SRN | | | | | | | | | | | |

PES UNIVERSITY

PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UC14MC572

MAY 2016: END SEMESTER ASSESSMENT (ESA) MCA IV SEMESTER

UC14MC572- DATA WAREHOUSING AND DATA MINING

Time: 3 Hrs | Answer All Questions | Max Marks: 100

1.	a)	What is Data warehouse? Explain various implementation steps for building a data warehouse.	1+9
	b)	Explain about Codd's ten desirable characteristics of an OLAP systems.	10
2.	a)	What is data mining? Explain about various types of attributes considered in data mining with an example.	2+8
	b)	What are the different similarity measures available to measure the objects?. Explain each with an example.	10
3.	a)	How do you mine frequent patterns without generating candidate itemsets? Mention briefly.	2
	b)	Explain about FP Growth algorithm in detail.	8
	c)	Apply Apriori algorithm to find the frequent patterns for the following set of transactions with 40 % of minimum support and 75 % of minimum confidence.	10

Transaction ID	Items
10	Milk, Coffee, Cheese, Juice
20	Newspaper, Coffee, Biscuits, Milk, Juice, Sugar, Tea
30	Cheese, Bread, Juice
40	Bread, Milk
50	Juice, Coffee, Bread, Cheese
60	Sugar, Tea, Coffee, Bread, Milk, Juice, Newspaper
70	Sugar, Newspaper, Coffee, Biscuits, Tea
80	Bread, Cheese
90	Milk, Bread
100	Bread, Cheese, Newspaper

4.	a)	How does the TreeGrowth algorithm build a decision tree? Explain it in detail.	10
	b)	Which algorithm is used to extract the rules directly from the dataset? Explain in detail.	10
5.	a)	What are the differences between classification and clustering?	4
	b)	Apply k_Median Clustering Technique to create two clusters for the following dataset. Mention the objects of Cluster 1 and Cluster 2 after applying the technique.	10

Students	Age	Mark1	Mark2	Mark3
Student 1	23	70	70	52
Student 2	18	79	85	75
Student 3	18	73	75	57
Student 4	22	85	86	87
Student 5	20	55	55	55
Student 6	22	85	86	87
Student 7	19	91	90	89
Student 8	20	70	65	60

	c)	Explain briefly about the methods available to find the distance between clusters.	6