Travel Credit Card Users Churn Prediction

Business Presentation

Contents

Business Problem Overview and Solution Approach

- The Thera bank recently saw a steep decline in the number of users of their credit card, credit cards are a good source of income for banks because of different kinds of fees charged by the banks like annual fees, balance transfer fees, and cash advance fees, late payment fees, foreign transaction fees, and others. Some fees are charged to every user irrespective of usage, while others are charged under specified circumstances.

- Customers' leaving credit cards services would lead bank to loss, so the bank wants to analyze the data of customers and identify the customers who will leave their credit card services and reason for same – so that bank could improve upon those areas.

Business Problem Overview and Solution Approach

- A need is required to come up with a classification model that will help the bank improve their services so that customers do not renounce their credit cards.

- The objective of the model is:
  - Explore and visualize the dataset.
  - Build a classification model to predict if the customer is going to churn or not
  - Optimize the model using appropriate techniques
  - Generate a set of insights and recommendations that will help the bank

greatlearning
Power Ahead

Data Overview

Variable	Description
CLIENTNUM	Client number. Unique identifier for the customer holding the account
Attrition_Flag	Internal event (customer activity) variable - if the account is closed then "Attrited Customer" else "Existing Customer"
Customer_Age	Age in Years
Gender	Gender of the account holder
Dependent_count	Number of dependents
Education_Level	Educational Qualification of the account holder
Marital_Status	Marital Status of the account holder
Income_Category	Annual Income Category of the account holder
Card_Category	Type of Card
Months_on_book	Period of relationship with the bank
Total_Relationship_Count	Total no. of products held by the customer
Months_Inactive_12_mon	No. of months inactive in the last 12 months
Contacts_Count_12_mon	No. of Contacts between the customer and bank in the last 12 months
Credit_Limit	Credit Limit on the Credit Card
Total_Revolving_Bal	The balance that carries over from one month to the next is the revolving balance
Avg_Open_To_Buy	Open to Buy refers to the amount left on the credit card to use (Average of last 12 months)
Total_Trans_Amt	Total Transaction Amount (Last 12 months)
Total_Trans_Ct	Total Transaction Count (Last 12 months)
Total_Ct_Chng_Q4_Q1	Ratio of the total transaction count in 4th quarter and the total transaction count in 1st quarter
Total_Amt_Chng_Q4_Q1	Ratio of the total transaction amount in 4th quarter and the total transaction amount in 1st quarter
Avg_Utilization_Ratio	Represents how much of the available credit the customer spent

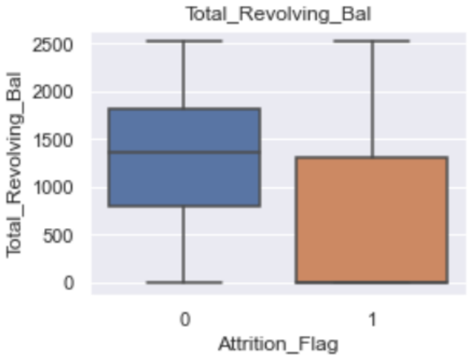
Observations	Variables
10127	21

Note:

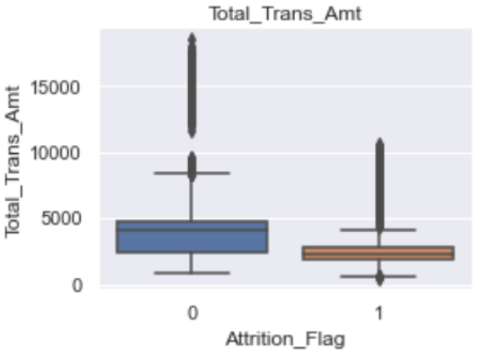
- CLIENTNUM column is removed.
- 'Unknown' categorical values in Education_Level, Marital_Status and Income_Category variables are treated with K-Nearest Neighbours Imputer.
- Income_Category variable values are formatted to be consistent by removing special characters like '+' or '$' and replacing capital K with small caps k.
- Attrition_Flag values changed to '1' for Attrited Customers and '0' for Existing Customers categorical values.
- Outliers are capped at lower and upper whiskers of IQR.
- Avg_Open_To_Buy, Total_Trans_Ct, Customer_Age and Avg_Utilization_Ratio variables are dropped due to potential multicollinearity.

EDA – Total_Revolving_Bal, Total_Trans_Amt, Total_Trans_Ct

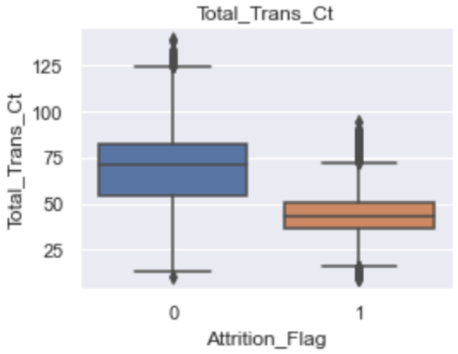
Attrition_Flag vs. Total_Revolving_Bal



Attrition_Flag vs. Total_Trans_Amt



Attrition_Flag vs. Total_Trans_Ct

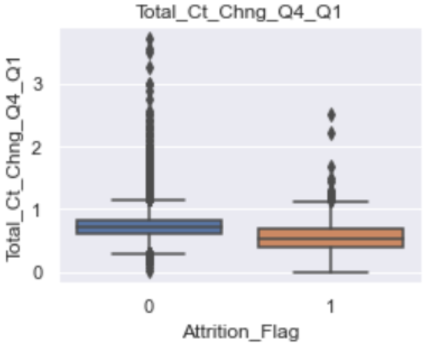


- Attrited customers tended to have a lower Total_Revolving_Bal than existing customers.
- This is in line with attrited customers tending to spend less on the bank's cards.

- Attrited customers displayed a significant lower transaction count and amount in Total_Trans_Amt and Total_Trans_Ct.

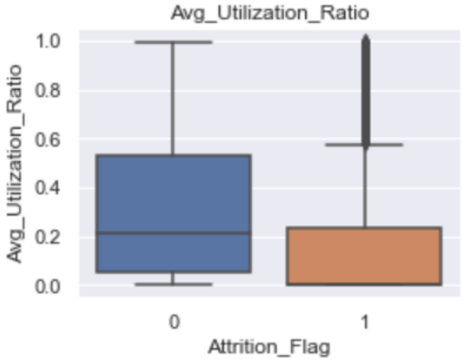
EDA – Total_Ct_Chng_Q4_Q1, Avg_Utilization_Ratio

Attrition_Flag vs. Total_Ct_Chng_Q4_Q1



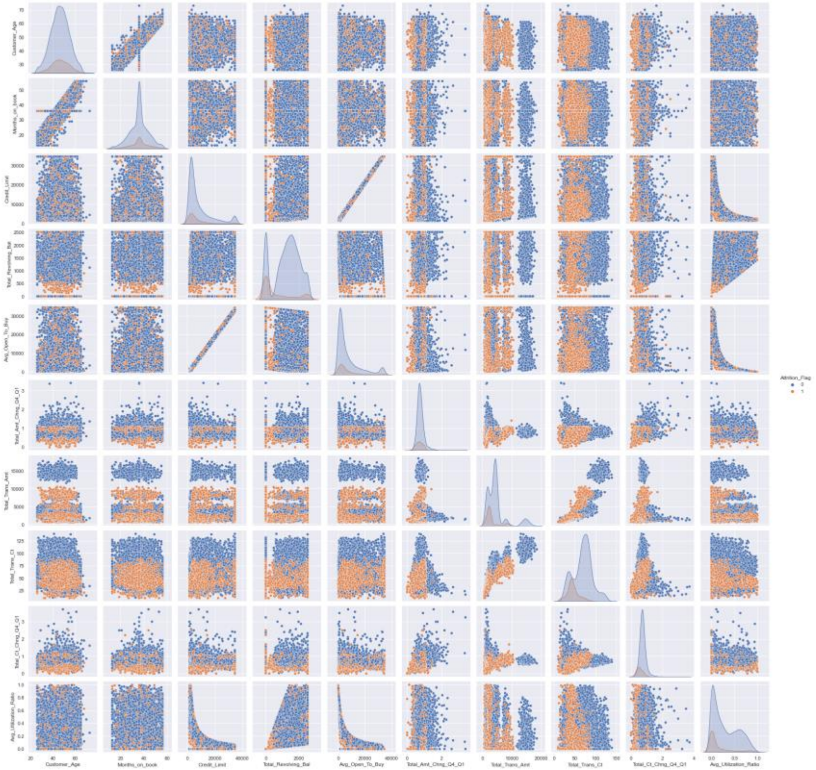
- Attrited customers also displayed a lower ratio of spending counts from Q4 last year to Q1 this year than existing customers in Total_Ct_Chng_Q4_Q1.

Attrition_Flag vs. Avg_Utilization_Ratio



- In line with attrited customers tending to spend less on the bank's cards, it is reinforced in the Avg_Utilization_Ratio where most attrited customers spend less than 0.2 of their credit limit.

EDA – Pairplot

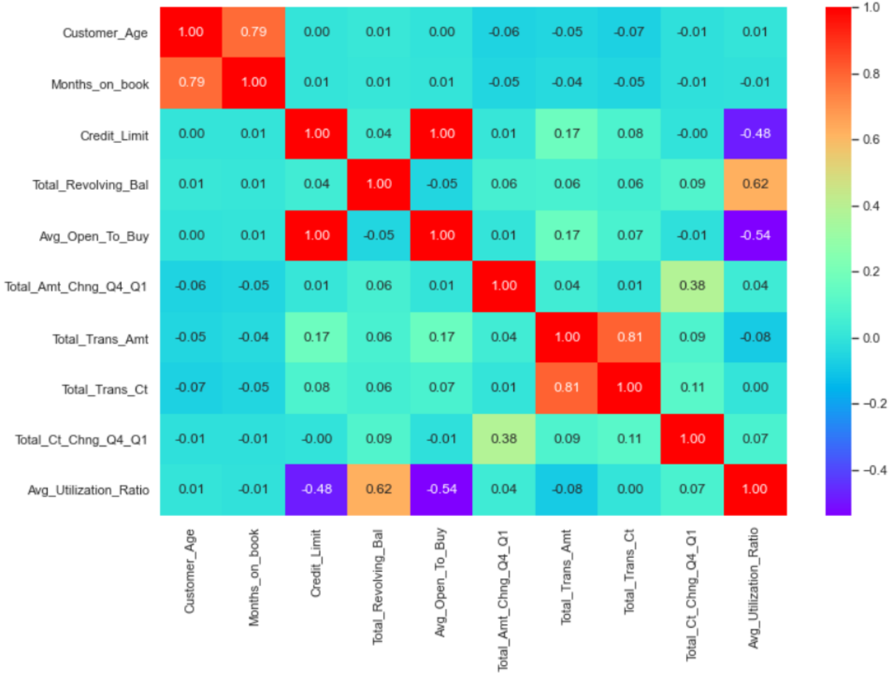


Observations

- From the pairplots, attrited customers display lower Total_Trans_Amt and Total_Trans_Ct which suggests prior to quitting the card services, they start to use the card much lesser and spend lesser on it before giving up the card.

- It can also be derived that many of the attrited customers start to use less of the bank's card services in Q1 vs. the previous quarter, Q4.

EDA – Correlation Matrix

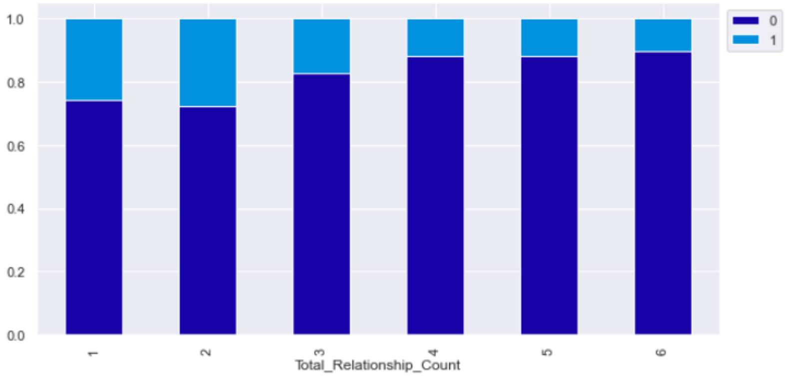


Observations

- Several pairs of variables are found to be highly correlated.
- Avg_Open_To_Buy, Total_Trans_Ct, Customer_Age and Avg_Utilization_Ratio variables are dropped due to potential multicollinearity after experimentation on model building.

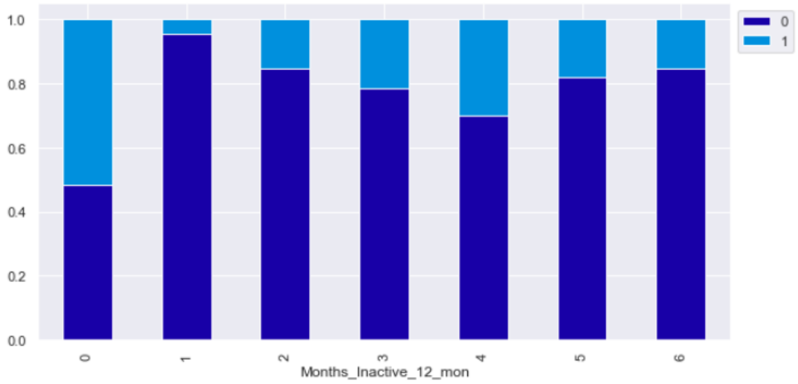
EDA – Total_Relationship_Count, Months_Inactive_12_mon

Attrition_Flag vs. Total_Relationship_Count



- Attrition tend to diminish to the least once customers have more than 3 products with the bank.
- The bank can consider cross selling more products to existing customers to lower the chance of customers attriting.

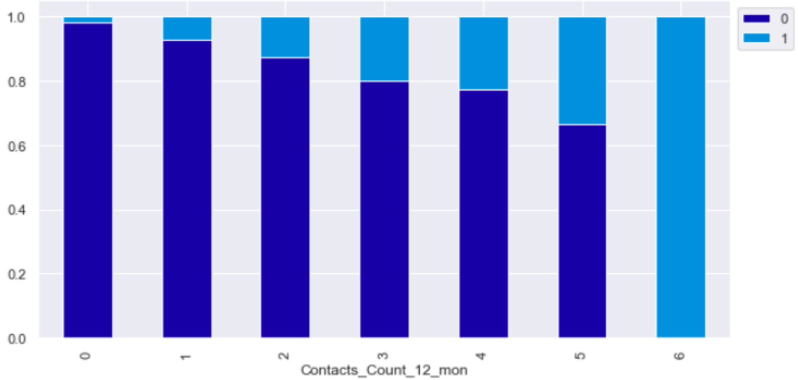
Attrition_Flag vs. Months_Inactive_12_mon



- There is a high ratio of attriting customers who are never inactive with the bank's card services but this can be a once off since the numbers are small (15 customers) and could indicate poaching by competitors currently.
- Over a larger time period, 4 months of inactivity among bank card users seem to have the highest attrition rates. It seems to suggest from 0 to 4 months of inactivity offers an early indication of possible customer attrition.
- The bank can consider offering incentives to encourage customer using their cards again.

EDA – Contacts_Count_12_mon, Card_Category

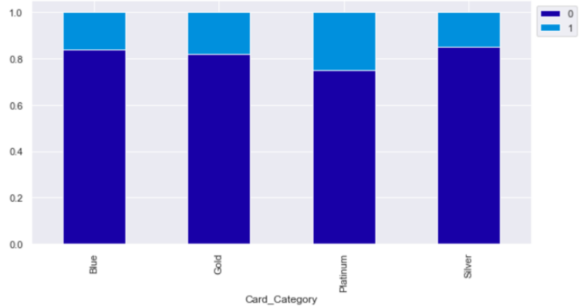
Attrition_Flag vs. Contacts_Count_12_mon



- It seems that the contacts between customer and bank is of the complaint nature as those with no contact have negligible attrition rates while attrition rate climbs with more customer contacts with the bank.
- 6 contacts in the last 12 months offers a 100% attrition rate.
- The bank should look to improve its card services and reduce complaints possibly indicating higher attrition chances in the future.

Attrition_Flag vs. Card_Category

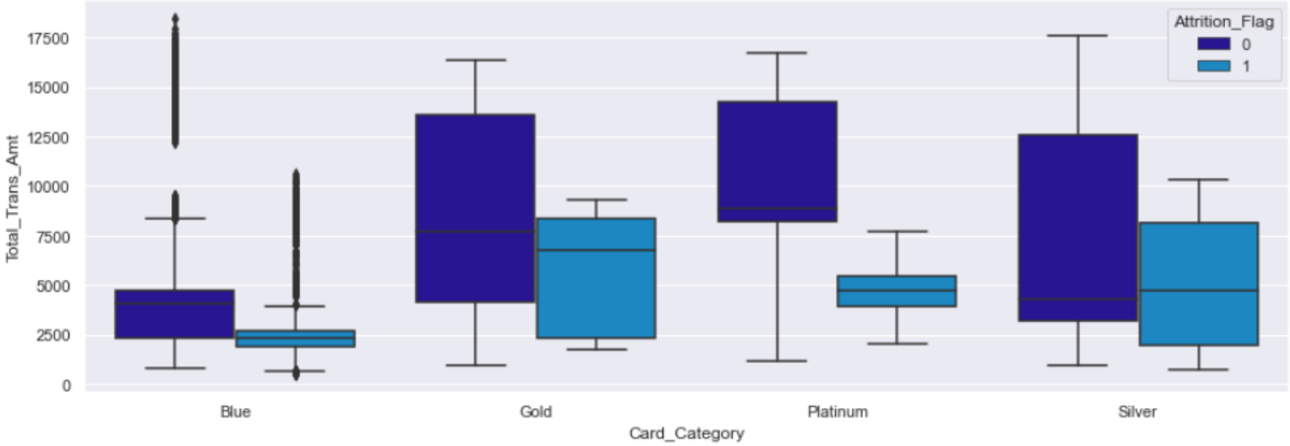
Attrition_Flag	0	1	All
Card_Category			
Blue	7917	1519	9436
Gold	95	21	116
Platinum	15	5	20
Silver	473	82	555
All	8500	1627	10127



- Platinum cards seem to be attriting at a higher rate than other cards even though their numbers are small.
- However they might be the bank's higher profit yielding products and preventing them from churning helps in keeping up profit margin.

EDA – Contacts_Count_12_mon, Card_Category

Attrition_Flag vs. Card_Category vs. Total_Trans_Amt



- It appears that all attriting customers of the card categories spend lower in total than customers who don't.
- This is especially more telling among platinum card holders who also on average spend higher than other card categories.

Model Performance Summary – Evaluation Criterion

- Model Evaluation Criterion
  - Model can make wrong predictions as:
    - False Positive: Predicting a customer is going to be churned but actually not so.
    - False Negative: Predicting a customer is staying but actually churning.

- Which case is more important?
  - Both the cases are important as:
  - If we predict a customer is going to be churned but actually not so then the wrong reasons for attrition will be derived.
  - If we predict a customer is not going to be churned but actually customer is going to be churned, then no effort will be spent on improving the services to the customer profiles to persuade them to stay with the bank's card services.

Model Performance Summary – Evaluation Criterion

- How to reduce losses?
  - We can use accuracy but since the data is imbalanced it would not be the right metric to check the model performance.
  - Therefore, f1_score should be maximized, the greater the f1_score higher the chances of identifying both the classes correctly.

Model Performance Summary – Approach

- The data set is split into 60% Training, 20% Validation and 20% Test.
  - Training data for model building
  - Validation data for evaluation in model building
  - Test data for final model evaluation

- Different models were constructed for comparison:
  - Logistic Regression, Logistic Regression with Oversampling, Logistic Regression with Regularization, Logistic Regression with Undersampling, DecisionTreeClassifier, BaggingClassifier, RandomForestClassifier, AdaBoostClassifier, GradientBoostingClassifier, XGBClassifier

Model Performance Summary – Approach

- Of these, the top 3 models were hyperparameter tuned to further improve on it:
  - AdaBoostClassifier with RandomSearchCV , AdaBoostClassifier with GridSearchCV, XGBClassifier with RandomSearchCV, XGBClassifier with GridSearchCV, GradientBoostingClassifier with RandomSearchCV, GradientBoostingClassifier with GridSearchCV

- Final evaluation is done on model with different metrics and confusion matrix with test dataset. F1-score is the main metric to differentiate the models.

Model Performance Summary – Performance Metrics

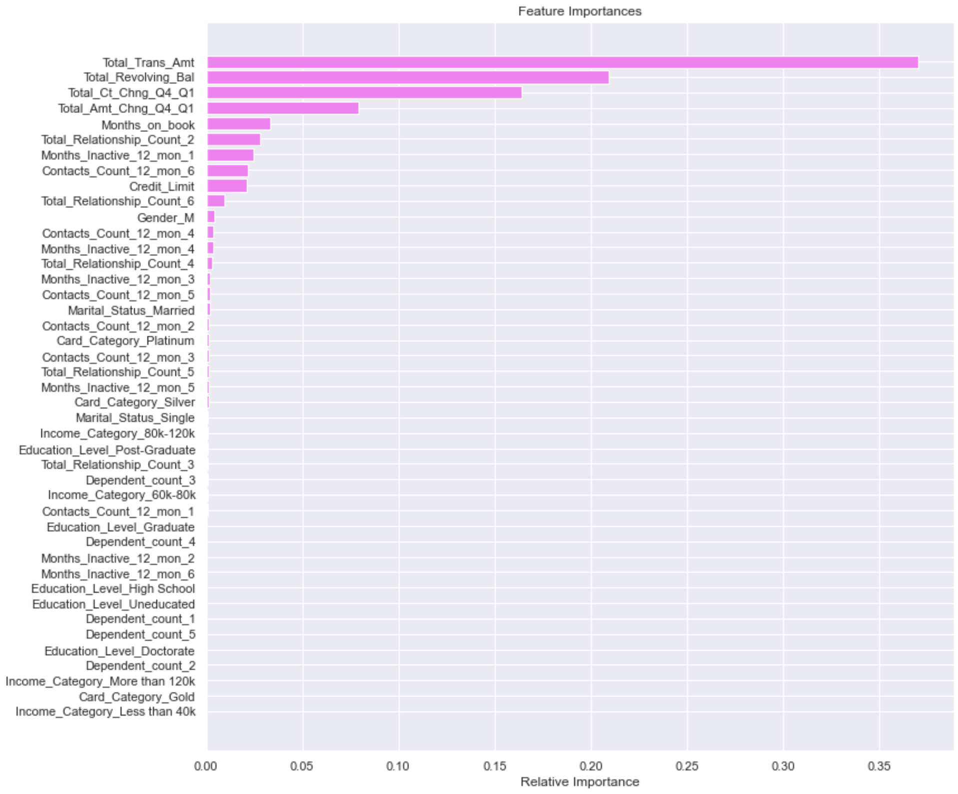
	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision	Train_F1-Score	Test_F1-Score
13	XGBoost GridSearchCV	0.999177	0.961007	1.000000	0.876923	0.994903	0.879630	0.997445	0.878274
9	XG Boost	1.000000	0.961500	1.000000	0.840000	1.000000	0.913043	1.000000	0.875000
12	XGBoost RandomSearchCV	0.998848	0.958539	1.000000	0.864615	0.992879	0.875389	0.996427	0.869969
15	GradientBoost GridSearchCV	0.976790	0.955577	0.897541	0.815385	0.955289	0.898305	0.925515	0.854839
14	GradientBoost RandomSearchCV	0.977778	0.955577	0.901639	0.812308	0.957563	0.901024	0.928760	0.854369
10	AdaBoost RandomSearchCV	0.979588	0.950642	0.920082	0.812308	0.951271	0.871287	0.935417	0.840764
11	AdaBoost GridSearchCV	0.979753	0.950642	0.923156	0.809231	0.949420	0.873754	0.936104	0.840256
8	Gradient Boost	0.956214	0.942744	0.788934	0.738462	0.927711	0.885609	0.852713	0.805369
5	Bagging	0.993745	0.933860	0.961066	0.698462	1.000000	0.863118	0.980146	0.772109
7	AdaBoost	0.929547	0.924975	0.686475	0.670769	0.845960	0.828897	0.757919	0.741497
4	Decision Tree	1.000000	0.913623	1.000000	0.713846	1.000000	0.738854	1.000000	0.726135
6	Random Forest	0.999835	0.924482	0.998975	0.563077	1.000000	0.943299	0.999487	0.705202
0	Logistic Regression	0.880988	0.878578	0.347336	0.323077	0.797647	0.801527	0.483940	0.460526
2	Logistic Regression Regularization	0.747325	0.735933	0.613730	0.593846	0.340922	0.323826	0.438346	0.419110
3	Logistic Regression Undersampling	0.690206	0.669793	0.676230	0.646154	0.296496	0.274869	0.412242	0.385675
1	Logistic Regression Oversampling	0.804609	0.801086	0.311475	0.310769	0.371184	0.360714	0.338719	0.333884

Observations

- The top 3 models by f1-score posed a tendency to drift to overfitting even though their f1-scores on test data, which is unseen data till the evaluation now, are about **87%.**

- The 4th model by ranking ,'Gradient Boost GridSearchCV', offers better model stability with **85.48%** score on test data and just **92.55%** on training data. This model will generalize better on future unseen data given its training and testing scores are quite high and are quite close to each other at **7.07%** points.

Model Performance Summary – Gradient Boost GridSearchCV Classifier Model Tuned Feature Importance



Observations

- In the Gradient Boost GridSearchCV model, it can be seen that Total_Trans_Amt variable is the most important in determining customer churn followed Total_Revolving_Bal and Total_Ct_Chng_Q4_Q1.

- What this means is total card expenditure, total card balance brought over from one month to the next and change in transaction count will give clear indication whether an individual customer will churn.

Business Insights and Recommendations

- Our analysis shows that customers in danger of attrition exhibited lower transaction amounts on their cards than other customers in the last 12 months. The bank can look into extending low usage card holders with merchant tie up discounts, fee waivers or rebates to encourage more spending with the bank's cards.

- Customers with less credit balance per month to roll over in Total_Revolving_Bal are more likely to churn. The bank can consider extending lower interest or interest free offers to customers with good credit to entice them to stay on.

- Customers who showed a much lower ratio of spending and transactions in Q1 compared to last year Q4 tend to be churned. Customers who exhibited this behavior should be actively engaged to retain them either with spending rebates or discounts.

Business Insights and Recommendations

- Customers are less likely to be churned when they have 3 or more product services with the bank. The bank can look into cross selling more products to existing customers to lower the chance of customers churning.

- Customers were also found to be more likely to churn and increasing starting from 0 – 4 months of inactivity. The bank can consider offering incentives to encourage customer utilizing their card services again. There is a one off anomaly of a higher ratio of attrition among customers with no months of inactivity. The suspicion is likely from a competitor's promotion campaign and the bank should look at how to retain and attract customers into its card services.

- Customers who raised more issues with the bank through more contacts are more likely to leave indicating unhappiness with the services. The bank should look into improving its card services to improve customer satisfaction through analyzing customer feedback thus mitigating the risk of customer attrition.

Business Insights and Recommendations

- Comments on additional data sources for model improvement
  - Additional data can be obtained from measured feedback of the data set a few months later, after corrective actions have been taken, in order to evaluate and strengthen the model in production.
  - Qualitative customer complaints can be grouped and analyzed to further target customer pain points and aspirations.

- Model implementation in real world and potential business benefits from model
  - The model implemented in the real world will help to reduce the cost of customer attrition and marketing efforts to attract new customers. Done right, the bank can target and intervene the right customer segments before customers decide to leave the bank's card services This will increase revenue and reduce both variable marketing costs and opportunity costs.

Business Insights and Recommendations

- Other Recommendations
  - Platinum card holders have the highest attrition rate among card categories. Though small in numbers, they can represent higher profit margin product stream for the bank and their customer base should be expanded and existing customers be interviewed to remove any concerns on the product.

  - A significant majority of customers are with the blue card of the bank instead of signing up on other card categories. The bank should explore expanding and retaining the customer base into these other card categories that can help in raising revenue and profit margin.

greatlearning

Power Ahead

Happy Learning !