greatlearning
Learning for Life

Project Summary

Batch details	PGPDSE-FT-CHENNAI SEP21
Team members	GREASH K, JISMY JOHN, PRUTHIV RAJAN K, SHIBIKANNAN TM, VIGNESH PRABAKARAN
Domain of Project	FINANCE
Proposed project title	THE USAGE OF MACHINE LEARNING TECHNIQUES TO PREDICT THE PROBABLITY OF DEFAULT OF CREDIT CARD CLIENTS.
Group Number	3
Team Leader	SHIBIKANNAN TM
Mentor Name	P.V.SUBRAMANIAN

Date: 08.01.2022

P.V. Subramanian

Signature of the Mentor

Signature of the Team Leader

Table of Contents

SL NO	Topic	Page No
1	Overview	1
2	Business problem goals	1
3	Topic survey in depth	3
4	Critical assessment of topic survey	3
5	Methodology to be followed	5
6	References	7

OVERVIEW

Risk analysis is an important aspect of running a business. Every company should analyze the risk before making a decision. And when it comes to giving credit to customers, credit risk or default risk analysis becomes a vital part which helps the lender to make the right decisions accordingly. Credit risk is the risk that the borrower of money will not be able to pay the interest and repay the principal on a debt as it becomes due. Machine learning models can help to improve the accuracy of credit risk analysis, providing a scientific method to identify potential debtors in advance.

1.BUSINESS PROBLEM UNDERSTANDING

Finance is critical in every business decision, from planning and budgeting and cash flow, management to the capital structure and how you control risks and costs.

Banks and other credit card issuing institutions benefit from issuing credit cards in tangible ways that directly increase their profitability through interest charges, credit card fees, merchant fees and also in intangible ways that increases your loyalty as a customer.

But if the credit card facility is provided to unqualified applicants such as those people with no repayment ability, heavy credit card debts will be created. So, it becomes very important to check the repaying ability of the credit card facility applicants as well as the credit card limit amount and has to be decided prior to credit card issuing.

2.BUSINESS OBJECTIVE

The main objective of the project is to determine whether the credit card holder becomes a defaulter. This helps the credit card issuing institutions for risk management and increase profitability.

3. APPROACH

The approach is to analyze the payment history, bill amount and certain other significant factors that will help to find the probability of default of credit card clients.

greatlearning
Learning for Life

4.CONCLUSIONS

The significant factors affecting the failure of repayment of credit card payment is analyzed based on the dataset and is used to create a model to understand the probable number of defaulters.

This will also help the card issuer to decrease the credit limit amount , hence reduce the credit risk in the coming months.

TOPIC SURVEY

1.PROBLEM UNDERSTANDING

Credit card accounts are revolving credit lines, and because of this, lenders and investors have more options to actively monitor and manage them compared to other retail loans. Consequently, managing credit card portfolios is a potential source of significant value to financial institutions. Better risk management could provide financial institutions with savings on the order of hundreds of millions of dollars annually.

As time passes, due to different reasons some customers become defaulter and hence unable to repay the amount. So, it becomes necessary to check the ability of the credit hold holders to repay the debt.

2.CURRENT SOLUTION TO THE PROBLEM

Each customer is analyzed on their ability to repay the debt on time and sort out those customers who is likely to be the defaulters in the following month.

3.PROPOSED SOLUTION TO THE PROBLEM

In order to reduce the risk of default to the bank or credit card issuing institutions, lenders could cut or freeze credit lines on accounts that are likely to go in default.

More accurate forecasts also improve macroprudential policy decisions, and reduce the likelihood of a systemic shock to the financial system.

4.REFERENCE TO THE PROBLEM

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

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

CRITICAL ASSESSMENT OF TOPIC SURVEY

1. Find the key area, gaps identified in the topic survey where the project can add value to the customers and business.

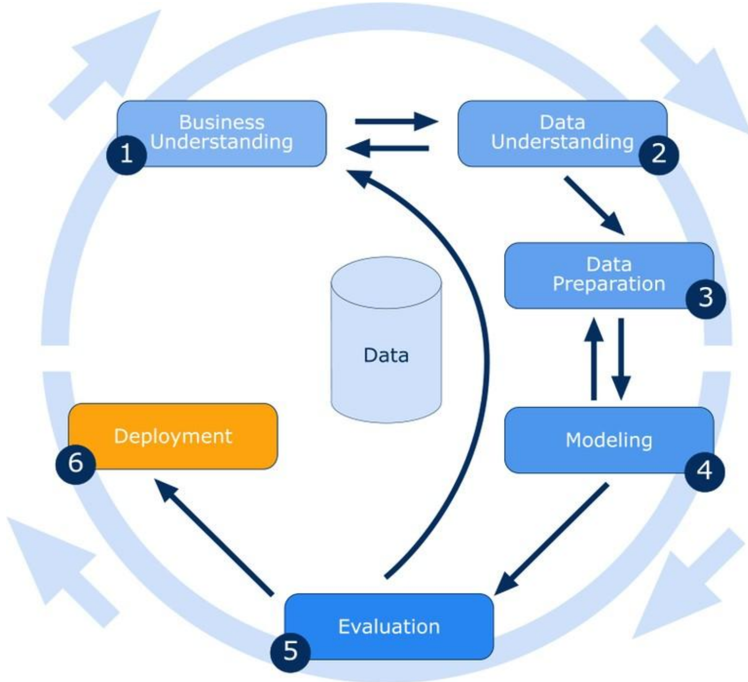
Critical assessment of the project will help us to know whether the customer is a defaulter or not but doesn't provide the accurate reasons for becoming a defaulter or cannot analyze whether there will be a betterment in the defaulter's financial status in the upcoming months. If we are able to accurately identify the customer, the business would not loose potential customers.

As suggested in the solution, if the business cut or freeze credit line on accounts that are likely to go into default, it will potentially avoid an increase in balances of accounts destined to default, known in the industry as 'run-up'. However, cutting these credit lines to reduce run-up also runs the risk of cutting the credits limits of accounts that will not default, thereby alienating customers and potentially forgoing profitable lending opportunities.

2.What key gaps are you trying to solve?

Our aim is to produce more accurate forecasts of delinquencies and defaults. This will reduce the likelihood of false positives which in turn helps both the parties that is the lender and the borrower

METHODOLOGY TO BE FOLLOWED



METHODOLOGY

1. Business Understanding:
To predict whether the customer will default on their credit card payment next month based on their previous payment characteristics.

The dataset consist of 30,001 rows and 25 columns

2. Data Understanding:
Dataset extracted from UCI Repository consists of a table comprising information of Bill amounts and Payment amounts of customers for the last six months along with personal details. With the help of these attributes, we will build our model accordingly.

This research employed a binary variable, default payment (Yes = 1, No = 0), as the response variable. The response variable distribution is 77.9% and 22.1% for 0 and 1 respectively.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.

3. Data Preparation:
Exploratory Data Analysis on the dataset is to be done to understand and process the data. Feature engineering techniques are to be performed in order to predict whether the customer will default on their credit card payment next month.

4. Modelling:
Since the final objective is to predict whether the customer will default on their credit card payment next month, Classification techniques can be used. Classification Models that act as base learners will be used for analysis. After evaluation of the base learner models, advanced classification models will also be used if required.

5. Evaluation:
Along with ‘Confusion Matrix’ of the model using both training and test data,
'AUROC', ‘f1_score’, ‘Recall’, ‘Precision’ performance measures will be used to assess and finalise the best-fit model as the dataset is not balanced. It is proposed to use SMOTE (synthetic minority oversampling technique),one of the most commonly used oversampling methods to solve the imbalance problem.

REFERENCES

- Lee, T. S., Chiu, C. C., Lu, C. J., & Chen, I. F. (2002). Credit scoring using the hybrid neural discriminant technique. Expert Systems with Applications, 23(3), 245–254.
- Lee, Y. S., Yen, S. J., Lin, C. H., Tseng, Y. N., Ma, L. Y. (2004). A data mining approach to constructing probability of default scoring model. In Proceedings of 10th conference on information management and implementation (pp. 1799–1813).
- Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science, 49(3), 312–329.
- Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society, 54(6), 627–635.
- Berry, M., & Linoff, G. (2000). Mastering data mining: The art and science of customer relationship management. New York: John Wiley & Sons, Inc.
- Chou, M. (2006). Cash and credit card crisis in Taiwan. Business Weekly, 24–27.
- https://arxiv.org/ftp/arxiv/papers/1802/1802.05326.pdf
- https://blog.crifhighmark.com/the-importance-of-credit-risk-management-in-banking/#:~:text=They%20need%20to%20manage%20their,reserves%20at%20any%20given%20time.
- https://www.cqf.com/blog/importance-of-machine-learning-for-risk-management#:~:text=One%20prominent%20use%20case%20for,is%20fraud%20detection%20and%20prevention.&text=As%20in%20the%20case%20of,questionable%20behavior%20has%20been%20discerned.
- https://www.researchgate.net/publication/333871592_Credit_Card_Risk_Assessment_Based_on_Machine_Learning

Original owner of data	I-Cheng Yeh Department of Information Management, Chung Hua University, Taiwan. (2) Department of Civil Engineering, Tamkang University, Taiwan.
Data set information	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.
Any past relevant articles using the dataset	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.
Reference	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.
Link to web page	https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients