**Group Proposal**

1. What problem did you select and why did you select it? What database/dataset will you use? Does it need to be cleaned?

We select Elo Merchant Category Recommendation. Our question is help understand customer loyalty

Database is: kaggle competitions download -c elo-merchant-category-recommendation.

This dataset need to be clean. Add some new features, such as ‘purchase month’.

1. What data mining algorithm will you use? Will it be a standard form, or will you have to customize it?

We will use decision tree, random forest, support vector machine, K nearest neighbor- Navie Bayes, clustering, K-means, hierarchical.

1. What softwares will you use to implement the network? Why?

Github and pycharm.

Use Github to collaborate.

Pycharm to do analysis.

1. What reference materials will you use to obtain sufficient background on applying the chosen network to the specific problem that you selected

We will use following materials:

An Introduction to Statistical Learning - Author: G. Casella- Free.

The Elements of Statistical Learning Data Mining - Author: Trevor Hastie.

Data Preprocessing in Data Mining - - Author: Garca, Salvador.

1. How will you judge the performance of your results? What metrics will you use? Provide a rough schedule for completing the project.

We will use accuracy to judge the performance of your results. We will use confusion metrics.

Schedule:

10/31 – 11/6 Data cleaning and preprocessing

11/7 – 11/14 Try decision tree and random forest

11/14 – 11/21 Try support vector machine and naïve bayes

11/21 – 11/28 Try clustering, K-means, and hierarchical

11/28 – 12/4 Judging the performance of the models and draw conclusion