**Name:** Janhvi Katakdhond

**Roll no:** 282067

**Batch:** B3

**ASSIGNMENT 4**

**Problem Statement:**

Apply appropriate ML algorithm on a dataset . Create confusion matrix based on above data and find :

a) Accuracy

b) Precision

c) Recall

d) F-1 score.

**Objective:**

The objective of this assignment is to apply a suitable machine learning classification algorithm on a selected dataset and evaluate its performance. The model's effectiveness will be assessed using a confusion matrix, from which key evaluation metrics such as Accuracy, Precision, Recall, and F1-score will be calculated. This process aims to understand the strengths and limitations of the model in correctly predicting the target variable and to demonstrate the practical application of classification techniques in machine learning.

Here I used admission dataset, to predict if a student will get admitted or rejected.

**S/W Packages and H/W apparatus used:**

Software used:

1. Python 3.x

2. Google Colab

Libraries and packages used: NumPy, Pandas

**Theory:**

One of the most common ML tasks is classification, where the goal is to assign input data into predefined categories or classes. Examples include predicting whether an email is spam or not, or whether a patient has a disease.

To evaluate the performance of a classification model, it is crucial to use appropriate metrics that reflect how well the model is predicting the outcomes. These metrics are typically derived from the confusion matrix.

**Confusion Matrix**

A confusion matrix is a table that summarizes the performance of a classification model by comparing the actual values with the predicted values. It includes four components:

* True Positives (TP): Correctly predicted positive class.
* True Negatives (TN): Correctly predicted negative class.
* False Positives (FP): Incorrectly predicted as positive (Type I error).
* False Negatives (FN): Incorrectly predicted as negative (Type II error).

**Evaluation Matrix:**

1. **Accuracy:**Measures the overall correctness of the model.

**Accuracy = TP + TN / FP + FN + TP + TN**

1. **Precision**  
   Indicates how many of the positively predicted cases were actually correct.

**Precision = TP /TP + FP**

1. **Recall**

Measures how many actual positive cases were correctly predicted.

**Recall = TP / TP + FN**

1. **F1 Score**  
   The harmonic mean of Precision and Recall. It balances the trade-off between the two.

**F1 Score = 2 \* Precision \* Recall / Precision + Recall**

These metrics are essential, especially in imbalanced datasets, where relying on accuracy alone can be misleading.

**Machine Learning Algorithm**

Depending on the nature of the dataset (binary or multiclass classification), various algorithms can be used, such as:

* **Logistic Regression**
* **Decision Trees**
* **Random Forest**
* **Support Vector Machines (SVM)**
* **K-Nearest Neighbors (KNN)**

Each of these has its own advantages depending on the complexity and size of the dataset, the number of features, and the interpretability requirements.

**Conclusion:**

In this assignment, Logistic Regression was employed on the *admission.csv* dataset to predict the likelihood of a student being admitted based on their GRE and TOEFL scores. As a binary classification problem, the model aimed to categorize students into two classes: *Admitted* and *Not Admitted*. The effectiveness of the model was assessed using a confusion matrix, from which key evaluation metrics—Accuracy, Precision, Recall, and F1-score—were derived.

These metrics provided insights into the model’s predictive power and its ability to correctly classify both admitted and non-admitted students. The results showed how academic indicators like GRE and TOEFL scores influence admission outcomes and how machine learning techniques like Logistic Regression can support data-driven decision-making in educational settings. Overall, this assignment highlighted the real-world application of classification algorithms and the importance of evaluating model performance through multiple metrics.