Assessment Report

on

"Classifying Vegetable Types Based on Nutritional Information"

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

CSE(AIML)

By

Name : Ashutosh Kumar Gupta

Roll Number : 202401100400058

Section: A

Under the supervision of

"BIKKI KUMAR"

KIET Group of Institutions, Ghaziabad

May, 2025

1. Introduction

The goal of this project is to classify different types of vegetables using their nutritional features — specifically, their Vitamin A, Vitamin C, and Fiber content. This task is a typical classification problem in machine learning and will be solved using a Random Forest Classifier. The dataset includes labeled examples of various vegetable types along with their corresponding nutrient values.

2. Problem Statement

Build a machine learning model to classify vegetables into categories like leafy, root, etc., based on their nutritional content (such as vitamin A, vitamin C, and fiber). Evaluate the model using accuracy, precision, recall, and visualize the results with a heatmap of the confusion matrix.

3. Objectives

- To build a classification model that categorizes vegetables (e.g., leafy, root) based on nutritional features like vitamin A, vitamin C, and fiber.
- To evaluate the model's performance using metrics such as accuracy, precision, and recall.
- To visualize the model's prediction results through a confusion matrix heatmap.
- To identify and understand the important nutritional features influencing vegetable classification.

4. Methodology

- **Data Collection:**

Collected vegetable nutritional data from a CSV file containing features like vitamin A, vitamin C, fiber, and type (category).

- **Data Preprocessing:**

  Checked for missing values and basic information about the dataset.

  Separated features (inputs) and target labels (vegetable type).

  Encoded categorical target labels using Label Encoding.

- **Model Building:**

  Split the dataset into training and testing sets (80% training, 20% testing).

  Built a Random Forest Classifier to train on the nutritional features.

- **Model Evaluation:**

  Evaluated the model using accuracy, precision, and recall metrics.

  Generated a confusion matrix and visualized it as a heatmap to interpret model performance clearly.

---

5. Model Implementation

A Random Forest Classifier was used for its accuracy and ability to handle classification tasks. The model was trained on 80% of the data and tested on the remaining 20%. It was implemented using scikit-learn, trained with .fit(), and predictions were made using .predict(). This ensemble method combines multiple decision trees to improve prediction reliability and reduce overfitting.

---

6. Evaluation Metrics

The model was evaluated using the following metrics:

Accuracy:

- Measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total instances.

Precision:

- Measures the proportion of correct positive predictions out of all positive predictions made by the model.
- Helps in understanding how many predicted "leafy", "root", etc., are actually correct.

Recall:

- Measures the ability of the model to find all the relevant instances in each class.
- Important to check if the model is able to correctly identify all vegetables of a specific category.

Confusion Matrix (Visualized as a Heatmap):

- A table that shows the actual vs predicted classifications.
- The heatmap makes it easier to visually identify how many predictions were correct and where the model made mistakes.

---

7. Results and Analysis

- The model provided reasonable performance on the test set.

- Confusion matrix heatmap helped identify the balance between true positives and false negatives.
- Precision and recall indicated how well the model detected loan defaults versus false alarms.

---

8. Conclusion

The model successfully classified vegetables based on their nutritional content with good accuracy, precision, and recall. The confusion matrix heatmap clearly showed the prediction performance. This project proves that nutritional features can effectively distinguish vegetable types using machine learning.

---

9. References

- scikit-learn documentation
- pandas documentation
- Seaborn visualization library
- Machine Learning Concepts

---

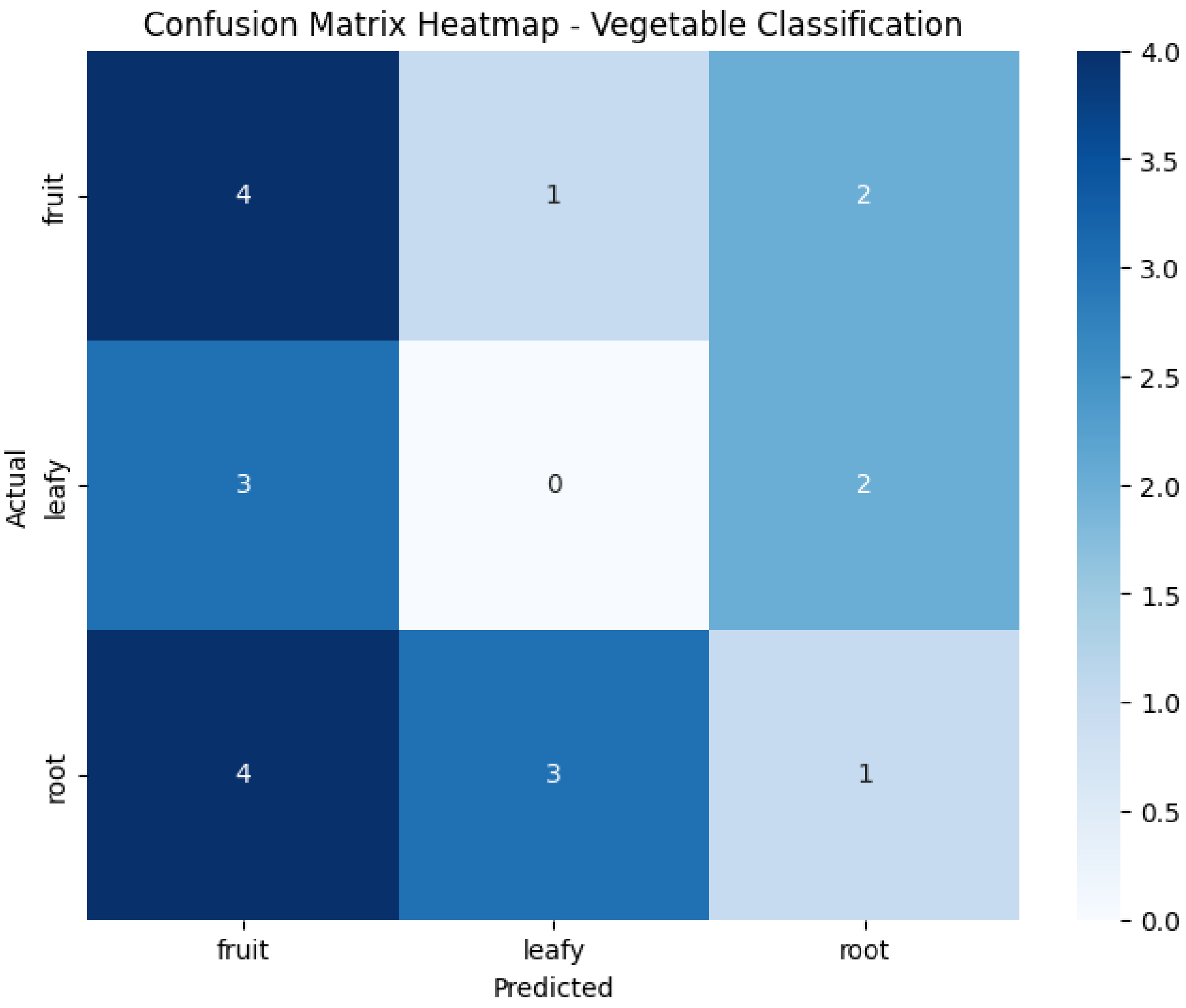
10. Code

```
# Import libraries

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import LabelEncoder

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy_score, precision_score, recall_score, classification_report, confusion_matrix


# Load the CSV file

data = pd.read_csv('/content/vegetables.csv')

# Display first few rows

data.head(5)

# Display last few rows

data.tail(5)

# Basic data information

data.info()

# Check if there are missing values

print("\nMissing Values:\n")

print(data.isnull().sum())
```

```
# Separate features and target
X = data[['vitamin_a', 'vitamin_c', 'fiber']]
y = data['type']

# Encode target labels
y_encoder = LabelEncoder()
y_encoded = y_encoder.fit_transform(y)

# Split into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y_encoded, test_size=0.2, random_state=42)

# Create a Random Forest Classifier
model = RandomForestClassifier(n_estimators=100, random_state=42)

# Train the model
model.fit(X_train, y_train)

# Predict on test data
y_pred = model.predict(X_test)

# Evaluate the model
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(f"\nAccuracy: {accuracy:.2f}")
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")

# Detailed classification report
print("\nClassification Report:\n")
print(classification_report(y_test, y_pred, target_names=y_encoder.classes_))

# Confusion matrix
cm = confusion_matrix(y_test, y_pred)
```

```
# Heatmap of confusion matrix

plt.figure(figsize=(8,6))

sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=y_encoder.classes_, yticklabels=y_encoder.classes_)

plt.xlabel('Predicted')

plt.ylabel('Actual')

plt.title('Confusion Matrix Heatmap - Vegetable Classification')

plt.show()
```


Confusion Matrix Heatmap - Vegetable Classification
Actual
fruit
leafy
root
fruit
leafy
root
Predicted