DREXEL UNIVERSITY
LeBow
College of Business

Customer Churn Analysis

Final Report

STAT-642-674

Data Mining for Business Analytics

Group 2

Aida Karimu, Jingxin Yao, Sunita Barik, and Zihan Huang

DREXEL UNIVERSITY
LeBow
College of Business

TABLE OF CONTENTS

DREXEL UNIVERSITY
LeBow
College of Business

EXECUTIVE SUMMARY

Customer Churn Analysis

Overview

Through the comprehensive analysis, we developed an anti-churn model to predict and prevent customer churn, better dealing with the adverse impact of customer churn on the company and use the results of customer churn analysis to help the company improve the service and optimize the revenue performance.

Problem Statement

- 27% Churn Rate: One third of customers are leaving the company.
- 88.6% Month-to-Month Contract: Almost all churners are the month-to-month plan.
- 49.4% Tenure: Half of the customers leave the company before the 9th month.

Proposed Solution

- Monitor Customers' Churn Behavior with the Anti-Churn Model
- Upgrade service & products
- Make Products Diversity: 3-, 6-, 9-month plan
- Acquire yearly plan customers using referral
- Discounted monthly price and unique services
- Build Customer Loyalty

Report Summary

The telecommunications industry is a service-oriented industry, and the quality of customer relationship management directly affects the economic benefits of enterprises. A research states that keeping one existing customer costs less than one-fifth of attracting a new customer.[1] And thus, the key to success for a company to keep an excellent operational performance for a long term is to pay attention to risks underneath the customer churn.

INTRODUCTION

Customer churn is one of the greatest fears of any industry, as it directly affects the revenues and growth of the companies. Although there are many reasons for customer churn, some of the major reasons are service dissatisfaction, costly services, and better alternatives. According to an article in Harvard Business Review (Gallo, 2014), it was studied that the cost of acquiring a new customer is anywhere from five to 25 times more expensive than retaining an existing one. [2] Hence, companies are increasingly focusing resources in retaining existing customers and accomplish effective churn management to predict the people with highest churn probabilities.

Our motivation to conduct this customer churn analysis is to get an experience of working in real world scenario and implement the acquired knowledge to find a practical solution for the real-world problems. Also, we believe that this would give us an opportunity to enhance our practical knowledge about data analysis and prediction models.

Our goal here is to find a solution to the customer churn problem by identifying the key factors leading to customer churn through data analysis and build an anti-churn model which can predict the customer susceptible to churn. Also, recommend some actionable retention strategies to prevent their customers from leaving or cancelling their services. The analysis methods will include exploratory data analysis, unsupervised (k-Means Clustering) and supervised (Decision Tree and Support Vector Machines) machine learning methods.

DREXEL UNIVERSITY LeBow College of Business

DATA

The data used for analysis is a dataset taken from a telecommunication company which contains the data of 7043 customers. These include customer demographic information, contract, and service types they signed in for and their account information.

Dataset Description

- It has 7043 rows which indicate customers and 21 columns which are the attributes of the customer.
- There are 18 categorical, 3 numerical variables in the dataset.
- The data contains 0.15% of missing data (values from 11 rows are missing).
  Missing value rows: "489", "754", "937", "1083", "1341", "3332", "3827", "4381", "5219", "6671", and "6755".

Predictor Variables: Gender, SeniorCitizen, MonthlyCharges, TotalCharges, PaymentMethod, tenure, InternetService, MultipleLines, Partner, StreamingMovies, Dependents, Phone Service, Contract, StreamingTV, OnlineSecurity, OnlineBackup.

Targeted Variable: Churn

Table 1: Attributes and Data types

#	Attribute Name	Data Type	Missing Data	Description
1	customerID	Nominal	No	unique customer identifier (Does not provide any insight)
2	gender	Nominal	No	gender of customer,
3	SeniorCitizen	Nominal	No	indicates if a customer is a senior citizen (1) or not (0)
4	Partner	Nominal	No	Indicates if the customer has a partner (Yes) or not (No)
5	Dependents	Nominal	No	Indicates if the customer has dependents (Yes) or not (No)
6	tenure	Numerical	No	the length of time that the customer has been a customer
7	PhoneService	Nominal	No	Indicates if the customer has phone service with the company (Yes) or not (No)
8	MultipleLines	Nominal	No	Whether the customer has multiple lines or not (Yes, No, No phone service)
9	InternetService	Nominal	No	Indicates if the customer has fiber optic, DSL or no internet service with the company
10	OnlineSecurity	Nominal	No	Whether the customer has online security or not (Yes, No, No internet service)
11	OnlineBackup	Nominal	No	Whether the customer has online backup or not (Yes, No, No internet service)
12	DeviceProtection	Nominal	No	Whether the customer has device protection or not (Yes, No, No internet service)
13	TechSupport	Nominal	No	Whether the customer has tech support or not (Yes, No, No internet service)
14	StreamingTV	Nominal	No	Whether the customer has streaming TV or not (Yes, No, No internet service)
15	StreamingMovies	Nominal	No	Whether the customer has streaming movies or not (Yes, No, No internet service)
16	Contract	Ordinal	No	The type of contract that the customer has with the company (Month-to-month, One year,Two year)
17	PaperlessBilling	Nominal	No	If the customer is enrolled in paperless billing (Yes) or not (No)
18	PaymentMethod	Nominal	No	The most recent payment method used by the customer to pay the company (Electronic check, Mailed check, Bank transfer (automatic), or Credit card (automatic)
19	MonthlyCharges	Numerical	No	The most recent amount that the customer is charged per month
20	TotalCharges	Numerical	Yes(11)	The total amount that the customer has been charged
21	Churn	Nominal	No	Whether the customer has left the company (Yes) or not (No) (Target Variable)

Descriptive Statistics

Descriptive analysis summarizes the data points in the dataset. The summary of the customer churn data is captured in the form tables.

Table 2: Descriptive Statistics for Numerical Variables

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Tenure	0	9	29	32.37	55	72
MonthlyCharges	18.25	35.5	70.35	64.76	89.85	118.75
TotalCharges	18.8	401.4	1397.5	2283.3	3794.7	8684.8

Table 3: Mode for Categorical Variables

Gender	**SeniorCitizen**	**Partner**	**Dependents**
Male	No	No	Yes
PhoneService	**MultipleLines**	**InternetService**	**OnlineSecurity**
Yes	Yes	Fiber optic	No
OnlineBackup	**DeviceProtection**	**TechSupport**	**StreamingTV**
No	No	No	No
StreamingMovies	**Contract**	**PaperlessBilling**	**PaymentMethod**
No	Month-to-month	Yes	Electronic check

Exploratory Data Analysis and Visualizations

Data analysis using three variables; Contract Types, Tenure in Months, and Internet Service.

Analysis 1

The plot (**Figure 1**) below shows that maximum churn can be seen among the group of people who have month-to-month contract type.

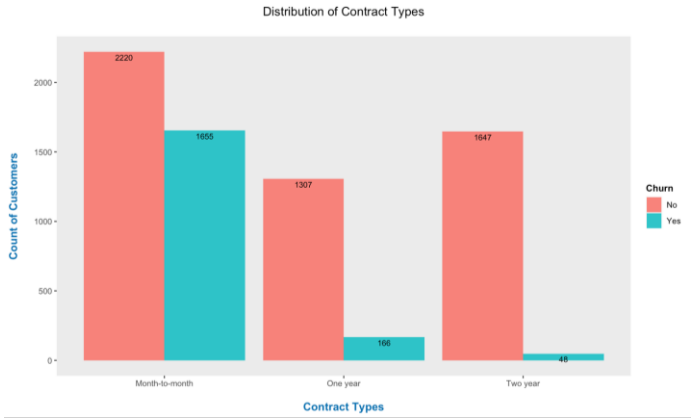

Figure 1: Distribution of Contract Types

Analysis 2

The plot (**Figure 2**) below indicates that majority of the customer churn occurred in less than an average period of 9 months. With increasing customer tenure, the churn rate decreases.

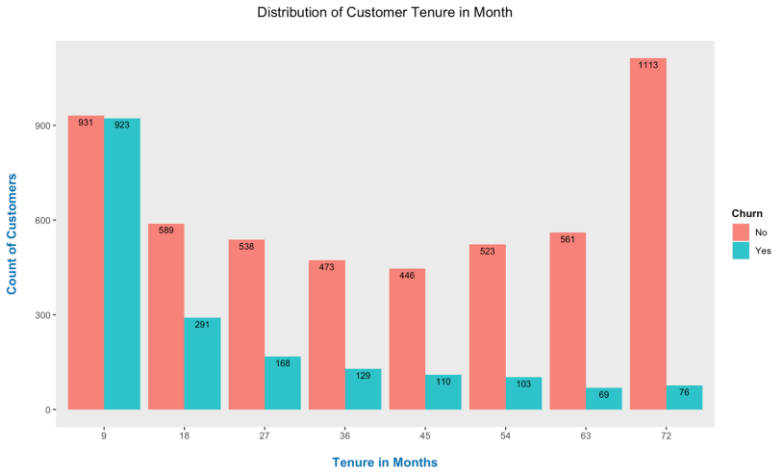

Figure 2: Distribution of Customer Tenure in Months

Analysis 3

The plot (**Figure 3**) below Churn is more significant for Fiber optic users than which is considered as the best internet service than the DSL internet services.

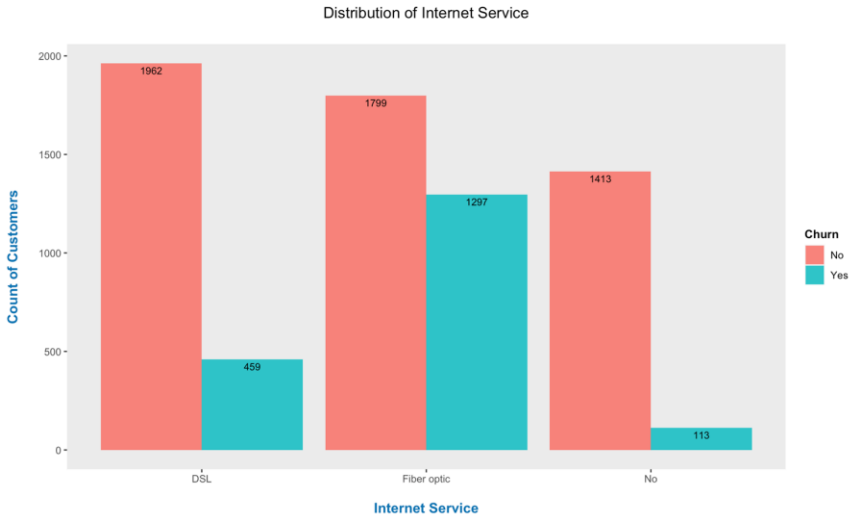

Figure 3: Distribution of Internet Service

Data Preparation

Data Quality Overview

In business point of view, data quality refers to data that are "fit for use" in their intended operational, decision-making, and other roles"[1] or that exhibits "conformance to standards" that have been set, so that fitness for use is achieved. [4] Common dimensions for data quality includes Relevance, Accuracy, Timeliness, Comparability, Completeness etc.

- In the dataset, we have 3 numerical variables, 1 ordinal variable and 17 nominal variables.
- Among the 17 nominal variables, CustomerID has high cardinality, and it cannot give us any useful insight. Therefore, we exclude CustomerID from our analysis.
- Among the 21 categorical variables, 7 nominal variables (include Churn) have 2 levels, 8 have 3 levels and 1 has 4 levels and all the numerical variables have positive values.
- There is 0.015% missing data which is negligible.
- Detected 0 outliers.

Data Cleansing

Data cleansing is the process in which corrupt or inaccurate records are detected and corrected or removed from a record set, table, or the database. It also identifies incomplete, incorrect, inaccurate, or irrelevant parts of the data. Then it replaces, modifies, or deletes the dirty or coarse data[5] as incorrect and inconsistent data can lead to wrong analysis conclusion. Moreover, certain modeling is vulnerable to missing data or noise, therefore data cleansing is necessary before analysis. In our case, we performed the following data cleansing methods:

Missing Data Imputation

There are 11 missing values in TotalCharges, which is 0.015% of the data. As we are going to use Support Vector Machines model in the analysis part, missing value must be removed or imputed. We imputed the missing values with median, as TotalCharges is skewed.

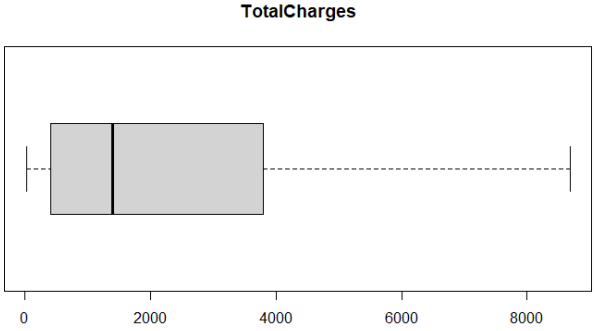

Figure 4: Distribution of TotalCharges

After imputation, the dataset still has 7043 valid rows, and no missing data exists.

Outlier Removal

Outlier is a data point that differs significantly from other observations.[6] The presence of outliers can give higher mean value and have other influence in statistical analysis. Some of the machine learning algorithms are robust to outliers, still we are identifying if there are any outliers in the dataset. We identified 0 outliers using box plots (**Figure 5)** and Z-score method after normalization. Therefore, Outlier removal is not required.

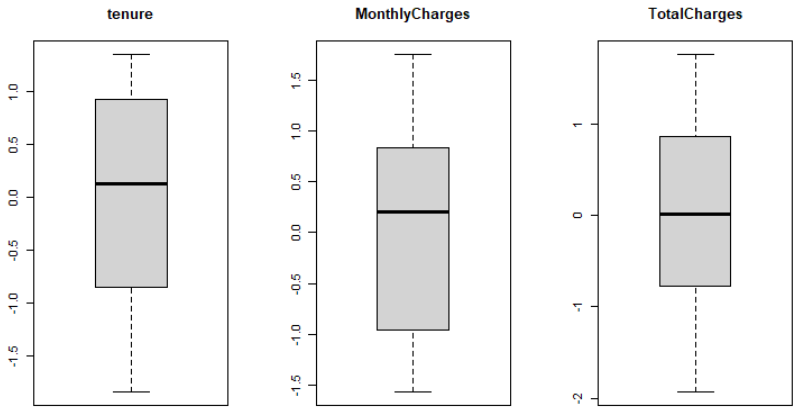

Figure 5: Distribution of transformed numerical variables

Data Transformation

Data transformation is a mathematical technique to transform data, so it can meet the statistical assumptions. Usually, we can use power transform to stabilize variance and make the data more like normally distributed. Also, in certain analysis we need to perform data standardization to make means and variances between 0 and 1.

In our case, we are going to perform k-means clustering on our numerical variables as k-Means uses Euclidean distance. Therefore, we first used Yeo-Johnson transformation to normalize the numerical variables, then standardize the normalized data.

Variable Conversion

As mentioned earlier in the data quality overview, there are 17 nominal variables and 1 ordinal variable. We converted the variables into factors for further analysis. For the ordinal variable "Contract," the levels are 'Month-to-month' < 'One year' < 'Two years' which was converted to 1, 2 and 3.

Data Discretization

Data discretization is a method that divides a continuous numerical variable into k discretized ranges. We wanted to find the relationship between tenure and churn, therefore we discretized churn into 8 groups, as 1 group stands for 9 months. The discretized variable is used in modeling.

Data Binarization and Recoding

Binarization is a process that transforms data into binary numbers which can make some of the modeling methods more efficient. We did binarization because Support Vector Machines requires nominal variables to be binarized and ordinal variables in numeric forms.

In our dataset, there are 6 nominal variables that have 2 levels, and we transformed them into 0 and 1. 9 nominal variables have 3 levels, and we transformed them by creating dummy variables. For example, the variable 'MultipleLines' has levels as 'No,' 'No Phone Service' and 'Yes,' and we created three dummy variables named as 'MultipleLines.No,' 'MultipliLines.Phone.Service' and 'MultipleLines.Yes' with 0 and 1 as values. For the ordinal variable 'Contract,' we transformed it into numeric form, which has value from 1 to 3. The processed dataset has 40 variables.

Data Splitting

Data splitting is to split the dataset into two parts, one is the training set which is used to train and develop models, and the other one is the testing set which is used to check if the model works correctly and the goodness of fit. We split the dataset with a ratio of 80/20, which means 80% of the data is in the training set and 20% of the data is in the testing set. After splitting, there are 5636 observations in the training dataset and 1407 observations in the testing dataset.

ANALYSIS RESULTS

Analysis Methods

Unsupervised Method: k-Means Clustering

K-Means Clustering is a partitional cluster method that divides observations to k clusters (the number of k needs to be predefined), based on similarity or the minimal distance to the cluster center. The reason we are using k-Means Clustering is that we want to find out if churn is related to charges, so that we can check if the current pricing policy is unreasonable.

To find out the optimal number of clusters, first we used the Silhouette method on the transformed numerical variables. The silhouette shows which objects lie well within the cluster and which ones are in between the clusters. It can give a relative quality of the clusters and an overview of the data configuration.[7] It can help in determining the optimal number of clusters.

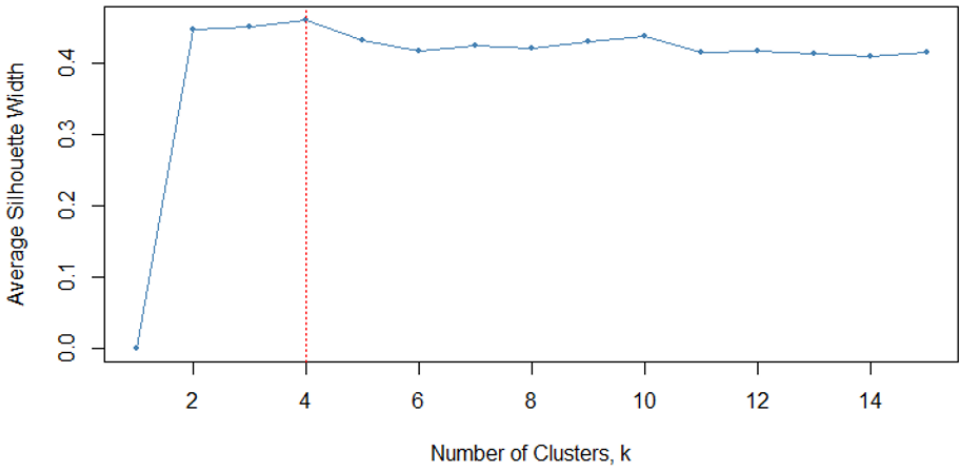

Figure 6: The Silhouette Plot

We have the maximum average silhouette width at number 4 (**Figure 6**), so we use 4 as our optimal number of clusters. The k-Means clustering gives the number of customers in each cluster and the cluster centroid information (**Figure 7)**.

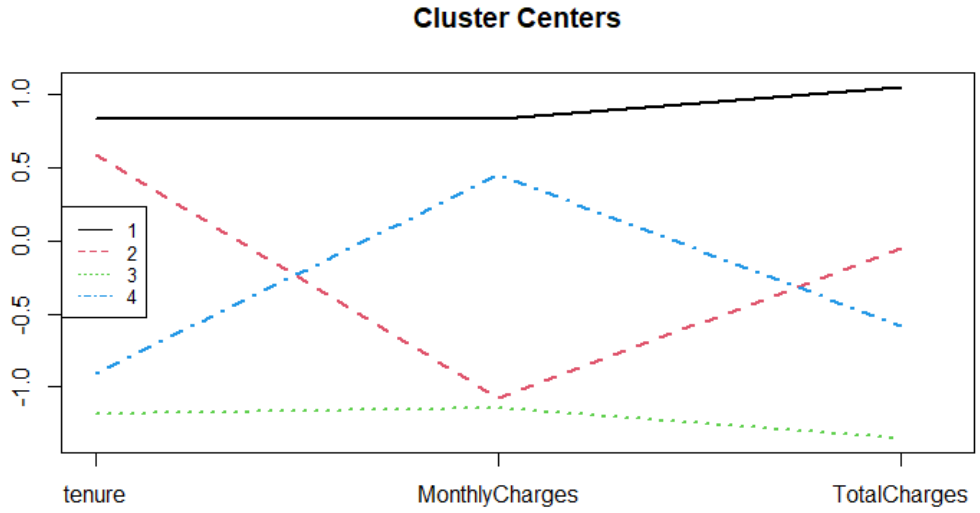

Figure 7: Cluster Centers

Table 4: Number of Customers in Clusters

Cluster 1	Cluster 2	Cluster 3	Cluster 4
2610	1459	1242	1732

The clustering result suggests that over 1700 customers (in Cluster 4) stay in the company for a short time and have the second highest monthly charges (**Table 4)**. This could be the reason they cancelled the contract and so the tenure is low. Also, 2610 customers (in Cluster 1) are choosing to stay in the company for a long time, but their monthly charges are also high. This suggests that Telco 's current pricing policy is very unreasonable.

To prove our insight, we performed both external and internal validation. For external validation, the number of churned and not churned customers are shown in the table below (**Table 5)**:

Table 5: Distribution of Churn and Not Churn Customers in Clusters

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
No	2109	1368	878	819
Yes	501	91	364	913

Cluster 1 and Cluster 4 have the highest churn number, which proves our assumption about unreasonable pricing as cause of customer churn.

For internal validation, we calculated sum of squared error (SSE) which is a measure of variation within a cluster (**Figure 8)**. The low SSE suggests the observations in a cluster are similar. Using the scree plot below, we identified an elbow at point 3 and 4, which suggested that our choice of cluster number was viable.

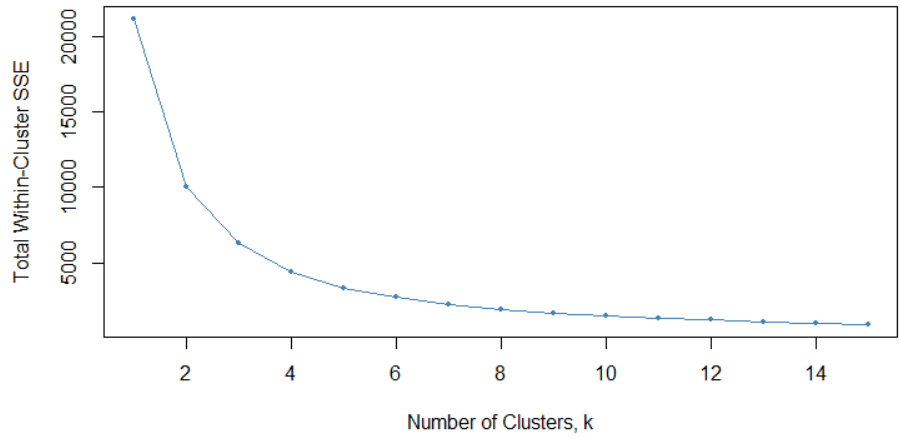

Figure 8: Within-cluster sum of squares

However, because k-Means Clustering can only be used on numerical variables, and we only have 3 numerical but 17 categorical variables, this method can only provide us with limited insights. Therefore, we performed supervised methods for further analysis.

Supervised Method: Decision Trees

Decision Tree is a data mining model used to classify an object or an instance into a predefined set of classes based on their attribute values.[8] The advantage of using decision tree is that it is self-explanatory and easy to interpret if small. It is also robust to irrelevant, redundant features, outliers, and missing values. Decision tree models are easy and fast to train and test, which could be a crucial factor that the company may consider. As we are building an anti-churn model to predict if a customer is staying with the company, we thought the decision tree could be an ideal model to use.

First, we built a basic decision tree as shown in **Figure 9**. The basic tree has 6 leaf nodes and 4 internal nodes. Among the 19 predictor variables, only 5 are involved in the model. About the model performance, we had 0.7981 training accuracy and 0.7903 testing accuracy, 0.9125 training specificity and 0.9110 testing specificity. However, the sensitivity value for training and testing model were only 0.4813 and 0.4558, the F1 value for training and testing were 0.5586 and 0.5354.

This suggests that although the model is balanced, our basic decision tree model predicted negative category which is customers staying in the company, but the performance of churn prediction was poor. Therefore, we will try the hyperparameter tuning model.

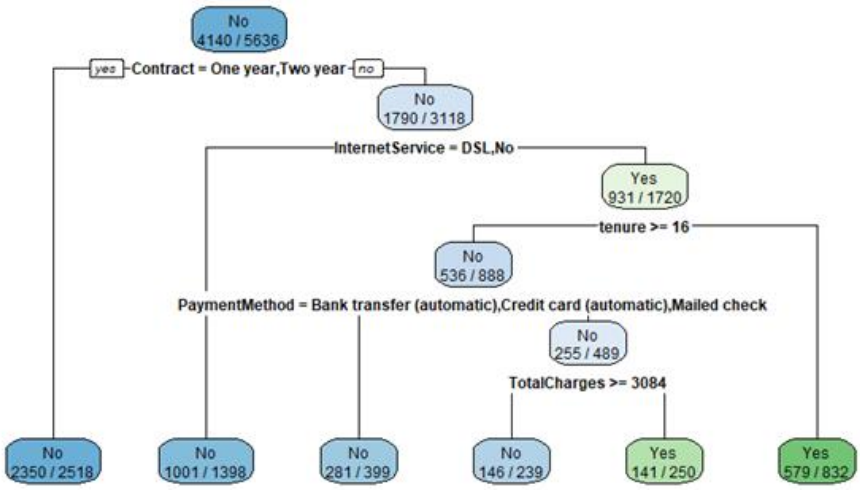

Figure 9: Basic Decision Tree

Hyperparameter tuning or optimization is to choose certain hyperparameter which can control the learning process for machine learning algorithm to optimize performance. It can reduce human effort necessary for applying machine learning, improve performance of machine learning algorithms and improve reproducibility and fairness of scientific studies.[9]

First, we wanted to conduct a grid search and choose the complexity parameter that is associated with the smallest cross-validation error. We searched the complexity parameter from the value of 0 to 0.2 by 0.005. Next, we set up a control object for a 10-fold cross validation on the grid search, repeated 3 times. We tuned the model with the grid search and the 10-fold cross validation control object, and the metric we chose to find the optimal model is accuracy. The hyperparameter tuned model gave us similar performance result with the basic decision tree model that had high accuracy and specificity but low sensitivity and F1 value. Therefore, we considered whether the class imbalance exists and whether it influenced the model performance.

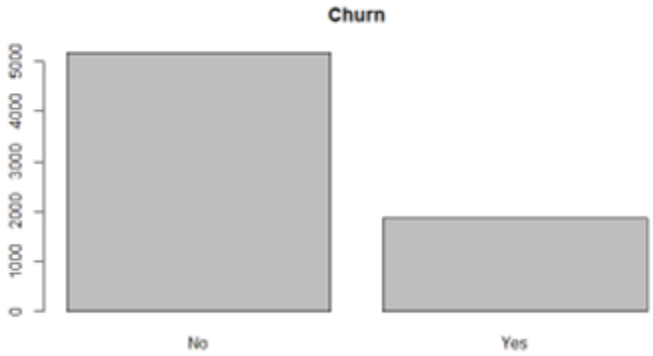

Figure 10: Distribution of Churn

As shown in **Figure 10**, there are over 5000 not churned customers and less than 2000 churned customers. In the real-world, the number of churned customers is always less. Therefore, we chose to use class weighting which will not change the size of the training set. The target variable 'Churn' has two levels, and we gave a weightage of 1.8837 to category 'Yes' and a weightage of 0.6807 to category 'No' as calculated.

We retuned the model with the same grid and control object but added weightages to the target variable. The retuned model had an optimal complexity parameter of 0.005, a small complexity parameter value suggested that the penalty of having many splits was low, and the decision tree was large. The performance measure values are in the table below **(Table 6):**

Table 6: Weighted Decision Tree Performance

	Training	Testing
Accuracy	0.7541	0.7655
Kappa	0.4556	0.4682
Sensitivity	0.7848	0.7641
Specificity	0.7430	0.7659
Precision	0.5245	0.5407
Recall	0.7848	0.7641
F1	0.6288	0.6333

With class weightage, the model improved a lot on sensitivity, which means the model's ability to predict churned customers improved. The specificity decreased due to low weightage assigned to negative class but still it exists moderately. The model also improved on the F1 value.

As a supplement and for comparison, we will use another supervised method in the next part.

Supervised Method: Support Vector Machines

Support Vector Machines are a set of supervised models that can be used for classification and regression, it involves constructing a hyperplane or a set of hyperplanes in a high dimensional space and to search for the optimal hyperplane, which is defined as the linear decision function with maximal margin between vectors of the two classes.[10] The hyperplane is the decision boundary to classify data. The strengths of Support Vector Machines are that they are robust to irrelevant and redundant variables and noises, they are not very prone to overfitting, and they are effective with high dimensional data.

However, the weakness is that SVMs are complex and can take a lot of time to train, model selection requires testing. The model does not give predicted probabilities and it can be difficult to interpret. Still, SVM is considered as one of the most powerful machine learning algorithms. We used SVM here in comparison with Decision Tree to offer the company the best model.

Similarly, we performed hyperparameter tuning. First, we wanted to conduct a grid search, but the performance metric we chose here is the cost and sigma. In Support Vector Machines, cost C acts as a regulation parameter associated with training error; larger cost means minimal training error while small cost means maximal margin. Sigma is a parameter that controls the level of non-linearity introduced in the model; larger sigma means the decision boundary tends to be linear, while small sigma means the decision boundary is highly nonlinear.[11]

We searched cost from 1 to 5 by 1, and sigma from 0.01 to 0.11 by 0.02. Next, we set up a control object for a 5-fold cross validation, repeated 3 times. For the modeling, we took class imbalance into consideration and gave weightage to both churned and not churned customers. In model tuning, we performed data standardization, and used Area Under the Curve as our metric.

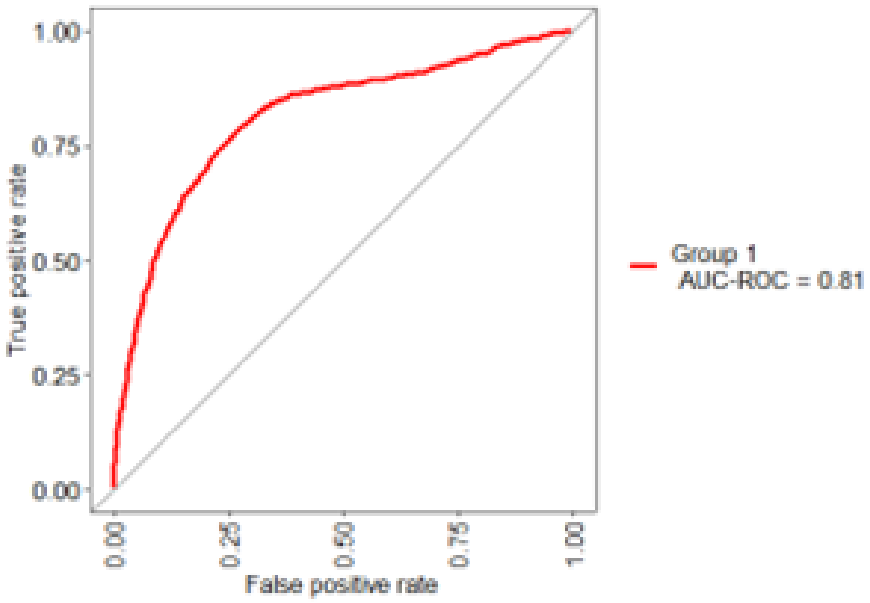

Figure 11: ROC Curve

The optimal model had a cost value of 1 and sigma value of 0.01. The largest Receiver Operating Characteristic (ROC) value was 0.8095. (**Figure 11)** This suggests that our model was focusing on maximizing the margin and the decision boundary is highly nonlinear.

Table 7 Support Vector Machines Model Performance

	Training	**Testing**
Accuracy	0.8078	0.7960
Kappa	0.4466	0.4159
Sensitivity	0.4666	0.4531
Specificity	0.9312	0.9197
Precision	0.7101	0.6706
Recall	0.4666	0.4531
F1	0.5631	0.5408

The model performance is shown in the table below **(Table 7).** Even with class weighting, the model still had low Sensitivity, Recall and F1 value, which means the model is classifying the churned customers poorly.

Analysis Method Summarization

- The k-Means Clustering suggests that the low tenure and high monthly charges associated are causing customer churn, and the current pricing is unsustainable.
- Performance comparison for supervised methods **(Table 8):**
  Among the models we built, the Support Vector Machines model had the highest accuracy, specificity, and precision, but even with class weighting it did not have high sensitivity which shows its ability to correctly predict churned customers. Support Vector Machines models took a long time to train, which could be expensive for the company. Also, this model is hard to interpret. Therefore, we would recommend the company use our weighted decision tree as the anti-churn model to track the customers' susceptibility to churn.
  The weighted decision tree model had overall high accuracy, sensitivity, specificity, this suggested that after class weighting the model could predict both churned and not churned customers appropriately. The Kappa values showed moderate agreement, and it is highest among the models we built. After class weighting, the model had a decrease in precision but recall and F1 value improved. Overall, the model is well-constructed based on real world data.
  For the goodness of fit, the model performed better on testing set for some of the measurements. However, as the performance measurements had consistency and were close between training and testing set, we can say that the model is balanced.

Table 8: Performance Comparison

	Basic Decision Tree		**Weighted Decision Tree**		**Support Vector Machines**	
	Training	Testing	Training	Testing	Training	Testing
Accuracy	0.7981	0.7903	0.7541	0.7655	0.8078	0.7960
Kappa	0.4302	0.4053	0.4556	0.4682	0.4466	0.4159
Sensitivity	0.4813	0.4558	0.7848	0.7641	0.4666	0.4531
Specificity	0.9126	0.9110	0.7430	0.7659	0.9312	0.9197
Precision	0.6654	0.6489	0.5245	0.5407	0.7101	0.6706
Recall	0.4812	0.4558	0.7848	0.7641	0.4666	0.4531
F1	0.5886	0.5354	0.6288	0.6333	0.5631	0.5408

- The weighted decision tree model provides variables of importance, Table **(Table 9)** below shows the variables had weightage larger than 50.
  The model shows that contract type is a key factor. The company should be concerned that the absence of various plan options is causing churn. Both monthly and total charges are important variables and based on our analysis in k-Means Clustering, pricing policy should be improved. The model also suggested that customers need high quality services, such as internet services, tech support and online security service.

Table 9: Variables of Importance

Variable	Weightage
Contract.L	650.9357
tenure	369.4142
Contract.Q	319.6861
TotalCharges	260.2167
MonthlyCharges	111.5918
InternetService Fiber optic	103.3431
TechSupportYes	84.4433
OnlineSecurityYes	73.5144

DISCUSSION & CONCLUSION

Solution & Recommendations

Based on our analysis below are some solutions and recommendations for the company:

Considering a long-term healthy operation and development in the telecom industry, the company must be alert to the potential risks that arise from the customer churn. To avoid customer churn in advance and to economize on expense for the customer retention, effective department-related changes should be made across different departments in the enterprise.

Customer Service Department should consider utilizing the anti-churn model based on which a list of users with a high churn rate can be used to monitor and track the customers' susceptibility to churn before they are leave the company. Regularly, the list should be updated and synchronized with relevant departments that participate in solving customer’s problems.

Meanwhile, the company should provide a high quality of service and attitude. For the basic network services, especially fiber optic services, users are taking fiber optic services instead of DSL services which means that they have a strong demand for network speed and quality, but the current service quality is not ideal. So, the customers are unhappy with other service quality such as online security, and technical support including the plan.

In the process of customer service, customers not only provide complaints, but also advice for improvements and other market information that play a positive role in elevating ideas for Research and Design Department and finding out the shortcomings or deficiencies of products in terms of quality and performance. And hence, with almost half of churners leaving within a short tenure, introducing a variety of plans, for example 3-, 6- or 9-month plans, should be considered as a method of providing diverse choices when customers initially purchase a plan.

Under the modern marketing concept, marketing activities come from the customers' needs and feedback. Reducing the customer churn rate and winning more new customers is particularly important. For the Marketing Department, on one side the customer loyalty should be enhanced, and on another side, they should launch luring marketing campaigns for existing customers such that month-to-month customers are willing to switch to a yearly plan.

Overall, the prediction model allows the company to identify customers with a considerable risk of churn in advance. Besides, making the product diverse and improving the quality of products and customer service should be improved by the telecom company.

Conclusion

Along with the journey competition among enterprises is always going to be aggressive, how to manage the enterprise scientifically and prevent unnecessary spending in relevant fields is an urgent and fatal topic for companies to think about. As an important indicator to measure the daily operation health, customer churn should be minimized by the enterprise. Effective management and monitoring of the customer churn can help the company save money to the greatest extent and improve overall revenue performance.

REFERENCES

[1] https://www.invespcro.com/blog/customer-acquisition-retention
[2] https://hbr.org/2014/10/the-value-of-keeping-the-right-customers
[3] Redman, Thomas C. (30 December 2013). Data Driven: Profiting from Your Most Important Business Asset. Harvard Business Press. ISBN 978-1-4221-6364-1.
[4] Herzog, T.N.; Scheuren, F.J.; Winkler, W.E. (2007). "Chapter 2: What is data quality and why should we care?". Data Quality and Record Linkage Techniques. Springer Science & Business Media. pp. 7–15. ISBN 9780387695020. Archived from the original on 31 July 2020. Retrieved 18 April 2020.
[5] Wu, S. (2013), "A review on coarse warranty data and analysis" (PDF), Reliability Engineering and System, 114: 1–11, doi: 10.1016/j.ress.2012.12.021
[6] Grubbs, F. E. (February 1969). "Procedures for detecting outlying observations in samples". Techno metrics. 11 (1): 1–21. doi:10.1080/00401706.1969.10490657. An outlying observation, or "outlier," is one that appears to deviate markedly from other members of the sample in which it occurs."
[7] Peter J. Rousseeuw (1987). "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis". Computational and Applied Mathematics. 20: 53–65. doi:10.1016/0377-0427(87)90125-7.
[8] Rokach, Lior; Maimon, O. (2014). Data mining with decision trees: theory and applications, 2nd Edition. World Scientific Pub Co Inc. doi:10.1142/9097. ISBN 978-9814590075. S2CID 44697571
[9] Matthias Feurer and Frank Hutter. Hyperparameter optimization. In: AutoML: Methods, Systems, Challenges, pages 3–4.
[10] Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vector networks" (PDF). Machine Learning. 20 (3): 273–297. CiteSeerX 10.1.1.15.9362. doi:10.1007/BF00994018. S2CID 206787478.
[11] Christos Theodoropoulos, "Support Vector Machines under the hood - An advanced explanation (Classification tasks) — Part 1", Towards Data Science, https://towardsdatascience.com/support-vector-machines-under-the-hood-c609e57a4b09