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 implementation completed for Apriori algorithm
   2. K-d tree implementation completed for Xmeans algorithm
2. Milestone 2: Sunday, March 13th
   1. Apriori mostly implemented
   2. ID3 tree implementation completed
3. Milestone 3: Sunday, March 20th
   1. Apriori algorithm implementation and testing completed
   2. Command line interface implemented
   3. Information gain implemented for ID3 algorithm
   4. ID3 algorithm implemented
4. Milestone 4: Sunday, March 27th
   1. Testing completed on ID3 algorithm
   2. XMeans core structure/design completed
   3. Progress Report completed
5. Milestone 6: Sunday, April 10th
   1. XMeans algorithm implementation completed
   2. Testing completed for XMeans algorithm
6. Mileston 7: Sunday, April 17th
   1. Final Report completed