AIM 5005 Machine Learning

**Project Proposal**

Due date: by 11:59 PM on Thursday, July 15, 2021

**Title: Telco** **Customer Churn Analysis and Prediction**

Name**: Xiaolan Li \_\_\_**

**1. Introduction**

* 1. Problems

- Specify the problems that you would like to solve from this project. i.e., Specify why you do this project

Problems:

**- Q1: Which services should company improve to avoid current customer leaving?**

**- Q2: What kind of services are more related to whether a customer leave?**

**- Q3: Who are more likely to leave and what the confidence of that predictions are?**

Reason**:**

**In this project, I’m going to explore the reason behind customer churn and predict who’s about to leave currently. This can help the company to maximize the customer experience for preventing customer attrition. Therefore, it can help companies reduce losses and even increase the number of customers to increase profits. Additionally, this can be applied in every company, with a high impact value.**

* 1. Related (prior) works

- Review the literatures

- Describe the current status or previous relevant works done by others

- Describe how your project will be different from the prior works

The paper “A hybrid classification model for churn prediction based on customer clustering” used a hybrid classification combined with multiple layer perceptron, k-means and gradient boosting decision tree to lower the dimension, seperate customer into 4 dataset and then predict the result.

I will use the PCA to do the dimension deduction then use the SVM, KNN, random forest to predict the result. If the result is not good, I will also add a cluster in the middle.

* 1. Possible outcomes of this project

- Final products (outcomes): Describe what you will produce when you finished this project

1. **Finding out the best model to predict if a customer is going to leave.**
2. **Finding out the features that lead to a customer leave and do an order of them.**
3. **Finding out the most important reason lead to current customer leave.**

**2. Technical Plans**

* 1. Dataset that you will use

**Downloaded from IBM community. Publish date May,2020.**

<https://community.ibm.com/accelerators/?context=analytics&query=telco%20churn&type=Data>

The data set is detailed in

[https://community.ibm.com/community/user/businessanalytics/blogs/stev en-macko/2019/07/11/telco-customer-churn-1113](https://community.ibm.com/community/user/businessanalytics/blogs/stev	en-macko/2019/07/11/telco-customer-churn-1113 )

**The data set includes information about:**

**-- 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**

* 1. Describe about learning methods that you will use for this project (if any)

**Logistic Regression**

**KNN model**

**Random Forest**

**PCA**

**SVM**