

Data Science Framework For Credit One

Agila Subash



Goal

Credit One needs to predict:

- If a customer can be approved or not for a loan
- If approved, what should be the credit limit for that customer.

Credit One : Problem Definition

Credit One, a credit scoring service could lose their customers because there is an increase in customers who have defaulted on their loans that they have secured.

Solution

We can check to see if default status can be predicted with 80% accuracy based on all the other available data

We can also check if the credit limit can be reduced or increased.

Available Data

Credit Limit: X1 (float) : Amount of given credit

Gender (int) : Male , Female (values : male =1; female =2)

Education (int) : Graduate school: 1, university : 2, high school : 3, other : 4

Marital Status X4 (int) : married, single, divorce, others

Age (int) : Age of the customer

Available Data

History of past payment : Past payment record for each month from Sep 2005- Aug 2005.

X6 = the repayment status in September, 2005; X7 = the repayment status in August, 2005; . . .; X11 = the repayment status in April, 2005

-2: No consumption; -1: Paid in full; 0: The use of revolving credit; 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.

Available Data

Amount of Bill (x12- X17) float : Bill amount for each month from Sept 2005 - April 2005.

Amount of Previous payment (x18- X 23) float : Amount Paid for each month for the past 6 months from Sept 2005- April 2005.

Client Behavior Y1 (int) : Default or not Default status. Y=0 then not default, Y=1 then default

DataFrame column

data	
ID	int64
LIMIT_BAL	int64
SEX	object
EDUCATION	object
MARRIAGE	int64
AGE	int64
PAY_0	int64
PAY_2	int64
PAY_3	int64
PAY_4	int64
PAY_5	int64
PAY_6	int64
BILL_AMT1	int64
BILL_AMT2	int64
BILL_AMT3	int64
BILL_AMT4	int64
BILL_AMT5	int64
BILL_AMT6	int64
PAY_AMT1	int64
PAY_AMT2	int64
PAY_AMT3	int64
PAY_AMT4	int64
PAY_AMT5	int64
PAY_AMT6	int64
DEFAULT...	object
CREDIT_LIM	cate...
Default	int64
Gender	int64
Edu_Level	int64

length: 29, dtype: object

Data Location

```
mysql+pymysql://@34.73.222.197/deepanalytics
```

Additional data needs

- There is no mention of how the credit limit is determined
- About the criteria for the loan approval process.
- There is no income data for the customer.

Data cleaning

- Duplicate rows need to be eliminated
- Irrelevant rows need to be eliminated.
- Data types need to be changed to integer to pass through classification algorithms

Building the Model

- The data is going to be split into test data and train data and passed through three classifier algorithms.
- The classifier algorithms would check if the default status of the customer can be predicted based on the other available data
- The classifier algorithms gives the prediction accuracy and the algorithm with best accuracy is picked for prediction

Evaluate the Model

The parameters in the algorithms can be assigned different values to try and increase the accuracy.

Presentation of the Results

- Results will be presented to the stakeholders.
- Depending upon the accuracy of the algorithms accuracy the confidence of the predictions can be determined.
- If the accuracy is less than 80% then the predictions can be overruled.

Maintenance of the Model

Continuous support and monitoring of the accuracy will be provided for the production run of the model.

Further new data would be used to test and train for increasing the accuracy of the prediction.

Flow Chart

