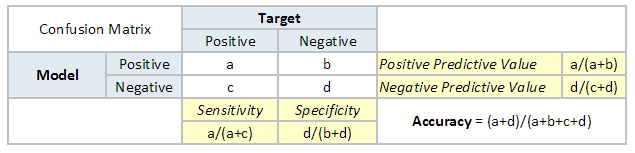
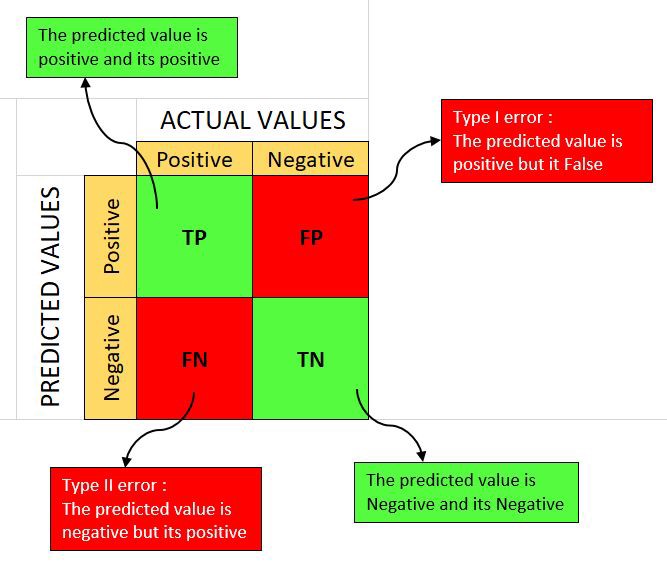
Evaluation Metrices

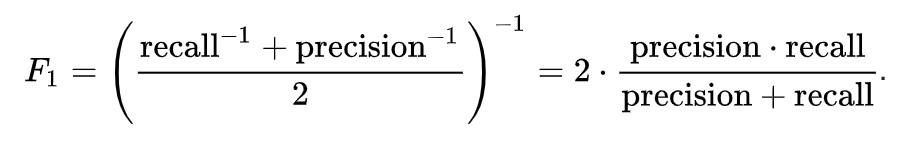
1. Confusion Matrix
2. F1 Score
3. Gain and Lift Charts
4. Kolmogorov Smirnov Chart
5. AUC – ROC
6. Log Loss
7. Gini Coefficient
8. Concordant – Discordant Ratio
9. Root Mean Squared Error

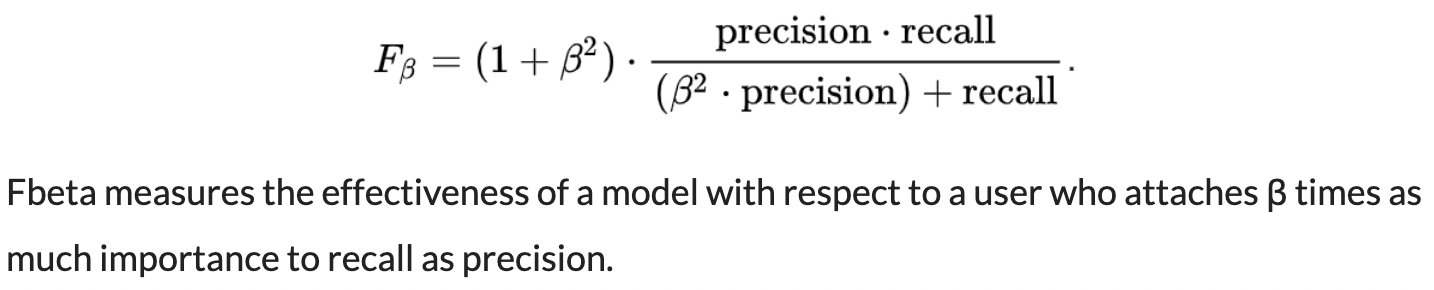
A confusion matrix is an N X N matrix, where N is the number of classes being predicted. It is a table that is used to describe the performance of a classification model (or "classifier" ) on a set of test data for which the true values are known





F1-Score is the harmonic mean of precision and recall values for a classification problem. We use HM because HM punishes extreme values more. Formula for F1-Score is as follows:





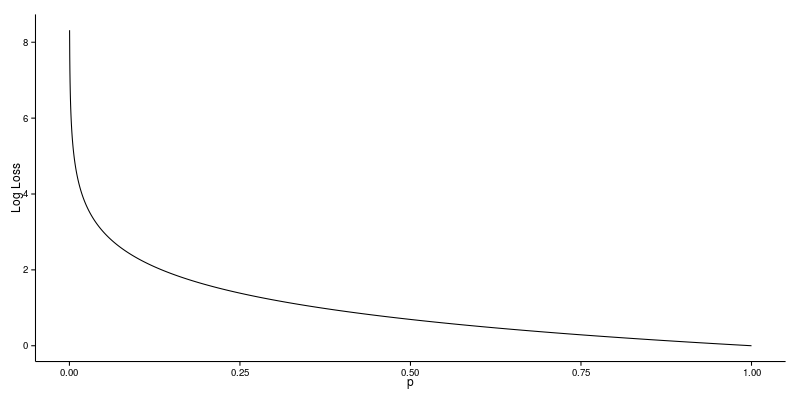
Gain and Lift chart are mainly concerned to check the rank ordering of the probabilities. Gain chart and lift chart are two measures that are used for measuring benefits of using the model and are used in business contexts such as target marketing, risk modelling, supply chain analytics, etc. They are two approaches used while solving classification problems with imbalanced data sets. It is the measures in logistic regression that help organizations to understand the benefits of using that model.

K-S or Kolmogorov-Smirnov chart measures performance of classification models. More accurately, K-S is a measure of the degree of separation between the positive and negative distributions. The K-S is 100, if the scores partition the population into two separate groups in which one group contains all the positives and the other all the negatives. On the other hand, If the model cannot differentiate between positives and negatives, then it is as if the model selects cases randomly from the population. The K-S would be 0. In most classification models the K-S will fall between 0 and 100, and that the higher the value the better the model is at separating the positive from negative cases.

Area Under the ROC curve (AUC – ROC) : This is again one of the popular metrics used in the industry. Biggest advantage of using ROC (receiver operating characteristic curve) curve is that it is independent of the change in proportion of responders. It is a performance measurement for the classification problems at various threshold settings. ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1. ROC curve is plotted with TPR against FPR where TPR is on y-axis and FPR is on the x-axis. true positive rate (TPR) against the false positive rate (FPR)



Log Loss : AUC ROC considers predicted probabilities for determining our model’s performance. Issue with AUC ROC is it only takes into account order of probabilities and hence it doesn’t take into account model’s capability to predict higher probability for samples more likely to be positive. In that case, we could use log loss which is nothing but negative average of the log of corrected predicted probabilities for each instance. Log Loss ramps up very rapidly as the predicted probability approaches 0. So, lower the log loss, better the model. However, there is no absolute measure on a good log loss and it is use-case/application dependent. Logarithmic loss indicates how close a prediction probability comes to the actual/corresponding true value. It is also called Binary Cross entropy



Gini coefficient is sometimes used in classification problems. It can be straight away derived from the AUC ROC number. Gini is nothing but ratio between area between the ROC curve and the diagonal line & the area of the above triangle. Following is the formulae used :

Gini = 2\*AUC – 1

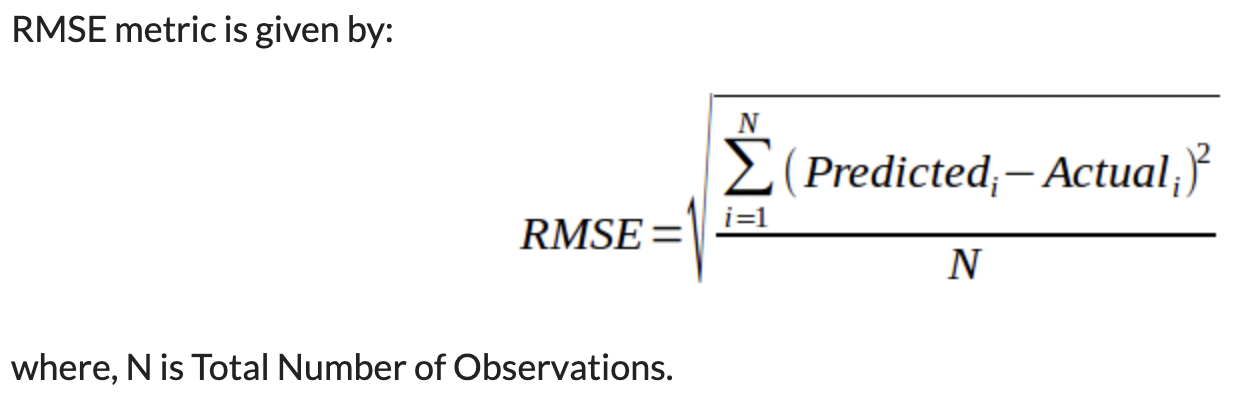
Gini above 60% is a good model. It is a measure of statistical dispersion intended to represent the income inequality or the wealth inequality within a nation or a social group.

Concordant – Discordant ratio : This is again one of the most important metric for any classification predictions problem. Concordant ratio of more than 60% is considered to be a good model. It is primarily used to access the model’s predictive power. In Concordance ratio, total number of Concordant pairs are counted & divided by the total number of pairs.

RMSE is the most popular evaluation metric used in regression problems. It follows an assumption that error are unbiased and follow a normal distribution. Here are the key points to consider on RMSE:

The ‘squared’ nature of this metric helps to deliver more robust results. It avoids the use of absolute error values which is highly undesirable in mathematical calculations.

It is highly affected by outlier values. Hence, make sure you’ve removed outliers from your data set prior to using this metric.



In case of Root mean squared logarithmic error, we take the log of the predictions and actual values. So basically, what changes are the variance that we are measuring. RMSLE is usually used when we don’t want to penalize huge differences in the predicted and the actual values, when both predicted and true values are huge numbers.

