Assignment 4.2

Name: Varun Borar

Enrolment Number: A2305219377

1. **Describe support and confidence.**

   Support: The number of instances in the dataset that contain all the items listed in the given itemset.

$$Support = \frac{no.of\ instances\ containing\ A}{Total\ no.of\ instances}$$

   Confidence: Given an association rule A->B (if A then B), confidence for rule is the conditional probability that B is true when A is true.

$$Confidence(\ A \rightarrow B) = \frac{no.of\ instances\ containing\ A\ and\ B\ both}{no.of\ instances\ containing\ A}$$

2. **Express the formula for confidence using support**.

$$Confidence(A \rightarrow B) = \frac{Support(AUB)}{Support(A)}$$

3. **Weather data set for association rule mining**

No.	Outlook	Temperature	Humidity	Windy	Play
1.	Sunny	Hot	High	False	No
2.	Sunny	Hot	High	True	No
3.	Overcast	Hot	High	False	Yes
4.	Rain	Mild	High	False	Yes
5.	Rain	Cool	Normal	False	Yes
6.	Rain	Cool	Normal	True	No
7.	Overcast	Cool	Normal	True	Yes
8.	Sunny	Mild	High	False	No
9.	Sunny	Cool	Normal	False	Yes
10.	Rain	Mild	Normal	False	Yes
11.	Sunny	Mild	Normal	True	Yes
12.	Overcast	Mild	High	True	Yes
13.	Overcast	Hot	Normal	False	Yes
14.	Rain	Mild	High	True	No

4. **Generate the frequent 1-itemsets (support count = 3)**

Items	Frequency
Sunny (Outlook)	5
Overcast (Outlook)	4
Rain (Outlook)	5
Hot (Temperature)	4
Mild (Temperature)	6
Cool (Temperature)	4
High (Humidity)	7
Normal (Humidity)	7
True (Windy)	6
False (Windy)	8
Yes (Play)	9
No (Play)	5

Since all the given items have support count greater than 3 thus all will be included in the dataset.

5. **Let φ = 3. Generate the frequent 2-itemsets.**

Item	Frequency	Included in Frequent 2-itemset
Sunny, Hot	2	NO
Sunny, Mild	2	NO
Sunny, Cool	1	NO
Sunny, High	3	YES
Sunny, Normal	2	NO
Sunny, False	3	YES
Sunny, True	2	NO
Sunny, No	3	YES
Sunny, Yes	2	NO
Overcast, Hot	2	NO
Overcast, Mild	1	NO
Overcast, Cool	1	NO
Overcast, High	2	NO
Overcast, Normal	2	NO
Overcast, False	2	NO
Overcast, True	2	NO
Overcast, No	0	NO
Overcast, Yes	4	YES
Rain, Hot	0	NO
Rain, Mild	3	YES
Rain, Cool	2	NO
Rain, High	2	NO
Rain, Normal	3	YES
Rain, False	3	YES
Rain, True	2	NO
Rain, No	2	NO
Rain, Yes	3	YES
Hot, High	3	YES
Hot, Normal	1	NO

Hot, False	3	YES
Hot, True	1	NO
Hot, No	2	NO
Hot, Yes	2	NO
Mild, High	4	YES
Mild, Normal	2	NO
Mild, False	3	YES
Mild, True	3	YES
Mild, No	2	NO
Mild, Yes	4	YES
Cool, High	0	NO
Cool, Normal	4	YES
Cool, False	2	NO
Cool, True	2	NO
Cool, No	1	NO
Cool, Yes	3	YES
High, False	4	YES
High, True	3	YES
High, No	4	YES
High, Yes	3	YES
Normal, False	4	YES
Normal, True	3	YES
Normal, No	1	NO
Normal, Yes	6	YES
False, No	2	NO
False, Yes	6	YES
True, No	3	YES
True, Yes	3	YES

6. Let $\varphi = 3$. Generate the frequent 3-itemsets.

Using frequent 2-Itemset, and applying A-priori algorithm to prune the dataset

Item	Included after pruning.
Sunny, High, False	YES
Sunny, High, No	YES
Sunny, False, No	NO
Rain, Mild, Normal	NO
Rain, Mild, False	YES
Rain, Mild, Yes	YES
Rain, Normal, False	YES
Rain, Normal, Yes	YES
Rain, False, Yes	YES
Hot, High, False	YES
Mild, High, False	YES
Mild, High, True	YES
Mild, High, Yes	YES
Mild, False, Yes	YES

Mild, True, Yes	YES
Cool, Normal, Yes	YES
High, False, Yes	YES
High, False, No	NO
High, True, Yes	YES
High, True, No	YES
Normal, False, Yes	YES
Normal, True, Yes	YES

Item	Frequency	Included in Frequent 3-itemset
Sunny, High, False	2	NO
Sunny, High, No	3	YES
Rain, Mild, False	2	NO
Rain, Mild, Yes	2	NO
Rain, Normal, False	2	NO
Rain, Normal, Yes	2	NO
Rain, False, Yes	3	YES
Hot, High, False	2	NO
Mild, High, False	2	NO
Mild, High, True	2	NO
Mild, High, Yes	2	NO
Mild, False, Yes	2	NO
Mild, True, Yes	2	NO
Cool, Normal, Yes	3	YES
High, False, Yes	3	YES
High, True, Yes	1	NO
High, True, No	2	NO
Normal, False, Yes	4	YES
Normal, True, Yes	2	NO

7. Using 75% minimum confidence and 20% minimum support, generate one-antecedent association rules for predicting *play*.

If Antecedent, Then consequent	Support	Confidence
Sunny -> No	(3/14) = 21.4%	(3/5) = 60%
Overcast-> Yes	(4/14) = 28.5%	(4/4) = 100%
Rain -> Yes	(3/14) = 21.4%	(3/5) = 60%
Mild -> Yes	(4/14) = 28.5%	(4/6) = 66.7%
Cool -> Yes	(3/14) = 21.4%	(3/4) = 75%
High -> Yes	(3/14) = 21.4%	(3/7) = 42.9%
High -> No	(4/14) = 28.5%	(4/7) = 57.1%
Normal -> Yes	(6/14) = 42.9%	(6/7) = 85.7%
False -> Yes	(6/14) = 42.9%	(6/8) = 75%
True -> No	(3/14) = 21.4%	(3/6) = 50%
True -> Yes	(3/14) = 21.4%	(3/6) = 50%

From the above table we can derive the following association rules with 75% min confidence and 20% min support,

1. If Outlook = Overcast then Play = yes. (Support*Confidence = 0.2857)
2. If Temperature = Cool then Play = yes. (Support*Confidence = 0.2142)
3. If Humidity = Normal then Play = yes. (Support*Confidence = 0.3673)
4. If Windy = False then Play = yes. (Support*Confidence = 0.3214)

8. Using 75% minimum confidence and 20% minimum support, generate two-antecedent association rules for predicting *play*.

If Antecedent, Then consequent	Support	Confidence
Sunny, High -> No	(3/14) = 21.4%	(3/3) = 100%
Rain, False -> Yes	(3/14) = 21.4%	(3/3) = 100%
Cool, Normal -> Yes	(3/14) = 21.4%	(3/4) = 75%
High, False -> Yes	(3/14) = 21.4%	(3/4) = 75%
Normal, False -> Yes	(4/14) = 28.6%	(4/4) = 100%

From the above table we can derive the following association rules with 75% min confidence and 20% min support,

1. If Outlook = Sunny and Humidity = High then Play = No. (Support*Confidence = 0.2142)
2. If Outlook = Rain and Windy = False then Play = Yes. (Support*Confidence = 0.2142)
3. If temperature = Cool and Humidity = Normal then Play = Yes. (Support*Confidence = 0.16)
4. If Humidity = High and Windy = False then Play = Yes. (Support*Confidence = 0.16)
5. If Humidity = Normal and Windy = False then Play = Yes. (Support*Confidence = 0.2857)

9. Multiply the observed support times the confidence for each of the rules in Exercises 7 and 8, and rank them in a table.

Rank	Rule
1	If Humidity = Normal then Play = yes. (Support*Confidence = 0.3673)
2	If Windy = False then Play = yes. (Support*Confidence = 0.3214)
3	If Outlook = Overcast then Play = yes. (Support*Confidence = 0.2857)
4	If Humidity = Normal and Windy = False then Play = Yes. (Support*Confidence = 0.2857)
5	If Temperature = Cool then Play = yes. (Support*Confidence = 0.2142)
6	If Outlook = Rain and Windy = False then Play = Yes. (Support*Confidence = 0.2142)
7	If Outlook = Sunny and Humidity = High then Play = No. (Support*Confidence = 0.2142)
8	If temperature = Cool and Humidity = Normal then Play = Yes. (Support*Confidence = 0.16)
9	If Humidity = High and Windy = False then Play = Yes. (Support*Confidence = 0.16)