IST 687 M004 Group 5

Milestone 2

**What we have done**

1. Get a summary of dataset, split it into subsets by airline partners
2. For each partner airline subset, we get the following data:

histogram of likelihood.to.recommend, number of observations, and descriptive statistics, and calculated NPS

1. Transformed categorical data into factors and then conduct a linear regression between likelihood.to.recommend and all other fields
2. Divide entire dataset into train set and test set
3. Cross validation on linear regression model
4. We create new categorical fields for
   1. type of customers(detractors, passive, or promoters)
   2. age group()
   3. arrival delay more than 5 minutes
   4. long duration trip

To find association rules between type of customers and other attributes

1. Create visualizations with likelihood.to.recommend vs other variables to see its relation individually, and eliminate unnecessary variables
2. Create correlation plot taking all the variables and selecting relevant attributes based on the plot.

**Our next step**

1. Support Vector Machine with type of customers
2. Try with 3-4 different models on our data and find the most efficient way

**Questions**

1. Which variables are not relevant in our prediction model
2. Determine whether text mining on free text will provide valuable insight
3. How to interpret the models regarding increasing our customer satisfaction

**How is team working**

Everyone is awesome!