

Suppose Mora Supermarket has four locations and has transaction data for customers. Propose an outline of an algorithm to mine the association rules without having to copy/transfer the data to one location.

[Font family](#)

[Font size](#)

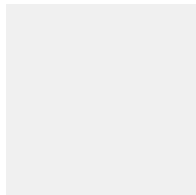
[Paragraph](#)



Path: [p](#)

## Question 2

Answer saved  
Marked out of 5.00



Flag question

### Question text

A customer transaction database contains the following transactions.

Cust_ID	TID	Items bought (in the form brand-item_category)
01	T100	{Elephant-Milk, Cargils-Meat, Cargils-Bread, Keels-Butter}
02	T200	{Nestle-Milk, Keels-Bread, Cargils-butter}
01	T300	{Keels-Bread, Nestle-Milk, Elephant-Water}

03	T400	{Nestle-Milk, Keels-Bread, Nestle-butter, Elephant-Water}
----	------	---

At the granularity of the item\_category (e.g. item could be Milk) find one rule with 60% support and 100% confidence with the following rule template.

Give your answer in the form of

Item<sub>1</sub>, Item<sub>2</sub>, Item<sub>3</sub>

i.e <item\_1><comma><space><item\_2><comma><space><item\_3>

for eg milk, butter, bread

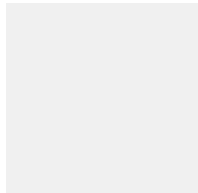
Answer:

milk, bread, butter

## Question 3

Answer saved

Marked out of 5.00



Flag question

### Question text

In data mining, the curse of dimensionality and the curse of cardinality are major challenges. Out of the two challenges which one is harder to handle? Briefly explain your answer.

[Font family](#)

[Font size](#)

[Paragraph](#)

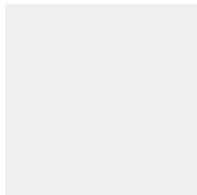


Path: [p](#)

## Question 4

Answer saved

Marked out of 1.00



Flag question

### Question text

What would be a good approach to find arbitrary shaped clusters?

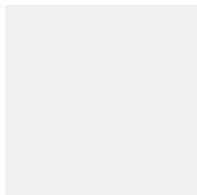
Select one or more:

- ☒ a. Density based clustering method such as OPTICS
- ☐ b. Partition based clustering method such as K-Means clustering
- ☐ c. Partition based clustering method such as PAM
- ☐ d. Partition based clustering method such as CLARANS
- ☒ e. Density based clustering method such as DENCLU

## Question 5

Answer saved

Marked out of 1.00



Flag question

### Question text

Following data is collected for 3 patients. Assuming all variables are symmetric and using a suitable binary distance metric without any weighting identify the patient that is most similar to P1

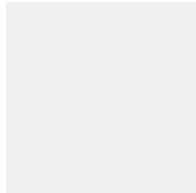
Attribute	Gender	Fever	Cough	T1	T2	T3	T4	T5	T6
P1	M	Y	P	N	N	N	P	N	P
P2	F	Y	P	N	P	N	P	P	P
P3	M	Y	P	N	N	N	N	P	P

Select one:

- ☐ a. P2
- ☒ b. P3

### Question 6

Answer saved  
Marked out of 1.00



Flag question

### Question text

What can we do with Decision Tree Induction?

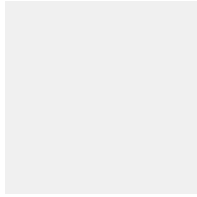
Select one or more:

- ☒ a. Classification
- ☐ b. Clustering
- ☐ c. Data Pre-processing: fill missing values
- ☒ d. Data Pre-processing: attribute relevance
- ☐ e. Data Normalization

### Question 7

Answer saved

Marked out of 1.00



Flag question

### Question text

What can be used to automatic concept hierarchy generation ?

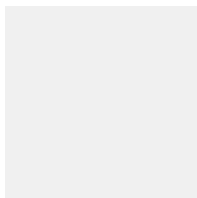
Select one:

- ☒ a. Hierarchical Clustering
- ☐ b. Analysing the number of distinct Values in the attributes
- ☐ c. Neural Networks
- ☐ d. Z-Score Normalization

## Question 8

Answer saved

Marked out of 1.00



Flag question

### Question text

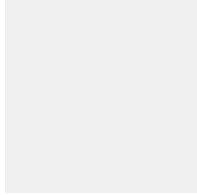
What are the characteristics that we use to measure the data quality?

Select one or more:

- ☐ a. Accuracy
- ☒ b. Timeliness
- ☒ c. Consistency
- ☒ d. Completeness

## Question 9

Answer saved  
Marked out of 1.00



Flag question

### Question text

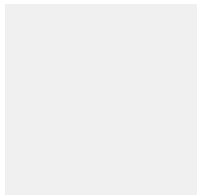
What is NOT considered as an Association Rule Mining algorithm?

Select one:

- ☐ a. Apriori
- ☐ b. FPGrowth
- ☐ c. VIPER
- ☒ d. DBScan
- ☐ e. Eclat

## Question 10

Answer saved  
Marked out of 1.00



Remove flag

### Question text

What types of patterns can be identified from Association Rule Mining

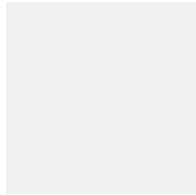
Select one or more:

- ☒ a. What products are purchased together in a Supermarket
- ☒ b. What would be the subsequent purchase after buying a mobile phone

- ☒ c. What would be the next web click after visiting set of links.
- ☒ d. What DNA sequences are sensitive to a given drug
- ☐ e. Clustering Call Detail Records to find moving users.

## Question 11

Answer saved  
Marked out of 1.00



Remove flag

### Question text

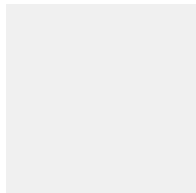
In the Apriori Algorithm why do we start looking for 1-item frequent item sets?

Select one:

- ☒ a. To take advantage of the downward closure property
- ☐ b. To take advantage of the upward closure property
- ☐ c. Because there are only a few 1-item frequent item sets
- ☐ d. 1-item set frequent patterns are easy to identify
- ☐ e. All 1-item set frequent patterns can be identified with one single scan of the database

## Question 12

Answer saved  
Marked out of 1.00



Flag question

### Question text

In the Decision Tree algorithm for what do we use Information Gain?

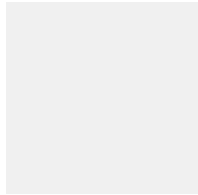
Select one:

- ☒ a. To find the attribute that can be used to split the data based on the class label
- ☐ b. To find the unknown class labels in the test dataset
- ☐ c. To fill missing values and remove noise from the data
- ☐ d. To find association rules among the attributes in the data
- ☐ e. To find non-spherical clusters in the data set.

## Question 13

Answer saved

Marked out of 5.00



Flag question

### Question text

Find the attribute that would become the first attribute on Decision Tree model based on Information Gain. You may use the formulae given below.

Select one:

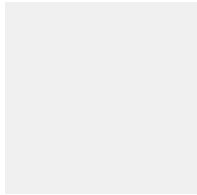
- ☒ a. Age
- ☐ b. Income



- ☐ c. Student
- ☐ d. Credit Rating
- ☐ e. Buys Computer

## Question 14

Answer saved  
Marked out of 1.00



Flag question

### Question text

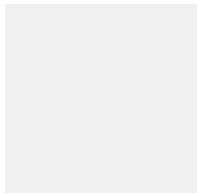
What could be a possible symptom of overfitting on Decision Tree?

Select one:

- ☒ a. Too many branches in the tree
- ☐ b. Test data results are showing very good accuracy
- ☐ c. Training data results are showing very good accuracy
- ☐ d. Leaf level nodes contain mixed collection of classes
- ☐ e. Leaf level nodes contain single classes

## Question 15

Answer saved  
Marked out of 5.00



Flag question

### Question text

Using Naive Bayes classification find out if a 25 year old, medium income student with fair credit rating would buy a computer. You may use the formulae given below.

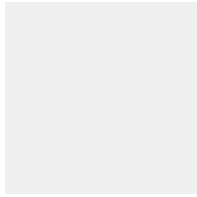
Select one:

- ☒ a. Yes
- ☐ b. no
- ☐ c. Can not calculate

## Question 16

Answer saved

Marked out of 1.00



Flag question

### Question text

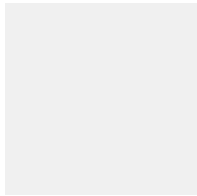
Using Manhaten Distance wh,at would be the distance beetween (1,0,5) and 5,6,7)?

Answer:

## Question 17

Answer saved

Marked out of 1.00



Remove flag

### Question text

What is a major differance between a Density based Clustering method and a Partition baed clustering method.

Select one:

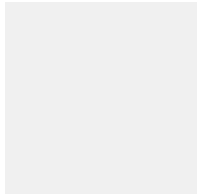
- ☐ a. Shape of the the clusters that are generated
- ☐ b. Computational Cost

- ☐ c. Supervised Vs Un-supervised learning
- ☐ d. The type of data that can be handled
- ☒ e. Accuracy of Density based methods is better

## Question 18

Answer saved

Marked out of 1.00



Flag question

### Question text

K-Medoid clustering is computationally more expensive than a K-means clustering algorithm.

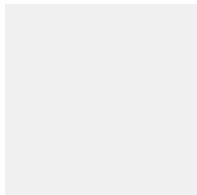
Select one:

- ☒ True
- ☐ False

## Question 19

Answer saved

Marked out of 1.00



Flag question

### Question text

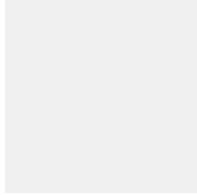
How many 3D cuboids can be created from a data cube with 4 dimensions?

Answer:

## Question 20

Answer saved

Marked out of 1.00



Flag question

### Question text

Example(s) for data cube operation(s) is/are ?

Select one or more:

- ☒ a. roll-up
- ☒ b. Drill-Down
- ☒ c. Slice
- ☒ d. Dice
- ☐ e. Cut