Customer Churn Analysis Using Association Rule Mining and Decision Tree Classifiers

By

Rifat Bin Alam Rohit 012202042

Submitted in partial fulfilment of the requirements
of the degree of Master of Science in Computer Science and Engineering

December 4, 2021

Department of Computer Science and Engineering
United International University

Abstract

Customer churn is a prominent issue facing companies. Therefore, preventing customer churn and retaining and retaining customers has become an essential issue for business operations and development. This paper aims to identify the reasons for customer churn for a prominent logistics company by using Apriori association rule mining. The expected output will be used by the business users to understand where they have the gaps in their business processes. The results from the Decision Tree and Apriori Algorithm shed light on which business feature was most prominent for causing churn.

Acknowledgements

This work would have not been possible without the input and support of many people over the last two trimesters. I would like to express my gratitude to everyone who contributed to it in some way or other.

First, I would like to thank my academic advisors, Dr.Swakkhar Shatabda, for giving me his guidance throughout the project duration.

Last but not the least, I owe to my family including my parents for their unconditional love and immense emotional support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter discusses the definition of churn and how different churned customers are classified based on their engagement with the business. This chapter also discusses the motivation and objective of the project.

1.1 Project Overview

Customers' liquidity is constantly expanding as a consequence of increased market competition brought on by free trade agreements and globalization of the economic system. A customer churn is a term used to describe a customer's decision to quit using or buying a company's goods or services. These three sorts of churners were discovered in a prior investigation. [1] :

1. Active churner: these clients are actively churning, which means that they are actively looking for a new provider.
2. Passive churner: Only if the originating firm ends the contract will these customers actively cease the business connection.
3. Potential churner: Without the company's knowledge, these clients will end their contracts. It is possible to anticipate the first two categories of churned clients using human approaches. Potential churn clients, on the other hand, are more difficult to forecast due to the complexity of the previous data. The third kind of churner is what the customer churn prediction model is aiming to forecast.

1.2 Motivation

Most businesses suffer from not having readily available solutions that can help them identify the reasons for customers churning. Here, 'reasons' mean the point in a customer journey with the business the customer decided to churn. From data alone, it is not readily understandable as to why the customer churned. In most cases, the data is not presented

in a properly structured way for non-technical users to understand, and the solutions in the market such as DataRobot[6], RapidMiner[7] that do provide these solutions but are expensive to scale and requires client companies to hand over user data for them to generate the model and give the prediction/reasons.

1.3 Objectives

The objective of this project is –

1. Build an easy to use interface for business managers to identify reasons for churn.
2. Create a machine learning model that uses Apriori Algorithm to identify churn reasons from user data.
3. Convert machine learning data into an understandable visualization for end business users.

1.4 Methodology

This project is done using the customer data set of 12,000 users gathered from a delivery logistics company. The data set was derived from the database server using BigQuery, and we subsequently cleaned the data to ensure there are no duplicates. We initially did some descriptive analytics on the data set to identify the number of features in the dataset and how it correlated with the churn data. Then we moved on to selecting the top correlated features and used that as our data set for further prediction analysis.

1.5 Project Outcome

An automated script that any non technical user can run on their own devices to do the churn reason analysis of their customers.

1.6 Organization of the Report

The material presented in this report is organized into 5 chapters. After this introductory chapter, the information in chapter is presented below -

1. **Chapter 2** describes the background behind the project by first explaining the prerequisites required to understand the report and then detailing the literature review and gap analysis of similar tools in the market.
2. **Chapter 3** details the methodology of the entire project covering the system architecture and the dataset details.This chapter also explains the two machine learning models used in this project.

3. **Chapter 4** explores the result output from each of the machine learning models and also gives a step by step guide for using the script to generate results from other datasets.

4. **Chapter 5** summarises the results of the project and provides the limitations faced during the project while outlining the scopes of future work that can be done.

Chapter 2

Background

This chapter explains in details the literature review done for the purpose of this project. As part of the literature review, existing solutions in the market were analysed along the criteria of cost, data security and usability.

2.1 Preliminaries

Basic understanding of what is meant by customer churn, supervised machine learning models (mainly decision trees), association rule mining and apriori algorithm is needed to understand the report.

2.1.1 Customer Churn

The churn rate, also known as the rate of attrition or customer churn, is the rate at which customers stop doing business with an entity. It is most commonly expressed as the percentage of service subscribers who discontinue their subscriptions within a given time period. It is also the rate at which employees leave their jobs within a certain period. For a company to expand its clientele, its growth rate (measured by the number of new customers) must exceed its churn rate.

2.1.2 Association Rule Mining

Association Rule Mining, as the name suggests, association rules are simple If/Then statements that help discover relationships between seemingly independent relational databases or other data repositories.

2.1.3 Apriori Algorithm

Apriori is an algorithm for frequent item set mining and association rule learning over relational databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used

to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.

2.2 Literature Review

Predicting client attrition is a two-way street, according to past study findings. One group of academics focuses on developing sophisticated algorithms in an effort to improve their predicting abilities. He et al. [2] presented an SVM and random sampling-based prediction model. Using a different sample distribution, random sampling may be used to correct for class imbalance. The prediction model is then built using the SVM. A classification technique based on the CSCUM chart was proposed by Chen Chen et al. [3]and colleagues to deal with class imbalance. The inter-arrival time (IAT) is all that is needed for this approach to predict churn for individual monitoring.In the paper by Gordini et al. [4] SVM was used to predict customer attrition using the AUC parameter-selection approach. The accuracy of the predictions was shown to be strongly influenced by parameter optimization in this study. The usual heuristic management technique is superior to the data-driven algorithm and retention plan.The Exhaustive Algorithm (EA), Genetic Algorithm (GA), Covering Algorithm (CA), and LEM2 Algorithm were used by Amin et al. [1] to develop an intelligent rule-based strategy for extracting churn customer choice criteria (LA). First, Stripling et al. [5] included the notion of profit maximization into the prediction of customer turnover using genetic algorithms to optimize the projected maximum profit measure (EMPC).Wang et al. [6] investigated how the GBDT predicts future attrition based on client activity in search ads. GBDT dynamic and static characteristics are first extracted using this approach. Based on these two characteristics, the GBDT prediction model is then built. It has been suggested by Amin et al. [7] that the classification certainty of distinct areas in the dataset may be estimated using the distance factor. Investigators are also looking into client turnover to find out what causes it. Weighted random forest (WRF) was utilized by Xie et al. [8] to forecast customer attrition. Not only does this approach better tackle the issue of class balance, but it also preserves its readability. In 2011, a prediction model based on rotation forest and Adaboost was presented by De Bock et al [9]. The Adaboost approach is used to enhance forecast performance while the rotating forest extracts consumer information. The experimental findings show that the prediction performance has been much improved. However, the model's interpret ability and the churn factors' comprehension are lacking. Because of this, De Bock et al. [10] cited in the previous line performed a second investigation that combined generalized additive models (GAM) with an ensemble classification technique to get their conclusions. Predictive accuracy was improved, and the method's outstanding interpretability was shown in the experiments.

Using data from social networks to forecast customer turnover was studied by Verbeke et al. [11]. This technique employs social network effects to deal with enormous networks, a time-dependent class label, and an uneven distribution of class members. Non-Markov

network effects were included into relational classifiers as part of this study, and a new parallel modeling technique combining relational and non-relational classifiers was also developed. Three causes are to blame for a poor usage rate of relevant information in social networks. Due to the complexity of networks and the absence of relevant tools, network characterisation is time-consuming. As a second point, the computational cost of extracting the structural properties of large-scale networks is rather significant. To round things out, the vast majority of a network's dynamic elements are simply ignored. Since this is the case, Mitrovic [12] have developed a method of learning based on panoptic representations that include both interactive and structural data. By splitting the data into distinct time periods, this technique may accommodate for various temporal granularities. A methodology based on interpretable user grouping and churn prediction was developed by Yang et al [13]. The initial step in this approach is to categorize users according to their daily activities and ego network topologies. An attention mechanism and long short-term memory (LSTM) deep learning pipeline are then devised. It was shown that this paradigm helps researchers understand user behavior and outperforms other predictive methodologies. There are two steps to this hybrid classification algorithm: segmentation and prediction. It was developed by Caigny et al. [14]. It is used to identify customer segments in the segmentation stage and in the prediction stage, each leaf of the tree is given a unique LR model.

2.2.1 Similar Applications

As mentioned in the earlier chapters two industry standard tools were researched for the purpose of this project. They are listed below -

1. DataRobot - DataRobot enables organizations to leverage the transformational power of AI by delivering the world's only AI Cloud platform.

2. RapidMiner - RapidMiner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. RapidMiner charges customers based on their use cases and data size however, they also require customer data to be sent to them for analysis.

2.2.2 Related Research

In Arivazhagan et al. [15] they have explored using regression modeling with Bayesian Boosting methods to predict which customers will churn from an existing data set. Using this method they had an accuracy rate of 89% using Logistic Regression and 95% using Bayesian Boosting with Logistic Regression.Makhtar et al. [16] uses Rough Set theory to classify customers as churned or not churned. Using this technique the authors have

gotten 94% accuracy in prediction customer churn. In Hong J et al. [17] uses the Apriori Association Mining rule to discover insights from truck crash characteristics.

2.3 Gap Analysis

This section explains the gaps identified with the two applications named above.

1. DataRobot - To achieve the full potential of DataRobot, the customer data has to be given to DataRobot. Additionally DataRobot charges USD 40,000 per year for their first tier of service.

2. RapidMiner - RapidMiner charges customers based on their use cases and data size however, they also require customer data to be sent to them for analysis.

As both services require user data to be sent to them it introduces a level of risk for the business as data security of those services can be compromised. Moreover, the yearly cost of using these services even at their most basic tier is a huge cost for many smaller companies especially from South Asian countries.

Chapter 3

Detailed Methodology

This chapter explains in detail the architecture of the python script developed for this project. This chapter also outlines the steps carried out to arrive at the final result.

3.1 System Architecture

The python script developed works in the method as outlined in the image below.

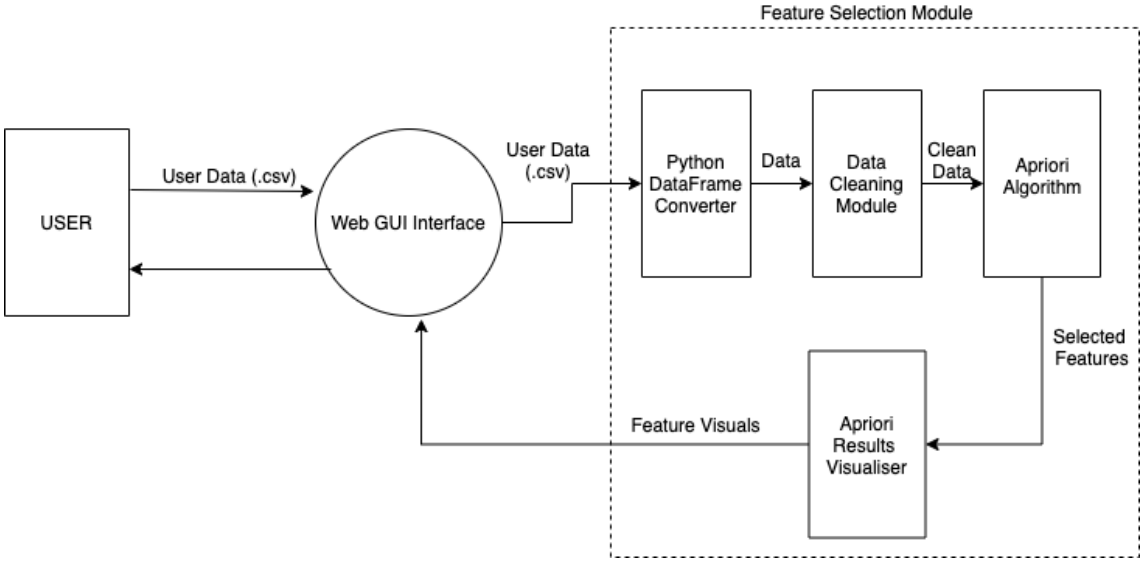

Figure 3.1: System Architecture

3.2 Dataset Details

The dataset used for this project was taken from a logistics company. The dataset contained churned customer data of **58,606** customers. Each individual customer had **29** features. The table below shows the data type of each feature in the dataset.

Feature Name	**Data Type**
active_length	Integer
is_churn	Integer
Parcel_count	Integer
Average_days_between_parcels	Float
ISD	Integer
SUBURBS	Integer
OSD	Integer
Success_Rate_of_ISD	Float
Success_Rate_of_SUBURB	Float
Success_Rate_of_OSD_HUB	Float
SLA_AVG	Float
Issue_parcels	Integer
number_of_times_issues_raised	Integer
Issue_resolution_tat	Float
second_mile_tat_hr	Float
E2E_TAT_day	Float
E2E_TAT_ISD_day	Float
E2E_TAT_SUB_day	Float
E2E_TAT_OSD_day	Float
Extend_of_breach_in_hr	Float
Extend_of_breach_in_hr_ISD	Float
Extend_of_breach_in_hr_SUB	Float
Extend_of_breach_in_hr_OSD	Float
return_tat_in_days	Float
issues_raised	Integer
issues_resolved	Integer
area_changes	Integer
damaged	Integer
creation_to_received_in_hours	Float

Table 3.1: Feature Data-types.

3.3 Project Methodology

Listed below is the step-by-step methodology on how the project will be completed.

- Data collection.
- Data Processing.
- Exploratory Data Analytics.
- Feature Selection
- Use decision tree and correlation mapping with churn to identify suitable features.
- Encode feature values into discrete values for algorithm to properly function.

- Configure Apriori Algorithm with selected features.
- Create a script for business users to use which automates the entire process

3.4 Dataset Description

In our project, we gave priority to the 'accuracy of the model. The dataset used had the following features relating to each customer -

- **Active Length** - This is the amount of time in days the user was active in the platform
- **is churn** - This is the label used to identify if the customer has churned. 1 denotes the customer has churned and 0 means the customer has not churned.
- **Parcel Count** - This is the number of parcels the customer has given to the company to deliver over their active length.
- **Average Days between parcels** - This is the number of days between a two parcel booking dates for a customer.
- **ISD, SUB, OSD** - These are the number of parcels delivered to each of the zones. ISD stands for INSIDE DHAKA, SUB stands for SUBURB Areas, OSD stands for Outside Dhaka.
- **SLA AVG** - This is the percentage of compliance of the Service Level Agreement with the customer.
- **Issue Parcels** - This is the number of parcels for which the customer has raised an issue.
- **Number of time issue raised** - This is the number of times an issue had to be raised for an user.
- **Issue Resolution TAT** - This is the amount of time it took for the issue to be resolved.
- **E2E TAT** - End-to-End TAT is the measurement in days of how long it took for the parcel to be delivered to its destination from the moment it was picked up.
- **Extend of Breach in Hour** - This is the measurement of how much extra time in hours it took for the parcel to be delivered, beyond the standard delivery time.
- **Return TAT in Hour** - This is the measurement in hours it took for the parcel to be returned to the user if the parcel is not delivered.
- **Damaged** - No. of parcels that were damaged.
- **Creation to Received** - Amount of time it took for the parcel to be picked from the customer's address.

3.5 Analysis

For analysing the existing data, we first imported the data into a pandas data frame. We did some descriptive analysis on the data to see the data types of each of the columns and gauge the balance of the dataset.

We identified the dataset to be imbalanced as shown in figure 3.2. Here, customers who have churned are classified as "1," and customers who have not churned are classified as "0."

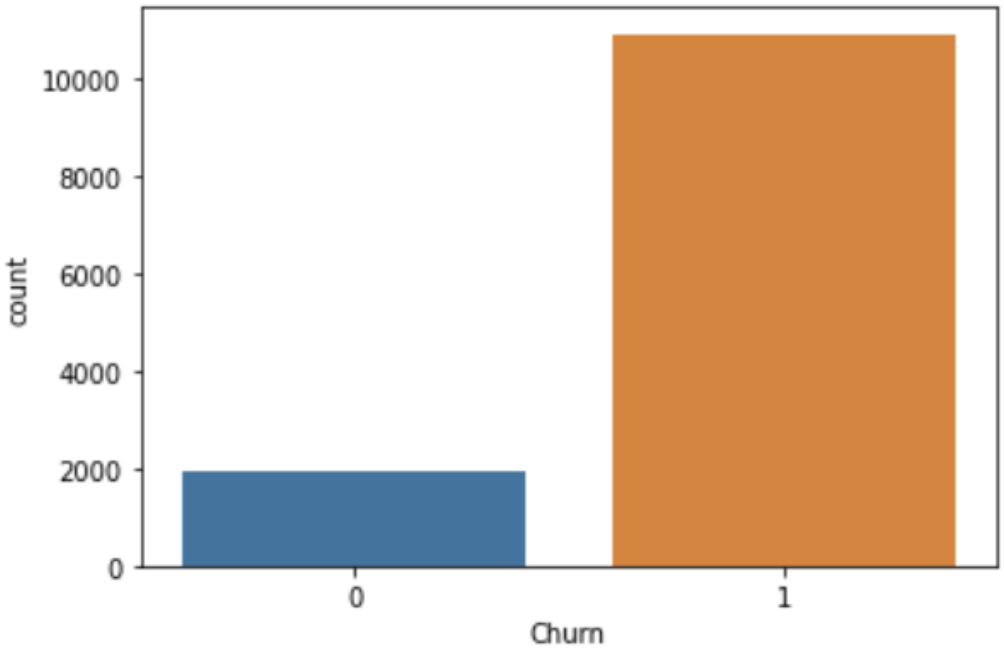

Figure 3.2: Distribution of customers across churned and not-churned labels

To reduce the computational complexity, we ran a correlation analysis of the feature in the dataset with the churn classification of the customers. The resulting correlation matrix gave us insight into the correlation distribution of each feature concerning churn as shown in figure 3.3.

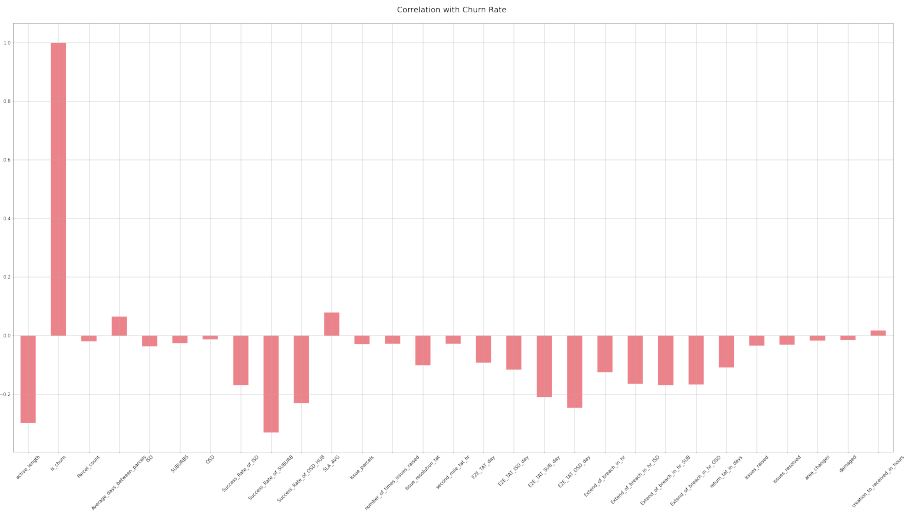

Figure 3.3: Feature Correlation matrix with churn

3.6 Decision Tree

Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split.

Mechanism

The decision trees use the CART algorithm (Classification and Regression Trees). In both cases, decisions are based on conditions on any of the features. The internal nodes represent the conditions and the leaf nodes represent the decision based on the conditions.

Impurity measurements

The cost function of a decision tree can be described by

$$J\left(k, t_k\right) = \frac{m_{left}}{m} I_{left} + \frac{m_{rig'h}}{m} I_{right}$$

where I_{node} is the impurity metric for the left and the right path in the tree, m_{node} is the number of samples on either the left or the right side of the subtree, and m is the total number of samples in the subtree.

Gini Impurity

The Gini impurity is defined as

$$G_i = 1 - \sum_{k=1}^{n} p_{i,k}^2$$

where p_k is the ratio of class k samples among all the training samples in the i:th node.

Shannon Entropy

The Shannon entropy is defined as

$$H_i = -\sum_{k=1, p_{i,k} \neq 0}^{n} p_{i,k} \log_2 (p_{i,k})$$

Mean Squared Error (MSE)

The mean squared error is given by

$$\mathrm{MSE}_{\mathrm{node}} = \begin{cases} \mathrm{MSE}_{\mathrm{node}} = \sum_{i \in node} \left(\hat{y}_{\mathrm{node}} - y^{(i)}\right)^2 \\ \hat{y}_{\mathrm{node}} = \frac{1}{m_{\mathrm{nade}}} \sum_{i \in \mathrm{node}} y^{(i)} \end{cases}$$

3.7 Apriori Algorithm

Association rules mining (ARM) is a data mining technique that identifies a group of objects that occur together in a single occurrence. Because of its exploratory and simple nature, the Apriori algorithm is well known for discovering association rules.

Using the Apriori method to find association rules requires a "bottom-up" strategy. The Apriori method is based on the idea that a k-itemset is frequent if and only if each item in the item-set is also frequent.

When extracting interesting rules from a dataset, there are two major phases required. The initial stage is to generate item-sets on a regular basis. The method initially examines the database for item-sets that meet a preset minimum level of *support*. The algorithm then creates rules that are more confident than a predetermined minimum *confidence* level. In ARM, *support* and *confidence* are critical notions for picking significant rules from a large set of potential rules. The support measures the credibility or strength of the rule by estimating the probability $P(A|B)$, which is interpreted as the share of cases in which the consequent occurs given that the antecedent has occurred. The confidence measures the credibility or strength of the rule by estimating the probability $P(A|B)$, which is interpreted as the share of cases in which the consequent occurs given that the antecedent has occurred. The following equations can be used to determine support and confidence:

$$\text{support } (A \Longrightarrow B) = P(A \cap B) = \frac{|A \cup B|}{|D|}$$

$$\text{confidence } (A \Longrightarrow B) = \frac{\text{support } (A \Longrightarrow B)}{\text{support}(A)} = \frac{P(A \cap B)}{P(A)}.$$

ARM methods may generate a vast collection of rules that satisfy the preset criteria for both and, depending on the dataset being analyzed. Confidence is inadequate when it fails to take the baseline frequency of the consequent into account. As a result, another measure called as lift was suggested to circumvent the aforementioned restrictions by adding the frequency of the consequent in its equation, as shown in the formula:

$$\text{lift}(A \Longrightarrow B) = \frac{\text{confidence } (A \Longrightarrow B)}{\text{support}(A)} = \frac{P(A \cap B)}{P(A)P(B)}$$

The rule $A \rightarrow B$ lift indicates how much the likelihood of B increases if A occurs. There are three examples. When $lift(A \rightarrow B)$ is more than one, there is a positive dependency between the antecedent and consequent, and the rule is considered useful. There is a negative dependency between the antecedent and the consequent with $lift(A \rightarrow B) < 1$. Finally, when $lift(A \rightarrow B) = 1$, A and B are independent and have no association. The higher the lift measure, the more interesting the produced rules are. We ordered the rules that fulfilled the minimal support and confidence requirements using this metric.

Chapter 4

Results and Discussion

This chapter explores the outputs from the decision tree classifier and the apriori algorithm. Along with the decision tree, we have been able to get the feature importance score. The feature importance score tells us which feature is the most important accordign to the dataset.

4.1 Decision Tree Output

From the dataset used the following diagram was extracted from the decision tree classifier used with a max depth of 3.

As shown in figure 4.1 it is observed that the primary feature which had the highest entropy value was parcel count and thus it was selected as the root node. From this root node based on the entropy value two branch nodes were observed. This decision tree tells us that creation to received time, active length, E2E tat are the major features that the business should consider monitoring in order to reduce customer churn. This recommendation is further reinforced by calculation the feature importance score as shown in table 4.1.

Variable	Feature Importance Score
Parcel count	0.667099
E2E TAT day	0.210618
active length	0.081136
E2E TAT OSD day	0.028442
creation to received in hours	0.012705
E2E TAT SUB day	0.000000

Table 4.1: Decision Tree Feature Importance Score

4.2 Apriori Algorithm

For the purpose of this paper Apriori Algorithm was used to give the association rules that accurately shows the user the reasons for customer churn. As defined in the Data

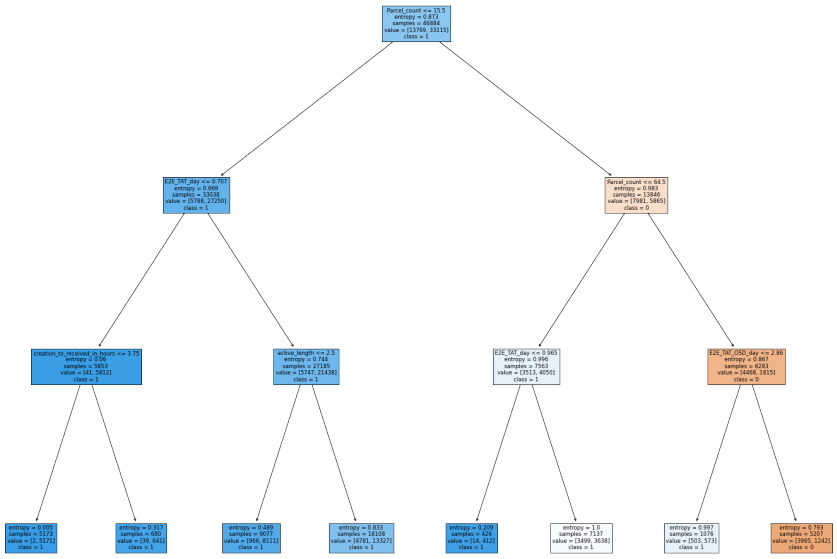

Figure 4.1: Decision Tree Diagram output

description section, *ischurn* is used to denote if a customer has churned or not, thus in the rules from the Apriori Model output, we have filtered results that exclusively contain *ischurn* as *consequent* as show in table 4.2

antecedents	**consequents**	**support**	**confidence**	**lift**
Success_Rate_of_ISD	is_churn	0.329969285	0.339503386	0.999648859
Success_Rate_of_OSD_HUB	is_churn	0.318121983	0.329695316	0.970769542
Success_Rate_of_ISD, Success_Rate_of_OSD_HUB	is_churn	0.308468627	0.329274005	0.969529014
Success_Rate_of_SUBURB	is_churn	0.308907416	0.328971963	0.968639668
Success_Rate_of_ISD, Success_Rate_of_SUBURB	is_churn	0.300131637	0.32837254	0.9668747
Success_Rate_of_OSD_HUB, Success_Rate_of_SUBURB	is_churn	0.292672225	0.320365034	0.943297043
Success_Rate_of_ISD, Success_Rate_of_SUBURB, Success_Rate_of_OSD_HUB	is_churn	0.283896446	0.319506173	0.940768176

Table 4.2: Customer Churn rules extracted from Apriori Algorithm.

4.3 Step By Step Guide for Script Operation

The python script can be executed very easily from either Windows Or MacOS systems using the command prompt on Windows or Terminal on MacOS. Below is a step by step guide on how to run the script on a MacOS system. The similar steps can be followed for Windows as well.

1. Put the dataset and python script in the same folder.Ensure that the dataset file is

named as “dataset” as shown in figure 4.2.

Figure 4.2: Folder and File Naming

2. Type in *cd Desktop/foldername* as shown in figure 4.3. For this example the folder in which all the files are kept is named as project. So the below code needs to be typed and then press enter.Here “Desktop” is used as the folder containing the script and dataset is kept in the desktop.

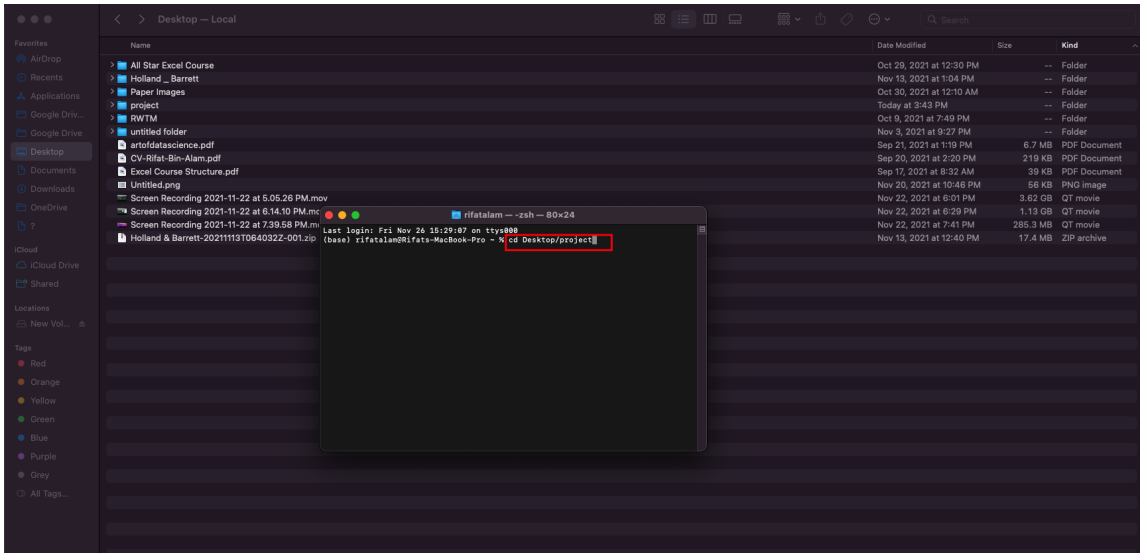

Figure 4.3: Opening folder in terminal

3. In the next line, type in the following code *python3 scr.py* and then press enter as shown in figure 4.4.

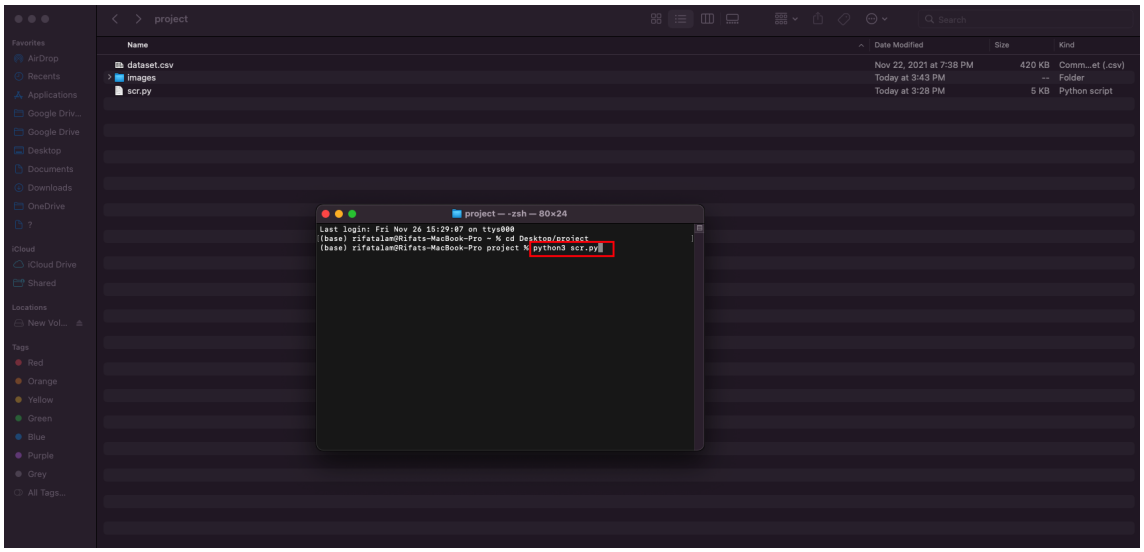

Figure 4.4: Activating The Script

4. After a few minutes all the files in the analysis will be generated in the folder as shown in figure 4.5.

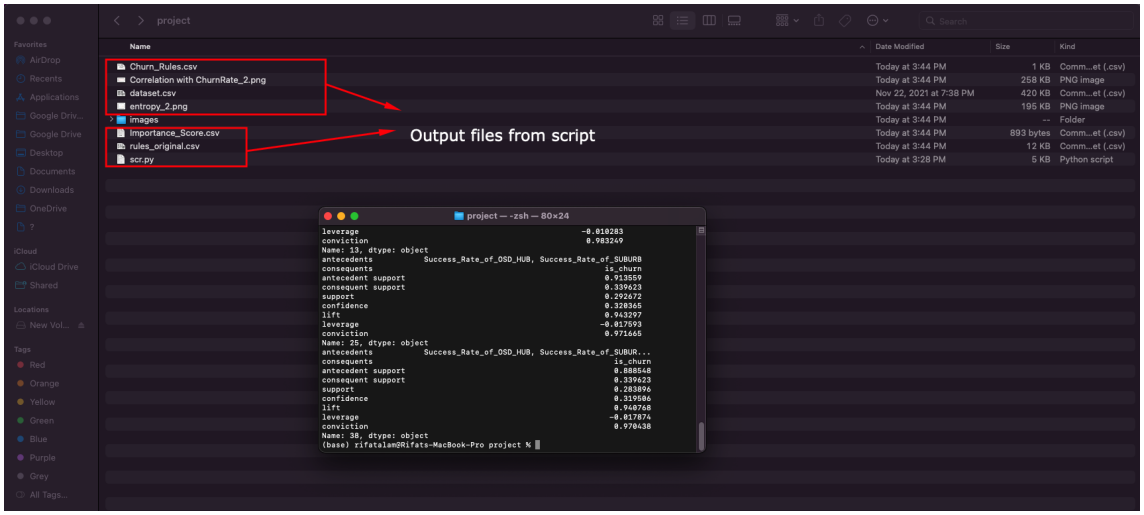

Figure 4.5: Script Output Files

Chapter 5

Conclusion

This chapter summarises the results from the previous chapter and provides the scope of the future works that can be done to improve the results of this project.

5.1 Summary

Comparison of the results from both Decision Tree and Apriori Alorithm shows different results. In the output from Decision Tree, we can see that the root was selected to be *parcelcount* however in none of the antecedents of *churn parcelcount* was not present. The same can be said for all of the nodes in the Decision Tree diagram.

This leads to the conclusion that in trying to decipher reasons of customer churn, these two algorithms independently will not give the business user enough insight to act upon. Since the features selected by both the decision tree and apriori algorithm contain features from the list of highly negatively correlated features from the correlation comparison diagram, we can consider both results to be indicative of which business functions to improve in order to reduce customer churn.

5.2 Limitation

Imbalanced Dataset

The dataset used for this paper was heavily imbalanced in favor of *churned* customers. This has a direct impact on the both the algorithms efficiency. For decision tree the presence of imbalanced data lead to most possible outcomes end in class 1 as there was not enough non *churned* user data present.

Discretizer Used

For Apriori Algorithm to process the data faster, $KBinsDiscretizer$ was used to discretize all data in the dataset with the $uniform$ parameter set. This generalised approach will have led to over fitted results.

5.3 Future Work

Imbalanced Dataset

One improvement scope here will be to use a balanced dataset, which has equal number of *churned* and $non-churned$ users. The tree diagram output from this dataset can be compared with the existing one to see if there is any fundamental change/improvement in the output.

Discretizer Used

The following are the different methods of discretization that can be explored in future renditions of the work :

- Equal Width - The simplest binning approach is to partition the range of the variable into k equal-width intervals. The interval width is simply the range [A, B] of the variable divided by k, where the value of k can be iterated.
- Equal Frequency - n equal-frequency binning we divide the range [A, B] of the variable into intervals that contain (approximately) equal number of points; equal frequency may not be possible due to repeated values.

References

[1] Adnan Amin, Sajid Anwar, Awais Adnan, Muhammad Nawaz, Khalid Alawfi, Amir Hussain, and Kaizhu Huang. Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, 237:242–254, 2017.

[2] Benlan He, Yong Shi, Qian Wan, and Xi Zhao. Prediction of customer attrition of commercial banks based on svm model. *Procedia Computer Science*, 31:423–430, 2014.

[3] Ssu-Han Chen. The gamma cusum chart method for online customer churn prediction. *Electronic Commerce Research and Applications*, 17:99–111, 2016.

[4] Niccolò Gordini and Valerio Veglio. Customers churn prediction and marketing retention strategies. an application of support vector machines based on the auc parameter-selection technique in b2b e-commerce industry. *Industrial Marketing Management*, 62:100–107, 2017.

[5] Eugen Stripling, Seppe vanden Broucke, Katrien Antonio, Bart Baesens, and Monique Snoeck. Profit maximizing logistic model for customer churn prediction using genetic algorithms. *Swarm and Evolutionary Computation*, 40:116–130, 2018.

[6] Qiu-Feng Wang, Mirror Xu, and Amir Hussain. Large-scale ensemble model for customer churn prediction in search ads. *Cognitive Computation*, 11(2):262–270, 2019.

[7] Adnan Amin, Feras Al-Obeidat, Babar Shah, Awais Adnan, Jonathan Loo, and Sajid Anwar. Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 94:290–301, 2019.

[8] Yaya Xie, Xiu Li, EWT Ngai, and Weiyun Ying. Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36(3):5445–5449, 2009.

[9] Koen W De Bock and Dirk Van den Poel. An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*, 38(10):12293–12301, 2011.

[10] Koen W De Bock and Dirk Van den Poel. Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. *Expert Systems with Applications*, 39(8):6816–6826, 2012.

[11] Wouter Verbeke, David Martens, and Bart Baesens. Social network analysis for customer churn prediction. *Applied Soft Computing*, 14:431–446, 2014.

[12] Sandra Mitrović, Bart Baesens, Wilfried Lemahieu, and Jochen De Weerdt. tcc2vec: Rfm-informed representation learning on call graphs for churn prediction. *Information Sciences*, 2019.

[13] Carl Yang, Xiaolin Shi, Luo Jie, and Jiawei Han. I know you'll be back: Interpretable new user clustering and churn prediction on a mobile social application. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 914–922, 2018.

[14] Arno De Caigny, Kristof Coussement, and Koen W De Bock. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2):760–772, 2018.

[15] B Arivazhagan and SDRS Sankara. Customer churn prediction model using regression with bayesian boosting technique in data mining. *Ijaema. Com*, 12(V):1096–1103, 2020.

[16] M Makhtar, S Nafis, MA Mohamed, MK Awang, MNA Rahman, and MM Deris. Churn classification model for local telecommunication company based on rough set theory. *Journal of Fundamental and Applied Sciences*, 9(6S):854–868, 2017.

[17] Jungyeol Hong, Reuben Tamakloe, and Dongjoo Park. Discovering insightful rules among truck crash characteristics using apriori algorithm. *Journal of advanced transportation*, 2020, 2020.