**Understanding Confusion Matrix**

Example to explain confusion matrix, Diabetes records

Path to the records: **C:\\Users\\rs\\Downloads\\01 Data Science Lab Copy\\02 Lab Data\\Python\\pima-indians-diabetes.txt**

Total no. of records = 308

**From the above 308 records we create train and test sets**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.4, random\_state=42)**

when **test\_size = 0.4** then automatically **train\_size = 0.6** and **random\_state = 42** (any number) to fix random pick of records.



**Basic Terms:**

**True Negative (TN):** Machine predicted as No and actual data is also No.

**True Positive (TP):** Machine predicted as Yes and actual data is also Yes.

**False Negative (FN):** Machine predicted as No when actual data is Yes.

**False Positive (FP):** Machine predicted as Yes and actual data is also Yes.

**In this example:**

total = 308

TN = 174

TP = 66

FN = 36

FP = 32

**List of rates computed from confusion matrix in classification report:**

**Accuracy:** Calculates accuracy of machine i.e., how often machine predictions are correct.

Accuracy = (TP + TN) / total

= (66 + 174) / 308 = 0.7792

**Error Rate (or) Miscalculation Rate:** how often machine predictions are wrong.

Error Rate = (FP + FN) / total

= (32 + 36) / 308 = 0.2207

**Recall (or) Sensitivity (or) True Positive Rate:** how often machine predicts yes when actual data is also yes.

Recall = TP / Actual Yes

= 66 / 102 = 0.6470

**False Positive Rate:** how often machine predicts yes when actual data is no.

False Positive Rate = FP / Actual No

= 32 / 206 = 0.1456

**Specificity:** how often machine predicts no when actual data is also no.

Specificity = TN / Actual No

= 174 / 206 = 0.8446

**Precision:** when machine predicts yes, how often it is correct.

Precision = TP / Predicted Yes

= 66 / 98 = 0.6734

**Prevalence:** how often yes condition actually occurs in our sample.

Prevalence = Actual Yes / Total

= 108 / 308 = 0.3311

**F1 Score (or) F Score (or) F-Measure:** Weighted Average of the Recall and Precision.

F1 Score = 2 .

= 2. = 2. = 0.6598