Home Improvement Bank Loan

Decision Making Analytics

# Introduction

The consumer credit department of a bank wants to automate the decision-making process for approval of home improvement bank Loan lines of credit. To do this, they will follow the recommendations of the Equal Credit Opportunity Act to create an empirically derived and statistically sound credit scoring model. The model will be based on data collected from recent applicants granted credit through the current process of loan underwriting. The model will be built from predictive modeling tools, but the created model must be sufficiently interpretable to provide a reason for any adverse actions (rejections).What if you can predict clients who default on their loans.

# Dataset Content

The Loan dataset (loan) contains baseline and loan performance information for 5,960 recent bank loans. The target (BAD) is a binary variable indicating whether an applicant eventually defaulted or was seriously delinquent. This adverse outcome occurred in 1,189 cases (20%). For each applicant, 12 input variables were recorded.

The data set loan reports characteristics and delinquency information for 5,960 bank loan loans. A home improvement bank loan is a loan where the obligor uses the loan of his or her home as the underlying collateral. The data set has the following characteristics:

◾ BAD: 1 = applicant defaulted on loan or seriously delinquent; 0 = applicant paid loan  
◾ LOAN: Amount of the loan request  
◾ MORTDUE: Amount due on existing mortgage  
◾ VALUE: Value of current property  
◾ REASON: DebtCon = debt consolidation; HomeImp = home improvement  
◾ JOB: Occupational categories  
◾ YOJ: Years at present job  
◾ DEROG: Number of major derogatory reports  
◾ DELINQ: Number of delinquent credit lines  
◾ CLAGE: Age of oldest credit line in months  
◾ NINQ: Number of recent credit inquiries  
◾ CLNO: Number of credit lines  
◾ DEBTINC: Debt-to-income ratio

# Exploratory Data Analysis (EDA)

Exploratory data analysis employs a variety of techniques (mostly statistical graphics) before making inferences from data. It is essential to examine all variables in the dataset to:

* Catch mistakes or adjusting labels
* Summary statistics interpretation
* See patterns in the data through data visualizations
* Extract important variables’ features
* Detect outliers, anomalies, and missing values
* Gain deep familiarity with the dataset

Refine selection of features that will be used to build the machine learning models.

Special attention to not skip the EDA process, because can generate inaccurate models or accurate models on the wrong data. This dataset contains 5960 observations. There are 12 attributes and one target

# Preprocessing the dataset

Before starting the process, it’s important to answer the following question:

* Is it clear what kind of problem we are dealing with?

because in many cases isn't so simple to identify it. A good understanding of the problem will help to choose the best machine learning algorithm then, right data mining and to make the right predictions or classifications. Thus, the first step, is preprocessing because, in real word, the raw data can be collect from many sources like sensors, websites, public data and many others.

# Variables transformations

Data transformation is the process of converting, cleansing, and structuring data into a usable format. The goal is to prepare data for analysis, reporting, storage, or data mining so it can be used to support decision making and drive organizational growth. Perform transformation if it is necessary

# Descriptive Analysis

The descriptive Analysis is used to simplify and summarize the mainly characteristics of the dataset. In other words, show what kind of information the dataset has. The R Summary method generates a descriptive statistic that summarize the central tendency, dispersion, and shape of the dataset. By using this method in Human Resource dataset, what important insights is possible to see?

# Summary of the Exploratory Data Analysis

Prepare a summary of your analysis about bank loan data

Bank Loan Decision Making Modeling Requirements

# Requirements

Project Name: Bank Loan Decision Making Analytics

Dataset: loan.csv

Method: Data Analytics Modeling

Models: Classifications and Predictions

Approach:

1. Project introduction
2. Business and analytics goals
3. Data preprocessing (such as attributes definition, data exploration, checking missing value, checking zero, and more)
4. Predictor analysis and relevancy
5. Dimension reduction (if needed)
6. Data transformation (if needed)
7. Data partitioning methods
8. Model selection
9. Model fitting, validation accuracy and test accuracy
10. Report models performance
11. Model evaluation (of the selected models)
12. Observation and conclusion

Deadline: Week 9

Delivery:

* Week 6: part 1, 2, and 3
* Week 7: part 4, 5, 6, 7, and 8
* Week 8: part 9, 10, and 11
* Week 9: complete work submission (documentation and presentation)

# Deliverable Policy

Demonstrate your analytics ability as an entry level data analytics professional. Your work should be delivered on time, late work will not be accepted. Do not share your work with other students. Avoid machine learning blunders and fundamental mistakes.