Tutorial 3B: Association Rule Mining

1. Association rules describe the relationships between certain variables in a large database.
   A. True
   B. False
2. What is Market Basket Analysis?
   A. The process of discovering association rules between variables in a dataset.
   B. Is the process of developing clusters in order to segregate data and discover the relevant categories of data.
   C. Market Basket analysis uses decision trees to predict outcomes.
   D. All of the above.
3. In market basket analysis; confidence is:
   A. The general measure of association between the two item sets.
   B. The conditional probability that a transaction contains item set B given that it contains item set A
   C. The probability that the two item sets occur together.
   D. The conditional probability that a transaction contains item set B given that it does not contain item set A.
4. Given a rule of the form IF X THEN Y, rule confidence is defined as the conditional probability that:
   A. Y is true when X is known to be true.
   B. X is true when Y is known to be true.
   C. Y is false when X is known to be false.
   D. X is false when Y is known to be false.
5. Association rule support is defined as
   A. the percentage of instances that contain the antecedent conditional items listed in the association rule.
   B. the percentage of instances that contain the consequent conditions listed in the association rule.
   C. the percentage of instances that contain all items listed in the association rule.
   D. the percentage of instances in the database that contain at least one of the antecedent conditional items listed in the association rule.

6. Apply the Apriori algorithm to the following data set:

Trans ID	Items Purchased
101	milk, bread, eggs
102	milk, juice
103	juice, butter
104	milk, bread, eggs
105	coffee, eggs
106	coffee
107	coffee, juice
108	milk, bread, cookies, eggs
109	cookies, butter
110	milk, bread

The set of items is {milk, bread, cookies, eggs, butter, coffee, juice}. Use 20% for the minimum support value.

7. Show two rules that have a support >= 0.2 and confidence >= 0.8 or greater for an itemset containing three items from question 6.