**MIS 545 Project Description**

**Introduction**

This project combines the design, preparation and evaluation steps necessary for a successful data mining task. You will need to choose and defend an interesting project that is of interest to your group but also to the community (*e.g.*, industry, academia, and healthcare *etc.*). The project should require the use of data mining techniques and a large dataset (N>20,000 after preprocessing).

Your team should comprise 6 to 7 people.

**Deliverables**

There will be three stages and correspondent deliverables in the context of this project.

**Stage 1, team work (10% of grade):** This stage will require the team to choose a problem, create a relevant dataset, suggest a data mining approach, and demonstrated on a subset of the data how the chosen approach will work. The team will build pilot data to demonstrate that the approach is appropriate or, if the results of the pilot study show the results will not be good, an alternative plan needs to be proposed. Assignment details will be posted separately.

Slides, data and text from this stage can be used in the final project report.

**Stage 2, individual work (10 % of grade):** In this stage, every member needs to complete an individual job that tests a different data (sub)set, or a different algorithm, or a different approach. The goal of this assignment is to ensure everyone works through a data mining cycle individually. Use this assignment to test something new! When one or more team members find interesting results, these can be leveraged for the next stage of the project. Assignment details will be posted separately.

Slides, data and text from this stage can be used in the final project report.

**Stage 3, team work (20 % of grade):** This final stage will require the team to complete the knowledge discovery processes: data preprocessing, data mining, and evaluation. Results from the pilot study and individual studies need to be taken into account. Assignment details will be posted separately.

**Presentations**

There will be two presentations given by each team in the class. One for stage 1 and one for stage 3. The class presentation should highlight those aspects of your solution likely to be of greatest interest to the class. The presentation should be focused on the particular stage being presented. This is also a good opportunity to get feedback on approach, evaluation, and interpretation. You are encouraged to share expertise with other teams.