iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42151

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Rainfall Predictions Using Data Visualization Techniques

Kaushal Kailas Sarda
Vellore, India

Abstract: ***Weather forecasting has now entered the era of Data Visualisation due to the advancement of climate observing systems like satellite meteorological observation and also because of the increase in the collection of weather data. So, the traditional intelligence models are not adequate to predict the weather precisely. Hence, machine learning-based techniques are importantto process massive datasets that can learn and make predictions more precise based on past data. This research paper provides a thorough review of different rainfall prediction techniques, along with the help of publicly available datasets. Statistical techniques for rainfall forecasting cannot predict well for long-term rainfall forecasting due to the constant change of climate phenomena every year. This paper delivers a precise classification of rainfall forecasting models and discusses potential future research methods in this area.***

Index Terms: Data Visualisation, Weather Prediction, Machine Learning Techniques, Data Analytics, Rainfall Pattern

I. INTRODUCTION

A forecast is calculation or estimation of future events, especially for economical trends or future weather. As an example, in the tropics regions which several countriesonly had two seasons in a year , many countries especially country which relies so much on agricultural field need to forecast rainfall in term to decide the best time to start the production and planning.Rainfall play important role informing of fauna and flora of natural life. It is not just significant for the human beings but also for other species like animals, plants and all living things. It plays a significant role in agriculture and farming and ; water is one of the most natural resources on earth. The changing climatic conditions in the country and the increasing global warming effects have made it difficult for the human beings and the planet to experience necessary amount of rainfall that is required to satisfy the human needs and its uninterrupted use . Therefore, it has become significant to analyze the changing patterns of the rainfall and try topredict the rain not just for the human needs but also to predict for natural disasters that could cause by the unexpected heavy rainfalls. The prediction of rainfall has serious importance in various dimensions and scopeReducing the impact of sudden and heavy rainfall can be very beneficial by taking appropriate safety measures before any natural disaster . To be more specific and aware of the devastating climatic changing and stay updated; predicting rainfall has been the focus of computer scientist and engineers. The dynamic approach and predictions are generated by physical models and Probability Mode; based on system of equations that predict the future Rainfall. The forecasting of weather by computer using equations are known as numerical weather predictions. Numerical weather prediction (NWP) uses mathematical models ofthe oceans to predict the weather based on current weather conditions To predict the weather bynumeric means, meteorologist has develop atmospheric models that approximate the change in humidity, temperature, etc using mathematical equations

1) Dataset 1 Description: The dataset1 consists of themeasurement of rainfall from year 1901-2015 for each state.
2) Dataset1 consists of 19 attributes (individual months, annual, and combinations of 3 consecutivemonths) for 36 sub divisions.
3) The data is available only from 1950 to 2015 for some of the subdivisions .The attributes are theamount of rainfall measured in mm.
4) Dataset 2 Description: The data refers to details on weather of district wise rainfall (in mm**)** calculated with the data for the period 1951-2000.

II. RELATED WORK

Many researchers had already worked on the rainfall prediction and their connection with climate change with different methods of machine learning. Several studies are being discussed in India to predict the climate and, researchers used different lengths of information and now studies have been reported using information over a long time. Various researches have explained that predicting andanalysis of monthly rainfall was suggesting different algorithms and methods. Some of them are:

1) Estimating Rainfall Prediction using Machine Learning Techniques on a Dataset R Vijayan, V Mareeswari, P Mohankumar, G Gunasekaran K Srika.The research paper involves prediction of rainfall in the city utilizing five methods of data mining: supporting vector machine, random forest, andmultilayer perceptron.

2) A survey on rainfall prediction techniques MR.Dhawal Hirani , Rr. Nitin Mishra (international journal of computer application (2250- 1797)).This paper reports a detailed survey on rainfall predictions using different rainfall prediction methods extensively used over last 20 years years. From the survey it has been found thatmost of the researchers uses artificial neural network for rainfall prediction and got preciseresults. The survey also gives a important fact that the forecasting techniques that use KMeans-Clustering, BPN, Multiple Linear regression, random forest regressor and SVM are suitable to predict rainfall.

3) Thirumalai Chandrasekar "Heuristic prediction ofrainfall using machine learning techniques",2017. discusses the amount of rainfall in past years according to the crop seasons in the country and predicts the rainfall for future years. Linear regression method is applied for the prediction. Rabi and kharif were taken as variables if one of the variable was given ,then the other can be predicted using linear regression. Standard deviation and Mean was also calculated for futureprediction of crop seasons in the country. This implementation prove to be useful for farmers to give an idea of which crop to grow during the particular time of the year due to rainfall prediction.

4) Rainfall prediction using data visualisation techniques (Yogesh Kumar Joshi, Udit Chawla, shipra shukla) 2020 10th internationalconference on cloud computing, data science & engineering . Discusses about all the varioustrends in rainfall in the states/union territories over the past century has been shown with the help of data all the various trends in rainfall in the states/union territories over the past century . The results obtained in this paper canbe used for the prediction of rainfall with the help of regression which can be of great benefit in the field of agriculture. It helps to identify the drought specific region and coastalrainfall trend.

5) Chakraborty S. , Nagwani n. , Dey l-"Weatherforecasting using incremental k-means clustering", 2017 have discussed about the forecasting of rainfall and weather with the help of incremental k-means clustering. In thispaper, a new method is proposed for weather forecasting by the help of clustering algorithm. The class KMeans(centroid-based algorithm**)** is imported from sklearncluster library. To find the optimal number of cluster for the dataset, the model was provided with differentnumbers of cluster ranging from varying numbers. The k-means++'method to passed to the init argument to avoid the random Initialization . The max_iteration and the n_initiation were passed with their default values.The new data is grouped into clusters. By doing the algorithm makes a strategy to predict the weather .Various Machine learning models can also help to predict the useful insights from the large datasets

III. METHODOLOGY

There are several steps involved in the visualisation ofdata. From the studying of the datasets to plotting it ingraphs, each step taken for the visualisation of the data is explained in this section.

1) Analysing the Problem Analysing the problem can be done by studying the data sets. By studying the datasets NaN values present in the datasets can be found. In this stage the type of visualisation which canbe used for the dataset is selected. Studying the dataset gives a clear view and better understanding of the problem statement.

2) Cleaning the Datasets Incorrect data leads to false and inaccurate conclusions. These inconsistencies in the datasets can be due to user error. To correct these inconsistencies data cleaning is used to identify and correcting(or removing) all the inaccurate records present in the dataset. Data cleaning can find out the inaccurate, incomplete, irrelevant and incorrect partsof the data and remove or replace them with median values. There are many tools available for data cleaning, in this paper pandas library is used for the cleaning of data. After removing these irregularities inthe datasets, Data Preprocessing is done.

3) Data Preprocessing Good quality of data needs to be accurate, precise, consistent and uniform. Data Processing is used to achieve the quality standards ofa dataset. Data Preprocessing is used to convert the raw data into a more refined and useful one. Data preprocessing is very important as it improves thequality of the raw data and helps in achieving better results. In this report, data preprocessing includes the replacement of NaN values with median values and then grouping of data in ascending order as well as grouping of data with respect to their respective states.These changes increase the reliability of the result and improve the efficiency of the program.

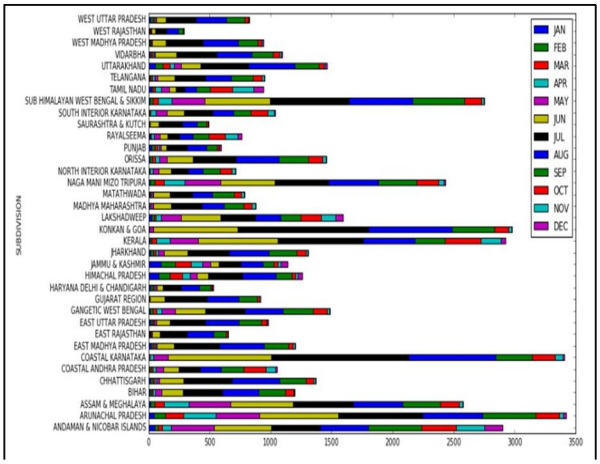
A. Algorithm and Methods

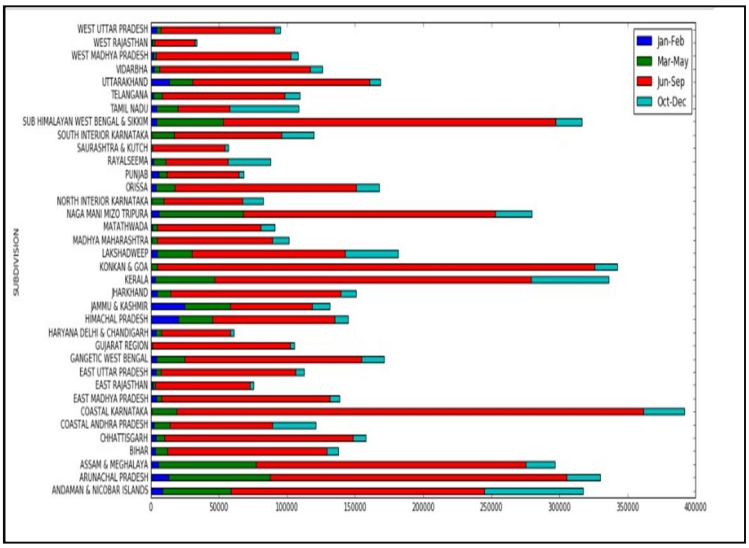
This predictive model is used for prediction of the rainfall. The first step is converting data in to the correct format and data cleaning to conduct analysis then observe variations in the patterns of rainfall. We predict the rainfall by separating the dataset into training set and testing set then we apply different machinelearning approaches and statistical techniques and compare and draw analysis over various approaches used in the project. With the helpof numerous approaches we attempt to minimize the error.

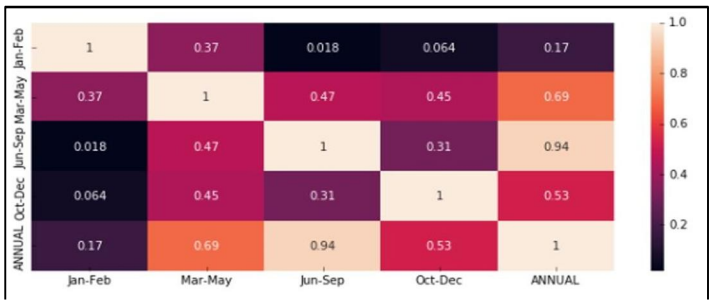
IV. RESULTS AND METHODS

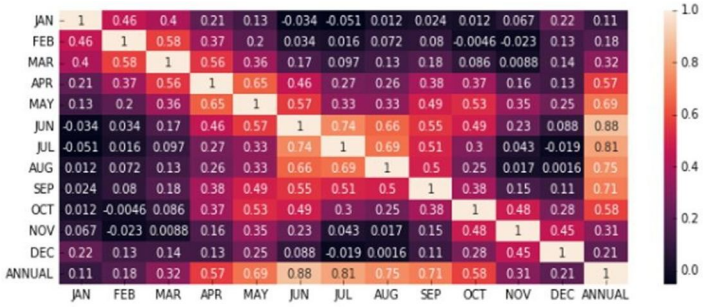
The following models were found to besuitable for studying rainfall patterns as permany reference papers:

1) Multiple Linear Regression
2) Random Forest Regression
3) SGD Regression
4) Support Vector Regression
5) After analysing the Random Forest Regression has the minimal mean absolute error and is found to predict all seasonal variations andpatterns of Rainfall in India compared to all other machine learning techniques.
6) First with this exploratory analysis, first import libraries and define functions for plotting the datausing matplotlib. Depending on the data, not all plots can be plotted as number of attributes can be more or less
7) Converting and cleaning the dataset in to the correct format to conduct precise predictions of rainfall
8) Make a good analysis of data and observe variation in the patterns of rainfall in every part of the country.
9) We predict the average rainfall by separatingdata into training sets and testing sets. Weapply various statistical and machine learning approaches(*SVM*,randomforest regressor,Linear Regression) in prediction and make analysis over various approaches done inthe project. By using various approaches wetry to minimize the error of the rainfallprediction in the country.
10) Bar graphs showing distribution of amount of rainfall. Distribution of amount of rainfall yearly, monthly, groups of months. Distribution of rainfall in subdivisions, districts form each month, groups of months.
11) Heat maps showing correlation betweenamount of rainfall between months.







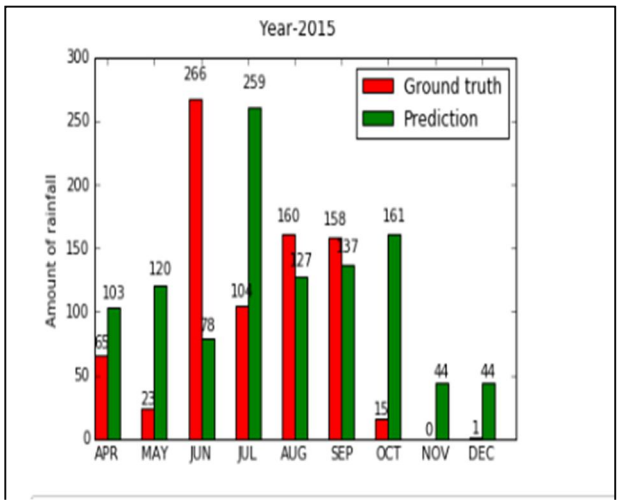


For prediction we organized data in such a way, given the rainfall in the last three months we predict the rainfall in the next consecutive month.

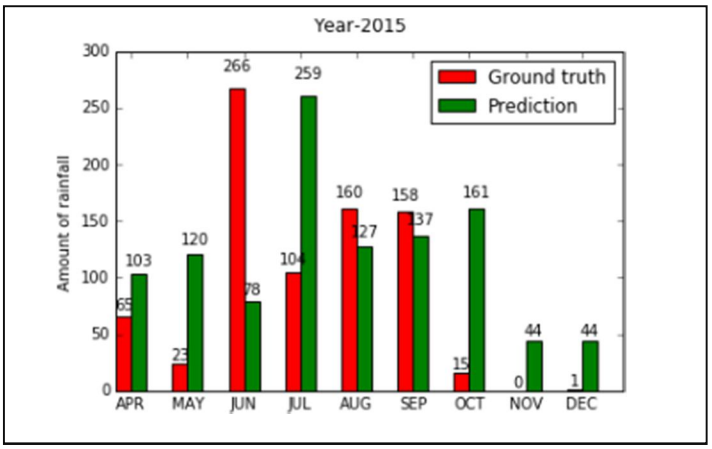
The important thing is that how well the training sets, testing sets and validation data sets describe thefeature space. If the number of points in the whole data set is large then any division may work fine but when the data set is limited, division ratio may playa crucial role.For all the experiments we used 80:20 training and test ratio.

a) Linear regression
b) Support Vector Regression
c) Random Forest Regressor
d) Testing metrics: We used Mean absoluteerror to train the models. In the research two types of trainings once training on completedataset and other with training with only 1state data
e) All means are standard deviation and observations are written, 1st one represents ground truth, 2nd one represents predictions.

This figure shows the represents ground truth, second one represents predictions using Linear Regression Model.



THIS figure shows the represents ground truth, secondone represents predictions using Support VectorRegression



V. CONCLUSION

This paper reports a detailed report on Rainfall Predictions using different machine learning techniques extensively used over last 20 years.

From the literature survey it has been found thatmost of the researchers used artificial neuralnetworks for rainfall prediction and got accurateresults. The research paper also gives a conclusion that the forecasting techniques that use machine learn techniques are suitable to predict rainfall than other statistical and numerical methods. The prediction of rainfall in the city utilizing three methods: supporting vector machine, random forest and linear regression. A classification system is used for successful prediction in which the input datagoes via a preprocessing stage and was cleaned and normalized until the classification process. Ten preparing and test information proportions are utilized from 10:90 to 90:10 to investigate the exhibition reliance of arrangementmethodologies on preparing information. Results show that the characterization procedures utilized performed well for no- downpour class however for downpour class,the systems didn't perform well. The purposes for the lower downpour class exactness mayincorporate missing qualities, absence ofsignificant climatic characteristics in the datasetand a lower in general precipitation rate in the district. For future work, it is proposed that further forecasts should be carried out by testingfurther techniques of classification and climate attributes on different weather dates. Henceforth, accuracy is based on random forest and logistic regression and future support vectormachine is used to estimate accuracy.

REFERENCES

[1] Estimating Rainfall Prediction using Machine Learning Techniques on a Dataset INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 06, JUNE 2020 ISSN 2277-8616 440 R Vijayan, V Mareeswari, P Mohankumar, G Gunasekaran K Srikar.

[2] International Journal of Computer Application (2250-1797) Volume 6– No.2, March April 2016 28 A Survey On Rainfall Prediction Techniques MR. DHAWAL HIRANI , DR. NITIN MISHRA

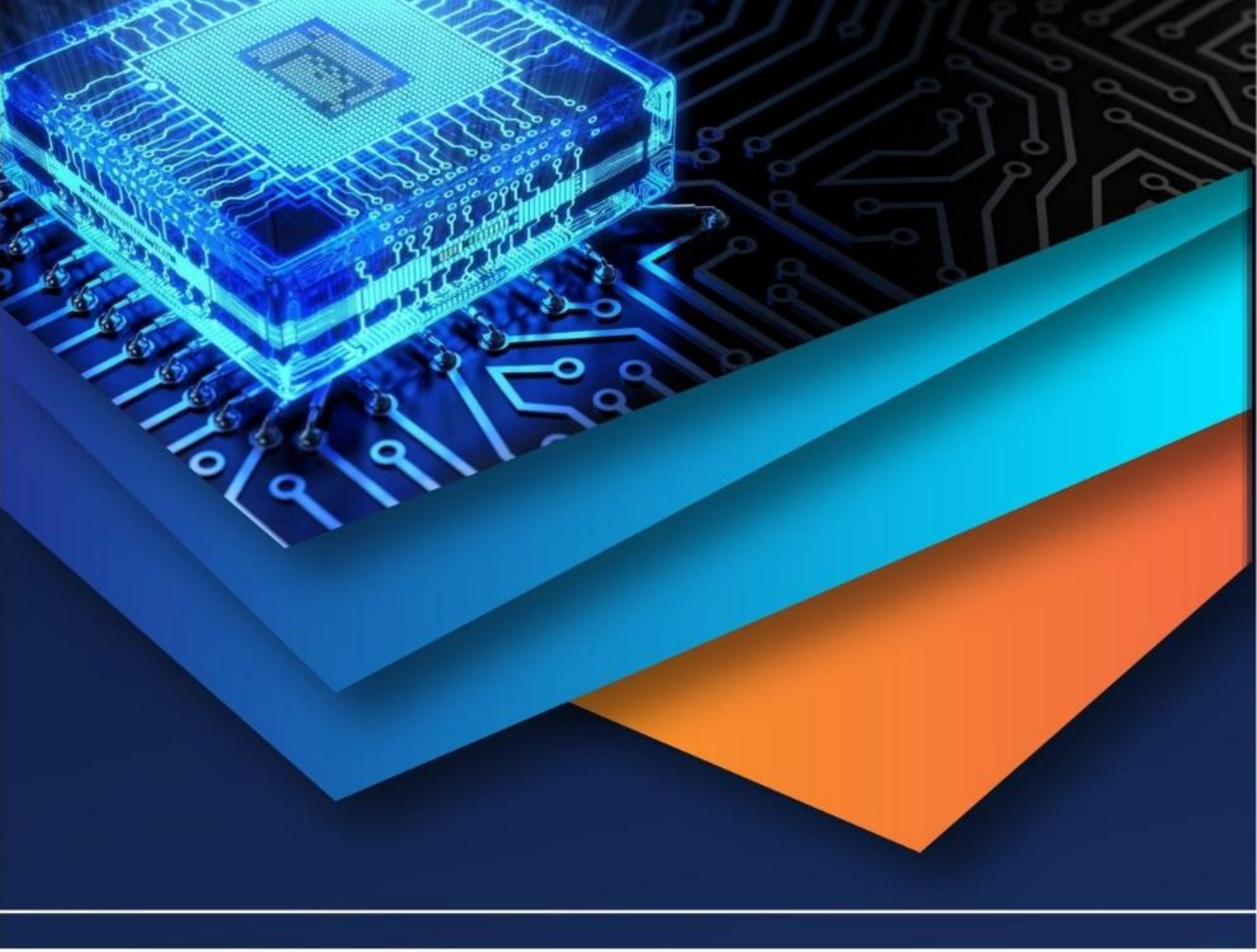
[3] G. Tripathi and R. Gupta, "A New Approach for Rainfall Prediction using Artificial Neural Network", International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 7, no. VII, July 2019.

[4] Weather Forecasting using Incremental K-
means Clustering SANJAY CHAKRABORTY Prof. N.K.NAGWANI LOPAMUDRA DEY

[5] Heuristic Prediction of Rainfall Using Machine Learning Techniques,May 2017,Conference: International Conference on Trends in Electronics and In formatics ICOEI 2017 and authors:Chandra Segar Thirumalai,K Sri Harsha
,M Lakshmi Deepak,K Chaitanya Krishna

[6] Rainfall Prediction Using Data Visualisation Techniques Yogesh Kumar Jo shi Department of CSE Amity University Uttar Pradesh, Noida, UP, India Udit Chawla Department of CSE Amity University Uttar Pradesh, Noida,
UP, Shipra Shukla Department of CSE Amity University Uttar Prades

[7] Trends in the rainfall pattern over India Article in International Journal of Climatology · September 2008. Pulak Guhathakurta,India Meteorological Department Pune and M. Rajeevan M.Sc, Ph.D



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY