Using Machine Learning to predict   
Customer Attrition in the Telecom Industry

Anish Mahapatra

Proposal  
MSc in Data Science

January 2021

Contents

[Abstract 3](#_Toc61723636)

[1. Background 4](#_Toc61723637)

[1.1 The need for Customer Churn Analysis 4](#_Toc61723638)

[1.2 Flagging customers and retention policies 5](#_Toc61723639)

[3. Problem Statement 6](#_Toc61723640)

[4. Related Works 6](#_Toc61723641)

[5. Aim and Objectives 7](#_Toc61723642)

[6. Significance of the study 8](#_Toc61723643)

[7. Scope of the study 8](#_Toc61723644)

[8. Research Methodology 9](#_Toc61723645)

[9. Required Resources 10](#_Toc61723646)

[9.1 Hardware Requirements 10](#_Toc61723647)

[9.2 Software Requirements 10](#_Toc61723648)

[10. Research Plan 10](#_Toc61723649)

[References 11](#_Toc61723650)

# 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. Background

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 retain the existing customer base and focus the rest of the budget on acquiring new customers.

The retention of the existing customer base in a focused and systemic manner is to be done, or its bottom line can be 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.

## 1.1 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 behaviour and mentality. 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.

In this post-pandemic age, where virtual presence via calls and mobile data is the top priority, customers streamline their monthly expenditure. Competitors are employing strategies such as offering low prices or value-add services to get consumers to switch. After acquiring a significant customer base, the companies monetize their customer base and turn a quick profit. Companies that can identify the bracket of people that are likely to leave and run targeted campaigns to showcase more value in their current offerings at a minimal budget are the companies that will be successful in the long run.

## 1.2 Flagging customers and retention policies

As service providers contend for a customer's rights, customers are free to choose a service- provider from an ever-increasing set of corporations based on customer need. This increase in competition has led customers to 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. Companies use retention strategies [(Jahromi et al., 2014)] to be able to maximize customer lifetime value by increasing the associated tenure. For telecom companies to reduce churn, it is vital to be able to predict specific metrics such as the high-risk customers, estimated time to attrite and likelihood to churn.

The learnings from multiple such exercises have been introduced as deployable machine learning algorithms that have been iterated over and refined based on the evolving need to be able to flag consumers more accurately. The selection of techniques to employ will depend on the performance of the model on the selected dataset, be it meta-heuristic, data mining, machine learning or even deep learning techniques. 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. Problem Statement

With the customer data acquired from the telecom company, we will accurately flag the bracket of customers likely to churn. This research will help telecom companies leverage their database to predict and actively target campaigns to customers that might churn. This can be a set standard in the industry where multiple machine learning algorithms can be run on a newer dataset, model monitoring can be done, and customers can be appropriately targeted.

The recommended model's primary users will be Telecom companies that wish to reduce customer attrition within the company by leveraging what Data Science has to offer. Given that the model predicts customers that will churn accurately, this can be done with limited hardware and on a regular cadence.

# 4. Related Works

The utilization of meta-heuristic models and machine learning was done to get accurate predictions on the telecom datasets. Some of the literature focused on enhancing the data itself via coherent pre-processing, efficient feature engineering and feature selection techniques[(Ahmed and Maheswari, 2017)] while the others focused on using more complicated algorithms such as Artificial Neural Networks, Support Vector Machine and so on. With a higher focus of most papers

# 5. Aim and Objectives

The paper aims 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 the factors and behaviour that leads to customer attrition in the telecom industry
* To be able to evaluate the performance of the models to be able to identify the appropriate models

# 6. Significance of the study

# 7. Scope of the study

# 8. Research Methodology

Introduction, dataset description, data pre-processing, transformation, models, evaluation metrics

# 9. Required Resources

Following are the required hardware and software requirements to be able to successfully and smoothly run the models.

### 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 8 GB (16 GB recommended for optimum performance)

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

### 9.2 Software Requirements

|  |  |
| --- | --- |
| Software | Minimum Version |
| Python | >= 3.5 |
| Jupyter Notebook | >= 6.0 |
| 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.

Ahmed, A.A. and Maheswari, D., (2017) Methods For Customer Retention In Telecom Industries. *2017 International Conference on Advanced Computing and Communication Systems*.

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.

Jahromi, A.T., Stakhovych, S. and Ewing, M., (2014) Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, [online] 437, pp.1258–1268. Available at: https://research.monash.edu/en/publications/managing-b2b-customer-churn-retention-and-profitability [Accessed 16 Jan. 2021].

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.