Classify Plants Based on Water Needs

Submitted By- Abhishek Kumar Maurya
Roll No.- 202401100300008
Branch -CSE(AI)-A
Guided By-Bikki Kumar Sir

Objective

To develop a machine learning model that classifies plants based on their water requirements. The aim is to assist in efficient gardening and agricultural planning by understanding water needs from plant features.

Problem Statement

Water management is critical in agriculture and horticulture. Overwatering or underwatering plants can lead to poor health or even plant death. By classifying plants based on their water needs, it becomes easier for gardeners, farmers, and urban planners to optimize irrigation systems and water usage. This project leverages a machine learning classifier to automate this classification based on plant features.

Dataset

The dataset includes various features describing plant characteristics. The target variable represents the water requirement category (e.g., Low, Medium, High). Categorical features are encoded and numerical data is used directly for training the model.

Tasks

1. Data Collection

- Load the dataset (plants.csv) into a pandas DataFrame.

2. Data Preprocessing

- Encode categorical variables using LabelEncoder.

- Separate features and target variable.

3. Model Implementation

- Split the dataset into training and testing sets.

- Train a classifier using scikit-learn.

4. Model Evaluation

- Evaluate the model using classification report metrics like precision, recall, and F1-score.

- Visualize the class distribution using seaborn.

5. Prediction and Interpretation

- Predict water needs for test data and interpret results.

6. Documentation

- Document all steps, code snippets, model outputs, and visualizations.

Python Code

```
import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import LabelEncoder

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import classification_report

import matplotlib.pyplot as plt

import seaborn as sns


file_path = '/content/plants.csv'
```

```
df = pd.read_csv(file_path)

print("First few rows:")

print(df.head())


label_encoders = {}

for column in df.columns:

    if df[column].dtype == 'object':

        le = LabelEncoder()

        df[column] = le.fit_transform(df[column])

        label_encoders[column] = le


target_column = df.columns[-1]

X = df.drop(columns=[target_column])

y = df[target_column]


X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = RandomForestClassifier(random_state=42)

model.fit(X_train, y_train)

y_pred = model.predict(X_test)


print("\nClassification Report:")

print(classification_report(y_test, y_pred))


plt.figure(figsize=(8, 5))
```

```
sns.countplot(data=df, x=target_column)

plt.title('Distribution of Plant Categories')

plt.xticks(rotation=45)

plt.tight_layout()

plt.show()
```

Reference:- Github, Data Analysis.