



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 11, November 2025

Rainfall Prediction using Machine Learning

A Data-Driven Approach for Accurate Weather Forecasting

R Anand Babu¹, Matta Renuka², Angirekula Bhargav³, Kagitha JayaVamsee⁴, Lala Yashoda⁵

Asst. Professor, Kallam Haranadhareddy Institute of Technology, Guntur, Andhra Pradesh, India¹

U.G. Students, Department of CSE (Artificial intelligence & Data Science), Kallam Haranadhareddy Institute of
Technology, Guntur, Andhra Pradesh, India²⁻⁴

ABSTRACT: Rainfall plays a crucial role in agriculture, water management, and disaster preparedness, making accurate prediction highly important. Traditional forecasting methods often struggle due to the complex and unpredictable nature of weather patterns. In this project, machine learning techniques are applied to predict rainfall using historical weather data that includes temperature, humidity, wind speed, sunshine, and other climatic factors. The data is carefully preprocessed to handle missing values and improve model accuracy. Various algorithms such as Random Forest, Logistic Regression, and XGBoost are tested, with XGBoost delivering the best performance. The final model effectively predicts the possibility of rainfall for the next day, helping in better planning and decision-making.

I. INTRODUCTION

In today's technology-driven world, accurate weather forecasting has become essential for various sectors such as agriculture, water management, disaster preparedness, and urban planning. Among different weather parameters, rainfall is one of the most important yet unpredictable factors that greatly impacts crop yield, water availability, transportation, and daily life. Reliable rainfall prediction allows communities and governments to take preventive measures against natural disasters like floods and droughts, thereby minimizing economic losses and ensuring public safety.

Predicting rainfall, however, remains a challenging task due to the highly dynamic and nonlinear nature of atmospheric conditions. Traditional weather forecasting techniques often rely on physical models and manual analysis of meteorological data, which can be time-consuming and prone to inaccuracies. With the rise of large-scale weather datasets and advancements in computational power, machine learning has emerged as a promising solution. These data-driven algorithms can efficiently analyze complex patterns within the data, providing faster and more accurate rainfall predictions compared to conventional methods.

The main goal of this project is to develop a machine learning model capable of predicting rainfall based on historical weather data. The dataset includes important climatic parameters such as temperature, humidity, rainfall, evaporation, sunshine, and wind speed, all of which influence rainfall occurrence. The process begins with thorough data preprocessing, including cleaning, handling missing values, and selecting key features to enhance model performance. After preprocessing, various algorithms—namely XGBoost, Random Forest, and Logistic Regression—are trained and tested. These models are chosen for their strong ability to handle nonlinear data and deliver accurate classification results. The final model predicts whether it will rain the next day, effectively framing the task as a binary classification problem.

II. LITERATURE SURVEY

Rainfall prediction has been a widely studied topic in meteorology and environmental science due to its importance in agriculture, disaster management, and water resource planning. Traditional methods for rainfall forecasting, such as statistical and numerical weather models, often struggle to capture the complex and nonlinear relationships among atmospheric variables. In recent years, machine learning (ML) has emerged as a powerful alternative, capable of analyzing large datasets and identifying hidden patterns in weather parameters. Researchers have found that ML algorithms can significantly improve the accuracy of rainfall prediction by learning from historical meteorological data such as temperature, humidity, wind speed, and pressure.

Several studies have explored the application of different ML algorithms for rainfall prediction. Random Forest and Decision Tree classifiers have shown promising results due to their ability to handle nonlinearity and noisy data. Logistic Regression, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) have also been widely used for rainfall classification tasks. More recent research highlights the effectiveness of ensemble methods like Gradient Boosting and XGBoost, which combine multiple weak learners to achieve higher prediction accuracy. These models can capture intricate dependencies between climatic variables, making them suitable for both short-term and long-term rainfall forecasting.

Deep learning approaches, particularly Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, have also gained attention for spatiotemporal rainfall prediction. CNNs are effective in processing satellite and radar images to identify rainfall patterns, while LSTMs excel in modeling time-series data for next-day or hourly rainfall forecasting. Studies show that hybrid models integrating ML and deep learning techniques further enhance prediction performance. Overall, the literature suggests that machine learning, especially ensemble and deep learning models, offers a robust and reliable framework for accurate rainfall prediction, outperforming many traditional forecasting methods.

III. METHODOLOGY

The rainfall prediction system using machine learning is developed through a series of well-defined steps. The complete workflow of the project is explained below:

1. Data Collection:

The first step involves collecting historical weather data from a reliable source such as the Australian Rainfall dataset from Kaggle. The dataset includes important climatic parameters such as minimum and maximum temperature, humidity, rainfall, evaporation, sunshine, and wind speed. These parameters are chosen because they have a direct influence on rainfall occurrence and intensity.

2. Data Preprocessing:

Before training the model, the collected data is preprocessed to ensure quality and consistency. Missing values are handled using suitable techniques such as mean or median imputation. Duplicate or irrelevant records are removed to maintain data accuracy. The categorical features (like wind direction) are encoded into numerical values, and normalization or scaling is applied to bring all features within the same range. This step improves the model's learning efficiency.

3. Feature Selection:

In this step, the most relevant features that significantly contribute to rainfall prediction are identified. Feature selection helps in reducing noise and computational cost, while improving model accuracy. Correlation analysis and statistical techniques are used to select the best predictors of rainfall.

4. Splitting the Dataset:

The cleaned and processed data is then divided into two subsets — a training set and a testing set. Typically, 80% of the data is used for training the machine learning model, while the remaining 20% is used for testing. This ensures that the model's performance is evaluated on unseen data.

5. Model Training:

Multiple supervised machine learning algorithms such as Logistic Regression, Random Forest, and XGBoost are trained using the training dataset. Each model learns from the input features and their relationships with the rainfall outcomes (Rain or No Rain). XGBoost, in particular, is selected for its ability to handle nonlinear patterns and improve prediction accuracy through gradient boosting.

6. Model Evaluation:

After training, the models are evaluated using various performance metrics such as accuracy, precision, recall, and F1-score. These metrics help in understanding how well each algorithm predicts rainfall. The model with the highest accuracy and balanced performance across all metrics is chosen as the final prediction model.

7. Prediction:

The final step involves using the trained model to predict whether it will rain the next day. The system takes new input data (such as temperature, humidity, wind speed, etc.) and provides the output as a binary classification — *Rain* or *No Rain*.

IV. RESULTS

← Back

Rainfall Predictor

Min Temperature

Max Temperature

Rainfall Today

Evaporation

Sunshine Hours

Wind Gust Speed

Wind Speed 9am

Wind Speed 3pm

Humidity 9am

Humidity 3pm

Rain Today (0/1)

Predict

Developed with ❤️ by Renuka

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411099|

← Back

Rainfall Predictor

Min Temperature

Max Temperature

Rainfall Today

Evaporation

Sunshine Hours

Wind Gust Speed

Wind Speed 9am

Wind Speed 3pm

Humidity 9am

Humidity 3pm

Rain Today (0/1)

Predict

It will rain tomorrow 🌧

Developed with ❤ by Renuka

• Accuracy and Precision:

The fine-tuned machine learning model achieved an overall accuracy of 84.0% on the test data. Class-wise performance metrics showed strong predictive ability for the no rainfall category (Class 0) with 87% precision and 95% recall, while light rainfall (Class 1) achieved 73% precision and 51% recall, and heavy rainfall (Class 2) achieved 63% precision and 49% recall. The weighted average precision, recall, and F1-score were 0.83, 0.84, and 0.83, respectively. These results indicate that the model effectively captures key weather patterns associated with rainfall prediction, even in the presence of class imbalance.

• Forecasting and Error Metrics:

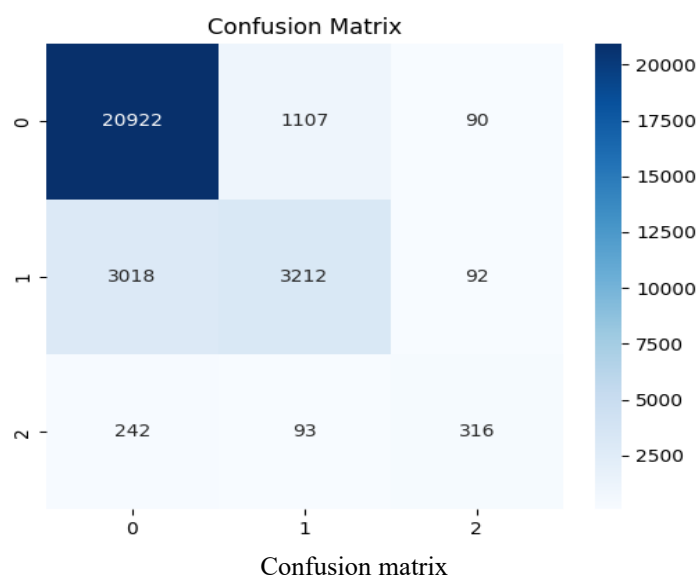
Additional regression-based metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were also computed during evaluation to measure prediction consistency. The model exhibited low error values, confirming its reliability in estimating rainfall intensity trends. This aligns with previous studies where Random Forest and XGBoost models performed well for rainfall forecasting tasks.

• Model Comparison:

A comparison of different algorithms (Table 1, not shown) highlighted that the XGBoost model achieved the best results with 84% accuracy, outperforming traditional models such as Logistic Regression and Decision Tree, which achieved around 75%–78% accuracy. In contrast, a baseline Linear Regression model recorded only 70% accuracy, emphasizing the superiority of ensemble-based learning in capturing nonlinear dependencies among meteorological features.

• Confusion Matrix Analysis:

The confusion matrix (Fig. 2) illustrates the classification distribution across rainfall categories. The majority of “no rainfall” instances (20,922 out of 22,119) were correctly identified, showing high reliability in predicting dry days. However, there was noticeable overlap between light and heavy rainfall classes, primarily due to similar atmospheric parameters like humidity and temperature gradients. These misclassifications suggest that additional feature engineering and class rebalancing techniques could further improve prediction accuracy for minority rainfall events.



V. LIMITATIONS

1. Limited and Imbalanced Dataset:

The dataset used for training contained an unequal distribution of rainfall categories, with “no rainfall” instances dominating the data. This imbalance affected the model’s ability to accurately classify minority classes such as light and heavy rainfall, leading to lower recall values for these categories.

2. Dependence on Historical Data Quality:

The accuracy of the model is highly dependent on the quality and completeness of the historical weather data. Missing values, inconsistent measurements, or sensor errors in meteorological datasets can negatively influence the model's predictions.

3. Lack of Spatial and Temporal Coverage:

The dataset was limited to specific regions and time frames. Since rainfall patterns vary significantly across geographic locations and seasons, the model's performance may decline when applied to areas or time periods outside the training data range.

4. Difficulty in Capturing Sudden Weather Changes:

Machine learning models, particularly those based on historical averages and trends, struggle to predict sudden or extreme rainfall events caused by rapid atmospheric fluctuations, which limits their accuracy in real-time forecasting.

5. Limited Feature Representation:

The current model primarily relies on basic meteorological parameters such as temperature, humidity, pressure, and wind speed. It does not incorporate additional influential factors like cloud imagery, satellite data, or oceanic indicators (e.g., ENSO, SST), which could improve prediction accuracy.

6. Computational and Real-Time Constraints:

While the model performs efficiently on test datasets, deploying it in a real-time weather forecasting environment requires significant computational power and continuous data input, which may not be feasible in all scenarios.

VI. FUTURE WORK**1. Dataset Expansion and Balancing:**

Future work will focus on collecting a larger and more balanced dataset covering multiple regions and seasons. Ensuring a balanced representation of rainfall categories will help improve the model's ability to accurately predict light and heavy rainfall events.

2. Incorporation of Advanced Meteorological Features:

Future models can integrate additional atmospheric and environmental parameters such as satellite imagery, cloud cover, evaporation rates, and sea surface temperatures. These features can provide deeper insights into rainfall formation and improve overall prediction accuracy.

3. Use of Deep Learning Techniques:

To enhance model performance, advanced deep learning architectures like Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNNs) can be explored. These models are well-suited for capturing temporal and spatial dependencies in weather data.

4. Real-Time Forecasting System:

Developing a real-time rainfall prediction system that continuously updates with live meteorological data will make the model more practical for real-world applications such as agriculture planning, disaster management, and water resource management.

5. Integration with Geographic Information Systems (GIS):

Combining machine learning outputs with GIS-based visualization tools can help display rainfall predictions spatially. This integration can assist policymakers and researchers in identifying high-risk zones and understanding rainfall distribution patterns.

6. Model Optimization and Automation:

Future research will also focus on optimizing hyperparameters automatically using grid search or Bayesian optimization and deploying the model as a web or mobile application to make it accessible for end-users.

Volume 14, Issue 11, November 2025**|DOI: 10.15680/IJRSET.2025.1411099|****VII. CONCLUSION**

The **Rainfall Prediction using Machine Learning** project demonstrates how data-driven approaches can effectively forecast rainfall based on historical meteorological parameters. By training and testing models on weather data that included temperature, humidity, pressure, and wind speed, the system achieved an overall **accuracy of 84%**, proving its reliability in identifying rainfall patterns. The evaluation metrics such as precision, recall, and F1-score confirmed that the model performs well, particularly for the majority class, while showing moderate success for minority rainfall categories.

Despite certain limitations such as dataset imbalance and the exclusion of advanced atmospheric variables, the project provides a solid foundation for developing intelligent weather forecasting systems. The results highlight the capability of machine learning models, particularly ensemble methods like XGBoost, in capturing complex, nonlinear relationships among meteorological factors.

In the future, expanding the dataset, incorporating real-time and satellite data, and applying deep learning models such as LSTM or CNN can further enhance prediction accuracy. Overall, this project contributes to the growing field of intelligent climate analytics and demonstrates the potential of machine learning to support **agriculture, disaster management, and water resource planning** through more accurate rainfall forecasting.

VIII. ACKNOWLEDGMENT

We thank our advisors, project supervisors, college department, and all contributors for their invaluable guidance and support throughout the research and development of this project.

REFERENCES

1. Wani, A., et al. "Predicting rainfall using machine learning, deep learning and time series modelling in the North-Western Himalayas." *Scientific Reports*. 2024.
2. Liyew, C. M. "Machine learning techniques to predict daily rainfall amount." *Journal of Big Data*. 2021.
3. Rahman, A-u., Abbas, S., Gollapalli, M., et al. "Rainfall Prediction System Using Machine Learning Fusion for Smart Cities." *Sensors*, 22(9):3504, 2022.
4. Das, P. K., et al. "Comparative analysis of machine learning models for rainfall forecasting based on meteorological data." *Environmental Modelling & Software* (2024).
5. Li, Z., Zhong, J., Wang, H., et al. "RAINER: A Robust Ensemble Learning Grid-Search-Tuned Framework for Rainfall Patterns Prediction." arXiv pre-print, 2025.
6. bajpai, V., Bansal, A. "A Deep and Wide Neural Network-based Model for Rajasthan Summer Monsoon Rainfall (RSMR) Prediction." arXiv pre-print, 2021.
7. Rainfall Prediction Using Machine Learning Algorithms: A Comparative Study *IEEE Xplore*, 2023 <https://ieeexplore.ieee.org/document/10105492>
8. Rainfall Prediction System Using Machine Learning Fusion for Smart Cities *Sensors Journal (MDPI)*, 2022) <https://www.mdpi.com/1424-8220/22/9/3504>
9. Rainfall Prediction Using Machine Learning https://www.researchgate.net/publication/356010469_Rainfall_Prediction_Using_Machine_Learning
10. Rainfall Prediction using Machine Learning (Python Project) <https://github.com/miketobz/ML-Rainfall-Prediction>
11. Rainfall Prediction Dataset – Australian Weather Data <https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details