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 breast cancer, chronic kidney disease, heart disease, voting records, and iris plants.

Proposed timeline and milestones.

1. Milestone 1: Friday, March 4th
   1. HashTree development completed
   2. K-d tree development completed
2. Milestone 2: Some date
   1. Apriori algorithm pseudocode and overall design completed
   2. XMeans algorithm pseudocode and overall design completed
3. Milestone 3: some date
   1. Apriori algorithm implemented
   2. XMeans algorithm implemented
4. Milestone 4: blah
   1. ID3 algorithm pseudocode and overall design completed
5. Milestone 5:
   1. Select and normalize relevant attributes from data sets
   2. Test algorithms using data sets

Any additional info.