**M.S. in Applied Data Science Project Portfolio**

**Milestone Requirement**

**Ian Ustanik**

**Customer Churn in the Airline Industry**

**Due Date: 4/20/2021 Date Submitted: 4/20/2021**

**File(s):** Algorithms\_updated.ows, DataMiningAirlines.pdf, Mitigate Missing Data.ipynb, Random Forest.ipynb, SVM & Neural Network.ipynb, The Brain Trust – Final Project.pptx, Vizualization and Text Analysis\_updated.ows

**Tool(s) Required For Viewing:** Anaconda – Orange for .ows files, Adobe Reader for .pdf files, Anaconda – Jupyter Notebook for .ipynb files, Microsoft PowerPoint for .pptx files

**Further Details:** This dataset contained the demographics and additional information for the passengers who had travelled with FlyFast, Cheapseats and Sigma Airlines for individual flights in the first quarter of 2014. The aim of the project was to analyze the reasons for the customer satisfaction ratings (NPS) achieved by the three airlines and suggest data driven insights to improve the ratings for FlyFast Airways, converting detractors and passives into promoters. After cleaning the dataset, unsupervised forms of machine learning such as Association Rule Mining and K-means Clustering Analysis were applied to understand interesting patterns in the data. The feedback provided by the customers had been analyzed by creating word clouds and sentence level sentiment analysis using the Liu Hu test to better understand and get to the root of the issues. Machine learning algorithms such as Neural Networks, Random Forests, and Support Vector Machines were applied to create supervised learning models used to predict and classify whether a passenger was likely to be a promoter, detractor or passive for various segments when exposed to new data.