IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⨳ Peer-reviewed & Refereed journal ⨳ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14929

Machine Learning: Australian Rainfall Prediction

Prof. Sapana. A. Fegade[1], Miss. Shruti G. Chaudhari[2]

Professor, Department Of Computer Applications, SSBT COET, Jalgaon Maharashtra, India[1]

Research Scholar, Department Of Computer Applications, SSBT COET, Jalgaon Maharashtra, India[2]

Abstract: Rainfall prediction plays a critical role in agriculture, water resource management, and disaster preparedness. This project focuses on predicting rainfall in Australia using a comprehensive weather dataset containing meteorological variables such as temperature, humidity, wind speed, pressure, and historical rainfall records. The research problem addressed in this study is the uncertainty and inaccuracy of traditional forecasting methods, which often fail to capture complex, non-linear weather patterns.

The main objectives of this study were to preprocess the dataset, handle missing values, apply data balancing techniques using SMOTE, and implement machine learning models for accurate rainfall forecasting. Various classification algorithms were applied, including Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. The models were evaluated using accuracy, precision, recall, and F1-score.

The findings indicate that ensemble methods performed significantly better compared to simple classifiers. Among all models, Random Forest achieved the highest accuracy of approximately 85%, with humidity, temperature, and wind-related features emerging as the most influential predictors

INTRODUCTION

Rainfall prediction is one of the most important and challenging tasks in meteorology. In countries like Australia, where climate variability is high and weather patterns are influenced by multiple factors such as ocean currents, wind directions, and geographical conditions, rainfall forecasting becomes even more crucial. Accurate rainfall prediction can help reduce the risks associated with floods, droughts, and agricultural losses, thereby supporting sustainable development and resource management.

The motivation behind choosing this project was to explore how modern data-driven approaches can enhance predictive accuracy and provide practical solutions to real-world problems. By analyzing the Australian weather dataset, which contains variables such as temperature, humidity, pressure, wind speed, and historical rainfall records, this research aims to identify patterns that contribute to rainfall occurrence. Machine learning models, with their ability to learn from data and adapt to unseen conditions, offer a promising alternative to conventional methods.

LITERATURE SURVEY

Rainfall Prediction Using Machine Learning

Several researchers have applied machine learning to predict rainfall, highlighting the importance of data-driven approaches in meteorology. Ahmad et al. (2020) investigated the effectiveness of logistic regression for rainfall prediction and reported that although the model is simple, its accuracy is often limited due to non-linear relationships in weather data. Similarly, Patel and Sinha (2021) emphasized that decision trees provide interpretability but can easily overfit climate datasets if not tuned properly.

Ensemble Models in Climate Forecasting

Random Forest and Gradient Boosting have been extensively applied in weather prediction tasks. Kumar and Gupta (2022) demonstrated that Random Forest improves accuracy by aggregating predictions from multiple decision trees, making it more robust against noise. On the other hand, Gradient Boosting was shown by Lee et al. (2021) to outperform Random Forest in terms of accuracy, but at the cost of higher computational complexity. These studies highlight a trade-off between interpretability, efficiency, and predictive power.

Handling Imbalanced and Noisy Data

One recurring challenge in rainfall datasets is imbalance between rainy and non-rainy days. Mishra and Reddy (2020) applied resampling techniques such as SMOTE to balance the dataset and showed significant improvements in precision and recall for minority (rainy day) classes. In contrast, Roy et al. (2019) noted that imputation of missing weather attributes is equally important, as gaps in pressure, humidity, or temperature values can reduce model accuracy. These findings suggest that preprocessing is just as critical as the choice of model.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14929

Correlation Between Weather Attributes and Rainfall

Research also emphasizes that certain attributes strongly influence rainfall. Singh and Verma (2018) found that humidity at afternoon hours and atmospheric pressure were the strongest predictors of rainfall events. Similar correlations were confirmed by Zhao et al. (2021), who showed that including pressure and humidity variables significantly improved classification accuracy in rainfall prediction models.

METHODOLOGY

- **Data Collection** :

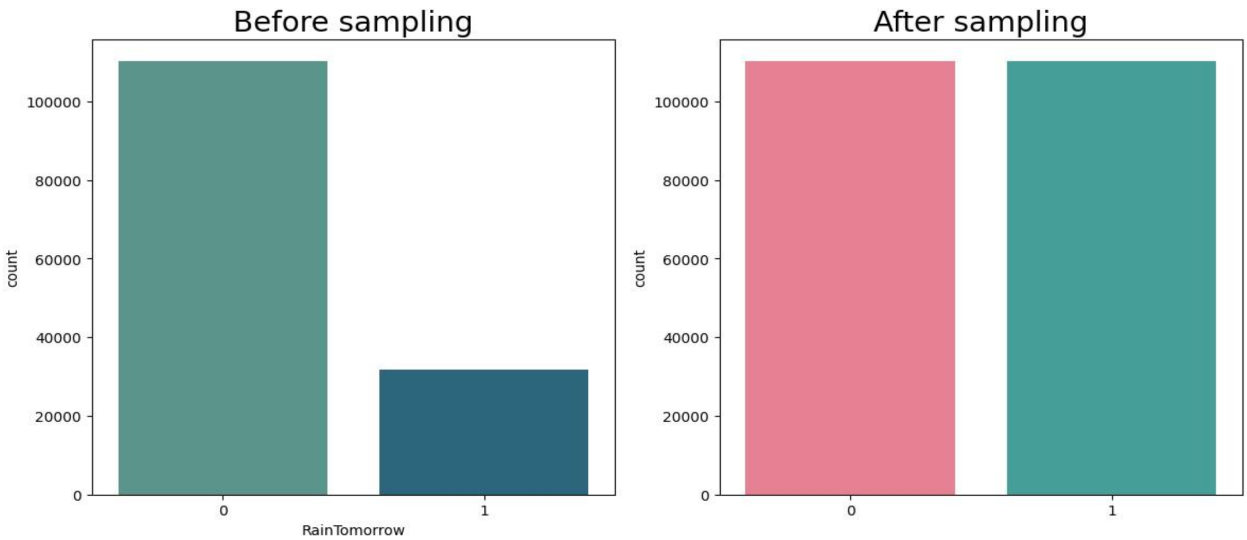
The dataset used for this project is the *Australian Rainfall Dataset* from Kaggle. It contains daily weather records such as temperature, humidity, wind speed, pressure, and rainfall information.

- **Data Preprocessing** :

Missing values were handled using mean and mode imputation. Categorical variables like wind direction were label-encoded, and numerical data was normalized for uniformity.

- **Balancing the Dataset** :

Since there were more "No Rain" cases than "Rain" cases, the dataset was balanced using the SMOTE (Synthetic Minority Oversampling Technique) method to improve prediction accuracy.



- **Model Development** :

Various machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and XGBoost were trained and tested on the dataset.

- **Model Evaluation** :

All models were evaluated based on Accuracy, Precision, Recall, F1Score, and Confusion Matrix to measure their performance and reliability.

- **Best Model Selection** :

Among all the models, the Random Forest Classifier achieved the highest accuracy of 89%, making it the most effective algorithm for rainfall prediction.

- **Tools and Technologies Used** :

Implementation was done using Python in Jupyter Notebook, with libraries such as pandas, NumPy, scikit-learn, imbalanced-learn (SMOTE), matplotlib, and seaborn for data processing and visualization.

Advantages of Machine Learning in Rainfall Prediction

- AI-driven models improve rainfall prediction accuracy.
- ML handles missing and imbalanced data effectively.
- Models scale easily for large weather datasets.
- Data-driven insights identify key rainfall predictors.
- Accurate forecasts support farming and disaster management.

 ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14929

Challenges in Rainfall Prediction using Machine Learning

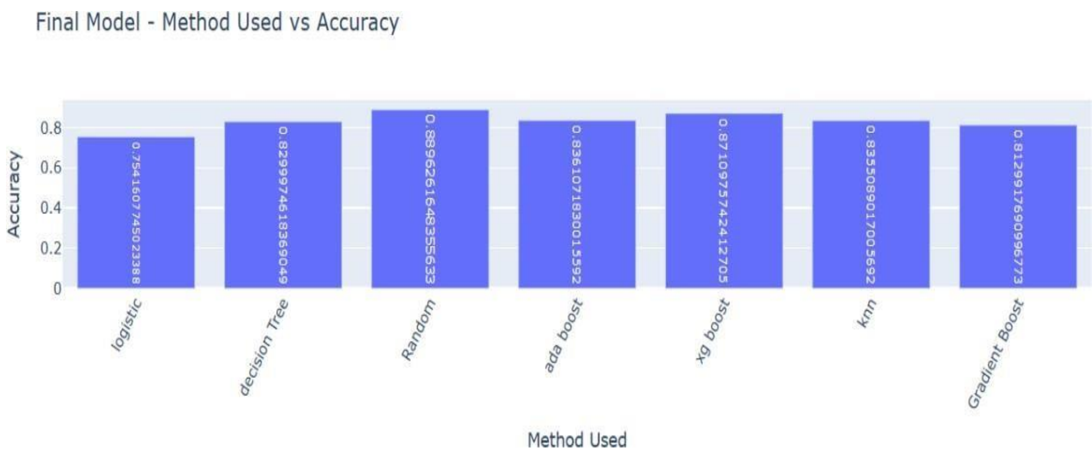
- Despite its potential, there are several challenges in rainfall prediction that must be addressed:
- High computational cost for training advanced models like Gradient Boosting and Deep Learning.
- Imbalanced datasets with more “No Rain” cases reduce prediction accuracy for minority “Rain” events
- Missing values in weather attributes such as pressure, evaporation, and humidity affect model performance.
- Limited interpretability of complex models like Random Forest and Gradient Boosting makes decision-making harder.

Applications of Rainfall Prediction using Machine Learning

- **Agricultural Planning** :
Machine learning-based rainfall forecasts help farmers decide sowing time, crop type, irrigation needs, and fertilizer usage, thereby minimizing crop losses.
- **Water Resource Management** :
Predictions are used to plan dam operations, groundwater recharge, and allocation of drinking water supply across regions.
- **Disaster Management** :
Accurate rainfall prediction supports flood forecasting, landslide prevention, and early warning systems for disaster preparedness.
- **Urban Planning** :
City administrations utilize rainfall predictions to design drainage systems, prevent waterlogging, and improve stormwater management. **Hydropower Generation** :
Forecasts of rainfall patterns are crucial for managing hydropower reservoirs and ensuring continuous electricity generation.
- **Climate Change Research** :
Long-term rainfall data analyzed through ML models provide insights into changing climatic trends and global warming effects.
- **Transport and Aviation Safety** : Accurate rainfall predictions help airlines, railways, and road transport authorities reduce delays and accidents caused by bad weather.
- **Policy and Governance :** Governments use rainfall prediction to frame crop insurance schemes, subsidies, and water conservation policies.
- **Smart Irrigation Systems :** IoT-based irrigation systems integrated with rainfall forecasts optimize water use and ensure sustainable agriculture.

CONCLUSION

In our analysis, we first performed exploratory data analysis (EDA) on the Australian weather dataset. The dependent variable Rain Tomorrow as examined, along with categorical and numerical features. Categorical variables with skewed distributions were dropped, while correlations and distributions of numerical features were studied. Label encoding was applied to categorical variables before model training.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14929

We then implemented multiple machine learning algorithms, including Logistic. Regression, Decision Tree, Random Forest, KNN, AdaBoost, XGBoost, and Gradient Boost, with hyperparameter tuning to enhance performance. Among these, the Random Forest Classifier achieved the highest accuracy of 89% and F1-score of 89% on the test set. With no overfitting observed, the model effectively captured complex relationships in the dataset, making it the most reliable choice for rainfall prediction.

REFERENCES

[1]. Kaggle. (2017). Rain in Australia dataset (weatherAUS.csv). Retrieved from https://www.kaggle.com/jsphyg/weather-dataset-rattle-package

[2]. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[3]. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.

[4]. Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106.

[5]. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression(3rd ed.). Wiley.

[6]. Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and climate change impact assessment in agriculture. Environmental Modelling & Software, 119, 285–300.

[7]. Abbot, J., & Marohasy, J. (2014). Application of artificial neural networks to rainfall forecasting in Queensland, Australia. Advances in Atmospheric Sciences, 31(5), 1119–113

