characteristics are key players of outlining the monthly and seasonal forecast.

The rainfall prediction is also emphasized for its significance for the prediction of flood and consequently takes the precautionary measure to save the people from devastating destructions that a flood can cause. There are studies that outlined the significance of rainfall prediction in forecasting flood on the regions where there is heavy rain every year. The areas with high risk for flood are the vulnerable areas that need the rainfall forecasting not just to save a human life but to safe agriculture, water reservation and livestock.

In comparison, the significance of rainfall prediction is also important for areas with high probability for the drought. The areas with high drought seasons are also vulnerable to high risk in terms of agriculture and livestock with an extreme threat to human life as a whole; the study conducted for Godavari river Of India. The artificial neural network model for rainfall prediction of 1to 6 hour a head tie is studied. The study outlined that within artificial neural networks, using six models utilizing rainfall parameters like humidity, air pressure, wind direction and wind speed can give more accurate and precise prediction when previous forecasting data is also used with these parameters as an input as well.Nevertheless, land sliding is another natural hazard that could be caused due to heavy rainfall. The rainfall prediction could assist in combating the devastation caused by land sliding. The rainfall prediction for the areas vulnerable to land sliding is an essential part of artificial intelligence within engineering and management fields. The metrological and hydrological centres are struggling hard to produce the more competitive and precise rainfall prediction in order to overcome these issues that the rainfall can cause and their efforts have marked quite an improvement in the rainfall prediction and forecasting data for many models using the neural networks. The prediction for extreme rainfalls is useful for not just the metrological departments in sharing in time alerts but also for the hydrological departments in order to form better safety measures for example the flood prediction in India.

The rainfall prediction systems are much popular with artificial neural networks and the rainfall prediction departments like the metrology and hydrology engineering with management. The rainfall prediction using the neural network aims at predicting more efficient and more accurate results and precise predictions for a more useful and reliable output that could be used by the management and engineering departments in designing the plans and policies that will not only increase efficiency but it will also enhance the management systems from a quality data produced by using the Artificial Neural Networks. The study conducted with the different networks highlighted different results by operating within same training functions and outlined that back propagation neural network is capable of obtaining more precise predictions. Also, that increased neurons can decrease errors (MSE). Neural networks have proved capability for the rainfall prediction and in obtaining accuracy with precision among the other networks with other modelling techniques .

.

**CHAPTER 3**

**RAINFALL**

# **3.1**Introduction to rainfall

Rainfall is one the most significant atmospheric occurrence that is not only useful for the environment itself but for all the living beings on the earth. It affects everything directly or indirectly and because it is one of the most important natural phenomena; it is also important for the human beings to ponder on the precipitation changes with the change in climate (Alpers & Melsheimer, 2004). The rainfall has a significant impact on the universal gauge of atmospheric circulation and it affects the local weather conditions as well. The rainfall helps in balancing the increasing temperature and in the survival of the human beings (Trenberth, 2011). The increasing temperature of the world is associated with the global warming and the water is one of the scarce and most useful resources which in the result of this increasing temperature are evaporating from the reserves. Rainfall is also compensation to all these reserves and it is necessary for the agriculture and its production as well. The phenomenon of rainfall differs with the difference in latitude and longitude. The rainfall phenomenon also differs with the difference of regions, planes, mountainous and plateaus.

Rainfall occurs as stratiform or convective rain; the high latitude areas experience stratiform rain which is quite a dominant form of the rain. These areas include the tropical and subtropical and they experience 50% to 80% of stratiform rain precipitation (Alpers & Melsheimer, 2004). It is important to measure the distribution of the rainfall on the global level and for that currently the remote satellite sensing techniques **are assisting in measuring the** distribution of the rain on the global level. Special Sensor Microwave Imager (SSM/I) onboard with the US Defense Metrological Satellite Program (DMSP) are used for gathering the information about the rainfall with other space-borne instruments like microwave instrument , flying aboard the US –Japanese Tropical Rainfall Measuring Mission (TRMM) and precipitation radar (PR) that operate on different frequencies and are assisting in the data collection and in getting the footprints accurately **.**

However, the precipitation of the rainfall is not constant and it changes every year. The rainfall is actually the evaporated water from the earth surface because of high temperature or heat that goes up and then comes down in the form of rain or snow. The rainfall is the most significant phenomenon has always been associated with the increasing demands of mankind. The human beings on the earth cannot live without water and there is no way that they can also prude it artificially. It is one of those precious resources that cannot be artificially produced and thus it is the focus of the most studies and researches going on in the world. the scientists and the engineers are collaborating researchers to find out the best and most effective way of measuring the rainfall and predicting the rainfall to compensate for the extreme water use around the globe and to be sufficient for the increasing demands in terms of agriculture, water reserves and in order to be safe and sound from the natural disasters like the flood and land sliding. There is a need to focus on the efficient use of the water and to make the accurate predictions about the rainfall so that the time and the resource could be saved.

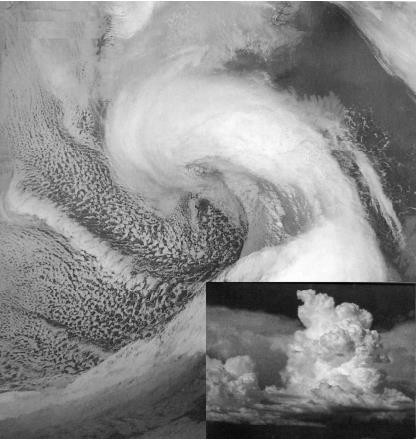
Similarly, the frequency, intensity and the amount of precipitation are changing with the changing temperatures and the effect of heat on the environment is also causing the changes in the precipitation levels (Kumar, Yang, Goddard, & Schubert, 2004). The rainfall can vary from the tropical storms to a thunderstorm, orographic rainfall and cyclones. The changing precipitation levels are observed by the Global Precipitation Climatology Project (GPCP) and presented the global changes in precipitation by the changing lands and the time period to impact it (Gu, Adler, Huffman, & Curtis, 2007). The rainfall affects the surface gravity waves in the upper water layer by generating turbulence and enhancing the roughness on the sea surface (Nystuen, 1990). The rainfall causes the notable changes to the environment and does not only assist in the everyday demands but also in cleaning the environment for the human beings. The air after the rainfall is fresh and clean and the pollution caused by the human beings is also controlled by the rainfall. The distribution and the amount of rainfall differ from region to region and area to area; therefore, with some precipitation in the different area it can also cause floods, tsunamis and land sliding. Therefore; the prediction of the rainfall and the forecasting of the precipitation is quite a significant field on which the scientists and the researchers are exploring ways to accommodate this natural phenomenon and to manage the resource in a more appropriate and useful manner to multiply the human life.

# 3.2 Types of rain

# 3.2.1 Conventional precipitation

The significant and the most dominant form of the rainfall is the convectional rainfall. It is experienced in the high latitude areas like the tropical and the subtropical. It is usually observed with lightning and thunderstorm. The conventional rainfall is a type of rainfall that is affected by the mountains and the mountainous regions; as it is the most dominant form of rainfall and it depends on the latitude. The formation of the conventional rainfall occurs when the air on the surface of the earth gets intense by the heat of the sun. The hot air is lighter than the cool air so it evaporates from the earth surface and forms clouds in the atmosphere. The further rise in the water vapours and gradually these vapours move upward direction towards the area of converging air and forms thick and heavy clouds. The heavy and unstable clouds rise further and the instability of these clouds then compel them to fall on the surface of the earth again in the form of conventional rain .

The conventional form of rainfall is observed mostly in West Africa and it is always followed by a thunderstorm and heavy lightening because of the heavy and unstable clouds rising upward in the atmosphere with converging air.The figure below shows the instability of the heavy clouds of convectional rainfall.



**Figure 3.1**: Heavy and unstable clouds of conventional rainfall (Collier, 2003).

**3.2.2 Orographic Rainfall**

The orographic rain is the form of rainfall that is formed by the moist air which usually can be observed above the mountains. The moist air above the mountains is evaporated or lifted upward direction. When the moist air is lifted and rises to a certain level it cools down; the orographic clouds are formed and then condenses and forms the precipitation. The orographic rainfall is formed by the midlatitude lands like the one with large mountain. The orographic rainfall has tiny water drops that are condensed. These small water drops from clouds and then these small clouds come together to form bigger clouds. These clouds also turn into snow over some period of time.

The orographic rainfall is observed on the midlatitude mountains with an axis perpendicular to the prevalent wind direction. These directions cause the sharp rainfall transitions and could be observed better with two adjacent ranges of the mountain to circulate the moist air more. The steadier and these are experienced mainly in afternoon of the summers with dynamic thunderstorms. The discrete formation of orographic precipitation is sometimes observed on the small mountains as well (Roe, 2005). Orographic rainfall is due to the uplift of masses of air by the wind.

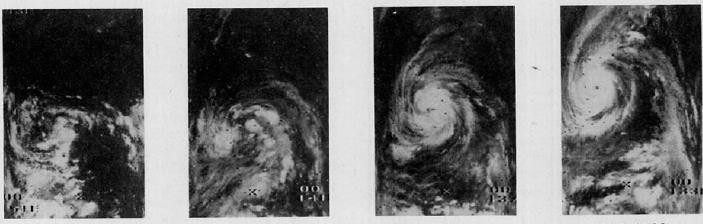
**Figure 3.2**: Wave cloud formation on Amsterdam Island in the far Southern Indian Ocean (NASA, 2005).

**3.2.3 Cyclonic or frontal rainfall**

The cyclonic or the frontal rainfall is the last and third type of the rainfall. The cyclonic by name represents the tempesting and occurs when the air masses with distinct characteristics collide with one another. The collision of light air that is warm and the cold air that is heavy occurs; the cold air encourages the warm air because it is lighter to rise. The rising air cools down by forming the water vapours. The condensation process initiates and forms the clouds (Thatcher, Takayabu, Yokoyama, & Pu, 2012).

The formation of clouds become heavy as they meet with other clouds and these heavy clouds become unstable and fall back on the earth as cyclonic rainfall. The cyclonic rainfall is common in Tropical areas with 23% of Degrees North Latitude and South of the equator with the temperate zone latitude of 66% degrees North and South. This is the reason that it is also known as the frontal rainfall (Thatcher, Takayabu, Yokoyama, & Pu, 2012).

The frontal/cyclonic rainfall has a specific period and time when it is more dominant and the precipitation is more rapid and concentrated. The cyclonic/frontal rain may have an extended precipitation that could be extended and keep the weather wet for log days. It is a lognormal type of distribution and its distribution depends on the area of precipitation (Cheng & Qi, 2001).



**Figure 3.3**: Cyclonic rainfall cloud formation (Rodgers & Adler, 1980)

**3.3 Measurement of Rainfall**

The rainfall is a natural phenomenon that is measured in mm. The measuring instrument is 203mm in diameter. This is a funnel that gathers the rainfall into a cylinder and has the capacity of measuring u to 25mm of rainfall (Alpers & Melsheimer, 2004). There are two techniques for measuring rainfall as described below:

# 3.3.1 Ordinary Rain Gauge

The ordinary rain gauge measurement is a less effective and less accurate technique of measuring the rainfall. It has been observed that the ordinary gauge is the non-automatic observation and uses a glass to measure the rain at regular intervals. It has a shell, a storage bottle with a storage vessel and a glass for measuring the rain. It is not effective for the heavier and substantial rainfall like the cyclonic/frontal rainfall (Carvalho, Assad, Oliveira, & Pinto, 2014). It uses a rainfall record book to compare and measure the rainfall for a particular period. It is less accurate and the data collected may not be precise (Agnihotri & Panda, 2014). The ordinary method of the rainfall measurement is helping in the local level measurements and those that are less accurate and less precise but this method of rainfall collection is appropriate for the measurement t record for a larger level. The observations performed in the ordinary rainfall are manual so the errors are not minimized.

# **3.3.2** Self-recording rain Gauge

The conventional self-recording rain gauge is more efficient and more effective in measuring the rain than that of the ordinary gauge for the measurement of rain. The traditional method of recording rain is inefficient and gives inaccurate results as it’s done manually so it can involve human errors (Beard, 1962). The self-recording rain gauge is observed to use simple technique and instruments to produce better measurements in order to have better probability and accuracy. The self-recording gauge consists of a tipping bucket and a lever balance that weighs the rain accordingly. It has a marking pen which marls each certain level recorded with the movement of the bucket and the precipitation for each hour is recorded by the gauge automatically without the support of human beings.

**3.3.3 Zonal distribution of Rain**

The patterns of the rainfall precipitation are not constant; it varies season to season and location to location. There are certain different zones that get more precipitation than few of them getting less precipitation. The precipitation of the rainfall as a mean global distribution studied to be affected by the latitudinal zones, land and sea surfaces and precipitation. The East Asian region precipitation including China, Korea and Japan evaluates that monsoon starts from mid-end of May to the end of July for China and September in case if Korea. It ends by the early August Peninsula and for Japan; it lasts longer from mid-September to end of October.

The Middle Eastern region experiences a severe autumn rainstorm. The countries along the red sea are the most significant to experience this kind of rainfall distribution and the Mediterranean countries are more unstable to these conditions. The experience intense thunderstorms but the weather for summer is more of hot and dry. These countries also experience hailing and may sometime encounter flooding. The North African region is also included within this area and experience the same weather. Both the weathers are extreme; in case of summers and winter being on the cold front.

The rainfall distribution in the region of East Africa is also studied as significantly more than that of the amount of rainfall studied. The patterns of zonal distribution are observed to be variable for each season. The rainfall prediction for each day is also changing with the changing trends of rainfall. The zonal distribution of the rainfall is open to seasonal change and the change in the sea level and earth surface due to the movement of plates (Johnson, 1962). The parameter like wind, atmosphere and humidity are also significant in determining these patterns. Air pressure and temperature also affect the zonal distribution of the rainfall annually and monthly. The changing temperature of the earth is also influencing the zonal distribution because it has more chances of being deteriorated. The high precipitation of the rainfall distribution recorded is that of an equatorial zone. Southeast Asia and the middle latitude areas experience the comparatively low amount of rain distribution and the deserts of subtropical regions experience even f or extremely insufficient rainfall annually .

# According to Haurwitz and Austin in 1944; there are six main regimes of the rainfall that are described as:

# 3.4.1Equatorial rainfall regime

It is the form of rain that is characterized by experiencing the rainfall throughout the year and in all seasons. This character of rainfall is not outlined by the rainfall in a particular season or in few specific months rather it is characterized by the consistent and continuous rainfall annually. The equatorial rainfall regime is experienced to have heavy rainfall in the month of March and September. The thermal air currents are generated by the heating effect and that contributes to the formation of heavy and unstable clouds resulting in the extreme rainfall within these months of the year. The equatorial regime of the rainfall extends between the zone of 10⁰ N and 10⁰ S latitude. The equatorial rainfall regime is followed by thunderstorm and heavy lighting because of the instability and huge size of the clouds. The rainfall is observed to be in the form of heavy showers but the time duration observed for these types of regimes is quite less as compared to the other regimes. These are the low-pressure belts and relatively high temperatures. Mostly they experience the conventional rainfall because of the heating effect. The distribution of the rainfall throughout the year is equal and uniform. These regions experience heavy rainfall sometimes with hailing and storms during the year.

# 3.4.2Tropical rainfall regime

The tropical rainfall regime is characterized by the heavy rainfall not throughout the year but only in the summers. The winters of the tropical areas are usually dry and they are not associated with much rainfall. The summer season not only experiences the heavy rainfall but is comparatively pleasant because of the consistent rainfall. The northern hemisphere and the Eastern hemisphere usually experience the maximum and minimum rainfall during the months of July and December. The tropical regime of the rainfall is under the influence of summer stagnation and thus it can be said that the winters are significantly dry. Therefore, the areas occupying the tropical regime may suffer in the winters but are ready to shine in the summers**.**

# 3.4.3 Monsoon rainfall regime

The monsoon rainfall regime is characterized as the zone 4 that is also under the stagnation of the summer season. The rainfall experienced in this regime is characterized as only in summers but slightly less as compared to that of tropical. The winters are also extremely dry as that of the tropical because this regime is at the sub-tropical high-pressure belt. The maximum rainfall is experienced in the month of July that is known to be the monsoon season and continues to the end of August. There are not much heavy showers of the rainfall and the thunderstorm and the lightning is also merely observed but the light shower continues for almost 2 months day to day and time to time. This regime is associated with the monsoon winds; as the monsoon winds start to blow; like the phenomenon experienced on the orographic rainfall; the wind elevates the water vapours and condenses them to form clouds and then into rainfall shower.The winds observed to move the water vapours in the upward direction are also associated with the formation of the cyclonic rainfall as described above. The air of different masses collides with one another and initiates the condensation process. The lighter air that is warm air collides with the heavy cold air to form the clouds and thunderstorms or lightening effects could be observed for the monsoon regime of the rainfall.

**3.4.4 Mediterranean rainfall regime**

The Mediterranean rainfall regime experiences the driest weather. For example: the weather conditions of Nicosia and Famagusta. The weather of these cities are extremely dry not only in winters but also in summers. The regime is associated with the sub-tropical high-pressure belt for the whole year. Although, the winters are cold and rainy they last for only 2 months that of January and February and the summer season is extremely hot and dry. They experience the comparatively heavy rainfall in the winter season opposite to the above-mentioned regime sand but dryness prevails throughout. The summers despite being very hot for the Mediterranean are still dry and hardly the rainfall is experienced. The city of Nicosia and Famagusta are also following this belt ad experience exactly the same weather as observed from the data collectedThe Mediterranean rainfall regime experiences the driest weather. For example: the weather conditions of Nicosia and Famagusta. The weather of these cities are extremely dry not only in winters but also in summers. The regime is associated with the sub-tropical high-pressure belt for the whole year. Although, the winters are cold and rainy they last for only 2 months that of January and February and the summer season is extremely hot and dry. They experience the comparatively heavy rainfall in the winter season opposite to the above-mentioned regime sand but dryness prevails throughout. The summers despite being very hot for the Mediterranean are still dry and hardly the rainfall is experienced. The city of Nicosia and Famagusta are also following this belt ad experience exactly the same weather as observed from the data collected from the metrological department. The Mediterranean rainfall regime is defined as zone 4 by Haurwitz and Austin, as extremely dry .

# 3.4.5 Continental rainfall regime

Similarly, the continental rainfall regime experiences the heavy rainfall in summers due to the convection. The high temperature forces the water to evaporate in the form of water vapors from the water surface and form clouds. The summer season continuously experiences the convection and the rainfall is experienced throughout the summer season opposite to that of the winter season formation of the clouds. Because the summer season experiences the heavy rainfall therefore, the winter season is usually dry it experiences only few slight showers of the precipitation. The weather conditions for such a regime is neither very dry nor are they rainy. The weather is moderate and the extreme seasons are not experienced. The high precipitation within the continental regime could experience the cyclonic rainfall and conventional rainfall as the precipitation depends upon the temperature. The high temperature will impact the precipitation and the formation of the clouds for heavy showers of the rainfall.

**3.4.6 Maritime rainfall regime**

The maritime regime is characterized by the mid-latitudes and along the western margins of the continent. The maritime regimes experience the rainfall in winter season and the summer is not very dry but due to the high precipitation in the winter season; the maritime regimes experience usually the dry summers. These regimes also experience the cyclonic rainfalls with several intervals and the process of rainfall continues throughout the year. There is no season in these regimes that is extremely dry or experiences no rainfall. These regimes also experience the winter monsoons despite of the summer monsoon winds and are expected to observe with the monsoon showers in the winters rather than that of the summers. The maritime air is conveyed to the cost through the dominant westerly.

**CHAPTER 4**

# **MACHINE LEARNING TECHNOLOGIES**

# **4.1** Introduction to machine learning

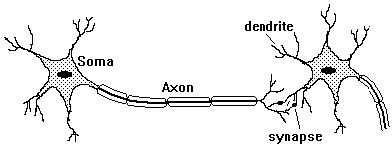
Machine learning is associated with the study of the algorithms that enhance the efficiency of the machines/computers automatically through the training and testing of the machine/computers with certainly different variables. The machine learning is among the most favourable and fastest growing areas of computer technology. The computers work efficiently with different algorithms and functions. The machine learning is the training the computer with certainly different algorithms to experience the machine in automatic smart data processing. The machine learning enhances the efficiency and accuracy of the data processing and is used in a wide range of fields. The machine learning is developed with effectual algorithms that utilize a certain set of tools and functions to solve the complex and huge data. The machine learning is assisting in the diverse field; these are normally artificial intelligence applications that are used for recognition and prediction like that of computer engineering and medical fields. The machine learning is popular among the modern computer technology and has many benefits. Machine learning develops some rules for the input data as discussed in the hybrid model that helps the machine to process the similar case each tie efficiently. It works on the prediction, and it is more important to understand that how variables the inputs into are moved into vectors. The machine learning has minimized the manual jobs for the people that also could have the space for errors and in accuracy.

**4.1.1.0 Artificial neural networks**

The artificial neural network is a computer networking system that can perform huge and intelligent tasks. It is a parallel and distributed processing system that can accomplish most complex tasks of recognition, prediction and detection without increasing the complexity of the problem. The artificial neural network has one input layer and one output layer between which are the hidden layers that process data. Each layer by processing the data forwards the result to the next hidden layer and finally the output layer obtains the result after the data processing. The artificial neural network is one the most popular machines of artificial intelligence that are used almost in every field nowadays. It uses certain different models for processing data like feed- forward back propagation, NARX model with different functions for each model. These are dynamic machines capable of solving complex to everyday problems and made the human life easy .

# 4.1.1.2 Neurons

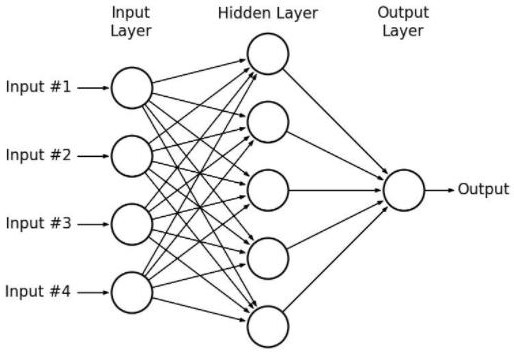
The artificial neural network performs these complex and huge tasks through the neurons that are fed into the layers and process the data just like the human brain. Neurons operate in the human brain to perform all the tasks and so are the artificial neural network does. The neural network is non-linear functions. The neurons in the neural network are first trained with the old data in order to get the new and predicted data.

**Figure 4.1**: Neuron scheme

After the training, testing is carried out to check the different results with different data and to obtain the comparison by feeding the system with a different number of neurons. The number of neurons fed to the system is varying and depends on the data and processing complexity. Therefore, the architectures may differ from one another depending on input/output complexity and the layers in the system.

# 4.1.1.3 Structure of ANN

The architecture of the ANN has three layers with a large number of neurons, the neurons are called the units and they are arranged in a sequence of layers.



**Figure 4.2**: Structure of ANN

a)INPUT LAYER is the first layer of the ANN structure is the input layer that takes the input for processing.

b)HIDDEN LAYER is the second layer that process the data transferred from the input layer for processing through neurons and, the weights are updated continuously for precision and validity of the output

c)OUTPUT LAYER is the third and the last layer through which the results are obtained from the, as shown in figure 4.2 above.

# 4.1.1.4 Weights

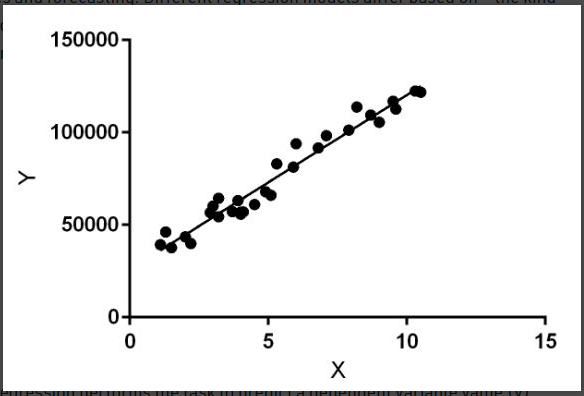
The weights in the ANN architecture are the memory storage that stores the information and data to get the desired results. The weight during the training, testing and validation are modified at every step so that the accuracy of the output is achieved; also, they store data for the future operations.

# 4.1.1.5 Feedforward neural network

The feed-forward neural network has multilayers for the processing of elements. Each layer processes the input data that it receives and forward the results obtained to the next layer. For this processing, each layer operates independently to generate the resulting that is forwarded to the next layer. The result obtained through processing of each layer is ultimately obtained from the output layer. Between input and output layer; there are hidden layers. The elements that process the input data work like the neurons in the human brain, these are called artificial neurons. The neurons in the layers send messages or information to other neurons through a channel called connections.

**4.2 Linear Regression**

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables getting used.



**Figure 4.3**  Linear Regression example

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.  
In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for our model.

4.2.1 **Hypothesis function for Linear Regression :**



**Figure 4.4** Hypothesis of Linear Regression

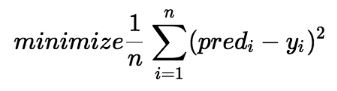
While training the model we are given :  
**x:** input training data (univariate – one input variable(parameter))  
**y:** labels to data (supervised learning)

When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best θ1 and θ2 values.  
θ1: intercept  
θ2: coefficient of x

Once we find the best θ1 and θ2 values, we get the best fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x.

How to update θ1 and θ2 values to get the best fit line ?

Cost Function (J):  
By achieving the best-fit regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum. So, it is very important to update the θ1 and θ2 values, to reach the best value that minimize the error between predicted y value (pred) and true y value (y).





**Figure 4.5** Cost Function J

Cost function(J) of Linear Regression is the **Root Mean Squared Error (RMSE)** between predicted y value (pred) and true y value (y).

**4.3 LASSO REGRESSOR**

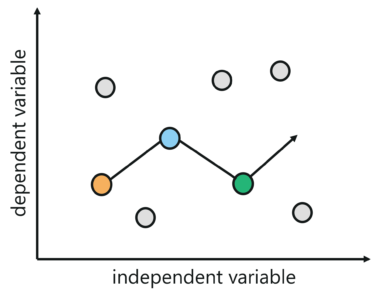
The word “LASSO” stands for ****L****east ****A****bsolute ****S****hrinkage and ****S****election ****O****perator. It is a statistical formula for the regularisation of data models and feature selection.

Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

Lasso Regression uses L1 regularization technique (will be discussed later in this article). It is used when we have more number of features because it automatically performs feature selection.

## 4.3.1 ****Lasso Regularization Techniques****

There are two main regularization techniques, namely Ridge Regression and Lasso Regression. They both differ in the way they assign a penalty to the coefficients. In this blog, we will try to understand more about Lasso Regularization technique.

- 

**Figure 4.6** Lasso Regression

## 4.3.2 ****Regularization****

Regularization is an important concept that is used to avoid overfitting of the data, especially when the trained and test data are much varying.

Regularization is implemented by adding a “penalty” term to the best fit derived from the trained data, to achieve a lesser variance with the tested data and also restricts the influence of predictor variables over the output variable by compressing their coefficients.

In regularization, what we do is normally we keep the same number of features but reduce the magnitude of the coefficients. We can reduce the magnitude of the coefficients by using different types of regression techniques which uses regularization to overcome this problem.

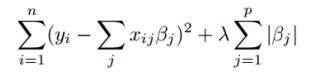
## ****4.3.3 L1 Regularization****

If a regression model uses the L1 Regularization technique, then it is called Lasso Regression. If it used the L2 regularization technique, it’s called Ridge Regression. We will study more about these in the later sections.

L1 regularization adds a penalty that is equal to theabsolute value of the magnitude of the coefficient. This regularization type can result in sparse models with few coefficients. Some coefficients might become zero and get eliminated from the model. Larger penalties result in coefficient values that are closer to zero (ideal for producing simpler models). On the other hand, L2 regularization does not result in any elimination of sparse models or coefficients. Thus, Lasso Regression is easier to interpret as compared to the Ridge.

## 4.3.4 ****Mathematical equation of Lasso Regression****

****Residual Sum of Squares + λ \* (Sum of the absolute value of the magnitude of coefficients)****



**Figure 4.7** Equation of Lasso Regression

Where,

1. λ denotes the amount of shrinkage.

b)λ = 0 implies all features are considered and it is equivalent to the linear regression where only the residual sum of squares is considered to build a predictive model

c)λ = ∞ implies no feature is considered i.e, as λ closes to infinity it eliminates more and more features

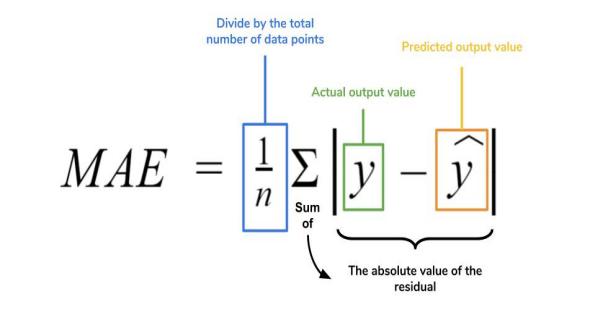
d)The bias increases with increase in λ

e)variance increases with decrease in λ

**4.4 Mean Absolute Error(MAE)**

We know that an error basically is the absolute difference between the actual or true values and the values that are predicted. Absolute difference means that if the result has a negative sign, it is ignored.

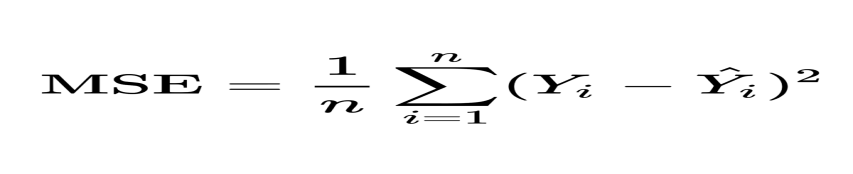
Hence, ****MAE = True values – Predicted values****



**Figure 4.8** MAE

**4.5 Mean Square Error(MSE)**

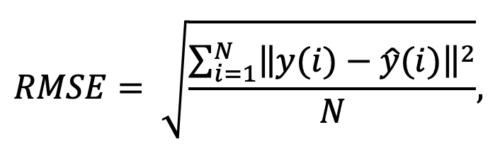
MSE is calculated by taking the average of the square of the difference between the original and predicted values of the data.



**Figure 4.9** MSE

**4.6 Root Mean Square Eroor(RMSE)**

RMSE is the standard deviation of the errors which occur when a prediction is made on a dataset. This is the same as MSE (Mean Squared Error) but the root of the value is considered while determining the accuracy of the model.



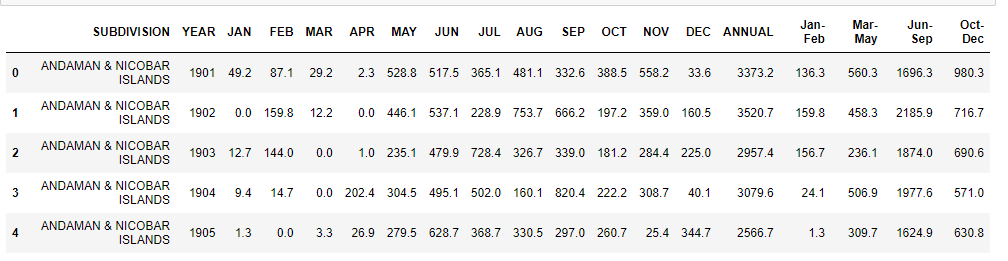
**4.10 Figure** RMSE

**CHAPTER 5**

**SIMULATION**

**5.1 Data processing**

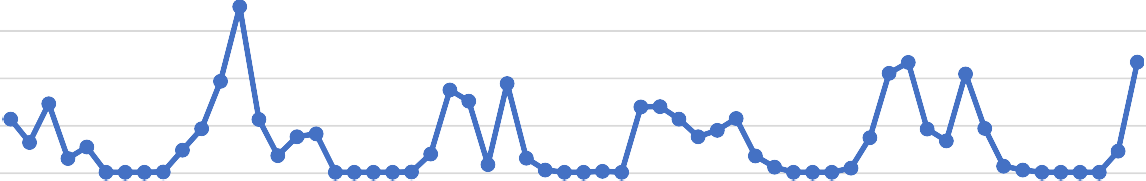
In this study, we are considering the prediction of rainfall. The main purpose and aim of the study are to utilize the machine learning for predicting the rainfall with accuracy and precision. Therefore, the study includes and examines the source data in the pre-processing phase and utilizes this data in the stimulation or processing phase further to predict the output that is forecasting effectively and efficiently. For this purpose, statistical data of INDIA is collected from the metrological departments. The data collected from the metrological departments is monthly, as this study will also focus on the monthly rainfall prediction. The data includes the significant parameters of the rainfall. Temperature, wind direction, wind speed, air pressure and humidity are considered. Each parameter of the rainfall is analyzed separately for INDIA. The statistical data studied and examined in this study will be used as input for processing the output as rainfall prediction. The data processing is done through machine learning techniques and different neural networks like ANN and Back-Propagation models.



**Figure 5.1** Data Processing

**5.2 Data PreProcessing**

The monthly rainfall data from for INDIA is collected from the metrological department to outline the trends and variations that are being observed within this time period and to highlight the rain cycles for each that will effective for prediction.



Rainfull

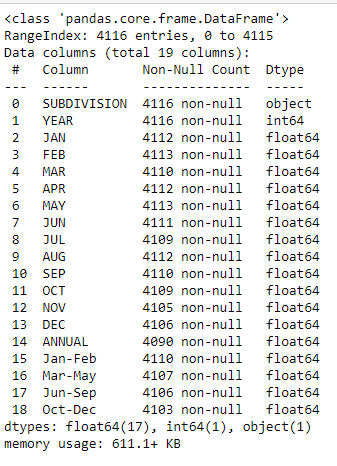
200

150

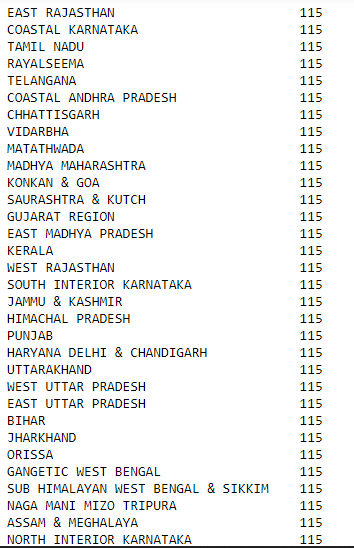
100

50

0



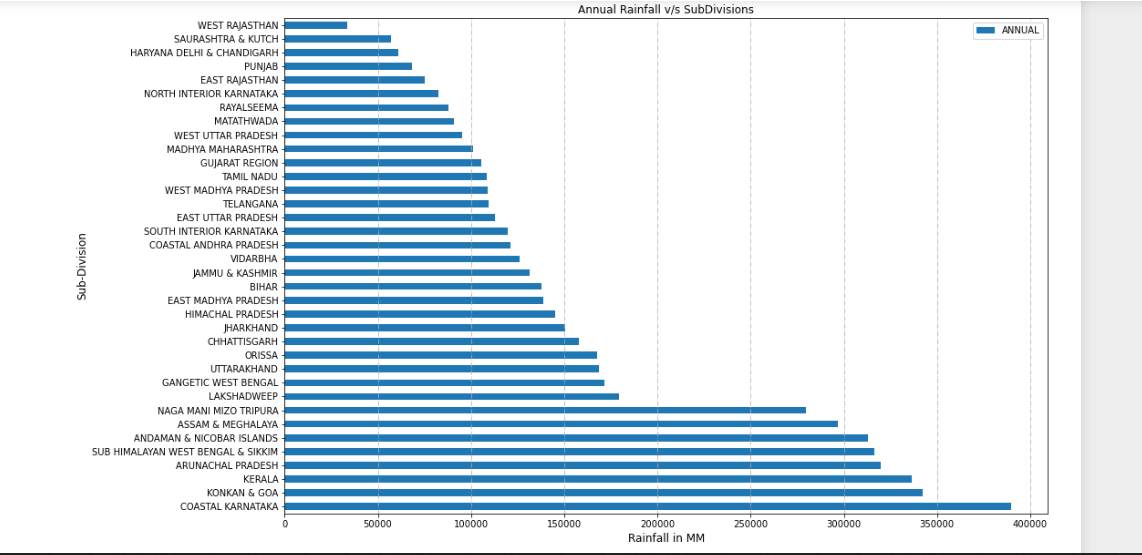
**Figure 5.2** Data Preprocessing



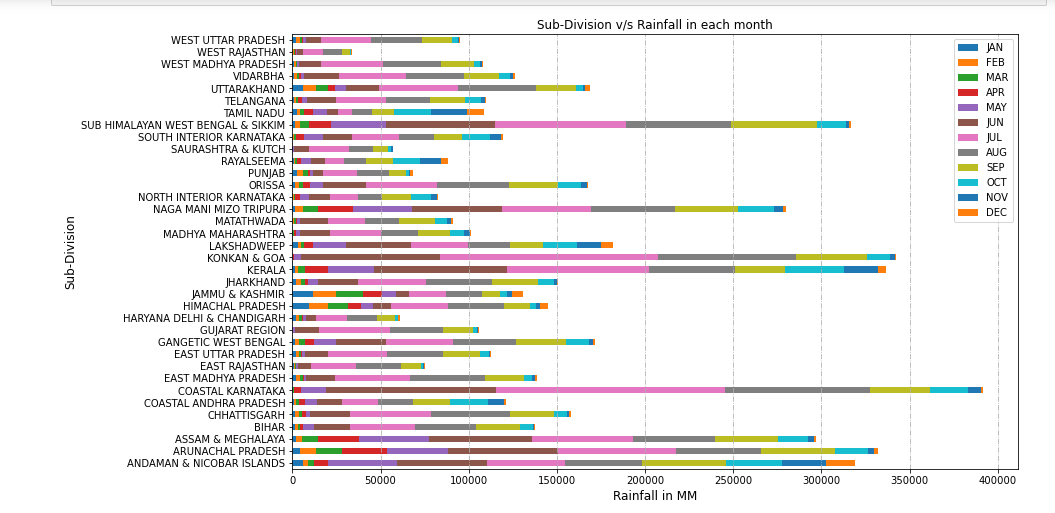
**Figure 5.3 Data set Of India State Rainfall**

**5.3 Data Visualization**

Data visualization is main concept for every project. We can easily understant the data in visualize part.



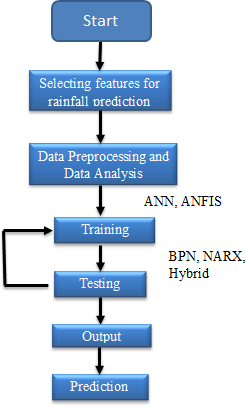
**Figure 5.4** Data Visualization-i



**Figure 5.5** Data visualization-ii

**5.4 Flowchart for rainfall prediction**

The below data flowchart outlines the selection of input and the procedure to get the output that in this study is the rainfall prediction. The flowchart illustrates the first step as a selection of inputs that are parameters, processing of these inputs and completing training, testing and validation for accurate and precise output that is rain forecast.



**Figure 5.5** Flow Chart

# 5.5Selection of the input and output data

In this study, the database is collected directly from the metrological departments to analyze the inputs and outputs. The significant inputs are highlighted as average temperature, air pressure, wind speed, wind direction and humidity and the output are rainfall for the years 1901-2015. The rainfall season for India is observed from November up to May; 7 months of rainfall in a year. The rainfall season for Nicosia as highlighted by the data starts from September and ends in May; that is around 9 months according to the data. For Famagusta, the rainfall season is similar to that of Nicosia. Therefore, to predict rainfall for these three cities above mentioned parameters will be considered as significant inputs to forecast precipitation.

# **5.6** Feature Extraction

This study is focusing on average temperature, air pressure, wind speed, wind direction and humidity. The parameters are defined as:

* TEMPERATURE

It is the degree to which something is hot or cold. It is affected by several factors like latitude, altitude, and distance from the sea, wind with ocean current. It affects the rainfall in a way that the amount of water vapour in the atmosphere determines the amount of rainfall as light or with a heavy shower. Thus, at high temperatures there is heavy rainfall.

* WIND DIRECTION

It is the origin of the wind from where it generates. The direction of the wind determines the precipitation. Therefore, it is also important to analyze the direction of the wind as a parameter for the rainfall prediction.

* WIND SPEED

It is the velocity of the air moving from high pressure to low pressure. The wind speed is in close correlation to the rainfall and it is necessary to consider it as an input for prediction. The increase in the wind speed decreases the intensity of the rainfall. Therefore, to experience more precipitation during the rainy season; the favourable wind speed is low rather than high.

* HUMIDITY

It is the amount of water vapour present in the air for a given temperature that is significant for saturation of vapours to form clouds. The increased moisture in the air will accommodate high saturation. The humid weathers experience more precipitation than the dry weather.

* AIR PRESSURE

The air pressure is the pressure within the atmosphere of the earth. The air pressure decreases with the increase in the elevation from the earth surface. For example: the air pressure on mountains is less as compared to plains.

**5.7 Training, Testing and Validation**

The creation of a database from inputs and outputs; the succeeding phase is the training the input through back propagation algorithm,LASSO and hybrid models. The training of the input data is done by the MATLAB software by using the NNTOOL and ANFISTOOL. Training is preceding step to testing of the input data with acceptable error is achieved. The NNTOOL automatically retains 30% of the input data for testing and validation and 70% of input data for the training. The ANFISTOOL requires manual retention of the input data for training, testing and validation. The ANFISTOOL manually fed with 60% of the input for training and 40% for testing.

**5.8 ANN**

The artificial neural networks are multipurpose and feasible set of the system. These are basically the computational methodologies that are used in diverse fields. The artificial neural networks are based on elements called neurons and these neurons receive inputs for processing the required results. The neural networks are multilayered nodes that are nonlinear and connected by the weighted lines. When the input is presented to the neural networks during the training of the data; the successful training then offers the neural networks that are capable of predicting the output. The neural networks are capable of classifying an object, recognizing a particular pattern of multifactorial data, approaching a function and even configuration . This study will focus on the learning algorithms like the feed- forward backpropagation and NARX model to predict the rainfall using the inputs described above.

**CHAPTER 6**

**DISCUSSIONS AND RESULTS**

The study focused on using the artificial neural networks for predicting the rainfall on monthly basis. The study proved that using back propagation and LASSO with hybrid models can be used to obtain accurate and precise forecast readings for rainfall, as demonstrated in this study. Although, the networks were trained before testing and validation; the foremost challenge of the study was to achieve maximum accuracy and efficiency in making the prediction.

Thus, the study trained and tested the monthly data for last 6 years of three different locations from two different regions. For this purpose, the study focused on neural networks collaborated with LINEAR REGRESSION. The study tested the rainfall data from 1901-2015.

# **6.1** Modelling

Modelling of rainfall dataset

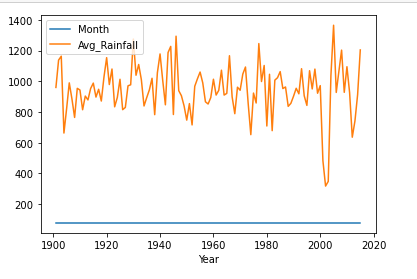


Figure 6.1 Modelling

**6.2 Comparing Results**

* THE MAE,MSE,RMSE,Training Accuracy and Testing Accuracy values for Rainfall Prediction Using **LINEAR REGRESSION**

**TEST DATA**

MAE: 36.693305772295616

MSE: 2707.377549592384

RMSE: 52.032466303187896

**TRAIN DATA**

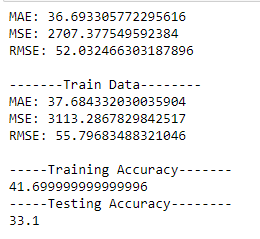
MAE: 37.684332030035904

MSE: 3113.2867829842517

RMSE: 55.79683488321046

TRAINING ACCURACY=41.69999992

TESTING ACCURACY=31.500



**Figure 6.2** Result of Linear Regression

* THE MAE,MSE,RMSE,Training Accuracy and Testing Accuracy values for Rainfall Prediction Using **LASSO REGRESSION**

**TEST DATA**

MAE: 41.77463317550605

MSE: 3011.4820490350985

RMSE: 54.87693755408

**TRAIN DATA**

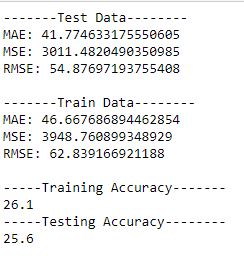
MAE: 46.667686894462854

MSE: 3948.760899348929

RMSE: 62.839166921188

TRAINING ACCURACY=26.1

TESTING ACCURACY=25.67



**Figure 6.3** Result of LASSO Regression

* THE MAE,MSE,RMSE,Training Accuracy and Testing Accuracy values for Rainfall Prediction Using **Ridge Model**

**TEST DATA**

MAE: 36.694264997117806

MSE: 2700.404122847211

RMSE: 51.96541275547815

**TRAIN DATA**

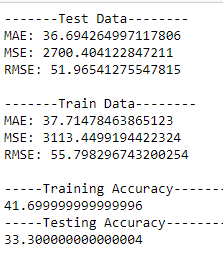
MAE: 37.71478463865123

MSE: 3113.4499194422324

RMSE: 55.798296743200254

TRAINING ACCURACY=41.699999999999996

TESTING ACCURACY=33.300000000000004



**Figure 6.3** Result of Ridge Model

* THE MAE,MSE,RMSE,Training Accuracy and Testing Accuracy values for Rainfall Prediction Using **Random Forest Model**

**TEST DATA**

MAE: 76.73671497584542

MSE: 9936.306763285023

RMSE: 99.68102509146374

**TRAIN DATA**

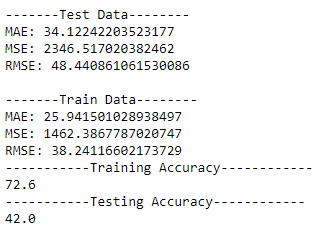
MAE: 25.941501028938497

MSE: 1462.3867787020747

RMSE: 38.24116

TRAINING ACCURACY=72.6

TESTING ACCURACY=42.0



**Figure 6.3** Result of Random Forest Model

**CHAPTER 7**

# ANALYSIS

# 7.1Introduction

The Systems Development Life Cycle (SDLC), or Software Development Life Cycle in [systems](http://en.wikipedia.org/wiki/Systems_engineering) [engineering,](http://en.wikipedia.org/wiki/Systems_engineering) [information systems](http://en.wikipedia.org/wiki/Information_systems) and [software engineering](http://en.wikipedia.org/wiki/Software_engineering), is the process of creating or altering systems, and the models and [methodologies](http://en.wikipedia.org/wiki/Methodologies) that people use to develop these systems. In software engineering theSDLC concept underpins many kinds of [software development methodologies](http://en.wikipedia.org/wiki/Software_development_methodologies). These methodologies form the framework for planning and controlling the creation of an information system the [software](http://en.wikipedia.org/wiki/Software_development_process) [development process.](http://en.wikipedia.org/wiki/Software_development_process)

### 7.2 Existing System

IIoT systems are vulnerable to single point of failure and malicious attacks, which cannot provide stable services. Due to the resilience and security promise of blockchain, the idea of combining blockchain and IoT gains considerable interest. However, blockchains are power-intensive and low-throughput, which are not suitable for power-constrained IoT devices. To tackle these challenges, we present a block chain system with credit-based consensus mechanism for IIoT.

### 7.2.1Disadvantages of Existing System:

1. These cannot provide stable services.

**7.3 Proposed System**

In this paper we propose a credit-based proof-of-work (PoW) mechanism for IoT devices, which can guarantee system security and transaction efficiency simultaneously. In order to protect sensitive data confidentiality, we design a data authority management method to regulate the access to sensor data. In addition, our system is built based on directed acyclic graph (DAG)-structured block chains, which is more efficient than the satoshi-style block chain in performance. We implement the system on Raspberry Pi, and conduct a case study for the smart factory. Extensive evaluation and analysis results demonstrate that credit-based PoW mechanism and data access control are secure and efficient in IIoT.

### 7.3.1 Advantages of Proposed System:

a) This protects the sensitive data confidentiality.

b)It provides system security and transaction efficiency.

**SDLC (Umbrella Model):**



DOCUMENT CONTROL

**Umbrella**

**Umbrella**

Business Requirement

Documentation

•

•

•

Requirements

Gathering

Feasibility Study TEAM FORMATION

Project Specification PREPARATION

ANALYSIS &

DESIGN

CODE

UNIT TEST

ASSESSMENT

INTEGRATION & SYSTEM

TESTING

DELIVERY/INS

TALLATION

ACCEPTANCE

TEST

TRAINING

**Umbrella**

**Figure 7.1** SDLC

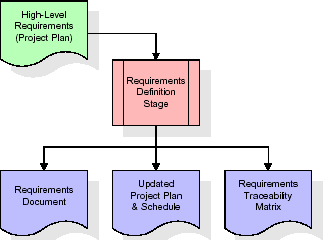
SDLC is nothing but Software Development Life Cycle. It is a standard which is used by software industry to develop good software.

**Stages in SDLC:**

* + Requirement Gathering
  + Analysis
  + Designing
  + Coding
  + Testing
  + Maintenance

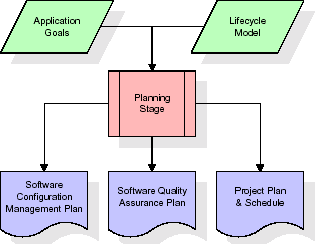
**A)Requirements Gathering stage:**

The requirements gathering process takes as its input the goals identified in the high-level requirements section of the project plan. Each goal will be refined into a set of one or more requirements. These requirements define the major functions of the intended application, define operational data areas and reference data areas, and define the initial data entities. Major functions include critical processes to be managed, as well as mission critical inputs, outputs and reports. A user class hierarchy is developed and associated with these major functions, data areas, and data entities. Each of these definitions is termed a Requirement. Requirements are identified by unique requirement identifiers and, at minimum, contain a requirement title and textual description.

**Figure 7.2** Requirements Gathering

**B)Analysis Stage**

The planning stage establishes a bird's eye view of the intended software product, and uses this to establish the basic project structure, evaluate feasibility and risks associated with the project, and describe appropriate management and technical approaches.

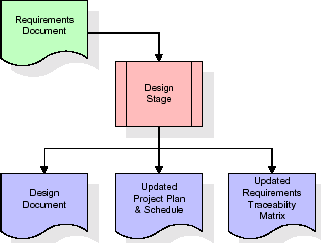


**Figure 7.3** Analysis Stage

The most critical section of the project plan is a listing of high-level product requirements, also referred to as goals. All of the software product requirements to be developed during the requirements definition stage flow from one or more of these goals. The minimum information for each goal consists of a title and textual description, although additional information and references to external documents may be included. The outputs of the project planning stage are the configuration management plan, the quality assurance plan, and the project plan and schedule, with a detailed listing of scheduled activities for the upcoming Requirements stage, and high level estimates of effort for the out stages.

**C)Design Stage**

The design stage takes as its initial input the requirements identified in the approved requirements document. For each requirement, a set of one or more design elements will be produced as a result of interviews, workshops, and/or prototype efforts. Design elements describe the desired software features in detail, and generally include functional hierarchy diagrams, screen layout diagrams, tables of business rules, business process diagrams, pseudo code, and a complete entity-relationship diagram with a full data dictionary. These design elements are intended to describe the software in sufficient detail that skilled programmers may develop the software with minimal additional input.

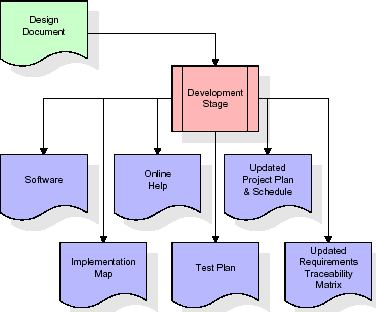


**Fig 7.4**: Designing Stage

When the design document is finalized and accepted, the RTM is updated to show that each design element is formally associated with a specific requirement. The outputs of the design stage are the design document, an updated RTM, and an updated project plan.

**Development Stage**

The development stage takes as its primary input the design elements described in the approved design document. For each design element, a set of one or more software artifacts will be produced. Software artifacts include but are not limited to menus, dialogs, data management forms, data reporting formats, and specialized procedures and functions. Appropriate test cases will be developed for each set of functionally related software artifacts, and an online help system will be developed to guide users in their interactions with the software.

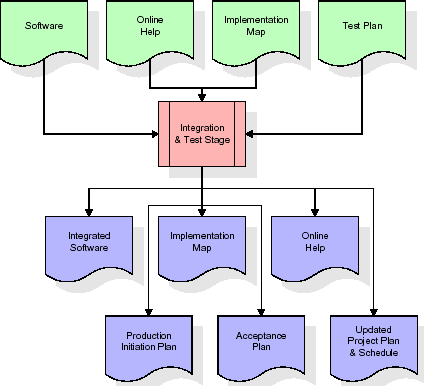


**FIG 7.5** DEVELOPMENT STAGE

The RTM will be updated to show that each developed artifact is linked to a specific design element, and that each developed artifact has one or more corresponding test case items. At this point, the RTM is in its final configuration. The outputs of the development stage include fully functional set of software that satisfies the requirements and design elements previously documented, an online help system that describes the operation of the software, an implementation map that identifies the primary code entry points for all major system functions, a test plan that describes the test cases to be used to validate the correctness and completeness of the software, an updated RTM, and an updated project plan.

**D)Integration and Test Stage**

During the integration and test stage, the software artifacts, online help, and test data are migrated from the development environment to a separate test environment. At this point, all test cases are run to verify the correctness and completeness of the software. Successful execution of the test suite confirms a robust and complete migration capability. During this stage, reference data is finalized for production use and production users are identified and linked to their appropriate roles. The final reference data (or links to reference data source files) and production user list are compiled into the Production Initiation Plan.

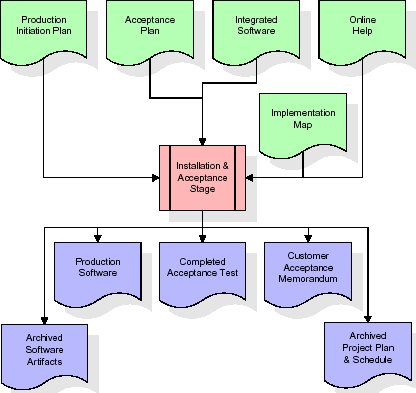


**FIG 7.6** Integration and Test Stage

**E)Installation & Acceptance Test:**

During the installation and acceptance stage, the software artifacts, online help, and initial production data are loa ded onto the production server. At this point, all test cases are run to verify the correctness and completeness of the software. Successful execution of the test suite is a prerequisite to acceptance of the software by the customer.

After customer personnel have verified that the initial production data load is correct and the test suite has been executed with satisfactory results, the customer formally accepts the delivery of the software.



**FIG 7.7** Installation & Acceptance Test

**7.4 Software Requirement Specification**

**7.4.1 Overall Description**

A Software Requirements Specification (SRS) – a [requirements specification](http://en.wikipedia.org/wiki/Requirements_specification) for a [software system](http://en.wikipedia.org/wiki/Software_system) is a complete description of the behavior of a system to be developed. It includes a set of [use cases](http://en.wikipedia.org/wiki/Use_case) that describe all the interactions the users will have with the software. In addition to use cases, the SRS also contains non-functional requirements. [Nonfunctional requirements](http://en.wikipedia.org/wiki/Non-functional_requirements) are requirements which impose constraints on the design or implementation (such as [performance engineering](http://en.wikipedia.org/wiki/Performance_engineering) requirements, [quality](http://en.wikipedia.org/wiki/Quality_%28business%29) standards, or design constraints).

System requirements specification: A structured collection of information that embodies the requirements of a system. A [business analyst](http://en.wikipedia.org/wiki/Business_analyst), sometimes titled [system analyst](http://en.wikipedia.org/wiki/System_analyst), is responsible for analyzing the business needs of their clients and stakeholders to help identify business problems and propose solutions. Within the [systems development lifecycle](http://en.wikipedia.org/wiki/Systems_development_life_cycle) domain, the BA typically performs a liaison function between the business side of an enterprise and the information technology department or external service providers. Projects are subject to three sorts of requirements:

a)[Business requirements](http://en.wikipedia.org/wiki/Business_requirements) describe in business terms what must be delivered or accomplished to provide value.

b)Product requirements describe properties of a system or product (which could be one of several ways to accomplish a set of business requirements.)

c)Process requirements describe activities performed by the developing organization. For instance, process requirements could specify .Preliminary investigation examine project feasibility, the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system. All system is feasible if they are unlimited resources and infinite time. There are aspects in the feasibility study portion of the preliminary investigation:

**F)ECONOMIC FEASIBILITY**

A system can be developed technically and that will be used if installed must still be a good investment for the organization. In the economical feasibility, the development cost in creating the system is evaluated against the ultimate benefit derived from the new systems. Financial benefits must equal or exceed the costs. The system is economically feasible. It does not require any addition hardware or software. Since the interface for this system is developed using the existing resources and technologies available at NIC, There is nominal expenditure and economical feasibility for certain.

**G)OPERATIONAL FEASIBILITY**

Proposed projects are beneficial only if they can be turned out into information system. That will meet the organization’s operating requirements. Operational feasibility aspects of the project are to be taken as an important part of the project implementation. This system is targeted to be in accordance with the above- mentioned issues. Beforehand, the management issues and user requirements have been taken into consideration. So there is no question of resistance from the users that can undermine the possible application benefits. The well-planned design would ensure the optimal utilization of the computer resources and would help in the improvement of performance status.

**H)TECHNICAL FEASIBILITY**

Earlier no system existed to cater to the needs of ‘Secure Infrastructure Implementation System’. The current system developed is technically feasible. It is a web based user interface for audit workflow at NIC-CSD. Thus it provides an easy access to .the users. The database’s purpose is to create, establish and maintain a workflow among various entities in order to facilitate all concerned users in their various capacities or roles. Permission to the users would be granted based on the roles specified. Therefore, it provides the technical guarantee of accuracy, reliability and security.

**CHAPTER 8**

**DESIGN**

**UML diagrams**

The Unified Modeling Language allows the software engineer to express an analysis model using the modeling notation that is governed by a set of syntactic semantic and pragmatic rules.

A UML system is represented using five different views that describe the system from distinctly different perspective. Each view is defined by a set of diagram, which is as follows.

* **User Model View**

1)This view represents the system from the users perspective.

2)The analysis representation describes a usage scenario from the end-users perspective.

* **Structural Model view**

1)In this model the data and functionality are arrived from inside the system.

2)This model view models the static structures.

* **Behavioral Model View**

It represents the dynamic of behavioral as parts of the system, depicting the interactions of collection between various structural elements described in the user model and structural model view.

* **Implementation Model View**

In this the structural and behavioral as parts of the system are represented as they are to be built.

* **Environmental Model View**

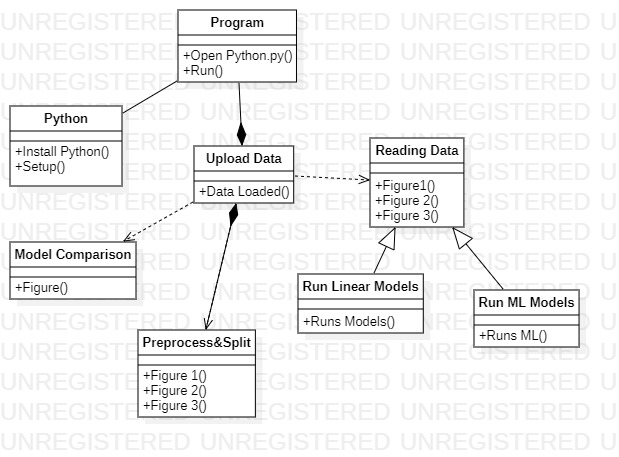
In this the structural and behavioral aspects of the environment in which the system is to be implemented are represented.

**8.1 CLASS DIAGRAM**

The class diagram is the main building block of object oriented modeling. It is used both for general conceptual modeling of the systematic of the application, and for detailed modeling translating the models into programming code. Class diagrams can also be used for data modeling.

The classes in a class diagram represent both the main objects, interactions in the application and the classes to be programmed. A class with three sections, in the diagram, classes is represented with boxes which contain three parts:

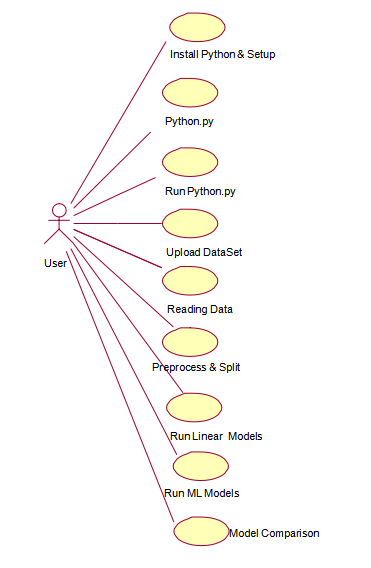
* The upper part holds the name of the class
* The middle part contains the attributes of the class
* The bottom part gives the methods or operations the class can take or undertake



**FIG 8.1** CLASS DIAGRAM

**8.2 USECASE DIAGRAM**

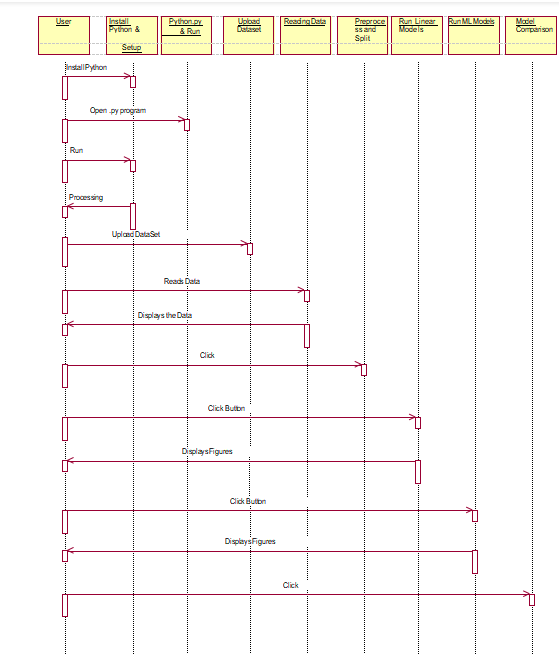
A **use case diagram** at its simplest is a representation of a user's interaction with the system and depicting the specifications of a use case. A use case diagram can portray the different types of users of a system and the various ways that they interact with the system. This type of diagram is typically used in conjunction with the textual use case and will often be accompanied by other types of diagrams as well.



**Fig 8.2** Usecase Diagram

**8.3 Seuence Diagram**

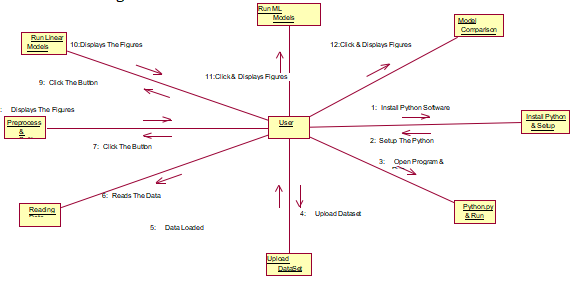
A **sequence diagram** is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the Logical View of the system under development. Sequence diagrams are sometimes called **event diagrams**, **event scenarios**, and timing diagrams.



**Fig 8.3** Sequence Diagram

**8.4Collaboration diagram**

A collaboration diagram describes interactions among objects in terms of sequenced messages. Collaboration diagrams represent a combination of information taken from class, sequence, and use case diagrams describing both the static structure and dynamic behavior of a system.

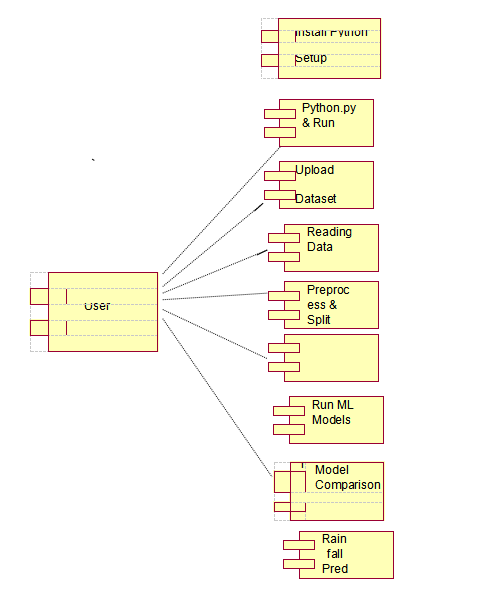


**Fig 8.4** Collaboration Diagram

**8.5 Component Diagram**

In the Unified Modeling Language, a component diagram depicts how components are wired together to form larger components and or software systems. They are used to illustrate the structure of arbitrarily complex systems

Components are wired together by using an assembly connector to connect the required interface of one component with the provided interface of another component. This illustrates the service consumer - service provider relationship between the two components



**Fig 8.5** Component Diagram

**8.6 Deployment Diagram**

A deployment diagram in the Unified Modeling Language models the physical deployment of artifacts on nodes. To describe a web site, for example, a deployment diagram would show what hardware components ("nodes") exist (e.g., a web server, an application server, and a database server), what software components ("artifacts") run on each node (e.g., web application, database), and how the different pieces are connected.

The nodes appear as boxes, and the artifacts allocated to each node appear as rectangles within the boxes. Nodes may have sub nodes, which appear as nested boxes. A single node in a deployment diagram may conceptually represent multiple physical nodes, such as a cluster of database servers.

**Fig 8.6** Deployment Diagram

Target

Data

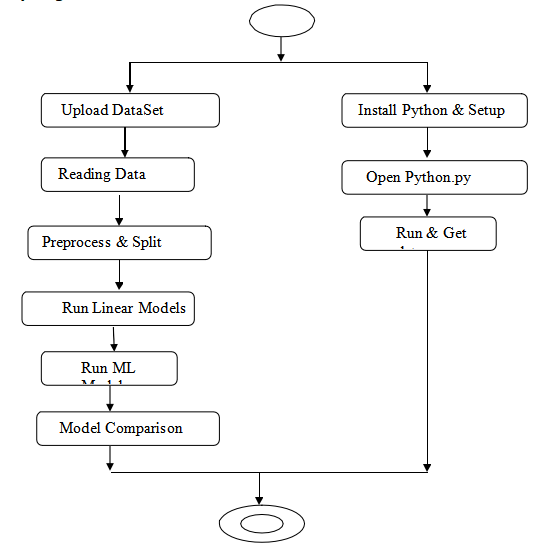
Meterological Data Set

Historical Weather Data

**8.7 Activity Diagram**

Activity diagram is another important diagram in UML to describe dynamic aspects of the system. It is basically a flow chart to represent the flow form one activity to another activity. The activity can be described as an operation of the system.

So the control flow is drawn from one operation to another. This flow can be sequential, branched or concurrent.

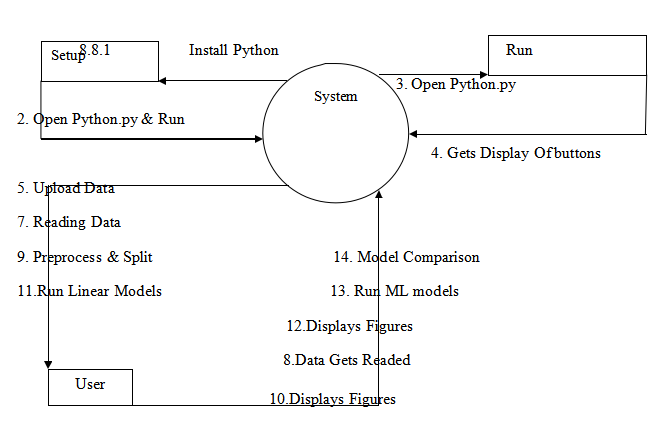


**Fig 8.7** Activity diagram

**8.8 Data Flow diagram**

[Data flow diagrams](http://www.edrawsoft.com/Data-Flow-Diagrams.php) illustrate how data is processed by a system in terms of inputs and outputs. Data flow diagrams can be used to provide a clear representation of any business function. The technique starts with an overall picture of the business and continues by analyzing each of the functional areas of interest. This analysis can be carried out in precisely the level of detail required. The technique exploits a method called top-down expansion to conduct the analysis in a targeted way.

As the name suggests, Data Flow Diagram (DFD) is an illustration that explicates the passage of information in a process. A DFD can be easily drawn using simple symbols. Additionally, complicated processes can be easily automated by creating DFDs using easy-to-use, free downloadable diagramming tools. A DFD is a model for constructing and analyzing information processes. DFD illustrates the flow of information in a process depending upon the inputs and outputs. A DFD can also be referred to as a Process Model. A DFD demonstrates business or technical process with the support of the outside data saved, plus the data flowing from the process to another and the end results.

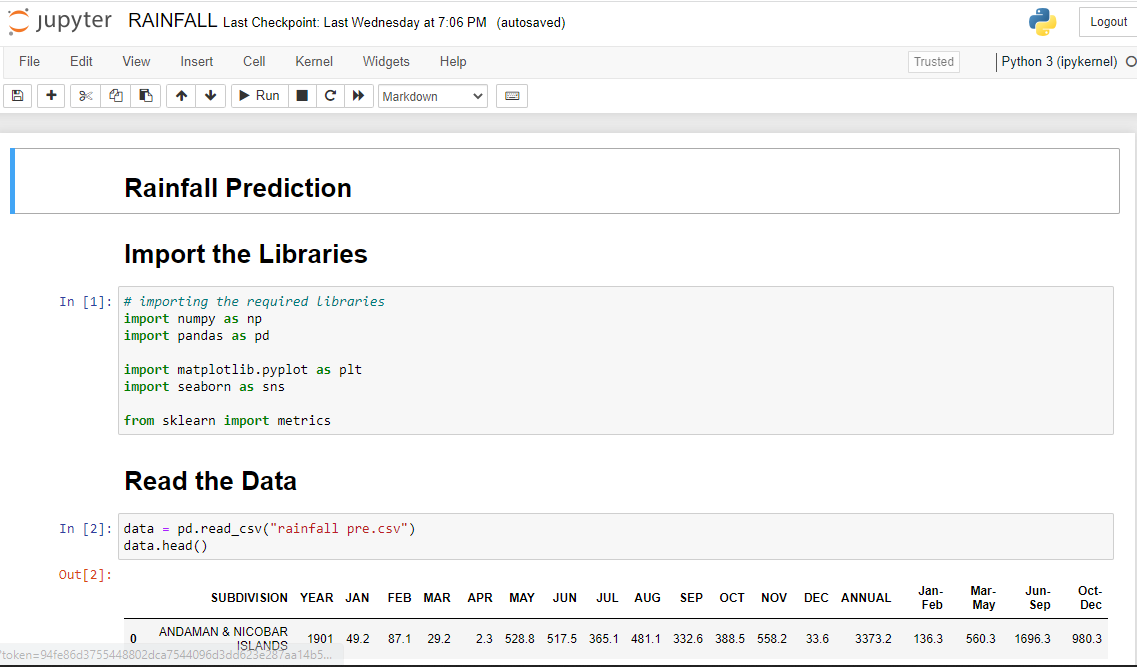


**Fig 8.8** Data flow Diagram

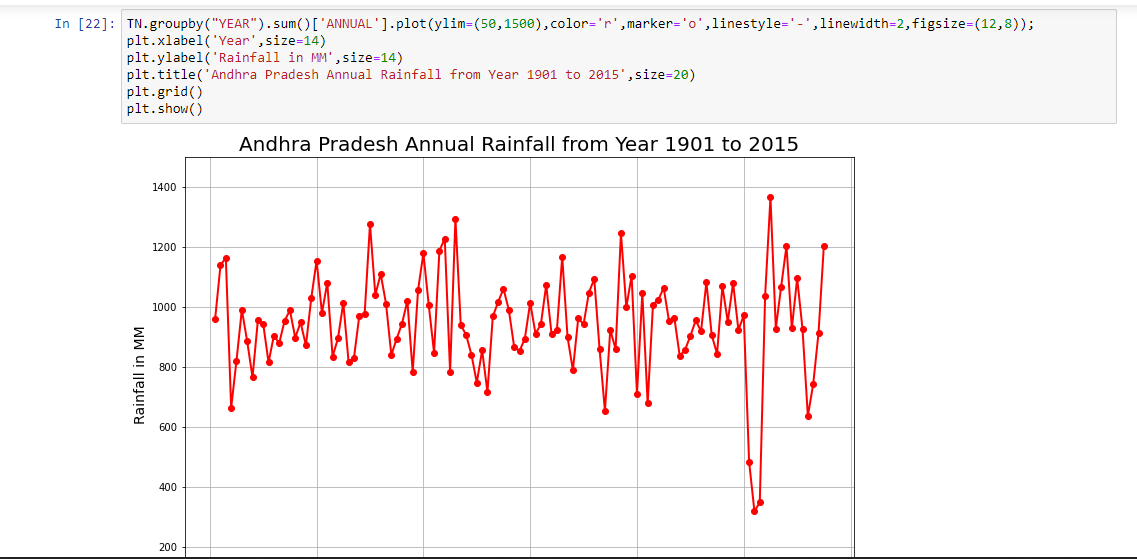
**CHAPTER 9**

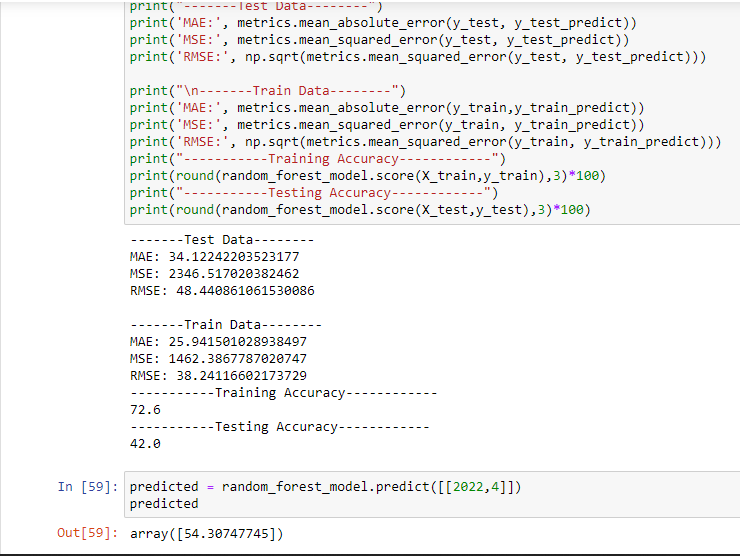
**SCREEN SHOTS**

**9.1 ML CODE SCREEN SHOTS**

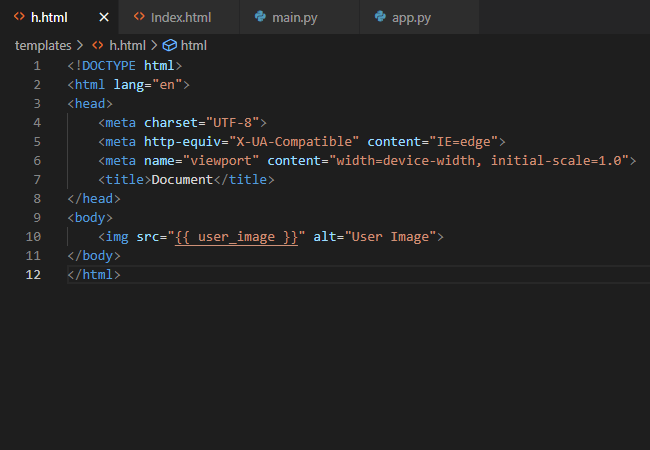


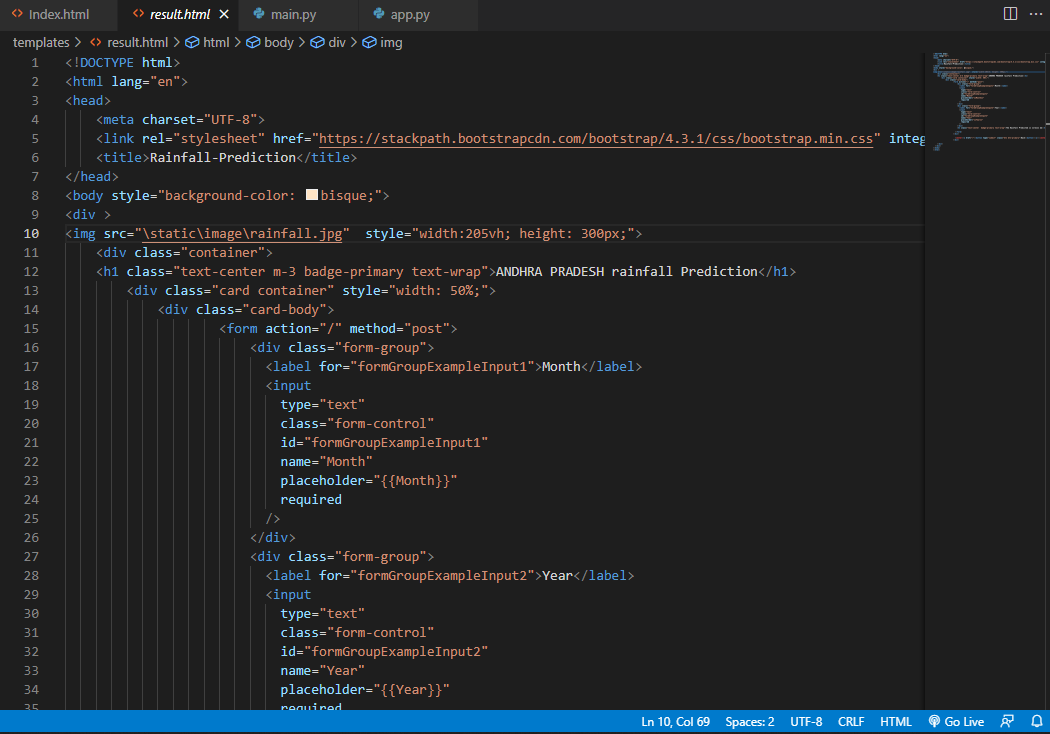


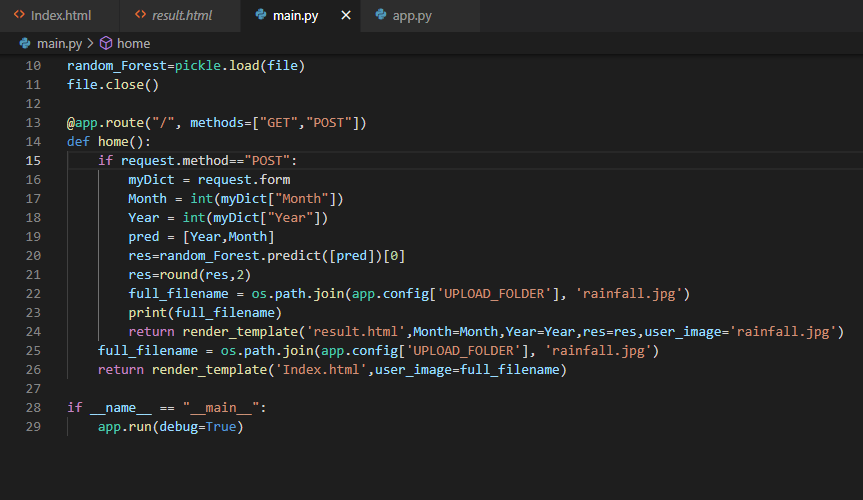


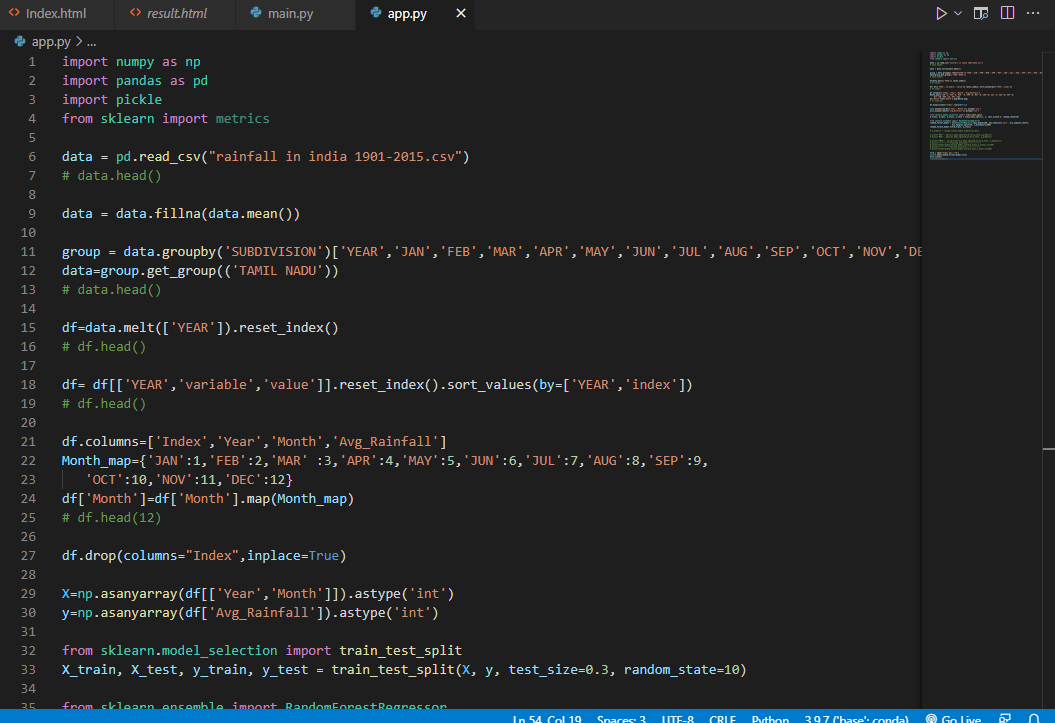


**9.2 GUI SCREEN CODE SCREEN SHOTS**

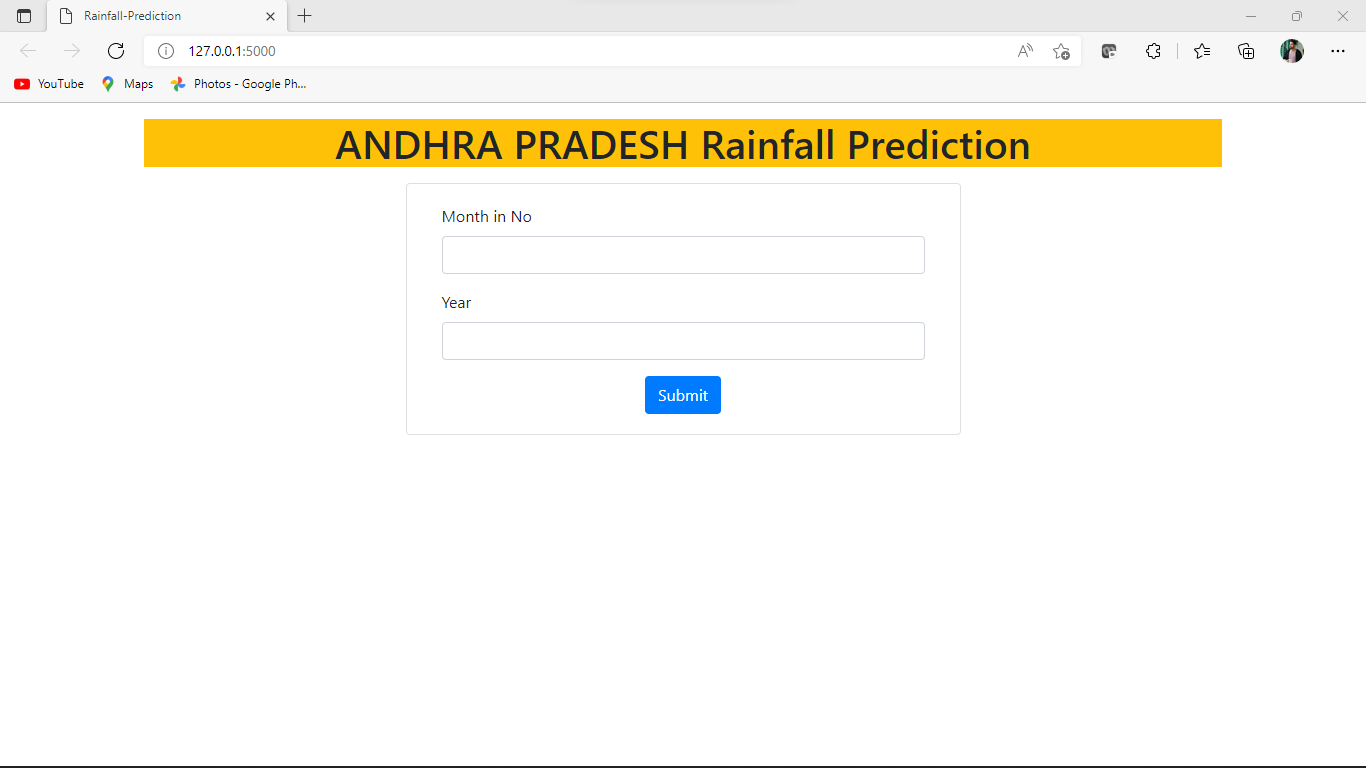


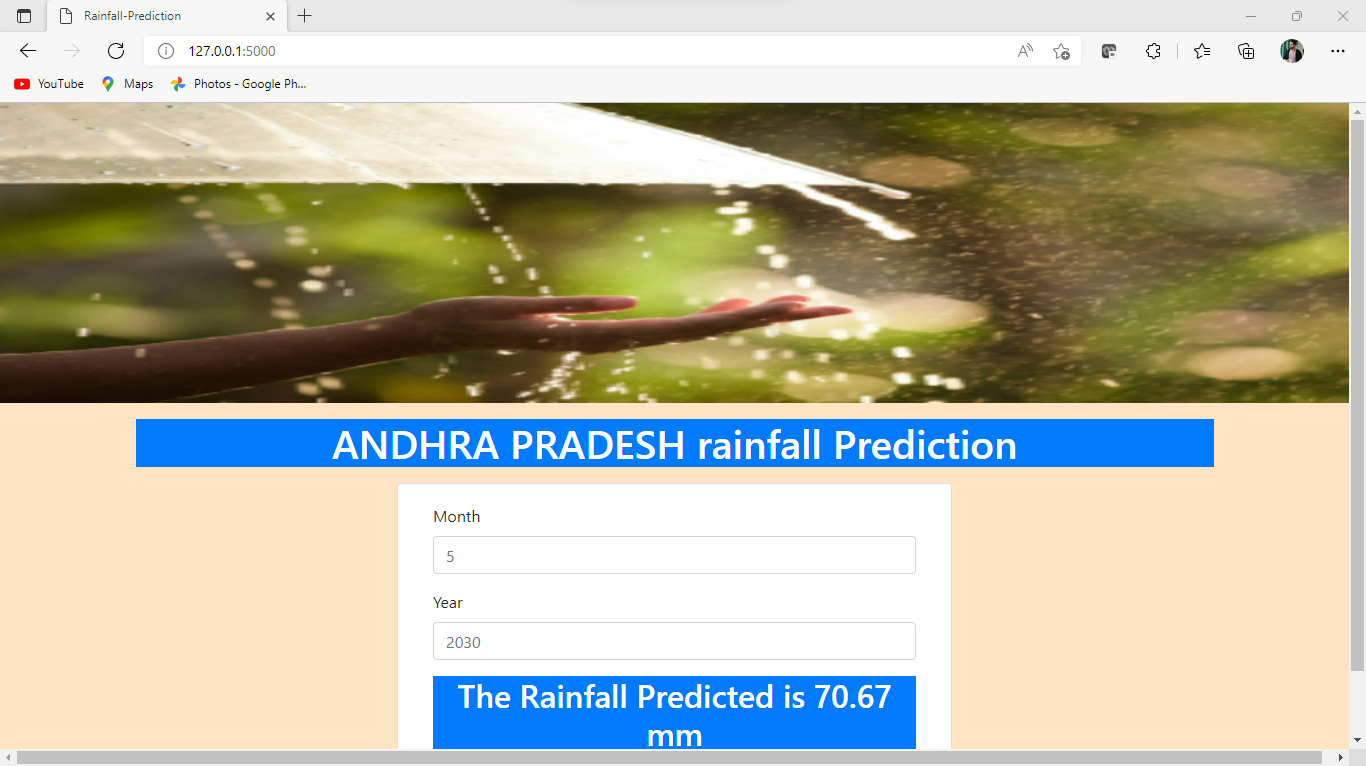






**9.3 GUI SCREENS PICTURES**





**CONCLUSION**

Rainfall is one the most significant natural phenomenon that is not only important for the human beings only but the living beings. Due to the changing climatic conditions, rainfall cycles are also changing and the temperature of the earth is rising. The changing temperature is also affecting the agriculture, industry and sometimes may cause flooding and land slide. Therefore, it is essential for the human beings to keep a check upon this natural phenomenon in order to survive. The water is a scarce natural resource without which human life is impossible and also there is no substitute to this natural resource. Thus, predicting the rainfall for agriculture and water reserves, also it also good for keeping human beings alert of natural disasters like flood and landslide. However, to overcome these issues and meet the demands, a system to forecast rainfall is essential using artificial intelligence of neural that is popular within the modern technology.

The study aimed at building a predicting system using neural networks that could predict monthly rainfall accurately and efficiently with minimum error. The study incorporated different areas and used their rainfall data with different neural networks like ANFIS and ANN, through training the networks with these inputs and outputs. The trained data is tested and then validated by making a comparison between actual and predicted data. The system used feature extraction to deduce the output prediction that could be more precise and accurate. The neural networks with different algorithms and functions were trained with rainfall parameters and the previous rainfall data to predict the results in this study. After training and testing; the results were compared to check the efficiency of the system; the RMSE’s were recorded to make sure that the system will operate not only to make the prediction but also the accurate data will be obtained. The study utilized back propagation, LASSO and Hybrid algorithms to forecast the rainfall.

Lastly, the rainfall predictions after training, testing are obtained that are quite accurate and through comparison outlined that the actual and predicted data for these areas illustrated finest results using the certainly different parameters of the rainfall that are different for different areas with minimum error observed using ANFIS only. The AFIS model outlined efficiency INDIA. The LASSO network performed not so well in the comparison for actual and predicted data.

**REFERENCES**

[1] Breiman, L. (1998a). Arcing classifiers (discussion paper). Annals of Statistics, 26, 801–824.

Google Scholar

[2] Breiman. L. (1998b). Randomizing outputs to increase prediction accuracy. Technical Report 518, May 1, 1998, Statistics Department, UCB (in press, Machine Learning).

[3] Breiman, L. 1999. Using adaptive bagging to debias regressions. Technical Report 547, Statistics Dept. UCB.

[4] Breiman, L. 2000. Some infinity theory for predictor ensembles. Technical Report 579, Statistics Dept. UCB.

[5] Dietterich, T. (1998). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization, Machine Learning, 1–22.

[6] Freund, Y. & Schapire, R. (1996). Experiments with a new boosting algorithm, Machine Learning: Proceedings of the Thirteenth International Conference, 148–156.

[7] Grove, A. & Schuurmans, D. (1998). Boosting in the limit: Maximizing the margin of learned ensembles. In Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98).

[8] Ho, T. K. (1998). The random subspace method for constructing decision forests. IEEE Trans. on Pattern Analysis and Machine Intelligence, 20(8), 832–844.

[9] Kleinberg, E. (2000). On the algorithmic implementation of stochastic discrimination. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(5), 473–490.

[10] Schapire, R., Freund, Y., Bartlett, P., & Lee,W. (1998). Boosting the margin:Anewexplanation for the effectiveness of voting methods. Annals of Statistics, 26(5), 1651–1686.

## APPENDIX

## A)SOURCE CODE

The following is the code of the project **“Rainfall Prediction”**.The code contains all the packages that are required for the project and the functions that are required also written in the code.

# importing the required libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn import metrics

data = pd.read\_csv("rainfall pre.csv")

data.head()

data.info()

data.isnull().sum()

data[["SUBDIVISION","ANNUAL"]].groupby("SUBDIVISION").sum().sort\_values(by='ANNUAL',ascending=False).plot(kind='barh',stacked=True,figsize=(15,10))

plt.xlabel("Rainfall in MM",size=12)

plt.ylabel("Sub-Division",size=12)

plt.title("Annual Rainfall v/s SubDivisions")

plt.grid(axis="x",linestyle="-.")

plt.show()

plt.figure(figsize=(15,8))

data.groupby("YEAR").sum()['ANNUAL'].plot(kind="line",color="r",marker=".")

plt.xlabel("YEARS",size=12)

plt.ylabel("RAINFALL IN MM",size=12)

plt.grid(axis="both",linestyle="-.")

plt.title("Rainfall over Years")

plt.show()

plt.figure(figsize=(10,6))

TN[['JAN', 'FEB', 'MAR', 'APR','MAY', 'JUN','JUL','AUG', 'SEP', 'OCT','NOV','DEC']].mean().plot(kind="bar",width=0.5,linewidth=2)

plt.title("Andhra Pradesh Rainfall v/s Months",size=20)

plt.xlabel("Months",size=14)

plt.ylabel("Rainfall in MM",size=14)

plt.grid(axis="both",linestyle="-.")

plt.show()

# analysis of rainfall data of rajasthan

Rajasthan = data.loc[((data['SUBDIVISION'] == 'WEST RAJASTHAN') | (data['SUBDIVISION'] == 'EAST RAJASTHAN'))]

Rajasthan.head()

#linear regression

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=10)

from sklearn.linear\_model import LinearRegression

LR = LinearRegression()

LR.fit(X\_train,y\_train)

# predicting

y\_train\_predict=LR.predict(X\_train)

y\_test\_predict=LR.predict(X\_test)

print("-------Test Data--------")

print('MAE:', metrics.mean\_absolute\_error(y\_test, y\_test\_predict))

print('MSE:', metrics.mean\_squared\_error(y\_test, y\_test\_predict))

print('RMSE:', np.sqrt(metrics.mean\_squared\_error(y\_test, y\_test\_predict)))

print("\n-------Train Data--------")

print('MAE:', metrics.mean\_absolute\_error(y\_train,y\_train\_predict))

print('MSE:', metrics.mean\_squared\_error(y\_train, y\_train\_predict))

print('RMSE:', np.sqrt(metrics.mean\_squared\_error(y\_train, y\_train\_predict)))

print("\n-----Training Accuracy-------")

print(round(LR.score(X\_train,y\_train),3)\*100)

print("-----Testing Accuracy--------")

print(round(LR.score(X\_test,y\_test),3)\*100)

from sklearn.linear\_model import Lasso

from sklearn.model\_selection import GridSearchCV

# create a lasso object

lasso = Lasso(max\_iter=100000)

# check for best alpha value using GridSearch

parameter={'alpha':[1e-15,1e-10,1e-8,1e-3,1e-2,1,5,1e1,1e2,1e3,1e4,1e5,1e6,1e7]}

lasso\_regressor=GridSearchCV(lasso,parameter,

scoring='neg\_mean\_squared\_error',

cv=5

)

print("Best Parameter for Lasso:",lasso\_regressor.best\_estimator\_)

lasso=Lasso(alpha=100.0,max\_iter=100000)

lasso.fit(X\_train,y\_train)

y\_train\_predict=lasso.predict(X\_train)

y\_test\_predict=lasso.predict(X\_test)

from sklearn import metrics

print("-------Test Data--------")

print('MAE:', metrics.mean\_absolute\_error(y\_test, y\_test\_predict))

print('MSE:', metrics.mean\_squared\_error(y\_test, y\_test\_predict))

print('RMSE:', np.sqrt(metrics.mean\_squared\_error(y\_test, y\_test\_predict)))

print("\n-------Train Data--------")

print('MAE:', metrics.mean\_absolute\_error(y\_train,y\_train\_predict))

print('MSE:', metrics.mean\_squared\_error(y\_train, y\_train\_predict))

print('RMSE:', np.sqrt(metrics.mean\_squared\_error(y\_train, y\_train\_predict)))

print("\n-----Training Accuracy-------")

print(round(lasso.score(X\_train,y\_train),3)\*100)

print("-----Testing Accuracy--------")

print(round(lasso.score(X\_test,y\_test),3)\*100)

from sklearn.linear\_model import Ridge

from sklearn.model\_selection import GridSearchCV

ridge=Ridge()

parameters={'alpha':[1e-15,1e-10,1e-8,1e-3,1e-2,1,5,10,20,30,35,40,45,50,55,100]}

ridge\_regressor=GridSearchCV(ridge,parameters,scoring='neg\_mean\_squared\_error',cv=5)

ridge\_regressor.fit(X\_train,y\_train)

print(ridge\_regressor.best\_params\_)

print(ridge\_regressor.best\_score\_)

**#GUI SCREENS CODE**

from flask import \*

import pickle

import os

app=Flask(\_\_name\_\_,template\_folder="templates")

PEOPLE\_FOLDER = os.path.join('static', 'image')

app.config['UPLOAD\_FOLDER'] = PEOPLE\_FOLDER

file=open("model.pkl","rb")

random\_Forest=pickle.load(file)

file.close()

@app.route("/", methods=["GET","POST"])

def home():

    if request.method=="POST":

        myDict = request.form

        Month = int(myDict["Month"])

        Year = int(myDict["Year"])

        pred = [Year,Month]

        res=random\_Forest.predict([pred])[0]

        res=round(res,2)

        full\_filename = os.path.join(app.config['UPLOAD\_FOLDER'], 'rainfall.jpg')

        print(full\_filename)

        return render\_template('result.html',Month=Month,Year=Year,res=res,user\_image='rainfall.jpg')

    full\_filename = os.path.join(app.config['UPLOAD\_FOLDER'], 'rainfall.jpg')

    return render\_template('Index.html',user\_image=full\_filename)

if \_\_name\_\_ == "\_\_main\_\_":

app.run(debug=True)

import numpy as np

import pandas as pd

import pickle

from sklearn import metrics

data = pd.read\_csv("rainfall in india 1901-2015.csv")

# data.head()

data = data.fillna(data.mean())

group = data.groupby('SUBDIVISION')['YEAR','JAN','FEB','MAR','APR','MAY','JUN','JUL','AUG','SEP','OCT','NOV','DEC']

data=group.get\_group(('TAMIL NADU'))

# data.head()

df=data.melt(['YEAR']).reset\_index()

# df.head()

df= df[['YEAR','variable','value']].reset\_index().sort\_values(by=['YEAR','index'])

# df.head()

df.columns=['Index','Year','Month','Avg\_Rainfall']

Month\_map={'JAN':1,'FEB':2,'MAR' :3,'APR':4,'MAY':5,'JUN':6,'JUL':7,'AUG':8,'SEP':9,

   'OCT':10,'NOV':11,'DEC':12}

df['Month']=df['Month'].map(Month\_map)

# df.head(12)

df.drop(columns="Index",inplace=True)

X=np.asanyarray(df[['Year','Month']]).astype('int')

y=np.asanyarray(df['Avg\_Rainfall']).astype('int')

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=10)

from sklearn.ensemble import RandomForestRegressor

random\_forest\_model = RandomForestRegressor(max\_depth=100, max\_features='sqrt', min\_samples\_leaf=4,

                      min\_samples\_split=10, n\_estimators=800)

random\_forest\_model.fit(X\_train, y\_train)

# y\_predict = random\_forest\_model.predict(X\_test)

# print('MAE:', metrics.mean\_absolute\_error(y\_test,y\_predict))

# print('MSE:', metrics.mean\_squared\_error(y\_test, y\_predict))

# print('RMSE:', np.sqrt(metrics.mean\_squared\_error(y\_test, y\_predict)))

# print("-----------Training Accuracy------------")

# print(round(random\_forest\_model.score(X\_train,y\_train),3)\*100)

# print("-----------Testing Accuracy------------")

# print(round(random\_forest\_model.score(X\_test,y\_test),3)\*100)

file = open("model.pkl","wb")

pickle.dump(random\_forest\_model,file)

file.close()

# print(y\_predict)