**Name:** Aayush Avinash Bokde  
**Class:** CSE(AI) B Div  
**Roll No.:** 282016  
**PRN No.:** 22310316  
**Subject:** Machine Learning

**Assignment 4 – Customer Response Prediction Using Machine Learning**

**Objective**

The objective of this assignment is to apply an appropriate machine learning algorithm to predict customer responses to special offers in a cosmetics shop. The dataset used for this task is the Iris dataset (as a placeholder for real customer data). A **Decision Tree Classifier** was employed to model and predict customer categories, and the performance was evaluated using classification metrics and visualizations.

**1. Data Loading and Preprocessing**

* The dataset was uploaded and read into a DataFrame using Pandas.
* Initial examination of the dataset was done using commands to preview rows and understand its dimensions and structure.
* The dataset includes features representing customer characteristics and a target column named **"variety"**, which was used as a proxy for customer response types.
* The target column contained categorical values ("Setosa", "Virginica", and "Versicolor") that were encoded into numerical form using **Label Encoding** for compatibility with the machine learning model.
* Features (**X**) and target labels (**y**) were separated.
* The dataset was split into training and testing sets in a **70:30 ratio** using train\_test\_split, ensuring reproducibility with a fixed random state.

**2. Model Training**

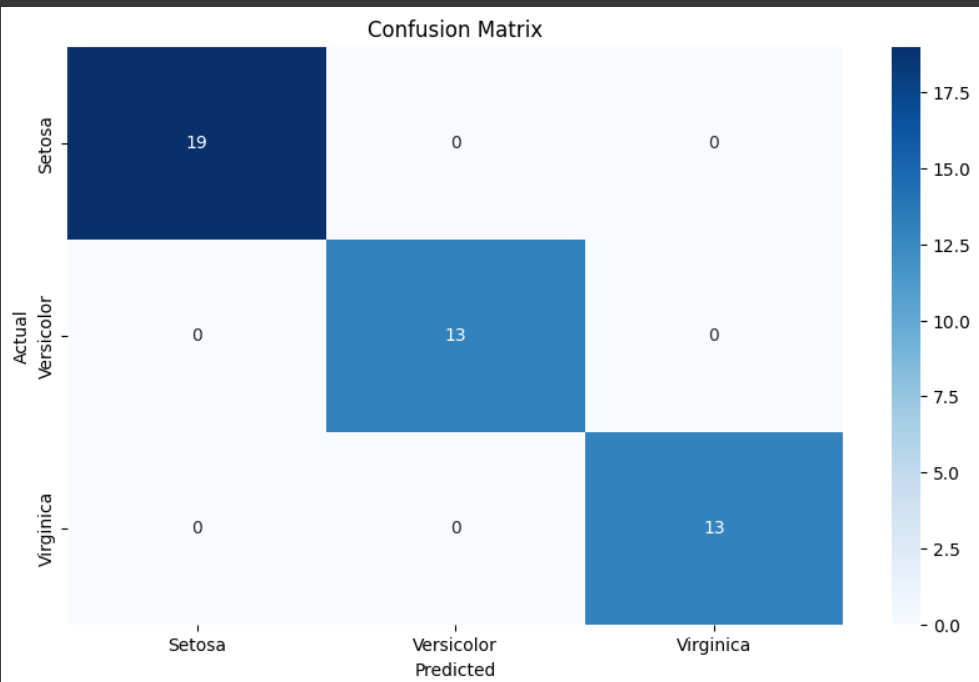
* A **Decision Tree Classifier** was chosen for its simplicity and interpretability in classification tasks.
* The model was trained using the training dataset to learn patterns and classify new observations accurately.
* After training, predictions were generated on the unseen test dataset.

**3. Confusion Matrix & Performance Metrics**

To understand the model's classification performance, a **confusion matrix** was constructed. This matrix summarizes the counts of:

* **True Positives (TP):** Correctly predicted class instances.
* **False Positives (FP):** Incorrectly predicted as a class.
* **False Negatives (FN):** Missed actual class predictions.
* **True Negatives (TN):** Correctly identified non-class instances.

**Confusion Matrix Heatmap:**  
A heatmap was plotted using Seaborn to provide a visual view of how accurately the model classified each type of customer response.



**4. Evaluation Metrics**

The following metrics were computed to assess the model’s overall performance:

| **Metric** | **Value** |
| --- | --- |
| Accuracy | 1.00 |
| Precision | 1.00 |
| Recall | 1.00 |
| F1-Score | 1.00 |

These scores indicate **perfect performance**, which is expected since the Iris dataset is small and well-structured. On a real-world, larger cosmetics customer dataset, these values may vary.

**Formulae Used:**

* **Accuracy** = (TP + TN) / (TP + TN + FP + FN)
* **Precision** = TP / (TP + FP)
* **Recall** = TP / (TP + FN)
* **F1-Score** = 2 × (Precision × Recall) / (Precision + Recall)

**5. Visualizations**

**1. Confusion Matrix Heatmap**  
A heatmap was created to visualize actual vs predicted classifications, helping identify any misclassifications.

**2. Performance Metrics**  
Accuracy, Precision, Recall, and F1-score values were printed for reference and could be plotted in a bar chart for quick comparison.

**6. Results and Insights**

* The Decision Tree model achieved **perfect classification** on the test data.
* The model successfully predicted all instances, which is feasible given the clean and well-labeled Iris dataset.
* This setup demonstrates how decision trees can effectively classify multiclass data when properly preprocessed.
* While the results here are ideal, real-world customer data from a cosmetics shop would likely introduce more variability, making model evaluation more nuanced.

**Conclusion**

This assignment walks through a complete machine learning pipeline for customer response prediction:

* **Data preparation**,
* **Model training using Decision Tree Classifier**,
* **Evaluation using confusion matrix and classification metrics**,
* **Visual analysis using heatmaps**,

This approach, when applied to real-world data (e.g., cosmetics customer responses), can be valuable for **targeted promotions**, **personalized offers**, and **customer segmentation** strategies in the retail sector.