STAT 517 Final Project Critique

Vignesh J Muralidharan: “Speed Dating Dataset”

Reviewed By Rohit Kumar Yadav

**Coverage of 3 course areas – Supervised/Cluster/Association:** All the three categories are explored and tested. Accuracy is tested with the models listed KNearest Neighbors, Naïve Bayes, Logistic Regression, Decision Tree, Random Forest, Gradient Boosting Classifier, SVM and Neural Networks. He did not used Regression because It will not work in this target because the variable is not continuous.

**Dataset size meets stated criterion:** The dataset used has over 8378 (rows) of data across 123 variables for a column which is total of 1,030,494. Well over the suggested 100,000 minimum requirement.

**Relevancy:** The dataset is relevant and pretty much interesting as dating become so popular nowadays in people life for choosing best partners in their life. As this dataset is an experimental gathered dataset from speed dating events.

**Difficulty:** I had difficulty in understanding the clusters where more number of clusters did not worked well to see the distinct group of clusters than less amount of clusters.

**Interestingness:** I thought the approach of match or not match based on the decisions taken in the event are interesting and his results in association rule with partner side preference is interesting where people gave importance to the race.

**Clarity:** Vignesh did a good job in explaining whole project with such a messy dataset I feel with lots of redundant features but he covered most of the models very well and explained very clearly except the clustering which I felt bit fuzzy which perhaps needs more observation and explanation.

**Originality:** I think the dataset is pretty unique and interesting to test all the models in which Vignesh did good job choosing such an interesting dataset to test all the models more precisely.

**Creativity:** I liked his creativity Associstion rule generation which is done based on the personal and partner reference and then he done with the specific interest attribute to get the match of the partners based on the specific attributes