**Rainfall Weather Forecasting Project**

1. Problem Definition

Weather forecasting is the application of science and technology to predict the conditions of the atmosphere for a given location and time. Accurate weather forecasting is crucial for planning daily activities, agriculture, disaster management, and various industries.

In this project, we focus on rainfall prediction. Specifically, we aim to address the following two problems:

Classification Problem: Predict whether it will rain tomorrow.

Regression Problem: Predict the amount of rainfall that will occur.

The dataset used for this project contains approximately ten years of daily weather observations from different locations in Australia. The main challenge is to build models that can accurately predict rainfall based on historical weather data.

**2. Data Analysis**

**Dataset Overview:** The dataset consists of 23 columns, including the date, location, temperature, humidity, wind speed, and other meteorological variables.

**Key Features:**

* **MinTemp**: Minimum temperature in degrees Celsius.
* **MaxTemp**: Maximum temperature in degrees Celsius.
* **Rainfall**: Amount of rainfall recorded for the day in mm.
* **Evaporation**: Class A pan evaporation (mm) in the 24 hours to 9am.
* **Sunshine**: Number of hours of bright sunshine in the day.
* **WindGustDir**: Direction of the strongest wind gust in the 24 hours to midnight.
* **WindGustSpeed**: Speed (km/h) of the strongest wind gust in the 24 hours to midnight.
* **Humidity**: Humidity percentage at 9am and 3pm.
* **Pressure**: Atmospheric pressure reduced to mean sea level at 9am and 3pm.
* **RainToday**: Boolean indicating if rainfall exceeds 1mm.
* **RainTomorrow**: Amount of next day rain in mm (target variable).

**3. EDA Concluding Remarks**

**Exploratory Data Analysis (EDA)** is crucial to understand the data distribution, identify patterns, and detect anomalies. Here are some key findings from the EDA:

* **Temperature Patterns**: Higher temperatures tend to correlate with higher evaporation and sunshine hours.
* **Rainfall Distribution**: Rainfall data is highly skewed with many days having zero rainfall.
* **Wind and Humidity**: Certain wind directions and speeds are more common, and humidity levels vary significantly between morning and afternoon.

**Visualizations**:

* **Histograms**: Show the distribution of temperature, humidity, and rainfall.
* **Box Plots**: Highlight the range and outliers in key features.
* **Heatmaps**: Display correlations between features.

**4. Building Machine Learning Models**

**Model Selection**: For this project, we used Random Forest models for both classification and regression tasks due to their robustness and ease of use.

This project demonstrates the application of machine learning to weather forecasting, specifically rainfall prediction. The models built provide valuable insights and reasonably accurate predictions for both classification and regression tasks.

**Key Takeaways**:

* **Preprocessing**: Proper handling of missing values, encoding categorical features, and scaling numerical features are critical steps.
* **Model Building**: Random Forest models offer robust performance for both classification and regression tasks.
* Top of Form
* Bottom of Form

By accurately predicting rainfall, such models can aid in planning and decision-making across various sectors, showcasing the power and potential of machine learning in real-world applications

Top of Form

**Challenges Faced**:

1. **Data Quality**: Handling missing values and outliers was a significant challenge.
2. **Feature Selection**: Determining the most relevant features required extensive experimentation and domain knowledge.
3. **Model Complexity**: Balancing model complexity and interpretability was crucial to ensure the models are not only accurate but also understandable.

Bottom of Form