Credit Card Payment Default Prediction

Credit card is a flexible tool by which a customer can use a bank's money for a short period of time.

Predicting accurately which customers are most probable to default represents a significant business opportunity for all banks. Bank cards are the most common credit card type in Taiwan, which emphasizes the impact of risk prediction on both the consumers and banks.

This would inform the bank’s decisions on criteria to approve a credit card application and also decide upon what credit limit to provide.



This dataset contains information on default payments, demographic factors, credit data, history of payment, and bill statements of credit card clients in Taiwan from April 2005 to September 2005.

Using the information given, predict the probability of a customer defaulting in the next month.

About Data Source:

Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.)

**Data Dictionary**

|  |  |
| --- | --- |
| **ID** | Unique ID of each client |
| **LIMIT\_BAL** | Amount of given credit (NT dollars):  It includes both the individual consumer credit and his/her family (supplementary) credit |
| **SEX** | Gender (1=male, 2=female) |
| **EDUCATION** | (1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown) |
| **MARRIAGE** | Marital status (1=married, 2=single, 3=divorced) |
| **AGE** | Age of the client |
| **PAY\_0** | Repayment status in September, 2005 (-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) |
| **PAY\_2** | Repayment status in August, 2005 (scale same as above) |
| **PAY\_3** | Repayment status in July, 2005 (scale same as above) |
| **PAY\_4** | Repayment status in June, 2005 (scale same as above) |
| **PAY\_5** | Repayment status in May, 2005 (scale same as above) |
| **PAY\_6** | Repayment status in April, 2005 (scale same as above) |
| **BILL\_AMT1** | Amount of bill statement in September, 2005 (NT dollar) |
| **BILL\_AMT2** | Amount of bill statement in August, 2005 (NT dollar) |
| **BILL\_AMT3** | Amount of bill statement in July, 2005 (NT dollar) |
| **BILL\_AMT4** | Amount of bill statement in June, 2005 (NT dollar) |
| **BILL\_AMT5** | Amount of bill statement in May, 2005 (NT dollar) |
| **BILL\_AMT6** | Amount of bill statement in April, 2005 (NT dollar) |
| **PAY\_AMT1** | Amount of previous payment in September, 2005 (NT dollar) |
| **PAY\_AMT2** | Amount of previous payment in August, 2005 (NT dollar) |
| **PAY\_AMT3** | Amount of previous payment in July, 2005 (NT dollar) |
| **PAY\_AMT4** | Amount of previous payment in June, 2005 (NT dollar) |
| **PAY\_AMT5** | Amount of previous payment in May, 2005 (NT dollar) |
| **PAY\_AMT6** | Amount of previous payment in April, 2005 (NT dollar) |
| **default\_payment\_next\_month** | Target Variable: Default payment (1=yes, 0=no) |

**Evaluation Metric**

Submissions are evaluated on [area under the ROC curve](http://en.wikipedia.org/wiki/Receiver_operating_characteristic) between the predicted probability and the observed target.

**Public and Private Split**

Test data is further divided into Public (40%) and Private (60%)

* Your initial responses will be checked and scored on the Public data.
* The final rankings would be based on your private score which will be published once the competition is over.

**Guidelines for Final Submission**

Please ensure that your final submission includes the following:

1. Solution file containing the predicted probability in the test dataset (format is given in sample submission csv)
2. Code file containing the following:

* **Code:** Note that it is mandatory to submit your code for a valid final submission

**Hackathon Rules**

1. *The final standings would be based on private leaderboard score*
2. Setting the final submission is recommended. Without a final submission, the submission corresponding to best public score will be taken as the final submission
3. Entries submitted after the contest is closed, will not be considered
4. The code file pertaining to your final submission is mandatory while setting final submission
5. Throughout the hackathon, you are expected to respect fellow hackers and act with high integrity.