Prediction of Customer Attrition in the   
Telecom Industry using Machine Learning

**Anish Mahapatra**  
Student ID 944563

Under the supervision of   
**Karthick Kaliannan Neelamohan**  
  
  
Mid-Thesis Report  
**Master of Science in Data Science**  
Liverpool John Moores University

MARCH 2021

Table of Contents

[Dedication 4](#_Toc64198023)

[Acknowledgements 5](#_Toc64198024)

[Abstract 6](#_Toc64198025)

[List of Tables 7](#_Toc64198026)

[List of Figures 7](#_Toc64198027)

[List of Abbreviations 7](#_Toc64198028)

[Chapter - 1 Introduction 8](#_Toc64198029)

[1.1 Background of the Study 8](#_Toc64198030)

[1.2 Problem Statement 8](#_Toc64198031)

[1.3 Aim and Objectives 8](#_Toc64198032)

[1.4 Research Questions 8](#_Toc64198033)

[1.5 Scope of Study 8](#_Toc64198034)

[1.6 Significance of Study 8](#_Toc64198035)

[1.7 Structure of Study 8](#_Toc64198036)

[Chapter - 2 Literature Review 9](#_Toc64198037)

[2.1 Introduction 9](#_Toc64198038)

[2.2 Data Analytics in the Telecom Industry 9](#_Toc64198039)

[2.3 Customer Attrition Analytics 9](#_Toc64198040)

[2.4 Predictive Modeling in Telecom Churn on different datasets 9](#_Toc64198041)

[2.5 Visual Analytics in the Telecom Industry 9](#_Toc64198042)

[2.6 Related Research Publications 9](#_Toc64198043)

[2.7 Discussion 9](#_Toc64198044)

[2.8 Summary 9](#_Toc64198045)

[Chapter – 3 Research Methodology 10](#_Toc64198046)

[3.1 Introduction 10](#_Toc64198047)

[3.2 Research Methodology 10](#_Toc64198048)

[3.2.1 Data Selection 10](#_Toc64198049)

[3.2.2 Data Pre-Processing 10](#_Toc64198050)

[3.2.3 Data Transformation 10](#_Toc64198051)

[3.2.4 Data Visualization 10](#_Toc64198052)

[3.2.5 Class Balancing 10](#_Toc64198053)

[3.2.6 Data Modelling 10](#_Toc64198054)

[3.2.7 Interpretation/Evaluation 10](#_Toc64198055)

[3.3 Proposed Model 10](#_Toc64198056)

[3.4 Summary 10](#_Toc64198057)

[References 11](#_Toc64198058)

[Appendix A – Research Plan 11](#_Toc64198059)

[Appendix B – Research Proposal 11](#_Toc64198060)

[Appendix C – Ethics Forms 11](#_Toc64198061)

# Dedication

This dissertation is dedicated to my family, whose unyielding love, support and encouragement have inspired me to pursue and complete this research.

# Acknowledgements

# 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 globally. 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 can help 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 is saved in large data warehouses where many features are present. Not all attributes are significant for churn predictions. Hence, feature engineering requires not only excessive computation but a substantial amount of time as well.

Through this research, we intend to find the model that can predict churn most accurately and the behaviour patterns that can indicate customer churn. The aim is to predict churn accurately and showcase the variation in performance of various algorithms. The dataset to be used for this research paper is the IBM Watson Dataset on customer churn in the Telecom industry.

# List of Tables

# List of Figures

# List of Abbreviations

|  |  |
| --- | --- |
| EDA | Exploratory Data Analysis |
| SVM | Support Vector Machine |
| KNN | K Nearest Neighbour |
| AUC | Area under ROC Curve |
| ROC | Receiver Operating Characteristics |
| XGBoost | Extreme Gradient Boosting |
| AdaBoost | Adaptive Boosting |

# Chapter - 1 Introduction

With the increase in the number of options consumers have in the Digital Age, for a company to be successful, it is vital 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 remaining budget on the acquisition of new customers.

# 1.1 Background of the Study

# 1.2 Problem Statement

# 1.3 Aim and Objectives

# 1.4 Research Questions

# 1.5 Scope of Study

# 1.6 Significance of Study

# 1.7 Structure of Study

# Chapter - 2 Literature Review

# 2.1 Introduction

# 2.2 Data Analytics in the Telecom Industry

# 2.3 Customer Attrition Analytics

# 2.4 Predictive Modeling in Telecom Churn on different datasets

# 2.5 Visual Analytics in the Telecom Industry

# 2.6 Related Research Publications

# 2.7 Discussion

# 2.8 Summary

# Chapter – 3 Research Methodology

# 3.1 Introduction

# 3.2 Research Methodology

## 3.2.1 Data Selection

## 3.2.2 Data Pre-Processing

## 3.2.3 Data Transformation

## 3.2.4 Data Visualization

## 3.2.5 Class Balancing

## 3.2.6 Data Modelling

## 3.2.7 Interpretation/Evaluation

# 3.3 Proposed Model

# 3.4 Summary

# References

# Appendix A – Research Plan

# Appendix B – Research Proposal

# Appendix C – Ethics Forms

## 