IT414 Assignment - 1

Name: Durga Supriya HL
Roll no.: 201IT121

The dataset matrix was created randomly:

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
T1	0	1	0	0	1	0	1	1	0	0
T2	0	0	0	1	0	0	0	0	1	1
T3	0	0	0	0	0	0	0	1	1	0
T4	1	1	0	1	1	0	1	1	0	0
T5	0	0	0	0	0	0	0	0	0	1
T6	1	0	0	1	1	1	0	0	1	1
T7	1	0	0	0	1	1	0	0	0	1
T8	1	0	0	1	0	0	1	0	0	1
T9	1	0	0	0	1	0	1	1	0	1
T10	0	1	1	0	1	1	1	0	1	0
T11	0	0	1	1	0	0	1	1	0	1
T12	0	0	0	0	0	1	1	0	1	0
T13	1	0	0	1	0	1	0	0	0	0
T14	0	0	0	0	1	0	1	0	1	1
T15	1	1	0	0	0	1	1	1	0	1
T16	0	1	1	0	0	1	0	0	0	0
T17	0	1	0	1	1	0	0	0	0	0
T18	0	1	0	0	1	0	1	1	1	0
T19	0	0	0	1	1	1	0	1	1	1
T20	1	0	1	0	1	0	0	1	0	0
T21	1	1	0	0	0	1	1	1	1	0
T22	0	1	0	1	1	0	0	0	0	0
T23	0	1	0	0	1	0	1	1	0	1
T24	1	0	1	1	1	1	0	0	1	0
T25	1	1	0	1	1	0	0	1	1	0

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
T26	0	0	0	0	1	1	0	1	0	0
T27	1	0	0	1	0	1	0	0	0	1
T28	1	1	0	1	0	0	0	0	0	1
T29	0	1	0	1	0	0	0	0	0	0
T30	1	1	1	1	0	0	1	0	0	1
T31	1	1	1	0	0	0	0	1	0	0
T32	0	0	1	0	1	0	1	1	1	1
T33	1	1	1	1	0	0	1	1	1	0
T34	1	0	0	0	0	0	1	0	1	0
T35	0	0	0	0	0	1	0	0	0	1
T36	0	1	1	0	0	0	1	0	1	0
T37	0	0	0	1	0	1	0	1	0	0
T38	0	0	0	1	1	0	0	0	1	1
T39	1	0	1	0	0	0	1	0	0	0
T40	0	0	0	1	0	0	1	1	1	1
T41	0	0	0	1	0	0	0	0	0	1
T42	0	1	1	0	0	0	0	0	0	0
T43	0	0	0	0	0	0	0	0	0	1
T44	0	0	1	0	0	0	0	0	0	0
T45	0	0	1	1	0	0	0	0	0	0
T46	1	0	1	0	1	0	0	1	1	0
T47	0	0	1	0	1	0	0	1	0	1
T48	0	1	0	0	1	1	0	1	0	1
T49	0	1	0	1	1	0	0	1	1	1
T50	0	0	0	0	0	1	1	0	0	0

Frequent item set

Support counts of individual items

```
Support of single items:
I1  I2  I3  I4  I5  I6  I7  I8  I9  I10
19  20  16  22  22  16  20  22  19  23
```

- k = 1

Minisup was assumed to be 6 so none of them got pruned in this step. Thus all the items are present in the frequent item set of k=1. The support counts are mentioned beside each item.

```
min-sup is assumed to be 6
Frequent item set fro k=1 :
I1 19
I2 20
I3 16
I4 22
I5 22
I6 16
I7 20
I8 22
I9 19
I10 23
```

- k = 2

The frequent item set for k=2 is calculated in the next step. The support counts are mentioned beside each item. The item sets with support count less than minsup are pruned. For example (I2, I6) is pruned since the pairs support count is 5.

```
Frequent item set for k=2 :
I1 I2: 8
I1 I3: 7
I1 I4: 10
I1 I5: 8
I1 I6: 7
I1 I7: 9
I1 I8: 9
I1 I9: 7
I1 I10: 8
I2 I3: 7
I2 I4: 9
I2 I5: 10
I2 I7: 10
I2 I8: 11
I2 I9: 7
I3 I7: 7
I3 I8: 7
I4 I5: 9
I4 I8: 8
I4 I9: 9
I4 I10: 12
I5 I6: 7
I5 I7: 8
I5 I8: 14
I5 I9: 11
I5 I10: 11
I6 I10: 7
I7 I8: 11
I7 I9: 10
I7 I10: 9
I8 I9: 10
I8 I10: 10
I9 I10: 8
```

- k = 3

The frequent item set for k=3 is then calculated. The support counts are mentioned beside each item.

```
Frequent item set for k=3 :
I2 I5 I8: 7
I2 I7 I8: 7
I5 I8 I10: 7
```

Rule generation

For rule generation, min_confidence was assumed to be 1.7.
We know that the confidence count will be as follows:

$$\text{Confidence}(A \rightarrow B) = \text{Support\_count}(A \cup B) / \text{Support\_count}(A)$$

$$= \frac{\text{Support\_count}(A) + \text{Support\_count}(B) + \text{Support\_count}(A,B)}{\text{Support\_count}(A)}$$

```
min_confidence is assumed to be: 1.7
Rules generated       Confidence count
I1-> I5:             1.73684
I1-> I10:             1.78947
I3-> I7:             1.8125
I3-> I8:             1.9375
I6-> I10:             2
```