Tutorial 7 : Association Rule Mining

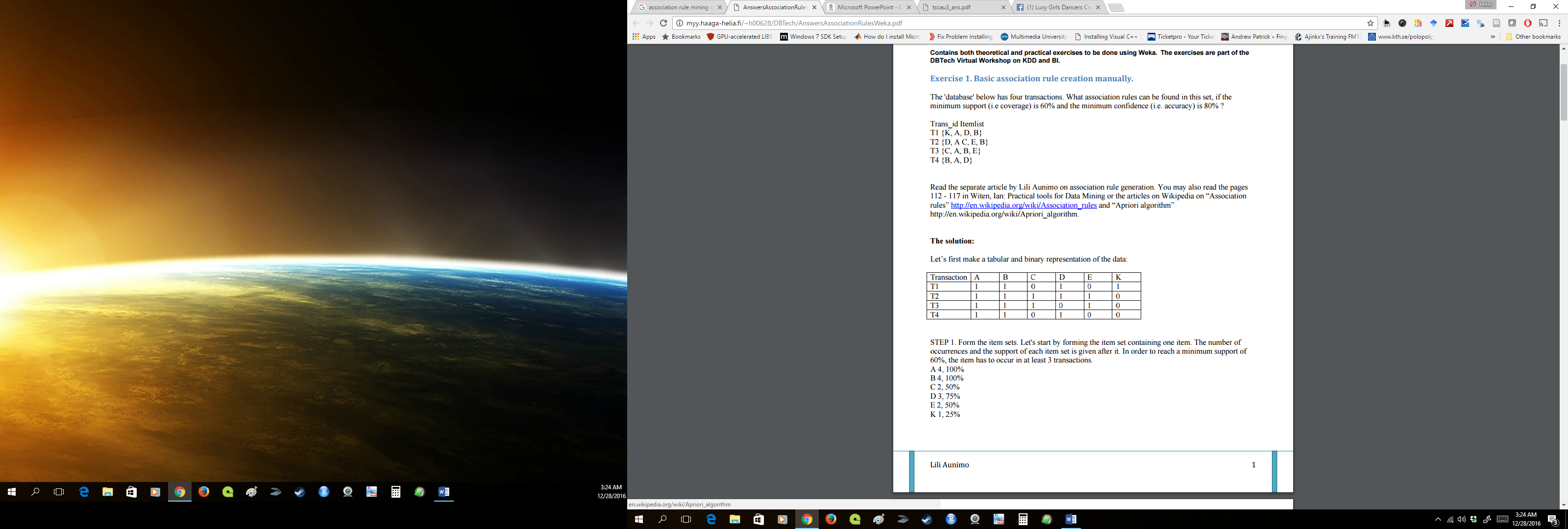
# Question 1

|  |  |
| --- | --- |
| Transaction ID | Itemlist |
| T1 | {K,A,D,B} |
| T2 | {D,A,C,E,B} |
| T3 | {C,A,B,E} |
| T4 | {B,A,D} |

What association rules can we find from the frequent item-sets above with minimum support of 0.6 and confidence of 0.8?

## Guided answer

STEP 1 : Let’s first make a tabular and binary representation of the data.



STEP 2: Let's start by forming the item set containing one item. The number of occurrences and the support of each item set is given after it. In order to reach a minimum support of **60**%, the item has to occur in at least 3 transactions.

|  |  |  |
| --- | --- | --- |
| Single Item | Occurrences | Support |
| A | 4 | 100 |
| B | 4 | 100 |
| C | 2 | 50 |
| D | 3 | 75 |
| E | 2 | 50 |
| K | 1 | 25 |

**What are the candidate single-itemsets?**

STEP 3: Continue by forming item set containing 2-items. Use the same format as before. Choose from the list in STEP 2 to form the permutations.

|  |  |  |
| --- | --- | --- |
| 2-item sets | Occurrences | Support |
| {A, B} | 4 | 100 |
| {A, D} | 3 | 75 |
| {B, D} | 3 | 75 |

STEP 4: Let’s continue by forming item-sets of 3 items from the previous step with support exceeding **60**%. What are the candidate itemsets at this point?

{A, B, D} 3 75%

STEP 5: We now choose calculate the confidence of the candidate item-sets from the previous steps.

|  |  |
| --- | --- |
| Candidate item-sets | Confidence |
| A->B  B->A  A->D  D->A  B->D  D->B  AB->D  D->AB  AD->B  B->AD  BD->A  A->BD | 100  100  100  75  100  75  100  75  75  100  100  100 |

We only choose the itemsets with confidence above 80 leaving…?

A->B B->A D->A D->B AB->D AD->B BD->A

# QUESTION 2

Given the following transaction data, answer the following questions.

|  |  |
| --- | --- |
| Transaction ID | Items |
| 1 | {Jacket, Boots} |
| 2 | {Milk, Cheese, Bread, Shoes} |
| 3 | {Cloth, Bread} |
| 4 | {Milk, Bread, Shoes, GreetingCard, Port, Apple, Soup} |
| 5 | {Cheese, Shoes, Beef} |
| 6 | {Jacket, Bread, SkiPants} |

a. What are the supports and confidences of the following two rules? Rule1: Milk → Bread Rule2: Bread → Milk

**Milk 2 33.3%**

**Bread 4 66.7%**

**Milk, Bread 2 33.3%**

**Milk -> Bread 100%**

**Bread -> Milk 50%**

b. If the minimum support is set at 30%, how many large itemsets will be found?

c. If the minimum support is set at 60%, how many large itemsets will be found?

d. Discuss what happens if the minimum support is set at 0%.