INDIA RAINFALL ANALYSIS PREDICTION

Major project report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology

In

Computer Science and Engineering

By

Tamanna (181427), Dikshant (181453),

UNDER THE SUPERVISON OF

Dr. Amol Vasudeva

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology, Waknaghat, 173234, Himachal Pradesh, INDIA

CERTIFICATE

This is to certify that the work which is being presented in the project report titled "**INDIA RAINFALL ANALYSIS PREDICTION"** in partial fulfilment of the requirements for the award of the degree of B.Tech in Computer Science And Engineering and submitted to the Department of Computer Science And Engineering, Jaypee University of Information Technology, Waknaghat is an authentic record of work carried out by "Tamanna (181427), Dikshant(181453)" during the period from January 2022 to May 2022 under the supervision of **Dr. Amol Vasudeva**, Assistant Professor (Senior Grade), Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat.

Tamanna   **Dikshant**

(181427)   **(181453)**

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

Supervised by:

Dr. Amol Vasudeva

Assistant Professor (Senior Grade)

Department of Computer Science & Engineering and Information Technology Jaypee University of Information Technology

Dated:

I

AKCNOWLEDGEMENT

Firstly, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the project work successfully.

I really grateful and wish my profound my indebtedness to Supervisor **Dr. Amol Vasudeva**, Assistant Professor (SG), Department of CSE Jaypee University of Information Technology, Wakhnaghat. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would like to express my heartiest gratitude to **Dr. Amol Vasudeva**, Department of CSE, for his kind help to finish my project.

I would also generously welcome each one of those individuals who have helped me straight forwardly or in a roundabout way in making the project a win. In this unique situation, I might want to thank the various staff individuals, both educating and non- instructing, which have developed their convenient help and facilitated my undertaking.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

Tamanna   **Dikshant**

TABLE OF CONTENT

Content   Page No.

LIST OF FIGURES

IV

ABSTRACT

India rainfall is a major concern in these days. Weather conditions have changed in the last three years. This has caused a drastic change in rainfall pattern. Effective rainfall factors are temperature, humidity, wind speed, pressure and precipitation. These factors are affecting the rainfall hugely. The study of rainfall is very important against the behavior and factors influencing rainfall. Only then we will be able to predict the current rainfall accurately. Machine Learning has made our work a lot easier. In machine learning there are lot of supervised and unsupervised algorithms. These algorithms are very helpful for India rainfall analysis project. The prediction of rainfall is done by the supervised machine learning using the regression concept. The aim of this project is to provide easily to access to known experts to the techniques and also used in the field of rainfall prediction and also study comparative between different types of machine learning techniques. Weather prediction models are mostly form on artificial neural networks. This research work makes relative studies of the two precipitation predictions tactics and discover a more precise one. Current techniques for predicting rainfall do not goes with the complex data that exists. The methods that we are using today are arithmetic systems and analytical methods, that do not work correctly if there is non-linear pattern. Prevailing system be unsuccessful every time dataset containing previous rainfall increasing in complexity. Rainfall is reviewed to be the main foundation of greatest of our country's economy. Likewise , agriculture is examined to be the main economy source. Appropriate estimation of rainfall is required to make correct speculation on agriculture. Prediction of rainfall is essential for agriculture as well as people of beach areas. Persons in beach areas are more prone to substantial rainfall and downpours, hence they must know about the rainfall well in advance. Parts which receive fewer rainfall and face water shortage should have rainwater harvesting. To begin a appropriate rainwater harvesting, rainfall approximation is required. Weather forecast is the easiest way to get maximum access. This research work can be used by all weather forecasting channels, so that the forecast news can be more accurate and spread to all parts of the country. Rainfall can cause many disasters due to heavy rainfall. Thus the aim of the project is to take prediction measures and more accurate predictions. The two basic predictions are short and long. The short term gives more accurate results. There the

major role in India is accuracy because the large population is dependent on heavy rainfall like agriculture and other fields. thus we used various machine learning algorithms to predict. The aim of the project is to find easy access to some higher technologies and go through the various algorithms

Chapter 01

INTRODUCTION

1.1   **Introduction**

1.2   **Problem Statement**

1.3   **Objectives**

1.4   **Methodology**

1.5   **Technical Tools and Technology**

1.1 Introduction

Rainfall prediction is used to predict the rainfall on the basis of given dataset as State, District, Month, Annual rainfall level and the grouped months precipitation. We develop this project based on linear regression algorithm using normal equation. First of all the regression is a first approach of linear model to find the actual relationship between the independent variable and dependent variable. We have a dataset for predict the coming precipitation. We predict the price with normal equation, gradient descent, classifications, and logistic regression etc. we are using the linear regression by using the normal equation. We use multiple linear regression algorithms for predict and when we run the model on Jupiter notebook then we display the results on I python notebook cells. As of today, rainfall is considered to be one of the sole responsible factors for most of the important things around the world. In India, agriculture is considered as one of the important factors to decide the economy of the country and agriculture is completely dependent on rainfall . Apart from this it is very necessary to find out the amount of rainfall in coastal areas around the world. In some areas where water is scarce, for setting up rainwater harvesting, rainfall must be predicted in advance. This project relates with prediction of rainfall using machine learning and neural networks. This project makes a relative study of machine learning approach and neural networks approach, then consequently delineates effective method for precipitation prediction. First preprocessing is done.

Preprocessing is the process of representing a dataset in the form of multiple graphs such as bar graphs, histogram's, etc.

Presently, rainfall forecasting is becoming major issues for most water preservation systems in the world. Most rainfall forecasting systems, now a days are not able to find a hidden layers or any non-linear patterns present in the classification. This project will help in finding all the hidden layers as well as non-linear patterns, which is valuable for making accurate predictions of rainfall. Rainfall Prediction is application to predict rainfall in a given area. We can do it in two ways. The first is to analyze the physical law influencing rainfall and the second is to create a system that will discover the hidden patterns or characteristics of the physical factors influencing it and the process involved in achieving it. The second is preferable because it does not involve any kind of mathematical calculations and can be useful for complex and non-linear data. Due to the presence of systems that do not accurately detect hidden layers and non-linear patterns, prediction results are inaccurate most of the time and can lead to huge losses. Therefore, the main objective of this research work is to find a system which can solve both the issues i.e., complexity as well as being able to find the hidden layers, which will give a fair and accurate prediction which will help the country in the development of agriculture. Will get and economy. In the context of climate change, it is relevant to find out whether the characteristics of the Indian summer monsoon are also changing. Indian Summer Monsoon (June-September) rainfall is very important for the country's economic development, disaster management, hydrology planning. Despite the growth in the service sector, India's economy is still dependent on agro-culture. More serious cases like crop failure, drought and famine due to weak or poor monsoon become very important for the country.

Rainfall plays a important role in the field and agriculture of india .we have to predict the rainfall on the basis of the accuracy such that to find the mean absolute error. if the mean absolute error is high in each district every month then it has high rainfall on the same district .If the mean absolute error is high it means some regions has low rainfall .so it can be concluded that the neutral approach is higher then all the machine learning approches because capability for finding entire nonlinear pattern is present inside any system, supervised and unsupervised can be accessed in some other countries like Australia. There is a relationship between linear and non-linear independent variables. It is basically deal with the collection of data and then use for further processing variables. Then the data is further processed and then there are the two approach of rainfall prediction which is regression and neural network . The mean absolute error is of neural network is low .Lower the absolute error higher the prediction .It states that the neural network performs better then the other regression algorithms. we have

to represent our dataset in the form of the graphs .one of the major challenge is the in rainfall data has much higher complexity .There are the most of the prediction system that are main goal to find hidden and non-hidden patterns .The first is the to analysis the affects of rainfall, Second one is make system using hidden patterns. but the mathematical approaches are more accurate in predicting the rainfall .The graph received a collection on the preprocess buttons there is value by month to month and year to year .firstly there is a input from the user the preprocess using lasso ,we have to splitting the data .then types are neural network and testing of data set .Then find results.

1.2 Problem Statement

Rainfall analysis is very important for these days. Why it is important for country like India. Indian economy is hugely relied on primary sector for employment and living including agriculture and secondary related to it. Rainfall therefore becomes necessary and thereby its prediction is also helpful for farmer to sow their crops accordingly. Heavy rainfall on the other hand can bring some disaster related scenario where timely and accurate prediction can helps provide disaster related management related works that's why we need to predict the rainfall intime. The rain forecast is very high important because heavy and irregular Rain can have many effects like destruction of crops and fields, property damage is better forecasting model needed. Early warning that can reduce risk Helps even more for life and property Management of agricultural farms in better way. cause heavy rain flood and for natural disasters like dry he square measure faced by individuals world every year. There are many models has been developed to assess rainfall and to estimate the probability of rain. These models are based on both supervised and unsupervised machine learning algorithms. take in overall rainfall considerations Don't Help Us Know If It's Raining specific conditions. accuracy is Major concern in machine learning. we're going to understand the data and then train the model accordingly predict whether it rains under the given conditions or not. Rainfall plays an important impact on agricultural crops, and farms. Heavy rainfall is a major cause of rainfall trending damage to property. So need a better forecasting model is required so that we can prevent the rainfall by using short-term and long-term predictions. The basic models are both supervised and unsupervised algorithms. The major concern in predicting the rainfall is the accuracy. We should understand the data and then train the model

accordingly. And then predict the rainfall under a given condition or not. The short term will predict the rainfall few days or weeks and the long term will predict the rainfall in the months and years. This means that the prediction of the rainfall is done by calculating the mean square error. If the mean square error is high means that algorithms will predict algorithm is more accurate as compared to others. The major concern is how to predict rainfall using such types of machine learning algorithms. Rainfall is not so important for the algaculture but also important for the other things like living things. then firstly we have to import all the libraries and then clean the dataset. Then we have to predict the rainfall with the help of the histograms and graphs. plotting a histogram shows that we have to build the graphs using these graphs we predict in which year or month the trend or rainfall is increased. Then we have to plot the bar graphs showing the trend in rainfall a few weeks or months in advance. Thus Plays the biggest impact..

1.3 Objective

The weather condition is plays important role in social and economic life in the country. Monsoon is difficult weather system in world which is the life line of all the countries. The monsoon also establish the unity of the country by continue the same seasoning system from south and north and from south and north and from east to west.

Ministry of earth sciences prepares Indian monsoon mission program.
Objective of the mission are: -
1. To forecast onset and quantity of monsoon
2. To improve the interaction between institutions and R&D organization to improve skill in forecast monsoon.
3. To develop modelling system for seasonal and extended range
4. To improve agriculture mechanism assistance
5. To study weather, ozone depletion levels.
6. To give warning of the disaster in the hilly areas in rainfall prone areas.

1.4 Methodology

The suggest ways is acted with the help of a simple flowchart. In this project we are using the three types of algorithms which includes linear regression, support vector machine, artificial neural networks. The project report shows how the factors vary the rainfall pattern. The prediction is totally based on the independent variables that bring the rainfall.

1.5   Technical Tools And Technology

Software Requirements And Hardware Requirements:

* ❖ OPERATING SYSTEM: We use Windows7 or newer for RAINFALL prediction web application.
* ❖ BROWSER: There is no single answer for this. Use whichever browser works best in your computer.
* ❖ However, I recommend the Chrome, Opera. PROCESSOR: 1.6 GHz
* ❖ CPU is minimal for web servers RAM: 4GB
* ❖ We run it on any browser.
* ❖ RAM is recommended.

Chapter 02

LITERATURE SURVEY

In this project there are various algorithms that is used for the prediction of rainfall. Basically, in this project the two most important techniques to predict Rainfall. One is Emperical and other one is Dynamic methods. The Emperical approach is based on the analysis of past weather data and its relationship to atmospheric variable is different. In a dynamical approach, predictions are generated based on numerical methods, such as using mathematical equations. There are different types of the algorithms that are very helpful for rainfall prediction such as, multiple linear regression, auto regression integrated moving average, adaptive splines threshold autoregressive, support vector machine, neural network. These are some of the algorithms that we are used in our project work to predicting the rainfall analysis

[1]. The authors have selected a supervised machine learning algorithm, the Naïve Bayes for predicting the rainfall analysis prediction. The algorithm that analyzes the past data related to temperature, zone area, and year, It takes historical data as input and output as a result. The algorithm is proved to simplify and reduce manual work and also provides smooth workflow

[2]. In this paper, the authors are used the Linear Regression algorithm along with neural network, and SVM algorithms is basically for predicting the rainfall analysis. The highest accuracy is getting by the LR algorithm

[3]. The authors is used time-series data for prediction. The data is temporal timeseries data as it is produced by scientific data, financial applications, GPS, weather data, etc Artificial Neural Network (ANN) are a widely used technique for modelling nonlinear and dynamic systems. The model has been developed to predict rainfall in between one months and two months.

[4]. In this paper, the authors have made their analysis using the correlation data and then prediction using the Multiple Linear Regression [MLR] model. The model has selected four parameters like cloud cover, precipitation, vapor pressure, average temperature. They have considered the **Andhra Pradesh** rainfall dataset for the prediction of rainfall.

[5] it is observed that in Andhra Pradesh, annual rainfall depends more in the months of January, February have been used for analyzing the pattern of rainfall. It also shows that there is rainfall in months march, April, May then there is lesser amount of rainfall in the months of June, July, August, September. It provides a complete analysis of the weather dataset of India

[6]. Shreekanth Parashar and Tanveer Hurra made an experiment using Data Mining Techniques for the prediction of rainfall analysis . They have used the Naïve Bayes algorithm, ANN, Decision Trees, Random Forest, and k-nearest neighbor Algorithm to find out which algorithm is best fit in the situation. They have further concluded that decision trees and Random forest-based models are best for predicting rainfall

[8]. Deepak Ranjan Nayak and his team have surveyed prediction of rainfall using ANN, this Method is for numerical methods and suitable than traditional methods. They have surveyed some of the commonly used neural networks like Back Propagation Network [BPN], Radial Basis Function Network [RBFN], Support Vector Machine [SVM], Self Organizing Map [SOM], artificial neural network is a special class. This paper can be more helpful for people who are using ANN for their predictions and the rainfall predicting algorithms that use MLR, SVM, BPN, and SOM are more beneficial

[9]. NawarajPaundel and Tekendra Nath Yogi conducted a machine learning algorithm for rainfall prediction in nepal. They have used classification algorithms like Decision-tree, Random Forest, and SVM. They have calculated recall, precision, accuracy, precision and F-measure values for all the algorithms, out of all, Random Forest has given 80% accuracy and it is higher. For early prediction linear regression method is used in this case. we have to predict the rainfall weather conditions like rainfall and cyclone Various techniques like data mining techniques are used to for model the decision tree. The algorithms are applied and performance is measured by calculating MSE ,RMSE , MASE. We also used k-mean nearest, support vector machines and performance of there algorithms can be calculated by using accuracy ,precision and recall .The knn have predictions in the various monsoon summer and post .We also used Gradient boosting Ada to measures the various accurate results. There are the various performance can be calculated by using confusion matrix.

Chapter 3

SYSTEM DEVELOPMENT

3.1 Analysis Of The Algorithms
3.2 About Data
3.3 About Data Scaling
3.4 Model Development
> ➢ **Use case diagram of the project**
> ➢ **Use case diagram of the admin**
> ➢ **Sequence diagram**
> ➢ **ER diagram**

3.1 Analysis Of The Algorithms

Linear Regression is a machine learning algorithm that is based supervised machine learning algorithms. Basically used for both classification as well as Regression but mainly used for Regression model.

It is a purpose of predicting the value based on the independent variables. It is mainly used for detecting the relationship between variables and forecasting many models like regression models that is based on the relationship between dependent and independent variables, they are considering the number of independent variables that are being used in project. It is the statistical model that defines linear relationship between the dependent variables and with a set of given variables. Linear relationship means value of one or more independent variable will change it may be increase or decrease with the given value of dependent variables. Majority is based on the response of values of the independent variables is assumed to be an function of those values less commonly the major conditional median and other percentile is used, such all formation of linear regression and regression attention on the reply of conditional probability distribution of the reply on the given values of the predictors, relatively the probability distribution of all of these variables which is the basic area of multivariate analysis . The Linear regression is the first type of regression to be studied carefully , and it is to be used vastly in real applications. This is model is depend linearly on their unknown variables the models is easily to be fit non-linearly connected to their variables. Since the statistical functions of the following calculators are very easy to decided. There are many

practical uses of Linear regression. Necessary applications are classified into one of the following categories:

-->Linear regression is used for predicting the data set values and the basic goal is predicting, reduction in error. Such a model is developed either if values of additional values are collecting with a response value. To make a prediction fitted model can be used.

-->Classification of variant can be attributed in response of the independent variables. It is used to strength the relationship between the independent variables and response. The details are unnecessary in response to subgroup of the independent variables.

3.2 About data

The problem was to build a large enough dataset. We have to divine the prices of the given apartment . We find a dataset done on web but nothing was available to what we are looking for. we are decide to implement "web scraping" basically used for extract information in the various websites. various features consists of of train and test data. There are the fixed size object of features in each row. feature can be accessed by its name and there are total eight features. And we take data from API. We use kaggle library for API. And the data is of States and in this there are many features and we have a target is to predict the fore coming rainfall levels. Level of rainfall is depending upon various features. And one feature will increase the possibility of rain and gives ascend to the graph. There are different ways but Linear regression models are best fit by using the least square application. Techniques such as reducing the "Lack of fit" and decreased least version of the penalized square as in the ridge and lasso. The linear regression is used to find the relational between a dependent variable with a given value of independent variable. The variable in vector is also called regression variables. There is relationship between the vectors of independent variables Thus we are trying to predict the vectors of independent variables. Using the training set such as linear function fit the training data. Linear regression is basically used to predict the forecast the dependent variables based on the regressor variables. We can use the straight line the equation as : $Y = \beta 0 + \beta 1x$. Hence the relational in the preceding diagram can also represented using same formula and basic task is to learn the value.

3.3 About Data Scaling

Data scaling is the method that is used for scale the data. When we use dataset for prediction then firstly we scale the data means we remove some values that are null or none. When null

values are occurs then we have a problem to predict or to use the model that is for prediction. In data science, outliers are the null values. According to the measurement of chart And data and there is no doubt about the presence of outliers. Abnormal values are taking very large and very small and even zero in one or more variables .Outliers can greatly effect the value of my learning algorithms. we know that straight line can be represented as: $Y = \beta 0 + \beta 1x$ .Hence the relationship can be represented as the same formula and the task of the linear regression model is to learn the value . Here is the code for removing outliers.

3.4 Model Development

> **Use case diagram of the User**
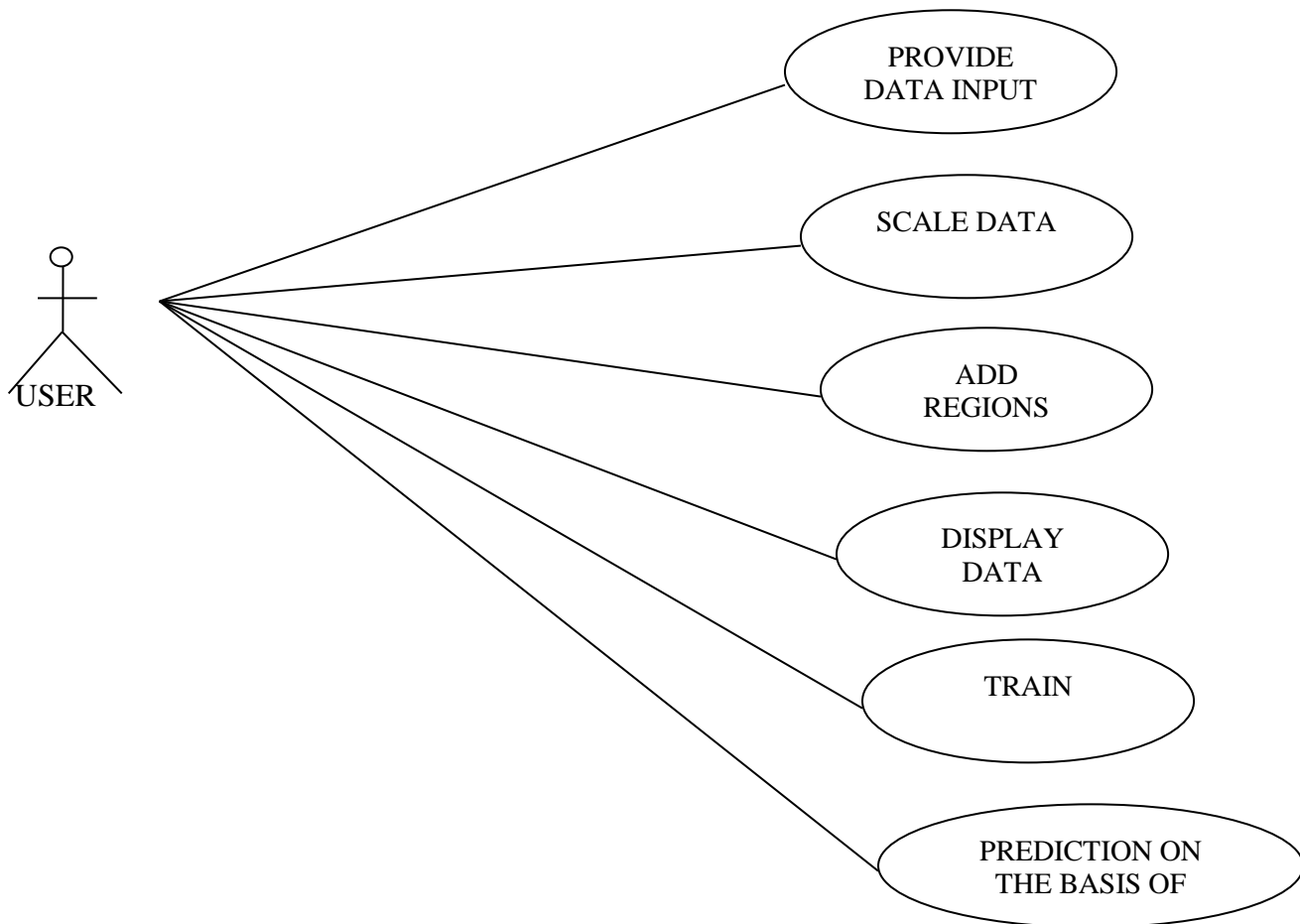
Fig 3.4.1 User diagram

➢ **Use case diagram of the admin**
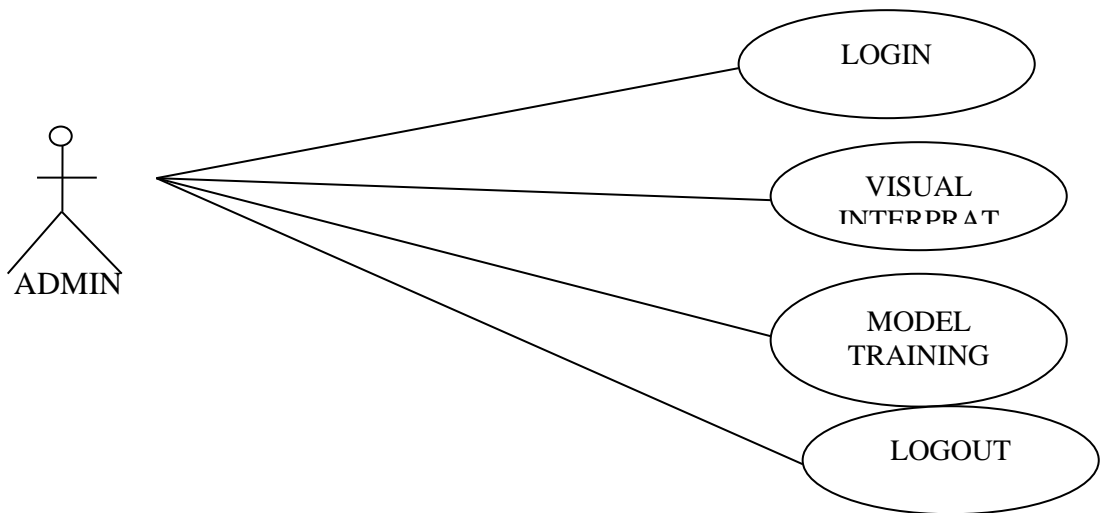
Fig 3.4.2 Admin diagram

➢ **Sequence diagram**
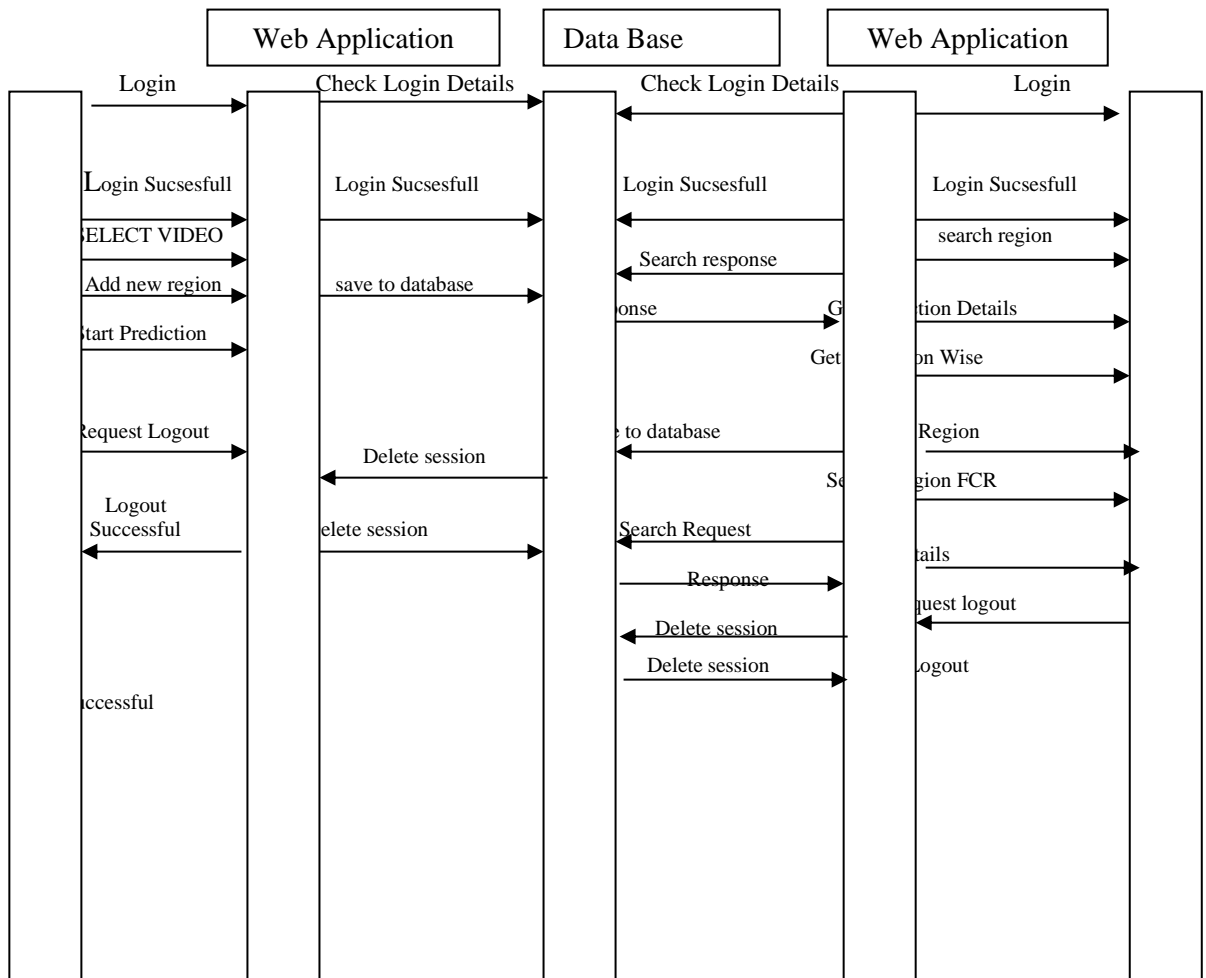
Fig 3.4.3 Sequence diagram
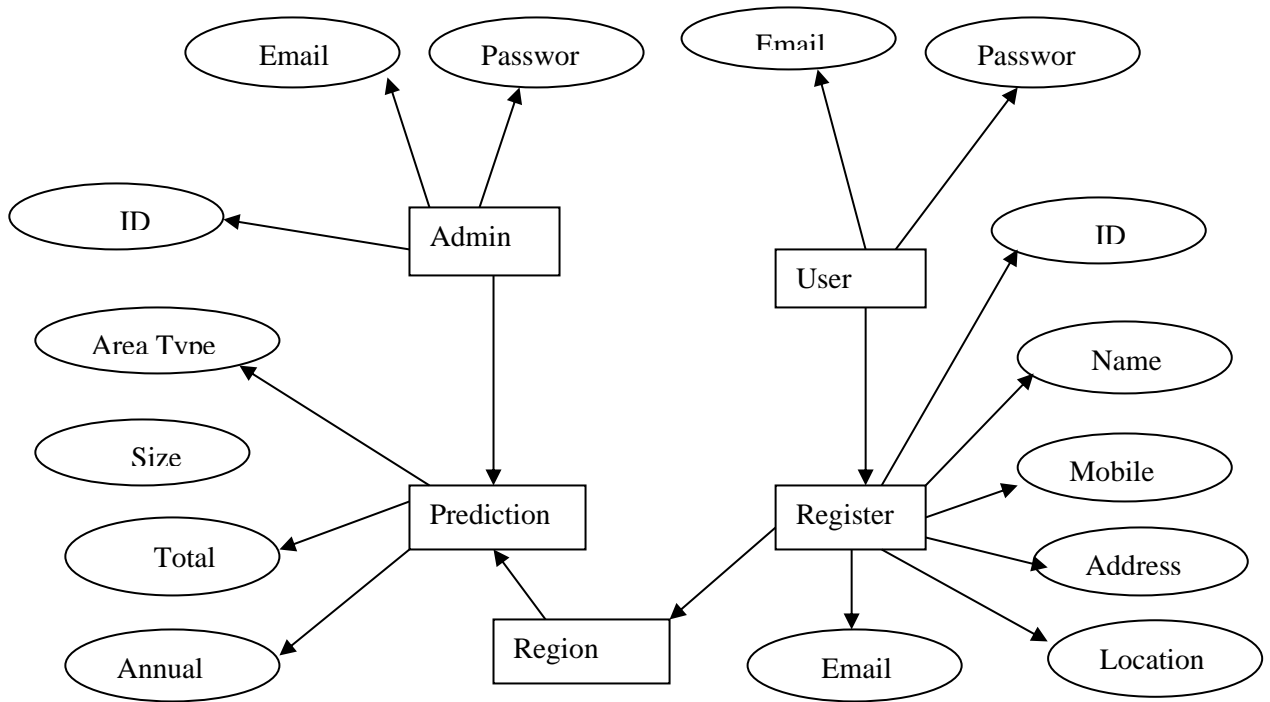
➢ **ER diagram**

Fig 3.4.4 ER diagram

Chapter 4

PERFORMANCE ANALYSIS

The execution of the project is split into seven sections . In the first part, we are going to import the necessary libraries and these libraries are very helpful for our project. Then next, we are going to prepare our given data set with required attributes then transformation is done on the data and then data analysis can be done using correlation, follow up by splitting the data set into train and test sets. Finally, model training is done by using the best fit algorithms which is best suitable in our project.

Step-1 Import Libraries

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Firstly we are imported NumPy, pandas, matplotlib, sklearn, and some important libraries . NumPy is used for mathematical computation and also used for N dimensional array . Pandas is used for suppose you want to access the csv, Excel file then you need to import the pandas and pandas is also used for data frames. Seaborn and matplotlib is used for data visualization and visualization can be done in the form of histograms, bar graphs, scatter plot and etc.

Step-2 Prepare Dataset

We have created our dataset different datasets taking the required features that are useful for our case study. we must have a basic first understanding of our dataset moving forward.

Step-3 Data Preprocessing

Data preprocessing is most important Steps when preparing our dataset for model training. data is often inconsistent, incomplete, and contains Noisy or unwanted data. therefore, Pre-

processing is required. that includes Some steps like handling missing Values, handling outliers, encoding Technique, scaling. Most important is to remove null values because the presence of null values will be Disturbances in the distribution of data, and may lead to false predictions. is very Low percentage of null values in the dataset.

Missing values:

Imputation is used to replace missing values. there are some types Mean-like imputation technique imputation, median imputation, mode imputation, random sampling allegation, etc. depending on the type of data we have, we can use necessary We have used the median Implantation to handle missing values.

Handling Outliers:

Outliers are nothing but an extreme value that deviates from the other Overview in the dataset. This Outsiders are either removed or replaced with their nearest limit value, either Upper limit or lower limit value.

Label Encoding:

Label encoding is one of a kind Encoding Techniques That Will Change categorical variable in numeric variable. important to convert label because our model only Understand numerical data.

Step-4: Visualization using the technique of Correlation

We need to understand our data. information Visualization is a powerful technique that Helps us to know about trends, Patterns that follow our data. there are Various techniques to visualize data, a Such a method is a correlation. Correlation Tells us how one or more are related. If two variables are correlated, then we can do Note that both are completely dependent with each other. variables that are strongly correlated with the target variable, They are said to have a greater effect target variable. Correlation can also be seen Using heatmap. heatmap is one of Visualizing graphs such as histograms, Boxplots that help us know our data Simplicity Because the human mind is so complex To understand data by numbers, they Can easily understand by using pictures.

Based on the color, we can easily differentiate how variables are correlated

Light color- Less correlated

Medium-light color- correlated

Medium color-less strongly correlated

Dark color- strongly correlated

Humidity is closely related to the target variable.

Next to humidity, even rain correlated with the target variable.

atmospheric pressure is variable negatively correlated with a goal, which This means that a price may increase Subtract the other value. In short, we can Let the two variables be opposite Proportional.

Correlation also helps us overcome some value, because it is one of the attribute extraction technology. if two independent variables are strongly correlated with each other, we can remove any one of these variable. there can be no reason Dataset disruption.

Step-6: Splitting Dataset

splitting a dataset into two sets should be corrected. dataset can 80% can be divided in proportion to the train set, 20% test set or 70% train set, 30% test set, or in any other way. split of also affects the accuracy of the dataset training model. Can do a slicing operation is used to separate the datasets.

We have taken care while dividing dataset, assure that the test set should hold Same features as train set and The dataset should also be statistically Meaningful. taking into account the independent variable in 'x' and hence the target variable in 'y'

x = df.iloc [:,:-1].values y = df.iloc [:, 7].values

80% training dataset

20% testing dataset

Make sure the dataset is split into Train and test set before training Ideal. sometimes we can get more accuracy for our model if we include both datasets.

Step-7: Model Training

It has many algorithms machine learning, but we have chosen . Only three of them to train our model. In regression, the accuracy can be measured by Using R2-Score or Mean Squared Error [MSE] or Root Mean Squared Error [RMSE]. model must be imported from sklearn package and then trained. We need to install Sklearn package and then import it. We have defined a

model method to call different models. Model method involves training and validation statements which will train and test the dataset. MSE Price, RMSE value, R2 score is also calculated Model training and testing. test dataset will check whether our trained The model is efficient for real-time data or No.
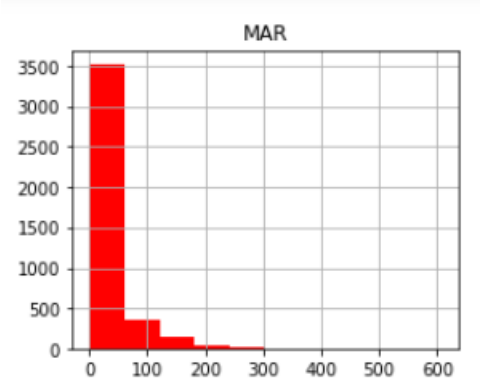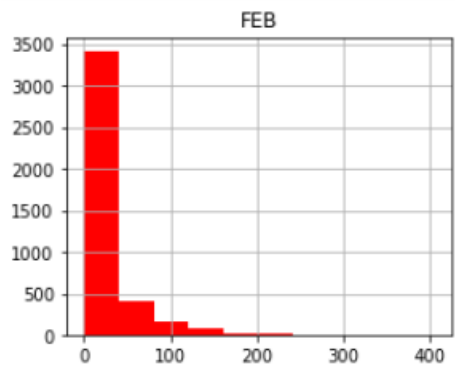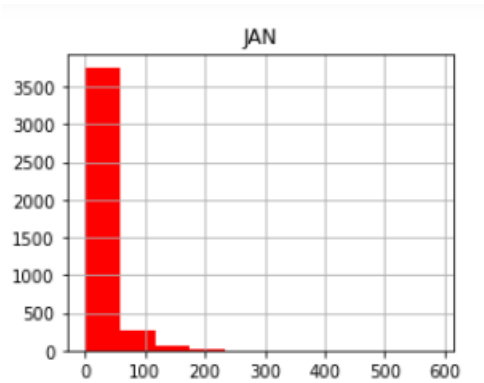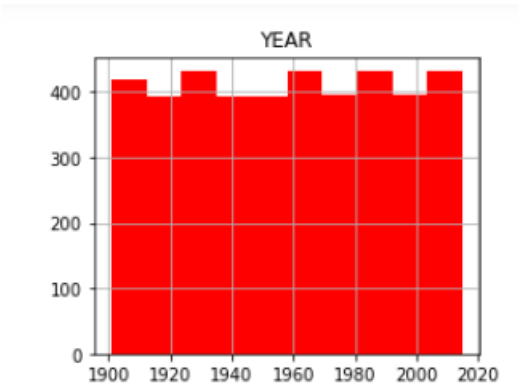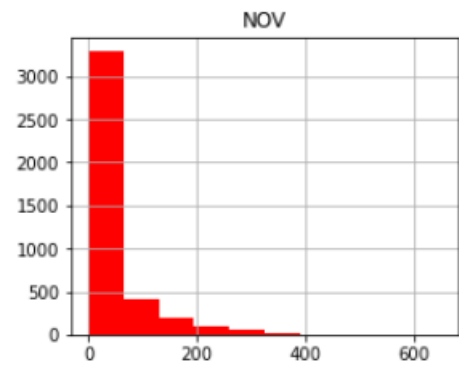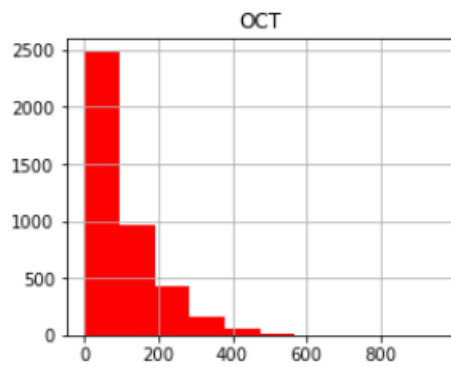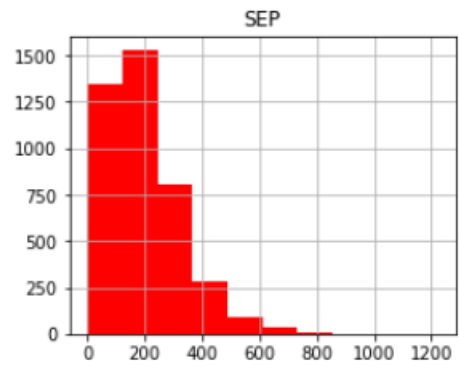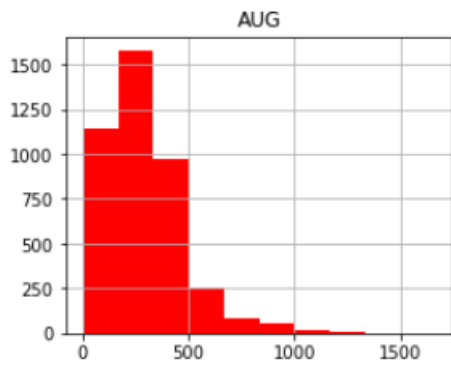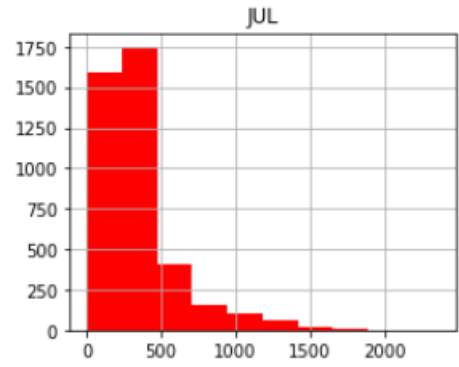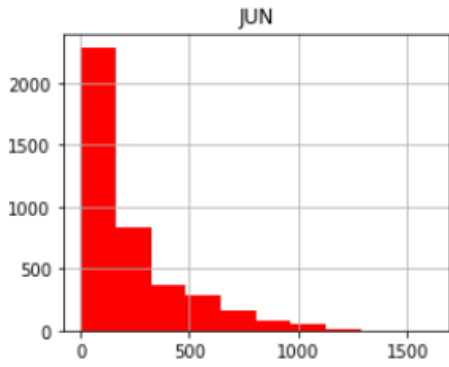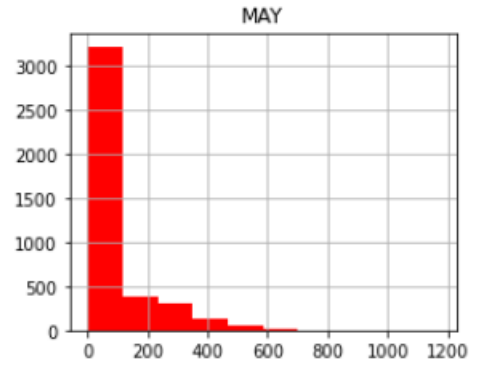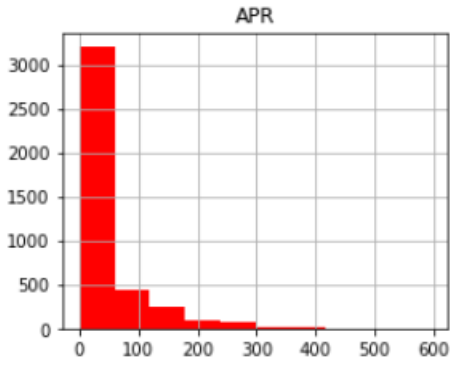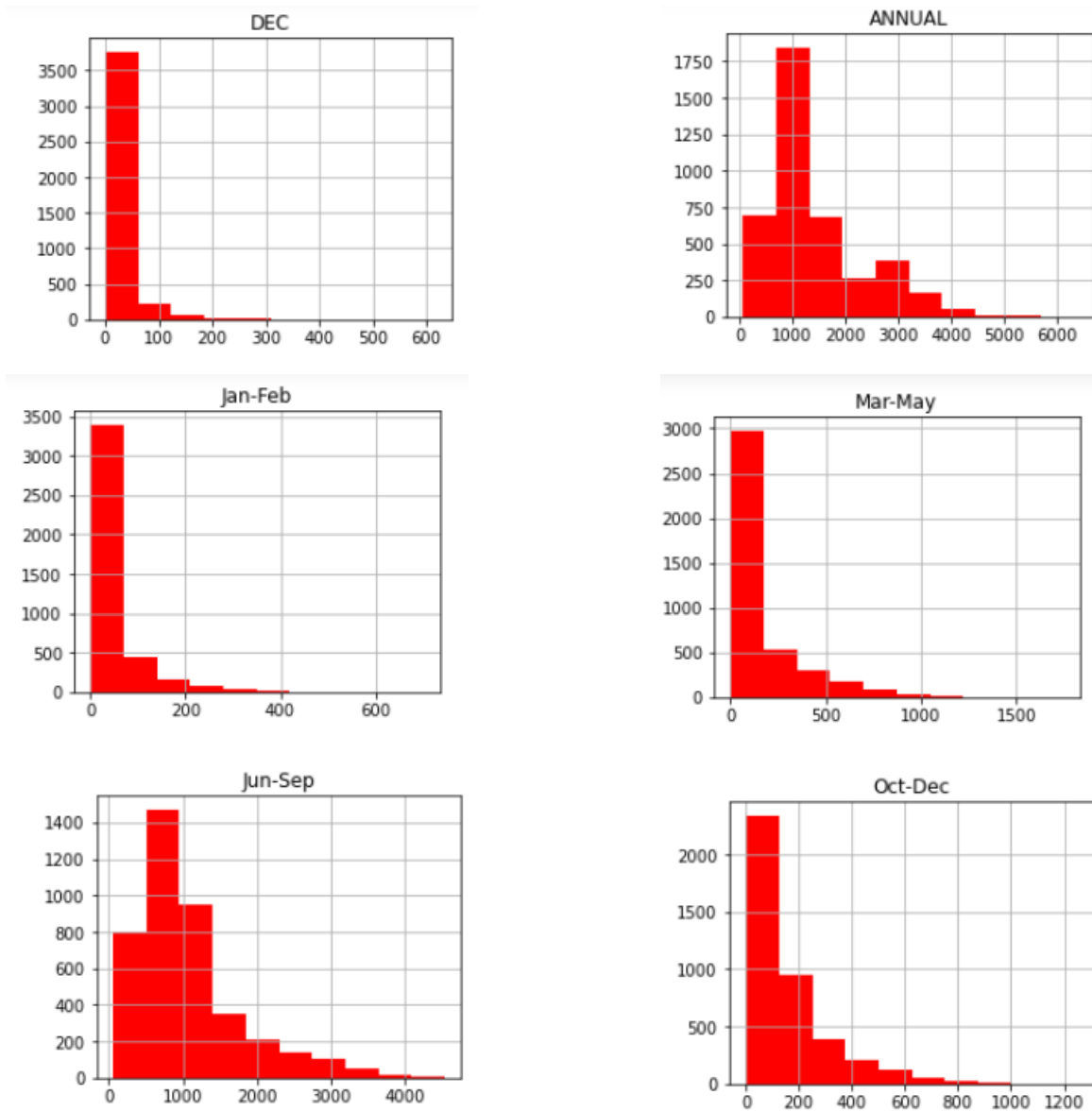
Fig 4.1: HISTOGRAMS

Discussion Of Fig4.1

Above all the histograms shows the distribution of Rainfall Overall Months. After distribution we observed that increasing the amount of rainfall in month of July, August, September.
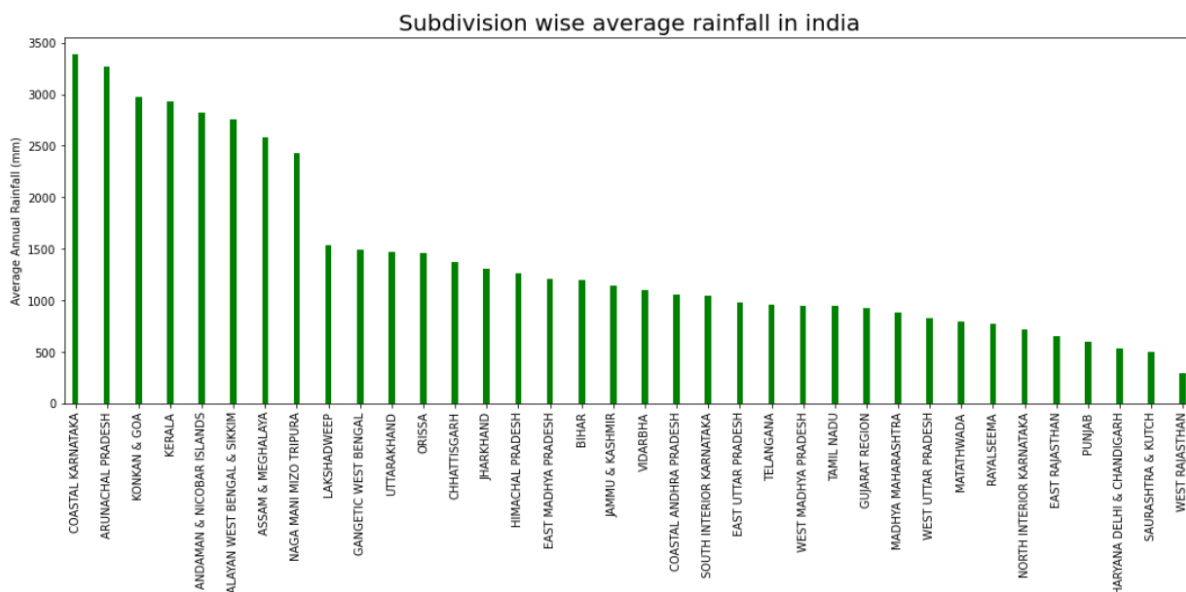
Fig 4.2: SUBIVISON WISE AVERAGE RAINFALL IN INDIA

Discussion Of Fig4.2

After Subdivisions we observed that highest annual rainfall are in the state of "Arunachal Pradesh", "Coastal Karnataka" and "Konkan & Goa" with the estimation annual rainfall of 3418mm, 3408mm and 2977mm respectively and lowest annual rainfall are in the state of "West Rajasthan", "Saurashtra & Kutch" and "Haryana Delhi & Chandigarh" with the estimation annual rainfall of 292mm, 495mm and 530mm respectively.
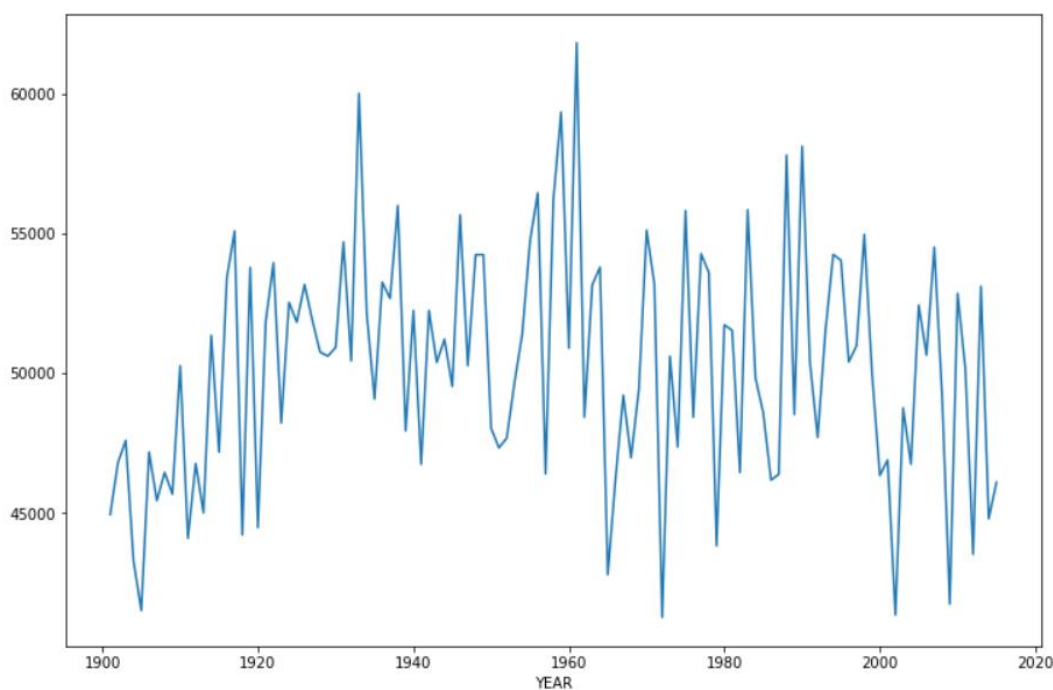
Discussion of Fig4.3

After implementing the line graph we observed that maximum of rainfall is all subdivision is 61815mm that is occurred in the year 1961. Minimum rainfall of all subdivision is 41273mm occurred in the year 1972. Average rainfall of all subdivision is 50182mm.
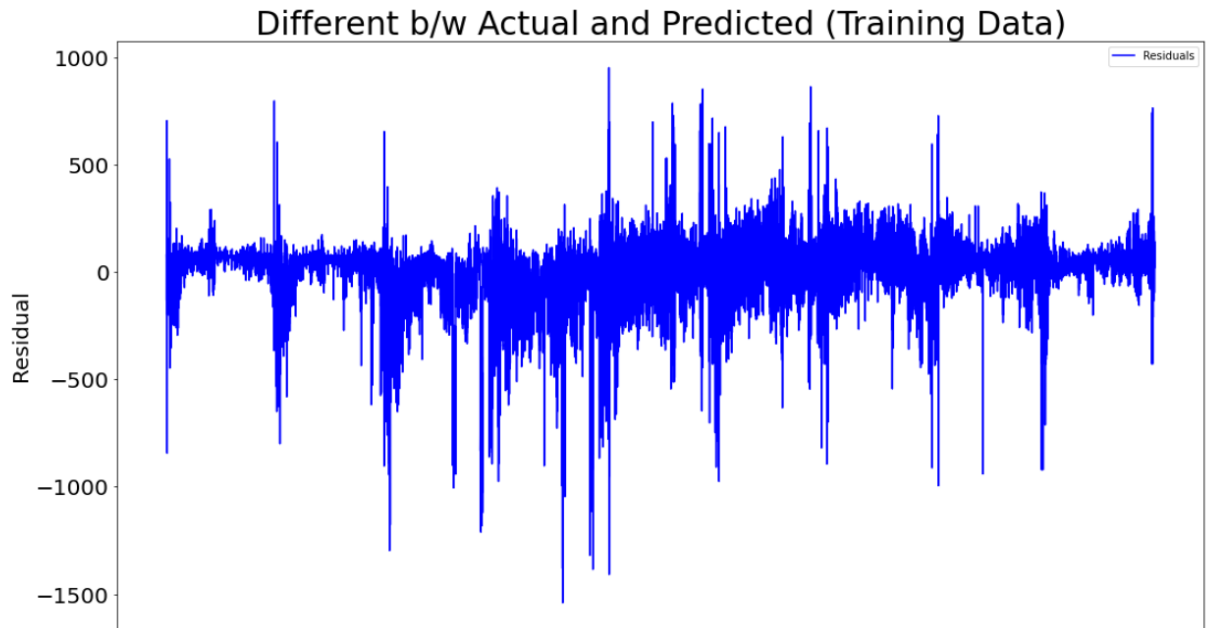

Fig4.4 Actual And Predicted Training Dataset

Discussion Of Fig4.4

To begin with, we will predict the rainfall for the current month with the predictor variable as rainfall over the last three months. The data is arranged in 36810 rows and 4 columns with the first three columns as the predictor variable and the last column being the dependent variable.
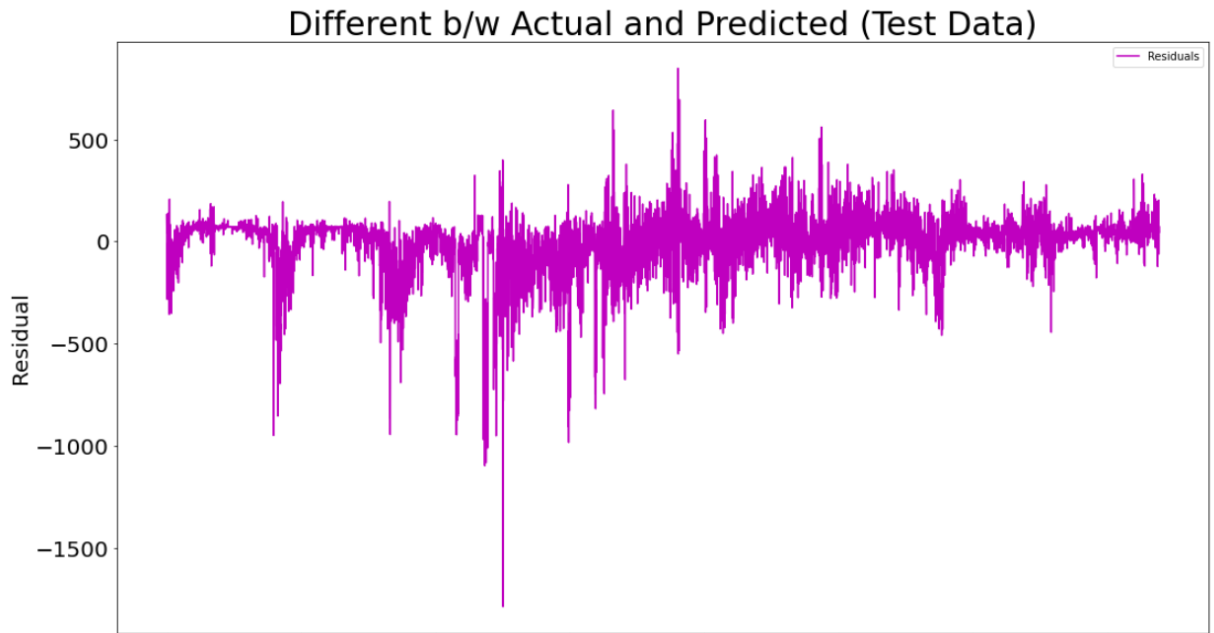
Fig4.5 Actual And Predicted Test Dataset

Discussion Of Fig4.5

For each month from April to December, four columns are additional to the bottommost of the data, producing 36810 rows and 4 columns of data. The training/testing is erratically divided in the ratio of 80:20.
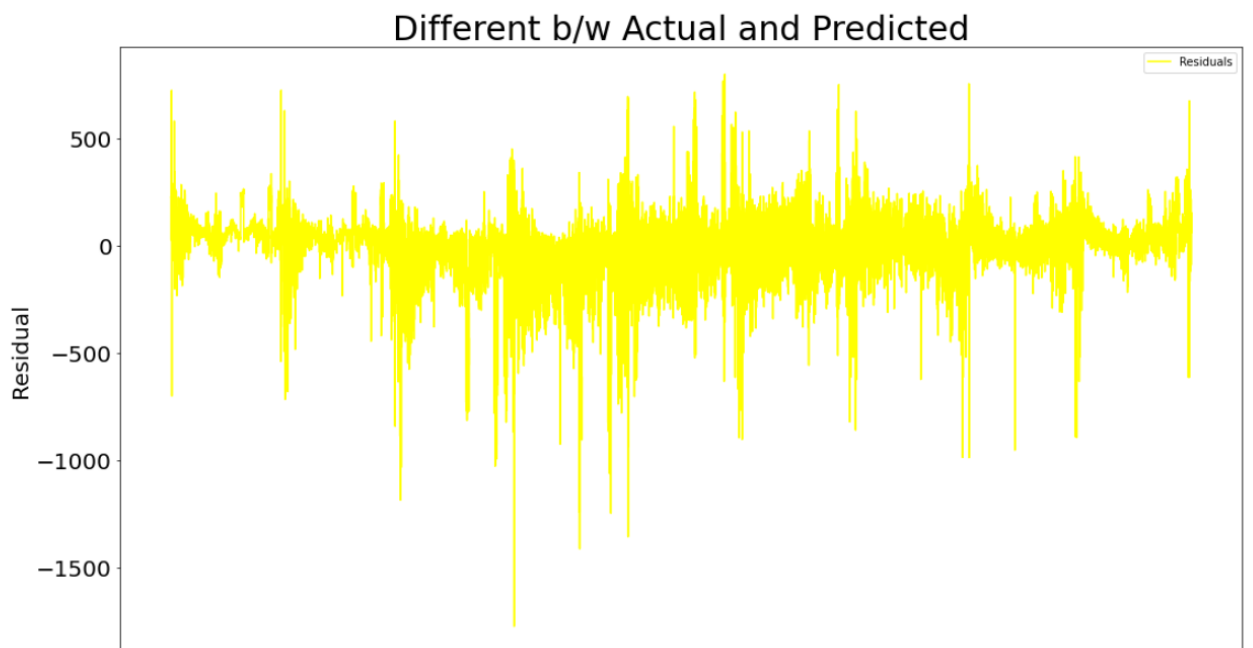
Discussion Of Fig4.6

In this section, we will expect the rainfall for the last three months and the present month with the interpreter variables as rainfall in the subsection. Since the subdivision is a unconditional variable consisting of 36 values, 36 dummy binary variables are added in place of the categorical variable to use in the deterioration.

MAD (Test Data): 90.30271570055083
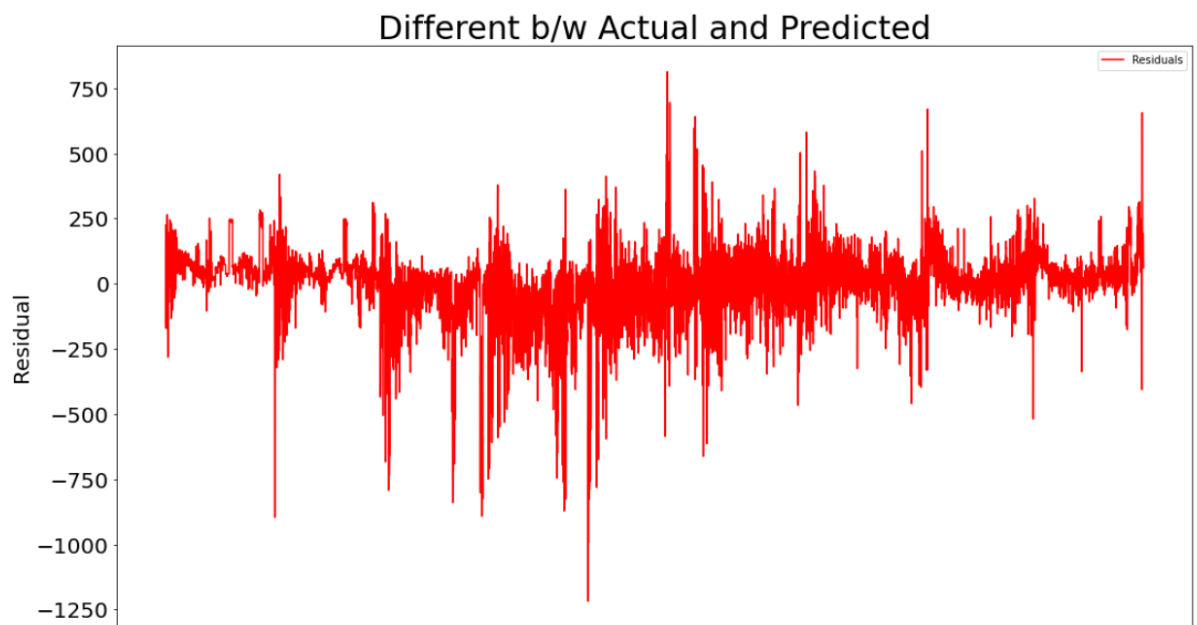
Fig4.7 Actual And Predicted Data

Discussion Of Fig4.7

The data is arranged in 36810 rows and 40 columns with 39 columns as the predictor variable and one column being the dependent variable. Of the 39 variables, three are the rainfall for the last three months and 36 are binary variables used for the subdivisions. For each month from April to December, columns are added to the bottom of the data to produce 36810 rows and 4 columns of data. The training/testing is randomly divided in the ratio of 80:20.

Training On Complete Dataset

Algorithm	MAE
Linear Regression	57.08862331011236
SVR	116.60671510825178
Artificial neural nets	44.329664907381066

Training On Telangana State

Algorithm	MAE
Linear Regression	31.249748674622477
SVR	59.35057496896855
Artificial neural nets	31.0601823988415

Chapter 5

CONCLUSION

5.1 Conclusion

5.2 Applications Of The Project

5.3 Future Scope Of The Project

5.4 Discussion

5.1 Conclusion

The main aim of the project is selection of fixed algorithms for predicting of the rainfall analysis related factors that affecting rainfall analysis it is provided that the random forest regression algorithms can be adaptable strategy for prediction in rainfall analysis. Machine learning provides the better intelligent models instead of traditional models. Regression is the best task of predicting the rainfall analysis and the processing techniques required to produced data set. The project includes various types of preprocessing techniques that is required for preparing our data set. In this project we are implementing different types of the preprocessing techniques that is required for preparing our data set. The data should be free from all kind's unpredictability, overfitting and other heterogenous which may affect the performance of the model. we also survey some regression algorithms that can after the prediction with disturbance in this input data. Relative study has helped us to understand different algorithms successfully. The study specifies large structural and material irregularity in rainfall analysis over India and no clear pattern noticed. A research work to this study could be a analysis of hydroclimatic adaptable for the different regions of the country.

There was a need for the development of a homogeneous spatially and (provisionally) rainfall series for all 36 meteorological subdivisions as well country as a whole. The newly formed rain chain is evenly feast across the nation state and it signifies all the existing districts. While Indian Monsoon rainfall as a whole does not show slightly important trend, with momentous rainfall Trends are observed in certain specific areas. This project is based on the estimation of the rainfall and the conclusion is the SVR is more valuable. The main reason is the helping

the client to manage the is a spread properties in some of the essential factors . For display the SVR the the decision is two fit with the normal .In this type of models we shown by using graph SVR is better than MLR.SVR can catch the non-linearity while on the other hand MLR an catch the linearly. we process MAE for both MLR and SVR run the execution of the process. Tuned expectation models has better expectation. Though, in the present-day study, we measured thirty-four atmospheric sub-divisions for the explore, but to be extra correct, micro near data like district intelligent data should be merged. Then the very tall accuracy micro close administration plan will be accomplished. Even, the lattice sensible rainfall education using identical advanced microwave isolated sensing expertise will be very useful for the developers. The collective machine learning practices, deep learning procedures similar long-short-term memory system can be used to accomplish very tall superiority prognostication statistics. The main reason for the maximum economy of India is the Rainfall. The primary reason for most of us is rainfall. Most of the time the prediction of rainfall is almost wrong and can be recognized from precise precipitation. One is the major machine learning approach. In the first one LASSO regression approach is required for the prediction. This one is more accurate was considered and the prediction was performed. Performing that shows lasso has more accurate as compared to the neural network. An algorithm for predicting rainfall is supposed to be lasso regression. The future scope of this project is as increase the percentage of the errors present. Also, one scope is to decrease the ratio to train the data into test data so that the level of accuracy is improved. When we test the accuracy there is an increase in the complexity of the algorithm. Hence the algorithm used for testing is the basis dataset is more important for further research.

5.2 Applications Of The Project

Rainfall analysis is very important for the information related to the crop. Planning and management for Indian agriculture. India Rainfall analysis is very necessary for crop planning and water resource management. The rainfall analysis suggests rain water harvesting to control soil erosion and maximization of water. India rainfall prediction plays a very important role in the agriculture of India. Rainfall forecasting plays a very important role in flood warning systems. The major variability of rainfall in the forecasting is very difficult. The variables in

the rainfall are due to the distribution of the temporal dimensions. The forecasting of the rainfall is due to the physically-based processes. Development in artificial intelligence and those techniques aimed at pattern recognition. Types of ANNs to forecast the special distribution of rainfall. Rainfall forecasting is a challenging task for flood management in the provision of a quantitative rainfall forecast.

5.3 Future Scope Of The Project

Rainfall Prediction simplify to the historical and present climate changes. Basically, It is a very important for the future forecasting to execute a planning in climate variables. Climate information for the future working on the basis of global scale, region computational techniques. various application on machine learning have gained attention towards which we can make various application regarding the various prediction like weather prediction and rainfall prediction & analysis.

5.1 Discussion

Recommended that the examination of old rainfall trend is very critical in several fields like water source organization, maintainable agricultural development, ecology supervision and strength area. The rainfall information of 115 centuries were used to inspect the inconsistency of yearly and cyclical rainfall in very feature and evaluated the trend in numerous ways for thirty-four atmospheric sub-divisions. The findings of 115 years annual and periodic rainfall difference investigation show that the maximum rainfall dissimilarity was initiate in the sub-divisions of Western India of yearly and seasonal, Western India and South Western India of wintertime, North India and Western India of monsoon and Western, North Western and some fragments of Easterly India of post rainy season rainwater. Although the bottommost distinction was institute in the sub-divisions of North Eastern and Eastern India. The matching product was testified by some previous trainings 32,45,66,71,105. The conclusions of test on twelve-monthly and periodic rainfall statement that thirteen meteorological sub-divisions stood experiential undesirable trend, whereas rest of the sub-divisions were noted constructive trend. Rajeevan and Guhathakurta and Rajeevan46 reported that Jharkhand, Chhattisgarh and Kerala were detected bad leaning, while eight sub-divisions like Gangetic West Bengal, Sea Andhra Pradesh, Kanakan & Goa,

North Interior Karnataka, Rayalessema, Jammu &Kashmir were noted falling trend over time. These grades are absolutely equal with the conclusions of the test of the contemporary study . Mirza testified that the once yearly and regular rainfall was supplementary or fewer dependable in sub-divisions sub-himalayan Bengal and Gangetic Bengal which was by and huge parallel with the discoveries of MK test in the existing training .

REFERENCES

[1]   N. Paudel and T. Nath Yogi, "Comparative Study of Machine Learning Algorithms for Rainfall Prediction A Case Study in Nepal," *International Journal of Advanced Research in Engineering and Technology (IJARET)*, vol. 11, no. 10, pp. 1582–1591, 2020, doi: 10.34218/IJARET.11.10.2020.151.

[2]   D. Ranjan Nayak, A. Mahapatra, and P. Mishra, "A Survey on Rainfall Prediction using Artificial Neural Network," 2013.

[3]   S. Parashar and T. Hurra, "A Study on Prediction of Rainfall Using different Data mining Techniques." [Online]. Available: https://www.researchgate.net/publication/341131681

[4]   A. Parmar, K. Mistree, and M. Sompura, "Machine Learning Techniques For Rainfall Prediction: A Review." [Online]. Available: https://www.researchgate.net/publication/319503839

[5]   B. Praveen *et al.*, "Analyzing trend and forecasting of rainfall changes in India using non-parametrical and machine learning approaches," *Scientific Reports*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-67228-7.

[6]   Ö. Terzi, "Monthly Rainfall Estimation Using Data-Mining Process," *Applied Computational Intelligence and Soft Computing*, vol. 2012, pp. 1–6, 2012, doi: 10.1155/2012/698071.

[7]   N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "Development and analysis of Artificial Neural Network models for rainfall prediction by using time-series data," *International Journal of Intelligent Systems and Applications*, vol. 10, no. 1, pp. 16–23, Jan. 2018, doi: 10.5815/ijisa.2018.01.03.

[8]   K. Sarvani, Y. Sai Priya, C. Teja, T. Lokesh, E. Bala, and B. Rao, "RAINFALL ANALYSIS AND PREDICTION USING MACHINE LEARNING TECHNIQUES," vol. 12, 2021, [Online]. Available: www.jespublication.com

[9]   M. Tanushri, R. C. Poovanna, and A. Professor, "Issue 6 www.jetir.org (ISSN-2349-5162)," 2020. [Online]. Available: www.jetir.org.