**CIS 9660 Data Mining**

**Qilan Lao**

**Project 1 Proposal Classification**

Airline Passenger Satisfaction

Due to the impact of coronavirus disease in 2019, the business volume of airlines around the world has been reduced. Despite the strength of the North American aviation industry, to maintain sustained growth and a sustained position as a cross regional industry leader, the company must always be vigilant to keep up with customer demand. Passengers' satisfaction with their air travel is a crucial metric for airlines, so airlines must understand what matters to their customers. Customers' evaluations of flight experiences are influenced by a variety of elements such as seat comfort, ease of online reservation, legroom, departure and arrival delay times, and so on. Therefore, my purpose of this project is to establish the classification model to predict passengers' evaluation of flight experience and determine of what are the important factors affecting their satisfaction. This project has the potential to really help airlines understand their advantages and enhance their business from a customer's standpoint. The dataset from Kaggle supplies us with a vast amount of data (over 100,000 rows) and several prediction variables (over 20). In this project, I'll clean up the data, do some exploratory data analysis, and create six prediction models (logistic regression, naive bayes, decision trees, random forests, support vector machines, and K-Nearest Neighbors) to compare their performance to determine the optimal model.

Dataset link:

Airline Passenger Satisfaction | Kaggle <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>