RAINFALL PREDICTION USING MACHINE LEARNING

[1] Sk Arshiya Julma, [2]Kotha Pravalika, [3]Bodla Sai Supriya, [4]K.Hari Chandhana, [5]Challa Harshini
[1]Assistant Professor, [2,3,4,5]UG Students, [1,2,3,4,5]Dept. of CSE (AI & ML), Malla Reddy Engineering College for Women (Autonomous), Hyderabad, India. E-Mail: arshiyashaik06@gmail. com

ABSTRACT
The prediction of rainfall is one of among the most significant methods of forecasting the weather in any country. This paper presents a rainfall forecast model using MLR (multiple linear regression) for the Indian dataset. As a result of the input data's multiple meteorological variables, forecasting rainfall is more accurate. Accuracy and correlation are the measures used to assess the suggested approach. The results show that the proposed model for machine learning outperforms the methods currently used in the literature.

INTRODUCTION

The Rainfall Prediction project is aimed at developing machine learning that can predict rainfall based on historical weather data. This documentation provides an overview of the project, including its objectives, dataset used, model architecture.

Rainfall prediction is significant to Indian culture and has a significant impact on people's daily lives. The meteorological department has the difficult task of making uncertain predictions about the frequency of rainfall. With the meteorological conditions changing, it is challenging to anticipate the rainfall with accuracy. Predicting the amount of rain during the summer and the rainy season is difficult. Researchers from around the world have created a variety of models to forecast rain fall, most of which use random numbers and are related to climatic data.

Data Pre-processing: Handle missing values will fill or remove missing values in the dataset. Normalized features will perform scaling and it will scale the numerical features to a common range. And it will preprocess the encode categorical features for convert categorical variables into numerical representations, if applicable

Feature Selection: Analyze the correlation between features and select relevant features for the model.

Model Training: Split the dataset into training and testing sets. Select an appropriate machine learning algorithm (e.g., Random Forest, Gradient Boosting, etc.). Train the model using the training set.

Model Evaluation: Use suitable metrics, for instance the mean absolute error (MAE) or room mean squared error (RMSE), to assess the performance of the model. Comparing the algorithm's predictions to the actual rainfall values for the testing set.

Model Deployment: Once the model achieves satisfactory performance, it can be deployed to make rainfall predictions against the actual rainfall values from the testing set.

There right now are no reliable ways to forecast whether it will rain today or not. Sometimes even the forecasts of the meteorological autonomy are wrong.

This problem is related to Rainfall Prediction using Machine Learning because machine learning models tend to perform better on the previously known task which needed highly skilled individuals to do so.

LITERATURE SURVEY

There are right now no accurate ways to predict whether or not it will rain today. Even the predictions made by the meteorological authority can occasionally be inaccurate.

The following keywords ("machine learning" OR "deep learning") AND ("precipitation prediction" OR "rainfall prediction" OR "precipitation nowcasting") were employed to gather articles from 2016 to 2020. Out of the almost 1240 findings, only supervised rainfall prediction papers using meteorological information from platforms such as radar systems, spacecraft, and observatories were chosen., while publications that used data from regular cameras, for example, pictures, were not included. Although the review's main focus is on rainfall forecasting, similar techniques can be used to predict other geophysical variables like temp and wind. As a result, the chapter's conclusions and debates are adaptable to other conditions. Except for one paper that was issued in 2015 and is an important contribution in this subject, there are a total of 66 reviewed publications, which are composed of conference and journal articles published from 2016 to 2020.

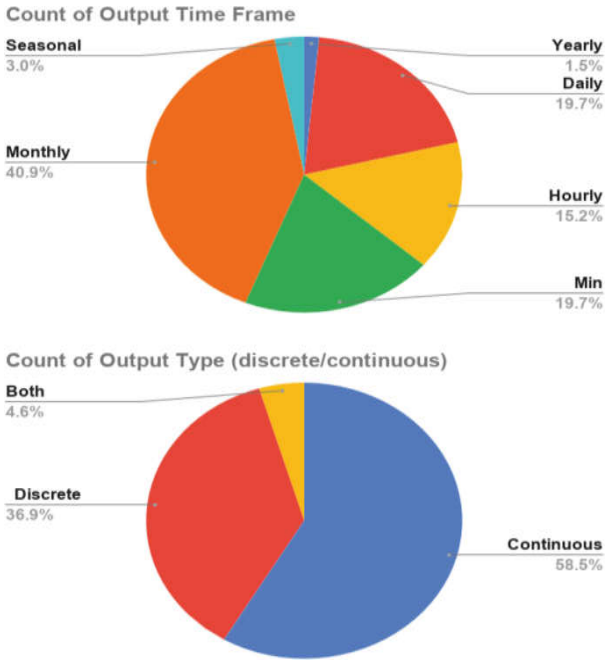


Figure 1 Pie Chart

EXISTING SYSTEM

Rainfall is triggered on by abrupt changes in the weather. The meteorological department was unable to provide precise forecasts. The two most popular techniques were statistical analysis and numerical analysis. Data on rainfall are not linear in nature. The three important features are distribution, precipitation, and amount of rainfall. These standards can differ from one location on earth to another. Traditional techniques have some drawbacks. These drawbacks affect predicting the precise amount of rainfall. Only linear procedures will be compatible with statistical methods.

DISADVANTAGES

Using statistical methods to predict the rainfall we will face many difficulties.

By using numerical analysis and precipitation the accurate results are not found.

PROPOSED SYSTEM

By using Random Forest and SVM classifier this predictor will be giving the accurate values. The rainfall predictor will analyze the dataset and give the accurate values according to the inputs. In this the data

visualization is also done together with the help of graphs.

SYSTEM ARCHITECTURE

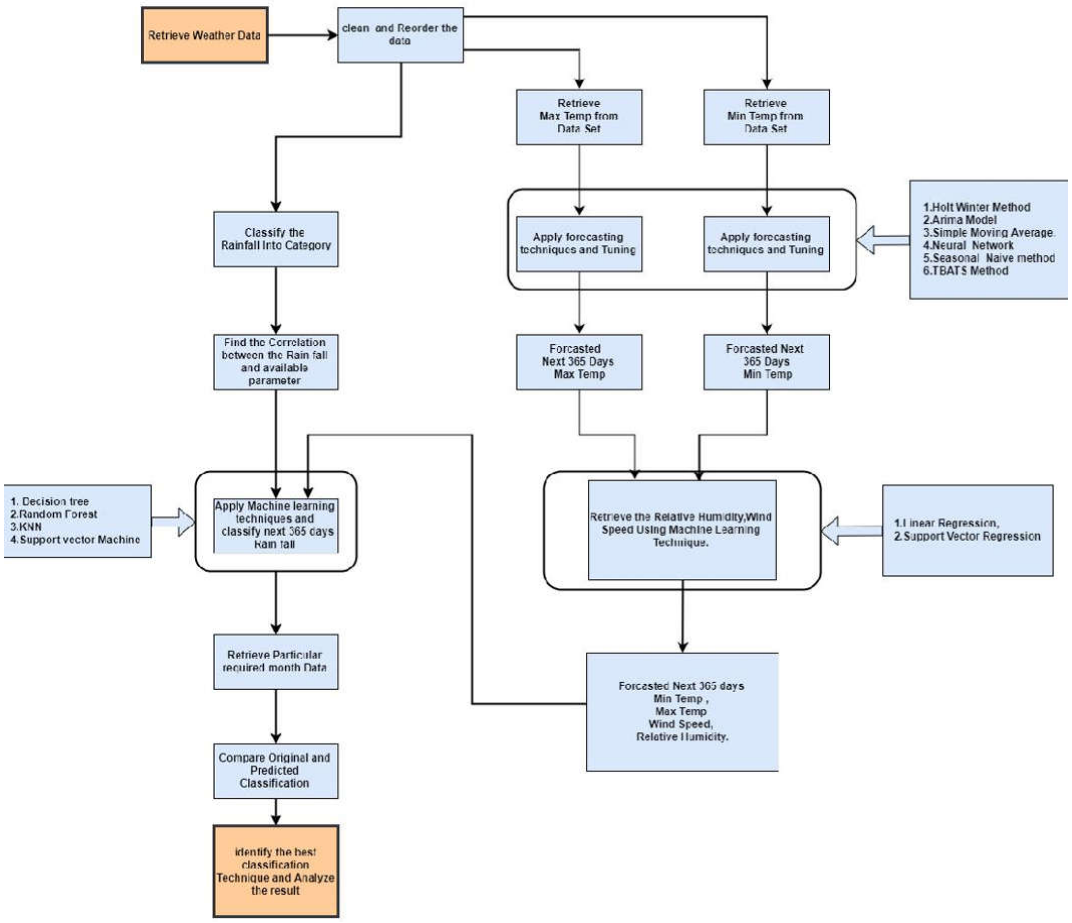


Figure 2 System Architecture

ADVANTAGES

- Accurate prediction of rainfall.
- The ability to recognise trends and patterns in past meteorological data is what machine learning for forecasting rainfall does.
- Such data can be used to create more precise climate models and gain a better understanding of the factors that influence weather patterns.
- Limit threats to lives and property as significantly as achievable, and enhance the management of farms..

DEFINE THE PROBLEM

A more accurate prediction system needs to be used over a timely notify to help minimise dangers to people and property and also help manage agricultural farms easier. Forecasting rainfall is crucial because heavy and inconsistent precipitation can have many effects such as the destruction of crops and farms and damage to property. Each year, people experience disasters like droughts and floods all across the world as a result of excessive rainfall. To assess the rainfall and forecast the likelihood of rain, many models have been devised. Both unsupervised and supervised algorithms for machine learning constitute the foundation of these models. Considering total rainfall won't enable us to determine if it rains under particular circumstances. The main issue

with machine learning is accuracy. To forecast whether it will rain given certain conditions, we must first comprehend the data and train the algorithm accordingly.

MODULES AND FUNCTIONALITIES

Three regional SVR and/or ANN made up the modular models. Two data-preprocessing methods, MA and SSA, were examined using a three-layer feed-forward ANN. The MA was superior to the SSA, according to the results. As test examples, four rainfall records from India and China—India, Zhongxian, Wuxi, and Zhenwan—were employed.

Decision trees, k-nearest neighbours, linear regression, and methods based on rules are examples of algorithms for machine learning used for rainfall prediction.

IMPLEMENTATION

LIBRARIES

Pandas:

Pandas is a sophisticated Python library that is freely available and provides outstanding performance. Applications for processing information and analysis. Python was mostly used for data munging and cleaning. On data analysis, it had little of an effect. Pandas found the answer. Whatever source where the data came from, we may use Pandas to carry out the five typical phases of data processing and analysis: prepare, modify, model, and analyse. Various professional and academic areas, such finance, economics, research, analytics, etc., use Python with Pandas.

Numpy:

NumPy is an universal library that manages arrays. It gives you a very fast multidimensional array object in addition to a way to interact with these arrays. This Python package is key to scientific computing..

It has a number of features, including these important ones:

- A strong object in an N-dimensional array.
- Complex (broadcasting) operations.
- C/C++ and Fortran code integration tools.
- Comprehensive knowledge of the Fourier transform, arbitrary numbers, and linear algebra.

An efficient multi-dimensional data store with numerous uses beyond study is NumPy. NumPy is able to quickly and simply connect with a variety of databases thanks to its ability to define any data-types..

Flask:

The Flask web-based application framework was developed in Python. Armin Ronacher, the team leader of Poocco, a global community of Python followers, developed it. Flask is built on the Jinja2 template engine and the Werkzeug WSGI framework.Both of them are Pocco initiatives. Web framework Flask is a Python tool that renders it simple to develop web apps. It's a microframework without an object relationship manager or any other comparable features, but its core is compact and simple to customise. Yes, it includes multiple outstanding characteristics such as a template engine and url routing. It is a WSGI web-based application framework.

Datatime:

Python comes with a built-in module called DateTime that can handle times and dates in a variety of ways. We'll look at some fundamental Python DateTime operations in this tutorial. The datetime module contains the following six primary object classes and their corresponding components:

1. datetime.timezone
2. datetime.datetime
3. datetime.tzinfo
4. datetime.time

5. datetime.timedelta
6. datetime. date

Pickle:

The pickle module implements binary protocols for serialising and deserializing Python object structures. Pickling is the process of converting a Python object structure onto a byte stream. Unpickling is the reversal process of Pickling in which a byte stream is turned into an object hierarchy..

ALGORITHMS

The Flask application uses a machine learning model to predict whether it will be a sunny or rainy day based on user input. However, the specific machine learning algorithm used for prediction.

The steps involved in the algorithm are as follows:

1. The Flask application is initialized with the necessary dependencies and template folder.
2. The pre-trained model is loaded from the "cat.pkl" file using the pickle.load() function and stored in the model variable.
3. The home route ("/") is defined, which renders the "index.html" template when accessed via a GET request.
4. The prediction route ("/predict") is defined, which handles both GET and POST requests. If a POST request is received, the code retrieves the input data from the form submitted by the user.
5. The input data is processed, including extracting the day and month from the provided date, and converting various form inputs to appropriate data types (float in this case).
6. The processed input features are stored in a list named input_lst.
7. The machine learning models predict () method is called with input_lst as the input to make the prediction. The predicted output is stored in the output variable.
8. Based on the predicted output, the application renders different templates ("after_sunny.html" or "after_rainy.html") to display the appropriate message to the user.
9. If the request method is not POST (e.g., it is a GET request), the application renders the "predictor.html"

TESTING THE SYSTEM

The goal of evaluation is to find flaws. Testing is the practise of attempting to find every possible flaw or vulnerability in a product of work. It additionally makes it feasible to test various components, subassemblies, assemblies, and/or complete products. Software testing is the method for guaranteeing that it complies with requirements, meets customer expectations, and does not malfunction in a way that is unacceptable. There are different types of tests. Every test type was developed to satisfy a particular testing requirement.

TEST CASE

TEST ID	TEST DESPRICTION	INPUT	EXPECTED OUTPUT	ACTUAL OUTPUT	RESULT
1	Collecting information	Temperature information	Rainfall	Rainfall	success
2	Collecting information	Temperature information	Rainfall	Rainfall	success

3	Collecting information	Temperature information	Rainfall	Windy	success
4	Collecting information	Temperature information	NoRainfall	Sunny	success
5	Collecting information	Temperature information	NoRainfall	Windy	success

Table.1. Test Case Report

Test Results: All of the abovementioned test cases were effective. There were however no errors detected..

RESULTS

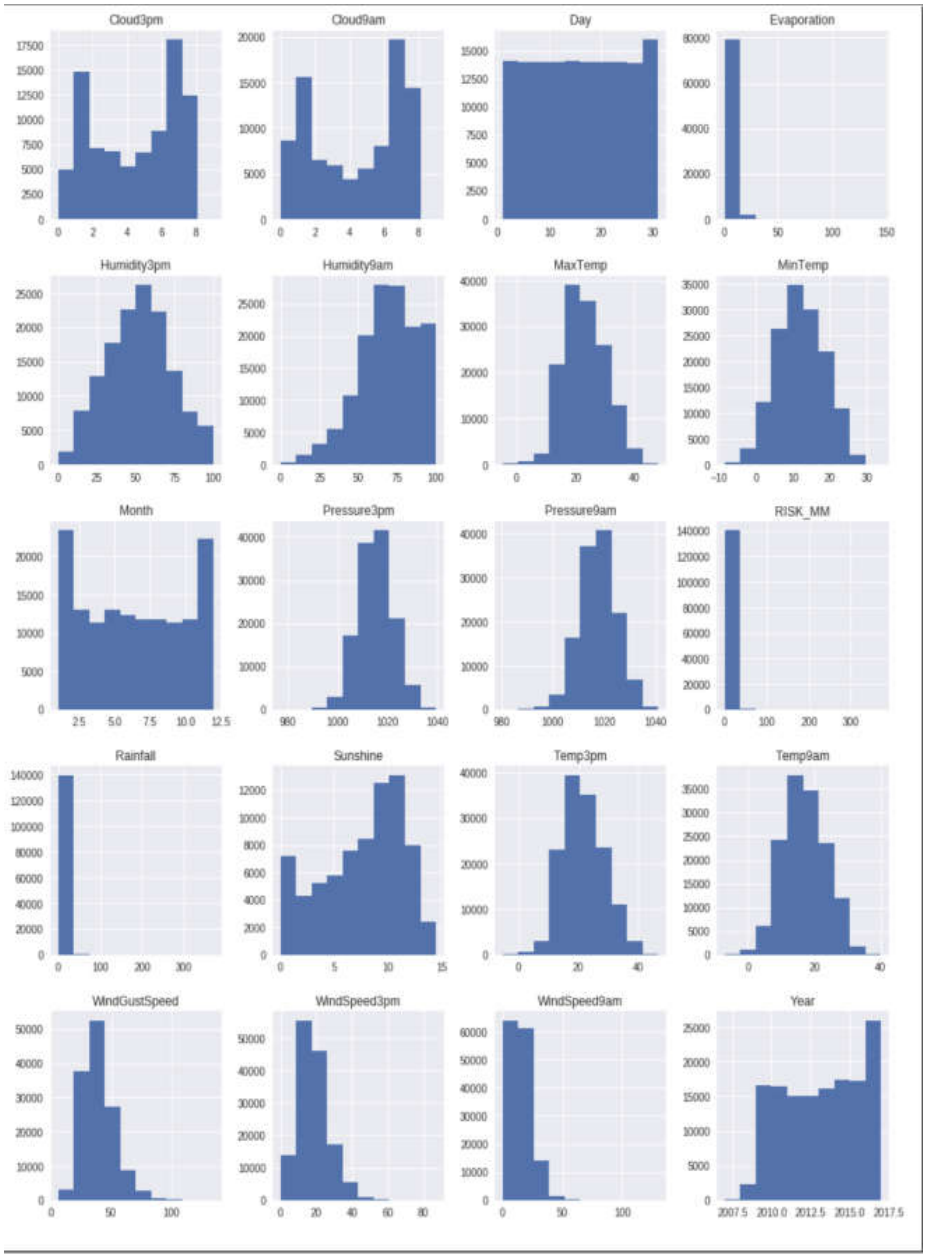


Figure.3. Univariate Visualization

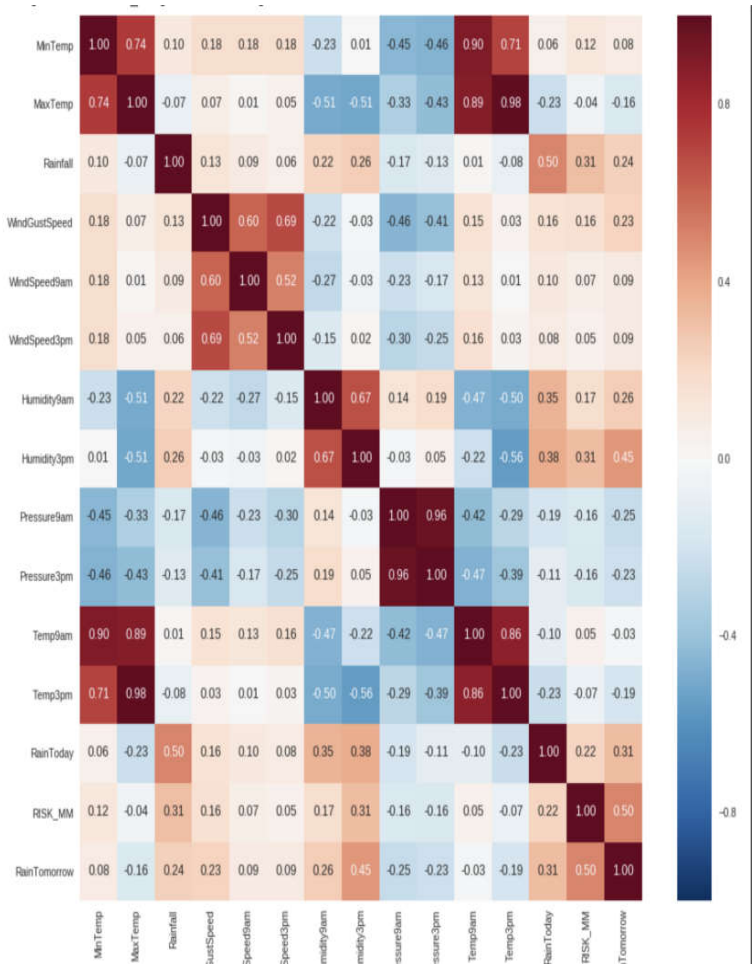
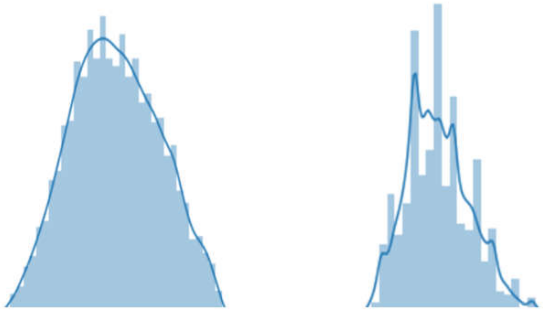


Figure.4. Heat Map



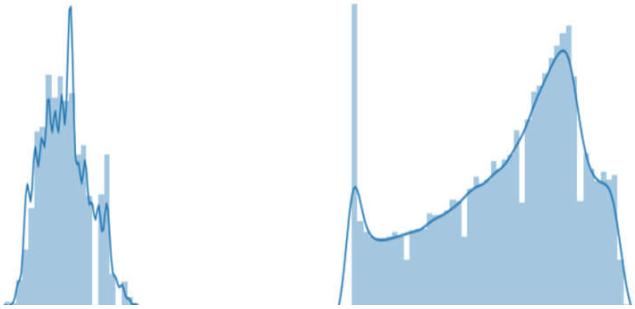


Figure 5 Weather

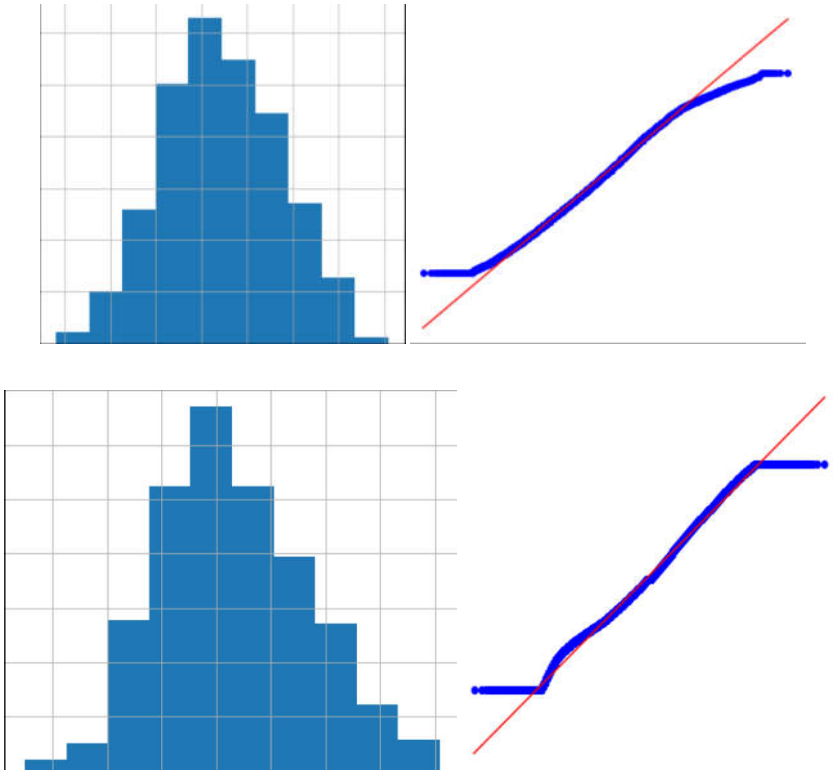


Figure 6 Rainfall

CONCLUSION

The general goal is to describe various machine learning approaches that can be used to predict rainfall. The purpose of this study is to create precise and effective models with fewer features and tests. The data is first pre-processed before being used in the modelling process. The most efficient categorization algorithms are K-Nearest Neighbour (87% efficiency) and Random Forest (88% efficiency). The Decision Tree classifier, on the other hand, has the lowest accuracy (73%). This research can be expanded to include other ML approaches such as time-series data, clustering and association rules, and other ensemble techniques. Given the limits of this research, there will be a need to develop more complicated and combined models in order to improve the accuracy of rainfall forecast systems. A study may also be made utilising more detailed monitoring for a specific

area and creating this type of model for a large dataset so that the computation rate can be enhanced with greater precision and accuracy.

FUTURE SCOPE

Rainfall prediction is critical due to high and unpredictable precipitation can have many consequences such as crop destruction and property damage, so an improved forecasting model is required for warning in advance that can minimise dangers to life and property while also better managing agricultural farms.As a result of early projections, we can even take preventative measures..

REFERENCES

[1] Manojit Chattopadhyay, Surajit Chattopadhyay, "Elucidating the role of topological pattern discovery and support vector machine in generating predictive models for Indian summer monsoon rainfall", Theoretical and Applied Climatology, pp. 1-12, July 2015, DOI: 10.1007/s00704-015-1544-5.

[2] Kumar Abhishek, Abhay Kumar, Rajeev Ranjan, Sarthak Kumar," A Rainfall Prediction Model using Artificial Neural Network", 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC 2012), pp. 82-87, 2012.

[3] Minghui Qiu, Peilin Zhao, Ke Zhang, Jun Huang, Xing Shi, Xiaoguang Wang, Wei Chu, "A Short-Term Rainfall Prediction Model using Multi-Task Convolutional Neural Networks", IEEE International Conference on Data Mining, pp. 395-400, 2017, DOI 10.1109/ICDM.2017.49.

[4] Aswin S, Geetha P and Vinayakumar R, "Deep Learning Models for the Prediction of Rainfall", International Conference on Communication and Signal Processing, April 3-5, 2018, India, pp. 0657-0661.

[5] Xianggen Gan, Lihong Chen, Dongbao Yang, Guang Liu, "The Research Of Rainfall Prediction Models Based On Matlab Neural Network", Proceedings of IEEE CCIS2011, pp. 45- 48.

[6] Sam Cramer, Michael Kampouridis, Alex A. Freitas and Antonis Alexandridis, "Predicting Rainfall in the Context of Rainfall Derivatives Using Genetic Programming", 2015 IEEE Symposium Series on Computational Intelligence, pp. 711 – 718.

[7] Mohini P. Darji, Vipul K. Dabhi, Harshadkumar B.Prajapati, "Rainfall Forecasting Using Neural Network: A Survey", 2015 International Conference on Advances in Computer Engineering and Applications (ICACEA) IMS Engineering College, Ghaziabad, India, pp.706 – 713

[8] Sandeep Kumar Mohapatra, Anamika Upadhyay, Channabasava Gola, "Rainfall Prediction based on 100 years of Meteorological Data",2017 International Conference on Computing and Communication Technologies for smart Nation, pp.162 – 166.

[9] Sankhadeep Chatterjee, Bimal Datta, Soumya Sen, Nilanjan Dey, "Rainfall Prediction using Hybrid Neural Network Approach", 2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), pp. 67 – 72.

[10] Mr. Sunil Navadia, Mr. Pintukumar Yadav, Mr. Jobin Thomas, Ms. Shakila Shaikh, "Weather Prediction: A novel approach for measuring and analyzing weather data", International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017), pp. 414.