A Deep Dive into Churn Analysis for Telecommunications

Lovro Kostelić  
*Business mathematics and economics*  
 *Zagreb School of Economics and Management*Zagreb, Croatia  
[lkosteli@student.zsem.hr](mailto:lkosteli@student.zsem.hr)

*Abstract* – Customer churn poses a significant challenge for telecommunications companies, leading to financial losses and a decline in market share. With the surge in data generated by telecommunication transactions, manual identification of churn patterns has become impractical. To address this, the project aims to conduct an in-depth churn analysis using machine learning techniques, providing valuable insights for proactive customer retention strategies.

Keywords – Churn analysis, Data analysis, Machine learning, Churn prediction.

# Introduction

Churn analysis is a method to calculate the attrition rate of each company's customers. It involves the identification of those customers who are most vulnerable to discontinuing use of services or products. In the development of a sustainable and robust strategy for customer retention, churn analysis plays an extremely important role. When a company is aware of the percentage of customers who end their relationship with them in each time period, they can come up with a detailed analysis of the causes for the churn rate using churn analysis. This contributes to the development of effective customer retention programs for the company. The churn rate is usually applied to a wide range of industries, in particular subscription services such as long-distance phone service, magazines and commercial banks. Churn analysis helps to understand the behavior of customers who have cancelled and moved their business to a competitor, as well as predict how likely it is that an event will happen. Other uses vary from calculating employee attrition in any given company.

Churn analysis has become a cornerstone in the telecommunications industry, playing a crucial role in understanding and managing customer attrition. With increasing competition and evolving consumer preferences, telecom companies are recognizing the significance of analyzing churn rates to maintain market share and profitability.

In today's telecommunications landscape, characterized by rapid technological advancements and shifting consumer expectations, churn analysis provides invaluable insights into customer behavior and preferences. By examining factors such as tenure, contract type, monthly charges and total charges, telecom companies can identify potential churn triggers and implement targeted retention strategies.

Customer retention stands as a key factor for long-term success within the telecom sector, which boasted a valuation of USD 1754.8 Billion in 2022. Projections indicate a promising growth trajectory, with expectations set to climb to USD 2652.5 Billion by 2030, reflecting a Compound Annual Growth Rate (CAGR) of 5.3% between 2023 and 2030. High churn rates not only result in revenue loss but also impose higher acquisition costs to replace lost customers. Through churn analysis, telecom providers can identify areas for improvement, whether it will be enhancing network coverage, streamlining billing processes, or improving customer service responsiveness.

Moreover, telecom firms can use churn research to divide up their clientele and target retention campaigns at particular behavioral or demographic categories. Businesses may implement more successful retention campaigns and increase client loyalty by knowing the particular requirements and preferences of various customer segments.

As telecommunications companies continue to innovate and adapt to changing market dynamics, churn analysis remains a fundamental tool for optimizing customer retention strategies and driving sustainable growth. By prioritizing customer satisfaction and leveraging insights derived from churn analysis, telecom providers can position themselves for success in an increasingly competitive landscape.

*About the dataset*

Dataset used in this project is Telco Customer Churn, taken from Kaggle. The dataset includes details about customer churn, indicating whether customers have left within the last month. It includes information on the services each customer has subscribed to, including phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies. Additionally, the dataset provides customer account information such as tenure, contract type, payment method, paperless billing status, monthly charges, and total charges. Demographic information, covering gender, age range, and the presence of partners and dependents, is also included. This comprehensive dataset enables an in-depth analysis of factors contributing to customer churn within the telecommunications domain. In the dataset we have 7043 rows and 21 features.

# Exploratory Data Analysis (EDA)

As already said, churn poses a significant challenge in the telecom industry, with average monthly rates between 1.9% and 2% among the top US wireless carriers.

Upon analyzing our dataset in Python, we removed 11 rows with missing values to ensure data accuracy. Our analysis revealed key correlations with churn are month-to-month contracts, absence of online security, and tech support which are all positively correlated, while tenure and two-year contracts showed negative correlations. Interestingly, services like online security, streaming TV, and tech support, even without internet connection, were negatively related to churn.

A graph showing a line of different colored lines

Description automatically generated with medium confidence

*Figure 1. Correlation among churn and other features*

*Data exploration*

Beginning our analysis, we first explore the dataset to uncover underlying patterns and potentially formulate hypotheses. Initially, we examine the distribution of individual variables, followed by a deeper investigation to reveal significant trends. Regarding gender distribution, our findings indicate a balanced representation, with approximately equal proportions of male and female customers within our dataset. Upon examining tenure, we observe a diverse distribution, with a notable concentration of customers exhibiting short tenure durations, contrasted with a considerable unit maintaining a long-standing relationship with the telecom company, spanning approximately 72 months which we can see at the graph below.

A graph of customer satisfaction

Description automatically generated

*Figure 2. Number of customers by their tenure*

This variance in tenure duration likely reflects the influence of disparate contractual arrangements, which may impact customers' inclination to retain or sever ties with the telecom provider.

Interestingly most of the monthly contracts last for 1-2 months, while the 2-year contracts tend to last for about 70 months. This shows that the customers taking a longer contract are more loyal to the company and tend to stay with it for a longer period of time. Now let's take a quick look at the relation between monthly and total charges. We can observe that the total charges increase as the monthly bill for a customer increases.

A graph of a graph showing the amount of charge

Description automatically generated with medium confidence

*Figure 3. Relationship between total charges and monthly charges*

Within our dataset, it is evident that 74% of customers don't churn. This skewness in the data aligns with expectations, as a substantial majority of customers are typically anticipated to remain loyal. However, it is important to acknowledge this skewness during our modeling actions, as it could potentially contribute to an elevated frequency of false negatives. Consistent with the correlation plot findings, customers with month-to-month contracts exhibit notably high churn rates which is seen on the following graph.

A graph of a number of different colored squares

Description automatically generated with medium confidence

*Figure 4. Customer churn by contract type*

# research - simulations

In our predictive modeling attempts, we did a comprehensive exploration of various machine learning algorithms, involving CatBoost, LightGBM, XGBoost, Decision Tree, and AdaBoost, all of which proved to work quite well for our analysis. Our initial approach involved splitting the dataset into training and testing subsets using scikit-learn library in Python. Through this method, we evaluated the performance of each model and found that AdaBoost yielded the most promising results, boasting an accuracy rate of 79.71%. This outcome underscored AdaBoost's effectiveness in capturing the underlying patterns within the data and making accurate predictions within a controlled testing environment.

Subsequently, we conducted another round of predictions using cross-validation techniques to ensure reliability in our findings. Once again, AdaBoost emerged as winner among the models considered, showcasing an impressive accuracy score of 80.76%. This superior performance reaffirmed AdaBoost's ability in accurately predicting outcomes, underscoring its viability as the preferred prediction method within our analytical framework. Through accurate evaluation and validation, our study explains the effectiveness of AdaBoost in addressing predictive challenges within the telecom industry landscape.

In our analysis of feature importance, using Shapash library, within our best-performing model, tenure emerges as the most influential factor, a result that aligns intuitively with the significance of customer tenure in determining churn behavior. Following closely behind is internet service with fiber optic, which utilizes a significant influence on churn prediction, indicating a strong correlation between the type of internet service and customer retention. Additionally, contractual agreements play a key role as well, with both two-year and one-year contracts featuring highly among the top predictors of churn. This finding underlines the impact of contractual terms on customer loyalty and stresses the importance of contractual flexibility in shaping churn dynamics within the telecom industry landscape.

A yellow bar chart with white text

Description automatically generated

*Figure 5. Feature importances*

Upon constructing the confusion matrix, our analysis reveals that the model accurately predicted 302 instances of churn (true positives) out of 561 individuals who actually churned, while incorrectly classifying 259 cases as churn (false positives). Furthermore, the model correctly identified 1380 cases of non-churn (true negatives) out of 1549 individuals who did not churn, while misclassifying 169 instances as non-churn (false negatives).

A chart of different colors

Description automatically generated with medium confidence

*Figure 6. Confusion matrix*

# Conclusion

In summary, our research shows that the predictive model we developed is pretty successful, with an accuracy rate of over 80%. It does a good job of guessing whether customers will stay with the telecom company or leave. By looking at things like contract types, how long customers have been with the company, and what kind of internet they use, the model can make these predictions accurately. This means telecom companies can use this model to figure out how to keep more customers happy. As we keep working on our model, we expect it to get even better at predicting customer behavior, helping telecom companies keep their customers and grow their business.

# references

1. Hancock, A. (2023) Telecommunication market size, Share & Trends Analysis Report by 2030, LinkedIn. Available at: <https://www.linkedin.com/pulse/telecommunication-market-size-share-trends-analysis-report-hancock-pa35f/> (Accessed: 18 March 2024).
2. Churn analysis: Predicting Churners. Available at: <https://www.researchgate.net/publication/295257947_Churn_analysis_Predicting_churners> (Accessed: 18 March 2024).
3. Bandiatindra (2019) Telecom churn prediction, Kaggle. Available at: [https://www.kaggle.com/code/bandiatindra/telecom-churn-prediction](https://www.kaggle.com/code/bandiatindra/telecom-churn-prediction%20) (Accessed: 18 March 2024).
4. B. Prabadevi (2023) Customer churning analysis using machine learning algorithms, International Journal of Intelligent Networks. Available at: <https://www.sciencedirect.com/science/article/pii/S2666603023000143> (Accessed: 18 March 2024).
5. Shapash report Shapash Report - Shapash 2.4.1 documentation. Available at: <https://shapash.readthedocs.io/en/latest/tutorials/generate_report/tuto-shapash-report01.html> (Accessed: 18 March 2024)