Experiment No:10	TE AI&DS
Date of Performance:	Roll No: 9696
Aim: To analyze and evaluate the performance of different Association mining techniques using WEKA tool	
Related CO5: To analyze and evaluate perform of data mining techniques applied on large dataset using open-source tool for data mining	
Objective: To Demonstrate Association Mining techniques on data sets using WEKA tool.	

Rubrics for assessment of Experiment:

Sr. No	Parameters	Exceed Expectations(EE)	Meet Expectations (ME)	Below Expectations (BE)
1	Timeline (2)	Early or on time (2)	One session late (1)	More than one session late (0)
2	Preparedness (2)	Knows the basic theory related to the experiment very well. (2)	Managed to explain the theory related to the experiment. (1)	Not aware of the theory to the point. (1)
3	Effort (3)	Done expt on their own. (3)	Done expt with help from other. (2)	Just managed. (1)
4	Documentation(2)	Lab experiment is	Documented in	Experiments not

		documented in proper format and maintained neatly. (2)	proper format but some formatting guidelines are missed. (1)	written in proper format (0.5)
5	Result (1)	Specific conclusion.(1)	Partially specific conclusion. (0.5)	Not specific at all. (0)

Assessment Marks:

Timeline(2)	Preparedness(2)	Effort(3)	Documentation(2)	Result(1)	Total(10)

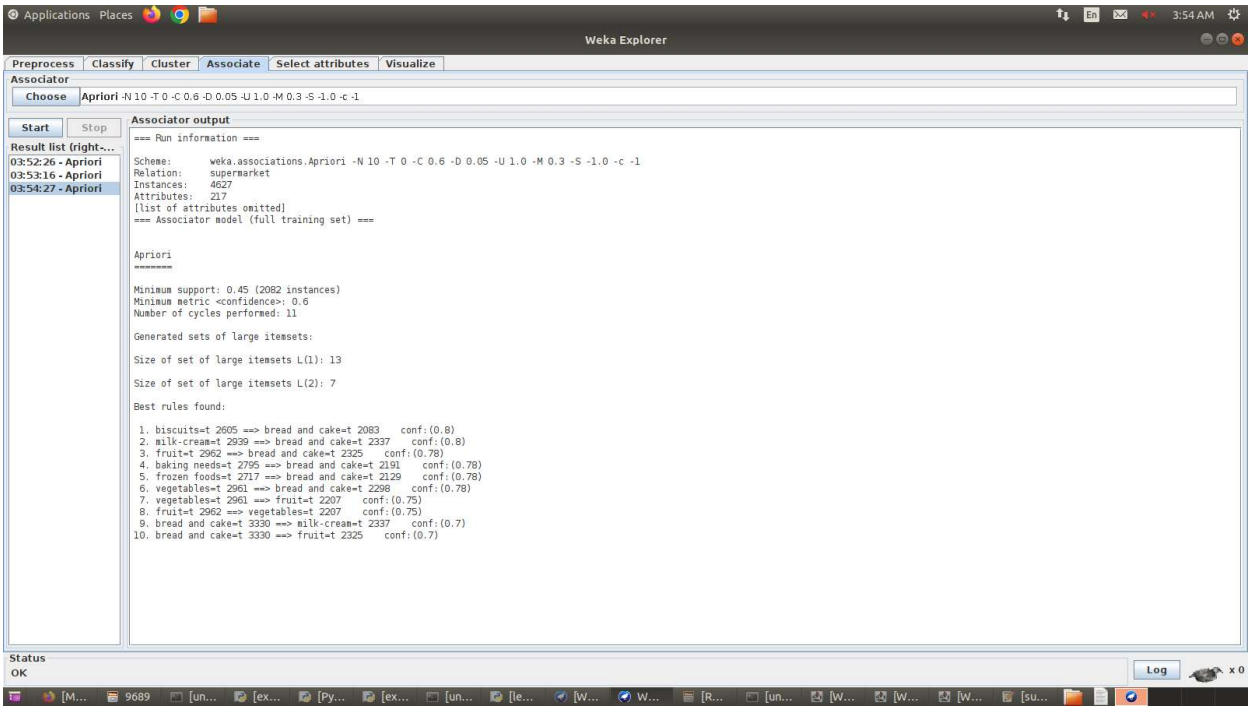
Theory : WEKA contains an implementation of the Apriori algorithm. The algorithm works only with discrete data. It can identify statistical dependencies between groups of attributes. Apriori algorithm can compute all rules that have a given minimum support and exceed a given confidence. Clicking on the "Associate" tab will bring up the interface for the association rule algorithms. The Apriori algorithm which we will use is the default algorithm selected. However, in order to change the parameters for this run (e.g., support, confidence, etc.) we click on the text box immediately to the right of the "Choose" button. Note that this box, at any given time, shows the specific command line arguments that are to be used for the algorithm. WEKA allows the resulting rules to be sorted according to different metrics such as confidence, leverage, and lift. We can also change the default value of rules (10) to be 20; this indicates that the program will report no more than the top 20 rules. The upper bound for minimum support is set to 1.0 (100%) and the lower bound to 0.1 (10%). Apriori in WEKA starts with the upper bound support and incrementally decreases support (by delta increments which by default is set to 0.05 or 5%). The algorithm halts when

either the specified number of rules are generated, or the lower bound for min. support is reached. Once the parameters have been set, the command line text box will show the new command line. We now click on start to run the program. This results in a set of rules. The panel on the left ("Result list") now shows an item indicating the algorithm that was run and the time of the run. You can perform multiple runs in the same session each time with different parameters. Each run will appear as an item in the Result list panel. Clicking on one of the results in this list will bring up the details of the run, including the discovered rules in the right panel. In addition, right-clicking on the result set allows us to save the result buffer into a separate file. Note that the rules were discovered based on the specified threshold values for support and lift. For each rule, the frequency counts for the LHS and RHS of each rule is given, as well as the values for confidence, lift, leverage, and conviction. In most cases, it is sufficient to focus on a combination of support, confidence, and either lift or leverage to quantitatively measure the "quality" of the rule. However, the real value of a rule, in terms of usefulness and action ability is subjective and depends heavily of the particular domain and business objectives.

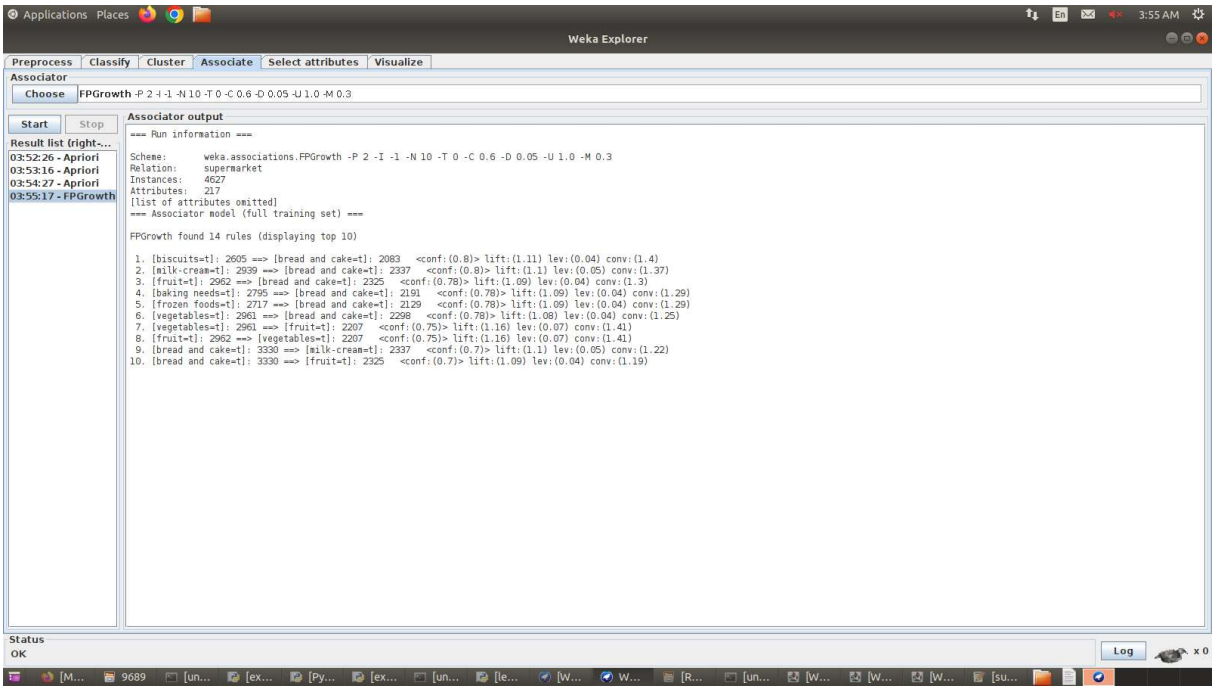
Practical Exercise:

Apply and evaluate the result for different Association Mining techniques on various datasets using WEKA.

1. Apriori Algorithm Output:



2. FP Growth



Postlab:

1) Apply apriori algorithm on the following dataset to extract frequent item sets and extract strong association rules for given :
Minimum support = 50%
confidence = 70%

Trans_id	Items
200	A,B,C
300	A,C
400	A,D
500	B,E,F

2) Give limitations of Apriori algorithms
3) List the name of packages/libraries used in Part 2 of this experiments
4) Apply FP tree Algorithm on above transactional dataset.

NAME:- SHIVPRASAD: GP.
ROLL NO:- 9696.

FR. CONCEICAO RODRIGUES COLLEGE OF ENGINEERING

Experiment 10 POSTLAB

(Q.1)

Itemset	Support count
A	3
B	2
C	2
D	1
E	1
F	1

min Support = $\frac{50}{100} \times 4 = 2$.

∴ reject itemset D, E, F.

Itemset	Support
A, B	1
A, C	2
B, C	1

Only itemset A, C satisfies min support count.

∴ Association Rules.

Rule	Support	Confidence	%
A ⇒ C	2	2/3 = 0.66	66%
C ⇒ A	2	2/2 = 1	100%

⇒ min confidence = 70%.

So valid rule: C ⇒ A.

Q.2).

→ Limitations

1) Computational complexity
2) Inability to handle numeric data
3) Biased minimum support threshold.
4) Differently handling sparse data.
5). Higher memory usage
6) Time and space overhead
7) Limited discovery of complex patterns
8) Lack of interpretation of context.

Q.3)

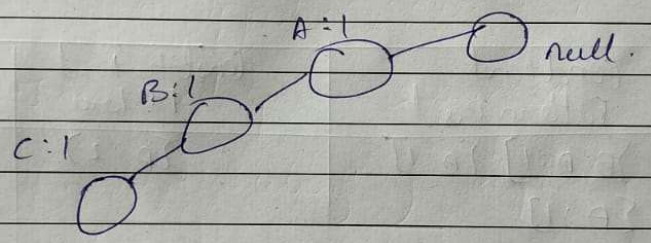
→ NAME of Packages / Library used

- apyori - Library for association rule mining using the Apriori algorithm. Can find association, frequent itemsets and generate rules from transaction data.

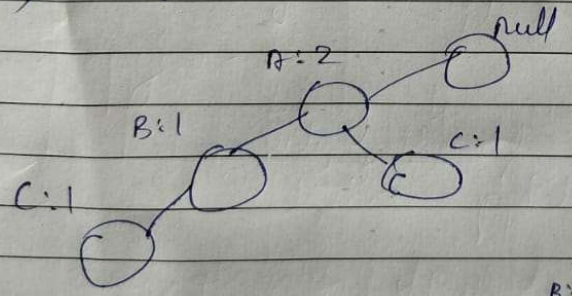
Q.4)

Itemset	Support
A	3
B	2
C	2
D	1
E	1
F	1

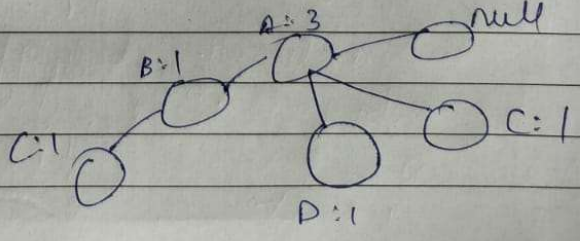
i) Tid (200) = A, B, C.

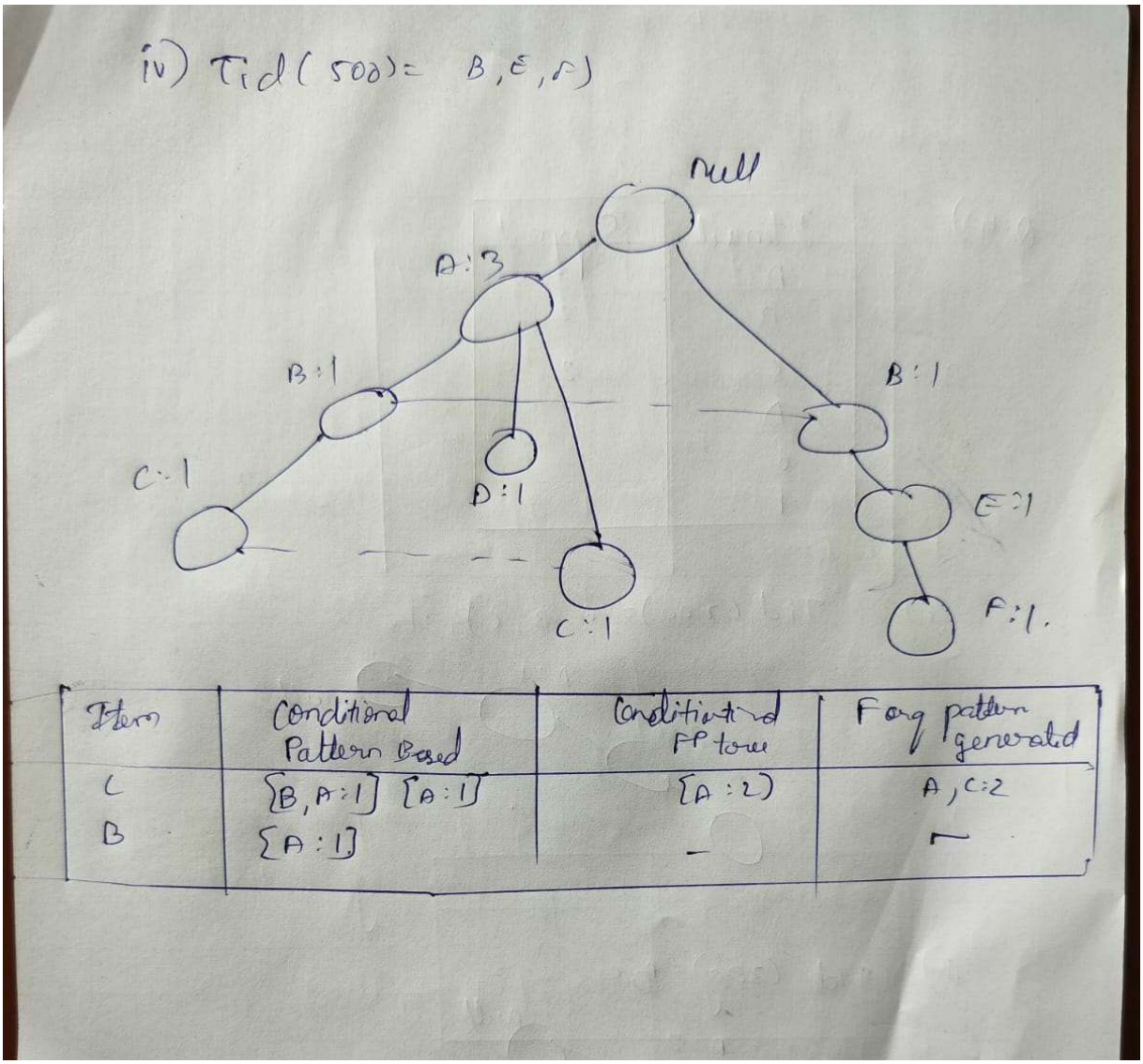


ii) Tid (300) = A, C.



iii) Tid (400) = D, D.





Item	Conditional Pattern Based	Conditional FP tree	Freq pattern generated
C	{B, A:1} {A:1}	{A:2}	A, C:2
B	{A:1}	-	-

Conclusion: I understood the algorithm and am able to implement the algorithm using WEKA tool.