Enhancing Customer Retention in Telecommunications: Leveraging Predictive Algorithms to Forecast Customer Churn in Telco

Project Proposal

Irene Sunny	Ru Yi	Shuqi Deng	Wenjing Zhao
CentraleSupélec	CentraleSupélec	CentraleSupélec	CentraleSupélec
Gif-sur-Yvette, France	Gif-sur-Yvette, France	Gif-sur-Yvette, France	Gif-sur-Yvette, France
irene.sunny@student-cs.fr	ru.yi@student-cs.fr	shuqi.deng@student-cs.fr	wenjing.zhao@student-cs.fr

1. PROBLEM DEFINITION

Forbes research reveals that according to what industry you are in, acquiring a new customer is anywhere from five to twenty-five times more expensive than retaining an existing customer. Keeping the right customer is valuable and one of the key metrics in understanding whether your company is retaining customers is customer churn rate. Churn rate describes the rate at which customers abandon a product or service. Churn is especially relevant to the telecommunications market, largely because it's a subscription-based sector. In the 2022 State of Customer Churn in Telecom survey, it was found that customer loyalty to telecom providers is down 22% post-pandemic, with customer stickiness being impacted more by the customer experience than ever. Competition is at an all-time high and many telcos are recognizing the need to improve customer experience and service in order to lower customer churn rates and compete effectively.

The primary problem this project seeks to address is to predict and understand customer churn patterns using dominant determinants within a fictional telecommunication company, Telco. By leveraging different predictive modelling, we aim to develop a robust solution that can foresee potential customer departures. Our study can facilitate the construction of an effective prediction tool for managers in the telecom sector to discover the underlying churn-risk customers which have a high probability of transferring to their competitors. Under such circumstances, the managers can take favourable measures to re-attract those underlying churners by promoting the dominant determinants revealed by this study through ML feature selection techniques.

2. METHODOLOGY

The focus of this project is on predicting customer churn based on several factors provided in the dataset. The comparison between four common prediction models would be used to determine the most suitable technique for customer churn prediction in the telecommunication industry.

2.1 Artificial Neural Networks (ANN)

ANN is a machine learning model suitable for learning non-linear patterns and relationships between features in complex datasets. It can be used for classification, regression analysis, and unsupervised learning. With the ability to automatically learn relevant features from smaller datasets, it reduces the need for manual quantifying of the importance of a variable on churn rate and is scalable to large datasets.

2.2 Decision Tree

Decision Trees (DTs) are tree-like structures used to establish sets of decisions that can generate classification rules tailored to a specific dataset. These tree models are also known as Classification Trees or Regression Trees, where the terminal nodes (leaves) represent class labels, and the branches represent combinations of features that lead to those class labels. While Decision Trees may not excel at capturing intricate and nonlinear relationships between attributes, they can still achieve high accuracy in addressing the customer churn problem, depending on the data's inherent characteristics.

2.3 Logistic Regression

Logistic regression is a statistical model used for binary classification in our prediction task, which, in this case, is to predict whether a customer will churn (leave). This model allows us to quantify the impact of each coefficient on the churn probability. It also allows for selective focus on important features that are suitable for our dataset with a large number of variables. Potential drawbacks of logistic regression include (i) potential overfitting when applied to small datasets, making it less suitable for future use with larger datasets, and (ii) linearity assumption which may limit its ability to handle complex interactions between independent and dependent variables.

2.4 Support Vector Machine (SVM)

Support Vector Machines (SVMs), also known as Support Vector Networks, are supervised learning models introduced by Boser, Guyon, and Vapnik. They are used for classification and regression analysis, based on structural risk minimization. Kernel functions are employed to enhance their performance, and ongoing research explores the best kernel selection. In churn prediction, SVMs often outperform Decision Trees (DT) and sometimes even Artificial Neural Networks (ANN), depending on the data type and transformations applied.

3. EVALUATION

The sample data obtained from GitHub tracks a fictional telecommunications company, Telco. The customer churn data sourced by the IBM Developer Platform. It consists of 7043 lines of customers, with 20 features. It includes the label whether the customer left within the last month, service the customer has signed up for and customer account information.

We will first clean the data, observe data characteristics through descriptive analysis, and delete outliers. Finally, the obtained results on the test set will be evaluated using a confusion matrix and AUC curve to get the model which has the best accuracy. We will select the model that demonstrates the highest accuracy and generalizability for prediction. Further, we will find out what factors influence customer churn rate to take advantageous measures. By doing so, we aim to provide valuable insights for the telecom industry, offering enhanced strategies for customer retention based on accurate predictive modelling for churn rates.

REFERENCES

[1] H. Jain, A. Khunteta, and S. Srivastava. 2020. Telecom churn prediction and used techniques, datasets and performance measures: a review. Telecommun Syst 76, 613–630 (2021). https://doi.org/10.1007/s11235-020-00727-0

[2] Jia Wertz. 2018. Don't Spend 5 Times More Attracting New Customers, Nurture The Existing Ones. Forbes. https://www.forbes.com/sites/jiawertz/2018/09/12/dont-spend-5-times-more-attracting-new-customers-nurture-the-existing-ones/?sh=badf6205a8e0

[3] T. Vafeiadis, K.I. Diamantaras, G. Sarigiannidis, K.Ch. Chatzisavvas. 2015. A comparison of machine learning techniques for customer churn prediction, Simulation Modelling Practice and Theory, Volume 55, 2015, Pages 1-9, ISSN 1569-190X https://www-sciencedirect-com.essec.idm.oclc.org/science/article/pii/S1569190X15000386

[4] B.E. Boser, I.M. Guyon, V.N. Vapnik. A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM (1992), pp. 144-152. https://dl.acm.org/doi/10.1145/130385.130401

[5] Prabadevi, B., Shalini, R. and Kavitha, B.R. 2023. Customer churning analysis using machine learning algorithms. https://doi.org/10.1016/j.ijin.2023.05.005