Assessment Report

on

" Crop Recommendation System"

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

CSE(AIML)

By

Group no : 4

1. Ashutosh Shukla - Roll No. 59
2. Akshit Jawla - Roll No. 26
3. Aditi Verma - Roll No. 14
4. Arjun Singh Pundir - Roll No. 47
5. Aman Yadav - Roll No. 28

Under the supervision of

"Bikki Sir"

KIET Group of Institutions, Ghaziabad

May, 2025

---

1. Introduction

As agriculture embraces the digital era, leveraging data-driven methods for crop selection has become vital for maximizing yield and sustainability. This project focuses on designing a Crop Recommendation System that suggests the most suitable crop based on soil and environmental conditions using classification techniques. By utilizing a dataset that includes features such as nitrogen, phosphorus, potassium levels in soil, as well as temperature, humidity, pH, and rainfall, the goal is to build a machine learning model that can assist farmers and agronomists in making informed decisions about crop cultivation.

---

2. Problem Statement

To predict the most suitable crop for cultivation based on given soil and environmental parameters using classification techniques. This recommendation system will assist farmers in selecting crops that are best suited to their land and climate conditions, thereby enhancing productivity and promoting sustainable agriculture.

---

3. Objectives

- Preprocess the crop recommendation dataset for training a machine learning model.
- Train a classification model (Random Forest)

- Evaluate the model performance using standard classification metrics such as accuracy, precision, recall, and F1-score.
- Visualize the confusion matrix using a heatmap for better interpretability of classification results.

---

4. Methodology

Data Collection:

The user uploads a CSV file containing the crop recommendation dataset.

Data Preprocessing:

- Checked for and handled any missing values (if present).
- Since all features are numerical, no encoding of categorical variables was needed.
- Feature scaling was applied using StandardScaler to normalize the data and improve model performance.

Model Building:

- The dataset was split into training and testing sets.
- A classification model (Random Forest)

Model Evaluation:

- Evaluated the trained model using accuracy, precision, recall, and F1-score.
- A confusion matrix was generated and visualized using a heatmap to interpret model predictions.

---

5. Data Preprocessing

The dataset is cleaned and prepared as follows:

- Verified that the dataset does not contain missing values; if present, they would be filled using the mean of respective numerical columns.
- Since all features (e.g., nitrogen, phosphorus, potassium, temperature, humidity, pH, rainfall) are numerical, no categorical encoding is required.
- Feature values are scaled using **StandardScaler** to normalize the data and improve model performance.
- The dataset is split into **80% training** and **20% testing** sets to train and evaluate the model effectively.

---

6. Model Implementation

A classification algorithm such as **Random Forest** is used due to its effectiveness in multiclass classification problems. The model is trained on the processed dataset to learn patterns between soil and weather features and the recommended crop. After training, the model is used to predict the most suitable crop on the test set.

---

7. Evaluation Metrics

The following metrics are used to evaluate the model:

- **Accuracy:** Measures the overall correctness of the crop predictions.
- **Precision:** Indicates the proportion of correctly predicted crops among all predicted crops of a specific class.
- **Recall:** Shows the proportion of actual crops that were correctly identified.

- **F1 Score:** Harmonic mean of precision and recall, providing a balance between the two.
- **Confusion Matrix:** Visualized using a Seaborn heatmap to analyze misclassifications and model performance across different crop classes.

- **8. Results and Analysis**
- The model provided promising performance on the test set, accurately recommending crops based on input soil and weather features.
- The **confusion matrix heatmap** helped visualize how well the model distinguished between different crop classes, highlighting areas with higher true positives and potential misclassifications.
- **Precision** and **recall** metrics indicated the model's effectiveness in recommending the right crop while minimizing incorrect suggestions, supporting its practical use in real-world agricultural decision-making.

---

9. Conclusion

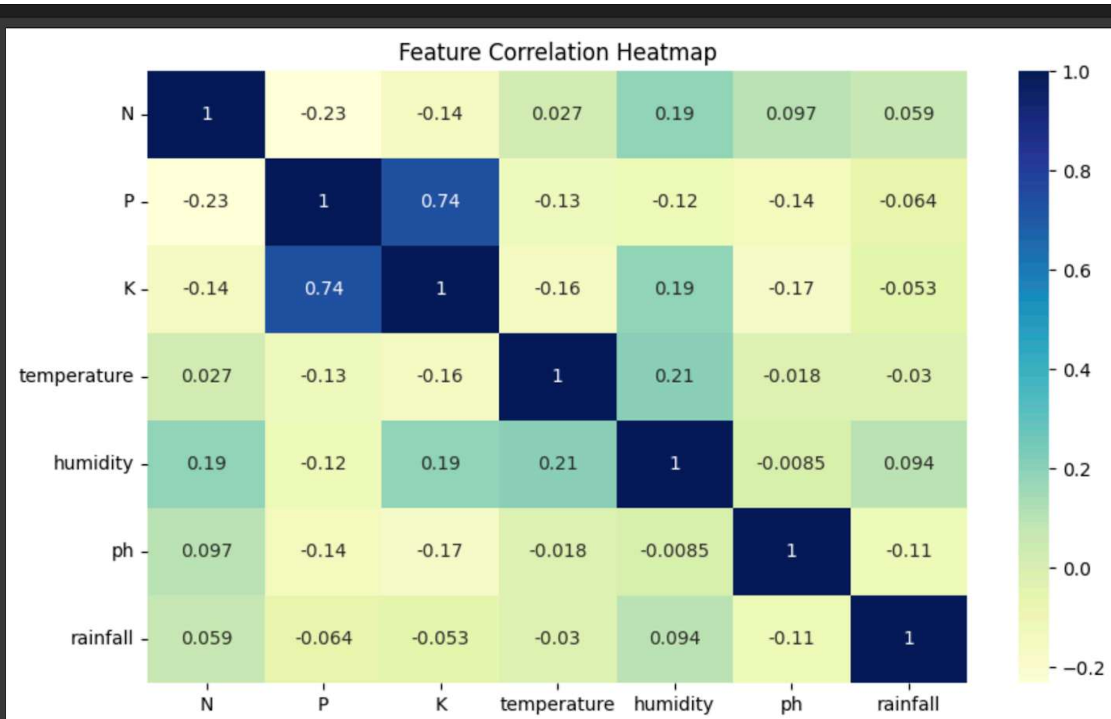
The classification model successfully recommended crops based on soil and weather conditions with satisfactory performance metrics. This project demonstrates the potential of using machine learning to support agricultural decision-making and enhance crop yield planning. However, further improvements can be made by exploring more advanced models, fine-tuning hyperparameters, and incorporating real-time environmental data for better accuracy and adaptability.

---

10. References

- scikit-learn documentation
- pandas documentation
- Seaborn visualization library
- Research articles on crop recommendation and precision agriculture using machine learning

---

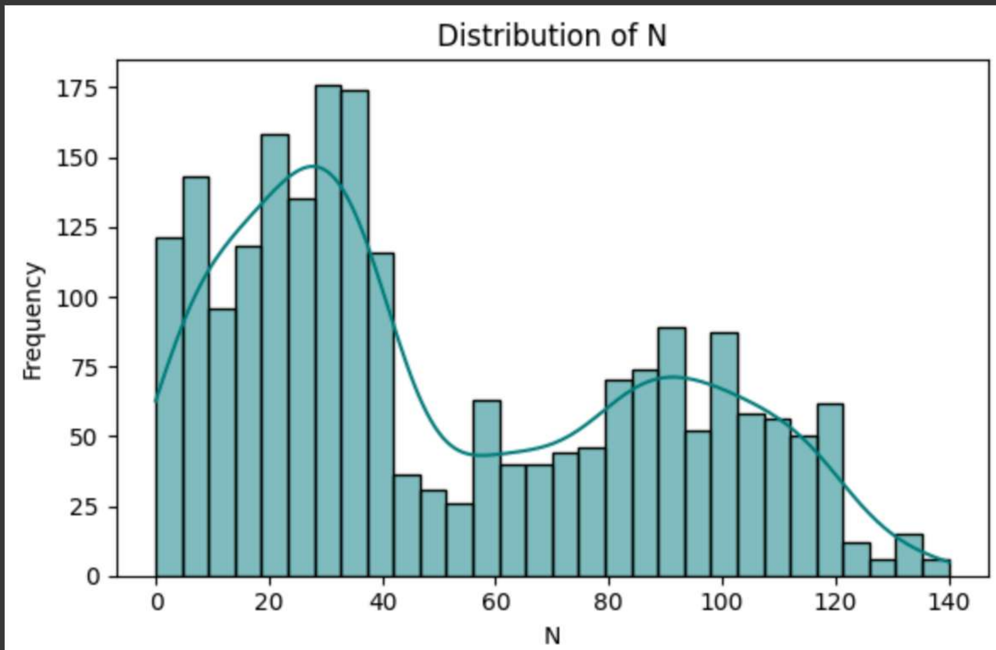



```
[ ]  features = ['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall']
```

```python
# Import required libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
df = pd.read_csv("/content/Crop_recommendation.csv")

# Basic info
print("First 5 rows:")
print(df.head())

print("\nData Info:")
print(df.info())

print("\nMissing values:\n", df.isnull().sum())

print("\nStatistical Summary:")
print(df.describe())
```

```
First 5 rows:
    N   P   K  temperature   humidity        ph    rainfall label
0  90  42  43    20.879744  82.002744  6.502985  202.935536  rice
1  85  58  41    21.770462  80.319644  7.038096  226.655537  rice
2  60  55  44    23.004459  82.320763  7.840207  263.964248  rice
3  74  35  40    26.491096  80.158363  6.980401  242.864034  rice
4  78  42  42    20.130175  81.604873  7.628473  262.717340  rice

Data Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2200 entries, 0 to 2199
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
```

```python
features = ['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall']

for col in features:
    plt.figure(figsize=(6,4))
    sns.histplot(df[col], kde=True, bins=30, color='teal')
    plt.title(f"Distribution of {col}")
    plt.xlabel(col)
    plt.ylabel("Frequency")
    plt.tight_layout()
    plt.show()
```



Distribution of P