

PSG COLLEGE OF ARTS & SCIENCE  
(AUTONOMOUS)

MSc DEGREE EXAMINATION DECEMBER 2023  
(First Semester)

Branch - **COMPUTER SCIENCE**  
**DATA MINING AND ANALYTICS**

Time: Three Hours

Maximum: 75 Marks

**SECTION-A (10 Marks)**

Answer ALL questions

ALL questions carry EQUAL marks

(10 × 1 = 10)

Module No.	Question No.	Question	K Level	CO
1	1	Data mining is a) The actual discovery phase of a knowledge discovery process b) The stage of selecting the right data for a KDD process c) A subject-oriented integrated time variant non-volatile collection of data in support of management d) None of these	K1	CO1
	2	Some telecommunication company wants to segment their customers into distinct groups in order to send appropriate subscription offers, this is an example of a) Supervised learning      b) Data extraction c) Serration                      d) Unsupervised learning	K2	CO2
2	3	You are given data about seismic activity in Japan, and you want to predict a magnitude of the next earthquake, this is in an example of a) Supervised learning      b) Unsupervised learning c) Serration                      d) Dimensionality reduction	K1	CO1
	4	Which of the following is not involve in data mining? a) Knowledge extraction      b) Data archaeology c) Data exploration              d) Data transformation	K2	CO2
3	5	Data set {brown, black, blue, green, red} is example of Select one: a) Continuous attribute      b) Ordinal attribute c) Numeric attribute              d) Nominal attribute	K1	CO1
	6	To detect fraudulent usage of credit cards, the following data mining task should be used Select one: a) Outlier analysis              b) Prediction c) Association analysis          d) Feature selection	K2	CO2
4	7	The number of item sets of cardinality 4 from the items lists {A, B, C, D, E} Select one: a) 2                                      b) 10 c) 20                                      d) 5	K1	CO1
	8	Classification is a) A subdivision of a set of examples into a number of classes b) A measure of the accuracy, of the classification of a concept that is given by a certain theory c) The task of assigning a classification to a set of examples d) None of these	K2	CO2

Cont...

5	9	For what purpose, the analysis tools pre-compute the summaries of the huge amount of data? a) In order to maintain consistency b) For authentication c) For data access d) To obtain the queries response	K1	CO1
	10	Which one of the following can be considered as the final output of the hierarchal type of clustering? a) A tree which displays how the close thing are to each other b) Assignment of each point to clusters c) Finalize estimation of cluster centroids d) None of the above	K2	CO2

**SECTION - B (35 Marks)**

Answer ALL questions

ALL questions carry EQUAL Marks

(5 × 7 = 35)

Module No.	Question No.	Question	K Level	CO
1	11.a.	Name methods of classification methods and explain any two.	K3	CO3
	(OR)			
	11.b.	Elucidate steps involved in data mining knowledge process?		
2	12.a.	Using example, write notes on Naive Bayes Algorithm.	K4	CO4
	(OR)			
	12.b.	Describe Star Schema and Snow Flake Schema.		
3	13.a.	Explain the different types of data repositories on which mining can be performed?	K3	CO3
	(OR)			
	13.b.	Describe multidimensional data model in detail.		
4	14.a.	Discuss how data warehousing is used in retail and telecommunication industry.	K4	CO4
	(OR)			
	14.b.	How data mining systems are classified? Discuss each classification with an example.		
5	15.a.	Explain the various data preprocessing tasks in data mining.	K3	CO3
	(OR)			
	15.b.	Explain as to how neural networks are used for classification of data.		

Cont...

**SECTION -C (30 Marks)**

Answer ANY THREE questions

ALL questions carry EQUAL Marks

(3 × 10 = 30)

Module No.	Question No.	Question	K Level	CO
1	16	Illustrate warehouse Architecture and its seven components.	K3	CO3
2	17	List out the OLAP operations in multidimensional data model and explain with an example.	K4	CO4
3	18	How to reduce the size of the fact table? Explain with an example.	K3	CO3
4	19	<p>Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit.</p> <p>(a) Enumerate three classes of schemas that are popularly used for modeling data warehouses.</p> <p>(b) Draw a schema diagram for the above data warehouse using one of the schema classes listed in</p> <p>(c) Starting with the base cuboid [day, doctor, patient], what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2004?</p> <p>(d) To obtain the same list, write an SQL query assuming the data are stored in a relational database with the schema fee (day, month, year, doctor, hospital, patient, count, charge).</p>	K4	CO4
5	20	Discuss about mining association rules using the apriori algorithm in detail.	K3	CO3

Z-Z-Z

END