

GOVERNMENT POLYTECHNIC, SADAR, NAGPUR

(An Autonomous Institute of Government of Maharashtra)

TERM EXAMINATION EVEN 2018

COURSE NAME : DATA MINING AND DATA WAREHOUSING

COURSE CODE : IT504E

PROGRAMME : Diploma in Information Technology

TIME : 3 Hours

MAXIMUM MARKS : 70

INSTRUCTIONS :

ENROLLMENT NO. :

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

- 1) All questions are compulsory.
- 2) Illustrate your answers with neat sketches wherever necessary.
- 3) Use of Non Programmable Electronic Pocket Calculators is permissible.
- 4) Figures to the right indicate full marks.
- 5) Assume suitable additional data if necessary.
- 6) Mobile phones & other similar electronic gadgets are strictly prohibited inside examination hall.



		Marks	COs
Q. 1	Attempt ANY FIVE	(10)	
1R2 a)	List any two functionalities of data warehouse.		CO1
1R2 b)	Define data warehouse.		CO1
4U2 c)	Describe Association Rule Mining.		CO4
6U2 d)	Describe Grid based method.		CO4
6R2 e)	Define cluster analysis.		CO5
3R2 f)	Define data mining.		CO1
4R2 g)	Define market basket analysis.		CO4
Q. 2	Attempt ANY THREE	(12)	
2R4 a)	Differentiate between OLTP and OLAP on the basis of definition, characteristics, size, orientation, user, functions, DB design and data.		CO2
1U4 b)	Describe the functionalities of data warehouse.		CO1
2U4 c)	Describe the different types of OLAP.		CO2
2U4 d)	Describe data warehouse architecture with the neat diagram.		CO1
2U4 e)	Suppose that a data warehouse consist of four dimension customers, product, salesperson and sales times and the three measure sales amt (in Rupee), VAT (in Rupee) and payment-type (in Rupee). Draw the star schema that is popularly used for modeling data warehouses and explain it.		CO6
Q. 3	Attempt ANY THREE	(12)	
3U4 a)	Describe the data mining techniques in detail.		CO5
4U4 b)	Describe why Association rule mining is necessary (Take example of Market Basket Analysis).		CO5
5R4 c)	Describe the issues regarding to the classification and prediction.		CO3
1R4 d)	List different data warehouse tools.		CO1
1U4 e)	Describe star-schema multidimensional data model with diagram.		CO6
Q. 4	Attempt ANY THREE	(12)	
5U4 a)	Illustrate an algorithm for classification using decision tree.		CO4
6U4 b)	Describe the K-mean clustering algorithm with example.		CO5
6U4 c)	Describe the categorization of major clustering methods.		CO5
3U4 d)	Describe the architecture of data mining with the help of suitable diagram.		CO6
4U4 e)	Describe Apriori algorithm with suitable example.		CO4
Q. 5	Attempt ANY TWO	(12)	
1A6 a)	Why data transformation is essential in the process of knowledge discovery? Describe it.		CO3

2A6	b)	Describe the various OLAP operations with the help of diagram.	CO4																																												
5A6	c)	A simple example from the stock market involving only discrete range has profit as categorically attribute with values { UP, DOWN} and training data is : <table><tr><th>Age</th><th>Competition</th><th>Type</th><th>Profit</th></tr><tr><td>Old</td><td>Yes</td><td>Software</td><td>Down</td></tr><tr><td>Old</td><td>No</td><td>Software</td><td>Down</td></tr><tr><td>Old</td><td>No</td><td>Hardware</td><td>Down</td></tr><tr><td>Mid</td><td>Yes</td><td>Software</td><td>Down</td></tr><tr><td>Mid</td><td>Yes</td><td>Hardware</td><td>Down</td></tr><tr><td>Mid</td><td>No</td><td>Hardware</td><td>Up</td></tr><tr><td>Mid</td><td>No</td><td>Software</td><td>Up</td></tr><tr><td>New</td><td>Yes</td><td>Software</td><td>Up</td></tr><tr><td>New</td><td>No</td><td>Hardware</td><td>Up</td></tr><tr><td>new</td><td>No</td><td>Software</td><td>Up</td></tr></table> <p>1. Apply decision tree algorithm – ID₃ and construct decision tree.</p> <p>2. Enlist rules from decision tree.</p>	Age	Competition	Type	Profit	Old	Yes	Software	Down	Old	No	Software	Down	Old	No	Hardware	Down	Mid	Yes	Software	Down	Mid	Yes	Hardware	Down	Mid	No	Hardware	Up	Mid	No	Software	Up	New	Yes	Software	Up	New	No	Hardware	Up	new	No	Software	Up	CO4
Age	Competition	Type	Profit																																												
Old	Yes	Software	Down																																												
Old	No	Software	Down																																												
Old	No	Hardware	Down																																												
Mid	Yes	Software	Down																																												
Mid	Yes	Hardware	Down																																												
Mid	No	Hardware	Up																																												
Mid	No	Software	Up																																												
New	Yes	Software	Up																																												
New	No	Hardware	Up																																												
new	No	Software	Up																																												
Q. 6	Attempt ANY TWO		(12)																																												
3A6	a)	Describe the following related to data mining: (i) Data specification (ii) Hierarchy specification (iii) Visualization specification	CO3																																												
4A6	b)	Find all frequent item sets for the given training set using Apriori. <table><tr><th>Transaction ID</th><th>Items Bought</th></tr><tr><td>2000</td><td>A, B, C</td></tr><tr><td>1000</td><td>A, C</td></tr><tr><td>4000</td><td>A, D</td></tr><tr><td>5000</td><td>B, E, F</td></tr></table> <p>Let minimum support 50% and minimum confidence 50%.</p>	Transaction ID	Items Bought	2000	A, B, C	1000	A, C	4000	A, D	5000	B, E, F	CO4																																		
Transaction ID	Items Bought																																														
2000	A, B, C																																														
1000	A, C																																														
4000	A, D																																														
5000	B, E, F																																														
6A6	c)	Consider the five points {x ₁ , x ₂ , x ₃ , x ₄ , x ₅ } with the following co-ordinates as a two dimensional sample for clustering : x ₁ = {0,2.5} x ₂ = {0,0} x ₃ = {1.5, 0} x ₄ = {5, 0} x ₅ = {5, 2} Illustrate the K-mean partitioning algorithm by using the above dataset.	CO4																																												

Course Outcomes :

CO1	Identify the concepts of data warehousing and data mining.
CO2	Identify difference between DBMS and Data warehouse.
CO3	Appreciate the issues underlying database implementation.
CO4	Perform various operation using data warehousing.
CO5	Perform query facilities to formulate queries and manipulate the database.
CO6	Create data warehouse.