**Assignment 4**

**Name: Ashutosh Shinde**

**Roll no: 282006**

**Batch: B1**

**Problem Statement**

Apply a machine learning algorithm to predict customer response (acceptance/rejection) for a special offer in a cosmetics shop. Evaluate the model using:

1. **Confusion Matrix**
2. **Accuracy**
3. **Precision**
4. **Recall**
5. **F1-Score**

**Objective**

1. To preprocess and encode categorical data for modeling.
2. To train a Random Forest classifier for binary classification.
3. To evaluate model performance using key metrics.

**Resources Used**

* **Software:** Google Colab
* **Libraries:**
  + Pandas (Data Manipulation)
  + Scikit-learn (Model Training & Evaluation)

**Methodology**

**1. Data Preprocessing**

* **Dataset:** Heart.csv (represents customer health data, analogous to cosmetic shop data).
* **Handled Categorical Variables:**
  + One-hot encoded ChestPain and Thal.
  + Binary encoded target (AHD → 1 for "Yes", 0 for "No").
* **Train-Test Split:** 80% training, 20% testing.

**2. Model Training**

* **Algorithm:** Random Forest Classifier (n\_estimators=100).
* **Baseline Comparison:** SVM (linear kernel) for reference.

**3. Evaluation Metrics**

* **Confusion Matrix:** TP, FP, TN, FN.
* **Accuracy:** Overall correctness.
* **Precision:** Proportion of true positives among predicted positives.
* **Recall:** Proportion of actual positives correctly predicted.
* **F1-Score:** Harmonic mean of precision and recall.

**Output:**

Confusion Matrix:

[[29 6]

[10 16]]

Accuracy: 0.74

Precision: 0.74

Recall: 0.74

F1 Score: 0.73

**Interpretation of Results**

**Confusion Matrix**

* **True Positives (TP):** 16
* **False Positives (FP):** 6
* **True Negatives (TN):** 29
* **False Negatives (FN):** 10

**Key Metrics**

1. **Accuracy (74%):** The model correctly predicts 74% of all cases.
2. **Precision (74%):** When the model predicts "Yes", it is correct 74% of the time.
3. **Recall (74%):** The model identifies 74% of all actual "Yes" responses.
4. **F1-Score (73%):** Balanced measure of precision and recall.

**Conclusion**

1. The Random Forest model achieved **moderate performance** (74% accuracy).
2. **Improvement Suggestions:**
   * Address class imbalance (if present) using resampling.
   * Tune hyperparameters (e.g., max\_depth, min\_samples\_split).
   * Try other algorithms (e.g., Gradient Boosting).
3. This framework can be directly applied to **cosmetic shop data** by replacing the dataset with customer purchase history and demographics.