# Assignment - 4

**Name:** Yashashree Nimbalkar
**Roll No:** 281040
**Batch:** A2
**PRN:** 22311423

---

## Statement

In this assignment, we aim to:

a) Apply an appropriate machine learning (ML) algorithm to a dataset collected from a cosmetics shop.
 b) Predict customer responses to a special offer using ML techniques.
 c) Evaluate the model using a confusion matrix and calculate:

- Accuracy

- Precision

- Recall

- F1-score

---

## Objective

1. Implement a classification model to analyze customer responses.

2. Use data preprocessing techniques to prepare the dataset for ML.

3. Evaluate model performance using key classification metrics.

---

## Resources Used

- **Software:** VS Code

- **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn

---

## Introduction to Machine Learning in Classification

Machine learning enables computers to learn patterns from data and make predictions. In this assignment, we implemented a classification model to predict customer responses based on their purchase history and demographics.

**Key Libraries Used:**

1. **Pandas & NumPy:** Data manipulation and preprocessing.

2. **Matplotlib & Seaborn:** Data visualization for insights.

3. **Scikit-learn:** ML model training, evaluation, and metric calculation.

---

## Methodology

**1. Data Collection and Preprocessing**

- **Dataset Used:** Customer Data from a Cosmetics Shop

- **Initial Steps:**

    - Loaded the dataset using Pandas.

    - Dropped unnecessary columns like Serial No.

    - Applied Label Encoding to categorical variables.

    - Standardized numerical features using StandardScaler.

**2. Splitting Data for Training and Testing**

- **Defined Features (X) and Target Variable (y):**

- ○ Features included customer details such as age, income, and purchase history.

  - ○ Target variable was whether the customer responded positively (1) or negatively (0) to the special offer.

- **Split Data:**

  - ○ 80% Training Set, 20% Testing Set using train_test_split().

## 3. Model Selection and Training

- **Algorithm Used:** Random Forest Classifier

  - ○ Chosen for its robustness and ability to handle non-linearity.

  - ○ Trained the model using the fit() method.

## 4. Model Evaluation using Confusion Matrix

- Generated a confusion matrix to compare predicted vs. actual values.

- Calculated key metrics:

  - ○ **Accuracy Score:** Measures overall correctness.

  - ○ **Precision Score:** Measures true positive rate.

  - ○ **Recall Score:** Measures sensitivity to actual positive cases.

  - ○ **F1 Score:** Harmonic mean of precision and recall for balanced evaluation.

---

# Advantages of Machine Learning in Classification

1. Enables automation of decision-making processes.

2. Provides data-driven insights for better marketing strategies.

3. Can be trained on large datasets for improved accuracy.

## Disadvantages

1. Requires extensive preprocessing for optimal performance.

2. Model performance depends on data quality and feature selection.

---

## Conclusion

This assignment focused on implementing a classification model using machine learning techniques. We trained a Random Forest Classifier on a dataset of customer responses to predict their likelihood of accepting a special offer. By evaluating the model using a confusion matrix and classification metrics, we assessed its effectiveness in predicting customer behavior. These skills are essential for real-world applications in marketing, sales, and customer relationship management.