**Problem statement: -**

Happy Customer Bank is a mid-sized private bank that deals in all kinds of banking products, like Savings accounts, Current accounts, investment products, credit products, among other offerings.

The bank also cross-sells products to its existing customers and to do so they use different kinds of communication like tele-calling, e-mails, recommendations on net banking, mobile banking, etc.

In this case, the Happy Customer Bank wants to cross sell its credit cards to its existing customers. The bank has identified a set of customers that are eligible for taking these credit cards.

Now, the bank is looking for your help in identifying customers that could show higher intent towards a recommended credit card, given:

* Customer details (gender, age, region etc.)
* Details of his/her relationship with the bank (Channel\_Code,Vintage, 'Avg\_Asset\_Value etc.)

**Approach: -**

I have applied decision tree and XG Boost algorithm and after comparing them finally selected XG boost algorithm to predict the probabilities for the class labels present in the target variable.

**XG Boost**

The XGBoost algorithm is effective for a wide range of regression and classification predictive modelling problems.

It is an efficient implementation of the stochastic gradient boosting algorithm and offers a range of hyperparameters that give fine-grained control over the model training procedure. Although the algorithm performs well in general, even on imbalanced classification datasets, it offers a way to tune the training algorithm to pay more attention to misclassification of the minority class for datasets with a skewed class distribution.

**Data pre-processing**

Data pre-processing is a data mining technique which is used to transform the raw data in a useful and efficient format. The pre-processing steps performed are: -

1. Label encoding: - converted all the labelled variable into numeric form.
2. Missing value analysis: - there were missing values present in the ‘Credict\_Product’ variable which is a class variable. So. I have applied mode to Impute and fill the missing values.
3. Data balancing: - plotted the count plot of the target variable and observed that the data is unbalanced.
4. Correlation analysis: - checked how each input variable is related to every other variable and observed that there is no variable that is highly positively or highly negatively correlated.

**Performed EDA**

Performed univariate, bivariate and multivariate analysis.

**Model Building**

1. Fitted decision tree on the unbalanced dataset and calculate its ROC\_AUC\_SCORE: - 0.6095
2. Fitted XG Boost on the unbalanced dataset and calculated its ROC\_AUC\_SCORE: - 0.7851
3. Fitted decision tree on balanced dataset after performing SMOTE and calculated its ROC\_AUC\_SCORE: - 0.6066
4. Fitted XB Boost on the balanced dataset after performing SMOTE and calculated its ROC\_AUC\_SCORE: -0.7835

**Since the value of ROC\_AUC\_SCORE is highest for XG Boost model applied on the unbalanced dataset. So, I have made the final probabilities prediction of both the classes using this model.**

**Submitted by: -**

**Ashmeet Kaur**

**ashkmeet@gmail.com**