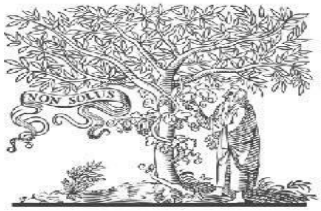
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 31st Mar 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 03

10.48047/IJIEMR/V12/ISSUE 03/108

Title **PREDICTION OF RAINFALL USING LINEAR REGRESSION, RANDOM FOREST AND KNN ALGORITHMS**

Volume 12, ISSUE 03, Pages: 765-771

Paper Authors
A. Akashnath Ganesh , K. Sowmya , BH. Prema Padmaja , A. Prasanna Kumari, A. Gowthami Priya

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

PREDICTION OF RAINFALL USING LINEAR REGRESSION, RANDOM FOREST AND KNN ALGORITHMS

A. Akashnath Ganesh [1], K. Sowmya [2], BH. Prema Padmaja [3], A. Prasanna Kumari[4], A. Gowthami Priya [5]

India Assistant Professor, Department Of Computer Engineering,
Kallam Haranadha Reddy Institute of Technology, India

ABSTRACT

In this paper,we present the Predicting rainfall is essential since excessive rains may cause a variety of calamities. Taking preventative actions is made easier with the aid of prediction. Moreover, the forecast should be correct. Because heavy precipitation is strongly connected with the economy and human lifespan, it might be a severe negative for earth science departments. Individuals all around the world square measure confronted with natural calamities like floods and droughts every year. For nations like India, whose economy is largely based on agriculture, the accuracy of rainfall estimates is of great importance. We are unable to accurately predict precipitation due to the nature of the atmosphere and the equations used. Regression may be used to forecast precipitation using machine learning approaches. Non-experts will have access to techniques and approaches used in precipitation prediction. A comparison study will be conducted among the various machine learning techniques.

Keywords:
Rainfall, Prediction, Machine Learning.

1. INTRODUCTION

Rainfall forecasting is critical because heavy and irregular rainfall can have a wide range of consequences, including crop destruction and property damage. A greater predictive model is required for early warning, which can reduce risks to life and property while also better managing agricultural farms. This forecast primarily benefits farmers, and water resources may be used more efficiently. Rainfall forecast is a difficult process, and the findings must be precise. There are several hardware devices available for forecasting rainfall based on meteorological variables such as temperature, humidity, and pressure. Because traditional approaches are inefficient, we can get accurate results by employing machine learning algorithms. We can simply accomplish that by analyzing past rainfall data and forecasting rainfall for future seasons. We may use different approaches, such as classification and regression, depending on the needs, and we can also determine the error between the real and predicted values, as well as the accuracy. Because different approaches generate varying degrees of accuracy, it is critical to select the appropriate algorithm and model it in accordance with the requirements.The benefits of regression approach is as follows, it is a strong approach for examining the connection between a single dependent variable and a large number of independent factors. It enables researchers to manipulate extraneous elements. It also helps to attain the measure of error using the regression line as a base for estimations. The regression line can also be used as a basis for estimates.

2. LITERATURE SURVEY

Due to the severe weather conditions, several farmers have taken harsh action such as committing suicide. 300000 farmers have committed suicide in India between 1995 and 2014. As a result of this, Thirumalai, Chandrasegar, et al. forecast the quantity of rainfall for future years based on the agricultural seasons of the past. Rabi, Kharif, and Zaid are the three main harvest seasons. In order to make early predictions, the linear regression approach is used. A linear regression was used to predict Rabi and Kharif's values if only one variable was known.

For the future forecast of agricultural seasons, the standard deviation and mean were also computed. In accordance with the crop season, farmers will be able to determine which crops they should harvest with the help of this tool, is a computer programme thatforecasts meteorological situations such as precipitation and thunderstorms, and allows users to take precautionary actions.

Data mining was used to model the algorithms. That would be the process of identifying patterns in huge data sets by using approaches at the confluence of machine learning, statistics and database systems. In computer science and statistics, data mining is the process of extracting information (using intelligent algorithms) from a data collection and turning it into an understandable format.

It is now possible to get a complete set of rainfall data with characteristics. Algorithm is used to the dataset once it has been split into training and testing sets. Actual and projected values are compared to calculate accuracy. In order to reach a high value, the accuracy can be improved by using better algorithms. The data is pre-processed and normalized. A training set of data up to 2010 was used, and a test set of data from 2011 to 2016 was used as a testing set. MAE, RMSE, and MASE were used to calculate the performance of the algorithms. Comparatively, the Linear Regression method has produced more accurate findings.

In this framework, 1. Introduction was discussed in section 2. Related work will be discussed in section. 3.Describes about our proposed work 4.Presents Results for our proposed work 5 will conclude this paper.

Author & Year	Proposed	Finding/Outcomes
Thirumalai , Chandra Segar,	The proposed method shows high accuracy in determining the type of skin lesion whether it is benign or malignant which will be very beneficial for diagnosis of melanoma skin cancer efficiently. Rainfall rates in prior years have been compared to several crop seasons (rabi, Kharif and zaid) in order to make predictions for future seasons. For an effective knowledge of agriculture in India, the study also assesses the different categories of data by linear regression approach using metrics.	Using the results of this programme, farmers may make an informed choice on when to harvest a particular crop. Regression linear aide à identifier.
Geetha, A., and G. M. Nasira	There are many weather events that may be predicted using a decision tree. This knowledge can save lives and be utilized by individuals from many walks of life to make sensible and informed decisions.	More relevant qualities may be utilized to predict the dependent variables in this model, allowing for further improvement.
Parmar, Aakash, Kinjal Mistree, and Mithila Sompura.	As a result of this research, it is possible to save and recover data that has been held for a longer length of time. However, it also reveals that neural networks are not able to make predictions about the data that has been stored.	As a result of this research, all the weather forecasting networks in India would be able to provide more accurate and timely information to their viewers.

Dash, Yajnaseni, Saroj K. Mishra, and Bijaya K. Panigrahi.	For summer monsoon (3.075) and post-monsoon (3.149) respectively, ELM approach has demonstrated higher performance with minimal mean absolute percentage error scores compared to KNN and ANN techniques.	These artificial intelligence techniques are capable of accurately predicting both summer monsoon and post-monsoon conditions in Kerala, India, with minimum prediction error scores, according to the results of this study.

3. PROPOSED SYSTEM

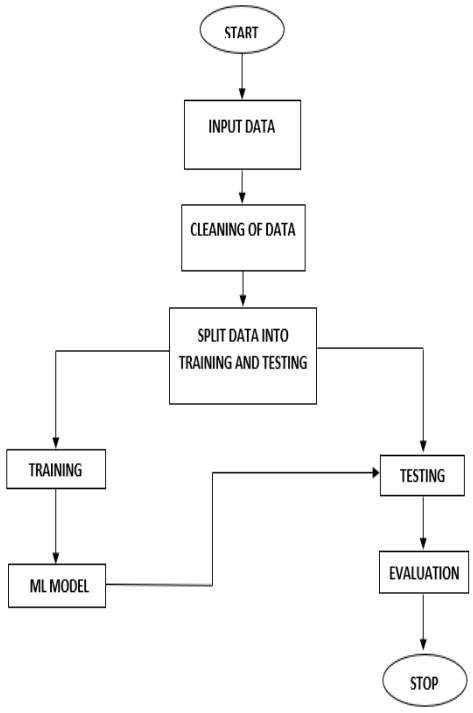
In this the study examines the use of machine learning in the prediction of rainfall to specifications status based on data produced. The model achieved best accuracy of using Linear Regression. The training and testing dataset were performed by the system from Scikit-learn module.

Precipitation is predicted using a predictive model. A good examination of the data and observation of variations in the patterns of rainfall are the initial steps in this process. Datasets are separated into training and testing sets before different machine learning and statistical techniques are applied to estimate the amount of rain. Innumerable techniques are used to reduce the mistake.

Techniques utilized in the production of this work include: By fitting an equation to determined information, multiple linear regression attempts to describe the relationship between two or more variables and a given answer. As far as I can see, it's nothing more than an extension of a simple regression toward the mean.

The general form of multivariable linear regression model is: $y=\alpha+\beta 1x1+ \beta 2x2+\ldots+ \beta kxk+\varepsilon$ where y = dependent variable and x1, x2… xk are independent variables, α, β are coefficients.

When one explicit variable isn't clear enough to map the relationship between the independent and also the variable quantity, multiple regression can be used to describe the relationship.



International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

The data which we have been collected will undergo for the cleaning process. We perform splitting of the input data into training and testing datasets with a desired ratio by using a powerful python module named Scikit-learn.

Machine learning algorithms are evaluated using the train-test split process when they are used to generate predictions on data that was not used to train the algorithm. You may compare the performance of different machine learning algorithms for your predictive modelling issue. If you have a tiny dataset or if further setup is necessary (such as if the dataset is not balanced), you should not utilize the technique.

4. RESULTS AND DISCUSSION

In this session we will discuss about the results that are obtained by performing the above method and how they are transformed by using the AI architecture.

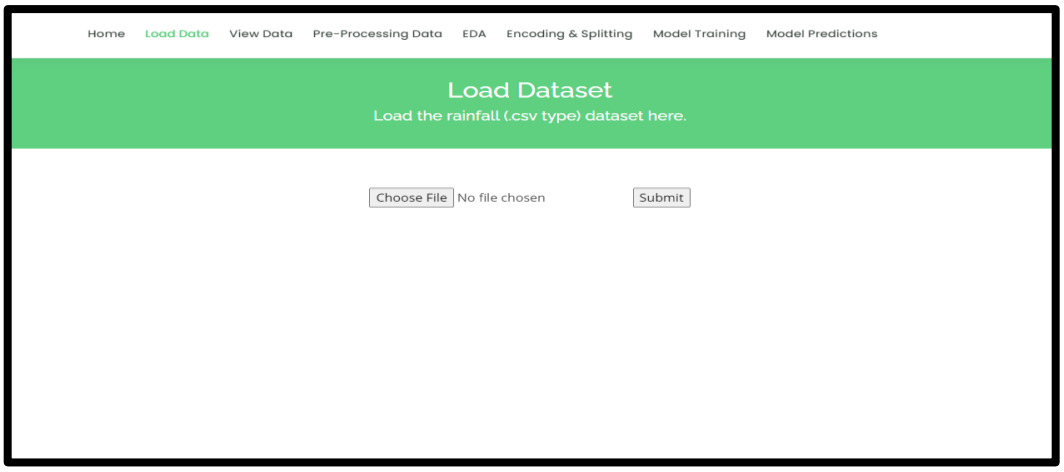
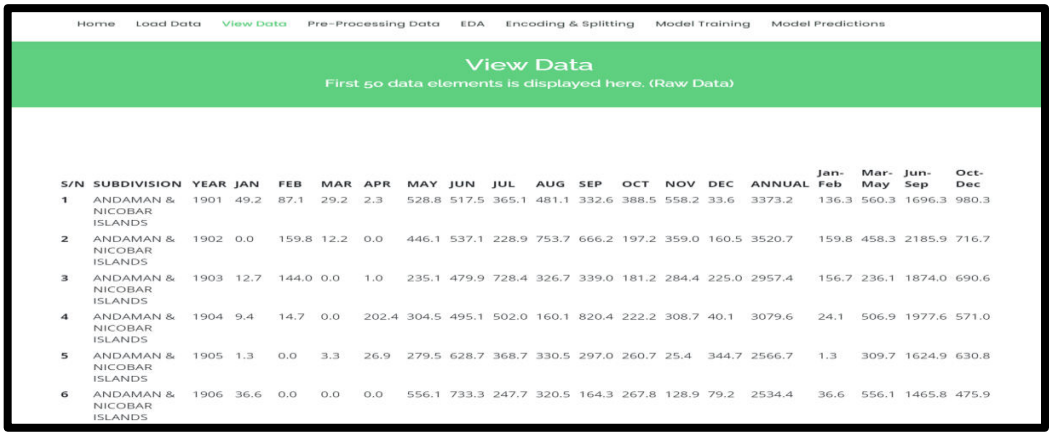


Fig 1. Input Images

The above images represent the loading of dataset to get the final response from the given data

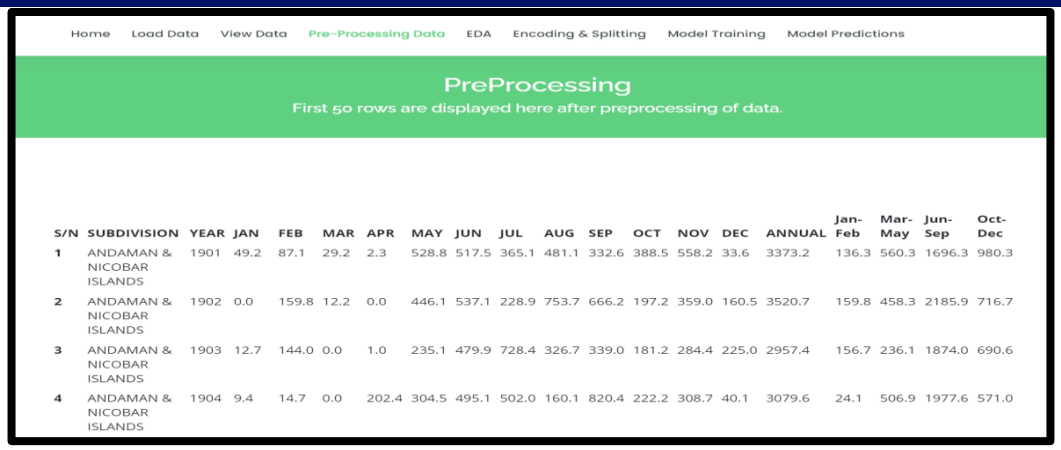


S/N	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
1	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	560.3	1696.3	980.3
2	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2	359.0	160.5	3520.7	159.8	458.3	2185.9	716.7
3	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2	284.4	225.0	2957.4	156.7	236.1	1874.0	690.6
4	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2	308.7	40.1	3079.6	24.1	506.9	1977.6	571.0
5	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7	25.4	344.7	2566.7	1.3	309.7	1624.9	630.8
6	ANDAMAN & NICOBAR ISLANDS	1906	36.6	0.0	0.0	0.0	556.1	733.3	247.7	320.5	164.3	267.8	128.9	79.2	2534.4	36.6	556.1	1465.8	475.9

Fig 2. View Data

The above image shows us the total available data in our project

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org



S/N	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
1	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	560.3	1696.3	980.3
2	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2	359.0	160.5	3520.7	159.8	458.3	2185.9	716.7
3	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2	284.4	225.0	2957.4	156.7	236.1	1874.0	690.6
4	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2	308.7	40.1	3079.6	24.1	506.9	1977.6	571.0

Fig 3. Preprocessing Data

The above shown figure represents the preprocessing of data in which the tasks like cleaning of data is performed.

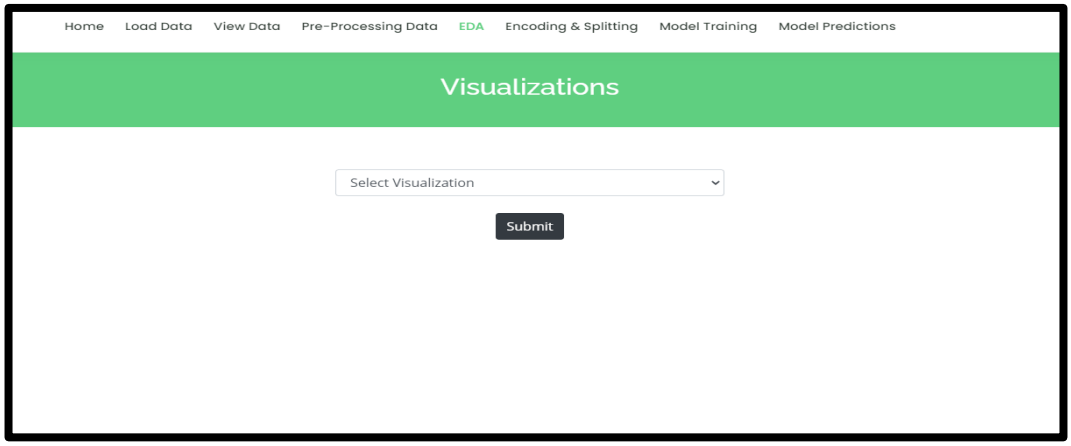


Fig 4. Visualizations

The above shown figure is for visualization which gives pie and histogram plots. More over examining the data using plotting is better than tabular form

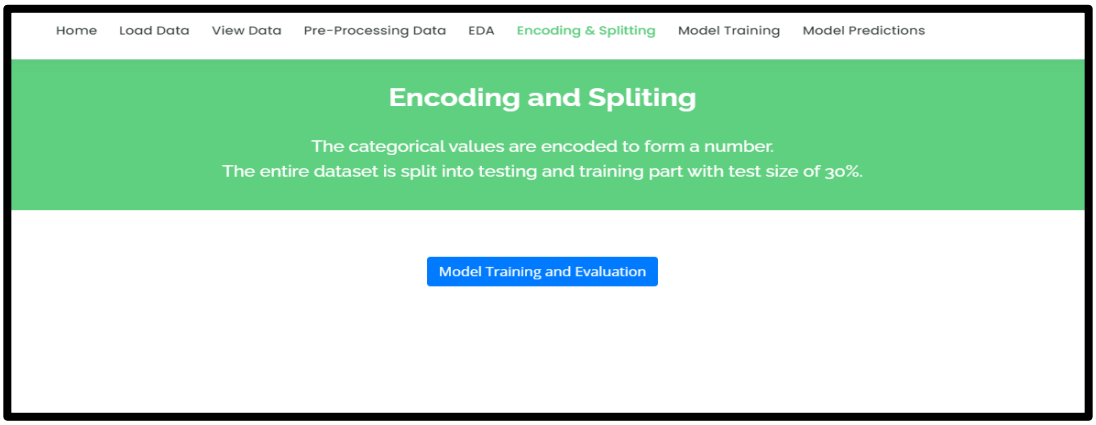


Fig 5. Encoding and Splitting Data

The above-shown figure is performing encodings and splitting of the dataset using Scikit-learn. As the encoding and splitting will be performed by using a powerful Scikit-learn module.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

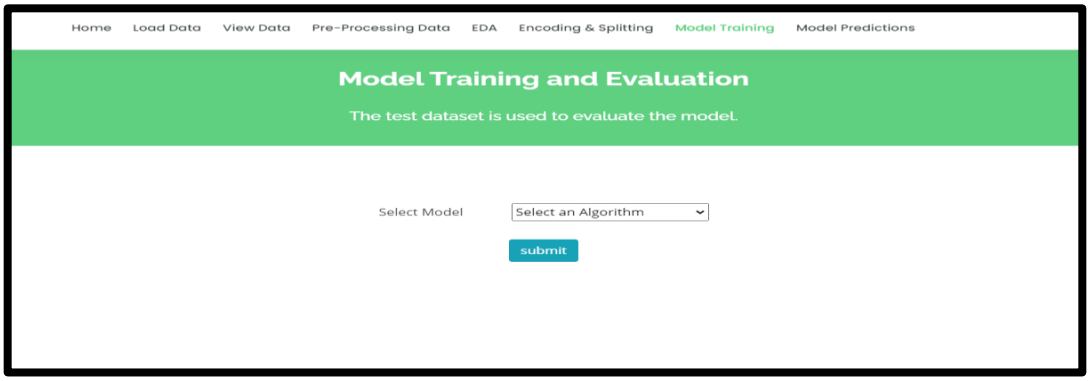


Fig 6. Model Training and Evaluation

The above-shown figure is for training our models and to visualize the evaluation.

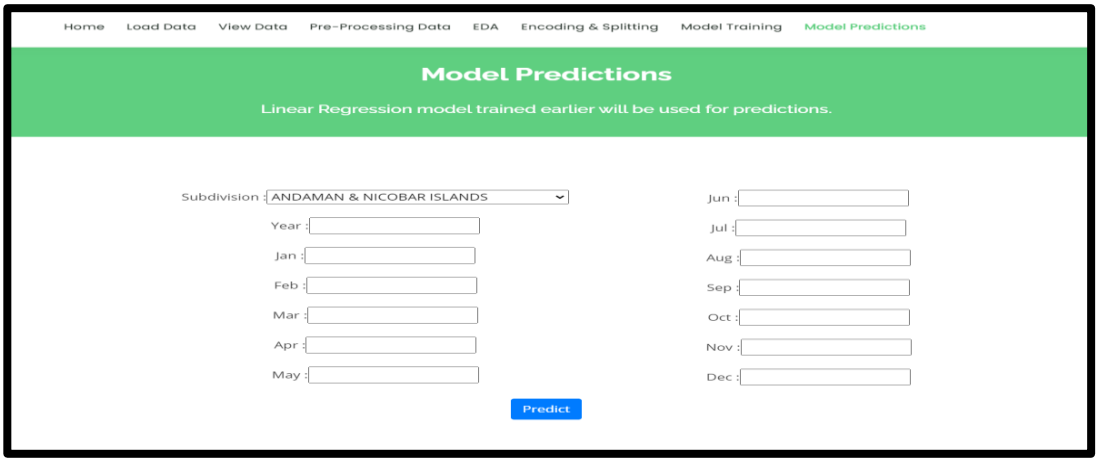


Fig 7. Input

The above-shown figure is for performing predictions to get the required output from the given inputs.

6. CONCLUSION

Machine learning algorithms are evaluated using the train-test split process when they are used to generate predictions on data that was not used to train the algorithm. You may compare the performance of different machine learning algorithms for your predictive modelling issue. If you have a tiny dataset or if the further setup is necessary (such as if the dataset is not balanced), you should not utilize the technique.

REFERENCES

[1] Thirumalai, Chandra Segar, et al. "Heuristic prediction of Rainfall using machine learning techniques." 2017 International Conference on Trends in Electronics and Informatics (ICEI). IEEE, 2017.

[2] Geetha, A., and G. M. Nasira. "Data mining for Meteorological applications: Decision trees for modeling Rainfall prediction." 2014 IEEE International Conference on Computational Intelligence and Computing Research.IEEE, 2014

[3] Parmar, Aakash, Kinjal Mistree, and Mithila Sompura."Machine learning techniques for rainfall prediction: A Review." 2017 International Conference on Innovations in Information Embedded and Communication Systems.2017.

[4] Dash, Yajnaseni, Saroj K. Mishra, and Bijaya K. Panigrahi. "Rainfall prediction for the Kerala state of India using Artificial intelligence approaches." Computers & Electrical Engineering 70 (2018): 66-73.

[5] Singh, Gurpreet, and Deepak Kumar. "Hybrid Prediction Models for Rainfall Forecasting." 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2019.

[6] Kar, Kaveri, Neelima Thakur, and Prerika Sanghvi. "Prediction of Rainfall Using Fuzzy Dataset." (2019).