32. Confusion Matrix

- Confusion Matrix is to model the difference b/w the output data generated from testing of a model and the output of the original data. i.e. matrix b/w predicted output and original output.
- Model with 90%, 95% or even 100% can give wrong predictions
- The problem with wrong predictions can be traced through **confusion matrix**
- Confusion matrix gives better analysis of the built model

- A confusion matrix is a simple and useful tool for understanding the performance of a classification model, like one used in machine learning or statistics.
- It helps you evaluate how well your model is doing in categorizing things correctly.
- It is also know as the **error matrix / evaluation matrix**.
- The matrix consists of predictions result in a summarized from, which has number of correct predictions and incorrect predictions.

 No description has been provided for this image

Interpretation of graphs

- TN = True Negative = Actual: 0, Predicted: 0 -> **True Negative**
- FN = False Negative = Actual: 1, Predicted: 0 -> **False Negative**
- FP = False Positive = Actual: 0, Predicted: 1 -> **False Positive**
- TP = True Positive = Actual: 1, Predicted: 1 -> **True Positive**

$$\text{Model Accuracy} = \frac{TN + TP}{TN + TP + FN + FP}$$

$$\text{Model Error} = \frac{FN + FP}{TN + TP + FN + FP}$$

False Negative: The model has predicted no (0), but the actual value was yes (1), it is also called as **Type-II error**

False Positive: The model has predicted yes (1), but the actual value was no (0), it is also called as **Type-I error**

- **False Negative is more dangerous**, depends on the situation

- if false negative is upto 5%, then reject this model

In [ ]:

32.1 Confusion Matrix (Sensitivity, Precision, Recall, F1-score)

Precision

Precision: It helps us to measure the ability to classify positive samples in the model.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- To increase the recall, False Positive value should be lower.

Recall

Recall: It helps us to measure how many positive samples were correctly classified by the ML model.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- To increase the recall, False Negative value should be lower.

F1-Score

- when we donot have information because of lack of knowledge in domain to whether improve Precsion or/and recall, then we will use **F1 Score**.

- It is the harmonic mean of precision and recall. It takes false positive and false negative into account.
- Therefore, it performs well on an imbalanced dataset.

$$\text{F1 Score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

- Should increase the value of F1-Score

In Confusion matix,

- Precsion should should be high
- Recall should be high
- F-Score should be high

In [ ]: