***First***

We should check the effect after FE. The following is the performance of each prior model that we simply use label encoder and standard scaler to process the original data. Among them, the best algorithms are marked as yellow: ACC, AUC, and Precision of the Cat Boost model, while F1, Kappa, and MCC of Light GBM.

***Given the excellent MCC performance of Light GBM and its efficient training speed***

We choose Light GBM to show the results of FE. Comparing them after 5-folds validation, the accuracy of the model after FE has been reduced and MCC has been extremely improved.

***Look at our confusion matrix and precision error.***

Here our confusion matrix is ​​reversed. Due to the large gap in the binary labels in the original data, over-sampling yes-label and under-sampling no-label at the same time to ensure the effectiveness of the training for the two labels. This makes our training results more balanced.

***The following is a cluster analysis of Boundary, Dimension & Manifold.***

After FE, the main feature dropped from 5 to 2, and the manifold & boundary between the two categories became more obvious, which also confirmed the success of our FE.

***The ROC curve is a graph showing the performance at all classification thresholds.***

This curve plots two parameters: **Precision & Recall**. And AUC stands for Area under the ROC Curve, which measures the entire two-dimensional area underneath the entire ROC curve.

Observing the ROC curve, the AUC of class 1 increases from 0.93 to 0.96 after doing FE.

***Threshold Plot.***

When analyze the ROC curve, the optimal threshold of the classifier can be directly calculated. To be more precise, it is possible to use a grid search to tune the threshold and locate the optimal value. The result clearly shows the optimal threshold is 0.38 for imbalanced classification, that means it become a yes if your prediction bigger than the threshold. After FE balanced the number of labels, threshold increases to 0.46.

***Evaluation***

This is a comparison of our prediction results before and after FE. Although MCC has decreased, but consider the real business scenario, as a marketing analyst for a bank, we want to identify users who may subscribe to a term deposit but have not yet subscribed. Such classes will belong to **False Positives** users who are expected to make a subscription but do not make it. Therefore, in addition to TP and TN, we also prefer FP and work on reducing them.

So, the **Recall** increased, which provides the percentage of correctly predicted target customers increased.

The **Specificity** gives us the percentage of non-target customers that are correctly predicted.

The **Precision** increased, it provides us with the increase of efficiency of prediction because we can get as much TP with as few resources as possible.

Thus our work performance and efficiency have been obtained a great improvement.