# Telecommunication Customer Attrition

Ruben Brionez Jr

Bellevue University

DSC630 - Predictive Analytics

Andrew Hua

June 16, 2024

## Telecommunication Customer Attrition

The telecommunication industry is an ever growing and changing industry in the Unites States. It has become one of the most vital industries that affects people's everyday lives. The telecommunications industry brings internet, voice, and cable T.V. services to millions of residential and commercial customers across the United States. One of the challenges that this industry faces is customer attrition. With many different providers, competition and promotions, there tend to be fluctuations in the number of customers a telecommunications company provides service to.

"The churn rate, also known as the rate of attrition or customer churn, is the rate at which customers stop doing business with an entity. It is most commonly expressed as the percentage of service subscribers who discontinue their subscriptions within a given time period" (Cheng, M., & Kvilhaug, S., 2024). The underlying goal of exploring and analyzing customer attrition is finding patterns and inferring or exploring possible causes as to why customers are leaving. Having a better understanding of customer churn allows a company to update their approach to keeping customers and improving revenue and sales.

## Data Selection

For this project I will be analyzing the data and using it to predict the churn for a fictitious telecommunications company. The initial dataset for this project was sourced from Kaggle.com and is titled "WA_Fn-UseC_-Telco-Customer-Churn.csv" (BlastChar,2018). The selected data set provides the data required to analyze and predict churn for a fictitious telecommunications company. It includes 7043 rows and 21 columns, or features, for use in analyzing and predicting churn. The rows of the data set represent individual customers, while the features of the data set include descriptive data of the customer, such as age, gender, tenure, and types of service.

Although this is the initial dataset, there may be more data needed as the project progresses. The process of cleaning and engineering may reduce the dataset to an undesirably small size so as I

begin working with the data and project it may be necessary to bring in additional data. Additional data may also be required as model selection and testing begin to take place.

## Model Selection

This project will require a model to predict the customer churn for the telecommunication company. The objective after cleaning and preparing the data set for modeling is to use the data set for a linear regression model. Since the target will be a single feature, churn, a simple linear regression model should be sufficient to use a prediction model. However, multiple models may be evaluated with various features to be sure the best model is selected to represent the churn predictions.

The other model that may be considered is a variation of the linear regression model, a multiple regression model. With this model, multiple features could be evaluated together to get a better understanding of how multiple features will affect the target of the model. Both models will probably be included in the final analysis of the project for comparison.

## Result Evaluation

Evaluation of the model will vary depending on the model used. The linear regression model will be evaluated using the RMSE and R-squared methods to check the predictions against actual values. A score using the R-squared metric of 0.70 or above is generally regarded as an indication of a good linear regression model.

When reviewing the results of the analysis of customer attrition, I want to find the factors that appear to correlate to an increase in higher customer attrition rates. Once the most influential factors are discovered, exploration of the causes can begin, and some inferences can be made as to why the factors are contributing to customer attrition.

## Learning Objectives

While working through this project I have multiple learning objectives and goals that I would like to meet:

- A better understanding of how customer attrition is calculated and evaluated within the telecommunication industry.
- A thorough understanding of how machine learning models can be utilized to in the telecommunication industry to predict and analyze customer attrition
- An understanding of what to do with the insights gained from predictive models and analysis regarding customer attrition.

Having worked for ten years in the telecommunications industry, this is a great opportunity to expand on my industry knowledge and gain further insights into a part of the industry I have a had little experience with.

## Risks and Ethical Considerations

There are multiple risks associated with the project. One of the biggest risks that I currently see is the risk of a small data set. With only roughly 7000 rows of customer data and not having cleaned and prepared the data yet, I have a concern that the data set might end up too small. Generally, from my experience in the telecommunications industry, service providers have many more customers than 7000. It would be ideal to have a larger data set to work with for the project.

An ethical consideration of the project is how the analysis and predictions may be used. Will the results of the predictions and analysis impact certain demographics' ability to attain high speed internet and cable services? Will the results of the analysis give a certain group better financial deals than other groups? These issues, at a minimum, should be discussed and evaluated.

## Contingency Planning

Contingency planning is something that should always be considered in any project, whether it is an academic project or a professional project. There should always be additional options to consider when completing a project. I have come up with a few for this project:

- **Finding Additional Data:** If after cleaning and preparing the data, the original data set becomes too small. Alternative sets of data will be sourced to supplement the existing data.
- **Model Selection and Target:** If after selecting a model, the results are not desirable or do not appear to make sense for the target, the target and features may need to be adjusted. This may change the scope of the project.
- **Alternative Project:** If for some reason there are issues and barriers that can not be resolved in a timely manner, I have a backup data set and project in mind. The biggest concern in using the backup project is coming up with a viable business problem or business question to justify the data set.

**References**

Cheng, M., & Kvilhaug, S. (Eds.). (2024, March 21). *Churn rate: What it means, examples, and calculations*. Investopedia.

https://www.investopedia.com/terms/c/churnrate.asp

BlastChar. (2018). *Telco Customer Churn* (WA_Fn-UseC_-Telco-Customer-Churn.csv)[Dataset]. Kaggle.

https://www.kaggle.com/datasets/blastchar/telco-customer-churn