

FALL 2023: CS 4435/5435 and DASE 4435 Data Mining

Homework 2, Due on **September 28, 2023**

1. A database has five transactions. Let  $min\ sup = 40\%$  and  $min\ conf = 60\%$ .

TID	Items Bought
T1	{A, C, D}
T2	{B, C, E}
T3	{A, B, C, E}
T4	{B, E}
T5	{E}

- (a) Find all frequent itemsets using Apriori and FP-growth, respectively. You are required to show the pattern generation process. Each process is worth some points. Compare the efficiency of the two mining processes. **(50 points)**
- (b) List all of the association rules (with support  $s$  and confidence  $c$ ) the frequent item sets and make a decision to reject or accept the rule. The result should be presented as shown below: **(40 points)**

Rules	Support(X, Y)	Support(X)	Confidence	Decision

- (c) Using **Weka**, list the top 10 association rules, using the Apriori algorithm with parameters provided above for the minimum support and confidence threshold. **(10 points)**.