**Objective**

The goal of the term project is to gain practical experience with a particular data mining task. Example tasks include frequent pattern analysis, association analysis, classification, clustering, outlier detection, and data warehouse/data cube design. You are free to choose any method or set of methods that are covered in class or any specialized method that applies to your particular problem

**Overview**

**At a high level, you are expected to( 1) select or produce a dataset that you are interested in exploring/mining; (2) formulate a data mining question (e.g. what are the groupings (clustering) in the dataset and how can we generate classes to train a classifier?); (3) conduct a literature survey of some of the most recent and successful solutions to the problem; (4) design a program or use an existing tool such as R, RapidMiner, KNIME, or Weka to answer your data mining question; (5) evaluate your analysis using an accepted validation method; and (6) report your findings in a written report and presentation.**

**Deliverables**

Project Proposal: The goal of the project proposal is for me to give you early feedback on your project plans and to determine whether the scope of your project is appropriate. The project proposal should include the following:

* A description of the problem area you which to investigate
* A preliminary list of papers that you plan to survey as part of your literature review.
* A description of the purpose of your experiments. What are you testing?
* A description of the dataset you will use to conduct your experiments. If you are using an existing dataset (recommended), provide a short description and a hyperlink so that I can review it.
* A description of the data mining techniques that you plan to use.
* A description of any risks associated with your proposed plan.

Project Report (30%): Your project report is the main deliverable and should generally follow the format of a template as a conference proceedings such as the following:

ACM (<http://www.acm.org/sigs/publications/proceedings-templates>)

IEEE (http://www.ieee.org/conferences\_events/conferences/publishing/templates.html)

The report should consist of the following sections:

* Introduction: What data mining problem are you investigating and why is it so important or useful? What is the purpose of your experiment? What are you testing? Why are your experiments interesting from a data mining perspective?
* Related Work: How have others attempted to solve the problem you are addressing? How does your chosen approach compare to these methods? Use ACM or IEEE guidelines for citations.
* Approach: What are the inner workings of your approach to the problem?
* Evaluation Methodology: How will you measure the performance of your data mining approach?
* Experimental Results: What results did you find?
* Discussion: Pick a few of the most meaningful or puzzling results from the experimental results section and try to determine why they happened
* Conclusion: What did you find and why is it important?

Project Presentation (10%): Your presentation should be between 10 - 15 minutes long. This is not a lot of time so you should only include a good introduction/motivation and the highlights of your methods and results. The motivation is very important part when presenting your research to an audience. You need to build a case for why is your presentation important.

**Important Dates**

3/2/2016 Project Proposal Deadline

5/11/2016 Final Report Submission Deadline

5/11/2016 Final Presentations