****

**Assessment Report**

on

**“Classify Plants Based on Water Needs”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AI)**

By

Name : ANUSHKA KAUSHAL

Roll Number : 202401100300058

Section: A

**Under the supervision of**

“BIKKY KUMAR”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

Efficient water management is vital in sustainable agriculture and horticulture. Determining the water requirements of various plants manually can be time-consuming and error-prone. This project aims to automate the classification of plant water needs using supervised machine learning. By analyzing key environmental attributes such as sunlight exposure, soil type, and watering frequency, a predictive model is developed to classify plants into different water need categories. This assists gardeners, farmers, and researchers in making data-driven decisions for optimized plant care and resource conservation.

**2. Problem Statement**

To classify plants based on their water requirements using environmental features such as sunlight exposure, soil composition, and watering frequency. This classification will help in optimizing irrigation practices and supporting water conservation strategies.

**3. Objectives**

● Preprocess the dataset for training a machine learning model.  
● Train a Random Forest Classifier to categorize plant water needs.  
● Evaluate model performance using standard classification metrics.  
● Visualize the confusion matrix using a heatmap for interpretability.

**4. Methodology**

**Data Collection**: The user uploads a CSV file containing the dataset.

**Data Preprocessing**:

* + Handling missing values using mean and mode imputation.
  + One-hot encoding of categorical variables.
  + Feature scaling using StandardScaler.

**Model Building**:

* + Splitting the dataset into training and testing sets.
  + Training a Logistic Regression classifier.

**Model Evaluation**:

* + Evaluating accuracy, precision, recall, and F1-score.
  + Generating a confusion matrix and visualizing it with a heatmap.

**5. Data Preprocessing**

* The dataset is cleaned and prepared as follows:
* ● Categorical columns such as soil type and water need are encoded using Label Encoding.  
  ● All unnamed or index columns are dropped.  
  ● Numerical features are scaled using StandardScaler.  
  ● The dataset is split into 80% training and 20% testing subsets for model training and evaluation.

**6. Model Implementation**

Random Forest is selected as the classification model due to its robustness, ability to handle mixed data types, and resistance to overfitting. The model is trained on the processed features and evaluated on unseen test data to predict the water requirement category of each plant.

**7. Evaluation Metrics**

* The following metrics are used to evaluate model performance:
* ● **Accuracy**: Proportion of correct predictions over total predictions.  
  ● **Precision**: Proportion of predicted positives that are actual positives.  
  ● **Recall**: Proportion of actual positives that were correctly predicted.  
  ● **F1 Score**: Harmonic mean of precision and recall.  
  ● **Confusion Matrix**: Visualized using Seaborn heatmap to interpret classification accuracy for each class.

**8. Results and Analysis**

● The trained model demonstrated strong performance across all classes.  
● The confusion matrix heatmap revealed clear classification with minimal misclassifications.  
● The precision and recall scores confirmed the model’s capability to correctly distinguish different water need levels.  
● The model can be scaled further by adding more features like humidity, seasonality, and temperature.

**9. Conclusion**

The Random Forest model effectively classified plant water needs using simple environmental features. This project showcases how machine learning can enhance resource planning and water management in agriculture. Future improvements may include exploring advanced models such as Gradient Boosting and handling more complex datasets involving temporal or geospatial data., improvements can be made by exploring more advanced models and handling imbalanced data.

**10. References**

● scikit-learn documentation  
● pandas documentation  
● Seaborn visualization library  
● Articles and research on plant irrigation modeling and environmental AI solutions





 