Ibiyengha Tobin - 301256083

Analytics Startup Plan

Project	*Predicting Default in Credit Card Payment for the next month*
Requestor	*Centennial College*
Date of Request	*July 2023*
Target Quarter for Delivery	*3rd Quarter of 2023*
Epic Link(s)	https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients
Business Impact	*This will help financial companies manage risk effectively, maximize profitability, develop customized customer strategies, and adhere to regulatory requirements.*

1.0 Business Opportunity Brief

Let me begin by explaining when an account can be said to be in credit card default. The account will become delinquent before it enters default. After 30 days without making a payment, this occurs. The credit card is severely past due if six consecutive months go by, during which the person fails to make at least the minimum payment required. At that point, default typically occurs.

According to Statistics Canada, Consumer Credit in Canada increased to 728312 CAD Million in April from 724113 CAD Million in March of 2023.

Credit card default has a significant impact on the financial institutions giving out these credit cards. Financial institutions suffer financial losses as a result of credit card defaults since they are unable to recoup the outstanding balances and interest fees from defaulting consumers, which has an effect on their balance sheets and profitability. The performance of a financial institution's loan portfolio as a whole can be severely impacted by credit card defaults, which can have an influence on important performance metrics like delinquency rates, charge-off rates, and loan loss reserves. Consistently high default rates can damage a financial institution's reputation and weaken consumers' trust, which could result in loss of customers and make it challenging to attract new clients.

To address this issue of credit card default and to preserve a solid financial position, financial institutions must implement effective risk management tactics and that is what this project seeks to do. I intend to show how to predict a default one month in advance.

The specific ask:

This project's objective is to use certain supervised machine learning algorithms to pinpoint the major factors that influence the probability of credit card default while highlighting the statistical foundations of the techniques employed.

The objective is to create an automated model that can both identify the important variables and predict a credit card default based on the client's information and previous transactions.

I will use classification analysis to examine details about the consumers, such as their age, marital status, education, and payment history in the past, in order to determine whether or not they will miss their credit card payment for the subsequent month.

1.1 Supporting Insights

This project stems from the fact that credit card default is a major source of concern for banks and financial organisations. This is because the credit card portfolio is regarded as an asset by these institutions and has a direct impact on their profitability.

The Vice-President of Advanced Analytics at Equifax Canada (Rebecca Oakes), said that "Financial stress is becoming a very real thing for many more Canadians. Its impact on consumer credit is not just visible in day-to-day credit card spending, but also in other non-mortgage debt like auto loans and lines of credit, where balances are on the rise."

Credit card balances increased throughout 1Q23 as some households leaned more on debt due to rising living costs and fell behind on payments. While the index remains lower than pre-pandemic levels (1.00% in March 2023 vs. 1.33% in February 2020), Fitch[1] anticipates the number of overdue accounts to rise because the economy has not to experience the full weight of the 2022 interest rate hikes.

It is important to mention that with the growing number of credit card users, banks have been facing an escalating credit card default rate. According to Canadian Bankers Association "76.2 million Visa and MasterCard cards are in circulation in Canada".

The number of customers who rely on credit cards for daily payment physical and online stores has grown steadily, as has the number of credit cards issued and the staggering amount of debt owed by cardholders. As a result, the majority of financial institutions have to deal with credit card default problems it turns out, data analytics can offer solutions to address the current situation and control credit risks.

This project offers a solution to the issue of credit card default because the ability to pay back the debt is now a significant factor in the finance sector. We can predict potential default accounts based on certain characteristics to solve this problem. The theory is that financial institutions will accept less losses if possible, default accounts are discovered earlier.

[1] Fitch Ratings(2023, May 11) *Canadian Credit Card Index - First-Quarter 2023*

1.2 Project Gains

- Credit default prediction will enable financial institutions to make the best lending decisions while reducing risk and exposure, which improves customer satisfaction and promotes solid business practises.
- Credit card issuers can take proactive measures to reduce risk and exposure by foreseeing which clients are most likely to fail on their credit card accounts.
- Many scholars have proved that machine learning has great practical value in solving various risk problems, including credit risk prediction.
- Two real-world credit card datasets were provided by the owen machine learning database of the University of California, and Luo et al.[2] applied the clustering-based classification method (CLC) to them. The efficiency of CLC over support vector machines has been demonstrated through practise. This project intends to build several models to further test which one has the best efficiency.
- Having an in-depth credit card default prediction system might provide you a distinct competitive advantage in the fiercely competitive credit card market. Customers are more likely to be drawn to and stayed with by financial institutions that can successfully manage default risk and keep a stronger credit card portfolio.
- Financial institutions can provide tailored support to clients at risk of default when they have a credit card default prediction model. Proactive action can be started to assist consumers in managing their debt, negotiating alternative payment plans, or investigating financial counselling

[2] Xiang Gao et al 2021 J. Phys.: Conf. Ser. 1828 012122

2.0 Analytics Objective

The main objective of this project is to develop a model that can be used to both identify the important variables and predict a credit card default based on client information and previous transactions.

Financial institutions will also have early warning indicators of probable default, enabling them to manage and reduce credit risk by acting pro-actively.

The project intends to test the hypotheses that: There will be lesser credit card default if there is a reliable model predicting the default.

KEY QUESTIONS:

1. What forecasting models are appropriate for predicting credit card default?
2. Which variables are most effective at predicting credit card default?
3. What methodology will be used to assess the effectiveness of the credit card default prediction model(s)?

2.1 Other related questions and Assumptions:

1. There may be an assumption that the features chosen for credit card default prediction remain constant and retain their predictive ability over time.
2. There might be a failure to consider that people who use credit cards occasionally face unforeseen financial problems that put a burden on their resources and make it challenging for them to fulfil their credit card payment commitments, such as medical expenses, home repairs, or automobile accidents.

2.2 Success measures/metrics

Accuracy

It gives a general measurement of the predictive capability of the model.

The KPI here is **Accuracy Rate.**

Confusion Matrix

This accuracy is broken down in the confusion matrix as true positives, true negatives, false positives, and false negatives. Together, these measures help comprehend the model's capacity for correctly classifying default and non-default cases as well as the likelihood that it will misclassify some instances.

K-Fold Cross Validation

K-fold cross-validation offers a more in-depth evaluation of the model's generalisation skills. This method offers a more thorough picture of the model's predictive capability and robustness, assisting in making defensible conclusions about its applicability for credit card default prediction. It does this by taking into account multiple training and validation subsets.

The KPI here is **K-Fold**

Stratified K-Fold Cross Validation

An improvement on K-fold that takes into account the class distribution within the dataset is called stratified K-fold cross-validation. Each fold in a stratified K-fold preserves a distribution of classes that is identical to the distribution in the original dataset since the class proportions are retained in each fold.

The KPI here is **Stratified K-Fold**

2.3 Methodology and Approach

Type of Analysis: *Classification and Regression*

Methodology: *I will start by removing missing values, outliers, and irregularities from the data. After that, I will conduct an exploratory data analysis. The response variable will therefore be defined as 1 in the case of a default and 0 in the absence of one. I will then use logistic regression to begin my classification analysis. I will carry out different classification analysis. The goal is to* *pinpoint the factors that are the main influencers in predicting credit card default and also predicting a credit card default a month in advance.*

Output: *The output will be a set of insights, rules and strategic recommendations that will help me to evaluate the best model that gives the important drivers in predicting credit card default, and credit card default a month in advance.*

3.0 Population, Variable Selection, considerations

Audience/population selection: 30000 Customers who defaulted in credit card payments in Taiwan

Observation window:

Inclusions: 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.

Exclusions:

Data Sources: https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients

Audience Level: Data Scientists, Risk Managers, Business Stakeholders(Executives in financial instititions), and compliance officers.

Variable Selection: There are 25 Variables

1. ID: ID of each client
2. LIMIT_BAL: Amount of given credit in New Taiwan(NT) dollars
3. GENDER: (1 = Male, 2 = Female)
4. EDUCATION (1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)
5. MARRIAGE : Marital status (1=married, 2=single, 3=others)
6. AGE: Age in years
7. REPAY_SEPT : Repayment status in September, 2005
8. REPAY_AUG: Repayment status in August, 2005
9. REPAY_JULY: Repayment status in July, 2005
10. REPAY_JUNE: Repayment status in June, 2005
11. REPAY_MAY: Repayment status in May, 2005
12. REPAY_APR Repayment status in April, 2005
13. BILL_AMT_SEPT: Amount of bill statement in September, 2005 (NT dollar)
14. BILL_AMT_AUG: Amount of bill statement in August, 2005 (NT dollar)

15. BILL_AMT_JULY: Amount of bill statement in July, 2005 (NT dollar)
16. BILL_AMT_JUNE: Amount of bill statement in June, 2005 (NT dollar)
17. BILL_AMT_MAY: Amount of bill statement in May, 2005 (NT dollar)
18. BILL_AMT_APR: Amount of bill statement in April, 2005 (NT dollar)
19. PREV_AMT_SEPT: Amount of previous payment in September, 2005 (NT dollar)
20. PREV_AMT_AUG: Amount of previous payment in August, 2005 (NT dollar)
21. PREV_AMT_JULY: Amount of previous payment in July, 2005 (NT dollar)
22. PREV_AMT_JUNE: Amount of previous payment in June, 2005 (NT dollar)
23. PREV_AMT_MAY: Amount of previous payment in May, 2005 (NT dollar)
24. PREV_AMT_APR: Amount of previous payment in April, 2005 (NT dollar)
25. DEFAULT_PAY_NEXTMNTH: (1=yes, 0=no)

Derived Variables:

Assumptions

- The model will predict the default.
- The model will provide the key variables for predicting credit card default.
- It is assumed that the features chosen for the model have predictive ability and are significant to the risk of credit card default.

Data Limitations

The project does not take into consideration future shifts in consumer behavior, the state of the economy, or other variables that can have an impact on credit card defaults.

4.0 Dependencies and Risks

Risk	Likelihood (based on historical data)	Delay (based on historical data)	Impact
The predicted accuracy of the model may not generalise effectively to new or unknown data.	*Low*		Complex relationships can be missed by the model, or it might not be able to adjust to

			shifting trends in credit card defaults.

5.0 Deliverable Timelines

Item	Major Events / Milestones	Description	Scope	Days	Date
1.	Kick-off / Formal Request	*The project will be formally initiated by meeting with my supervisor*		*1*	*4 July 2023*
2.	Assessment / Triage				
3.	Prioritization				
4.	Data Exploration & Analysis • Issues with duplicates • Issues with Spend data	Exploratory data analysis (EDA) is a crucial step in data analysis that makes use of statistical summaries and graphical representations to find trends, patterns, or verify assumptions in data.		7	*11th July -17th July 2023*
5.	Story Board 1				
6.	QA Output				
7.	Internal team Presentation				*8th August 2023*
8.	Go/No Go				
9.	Story Board 2				
10.	Pilot				
11.	Delivery & sign-off				*16th August 2023*

References

Equifax Canada Market Pulse Quarterly Credit Trends Report (2022, Sept 6) *Financial Stress Mounts, Credit Card Demand and Debt Rise*
https://www.consumer.equifax.ca/about-equifax/press-releases/-/blogs/financial-stress-mounts-credit-card-demand-and-debt-rise/

Fitch Ratings (2023, May 11) *Canadian Credit Card Index - First-Quarter 2023* –
https://www.fitchratings.com/research/structured-finance/canadian-credit-card-index-first-quarter-2023-11-05-2023#:~:text=Late%2DStage%20Delinquencies%20Edge%20Up,level%20since%201Q22%20(1.07%25).

Canadian Bankers Association (2023, March 31) *Focus: Credit Cards: Statistics and Facts* –
https://cba.ca/credit-cards

Xiang Gao et al 2021 J. Phys.: Conf. Ser. 1828 012122
https://iopscience.iop.org/article/10.1088/1742-6596/1828/1/012122/pdf