63/2 (SEM–4) MCA 4.2

2024

COMPUTER SCIENCE AND TECHNOLOGY

Paper : MCA 4.2

(Data Mining and Warehousing)

Full Marks : 75

Pass Marks : 30

Time : Three hours

The figures in the margin indicate full marks for the questions.

1. Choose the correct answer : 1×5=5

   (a) _____ predicts future trends and behaviour allowing business managers to make proactive knowledge-driven decision.

   (i) Data warehouse

   (ii) Data mining

   (iii) Meta data

   (iv) Data Mart

(b) A operational system in which of the following

  (i) A system that is used to run the business in real-time and is based on historical data.

  (ii) A system that is used to run the business in real-time and is based on current data.

  (iii) A system that is used to support decision-making and based on current data.

  (iv) A system that is used to support decision-making and is based on historical data.

(c) From where are classification rules extracted ?

  (i) Branches

  (ii) Decision tree

  (iii) Siblings

  (iv) Root node

(d) Classification is

  (i) a sub-division of a set of examples into a number of classes

  (ii) a measure of the accuracy of the classification of a concept that is given by a certain theory

  (iii) the task of assigning a classification to a set of examples

  (iv) None of the above

(e) Euclidean distance measure is

  (i) a stage of the KDD process in which new data is added to the existing selection

  (ii) the process of finding a solution for a problem simply by enumerating all possible solutions accoring to some predefined order and then testing them

  (iii) the distance between two points as calculated using the Pythagoras theorem

  (iv) None of the above

2. Answer the following questions : *(any five)*

                                    2×5=10

(a) Differentiate between base cuboid and apex cuboid.

(b) What is concept hierarchy ?

(c) List out the major issues in data mining.

(d) Why organizations consider data warehousing a critical need?

(e) Explain Association rule in mathematical notations.

(f) Define nominal, ordinal and ratio scaled variables.

(g) Differentiate agglomerative and Divisive hierarchical clustering algorithm.

3. Answer the following questions : *(any six)*

$5 \times 6 = 30$

(a) What are the advantages and disadvantages of data warehouse system?

(b) Write the difference between OLAP and OLTP.

(c) Explain Metadata repository.

(d) Write in brief about schemas syntax in multidimensional data model.

(e) Write *any one* partitioning based clustering method.

(f) Write DBSCAN clustering algorithm and estimate time and space complxity.

(g) Describe the various phases in KDD with diagram.

(h) What is the importance of data marts in data warehouse?

4. Answer the following questions : *(any three)*

$10 \times 3 = 30$

(a) Define data warehouse. List the application area of data warehouse system. Explain in detail about multitier architecture of data warehousing system with a neat diagram.

(b) Define multidimensional data model, fact table and dimension table with example. Explain the OLAP operation in multidimensional data model.

(c) Define support and confidence in association rule mining. A database has five transactions. Let min sup = 60% and minc of = 80%.

TID	Item bought
T_{100}	{M, O, N, K, E, Y}
T_{200}	{D, O, N, K, E, Y}
T_{300}	{M, A, K, E}
T_{400}	{M, U, C, K, Y}
T_{500}	{C, O, O, K, I, E}

(i) Find all frequent item sets using apriori algorithm.

(ii) List all the strong association rules.

(d) What is the main objective of clustering? Describe briefly about the requirements of clustering. Compare K-means with K-medoids algorithms for clustering.