Classify Plants Based on Water Needs

Name : Aryan Khohar

Roll No: 202401100300068

Project Name: Classify Plants Based on Water Needs

Objective

The objective of this project is to classify the **water requirements of plants** (e.g., low, medium, high) based on their attributes such as **soil type** and other features. The classification model can be useful for agriculture planning, irrigation automation, and plant care systems.

Dataset Overview

- The dataset used contains information about different plant features including **soil type** and others (not named explicitly here).
- The **target variable** is water_need, which represents how much water a plant typically requires. This is a **categorical feature** with classes like Low, Medium, and High.

Steps Performed

1. Data Preprocessing

- Categorical feature soil_type was encoded using LabelEncoder.
- Target variable water_need was also encoded using LabelEncoder for model compatibility.

2. Feature Scaling

- All features were scaled using StandardScaler to ensure they are on the same scale, which helps improve model performance and convergence.

3. Train-Test Split

- The dataset was split into **80% training** and **20% testing** using train_test_split.

4. Model Training

- A **Random Forest Classifier** with 100 trees was trained on the scaled training data.
- Random Forest was chosen for its high accuracy, robustness, and ability to handle both linear and non-linear patterns.

5. Model Evaluation

- The model's predictions were compared against actual labels using:
  - **Confusion Matrix**: To visualize classification performance.
  - **Classification Report**: To show precision, recall, and F1-score per class.

6. Feature Importance

- A bar chart of feature importances was generated to understand which features contributed most to the predictions.

Results

✅ Classification Report

- The model provided a detailed breakdown of precision, recall, and F1-score for each class (low, medium, high).
- It shows how well the model distinguishes between different water needs.

📊 Confusion Matrix

- The heatmap of the confusion matrix allowed easy identification of misclassified instances.
- High diagonal values indicated strong performance in correctly predicting the classes.

🔍 Feature Importance

- The bar plot of feature importance revealed which features had the greatest impact on determining the water need.
- This insight is valuable for focusing on the most influential factors during cultivation or irrigation planning.

Conclusion

The Random Forest model effectively classifies plant water needs based on given features. With high accuracy and clear visualization tools, this project demonstrates a practical application of machine learning in agriculture and environmental management.

Future Improvements

- Include more detailed plant features like leaf type, sunlight exposure, temperature tolerance, etc.
- Try other classification models (e.g., SVM, Gradient Boosting) for comparison.
- Deploy the model using a simple web interface with tools like **Gradio** or **Streamlit**.