Initial Performance Metric Exploration

1 Different Methods: Benefits and Limitations

Accuracy: Can be misleading for imbalanced datasets (e.g., predicting the majority class can give high accuracy even if the model ignores the minority class).

Precision: Suitable when the cost of false positives is high. Not a priority for our scenario as not too big of a cost if we count someone who stayed as exited.

Recall: Suitable when the cost of false negatives is high. Might be important for us as bigger cost when we believe that someone who has exitted has been counted as retained.

F1-Score: Suitable when there is a trade-off between precision and recall, especially in imbalanced datasets. F1-score is helpful when you care equally about false positives and false negatives.

ROC Curve: Plots the True Positive Rate (Recall) against the False Positive Rate (FPR = FP / (FP + TN)) for different classification thresholds. AUC-ROC: Measures the area under the ROC curve. Can be overly optimistic for highly imbalanced datasets where one class dominates.

AUC-PR: Measures the area under the precision-recall curve. Use case: More informative than AUC-ROC for imbalanced datasets, where the focus is on the minority class. The AUC-PR curve emphasizes the model's performance with respect to the positive (minority) class. Limitations: Does not capture performance on negative (majority) class.

Logarithmic Loss: Limitations: It requires probabilistic outputs and is more complex to interpret than other metrics.

Matthews Correlation Coefficient (MCC) Use case: MCC gives a balanced measure even when the classes are imbalanced. It takes into account all four confusion matrix elements (TP, TN, FP, FN). Limitations: More difficult to interpret compared to precision, recall, or accuracy.

Balanced Matthews Correlation Coefficient: MCC is known to to mitigate the imbalance of a test set, however it still remains dependent on the prevalence of the predicted categories, this can lead to MCC being underestimated at extremely high or low positive prevalence. Balanced MCC[**?**] is an extension of balanced accuracy which is a performance metric that is calibrated to a test set with a positive prevalence of 50%. This means we can have an MCC metric independent of prevalence. This can help us see if our model is performing well on predicting those that stay with the bank but also on our minority class of those who churn. We are able to compute this from the confusion matrix.

Balanced Accuracy: Useful when the dataset is imbalanced, as it gives equal weight to the performance on both the positive and negative classes.