1) **Which 3 methods of classification are chosen and why?**

- Logistic Regression
- Decision Tree
- Random Forest

These are the three classification models used,

Logistic regression is easier to implement and interpret, and very efficient to train. It is very fast at classifying unknown records. It performs well when the dataset is linearly separable. It can interpret model coefficients as indicators of feature importance.

Decision Tree is its ability to use different feature subsets and decision rules at different stages of classification. It forces the consideration of all possible outcomes of a decision and traces each path to a conclusion.

Random Forest provides the highest accuracy. The random forest technique can also handle big data with numerous variables running into thousands. It can automatically balance data sets when a class is more infrequent than other classes in the data.

2) **Which method among the 3 chosen is working better than the other 2?**

Logistic Regression model which was considered the baseline model looked better than the other 2 models.

3) **How is feature engineering changing the results of the evaluation metrics?**

Feature engineering refers to the process of designing artificial features into an algorithm. These artificial features are then used by that algorithm in order to improve its performance, or in other words, reap better results. Data scientists spend most of their time with data, and it becomes important to make models accurate. When feature engineering activities are done correctly, the resulting dataset is optimal and contains all of the important factors that affect the business problem. As a result of these datasets, the most accurate predictive models and the most useful insights are produced.