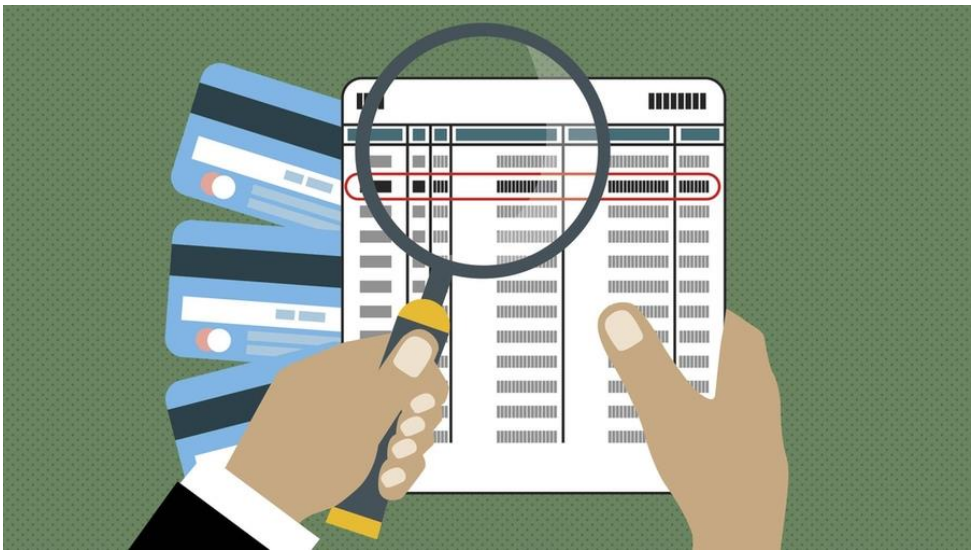

Default Credit Card Clients Prediction

CAPSTONE PROJECT REPORT

Dasun Kehelwala (DSA_0392)

Machine Learning Foundations Training
2022-11-20

Introduction (Problem Definition)

This project is focusing on predicting credit card clients who will default on their next month payment based on their demographic characteristics, past spending patterns and past repayment patterns. This is one of important business problems for banks which provide credit card facilities for Customers. This will be specifically useful to manage credit risks. This challenge is addressed as classification problem in this project. After deployment, customers should be able to access the service through API and also by sending batch input as csv file. Derived machine learning model predicts whether customer is going to default next month payment or not.

Dataset

The dataset used in this project was named as "Default of credit card clients Data Set". This dataset contains the default payment details in Taiwanese banking industry in year 2005. Dataset is available to download downloaded from UCI Machine learning repository through following link.

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

This dataset is Multivariate dataset with 24 attributes and 30,000 instances. All the attributes available in dataset is converted to Real Integer values. This dataset was donated to public access in 2016-01-26. Most notable observation about this dataset is its class imbalance.

Table 1: Attribute Details

Attribute	Description
ID	Identifier for data entry
X1 (LIMIT_BAL)	Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit. (numerical)
X2 (SEX)	Gender (1 = male; 2 = female). →Categorical variable mapped to integers

X3 EDUCATION	Education (1 = graduate school; 2 = university; 3 = high school; 4 = others). →Categorical variable mapped to integers
X4 MARRIAGE	Marital status (1 = married; 2 = single; 3 = others). →Categorical variable mapped to integers
X5 AGE	Age (year) →Numerical
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: -2= , -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. → Categorical variables mapped to integers, but have ordinal nature as per definition
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. →Numerical
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. →Numerical
Y	default payment (Yes = 1, No = 0)

Positive class (Defaulted CX or value 1) represents only 22.1% of whole dataset. Figure 1 depicts the imbalanced Nature of Dataset.

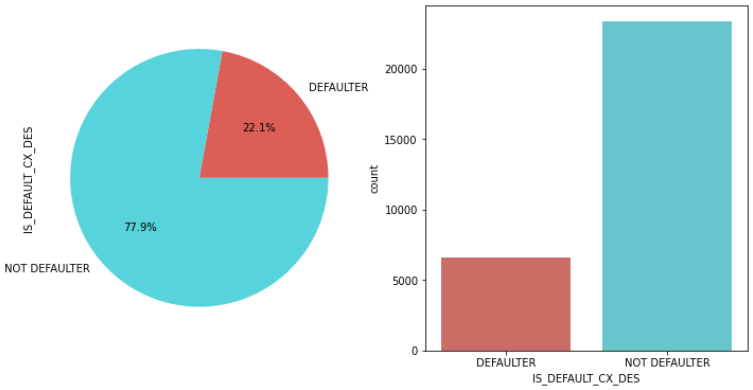

Figure 1 : Imbalanced Nature of Dataset

Methodology (Solution Approach, Tools used)

As per problem definition, ML model need to predict This solution is approached as binary classification problem.

Results

aaaa

Conclusion

aaaa

Discussion

References

- ✓ https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients#
- ✓ Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications, 36(2), 2473-2480.
- ✓ https://www.kaggle.com/datasets/uciml/default-of-credit-card-clients-dataset

✓