ADMN5016: Applied Artificial Intelligence and Machine Learning

Maria Isabel Dio Grau

Project Proposal

**Dataset**: Telecom Churn

**Link**: <https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets>

**Description**: This dataset includes account and activity information from over 3,000 customers of Orange Telecom, including state, area code, mobile plan, total day calls, etc. The target variable is whether or not these customers churned or terminated their telecom subscription.

**Machine learning algorithm**: Classification

**Business case**: Mobile company ADMN has built a wide customer base over the past few years. The company now has enough data to analyze in order to maintain good relationships with its customers since it costs less to retain an existing customer than to acquire a new customer. Using classification algorithms, this project aims to predict whether or not a customer will churn. If a customer is predicted to churn, the company can take corrective actions to improve its services, such as increasing customer support staff or offering mobile plan promotions. This will also make the company more competitive against other mobile plan providers because of its proactive approach in providing quality service.

**Other information**: 86% of customers in the dataset did not churn and only 14% churned. To address the imbalance, bootstrap aggregation or bagging will be used to avoid overfitting.

**Backup dataset**: Bank Churn

**Link**: <https://www.kaggle.com/datasets/shrutimechlearn/churn-modelling>

**Background**: This project proposal has the same concept as the Telecom Churn dataset, where the goal is to predict whether or not a customer will churn or leave a bank using classification algorithms. The dataset includes bank account information from 10,000 customers in France, Germany, and Spain. 80% of customers in the dataset did not churn, while 20% left the bank. Bagging will be used to avoid overfitting.

*\*Since the algorithms and datasets are similar, my final project will depend on which one provides a stronger business case (e.g., cost savings).*