**APPROACH**

* For solving this problem first I have uploaded all the test set and train set and check for value count of dependent variable which is Is\_Lead. When I plot the bar chart for dependent variable I came to know that this is the problem of imbalance dataset. Since I am in my last semester I have been taught how to deal with Imbalance datasets.
* Then I started exploring different variables and find out that average account balance have outliers and to handle outliers I have used capped using IQR and removing outliers since both were giving me same roc-auc score hence I prefer to remove outliers.
* After that I found out about missing value in credit product and I put the most frequent category in missing place.
* After that I tried to find any correlated variables which can cause dimensionality issue but since the correlation was not too high hence I ignored the correlation problem.
* After I have split the dataset into train and test split and scales the dataset because average account balance variable has very huge value which can dominate other features.
* Then First for categorical variable I have used get dummy method but then I found better accuracy by using one hot encoding as it is much better method as compared to get dummy which given k variables while one hot encoding give k-1 variables.
* Then I use SMOTE technique which is oversampling technique to balance the dataset I was getting around 75 % roc-auc score. However when I tried other techniques such as under sampling and SMOTE-Kmean I came to know that kmean smote was giving better score hence I go with Kmean Smote.
* After that I tried Discretization but it didn’t improve the accuracy score so I remove that part
* Then I tried models like random forest,svm and logistic regression but I wasn’t able to get better score so I tried Xgboost,Catboost and LGBM classifier which I have used in my last semester project so I tried and I get better accuracy as these classifier are very powerful and Catboost even has over fitting stopper which can help to not overfit the data.
* I also tried feature selection technique like Recursive Feature Engineering but after reducing the feature my score got reduced as there are less data columns.

It was very nice experience to participate in this competition as a fresher it was opportunity for me to get more practice and learn more .Thank You