Experiment 7

Title: Apply APRIORI algorithm on “weather.nominal” dataset and analyze the results.

Dataset: weather.nominal

Step-1: Upload dataset in Weka.

Fig.7.1

The image displays an interface for analyzing a dataset named "weather.nominal". This dataset appears to be designed for predictive modeling, specifically to understand and predict the "play" attribute based on weather conditions. It contains 14 instances & 5 Attributes, each representing a different day's weather conditions and whether an activity was played.

- **Outlook (Nominal):** Describes the general weather outlook, with distinct values such as 'Sunny', 'Overcast', and 'Rainy'. The selected attribute shows 5 instances for each of these outlook types.
- **Temperature (Nominal):** Likely represents the temperature conditions (e.g., 'hot', 'mild', 'cool').
- **Humidity (Nominal):** Indicates the humidity levels (e.g., 'high', 'normal').
- **Windy (Nominal):** Denotes whether there is wind (e.g., 'true', 'false').
- **Play (Nominal):** This is the class attribute, indicating whether a certain activity was played or not (e.g., 'yes', 'no'). The bar chart at the bottom right visually represents the distribution of the 'play' attribute across different 'outlook' categories, with red and blue bars likely representing the two possible outcomes for 'play.

Step-2: Apriori Algorithm Configuration

Fig.7.2

The image displays a data mining interface focused on the "Associate" tab, specifically for running an **Apriori algorithm**. This setup is designed for **association rule mining**, which is used to discover interesting relationships or co-occurrence patterns among items in large datasets.

Step-3: Configure the parameters

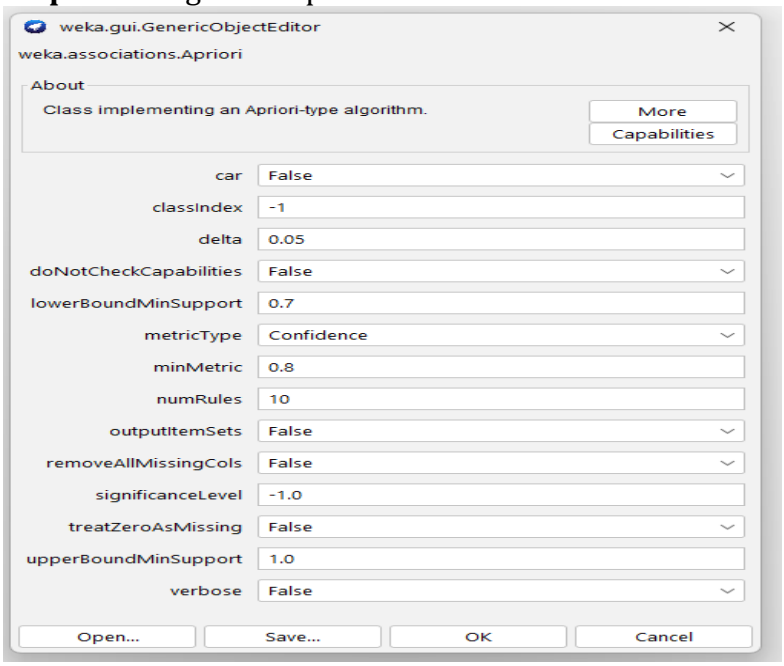


Fig.7.3

This image displays the configuration window for the **Apriori algorithm** within the Weka data mining software (specifically, weka.associations.Apriori). This step is about setting the detailed parameters that control how the Apriori algorithm will discover association rules from a dataset.

Step-4: Associator Output and Result (min.confidence= 80%, min.support= 70%)

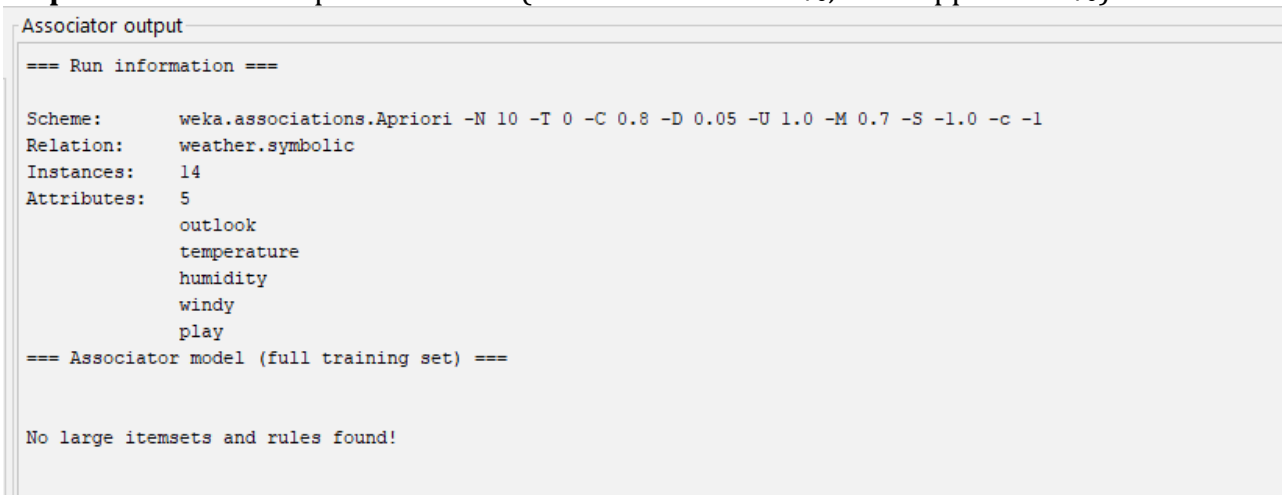


Fig.7.4

This image displays the "Associator output" section, which provides the results and run information after executing the Apriori association rule mining algorithm.This section presents a summary of the Apriori algorithm's execution and its findings. Specific parameters used for the run: N 10 (maximum 10 rules), -C 0.8 (minimum confidence of 80%), -M 0.7 (minimum support of 70%). However, the crucial output for this specific run is "No large itemsets and rules found!". So under these conditions, we are unable to generate Association Rule.

Step-5: Configure the parameters

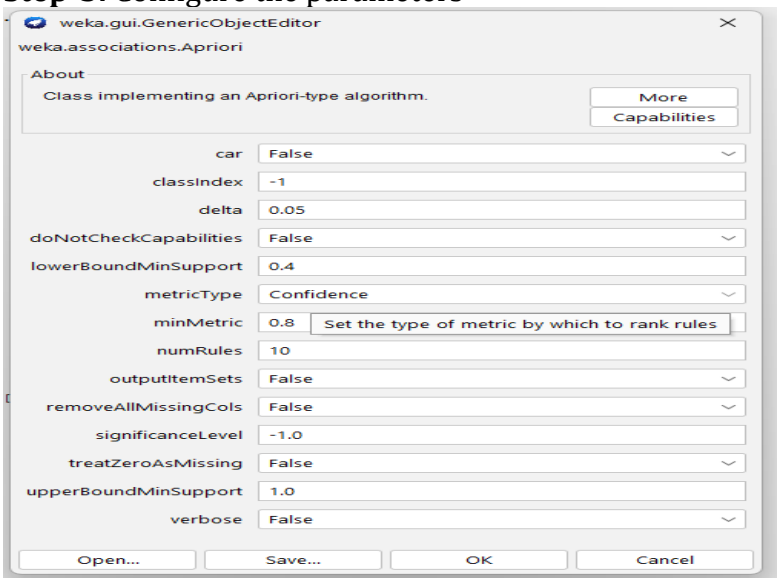
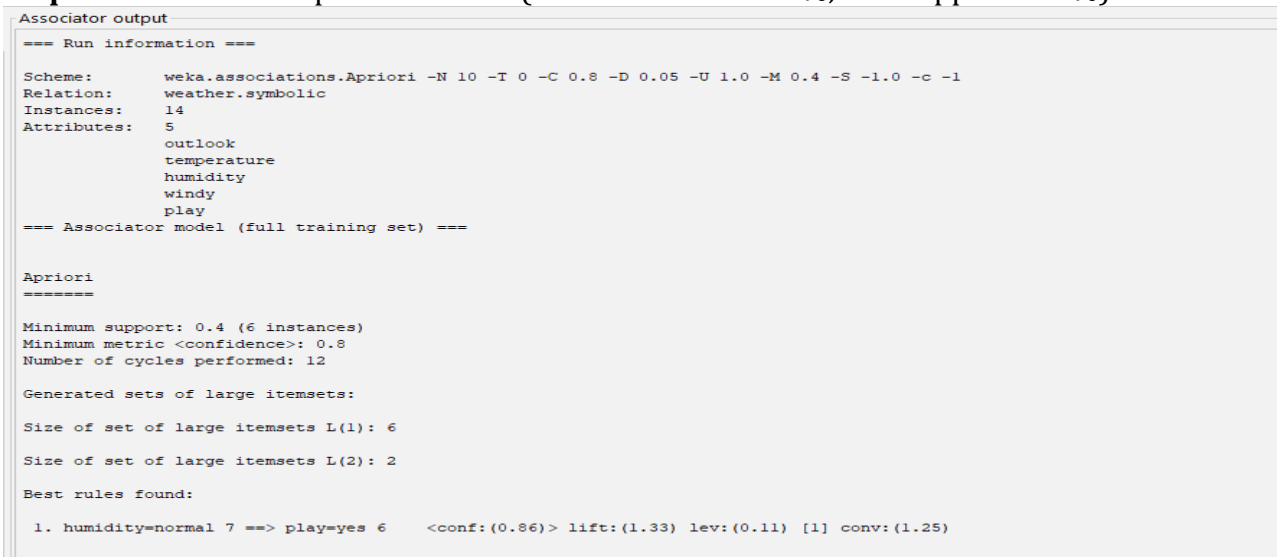


Fig.7.5

This image displays the configuration window for the **Apriori algorithm** within the Weka data mining software (specifically, weka.associations.Apriori). This step is about setting the detailed parameters that control how the Apriori algorithm will discover association rules from a dataset.

Step-6: Associator Output and Result (min.confidence= 80%, min.support= 40%)



Associator output

```
=== Run information ===

Scheme:       weka.associations.Apriori -N 10 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.4 -S -1.0 -c -1
Relation:     weather.symbolic
Instances:    14
Attributes:   5
              outlook
              temperature
              humidity
              windy
              play
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.4 (6 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 2

Best rules found:

 1. humidity=normal 7 ==> play=yes 6    <conf:(0.86)> lift:(1.33) lev:(0.11) [1] conv:(1.25)
```

Fig.7.6

This image displays the "Associator output" section, which provides the results and run information after executing the Apriori association rule mining algorithm.This section presents a summary of the Apriori algorithm's execution and its findings. Specific parameters used for the run: N 10 (maximum 10 rules), -C 0.8 (minimum confidence of 80%), -M 0.4 (minimum support of 40%). However, this execution successfully found "large itemsets and rules."

Under "Generated sets of large itemsets:", it reports:

"Size of set of large itemsets L(1): 6" (meaning 6 frequent itemsets of size 1 were found).

"Size of set of large itemsets L(2): 2" (meaning 2 frequent itemsets of size 2 were found).

Step-7: Configure the parameters

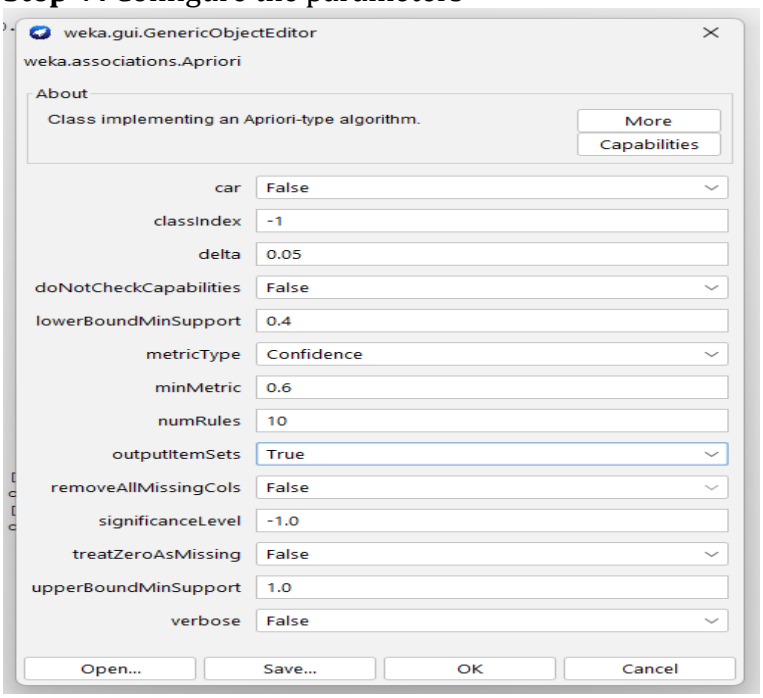


Fig.7.7

This image displays the configuration window for the **Apriori algorithm** within the Weka data mining software (specifically, weka.associations.Apriori). This step is about setting the detailed parameters that control how the Apriori algorithm will discover association rules from a dataset.

Step-8: Associator Output and Result (min.confidence= 60%, min.support= 40%)

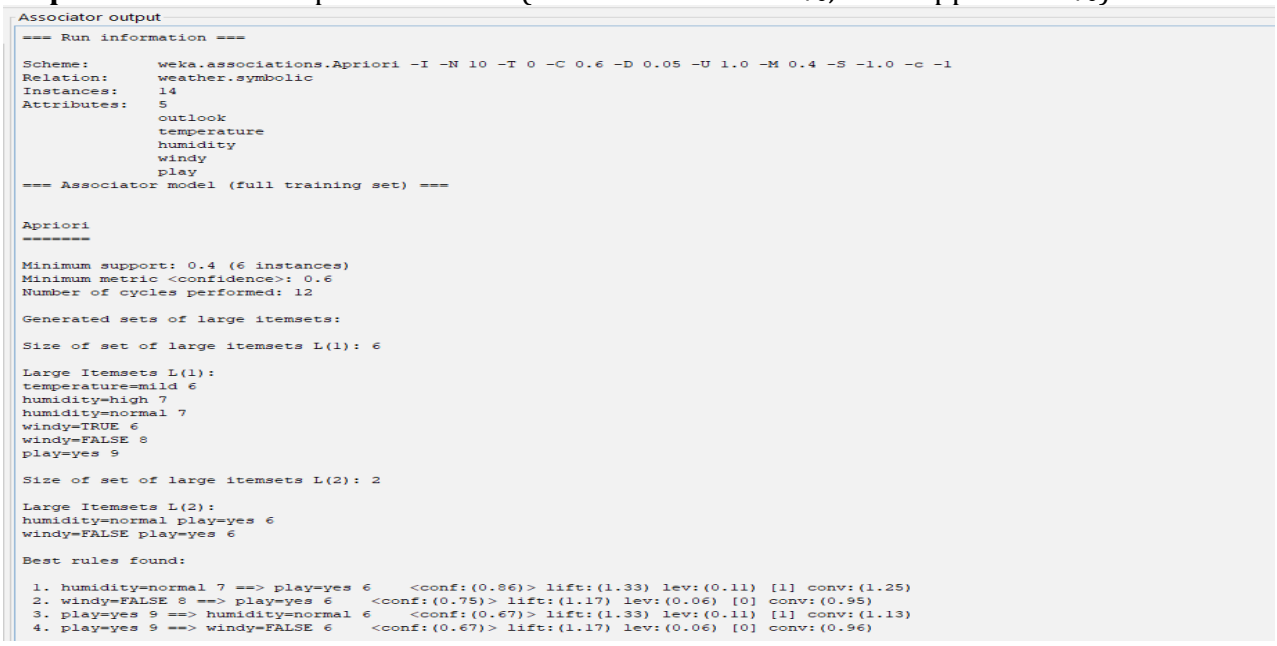


Fig.7.8

This image displays the "Associator output" section, which provides the results and run information after executing the Apriori association rule mining algorithm.This section presents a summary of the Apriori algorithm's execution and its findings. Specific parameters used for the run: N 10 (maximum 10 rules), -C 0.6 (minimum confidence of 60%), -M 0.4 (minimum support of 40%). However, this execution successfully found "large itemsets and rules."
The "Generated sets of large itemsets" now lists the actual itemsets found:
Size of set of large itemsets L(1): 6

- temperature=mild 6
- humidity=high 7
- humidity=normal 7
- windy=TRUE 6
- windy=FALSE 8
- play=yes 9

Size of set of large itemsets L(2): 2

- humidity=normal play=yes 6
- windy=FALSE play=yes 6

Finally, the "Best rules found:" section now lists four association rules:

- humidity=normal ➔ play=yes
- windy=FALSE ➔ play=yes
- play=yes ➔ humidity=normal
- play=yes ➔ windy=FALSE

Step-9: Configure the parameters

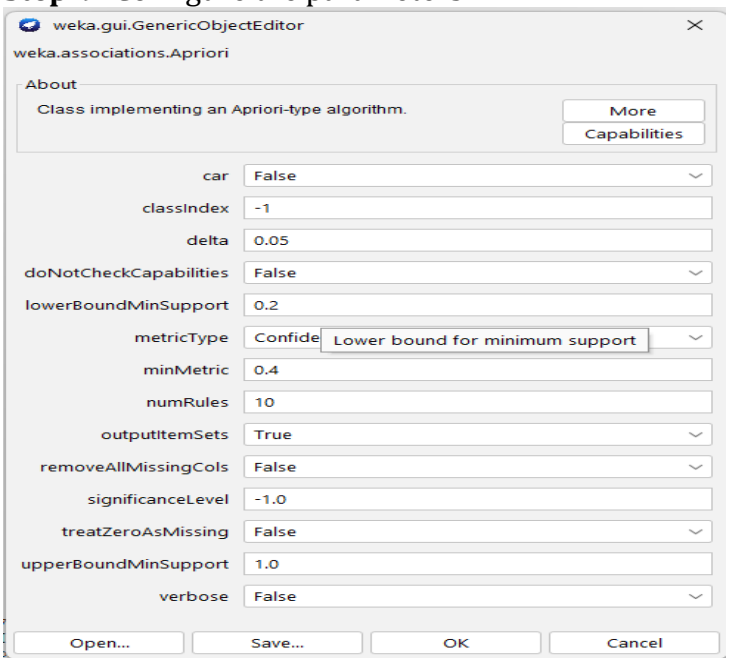


Fig.7.9

This image displays the configuration window for the **Apriori algorithm** within the Weka data mining software (specifically, weka.associations.Apriori). This step is about setting the detailed parameters that control how the Apriori algorithm will discover association rules from a dataset.

Step-10: Associator Output and Result (min.confidence= 40%, min.support= 20%)

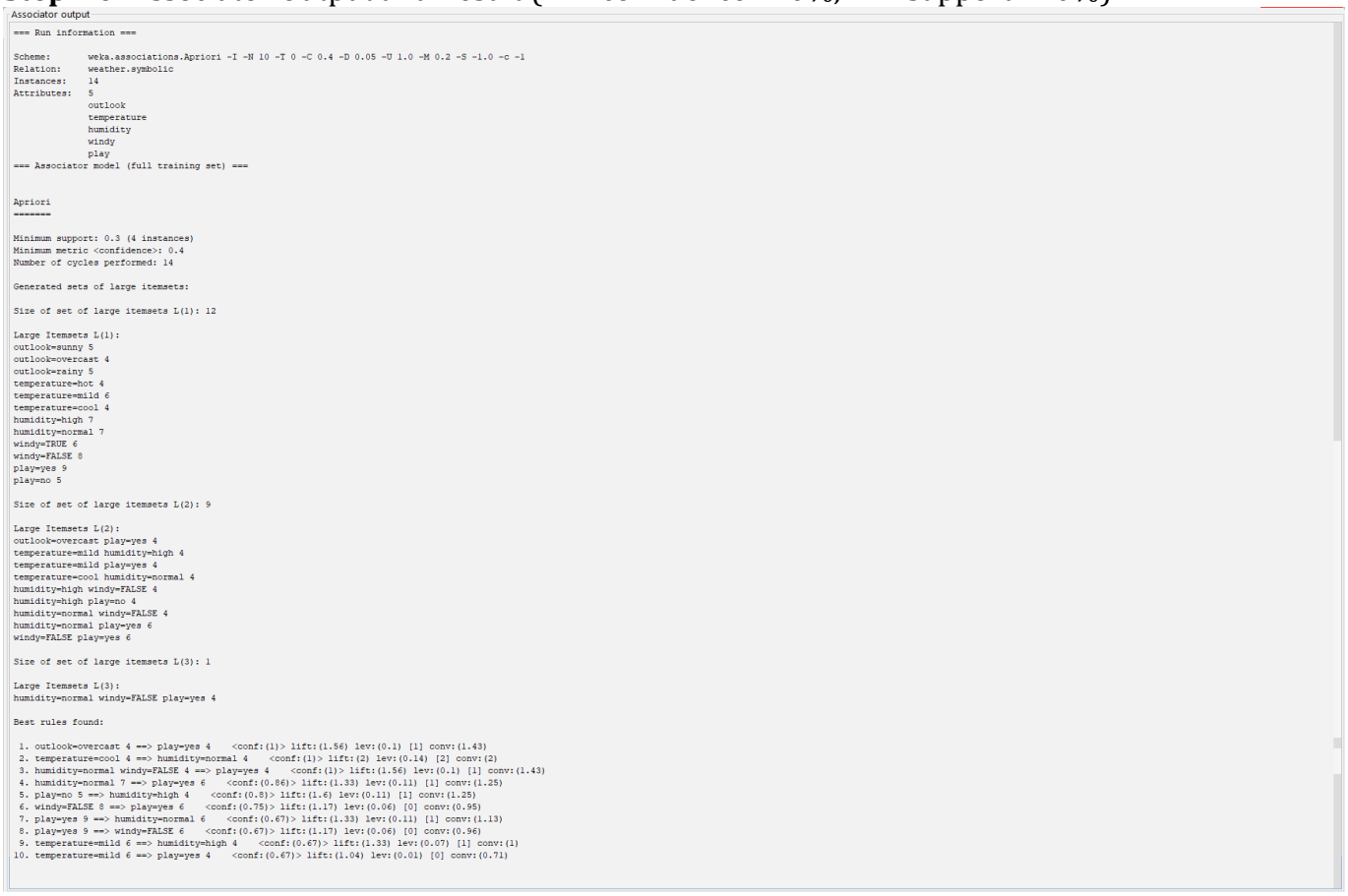


Fig.7.10

This image displays the "Associator output" section, which provides the results and run information after executing the Apriori association rule mining algorithm.This section presents a summary of the Apriori algorithm's execution and its findings. Specific parameters used for the run: N 10 (maximum 10 rules), -C 0.4 (minimum confidence of 40%), -M 0.2 (minimum support of 20%). However, this execution successfully found "large itemsets and rules."

The "Generated sets of large itemsets" are now much more numerous:

- Size of set of large itemsets L(1): 12
  - {outlook=sunny}, {outlook=overcast}, {outlook=rainy}, {temperature=mild}, {temperature=hot}, {temperature=cool}, {humidity=high}, {humidity=normal}, {windy=TRUE}, {windy=FALSE}, {play=yes}, {play=no}

- Size of set of large itemsets L(2): 9
  - {outlook=sunny}, {play=no}, {outlook=overcast play=yes}, {temperature=mild humidity=high } etc……
- Size of set of large itemsets L(3): 1
  - humidity=normal windy=FALSE play=yes}

Finally, the "Best rules found:" section lists 10 association rules, indicating that the algorithm successfully identified.

- outlook=overcast ➔ play=yes (conf:(1.0))
- temperature=cool ➔ humidity=normal (conf:(0.67))
- humidity=normal windy=FALSE ➔ play=yes (conf:(1.0))
- outlook=sunny ➔ play=no (conf:(0.8))
- windy=FALSE ➔ play=yes (conf:(0.75))
- play=yes ➔ humidity=normal (conf:(0.67))
- play=yes ➔ windy=FALSE (conf:(0.67))
- temperature=hot ➔ play=no (conf:(1.0))
- temperature=mild ➔ humidity=high (conf:(1.0))
- temperature=mild ➔ play=yes (conf:(1.0))

Step-11: Configure the parameters

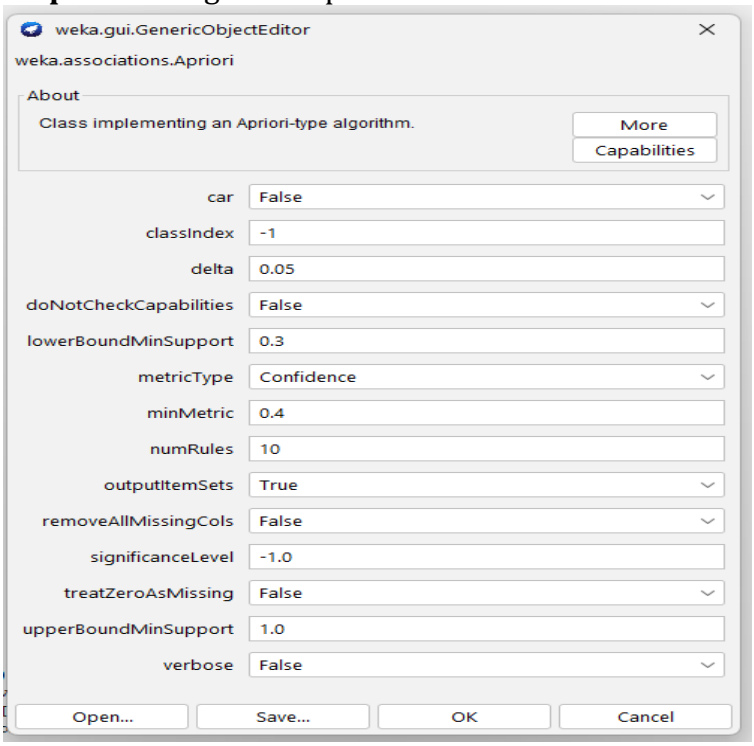


Fig.7.11

This image displays the configuration window for the **Apriori algorithm** within the Weka

data mining software (specifically, weka.associations.Apriori). This step is about setting the detailed parameters that control how the Apriori algorithm will discover association rules from a dataset.

Step-12: Associator Output and Result (min.confidence= 40%, min.support= 30%)

Associator output

```
=== Run information ===

Scheme:       weka.associations.Apriori -I -N 10 -T 0 -C 0.4 -D 0.05 -U 1.0 -M 0.3 -S -1.0 -c -1
Relation:     weather.symbolic
Instances:    14
Attributes:   5
              outlook
              temperature
              humidity
              windy
              play
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.3 (4 instances)
Minimum metric <confidence>: 0.4
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Large Itemsets L(1):
outlook=sunny 5
outlook=overcast 4
outlook=rainy 5
temperature=hot 4
temperature=mild 6
temperature=cool 4
humidity=high 7
humidity=normal 7
windy=TRUE 6
windy=FALSE 8
play=yes 9
play=no 5

Size of set of large itemsets L(2): 9

Large Itemsets L(2):
outlook=overcast play=yes 4
temperature=mild humidity=high 4
temperature=mild play=yes 4
temperature=cool humidity=normal 4
humidity=high windy=FALSE 4
humidity=high play=no 4
humidity=normal windy=FALSE 4
humidity=normal play=yes 6
windy=FALSE play=yes 6

Size of set of large itemsets L(3): 1

Large Itemsets L(3):
humidity=normal windy=FALSE play=yes 4

Best rules found:

 1. outlook=overcast 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
 2. temperature=cool 4 ==> humidity=normal 4    <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
 3. humidity=normal windy=FALSE 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
 4. humidity=normal 7 ==> play=yes 6    <conf:(0.86)> lift:(1.33) lev:(0.11) [1] conv:(1.25)
 5. play=no 5 ==> humidity=high 4    <conf:(0.8)> lift:(1.6) lev:(0.11) [1] conv:(1.25)
 6. windy=FALSE 8 ==> play=yes 6    <conf:(0.75)> lift:(1.17) lev:(0.06) [0] conv:(0.95)
 7. play=yes 9 ==> humidity=normal 6    <conf:(0.67)> lift:(1.33) lev:(0.11) [1] conv:(1.13)
 8. play=yes 9 ==> windy=FALSE 6    <conf:(0.67)> lift:(1.17) lev:(0.06) [0] conv:(0.96)
 9. temperature=mild 6 ==> humidity=high 4    <conf:(0.67)> lift:(1.33) lev:(0.07) [1] conv:(1)
10. temperature=mild 6 ==> play=yes 4    <conf:(0.67)> lift:(1.04) lev:(0.01) [0] conv:(0.71)
```

Fig.7.12

This image displays the "Associator output" section, which provides the results and run information after executing the Apriori association rule mining algorithm.This section presents a summary of the Apriori algorithm's execution and its findings. Specific parameters used for the run: N 10 (maximum 10 rules), -C 0.4 (minimum confidence of 40%), -M 0.2 (minimum support of 20%). However, this execution successfully found "large itemsets and rules."

The "Generated sets of large itemsets" are now much more numerous:

- Size of set of large itemsets L(1): 12
  - {outlook=sunny}, {outlook=overcast}, {outlook=rainy}, {temperature=mild}, {temperature=hot}, {temperature=cool}, {humidity=high}, {humidity=normal}, {windy=TRUE}, {windy=FALSE}, {play=yes}, {play=no}
- Size of set of large itemsets L(2): 9
  - {outlook=sunny}, {play=no}, {outlook=overcast play=yes}, {temperature=mild humidity=high } etc…..

- Size of set of large itemsets L(3): 1
  - {humidity=normal windy=FALSE play=yes}

Finally, the "Best rules found:" section lists the same 10 association rules as the previous output, including examples like:

- outlook=overcast ➔ play=yes (conf:(1.0))
- temperature=cool ➔ humidity=normal (conf:(0.67))
- humidity=normal windy=FALSE ➔ play=yes (conf:(1.0))
- outlook=sunny ➔ play=no (conf:(0.8))
- windy=FALSE ➔ play=yes (conf:(0.75))
- play=yes ➔ humidity=normal (conf:(0.67))
- play=yes ➔ windy=FALSE (conf:(0.67))
- temperature=hot ➔ play=no (conf:(1.0))
- temperature=mild ➔ humidity=high (conf:(1.0))
- temperature=mild ➔ play=yes (conf:(1.0))

Experiment Outcome:

The experiment utilized the Apriori algorithm on the weather.nominal dataset to discover association rules between weather conditions and the likelihood of "play." By progressively lowering the minimum support and confidence thresholds, the analysis transitioned from finding no rules to identifying a single strong rule, and finally, to a comprehensive set of ten rules and their supporting itemsets. This iterative process effectively demonstrated how adjusting parameters influences the discovery of patterns, ultimately revealing key conditions (e.g., "outlook=overcast" or "temperature=hot") that strongly correlate with whether an activity is played or not.