**Credit Card Lead Prediction**

**Approach Followed:**

1. **Understanding** the data.
2. **Check for missing values if any**: Credit\_Product variable has missing values and it was imputed with ‘mode’.
3. **Exploratory Data Analysis**

Interpretation: Three variables such as Vintage, Age, Occupation, Channel\_code were more correlated with the target variable.

1. **Check for Class Imbalance**
2. **Data preparation:** Create Dummy columns for categorical variables and concatenate them with numerical variables.
3. **Split the data into train and test** (80:20 )
4. Used **SMOTE** technique on train data to resample the minor class labels
5. **Standardization**
6. **Model Building:** 3 classifiers such as Logistic Regression, Random Forest Classifier, XGBoost Classifier were used to train the data. Out of these models, the XGBoost classifier achieved the best results on training data, i.e 86% with roc\_auc score=0.85. Recall value= 0.98 for identifying customers who are not Lead(for class=0) and Precision value =0.93 (for class=1).
7. **Threshold Modification on ROC\_AUC**: When the threshold=0.6 for classifying the labels, recall value(for class=0 ) improved to 0.99 and precision value(for class=1) improved to 0.98, area of the curve remains the same.
8. **Test Model on test data**: XGBoost model gave better test results 79% with recall 0.98 (class=0) than Random Forest Classifier.
9. **Final Model:** XGBoost Classifier, since this model performed well on train and test data. Hence. XGBoost classifier is used as the final model for predicting Credit-card Lead customers.
10. Predicting the given **test.csv file** using XGBoost classifier