Credit Card Lead Prediction

# Imported important libraries

# Loading files

# Exploratory Data Analysis

## Univariate Analysis

# Feature Engineering

# Feature selection

# Machine Learning Algorithm

## Classification

# Approach

* I imported all the necessary Libraries and loaded the data
* Given dataset consist of 245725 rows and 11 column
* I also discovered the statistical metrics in the data such as quartiles mean max etc

# Handling Missing value

* I found that Credit History has null values and treted them with mode function

# Handling Outliers

* Here we came to find that are no outliers if there I t will be handled by our model

# Categorical encoding

* For categorical encoding I did with one hot encoding where the rank do not matter to our categories

# Bivariate Analysis of categorical column

* In this we consider gender and we computed with target by this we came to know that
* Female=79% of women are not interested in credit card only 20% of women are interested in loan amount
* Male=73% of men’s are not interested in the credit card only 26% of men are interested in the credit card
* Male are showing more interest in credit card

# Normalization and Standardization

* Given data is normally distributed so we did not go with these

# Feature Selection

* We are having only 11 features all the features having their own weightage so we consider all the feature and we are seen that there is no correlation between the features

# Cross Validation Technique

* Here I used stratified K-fold technique
* The most basic example is that we want the same proportion of different classes in each fold. Most of the times it happens by just doing it randomly, but sometimes, in complex datasets, we have to enforce a correct distribution for each fold.
* So I used stratified validation

# Model Building

* Here I used 3 types of model
* Logistic regression
* XG boost
* Cat boost
* With the help of training data I built my logistic regression model and I got 77% of accuracy
* Because of classification problem I preferred XG-BOOST
* XG Boost is an implementation of gradient , decision trees designed for speed and performance
* **Cat Boost** is based on gradient boosted decision trees. During training, a set of decision trees is built consecutively. Each successive tree is built with reduced loss
* for XG boost I consider 9 fold and 20000 estimate I got an ROC\_AUC\_SCORE as 78.98%
* for cat boost I consider 7 fold and I got an ROC\_AUC\_SCORE as78%
* and I find the average of both the model
* I predicted on test data

# Performance Matrices (ROC\_AUC\_SCORE)

* By these models I got the 78.9% of score for our data
* By these we came to know that our 78% of customer are not interest in credit card
* More than 70% of customer are not active in previous 3 month
* Hence only 20-30% of the customer are shown there interest in credit card
* Probability is calculated for our Target column

### 

## A heartfelt gratitude to the Analytics Vidya team for the Oppurtunity.