

Classification Report

1. Introduction

- **Dataset Description:** Describe the `weather_classification_data.csv` dataset, including features like Temperature, Humidity, Wind Speed, and the target variable Weather Type.
- **Objective:** State the objective of the analysis, which is to classify weather types based on the provided features using various classification algorithms.

2. Data Cleaning and Preparation

- **Loading the Data:** Load the dataset using `pandas` and display the first few rows to understand its structure.
- **Handling Missing Values:**
 - Identify and print the missing values for each column.
 - For numerical columns, fill missing values with the mean of that column.
 - For categorical columns, fill missing values with the mode (most frequent value).
- **Encoding Categorical Variables:** Convert categorical features into numerical values using Label Encoding.
- **Scaling Numerical Features:** Standardise numerical features using `StandardScaler` to normalise the data.

3. Data Analysis and Visualization

- **Summary Statistics:** Generate and display summary statistics (mean, median, standard deviation, etc.) of the dataset.
- **Histograms:** Create histograms to visualise the distribution of numerical features.
- **Scatter Plots:** Generate scatter plots to explore relationships between pairs of features.
- **Box Plots:** Use box plots to visualise the distribution of features and identify outliers.
- **Correlation Heatmaps:** Create a heatmap to show the correlation between numerical features.

4. Model Building

- **Splitting Data:** Split the data into training and testing sets.
- **Applying Classification Models:**
 - **Logistic Regression:** Train and evaluate a Logistic Regression model.
 - **Decision Tree:** Train and evaluate a Decision Tree Classifier.
 - **Random Forest:** Train and evaluate a Random Forest Classifier.

5. Model Evaluation

- **Classification Reports:** Generate classification reports for each model to evaluate performance metrics like accuracy, precision, recall, and F1-score.
- **Performance Comparison:**

- Create a DataFrame to compare the performance of different models.
- Plot a bar chart to visualise the comparison of model performance.

6. Conclusion

- **Summary of Findings:** Summarise the findings from the classification models, including the strengths and weaknesses of each model.
- **Future Work:** Suggest possible improvements, such as tuning hyperparameters, exploring additional features, or trying other classification algorithms.

I am also adding the graphs made from the visualisation codes:





