How do I evaluate my classification model with an imbalanced dataset?

Accuracy alone cannot be the right metric for classification problems.
Other metrics that can be checked are Confusion Matric, Precision, Recall, Roc /Auc curve. If you choose the wrong metric to evaluate your models, you are likely to choose a poor model, or in the worst case, be misled about the expected performance of your model.

$$\frac{TP+TN}{TP+TN+FP+FN}$$

Accuracy

Accuracy is one metric which gives the fraction of predictions our model got right.
Accuracy = Number of correct predictions / Total number of predictions.

Confusion matrix

		Predicted No	Predicted Yes
True	No	1320 TN	36 FP
True	Yes	34 FN	110 TF

It is a summary table showing how good our model is at predicting various classes.
True Positive(**TP**):- When you predict an observation belongs to a class and it actually does belong to that class. eg. Model predicts Customer would buy a loan and he does. (110)
True Negative(**TN**):-When you predict an observation does not belong to a class and it actually does not belong to that class. eg. Model predicts Customer would not buy a loan and he does not. (1320)
False Positive(**FP**):-When you predict an observation belongs to a class and it actually does not belong to that class. eg. Model predicts Customer would buy a loan and he does not. (36)
False Negative(**FN**):- When you predict an observation does not belong to a class and it actually does belong to that class. eg. Model predicts Customer would not buy a loan and he does(34)

$$\frac{TP}{TP+FN}$$

Recall/Sensitivity

Sensitivity is also known as the True Positive rate
Sensitivity = No. of True Positives / (No. of True Positives + No. of False Negatives).
Higher recall lesser chances of false negatives.
For eg.In the Loan example above. The potential customer is missed by the sales team if the false negative is higher. This is a loss of opportunity.

$$\frac{TP}{TP+FP}$$

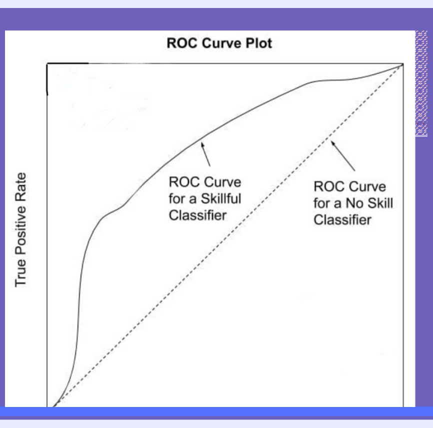
Precision

Precision — Also called Positive predictive value
TPrecision=No. of True Positives/(No. of True Positive + No. of False positives)
Higher precision, lesser chances of false positives.
In Loan example . The sales team would be unnecessary calling customers to pitch their loan products if the false positive is higher. This is a loss of resources.

$$\frac{2X\ Precision\ X\ Recall}{Precision+\ Recall}$$

F1 score

The F1 score takes into account both Precision and Recall.F1 score is a weighted average of precision and recall. For eg in customer loan, if the bank decides they don't want to spend many resources and also not lose customers F1 score would give a balance between recall and precision



ROC-AUC Curve

ROC(Receiver Operator Characteristic) curve plots the performance of a binary classifier under various threshold settings; this is measured by true positive rate and false-positive rate. The AUC (Area Under Curve) is the area enclosed by the ROC curve. It provides a single score to summarize the plot that can be used to compare models.

Created by www.linkedin.com/in/yogitadarade