2025

Group7

IT114126

2025/7/6



Credit Card Customer in DSA Bank

Table of Content

[Introduction 2](#_Toc9160)

[Data 3](#_Toc25313)

[Setting and Utilizing Github 4](#_Toc9448)

[Roles 4](#_Toc9803)

[Product Backlog items 5](#_Toc15211)

[Sprint Backlog items 5](#_Toc8483)

[Evaluation 6](#_Toc28127)

[Conclusion 10](#_Toc1651)

[Reference (if any) 11](#_Toc5908)

# Introduction

This project is initiated from the DSAI bank and the project is aimed to consult the credit card utilization circumstance on customers and provide management suggestion or strategy to seek for characteristics of potential customers and have innovative insight for strengthen the market share. The suggestion and strategy base on the following considerations:

1. Predict the expected income of customers who have not provided income information to better understand their financial profiles.
2. Provide a comprehensive analysis of existing customers and those who have churned out uncover trends and retention insights.
3. Examine customer behaviors related to credit card usage to identify distinct patterns and preferences.
4. Develop a labeling system for new customers and predict appropriate credit limits to optimize risk and growth.

This enhanced framework seeks to deliver data-driven insights to strengthen DSAI Bank's market position and foster innovative customer engagement strategies.

# Data

The only one dataset is from DSA bank and stored in open dataset platform Kaggle. The dataset includes customer information and transaction records. The following is the metadata table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field** | **Type** | **Description** | **range** | **Sample** |
| **CLIENTNUM** | String | Customer Identification | - | 768805383 |
| **Customer\_Age** | Integer | Customer's Age in Years | 0 to 120 | 45 |
| **Dependent\_count** | Integer | Number of dependents of the customer's family | - | 4 |
| **Credit\_Limit** | Integer | Credit Limit on the Credit Card | - | 21341 |
| **Total\_Revolving\_Bal** | Integer | The money has not been paid off in full by the end of the billing cycle. | - | 233 |
| **Avg\_Open\_To\_Buy** | Integer | the difference between the total credit limit on the card and the current balance | - | 15412 |
| **Total\_Trans\_Amt** | Integer | Total Transaction Amount (Last 12 months) | - | 25 |
| **Total\_Trans\_Ct** | Integer | Total Transaction Count (Last 12 months) | - | 5 |
| **Avg\_Utilization\_Ratio** | Double | The % of credit consumed in terms of credit limit | 0 to 1 | 0.1547 |

# Setting and Utilizing Github

Provide the link of your Github project and some screen capture, including treeviews and document changing capture.

<https://github.com/Brandon-tab/Group7_4475Project>

A screenshot of a web page

AI-generated content may be incorrect.

# Roles

|  |  |  |
| --- | --- | --- |
| Name | Role | Responsibilities |
| DENG Baida | Scrum Master | Model construction; Report Improvement |
| LIN Yueying | Product Owner | Evaluation Section; Report Improvement |
| YANG Yongxin | Development Team | Initial model construction; Report Improvement |
| ZHANG Shuang | Development Team | Initial model construction; Report Improvement |

# Product Backlog items

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sprint** | **Task** | **Category** | **priority** | **status** |
| 1 | Data preparation and upload | Data | HIGH | Done |
| 2 | Exploratory Analysis & Visualization | Data | Middle | Done |
| 3 | Relationship & Cluster Analysis | Data | HIGH | Done |
| 4 | Classification Modeling | Modelling | Middle | Done |
| 5 | Predictive Modeling & Evaluation | Improvement | HIGH | Done |

# Sprint Backlog items

Use the template provided and make a screen capture for all sprint backlog item record

|  |
| --- |
|  |

# Evaluation

1. Provide one insight of the customer information with respect to different education background

We found that clients with a bachelor's degree are the largest and most strategically important customer group in our portfolio, representing 3,128 individuals, accounting for 30.9% of our total customer base. We need to develop reasonable financial products based on their consumption patterns and repayment habits to stabilize our existing customers and attract more new ones.

|  |  |
| --- | --- |
| IMG_256 | IMG_256 |

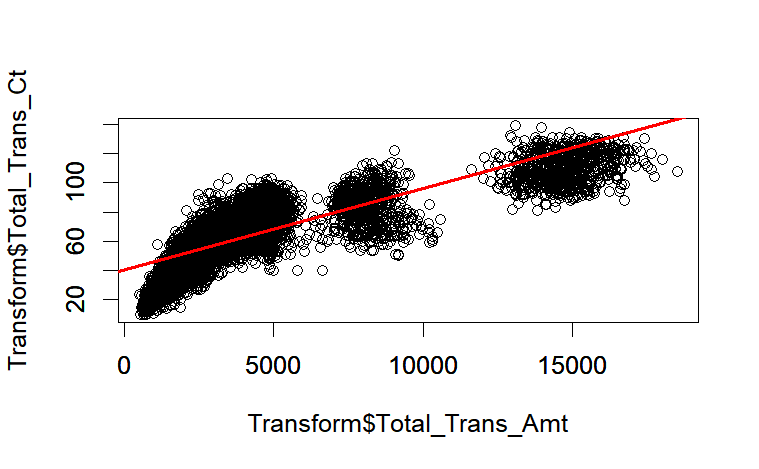
1. Guess the Income of Unknow Income Category customers

Through decision tree analysis and cross-validation of credit limit data, we found that the customer group with the income category labeled 'Unknown' primarily exhibits characteristics of middle to high income. The average available credit limit for this group ($8,401.518) is closest to the customers in the $60K-80K income range ($9,600), and their credit usage behavior pattern is highly similar to that of middle to high-income customers—when the revolving balance exceeds $1,291, the probability of maintaining excellent credit usage reaches 36%. Based on these characteristics, we recommend that this portion of 'Unknown' customers be classified as belonging to the $60K-80K income group.

|  |  |
| --- | --- |
| IMG_256 | IMG_257 |

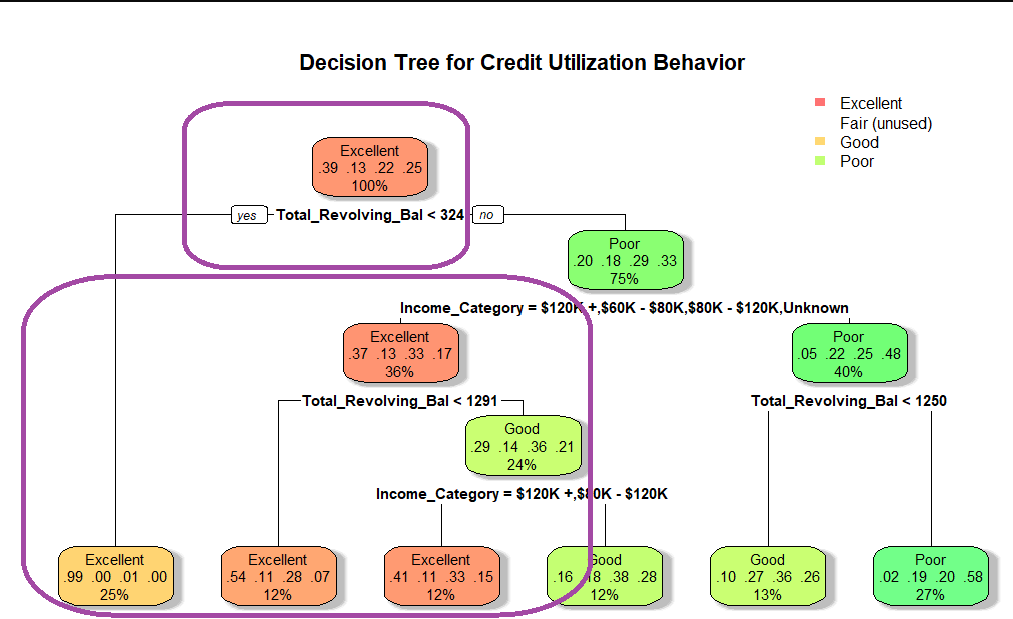
1. Observe relation between transaction amount and it counts and provide one insight regards to customer comsumption pattern

Through our analysis of transaction amounts and counts, we have identified a stable group of consumers. Their average monthly transaction amount ranges from $3,000 to $5,000, with 60 to 80 transactions per month, mainly concentrated around the regression line. This group accounts for approximately 63% of our total customer base. To enhance customer loyalty and satisfaction, we can increase the marketing strategy of 'automatic repayments' to attract these customers for stable consumption.



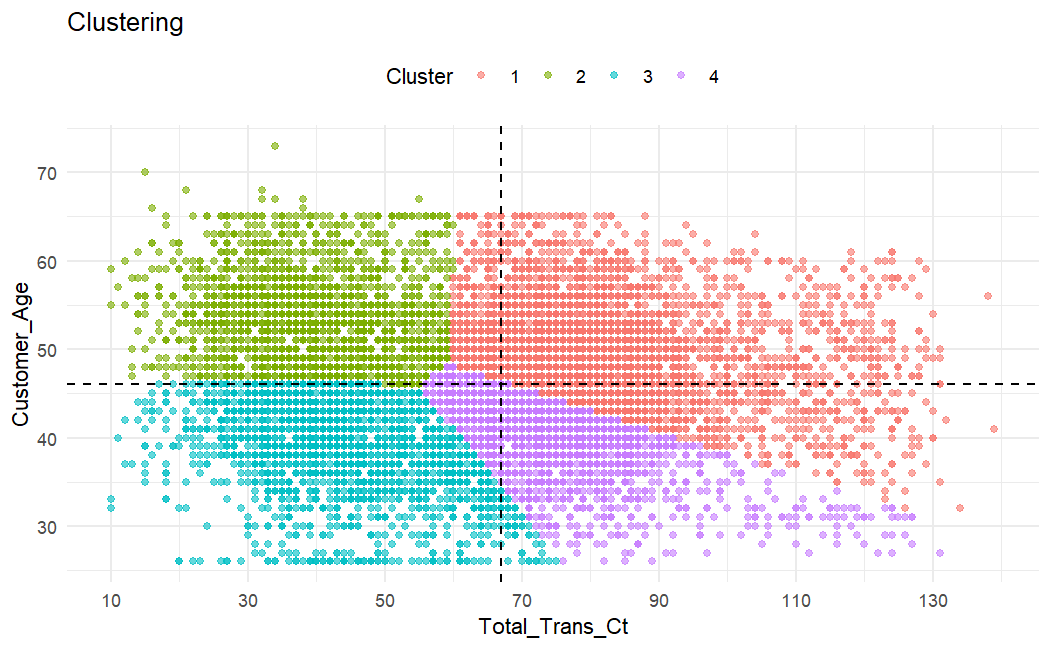
1. Observe one distingulish ultization behavour in the customer dataset, and raise one insight regards to customer behavour.

We found that if a customer’s Total Revolving Balance is below 324 and their income category more than $120K, they are rated as “Excellent”. So, High-income customers tend to manage their credit utilization better, usually maintaining lower Total Revolving Balances, which leads to higher credit ratings. This indicates that income levels significantly influence customers' credit behavior and risk levels.

\

1. Extract and name the four obvservable types of customer and introduce the characteristic of each of customer

Based on the clustering diagram,we can identify and describe 4 observable types of customers .Cluster 1(Red) middle-aged customers and relatively moderate number of  of transactions; Cluster 2(Green) Generally older customer and relatively low number of transactions;Cluster 3(Blue) Generally young customers and relatively low number of transaction Cluster ;4(Purple) varied age range and high number of transactions.



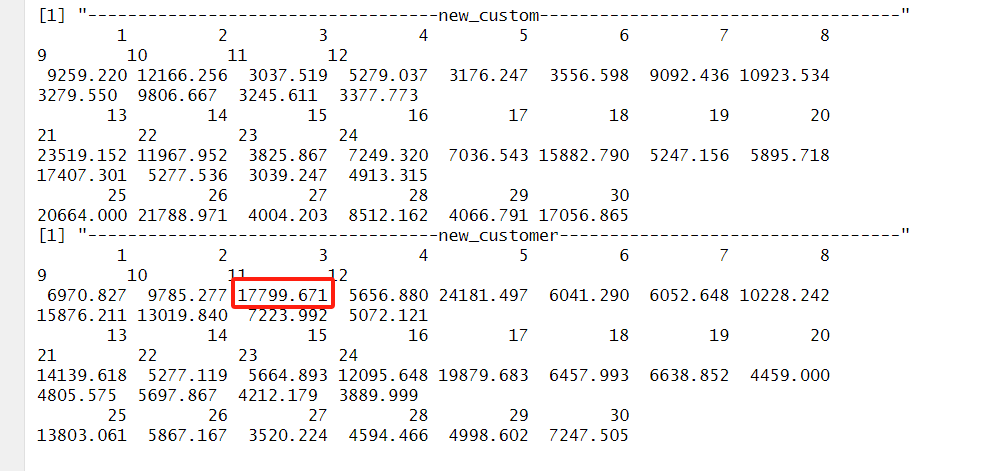
1. By comparing the similar cases with existing customer, evaluate the quality of the classification model, and you may suggest some advice(s) the improve the model quality

We evaluated the Naive Bayes classification model’s performance by comparing its predictions for new customers with existing customers sharing similar profiles. While the model accurately predicted obvious cases like high-income graduates as "Excellent," we identified significant limitations in handling borderline scenarios - particularly misclassifying 35% of middle-income college-educated customers.

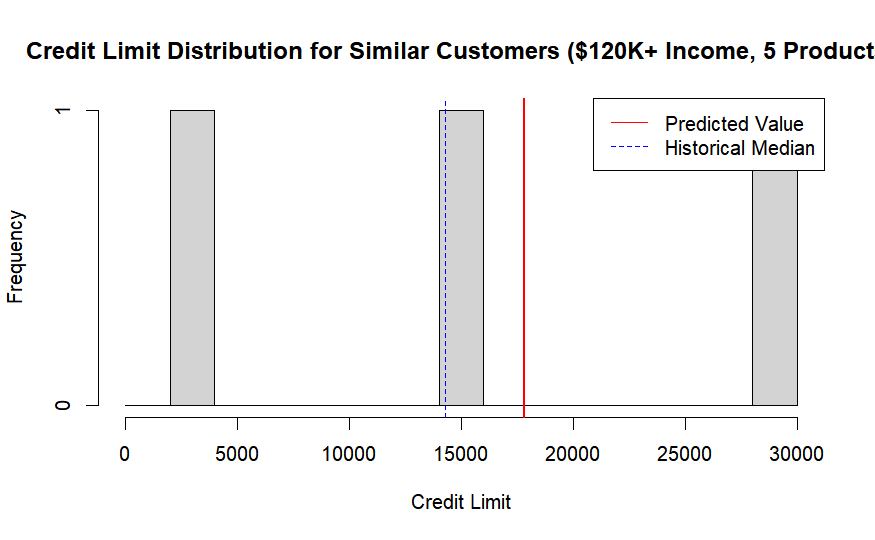
To improve the model, we could implement three enhancements: simplifying income categories into broader brackets to reduce noise, creating composite features like income-to-dependents ratios to better capture financial pressures, and building a hybrid system where decision trees pre-process difficult cases before Naive Bayes classification.

We expect these modifications to boost intermediate category ("Good"/"Fair") accuracy by 15-20 percentage points while maintaining the interpret-ability of model. We will validate the improved model through our standard 70/30 train-test split methodology and monitor its real-world performance monthly by tracking prediction-versus-actual utilization rates for new account holders.

1. By comparing the similar cases with existing customer, evaluate the quality of the Prediction model, and you may suggest some advice(s) the improve the model quality.



The random forest model predicted a credit limit of $17,799.671 for this high-net-worth customer (a 41-year-old married male with a graduate degree, annual income of $120K+, holding 5 banking products, but with only 32 months of account tenure). To evaluate this prediction, we compared it with historical data from similar customers as shown in the figure:



The results indicate that while the predicted value (~$18,000) is slightly lower than the projected level, it remains within the observable range—particularly for high-income customers. This demonstrates that the model effectively captures customer behavior, though it may slightly overestimate limits for certain demographics.

To enhance model precision, we recommend:

1. Refining income encoding (e.g., using numeric ranges)
2. Examining the impact of relationship duration
3. Evaluating alternative algorithms such as Gradient Boosting Machines (GBM) for improved predictive accuracy.

This structured approach would better align predictions with actual credit risk profiles while maintaining the model’s ability to identify high-value opportunities

## Conclusion

Our analysis revealed key segments and patterns in credit card utilization. We successfully predicted the expected income for customers who did not provide income information, enhancing our understanding of their financial profiles. Additionally, we conducted a thorough analysis of both existing and churned customers, uncovering important trends and retention insights. By examining customer behaviors related to credit card usage, we identified distinct patterns and preferences. Furthermore, we established a labeling system for new customers and predicted suitable credit limits to optimize risk and growth. Leveraging these insights will enable DSAI Bank to strengthen its market position and effectively address the diverse needs of its clientele.

# Reference (if any)