1. Data Loading and Exploration

The goal of this project is to there is a rainfall or not. We will be using Decision Tree Classifier and Random Forest models and compare them.

2. Data Preprocessing

Data Cleaning

• autoclean() and klib.datacleaning() were not used

• Removed unnecessary columns which are non-predictive such as id and day

• Handled missing values by filling them as appropriate (mode, median).

• encoding was unnecessary because columns were all numerical.

• scaling was done only on features assigning them to x variable with target value being exempted to prevent it from becoming continuous value

3. Model Training

Splitted the dataset into training and test sets using an 80-20 split

Model Selection

Two machine learning model was considered:

• Decision Tree Classifier  
• Random Forest Classifier

4. Evaluation

The following metrics were used to compare the models:

• Classification report

• Accuracy