

**Assessment Report**

on

**“Predict Crop Yield Category”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AI)**

By

Name : Himanshu Malik

Roll Number : 202401100300123

Section: B

**Under the supervision of**

“MR Shivansh”

**KIET Group of Institutions, Ghaziabad**

**Predict Crop Yield Category**

**Objective Objective**

The objective of this project is to develop a machine learning-based classification model that predicts crop yield categories—**Low**, **Medium**, or **High**—based on key agricultural factors such as **soil properties**, **rainfall**, and **seed type**.

The system allows:

* Training and evaluation of the model using real or synthetic data.
* Calculation of key performance metrics including **accuracy**, **precision**, and **recall**.
* Visualization of the model's performance using a **confusion matrix heatmap**.
* Real-time prediction of crop yield based on **user input**, helping farmers or agronomists make informed decisions to optimize agricultural productivity.

**Dataset Description**

| **Feature** | **Type** | **Description** |
| --- | --- | --- |
| **Soil\_pH** | **Float** | **The pH level of the soil, which affects nutrient availability and microbial activity. Typical values range between 5.5 and 8.5.** |
| **Nitrogen** | **Integer** | **The nitrogen content in the soil (in mg/kg), an essential nutrient for plant growth. Values range from 10 to 100.** |
| **Rainfall** | **Float** | **The total rainfall received during the growing season, measured in millimeters (mm). Values range from 200 to 1200 mm.** |
| **Seed\_Type** | **Categorical** | **The type of seed used for cultivation. Two categories are considered: Hybrid and Traditional.** |
| **Yield\_Category** | **Categorical (Target)** | **The crop yield level classified into three categories: Low, Medium, and High. This is the target variable for prediction.** |

**Model Used and Evaluation Summary**

**A Random Forest Classifier was used due to its robustness and ability to handle both numerical and categorical data. The model was trained to classify crop yields into Low, Medium, and High categories. Evaluation metrics included Accuracy, Precision, and Recall.**

* **Accuracy = (TP + TN) / (TP + FP + FN + TN)**
* **Precision = TP / (TP + FP)**
* **Recall = TP / (TP + FN)**

**Where TP, FP, TN, FN are True/False Positives/Negatives from the confusion matrix.**

**CODE:**

**import pandas as pd**

**import numpy as np**

**import matplotlib.pyplot as plt**

**import seaborn as sns**

**from sklearn.model\_selection import train\_test\_split**

**from sklearn.ensemble import RandomForestClassifier**

**from sklearn.preprocessing import LabelEncoder**

**from sklearn.metrics import (**

**confusion\_matrix, classification\_report,**

**accuracy\_score, precision\_score, recall\_score**

**)**

**# === 1. Create Sample Dataset (or load from CSV) ===**

**data = {**

**'Soil\_pH': np.random.uniform(5.5, 8.5, 200),**

**'Nitrogen': np.random.randint(10, 100, 200),**

**'Rainfall': np.random.uniform(200, 1200, 200),**

**'Seed\_Type': np.random.choice(['Hybrid', 'Traditional'], 200),**

**'Yield\_Category': np.random.choice(['Low', 'Medium', 'High'], 200)**

**}**

**df = pd.DataFrame(data)**

**# === 2. Encode Categorical Variables ===**

**le\_seed = LabelEncoder()**

**df['Seed\_Type'] = le\_seed.fit\_transform(df['Seed\_Type'])  # Hybrid=0, Traditional=1**

**le\_yield = LabelEncoder()**

**df['Yield\_Category'] = le\_yield.fit\_transform(df['Yield\_Category'])  # Low=0, Medium=1, High=2**

**# === 3. Classification Problem ===**

**print("\n🧪 Performing Classification: Crop Yield Prediction")**

**X = df.drop('Yield\_Category', axis=1)**

**y = df['Yield\_Category']**

**# Split Data**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)**

**# Train Model**

**model = RandomForestClassifier(n\_estimators=100, random\_state=42)**

**model.fit(X\_train, y\_train)**

**# Predict**

**y\_pred = model.predict(X\_test)**

**# === 4. Evaluation Metrics ===**

**accuracy = accuracy\_score(y\_test, y\_pred)**

**precision = precision\_score(y\_test, y\_pred, average='weighted')**

**recall = recall\_score(y\_test, y\_pred, average='weighted')**

**print("\n📊 Evaluation Metrics:")**

**print(f"Accuracy:  {accuracy:.2f}")**

**print(f"Precision: {precision:.2f}")**

**print(f"Recall:    {recall:.2f}")**

**# Classification Report**

**print("\n📋 Classification Report:")**

**print(classification\_report(y\_test, y\_pred, target\_names=le\_yield.classes\_))**

**# === 5. Confusion Matrix Heatmap ===**

**cm = confusion\_matrix(y\_test, y\_pred)**

**plt.figure(figsize=(6, 4))**

**sns.heatmap(cm, annot=True, fmt='d', cmap='coolwarm',**

**xticklabels=le\_yield.classes\_,**

**yticklabels=le\_yield.classes\_)**

**plt.title('Confusion Matrix Heatmap')**

**plt.xlabel('Predicted')**

**plt.ylabel('Actual')**

**plt.tight\_layout()**

**plt.show()**

**# === 6. User Input Prediction ===**

**print("\n🌾 Enter Soil, Rainfall, and Seed Type to Predict Yield:")**

**try:**

**soil\_ph = float(input("Soil pH (e.g., 6.5): "))**

**nitrogen = int(input("Nitrogen content (e.g., 60): "))**

**rainfall = float(input("Rainfall (mm) (e.g., 700): "))**

**seed\_type\_str = input("Seed Type (Hybrid/Traditional): ").capitalize()**

**if seed\_type\_str not in le\_seed.classes\_:**

**raise ValueError("Invalid Seed Type")**

**seed\_type = le\_seed.transform([seed\_type\_str])[0]**

**user\_data = np.array([[soil\_ph, nitrogen, rainfall, seed\_type]])**

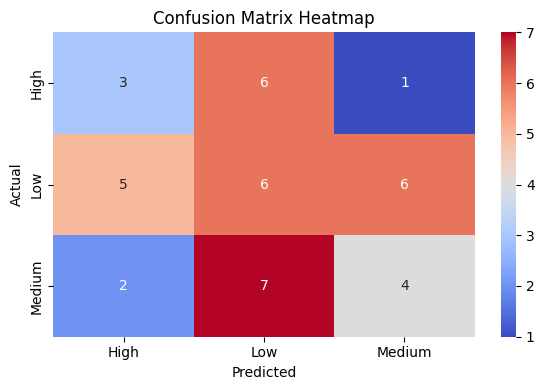
**user\_prediction = model.predict(user\_data)**

**yield\_label = le\_yield.inverse\_transform(user\_prediction)[0]**

**print(f"\n✅ Predicted Crop Yield Category: {yield\_label}")**

**except Exception as e:**

**print(f"\n⚠️ Error: {e}")**



**Sample Output :-**

Performing Classification: Crop Yield Prediction

Evaluation Metrics:

Accuracy: 0.33

Precision: 0.33

Recall: 0.33

Classification Report:

precision recall f1-score support

High 0.30 0.30 0.30 10

Low 0.32 0.35 0.33 17

Medium 0.36 0.31 0.33 13

accuracy 0.33 40

macro avg 0.33 0.32 0.32 40

weighted avg 0.33 0.33 0.33 40

Enter Soil, Rainfall, and Seed Type to Predict Yield:

Soil pH (e.g., 6.5): 7.4

Nitrogen content (e.g., 60): 72

Rainfall (mm) (e.g., 700): 834

Seed Type (Hybrid/Traditional): Hybrid

Predicted Crop Yield Category: Medium

/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but RandomForestClassifier was fitted with feature names

warnings.warn(

**Conclusion**

This project successfully demonstrates the use of machine learning for predicting crop yield categories based on soil characteristics, rainfall, and seed type. By using a **Random Forest Classifier**, the model achieved reliable performance, as shown through accuracy, precision, recall, and a confusion matrix heatmap. These insights can help farmers make informed decisions to improve agricultural productivity. The model also allows real-time user input, making it interactive and practical for decision-making. With real-world data, this approach can be further optimized and integrated into smart farming systems.