

Good and Bad Customers for Granting Credit

Problem Statement

The banks with the intent of credit card were more focused on the number of customers using their credit service but the drawback of them not being able to pay back the credit in time was an issue that soon followed, a system was in need to effectively decide the credit limit to be allowed to a person based on his previous credit history. You will learn how to apply Logistic Regression to credibility of the customer. Also learn how to evaluate Logistic Regression model using various parameter like on Accuracy, Sensitivity, Specificity and area under the ROC curve.

Build a classification model using logistic regression to predict the credibility of the customer, in order to minimize the risk and maximize the profit of a bank.

Data Description

- **Customer ID:** Unique identification of each customer
- **Credit_Amount:** Credit amount in dollars
- **Gender:** 1=Male, 2=Female
- **Academic_Qualification:** 1=Undergraduate, 2=Graduate, 3=Postgraduate, 4=Professional, 5=Others, 6=Unknown
- **Marital:** 1=Married, 2=Single, 3=Do not prefer to say
- **Age_Years:** Age in years
- **Repayment_Status_Jan:** Repayment status in Jan (0=Paid on time, 1=Payment delay for one month, 2=Payment delay for two months, ... 6=Payment delay for six months)
- **Repayment_Status_Feb:** Repayment status in Feb (Scale same as above)
- **Repayment_Status_March:** Repayment status in March (Scale same as above)
- **Repayment_Status_April:** Repayment status in April (Scale same as above)
- **Repayment_Status_May:** Repayment status in May (Scale same as above)
- **Repayment_Status_June:** Repayment status in June (Scale same as above)
- **Jan_Bill_Amount:** Amount of bill statement in Jan (In dollars)
- **Feb_Bill_Amount:** Amount of bill statement in Feb (In dollar)
- **March_Bill_Amount:** Amount of bill statement in March (In dollar)
- **April_Bill_Amount:** Amount of bill statement in April (In dollar)
- **May_Bill_Amount:** Amount of bill statement in May (In dollar)
- **June_Bill_Amount :** Amount of bill statement in June (In dollar)
- **Previous_Payment_Jan:** Amount of previous payment in Jan (In dollar)
- **Previous_Payment_Feb:** Amount of previous payment in Feb (In dollar)
- **Previous_Payment_March:** Amount of previous payment in March (In dollar)
- **Previous_Payment_April:** Amount of previous payment in April (In dollar)
- **Previous_Payment_May:** Amount of previous payment in May (In dollar)
- **Previous_Payment_June:** Amount of previous payment in June (In dollar)
- **Default_Payment:** Default payment of next month (1=yes, 0=no)

KPMG Data Science Prodegree

Logistic Regression: Problem Statement



Evaluation Parameters

Evaluation will be based on:

- Data Preparation
- Model Comparison
- Model Selection

Data Preparation

Analyze the data statistically and treat the multicollinear variables.

Model Comparison

Apply logistic regression algorithms for every change made in the datasets and compare results.

Model Selection

Select the best model. Model selection to be based on Accuracy, Sensitivity & Specificity and area under the ROC curve.

Expected Outcome

Higher accuracy in predicting the outcome using test data.