

**SUPERVISED BY DONE BY**

PRADEEPTHI MADAM P.SANTHOSH

K.ANISH

M.GIRISH KUMAR

T.NIKHIL TEJA

**CERTIFICATE**

This is to certify that P.Santhosh,T.Nikhil Teja,M.Girish Kumar,K.Anish enrolled in the B.tech degree programme(Computer Science Engineering) of the JNTUH College of Engineering ,Jagtial and Arora Engineering College respectively ,has successfully completed the four week internship cum hands-on training program conducted by Smartbridge at stanley engineering college in ‘Introduction to Machine Learning using Python 3’ during the time period from 3rd June, 2019 to 21st June, 2019 under the guidance of Pradeepthi madam, Speaker. During this period of internship with us he was found punctual, hardworking and inquisitive.

**MENTORS**

Raviteja

Manthan

**ABSTRACT**

Rainfall becomes a significant factor in agricultural countries

Like India. Rainfall prediction has become one of the most

Scientifically and technologically challenging problems in the

World. A wide variety of rainfall forecast methods are

available. There are mainly two approaches to predict

rainfall.They are Empirical method and dynamical method.

The empirical approach is based on analysis of historical data

of the rainfall and its relationship to a variety of atmospheric

and oceanic variables over different parts of the world. The

most widely  used  empirical approaches used  for  climate prediction are regression, artificial neural network, fuzzy logic

and group  method of  data  handling.  This paper uses machine learning techniques for rainfall prediction.

Rainfall Prediction

**INDEX**

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 (Figures &amp; Tables)

4.1.2 (Figures &amp; Tables)

4.1.3 (Figures &amp; Tables)

4.1.4 (Figures &amp; Tables)

5. Findings and Suggestions

6. Conclusion

7. Bibliography and References

**INTRODUCTION**

Accurate forecasting of rainfall has been one of the most important issues in hydrological research because early warnings of severe weather can help prevent casualties and damages caused by natural disasters, if timely and accurately forecasted.

To construct a predictive system for accurate rainfall, forecasting is one of the greatest challenges to researchers from diverse fields such as weather data mining ,environmental machine learning , operational hydrology (Li and Lai), and statistical forecasting . A common question in these problems is how one can analyse the past and use future prediction. The parameters that are required to predict rainfall are enormously complex and subtle even for a short term period.

Physical processes in rainfall are generally composed of a number of sub-processes. A accurate modelling of rainfall by a single global model is sometimes not possible.

To overcome this difficulty, the concept of modular modelling and combining different models has attracted more attention recently in rainfall forecasting. In modular models, several sub-processes are first identified, and then separate models (also called local or expert models) are established for each of them .

So far, various modular models have been proposed, depending on soft or hard splitting of training data. Soft splitting means that the dataset can be overlapped, and the overall forecasting output is the weighted average of each local model .

In the hard splitting, there is no overlap of data and the final forecasting output is derived explicitly from only one of the local models . The approach of combining several models is also knownas ensemble modelling. The basic idea behind the ensemble model is to build several different models for the same process and to integrate them together

In this article, we make a comparison of several machine learning methods of forecasting an average daily and monthly rainfall in particular area . All the methods are coupled with data-preprocessing techniques. Prior to applying the methods, an input selection techniques are used. For the modelling of the rainfall, a Random forest regression is proposed.

**Random forests** or **random decision forests** are an ensembel learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

**OBJECTIVES OF RESEARCH**

The main objective of research is to estimate the rainfall prediction using machine learning algorithms.

**Primary Objectives**

* To study and understand the process of estimation of rainfall
* To know the methodologies applied for the estimation
* To make user easily understand its implementation

**Scope of the study**

The project scope involves the study of climate conditions at a particular area for a long period of time and predicting the chance of rainfall at that particular area.

This project will not only predict rainfall, it will also predict’s the rate in which it would rain .

**Managerial Usefullness of Study**

* It is usefull to farmers for deciding the crop that suits the predicted rainfall
* Informs the costal area people not to go in to sea
* Warns government if it is going to rain in high rate
* It is also used in air traffic

**PROBLEM STATEMENT**

The proposed research aims to develop a prediction model of the rain induced specific attenuation, in order to enhance the capability of communication services ,and for future systems at higher frequency bands. The mathematical models to be developed will be based on the experimental data on rain drop size distribution collected by disdrometer. This problem can be divided into following subproblems.

* To Analyse a Rain Rate Exceedance Characteristics
* Monthly basis
* Annual basis
* To Develop Rain Drop-Size Distribution (DSD) Model.
* Station wise Analysis
* Combined Analysis for all stations (Integrated DSD Model)
* To Develop Extinction Cross-Section Model
* Modeling of Extinction Cross-Section
* Modeling of Coefficients of Extinction Cross-Section
* To Propose Rain Induced Specific Attenuation model.

The outcome of the research will provide necessary input for design ingreliable communication links at millimeter wave frequencies in Indian region. The main contribution of this research is to develop specific rain attenuation model using integrated DSD model and extinction cross-section model with different rain rate exceedance characteristics.

**INDUSTRY PROFILE**

**Weather forecasting** is the application of science and technology to predict the conditions of the atmosphere for a given location and time. People have attempted to predict the weather informally for millennia and formally since the 19th century. Weather forecasts are made by collecting quantitative data about the current state of the atmosphere at a given place and using meteorology to project how the atmosphere will change.

In the past, the human forecaster was responsible for generating the entire weather forecast based upon available observations. Today, human input is generally confined to choosing a model based on various parameters, such as model biases and performance. Using a consensus of forecast models, as well as ensemble members of the various models, can help reduce forecast error.However, regardless how small the average error becomes with any individual system, large errors within any particular piece of guidance are still possible on any given model run.Humans are required to interpret the model data into weather forecasts that are understandable to the end user. Humans can use knowledge of local effects that may be too small in size to be resolved by the model to add information to the forecast. While increasing accuracy of forecast models implies that humans may no longer be needed in the forecast process at some point in the future, there is currently still a need for human intervention

**AIR TRAFFIC**

Because the aviation industry is especially sensitive to the weather, accurate weather forecasting is essential. Fog or exceptionally low ceilings can prevent many aircraft from landing and taking off. Turbulence and icing are also significant in-flight hazards.Thunderstorms are a problem for all aircraft because of severe turbulence due to their updrafts and outflow boundaries, icing due to the heavy precipitation, as well as large hail, strong winds, and lightning, all of which can cause severe damage to an aircraft in flight. Volcanic ash is also a significant problem for aviation, as aircraft can lose engine power within ash clouds. On a day-to-day basis airliners are routed to take advantage of the jet stream tailwind to improve fuel efficiency. Aircrews are briefed prior to takeoff on the conditions to expect enroute and at their destination.Additionally, airports often change which runway is being used to take advantage of a headwind. This reduces the distance required for takeoff, and eliminates potential crosswinds.

**Marine**

Commercial and recreational use of waterways can be limited significantly by wind direction and speed, wave periodicity and heights, tides, and precipitation. These factors can each influence the safety of marine transit. Consequently, a variety of codes have been established to efficiently transmit detailed marine weather forecasts to vessel pilots via radio, for example the MAFOR (marine forecast) Typical weather forecasts can be received at sea through the use of RTTY, Navtex and Radiofax.

### Agriculture

Farmers rely on weather forecasts to decide what work to do on any particular day. For example, drying hay is only feasible in dry weather. Prolonged periods of dryness can ruin cotton, wheat, and corncrop. While corn crops can be ruined by drought, their dried remains can be used as a cattle feed substitute in the form of silage.Frosts and freezes play havoc with crops both during the spring and fall. For example, peach trees in full bloom can have their potential peach crop decimated by a spring freeze. Orange groves can suffer significant damage during frosts and freezes, regardless of their timing.

**Review of literature**

**APPROACHES FOR RAINFALL PREDICTION**

Numerical rainfall prediction is the prediction of weather phenomena by the numerical solution of the equations governing the motion and changes of condition of the atmosphere. Numerical rainfall prediction techniques, in addition to being applied to short-range rainfall prediction, are used in such research studies as water-pollutant.

The first operational numerical rainfall prediction model consisted of only one layer and therefore it could model only the temporal variation of the mean vertical structure of the atmosphere. Computers now permit the development of multilevel (usually about 10–20) models that could resolve the vertical variation of the wind, temperature and moisture. These multilevel models predict the fundamental meteorological variables for large scales of atmosphere

**GAPS IN EXISTING RESEARCH**

After a comprehensive study made on the existing literature, a lot of limitations/gaps have been found in the area of rainfall prediction

• Majority of work reported for rainfall prediction problems has been done using various statistical methods like Curve Fitting, Regression Analysis, ARIMA model etc. which have their own limitations. Hence a more attention is required towards a new approach for rainfall prediction.

• It may conclude from the literature survey that a lit bit attention has been paid in South Western monsoonal seasonal precipitation prediction. Hence a more research is required towards this.

• Most of the works reported on rainfall prediction has paid focus on objective rainfall prediction system, not so much attention has been given for the subjective rainfall prediction which produces more accurate results.

• Most of work with the rainfall prediction using soft computing is done. There is limited work towards hybridization of Neural Network and Fuzzy System in weather forecasting. Hence more emphasis is required towards it.

**CONCLUSION OF LITERATURE REVIEW**

From the survey of literature, it is concluded that soft computing techniques especially Neural Network and Fuzzy System has become interesting preference for researchers to solve rainfall prediction problems. Development of hybridization of Neural Network and Fuzzy System are still the major issues related to rainfall prediction which include forecasting of Local Monsoonal Precipitation (LMP). Therefore, in the present work, weather forecasting system for LMP problem with very good performance measures including very less Root Mean Square Error (RMSE) have been considered. An attempt has been made to develop hybrid algorithm that is based on combination of powers of two algorithms Neural Network and Fuzzy System for weather forecasting of Local Monsoonal Precipitation has been done.

**Data Collection**

We are predicting rainfall on basis of Temperature, Dew Point, Humidity, Sea Level Pressure, Visibility, Wind of all average values.

**Temperature**

The average temperature of the air as indicated by a properly exposed thermometer during a given time period, usually a day, a month, or a year.

Atmospheric temperature is a measure of temperature at different levels of the Earth's atmosphere. It is governed by many factors, including incoming solar radiation, humidity and altitude. Air temperature is the intensity aspect of sun's energy that strikes the earth's surface. Because the amount of energy from the sun reaching the earth varies from day to day, from season to season, and from latitude to latitude, temperatures also vary. The earth as a whole receives a constant flow of radiant short-wave energy from the sun. The earth also radiates long-wave energy to space. During the day, the flow of short-wave radiation absorbed exceeds long -wave energy emitted, and the surface temperature increases.

**Wind**

Wind speed is measured using an anemometer, which is then recorded on the Beaufort scale.

The perceptible natural movement of the air, especially in the form of a current of air blowing from a particular direction.

Wind is the flow of gases on a large scale. On Earth, wind consists of the bulk movement of air. Wind is caused by differences in pressure. When a difference in pressure exists, the air is accelerated from higher to lower pressure. Wind speed is affected by a number of factors and situations, operating on varying scales. These include the pressure gradient, Rossby waves and jet streams, and local weather conditions. There are also links to be found between wind speed and wind direction, notably with the pressure gradient and surfaces over which the air is found

**Humidity**

A quantity representing the amount of water vapor in the atmosphere

Humidity is a term used to describe the amount of water vapor in a mixture of air and water vapor. It is defined as the ratio of the partial pressure of water vapor in the air water mixture to the saturated vapor pressure of water at the prescribed temperature. The relative humidity of air depends not only on temperature but also on the pressure of the system of interest. Relative humidity is often used instead of absolute humidity in situations where the rate of water evaporation is important, as it takes into account the variation in saturated vapor pressure.

**Sea Level Pressure**

The atmospheric pressure, at any elevation, reduced by formula to a value approximating the pressure at sea level.

**Visibility**

The distance one can see as determined by light and weather conditions.

**Dew Point**

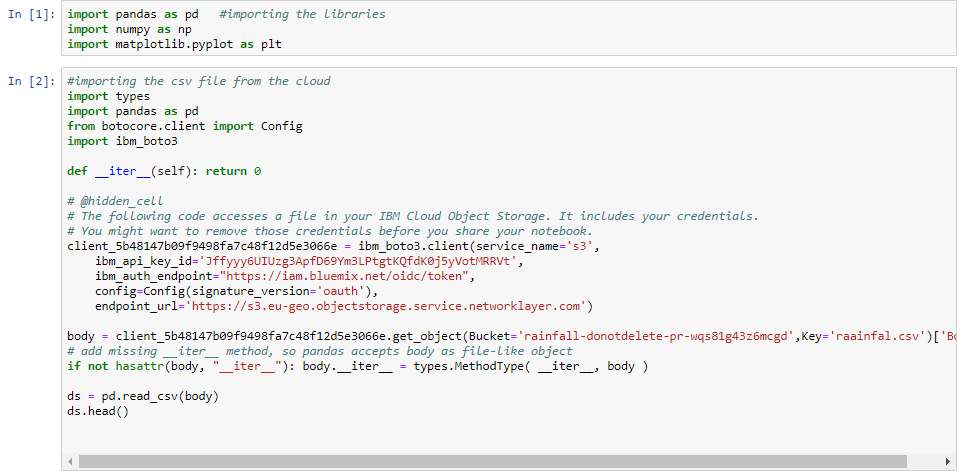
The atmospheric temperature (varying according to pressure and humidity) below which water droplets begin to condense and dew can form.

**Methodology**

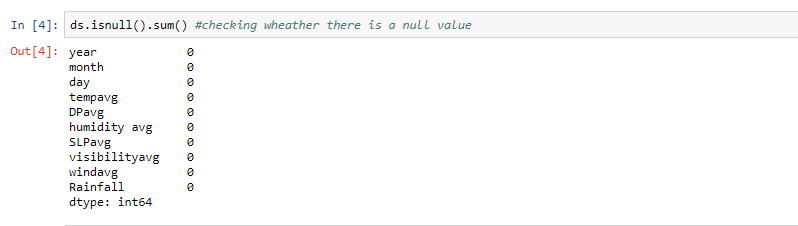
Rainfall prediction  is the application of science and technology to predict the conditions of the atmosphere for a given date and year. People have attempted to predict the weather informally for millennia and formally since the 19th century .

Rainfall predictions are made by collecting quantitative data about the current state of the atmosphere at a given place and using meteorology to project how the atmosphere will change.

**Explanatory Data Analysis**

****

* Here we are importing all packages like “pandas” and “numpy” for data manipulation and “matplotlib” for visualization
* We have also inserted dataset which we collected ,to the program

****

* Here we are checking weather there are any null values in our dataset
* We didn’t have any null values in our dataset
* If there are any null values we have to fill them with some value
* That datatype of value should match the datatype of other element of it’s column.

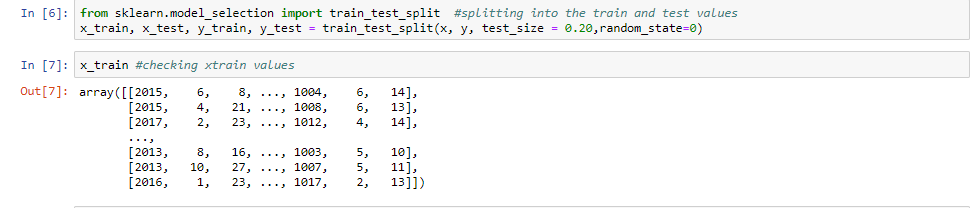


* We are separating all independent data and dependent data.
* In this case dependent data is rainfall.
* Independent elements in this dataset are parameters
* Temperature
* Humidity
* Dew point
* Sea level pressure
* Wind

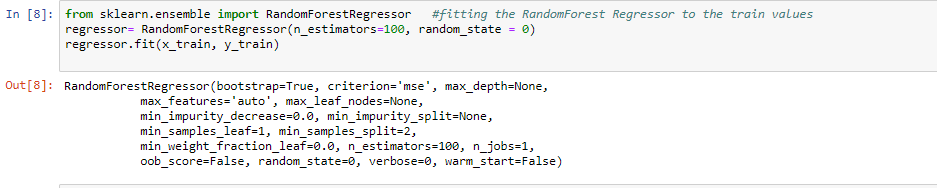
**Model Building**

We have three steps in model building .They are

1. Analysing data.
2. Checking any null values in the dataset or not.
3. Performing train, test, split.
4. Model selection.
5. Predicting the datavalue.
6. Checking the accuracy.
7. Visualization.



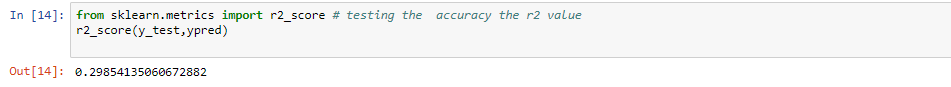
* Here we are dividing the data into test data and train data
* For this we are importing model\_selection package
* At first we train our model with test data
* Then we test how efficiently model got trained using test data



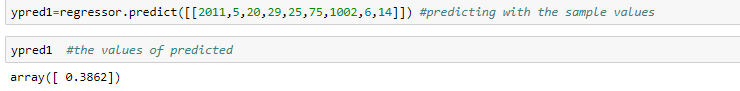
* Training the data to the good model is important
* For our data the best fitting model is RandomForestRegressor
* RandomForestRegressor library is available in sklearn package

Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap Aggregation, commonly known as bagging.

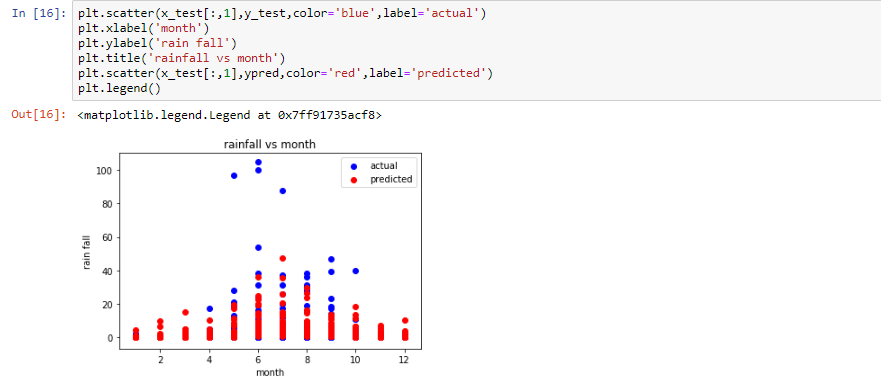
**Evaluating Metrices**



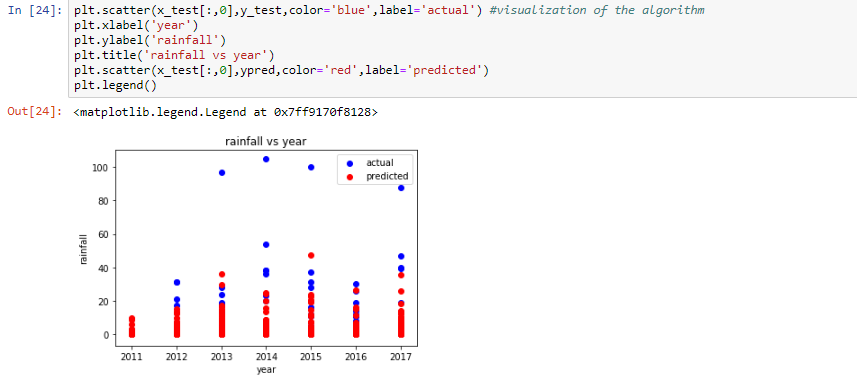
* Using r2\_score we could predict the accuracy of the prediction
* This accuracy shows how accurately our data have got trained to the model
* r2\_score library is imported from sklearn.metrices package

****

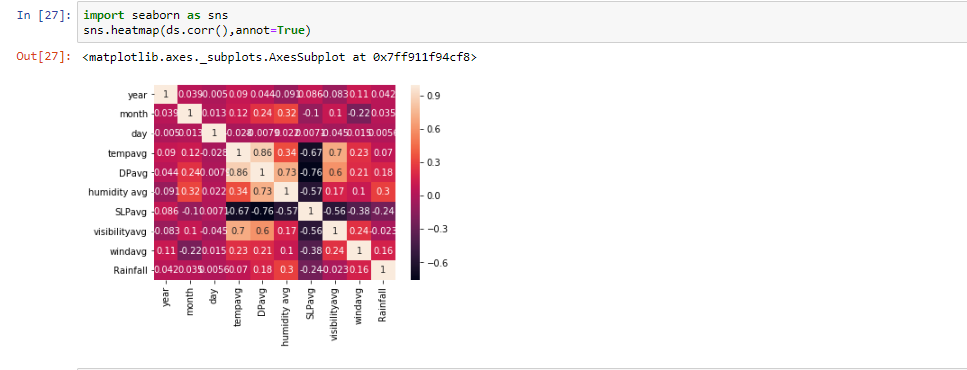
* using predict function we are predicting rainfall
* to predict this we are giving 9 parameters as inputs they are
* year
* month
* date
* humidity
* dew point
* sea level pressure
* wind
* visibility



* here we are ploting the scatter graph between different months and rainfall in those months
* on x-axis there are months
* on y-axis there are rainfall in cm
* red dots describes the predicted rainfall
* blue dots represents the actual rainfall
* if we observe we could find many red dots and blue dots were merged
* this means we had accurate prediction

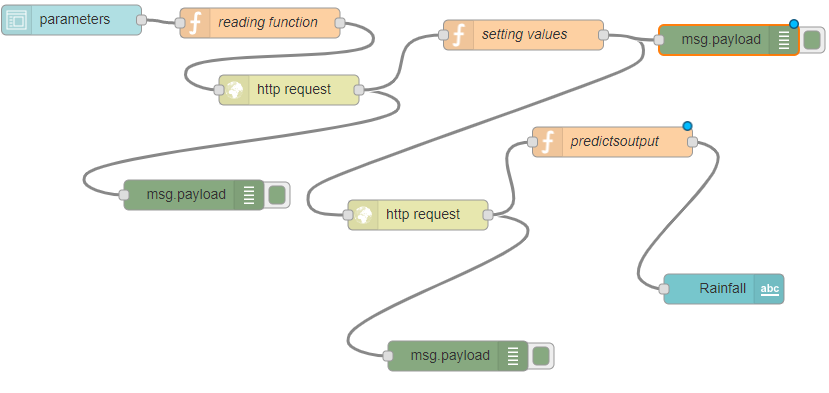


* here we are ploting the scatter graph between different months and rainfall in those years
* on x-axis there are years
* on y-axis there are rainfall in cm
* red dots describes the predicted rainfall
* blue dots represents the actual rainfall
* if we observe we could find many red dots and blue dots were merged
* this means we had accurate prediction



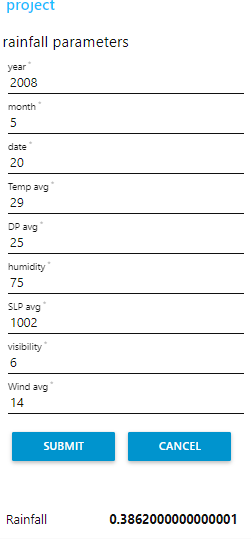
* Here we are plotting heat map for the given data.
* A **heat map** (or **heatmap**) is a graphical representation of data where the individual values contained in a matrix are represented as colors. "Heat map" is a newer term but shading matrices have existed for over a century.
* Many different color schemes can be used to illustrate the heat map, with perceptual advantages and disadvantages for each.

**FLOW**



* To make our project easy access we are using flow
* In this flow we had different types of nodes
* Function node
* Debug node
* Label node
* https request node
* form node
* Flow is done using the soft ware NODE-Red
* **Node-RED** is a flow-based development tool for visual programming developed originally by IBM for wiring together hardware devices, APIs and online services as part of the Internet of Things.

**USER INTERFACE**

****

* This is how our user interface looks like
* If all the parameters are filled and submitted, the predicted rainfall will be previewed

**Findings And Suggestions**

* We found the data from the books we collected and from the weather forecast centre.
* Some of the models we searched from the internet.
* The front end development is done on the IBM CLOUD.
* There is a lack of competent software professionals and scientists working with Indian Meteorological Department (IMD) is big problem in rainfall prediction.
* Farmers demand accuracy in long-range forecast that could help them figure out whether or not to sow in the season, what they need more is forecasts at block level. A block is an administrative unit within a district.
* So we suggest to overcome these kind of problems, so that we can help our farmers in crop production.

**CONCLUSION**

This project investigates the use of machine learning methods and particularly suggests in improving the rainfall forecasting.The rainfall data include the years of 2011-2017.

For a reasonable evaluation of the random forest regression model we conclude that the efficient method among the two input techniques due to simplicity in computation and superior capability of the rainfall forecasting.

**BIBLIOGRAPHY AND REFERENCES**

**REFERENCES**

**Abrahart, R.J. and See, L. (2002)** : Multi-model data fusion for river flow forecasting: An evaluation of six alternative methods based on two contrasting catchments, Hydrology and Earth System Sciences 6(4): 655–670.

**Baruque, B., Porras, S** : and Corchado, E. (2011). Hybrid classification ensemble using topology-preserving clustering, New Generation Computing 29(3): 329–344.

**Chalimourda, A., Sch ̈olkopf, B** & **Smola, A.J. (2004).** Experimentally optimal ν in support vector regression for different noise models and parameter settings, Neural Net-works. The Official Journal of the International Neural Network Society 17(1): 127–41.