Confusion Matrix:-

A Confusion Matrix helps us understand how well a classification model is performing by comparing its predictions with the actual results.

Actual \ Predicted	**Positive (Yes)**	**Negative (No)**
Positive (Yes)	True Positive (TP)	False Negative (FN)
Negative (No)	False Positive (FP)	True Negative (TN)

Explanation of Terms :-

1. **True Positive (TP) →** The model correctly predicted positive (e.g., predicted "pass" and the student actually passed).
2. **False Positive (FP) [Type I Error] →** The model incorrectly predicted positive (e.g., predicted "pass" but the student actually failed).
3. **False Negative (FN) [Type II Error] →** The model incorrectly predicted negative (e.g., predicted "fail" but the student actually passed).
4. **True Negative (TN) →** The model correctly predicted negative (e.g., predicted "fail" and the student actually failed).

Example: Predicting Whether a Student Will Pass an Exam

Actual/predicted	**Pass (Positive)**	**Fail (Negative)**
Pass (Actual)	TruePositive=50	FlaseNegative=10
Fail (Actual)	FalsePositive=5	TrueNegative=35

→ **True Positive (TP = 50):** 50 students correctly predicted as passing**.**

→**False Positive (FP = 5):** 5 students wrongly predicted as passing but actually Failed.

→**False Negative (FN = 10):** 10 students wrongly predicted as failing but actually passed.

→**True Negative (TN = 35):** 35 students correctly predicted as failing.

Accuracy :-

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

Accuracy=50+35/50+35+5+10

=85/100

=85%

Precision:-

Precision=TP/FN+TP Precision

=50/50+10=50/60=90.9%

Columns:

Columns are the **structural components** of a dataset (typically in tabular format). Each column represents a **specific type of information** and stores **values of the same attribute** for different rows (records).

Employee_ID	Name	Age	Department	Salary	Joining_Date
101	Alice	28	HR	50,000	2020-06-15
102	Bob	35	IT	80,000	2018-09-01
103	Charlie	30	Finance	70,000	2019-11-20

- **Columns** = Employee_ID, Name, Age, Department, Salary, Joining_Date
- **Rows** = Individual records (employees)

Key characteristics of columns:

- **Hold raw data** and maintain a structured format.
- Each column represents a single attribute (e.g., Age, Department).
- Different data types:
    - Age → Numeric

- Name → Text
- Joining_Date → Date

Features :-

Features are **input variables** used to train a machine learning model. They represent **meaningful** or **relevant** information extracted from the dataset, often preprocessed for better accuracy.

Employee_ID	Name	Age	Department	Salary	Joining_Date	Attrition
101	Alice	28	HR	50,000	2020-06-15	No
102	Bob	35	IT	80,000	2018-09-01	Yes
103	Charlie	30	Finance	70,000	2019-11-20	No

Relevant Features:

- Age
- Salary
- Department (converted to numbers via encoding)
- **New Feature**: Years_at_Company (calculated from Joining_Date)

Non-relevant Features:

- Employee_ID (Just an identifier, does not impact attrition)
- Name (Personal data, not useful for prediction)

Difference between Columns and features:-

Aspect	Columns (Data Table)	Features (Machine Learning)
Purpose	Stores raw data	Used for predictions
Scope	Any dataset	Specific to ML models
Example	`Customer_ID`, `Name`, `Purchase_Amount`	`Purchase_Amount`, `Customer_Lifetime_Value`
Processing	Stored as-is	Transformed (e.g., scaled, encoded)