

# **CPE213 Data Models Project Proposal**

## **Telco Customer Churn**



### **Team Member:**

Mr. Chanon Khanijoh	63070503408
Mr. Tunwa Satianrapapong	63070503419
Mr. Napas Vinitnantharat	63070603422
Mr. Pechdanai Saepong	63070503434
Ms. Fasai Sae-Tae	63070503436

Department of Computer Engineering  
King Mongkut's University of Technology Thonburi

## **Title: Telco Customer Churn**

### **Abstract**

The objective of this work is to predict the churn of customers from the selected dataset which is the Telecom customer churn by using the different classification models. This research has been focused on how the data can be used to analyze the churn rate and the customers' attributes, who are likely to churn.

The dataset is chosen from Kaggle called 'Telco Customer Churn', which is the telecommunication that provides both home phone and internet service about 7043 customers. Each customer has attributes such as Tenure, MultipleLines, DeviceProtection, etc. Those attributes can be used in Exploratory Data Analytics and Data Visualization.

In this research, random forest, decision tree, and naive bayes have been used to classify customer churn.

## Introduction

In terms of the business telecommunication industry, one of the important problems is to understand the behaviour of each customer, then try to predict who is going to leave the company service or going to discontinue the subscription in the other words is to minimize the customer loss rate, which is known as “Churn”, since the churn causes the company to drop in revenues. The churn can be motivated by different factors such as the company selling the wrong product type to the customer, the customer didn't achieve their desired outcome, wrong price structuring, etc. So, the telecommunication service provider can offer the new subscription or package to the customers, who are likely to churn, to prevent the customer loss rate.

In this project, the telco customer churn dataset has been used to understand the behavior of churn customers and try to predict who is going to be the next churn customer. The Telco customer churn data contains information about a telco company that provided home phone and Internet services to 7043 customers in California in 2019. In this research, Random forest, decision tree, and naive bayes have been used to classify the customer churn.

## Data source description

Data source URL: <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

Data advanced description:

<https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113>

The data has been chosen from Kaggle called 'Telco Customer Churn' which is the IBM Samples Data Sets. Each row represents a customer, each column contains the customer's attributes described in the column Metadata. The raw data contains 7043 rows (customers) and 21 columns (features). The 'Churn' column is the target to predict.

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

## Problem statement

One of the main problems that most telecom companies' service providers are facing is the problem of the ability of clients to switch their services to other telecom companies. As a result, churning is the problem that needs to be solved. To solve this problem, the churn rate can be minimized by analyzing the past behavior of customers that churned using different data analytic techniques. This helps the company to improve its revenue and reduce the rate of customer loss.

## **Proposed analytic technique**

1. Understand business needed
2. Exploratory Data Analysis (EDA)
  - Data Visualization
  - Cross-tabulation
3. Data preparation (Depend on dataset)
  - Data Cleaning
  - Manage the categorical data
  - Feature Engineering
  - Split the training set and testing set
4. Classification
  - Random Forest
  - Decision Tree
  - Naive Bayes
5. Evaluation of each classification model
  - Accuracy score
  - F1 score
  - Precision score
  - Recall score
  - Confusion matrix
  - P-value