Approach:

It was an imbalanced dataset with categorical variables except for one (‘Balance’). We converted Age to bin column using feature engineering therefore it also converted to ordinal categorical variable.

We needed to choose the algorithm which works well with categorical variable therefore logistic regression was not working well with this data.

I tried decision tree, random forest and XGboost with different parameters(using hyperparameter tuning). We needed to choose the probability cut off to maximize ‘macro f1’ score .

XGBoost performed best among them with certain parameters which I got using hyperparameter tuning.