***RAINFALL PREDICTION***

**Group members:**

1. **Selam Kiflay 2. Ketim Teklu 3. Abdi Gonfa 4. Hilina Amare**

**1. Project Idea:**

This project aims to predict the probability of rainfall for a specific day and location (e.g., Ethiopia, Addis Ababa) using machine learning. The goal is to provide accurate, user-friendly forecasts to support agricultural planning and water management by developing a system that takes country, city, and date as inputs and outputs a rainfall probability (e.g., 68.5%) and outcome (Rain/No Rain).

**2. Relevance to Sustainable Development Goals (SDGs):**

The project aligns with SDG 2 (Zero Hunger) and SDG 13 (Climate Action). Accurate rainfall predictions enable farmers to optimize planting and irrigation, enhancing food security. By forecasting weather patterns, the system supports climate resilience, helping communities adapt to variable rainfall in regions like Ethiopia, where agriculture depends on timely rains.

**3. Literature Examples:**

1. Manandhar et al. (2019), ”Machine Learning for Rainfall Prediction”: This study uses Random Forest to predict rainfall in Nepal, achieving high accuracy with weather features like humidity and temperature. It inspires feature selection and model choice for this project.

2. Pham et al. (2020), ”Deep Learning for Weather Forecasting”: This paper applies neural networks to rainfall forecasting, highlighting data preprocessing needs. It informs the project’s data cleaning approach but suggests simpler models may suffice.

**4. Data Description:**

The project uses **train.csv** (2190 rows, CSV format) with daily weather data, including features: day (1–365), pressure (hPa), max/min/temperature (°C), dewpoint (°C), humidity (%), cloud (%), sunshine (hours), wind direction (°), wind speed (km/h), and binary rainfall (0/1). **Rainfall.csv** supplements this data. Test data (**test.csv**) has similar features. Preprocessing involves missing value imputation (forward fill, interpolation), outlier removal, and scaling. Future predictions use WeatherAPI forecasts.

**5. Approach (Machine Learning or Deep Learning):**

We use a machine learning approach using Random Forest is chosen for its robustness and high accuracy (84.77%) on this dataset. The task’s complexity—binary classification with structured weather data—suits Random Forest, which handles non-linear patterns and class imbalance (86% rain) better than deep learning. Deep learning (e.g., neural networks) was tested but underperformed (80.27%) due to the dataset’s moderate size and feature structure.