Code No: **R41053** **R10** **Set No. 1**

IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2015

DATA WAREHOUSING AND DATA MINING

(Common to Computer Science & Engineering and Information Technology)

Time: 3 hours **Max. Marks: 75**

Answer any FIVE Questions
All Questions carry equal marks
****

1	a)	What are the origins of data mining? Explain with the help of a diagram.	[7]
	b)	Explain the steps in the data mining process giving example for each step.	[8]
2	a)	Define sampling. What are different types of sampling?	[7]
	b)	Explain how Principal component Analysis is used for dimensionality reduction.	[8]
3	a)	Explain the star schema of data ware house model with the help diagram.	[8]
	b)	Differentiate between data ware house and operational database.	[7]
4	a)	Explain how cross validation is useful in classifiers of data mining.	[8]
	b)	Explain the bayes theorem.	[7]
5	a)	What is classification? Explain decision tree classifier with the help of diagram.	[8]
	b)	What are the methods for expressing attribute test conditions.	[7]
6	a)	What is support counting? How it is done with hash trees?	[8]
	b)	Explain pruning?.	[7]
7	a)	What are the advantages of k-means algorithm?	[7]
	b)	What are the additional issues of k-means algorithm?	[8]
8	a)	Explain DIANA clustering algorithm.	[8]
	b)	Explain BIRCH clustering algorithm.	[7]

Code No: **R41053** **R10** **Set No. 2**

IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2015

DATA WAREHOUSING AND DATA MINING

(Common to Computer Science & Engineering and Information Technology)

Time: 3 hours **Max. Marks: 75**

Answer any FIVE Questions
All Questions carry equal marks
****

1	a)	Explain various data mining tasks giving example for each.	[8]
	b)	What is the need for preprocessing? List various preprocessing techniques.	[7]
2	a)	What is correlation? Derive an expression for correlation.	[8]
	b)	Explain how weights are useful for combining similarities.	[7]
3	a)	Explain fact constellation schema with the help of diagram.	[8]
	b)	Explain partial materialization in data cube implementation.	[7]
4	a)	What are split points and how to find out correct split points? Explain with example.	[8]
	b)	Explain the criterion for finding the correct tree size.	[7]
5	a)	Explain the nearest neighbor algorithm for classification.	[8]
	b)	Explain the parameters for the evaluation of classifier.	[7]
6	a)	What are different types of association rules? Give examples.	[7]
	b)	Explain FP growth algorithm for the generation of frequent item sets.	[8]
7	a)	List different types of clustering approaches. Mentioning example for each.	[7]
	b)	Explain the k-means algorithm for distance based clustering.	[8]
8	a)	What are the advantages of hierarchical clustering approach?	[7]
	b)	Explain the basic agglomerative clustering algorithm.	[8]

Code No: **R41053** **R10** **Set No. 3**

IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2015

DATA WAREHOUSING AND DATA MINING

(Common to Computer Science & Engineering and Information Technology)

Time: 3 hours **Max. Marks: 75**

Answer any FIVE Questions
All Questions carry equal marks

*****

1	a)	What are the challenges in data mining that motivate the mining tasks?	[8]
	b)	Enumerate the applications of data mining.	[7]
2	a)	Explain similarity between binary vectors.	[7]
	b)	What is Jaccard coefficient? Derive an expression for Jaccard coefficient.	[8]
3	a)	Draw a data cube and explain OLAP operations.	[8]
	b)	Explain effective cube computation.	[7]
4	a)	Define the terms Entropy, information gain and gini index. How they are useful for attribute selection?	[9]
	b)	Explain decision tree induction.	[6]
5	a)	Explain the of naïve Bayesian classifier.	[8]
	b)	Draw a Bayesian belief network.	[7]
6	a)	Explain any association mining algorithm without generating candidate item sets.	[8]
	b)	Define support and confidence. What is support threshold?	[7]
7	a)	What are the strengths and weaknesses of k-means clustering?	[7]
	b)	Justify how K- mediods and PAM clustering algorithms can replace k-means algorithm to overcome its limitations.	[8]
8	a)	What is the basic principle of density based clusteing? How it is advantageous over others?	[8]
	b)	What are the strengths and weaknesses of DBSCAN?	[7]

Code No: **R41053** **R10** **Set No. 4**

IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2015

DATA WAREHOUSING AND DATA MINING

(Common to Computer Science & Engineering and Information Technology)

Time: 3 hours **Max. Marks: 75**

Answer any FIVE Questions
All Questions carry equal marks
****

1	a)	Explain different types of data attributes and their measurements.	[8]
	b)	Explain different types of datasets giving example for each.	[7]
2	a)	Define similarity and dissimilarity of attributes.	[5]
	b)	Explain Euclidian distance, Minkowski distance and Mahalanobis distance.	[10]
3	a)	List OLAP operations and explain each with examples.	[8]
	b)	Explain the 3- tier data ware house architecture with the help of diagram.	[7]
4	a)	Explain the decision tree classifier with the help of a diagram.	[8]
	b)	Explain overfitting and tree pruning.	[7]
5	a)	Explain Bayes theorem. How it is useful for classification in data mining?	[8]
	b)	What is Bayes error rate?	[7]
6	a)	What is Apriori property? Explain Apriori algorithm with an example.	[8]
	b)	What is frequent item set generation? What are candidate itemsets?	[7]
7	a)	Explain a basic k-means algorithm with example.	[8]
	b)	What are the applications of clutering?	[7]
8	a)	Explain the basic hierarchical clustering strategies, AGNES and DIANA.	[7]
	b)	Explain DB SCAN clustering algorithm in detail.	[8]