

Predicting Customer Churn in Telecom using Machine Learning Techniques

Request for Proposal

Anish Mahapatra
MSc in Data Science

Contents

| | |
|---|---|
| Abstract..... | 3 |
| 1. Aim and Objectives | 4 |
| 2. Introduction | 5 |
| 3. Background and Related Research..... | 5 |
| 3. Problem Statement..... | 7 |
| 4. Research Questions | 7 |
| 5. Aim and Objectives | 7 |
| 6. Significance of the study | 7 |
| 7. Scope of the study | 7 |
| 8. Research Methodology | 8 |
| 10. Required Resources | 9 |
| 10.1 Hardware Requirements | 9 |
| 10.2 Software Requirements | 9 |
| 11. Research Plan | 9 |
| References | 9 |

Abstract

With the advent of increasing competition in various segments of the market, companies must be to retain customers to maximize profits. Customer retention policies can affect the annual turnover drastically depending on the rate of churn. In fact, the cost of customer churn to the Telecom industry is approximated to be around \$10 billion per year. Studies show that the cost of customer acquisition is 5-10 times higher than the cost of customer retention. Companies on average can lose 10-30% of their customer annually. This leads to the understanding that 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 be able to understand what the indicators are that makes a customer likely to churn.

Datasets for customer churn are quite large and are stored in large data warehouses where many features can be found. Not all the features are significant for churn prediction. Hence, feature engineering requires not only computation but a great 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 be able to predict the churn accurately and show the performance difference of various algorithms.

1. Aim and Objectives

The main aim of the research paper is to develop a model to predict the customers that are likely to churn from a Telecom Company based on historical customer 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 an appropriate balancing technique that can be used on the imbalanced dataset
- To compare the predictive models to identify the most accurate model to be able to classify 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

What is the need? Background of the domain, relevance to previous research, problem(s), purpose of the study

The ability to retain customers showcases the company's ability to run the business. With the digital age now where everything is online, it is important to any business to be able to understand customer sentiment virtually. The cost of customer churn in the Telecom Industry is an approximated \$10 billion annually [1].

3. Background and Related Research

4. Problem Statement

5. Research Questions

6. Aim and Objectives

7. Significance of the study

8. Scope of the study

9. Research Methodology

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

10. Required Resources

10.1 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)

10.2 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 |

11. 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

1. 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.