

Name: Rohit Bhabire

Roll No: 281042 Batch: A2

Assignment 4

Problem Statement

Apply an appropriate Machine Learning (ML) algorithm on a dataset collected in a cosmetics shop, showing details of customers, to **predict customer response for a special offer**. Create a **confusion matrix** and calculate the following metrics:

- a) Accuracy
 - b) Precision
 - c) Recall
 - d) F1-Score
-

Objectives

1. Apply a supervised machine learning algorithm to predict customer responses.
 2. Analyze and preprocess the dataset for better model performance.
 3. Evaluate the model performance using a confusion matrix.
 4. Calculate classification metrics (Accuracy, Precision, Recall, F1-Score).
-

Resources Used

- **Software:** Visual Studio Code
 - **Libraries:** pandas, matplotlib, seaborn, sklearn
-

Theory

Classification is a supervised learning method where the model learns to predict a label from input features. The task involves binary classification – predicting whether a customer will respond (Yes) or not (No).

Confusion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Evaluation Metrics

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

Methodology

1. Data Preprocessing

- Load the dataset using pandas.
- Handle missing values through imputation or removal.
- Encode categorical features (e.g., Gender) using one-hot encoding.

- Normalize numerical features using MinMaxScaler or StandardScaler.
- Split the dataset into training (75%) and testing (25%) sets.

2. Model Selection

For binary classification, suitable ML algorithms include:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)

3. Model Training and Prediction

- Train the chosen model on the training set.
- Predict on the test dataset.

4. Evaluation using Confusion Matrix

- Generate the confusion matrix.
 - Calculate:
 - Accuracy
 - Precision
 - Recall
 - F1-Score
-

Conclusion

- The chosen machine learning model successfully predicted customer responses with reasonable accuracy.
- Evaluation metrics help in assessing model performance and identifying potential improvements.
- Further enhancement can be done using advanced feature engineering and hyperparameter tuning.