Implementation Project

We will be implementing three different data mining algorithms covering frequent itemset mining, classification, and clustering: the Apriori algorithm will be used for frequent itemset mining; the ID3 algorithm will be used for classification; and the X-means algorithm will be used for clustering.

Various data sets will be used to test each implementation. The data sets contain information regarding belgium and ta-feng retail market transactions, online news portal click-streams, breast cancer, chronic kidney disease, heart disease, voting records, and iris plants.

Proposed timeline and milestones.

1. Milestone 1: Sunday, March 6th
   1. HashTree development completed
   2. K-d tree development completed
2. Milestone 2: Sunday, March 13th
   1. Apriori mostly implemented
   2. XMeans overall design completed
3. Milestone 3: Sunday, March 20th
   1. Apriori algorithm implementation and testing completed
   2. Command line interface implemented
   3. XMeans algorithm implemented
4. Milestone 4: Sunday, March 27th
   1. Test data preprocessed/normalized for XMeans algorithm
   2. Testing completed on XMeans algorithm
   3. Information gain implemented for ID3 algorithm
   4. Progress Report completed
5. Milestone 5: Sunday, April 3rd
   1. …..unless you think of something John, I’ll delete this milestone. I just kept it in here to make it obvious that a week was skipped.
6. Milestone 6: Sunday, April 10th
   1. ID3 algorithm implementation completed
   2. Testing completed for ID3 algorithm
7. Mileston 7: Sunday, April 17th
   1. Final Report completed