END TERM EXAMINATION

SIXTH SEMESTER [BCA] JULY 2023

Paper Code: BCA-302   Subject: Data Warehouse and Data Mining

Time: 3 Hours   Maximum Marks: 75

Note: Attempt five questions in all including Q.No. 1 which is compulsory. Select one question from each unit.

Q1 Answer any **ten** of the following.   (2.5x10=25)
- a) A dimension table is wide; the fact table is deep. Justify.
- b) Define the term support and confidence used in association rule mining
- c) What is meant by spatial database?
- d) Why data transformation is essential in the process of knowledge discovery?
- e) List and describe the five primitives for specifying a data mining task.
- f) Data warehouse is the only viable means to resolve the information crisis and to provide strategic information. Justify
- g) What is meant by slice and dice in context of OLAP, with example?
- h) How operational system makes the wheels of the business turn? Comment.
- i) Compare OLTP and OLAP systems.
- j) Define support and confidence
- k) What is meta data? Explain the use of meta data in data warehouse?
- l) Differentiate between supervised and unsupervised learning.

UNIT-I

Q2 a) Explain the steps involved in data mining when viewed as a process in knowledge discovery.   (6.5)
   b) Imaging that you need to analyze 'All Electronics' sales and customer data (Data related to the sales of electronic items). You note that many tuples have no recorded value for several attributes such as customer income. How can you go about filling in the missing values for this attribute? Explain some of the methods to handle the problem.   (6)

OR

Q3 a) Discuss data integration and data transformation in detail.   (6.5)
   c) What are the major issues and challenges faced in data mining? How do these issues impact the accuracy, efficiency, and reliability of the results obtained from data mining?   (6)

P.T.O.

UNIT-II

Q4 a) Draw and explain various components of three tier data warehouse architecture. (6.5)

b) What is OLAP (Online Analytical Processing) and how does it relate to data warehousing? How does OLAP enable interactive and multidimensional analysis of data stored in a data warehouse? (6)

OR

Q5 Consider the following transactional data. Find frequent item sets using **Apriori and FP growth**. Apply the association rule mining to get the association rules with min support of 2 and confidence of 50%. (12.5)

T.ID.	List of Items IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

UNIT-III

Q6 a) What is cluster analysis? Discuss major clustering methods. (6.5)

b) What is prediction accuracy and how is it measured in classification and prediction tasks? Can you explain different accuracy measures, such as accuracy, precision, recall, F-measure, and ROC curve? How do these measures help in evaluating the performance of classification models? (6)

OR

Q7 a) What is classification by back propagation and how does it apply to neural networks? Can you explain the concept of feed-forward neural networks and the training process using back propagation algorithm? What are the benefits and challenges of using neural networks for classification tasks? (6.5)

b) Can you explain the concept of partitioning methods in cluster analysis? How do algorithms like k-means, k-medoids, and CLARA divide the dataset into distinct partitions or clusters based on similarity or distance measures? What are the advantages and limitations of partitioning methods? (6)

P.T.O.

UNIT-IV

Q8   a)   How does data mining enable multidimensional analysis and descriptive mining of complex data objects? Can you explain how data mining techniques can be applied to analyze data with multiple dimensions, such as time, geography, and product, and uncover meaningful patterns and insights?   **(6.5)**

b)   What is the process of mining spatial databases? Can you discuss how spatial data, such as geographic coordinates and spatial relationships, can be analyzed using data mining techniques to discover patterns, trends, and relationships in spatial datasets? What are some common algorithms and methods used for mining spatial databases?   **(6)**

OR

Q9   a)   What are some real-world applications of data mining across various industries and domains? Can you provide examples of how data mining has been successfully applied in areas such as finance, healthcare, marketing, fraud detection, and customer relationship management? How has data mining contributed to decision-making and improving business processes in these applications?   **(6.5)**

b)   How does data mining apply to multimedia databases? Can you explain how data mining techniques can be used to analyze multimedia data, including images, videos, and audio files, and extract meaningful patterns and knowledge from these complex data types? What are the challenges and opportunities in mining multimedia databases?   https://www.ggsipuonline.com   **(6)**

****