Data Mining

<u>Assignment 1</u>

All Questions are of 1 mark.

1. Which of the following is usually the last step in the data mining process?
a) Visualization
b) Preprocessing
c) Modelling
d) Deployment

Ans: d

Explanation: The last step in the data mining process is to deploy the models to a production environment. Deployment is important because it makes the models available to users for use.

2. Sales database of items in a supermarket can be considered as an example of:
a) Record data
b) Ordered data
c) Graph data
d) None of the above

Ans: a

Explanation: The most basic form of record data has no explicit relationship among records or data fields, and every record (object) has the same set of attributes. Record data is usually stored either in flat files or in relational databases.

3. HTML links are an example of:
a) Record data
b) Ordered data
c) Graph data
d) None of the above

Ans: c

Explanation: HTML links are an example of graph data.

4. Name of a place, can be considered an attribute of type?
a) Nominal
b) Ordinal
c) Interval
d) Ratio

Ans: a

Explanation: Nominal-related to names. The values of a Nominal attribute are name of things, some kind of symbols. There is no order (rank, position) among values of nominal attribute.

5. A store sells 10 items. Maximum possible number of candidate 3-itemsets is:
a) 120
b) 6
c) 15

d) 56

Ans: a

Explanation: Number of ways of choosing 3 items from 10 items is 10C3 = 120

6. If a record data matrix has reduced number of columns after a transformation, the transformation has performed:
a) Data Sampling
b) Dimensionality Reduction
c) Noise Cleaning
d) Discretization

Ans: b

Explanation: Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principal variables.

Answer Q7-Q10 based on the following table:

Transaction ID	Itemsets
1	{1, 2, 4, 5}
2	{2, 3, 5}
3	{1, 2, 4, 5}
4	{1, 2, 3, 5}
5	{1, 2, 3, 4, 5}
6	{2, 3, 4}

7. Support of rule {4,5} -> {1}
a) 1
b) 0.5
c) 0.25
d) 0

Ans: b

Explanation: support of X -> Y is support({X,Y})/|T| = 3/6 = 0.5.

8. confidence of rule {4,5} -> {1} is:
a) 1
b) 0.5
c) 0.25
d) 0.75

Ans: a

Explanation: Confidence measures the occurrence of products together in a dataset.

Confidence(X->Y) = support({X,Y})/support({X})=(3/6)/(3/6)=1.

9. Support of {1} -> {2,5} is:
a) 2/3
b) 2/2
c) 1/4
d) 3/4

Ans: a

Explanation: support of X -> Y is support({X,Y})/|T| = 4/6 = 2/3.

10. Confidence of {1} -> {2,5}
a) 2/3
b) 1
c) 0
d) 0.5

Ans: b

Explanation: Confidence measures the occurrence of products together in a dataset.

Confidence(X->Y) = support({X,Y})/support({X})=(4/6)/(4/6)=1.