

Project Proposal

Abstract

Defaults of credit cards are one of the biggest problems for many companies and banks. On time and efficient identification of probability of default plays a key role in saving banks from any crisis, particularly in the field of payments. In this project, I will build an efficient and accurate model to detect the default credit card of the customers and the model is based on machine learning techniques. Particularly supervised learning algorithms. The model is developed based on classification algorithms including Support vector machine and Decision tree. Will apply the models after exploring the best features. After I train the models I will evaluate each one using an accuracy metric.

Dataset

This data is set to the case of customers' default payments in Taiwan and compares the predictive accuracy of the probability of default. The data includes 24 columns. 23 columns contain independent variables and one dependent variable which will predict default payment next month. And 30000 instances. the dataset columns mentioned as X_n . In following what each one refers to. as the UCI site mentioned

X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.

X2: Gender (1 = male; 2 = female).

X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).

X4: Marital status (1 = married; 2 = single; 3 = others).

X5: Age (year).

X6 - X11: History of past payment. We tracked the past monthly payment records (from April to September, 2005) as follows: X6 = the repayment status in September, 2005; X7 = the repayment status in August, 2005; . . .; X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.

X12-X17: Amount of bill statement (NT dollar). X12 = amount of bill statement in September, 2005; X13 = amount of bill statement in August, 2005; . . .; X17 = amount of bill statement in April, 2005.

X18-X23: Amount of previous payment (NT dollar). X18 = amount paid in September, 2005; X19 = amount paid in August, 2005; . . .; X23 = amount paid in April, 2005.

Tools

There are many software's and tools that are used to run and implement machine learning algorithms such as TPU and GPU from google. These tools help to save time during training models, many software such as Colab which is an online cloud based platform contains the Jupyter Notebook framework which I will use in my project. To build a machine learning classifier will use some important Python libraries: pandas , numpy and different methods.