Course Code: 20CS0517 **R20**

SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY:: PUTTUR
(AUTONOMOUS)
Siddharth Nagar, Narayanavanam Road–517583

QUESTION BANK (DESCRIPTIVE)

Subject with Code: Data Warehousing and Data Mining (20CS0517) **Course & Branch:** CSE

Year & Sem: III B.Tech & I Sem **Regulation:** R20

UNIT–I
INTRODUCTION TO DATA MINING AND DATA PREPROCESSING

1	a	Define Data mining? What are all points to be discussed to motivate data mining?	[L1][CO1]	[6M]
	b	Explain Data mining as a step in the process of knowledge discovery.	[L2][CO1]	[6M]
2	a	Explain about data mining. On what kind of data.	[L2][CO1]	[6M]
	b	Compare Data Warehousing and Data Mining.	[L2][CO1]	[6M]
3	a	Define KDD?	[L1][CO1]	[2M]
	b	Explain about data mining as a step in the process of Knowledge discovery.	[L2][CO1]	[10M]
4	a	How to classify data mining systems? Discuss in detail.	[L2][CO1]	[6M]
	b	Explain about Dimensionality reduction methods?	[L2][CO1]	[6M]
5		Explain in detail about Data Mining Functionalities with example.	[L2][CO1]	[12M]
6		Discuss about Data Mining Task primitives with examples.	[L2][CO1]	[12M]
7	a	Discuss the Major issues in Data mining.	[L2][CO1]	[6M]
	b	Why do we pre-process the data? Discuss?	[L2][CO2]	[6M]
8	a	Classify different data pre-processing techniques used to improve the overall quality of the mined data.	[L2][CO1]	[6M]
	b	What is data cleaning? Describe in detail the different methods for data Cleaning.	[L1][CO1]	[6M]
9	a	Explain about Data Transformation in data Mining.	[L2][CO2]	[6M]
	b	What is Data Reduction? Discuss in brief.	[L1][CO1]	[6M]
10	a	Illustrate the concept of Data discretization.	[L2][CO1]	[6M]
	b	Determine the concept hierarchy generation for categorical data.	[L3][CO1]	[6M]

Course Code: 20CS0517

R20

UNIT–II
DATA WAREHOUSE AND OLAP TECHNOLOGY

1	a	Discuss in detail about Data Warehouse Implementation.	[L2][CO2]	[6M]
	b	Explain in detail about the key feature of Data Warehousing.	[l2][CO2]	[6M]
2	a	Discuss in detail about different types of Data ware Housing.	[L6][CO2]	[6M]
	b	Distinguish between OLTP and OLAP.	[L2][CO2]	[6M]
3	a	Discuss in brief about Multi-dimensional data model.	[L2][CO2]	[6M]
	b	Why have a Separate Data Warehouse?	[L1][CO2]	[6M]
4		Discuss the following data warehouse Model: i) Enterprise Warehouse ii) Data Mart iii) Virtual Warehouse	[L2][CO2]	[12M]
5	a	What is data Cube? How will data cube allows data to be modeled and viewed in multiple dimensions.	[L1][CO2]	[6M]
	b	Discuss in brief about schemas in multidimensional data model.	[L2][CO2]	[6M]
6	a	Explain the Role of Concept Hierarchies in dimension.	[L2][CO2]	[6M]
	b	Classify a Starnet Query Model. How will involve in Multidimensional Databases.	[L2][CO2]	[6M]
7		Analyze the OLAP operation in multidimensional data.	[L4][CO2]	[12M]
8		Explain about the Three-tier data warehouse architecture with a neat Diagram.	[L2][CO2]	[12M]
9		Define OLAM? Construct the architecture of OLAM	[L2][CO2]	[12M]
10	a	Explain in brief about ROLAP, MOLAP and HOLAP servers.	[L2][CO2]	[6M]
	b	Elaborate about Attribute Oriented Induction with example.	[L5][CO2]	[6M]

Course Code: 20CS0517

R20

UNIT–III

MINING FREQUENT PATTERNS, ASSOCIATIONS AND CORRELATIONS

1	a	Explain about market basket analysis in Association Mining.	[L2][CO3]	[6M]
	b	Explain support, confidence and lift measures with respect to Association Mining.	L2][CO3]	[6M]
2	a	Discuss about Basic Concepts of Frequent Item set mining.	[L2][CO3]	[6M]
	b	What are the advantages of FP-Growth algorithm?	[L1][CO3]	[6M]
3	a	Explain Multilevel Association rules for mining data.	[L2][CO3]	[6M]
	b	Discuss in details Multidimensional association rules for mining data.	[L2][CO3]	[6M]
4		Describe the steps involved in improving the efficiency of the Apriori algorithm.	[L2][CO3]	[12M]
5		Explain about the Apriori algorithm for finding frequent item sets with an example	[L2][CO4]	[12M]

TID	*List of item_IDs*
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

Generate the list of frequent item-set ordered by their corresponding Suffixes, where the minimum support count is 2 and Minimum Confidence is 60%.

6	a	What are the Draw backs of Apriori Algorithm?	[L1][CO4]	[6M]
	b	Explain about FP Growth Concept in Detail?	[L2][CO4]	[6M]
7		Make use of the database which has five transactions. Let minimum Support = 60% and minimum confidence = 80%.	[L3][CO4]	[12M]

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 FP-growth.

8		Explain about Apriori Algorithm with an example	[L2][CO4]	[12M]
9	a	Outline FP growth algorithm with an example.	[L2][CO4]	[6M]
	b	How will measure from Association Analysis to Correlation Analysis.	[L2][CO4]	[6M]
10		Explain about Constraint based Association mining	[L2][CO4]	[12M]

Course Code: 20CS0517

R20

UNIT –IV

CLASSIFICATION AND PREDICTION

1		Explain the following terms i) Classification ii) Prediction	[L2][CO5]	[12M]
2	a	Distinguish between supervised and unsupervised learning.	[L2][CO5]	[6M]
	b	What are the Issues regarding Classification and Prediction? Explain.	[L1][CO5]	[6M]
3	a	Define Decision Tree. Why are decision tree classifiers so popular?	[L2][CO5]	[6M]
	b	Outline the concept of Classification by Decision Tree Induction.	[L2][CO5]	[6M]
4		Discuss the following terms i) Gini Index ii) Gain ratio iii) Information Gain	[L2][CO5]	[12M]
5	a	Define Bayes theorem.	[L2][CO5]	[4M]
	b	Explain the Naïve Bayesian Classification with an example	[L2][CO5]	[8M]
6		Summarize about attribute selection measures.	[L2][CO5]	[12M]
7		Explain about Bayesian belief networks with an example.	[L2][CO5]	[12M]
8	a	Discuss about Rule based Classification method.	[L2][CO5]	[6M]
	b	Define Neural Network. Explain the Classification by Back Propagation.	[L1][CO5]	[6M]
9		What is prediction? Explain about Linear regression method.	[L1][CO5]	[12M]
10	a	Discuss about Accuracy and Error measures	[L1][CO5]	[6M]
	b	Evaluating the accuracy of a classifier in data mining	[L5][CO5]	[6M]

Course Code: 20CS0517

R20

UNIT–V

CLUSTER ANALYSIS

1	a	Define Clustering. List basic requirements of cluster analysis.	[L1][CO6]	[6M]
	b	Describe the working of PAM algorithm.	[L2][CO6]	[6M]
2	a	Explain the various types of data in Cluster analysis	[L2][CO6]	[6M]
	b	Inference the working of k-means clustering.	[L4][CO6]	[6M]
3		Explain K-Means and K-Medoids partitioning methods in detail.	[L2][CO6]	[12M]
4		Discuss in detail about Partitioning methods in clustering with Examples.	[L2][CO6]	[12M]
5		Discuss the key issues in hierarchical clustering algorithm.	[L2][CO6]	[12M]
6	a	Compare Agglomerative and Divisive hierarchical clustering.	[L5][CO6]	[6M]
	b	What are the basic approaches for generating an agglomerative hierarchical clustering? Explain the algorithm.	[L1][CO6]	[6M]
7	a	Explain the following clustering methods in detail: i) BIRCH ii) CURE	[L2][CO6]	[6M]
	b	How clusters are identified using DBSCAN algorithm?	[L1][CO6]	[6M]
8		Influence the importance of Grid-based and Model-Based methods in detail.	[L1][CO6]	[12M]
9		What is outlier analysis? Name the methods for detecting outliers	[L1][CO6]	[12M]
10		Discuss in detail about the Applications and trends in Data Mining.	[L2][CO6]	[12M]