**Project Title**: Types of Professors at UT Arlington

**Objective and Overview**: The objective is to collect rating information (including star ratings and comments) about professors at UT Arlington and put them into distinct groups based on their ratings and performance at the university. The professors will be sorted into categories such as “Interesting”, “Easy” or “Unknowledgeable”. Professors will be able to belong to more than one category

**Significance**: This project will help better understand the kind of professors found at UT Arlington and identify the common complaints made by students.

**Datasets to be collected**:

UT Arlington Student Feedback Survey

Instructor ratings from RateMyProfessor.com

**Data Mining Tasks To Be Performed**

1. Downloading appropriate webpages from the websites
2. Mining the web pages downloaded for rating information
3. Cluster the data based on its attributes
4. Calculate the similarity between members of the same cluster and dissimilarities between members of different clusters.

**End of Semester Deliverables**:

At the end of the semester, the data to be delivered will be:

1. A description of the various groups that professors can be divided into.
2. Graphs showing the distribution of professors into the various divisions.
3. A scatter plot graph showing the distance between individual professors on separate attributes.

This same data will be put on the project website and presented before the class in a demo.

**Project Challenges**:

The major challenge in the project will be getting the information from the webpages. The UT Arlington SFS dataset is a simple table, but the RateMyProfessor data is more complex and will require more time to accurately parse the download. Compared to this, clustering the data and dividing the professors into classes will be easy.

**Addressing the Challenge**:

This challenge will be best addressed by starting as soon as possible on the project and getting information on parsing webpages. Consequently, the downloading and parsing should be done by the progress report deadline on April 2nd.

**Design and Implementation**:

As stated previously, data from the two datasets will be gotten from their respective websites. Afterwards, the professors will be clustered based on each different attribute e.g. “Easiness” in order to put them into groups. Keywords will also be selected from the comments connected with each rating on ratemyprofessor.com and used as categories in which to cluster professors.

**Evaluating the Solution**:

Evaluating the solution will be done by comparing the similarity and dissimilarity of professors inside and between clusters. Higher similarity inside clusters and higher dissimilarity between clusters will indicate a good clustering.