# Credit Card Lead Prediction

## Overview:

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 Code etc.)
- Details of his/her relationship with the bank (Channel Code, Avg Account Balance etc.)

## Data and Findings:

Total Records = 245725

| **Variable** | **Definition** |
|---|---|
| ID | Unique Identifier for a row |
| Gender | Gender of the Customer |
| Age | Age of the Customer (in Years) |
| Region Code | Code of the Region for the customers |
| Occupation | Occupation Type for the customer |
| Channel Code | Acquisition Channel Code for the Customer (Encoded) |
| Vintage | Vintage for the Customer (In Months) |
| Credit Product | If the Customer has any active credit product (Home loan, Personal loan, Credit Card etc.) |
| Avg Account Balance | Average Account Balance for the Customer in last 12 Months |
| Is Active | If the Customer is Active in last 3 Months |
| Is Lead (Target) | If the Customer is interested for the Credit Card<br>0: Customer is not interested<br>1: Customer is interested |

## Exploratory Data Analysis:

Performed EDA using pandas_profiling and found the following details-

- Categorical – 6
- Numerical – 3
- Binary – 2
- Missing data – 29325 in Credit_Product

Detailed EDA Report of various Histograms, Interactions and Correlations are present in 'Exploratory_Data_Analysis.html' file.

## Data Preprocessing:

- Combined the Train and Test dataset using 'source' column
- Replaced missing data in Credit_Product with a new category 'NA'
- Avg_Account_Balance feature is transformed to log-normal distribution as ideally bank balance/salary will follow log-normal
- Performed Label Encoding for 6 Categorical features - Gender, Region_Code, Occupation, Channel_Code, Credit_Product, Is_Active.
- Separated the Train and Test dataset from the combined dataset.
- Created input variables as X and target variable as y i.e. Is_Lead.
- Checked the feature importance of different features and dropped unnecessary columns such as ID, source and Is_Lead from X.
- Splitting of the Train dataset into 75% training and 25% validation set.

## Model Building and Evaluation:

For training and evaluation, 3 models have been used – LightGBM, XGBoost and CatBoost.

- Checked the ROC AUC Score for default models.
- Hyperparameter Optimization with Randomized Search CV for finding the best model parameters and using early fitting detector to avoid overfitting.
- Use of Cross-Validation with Stratified K-Fold as target variable is unbalanced.

Finally, the best model used for prediction on Test dataset is a blend (average) of predictions from LightGBM, XGBoost and CatBoost classifiers.

## Conclusion:

By using this analysis, we can predict whether the Credit Card Lead will be converted for a customer from future. So, that we can group those customers into segments and give them some surprise offers so that the lead conversion rate will improve, which in terms increase the profit of the Bank.