CREDIT CARD CUSTOMER CHURN ANALYSIS

GROUP 8

SUJAY KOTWALE(SXK220067)

RYAN DUONG(RXD180030)

BHUSHAN RAVINDRA BAMBLE(BXB220006)

ADITYA JAIN(AXJ220043)

MEHAK DHAWAN(MXD210073)

GUIDED BY

PROFESSOR ZHE ZHANG

MIS 6356.006 - BUSINESS ANALYTICS WITH R

Fall 2022

December 7, 2022

December 7, 2022

Contents

List of Figures

1 Motivation

All bank companies have credit card facilities, to keep customers loyal and move to other credit card a challenge for each bank companies. Current situation, In a bank, Manager is notified with an alarming number of customers leaving their credit card services. Now the problem is how to identify and retain the customers who is going to leave the bank credit cards in the coming months, So that Bank can take actions to retain them. Here comes the role of business intelligence Engineer, who will help the bank to predict, who is gonna leave their company so they can proactively reach to the customer to provide them better services and turn customers decisions in the favour of the bank.

2 Problem Introduction

Customer attrition, also known as customer churn, customer turnover, or customer defection, is the loss of clients or customers. Telephone service companies, Internet service providers, pay TV companies, insurance firms, and alarm monitoring services, often use customer attrition analysis and customer attrition rates as one of their key business metrics. Likewise, financial institutions providing credit card services also look up to customer attrition with due diligence. The process of approving a new customer for a credit card is time-consuming and costly. On the other hand, the cost of retaining an existing customer is significantly lower than acquiring a new customer. An existing customer, when satisfied, can also be thought of as a potential source of organic brand promotion and attract new customers through 'word of mouth' without the financial institution having to spend money on marketing for new customer acquisition. In this project, we have attempted to build a model that can predict whether a customer will churn or not.Input and output of model This model can be helpful for the financial institutions to reduce the customer churn rate by giving special attention and offering targeted promotional offers to customers expected to churn.

3 Data Description

We found relevant dataset about bank customer information from Kaggle, which consists of more than 10127 pieces and includes 23 features such as age, income, marital status, credit card limit and so on. 16 of them are numerical while 5 of them are categorical and 2 were the results so we have removed that columns.

In the dataset we defined the three main classes of features,

Categorical features: Customer_Age, Gender, Education_Level, Marital_Status, Income_Category.
customer-bank relationship features: -Dependent_count: Number of people uses that specific account
-Card_Category: Is the card Premium or a Basic account?
-Months_on_book: the duration of the relationship in the present condition.

	dataFeatures	dataType	null	nullPct	unique	uniqueSample
0	CLIENTNUM	int64	0	0.0	10127	[720274158, 788675658]
1	Attrition_Flag	object	0	0.0	2	[Attrited Customer, Existing Customer]
2	Customer_Age	int64	0	0.0	45	[49, 42]
3	Gender	object	0	0.0	2	[F, M]
4	Dependent_count	int64	0	0.0	6	[3, 5]
5	Education_Level	object	0	0.0	7	[Uneducated, Unknown]
6	Marital_Status	object	0	0.0	4	[Married, Unknown]
7	Income_Category	object	0	0.0	6	[$40K - $60K, $120K +]
8	Card_Category	object	0	0.0	4	[Silver, Platinum]
9	Months_on_book	int64	0	0.0	44	[37, 35]
10	Total_Relationship_Count	int64	0	0.0	6	[6, 5]
11	Months_Inactive_12_mon	int64	0	0.0	7	[3, 5]
12	Contacts_Count_12_mon	int64	0	0.0	7	[3, 5]
13	Credit_Limit	float64	0	0.0	6205	[3414.0, 1808.0]
14	Total_Revolving_Bal	int64	0	0.0	1974	[2260, 1666]
15	Avg_Open_To_Buy	float64	0	0.0	6813	[5267.0, 1555.0]
16	Total_Amt_Chng_Q4_Q1	float64	0	0.0	1158	[0.382, 1.126]
17	Total_Trans_Amt	int64	0	0.0	5033	[14230, 2508]
18	Total_Trans_Ct	int64	0	0.0	126	[25, 86]
19	Total_Ct_Chng_Q4_Q1	float64	0	0.0	830	[0.807, 1.023]
20	Avg_Utilization_Ratio	float64	0	0.0	964	[0.846, 0.478]

Figure 1: Dataset variables description

-Total_Relationship_Count: Total number of products held by the customer or client could have other products like debit card, loans, and so on.
-Contacts_Count_12_mon: the number of contacts between the customer and the bank in the last 12 months.
(It could be a key indicator of the satisfaction level of the client: the more contacts, the higher the probability that there is something that causes attrition)

Credit Card utilization features: Months_Inactive: it determines how many months the client has been inactive.
Credit_Limit: this is the maximum amount the client is allowed to use;
Total_Revolving_Bal: the debt amount. For example the revolving balance value in February is determined by:
Debt_February = Debt_January + CreditUsed_February - DebtPaid_February
Avg_Open_To_Buy: It is the average over the last 12 months of the Open To Buy value.
Total_Trans_Amt: total transactional amount in the last 12 months.
Total_Amt_Chng_Q4_Q1: the ratio between transactional amount of first quarter and the same amount for fourth quarter. Hence, a value smaller than 1 means that the customer has spent less in this quarter with respect to the last one.
Total_Trans_Ct, Total_Ct_Chng_Q4_Q1: their meaning is analogous to the last two variables. Of course, thee differ on the underlying reference variable, since in this case it is the number of transactions instead of the amount.

Avg_Utilization_Ratio: It is the average proportion of the credit used with respect to the credit_limit in the last 12 months.

There are two more features:

CLIENTNUM: primary key of the dataset. Attrition_Flag: whether the customer is an attrited one or not. Attrited customer are referred to closed accounts, thus there might be clients near to churn in the dataset but classified as normal clients (this will cause some errors, like we will see below).

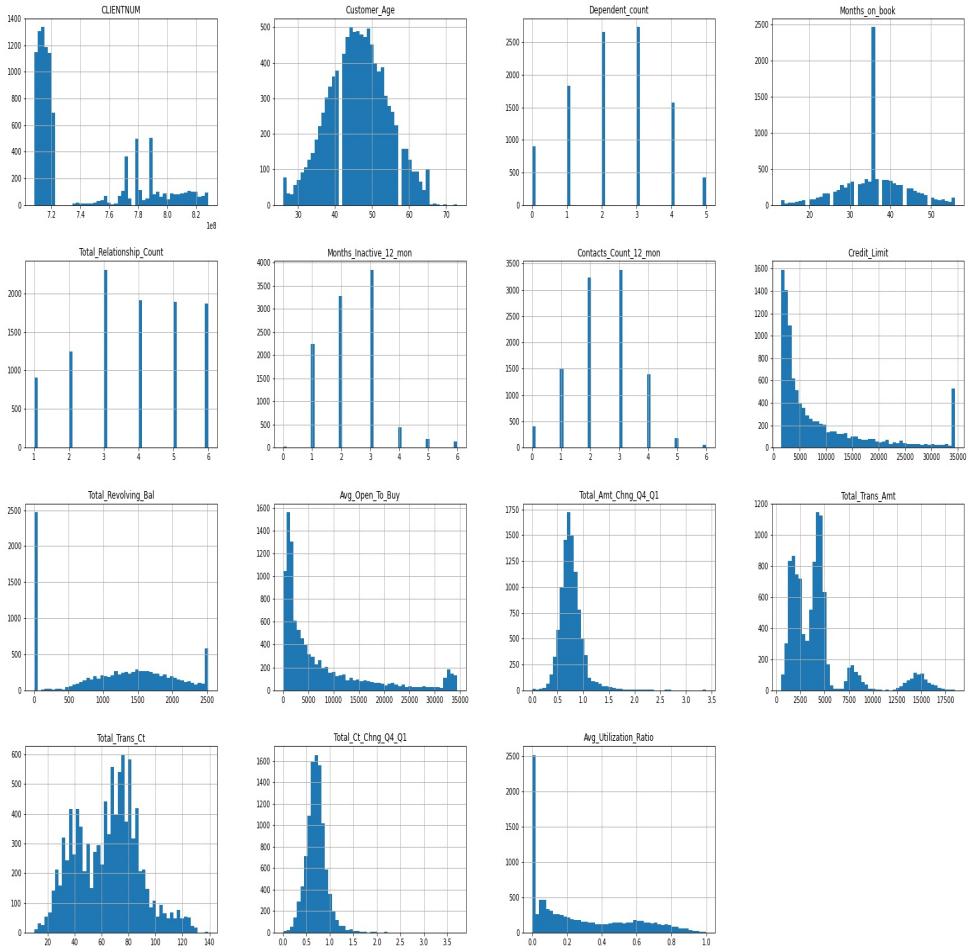


Figure 2: Data distribution plot

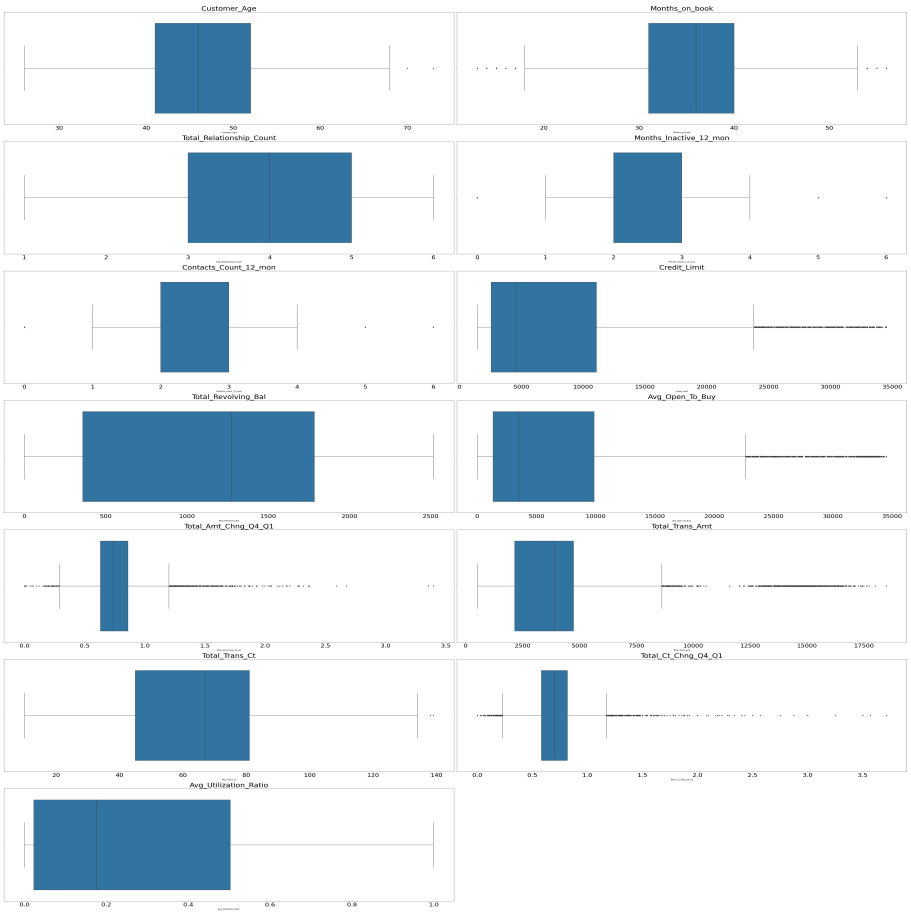


Figure 3: outliers plots of the features

4 Exploratory Data Analysis

We conduct exploratory data analysis (EDA) to have a better understanding of the data by checking for missing and duplicated values, handling outliers, visualizing distributions and plotting graphs to see the relationships between features and our target, which is whether the customer get churned. Here are some important features that needed to illustrate.

We can observe that it is an unbalanced distribution (about 84% of the data belong to the class Existing Customer, whereas only 16% of clients are in the class of churned ones).As we have shown this in the figure 4.

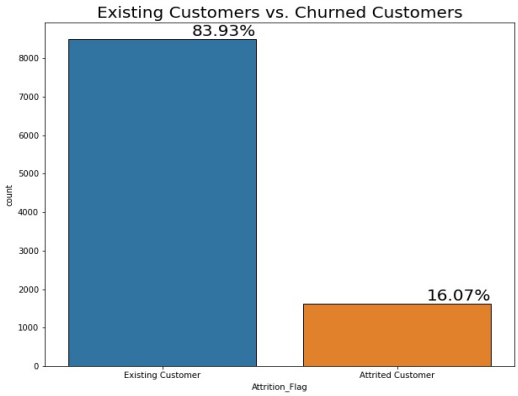
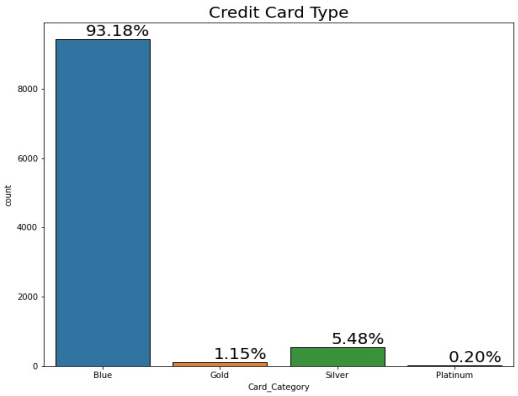


Figure 4: Analysis of total customers

It can be seen from the below table that the type of card held by majority of people is blue card with 93.2%.



4

Figure 5: Analysis on the basis of credit card type

In the figure 5, we split the data into two parts, thus clearly, we can visualize the relationships between card type and both currently existing customer and left customers. These two follows the same pattern, with the amount of blue card holders extremely surpasses the others.

In the figure 6, Even we have analyzed the credit card type correlation of churn customers with their genders.

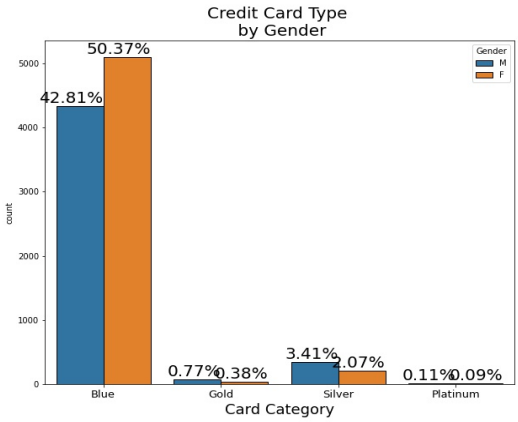


Figure 6: Analysis on the basis of credit card type based on gender

In the figure 7, We have analyzed the churn customers on the basis of gender.

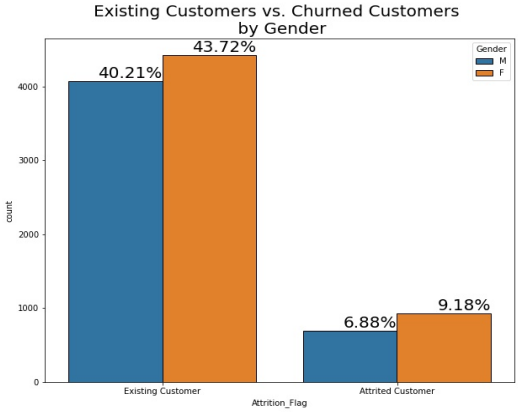


Figure 7: Analysis of customers on the basis of gender

In the figure 8, We have analyzed the percentage of credit use on the basis of the education of the customers.

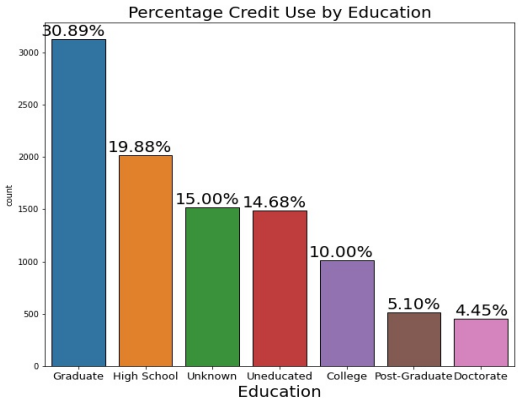


Figure 8: Analysis of percentage credit use by education

In the figure 9, We have analyzed the percentage of credit use by status of the customers.

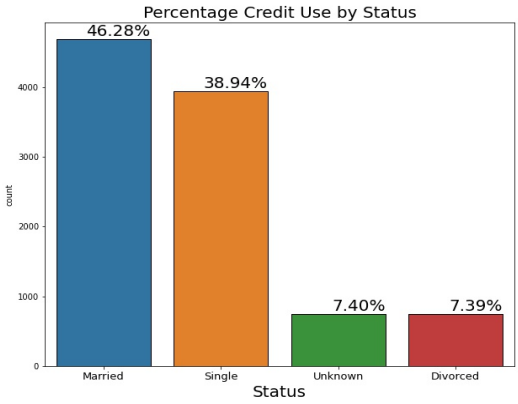


Figure 9: Analysis of percentage credit use by status

5 Data Preprocessing and Model Evaluation

Before we create a machine learning model, first we have to preprocess the data. Preprocessing data is done so that the machine can read our data correctly.

5.1 Missing Values

Missing values will reduce our machine learning model reliability, our model could be biased. That is why we have to handle missing values before creating a machine learning model. There are several ways to handle missing values, including: -Drop data point: This method is not recommended especially if we don't have many data points, it could lead to information loss.
-Fill it with median, mode, or mean. This method is quite simple but use it with precautions and make sure we choose the correct statistics with good reason.

-Fill it with the same value as similar data points.
-Fill it with assumptions based on business knowledge. This is to me the best method, for example, if we have missing values of rating of an item it could be because nobody has rated that item, or maybe nobody has bought that item yet, if we fill the missing values with 0 it would mean that we assume that the item is really bad but if we fill the missing value with 5 it would mean that we assume that the item is really good, so it would be better if we fill the missing value with 3 or 4.
In our case We have checked for the missing values. As in the figure 10, we have shown that there is no missing values in the data.

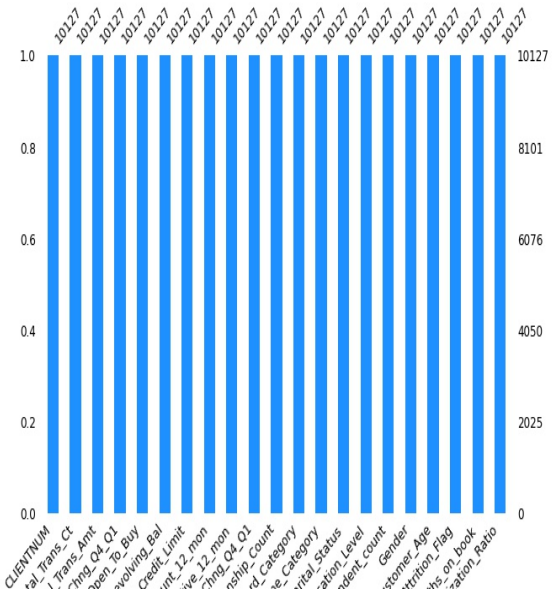


Figure 10: Check Missing values

5.2 Encoding

Feature Encoding Feature encoding is done so that our machine can read our categorical data. So far, machines can only read numbers which is why we have to turn our categorical data into numerical data.

Label encoding: we have used label encoding when the feature has ordinal values for example in our dataset. We have Card_Category, this feature has ordinal values where Blue is the lowest level and Platinum is the highest level. Label encoding can also be used for a categorical feature that only has 2 unique values.

We have manually encode feature in a following way :

```
mapAttrition = ‘{Attrited ’Customer:1, ‘Existing ’Customer:0}
mapGender = ‘{’M:1, ‘’F:0}
mapCard = ‘{’Blue:0, ‘’Silver:1, ‘’Gold:2, ‘’Platinum:3}
dfLabel‘[’Attrition_Flag] = dfLabel‘[’Attrition_Flag].map(mapAttrition)
dfLabel‘[’Gender] = dfLabel‘[’Gender].map(mapGender)
dfLabel‘[’Card_Category] = dfLabel‘[’Card_Category].map(mapCard)
dfLabel‘[[’Attrition_Flag’,’Gender’,’Card_Category]].sample(5)
```

5.3 Correlation matrix

Pearson correlation matrix can give us the information about relationships between features. It can help us do feature selection by removing some highly correlated features in the model training step. Here we plot correlation graphs of categorical features in our dataset.

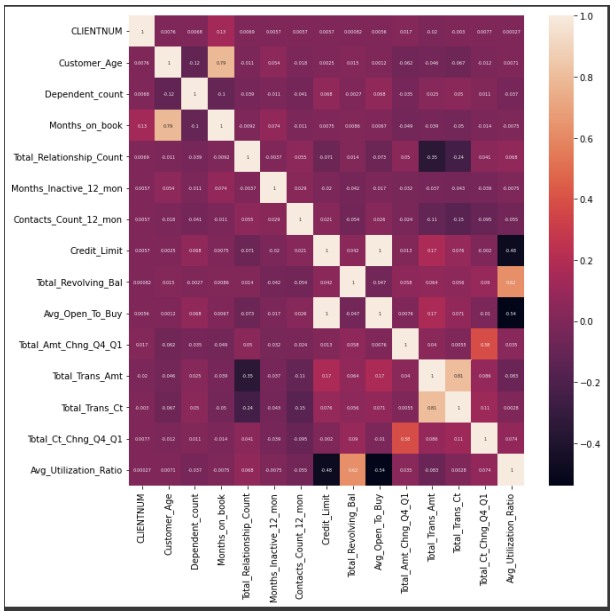


Figure 11: Correlation graph

Values on both the vertical and perpendicular axis are our features, with the data in the middle shows the relationships between the corresponding axes. The color, obviously, reflects the strength of the correlation, and the lighter the color, the higher the positive correlation between two features, while the darker the color, the higher the negative correlation.

5.4 Model Evaluation

5.4.1 Confusion Matrix

The confusion matrix, also called error matrix, is a standard format for the accuracy evaluation and is represented by a matrix of 2 rows and 2 columns as it shown in figure 12.

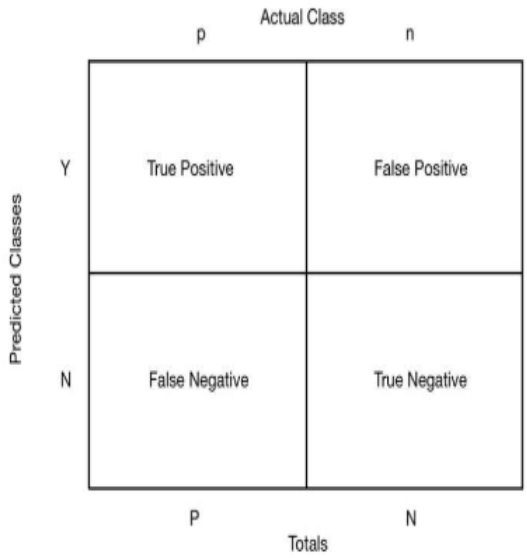


Figure 12: Confusion matrix

Positive and negative refer to the result of model on whether the customer gets churned, while true and false indicate whether our model predicts correct. For example, true positive in the upper left means our model predicts this customer will leave our services and it is true. Similarly, we can define the other three matrices using the same rule. We choose recall ratio to analysis model performances. It demonstrates how many churned customers are successfully predicted.

$$Recall = \frac{Truepositive}{Truepositive + Falsepositive}$$

Since our goal is to predict consumer churn as many as possible, the recall ratio is the best matrix among others. Large proportion reflects high prediction accuracy.

5.4.2 Receiver Operating Characteristics Curve (ROC curve)

Threshold is an important hyperparameter in ROC curve. It is a standard which helps to determine whether a customer will get churned. The result of our model is a numerical number lies in the range of 0 to 1, therefore, those under the threshold will be predicted staying while those surpass will be predicted churned. Since it is a hyperparameter, we can set its value to adjust the strictness of this method. Specifically, the higher the threshold, the more difficult a customer to be assigned to lost customer, the higher the accuracy of this result is. Points on the ROC curve show the true positive rate and false positive rate achieved by specific decision thresholds, and the monotone curve is obtained by scanning all possible decision thresholds, and the area under the curve (AUC) corresponds to the proportion of correctly arranged positive and negative sample pairs . It is defined as that if we want to consider only the optimal threshold, then the higher AUC is, the stronger the predicted ability the model will have.

6 BI Model

6.1 Logistic Regression(LR)

Logistic Regression is a linear model, which connects $X_1, ..., X_p$ to the conditional probability $P(Y = 1|X_1, ..., X_p)$ through this formula:

$$P(Y = 1|X_1, ..., X_p) = \frac{\exp(\beta_0 + \beta_0 X_1 + ... + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_0 X_1 + ... + \beta_p X_p)}$$

$$\beta_0, \beta_1, ..., \beta_p$$

are regression coefficients, which are derived from a method called maximum-likelihood using the dataset.If the probability P is greater than a threshold value, the new instance will then be assigned to Y=1 class accordingly. Otherwise, it will be assigned to Y=0. Usually, the threshold is set to be 0.5 and therefore is so- called Bayes Classifier.

6.2 KNeighbors(KNN)

K Nearest Neighbors takes less time than the other and therefore is a relatively simple model. It conducts predictions straightforward from training set data, by calculating the closest k objects on distance of dataset to the input data, where k is the hyperparameter and can be adjusted to affect the classifier performance, and it then assigns classification based on maximum voted classes out these adjacent classes.

There are many ways to calculate the distance, for example, Euclidean distance and distance Manhattan, with the former one the most popular. The distance d between two points a and b can be calculated through the formula below:

$$d(a, b) = \sqrt{\sum (X_i - \bar{X})^2}$$

The picture below shows the principle of k nearest neighbor.

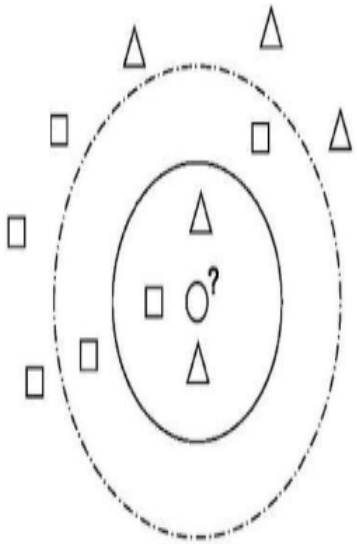


Figure 13: Sketch map of K Nearest Neighbors

6.3 Random Forest(RF) and Decision Tree(DT)

Random forest is a supervised learning model. It was proposed by Breiman and Cutler in 2001, and is based on decision tree and ensemble learning. Decision tree can describe complicated relationships between x and y rather than simple linear relationship, thus has a stronger modeling strength. However, a single tree model is very sensitive to the training set data, therefore is very likely to cause overfitting problem. Ensemble learning, however, can solve this problem by a method called bagging, which is to train multiple learners, with each one's training data comes from a collection of bootstrapped samples selected randomly from original dataset with replacement. It decreases variance by introducing randomness into model framework, making the model more robust and the result more accurate and convincing. Specifically, each tree learns independently from random sub-dataset and sub-features, and the final outcome is drawn with the help of deterministic averaging process, or in other words, the average of predictions of individual trees. A simple example of random forest tree is shown below as figure .

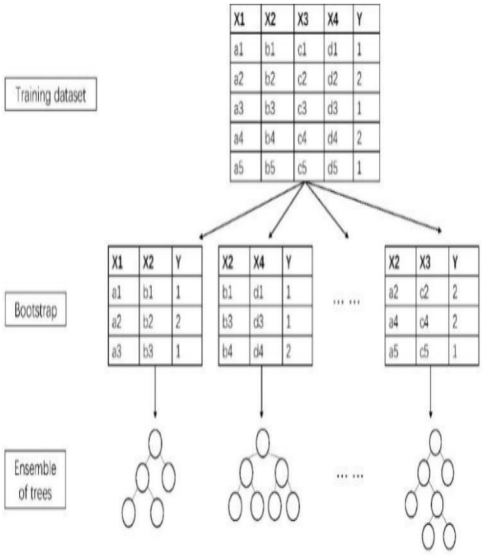


Figure 14: Description of Random Forest. X1 to X4 are all features, with Y is the outcome that needs to be predicted. The original dataset undergoes splits to several sub-sets with each one contains less features and less data. Then, each sub-set will be used to train a tree to make prediction, and deterministic averaging process is applied to determine the final result

6.4 Neural Network(NN)

Neural networks are a set of algorithms modeled loosely after the human brain, that are designed to recognize patterns. A neural net consists of thousands or even millions of simple processing nodes that are densely interconnected. Most of today's neural nets are organized into layers of nodes, and they're "feed-forward," meaning that data moves through them in only one direction. An individual node might be connected to several nodes in the layer beneath it, from which it receives data, and several nodes in the layer above it, to which it sends data.

7 Results

	precision	recall	f1-score	support
Attrited Customer	0.79	0.50	0.61	496
Existing Customer	0.91	0.97	0.94	2543
accuracy			0.90	3039
macro avg	0.85	0.74	0.77	3039
weighted avg	0.89	0.90	0.89	3039

Figure 15: Logistic regression

	precision	recall	f1-score	support
Attrited Customer	0.52	0.28	0.36	496
Existing Customer	0.87	0.95	0.91	2543
accuracy			0.84	3039
macro avg	0.70	0.62	0.64	3039
weighted avg	0.81	0.84	0.82	3039

Figure 16: K-Nearest Neighbour

	precision	recall	f1-score	support
Attrited Customer	0.80	0.78	0.79	496
Existing Customer	0.96	0.96	0.96	2543
accuracy			0.93	3039
macro avg	0.88	0.87	0.88	3039
weighted avg	0.93	0.93	0.93	3039

Figure 17: Decision tree

	precision	recall	f1-score	support
Attrited Customer	0.92	0.77	0.84	496
Existing Customer	0.96	0.99	0.97	2543
accuracy			0.95	3039
macro avg	0.94	0.88	0.91	3039
weighted avg	0.95	0.95	0.95	3039

Figure 18: Random Forest

	precision	recall	f1-score	support
0	0.60	0.89	0.72	496
1	0.98	0.88	0.93	2543
accuracy			0.88	3039
macro avg	0.79	0.89	0.82	3039
weighted avg	0.91	0.88	0.89	3039

Figure 19: Neural network

	model	f1_score	f1_score_Mean	f1_score_Std	roc_auc	roc_auc_Mean	roc_auc_Std
0	Logistic Regression	[0.6649616368286445, 0.6700767263427...	0.644201	0.026226	[0.9383774509884177, 0.9129564556641644, 0.921...	0.921480	0.011112
1	KNN	[0.6353591160220994, 0.6324786324786326, 0.578...	0.613872	0.032480	[0.8919001069478583, 0.8935668606112915, 0.887...	0.886052	0.009502
2	Decision Tree	[0.8286692921348314, 0.75, 0.7899115044247788,...	0.795195	0.031366	[0.8966564722217012, 0.8389290203951443, 0.867...	0.875808	0.023121
3	Random Forest	[0.8909512761020882, 0.8316831683168316, 0.831...	0.858483	0.025571	[0.9881925811959398, 0.9811790532300013, 0.9820...	0.985179	0.003017

Figure 20: Analysis of percentage credit use by status

- To summarize our findings, some of the attributes that played a large part in helping the model predict customer churning included transactions count, total revolving balance, and average utilitzation ratio. This makes sense because customers who use their card less often are less eager to stay with the bank.

- We found that a very large portion of the customer base are Blue cardholders. It would be interesting to answer whether promotional efforts should be directed to the Blue cardholder segment that is volume heavy or toward the Platinum cardholders that is sales heavy. We also found that a large segment of the customer base has a middle class income making less than $40,000 and a large segment also holds a graduate degree.

- As mentioned, we found that the best performing model was the Random Forest model which yielded a 95% accuracy with the highest ROC score. The worst performing model was the K-Nearest-Neighbors Model which yielded a 84% accuracy with the lowest F-score and ROC score out of all 5 models.

8 Conclusion

- We presented how a bank was eager to provide better services to turn customers' decisions in the opposite direction and prevent them from churning. By using a classification model to predict those who are likely to churn and then reacting proactively, we reduce the overwhelming cost of offering special promotions to every customer.

- Ultimately, the best predictor of a customer who is most likely to churn is a customer that uses their card very little. While we can easily segment these customers into a cluster where promotional efforts are issued by the number of transactions, a more comprehensive method would be to identify segments of customers meeting varying criteria and determining the return on investment of promotional efforts toward these segments.

9 References

[1] Xinyu Miao1, Haoran Wang, "Customer Churn Prediction on Credit Card Services using Random Forest Method,"Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)

[2] Trie sony kusumwoibowo, "credit-card-customer-churn-predictive-analytics-", Dev genius(medium,2022)
(https://blog.devgenius.io/credit-card-customer-churn-predictive-analytics-b012ff8c385d)

[3] Jarar Zaidi, " Project: Modeling Predicting of Churning Customers (in R)" Towards data Science 2020

10 PRESENTATION CODE LINK

Presentation Recording Link
Presentation Slides
Python Code