Customer Churn Prediction Analysis

19CSE305 - Machine Learning

1st Abhinav S
Amrita School of Engineering
CB.EN.U4CSE20301

2nd Jietthesh Balaji M
Amrita School of Engineering
CB.EN.U4CSE20322

3rd Kausalyaa Sri
Amrita School of Engineering
CB.EN.U4CSE20326

4th Kishore Kumar V S
Amrita School of Engineering
CB.EN.U4CSE20328

5th Sujith Roshan N
Amrita School of Engineering
CB.EN.U4CSE20363

***Abstract*—Customers are a company's most valuable asset, and keeping customers is critical for any organisation looking to increase revenue and develop long-term meaningful relationships with customers. Furthermore, the cost of obtaining a new client is five times that of keeping an existing customer. Customer Churn/Attrition is one of the most well-known business difficulties in which consumers or subscribers drop doing business with a service or a firm. Ideally, they will no longer be a paying customer. A client is considered to have been churned if a certain length of time has passed since the consumer last interacted with the company. Identifying whether or not a client will churn and offering relevant information aimed at customer retention are crucial to lowering churn. Our brains cannot anticipate customer turnover for millions of clients; here is where machine learning may assist.**

Keywords—Machine Learning, Customer churn prediction, Attrition

I. Introduction

The development and digitalization of the world has led to new ways of doing business and companies all over the globe have been forced to adapt. Subscription based services are one of the outcomes of the explosive digitalization that has taken the world by storm and with this comes both possibilities and challenges that require modern day solutions. The digitalization has also brought forward an ongoing trend to improve current data processing activities as a part of customer relationship management strategies. In a subscription-based business model, a fundamental part of success is to minimize the rate of customers ending their subscriptions, in other words, to minimize churn [2].

Customer churning refers to the action of when a customer chooses to abandon their service provider. The term is relatively new and has gained more relevance with the emergence of online services. Firms across the globe recognize customer churning as a great loss since they have already invested in attracting these customers. This is one of the major reasons that customer retention is beneficial for a firm. Customers can churn for many reasons and it is hard to pinpoint a general reason for churning. The availability of information has given consumers a bargaining power, and nowadays customers can easily find the service provider, which provides the same product with a more satisfying deal. To manage this, firms invest in customer churn prediction, which means that companies try to predict which of their customers will churn, so that they can apply preventative measures. These preventive measures could differ depending on the reason a customer might churn, and could be for example, offering a lower price or including an extra service.

II. Problem Formulation

The subscription-based business model is continuously growing due to digitalization and offers companies an innovative way of conducting their business [2]. At the same time, more and more services are being digitalized and data has become much easier to collect, store, and process [3]. There is an abundance of different service providers to choose from, which has increased competition and made it more difficult to retain customers, in this modern-day service market [2]. Due to the availability of data and substitutes, subscription-based businesses must adapt by focusing more on Customer Relationship Management, specifically customer churn management [3]. The key to success within a subscription-based business is to keep a low churn rate, which is defined as the number of customers leaving their service provider during a given period of time. As a service provider, there is a greater chance of selling to an existing customer rather than a completely new one. This can be highlighted by the cost of attracting new customers exists in machine learning is the regression type. attracting new customers can cost somewhere between five to six times more relative to retaining customers, which states the importance of preventing churn.

Churn prediction and prevention can be very beneficial for a firm's reputationSince profits are derived from customers, a decrease in the churn rate has a significant impact on the firm's profits. Customer relationship management can lead to a better and more valuable relationship to the customers and create some sort of loyalty towards the company, which could increase revenue streams over time.

III. Methodology

A. Dataset

This dataset contains customer information from a fictional telco company. This company provides various services such as streaming, phone, and internet services.

i) customerID - ID number of the customer (Identifier).
ii) Churn - Churn status, whether the customer churned or not (Target Variable).
iii) gender - Whether the customer is a male or a female.
iv) SeniorCitizen - Whether the customer is a senior citizen or not.
v) Partner - Whether the customer has a partner or not.
vi) Dependents - Whether the customer has dependents or not.
vii) tenure - Number of months the customer has used the service.
viii) Contract - The contract term of the customer.
ix) PaperlessBilling - Whether the customer has paperless billing or not.

x) PaymentMethod - The customer's payment method.
xi) MonthlyCharges - The amount charged to the customer monthly.
xii) TotalCharges - The total amount charged to the customer. The values are of the type string object in the dataset.
Services that each customer has signed up for.
xiii) PhoneService - Whether the customer has a phone service or not.
xiv) MultipleLines - Whether the customer has multiple lines or not.
xv) InternetService - Customer's internet service provider
OnlineSecurity - Whether the customer has online security or not.
xvi) OnlineBackup - Whether the customer has online backup or not.
xvii) DeviceProtection - Whether the customer has device protection or not.
xviii) TechSupport - Whether the customer has tech support or not.
xix) StreamingTV - Whether the customer has streaming TV or not.
xx) StreamingMovies - Whether the customer has streaming movies or not.

B. Pre-Processing

After we load the data, we analyze the customer data and drop the unnecessary column CustomerID as it does not contribute to the learning process. The values in the column TotalCharges are converted from string to numeric. If there is any null value, then we will remove that row since the data will not be valid for the analysis and data cleaning is also done at the same time.

1. *Normalisation:* All the rows in the dataset is normalised using the StandardScaler() method, which uses z-score normalisation.

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService
0	-1.008596	-0.440921	1.032482	-0.653764	-1.285566	-3.051036	0.059553	-1.181912
1	0.991477	-0.440921	-0.968540	-0.653764	0.060346	0.327757	-0.994654	-1.181912
2	0.991477	-0.440921	-0.968540	-0.653764	-1.244781	0.327757	-0.994654	-1.181912
3	0.991477	-0.440921	-0.968540	-0.653764	0.508983	-3.051036	0.059553	-1.181912
4	-1.008596	-0.440921	-0.968540	-0.653764	-1.244781	0.327757	-0.994654	0.175873
...	...	...	...	...	...	...	...	...
7005	0.991477	-0.440921	1.032482	1.529603	-0.347506	0.327757	1.113760	-1.181912
7006	-1.008596	-0.440921	1.032482	1.529603	1.610184	0.327757	1.113760	0.175873
7007	-1.008596	-0.440921	1.032482	1.529603	-0.877714	-3.051036	0.059553	-1.181912
7008	0.991477	2.267980	1.032482	-0.653764	-1.163210	0.327757	1.113760	0.175873
7009	0.991477	-0.440921	-0.968540	-0.653764	1.365473	0.327757	-0.994654	0.175873

Figure 3.1 Customer Churn prediction data after normalisation

2. *Data Balancing:* The given dataset is imbalanced, favouring the outcome 'YES' in the ratio 70:30. This is balance by using SMOTE (Synthetic Minority Oversampling Technique).

C. Models

Our customer churn Prediction has made a comparison of accuracy in Pointing the outcome of Testing set between 11 different models. That includes Logistic Regression, K – Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree, Naïve Bayes, Random Forest, Perceptron, Principal Component Analysis (PCA), K Means Clustering, ADA Boost, XG Boost.

D. Data Visualization:

The Figure 3.2 shows the number of customers who opted for various services from the telecommunication company according to the dataset. This also shows customers who did not opt for the service(s).

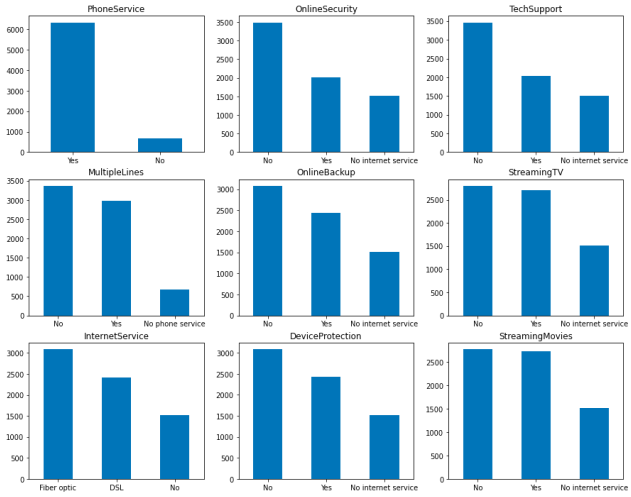

Figure 3.2 Feature Plot

IV. MODELS

With the help of data visualization, we can see how the data looks like and what kind of correlation is held by the attributes of data. It is the fastest way to see if the features correspond to the output.

A. K – Nearest Neighbours (KNN)

The k-nearest neighbours' algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

The accuracy of the model negligibly improved after dimensionality reduction using PCA and continued to improve after parameter tuning. But the improvement in performance was only by around 1%.

B. Support Vector Machine (SVM)

SVMs are used in applications like handwriting recognition, intrusion detection, face detection, email classification, gene classification, and in web pages. This is one of the reasons we use SVMs in machine learning. It can handle both classification and regression on linear and non-linear data.

The SVM classifier is one of the best performing models with a highest accuracy of 0.833 which is obtained after parameter tuning.

C. Decision Tree

A decision tree is a type of supervised machine learning used to categorize or make predictions based on how a previous set of questions were answered. The model is a form of supervised learning, meaning that the model is trained and tested on a set of data that contains the desired categorization.

It is observed that the accuracy of the model dropped after dimensionality reduction using PCA. But the drop in performance is only by 0.7%.

D. Logistic Regression

Logistic regression is a statistical method that is used for building machine learning models where the dependent variable is dichotomous that is binary. Logistic regression is used to describe data and the relationship between one dependent variable and one or more independent variables.

This model is not affected by both PCA and parameter tuning and the accuracy score is observed to be the same.

E. Naïve Bayes (NB)

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset.

The NB model comparatively scores better with its best accuracy score being 0.836, which is achieved after dimensionality reduction. It is an improvement of approximately 5.5% over the standard model, without PCA.

F. Random Forest

Random Forest is a powerful and versatile supervised machine learning algorithm that grows and combines multiple decision trees to create a "forest." It can be used for both classification and regression problems in R and Python.

Random forest model gained performance by 5.5% after PCA.

G. Multi-LayerPerceptron

A neural network link that contains computations to track features and uses Artificial Intelligence in the input data is known as Perceptron. This neural links to the artificial neurons using simple logic gates with binary outputs.

The model is one of the best performing models to predict customer churn attrition. The model performs better after PCA and gains more performance after parameter tuning.

H. K-Means clustering

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabelled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

This clustering model does not fit well with the given data. The average accuracy is only around 60%. This is because of how the data is clustered. The silhouette score verifies this observation in the dataset. The silhouette score is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighbouring clusters [4]. The silhouette score is only 0.16, which means the data points are not well matched with its cluster.

I. Dimensionality Reduction using Principal Component Analysis (PCA)

Principal component analysis, or PCA, is a dimensionality-reduction method that is used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one.

PCA has helped to improve accuracies of some models while the accuracies of other models have stayed the same.

V. Comparison in accuracy of Different Model in Classification

Model	Before PCA and Parameter Tuning	After PCA	After Parameter Tuning
Logistic Regression	0.778	0.778	0.778
K - Nearest Neighbours	0.760	0.769	0.784
Naive Bayes	0.792	0.836	0.781
Decision Tree	0.763	0.757	NA
Random Forest	0.792	0.836	NA
SVM	0.802	0.802	0.833
K - Means Clustering	0.659	0.660	0.660
ADA Boost	0.819	0.782	NA
XG Boost	0.843	0.826	NA
Multi-Layer Perceptron	0.815	0.819	0.847

VI. Conclusion

Our results indicate that customer churning can be predicted with a respectable accuracy using machine learning. Furthermore, from our results we can conclude that ensemble learners, using boosting, improved the performance of our churn prediction model, relative to the studied single learner Naïve Bayes. In particular, the XGBoost classifier performed best based on overall accuracy before parameter tuning and PCA. It should be emphasized that the dataset used for this study was imbalanced, and we applied different sampling methods in order to investigate how balancing the training dataset would affect the result. We conclude that balancing, using sampling methods, has a positive influence for the studied problem.

VII. References

[1] F. Khodakarami and Y. Chan, "Exploring the role of customer relationship management (CRM) systems in customer knowledge creation," Information & Management, vol. 51, pp. 27-42, 2014.

[2] D. Buö and M. Kjellander, "Predicting Customer Churn at a Swedish CRM-system Company," Linköpings Universitet, Linköping, 2014.

[3] M. Sergue, "Customer Churn Analysis and Prediction using Machine Learning for a B2B SaaS company," KTH, Stockholm, 2020.

[4] https://en.wikipedia.org/wiki/Silhouette_(clustering)