Developing a Smart Data Visualization Model to Enhance the Efficacy of Predicting Rainfall

Teesha Ahuja

Bharati College, University of Delhi, New Delhi, India

ABSTRACT

Due to the development of climate observation systems like satellite meteorological observation and the expansion of weather data collection, weather forecasting has entered the Data Visualization era. Therefore, the traditional intelligence models must be better able to predict the weather accurately. Therefore, machine learning-based methods are essential for processing huge datasets that can learn from previous data and make more accurate predictions. Using datasets that are freely accessible, this study provides an in-depth analysis of various methods for predicting rainfall.

Due to the annual change in climate phenomena, statistical forecasting methods cannot accurately forecast long-term rainfall. This paper discusses potential future research approaches and provides a precise classification of rainfall forecasting models.

INTRODUCTION

A forecast computes or estimates upcoming events, particularly economic trends or weather. For instance, many nations, particularly those that rely so heavily on agricultural fields, in tropical regions with only two seasons must forecast rainfall to determine when production and planning should begin. Natural fauna and flora are formed in large part by rainfall. It is important to humans and other species, including plants, animals, and all living things. It significantly impacts farming and agriculture; One of the most abundant natural resources on Earth is water. The country's changing climate and the growing effects of global warming have made it hard for people and the planet to get the rain they need to meet their needs and use it all the time. As a result, it has become increasingly important to examine the shifting patterns of rainfall and attempt to predict it, not only for human requirements but also for the prevention of potential natural disasters brought on by unexpectedly heavy rains. In various spheres and dimensions, accurate rainfall forecasting is crucial; By taking appropriate safety precautions before any natural disaster, you can lessen the impact of sudden and heavy rain. Predicting rainfall has been the focus of computer scientists and engineers, to be more specific, aware of the devastating climate change and up to date. Physical models and Probability Mode generate the dynamic approach and predictions; based on a set of equations that can predict rain in the future. Numerical weather predictions are computerized weather forecasts based on equations. Based on the current weather conditions, numerical weather prediction (NWP) uses mathematical ocean models to predict the weather. Meteorologists have developed atmospheric models that approximate the change in temperature, humidity, and other variables to predict the weather numerically. Using equations from math.

1) Description of Dataset 1: From 1901 to 2015, each state's rainfall measurement is included in Dataset 1.

2) The first dataset contains 19 attributes for 36 sub-divisions (individual months, annual, and combinations of three consecutive months).

3) For some subdivisions, the data are only available from 1950 to 2015. The amount of precipitation expressed as mm is the attribute.

4) Description of Dataset 2: District-by-district Rainfall (in mm) calculated using data for 1951-2000 is referred to in the data.

METHODOLOGY

It involves several steps in the visualization of data. This section explains each step in data visualization, from studying the datasets to plotting them in graphs.

1) Examining the data sets is one way to examine the issue. Identifying the NaN values in the datasets is possible by studying them. The type of visualization for the dataset is chosen at this point. The problem statement becomes clearer and easier to comprehend when studying the dataset.

2) Cleaning the Datasets Errors in the data lead to incorrect conclusions. User error may be the cause of these dataset inconsistencies. Data cleaning is used to correct (or remove) the inaccurate records in the dataset to fix these inconsistencies. The incorrect, incorrect, incomplete, irrelevant, and faulty parts of the data can be removed or replaced with median values through data cleaning. There are numerous data cleaning tools available. Data is cleaned using the panda's library in this paper. Data Pre-processing is completed after these irregularities in the datasets have been removed.

3) Data Pre-processing: High-quality data must be consistent, accurate, and precise. To meet a dataset's quality standards, data processing is used. Preprocessing data is very important because it helps get better results and improves the raw data quality. The raw data is transformed into one that is more refined and useful through data preprocessing. Data preprocessing in this report includes grouping data according to their states, replacing NaN values with median values, and grouping data in ascending order. The program's efficiency and reliability are both enhanced by these modifications.

A. Algorithm and Methods

Rainfall prediction is made possible by this predictive model. To carry out the analysis, the first thing that needs to be done is to clean and format the data. After that, variations in the patterns of rain will be observed. By dividing the dataset into training and testing sets, we can predict the amount of rain. We compare and analyze the project's methods using various statistical and machine-learning approaches. We reduce errors as much as possible by employing various strategies.

RESULT

According to numerous reference papers, the following models were found to be suitable for studying rainfall patterns:

1) Multiple Linear Regression

2) Random Forest Regression

3) SGD Regression

4) Support Vector Regression

5) After analyzing the Random Forest Regression, it was discovered that, in comparison to all other machine learning methods, it accurately predicted all seasonal variations and patterns of rainfall in India.

6) Import libraries and define functions using matplotlib to plot the data in this exploratory analysis first. Not all plots can be plotted depending on the data because the number of attributes can be greater or lesser.

7) Converting and cleaning the dataset into the appropriate format to make accurate rainfall predictions.

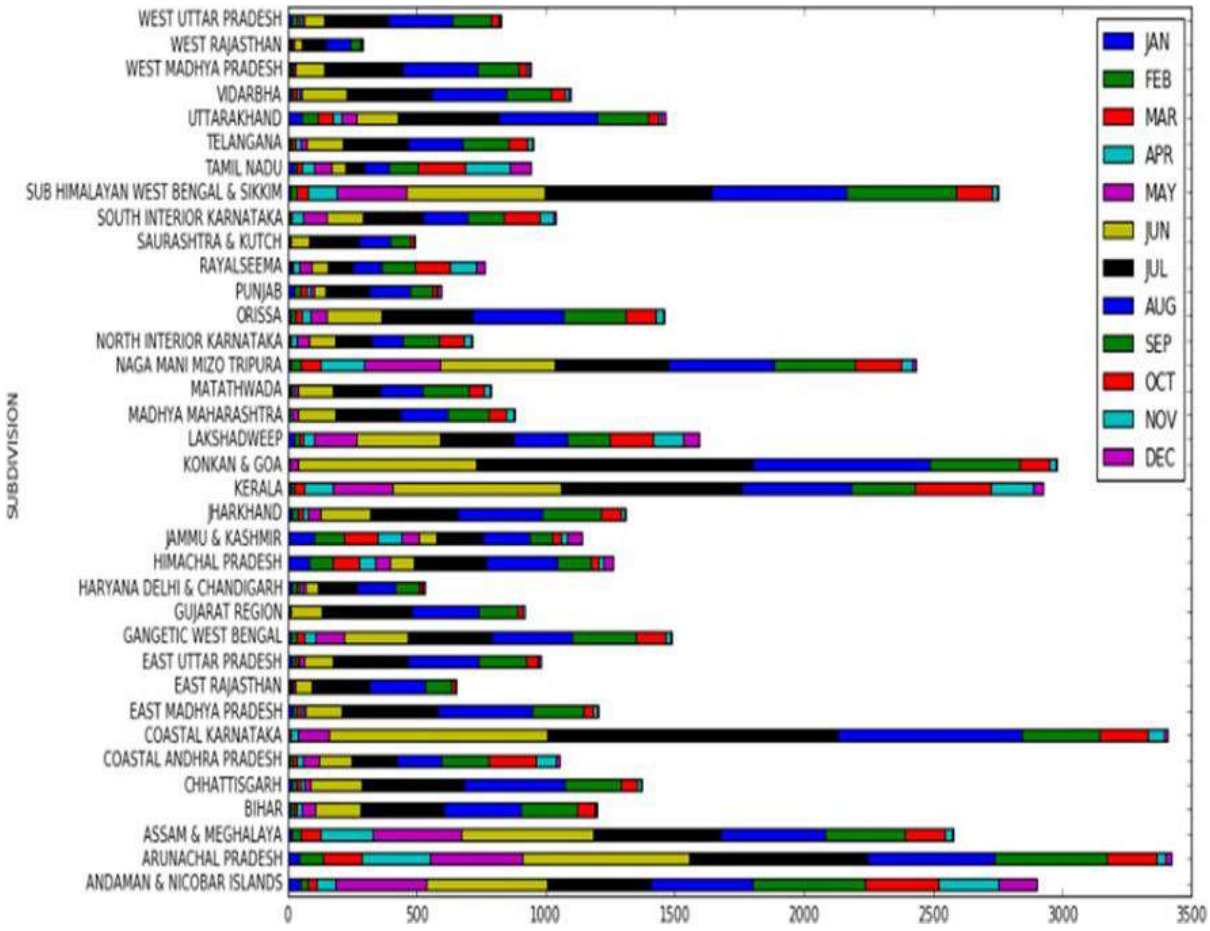
8) Conduct a thorough analysis of the data and observe variations in rainfall patterns across the country.

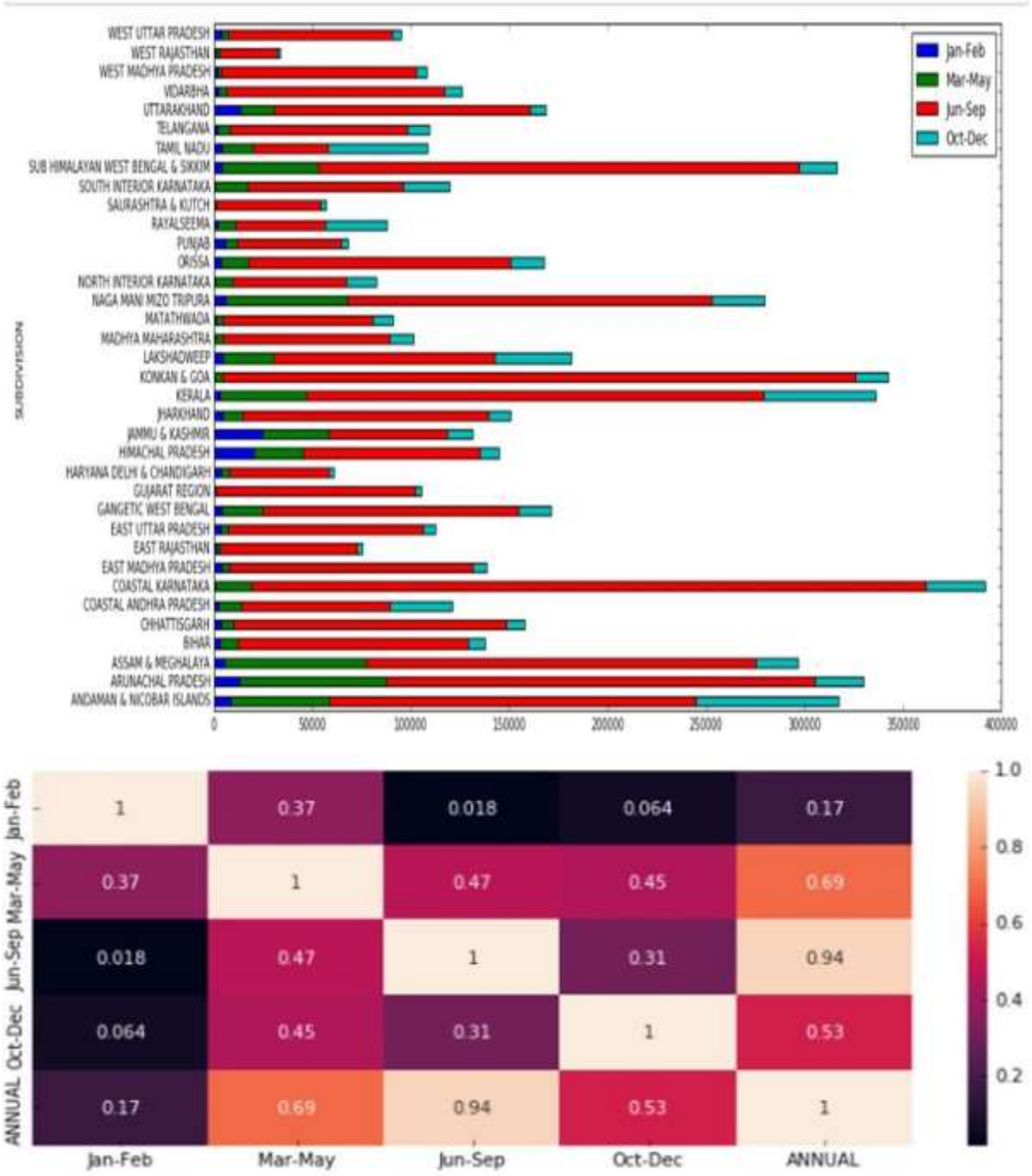
9) By dividing the data into training and testing sets, we can predict the typical amount of rain. In the project's prediction and analysis, we use various statistical and machine learning methods (SVM, random forest regressor, linear regression). We reduce the country's rainfall prediction error using various methods.

10) Bar graphs that show how much rain falls in different places. The distribution of annual, monthly, and month-by-month rainfall amounts.

Distribution of rain in monthly districts, subdivisions, and groups of months

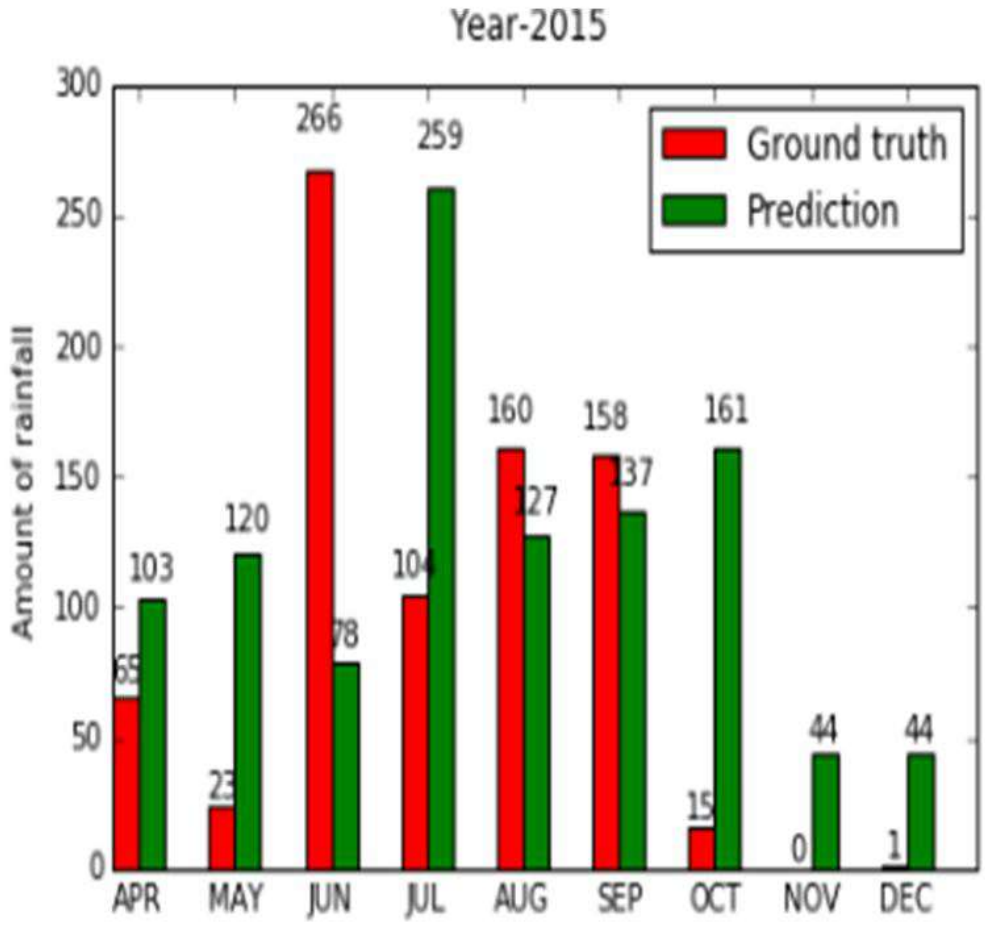
11) Heat maps show a link between the amount of rain that falls in different months.



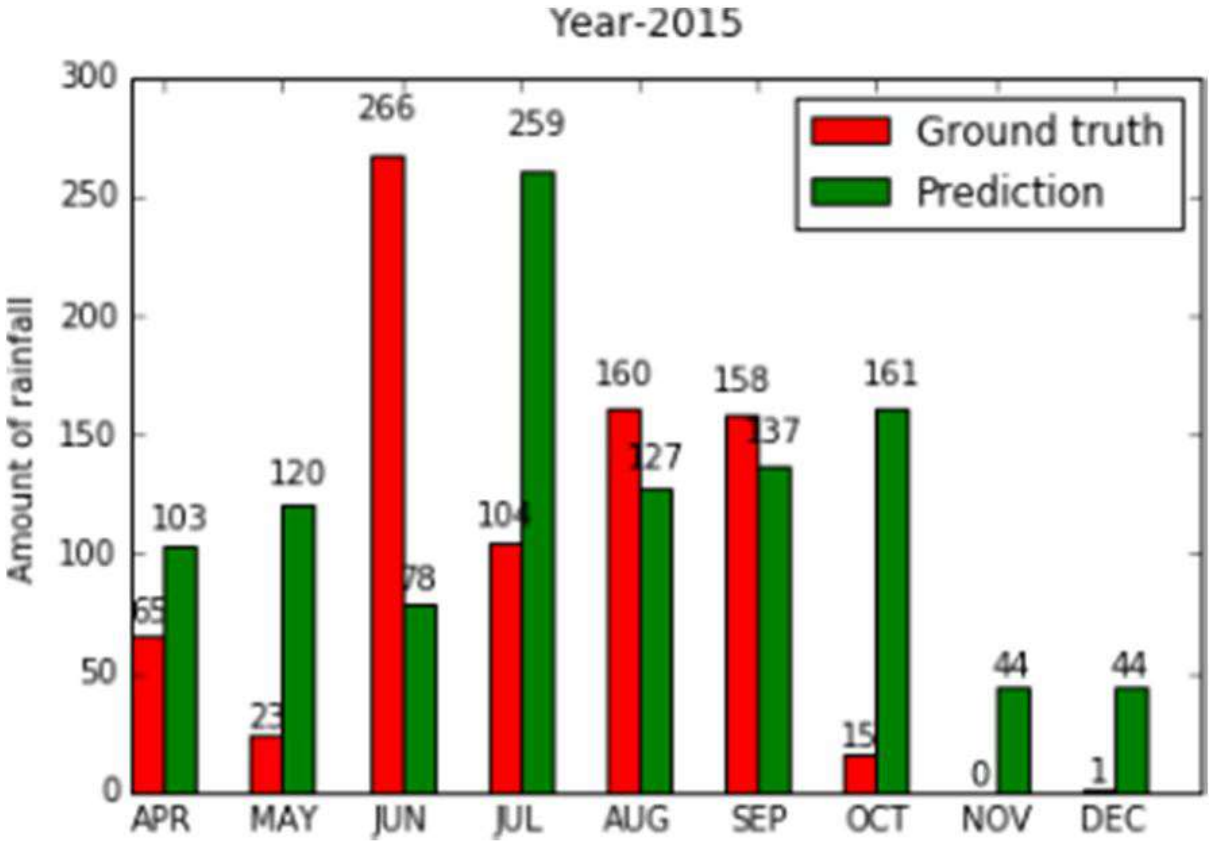

WEST UTTAR PRADESH
WEST RAJASTHAN
WEST MADHYA PRADESH
VIDARBHA
UTTARAKHAND
TELANGANA
TAMIL NADU
SUB HIMALAYAN WEST BENGAL & SIKKIM
SOUTH INTERIOR KARNATAKA
SAURASHTRA & KUTCH
RAYALSEEMA
PUNJAB
ORISSA
NORTH INTERIOR KARNATAKA
NAGA MANI MIZO TRIPURA
MATATHWADA
MADHYA MAHARASHTRA
LAKSHADWEEP
KONKAN & GOA
KERALA
JHARKHAND
JAMMU & KASHMIR
HIMACHAL PRADESH
HARYANA DELHI & CHANDIGARH
GUJARAT REGION
GANGETIC WEST BENGAL
EAST UTTAR PRADESH
EAST RAJASTHAN
EAST MADHYA PRADESH
COASTAL KARNATAKA
COASTAL ANDHRA PRADESH
CHHATTISGARH
BIHAR
ASSAM & MEGHALAYA
ARUNACHAL PRADESH
ANDAMAN & NICOBAR ISLANDS
SUBDIVISION
Jan-Feb
Mar-May
Jun-Sep
Oct-Dec
0
50000
100000
150000
200000
250000
300000
350000
400000

Jan-Feb
Mar-May
Jun-Sep
Oct-Dec
ANNUAL
1 0.37 0.018 0.064 0.17
0.37 1 0.47 0.45 0.69
0.018 0.47 1 0.31 0.94
0.064 0.45 0.31 1 0.53
0.17 0.69 0.94 0.53 1
1.0
0.8
0.6
0.4
0.2

We organized the data so that, given the rainfall over the previous three months, we could predict the rainfall for the following month.



The accuracy with which the feature space is described in the training, testing, and validation data sets is crucial. Any division may be sufficient if the total number of points in the data set is large; however, the division ratio may be crucial if the data set is small. We used training and test ratios of 80:20 for all of the experiments.



CONCLUSION

The purpose of this paper is to provide a comprehensive account of how various machine learning methods have been utilized extensively over the past two decades to predict rainfall.

According to the literature review, most researchers used artificial neural networks to predict rainfall accurately. The research paper also concludes that machine learning forecasting strategies are superior to other statistical and numerical strategies for predicting rainfall. The city's rainfall prediction using three methods: supporting linear regression, random forest, and the vector machine. For successful prediction, the input data are cleaned and normalized in a pre-processing stage before being classified by a classification system.

REFERENCES

[1] R Vijayan, V Mareeswari, P Mohankumar, G Gunasekaran K Srikar, (2020), 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

[2] Dhawal Hirani, Dr. Nitin Mishra (2016), A Survey On Rainfall Prediction Techniques, *International Journal of Computer Application* (2250-1797) Volume 6– No.2, March April 2016, 28

[3] G. Tripathi and R. Gupta, (2019), 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] Sanjay Chakraborty, Prof. N.K.Nagwani, Lop Amudra Dey, (2012), Weather Forecasting using Incremental K means Clustering, *Data Mining and Knowledge Engineering*, Vol 4, No. 5

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

[6] Yogesh Kumari Joshi, Uditi Chawla, Shipra Shukla, (2020), Rainfall Prediction Using Data Visualisation Techniques, *10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*

[7] Pulak Guhathakurta, M. Rajeevan (2008), Trends in the rainfall pattern over India, *International Journal of Climatology*, September 2008.