MODEL EVALUATION

• INTRODUCTION TO MODEL EVALUATION

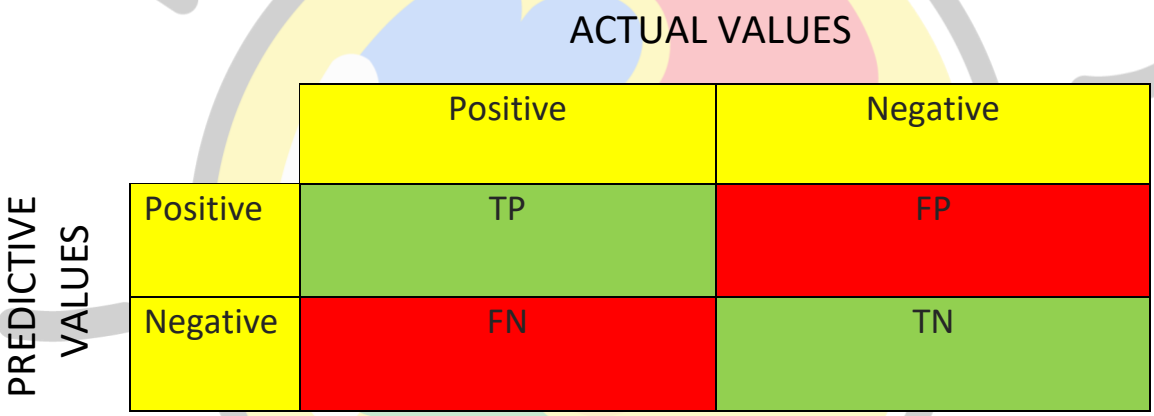
Model selection is the process of selecting one final machine learning model from among a collection of candidate machine learning models for a training dataset.
Model selection is a process that can be applied both across different types of models (e.g. logistic regression, SVM, KNN, etc.) and across models of the same type configured with different model hyperparameters (e.g. different kernels in an SVM).

For example, we may have a dataset for which we are interested in developing a classification or regression predictive model. We do not know beforehand as to which model will perform best on this problem, as it is unknowable. Therefore, we fit and evaluate a suite of different models on the problem.

• CONFUSION MATRIX

A Confusion matrix is an N x N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model. This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making.

For a binary classification problem, we would have a 2 x 2 matrix as shown below with 4 values:



Let's decode the matrix:

- The target variable has two values: Positive or Negative
- The columns represent the actual values of the target variable
- The rows represent the predicted values of the target variable

What are TP, FP, PN, FN then?

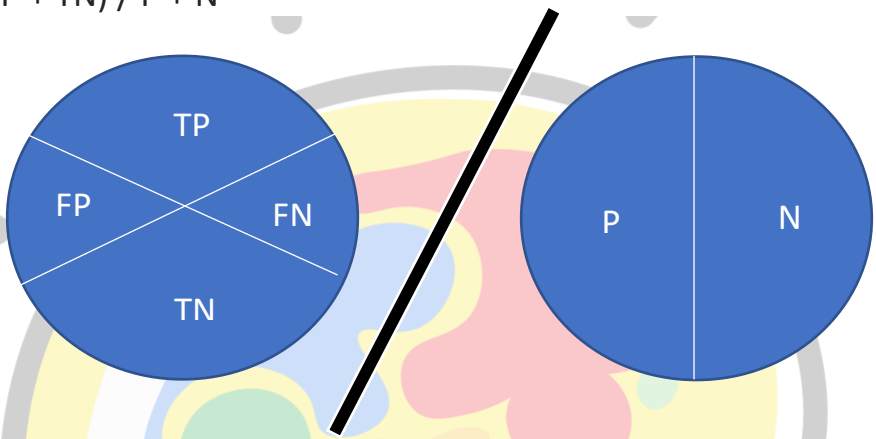
- TP – TRUE POSITIVE – The predicted positive is actually positive.
- FP – FALSE POSITIVE – The predicted positive is actually false.
- FN – FALSE NEGATIVE – The predicted negative is actually false.
- TN – TRUE NEGATIVE – The predicted negative is actually true.

For example – Someone tests for Covid-19 and he gets his report as Negative but he is actually positive, hence it is a false negative.

• ACCURACY

Accuracy (ACC) is calculated as the number of all correct predictions divided by the total number of the dataset. The best accuracy is 1.0, whereas the worst is 0.0. It can also be calculated by 1 – ERR.

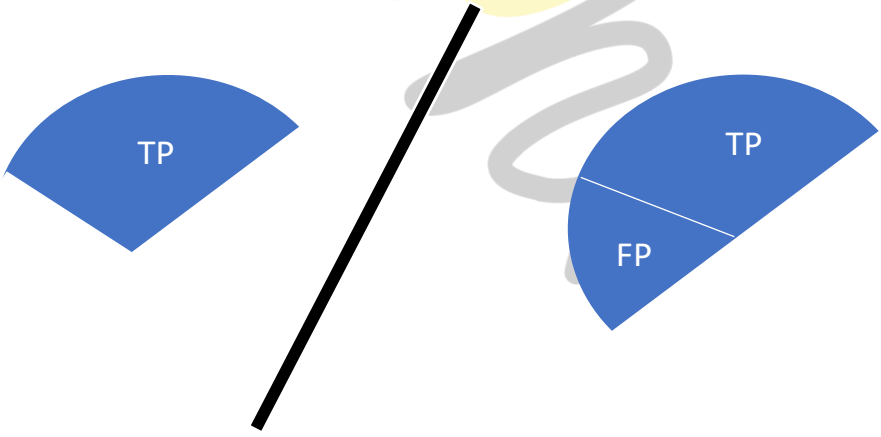
Formula – (TP + TN) / P + N



• PRECISION (POSITIVE PREDICTIVE VALUE)

Precision (PREC) is calculated as the number of correct positive predictions divided by the total number of positive predictions. It is also called positive predictive value (PPV). The best precision is 1.0, whereas the worst is 0.0.

Formula – TP / (TP + FP)



• SENSITIVITY/RECALL

Sensitivity (SN) is calculated as the number of correct positive predictions divided by the total number of positives. It is also called recall (REC) or true positive rate (TPR). The best sensitivity is 1.0, whereas the worst is 0.0.

• F1 SCORE

Having a precision or recall value as 0 is not desirable and hence it will give us the F1 score of 0 (lowest). On the other hand, if both the precision and recall value is 1, it'll give us the F1 score of 1 indicating perfect precision-recall values. All the other intermediate values of the F1 score ranges between 0 and 1.