

School of Computer Science Engineering and Technology

Course- BTech

Type- Specialization Core II

Course Code- CSET228

Course Name- Data Mining and Predictive Modelling

Year- 2024-25

Semester- Even

Date-22-01-2025

Batch- IV Semester (All)

Lab .5.1

CO Mapping

Exp No	Name	CO1	CO2	CO3
1	Association rule mining and implementation of the Apriori algorithm	✓		

- 1.** Do the following operations on given item dataset: (25 Minutes)
 - a) Preprocess the data using mlxtend library and list the products using apply function.
 - b) Using TransactionEncoder function transform the list and shows the obtained dataframes.
- 2.** Import apriori and association rules to do the following operations: (25 Minutes)
 - a) Using apriori, find the frequent items when minimum support is 0.20 at least.
 - b) Find the association rule on all items when min support and confidence is 50%.
 - c) Using the lift metric and minimum threshold = 0.65, find the antecedent, consequent supports for each item.
- 3.** Use the **retail_dataset.csv** for doing the following tasks. (40 Minutes)
 - a) To be able to use Apriori algorithm first convert the dataset into binary format using one hot encoding technique. For doing it find out the common items in the dataset and using iterative statements set the uncommon items as 0 and the common items as 1.
 - b) Apply the apriori algorithm to the modified dataset with the following attribute

values. The minimum support should be 0.2. Print the frequent itemsets with their supports.

	support	itemsets
0	0.869841	(nan)
1	0.501587	(Milk)
2	0.425397	(Bagel)
3	0.476190	(Meat)
4	0.501587	(Cheese)
5	0.438095	(Eggs)
6	0.438095	(Wine)

- c) Next use the association rule mining using appropriate library in python to generate association rules. The minimum threshold should be 0.6. Print the antecedents and the consequents.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Milk)	(nan)	0.501587	0.869841	0.409524	0.816456	0.938626	-0.026778	0.709141
1	(Bagel)	(nan)	0.425397	0.869841	0.336508	0.791045	0.909413	-0.033520	0.622902
2	(Meat)	(nan)	0.476190	0.869841	0.368254	0.773333	0.889051	-0.045956	0.574230
3	(Cheese)	(nan)	0.501587	0.869841	0.393651	0.784810	0.902245	-0.042651	0.604855
4	(Eggs)	(nan)	0.438095	0.869841	0.336508	0.768116	0.883053	-0.044565	0.561310

- a. Scatter plot the support vs confidence using the data.

