# TELECOM CUSTOMERS CHURN PREDICTION SYSTEM

By

Group KWKB(initials of group members)

DEPARTMENT OF NETWORK
SCHOOL OF COMPUTING AND INFORMATICS TECHNOLOGY

A Concept Paper submitted to the School of Computing and Informatics Technology

Supervisor

Mr. Grace Kamulegeya
Department of Networks
School of Computing and Informatics Technology, Makerere University
gkamuleg@cit.ac.ug, +256-41-540628, Fax: +256-41-540620

July 2019

## GROUP MEMBERSHIP:

| # | Names | Registration Number | Signature |
|---|---|---|---|
| 1 | KYANZI HASSAN MUSISI | 17/U/19979/EVE | |
| 2 | WAMALA EDGAR WATSON | 17/U/20633/EVE | |
| 3 | KASOZI JAFFAR | 17/U/4741/EVE | |
| 4 | BWAYO JOEL | 17/U/3777/EVE | |

## 1.0 Introduction

Nowadays as we talk about subscription to services offered by telecom companies, Customer retention is the key to look at. **Customer churn (or customer attrition)** is a tendency of customers to abandon a brand and stop being a paying client of a particular business. The percentage of customers that discontinue using a company's products or services during a particular time period is called a *customer churn (attrition) rate.*

## 2.0 Background to the Problem

Churn is one of the largest problems facing most businesses. Harvard Business review, it costs between 5 times and 25 times as much to find a new customer than to retain an existing one. In other words, your existing customers are worth their weight in gold! And so predicting so that you can avoid customer churn is an important business function.

We have used a data set called Telco customer churn from Kaggle.com that included 7,033 unique customer records for a telecom company called Telco. Each entry has information about the customer, which included features such as:

1. Services—which services the customer subscribed to (internet, phone, cable, and extra.)
2. Tenure—How long they had been a customer
3. Basic demographic info—whether they were elderly, had dependents, extra.

For the purposes of our system, the dependent variable is whether or not the customer will "churn" or not within the next month. In fact, a full 27% of customers will be labeled as will have to leave the company within the next month. With rates of attrition this high, it would only be a matter of months before the company loses most of its customers—if they don't adapt our customer churn prediction system.

Assumptions

For the purpose of our system, we have made two assumptions:

- For each customer that leaves the company, it will cost Telco $500 to replace that customer. Marketing, ads, campaigns, and outreach thus the cost adds up.
- We can help Telco retain each customer who is likely to churn by investing $100 in them. This could be through discounts, improving service (upping their internet speed, for example), or offering perks.

## 3.0 Problem Statement

Customer churn is a major problem and one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecom field we have chosen, it is seeking to develop means to predict potential customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn.

The main contribution of our work is to develop a churn prediction model which assists telecom operators to predict customers who are most likely subject to churn. The model developed in this work uses machine learning techniques on big data platform and builds a new way of features' engineering and selection.

# 4.0 Objectives

1. **Main Objective**

The main objective of our system is to develop a churn prediction model which will assist telco company operators to predict customers who are most likely subject to churn.

2. **Other Objective**

   Another objective is to use customer social network in the prediction model by extracting Social Network Analysis (SNA) features. The use of SNA will enhance the performance of the model from 84 to 93.3% against Area Under Curve(AUC) standard.

## 5.0 Methodology

For the model, we are to develop, it will look first at several different machine learning algorithms to see which ones to move forward with. Our first step will be data loading followed by data wrangling, data visualization, feature engineering that will involve breaking down data into training and test sets using train-test-split, which will allow us to cross-validate our results later. We shall also stratify the train-test-split,to ensure that the same proportion of our target variable is found in both our training and test sets.

We shall use also some minority oversampling in order to balance our data set. Since only ~27% of the records in our csv file were marked as "churned", feeding our data into our algorithm without oversampling will lead it to under classifying our target variable.

We shall also use imblearn package's SMOTE to bring our minority class up to 50% of our dataset and when our data is balanced, we will then perform yet another train-test-split—this time just on our training data. The reason for doing it this way will be avoiding violation of the cardinal rule of cross-validation—basing decisions off of the results that test data provided.

After all this glorious data munging, we shall plot a ROC curve to compare how each algorithm will do on identifying true positives (sensitivity) vs. false positives (specificity).

Knowing this, we shall then use Sci-Kit Learn's GridSearchCV function, which allows us to tune our model. We shall set recall as the scoring metric to optimize on, and then use combinations of different hyper parameters to find the model with the best fit. Our goal is to squeeze out every last ounce of recall we could out of our models, and nothing less!

GridSearchCV also includes a handy cross-validation function (that's what the CV stands for!), so we shall perform Stratified K-Folds cross-validation on every model's pass through new parameters. Needless to say, we were being quite thorough in our attempts to avoid over fitting our models.

Finally, a big part of our analysis will have to do with creating a "net dollars saved" function that would determine how much money we spent on retaining customers, versus how much we shall save by not having to replace them. This, along with recall, will make up the decision criteria upon which we will judge whether we will have a successful model.

## 6.0 Outcomes

After tuning our models, it will come down to 3 models that are neck-and-neck. Our final step will be to adjust the probability threshold for each model (between i =.01 and i = 1).This will allow us to optimize our “net dollars saved” function. Essentially, for each probability threshold. We are asking our model to predict whether the customer would churn or not—even very low and very high i-values. Of course, as approached 0, our models will essentially predict that will churn—and conversely, as it approaches 1, that no one would churn.

## 7.0 References


[1],**Mark Peterson** Senior Data Scientist at Alliance Data,Predicting customer churn.Accessed on: July10, 2019. [Online]. Available:https://www.datacamp.com/courses/predicting-customer-churn-in-python

[2],**Altex soft** Customer Analyst and machine learning expert,Customer churn prediction using machine learning main approaches and models. Accessed on: July11, 2019. [Online]. Available:https://www.kdnuggets.com/2019/05/churn-prediction-machine-learning.html

[3],**Colleen Farelly**,Data Scientist/Poet/Social Scientist/Topologist. Data pipelining methods. Accessed on: July11, 2019. [Online]. Available:https://www.quora.com/What-is-a-data-science-pipeline