

**COMP 4420 Introduction to Data Mining**  
**Project#1 –Apriori Implementation**  
**Due Date: 10:00 AM, 04/13/16**

**Overview:**

For this project, you will write a command line based C++ program to generate all large itemsets using the *Apriori* algorithm. The content of a sample data file is given below:

```
A B C D E
A B
B D
B C
A B D
B C
A C
A B C E
A B C
```

The first line represents all the items and the following lines represent all the transactions. Given a data file, the program should generate all the large itemsets using a minimum support threshold provided from the command line. A sample execution is given next:

```
./g++ apriori.cpp -o apriori
./apriori data.txt -s .2
Large 1-itemsets: A B C D
Large 2-itemsets: AB AC BC BD
Large 3-itemsets: ABC
Large 4-itemsets:
Large 5-itemsets:
```

**Deliverables:**

You will need to turn in your source file *apriori.cpp* in Blackboard and a printed report in class by 10:00 AM on April 13<sup>th</sup>. The report should contain (a) a title page whose format can be found in Blackboard and (b) a comparison of the large itemsets generated by your program and the Apriori implementation in Weka. For the comparison, you will be using the data file provided in Blackboard.