# Machine Learning for Sustainable Development Goal 13: Climate Action

## 1. Introduction

Project Objective: To leverage machine learning to tackle challenges related to climate action, aiming to support SDG Goal 13 by predicting rainfall patterns, assessing climate impacts, and enhancing water resource management in vulnerable regions. This project focuses on improving forecasting accuracy to inform agricultural practices, disaster preparedness, and water conservation efforts in areas affected by climate variability.

Motivation: Access to clean water is critical for health and well-being, particularly in the face of climate change. By leveraging machine learning to improve rainfall prediction, we aim to enhance resource allocation, optimize maintenance of water infrastructure, and support effective sanitation efforts.

## 2. Data Collection

Dataset Description:

Features: Historical weather data, including parameters such as temperature, humidity, wind speed, atmospheric pressure, and precipitation levels.

Size: X rows by Y columns

Target Variable: Rainfall occurrence (binary: yes/no, or multiclass for varying levels of rainfall.

## 3. Exploratory Data Analysis (EDA)

Summary Statistics:

Mean, Median, and Distribution: Analyze the mean, median, and distribution of rainfall data over time to understand average precipitation levels and variability.

Visualizations:

Correlation Heatmap: Create a correlation heatmap to identify relationships between rainfall and relevant variables, such as temperature, humidity, and atmospheric pressure.

Boxplots: Use boxplots to detect outliers in rainfall data, providing insights into extreme weather events.

Histograms: Generate histograms to assess the distribution of rainfall amounts, revealing patterns in rainfall intensity and frequency.

Insights: Examine key trends in rainfall patterns, including seasonal variations and anomalies, to assess potential impacts of climate change on precipitation levels and to support efforts toward achieving Sustainable Development Goal 13 (Climate Action).

## 4. Data Preprocessing

Handling Missing Values: Used median imputation for features with missing values.  
Encoding Categorical Variables: One-hot encoding for any categorical features.  
Feature Scaling: Standardized features using `StandardScaler` for better performance in machine learning models.

## 5. Machine Learning Model Selection

Model Choices:  
- Logistic Regression (for binary classification).  
- Random Forest Classifier (for handling non-linear relationships and feature importance).  
- Support Vector Machine (SVM) for optimal margin separation.  
Why Scikit-Learn: Easy implementation, variety of algorithms, and effective performance metrics.  
Evaluation Metric: Accuracy, Precision, Recall, and F1-Score due to the critical nature of accurately identifying contamination.

## 6. Model Implementation

Data Splitting: The dataset was divided into training and testing sets, allocating 80% of the data for training and 20% for testing using the train\_test\_split function from Scikit-Learn.

Hyperparameter Tuning: To enhance the accuracy of the rainfall prediction model in alignment with SDG Goal 13 (Climate Action), This process aimed to identify the optimal number of estimators and maximum depth of the trees. Additionally, cross-validation with 5 folds was implemented to ensure robust model generalization and reliability in predicting rainfall patterns.

7. Code Example

from sklearn.model\_selection import train\_test\_split

from imblearn.over\_sampling import RandomOverSampler

# Example: If 'rainfall' is your target column and the rest are features:

features = df.drop('rainfall', axis=1) # Select all columns except 'rainfall'

target = df['rainfall'] # Select the 'rainfall' column as target

X\_train, X\_val, Y\_train, Y\_val = train\_test\_split(features,

target,

test\_size=0.2,

stratify=target, random\_state=2)

# As the data was highly imbalanced we will

# balance it by adding repetitive rows of minority class.

ros = RandomOverSampler(sampling\_strategy='minority',

random\_state=22)

X, Y = ros.fit\_resample(X\_train, Y\_train)

## 7. Results and Evaluation

Model Performance:

The Random Forest model achieved an accuracy of X%, with an F1-score of Y%, demonstrating its effectiveness in predicting rainfall patterns and variability related to climate change.

Feature Importance:

Analysis revealed the key features influencing rainfall predictions, such as temperature, humidity, atmospheric pressure, and historical precipitation data, highlighting their significance in understanding climate dynamics.

Confusion Matrix:

A confusion matrix was utilized to visualize the true versus predicted rainfall outcomes, enabling the identification of common misclassifications and areas for improvement in the model’s predictive capabilities.

## 8. Conclusion and Future Work

Key Takeaways: Machine learning models have shown effectiveness in predicting rainfall patterns based on various meteorological data, including temperature, humidity, and historical precipitation levels. This project highlights the potential for improving climate resilience and resource management in response to changing weather conditions.

Future Improvements:

Incorporating real-time data for continuous learning and more accurate forecasts. Expanding the dataset to include a wider geographical range and diverse climatic conditions. Implementing models on edge devices for on-site analysis in remote areas to enhance accessibility and response time.

## 9. References

- Kaggle Dataset  
- Scikit-Learn Documentation