

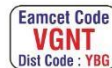


VIGNAN INSTITUTE OF TECHNOLOGY AND SCIENCE

Near Ramoji Film City, Deshmukhi Village, Pochampally Mandal, Yadadri Bhuvanagiri Dist.

(Approved by AICTE, New Delhi, Affiliated to JNTUH, Hyderabad)

AN AUTONOMOUS INSTITUTION



A.Y. 2024-25 Data Mining IV Year I Sem

Department of Computer Science and Engineering

UNIT-1

S.NO	QUESTION	MARKS	CO	PO	BTL
1.a	What is Data Warehouse? Define data mining	2	1	1	1
1.b	Discuss the types of the data can be Mined.	3	1	1	2
1.c	Describe the various phases in knowledge discovery process with a neat diagram.	5	1	1	1
1.d	Explain about Data Mining Functionalities.	5	1	1	2
2.a	Discuss about Interestingness Patterns	2	1	1	2
2.b	List out the applications of data mining.	3	1	3	1
2.c	Explain about Classification of Data mining Systems.	5	1	1	2
2.d	Describe about data mining task primitives	5	1	2	1
3.a	Discuss about subset selection	2	2	1	2
3.b	Discuss about the Issues to be considered during Data Integration	3	1	2	2
3.c	Explain how to integrate data mining system with a data warehouse.	5	1	2	2
3.d	Discuss the major issues in data mining.	5	1	2	2
4.a	Define predictive mining	2	1	1	1
4.b	Discuss the challenges associated with data mining	3	1	2	2
4.c	Explain the need or significance of data preprocessing and various forms of preprocessing. Or Explain the significance of data preprocessing in data mining.	10	1	2	2
5.a	Define outlier analysis?	2	1	1	1
5.b	List the methods of filling missing values.	3	1	2	1
5.c	Differentiate between data mining and data warehouse.	5	1	1	2
5.d	Explain about the Data Cleaning techniques in detail.	5	1	2	2
6.a	Define Stepwise forward selection	2	1	1	1
6.b	State binning can handle noisy data?	3	1	2	1
6.c	Describe Data Normalization methods in brief	5	1	1	1
6.d	Explain briefly about Dimensionality Reduction.	5	1	1	2
7.a	How do task primitive contribute to the overall effectiveness of data mining? Explain	2	1	1	1
7.b	List the Numerosity Reduction methods	3	1	1	1
7.c	Explain Briefly about Data Cube aggregation, attribute subset selection, Data Discretization and Concept hierarchy generation.	5	1	2	2
7.d	“Data preprocessing is necessary before data mining	5	1	2	2

	process”. Justify your answer.				
--	--------------------------------	--	--	--	--

Unit-2

S.N O	DESCRIPTION OF QUESTION	MARKS	C O	PO	BTL																				
1.a	Define frequent itemset?	2	2	1	1																				
1.b	Show how to compute confidence of an association rule? Give example	3	3	2	2																				
1.c	Explain the process of finding frequent patterns	10	3	2	2																				
2.a	Define maximal frequent item set.	2	2	1	1																				
2.b	Describe about association rule mining?	3	3	3	1																				
2.c	Give an overview of correlation analysis.	5	3	1	2																				
2.d	Explain the measures of association rule mining? Explain	5	3	3	2																				
3.a	Describe maximal frequent itemset?	2	2	1	1																				
3.b	Differentiate frequent subsequence and frequent substructure.	3	2	1	2																				
3.c	Compute all the frequent item sets using Apriori algorithm for the given data where min-sup = 2. Transactional Data for an <i>AllElectronics</i> Branch <table><tr><th><i>TID</i></th><th><i>List of item IDs</i></th></tr><tr><td>T100</td><td>I1, I2, I5</td></tr><tr><td>T200</td><td>I2, I4</td></tr><tr><td>T300</td><td>I2, I3</td></tr><tr><td>T400</td><td>I1, I2, I4</td></tr><tr><td>T500</td><td>I1, I3</td></tr><tr><td>T600</td><td>I2, I3</td></tr><tr><td>T700</td><td>I1, I3</td></tr><tr><td>T800</td><td>I1, I2, I3, I5</td></tr><tr><td>T900</td><td>I1, I2, I3</td></tr></table>	<i>TID</i>	<i>List of item IDs</i>	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	10	3	3	3
<i>TID</i>	<i>List of item IDs</i>																								
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																								
4.a	Define Support of an association rule.	2	3	1	1																				
4.b	Define Association rule mining two step process	3	3	1	1																				
4.c	Apply FP-Growth algorithm to the following data for finding frequent item sets, consider support threshold as 30%. <table><tr><th>TID</th><th>List of ItemIDs</th></tr><tr><td>1</td><td>I1, I2, I4, I5</td></tr><tr><td>2</td><td>I2, I4, I7</td></tr><tr><td>3</td><td>I2,I3,I4,I5</td></tr><tr><td>4</td><td>I1,I3,I4,I7</td></tr><tr><td>5</td><td>I1,I2,I3,I4,I5</td></tr><tr><td>6</td><td>I3,I4,I5,I6</td></tr></table>	TID	List of ItemIDs	1	I1, I2, I4, I5	2	I2, I4, I7	3	I2,I3,I4,I5	4	I1,I3,I4,I7	5	I1,I2,I3,I4,I5	6	I3,I4,I5,I6	5	3	3	3						
TID	List of ItemIDs																								
1	I1, I2, I4, I5																								
2	I2, I4, I7																								
3	I2,I3,I4,I5																								
4	I1,I3,I4,I7																								
5	I1,I2,I3,I4,I5																								
6	I3,I4,I5,I6																								
4.d	Explain in detail about multilevel association rules.	5	3	1	2																				
5.a	Explain the purpose of Apriori algorithm.	2	3	3	2																				
5.b	Give a note on Closed Frequent Item	3	3	1	2																				
5.c	Describe various types of association rules.	5	3	1	1																				
5.d	Write a short notes on constraint based association mining.	5	3	1	2																				

6.a	Quote an example for quantitative association rule.	2	3	3	1
6.b	Write the FP-graph algorithm	3	3	1	1
6.c	Explain about the identification of sub graphs in a graph	5	3	3	2
6.d	Write a short note on SPM algorithm?	5	3	1	2
7.a	What is the advantage of rule based classification system.	2	3	3	1
7.b	Describe the role of correlation analysis in identifying relationships within dataset.	3	3	1	1
7.c	How constraint based method does influences the discovery of association rule	5	3	3	2
7.d	Explore different mining methods used in various association rule mining and discuss advantages and limitations	5	3	1	2

Unit-3

S.NO	DESCRIPTION OF QUESTION	MARKS	CO	PO	BTL
1.a	Define classification?	2	3	1	1
1.b	Compare the methods of classification and prediction	3	3	3	2
1.c	Explain classification as a two step process	5	3	1	2
1.d	Explain the way of evaluating performance of classification model? Explain.	5	3	3	2
2.a	Why are decision trees useful?	2	4	1	1
2.b	List the advantages of using decision trees.	3	4	2	1
2.c	Discuss about Decision tree induction algorithm with an example	10	3	2	2
3.a	Define the significance of information gain?	2	3	1	1
3.b	Explain the need for tree pruning in decision tree induction?	3	3	2	2
3.c	Describe Bayesian Belief Network with an example.	10	3	1	1
4.a	Define Pre pruning and post pruning.	2	3	1	1
4.b	Explain the merits and de-merits of the lazy learning method.	3	4	2	2
4.c	Discuss in detail about rule-based classification.	5	3	2	2
4.d	Briefly explain classification problems and general approaches to solve them.	5	3	2	2
5.a	Define associative classification?	2	3	1	1
5.b	Discuss about attribute selection measure Information Gain.	3	3	2	2
5.c	Explain the way of solving a classification problem using k-nearest neighbor algorithm?	5	3	3	2
5.d	Describe about Lazy Learners classification. Explain the merits and de-merits of the lazy learning method.	5	3	2	1
6.a	Define test data?	2	3	1	1
6.b	List the characteristics of k-nearest neighbor algorithm	3	3	1	1

6.c	How does basian classifier handle uncertainty.	5	3	2	2
6.d	Discuss decision tree over fitting and pruning techniques	5	3	2	2

Unit-4

S.N O	DESCRIPTION OF QUESTION	MARKS	CO	PO	BTL
1.a	Define Clustering?	2	3	1	1
1.b	Explain the applications of cluster analysis.	3	4	3	2
1.c	Explain different Clustering Methods?	10	3	1	2
2.a	Define outlier?	2	4	1	1
2.b	Describe K-Means Additional issues?	3	4	3	1
2.c	Illustrate K-mean algorithm with an example.	10	3	3	2
3.a	Discuss the two approaches to improve quality of hierarchical clustering.	2	4	3	2
3.b	List out all partitioning methods for clustering data.	3	5	1	1
3.c	Give a brief note on PAM(K-Medoids) Algorithm with example	10	5	3	2
4.a	Discuss the two approaches to improve quality of hierarchical clustering	2	4	2	2
4.b	Explain Hierarchical clustering.	3	5	1	2
4.c	Discuss hierarchical methods for clustering and contrast agglomerative and divisive approaches.	6	5	1	2
4.d	Discuss the merits and demerits of hierarchical approaches for clustering	4	4	2	2
5.a	Define categorical variable?	2	4	1	1
5.b	Compare agglomerative and divisive methods.	3	4	1	2
5.c	How density based method is used for clustering?	5	5	2	2
5.d	Discuss about the drawbacks of k-means algorithm? How can we modify the algorithm to diminish that problem? Or Illustrate K-mean algorithm with an example.	5	5	2	2
6.a	Define interval-scaled variables?	2	5	1	1
6.b	Differentiate between clustering and classification	3	5	1	2
6.c	Explain the following. a) Cluster analysis. b) Grid-based methods.	5	5	1	2
6.d	Compare and contrast rule based classification systems with decision tree induction.	5	5	1	2

UNIT -5

S.NO	DESCRIPTION OF QUESTION	MARKS	CO	PO	BTL
------	-------------------------	-------	----	----	-----

1.a	Define data stream mining?	2	5	1	1
1.b	Explain the applications of Web Mining?	3	5	3	2
1.c	Explain in detail about Time series Mining.	10	5	1	2
2.a	Define Structured data?	2	6	1	1
2.b	Describe the Terminology of Web Structure?	3	5	1	1
2.c	Explain the following: i) Spatial data mining. ii) Mining sequence patterns in transactional databases	10	5	3	2
3.a	Define Text Mining?	2	5	1	1
3.b	Explain the methods under Text Transformation.	3	6	1	2
3.c	Explain the way of applying the mining techniques to unstructured text database? Explain with example	10	5	3	2
4.a	Give examples for unstructured text.	2	5	3	2
4.b	Give the taxonomy of web mining.	3	6	2	2
4.c	Discuss various kinds of patterns to be mined from web/server logs in web usage mining.	5	5	2	2
4.d	Discuss in detail about web content mining	5	5	2	2
5.a	Define stop words?	2	5	1	1
5.b	Explain the Advantages & Disadvantages of Text Mining?	3	6	2	2
5.c	Explain about the Multimedia data mining	10	5	2	2
6.a	Define NLP?	2	5	1	1
6.b	List the applications of web usage mining.	3	6	3	1
6.c	Explain about the World wide web mining	10	5	1	2