PROBLEM SET V – FREQUENT PATTERN MINING IMPLEMENTATIONS

(1) Test drive the basic version of Apriori and FP Growth algorithms for Frequent Itemset Mining using the package / library support in the platform of your choice. Test it with various support and confidence measures and generate a time comparison for varied data set sizes. To do the performance comparison you may use benchmark datasets provided for FIM such as the FIMI workshop or other sources.

(2) Extend the Apriori Algorithm discussed in the class supporting Transaction Reduction approach to improve the time complexity issue as a result of the repeated scans limitation of Apriori. You may compare this extended version with the earlier implementations in (1) over the same benchmark dataset.

(3) Test drive any one implementation in (1) or (2) adopting a Vertical Transaction Database format.

(4) Using a vertical transaction database notation, generate the FI's following the intersection approach (basic ECLAT) discussed in the class. Use earlier benchmark datasets in (1).

(5) Extend the basic Apriori algorithm to generate Frequent Patterns which differentiate ab from ba (ordered patterns generation).

(6) Implement following extensions to Apriori Algorithm (discussed / to be discussed in the class): Hash based strategy, Partitioning Approach and Sampling strategies. You may refer to online tutorials for a formal pseudocode description.

(7) Implement the Dynamic Itemset Counting Algorithm for Frequent Itemset Generation.

(8) Test drive (download exe's or generate one using open source versions) any three algorithm for FIM not discussed in the class.

(9) For any of the real time dataset shared by my TA earlier or some other online challenge dataset, implement pre-processing strategies required and generate associations amongst the attributes of interests using any one FP implementation.