COMPREHENSIVE ANALYSIS AND FUTURE PROJECTIONS OF INDIAN RAINFALL PATTERNS (2020-2023)

Raina Ann Baby
PG Scholar
PG Department of Computer Science
Sree Sankara Vidyapeetom college
Valayanchirangara, Perumbavoor
Ernakulam , Kerala
rainaannbaby@gmail.com
Dr. Manusankar C
Assistant Professor and Head
PG Department of Computer Science
Sree Sankara Vidyapeetom college
Valayanchirangara , Perumbavoor
Ernakulam , Kerala
manusankarc@ssvcollege.ac.in

ABSTRACT

In meticulous analysis of rainfall patterns in the Indian states during the period spanning January 2020 to July 2023.Employing data sourced from the Open Government Data Platform, specifically curated by the India Meteorological Department (IMD GRID MODEL) and the National Remote Sensing Centre (NRSC VIC MODEL), this investigation aims to discern the monthly average rainfall trend across diverse region of India. The study's findings offer information about historical rainfall distribution and variation. A paramount outcome of this endeavor is the computation of anticipated average rainfall for each state in India over the imminent five months of 2023. By employing sophisticated statistical and analytical methodologies, this study provides a thorough comprehension of the prevalent rainfall patterns across the Indian states. The research unveils intricate precipitation trends, discerning variations among diverse geographical locales and temporal segments.

Keywords: Rainfall analysis, Indian states, India Meteorological Department, National Remote Sensing Centre, Year, Month, Average monthly rainfall, Machine learning, Linear Regression Algorithm.

1. INTRODUCTION

The study undertakes a rigorous examination of the intricate shade of rainfall patterns across the diverse state of India over the time period from January 2020 to July 2023. The empirical foundation of this inquiry is drawn from the Open Government Data Platform, substantively stoked by data benefactions from the recognized India Meteorological Department (IMD GRID MODEL) and the complete National Remote seeing Centre (NRSC VIC MODEL). The core of this research focuses on the comprehensive examination of average monthly precipitation throughout the whole spectrum of Indian states.

Embedded within the framework of this research is the pivotal role of machine learning, an epochal facet of artificial intelligence. Its cardinal ideal resides in the emulation of mortal cognitive processes, endowing machines with the capacity to autonomously ripen perceptivity and foster informed decision making through the illumination of data-driven exposures. A major goal of machine learning is to automate the creation of analytical models, thereby enabling computers to extract advanced insights from data without the need for explicit programming. Central to this effort is a training phase that coordinates model refinement to uncover complex patterns and potential relationships hidden in the dataset.

This study unravels the conspiracy of machine learning machines to predict average rainfall trends within Indian states. The most important of these methods exploits the power of linear regression algorithms. This is the key arrow of machine learning. In order to understand the complicated interactions between dependent and independent variables, this algorithm uses a linear model as its fulcrum. In this study, it aids in forecasting future average rainfall in Indian states, contributing to a comprehensive understanding of forthcoming precipitation dynamics.

It’s worth nothing that data repositories have undergone a transformation, state and district naming conventions have been coded, and strategic efforts have been made to increase analytical power. The basic features used for this analysis include strategic coding of states and time series breaks of months. The interdisciplinary approach culminates in the formulation of an avant-grade framework aimed at providing prophetic insights into upcoming average precipitation at monthly intervals for each state of India. This lucky foresight expands our cognitive toolkit and greatly improves our ability to predict and understand how average rainfall will evolve over the next five months in 2023.

2. LITERATURE REVIEW

In this paper [1] “A Comparison and Ranking Study of Monthly Average Rainfall Datasets with IMD Gridded Data in India” by Vasala Saicharan addresses the critical need for accurate rainfall measurement in hydrological applications by evaluating the performance of seven rainfall datasets widely used in India at different spatial and temporal resolutions. Using smart guage analysis and rule-based decision trees as benchmarks and data from the India Meteorological Department, the study identifies optimal datasets for different levels of analysis. In particular, GPM seems to be strong, especially in pixels and meteorological subdivision, while TerraClimate lags behind. At the subdivision level, combined datasets are preferred. The effectiveness of GPM is also confirmed by time analyses. Overall, this study will help select materials for hydrological and agricultural applications that contribute to sustainable development.

In this paper [2] “Predictive Analytics for Rainfall Prediction”, by Lakshmi S R, Asha N, K C Gouda discussed the critical role of rainfall in food production and water management, especially in agrarian economies like India. Economic links to agriculture recognize that accurate rainfall is becoming central, reinforced by concerns about climate change and advances in high-performance computing. Users from farmers to decision makers need more accurate forecasts for planning, using time and space information. Recent record expansions allow novel methods, which includes predictive analytics, facilitating versatile multi-scale forecasting. A predictive device has been crafted for real-time India-wide rainfall forecasts. In summary, the text emphasizes the importance of forecasting ability using large-scale climate data and presents a forecast analysis package for low-cost periodic forecasting using cloud technology and big data, which represents a breakthrough in accurate precipitation forecasting.

In this project [3] “A Study on Rainfall Prediction Techniques” by Dr. C K Gomathy focuses on rainfall, which is an important aspect because it can lead to disasters. Accurate forecasts help prevent disasters because short-term forecasts are more accurate than long-term forecasts. This challenge is accurate long-term forecasting, especially for heavy rainfall events that effect the economy and human well being. This venture aims to make rainfall prediction strategies handy to non-experts, in particular through machine learning processes like Support Vector Machine (SVR). The find out about concludes that SVR is a precious and adaptable strategy, addressing problems associated to records distribution, model geometry, and overfitting SVR outperforms Multiple Linear Regression (MLR) in taking pictures non-linear relationships in data, as confirmed through Mean Absolute Error (MAE) evaluation. The study recommends using linear and RBF kernels for linear and non-linear relationships. Adjusted SVR provides the best predictions of all comparable models, including simple linear regression (SLR) and standard SVR.

In this paper [4] “A comparative Study of Long Short-Term Memory for Rainfall Prediction in India” by Chawngthu Zoremsanga and Jamal Hussain focus on forecasting the monthly average rainfall in India due to its important environmental and social implications. The study uses four separate LSTM models to improve forecast accuracy and compare their performance with a recognized reference model. LSTM models are trained and evaluated on all India historical monthly rainfall data from 1871 to 2016. The models are fine-tuned using the MSE (Mean Squared Error) loss function and Adam’s optimization technique. Performance evaluation is based on statistical measures such as mean absolute error (MAE) and root mean square error (RMSE). They learn about displays that augmenting the number of neurons and making use of stacked LSTM layers leads to improved model performance. Notably, LSTM Model_4 attains an RMSE of 245.30, surpassing the benchmark model’s RMSE of 251.63, signifying the effectiveness of the proposed procedures in rainfall prediction and the use of deep learning techniques.

In this paper [5] “Short to medium range impact based forecasting of heavy rainfall in India” by Mrutyunjay Mohapatra discussed significant progress has been made in understanding heavy rainfall during the monsoon season due to advances in observational and numerical modeling of the monsoon phenomenon. These advances have resulted in significant improvements in short and medium-range (up to five days) heavy precipitation forecasts, with a 0% improvement in the accuracy of heavy rainfall forecasts over the last five years (2018-2022) compared to the previous ones. Despite improved forecasting and warning capabilities, reducing loss of life and property requires a broader approach that includes hazard forecasting systems, effective assessment models and stakeholder engagement to implement risk warnings and effective actions. To address this issue, the India Meteorological Department (IMD) launched Impact Based Forecasting (IBF) for heavy rainfall. Initially at the meteorological sub-regional level in July 2023 and then in August 2019 at the regional and city levels. This approach estimates the potential impact of heavy rain in different areas and describes the necessary response measures. The IBF has evolved over time and currently includes meteorological, geophysical, geospatial and socio-economic components, facilitated by a web-based GIS-based decision support system (DSS). This find presents an overview of the improvement degrees of IBF for heavy rainfall in India, highlighting its possible to beautify useful resource management, disaster preparedness, and city planning, finally leading to a reduction in loss of lifestyles and property.

In this paper [6] "A machine learning approach for probabilistic multi-model ensemble predictions of Indian summer monsoon rainfall" by Nachiketa Acharya and Kyle Joseph Chen Hall focuses on improving the accuracy of Indian Summer Monsoon Rainfall (ISMR) forecast by adopting a Probabilistic Multi-Model Ensemble (PMME) approach. Although traditional methods suffer from limitations such as overconfidence or imprecise distributional assumptions, this study proposes an innovative solution using a modified version of an extreme learning machine (ELM) that performs probabilistic predictions, the probabilistic output extreme learning machine (PO-ELM). PO-ELM-based PMME is rigorously evaluated over a 37-year period using sigmoid additive neurons and adaptive linear programming. The results demonstrate the effectiveness of this machine learning-based strategy in producing skillful and reliable ISMR multi-model ensemble forecasts, contributing to improved decision-making in the Indian monsoon situation.

3. DATA AND METHODOLOGIES

The construction of an analytical model, this approach enabled accurate forecasts for the next five months until 2023, providing valuable information to sectors dependent on rainfall patterns and helping to make informed decisions in sectors.

3.1. Dataset Description

The data utilized historical rainfall data sourced from the India Meteorological Department and National Remote Sensing Centre spanning January 2020 to July 2023.Data for each month from January 2020 to July 2023 have been aggregated into a single data set.

The tabulated data, structured the use of the pandas library in Python, captures rainfall records for a variety of Indian states and districts. The DataFrame contains columns such as 'state' indicating the state, 'district' indicating the area, 'date' indicating the date of measurement, 'year' indicating the year, ‘month’ indicating the month of data collection, and average rainfall measured average precipitation. This dataset includes 911180 rows, imparting giant archives of rainfall observations. The ‘Avg_rainfall’ column might contain missing values marked as NaN. The investigation and analysis of rainfall trends across various Indian regions and periods are made simpler by this hierarchical approach.

3.1.1 Data preprocessing

To improve the dataset’s completeness and consistency, a pivotal information preprocessing system was once done. A strategic decision was made to replace missing data with 0s. This clever approach not only ameliorated the potential pitfalls of incomplete records, but also strengthened the overall reliability of the dataset. This procedure is crucial to the research’s framework since it makes sure that any future analyses and conclusions generated from the data are accurate and unaffected by missing data.

A crucial step was taken to enhance the computational analysis of categorical data. This system is instrumental in making ready categorical data for integration into analytical frameworks and computing machine learning models. With the advent of LabelEncoder, the categorical columns of the dataset, especially ‘state’ and ‘district’, were systematically encoded into numerical equivalents. This transition from categorical to numerical form is extremely important to facilitate later analysis, as it provides the data with the necessary coherence for machine learning algorithms. This preprocessing approach optimizes the dataset for useful insights and predictions inside the research.

3.2. Analysis of Dataset

The code fragment is an important data processing. Using the encoded columns in the DataFrame, the code effectively filters and extracts relevant data related to a particular combination of state and district using the code. The resulting subset, stored in a variable, offers a refined dataset tailored to the specified geographic area. Perform a data aggregation process, which calculates the mean average rainfall for each month. This aggregation aids in understanding the temporal variation of rainfall patterns within the specified geographic area.

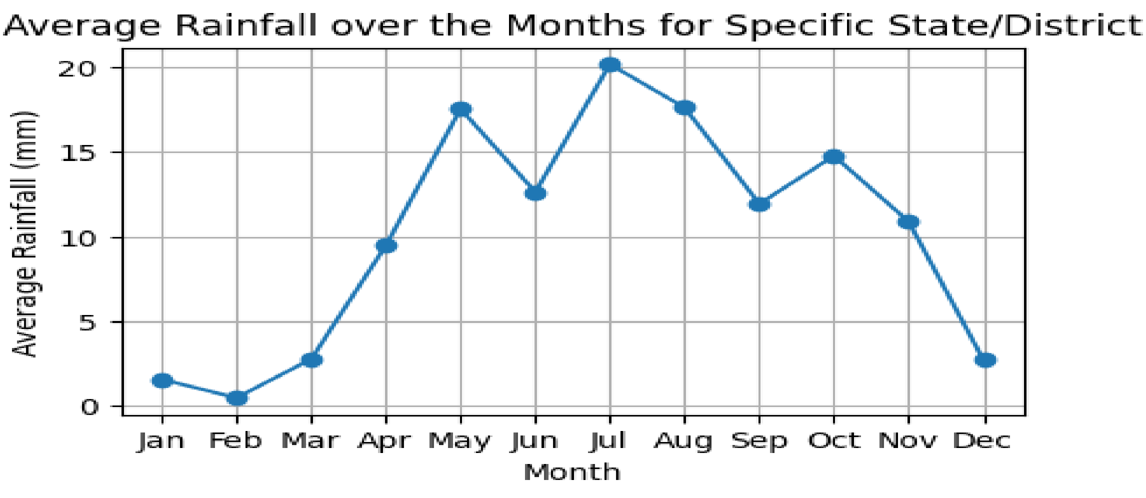

Fig.1 Average rainfall over the months for Ernakulam, Kerala

The dataset contains common rainfall information and first groups it through 'State' and 'Month', calculating the suggested rainfall for every combination. It then creates a pivot table with states as rows, months as columns, and values as average rainfall. This table provides a clear view of how rainfall patterns vary by state and month, helping to visualize and analyze this data.

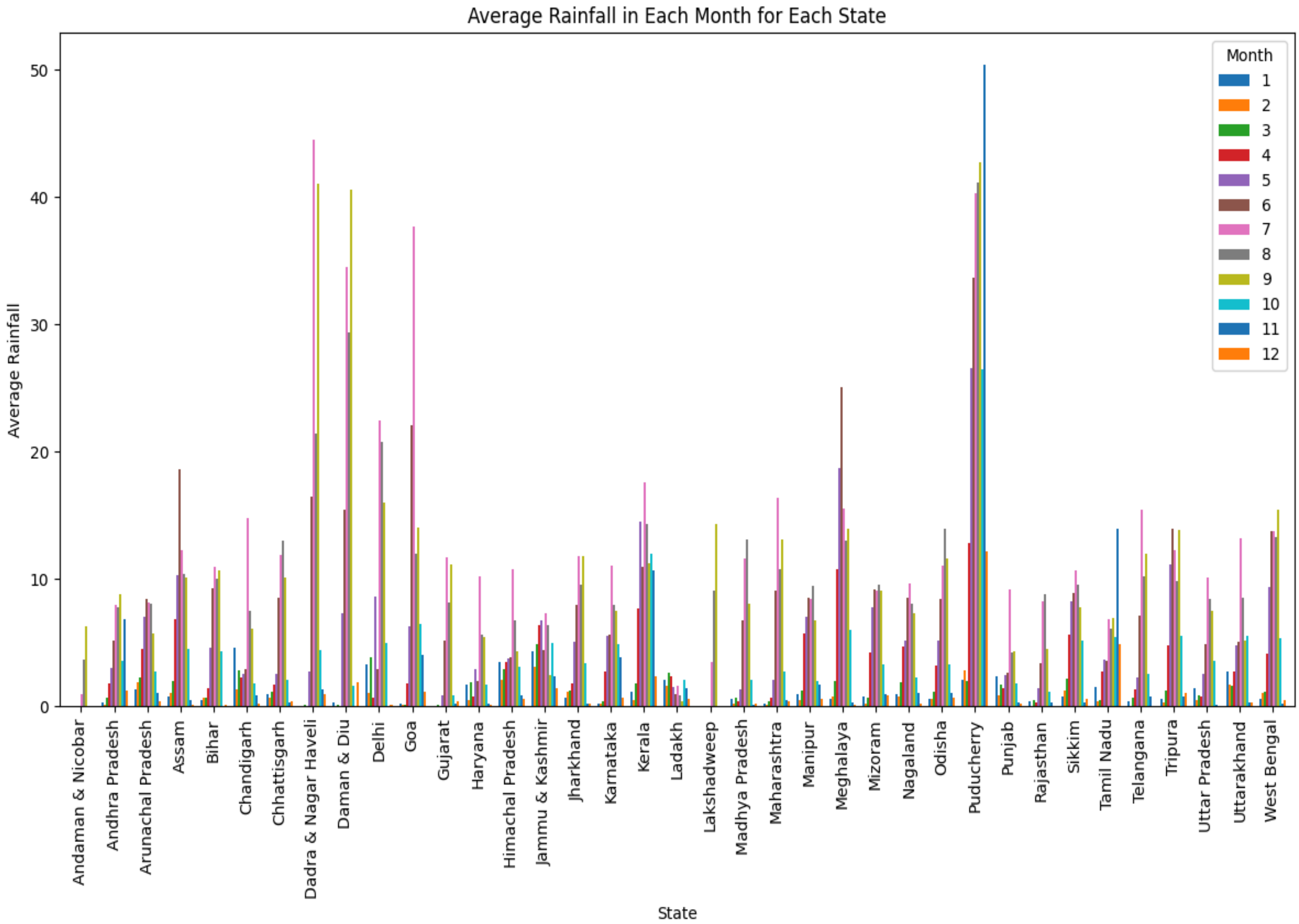

Fig.2 Total average rainfall in each month for each state

3.3 Linear Regression Algorithm

Machine learning is a subset of artificial intelligence decided to create algorithms and statistical models that can learn from data to make predictions. Linear regression is a supervised machine learning technique that maps points in a labeled dataset to optimized linear functions, enabling prediction from new data.

Linear regression is a supervised machine learning technique that calculates a linear relationship between a dependent variable and one or more independent characteristics. It is called univariate linear regression if there is one independent characteristic and multivariate linear regression if there are more. The goal is to decide the most useful linear equation for predicting the based variable's value based totally on the impartial variables. This equation forms a straight line, which means interdependence. The slope of the line describes the change in the independent variable per unit change in the independent variables encapsulating its predictive nature.

3.3.1 Creating Model

Imports critical functionalities from the scikit-learn library for constructing and evaluating machine studying models. The split the data into training and test sets, which is essential for model evaluation. The Linear Regression class provides tools for creating linear regression models that capture relationships between input characteristics and an outcome variable. Defines two variables ‘x’, and ‘y’, where ‘x’ is a specified subset of DataFrame, specifically the ‘State’ and ‘Month’ columns. These columns represent the relevant features used to train the machine learning model. On the other hand, ‘y’ is assigned the values from the ‘Average rainfall’ column of DataFrame which serves as the goal variable for the model. This separation of features and target values forms the base for training a model to predict average rainfall using the decoded state and month information as input features. The split is 80-20, where 80% of the data is reserved for training and 20% for testing. The training set is used to build the model and the test set analyzes. A trained Linear Regression model is developed. An instance of the Linear Regression class from scikit-learn is allocated to the variable. This classification represents a Linear Regression model. Using the training data and target data, the appropriate method is called to train the model. During training, the model is called to train the relationship between input features and target values, trying to find the best-fitting linear equation that can predict the target variable based on the given features. After a model instance is trained, it retains the learned association and can be used to predict new data.

3.3.2 Prediction

A trained linear regression model stored in the variable is used to make predictions based on the test data. The prediction method is applied to the input features, ensuring in the predicted variable. This array includes the predicted values for the goal variable primarily based on the points in the checking out set. Each prediction corresponds to an instance in the checking out information and represents the model’s estimation of the common rainfall, leveraging the discovered relationships between ‘State’ and ‘Month’ features and the ‘Average rainfall’ goal variable from the training process. Then we focus on forecasting the average rainfall for a specific month for a specific state. It first defines a list of target months (August-December) and creates a data container to hold the future predictions. Then narrows down the dataset to data specific to a special state and proceeds to simulate new data entries for the upcoming 5 months in 2023. This simulated information is developed by the usage of the encoded state value and the predefined months list. The simulated data is used to create a new DataFrame. The trained Linear Regression model is subsequently used to forecast rainfall for the fresh data. The model’s predictions for the given months and state are contained in the array. This summarizes the process of generating forecasts of average rainfall over the following months using a trained machine learning model.

4. RESULT

The month which contains the values from August to December, and states which contains the names of the various states and union territories of India. This prepares the list to be created to store future predicted data. This stage for later data processing and analysis related to forecasts for specific months and states.

In accordance with the analytical methodology targeted inside this research study, prognostications related to the suggested precipitation degrees for every individual Indian states throughout the temporal span spanning August through December 2023 are ascertained through the utilization of a pre-existing machine learning model. This effort is driven by a complex set of nested iterative procedures expertly tailored to carefully traverse each month and state. Encoded representations of the states are extracted during these iterations, synchronized with the respective month designations, and then combined into improved model inputs. Employing this meticulously curated dataset, the machine learning model seamlessly computers forecasts of the anticipated average rainfall values. These calculated predictions, elegantly complemented by the accompanying state nomenclatures, are systematically integrated into a cumulative collection. This aggregation of effects culminates in a complete repository of insights, casting illuminating views upon the expected climatic trends for the latter half of 2023. This complex process, guided by the paradigms of historical data and the deep predictive power of the trained model, is evidence of a careful scientific approach that advances our understanding of future climate dynamics. Create a DataFrame that acts like a table where we can store and organize data. The reason for this precise DataFrame is to maintain the predicted information that we've generated. This structure is similar to a spreadsheet, where each row corresponds to a set of predicted values and each column represents a specific aspect of those predictions. This makes it convenient to process and present forecast data in a structured and accessible way.

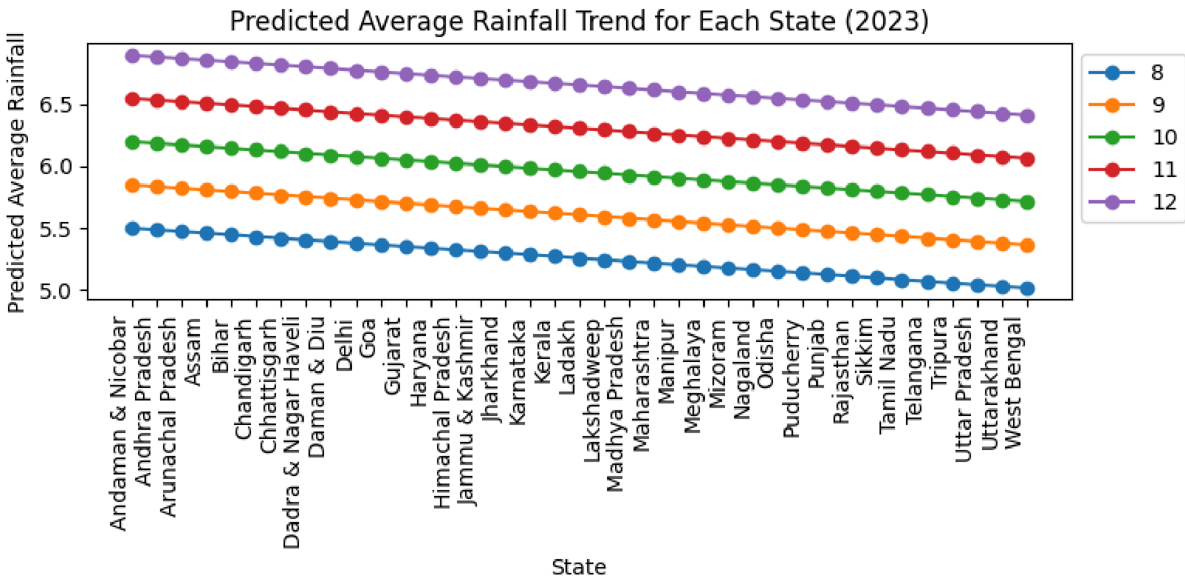

Fig.3 Predicted average rainfall trend in each state for next five months

5. CONCLUSION

This study has ushered in a plethora of discoveries that go beyond the realm of climate science in its quest to unravel the complex story of rainfall patterns across the several states of India between January 2020 and July 2023.This study has provided a high-level understanding of the temporal variability and regional differences that shape the Indian meteorological landscape by examining the average monthly rainfall in each state. Through rigorous examination and integration of diverse features encompassing state name, district names, months, years, and average rainfall data, we have sought to extract meaning trends from the intricate fabric of climate data. The application of a linear regression model carefully trained on the data from 2020 to 2023 has enabled us to embark on our predictive analytics journey. The model's ability to predict average rainfall for each state over the next five months of 2023 highlights the potential of machine learning techniques to make sense of complex weather phenomena. Through the visualizations strictly crafted, the predicted trends have been converted into pictorial representations that transcend bare figures. These visual aids have unlocked the power to discern seasonal shifts, geographic variations, and potentially imperative anomalies. Unfortunately, the lack of real precipitation data for comparison in the study posed challenges in directly assessing the accuracy of the model's predictions, but despite this limitation, the presented methods and results provide a basis for further investigation, refinement and validation in future studies.

REFERENCE

[1] Vasala Saicharan and Shwetha Hassan Rangaswamy" A Comparision and Ranking Study of Monthly Average Rainfall Datasets with IMD Gridded Data in India" Department of Water Resources and Ocean Engineering, National Institute of Technology Karnataka,Surathkal 575025, India. Sustainability , Vol.15 ,Page 5758, 2023

[2] Lakshmi S R, Asha N, K C Gouda "Predictive Analytics for Rainfall Prediction" Department of Computer Science and Engineering, Kanakpura, Bangalore and CSIR Fourth Paradigm Institute NAL Belur Campus, Bangalore, India. International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTCONV4IS29019, Vol. 4 Issue 29, April-2018

[3] Dr. C K. Gomathy, Annapareddy Bala Narasimha Reddy, Aravapalli Pavankumar, Aile Lokesh "A Study on Rainfall Prediction Techniques" Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram. International Journal of Scientific in Engineering and Management (IJSREM) ISSN: 2582-3930, Vol. 05 Issue 10, October-2021

[4] Chawngthu Zoremsanga and Jamal Hussain "A Comparative Study of Long Short-Term Memory for Rainfall Prediction in India" Proceedings of the NIELIT's International Conference on Communication, Electronics and Digital Technology, Mizoram University, Page 547-558, https://www.researchgate.net/publication/371881610

[5] Bushair M T, M.Mohapatra "Short to medium range impact based forecasting of heavy rainfall in India" Indian Meteorological Department, MAUSAM ISBN: 0252-9416, Vol. 74 Issue 2 ,Pages 311-344, March-2023.

[6] Nachiketa Acharya and Kyle Joseph Chen Hall “A machine learning approach for probabilistic multi-model ensemble predictions of Indian summer monsoon rainfall” Cooperative Institute for Research in Environmental Science (CIRES),University of Colorado Boulder,Boulder,CO,United States and National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory Boulder,CO,United States. MAUSAM IMD Vol.74 Issue 2, Page 421-428, April-2023

[7] Sreejith O. P. Kelkar R. R “Meteorological sub-divisions of India and their geopolitical evaluation from 1875-2020” MAUSAM, Vol. 71, Page 571-584, August 2021

[8] Adnan M. Abdulazeez Dastan Maulud “A review on linear regression comprehensive in machine learing.2020

[9] Zhenya Liu, Zhengtong Zhu, Jing Goa, Cheng Xu “Forecast methods for time series data: A survey-2021” IEEE Access. Vol. 9, Page 91896-91912, 2021