Individual Project 1:

Weblog Mining

Due Date: **21 December, 2019**

Marks: 20%

Objectives

- Learn to build a data mining application using a tool, e.g. PASW Modeler, Weka 3, or SciKit , or by programming using a familiar programming language, e.g., Python/Java/Visual C++/Visual Basic
- Explore the beauty of sequential association analysis and other data mining methods
- Understand and implement essential data mining techniques for a particular application
- Practice your ANALYTICS!

Brief Description

Based on a selected dataset, e.g. Microsoft Anonymous Web Data (publicly available from http://kdd.ics.uci.edu/databases/msweb/msweb.html), implement a data mining algorithm from ARM, Classification and Clustering (e.g. sequential ARM or ID3) and report the interesting patterns mined. The marking criteria are as follows.

Implementation	40%
Innovation (New Formulation, Idea, Algorithm, etc.)	30%
Report and Analysis	30%

Note:
If you don't want to write programs on your own, you may choose to use PASW Modeler (Clementine) or other public domain software like Weka 3 to carry out the mining task.

What to hand in

- ~~A video demo (no longer than 3 minutes) of your achievement~~.
- A short (<=8 pages) written report summarizing the work you have done and the **ANALYSIS** you have made.
- The source codes (if any) or the design (if using tools) of your system.

References

1. R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules, Proc. of Very Large Databases (VLDB'94), pp.487-499, Santiago, Chile, 1994.
2. R. Agrawal and R. Srikant. Mining Sequential Patterns, Proc. of IEEE Int. Conf. on Data Engineering (ICDE'95), pp.3-14, Taipei, 1995.