A PREDICTIVE ANALYSIS OF CREDIT CARD CUSTOMER CHURN USING MACHINE LEARNING AND ENSEMBLE TECHNIQUES

by

Ayodotun Ojo, BSc., University of Windsor, 2017

A Major Research Project presented to Ryerson University
in partial fulfillment of the requirements for the degree of

Master of Science in the Program of
Data Science and Analytics

Toronto, Ontario, Canada, 2022

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A MAJOR RESEARCH PROJECT

I hereby declare that I am the sole author of this Major Research Paper (MRP). This is a true copy of the MRP, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this MRP to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

Ayodotun Ojo

A PREDICTIVE ANALYSIS OF CREDIT CARD CUSTOMER CHURN USING MACHINE LEARNING AND ENSEMBLE TECHNIQUES

Ayodotun Ojo

Master of Science 2022

Data Science and Analytics

Ryerson University

ABSTRACT

Customer relationship management and customer churn prediction have received growing attention over the past decade. In this research paper, we conducted a predictive analysis leveraging Machine Learning techniques such as Random Forest, Support Vector Machine, K-Nearest Neighbours, XGBoost and Artificial Neural Networks, as well as Ensemble techniques such as Stacking using Logistic Regression as a meta-learner and Voting to classify credit card customer churn. Since foregone profits of attrited customers and the cost of attracting new customers can be significant, an increase in retention rate can be very profitable to banks and service providers. By ascertaining potential churners, service providers can design a more effective client relationship management strategy and offer more tailored services to customers. An exploratory data analysis is conducted, and we also determine which attributes are more relevant in predicting customer churn to support the interpretability of the analysis.

Keywords:

Churn Classification, Credit Card Customers, Ensemble, Attrition, Customer Relationship Management

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to my supervisor for this MRP, Dr. Farid Shirazi, for his support, guidance and encouragement through all the stages of writing my research project. I am very thankful for his guidance and assistance throughout the term and for providing me with valuable research materials and feedback needed to complete this report.

I would also like to thank God for seeing me through my studies, and my loving family and friends for their continued support and enthusiasm throughout the course of my program at the University. For this, I am entirely grateful.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The aim of this research is to develop and compare predictive models to detect credit card customer attrition in order to better help financial institutions reduce the severity of loss to their businesses incurred by customers exiting the company. We also examine if there any improvements by combining base learners, as many studies in this domain do not experiment with ensemble techniques. First of all, we give a brief background on the domain in study, and then we define the problem and research question. We then provide a literature review and conduct a detailed exploratory analysis of the dataset. Next, we go over the methodology and experimental set up, as well as performance metrics for the classifiers. Finally, the result of the experiment is evaluated and the paper is concluded with a final remark and recommendations for future work.

1.1. Background

Customer Churn (as known as Customer Attrition) is one of the most important performance metric for any customer-centric business to evaluate. Customer churn measures the number customers who decide to discontinue using a company's product or service. It is a significant aspect of study as it costs more for an institution to acquire new customers than it does to retain existing customers.

Customer Churn Prediction is a form of Customer Relationship Management (CRM) in which a company develops machine learning models that predicts if a customer is planning on leaving or reducing its purchases from a company. These techniques can be applied to credit card customers in order for banks to improve their customer retention.

Being proactive and detecting in advance if a customer is planning to leave, and reacting in time to convince them to stay can result in a more satisfied customer base. Therefore, such predictive analysis can help financial institutions reduce churn and offer more tailored products to their customers.

For this research project, Machine Learning techniques including Random Forest, Support Vector Machine, K-Nearest Neighbours, XGBoost and Artificial Neural Networks, as well as Ensemble techniques such as Stacking using Logistic Regression as a meta-learner and Voting are experimented for credit card customer churn.

1.2. Research Questions

The aim of the research is to provide an effective method of predicting whether or not a credit card customer will remain or exit the company. Based on the objective and the domain of study, the research questions are formulated below.

1. What is the most suitable Machine Learning or Ensemble model to be used to predict future customer churn for the given dataset?
2. How do the different methods compare to one another?
3. Can the applied methods perform better by utilizing feature selection techniques to remove noise from the data?
4. Which variables are the strongest predictor of a customer attrition?
5. How can the results be applied in future works to improve customer retention? How do the results compare to the current methodology used for customer churn prediction.

2. LITERATURE REVIEW

Customer relationship management, and customer churn prediction in particular, have received a growing attention during the last decade [1]. Churn management can be described as the concept of identifying those customers who are intending to move their custom to a competing service provider [2].

The primary focus of this project is to assess which machine learning or ensemble method is best at predicting credit card customer churn for banks and other credit card service providers. The results is used to improve the customer retention strategy of the bank. According to Athanassopoulos, losing customers not only leads to opportunity costs because of lack of sales revenue, but also to the cost of attracting new customers, which includes promotion, discounts, effort to know customer needs, and time to build sustainable relationships [3]. This research aims to help banks mitigate these cost associated with losing credit card customers. By predicting if a customer intends on leaving, the bank can proactively react in time and avoid the potential cost.

There are several studies that discuss the 4 dimensions of Customer Relationship Management [4,5,6]. These dimensions can be viewed as complete cycle of customer management [7,8,9]. They include:

1. Customer Identification
2. Customer Attraction
3. Customer Retention
4. Customer Development

According to Ngai et el., these dimensions share the common goal of creating a deeper understanding of customers to maximize customer value to the organization in the long term. Machine learning and deep learning techniques, therefore, can help to accomplish such a goal by extracting or detecting hidden customer characteristics and behaviours from large databases. Among the 4 major dimensions mentioned above, customer retention is the central concern for CRM and accounts for more than 62% of customer relationship studies [10], it is the main focus of this research project.

Although churn prediction modelling has been extensively researched, no general consensus exists on the performance of churn prediction modelling techniques [1]. For instance both Hung et al. [11] and Huang et al. [12] conducted a predictive analysis for customer churn on telecommunication data. Both studies, in addition to other modelling techniques, incorporated Decision Trees and Artificial Neutral Network. The first study found that the Artificial Neural Network gave the best predictive results at 99% and the decision trees gave the second best at 90%. Conversely, the second study found that Decision Tree and SVM gave the best predictive results outperforming other modelling techniques including Artificial Neural Network.

Another study by Buckinx and Van den Poel [13] applied Artificial neutral networks (ANN), Logistic regression (LR) and Random forest (RF) on retailing dataset, they concluded that random forest outperformed the ANN and LR. Widely varying methodologies and experimental setups have been carried to cross compare different various techniques for customer churn. For this section, an overview of related works will be discussed.

In a study by Miao and Wang [14], three models including Random Forest, Linear Regression and K-Nearest Neighbor (KNN) are applied to a Credit card holders' information dataset, using a 5-fold cross validation. The dataset contains more than 10000 instances and 21 features variables. By tuning hyperparameters and evaluating models based on ROC & AUC and confusion matrix, it is concluded that Random Forest has the best performance with its accuracy reaching 96.25%. The researchers found that Total transaction amount in the last 12 months, total transaction count in the last 12 months and total revolving balance are the top three important features which have the significant impacts on the customer churn prediction. It shows that the more frequent customers use their credit cards, the less likely they are to leave, and by using this model, bank managers can proactively take actions to fight against customer churn.

Using the same credit card holder dataset, Guliyev and Tatoğlu [15] focuses especially on explainable Machine Learning models and uses SHapely Additive exPlanations (SHAP) values to support the machine learning model evaluation and interpretability for customer churn analysis. The effects of various factors such as age, income, gender, credit card status, and discount opportunities offered by banks on customer churn were examined with Logistic Regression, Decision Tree, Random Forest and XGBoost classification models. They concluded that XGBoost model outperformed other machine learning methods in classifying churn customers

Ünlü and Demirberk [16], using the same credit card dataset, employed a hybrid machine learning algorithm to predict credit card customer churn. The proposed model is Support Vector Machine (SVM) with Bayesian Optimization (BO). BO is used to optimize the hyper-parameters of the SVM. The performance metrics used are accuracy, precision, recall and F1-score. It is shown that the best kernel to predict churn behaviour of the customers is the SVM with linear kernel. The researchers concluded that although the data set is complex and contains many explanatory variables, a linear model fits the data better than the non-linear ones.

Several studies focuses on the analysis of churn drivers, a 2009 study by Busckinx & Van den Poel [13] illustrates the need to gain insights in the causes of churn, since foregone profits of (partially) defected customers can be significant, an increase of the retention rate can be very profitable. The paper focuses on the treatment of a company's most behaviourally loyal customers in a non-contractual setting. The model is built in order to predict partial defection by behaviourally loyal clients using three classification techniques: Logistic regression, automatic relevance determination (ARD) Neural Networks and Random Forests. They concluded that predicting partial defection of behaviourally loyal customers is a successful strategy as behavioral variables are much more important than demographic variable.

Another study by Shirazi and Mohammadi [17], examines the effect of different aspects of customers' behavior on churning decisions, with emphasis on the retired segment. The study constructs a predictive churn model by utilizing big data, including the structured archival data, integrated with unstructured data from sources such as online web pages, the number of website visits and phone conversation logs, for the first time in the financial industry. Similar to Busckinx & Van den Poel, the researchers concluded that clients' behavior has a significant correlation with churning decisions. They noted that the finding has a considerable impact on designing effective client relationship management strategies in the future because the higher the rate of accuracy in detecting potential churners, will result in higher rates of success for more effective and efficient retention strategies.

Since customer churn is often a rare event for service providers including telecommunication, finance and retail, we can expect to observe high class imbalance in the dataset. A study conducted by Burez and Van den Poel [18] investigates how we can better handle class imbalance in churn prediction. Using appropriate evaluation metrics such as AUC and lift, the researchers investigates the increase in performance by using sampling and 2 modelling techniques including gradient boosting and weight random forest. The results showed that there is no increase in predictive performance when using advanced under sampling technique. This is in line with a study conducted by Japkowics [19] who noted that using sophisticated sampling techniques do not give any clear advantage. Burez and Van den Poel found that weighted random forests, as a cost-sensitive learner, performs significantly better compared to Random forest. According to Chen and Breiman [20], since the Random forests classifier tends to be biased towards the majority class, one can place a heavier penalty on misclassifying the minority class. A weight is assigned to each class, with the minority class given larger weight.

According to Burez and Van den Poel, the basic sampling methods include under-sampling and over-sampling. Under-sampling eliminates majority-class examples while over-sampling, in its simplest form, duplicates minority-class examples. Because over-sampling introduces additional training cases, it can increase the time necessary to build a classifier. Worse yet, because over-sampling often involves making exact copies of examples, it may lead to overfitting [21]. A well renowned research by Chawla et al. [21], examined how Synthetic Minority Over-sampling Technique (SMOTE) approach can improve the accuracy of classifier for the minority class. The minority class is over-sampled by creating “synthetic” examples rather than by over-sampling with replacement. This approach was inspired by Ha & Bunke [22] study on handwritten character recognition. They successfully created additional training data by performing certain operation on real dataset such as rotation and skew.

There exist several studies which has made use of SMOTE for customer churn prediction. One which is by Lal and Kumar [23], the researchers proposed a predictive model on churn customers using SMOTE and XGBoost additive model and machine learning techniques in Telecommunication Industries. SMOTE and XGBoost technique was used to handle the imbalanced dataset and gave high accuracy for predictive model to identify where or not a customer will churn. The researcher concluded that the proposed predictive model is more accurate and capable to handle imbalance dataset. Another paper by Eka Pura Hartati et al. [24] investigates handling class imbalance in churn prediction using combined Synthetic Minority Over-Sampling (SMOTE) and Random Under-Sampling (RUS) with Bagging method for a better churn prediction performance's result.

Another important step in data processing for customer churn dataset is feature selection. According to Hadden et al. [2] Feature selection is the process of identifying the fields which are the best for prediction. Sun et al. [25] emphasizes that it an important stage because it helps with both data cleansing and data reduction, by including the important features and excluding the redundant, noisy and less informative ones.

Niang et el [26] also describes feature selection as the most crucial process in achieving a high-performance churn prediction model. They mentioned that the feature selection process is not highlighted in many studies within this domain, unfortunately. According to the researchers, It was found that few studies did not even include the feature selection process as part of their data preparation from several articles published between 2016 and 2021.

Yin et al. [27] stated that when the customer churn prediction model is built, a large number of features bring heavy burdens to the model and even decrease the accuracy. Therefore it is important to perform feature selection prior to modelling to achieve high-level accuracy. This research paper aims to use minority class balancing techniques as well as feature selection to develop models to detect banks' customer churn. This will better help to develop more effective and efficient retention strategies for banks.

3. DESCRIPTIVE ANALYTICS | EXPLORATORY DATA ANALYSIS

3.1. Data Acquisition

The Credit Card Customers Dataset is available on Kaggle. The Dataset was originally obtained from Analyttica Datalab Inc, a Data Analytics and AI Solution Company, which provides real business cases, while retaining the privacy of the company.

3.2. Data and Features

This dataset consist of 10,127 credit card consumer observations. All customers in the dataset have been with the bank for at least 12 months. After the data clean-up and pre-processing, the dataset is made up of 20 predictor variables and a target variable to indicate a customer's decision to leave or remain with the credit card provider.

Tables 1 below summarizes the definition of each variable in the dataset.

Table 1*: Definition of Features*

	Feature	**Definition**
1	Attrition Flag	This is a flag to indicate if the customer is an existing customer or has exited
2	Client Number	Unique identifier for the customer holding the account
3	Customer Age	The customer's age in years
4	Gender	The gender of the customer, M=Male and F=Female
5	Dependent Count	This is the number of dependents of the customer
6	Education Level	Educational Qualification of the account holder (example: high school, college graduate, etc.)
7	Marital Status	The marital status of the customer e.g. Married, Single, Unknown
8	Income Category	Annual Income Category of the account holder, values include < $40K, $40K - 60K, $60K - $80K, $80K-$120K, > $120K, Unknown
9	Card Category	Type of Card held by the customer (Blue, Silver, Gold, Platinum)
10	Months on book	Length of the customer/bank Relationship

11	Total Relationship Count	Total number of products held by the customer
12	Months Inactive 12 months	Number of months inactive in the last 12 months
13	Contacts Count 12 months	Contacts to the bank in the last 12 months
14	Credit Limit	Credit Limit on the Credit Card
15	Total Revolving Balance	Revolving Balance on the Credit Card
16	Average Open To Buy	Open to Buy Credit Line (Average of last 12 months)
17	Total Amount Change Q4 Q1	This is the change in Transaction Amount (Q4 over Q1)
18	Total Transaction Amount	Total Transaction Amount (Last 12 months)
19	Total Transaction Count	Total Transaction Count (Last 12 months)
20	Total Count Change Q4 Q1	Change in Transaction Count (Q4 over Q1)
21	Average Utilization Ratio	Average Card Utilization Ratio

3.3. Exploratory Data Analysis

An exploratory data analysis is conducted to gain a deeper insight of the dataset. First we split the data into categorical and numerical features. The categorical features include attrition flag, gender, education level, marital status, income category and card category. The numerical features consist of customer age, dependent count, months on book, total relationship count, months inactive 12 month, contacts count 12 month, credit limit, total revolving balance, average open to buy, total amount change q4 to q1, total transaction amount, total transaction count, total count change q4 to q1 and average utilization ratio.

a. Categorical Features

In this section, we examine the categorical variables. Figure 1 through 6 below show the distribution of the categorical variables. We find that only 16.1% of the customers churn, this is expected as churning is typically a rare occurrence. The dataset is fairly evenly distributed between male and female customers. We also find that a good majority (more than 70%) of the customers have a high school diploma or higher education. Almost half of the customers are married and 39% of the customer are single. More than half of the customers earn less than 60k annually and vast majority of 93.2% of the customers own a blue card.

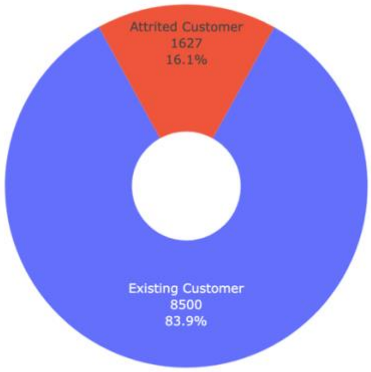

***Figure 1**: Distribution between Churned and Existing Customers*

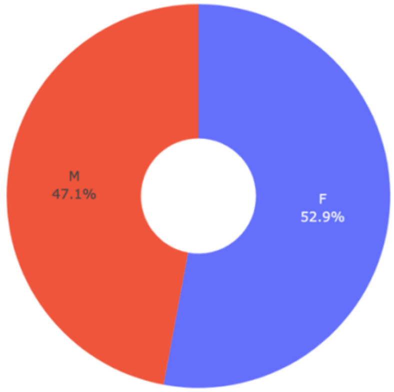

***Figure 2**: Distribution of Male and Female customers in dataset*

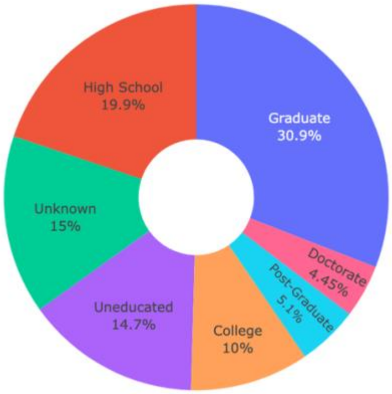

***Figure 3**: Distribution of Education Level of all customers*

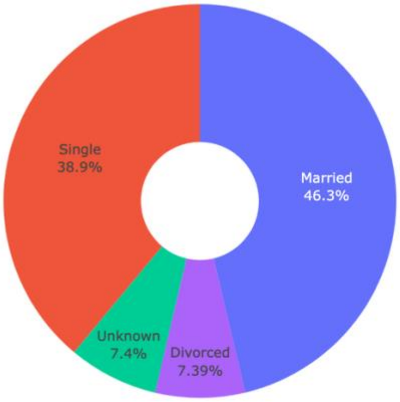

***Figure 4**: Distribution of Marital Status of all customers*

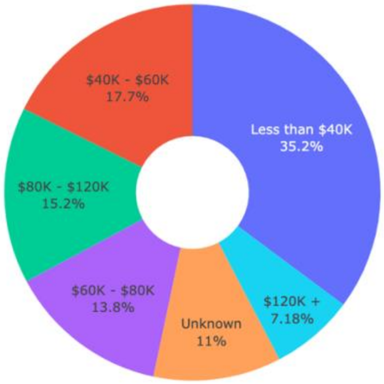

***Figure 5**: Distribution of Income category of all customers*

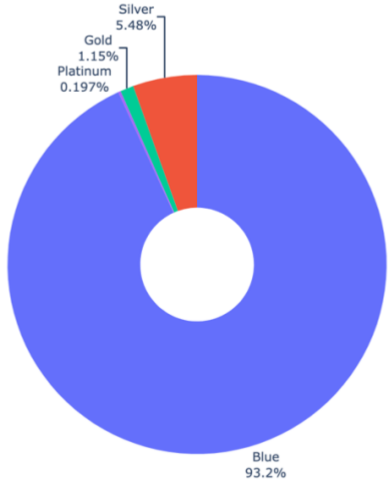

***Figure 6**: Distribution of Card types held by all customers*

Furthermore, we examine the categorical variables by splitting the data into 2 groups, the existing customers and the attrited customers. By splitting the data, we can search for patterns in each class by visualizing the data.

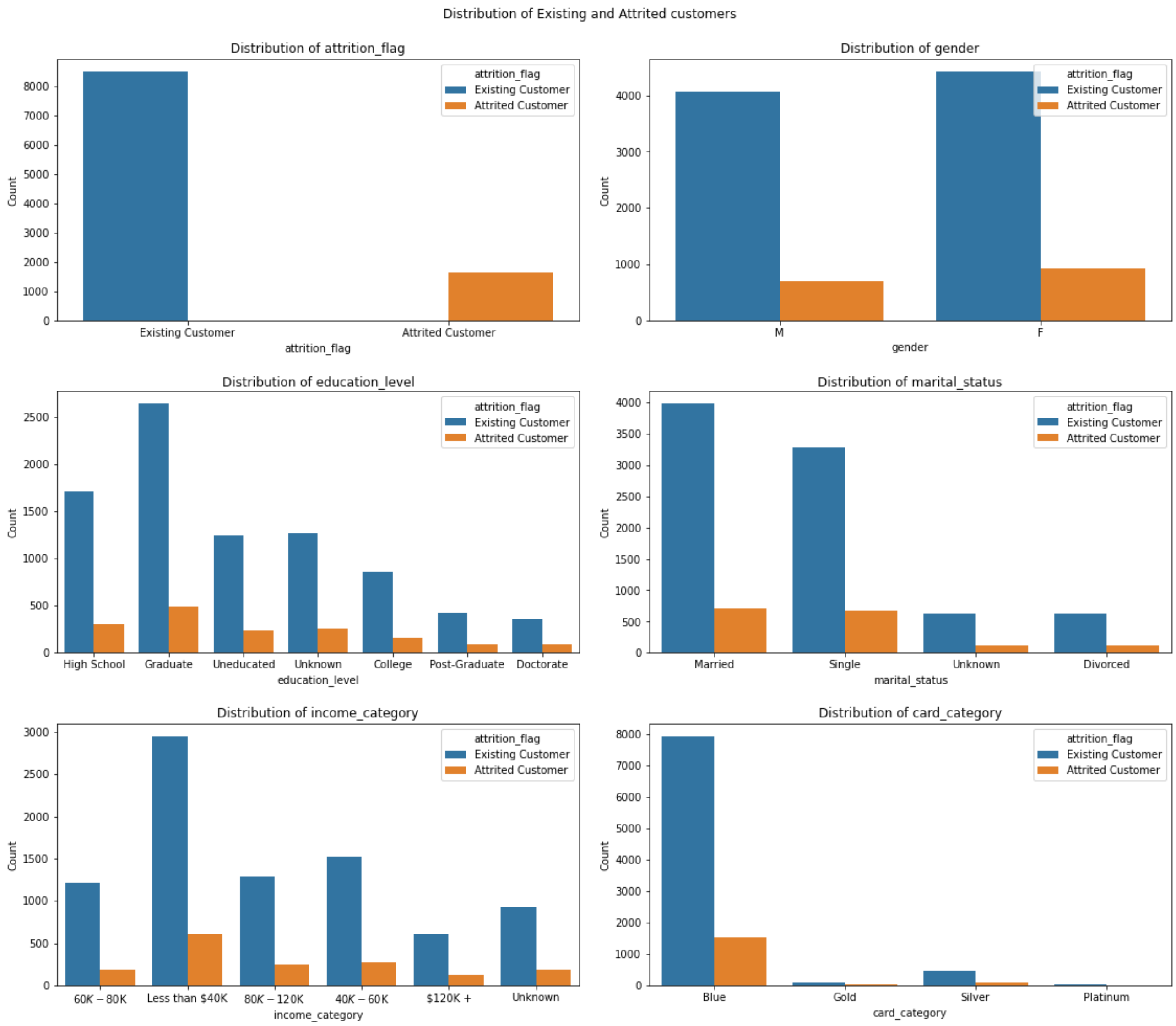

***Figure 7**: Combined distribution of all categorical variable grouped by class*

Figure 7 above show the distribution between the churned group and the existing group for all the categorical attributes using a combined bar chart. We see that both classes have indistinguishable patterns for all the categories, therefore, we are unable to gain much information by simply visualizing the data.

b. Numerical Features

Next, we examine the numerical variables in the dataset. The correlation among the numerical features is examined using a heatmap. Figure 8 shows the Pearson correlation between these numerical features.

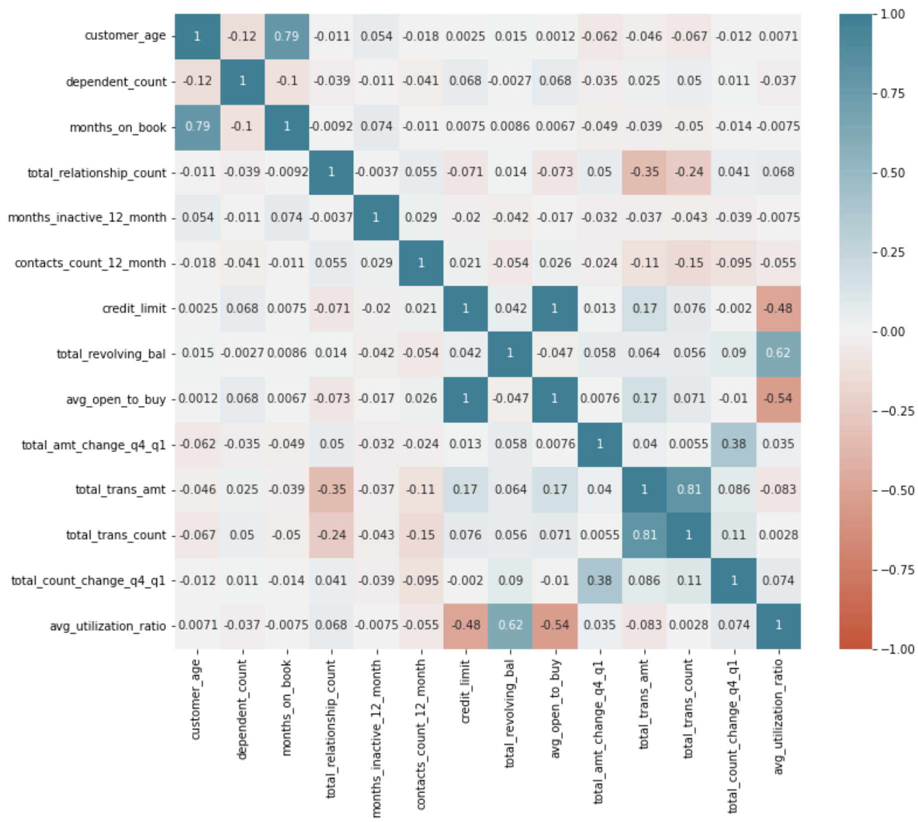

***Figure 8**: Pearson correlation heatmap of numerical attributes*

As expected, attributes such as months on the book and customer age, total revolving balance and average utilization ratio, total count of transaction and total transaction amount have high positive correlation. We find that average open to buy and credit limit have a perfect positive correlation. Conversely, we see a negative correlation between average utilization ratio and average open to buy of 0.54.

We then examine the distribution of the numerical variables in dataset. Figure 9 below shows the distribution using a combined histogram plots for the variables.

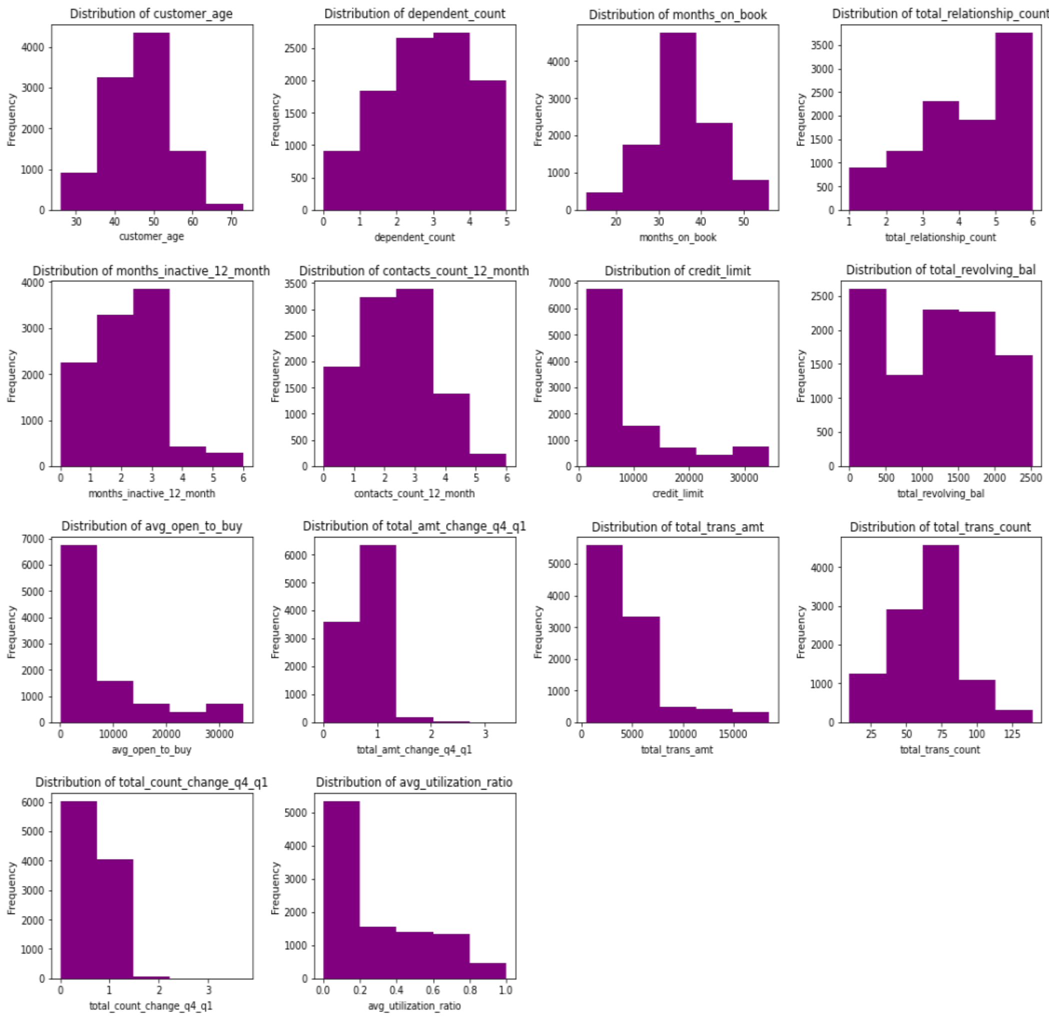

Figure 9*: Combined histogram showing distribution of all numerical features in dataset*

We find that features such as months inactive in 12 months, credit limit, average open to buy, total amount changes from Q4 to Q1, total transaction amount, total count change from Q4 to Q1 and average utilization ratio are skewed to the right. Total relationship count is left skewed as majority of the customers hold at least 3 products, while the other numerical variables have a normal distribution.

By splitting the numerical features into 2 groups of existing customers and attrited customers, we can futher examine the groups for patterns. The result is highlighted in section 5.1.

4. METHODOLOGY AND EXPERIMENT

4.1. Aim of the study

The aim of this study is to evaluate which Machine Learning, Deep Learning or Ensemble techniques is most efficient at predicting credit card customer churn for banks. Widely varying methodologies and experimental setups have been carried to compare various techniques for customer churn, however Ensemble techniques are not experimented in many studies within this domain. In this study, Ensemble techniques such as Voting and Stacking is applied to combine several base model in order to produce an optimal model.

The study aims to help banks mitigate the cost associated with losing credit card customers. By predicting if a customer intends on leaving, the bank can act proactively and mitigate these cost.

4.2. Experimental design

The dataset contains more than 10,000 credit card consumer records, it is made up of 20 feature variables (after data pre-processing) and a target variable to indicate a customer's decision to leave or remain with the credit card provider.

There is only a small percentage (16.07%) of the customers who have decided to churn, therefore class balancing techniques will be required to improve the modelling performance. Below is the experimental design and set up for this research project.

a. Feature Selection

Feature selection is the process of identifying the fields which are the best for prediction [2]. Sun et al. [25] describes it as a crucial stage because it helps with both data cleansing and data reduction. To achieve a high-performance churn prediction model, feature selection is performed as part of the data preparation. Feature selection is applied to the 22 feature variables, with the aim of eliminating noise and redundancy in the dataset and ideally, improve the overall performance of the models. As part of the initial data clean up, the last 2 columns named "Naive_Bayes_Classifier" are eliminated as recommended by the data provider. Following that, the client number column which is unique numeric values assigned to each customer is eliminated as it holds no relevant information.

Multicollinearity of the features is then examined using Variance Inflator Factor (VIF) and Correlation heatmap. Pearson correlation heatmap gives information on the relationships between the features variables and can help reduce dimensionality by removing highly correlated features during data processing. Multicollinearity among the feature variable are examined and features with high correlation are eliminated to remove redundant variables. Since the proportion of cases in the category is small, the indicator variables will necessarily have high VIFs, even if the categorical variable is not associated with other variables in the model, therefore a VIF of 15 and less is accepted in order to not loss valuable information. After examination, the following 3 attributes are eliminated one at a time, in a sequentially order: average open to buy, customer age and total transaction count. The performance of the base models are evaluated on the full dataset and the partial dataset using a 10 fold cross validation to determine if there any improvements using feature selection.

b. Cross Validation

The models are developed using a 10-fold cross-validation. By using cross validation, we can verify how accurate the model is performing on multiple and different subsets of data, therefore, ensuring that it generalizes well to the new samples.

c. Class Balancing Techniques

Since churning is often a rare occurrence, we find that the dataset is highly imbalanced, only approximately 16% of the customers discontinue using the service. Balancing techniques are applied to improve the performance of the model. The experiment is carried out on 2 datasets using the following balancing methods.

i. **SMOTE Oversampling**: Synthetic Minority Oversampling Technique, or SMOTE oversampling method is used to increase the label churn example, balancing the number of churn and attritted customer examples of label samples.

ii. **Stratified Random Sampling**: The training/validation dataset is created to balance the data for more accurate results using a stratified random sampling. The training and validation set is randomly selected for 1000 observations, using a 70:30 ratio for Existing to Churned observations. The remaining observations in dataset is used as the test data.

d. Train/Validation/Test Set Split

For this experiment, Train/Validation/Test set split is carried out for the SMOTE dataset and the Randomly Sampled Dataset, the set-up is as follows.

i. **SMOTE Dataset**

For Stacking Ensemble technique, the dataset is divided into training-validation-test sets based on a 70-15-15 split, this makes up the level 0 data. The level 0 models/base learners (RF, SVM, KNN, ANN, XGBoost) are fitted on the training dataset and are used to predict the labels of the instances in the validation set and test set. The predictions of the base learners are combined to make up the level 1 dataset i.e. the prediction on the validation set makes up the level 1 training data and the predictions on the test set makes up the level 1 test data. The meta-learner (LR) learns to make predictions from level 1 train data and then predicts the label of the level 1 test set. The results of the Ensemble techniques is compared with results attained from using the base learners only on the test set.

For Majority Voting Ensemble technique, the base models are fitted on the training data and predictions are made on the test data. The predictions of the models are combined, the prediction of the Voting Ensemble method is the majority voting of the contributing models. This result is compared with the results attained from using the single models only to predict the test set.

ii. **Randomly Sampled Dataset**

Using a stratified random sampling, the training/validation dataset is created to balance the data. The training/validation set is randomly selected for 1000 observations by a stratified random sampling. 700 samples are randomly selected from the Non-Churn label and 300 from the Churn observations, making a 70:30 non-churn to churn ratio. The remaining data is used as the test set.

4.3. Modelling Techniques

For this experiment, the performance of the utilized machine learning, deep learning and ensemble techniques are evaluated to assess which model best predicts customer churn. These modelling techniques are briefly described below.

a. Random Forest

Random forests is a supervised learning method proposed by Leo Breiman in 2001 for building a predictor ensemble with a set of decision trees that grow in randomly selected subspaces of data [28]. Figure 10 below shows the basic architecture. It is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [29]. In Random Forest, the decision trees are built independently and the final outcome is drawn based on the average of the individual decision tress. Random forest algorithm provides a higher level of accuracy in predicting outcomes over the decision tree algorithm. It is an ideal algorithm as it solves the problem of overfitting by applying a method known as bootstrapping, on randomly selected independent subset of the dataset.

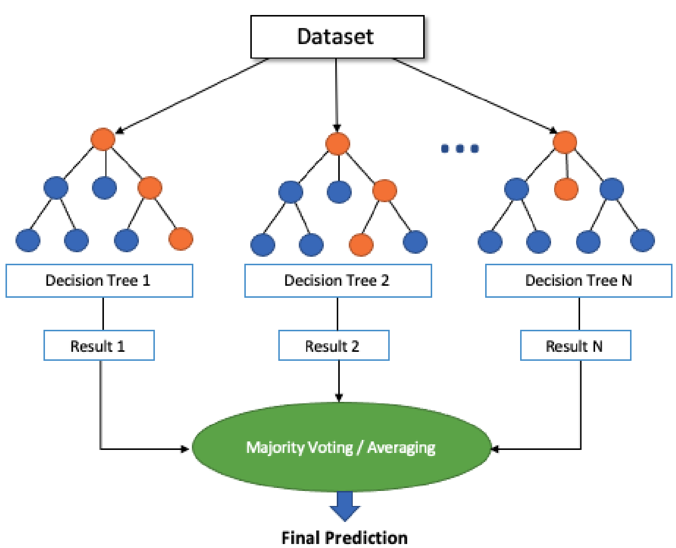

***Figure 10**: Random Forest Algorithm*

b. Extreme Gradient Boosting

Extreme Gradient Boosting, or XGBoost similar to Random Forest is a Decision tree ensemble learning algorithm. It is described by Chen and Guestrin [30] as a scalable machine learning system for tree boosting. It is a supervised learning method which uses several optimization technique to attain improved result. XGBoost builds decision tree sequentially, considering the gradient of the previous model for subsequent trees in efforts to minimize error.

c. K-Nearest-Neighbors

K-nearest neighbors (KNN) is a simple method of machine learning. KNN is a non-parametric classification method used to classify unlabeled observations by assigning them to the class of the labeled examples with having the closest distance to unlabeled observation, the KNN algorithm assigns a category to observations in the test dataset by comparing them to the observations in the training dataset [31].

For instance, let ‘x’ in Figure 11 below be a new instance. To classify the new sample, KNN finds the set of k objects from the training data closest to input data instance by using distance calculation, most often, the Euclidean distance. It assigns the maximum voted classes out of these neighboring classes to the new sample, for k = 2, the assigned class is yellow circle, and for k = 3, the assigned class in blue triangle.

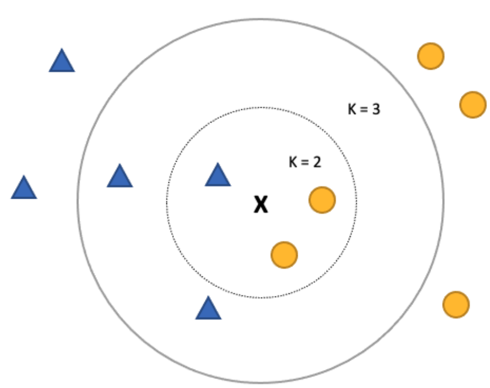

***Figure 11**: K-Nearest Neighbors Example*

d. Logistic Regression

Logistic regression, or logit regression is a statistical methods that can be implemented in modelling binary dependent variables. It is a mathematical modelling approach used to define the relationship between such independent variables as X1, X2, …, Xn, and Y binary dependent variable which is coded as 0 or 1 for two possible classes [32].

Logistic regression analyzes the relationship between multiple independent variables and a categorical dependent variable, and estimates the probability of occurrence of an event by fitting data to a logistic curve [33].

e. Support Vector Machine

A Support Vector Machine, or SVM, is a non-parametric supervised learning model based on statistical learning theory, it was first introduced by Vapnik.

SVM is a two dimensional description of the optimal surface evolved from the linearly separable case. The objective is to find a hyperplane in an N-dimensional space (N being the number of features) that distinctly classifies the data points. Figure 12 shows the basic idea. Two classes are separated by hyperplane H without errors. H1 and H2 are planes with a center point of H. The distance of H1 and H2 is called the class interval. Optimal separating surface ensures error-free separation of the two types of classes, this is known as the largest class interval [34].

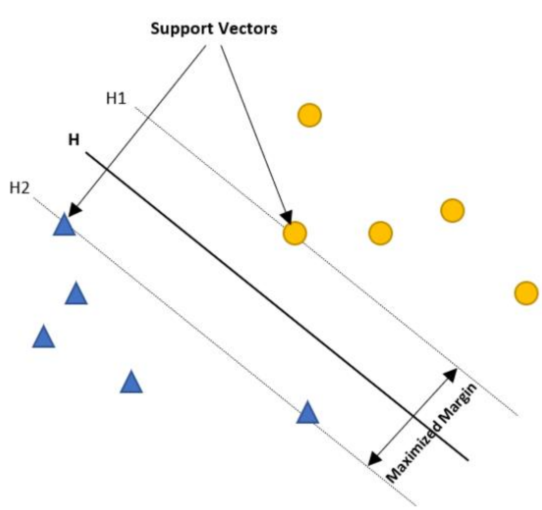

***Figure 12**: Support Vector Machine Example*

f. Artificial Neutral Network

Artificial Neural Networks (ANN) is a calculation method that builds several processing units based on interconnected connections. The network consists of an arbitrary number of cells or nodes or units or neurons that connect the input set to the output [35]. The basic artificial neural network architecture is seen in Figure 13 below.

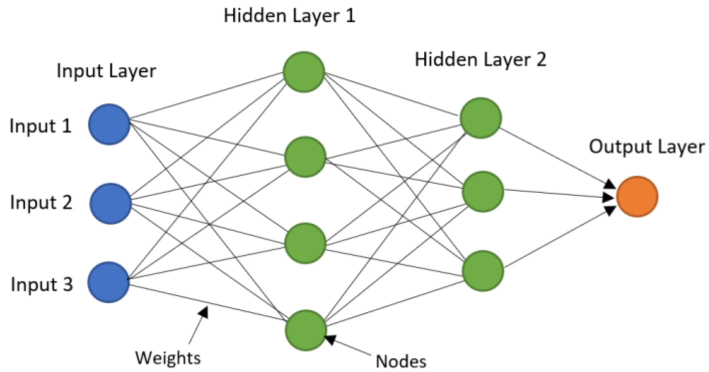

***Figure 13**: Artificial Neural Network Architecture*

Each connection has a connection weight, and each neuron has a threshold value and an activation function [36]. The weights continually adjust according to the learning rule and error values, as learning proceeds. ANN has the ability to find hidden pattern in the dataset which may not be identified by traditional machine learning techniques. For this experiment, the Artificial Neural Network is made up 4 hidden layers with a rectified linear activation function or ReLU. Each layer is followed by a drop out layer to avoid overfitting the training data. The output layer uses a sigmoid activation function and the model is compiled using Adam optimizer.

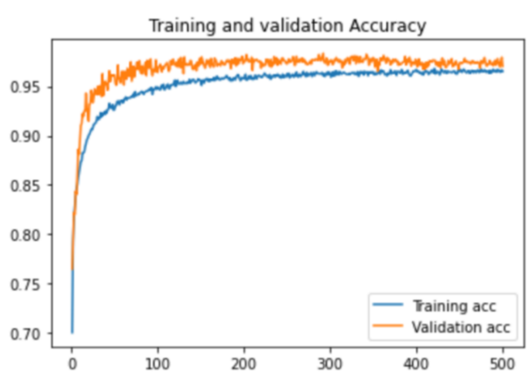

***Figure 14**: ANN training and validation accuracy over 500 epoch*

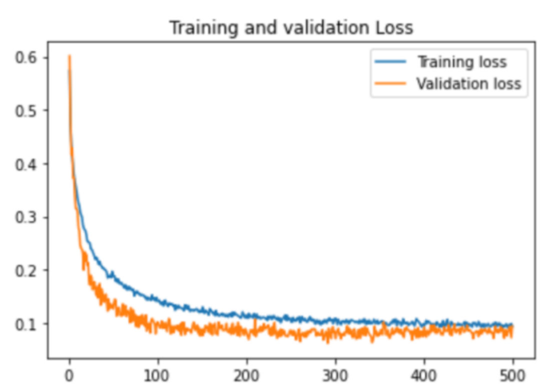

***Figure 15**: ANN training and validation loss over 500 epoch*

After conducting a search for an optimal epoch value having minimal loss, we chose to use 250 as an appropriate epoch value as shown in figure 14 and 15 above. The model is fitted with a batch size of 128 and validation split of 20%.

g. Stacking Ensemble Technique

Ensemble modelling is the process of combining multiple base models to make a final prediction, in order to obtain an improved performance. Ideally, applying an ensemble method, achieves a better performance than any single model used in the ensemble. Incorporating ensembles will reduce the variance element of the prediction error and improve performance. Stacking and Voting ensemble techniques are experimented for this project.

Stacking ensemble techniques combines the prediction of multiple base learner. It creates a new dataset from the combined prediction which is then used as input for the meta-learner. The architectural design is seen in Figure 16 below.

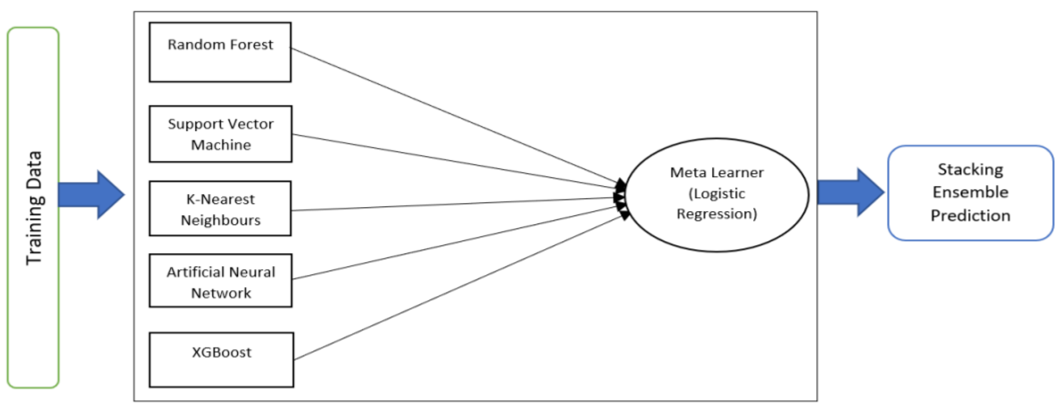

Figure 16: *Stacking Ensemble Technique Architecture*

h. Voting Ensemble Technique

Voting ensemble is another method of combining base learners. Voting ensemble techniques combines multiple models to make a final prediction by using the majority voting of the contributing models as illustrated in figure 17 below.

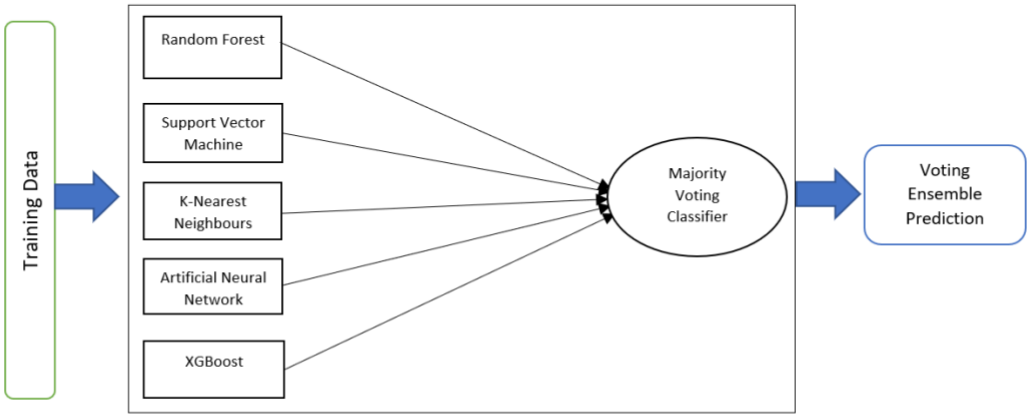

***Figure 17**: Voting Ensemble Technique Architecture*

Since neither keras nor scikit-learn have a function to create a stacking and voting ensemble model for Artificial neural networks which is one of our base learner for this experiment, a custom voting and stacking ensemble function is created and utilized for modelling. Tables 2 below shows sample training and test data for the Ensemble models.

Table 2: *Sample Train/Test data for Ensemble models*

label	Random Forest	SVC	K-Nearest Neighbours	XGBClassifier	ANN
0	0	0	0	0	0
0	0	0	0	0	0
1	1	1	1	1	1
1	1	1	0	1	1
0	0	0	0	0	0
0	0	1	1	0	0
0	0	0	1	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	1	0	0

4.4. Evaluation Metrics

The results of the models are evaluated using Confusion Matrix, Accuracy, Recall, Precision and F1-Score. These metrics are calculated as below.

a. *Confusion Matrix*

The confusion matrix allows for more detailed performance of the classification models. It reports how well the performed by showing the number of true positives, false negatives, false positive and true negative. Table 3 below shows the typical format of the confusion matrix and it explained below

Table 3: *Confusion Matrix Format*

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

True Positive: Indicates the observations with actual class of Positive, predicted as Positive
True Negative: Indicates the observations with actual class of Negative, predicted as Negative
False Positive: Indicates the observations with actual class of Negative, predicted as Positive
False Negative: Indicates the observations with actual class of Positive, predicted as Negative

b. *Accuracy*

Accuracy measure the number of correctly predicted instance by the model. It is calculated as the ration between the number of correct prediction to the total number of observations. It is mathematically expressed as:

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \tag{1}$$

Accuracy can be a misleading metric when working with a dataset that is highly imbalanced, therefore it is important that we take record of other performance metrics for this analysis.

c. *Precision*

Precision measures the ratio between true positives and all positives prediction made by the model, it gives a measure of relevant data points retrieved by the model. For this analysis, It gives a ratio of customers who actually churn to all the customers predicted to churn by the model. It is mathematically expressed as:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

d. Recall

Recall is the measure of the model's ability to correctly identify True Positive. It is the ratio of True Positives retrieved by the model to all actually Positives observations. For this experiment, it gives a measure of the customers the model correctly predicts to churn to all the customers who actually churn. Mathematically:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

e. F1-Score

F1-Score combines recall and precision by taking the harmonic mean, balancing out both precision and recall. it mathematically calculated as:

$$F1 - Score = 2 * \frac{Recall * Precision}{Recall + Precision} \tag{4}$$

4.5. Feature Importance

Feature importance describes which variables are more relevant for predicting credit customer churn in the dataset. Evaluating which attributes are more relevant in predicting if a customer intends on churning is a top priority for service provider, as it enables them provide a more tailored product and service, helps them improves customer relationship, and promotes customer retention. For this analysis, the feature importance of the predictor variables are determined using XGBoost and verified with Random Forest.

5. RESULTS AND DISCUSSION

5.1. Exploratory Analysis Results

An exploratory data analysis is conducted to gain an in depth understanding of the data. In this section, we will highlight some notable EDA findings. By visualizing the data, we see that the dataset is the highly imbalanced between the churned class and the existing class, with majority of the customers staying with the company. Figure 1 shows that only 16% of customers discontinue using the credit card services provided by the bank.

Similarly, we find that majority of the customers (more than 93%) are Blue card holders. The remaining 7% hold either Silver, Gold or Platinum cards as shown in Figure 6. Since the Blue card tends to be very popular among the customers, promoting this package to potential customers could be a good marketing strategy for the credit card company.

We further examine some behavioural attributes by splitting the data set into churned and existing customers to gain deeper insights. Both classes are examined using a stacked histogram.

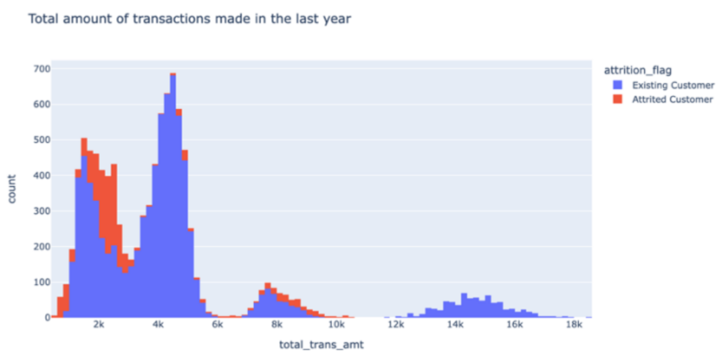

Figure 18*: Distribution of total amount of transaction among Churned and Existing Class*

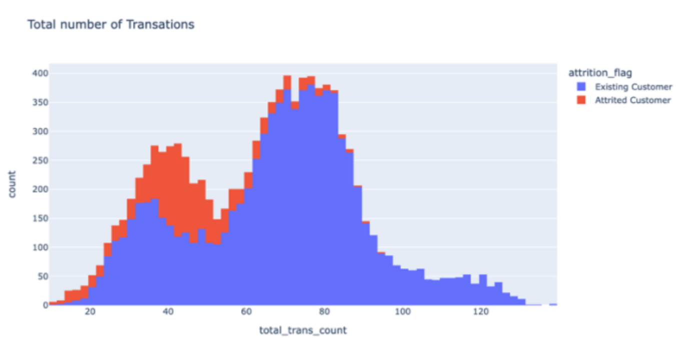

Figure 19*: Distribution of total transaction count among Churned and Existing Class*

Figure 18 and 19 above shows that churning customers tend to spend less amount on transactions in total than the existing customers, they also have a lower transaction count than the existing customers. This implies that behavioural attributes can be significant in determining whether a customer will churn.

Contrarily, demographic attributes such as age, gender, marital status, education level and income category showed no significant difference between the classes. Both classes have similar patterns and therefore visualizing the data did not give much information.

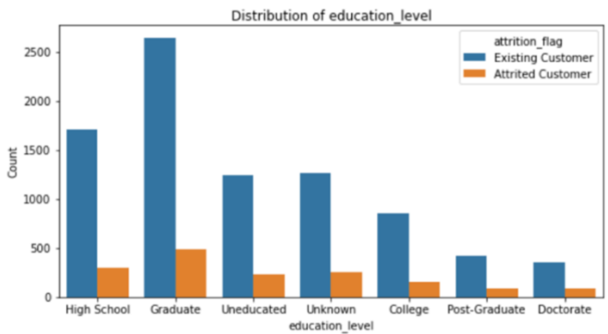

Figure 20: *Distribution of Education Level between Churned and Existing Class*

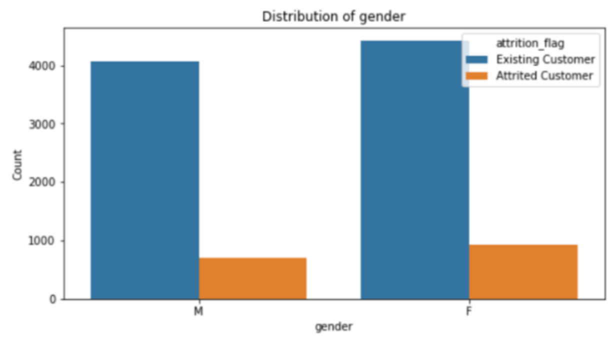

Figure 21: *Distribution of Gender between Churned and Existing Class*

For instance, Figure 20 and 21 above shows the distribution of educational level and gender attributes among the churn and existing classes. We find that there is no significant difference in the patterns shown by existing class and the churned classes, therefore we are unable to draw conclusions.

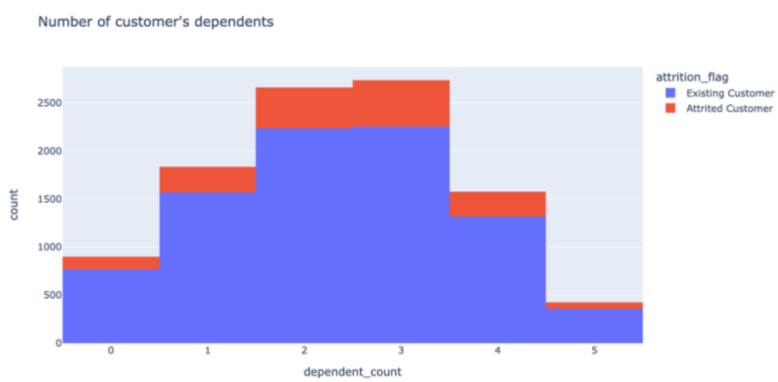

Figure 22: *Stacked histogram showing distribution of dependent count by class*

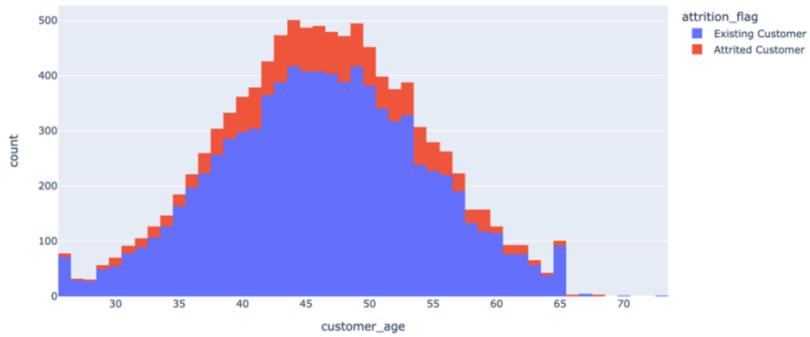

Figure 23*: Stacked histogram showing distribution of customer's age by class*

Similary, Figure 22 and 23 above illustrates that demographic features such as the customer's dependent count and age, have very similar patterns for both the existing customers and attrited customers. This implies that the demographic attributes are weak predicator of credit customer churn.

5.2. Machine Learning Results

a. Cross Validation

A 10 fold cross validation is carried out to verify the performance of the machine learning methods. The cross validation is carried on the full dataset and partial dataset to determine if feature selection showed any improvement in performance.

Table 4*: Average performance results of 10 fold cross validation with feature selection*

Model	Accuracy	Precision	Recall	F1-Score
ANN	0.904812	0.745803	0.520109	0.608811
K-Nearest Neighbours	0.892661	0.705268	0.571609	0.630488
Logistic Regression	0.877752	0.766837	0.346024	0.475252
Random Forest	0.950626	0.908551	0.770719	0.833494
SVC	0.839341	0.000000	0.000000	0.000000
XGBClassifier	0.963266	0.911175	0.854961	0.881958

Table 5: *Average performance results of 10 fold cross validation without feature selection*

Model	Accuracy	Precision	Recall	F1-Score
ANN	0.910043	0.806796	0.583299	0.674828
K-Nearest Neighbours	0.891675	0.703130	0.564872	0.626104
Logistic Regression	0.891280	0.731923	0.510766	0.600406
Random Forest	0.962477	0.927931	0.831610	0.876698
SVC	0.839341	0.000000	0.000000	0.000000
XGBClassifier	0.972647	0.928107	0.899807	0.913468

Using performance metrics such as accuracy, precision, recall and f1-score, Table 4 and 5 above show the results of the cross validation on the full dataset and a subset of the data features. The results shows that the models perform better when the full dataset is used, rather than a subset of the features. This was unexpected as feature selection performed by examining correlation and collinearity of the variables typically improves performance. Since we saw no improvement with feature selection, the rest of the experiment is carried out on the full dataset.

On evaluating the results, although the recall values are relatively low, we observed that all the model except SVM performed quite well, given that no balancing technique is applied. XGBoost outperformed the other models with an accuracy of 97.26%, Precision of 92.81%, Recall of 89.98% and F1-Score of 91.34%. Random Forest follows closely with an accuracy of 96.24%, Precision of 92.79%, Recall of 83.16% and F1-Score of 87.67%. By experimenting with SMOTE balancing techniques, we were able to boost the recall scores of the models. The below sections shows the results.

b. Modelling on SMOTE Dataset

This experiment focuses on determining which modelling technique is best at predicting credit customer churn. SMOTE balancing technique is applied on the train data to generate more sample for the churned class. Modelling is then performed using Machine Learning and Ensemble methods including Artificial Neural Networks, Random Forest, Support Vector Machine, K-

Nearest Neighbour, XGBoost Classifier, Stacking Ensemble using Logistic Regression as a meta-learner and Voting Ensemble. These models are used to make predictions on the same test data. The experiment is repeated over 10 iterations and evaluated based on the accuracy, precision, recall and f1-score. The performance results are recorded at each iteration.

***Table 6**: Average Performance of models on Testing Data using SMOTE*

Model	Accuracy	Precision	Recall	F1 Score
ANN	0.943026	0.810454	0.855835	0.831565
K-Nearest Neighbours	0.802829	0.443667	0.773754	0.563462
Random Forest	0.957171	0.861327	0.880930	0.870859
SVC	0.902961	0.657117	0.859329	0.744346
Stacking_estimator	0.970329	0.916928	0.900502	0.908580
Voting_estimator	0.955921	0.903555	0.819886	0.859350
XGBClassifier	0.970592	0.906521	0.915123	0.910645

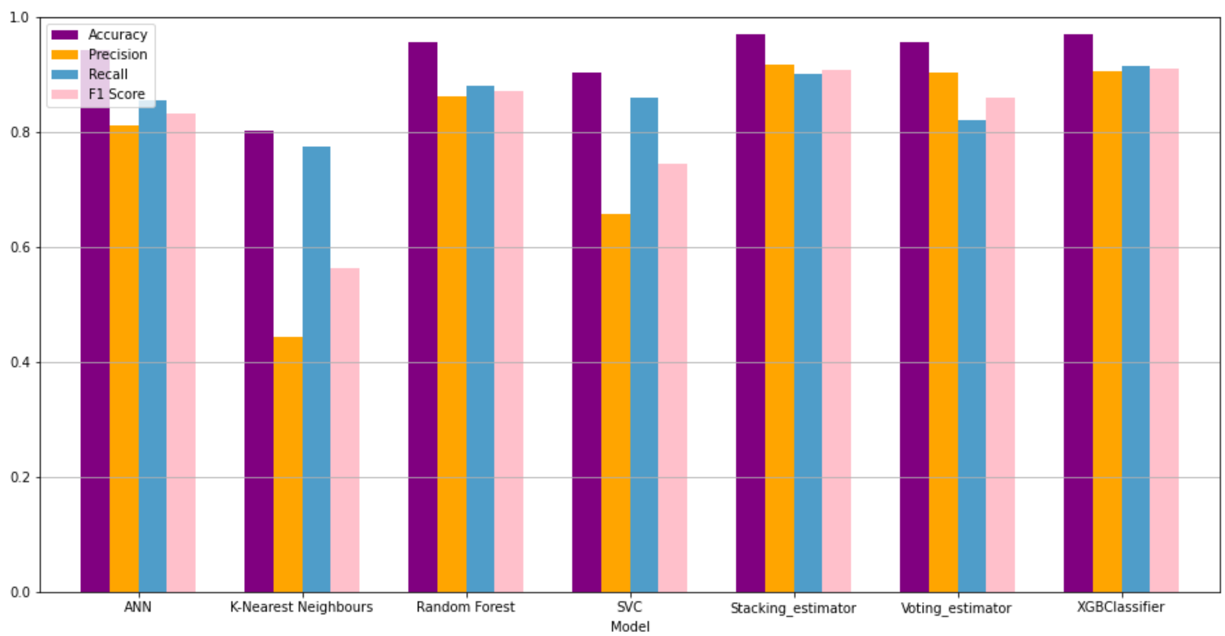

***Figure 24**: Average Performance of models on Testing Data using SMOTE*

Table 6 and Figure 24 above shows the average performance of the models. By applying SMOTE balancing technique on the training dataset, we were able to increase the recall score significantly.

The XGBoost and Stacking model outperformed the other model. On average, XGBoost had the highest accuracy of 97.06%, followed by the Stacking Ensemble with an accuracy of 97.03%. Since the dataset is highly imbalanced, accuracy could be a misleading metric for evaluating the model. Recall and Precision are more relevant evaluation metric for this analysis. The Stacking Ensemble method consistency had the highest precision values with an average of 91.69% and values reaching 94.27% over the 10 iterations. While XGBoost consistently gave higher recall values with an average of 91.51% and a maximum value of 92.64%.

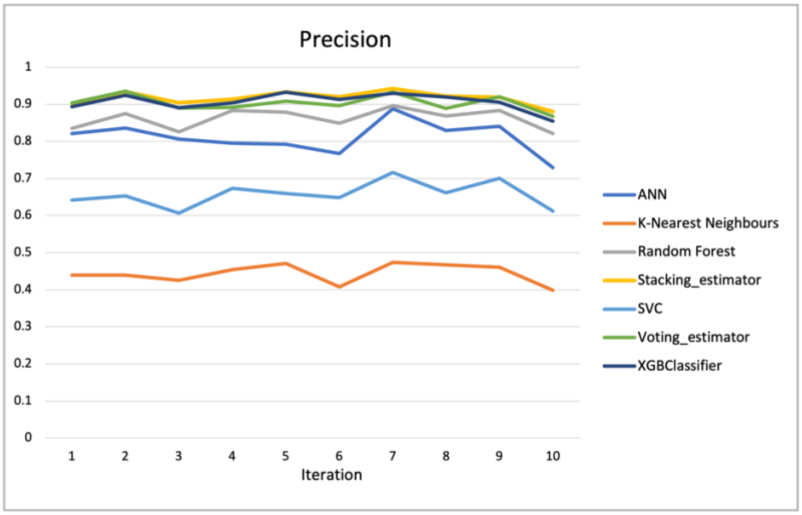

Figure 25: *Precision scores of the models on test data over 10 iterations*

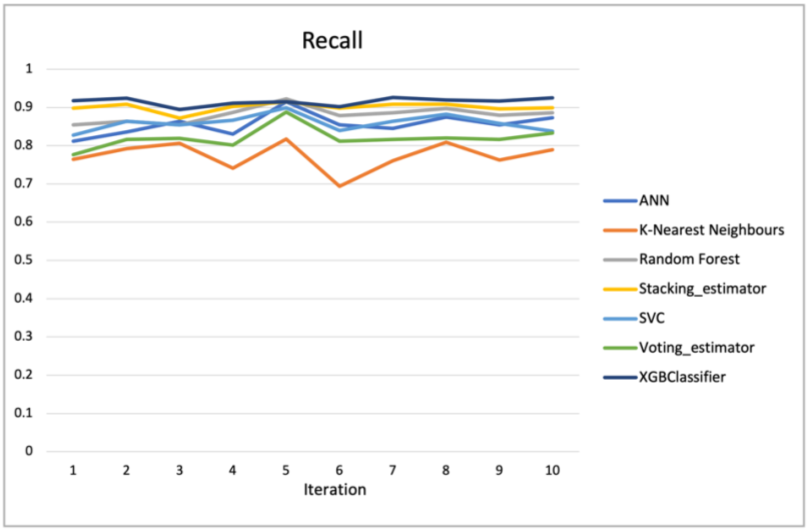

Figure 26: *Recall scores of the models on test data over 10 iterations*

Figure 25 and Figure 26 above shows the recall and precision values for the machine learning and ensemble models over 10 iterations. XGBoost and Stacking ensemble method outperform the other models, performing competitively with one another. Although the Voting ensemble method showed high precision score, the recall values were relatively low, and therefore, there will be no further examination of its performance.

An interesting observation is the trade-off between precision and recall for the XGBoost and the Stacking Ensemble model. While both models have very similar F1-score, which is a harmonic average of recall and precision, XGBoost is highly dominant in terms of the recall score and Stacking Ensemble performs best in terms of precision. Furthermore, we compare these 2 models with the use of a confusion matrix.

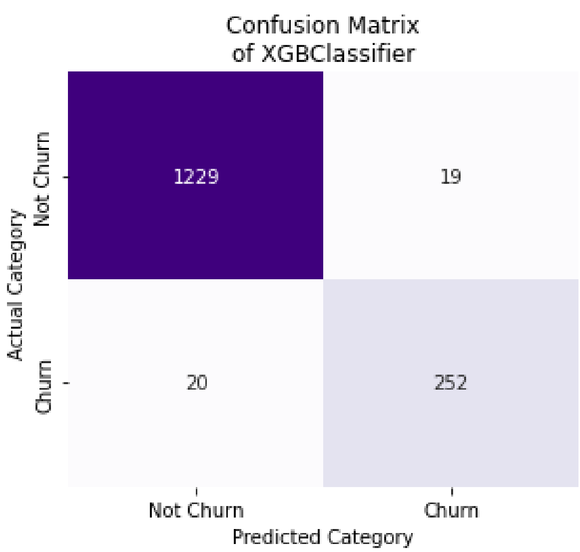

***Figure 27**: Confusion Matrix for XGBoost Classifier Model*

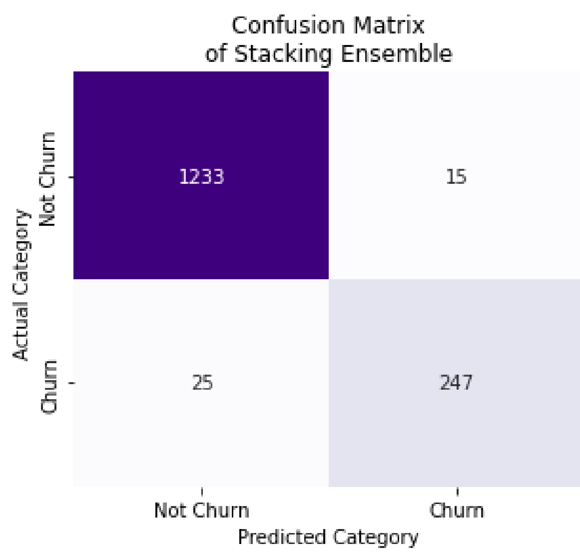

***Figure 28**: Confusion Matrix for Stacking Classifier Model*

The XGBoost model predicted more False Positive than the Stacking Ensemble model, while the Stacking model predicted more False Negatives than the XGBoost model as shown in Figure 27 and Figure 28. In practical terms, if the credit card company decides to run a marketing campaign in order to promote customer retention, they could benefit from using the XGBoost model since it has the ability to correctly capture more of the churning customers than the Stacking Ensemble.

However, if the company was running a more costly marketing campaign and is only interested in capturing a portion of the churning customer, then they could be better off using the Stacking Ensemble model since the model showed a higher precision and thus a higher confidence in its positive predictions, predicting less false positives than the XGBoost model.

c. Modelling on Randomly Sampled Data

A similar experimentation is carried out on the random sampled dataset. Table 7 shows the average scores over 10 iterations. Again, XGBoost and Stacking Ensemble perform best at predicting customer churn. However, we see that there is no significant improvement compared to the SMOTE dataset.

Table 7: *Average Performance of models on Randomly Sampled Data*

Model	Accuracy	Precision	Recall	F1 Score
K-Nearest Neighbours	0.885517	0.564764	0.413632	0.477341
ANN	0.900603	0.580855	0.781207	0.665603
SVC	0.914249	0.657734	0.672444	0.664584
Random Forest	0.947656	0.767944	0.840604	0.802420
XGBClassifier	0.952615	0.772004	0.888413	0.825933
Stacking_estimator	0.952959	0.784829	0.867770	0.823629
Voting_estimator	0.938563	0.831252	0.645862	0.726522

5.3. Feature Importance Results

In order for credit card companies to have a better understanding of their customer base, Feature Importance is determined using XGBoost and verified using Random Forest. Evaluating which attributes are more relevant in predicting churn is a top priority to service provider, as it helps understand the customer base and improve retention.

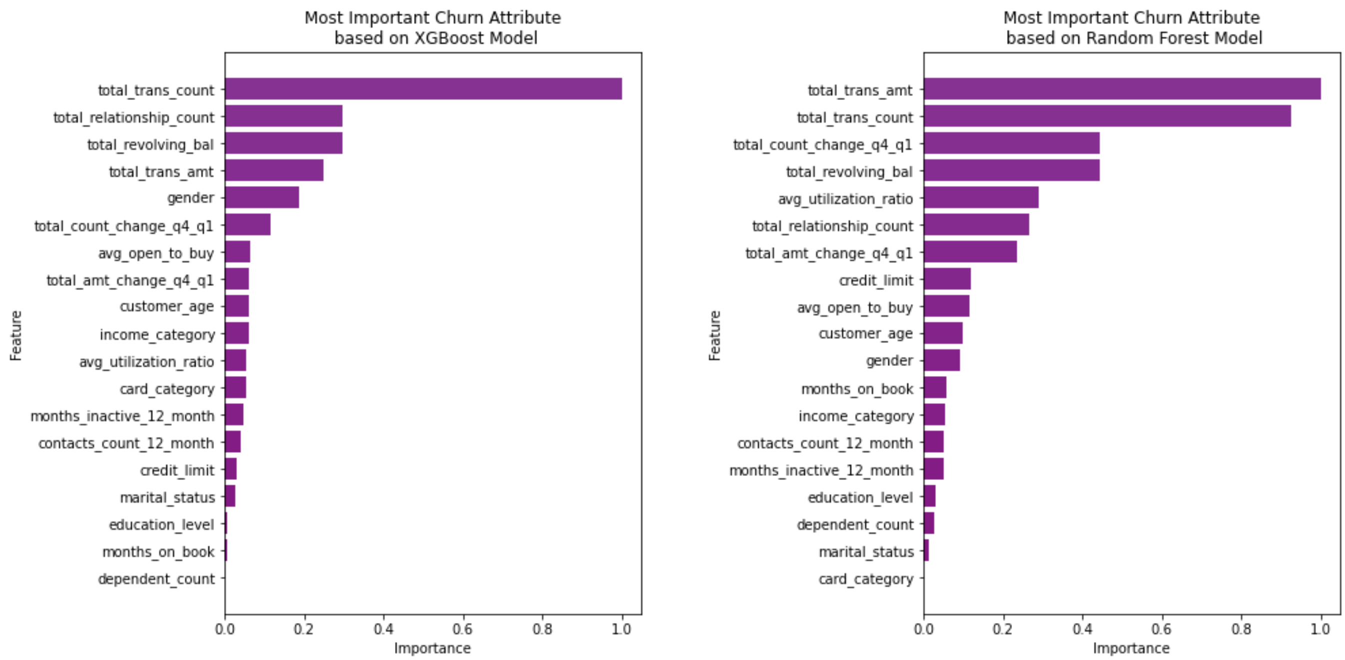

***Figure 29**: Feature importance determined by XGBoost and Random Forest Model*

Figure 29 shows that behavioural and interactive attributes such as total transaction amount in the past year, total revolving balance and total number of products owned by in the customer are most relevant in predicting if a credit card customer will churn. We also find that demographic attributes such as income category, education level, dependent count and marital status are the least relevant in determining customer churn. This is parallel to the studies conducted by Shirazi and Mohammadi, as well as Busckinx and Van den Poel, which both concluded that the clients' behavior is more significant in ascertaining their churning decisions than demographics attributes.

6. CONCLUSION AND FUTURE WORKS

The credit customer churn data has been fully exploited by conducting a predictive analysis using Machine Learning and Ensemble techniques including Artificial Neural Networks, Random Forest, Support Vector Machine, K-Nearest Neighbour, XGBoost Classifier, Stacking Ensemble using Logistic Regression as a meta-learner and Voting Ensemble. Since the data set is highly imbalanced, the analysis is conducted using 2 balancing technique to create a SMOTE dataset and a Randomly sampled dataset.

We were able to significantly improve the recall score (up to 27% for ANN) by applying SMOTE balancing technique to the training data and making predictions on the raw data. Overall, XGBoost classifier and Stacking Ensemble outperformed all other models. After 10 iterations, XGBoost classifier achieved an average accuracy of 97.06%, precision of 90.65%, recall of 91.51% and f1-score of 91.06%. The Stacking Ensemble model performed competitively with an accuracy of 97.03%, precision of 91.69%, recall of 90.05% and f1-score of 90.86% on average. Although the Voting Ensemble showed high precision values of 90.36%, the recall values were relatively low and therefore is would not be a suitable model to utilize. Compared to the SMOTE dataset, the stratified randomly sampled dataset showed no improvements.

Over the 10 iterations, we observed a trade-off between precision and recall for the XGBoost and the Stacking Ensemble model. XGBoost consistently achieved the highest recall score while Stacking Ensemble performed best in terms of precision. When deciding which model to use, the credit card company should consider its goals as well as the cost of the marketing strategy. XGBoost model has the ability to correctly capture more of the churning customers than the Stacking Ensemble. However, if the marketing campaign is more costly and if the company is only interested in capturing a portion of the churning customer, they could consider utilizing the Stacking Ensemble model as the model is able to achieve a higher precision and predict less false positives than the XGBoost model.

On evaluating the features importance, we saw that behavioural and interactive attributes such as total transaction count over the past year, total number of products held by the customer and total revolving balance are more relevant in predicting if a credit card customer will churn rather than demographic attributes such as income category, education level, dependent count and marital status.

As we have seen in this research, we have applied a novel idea and seen that Ensemble techniques such as Stacking and Voting can be very useful in predicting customer churn and have shown commendable results, however they are not well experimented in many studies within this domain. By using Stacking Ensemble technique, we were able to achieve an higher precision scores than any other model. In future works, various ensemble techniques can be experimented using other combination of machine learning techniques as well as other class balancing techniques.

7. APPENDIX

7.1. Sample File

https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers

https://leaps.analyttica.com/sample_cases/9549

7.2. Github link

https://github.com/Didi-Ojo/Credit-Card-Customer-Churn-Analysis

8. REFERENCES

[1] Verbeke, Wouter & Martens, David & Mues, Christophe & Baesens, Bart. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Syst. Appl.. 38. 2354-2364. 10.1016/j.eswa.2010.08.023.

[2] Hadden, John & Tiwari, Ashutosh & Roy, Rajkumar & Ruta, Dymitr. (2007). Computer assisted customer churn management: State-of-the-art and future trends. Computers & Operations Research. 34. 2902-2917. 10.1016/j.cor.2005.11.007.

[3] Athanassopoulos, Antreas. (2000). Customer Satisfaction Cues To Support Market Segmentation and Explain Switching Behavior. Journal of Business Research. 47. 191-207. 10.1016/S0148-2963(98)00060-5.

[4] Swift, Ronald S. *Accelerating Customer Relationships: Using Crm and Relationship Technologies*. Upper Saddle River, NJ: Prentice Hall PTR, 2000. Internet resource.

[5] Parvatiyar, Atul & Sheth, Jagdish. (2001). Customer relationship management: Emerging practice, process, and discipline. Journal of Economic and Social Research. 3. 1-34.

[6] Kracklauer, Alexander & Mills, D. & Seifert, Dirk. (2004). Collaborative Customer Relationship Management: Taking CRM to the Next Level. 10.1007/978-3-540-24710-4, pp. 3-6

[7] Au, Wai-Ho & Chan, Keith & Yao, Xin. (2004). A novel evolutionary data mining algorithm with applications to churn prediction. Evolutionary Computation, IEEE Transactions on. 7. 532 - 545. 10.1109/TEVC.2003.819264.

[8] Kracklauer, Alexander & Mills, D. & Seifert, Dirk. (2004). Customer Management as the Origin of Collaborative Customer Relationship Management. Collaborative Customer Relationship Management - Taking CRM to the Next Level. 3-6. 10.1007/978-3-540-24710-4_1.

[9] Ling, R. & Yen, D.C.. (2001). Customer relationship management: An analysis framework and implementation strategies. Journal of Computer Information Systems. 41. 82-97.

[10] Ngai, Eric & Xiu, Li & Chau, Dorothy. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert Syst. Appl.. 36. 2592-2602. 10.1016/j.eswa.2008.02.021.

[11] Hung, Shin-Yuan & Yen, David & Wang, Hsiu-Yu. (2006). Applying data mining to telecom chum management. Expert Systems with Applications. 31. 515-524. 10.1016/j.eswa.2005.09.080.

[12] Huang, Bingquan & Kechadi, Tahar & Buckley, Brian. (2012). Customer churn prediction in telecommunications. Expert Systems with Applications: An International Journal. 39. 1414-1425. 10.1016/j.eswa.2011.08.024.

[13] Buckinx, Wouter & Van den Poel, Dirk. (2005). Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. European Journal of Operational Research. 164. 252-268. 10.1016/j.ejor.2003.12.010.

[14] Miao, Xinyu & Wang, Haoran (2022) Customer Churn Analysis in Banking Sector: Evidence from Explainable Machine Learning Models

[15] Guliyev, H., and F. Yerdelen Tatoğlu. "Customer Churn Analysis in Banking Sector: Evidence from Explainable Machine Learning Models". *JOURNAL OF APPLIED MICROECONOMETRICS*, vol. 1, no. 2, Dec. 2021, pp. 85-99, doi:10.53753/jame.1.2.03.

[16] Ünlü, Kamil Demirberk. (2021). Predicting credit card customer churn using support vector machine based on Bayesian optimization. Communications Faculty Of Science University of Ankara. 70. 827-836. 10.31801/cfsuasmas.899206.

[17] Shirazi, Farid & Mohammadi, Mahbobeh (2018). A Big Data Analytics Model for Customer Churn Prediction in the Retiree Segment. International Journal of Information Management. 48. 10.1016/j.ijinfomgt.2018.10.005.

[18] Burez, Jonathan & Van den Poel, Dirk. (2008). Handling class imbalance in customer churn prediction. Expert Systems with Applications. 36. 4626-4636. 10.1016/j.eswa.2008.05.027.

[19] Japkowicz, Nathalie. (2000). The Class Imbalance Problem: Significance and Strategies. Proceedings of the 2000 International Conference on Artificial Intelligence ICAI.

[20] Chen, Chao & Breiman, Leo. (2004). Using Random Forest to Learn Imbalanced Data. University of California, Berkeley.

[21] Chawla, Nitesh & Bowyer, Kevin & Hall, Lawrence & Kegelmeyer, W.. (2002). SMOTE: Synthetic Minority Over-sampling Technique. J. Artif. Intell. Res. (JAIR). 16. 321-357. 10.1613/jair.953.

[22] HA, Thien & Bunke, Horst. (1997). Off-line, handwritten numeral recognition by perturbation method. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 19. 535 - 539. 10.1109/34.589216.

[23] Lal, Bechoo & Kumar, Suraj. (2021). Predictive Model on Churn Customers using SMOTE and XG-Boost Additive Model and Machine Learning Techniques in Telecommunication Industries. International Journal of Scientific Research in Science and Technology. 646-661. 10.32628/IJSRST218498.

[24] Hartati, Eka & Adiwijaya, Kang & Bijaksana, Moch. (2018). Handling imbalance data in churn prediction using combined SMOTE and RUS with bagging method. Journal of Physics: Conference Series. 971. 012007. 10.1088/1742-6596/971/1/012007.

[25] Sun, Zehang & Bebis, George & Miller, Ronald. (2004). Object Detection Using Feature Subset Selection. Pattern Recognition. 37. 2165-2176. 10.1016/j.patcog.2004.03.013.

[26] Naing, Yuwadi & Raheem, Mafas & Batcha, Nowshath. (2022). Feature Selection for Customer Churn Prediction: A Review on the Methods & Techniques applied in the Telecom Industry. 1-5. 10.1109/ICDCECE53908.2022.9793315.

[27] Wu, Yin & Qi, Jiayin & Wang, Chen. (2009). The Study on Feature Selection in Customer Churn Prediction Modeling. Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics. 3205-3210. 10.1109/ICSMC.2009.5346171.

[28] Biau, Gérard. (2010). Analysis of a Random Forests Model. Journal of Machine Learning Research. 13.

[29] Breiman, L. (2001). Random Forests. Machine Learning. 45. 5-32. 10.1023/A:1010950718922.

[30] Chen, Tianqi & Guestrin, Carlos. (2016). XGBoost: A Scalable Tree Boosting System. 785-794. 10.1145/2939672.2939785.

[31] Zhang, Zhongheng. (2016). Introduction to machine learning: K-nearest neighbors. Annals of Translational Medicine. 4. 218-218. 10.21037/atm.2016.03.37.

[32] Korkmaz, M., Güney, S., & Yiğiter, Ş.Y. (2012). THE IMPORTANCE OF LOGISTIC REGRESSION IMPLEMENTATIONS IN THE TURKISH LIVESTOCK SECTOR AND LOGISTIC REGRESSION IMPLEMENTATIONS/FIELDS.

[33] Park, Hyeoun-Ae. (2013). An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain. Journal of Korean Academy of Nursing. 43. 154-164. 10.4040/jkan.2013.43.2.154.

[34] Zhang, Yongli. (2012). Support Vector Machine Classification Algorithm and Its Application. 179-186. 10.1007/978-3-642-34041-3_27.

[35] Dastres, Roza & Soori, Mohsen. (2021). Artificial Neural Network Systems. International Journal of Imaging and Robotics. 21. 13-25.

[36] N B, Harikrishnan & Kathpalia, Aditi & Saha, Snehanshu & Nagaraj, Nithin. (2019). ChaosNet: A chaos based artificial neural network architecture for classification. Chaos: An Interdisciplinary Journal of Nonlinear Science. 29. 113125. 10.1063/1.5120831.