Approach Document

Credit card lead prediction

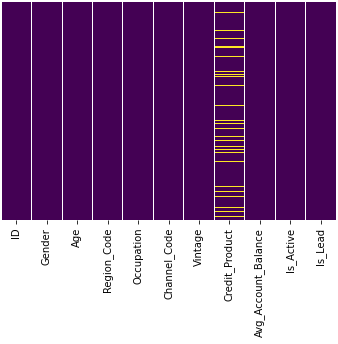
By: Madhavi Patel

Contents

1. EDA
2. Data Cleaning
3. Feature Engineering
4. Oversampling(class imbalance)
5. Modelling

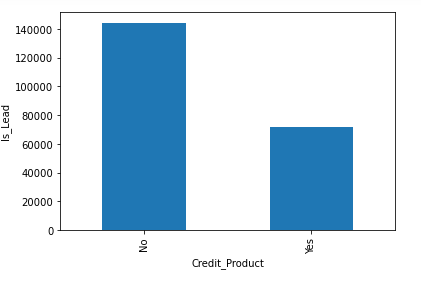
EDA, Data cleaning and Feature Engineering

In EDA first I have found whether there is a missing value in the data or not.



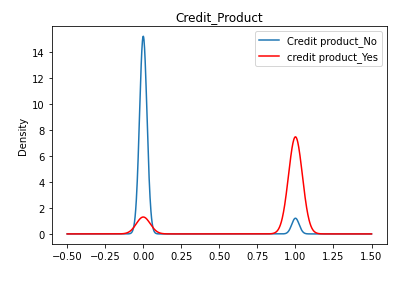
By this I have found out that there is a missing value in one variable that is Credit\_product

As, It is a categorical variable I have checked its frequent category



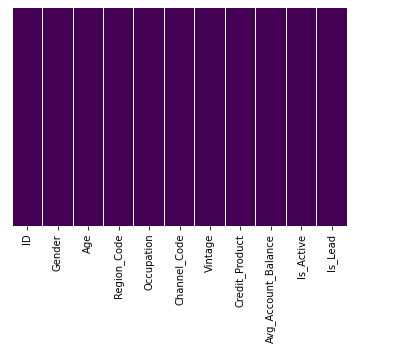
Which came ‘No’ so we can consider it as a mode for the variable

But before replacing the variable with its mode value I have checked the relation of both the categories with dependent variable,



And from the above plot its not much clear about the impact of the category. So I have kept it as a missing category.

Replaced null value with that.



Missing values have been replaced

Then I have checked for the outlier only one variable has minor outlier which doesn’t have much impact on dependent variable so I didn’t do anything on that.

Class Imbalance

When I checked for the class imbalance for the dependent variable, I found out there is a class

imbalance so I have done up sampling for that.

Model Building

I have applied logistic regression

In which I got 86% accuracy and then for the validation I have done auc roc score check