

Bank Credit Card Default Prediction



Problem Statement

The banks with the invent 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.

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 German Credit Bank.

Data Description

- Creditability: (1-good credit risk, 0-bad credit risk)
- Status_of_existing_account : Status of existing checking account

1: ... < 0 DM
2: 0 <= ... < 200 DM



3: ... \geq 200 DM / salary assignments for at least 1 year
4: no checking account

- Duration_of_Credit_month : Duration in month
- Payment_Status_of_Previous_Credit: Credit history of the customer

0: no credits taken/all credits paid back duly
1: all credits at this bank paid back duly
2: existing credits paid back duly till now
3: delay in paying off in the past
4: critical account/other credits existing (not at this bank)

- Purpose_of_loan: Purpose
 - 0: car (new)
 - 1: car (used)
 - 2: furniture/equipment
 - 3: radio/television
 - 4: domestic appliances
 - 5: repairs
 - 6: education
 - 7: (vacation - does not exist?)
 - 8: retraining
 - 9: business
 - 10: others

- Credit_Amount: Credit amount

- Value_of_Savings_account&bonds: Savings account/bonds
 - 1: ... < 100 DM
 - 2: $100 \leq \dots < 500$ DM
 - 3: $500 \leq \dots < 1000$ DM
 - 4: ... ≥ 1000 DM
 - 5: unknown/ no savings account

- Years_of_Present_Employment: Present employment since



- 1: unemployed
- 2: ... < 1 year
- 3: 1 <= ... < 4 years
- 4: 4 <= ... < 7 years
- 5: .. >= 7 years

- Percentage_of_disposable_income: Installment rate in percentage of disposable income
- Sex_&MaritalStatus: Personal status and sex
 - 1: male (divorced/separated)
 - 2: female (divorced/separated/married)
 - 3: male (single)
 - 4: male (married/widowed)
 - 5: female (single)
- Attribute 10: Guarantors/Debtors: Other debtors / guarantors
 - 1: none
 - 2: co-applicant
 - 3: guarantor
- Duration_in_Present_Residence: Present residence since
- Property: Property
 - 1: real estate
 - 2: if not A121: building society savings agreement/ life insurance
 - 3: if not A121/A122: car or other, not in attribute 6
 - 4: unknown / no property
- Age_in_years: Age in years
- Concurrent_Credits: Other installment plans
 - 1: bank
 - 2: stores
 - 3: none



- Housing: Housing
 - 1: rent
 - 2: own
 - 3: for free
- No_of_Credits_at_this__Bank: Number of existing credits at this bank
- Occupation: Job
 - 1: unemployed/ unskilled - non-resident
 - 2: unskilled - resident
 - 3: skilled employee / official
 - 4: management/ self-employed/highly qualified employee/ officer
- No_of_dependents: Number of people being liable to provide maintenance for
- Telephone: Telephone
 - 1: none
 - 2: yes, registered under the customer's name
- Foreign_Worker: foreign worker
 - 1: yes
 - 2: no

Evaluation Parameters

Evaluation will be based on:

- Data Preparation
- Model Selection.
- Presentation.

Data Preparation

Analyze the data statistically and treat the multicollinear variables



Model Comparison.

Apply logistic regression algorithms for every changes 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.

