Machine Learning Application and Report

Module Code: CIS4035-N

Credit Card Users Churn Prediction

Classifying Customer's Behaviour: Attrite vs Not Attrite

MSc Data Science

Name: Chinedu Emmanuel Agwunobi

Student ID: B1555134

Table of Contents

1 Abstract

The gradual but consistent decrease in the number of customers retained over time is referred to as "customer churn," and it is a word that is frequently used in the business and financial sectors (Agarwal et al, 2022). Customer churn prediction models aim to detect customers with a high propensity to attrite. An accurate model permits to correctly target future churners in a retention marketing campaign, while a comprehensible and intuitive rule-set allows to identify the main drivers (feature importance) for customers to churn, and to develop an effective retention strategy in accordance with domain knowledge (Verbeke et al, 2011). The objective of this article is to train machine learning algorithms that will perform well to identify banking customers who may be considering switching financial institutions.

2 Introduction

In the era of advancement of technology and businesses, the market has become competitive and challenging for credit card companies in retaining their customers. One of the major reasons for customer attrition or churn is dissatisfaction with the services provided by the credit card company. Predicting customer churn in advance can help credit card companies take proactive measures to retain their customers and improve customer satisfaction. A small improvement in customer retention hence can lead to a significant increase in profit (Van den Poel and Larivière, 2004). That is why both accurate and comprehensible churn prediction models are needed, in order to identify respectively the customers who are about to churn and their reasons to do so (Verbeke et al, 2011). Predicting a customer will attrite and the customer doesn't attrite or Predicting a customer will not attrite and the customer attrite, which is more important to a financial institution?

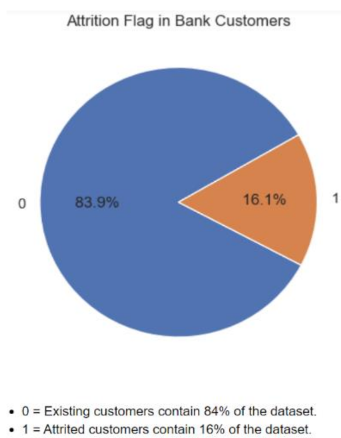


Fig.2. Percentages of Existing and Attrited Customers

By leveraging the power of machine learning, this research can help credit card companies to answer the questions above and improve their customer retention strategies and ultimately enhance their profitability.

3 Methodology

3.1 Dataset Description

The data was obtained from [Kaggle] an online open source platform that hosts data science and machine learning competitions. The dataset, [BankChurners] consist of 23 columns with 10127 entries, and downloaded as CSV file.

	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Level	Marital_Status	Income_Category	Card_Category	Months_on_book
0	768805383	Existing Customer	45	M	3	High School	Married	60K−80K	Blue	39
1	818770008	Existing Customer	49	F	5	Graduate	Single	Less than $40K	Blue	44
2	713982108	Existing Customer	51	M	3	Graduate	Married	80K−120K	Blue	36
3	769911858	Existing Customer	40	F	4	High School	Unknown	Less than $40K	Blue	34
4	709106358	Existing Customer	40	M	3	Uneducated	Married	60K−80K	Blue	21

Fig.2. BankChurners

3.2 Data Preprocessing

The dataset was imported as a CSV file. No duplicate, no missing values, unknown values were converted to "nan" and imputed back using mode in SimpleImputer. These 3 columns were dropped; "CLIENTNUM", "Naive_Bayes_1", "Naive_Bayes_2".

3.3 Exploratory Data Analysis (EDA)

The exploratory data analysis was carried out in two formats; univariate and bivariate analyses to investigate the dataset and derive insights.

3.3.1 Univariate Analysis

3.3.1.1 Customer_Age and Month_on_Book

Fig.3. Customer_Age Distribution

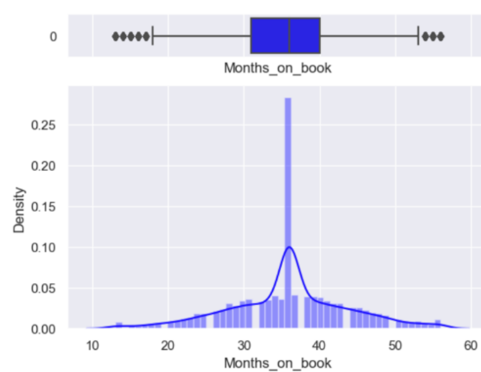


Fig.4. Month_on_Book Distribution

- The distribution of Customer_Age is normally distributed with mean and median at 46 years.
- Most customers are with the bank between the range of 30 - 40 months, which is approximately for 3 years.

3.3.1.2 Credit_Limit and Total_Revolving_Bal

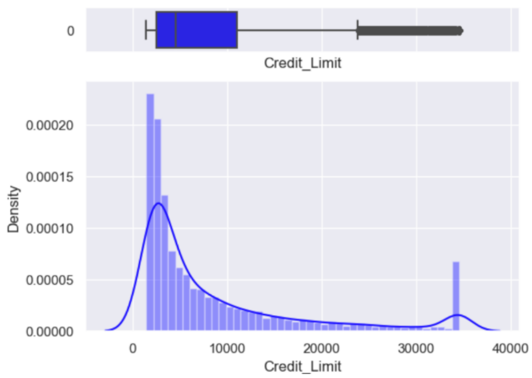


Fig.5. Credit_Limit Distribution

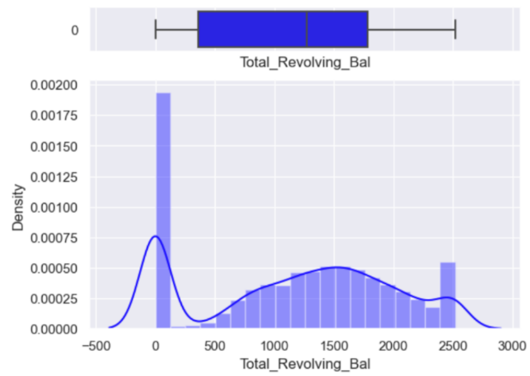


Fig.6. Total_Revolving_Bal Distribution

- 50% of the customers of the bank have a credit limit of less than <5000.
- Most customers pay the complete dues of credit card and have 0 revolving balance.

3.3.1.3 Total_Trans_Amt and Avg_Utilization_Ratio

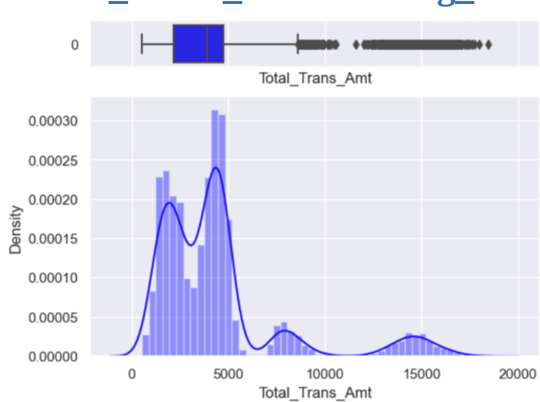


Fig.7. Total Transaction Distribution

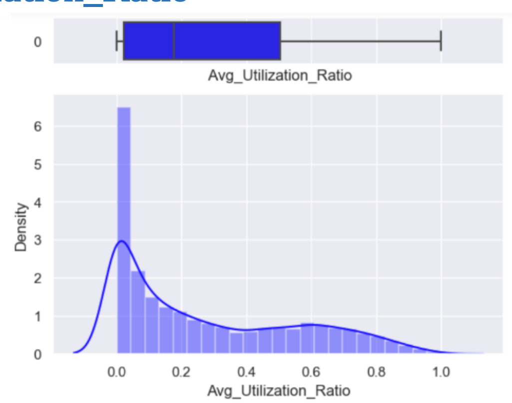


Fig.8. Average Utilization Distribution

- There are two peaks in data at total transaction amounts of one around 2500 and the second around the mean value of ~4500.
- From the boxplot in Total_Trans_Amt, we can see that there are outliers - customers with more than ~8000 total transaction amounts are being considered as outliers.

Observations:

- Open to buy means how much credit a customer is left with
  - Low values of Open to buy could represent either
    - Customers have low credit limits
    - Customers are spending a lot so they are left less open to buy
- Average utilization ratio = (1 - (open to buy/credit limit))

- Low values of the Average utilization ratio represents
    - (Open to buy/credit limit) is nearly equal to 1 -> Open to buy is nearly equal to the credit limit -> customers are spending less using their credit cards
- Credit limit is also right-skewed which represents - most of the customers have low credit limits

- Looking at the 3 variables, we can conclude that most of the customers have low credit limits and are not utilizing their credit cards much.

3.3.2 Bivariate Analysis

Here we compared the Attrition_Flag with few other columns.

3.3.2.1 Heatmap

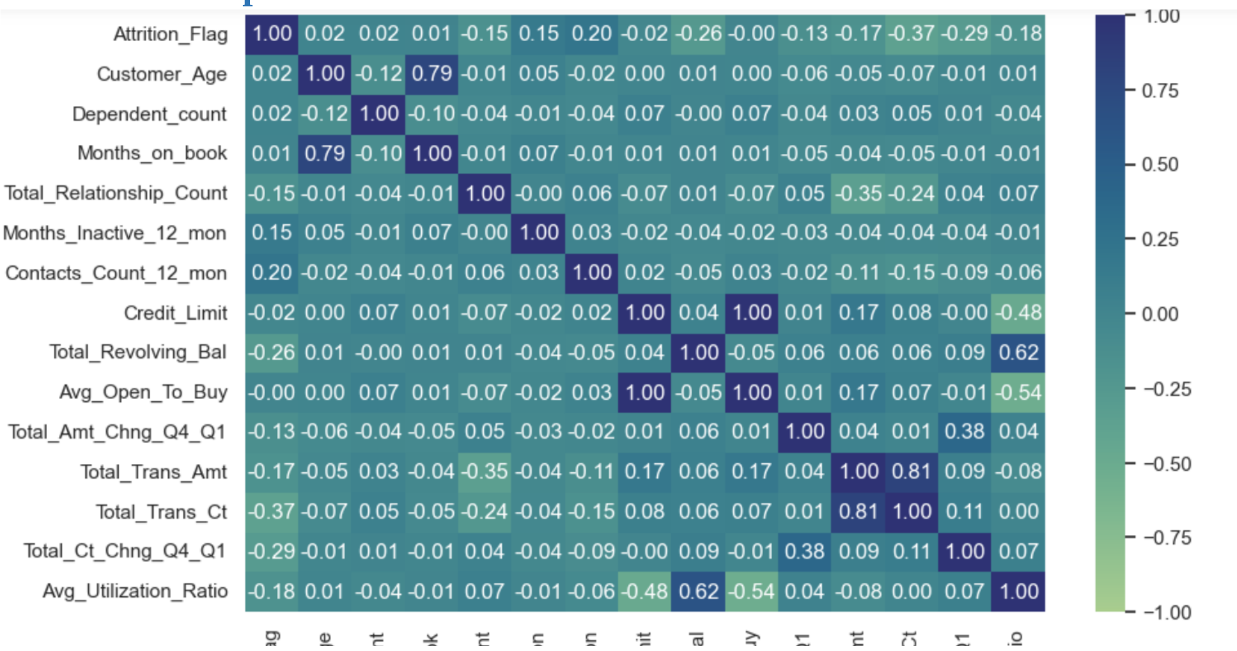


Fig.9. Correlation_Chart

Observations

- Attrition_Flag shows a bit of a negative correlation with Total_Trans_Ct (total transactions) and Total_Trans_Amt (total transaction amount).
- There's a strong positive correlation between Months_on_book and Customer_Age, Total_Revolving_Bal and Avg_Utilization_Ratio, Total_Trans_Amt and Total_Trans_Ct.

3.3.2.2 Attrition_Flag vs Contacts_Count_12_mon

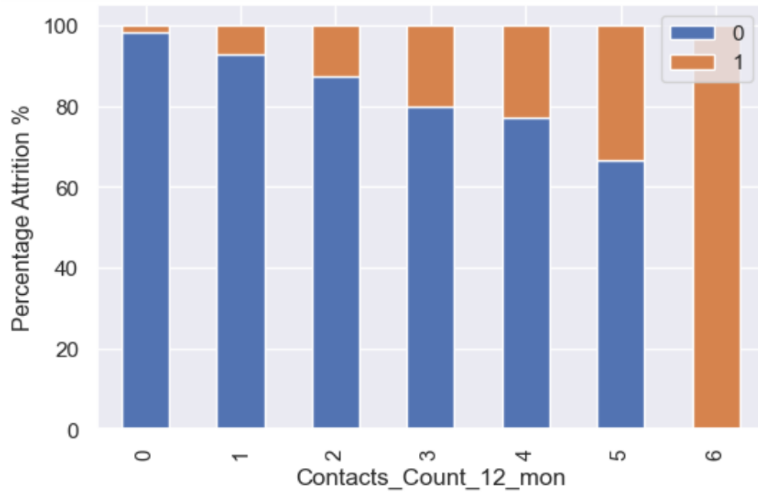


Fig.10. Contacts_Counts

- Highest attrition is among the customers who interacted the most with the bank, probably due to unresolved issues.

3.3.2.3 Attrition_Flag vs Months_Inactive_12_mon

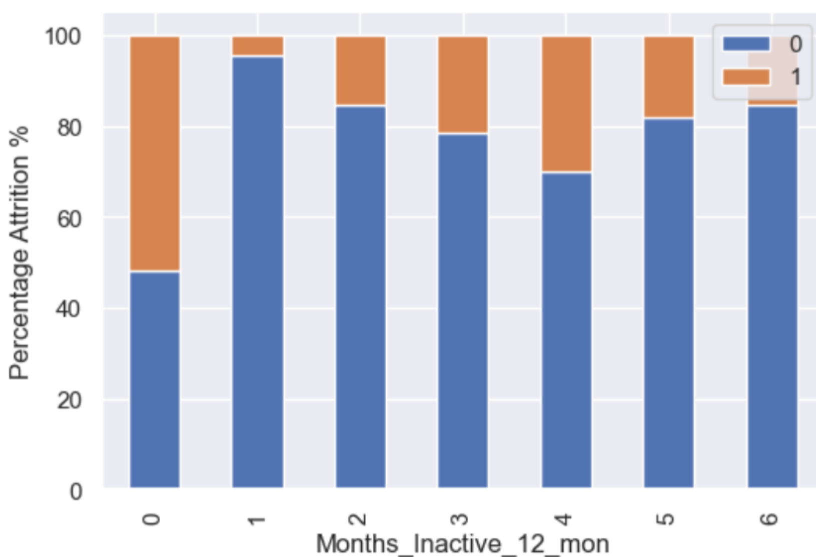


Fig.11. Inactivity of Customers with the Bank

- As inactivity increases attrition also increases (2-4 months)
- The interpretation from here for 0 months and 6 months is difficult as customers who recently used the card attrited the most while those who were inactive for 6 months attrited less.

3.3.2.4 Attrition_Flag vs Total_Relationship_Count

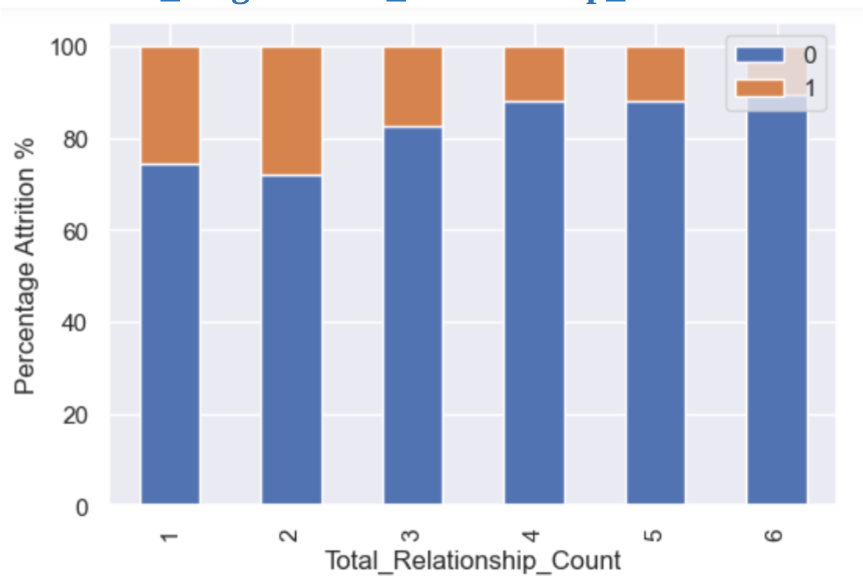


Fig.12. Relationship_counts

- Attrition is highest among the customers who are using 1 or 2 products offered by the bank - together they constitute ~55% of the attrition.
- Customers who use more than 3 products are the ones least attiring, such customers might be more financially stable and actively invest in different services provided by the bank.

3.4 Experiments

We observed that the dataset has few outliners during the data preprocessing and the "nan" values were treated using the "most frequent(mode)" in SimpleImputer. Further preprocessing like handling imbalanced dataset due to the skewness with the classes and normalization were not handled. The normalization was only used with the SVM model to handle the non-linearity of the dataset.

3.4.1 Split Data

The dataset is divided into training set and testing set using train_test_split. The training set consists of 70% of data and the remaining 30% data is kept for testing.

3.4.2 Encoding Categorical Variables

Many of the columns in the dataset contain categorical values in various ways. The Card category, for example, contains four subcategories: blue, silver, gold, and platinum (Panduro-Ramirez et al, 2022). We used get_dummies to encode the categorical variables so that the model will understand and extract valuable information.

Gender_M	Education_Level_Doctorate	Education_Level_Graduate	Education_Level_High School	Education_Level_Post-Graduate	Education_Level_Uneducated	Marital_Status_Ma
1	0	0	0	1	0	
0	0	0	1	0	0	
1	0	0	0	0	0	
0	0	1	0	0	0	
0	0	0	1	0	0	

Fig.13. Encoded_Variables

3.4.3 Building the Models

A model can make wrong predictions as:

- Predicting a customer will attrite and the customer doesn't attrite.
- Predicting a customer will not attrite and the customer attrite.

Predicting that customer will not attrite but he attrite, implies losing on a valuable customer or asset.

How to reduce this loss is by reducing False Negatives?

Bank would want `**Recall**` to be maximized, greater the Recall higher the chances of minimizing false negatives. Hence, the focus should be on increasing Recall or minimizing the false negatives or in other words identifying the true positives (i.e. Class 1) so that the bank can retain their valuable customers by identifying the customers who are at risk of attrition.

Checking Model Performance

The reported average includes the macro average which averages the unweighted mean per label, and the weighted average or the averaging the support-weighted mean per label.

In classification, the class of interest is considered the positive class. Here, the class of interest is 1, that is identifying the customers who are at risk of attrition.

Reading the confusion matrix (clockwise):

True Negative ((TN) Actual=0, Predicted=0): Model predicts that a customer would not attrite and the customer does not attrite

False Positive ((FP) Actual=0, Predicted=1): Model predicts that a customer would attrite but the customer does not attrite

False Negative ((FN) Actual=1, Predicted=0): Model predicts that a customer would not attrite but the customer attrite.

True Positive ((TP) Actual=1, Predicted=1): Model predicts that a customer would attrite and the customer actually attrite.

3.4.3.1 Confusion Matrix

Confusion matrix is a standard output form of classification (Jialu et al, 1998). It gives the correspondent relation between the predicted category and the inherent category of samples (Xiong, 2012).

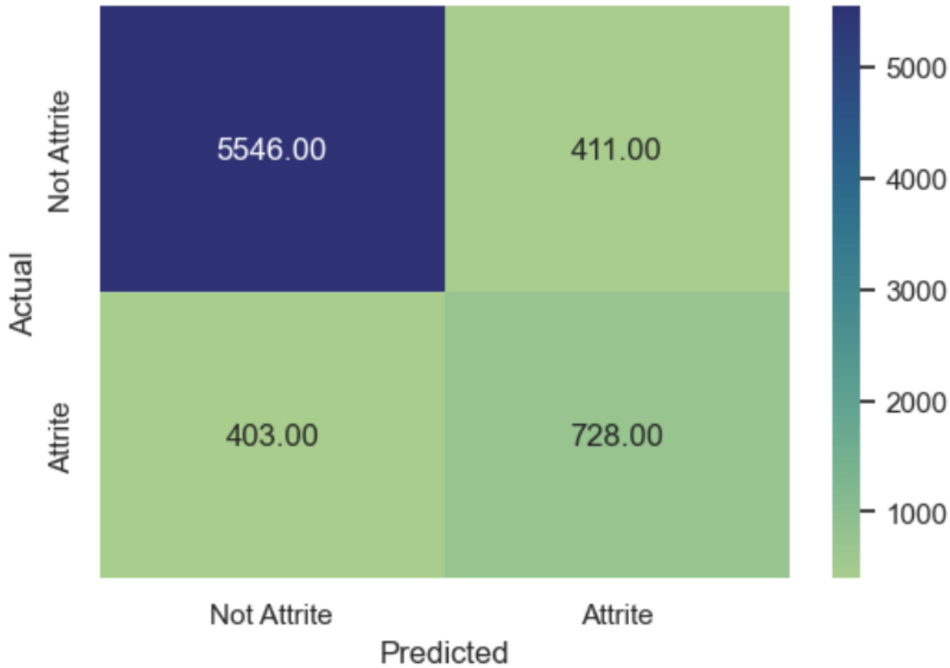


Fig.14. Confusion Matrix: Actual vs Predicted

Here confusion matrix will be used to evaluate our predictions.

3.4.3.2 Accuracy

Accuracy measures the overall correctness of the model's predictions.

$$Accuracy = \frac{Correct\ Predictions}{Total\ Predictions}$$

3.4.3.3 Precision

This metric measures the ability to correctly detect or classify cases belonging to the positive class. The higher the ratio, the better the precision of the classification model (Oduro, Yu and Huang, 2022).

$$Precision = \frac{TP}{TP + FP}$$

3.4.3.4 Recall

This metric specifies the number of positive cases correctly predicted from the total number of positive cases.

$$Recall = \frac{TP}{TP + FN}$$

Here we will utilize recall to justify the actual and predicted by optimizing the model to minimize the false negatives.

3.4.3.5 F1 Score

F1 score is the harmonic mean of precision and recall, and it provides a single metric that balances both precision and recall.

$$F1\ Score = 2\ x\ \frac{Precision\ x\ Recall}{Precision\ x\ Recall}$$

F1 score is a suitable metric to also use to evaluate imbalanced data. F1 score, precision and recall are rated from 0(bad) to 1(good).

3.4.3.6 Hyper-Parameter Tuning

Hyper-parameter tuning was utilized to optimize the model during training to improve performance. Hyper-parameter tuning plays a vital role in the optimal performance of any machine learning algorithm (Probst, Boulesteix and Bischl, 2019).

3.4.3.7 Cross-validation

Cross validation of type k-fold (10 n-split) was used to improve model performance and reduce overfitting.

3.4.4 Model Selection

Building a churn prediction model is the key objective of this research. Most of the time, specialists train multiple models, tune, evaluate, and test them to find the one that finds potential churners with the level of accuracy they want on training data (Manoj, Bharath and Mudhol, 2022).

My motivation is based on the work of (Bhujbal and Bavdane, 2021) and (Manoj, Bharath and Mudhol, 2022), both utilized the same dataset, similar models and related topics.

The machine learning models often used to predict customer attrition that were utilized here are; Logistic Regression, Gaussian Naïve Bayes', Decision Tree, and Support Vector Machine (SVM).

3.4.4.1 Logistic Regression

A classification algorithm that works well for binary classification problems. It models the probability of a sample belonging to a particular class using a logistic function. It measures the relationship between a dependent variable and one or more independent variables to figure out how likely an event is to happen (features) (Manoj, Bharath and Mudhol, 2022). Logistic regression is often used as a baseline model to compare the performance of more complex algorithms.

3.4.4.2 Gaussian Naïve Bayes

This is a probabilistic classification algorithm that is based on Bayes' theorem. It assumes that the features are independent of each other and that the probability distribution of each feature is Gaussian. Gaussian Naive Bayes is a fast and simple algorithm that works well for high-dimensional datasets.

3.4.4.3 Decision Tree

Decision Trees are tree-shaped structures representing sets of decisions capable of generating classification rules for a specific dataset (Nie et al, 2011). The idea behind this algorithm is simple: it splits the training set into two subsets based on Gini index that produces the purest subsets which is given by the equation below:

$$Gini = 1 - \sum_{i=1}^{class} p(i|t)^2$$

where p stands for the probability of finding the i^{th} class after a node (Wang, Nguyen and Nguyen, 2020).

3.4.4.4 Support Vector Machine

A classification algorithm that works by finding the best hyperplane that separates the data into different classes. SVMs can handle both linear and nonlinear data by using kernel functions. They are often used in problems with high-dimensional data and can handle noisy data well.

4 Results

4.1 Logistic Regression

	precision	recall	f1-score	support
0	0.93	0.93	0.93	5957
1	0.64	0.64	0.64	1131
accuracy			0.89	7088
macro avg	0.79	0.79	0.79	7088
weighted avg	0.89	0.89	0.89	7088

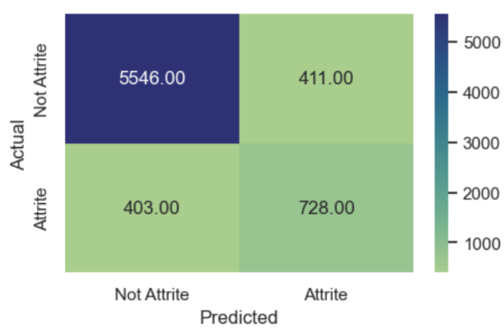


Fig.15. Optimized Logistic Reg_Train

	precision	recall	f1-score	support
0	0.93	0.93	0.93	2543
1	0.65	0.62	0.64	496
accuracy			0.88	3039
macro avg	0.79	0.78	0.78	3039
weighted avg	0.88	0.88	0.88	3039

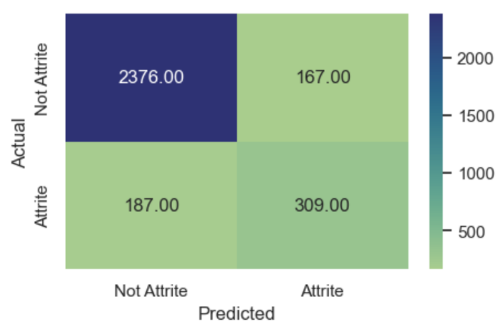
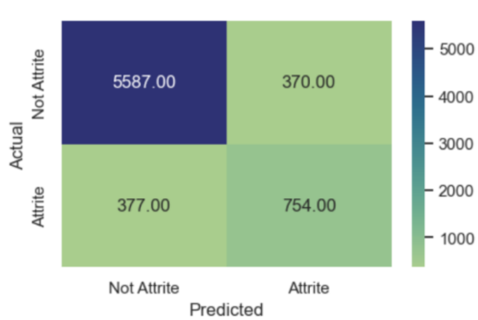


Fig.16. Optimized Logistic Reg_Test

4.2 GuassianNB

	precision	recall	f1-score	support
0	0.94	0.94	0.94	5957
1	0.67	0.67	0.67	1131
accuracy			0.89	7088
macro avg	0.80	0.80	0.80	7088
weighted avg	0.89	0.89	0.89	7088



	precision	recall	f1-score	support
0	0.92	0.95	0.93	2543
1	0.67	0.57	0.62	496
accuracy			0.89	3039
macro avg	0.80	0.76	0.78	3039
weighted avg	0.88	0.89	0.88	3039

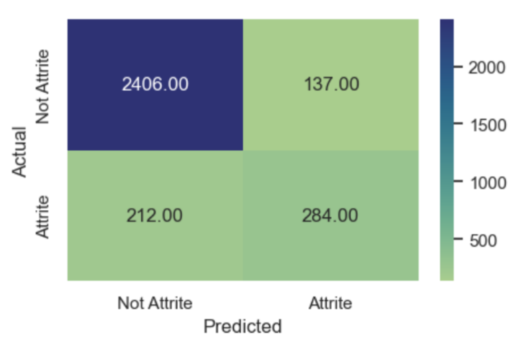
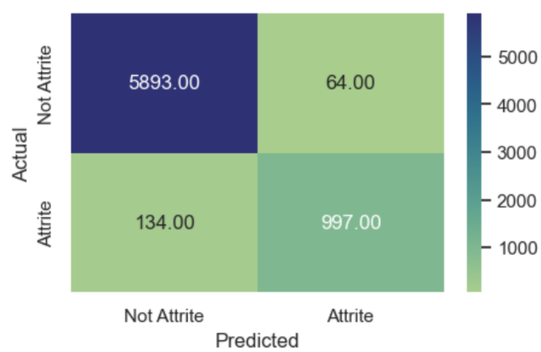


Fig.17. Optimized Naïve Bayes_Train Fig.18. Optimized Naïve Bayes_Test

4.3 Decision Tree

	precision	recall	f1-score	support
0	0.98	0.99	0.98	5957
1	0.94	0.88	0.91	1131
accuracy			0.97	7088
macro avg	0.96	0.94	0.95	7088
weighted avg	0.97	0.97	0.97	7088



	precision	recall	f1-score	support
0	0.95	0.97	0.96	2543
1	0.84	0.76	0.80	496
accuracy			0.94	3039
macro avg	0.90	0.87	0.88	3039
weighted avg	0.94	0.94	0.94	3039

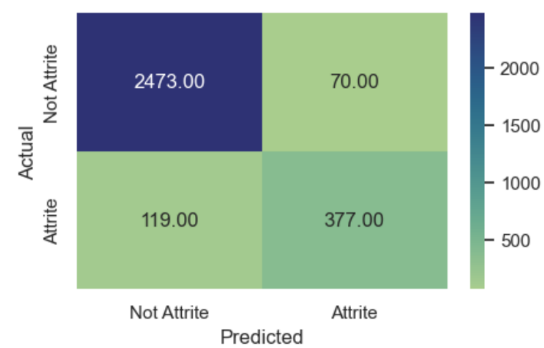
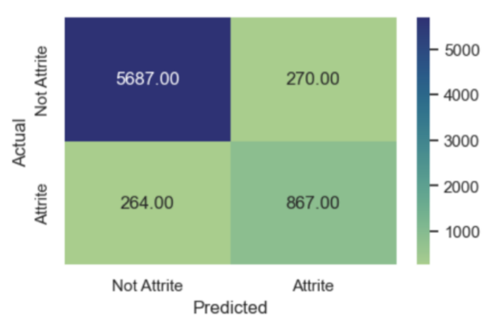


Fig.19. Optimized DecisionTreeClassifier_Train Fig.20. Optimized DecisionTreeClassifier _Test

4.4 SVM

	precision	recall	f1-score	support
0	0.96	0.95	0.96	5957
1	0.76	0.77	0.76	1131
accuracy			0.92	7088
macro avg	0.86	0.86	0.86	7088
weighted avg	0.92	0.92	0.92	7088



	precision	recall	f1-score	support
0	0.94	0.95	0.94	2543
1	0.71	0.67	0.69	496
accuracy			0.90	3039
macro avg	0.82	0.81	0.82	3039
weighted avg	0.90	0.90	0.90	3039

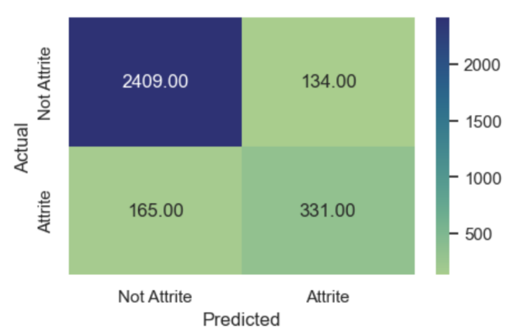


Fig.21. Optimized SVM_Train Fig.22. Optimized SVM _Test

The models performed better after hyper-parametric tuning or setting the optimal threshold. We set the optimal threshold by determining the point at which precision and recall are equal as shown below.

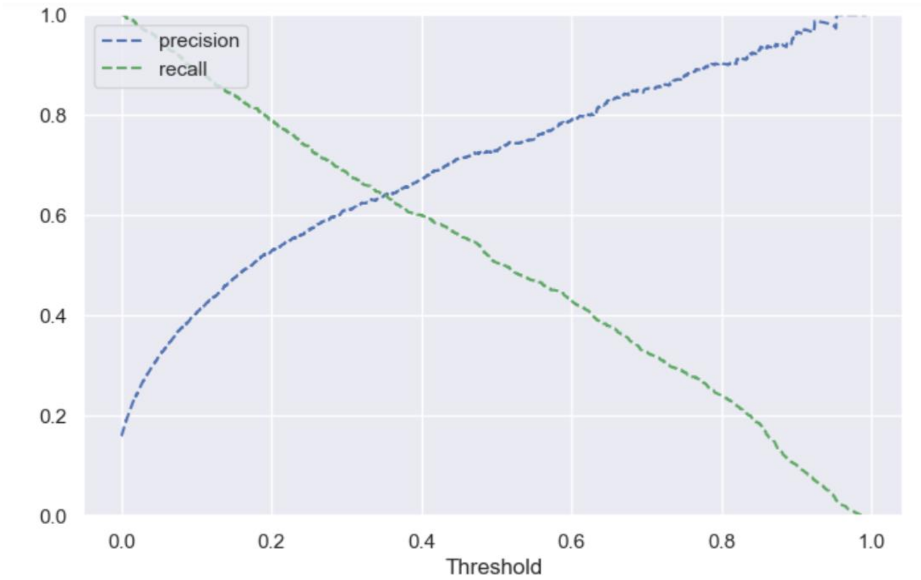


Fig.23. Optimal Threshold for Logistic Regression

5 Discussion

5.1 Feature Importance

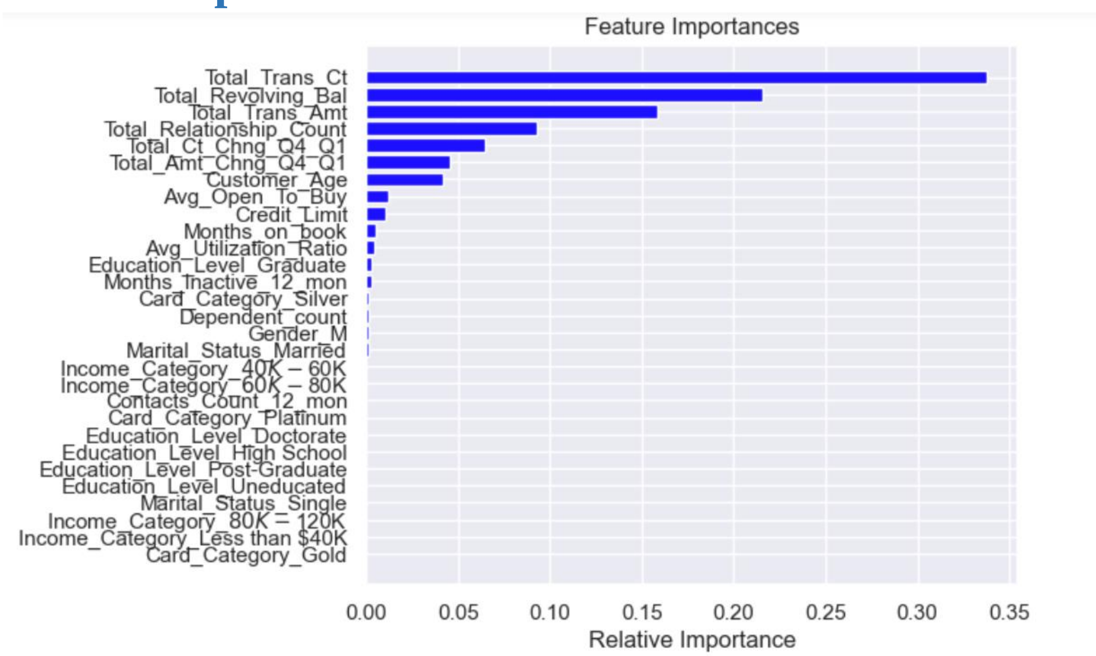


Fig.24. Feature Importance

- **Total_Trans_Ct** is the most important feature followed by **Total_Revolving_Bal** and **Total_Trans_Amt** which makes **sense.Customers** who are doing more transactions with the bank have lower chance of attrition.
- **Total_Ct_Chng_Q4_Q1**, **Total_Relationship_Count**, **Total_Amt_Chng_Q4_**Q1 are also important factors.

5.2 Model Comparison

	CrossValMeans	CrossValerrors	Algorithm
0	0.90477	0.01232	LogisticRegression
1	0.87514	0.01269	GaussianNB
2	0.93496	0.00822	DecisionTreeClassifier
3	0.90632	0.01054	SVC

Fig.25. Average Accuracy/Errors

5.3

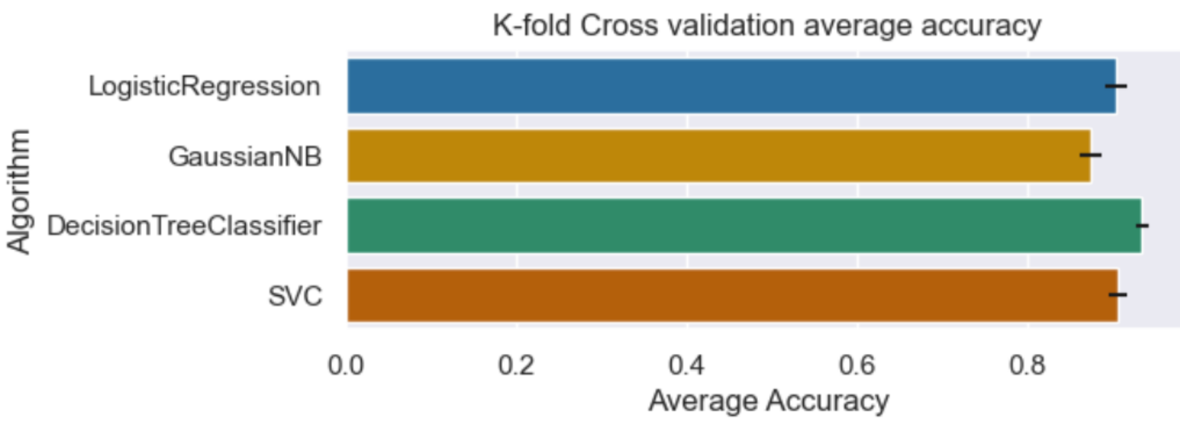


Fig.26. Average Accuracy

- Logistic Regression and GaussianNB had an average accuracy/errors of 0.9048/0.01232 and 0.8751/0.1269 respectively. This shows that the GaussianNB have the least accuracy.
- DecisionTreeClassifier and SVC had an average accuracy/errors of 0.9350/0.00822 and 0.9048/0.01232 respectively. DecisionTreeClassifier had the best performance. These models will perform well in differentiating out those customers who have high chances of leaving the bank, meaning it will eventually help in reducing the attrition rate.

6 Conclusion

This paper have been able to show extensively that we can maximize the recall by minimizing the false negatives. Few factors might have affected the outcome, so there is a need to handle imbalanced data. The paper was able to identify the key influencers of customer attrition.

7 References

Agarwal, V., Taware, S., Tadav, S.A., Gangodkar, D., Rao, A.L.N., and Srivastav, V.K. (2022). Customer - Churn Prediction Using Machine Learning. *2nd International Conference on Technological Advancements in Computational Sciences (ICTACS).*

Verbeke,W., Martens, D., Mues, C., and Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, (38)3, pp.2354-2364.

Panduro-Ramirez, J., Akram, S.V., Reddy, C.S., Ruiz-Salazar, J.M., Kanwer, B. and Singh, R. (2022).Implementation of Machine Learning Techniques for predicting Credit Card Customer action. *International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*

Van den Poel, D. and Larivière, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 157, pp. 196-217

Jialu, Z., Shiqian, Q. and Ge, Y. (1998). Assessment methods of speech synthesis systems for Chinese, *Acta Acustica*, (23)1, pp.19-30.

Xiong, Y. (2012). Building text hierarchical structure by using confusion matrix. *5th International Conference on BioMedical Engineering and Informatics, Chongqing, China*, pp. 1250-1254.

Oduro, M.S., Yu, H. and Huang, H. (2022). Predicting the Entrepreneurial Success of Crowdfunding Campaigns Using Model-Based Machine Learning Methods. *International Journal of Crowd Science*, (6)1, pp.7-16.

Probst, P., Boulesteix, A.L. and Bischl, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. *Journal of Machine Learning Research*, (20), pp. 1-32.

N V, M. K., K K, B. K. and Mudhol, A.C. (2022). Machine Learning based Prediction of Customer Churning in Banking Sector. *International Conference on Augmented Intelligence and Sustainable Systems (ICAISS),* Trichy, India, pp. 474-481.

Bhujbal, N.S and Bavdane, G.P. (2021). Leveraging the efficiency of Ensembles for Customer Retention. *Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC),* Palladam, India, pp.1675-1679.

Wang, X., Nguyen, K. and Nguyen, B.P.(2020). Churn Prediction using Ensemble Learning. *Proceedings of the 4th International Conference on Machine Learning and Soft Computing*, pp.56–60.