Question Bank

Francis Xavier Engineering College
(An Autonomous Institution)
Department of Computer Science and Engineering
Year/Sem : IV/VII
Degree/Branch : B.E. – Computer Science and Engineering
19CS7701 DATA MINING

Question Bank – Unit 1

Part A

Q.No.	Question	CO-K Level	PO-PI Code
1.	What is the primary purpose of building the multidimensional model?	CO1-K2	4.6.3
2.	What is the need of data warehouses?	CO1-K2	4.6.3
3.	How can you tell the difference between fact and dimension tables?	CO1-K4	4.4.2
4.	How is multidimensional data model used in data warehouse?	CO1-K1	4.6.4
5.	Differentiate between a data warehouse and a data mart.	CO1-K4	4.6.2
6.	Give the differences between a database and a data warehouse.	CO1-K4	4.4.2

Part B

Q.No.	Question	CO-K Level	PO-PI Code
1.	Summarize the various OLAP operations in the Multidimensional Data Model.	CO1-K1	4.4.1
2.	Explain multidimensional data model with a neat diagram.	(13)	4.6.3
3. i)	Differentiate between star schema and snowflake schema in the context of data warehousing.	CO1-K4	4.6.2

3. ii)	Define metadata and explain the types of metadata.	CO1-K2	4.4.1
4.	Compare OLAP vs OLTP and explain what will happen if we use the same database for both OLAP and OLTP?	CO1-K4	4.6.2
5.	Discuss about the purpose of various Data warehousing Components with the diagram.	CO1-K2	4.6.3

Question Bank – Unit 2

Part A

Q.No.	Question	CO-K Level	PO-PI Code
1.	What are major issues that will be faced in data warehouse implementation?	CO2-K2	4.6.3
2.	What are the challenges in ETL process?	CO2-K4	4.6.2
3.	What is the difference between ROLAP and MOLAP?	CO2-K2	4.6.3
4.	Why is it that tuning is difficult in a data warehouse?	CO2-K2	4.5.1
5.	What are the major applications of data warehousing?	CO2-K1	4.6.3
6.	What are the three tiers of data warehouse architecture?	CO2-K1	4.4.1

Part B

Q.No.	Question	CO-K Level	PO-PI Code
1.	How does data warehouse contribute to business intelligence?	CO2-K2	4.6.1

2.	What do you understand by data staging and ETL explain with suitable structure?	CO2-K2	4.5.1
3.	Define data warehouse. Draw the architecture of data warehouse and explain the three tiers in detail.	CO2-K4	4.5.1
4.	What are the challenges in tuning the data warehouse? How do you optimize a data warehouse?	CO2-K4	4.4.3
5.	Explain in detail about the implementation of a data warehousing.	CO2-K2	4.6.3

Question Bank – Unit 3

Part A

Q.No.	Question	CO-K Level	PO-PI Code
1.	Why is feature selection important in data mining?	CO3-K4	4.6.2
2.	What do you understand by Data Mining?	CO3-K2	4.4.1
3.	What are the major issues in data mining?	CO3-K4	4.6.2
4.	What is data mining query languages?	CO3-K1	4.6.2
5.	Why pre-processing is needed in data mining?	CO3-K4	4.6.2
6.	Define association and correlations.	CO3-K1	4.1.1

Part B

Q.No.	Question	CO-K Level	PO-PI Code
1.	How does data mining work? Discuss the different stages in data mining process.	CO3-K4	3.6.3
2.	Discuss in detail about Mining frequent patterns.	CO3-K2	4.6.2
3.	What is data Pre-processing? Explain the various data pre-processing techniques.	CO3-K2	3.4.2
4.	Generalize and discuss about association rule mining with examples and state how association mining to correlation analysis is dealt with.	CO3-K2	2.4.4
5.	Generalize and discuss about association rule mining with examples and state how association mining to correlation analysis is dealt with.	CO3-K2	2.4.4

Question Bank – Unit 4

Part A

Q.No.	Question	CO-K Level	PO-PI Code
1.	List the major clustering methods.	CO4-K1	4.1.2
2.	List the major classification methods.	CO4-K1	4.1.2
3.	Illustrate support vector machine with example.	CO4-K3	4.3.3
4.	What are ways of reducing dimensionality?	CO4-K1	2.2.3
5.	Define Lazy learners with an example.	CO4-K1	4.1.1
6.	List the major clustering methods.	CO4-K1	4.1.2

Part B

Q.No.	Question	CO-K	PO-

		Level	PI Code
1.	Describe in detail about the following Classification methods. ● Bayesian classification ● Classification by Back propagation.	CO4-K 2	2.4.4
2.	What is grid-based clustering? With an example explain an algorithm for grid-based clustering.	CO4-K2	3.2.2
3.	Define classification? With an example explain how support vector machines can be used for classification.	CO4-K 2	4.2.1
4.	Explain the following: ● Hierarchical based method. ● Density based methods.	CO4-K2	4.2.1
5.	What is outlier mining important? Briefly describe the different approaches behind statistical–based outlier detection, distance-based outlier detection and deviation-based outlier detection.	CO4-K4	2.4.4

Question Bank – Unit 5

Part A

Q.No.	Question	CO-K Level	PO-PI Code
1.	What is temporal database? Explain temporal database with example.	CO5-K4	3.1.5
2.	What are the advantages of spatial database?	CO5-K2	3.1.1
3.	Explain about text mining and discuss about the challenges in text mining.	CO5-K1	2.4.3
4.	What are the advantages and disadvantages of logistic regression?	CO5-K4	3.1.1
5.	How does temporal database differ from regular database?	CO5-K4	2.2.4
6.	What is the difference between time series and sequential data?	CO5-K4	2.2.4

Part B

Q.No.	Question	CO-K	PO-

		Level	PI Code
1.	Discuss about spatial databases with example.	CO5-K2	4.2.1
2.	What is temporal database? Explain temporal database with example.	CO5-K2	4.2.1
3.	What is Multimedia database? Explain about multimedia database with example.	CO5-K2	4.2.1
4.	What is web mining? Explain the various types of web mining methods.	CO5-K2	4.2.1
5.	Explain about text mining and discuss about the challenges in text mining.	CO5-K4	2.4.3

PART – C

Q.No.	Question	CO-K Level	PO-PI Code
1. i)	Analyze and elaborate the current trends in data mining in the following fields. i. Financial data analysis ii. Biological data analysis iii. Intrusion detection	K4	2.2.4
1. ii)	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. Draw a schema diagram for the above data warehouse using one of the schemas. [star, snowflake, fact constellation]	K4	1.3.1
2.	Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9) Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2). The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as- P(a, b) = |x2 – x1| + |y2 – y1|	CO4-K3	1.3.1

<table>
<tr><td></td><td>Use K-Means Algorithm to find the three cluster centers after the second iteration.</td><td></td><td></td></tr>
<tr><td>3.</td><td>Explain and Apply the Apriori algorithm for discovering frequent item sets of the table.
<table><tr><td>Trans ID</td><td>Items Purchased</td></tr><tr><td>T1</td><td>I1, I2, I5</td></tr><tr><td>T2</td><td>I2, I4</td></tr><tr><td>T3</td><td>I2, I3</td></tr><tr><td>T4</td><td>I1, I2, I4</td></tr><tr><td>T5</td><td>I1, I3</td></tr><tr><td>T6</td><td>I2, I3</td></tr><tr><td>T7</td><td>I1, I3</td></tr><tr><td>T8</td><td>I1, I2, I3, I5</td></tr><tr><td>T9</td><td>I1, I2, I3</td></tr></table>minimum support count is 2 and minimum confidence</td><td>CO3-K3</td><td>1.3.1</td></tr>
</table>