CUSTOMER CHURN ANALYSIS AND PREDICTION

A PROJECT REPORT

Submitted by

PAVANI RAVIRALA (21203863)

SHASANK PYATE (21200341)

Under the Supervision of

Dr. Sarp Akcay

School of Mathematics and Statistics

UNIVERSITY COLLEGE DUBLIN

Belfield, Dublin 4, Ireland – D04 V1W8

JULY 2022

Contents

List of Figures

List of Tables

1 Abstract

The tall increment in the number of companies competing in developing markets makes client maintenance an important factor for any company to outlive. It has become very difficult for them to retain existing customers. Since the cost of acquiring new customers is much higher than the cost of retaining the existing customers, industries need to take necessary steps to retain the customers to stabilize their market value. For this reason, large telecommunications corporations seek to develop models to predict which customers are more likely to change and act accordingly. Hence, numerous strategies have been proposed to dissect and study customer maintenance. The legitimacy of such strategies isn't however demonstrated though. This project tries to fill this crevice by observationally comparing two strategies: Customer churn using decision trees and logistic regression models.

2 Introduction

It is quite challenging for an organization to retain customers and also ensure growth by acquiring new customers. However, this challenge is mostly perceived by the telecommunication industry, where the customers are prone to frequent switching among the different providers in the telecom sector, owing to the better performance of one provider over another. This fuels the need to establish a definite mechanism or a model that can identify the limitations of the organization due to which the organization fails to retain its customers and improve the sales revenue and growth profits as presented (Jayaswal et al. 2016). In such a scenario Data Science and Analytics has become a vital and integral part of every business plan.

The churn prediction process is a key data analysis and classification process that enables any trade to analyze if the churn rate among the clients picked up over a positive period. With the "Organizations within the telecom division" looking forward to the maintenance of "faithful clients" for their trade development, the endeavours are moreover sharp on procuring "modern clients" at the same time. According to the progress of the conceivable outcomes of picking up more clients and holding them to earn a more prominent benefit, the organizations build up procedures to anticipate the churn rate and distinguish their drawbacks to make strides in the circumstance by utilizing diverse data analysis mechanisms and classifiers. This gives a more noteworthy opportunity for regulating "client relationship management."

Recently, Customer Relationship Management (CRM) has gotten a lot of attention from companies where customer retention is considered to be its main factor to be investigated as it focuses on developing and controlling loyal, profitable and lasting relations with customers. Developing successful retention techniques is important for businesses in general, and mobile operators. To do so, they need to have a clear picture of what is happening around the customers and why they are not able to retain a particular group of customers.

Examining customer churn has been a subject of numerous strategies and procedures counting factual methods and data mining. The predominance of utilizing data mining is well set up to examine customer churn compared with conventional advertising inquiries about overviews. Market Surveys based on running questionnaires or views of customers suffer from high cost and time taking. On contrary, data mining provides knowledge of the entire client population based on an investigation of their current and verifiable information. Increasingly, Data mining gets to be a fundamental strategy in client maintenance to anticipate future customers' demeanours and detect patterns inside verifiable information Liu and Fan (2014). Two of the more Data mining techniques are Regression Analysis and Decision Trees.

First, logistic regression predicts the occurrence probability of customer churn by formulating a set of equations, input field values, factors affecting customer churn and the output field (Ahn, Han, and Lee 2006; Burez and Van den Poel 2007). Second, decision trees, are used to solve classification problems where the instances are classified into one of two classes (i.e., positives and negatives).(Burez and Van den Poel 2007)

2.1 Aim, Objective and Limitation

Aim: Precisely forecast and analyze the future churn to help authoritative choices for the company's telecom customers. Also, develop a Churn prediction model which includes machine learning and statistical algorithm to predict consumer churn.

Objective: Comparing two Machine learning techniques to figure out which among both would be better to Predict Churn in the telecom domain.

Limitation: Telecom companies won't share their private information since it is secret, and data spill may be a genuine offence.

An exact and tremendous sum of information will be supportive in training the Machine Learning algorithm more absolutely.

3 Modelling techniques and Algorithms

3.1 Logistic Regression

Within the many machine learning models, Logistic regression is a supervised machine learning model. It is also a discriminative model because it attempts to differentiate between various classes or categories. Logistic Regression is one of the Machine Learning techniques which is used to predict the probability of the Churn of a customer Faris, Al-Shboul, and Ghatasheh (2014). It is based on a statistical approach which would analyze the effect of variables on the variable to be predicted. Prediction is made by forming a set of equations connecting input values. The odds - that is the probability of success divided by the failures are calculated first and a logit transformation is applied. This step is known as the log odds or natural logarithm of the odds. The logistic regression model is described with the help of the equations (1),(2) and (3)

$$P(y=1|x_1,...,x_n) = f(y) \tag{1}$$

$$f(y) \quad = \tfrac{1}{1+e^{-y}}(\text{logit function}) \tag{2}$$

$$y \quad = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n \tag{3}$$

Where:

- y is the target variable for each j (customer in churn modelling), y is a binary class label (0 or 1);
- β_0 is a constant;
- β_i is the weight given to the specific variable x_i associated with each customer j (j=1,. . . ,m);
- $x_1, ..., x_n$ are the predictor variables for each customer j, from which y is to be predicted.

The beta parameters in the logistic regression formula are commonly estimated via the method of maximum likelihood estimation (MLE). Different values of beta are plugged in through many iterations such that the best fit of log odds can be optimized. With these optimized parameters at hand, we can calculate the conditional probability of each observation, find the log and sum them together to give us a predicted probability. In general, for binary classification problems, a predicted probability of less than 0.5 will be labelled as 0 while a probability greater than 0.5 will be labelled as 1.

3.2 Decision Trees

Decision trees, the foremost popular predictive models, maybe a tree chart displaying the variables' connections. Utilized to illuminate classification and prediction issues, decision tree models are represented and assessed in a top-down way.

The two fundamental steps in building decision trees are tree building and tree pruning. A decision tree is built starting from the root node that represents a feature to e classified. Selecting a feature is done by evaluating entropy. The root nodes have the lowest entropy while the lower level-nodes have higher entropies. The lower-level nodes are then constructed in a similar way to the divide and conquer strategy. Improving predictive accuracy and reducing complexity, the pruning process is applied to decision trees to produce a smaller tree and guarantee a better generalization by removing branches containing the largest estimated error rate.

The decision about a given case regarding which of the two classes it belongs to is thus made by moving from the root node to all leaves. On one hand decision trees have several advantages. First, they are easy to visualize and understand. Second, since it is a nonparametric approach, no prior assumptions about the data are needed. Third, decision trees can process any kind of data regardless of whether it is numerical and categorical. On the other hand, decision trees suffer from few disadvantages. First, the performance of a decision tree can suffer due to complex interactions between variables. Second, if a complex decision tree is built, then it is hard to visualize and interpret the tree.

Algorithm of a Decision Tree:

- It begins with the original set S as the root node.
- On each iteration of the algorithm, it iterates through the very unused attribute of the set S and calculates the Entropy(H) of this attribute. The Entropy is calculated using the formulas below:

$$Entropy(H) = -\sum_{i=1}^{n} p(c_i) log(p(c_i))$$

  Where c_i is the i th category/class and $p(c_i)$is the probability of i th class.
- It then selects the attribute which has the smallest Entropy.
- The set S is then split by the selected attribute to produce a subset of the data.
- The algorithm continues to recur on each subset, considering only attributes never selected before.

		Actual Values	
		Positive (1)	Negative (0)
Predicted	Positive (1)	TP	FP
Values	Negative (0)	FN	TN

Table 1: Confusion Matrix

4 Performance Evaluation

4.1 Classification Accuracy

Classification accuracy is commonly used as a metric to measure the performance of a classifier or a classification model. It is calculated as the number of correct predictions divided by the total number of predictions in the data. It is a widely used metric to measure the performance of a classifier because it is intuitive and very easy to calculate.

"Accuracy and its complement error rate are the most frequently used metrics for estimating the performance of learning systems in classification problems".(Branco, Torgo, and Ribeiro 2016)

As a first step in calculating the classification accuracy of a model, a classification model is built to make predictions for each observation in the unseen data (test data). The predictions are then compared to the actual labels of the test data. Finally, accuracy is calculated as the number of observations correctly predicted in the test data divided by the overall number of observations in the test data.

A confusion matrix is a visual representation of information about actual and predicted classifications produced by a classification model (Nisbet, Elder, and Miner 2009). Table 1 depicts a confusion matrix for a binary classifier.

$$Accuracy = \frac{CorrectPredictions}{TotalPredictions} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$\text{Error rate} = 1 - Accuracy \tag{5}$$

Word of caution: Classification Accuracy is an ambiguous marker, especially in the case of imbalanced data. To clarify, a base contains 9,990 churn clients and 10 non-churn ones. If a demonstration succeeds to anticipate that all 10,000 customers are at chance of churn, the precision of classification will be 99.9%. The high accuracy rate erroneously demonstrates that the model is exceptionally precise in foreseeing client churn since the model does not identify any non-churn customers. Correctly anticipating churn cases is continuously more vital than anticipating non-churn cases as the probability of mis-predicting churn is higher than that of mis-predicting non-churn.

4.2 Sensitivity and Specificity

Sensitivity and specificity scientifically portray the exactness of a test which reports the presence or absence of a condition. People for which the condition is fulfilled are considered "positive" and those for which it isn't are considered "negative".

Sensitivity (true positive rate) alludes to the probability of a positive test, conditioned on truly being positive.

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

Specificity (true negative rate) alludes to the probability of a negative test, conditioned on truly being negative.

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

Mobile administrators favour models with high sensitivity instead of models with high specificity since the cost associated with the erroneous classification of churners are higher than the cost related to the incorrect classification of a non-churner. A compromise between high sensitivity combined with sensible specificity should be continuously made so mobile administrators can successfully manage their promoting budget to attain high customer maintenance.

4.3 ROC Curve

The Receiver Operating Characteristic (ROC) curve is a depiction of the relations between the true positive rate (i.e., benefits) and false positive rate (i.e., costs), drawn on the y and x-axis respectively (Karahoca, Karahoca, and Aydin 2007). In our case, true positive rate (TPR) is the churners ratio correctly predicted as churners and false positive rate (FPR) is the non-churners ratio wrongly predicted as churners. The ROC curve also tells us if there are any compromises between benefits to costs. The ROC curve is plotted with the help of points corresponding to the predicted observations.

The best performing model is when the ROC curve passes through or close to (0, 1) as shown in Figure 1. The model sensitivity and specificity will then be 100% (i.e., no false negatives and no false positives respectively). In the case of logistic regression, probabilities are produced instead of a binary class decision (i.e. churn or not). To build a binary classifier we use threshold values in the range of (0,1). If the classifier probability is greater than the said threshold then it is assigned to class churn. Otherwise, it is non-churn.

5 Experiments and Results

5.1 Data set Description

This section highlights the processes involved in the data understanding phase. A single data set has been used to explore and analyze the churn rate based on different variables. The data set was obtained from the IBM Developer Platform and can be downloaded from here. The sample dataset tracks the customers of a fictional telecommunication company called Telco. The analysis is based on 7043 Customers respectively containing 21 variables each. The variables in the dataset can be broken down into the following sections:

- Customers who left within the last month – the column is called Churn.
- Services Information - Services that each customer has signed up for (phone, multiple lines,

Figure 1: ROC Curve
source by:Wikipedia

internet, online security, online backup, device protection, tech support, and streaming TV and movies.)
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges.
- Demographics about customers – gender, age range, and if they have partners and dependents.

5.2 Exploratory Data Analysis and Data Cleaning

In this step, we mainly analyze the important characteristics of the data with the help of visualization techniques or summary statistics. We try to understand the data, discover anomalies or patterns of interest or check if there are any other conditions to be cautious of.

As the first step of EDA, we would want to know a concise summary of the dataset like the column names and data type of each column. This is shown below.

```
## 'data.frame':    7043 obs. of  21 variables:
##  $ customerID      : chr  "7590-VHVEG" "5575-GNVDE" "3668-QPYBK" "7795-CFOCW" ...
##  $ gender          : chr  "Female" "Male" "Male" "Male" ...
##  $ SeniorCitizen   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Partner         : chr  "Yes" "No" "No" "No" ...
##  $ Dependents      : chr  "No" "No" "No" "No" ...
##  $ tenure          : int  1 34 2 45 2 8 22 10 28 62 ...
##  $ PhoneService    : chr  "No" "Yes" "Yes" "No" ...
##  $ MultipleLines   : chr  "No phone service" "No" "No" "No phone service" ...
##  $ InternetService : chr  "DSL" "DSL" "DSL" "DSL" ...
##  $ OnlineSecurity  : chr  "No" "Yes" "Yes" "Yes" ...
##  $ OnlineBackup    : chr  "Yes" "No" "Yes" "No" ...
##  $ DeviceProtection: chr  "No" "Yes" "No" "Yes" ...
##  $ TechSupport     : chr  "No" "No" "No" "Yes" ...
```

```
##  $ StreamingTV     : chr  "No" "No" "No" "No" ...
##  $ StreamingMovies : chr  "No" "No" "No" "No" ...
##  $ Contract        : chr  "Month-to-month" "One year" "Month-to-month" "One year" ...
##  $ PaperlessBilling: chr  "Yes" "No" "Yes" "No" ...
##  $ PaymentMethod   : chr  "Electronic check" "Mailed check" "Mailed check" "Bank transfer (a
##  $ MonthlyCharges  : num  29.9 57 53.9 42.3 70.7 ...
##  $ TotalCharges    : num  29.9 1889.5 108.2 1840.8 151.7 ...
##  $ Churn           : chr  "No" "No" "Yes" "No" ...
```

5.2.1 Missing values

We would want to know the number of null values in each column and remove or replace them so that machine learning algorithms can perform better.

```
##       customerID           gender    SeniorCitizen          Partner
##                0                0                0                0
##       Dependents           tenure     PhoneService    MultipleLines
##                0                0                0                0
##  InternetService   OnlineSecurity     OnlineBackup DeviceProtection
##                0                0                0                0
##      TechSupport      StreamingTV  StreamingMovies         Contract
##                0                0                0                0
## PaperlessBilling    PaymentMethod   MonthlyCharges     TotalCharges
##                0                0                0               11
##            Churn
##                0
```

The Total charges column has 11 null values even though Monthly Charges are not null for these entries. So, the rows with which these NA values are associated were deleted from the dataset. Initially, we had 7043 observations and 21 variables. After deletion of those 11 null values, we now have 7032 observations and 21 variables. The dimensions of the dataset after deletion is printed below.

```
## [1] 7032   21
```

5.3 Some insights into data

In this section, we try to gain some insights into data that a company would want to know in the form of visualizations.

5.3.1 Response variable - Churn

Figure 2 presents the churn percentage of the company. The customer churn percent is around 28% and retention is around 75%. We clearly have an imbalanced dataset, where customer retention cases are more than customer churn.

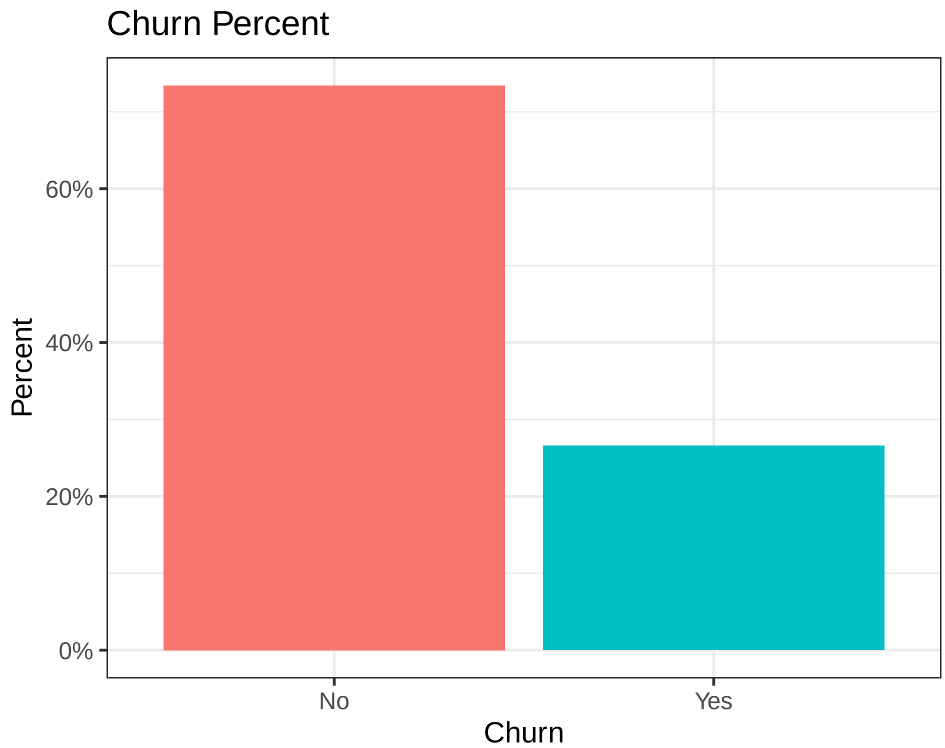

Figure 2: Percentage of Churn

5.3.2 Demographics of Customer

Figure 3 shows the percentage of Churn with respect to each attribute in the Demographics category.

The percentage of churn is the same regardless of whether the gender is male or female. The churn rate of senior citizens is almost twice higher than that of other age groups. Customers with Partners or Dependents tend to have lower churn rates than those who do not have a Partner or a dependent.

5.3.3 Services Information

Similar to the plots in the demographic section, percentage of churn is plotted against each and every category in Services Information which is shown in Figure 4

PhoneService and MultipleLines columns do not contribute to predicting Churn rate as the percentages of Churn are the same. Churn rates are higher in the case of Fiber Optic Internet Service. Clients with no online security, no internet service, no device protection and no tech support have left this telco company last month. The churn rate is the same for clients with or without streaming TV. In all the cases, the churn rate seems to be the same when any client has no internet service.

5.3.4 Customer Information

The columns involved here are: StreamingTV, StreamingMovies, Contract, PaperlessBilling, PaymentMethod, MonthlyCharges, TotalCharges

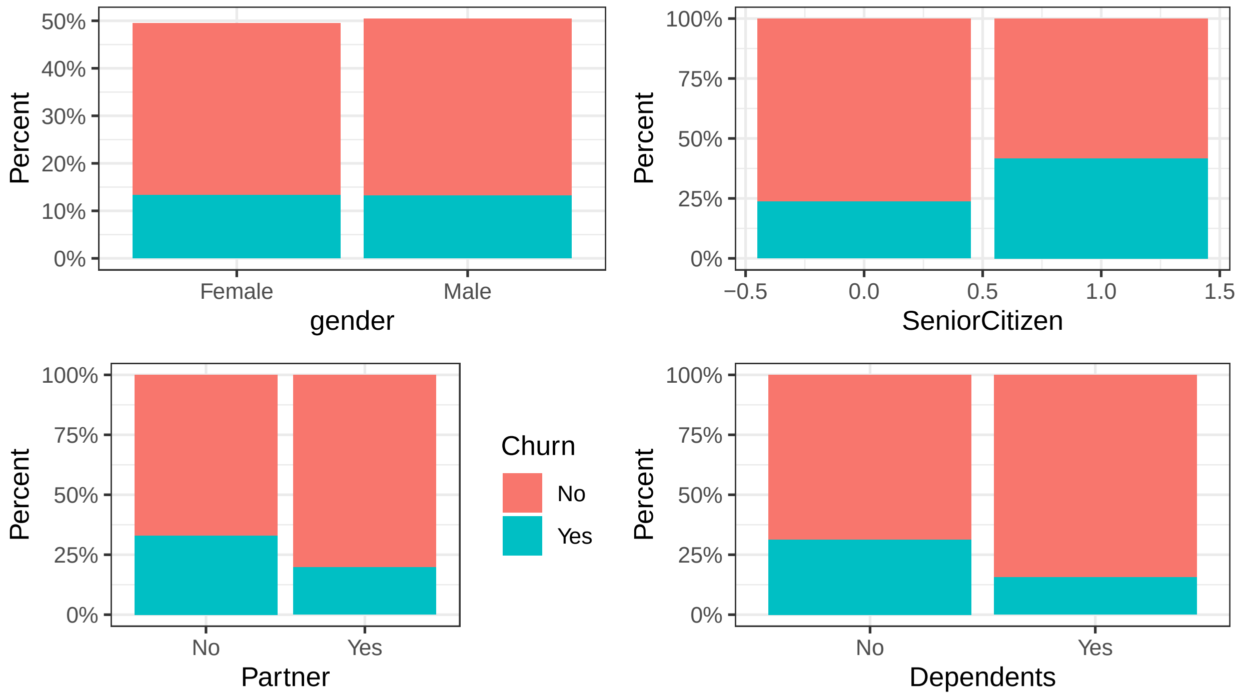

Figure 3: Churn percent with respect to demographics of a customer

It can be seen from Figure 5 that clients with monthly contract tend to leave the company more often than those with yearly contract. Churn rate is higher for clients who subscribed to Paperless Billing. The churn rate is lower when clients have opted for electronic check as their payment method than any other form of payment.

It can be inferred from Figure 6 that the median time period for customer to leave is around 10 Months. Clients who pay higher monthly charges and low total charges are likely to leave the company.

5.4 Modelling

5.4.1 Logistic Regression

A model can be marked 'good' when having a 25% higher classification accuracy rate than the proportional by chance accuracy rate (Costea and Eklund 2003). The analysis was started by computing the proportional by chance accuracy rate based on calculating the proportion of cases for each group (churn and non-churn). The proportions of the non-churn and the churn groups are presented in Table 2. The proportional by chance accuracy rate was then computed by squaring and summarizing the proportion of cases in each group. The chosen model's accuracy rates are 25% higher than the proportional by chance accuracy rate. As a result, the classification accuracy criteria are satisfied, and the models perform better than a random guess. Additionally, the confusion matrix of the optimal logistic regression model is presented in Table 3.

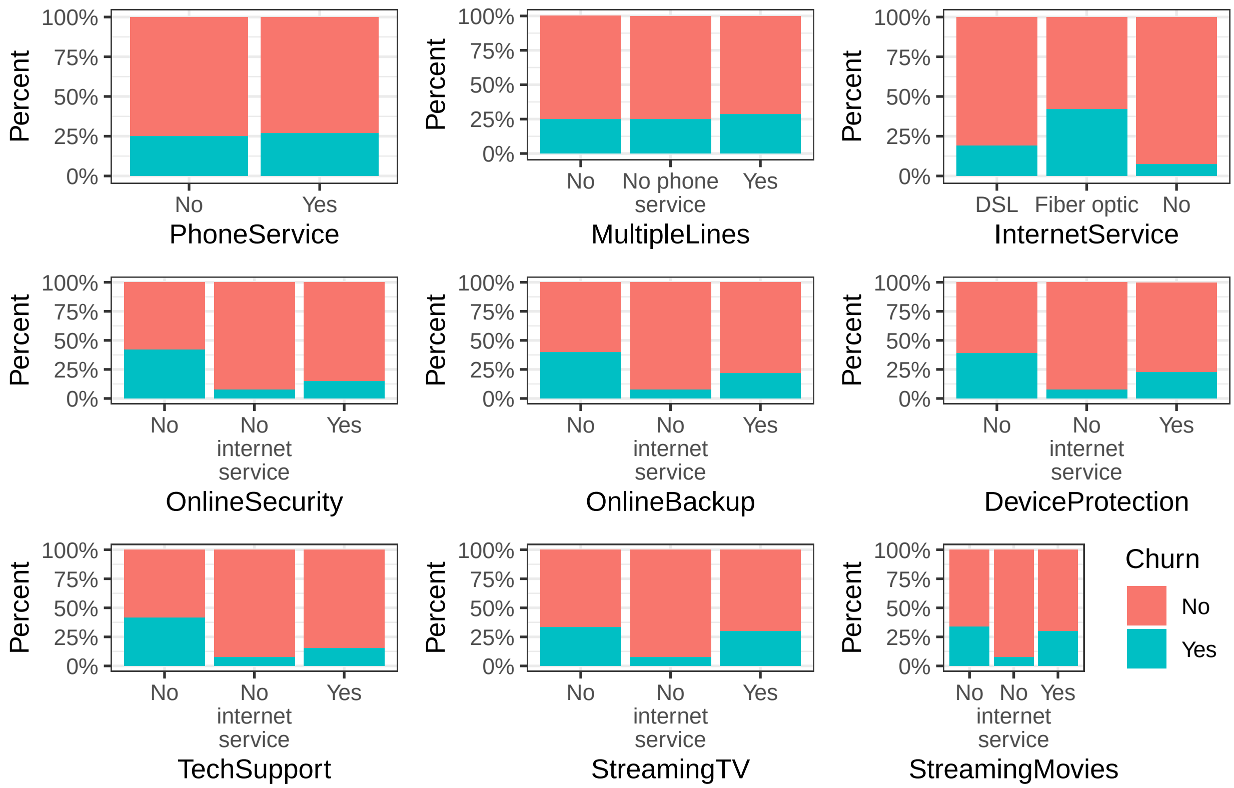

Figure 4: Percentage of churn for the services customers subscribed to

Proportion of churn	0.266
Proportion of non-churn	0.734
Proportion by chance	0.61
Accuracy of logistic regression	0.814

Table 2: Proportional by chance accuracy rate

5.4.2 Decision Tree

Customer churn behaviour was investigated by creating a decision tree model. The classification table for the best decision tree with fine tuned parameters is presented in Table 4.

The Decision Tree built is shown in Figure 7. In this tree, tenure is the root node. If an observation has $tenure \geq -0.646$ then the observation is classified as not-churn with a probability of around 0.8. If an observation has a $tenure < -0.646$, it checks if the `InternetService=Fiber Optic` < 0.5. If the answer is yes, then the observation is classified as not-churn with a probability of around 0.7. If the answer is no then, the observation is classified as churn with a probability of 0.7 approximately.

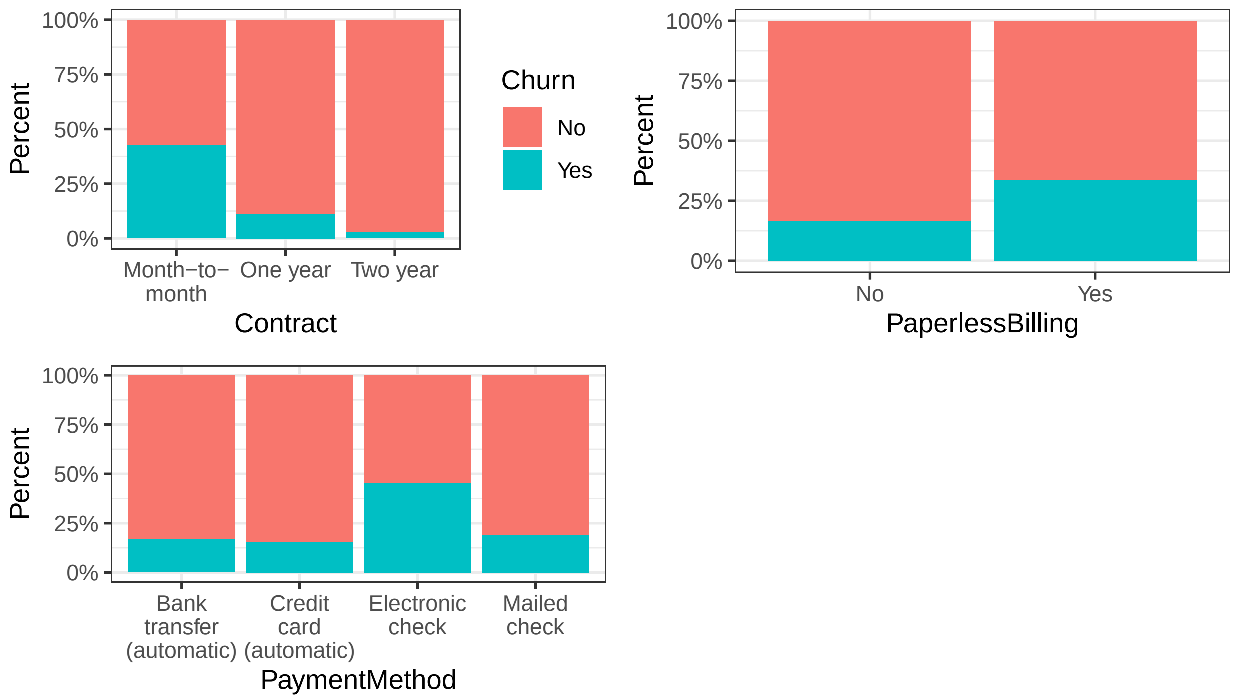

Figure 5: Percentage of churn with respect to Customer Churn Information (categorical data)

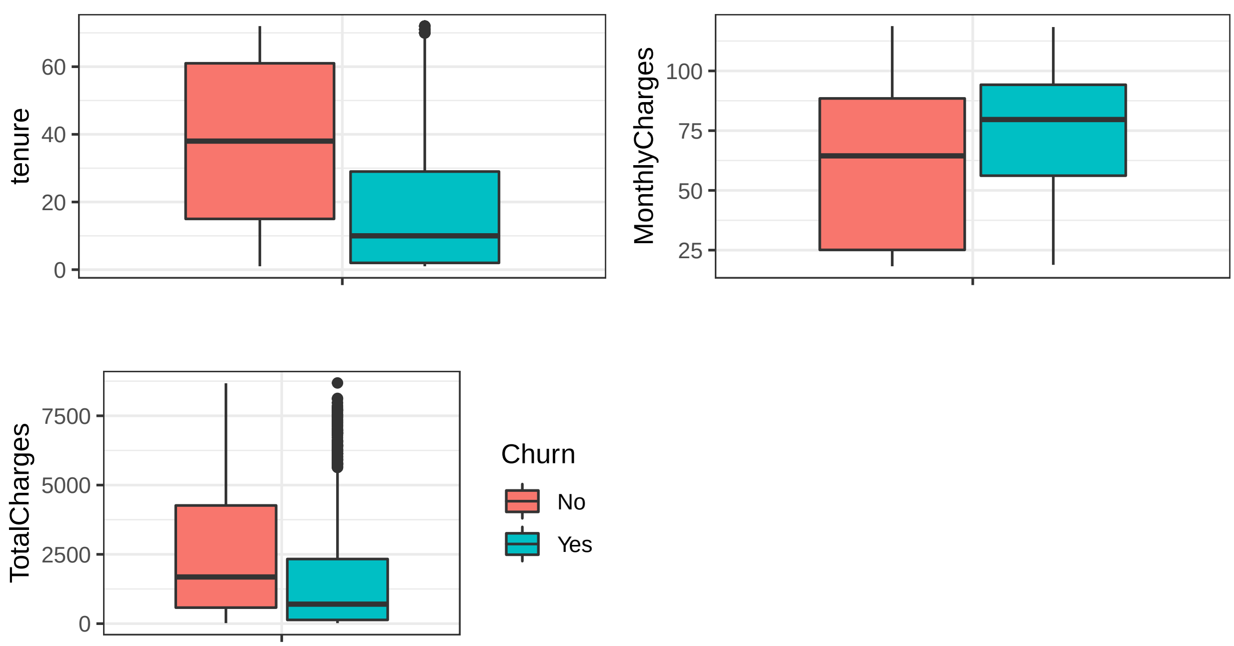

Figure 6: Percentage of churn with respect to Customer Churn Information (numerical data)

		Actual Values No	Actual Values Yes
Predicted Values	No	1385	229
	Yes	163	331

Table 3: Confusion Matrix of logistic regression

		Actual Values No	Actual Values Yes
Predicted Values	No	1440	108
	Yes	328	232

Table 4: Confusion Matrix of decision trees

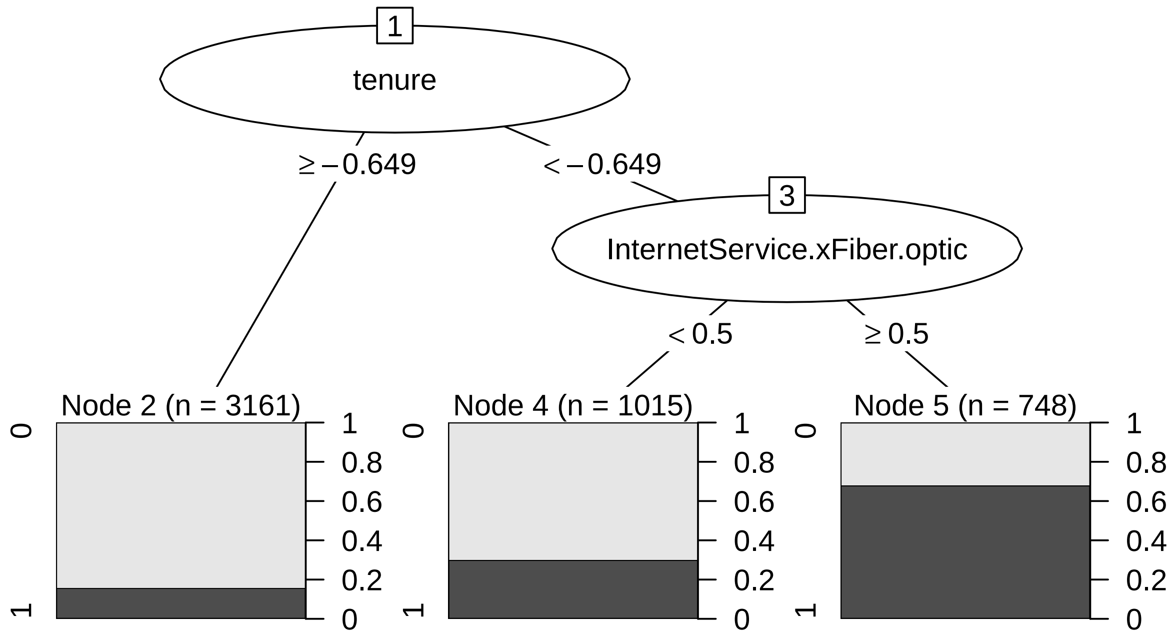

Figure 7: Decision Tree model

5.5 Model Evaluation

From a data-analysis point of view, the models displayed within the segment were chosen based on the quality of their yields. A choice on which demonstrating technique to be embraced for conveying the ultimate customer churn model should be taken as the conclusion at this stage. ROC curve, Sensitivity,Specificity and Overall Accuracy are used here to compare the performance of the modelling techniques and finally the model selection for Churn Prediction is done.

Figure 8 shows the ROC curve for both the decision tree and Logistic Regression. Table 5 summarizes the metrics used to compare performance of the models. It can be seen from the results that Logistic Regression is quite ahead of the Decision tree in terms of accuracy,sensitivity and area under the curve (AUC) of ROC . We need models with high sensitivity than high specificity as discussed in Section 4.2. Hence, Logistic Regression was used to predict customer churn.

5.6 Prediction using R Shiny App

The trained logistic regression model was saved to build an interactive web application on Shiny. Shiny is an open-source package in R that is used to build interactive web apps and dashboards in

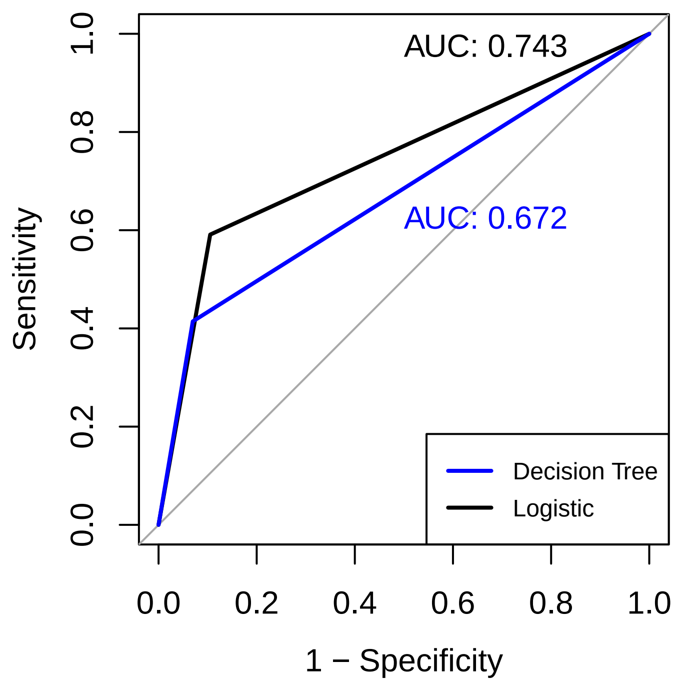

Figure 8: ROC-AUC Curve of logistic and decision tree models

Evaluation metrics	Logistic Regression	Decision Tree
Accuracy	0.814	0.793
Sensitivity	0.895	0.814
Specificity	0.591	0.682

Table 5: Comparing evaluation metrics

R that runs on the backend.

The first tab in the Shiny App predicts if an individual customer would want to leave the company or not. Initially, the company has to answer a set of questionnaires about the customer. Some examples of the questionnaires are: gender, if they are senior citizens, if they have dependents and partners, do they have a Phone Service, have they subscribed to services like Streaming TV, Streaming Movies, Online Backup, Tech Support, etc., how long have they been a customer in the company and so on. With these inputs and the loaded trained model for prediction, it becomes easier for the company to know if the customer would leave or not. The design of this tab is shown on the left side of Figure 9.

The second tab in the Shiny App predicts the percentage of Churn when the company has a bulk amount of data instead of a single customer. The input is a .csv file that consists of rows similar to the inputs of the questionnaires in the first tab. Each row speaks about an individual customer. The input file with the help of the loaded trained model can be used to give the percentage of customers who leave or who wish to remain. The design of this tab is shown on the right side of Figure 9.

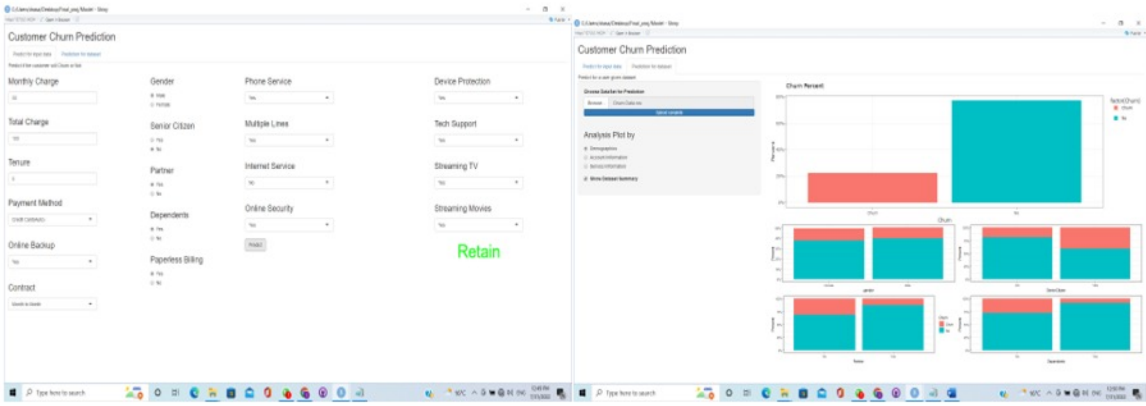

Figure 9: Prediction and Analysis in R Shiny App

6 Future Scope and Limitations

The applied methods are not the only ones which can predict the Churn but they are only some of the methods that can be used. There is no straightforward explanation why only particular methods are to be used.

- Not all the complex interactions have been included in the data set and there might be even more interactions that can be included in the future as part of the analysis.
- The data which is used cannot really stay constant forever and there is always a new feature being added and hence there is a scope for modifications over time.
- Especially customer interaction has not been included in the data set which makes the data set impure and organization oriented but not customer-centric.
- The current Data science models have a relatively short life span, and the mobile market faces new technologies on a daily basis. As a result, historical data become less valuable for predictions.

7 Conclusion

From the above analysis and modelling, one can assess the per cent of customers' churn and the losses due to Churn. At the same time, they can come up with a better structure to reduce the churn rate of the organisation. The above methodologies are not the only ones which can help in predicting the churn rate rather there are multiple and even better models available.

Having said the above, there are still limitations in predicting the churn rate exactly, as there may be a sudden change in the market due to which the mentioned internal and external factors may vary with which the data which is used for the model might be not much useful. So, it would be always better to have real-time updated data for performing analysis for better and more accurate outcomes.

Also, the telecom domain is not the only domain where churn analysis is helpful rather it can be applied in multiple domains like Education, Financial, Transport Sector etc. to understand a few factors like why students are leaving schools, why the customers are shifting among the banks, why the public is avoiding using public transport and many others.

8 References

Ahn, Jae-Hyeon, Sang-Pil Han, and Yung-Seop Lee. 2006. "Customer Churn Analysis: Churn Determinants and Mediation Effects of Partial Defection in the Korean Mobile Telecommunications Service Industry." *Telecommunications Policy* 30 (10-11): 552–68.

Branco, Paula, Luis Torgo, and Rita P Ribeiro. 2016. "A Survey of Predictive Modeling on Imbalanced Domains." *ACM Computing Surveys (CSUR)* 49 (2): 1–50.

Burez, Jonathan, and Dirk Van den Poel. 2007. "CRM at a Pay-TV Company: Using Analytical Models to Reduce Customer Attrition by Targeted Marketing for Subscription Services." *Expert Systems with Applications* 32 (2): 277–88.

Costea, Adrian, and Tomas Eklund. 2003. "A Two-Level Approach to Making Class Predictions." In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*, 9–pp. IEEE.

Faris, Hossam, Bashar Al-Shboul, and Nazeeh Ghatasheh. 2014. "A Genetic Programming Based Framework for Churn Prediction in Telecommunication Industry." In *International Conference on Computational Collective Intelligence*, 353–62. Springer.

Farquad, Mohammed Abdul Haque, Vadlamani Ravi, and S Bapi Raju. 2014. "Churn Prediction Using Comprehensible Support Vector Machine: An Analytical CRM Application." *Applied Soft Computing* 19: 31–40.

Han, Shui Hua, Shui Xiu Lu, and Stephen CH Leung. 2012. "Segmentation of Telecom Customers Based on Customer Value by Decision Tree Model." *Expert Systems with Applications* 39 (4): 3964–73.

Jayaswal, Pretam, Bakshi Rohit Prasad, Divya Tomar, and Sonali Agarwal. 2016. "An Ensemble Approach for Efficient Churn Prediction in Telecom Industry." *International Journal of Database Theory and Application* 9 (8): 211–32.

Karahoca, Adem, Dilek Karahoca, and Nizamettin Aydin. 2007. "GSM Churn Management Using an Adaptive Neuro-Fuzzy Inference System." In *The 2007 International Conference on Intelligent Pervasive Computing (IPC 2007)*, 323–26. IEEE.

Liu, Dong-sheng, and Shu-jiang Fan. 2014. "A Modified Decision Tree Algorithm Based on Genetic Algorithm for Mobile User Classification Problem." *The Scientific World Journal* 2014.

Nisbet, Robert, John Elder, and Gary D Miner. 2009. *Handbook of Statistical Analysis and Data Mining Applications*. Academic press.