**Rainfall Prediction Using Machine Learning: A Comparative Analysis of Classifier Performance**

Project Report in partial fulfillment of the degree

# Bachelor of Technology

in

# Computer Science & Engineering

## By

**2203A51073** CHARUL PAREEK

**2203A51090** SHIFA IMUN

**2203A51100** ZOYA NAUSHEEN

**Under the Guidance of**

# Dr. Mamta Pandey

# Associate Dean (Data Science)

# &

# Associate Professor CS & AI.

**Submitted to**



# DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

## SR UNIVERSITY, ANANTHASAGAR, WARANGAL





**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**CERTIFICATE**

This is to certify that the Project Report entitled “Rainfall Prediction Using Machine Learning: A Comparative Analysis of Classifier Performance” is a record of Bonafide work carried out by Charul Pareek, Shifa Imun, Zoya Nausheen, bearing Roll No(s) **2203A51073, 2203A51090, 2203A51100** during the academic year 2023-2024 in partial fulfillment of the award of the degree of ***Bachelor of Technology*** in **Computer Science Engineering** by the SR UNIVERSITY, WARANGAL.

# Supervisor Head of the Department

Dr. Mamta Pandey, Dr. M. Sheshikala

Professor & HoD

SR University

Associate Dean

(Data Science),

&

Associate Professor

CS & AI, (CSE)

SR University

**ACKNOWLEDGEMENT**

We express our thanks to course coordinator Dr. Mamta Pandey, Asst. prof. for guiding us from the beginning through the end of the course project. We express our gratitude to head of the department CS & AI, Dr. M. Sheshikala, Associate Professor for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, Dr. C. V. Guru Rao, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thank to all teaching and non-teaching staff of the department for their suggestions and timely support.

# ABSTRACT

Rainfall prediction plays a crucial role in various fields such as agriculture, water resource management, and disaster preparedness. In this project, we explore the application of machine learning algorithms for predicting rainfall categories based on weather-related features. The dataset used contains information on factors like temperature, humidity, and wind speed, along with corresponding rainfall measurements.

Our study focuses on comparing the performance of three popular classifiers: Logistic Regression, XGBoost, and Support Vector Machine (SVM). We employ a comprehensive pipeline involving data preprocessing, feature engineering, model training, and evaluation to assess the predictive capabilities of each classifier. Evaluation metrics including accuracy, precision, recall, and F1-score are utilized to quantify the performance of the models.

Results demonstrate that XGBoost outperforms the other classifiers, exhibiting higher accuracy and better precision-recall trade-offs. We discuss the significance of each classifier's performance metrics and provide insights into the factors contributing to their effectiveness in rainfall prediction.

This documentation serves as a comprehensive analysis of rainfall prediction using machine learning techniques, offering valuable insights for researchers, practitioners, and policymakers involved in weather forecasting and related domains.

# Table of Contents

|  |  |  |
| --- | --- | --- |
| **S.NO** | **Content** | **Page No** |
| 1 | Introduction | 1 |
| 2 | Literature Review | 2 |
| 3 | Design | 3 |
| 4 | Methodology | 3 |
| 5 | Data Pre-processing | 5 |
| 6 | Results | 7 |
| 7 | Conclusion | 10 |
| 8 | Future scope | 11 |
| 9 | References | 12 |

1. **INTRODUCTION:**

# Rainfall prediction is a critical aspect of weather forecasting with far-reaching implications for agriculture, water resource management, and disaster preparedness. Accurate forecasts enable stakeholders to make informed decisions, mitigate risks, and optimize resource allocation. Traditional methods of rainfall prediction often rely on meteorological models, which, while valuable, may not capture the complex interplay of factors influencing precipitation patterns.

# In this project, we delve into the application of ML algorithms for rainfall prediction, focusing on the comparison of three widely used classifiers: Logistic Regression, XGBoost, and Support Vector Machine (SVM). These classifiers are chosen for their distinct characteristics and performance in handling classification tasks.

# Our objective is twofold: first, to evaluate the efficacy of each classifier in predicting rainfall categories based on various weather-related features, and second, to provide insights into the factors influencing their predictive performance. Through a systematic analysis of model performance metrics, including accuracy, precision, recall, and F1-score, we aim to identify the most suitable classifier for rainfall prediction tasks.

# This documentation presents a comprehensive overview of the project methodology, including data preprocessing, feature engineering, model training, and evaluation. We discuss the implications of our findings for weather forecasting applications and highlight potential avenues for future research in the field of ML-based rainfall prediction.

# 2.LITERATURE REVIEW

# P. Goswami and Srividya [1] have combined RNN and TDNN features and conclusion of their work was that composite models gives better accuracy than the single model. C.Venkatesan et al. [2] used Multilayer Feed Forward Neural Networks (MLFNN) for predicting Indian summer monsoon rainfall. Error Back Propagation (EBP) algorithm is trained and applied to predict the rainfall. Three network models with two, three and ten input parameters have analyzed. They also compared the output result with the statistical models. A.Sahai et al. [3] used error back propagation algorithm for Summer Monsoon Rainfall prediction of India on monthly and seasonal time series. They used data of previous five years of monthly and seasonal mean rainfall values for rainfall prediction. N.Philip and K.Josheph [5] used ABF neural network for yearly rainfall forecasting Kerala region. Their work suggests that ABFNN performs better than the Fourier analysis. V. Somvanshi et al. [7] predictied rainfall of Hyderabad, INDIA region using ANN model. They also compared ANN with ARIMA technique. They used past four months rainfall data as inputs to neural network model. S. Chattopadhyay and M. Chattopadhyay [9] have used two parameters minimum temperature and maximum temperature for rainfall forecasting. S. Chattopadhyaya and G. Chattopadhyaya [10] used Conjugate Gradient Decent (CGD) and Levenberg–Marquardt (LM) learning algorithm for training. Performances of both algorithms were same in prediction task. C. Wu et al. [12] predicted the rainfall of India and China. Artificial Neural Network (MANN). MANN’s performance was compared with LR, K-NN and ANN. K. Htike and O. Khalifa [13] used yearly, biannually, quarterly and monthly rainfall data for rainfall prediction. They trained four different Focused Time Delay Neural Networks (FTDNN) for rainfall forecasting. Highest prediction accuracy was provided by the FTDNN model when yearly rainfall data is taken for training. S. Kannan and S. Ghosh [14] contributed towards developing K- mean clustering technique combined with decision tree algorithm, CART, is used for rainfall states generation from large scale atmospheric variables in a river basin. Rainfall state on daily basis is derived from the historical daily multi-site rainfall data using K-mean clustering. M. Kannan et al. [15] predicted short term rainfall. Empirical method technique is used for prediction task. Data of three specific months for five years is analyzed for particular region. Clustering is used for grouping the elements. G. Geetha and R. Selvaraj [16] used ANN model for predicting monthly rainfall of Chennai region. M. Sharma and J. Singh [17] considered parameters such as rainfall, maximum and minimum temperature, and relative humidity. They predicted weekly rainfall over Pantnagar region. ANN obtained higher prediction accuracy than multiple linear regression model. J. Abbot and J. Marohasy [18] used Time Delay Recurrent Neural Network (TDRNN) for monthly rainfall prediction over Australia region. A. Kumar et al. [19] predicted average rainfall over Udipi district of Karnataka. They used ANN models for prediction task of rainfall. They concluded that Back Propagation Algorithm (BPA) was better than the layer recurrent and cascaded back propagation. Soo-Yeon Ji et al. [21] predicted the hourly rainfall.

# 3. DESIGN:

**Requirement Specifications**

## Hardware Requirements

## System

## RAM

## Hard Disk

## Input

## Output

## Software Requirements

* + - **OS**
    - **Platform**
    - **Program Language**

# 4. METHODOLOGY:

After Data-processing and data visualization the next step is to apply the models on the dataset. Our dataset comes under supervised learning as it contains the labeled data (target variables, feature variables). First the dataset is splitted into training set and testing set. Then the model is trained on training set and then tested on testing set.

**4.1 Logistic Regression Algorithm:**

Logistic regression is a machine learning algorithm which comes under supervised learning. It is a parametric method, where an equation is formed to solve. The equation returns continue values. These continues values should converted to categorical values.so, we use an activation function called “sigmoid”.by using log error function we calculate the error.

* from sklearn.linear\_model import LogisticRegression
* lr=LogisticRegression()
* mm=lr.fit(x\_resem\_train,y\_resem\_train)

**4.2** **XGBoost Classifier:**

XGBoost, short for eXtreme Gradient Boosting, is a powerful machine learning algorithm that belongs to the class of gradient boosting algorithms. It is designed to efficiently handle structured/tabular data and is widely used for both regression and classification tasks. XGBoost has gained popularity for its scalability, speed, and superior performance in various machine learning competitions and real-world applications.

* import numpy as np
* import pandas as pd
* from sklearn.model\_selection import train\_test\_split
* from sklearn.metrics import accuracy\_score
* import xgboost as xgb

# 4. 3 Support Vector Machine algorithm:

# Support vector machine algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression problems. SVM works by constructing a hyperplane or a line that separates the different classes of data points. SVM has support vectors. The distance between positive hyperplane and negative hyperplane is called margin.

* from sklearn.svm import SVC
* svm\_model=SVC(kernel='linear')
* svm\_model.fit(x\_resem\_train,y\_resem\_train)

# 5. DATASET PRE-PROCESSING:

# DATASET DESCRIPTION

# Attributes:

# DATE

# UI TIME

# Temperature (K)

# Relative Humidity (%)

# Pressure (hPa)

# Wind speed (m/s)

# Rainfall (kg/m2)

# Short-wave irradiation (Wh/m2)

**5.1 Sample Dataset**

**A screenshot of a data

Description automatically generated**

**5.2 Exploratory Data Analysis and Model Evaluation**

**Histogram and Kernel Density Plot of Rainfall**

A graph of a graph

Description automatically generatedA graph of a graph

Description automatically generated

**5.2.1 5.2.2**

**Distribution Plots of Numeric Features**

A group of blue graphs

Description automatically generated

**5.2.3**

**Boxplots of Numeric Features**

A group of blue and black graphs

Description automatically generated

**5.2.4**

# 6. RESULTS:

# 6.1 Data Exploration:

**A screenshot of a computer

Description automatically generated**

**Logistic Regression:**

**A screenshot of a computer

Description automatically generated**

**6.1.1**

**A screenshot of a computer

Description automatically generated**

**6.1.2**

(From fig 6.1.1), 0: Represents the absence or negative outcome of a particular category or event. For example, in the context of rainfall prediction, it could represent the absence of rainfall or the classification of a day as "No Rain" or "Not Rainy".

(From fig 6.1.1), 1: Represents the presence or positive outcome of a particular category or event. In the rainfall prediction project, it could represent the presence of rainfall or the classification of a day as "Rain" or "Rainy".

**XGBoost:**

**A screenshot of a computer

Description automatically generated**

**6.1.3**

**Support Vector Machine:**

**A screenshot of a computer

Description automatically generated**

**6.1.4**

**6.2 Result Analysis:**

From figure 6.1.1, 6.1.2, 6.1.3, and 6.1.4 we observe,

|  |  |  |
| --- | --- | --- |
| **S.NO** | **MACHINE LEARNING MODEL** | **ACCURACY** |
| 1 | Logistic regression | 0.65 |
| **2** | XGBoost | 0.906666 |
| **3** | Support vector Machine | 0.7867 |

# 7. CONCLUSION:

Among worldwide, agriculture has the major responsibility for improving the economic contribution of the nation. However, still the most agricultural fields are underdeveloped due to the lack of deployment of ecosystem control technologies. To prevent this problem, Agricultural sectors have to predict the crop from given data set using machine learning techniques. The estimation of rainfall is of great importance in terms of water resources management, human life and their environment. It can be met with the incorrect or incomplete estimation problems because rainfall estimation is affected from the geographical and regional changes and properties. This paper presented review of different methods used for rainfall prediction and problems one might encounter while applying different approaches for rainfall forecasting. Due to nonlinear relationships in rainfall data and ability of learning from the past makes XGBoost a preferable approach from all available approaches.

**8. FUTURE SCOPE:**

# The project holds promise for further advancement in rainfall prediction. Exploring advanced machine learning models like deep learning algorithms and ensemble techniques could enhance accuracy and robustness. Additionally, fine-tuning hyperparameters and optimizing feature engineering can improve performance. Integrating real-time data feeds, developing a user-friendly interface, and ensuring scalability and deployment optimization are key areas for enhancement. Lastly, integrating with IoT devices offers potential for localized and personalized predictions. These avenues for improvement will elevate the project's usability, accuracy, and practical applications in various fields.

# REFERENCES:

[1] P. Goswami and Srividya, “A novel Neural Network design for long range prediction of rainfall pattern,” Current Sci. (Bangalore), vol. 70, no. 6, pp. 447-457, 1996.

[2] C. Venkatesanet, S. D. Raskar , S. S. Tambe , B. D. Kulkarni , and R. N. Keshavamurty , “Prediction of all India summer monsoon rainfall using Error-Back-Propagation Neural Networks,” Meteorology and Atmospheric Physics, pp. 225-240, 1997.

[3] A. K. Sahai, M. K. Soman, and V. Satyan, “All India summer monsoon rainfall prediction using an Artificial Neural Network,” Climate dynamics, vol. 16, no. 4, pp. 291-302, 2000.

[4] N. S. Philip and K. B. Joseph, “On the predictability of rainfall in Kerala-An application of ABF neural network,” Computational ScienceICCS, Springer Berlin Heidelberg, pp. 1-12, 2001.

[5] N. S. Philip and K. B. Joseph, “A Neural Network tool for analyzing trends in rainfall,” Comput. & Geosci., vol. 29, no. 2, pp. 215-223, 2003.

[6] N. Chantasut, C. Charoenjit, and C. Tanprasert, “Predictive mining of rainfall predictions using artificial neural networks for Chao Phraya River,” 4th Int Conf. of the Asian Federation of Inform. Technology in Agriculture and the 2nd World Congr. on Comput. in Agriculture and Natural Resources, Bangkok, Thailand, pp. 117-122, 2004.

[7] V. K. Somvanshi, O. P. Pandey, P. K. Agrawal, N.V. Kalanker1, M.Ravi Prakash, and Ramesh Chand, “Modeling and prediction of rainfall using Artificial Neural Network and ARIMA techniques,” J. Ind. Geophys. Union, vol. 10, no. 2, pp. 141-151, 2006.

[8] S. Chattopadhyay, “Anticipation of summer monsoon rainfall over India by Artificial Neural Network with Conjugate Gradient Descent Learning,” arXiv preprint nlin/0611010, pp. 2-14, 2006.

[9] S. Chattopadhyay and M. Chattopadhyay, “A Soft Computing technique in rainfall forecasting,” Int. Conf. on IT, HIT, pp. 19-21, 2007.

[10] S. Chattopadhyay and G. Chattopadhyay, “Comparative study among different neural net learning algorithms applied to rainfall time series, "Meteorological applicat., vol. 15, no. 2, pp. 273-280, 2008.

[11] P.Guhathakurta, “Long lead monsoon rainfall prediction using deterministic Artificial Neural Network model,” Meteorology and Atmospheric Physics 101, pp. 93-108, 2008.

[12] C. L. Wu, K. W. Chau, and C. Fan, “Prediction of rainfall time series using Modular Artificial Neural Networks coupled with datapreprocessing techniques," J. of hydrology, vol. 389, no. 1, pp. 146- 167, 2010. [13] K. K. Htike and O. O. Khalifa, “Rainfall forecasting models using Focused Time-Delay Neural Networks,” Comput. and Commun. Eng. (ICCCE), Int. Conf. on IEEE, 2010

[14] S. Kannan, Subimal Ghosh, “Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output”, Springer-Verlag, July- 2010.

[15] M.Kannan, S.Prabhakaran, P.Ramachandran, “Rainfall Forecasting Using Data Mining Technique”, International Journal of Engineering and Technology Vol.2 (6), 397-401, 2010.

[16] G. Geetha and R. S. Selvaraj, “Prediction of monthly rainfall in Chennai using Back Propagation Neural Network model,” Int. J. of Eng. Sci. and Technology, vol. 3, no. 1, pp. 211 213, 2011.

[17] M. A. Sharma and J. B. Singh, “Comparative Study of rainfall forecasting models,” New York Sci. J., pp. 115-120, 2011.

[18] J. Abbot and J. Marohasy, “Application of Artificial Neural Networks to rainfall forecasting in Queensland, Australia,” Advances in Atmospheric Sci.,vol. 29, no. 4, pp. 717-730, 2012.

[19] A. Kumar, A. Kumar, R. Ranjan, and S. Kumar, “A rainfall prediction model using artificial neural network,” Control and Syst. Graduate Research Colloq. (ICSGRC), pp. 82-87, 2012.

[20] R. R. Deshpande, “On the rainfall time series prediction using Multilayer Perceptron Artificial Neural Network,” Int. J. of Emerging Technology and Advanced Eng., vol. 2, no. 1, pp. 148-153, 2012.

[21] Soo-Yeon Ji, Sharad Sharma, Byunggu Yu, Dong Hyun Jeong, “Designing a Rule-Based Hourly Rainfall Prediction Model”, IEEE IRI 2012, August – 2012.