



VIGNAN INSTITUTE OF TECHNOLOGY AND SCIENCE

Near Ramoji Film City, Deshmukhi (V), Yadadri Bhuvanagiri Dist, Telangana - 508 284.

Approved by AICTE, New Delhi, Affiliated to JNTU Hyderabad

EAMCET CODE : VGNT



Department of Computer Science & Engineering(Data Science)

Question Bank

III Year I Semester – 2022-23

DATA MINING

UNIT -1

S.NO	QUESTION	MARKS	CO	PO	BTL
1.a	Define data mining?	2	1	1	1
1.b	Explain the types of the data can be Mined.	3	1	1,2,3	1
1.c	Discuss briefly about data cleaning techniques.	10	2,3	1,2	1
2.a	Define Data Mining Task primitives.	2	1	1,2	4
2.b	Explain about Data Mining Functionalities.	3	2,3	1,2	1
2.c	Explain Briefly about Data Cube aggregation and attribute subset selection?	10	1	2,3	2
3.a	What is data cleaning?	2	2	1,2	2
3.b	Explain about Classification of Data mining Systems	3	3	1,2	1
3.c	Explain Briefly about Data Transformation Techniques.	10	2	2,3	2
4.a	Define dimensionality reduction	2	2	2,3	1
4.b	Describe the various phases in knowledge discovery process with a neat diagram.	3	1	2	2
4.c	What are the typical methods for Data Discretization and Concept Hierarchy Generation?	10	1	2	2
5.a	What is Data Binaryzation?	2	1	1,2	1
5.b	Explain about Data Mining Task primitives.	3	2,3	2	6
5.c	Explain briefly about Dimensionality Reduction.	10	2	2	1
6.a	What is Data Preprocessing?	2	1	2	1
6.b	Explain about data integration.	3	2,3	2	2
6.c	Explain about Numerosity Reduction.	10	1	2	2

UNIT -2

S.N O	DESCRIPTION OF QUESTION	MARK S	C O	P O	BT L
1.a	What is Apriori principle?	2	2	1	1
1.b	Define itemset and explain about Frequent Item Set Generation.	3	2	2	2
1.c	Demonstrate Apriori algorithm with example.	10	2	4	3
2.a	Explain about the Support and Confidence Measures.	2	2	1	2
2.b	Explain about Association Rule Generation.	3	2	4	2
2.c	Explain in detail about multilevel association rules.	10	2	2	2
3.a	What is frequent itemset?	2	2	1	1
3.b	Describe maximal frequent itemset?	3	2	2	1
3.c	Give an example for finding frequent itemsets by using candidate generation.	10	2	4	3
4.a	Define Association rule mining two step process.	2	2	1	1
4.b	What are frequent patterns? Give an example.	3	2	2	2
4.c	Explain the FP-Growth algorithm with example.	5	2	4	5
5.a	Define FP-Growth algorithm.	2	2	1	1
5.b	Describe Closed Frequent Itemset?	3	2	2	1
5.c	Explain how can we find frequent itemsets without using candidate generation.	10	2	4	5
6.a	How to compute confidence measure for an association rule?	2	2	2	3
6.b	Define the terms Frequent Itemset, Maximal Frequent Item Set, Closed Frequent Item Set.	3	2	1	1
6.c	Make a comparison of Apriori and FP-Growth algorithms for frequent item set mining in transactional databases. Apply these algorithms to the following data: TID LIST OF ITEMS	10	2	4	6

1	Bread, Milk, Sugar, TeaPowder, Cheese, Tomato				
2	Onion, Tomato, Chillies, Sugar, Milk				
3	Milk, Cake, Biscuits, Cheese, Onion				
4	Chillies, Potato, Milk, Cake, Sugar, Bread				
5	Bread, Jam, Mik, Butter, Chilles				
6	Butter, Cheese, Paneer, Curd, Milk, Biscuits				
7	Onion, Paneer, Chilies, Garlic, Milk				
8	Bread, Jam, Cake, Biscuits, Tomato				

UNIT -3

S.N O	DESCRIPTION OF QUESTION	MARKS	CO	PO	BTL
1.a	What is rule classification?	2	3	1	1
1.b	Explain about the classification and prediction with example.	3	3	4	5
1.c	Explain about the decision tree construction with example.	10	3	4	6
2.a	How prediction is different from classification?	2	3	4	4
2.b	Summarize Classification techniques.	3	3	2	2
2.c	How does the Naïve Bayesian classification works? Explain in detail.	10	3	4	5
3.a	Define information gain.	2	3	1	1
3.b	Write about the Measures for selecting the Best Split?	3	3	2	2
3.c	Explain about the K-Nearest neighbor classification	10	3	2	1
4.a	Define Pre pruning and post pruning.	2	3	1	1

4.b	Explain the steps Join and Prune.	3	3	2	2
4.c	Explain classification by using decision tree induction.	10	3	5	3
5.a	What is Decision tree? Give an example.	2	3	1	1
5.b	Explain about the Naïve-Bayes Classifier.	3	3	5	3
5.c	Discuss about attribute selection measure Information Gain.	10	3	4	5
6.a	Define Prediction?	2	3	2	4
6.b	Write the need for tree pruning in decision tree induction?	3	3	4	5
6.c	Write about Lazy Learners for classification.	10	3	4	5

UNIT -4

S.N O	DESCRIPTION OF QUESTION	MARKS	CO	PO	BTL
1.a	Define Clustering?	2	4	1	1
1.b	Discuss various types of data in cluster analysis.	3	4	4	5
1.c	Explain different Clustering Methods?	10	4	4	6
2.a	What is an outlier?	2	4	1	1
2.b	Describe K-Means Additional issues?	3	4	2	4
2.c	Explain about the Partitioning Clustering-K-Means Algorithm.	10	4	4	5
3.a	List the demerits of k-means algorithm.	2	4	1	1
3.b	Explain about Agglomerative Methods and Divisive methods.	3	4	4	5

3.c	Explain about the PAM Algorithm.	10	4	4	5
4.a	Mention the various types of clustering methods.	2	4	1	1
4.b	Explain Hierarchical clustering.	3	4	4	5
4.c	Explain K-Medoids clustering.	10	4	4	5
5.a	Give a note on k means algorithm.	2	4	1	1
5.b	What are the key issues in Hierarchical Clustering?	3	4	2	4
5.c	Write about the basic Agglomerative Hierarchical Clustering Algorithm?	10	4	2	2
6.a	List the Key Issues in Hierarchical Clustering.	2	4	1	1
6.b	Write about the strengths and weakness of Hierarchical Clustering?	3	4	2	2
6.c	Explain about DBSCAN clustering method.	10	4	2	2

UNIT -5

S.N O	DESCRIPTION OF QUESTION	MARKS	CO	PO	BTL
1.a	Define Web Mining?	2	4	1	1
1.b	What are the applications of Web Mining?	3	4	4	5
1.c	Explain in detail about 3 phases of web mining.	10	4	4	6
2.a	What is OLAP?	2	4	1	1
2.b	Describe the Terminology of Web Structure?	3	4	2	4
2.c	Distinguish between Data Mining & Web Mining?	10	4	4	5
3.a	What is Text Mining?	2	4	1	1
3.b	Explain the methods under Text Transformation.	3	4	4	5
3.c	Explain Text categorization in detail.	10	4	4	5

4.a	What is Unstructured Text?	2	4	1	1
4.b	Explain Text Pre-Processing.	3	4	4	5
4.c	Explain the steps of Text Mining Process with a neat Diagram	10	4	4	5
5.a	What is Episode Rule Discovery?	2	4	1	1
5.b	What are the Advantages & Disadvantages of Text Mining?	3	4	2	4
5.c	Write about Text Clustering in detail?	10	4	2	2
6.a	What is NLP?	2	4	1	1
6.b	What are the applications of Text Mining?	3	4	2	2
6.c	Explain Web Structure Mining in Detail	10	4	2	2