COSC 4335 *“Data Mining”*

Review Sheet for Fall 2018 Final Exam

The final exam has been scheduled for Monday, December 10, 2p in our classroom and will be “open everything” and will take about 110 minutes. The use of computers is not allowed! The exam counts approx. 20% towards the final course grade. The exam will cover the following topics (book pages refer to the second edition of the Tan book:

1. \*\*\* Agglomerative Hierarchical Clustering
   1. Transparencies
   2. Book pages 554-565 (Section 7.3)
2. \*\*\*\*\*\* Classification Techniques
   1. all transparencies covered in the lecture which cover decision trees, overfitting, , ensembles, neural networks and n-fold cross-validation.
   2. book pages: 119-137, 147-156. 249-250, 254-262, 296-300, 305-310
3. \*\*\* Association Analysis (just centering on basics and frequent item set generation/Apriori algorithm interestingness measures such as confidence, support)
   1. book pages: 357-373)
4. \*\*\*\*\* Outlier Detection
   1. Transparencies covered in the lecture
   2. Textbook pages 703-717, 719-728
5. \*\* Preprocessing (does not cover data quality)
   1. Preprocessing transparencies
   2. Textbook pages 50-65
6. \*\* Slides of the Data Science Lectures (will only ask basic questions; maybe one question will be an essay-style question)

Algorithms and techniques you should know well include: decision tree induction, Adaboost, how neural networks work and how they learn their weights; APRIORI algorithm for frequent item set creation, k-nearest neighbor based and model-based outlier detection.