**SEIS 763: Machine Learning**

**Project Team:**

**Tingting Wang, Kutman Alymbek Uulu, Adeola Olusi, Ashenafi Darihun, & Maxine Thao**



**Project Title:**

**Credit Card Fraud Detection**

**--Anonymized credit card transactions labeled as fraudulent or genuine**

**1-Description of data source and web link(s).**

**Data Source:**

Name: I-Cheng Yeh  
email addresses: (1) icyeh **'@'** chu.edu.tw (2) 140910 **'@'** mail.tku.edu.tw  
institutions: (1) Department of Information Management, Chung Hua University, Taiwan. (2) Department of Civil Engineering, Tamkang University, Taiwan.  
other contact information: 886-2-26215656 ext. 3181

**Dataset link:**

<https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients>

**2.Size/# of records of the dataset or files.**

|  |  |
| --- | --- |
| **Number of Instances:** | 30000 |

**3.# of attributes of the dataset.**

|  |  |
| --- | --- |
| **Number of Attributes:** | 24 |

**4.Description of each attribute.**

**Attribute Information:**

This research employed a binary variable, default payment (Yes = 1, No = 0), as the response variable. This study reviewed the literature and used the following 23 variables as explanatory variables:

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; . . .

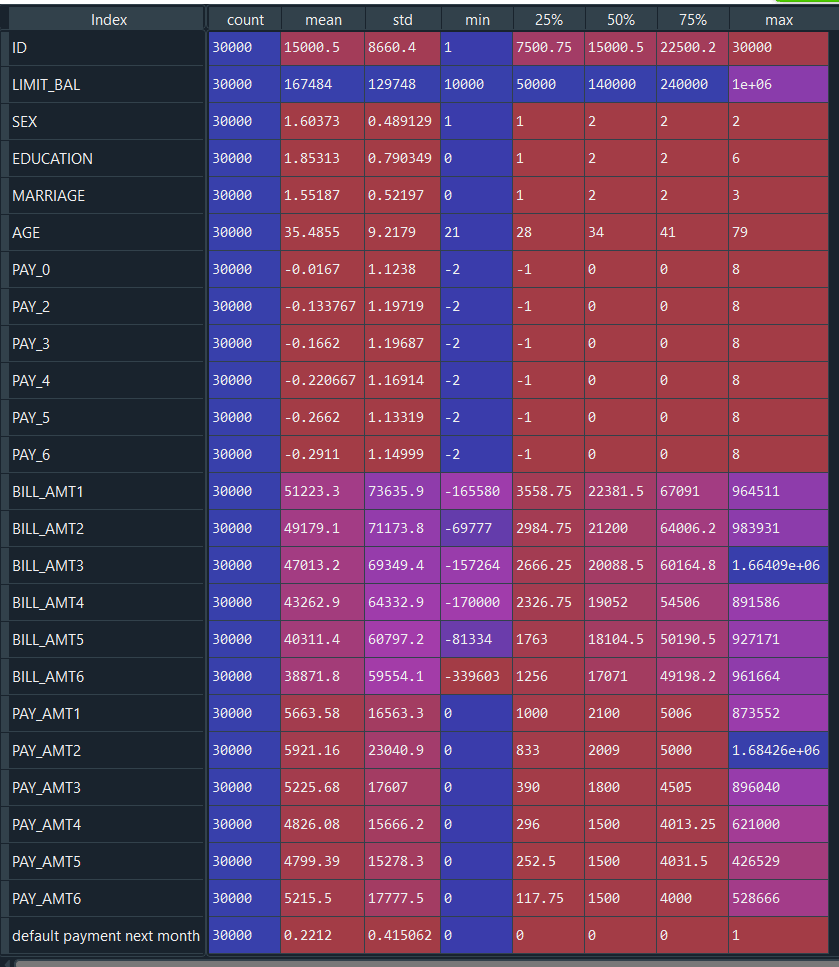
X17 = amount of bill statement in April 2005.  
X18-X23: Amount of previous payment (NT dollar).

X18 = amount paid in September 2005.

X19 = amount paid in August 2005; . . .

X23 = amount paid in April 2005.

**5.Some general statistics of the dataset.**



**6.Tools / methods you plan to use in your study.**

**Tools:** MATLAB, Python package.

**Methods:**

Use linear regression or logistic regression, regularization, and so on.

**7.Exactly what problems/questions your team plans to predict / study.**

By using pre-trained neural network to perform classification on a new collection of images

Data Features

Process flow diagram

Final Machine Learning Model

Training Dataset

Build a model

Cross validation

Machine Learning

Choose The best feature Set

Table

Feature Extraction Method

Dataset