## **DATS 6103 Final Project**

Jeongmin An, Michael Ehart, Nikhil Obuleni, and Diluni Peiris

## **Part 1: Topic Proposal**

1. *The research topic your team comes up with*

Predicting customer churn at a California telecommunications company.

1. *The SMART question(s) of your research*

What factors influence customers’ churn rate? Specifically:

* Are customers with longer-term contracts (one- or two-year) less likely to churn compared to those with month-to-month contracts?
* Are particular products (phone, cable internet, fiber-optic internet, streaming) more associated with churn or retention?
* What demographics (age groups, genders, etc.), if any, are more likely to churn?
* Does the churn rate show geographic patterns?

How can the company use this information to improve customer retention?

1. *The source of your data set(s)*

The dataset was created for the Maven Churn Challenge in 2022:

<http://mavenanalytics.io/blog/maven-churn-challenge>

We extracted the data from Kaggle:

<https://www.kaggle.com/datasets/shilongzhuang/telecom-customer-churn-by-maven-analytics/data>

1. *The link to your team’s GitHub repo*

<https://github.com/nik-hill-323/Data_mining_Final_project.git>

1. *The modeling methods you propose to use*

* A range of classification machine learning techniques, including logistic regression, decision trees, and ensemble techniques such as random forest and boosting.
* A clustering model to understand different groups of customer profiles of the company.