Credit Card Customer Churn Prediction

Anusha S Rao
Computer Science Engineering
PES University,BSK – 3 Campus
Bengalure, India
ctckm7@gmail.com

Moyank Giri
Computer Science Engineering
PES University,BSK – 3 Campus
Bengalure, India
moyank110@gmail.com

Moulya T
Computer Science Engineering
PES University,BSK – 3 Campus
Bengalure, India
moulyat9353@gmail.com

Ganesh Tarun S R
Computer Science Engineering
PES University,BSK – 3 Campus
Bengalure, India
ganeshtarunsr101@gmail.com

I. PROBLEM STATEMENT

The main problem this paper intends to deal with is to analyze and come up with a predictive model for predicting Credit Card Customer Churn in order to support companies suffering from customer churn such that they can detect Churning Customers as early as possible and can take necessary action in order to retain the Churning Customers.

II. INTRODUCTION

Banks fall under the Service Sector have a huge number of customers interacting with them on daily basis. Banks provides many services such as debit cards, credit cards , ATM, net banking ,UPI transactions etc. Credit card customer churn is common problem faced by any bank where the customers start to leave the credit card service provided by the bank due to multiple reasons and factors. Predicting the causes of the churn is important as it helps the banks in identifying which customer is likely to churn and what actions can be taken in order to satisfy the needs of the customer. Also, this helps in increasing the quality of service for customers leading to the retention of the customers. This prediction is vital for the banks because the customers are prioritizing quality of service provided and the main obstacle for every bank is the competition amongst them. Customers have all the liberty to choose any of the services provided by any banks based on how various factors such as how customer-friendly the bank is in terms of technology, interest rates, accessibility to the bank etc. and various other factors. It becomes the responsibility of the banks to retain their customers through various agendas for which predicting the causes for churn becomes the root. The churning of the customers is not just bounded to bank but is proliferated in almost all the service sectors. The main reason for this could be the liberty possessed by the customers or the competition among the service providers. Churning can be viewed in many ways such as it might be the number of customers dropped, ratio or percentage of customers lost in comparison with total customers a bank has and so on. Churn can be calculated on a period of a particular quarter or annually.

For Banks, predicting customer churn precisely is critical to long-term success rate. Prediction of customer churn which is accurate drives many facets. Therefore appreciable improvements in accuracy of prediction can lead to drastic improvements in Banking sector in terms of the total revenue or the profit. Another considerable challenge is the data we obtain to predict the model which may have missing data, inconsistent data or may be the feature selection which accounts for the timely data. Training and testing the model built also plays vital role in contributing to the accuracy of the model. Initially many credit card customer churn prediction models have existed, these models may not address the growth in complexity of the present world or may not be accurate enough to meet the needs of present requirements due to advancement in technology which may have led to different problem scenario. Present day models make use of automation algorithms such as random forest classifiers, SVM (support vector machine), extreme gradient boosting (XGBoost) and many more can be implemented to create extremely accurate models.

The dataset used is taken from Kaggle website. Thus, summarizing all the above points, the main aim of the project is to use effective methods to build an accurate model which can have multiple benefits to a bank or any such organization which suffers from customer churn. According to some statistics provided "A bank can increase its profits by up to 85 % by improving the retention rate by up to 5 % “. It is also much more cost effective to retain a customer than attracting new customers.

III. KEYWORDS

Churn Prediction; Logistic Regression (LR); Decision Tree (DT); Random Forest Classifier (RF); Support Vector Machine (SVM); SMOTE; Oversampling; Under sampling; XGBoost; KNN; Hybrid approach; Naive Bayes Classifier.

IV. RELATED WORK (LITERATURE SURVEY)

Hybrid data mining models approach [1]: This paper describes about working on UCI data. Primarily the data used had a lot of variation which is normalized. The data set is divided into clusters using unsupervised methods like K-means and rough k-means algorithms. The performance is measured in terms of precision, sensitivity, specificity, accuracy, and misclassification error. The insights from this paper is hybrid system works well when compared to single classifier model. Hybrid model has higher accuracy and lower misclassification error over using a single model. Of all the hybrid models SVM combined with rough k-means clustering algorithm works well with better accuracy.

One Class support vector machine (OCSVM) based undersampling [2]: This paper reflects the work on Automobile Insurance fraud dataset and Credit card customer churn dataset. Decision Tree (DT), SVM, LR, Group Method of Data Handling (GMDH), Probabilistic Neural Network (PNN) approaches are used for the classification. For Credit card customer churn dataset, undersampling with the radial basis kernel yielded significant performance with respect to DT. The paper recommends DT over other classifiers as it also yields “if-then” rules, while achieving high AUC. It also demonstrated OCSVM based undersampling. Proposed Under-sampling methodology reduced the complexity of building the system and at the same time, yielded significantly accurate results. Also, the paper gives the insight of preferring DT over SVM as there is no statistically significant difference between the two.

Data mining techniques to predict the churn: This paper works on 97% non-churned and 3% churned dataset and have used over-sampling, undersampling and Synthetic Minority Over-sampling Technique (SMOTE) to balance it. The paper is inclusive of LR, RF, SVM as the constituents of the model. Through experimentations the best results were concluded when the unbalanced original data is SMOTED, RF was implemented and for combination of undersampling and oversampling [3].

In Guangli Nie and team’s churn prediction of credit card in China’s banking industry, the paper focuses on the execution and the understanding of the model rather than building a new one [4]. The paper proposes the development of a criterion measure called misclassification cost. The paper proposes logistic regression and Decision tree classification model. Some selected variables shows that the demographic information makes little contribution to the churn prediction. This idea can be implemented in our model also. The test results shows that LR performs better than the DT. Multicollinearity has not been discussed in decision tree application.

Machine Learning approach to resolve gap of churn and non-churn customers: The paper infers on accuracy levels that can be achieved by classifiers. A novel approach KNN is proposed for grouping the data into training and testing sets depending on weighted scales along with XGBooster algorithm [5], aiming for high accuracy in model. The experiment concludes that XGBoost gives the best result in terms of accuracy, sensitivity and specificity. This brings the model designer to the conclusion that XGBoost can be used while building the model for accurate results and better forecasts.

Churn of customers in telecommunication sector can be used as backing to our problem statement. Hybrid approach [6] discussed in the paper focuses on hybrid methodology rather than the non-hybrid ones to increase the accuracy of the classifiers. Algorithms such as LOLIMOT and C5.0 are proposed in the paper. Also approaches to ANN and ICA is proposed for building a better schema. The paper concludes with a conclusion that the number of features and their subsets majorly affect the prediction accuracy. Thus, better feature selection leads to better model development is one of the major inferences.

Another paper describes a study on bank customers in India [7]. It speaks about converting raw customer data into meaningful and useful data that suits modelling. It deals with 2 classification techniques namely CART and C 5.0. While CART yielded 95.01% classification rate on training data and 91.22 per cent on test data, C5.0 yielded 69.3% and 68.4% classification rate on training data and on test data respectively. The results obtained on Churned class by CART is quite high but C 5.0 had shown poor results in predicting churned customers. Also, the paper also infers on future predictions of churns by formulating intervention strategies.

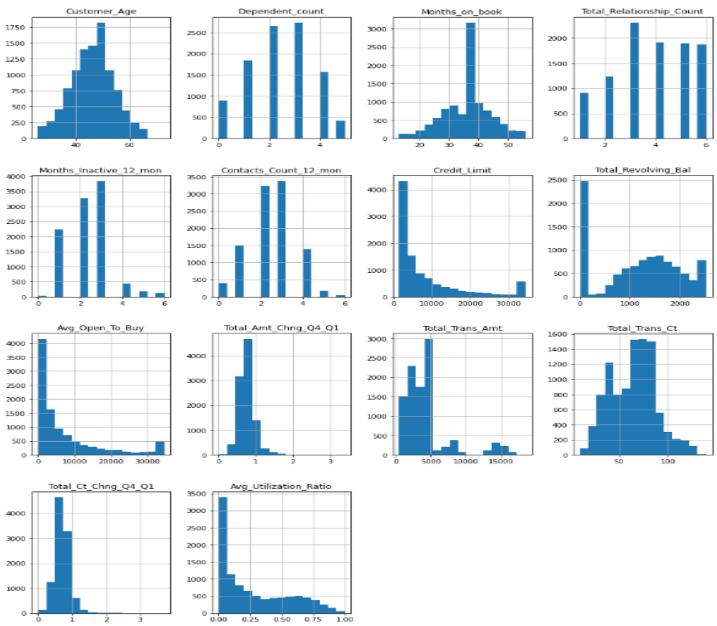
Rule extraction from Support Vector Machine is another proposed method [8]. This paper speaks about Tree being generated using Naïve Bayes Tree (NBTree) resulting in the SVM + NBTree hybrid. The data set consists of 93.11% non-churned and 6.89% churned customers. Using the original unbalanced data only the observation proposed hybrid SVM + NBTree yielded the best sensitivity compared to other classifiers. The paper recommends that it is better to support vectors and use case-sp (sensitivity-68.33%, specificity-74.38%, accuracy- 75.18%) to generate rules

Deep Learning model [9] is also one of the proposed solutions. Techniques such as LSTM can be employed for time series data prediction. Bi-LSTM allows sequence time step information in both forward and backward direction. Customers transaction details are created as features and are passed to the model. For each feature Recency, frequency, and monetary features are extracted and the model is allowed to learn from the pattern. The paper gives the insight on improved churn prediction. The idea of Down sampling and up sampling is being implemented for raw data set using REHC and SMOTE. Results shows better performance when compared with other deep learning models.

V. INITIAL INSIGHTS

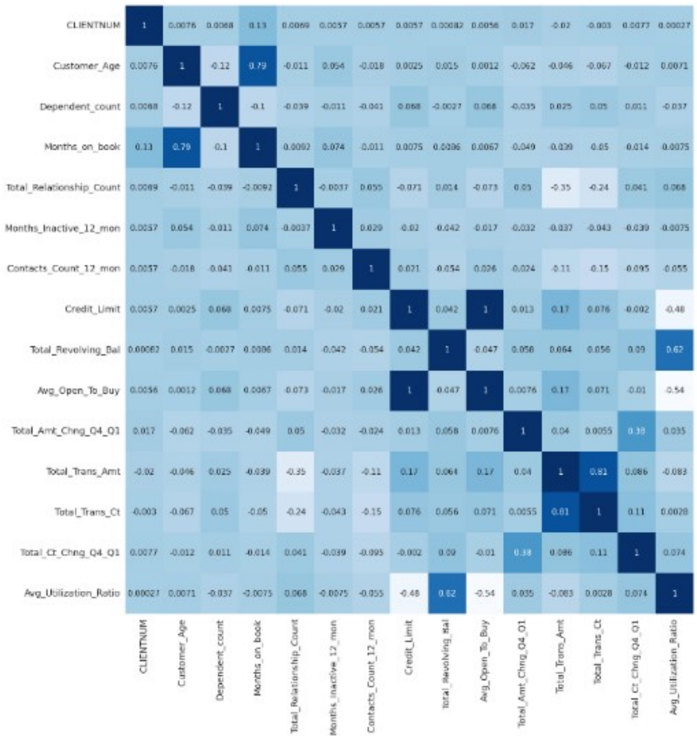
In order to completely understand the problem and to tackle it effectively we need to initially analyse our data and come up with useful insights in order to make the best use of the data at hand.

The first step for our analysis was to check the distribution of each Numerical Column in the dataset. The numerical columns include Customer Age, Dependent count, Months on book, Total Relationship Count etc. The following shows the distribution of each numerical attribute using a histogram:



As seen above that not all attributes are normally distributed and many columns seem to be skewed either towards the left or the right. Hence as an initial insight to the data we can infer that the data is not normally distributed and might require some transforms to analyze the data efficiently as a normal distribution.

A correlation analysis was also conducted over the complete dataset in order to get insights to check which variables are significantly correlated to each other. For this we make use of a heatmap representation where we annotate each cell with their correlation value as shown below



The above heatmap can be interpreted as for any cell, row variable correlation with the column variable is the value of the cell. The correlation of a variable with itself is 1. For that reason, all the diagonal values are 1.00.

For the analysis, we consider that a correlation of above 0.3 is a significant positive correlation while a correlation of values between 0 and 0.3 is not a significant positive correlation. Similarly, we consider that a correlation of below -0.3 is a significant negative correlation while a correlation of values between 0 and -0.3 is not a significant negative correlation.

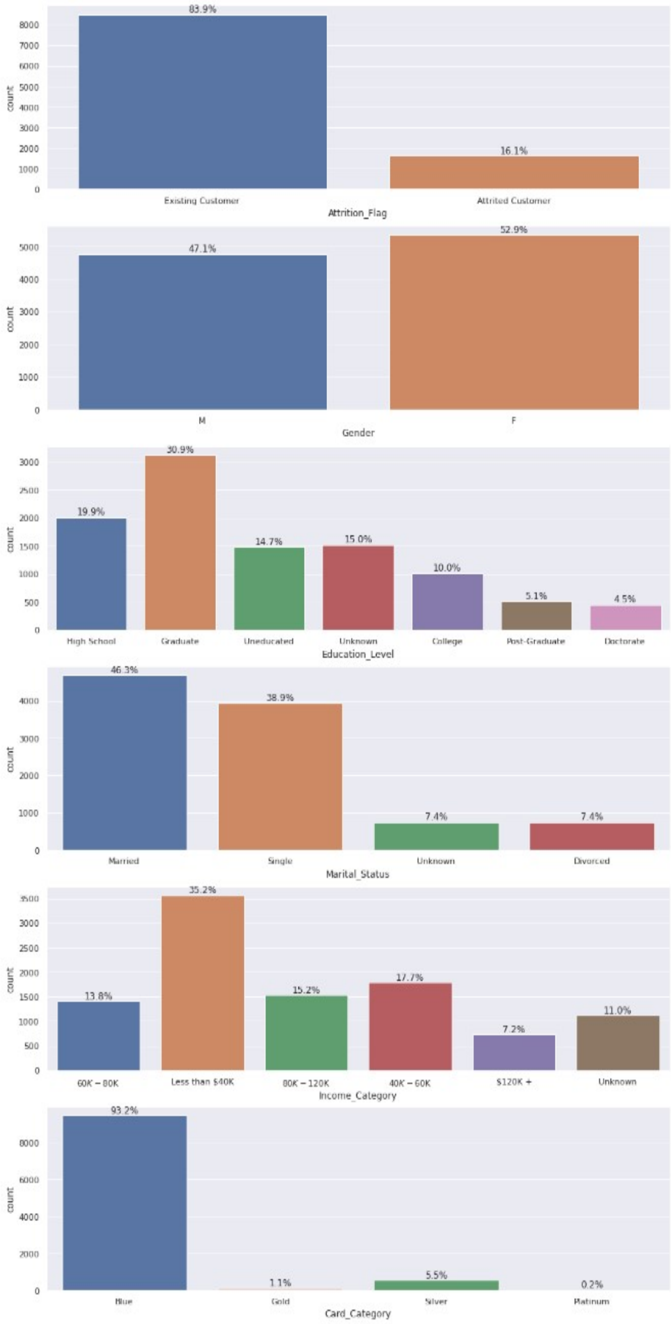
For the above correlation analysis, we can infer the following insights in the data:

1. There is a High Positive correlation between the attribute total transaction amount and total transaction count
2. There is a High Positive Correlation between Customer Age and Period of Relationship with Bank.
3. There is a High Positive Correlation between Total Revolving Balance on Credit Card and Average Card Utilization Ratio.
4. There is a Moderately High Positive Correlation between Change in Transaction Amount (Q4 over Q1) and Change in Transaction Count (Q4 over Q1)
5. There is a Moderate negative correlation between credit limit and average utilization ratio

6. There is a Moderate Negative correlation between Open to Buy Credit Line (Average of last 12 months) and average utilization ratio.

7. There is a Moderate Negative correlation between Total Transaction Count (Last 12 months) and Total no. of products held by the customer.

In order to analyse the categorical attributes, we make use of Count plots to show its distribution over different classes within the attribute:



VI. PROPOSED APPROACH

Based on the above insights and the literature survey we can come up with different approaches to build our prediction model. We can go for different classification techniques and adapt the one which gives us the best results. Classification techniques like Decision tree, Random Forest Classifier, Naïve Bayes Classifier etc yield accurate results when worked upon individually or when combined together. From the literature survey made we can infer those different classifiers have yielded different accuracy. This shows that the data set we use also has the impact on the classifiers we tend to implement.

Our Approach intends to make use of an ensemble model for prediction which intends to provide predictions based on multiple models rather than just using a single model.

The main advantage of using an ensemble method is that an ensemble model in general can give high accuracies while utilizing simpler models for its prediction. Making use of simpler models also helps in increasing the speed at which predictions are made while improving upon accuracy of the prediction if used only a single model. Ensemble models combine predictions from all models using a "vote" based method where predictions from all models are taken and combined to produce a single better prediction. Our ensemble model approach could consist of models such as Decision Tree, XGboost, Random Forest, SVM, Logistic Regression, Naive Bayes and many more to yield better and more accurate results within shorter durations.

Here are few advantages that can be seen with each of the above models mentioned above:

XGBoost: The benefit of gradient boosting is that they can automatically provide estimates of feature importance from a trained predictive model. XGboost is one of the implementations of gradient boosting concept, it uses a more regularized model formalization to control over-fitting, which gives it better performance. XGBoost can be used to train a standalone random forest. XGB consists of a number of hyper-parameters that can be tuned, has an in-built capability to handle missing values. It provides various intuitive features, such as parallelization, distributed computing, cache optimisation and more. After a boosting tree is constructed, it retrieves feature importance scores for each attribute. The feature importance contributes a score which indicates how much valuable each feature was in the construction of the boosted decision trees within the model.

Decision Tree: This algorithm is one of the supervised learning algorithms which can be used for both classification and regression problems. In this algorithm a training model is created that predicts the class or value of target variable by learning decision rules gathered from training data. It uses a flowchart kind of tree structure which shows the predictions from a chain of feature-based splits. The approach for construction decision tree is usually top-down in which a variable is chosen at each step that splits the set of items. In Decision tree there are two types of nodes namely Decision Node and Leaf Node. Leaf Nodes are the outputs of decision nodes and do not contain branches whereas decision nodes can be used to make any decisions and take branch.

SVM: This is supervised machine learning algorithm that can be used for both classification or regression challenges. SVM algorithm has a technique called the kernel trick. The SVM kernel function takes low dimensional input space and transforms it to a higher dimensional space. SVM offers good accuracy and perform faster prediction. They also use less memory because they use a subset of training points in the decision phase. SVM works well with a clear margin of separation and with high dimensional space. It constructs a hyperplane in an iterative manner, which is used to minimize an error. The core idea of SVM is to find maximum marginal hyperplane that best divides the dataset into classes.

Naive Bayes algorithm: The naïve Bayes Classifier is a probabilistic model which functions on the Bayes theorem. This algorithm makes an assumption that all the variables in the dataset is naive i.e., not correlated to each other. It performs well in multi-class prediction, categorical input variables compared to numerical variables. The assumption that all features are independent makes naive bayes algorithm very fast compared to complicated algorithms. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Random forest Classifier: This is a supervised machine learning algorithm which can be used for both regression and classification. It is an ensemble method which combines the results of small decision trees called estimators and combines the predictions of these estimators to produce more accurate prediction. These are generally used as black box models as they produce predictions across a wide range of data with little configuration. There are two methods for random forest to ensure that the behaviour of each individual tree is not correlated with behaviour of other trees in the model. They are Bagging and Feature randomness. Bagging is a process of allowing each individual tree to randomly sample from the dataset with replacement. In Feature randomness each tree in a random forest can pick only from a random subset of features.

REFERENCES

[1] R. Rajamohamed ,J. Manoaran , ' *Improved credit card churn prediction based on rough clustering and supervised learning technique*',2018 https://link.springer.com/article/10.1007/s10586-017-0933-1

[2]G. Ganesh Sundarkumar, Vadlamani Ravi, V. Siddeshwar,' *One-Class Support Vector Machine based undersampling: Application to Churn prediction and Insurance Fraud detection* ',2015 https://ieeexplore.ieee.org/abstract/document/7435726

[3] Dudyala Anil Kumar and V. Ravi *,'Predicting credit card customer churn in banks using data mining'*, 2008 https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1086.1056&rep=rep1&type=pdf

[4] Guangli Nie , Wei Rowe , Lingling Zhang , Yingjie Tian , Yong Shi , ' *Credit card churn forecasting by logistic regression and decision tree* ',2011 https://www.researchgate.net/profile/Xingsen-Li/publication/224759968_The_Analysis_on_the_Customers_Churn_of_Charge_Email_Based_on_Data_Mining_Take_One_Internet_Company_for_Example/links/5891e87a92851cda256a0358/The-Analysis-on-the-Customers-Churn-of-Charge-Email-Based-on-Data-Mining-Take-One-Internet-Company-for-Example.pdf

[5] Hemlata Dalmia, Ch V S S Nikil , Sandeep Kumar ,' *Churning of Bank Customers Using Supervised Learning* ',2020 www.researchgate.net/profile/Sandeep-Kumar-249/publication/340855263_Churning_of_Bank_Customers_Using_Supervised_Learning/links/5f54b5ea92851c250b96c697/Churning-of-Bank-Customers-Using-Supervised-Learning.pdf

[6] Elham Jamalian, Rahim Foukerdi,' *A Hybrid Data Mining Method for Customer Churn Prediction* ',2018 https://etasr.com/index.php/ETASR/article/view/2108/pdf

[7] Dr. U. Devi Prasad, S. Madhavi,' *PREDICTION OF CHURN BEHAVIOR OF BANK CUSTOMERS USING DATA MINING TOOLS* ',2012 https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.456.4094&rep=rep1&type=pdf#page=100

[8] M.A.H. Farquad1, V. Ravi1,S. Bapi Raju,' *Data Mining Using Rules Extracted from SVM: An Application to Churn Prediction in Bank Credit Cards*',2009 https://link.springer.com/chapter/10.1007/978-3-642-10646-0_47

[9] Mr. M. John Britto, Dr. R. Gobinath,' *Improved Churn Prediction Model In Banking Industry And Comparison Of Deep Learning Algorithms* ',2021 http://www.journal-aquaticscience.com/article_133719_d883304ae722b3c710a2722f1f0d3e7a.pdf