**MACHINE LEARNING with PYTHON and IBM Watson Studio**

A summer internship report Submitted in partial fulfillment of the requirements for the award of degree

of

**BACHELOR OF TECHNOLOGY**

in

**COMPUTER SCIENCE AND ENGINEERING**

by

KOTTAVALASA YELLAM NAIDU(16B61A0519)

Under the esteemed guidance of

**Mr.Mohan P**

(M TECH)

SMART BRIDGE





**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SATYA INSTITUTE OF TECHNOLOGY AND MANAGEMENT**

*Vizianagaram,AndhraPradhesh 535001*

*(Under the Management of GayatriVidhyaParishad, Visakhapatnam)*



**CERTIFICATE**

This is to certified that the summer internship entitled “**MACHINE LEARNING with PYTHON and IBM WATSON Studio”** which is a practical ,theoretical work carried out by KOTTAVALASA YELLAM NAIDU (16B61A0519) in partial fulfillment for the award of the degree of **BACHELOR OF TECHNOLOGY** in Department of Computer Science and Engineering ,during the year 2018-2019.The summer internship has been approved as it satisfies the academic requirements.

N.Krishna

Associative Professor &Head Of the Department,CSE

Satya Institute Of Technology and Management.

**Acknowledgement**

I express my sincere thanks and gratitude Head of the Department, **DR.S.PHANIKUMAR**, for his supervision and guidance for internship.

I also express my sincere thanks and gratitude to **MR.P.MOHAN** of smart bridge for sharing their knowledge with us and for their guidance during internship.

**Project Associates**

**Topic Page No**

**1.Title of the project**

1.1 Introduction

1.2 Objectives of Research

1.3 Problem Statement

1.4 Industry Profile

**2.Review of Literature**

**3.Data Collection**

**4.Methodology**

4.1 Exploratory Data Analysis

4.1.1 Lambda Function and Replacing

4.1.2 Histographs

4.1.3 Normalization and Splitting the data

4.1.4 Bargraphs

**5.Conclusion**

**6.Bibilography and Reference**

****

**Summer Internship Program**

**A Case study on Flood Prediction in Kerala**

**Submitted By- Supervisors-**

K.Yellam Naidu

Ch.Dhana Lakshmi

Ch.Pravallika

G.Deepika

Ch.Pavya Sree

Signature of the Supervisor

**ABSTRACT:-**

Floods are the most destructive natural disasters which are highly complex. The research on flood prediction has been contributing to risk reduction, minimizing loss of human lifes and reducing the property damage associated with floods. To minimize the complexicity of mathematical expressions of physical process of floods during past two decades machine learning methods have highly contributed in the advancement of production systems providing better performance and cost effective solutions. Due to the benefits and potentials of ML its popularity has increased researchers by introducing Machine Learning methods have been aiming at discovering more accurate and efficient prediction models. The main contribution is to demonstrate ML models in flood prediction.ML models are used to do the qualitative analysis in order to find the accuracy, effectiveness and speed of various algorithms. As a result, ML models introduces the most prediction methods for both long term and short term floods. Further more the major trends in improving the quality of flood prediction models are investigated. Among them algorithm ensemble and model optimization are reported as most effective strategies in ML models.

**1.1 INTRODUCTION:**

**Machine Learning:**

Machine Learning is a sub-area of Artificial Intelligence. It is a category of algorithm that allows software applications to become more accurate in predicting the outcomes without being explicitly programmed. It mainly focuses on development of computer programs.

The primary aim is to allow the computers automatically.Machine Learning algorithms build a mathematical models based on sample data known as training data, in order to make the predictions or decisions we use Machine Learning.

**Examples of Machine Learning:**

1) Google Search Engine

2) Online Ticket Booking

3) In Education Field

4) Transportation

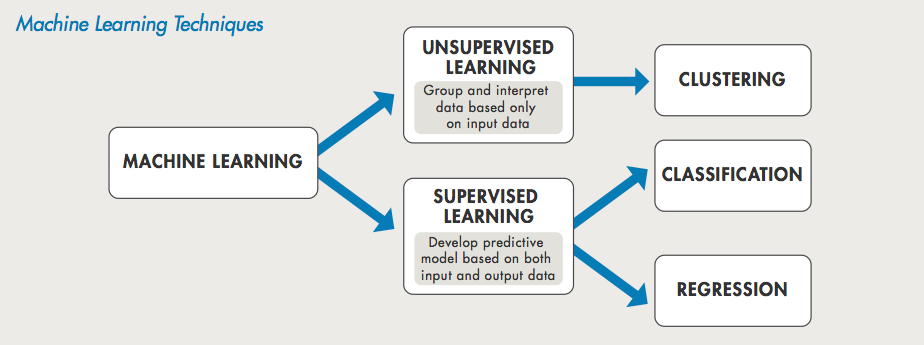
5) Social Media Services

**Need for Machine Learning:**

Machine Learning is a field which is raised out of Artificial Intelligence(AI). But except for few mere tasks such as finding the shortest path between point A and B, we were unable to program more complex and constantly evolving challenges.There was a realisation that the only way to be able to achieve this task was to let machine learn from itself. So machine learning was developed as a new capability for computers. And now machine learning is present in so many segments of technology, that we don’t even realise it while using it.

The data being very massive, the time taken to compute is increased, and this is where Machine Learning comes into action, to help people with large data in minimum time.

The techniques we use for data mining have been around for many years, but they were not effective as they did not have the competitive power to run the algorithms. If you run deep learning with access to better data, the output we get will lead to dramatic breakthroughs which is machine learning.

****

**Types of Machine Learning:**

a)Supervised Machine Learning

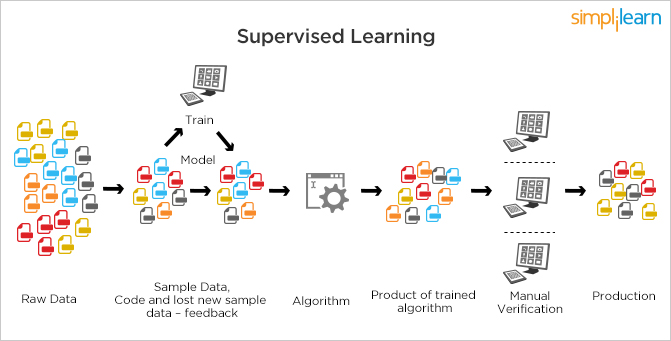
b)Unsupervised Machine Learning

c)Reinforcement Machine Learning

**Supervised Machine Learning:**

A majority of practical machine learning uses supervised learning.

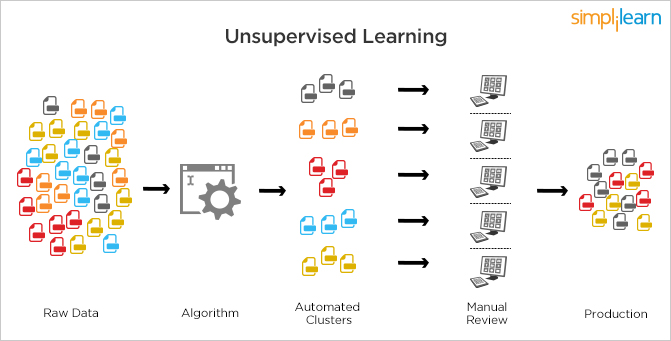
In supervised learning, the system tries to learn from the previous examples that are given. Speaking mathematically, supervised learning is where you have both input variables (x) and output variables(Y) and can use an algorithm to derive the mapping function from the input to the output.The mapping function is expressed as Y = f(X).



Unsupervisied Machine Learning:-

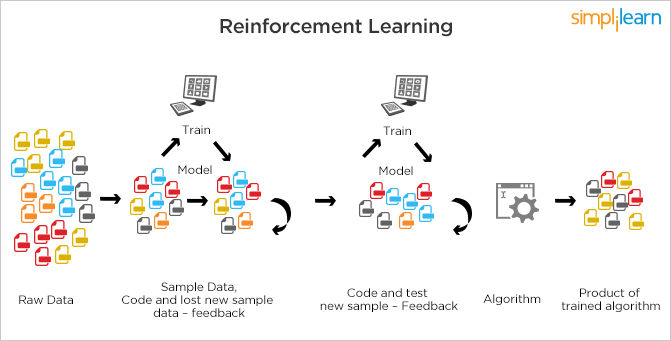
Unlike supervised learning, unsupervised learning is used with data sets without historical data. An unsupervised learning algorithm explores surpassed data to find the structure. This kind of learning works best for transactional data; for instance ,it helps in identifying customer segments and clusters with certain attributes – this is often used in content personalization.

Popular techniques where unsupervised learning is used also include self-organizing maps, nearest neighbor mappig, singular value decomposition, and k-means clustering. Basically, online recommendations, identification of data outliers, and segment text topics are all examples of unsupervised learning.



Reinforcement Machine Learning:-

This is a bit similar to the traditional type of data analysis; the algorithm discovers through trial and error and decides which action results in greater rewards. Three major components can be identified in reinforcement learning functionality: the agent, the environment, and the actions. The agent is the learner or decision-maker, the environment includes everything that the agent interacts with, and the actions are what the agent can do.



Reinforcement learning occurs when the agent chooses actions that maximize the expected reward over a given time. This is best achieved when the agent has a good policy to follow.

Flood forecasting is the use of forecasted precipitation and streamflow data in rainfall-runoff and streamflow routing models to forecast flow rates and water levels for periods ranging from a few hours to days ahead, depending upon the size of the watershed of river basin. Flood forecasting can also make use of forecasts of precipitation in an attempt to extend the time available.

Flood forecasting is an important component of flood warning, where the distinction between the two is that the outcome of flood forecasting is a set of forecast time-profiles of channel flows or river levels at various locations, while “flood warning” is the task of making use of these forecasts to tell decisions on warning of floods.

Real-time flood forecasting at regional area can be done within seconds by using the technology of artificial neural network. Effective real-time flood forecasting models could be useful for early warning and disaster prevention.

**1.2 Objectives of Research**

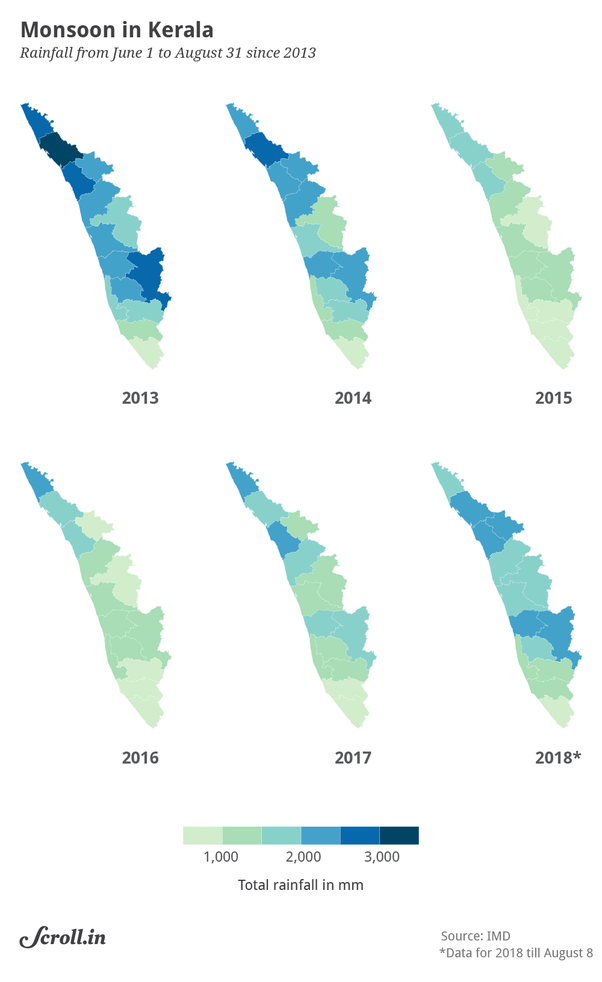
Here we are considering the year 2018, Extreme precipitation events and floods caused losses to human lives and infrastructure have increased under the warming climate. In August 2018, the state of Kerala (India) witnessed large-scale flooding, which affected millions of 10 people and caused 400 or more deaths.The return period of extreme rainfall and the potential role of reservoirs in the recent flooding in Kerala. We show that Kerala experienced 53% above normal rainfall during the monsoon season (till August 21st) of 2018.

Moreover, 1, 2, and 3-day extreme rainfall in Kerala during August 2018 had return periods of 75, 200, and 100 years. Six out of seven major reservoirs were at more than 90% of their full capacity on August 8, 2018, before extreme rainfall in Kerala. Extreme rainfall at 1-15 days duration in August 2018 in the catchments upstream 15 of the three major reservoirs (Idukki, Kakki, and Periyar) had the return period of more than 500 years. Extreme rainfall and almost full reservoirs resulted in a significant release of water in a short span of time. Therefore, above normal seasonal rainfall (before August 8, 2018), high reservoir storage, and unprecedented extreme rainfall in the catchments where reservoirs are located worsened the flooding in Kerala. Reservoir operations need to be improved using a skilful forecast of extreme rainfall at the longer lead time (4-7 days).

Frequent extreme precipitation events cause flooding, which have become common in India. The frequency of great floods and extreme precipitation events has substantially increased under the warming climate, 25 which is consistent with the observations as well as climate model projections. India has witnessed some of the most unprecedented extreme precipitation events that caused flooding and loss of lives in the recent past.

The recent extreme rainfall and widespread flooding in Kerala exemplify the enormity of extreme rainfall and large-scale floods in India. The persistent and extreme rainfall occurred in August 2018 in Kerala affected all the aspects of human lives 10 including socioeconomic conditions, transportation, infrastructure, agriculture, and livelihood. The Kerala flood of 2018 has already attracted attention from the media, scientific community, and policymakers, which is probably the worst flood in a century (The Independent, 16 August 2018). As per the preliminary estimates, the Kerala flood caused the death of more than 440 people (Gulf News, 30th August 2018) and economic damage exceeding $3 billion (News18, 17 August 2018). Despite the state-wide extreme rainfall in Kerala in August 2018, potential causes (heavy rain and reservoir operations) of 15 floods have been greatly debated.

We obtained daily observed rainfall data from India Meteorological Department (IMD) at 0.25° for the period 1901-2018 (till 21st August). IMD rainfall dataset is developed using more than 6000 observing gauge stations across India, and a substantial number of stations are located in Kerala. We took the average rainfall rate of every month from January to December. Based on the average rainfall rates, the target variable i.e; the flood is predicted by using machine learning techniques.



**1.3 PROBLEM STATEMENT**

Here we have considered a dataset which is related to flood prediction. It consists of features (inputs or attributes) of every month average rainfall rate which are obtained from IMD. Depending on these values we have to predict whether the flood occurs or not. In order to solve this problem we are using the Machine Learning techniques such as regression and classification.

**1.4 INDUSTRY PROFILE**

Rainfall over Kerala during southwest monsoon season 2018 (1 June to 19 August, 2018) has been exceptionally high. Kerala so far received 2346.6 mm against normal of 1649.5 mm (above normal by 42%). The spatial distribution of district-wise seasonal rainfall is shown in Annexure 1. It indicates that highest excess rainfall is recorded over Idukki District (92% above normal) followed by Palakkad (72% above normal).

**2.REVIEW FROM LITERATURE**

**2.1 Floods**

A flood is an overflow of an expanse of water that submerges land. The EU Floods directive defines, a flood as a temporary covering by water of land not normally covered by water in the sense of flowing water", the word may also be applied to the inflow of the tide. Flooding may result from the volume of water within a body of water, such as a river or lake, which overflows or breaks levees, with the result that some of the water escapes its usual boundaries. While the size of a lake or other body of water will vary with seasonal changes in precipitation and snow melt, it is not a significant flood unless the water covers land used by man like a village, city or other inhabited area, roads, expanses of farmland, etc. Floods are common and costly natural disasters. Floods usually are local, short-lived events that can happen suddenly, sometimes with little or no warning. They usually are caused by intense storms that produce more runoff than an area can store or a stream can carry within its normal channel. Rivers can also flood when dams fail, when ice jams or landslides temporarily block a channel, or when snow melts rapidly. In a broader sense, normally dry lands can be flooded by high lake levels, by high tides, or by waves driven ashore by strong winds. Small streams are subject to floods (very rapid increases in runoff), which may last from a few minutes to a few hours. On larger streams, floods usually last from several hours to a few days. A series of storms might keep a river above flood stage (the water level at which a river overflows its banks) for several weeks.

**2.2 Weather Patterns**

Floods can occur at any time, but weather patterns have a strong influence on when and where floods happen. Cyclones, or storms that bring moisture inland from the ocean, can cause floods. Thunderstorms are relatively small but intense storms that can cause floods in smaller streams. Frontal storms form at the front of large, moist air masses moving across the country and can cause floods. Hurricanes are intense tropical storms that can cause floods.

**2.3 Very Large Scale Floods**

The size, or magnitude, of a flood is described by a term called recurrence interval. By studying a long period of flow records for a stream, it is possible to estimate the size of a flood that would, for example, have a 5-year recurrence interval (called a 5-year flood). A 5-year flood is one that would occur, on the average, once every 5 years. Although a 100-year flood is expected to happen only once in a century, there is a one per cent chance that a flood of that size could happen during any year. The magnitude of floods can be altered if changes are made in a drainage basin. Harvesting timber or changing land use from farmland to housing developments can cause the runoff to increase and cause an increase in the magnitude of flooding. Building dams that store water can reduce the magnitude of floods.

**2.4 Causes of Floods**

Flooded areas of land usually start off as very dry land. Floods are caused by heavy rains that pour too much water into rivers and other waterways. Making these natural channels unable to carry all the water rising water flows over or breaks the banks to the waterways causing the surrounding land to be flooded. Different causes of floods can come from masses of snow melting of tidal waves. Floods are causes not only by rain but also by human changes to the surface of the earth. Forming deforestation, and urbanization increase the runoff from rains thus storms that previously would have caused no flooding today inundate vast areas. The reckless building in vulnerable areas, poor watershed management, and failure to control the flooding also help create the disaster condition.

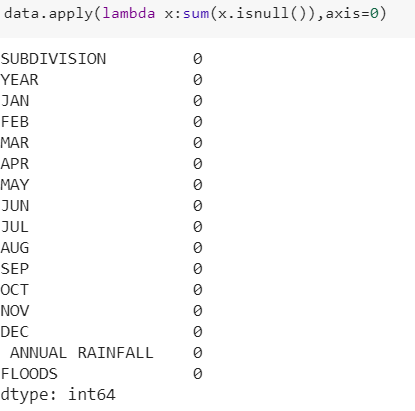
**3.DATA COLLECTION**

The data has been collected from IMD weather forecasting. The attributes of the data are months and their average rainfall rate according to the corresponding years. The predictor of the dataset is whether the flood occurs or not represented in the form of “yes” or “no”. Depending on the each month particular average rainfall rates, the annual rainfall rate is taken. Weather forecasting is the application of science and technology to predict the conditions of the atmosphere for a given location and time.  Weather forecasts are made by collecting quantitative data about the current state of the atmosphere at a given place and using meterology to project how the atmosphere will change. Depending upon the different parameters of weather the average rainfall has been noted.

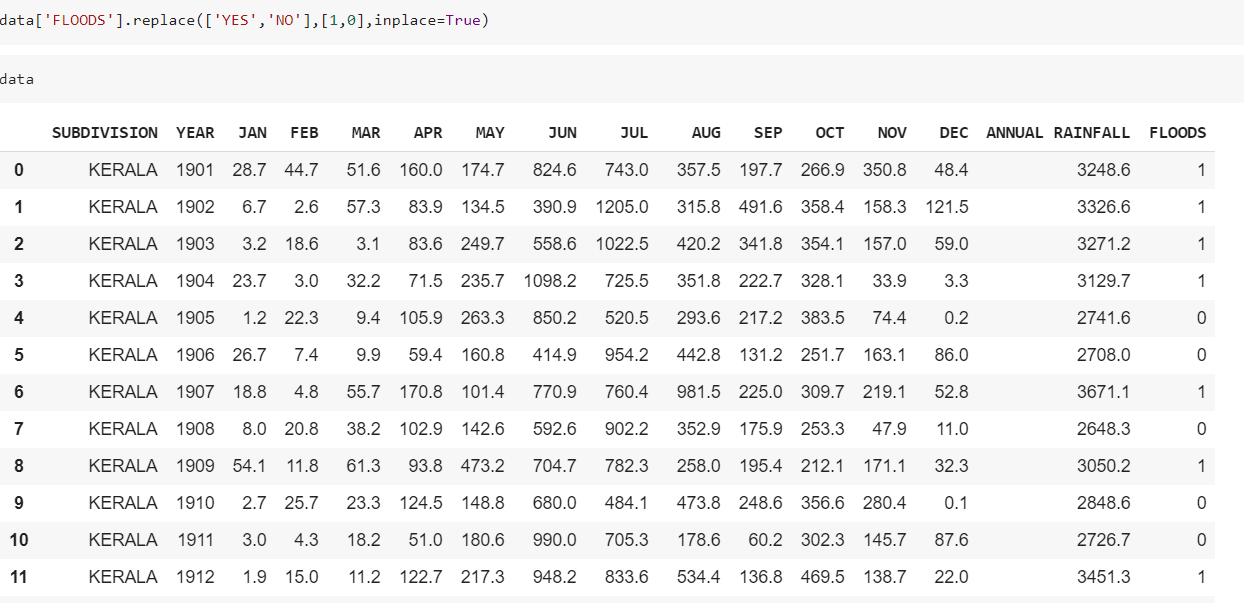
**4. METHODOLOGY**

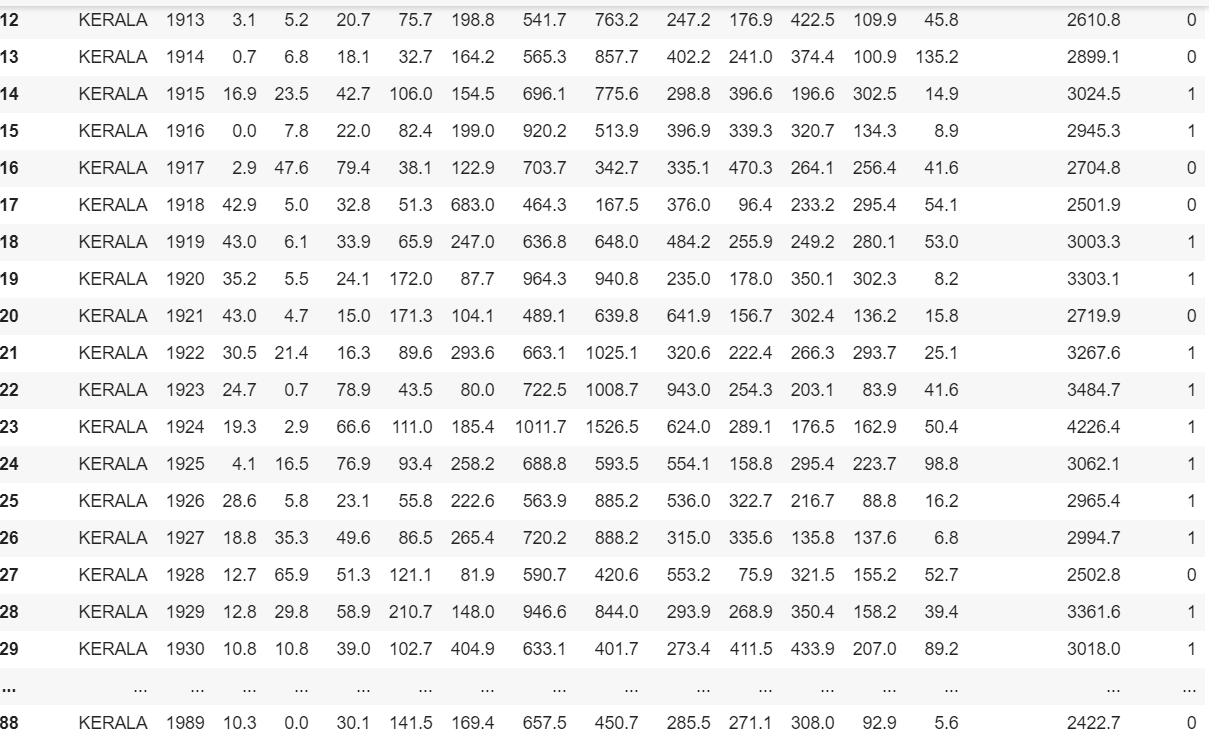
**4.1 Exploratory Data Analysis**

4.1.1



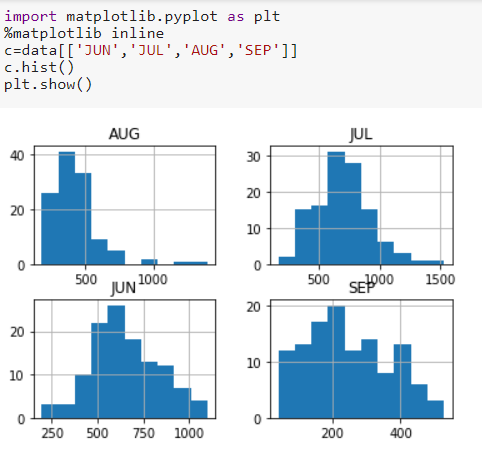
In the data we are applying a lambda, lambda means we have to find a unknown values, isnull function returns Boolean if the data contains missing values then return true otherwise it return false





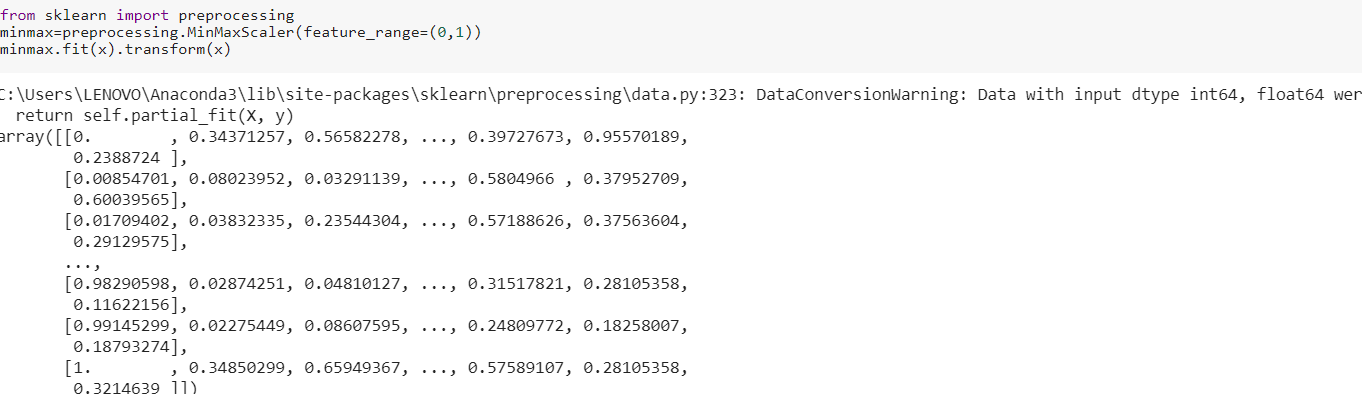
In this we are replacing “yes” with 1 and “no” with 0.The syntax inplace is equal to true checks if the value is yes then it is replaced with 1 or else with 0.

4.1.2

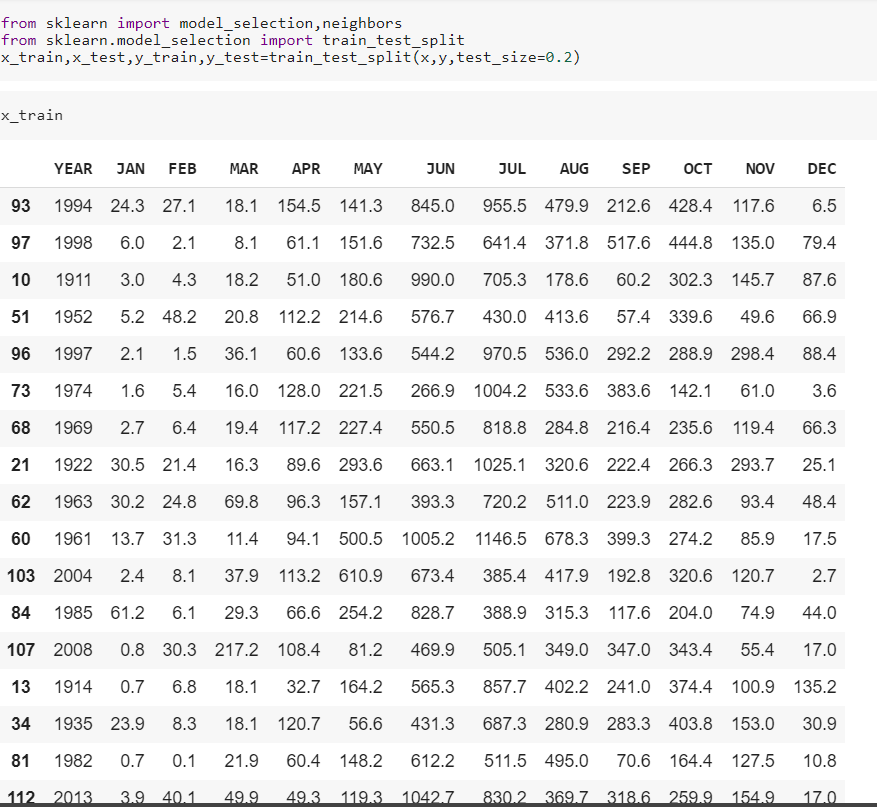


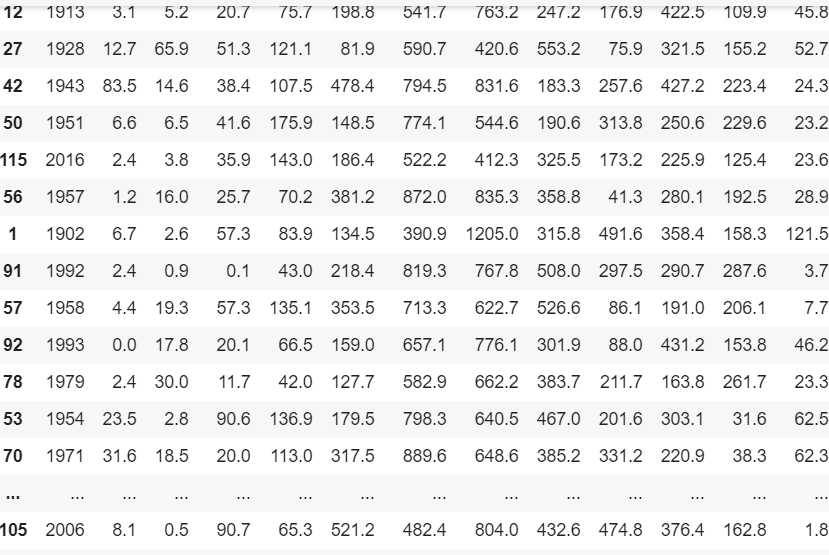
In this data we have imported the package matplotlib.We have drawn histograms for four months(Aug,Jul,Jun,Sep) because these four months cointains a heavy rainfall when compared to the other months

4.1.3



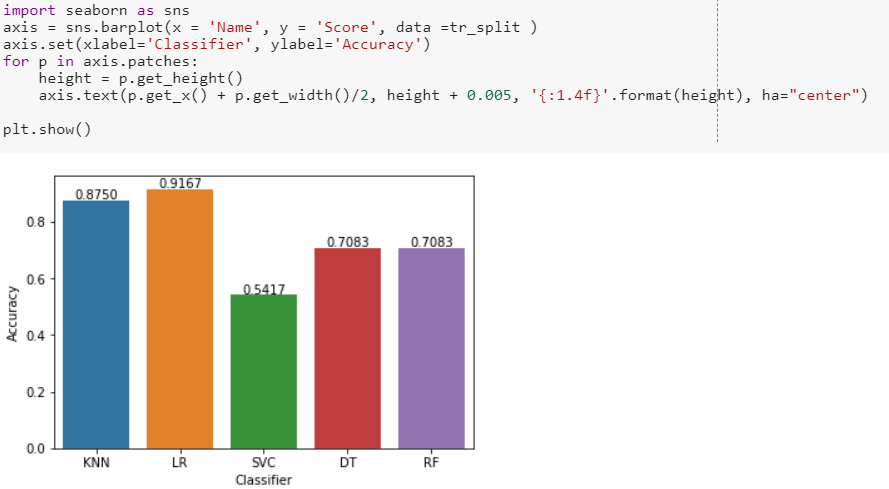
Here we are doing the Normalization process because the data we have taken is the raw data.Whenever we plot this we will get a graph with rises and falls. In order to get the bell shaped curve, we do the Normalization process.Here we did the Normalization process using the MinMax scaler.





In this ,we had splited the data into two parts like training data and testing data we have to take a variables like x and y so we have to train and test the data. First we have to the import the model\_selections from sklearn and also import the test\_train\_split.

4.1.4



From the above graph logistic regression has a more accuracy than others. So the logistic regression is the better model for the prediction.

**5. CONCLUSION**

The above data set contains annual rainfall values of kerala in different years. As we have observed that the annual rainfall in 2018 is very high as compared to other years. We have drawn histograms by taking different months to observe the variations in annual rainfall in different years. We have performed different classifications for the given data set. They are KNN Classifier, Logistic Regression, Support Vector Classification, Decision Tree, Random forest. We calculated the accuracy by using different classifications. We have drawn one barplot to compare accuracy of different classifications. From that barplot we have concluded that the accuracy obtained from the logistic regression is high .So we prefer logistic regression to predict the value of annual rainfall by using previous values.

**6. BIBILOGRAPHY AND REFERENCES**

1.WilliRichert and Luis Pedro Coelho “To get the knowledge about machine learning using python”

2.<https://www.coursera.org/learn/machine-learning-with-python>

3.<https://www.datacamp.com/tracks/machine-learning-with-python>

4.<https://www.geeksforgeeks.org/introduction-machine-learning-using-python>

5.RaulGaretta,GuillermoMoncecchi “this book is used for learning scikit-learn.

6.https://www.drivespark.com/off-beat/kerala-floods-details-facts-precautions-while-driving-026190.html7.John Paul Mueller and Luca Massaron “To learn Machine learning.