SIDDHARTH GROUP OF INSTITUTIONS :: PUTTUR
Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK (DESCRIPTIVE)

Subject with Code : Data Warehousing and Data Mining(18CS0521) **Course & Branch**: B.Tech.- CSE

Year &Sem: III B.Tech. & II - Sem **Regulation:** R18

UNIT –I

INTRODUCTION TO DATA MINING AND DATA PREPROCESSING

1.	a)	Define Data mining.	[L1][CO1]	**[2M]**
	b)	Classify Data pre-processing methods?	[L4][CO1]	**[2M]**
	c)	Determine the Data mart.	[L4][CO1]	**[2M]**
	d)	Define Data normalization.	[L1][CO1]	**[2M]**
	e)	Distinguish the data reduction methods.	[L4][CO1]	**[2M]**
2.	a)	Define Data mining? Explain about data mining on what kind of data.	[L1][CO1]	**[5M]**
	b)	Compare Data Warehousing and Data Mining	[L5][CO1]	**[5M]**
3.	a)	What is KDD? Explain about data mining as a step in the process of knowledge discovery.	[L1][CO1]	**[5M]**
	b)	How to classify data mining systems? Discuss	[L1][CO1]	**[5M]**
4.	a)	What motivated Data mining? Explain .	[L1][CO1]	**[5M]**
	b)	Explain Data mining as a step in the process of knowledge discovery.	[L5][CO1]	**[5M]**
5.		Discuss about Data Mining Task primitives with examples.	[L6][CO1]	**[10M]**
6.	a)	Discuss the Major issues in Data mining.	[L6][CO1]	**[5M]**
	b)	Why do we preprocess the data? Discuss?	[L1][CO2]	**[5M]**
7.		Explain in detail about Data Mining Functionalities with example.	[L5][CO1]	**[10M]**
8.	a)	Classify different data preprocessing techniques used to improve the overall quality of the mined data.	[L4][CO1]	**[5M]**
	b)	Explain about Data Transformation.	[L2][CO2]	**[5M]**
9.	a)	What is Data Reduction? Discuss in brief.	[L1][CO1]	**[5M]**
	b)	Determine the concept hierarchy generation for categorical data	[L4][CO1]	**[5M]**
10.	a)	Illustrate the concept of Data discretization.	[L2][CO1]	**[5M]**
	b)	Explain about Dimensionality reduction methods?	[L2][CO1]	**[5M]**

UNIT –II

DATA WAREHOUSE AND OLAP TECHNOLOGY:AN OVERVIEW

1.	a)	What is Data Warehousing?	[L1][CO2]	**[2M]**
	b)	Compare the differences between ROLAP and MOLAP server.	[L2][CO2]	**[2M]**
	c)	Classify the major difference between Star and Snowflake schema.	[L2][CO2]	**[2M]**
	d)	Define Base and Apex Cuboids with appropriate example	[L1][CO2]	**[2M]**
	e)	What is AOI?	[L1][CO2]	**[2M]**
2.		Discuss in brief about schemas in multidimensional data model.	[L6][CO2]	**[10M]**
3.	(a)	Compare OLTP and OLAP.	[L4][CO2]	**[5M]**
	(b)	Construct lattice of cuboids given 4 dimensions: time, location, product and supplies.	[L6][CO2]	**[5M]**
4.		Elaborate about Attribute Oriented Induction with example.	[L6][CO2]	**[10M]**
5.		Explain about the Three-tier data warehouse architecture with a neat diagram.	[L5][CO2]	**[10M]**
6.	(a)	What is OLAM? Draw the architecture of OLAM	[L1][CO2]	**[5M]**
	(b)	Define Data warehouse? Discuss Design principles.	[L1][CO2]	**[5M]**
7.		Discuss in detail about Data Warehouse Implementation	[L6][CO2]	**[10M]**
8.		Examine the process of conversion from Data Warehouse to Data Mining.	[L4][CO2]	**[10M]**
9.	(a)	Explain in detail about Fact constellation schema with an example.	[L5][CO2]	**[5M]**
	(b)	Explain any four OLAP operations with appropriate examples	[L5][CO2]	**[5M]**
10.	(a)	How are concept hierarchies useful in OLAP? Explain.	[L1][CO2]	**[5M]**
	(b)	Explain in brief about ROLAP, MOLAP and HOLAP servers.	[L2][CO2]	**[5M]**

UNIT –III

MINING FREQUENT PATTERNS,ASSOCIATIONS AND CORRELATIONS

1.	a)	What is Association rule mining?	[L1][CO3]	**[2M]**
	b)	Define the concept of Support and Confidence.	[L1][CO3]	**[2M]**
	c)	Illustrate the frequent itemset mining?	[L2][CO3]	**[2M]**
	d)	Analyze the curse of dimensionality?	[L4][CO3]	**[2M]**
	e)	What are the draw backs of Apriori Algorithm?	[L1][CO3]	**[2M]**
2.	a)	Discuss about Basic Concepts of Frequent Itemset mining.	[L6][CO3]	**[5M]**
	b)	What are the advantages of FP-Growth algorithm?	[L1][CO3]	**[5M]**
3.		Explain Multilevel Association rules and Multidimensional association rules for mining data.	[L5][CO3]	**[10M]**
4.		Explain about the Apriori algorithm for finding frequent item sets with an example.	[L5][CO3]	**[10M]**

TID	**T100**	**T200**	**T300**	**T400**	**T500**	**T600**	**T700**	**T800**	**T900**
ITEM IDS	I1,I2,I5	I2,I4	I2,I3	I1,I2,I4	I1,I3	I2,I3	I1,I3	I1,I2,I3,I5	I1,I2,I3

Generate the list of frequent item-set ordered by their corresponding suffixes, where the minimum support count is 2.

5.		What are the Draw backs of Apriori Algorithm? Explain about FP Growth Concept in Detail?	[L4][CO3]	**[10M]**
6.		Make use of the database which has five transactions. Let minimum support = 60% and minimum confidence = 80%.	[L3][CO3]	**[10M]**

Transaction	Items
T10	M, O, N, K, E, Y
T20	D, O, N, K, E, Y
T30	M, A, K, E
T40	M, U, C, K, Y
T50	C, O, O, K, I, E

Find all frequent item sets using Apriori and FP-growth, respectively.

7.		Explain about Apriori Algorithm with an example	[L5][CO3]	**[10M]**
8.		Outline FP growth algorithm with an example.	[L2][CO3]	**[10M]**
9.	a)	Explain about Constraint based Association mining	[L5][CO3]	**[5M]**
	b)	Discuss about the criteria for classifying the frequent itemset.	[L6][CO3]	**[5M]**
10.		Describe the steps involved in improving the efficiency of the Apriori algorithm	[L2][CO3]	**[10M]**

UNIT –IV

CLASSIFICATION AND PREDICTION

1.	(a)	Define the concept of classification.	[L1][CO4]	**[2M]**
	(b)	What is Regression?	[L1][CO4]	**[2M]**
	(c)	Define Bayes theorem.	[L1][CO4]	**[2M]**
	(d)	How to evaluate the accuracy of a Classifier?	[L1][CO4]	**[2M]**
	(e)	What is Gain Ratio?	[L1][CO4]	**[2M]**
2.		What are the Issues regarding Classification and Prediction? Explain.	[L1][CO4]	**[10M]**
3.		Outline the concept of Classification by Decision Tree Induction.	[L2][CO4]	**[10M]**
4.		Define Bayes theorem. Explain the Naïve Bayesian Classification with an example	[L1][CO4]	**[10M]**
5.		Discuss about Rule based Classification method.	[L6][CO4]	**[10M]**
6.		Illustrate about Naïve Bayes Classification with an example.	[L2][CO4]	**[10M]**
7.		Define Neural Network. Explain the Classification by Back Propagation	[L1][CO4]	**[10M]**
8.		Evaluate the Classification process of back propagation model with an example	[L5][CO4]	**[10M]**
9.	(a)	Explain about Bayesian belief networks with an example.	[L5][CO4]	**[5M]**
	(b)	Summarize about attribute selection measures.	[L2][CO4]	**[5M]**
10.	(a)	Discuss about Accuracy and Error measures.	[L6][CO4]	**[5M]**
	(b)	What is prediction? Explain about Linear regression method.	[L1][CO4]	**[5M]**

UNIT –V

CLUSTER ANALYSIS

1.	a)	Define Clustering. List basic requirements of cluster analysis.	[L1][CO5]	**[2M]**
	b)	Illustrate the outlier analysis?	[L2][CO5]	**[2M]**
	c)	Write down some typical applications of clustering	[L2][CO5]	**[2M]**
	d)	Give a brief note on PAM Algorithm.	[L2][CO5]	**[2M]**
	e)	Classify various Clustering methods.	[L4][CO5]	**[2M]**
2.	a)	Inference the working of k-means clustering.	[L4][CO5]	**[5M]**
	b)	Compare Agglomerative and Divisive hierarchical clustering.	[L5][CO5]	**[5M]**
3.	a)	What are the basic approaches for generating an agglomerative hierarchical clustering? Explain the algorithm.	[L1][CO5]	**[5M]**
	b)	What is outlier analysis? Discuss.	[L1][CO5]	**[5M]**
4.		Discuss in detail about Partitioning methods in clustering with examples.	[L6][CO5]	**[10M]**
5.		Explain the following clustering methods in detail: (a) BIRCH. (b) CURE	[L5][CO5]	**[10M]**
6.		How clusters are identified using DBSCAN algorithm?	[L1][CO5]	**[10M]**
7.		What is clustering analysis? Explain different types of data in clustering with an example	[L1][CO5]	**[10M]**
8.	a)	Explain k-Means and k-Medoids partitioning methods in detail.	[L5][CO5]	**[5M]**
	b)	Discuss the key issues in hierarchical clustering algorithm.	[L6][CO5]	**[5M]**
9.		Influence the importance of Grid-based and Model-Based methods in detail.	[L5][CO5]	**[10M]**
10.		Discuss in detail about the Data Mining Applications.	[L6][CO5]	**[10M]**