

--	--	--	--	--	--	--	--	--	--



P.E.S. College of Engineering, Mandya - 571 401

(An Autonomous Institution affiliated to VTU, Belgaum)

First Semester - M.Tech, Computer Science and Engineering (MCSE)

Make-up Examination; Feb - 2017

Data Warehousing and Data Mining

Time: 3 hrs

Max. Marks: 100

Note: Answer FIVE full questions, selecting ONE full question from each unit.

UNIT - I

- 1 a. Define data warehouse. Compare the features of OLTP and OLAP systems. 10
- b. With block diagram explain the architecture of data warehouse. 10
- 2 a. Distinguish between predictive task and descriptive task. 2
- b. Define data mining. Explain core data mining tasks with example. 8
- c. Explain any three data processing techniques. 10

UNIT - II

- 3 a. Explain characteristics of decision tree induction algorithm. 10
- b. Consider the training samples given in table 3(b) for a binary classification problem. Using the table compute the Gini index for all three attributes, which attribute is better? 10
- 4 a. Compare rule ordering scheme and class based ordering scheme. 4
- b. Write the characteristics and algorithm of K-nearest neighbor classifier. 10
- c. Consider a football game between two rival teams. Team 0 and Team 1. Suppose Team 0 wins 65% of the time and team 1 remaining matching. Among the games won by team 0, only 30% of them come from playing on team1's football field. On the other hand 75% of victories of Team 1 are obtained at home. If team 1 is to host the neat match between the two teams, which team will most likely emerge as a winner? 6

UNIT - III

- 5 a. Write and explain Apriori algorithm. 10
- b. For a set of transaction given in table 5b find the frequent item set. Consider minimum support count as 3. 5
- c. Write an algorithm to extract closed frequent item sets from a given data set. 5
- 6 a. Explain FP-growth approach for discovering frequent item set. 10
- b. Define concept hierarchy. Explain how a concept hierarchy can be represented for a market basket analysis. Also write the advantages of incorporating concept hierarchy into association analysis. 10

UNIT - IV

- 7 a. Explain different types of clustering. 10
- b. Write and explain K-means algorithm. 10
- 8 a. Explain spatial data mining with an example. 10
- b. List the important parameters affecting DBSCAN algorithm. Write DBSCAN algorithm and explain the time and space complexities involved. 10

UNIT - V

- 9 a. Explain web mining techniques. 10
- b. Explain different approaches to choose a good data mining system. 10
- 10 a. Explain trends in data mining. 10
- b. Explain theoretical foundation of data mining. 10

TABLE 3(b)

Customer ID	Hair color	Height	Dress size	class
1	Black	Tall	Small	Co
2	Black	short	Medium	Co
3	Black	short	Medium	Co
4	Black	short	Large	Co
5	Black	short	Extra large	Co
6	Black	short	Extra large	Co
7	Brown	short	Small	Co
8	Brown	short	Small	Co
9	Brown	short	Medium	Co
10	Brown	average	Large	Co
11	Black	Tall	Large	C1
12	Black	Tall	Extra large	C1
13	Black	Tall	Medium	C1
14	Black	average	Extra large	C1
15	Brown	average	Small	C1
16	Brown	average	Small	C1
17	Brown	average	Medium	C1
18	Brown	average	Medium	C1
19	Brown	average	Medium	C1
20	Brown	average	large	C1

TABLE 5(b)

TID	Items			
1	I ₂ ,	I ₅		
2	I ₂ ,	I ₄ ,	I ₁ ,	I ₆
3	I ₅ ,	I ₄ ,	I ₁ ,	I ₃
4	I ₂ ,	I ₅ ,	I ₄ ,	I ₁
5	I ₂ ,	I ₅ ,	I ₄ ,	I ₃

* * *