Task 1 – Part 1

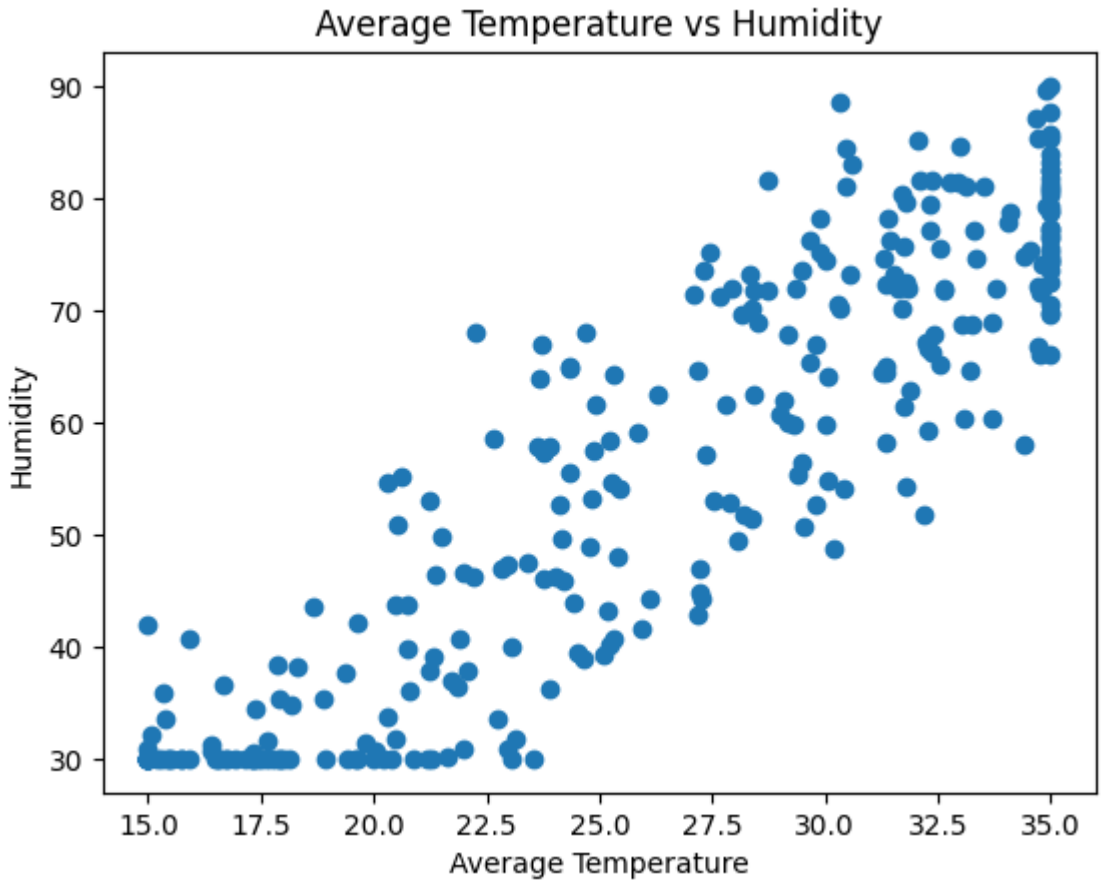
Team Name - OptiMinders

Data Preprocessing

To preprocess the data, rows with null values were removed. As a result, 15 rows were removed. The 'rain_or_not' feature was encoded using binary encoding.

Exploratory Data Analysis

To perform exploratory data analysis, a correlation matrix between the features was created. It showed that 'avg_temperature' and 'humidity' have a strong positive correlation. Boxplot were created to get an idea about data distribution.



Model Comparison

Several models were tested for the classification task.

- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost Classifier
- Majority Voting
- Multi-Layer Perceptron

Logistic regression was the best performing model with 0.7166 accuracy. Features were normalized and the models were tested again. But there was no improvement in accuracy.

Hyperparameter Tuning

Grid search was used to tune hyperparameters.

Time Series Forecasting

LSTM was used to predict the features for the next 21 days. Logistic regression was used to predict the probability of rain using the predicted parameters.