COSC 3337 *“Data Science I”*

Review Sheet for Fall 2019 Final Exam

The final exam has been scheduled for Friday, December 6, 2p in our classroom and will be “open everything” and will take about 115 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. \*\*\*\*\*\*\*\*\*\* Clustering
   1. Transparencies on K-Means, PAM, DBSCAN, Hierarchical Clustering, Cluster Validation
   2. Book pages 5.25-529, 534-539, 548—552, 554-576, Silhouette Coefficient(581-582)
2. \*\*\*\* Classification Techniques
   1. all transparencies covered in the lecture which cover ensembles, Naïve Bayes and Neural networks
   2. Book pages 212-223, 225, 254-262, 296-300, 302-313
3. \*\* Preprocessing (does not cover data quality)
   1. Preprocessing transparencies
   2. Textbook pages 50-65
4. \*\* Slides of the Data Science Lectures (will only ask basic questions; maybe one question will be an essay-style question)
5. \* Outlier Detection
   1. Covered transparencies
   2. 703-705

Algorithms and techniques you should know well include: K-Means, PAM (not covered in textbook), DBSCAN, Agglomerative hierarchical clustering, Adaboost, how neural networks learn weights and backpropagration.