Using Machine Learning to predict   
Customer Churn in the Telecom Industry

Anish Mahapatra

Proposal  
MSc in Data Science

January 2021

Contents

[Abstract 4](#_Toc61709565)

[1. Aim and Objectives 5](#_Toc61709566)

[2. Introduction 6](#_Toc61709567)

[2.1 The need for Customer Churn Analysis 6](#_Toc61709568)

[3. Background 8](#_Toc61709569)

[4. Problem Statement 9](#_Toc61709570)

[5. Research Questions 9](#_Toc61709571)

[6. Aim and Objectives 9](#_Toc61709572)

[7. Significance of the study 9](#_Toc61709573)

[8. Scope of the study 9](#_Toc61709574)

[9. Research Methodology 10](#_Toc61709575)

[10. Required Resources 11](#_Toc61709576)

[10.1 Hardware Requirements 11](#_Toc61709577)

[10.2 Software Requirements 11](#_Toc61709578)

[11. Research Plan 11](#_Toc61709579)

[References 12](#_Toc61709580)

# Abstract

With the advent of increasing competition in various market segments, companies must retain customers to maximize profits. Customer retention policies can affect the annual turnover drastically depending on the rate of churn. The cost of customer churn to the Telecom industry is about $10 billion per year. Studies show that customer acquisition cost is 5-10 times higher than the price of customer retention. Companies, on average, can lose 10-30% of their customer annually. Developing processes and efficient consumer-centric policies to reduce customer churn can reduce spend on customer relations. For this, one would need to understand and track customer behaviour to understand the indicators that make a customer likely to churn.

Datasets for customer churn are quite large and saved in large data warehouses where many features are present. Not all attributes are significant for churn prediction. Hence, feature engineering requires not only computation but a substantial amount of time as well. Through this paper, we will find and the features that that will be significant for churn prediction. The aim is to predict churn accurately and showcase the variation in performance of various algorithms.

# 1. Aim and Objectives

The aim is to develop a trustworthy and interpretable model that will predict the customers that will churn from a Telecom Company based on historical customer telecom data. The identification of the customers that churn will aid telecom companies in significantly reducing expenditure on customer relations.

The objectives of the research are based on the above aim and are as follows:

* To analyze the relationship and visualize patterns of customer behaviour to be able to indicate to the telecom company if a customer is going to churn
* To suggest suitable feature engineering steps to be able to extract the most value from the data including picking the most significant features
* To find appropriate balancing techniques to enhance the model performance on the dataset
* To compare the classification or predictive models to be able to identify the most accurate model to determine the customers that will churn
* To understand and document the factors that lead to customer churn in the telecom industry
* To be able to evaluate the performance of the models to be able to identify the appropriate models

# 2. Introduction

With the increase in the number of options consumers in the telecom space have with the advent of the Digital Age, for a company to be successful, it is vital to be able to keep costs low and profits high. One of the most effective ways to do this is to be able to retain the existing customer base and focus the rest of the budget on acquiring new customers.

The retention of the existing customer base must be done in a focused and systemic manner, or else the bottom line of the company can get affected. A targeted way to approach the end goal of customer retention is to flag customers that have a high probability to churn. Based on customer behaviour and attributes, if we can flag the customers that are likely to churn, we can run targeted campaigns to retain customers.

## 2.1 The need for Customer Churn Analysis

The ability to retain customers showcases the company's ability to run the business. With the digital age now where everything is online, any business needs to understand customer sentiment virtually. The cost of customer churn in the Telecom Industry is approximately $10 billion annually [(Castanedo et al., 2014)]. Customer acquisition costs are higher than customer retention by 700%; if we were to increase customer retention rates by just 5%, profits could see an increase from 25% to even 95% [(Hadden et al., 2006)]. For a company to be profitable, it is thus essential to be able to take pre-emptive action to be able to retain customers that may churn. Churn in telecom companies is defined as the customers who stop using their specific services and plans for long periods.

As service providers contend for the rights to a customer, customers are free to choose a service- provider from an ever-increasing set of corporations depending on customer need. This increase in competition has led to customers being able expect tailor made products at a fraction of the price [(Kuo et al., 2009)]. Churned customers those customers that move from one service provider to another [(Ahmad et al., n.d.)] [(Andrews, 2019)]. Churn can be due to the non-satisfaction of current services, better offerings from other service providers and even lifestyle changes. In the customer's behaviour patterns, there is likely to be a few significant indicators as to why the customer is willing to take the active step of moving across service providers. We shall identify the attributes that can indicate churn in our methodology through this research.

# 3. Literature Survey

The current methodology employed to

# 4. Problem Statement

# 5. Aim and Objectives

gibberish

# 6. Significance of the study

Orem ummah

# 7. Scope of the study

# 8. Research Methodology

Introduction, dataset description, data preprocessing, transformation, models, evaluation metrics

# 9. Required Resources

### 9.1 Hardware Requirements

Based on the defined scope of the proposed thesis, the following are the required resources:

**NOTE:** Please ensure you have Administrator access in the machine (Windows/Ubuntu/macOS)

The minimum hardware requirements for this project are:

**RAM:** Minimum 4 GB (8 GB recommended for optimum performance)

**Disk space:** Minimum of 4GB free space needs to be allocated (Depends on the Simulation Size)

### 9.2 Software Requirements

|  |  |
| --- | --- |
| Software | Minimum Version |
| R | >= 3.6 |
| Python | >= 3.5 |
| NetLogo | >= 6.1.1 |
| R Studio | >= 1.2.0 |
| Java SDK | >= 8 |
| Excel | >= 2007 |

# 10. Research Plan

The following GANTT chart proposes the timeline for the research and implementation of the project.

Based on the complexity of the different phases, the timelines are subject to minor adjustments. Regardless, the candidate shall pledge to stick to the timeline as closely as possible.

# References

Ahmad, A.K., Jafar, A. and Aljoumaa, K., (n.d.) Customer churn prediction in telecom using machine learning in big data platform. [online] Available at: https://doi.org/10.1186/s40537-019-0191-6.

Andrews, R., (2019) Churn Prediction in Telecom Sector Using Machine Learning. *International Journal of Information Systems and Computer Sciences*, 82, pp.132–134.

Castanedo, F., Valverde, G., Zaratiegui, J. and Vazquez, A., (2014) Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network Federico. pp.1–8.

Hadden, J., Tiwari, A., Roy, R. and Ruta, D., (2006) Churn Prediction: Does Technology Matter. *International Journal of Intelligent Technology*, 1, pp.104–110.

Kuo, Y.-F., Wu, C.-M. and Deng, W.-J., (2009) The relationships among service quality, perceived value, customer satisfaction, and post-purchase intention in mobile value-added services. *Computers in Human Behavior*, 25, pp.887–896.