Imaduddin Sheikh

CS 422 - Homework 5

1. Exercises

1.1 Tan, Ch. 5 (Association Analysis)

Q.1)

a) Eggs → Bread

It is common for customers to buy items for breakfast. If a customer buys eggs, which they have a high chance of doing so (High Support), they have a high chance of also buying bread. So, there is high confidence In buying bread with the eggs.

This type of rules is uninteresting because they are quite well-known.

b) Eggs → Hair Shampoo

Eggs have high support as many people buy eggs. But the rule demonstrates low confidence because it is not necessary that a customer who buys eggs will also buy a hair shampoo.

This type of rule/pattern is also not interesting due to low confidence.

c) Coconut Oil → Laundry Detergent

Coconut Oil doesn't appear frequently in customer transactions so there is low support. Someone who buys coconut oil will not necessarily buy laundry detergents.

This type of rules is also not interesting.

d) Strawberries → Chocolate

Customer don't typically buy strawberries (low support) but when they do chocolate also tends to be in the transaction alongside strawberries (high confidence).

Such a rule in interesting and quite useful.

Q.2)

a) There are 10 distinct baskets/transactions.

- {e}: $s = \frac{8}{10} = 0.8$
- {b, d}: $s = \frac{2}{10} = 0.2$
- {b, d, e}: $s = \frac{2}{10} = 0.2$

b)

- {b, d} → {e}: $c = \frac{0.2}{0.2} = 1$
- {e} → {b, d}: $c = \frac{0.2}{0.8} = 0.25$

Hence, confidence is not symmetric.

c) There are 5 customers/baskets in total.

- {e}: $s = \frac{4}{5} = 0.8$
- {b, d}: $s = \frac{5}{5} = 1$
- {b, d, e}: $s = \frac{4}{5} = 0.8$

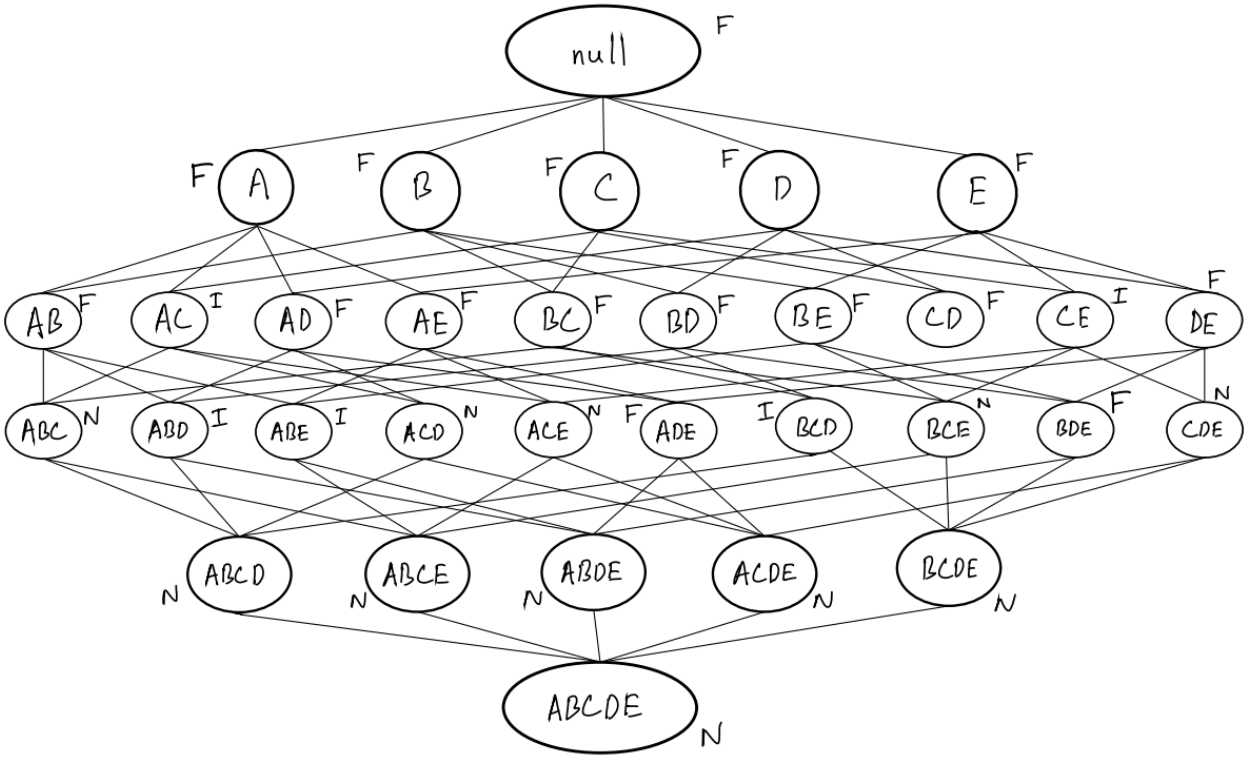
d)

- {b, d} → {e}: $c = \frac{0.8}{1} = 0.8$
- {e} → {b, d}: $c = \frac{0.8}{0.8} = 1$

e) No relationship between c_1, c_2, s_1 and s_2 as support for {e} remained the same, nothing can be said about support for {b, d} and {b, d, e} (even though it increased significantly for Customer ID). Increase in support is also not reflected in changes in confidence of the rules.

Q. 9

a)



b) $Percentage\ of\ frequent\ itemsets = \frac{16}{32} \times 100 = 50\%$