K. N. Toosi University of Technology

Faculty of Physics
Educational Group of
Atomic-Molecular and Astronomy

Special Topics I Final Projects
(Project 2 - Good Customer, Bad Customer)

by

Ali Bagheri Koochaksaraei

Teacher

Dr. Mohammad Hossein Zhoolideh

Academic Year 1402-1403

(First Semester)

Overview

Scoring credit cards is one of the traditional methods of banks to manage risk. Banks use personal information and data submitted by credit card applicants to predict the likelihood of loan repayment by applicants. The bank can decide whether to issue a credit card to the applicant or not.

In this project, we ask you to build a model to predict whether an applicant is a *good* or *bad* customer. To create your label, you need to use **vintage analysis**. Imbalanced data is another major problem. You can download the data needed for this project from this link.

Two datasets named *credit.csv* and *application.csv* have been provided to you in this project (these datasets could be merged by *ID*), the explanation of the columns of each of these datasets is as follows:

application.csv	
Name	**Description**
Id	Client number
Gender	Gender
Own Car	Is there a car
Own Realty	Is there a property
Children	Number of children
Annual Income	Annual income
Income Type	Income category
Education Type	Education level
Family Status	Marital status
Housing Type	Way of living
Days Birthday	Birthday
Days Employed	Start date of employment
Mobile	Is there a mobile phone
Work Phone	Is there a work phone
Phone	Is there a phone
Email	Is there an email
Occupation Type	Occupation
Family Size	Family size

credit.csv		
Name	**Description**	**Remarks**
ID	Client number	-
Months Balance	Record month	The month that the data was collected from is the beginning point; going backwards, the current month is 0, the prior month is -1, and so on.
Status	Status	**0** 1-29 days past due **1** 30-59 days past due **2** 60-89 days overdue **3** 90-119 days overdue **4** 120-149 days overdue **5** Bad or past-due debts, write-offs lasting more than 150 days **C** Paid off that month **X** No loan for the month

Note: The given data is raw. To answer this question, you must first preprocess the data using the Pandas package.

Important Points

Be sure to

- Leave appropriate comments for different parts of your code.

- Completely explain about the algorithm(s) you use to answer this question.

- Use **model selection**, **feature engineering** and **feature scaling** in your code.

- Measure your model performance using model evaluation metrics and interpret the obtained result(s).

- If you used a specific book or article in your project, mention it in your notebook.

A part of your score will be allocated to these items.

* You should write all the steps of your project in the **Jupyter notebook** and upload it as a file with the **.ipynb** extension on the vc site.