

Instructor: I. Kotsireas e-mail ikotsire@wlu.ca

Project due date: **December 4, 2016**

1 Introduction

The objective of this project is to implement two Data Mining algorithms and test your implementations with various reasonably large datasets. The term *reasonably large* is to be interpreted as anywhere from a few hundreds to a few thousands instances (tuples).

2 Algorithms

You need to implement two out of the three Data Mining algorithms seen in class:

1. the ID3 algorithm, based on entropy/information gain, to build decision trees;
2. the Apriori algorithm for mining association rules;
3. the k-Means clustering algorithm;

3 Datasets

You need to test your implementations for each Data Mining algorithm with two datasets:

1. one dataset provided by the instructor;
2. one (reasonably large) dataset that you will find on their own.
A large selection of datasets is available for instance at the Machine Learning Repository <http://archive.ics.uci.edu/ml/>

4 Data Mining Resources

This is a selection of some useful on-line Data Mining resources:

- A portal for KDD/DM tools, news etc. <http://www.kdnuggets.com>
- Data Mining tutorials and other goodies: <http://www.the-data-mine.com/>
- ACM Special Interest Group (SIG) on KDD <http://www.sigkdd.org/>

5 Submission of your work

By the project due date, you need to to submit the following items:

1. The `cover sheet`, (Names/Student IDs of group members, Course Number, Date)
2. The `design document`, which should include:
 - (a) description of what features work and what features don't.
 - (b) description and justification of the design choices made.
 - (c) description of data structures used.
 - (d) instructions on how to compile and run the code.
3. The `test document` which should show the results of running (with your implementation) each algorithm on the two datasets, as specified above. Include a copy of the dataset you found on your own.
4. The `code` in printed version as well as in electronic form. Well-commented code will count for more marks than code without comments.