

Total No. of Questions: 6

Total No. of Printed Pages:3

Enrollment No.....



Faculty of Engineering
End Sem (Odd) Examination Dec-2018
CA5CO15 Data Warehousing and Mining

Programme: MCA

Branch/Specialisation: Computer
Application

Duration: 3 Hrs.

Maximum Marks: 60

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d.

- Q.1 i. The KDD is abbreviation for 1
(a) Knowledge Database Definition
(b) Knowledge Discovery in Databases
(c) Knowledge Discovery Definition
(d) Knowledge Data Definition
- ii. Data can be updated in ____ environment. 1
(a) Data warehouse (b) Data mining
(c) Operational (d) Informational
- iii. The second phase of A Priori algorithm is _____. 1
(a) Candidate generation (b) Item set generation
(c) Pruning (d) Partitioning
- iv. data are noisy & have many missing attribute values 1
(a) Pre-processed (b) Cleaned
(c) Real-World (d) Transformed
- v. _____ is a clustering procedure characterized by the 1
development of a tree-like structure.
(a) Non-hierarchical clustering
(b) Hierarchical clustering
(c) Divisive clustering
(d) Agglomerative clustering
- vi. Decision Tree induction is an example of _____ 1
(a) Classification (b) Association
(c) Clustering (d) Prediction

P.T.O.

[2]

- vii. Snow flake schema has _____ table(s). **1**
 (a) Sub division (b) Sub-dimension
 (c) Index (d) None of these
- viii. Data is loaded, accessed and Scanned in _____ **1**
 (a) OLTP (b) OLAP (c) SMTP (d) FTP
- ix. GA stands for _____. **1**
 (a) Genetic algorithm (b) Gene algorithm.
 (c) General algorithm. (d) Geo algorithm
- x. _____ mining is concerned with discovering the model **1**
 underlying the link structures of the web.
 (a) Data structure (b) Web structure
 (c) Text structure (d) Image structure

- Q.2 i. What is data mining? **2**
- ii. Explain KDD process with the help of a diagram. **8**
- OR iii. Describe various data mining functionalities in detail. **8**

- Q.3 i. Describe the strategies for data reduction. **3**
- ii. Use the two methods below to normalize the following group of **7**
 data:
 200, 300, 400, 600, 1000
 (a) Min-max Normalization (b) Z-score Normalization

- OR iii. A database has five transactions. Let minimum support=60% and **7**
 minimum confidence=80%.
 TID ITEMS_BOUGHT
 T100 {M, O, N, K, E, Y}
 T200 {D, O, N, K, E, Y}
 T300 {M, A, K, E}
 T400 {M, U, C, K, Y}
 T500 {C, O, R, K, I, E}
 Find all frequent itemsets using Apriori algorithm.

- Q.4 i. Differentiate between Classification and Clustering. **3**
- ii. The following table represents the training database that gives **7**
 information of whether or not to play golf, given a set of climatic

[3]

condition. Induct a decision tree and write the decision rules using information gain attribute selection measure:

Outlook	Temp	Humidity	Windy	Class:Play?
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Overcast	Hot	High	False	P
Rain	Mild	High	False	P
Rain	Cool	Normal	False	P
Rain	Cool	Normal	True	N
Overcast	Cool	Normal	True	P
Sunny	Mild	High	False	N
Sunny	Cool	Normal	False	P
Rain	Mild	Normal	False	P
Sunny	Mild	Normal	True	P
Overcast	Mild	High	True	P
Overcast	Hot	Normal	False	P
Rain	Mild	High	True	N

- OR iii. Write and explain k-Means clustering algorithm. Write strengths **7**
 and weakness of k-Means clustering method.

- Q.5 i. What is Data Warehouse? **3**
- ii. Explain snow flake schema with diagram. What is the difference **7**
 between star and snow flake schema.
- OR iii. Describe typical OLAP operations with diagram. **7**

- Q.6 Attempt any two: **5**
- i. Write short note on: **5**
 (a) Mining WWW (b) Spatial Mining
- ii. Discuss social impact of data mining. **5**
- iii. What are the applications of Data mining? Discuss any two **5**
 applications.

Marking Scheme
CA5CO15 Data Warehousing and Mining

Q.1	i.	The KDD is abbreviation for (b) Knowledge Discovery in Databases	1
	ii.	Data can be updated in _____environment. (c) Operational	1
	iii.	The second phase of A Priori algorithm is _____. (c) Pruning	1
	iv.data are noisy & have many missing attribute values (c) Real-World	1
	v.	_____ is a clustering procedure characterized by the development of a tree-like structure. (b) Hierarchical clustering	1
	vi.	Decision Tree induction is an example of _____ (a) Classification	1
	vii.	Snow flake schema has _____ table(s). (b) Sub-dimension	1
	viii.	Data is loaded, accessed and scanned in _____ (b) OLAP	1
	ix.	GA stands for _____. (a) Genetic algorithm	1
	x.	_____ mining is concerned with discovering the model underlying the link structures of the web. (b) Web structure	1
Q.2	i.	Data mining definition	2
	ii.	Diagram.KDD process Explanation 1 mark for each step (1 mark *5)	3 marks 5 marks
OR	iii.	Data mining functionalities 2 Mark for each functionality	8 (2 marks * 4)
Q.3	i.	Strategies for data reduction. 1 Mark for each strategy	3 (1 mark * 3)
	ii.	(a) Min-max Normalization (b) Z-score Normalization	3.5 marks 3.5 marks
OR	iii.	Item generation	4 marks 7

		Pruning	3 marks
Q.4	i.	Difference between Classification and Clustering. 0.75 Mark for each difference	3 (0.75 mark * 4)
	ii.	Calculations of Information Gain Tree Construction	4 marks 3 marks
OR	iii.	Algorithm Strength and weakness	4 marks 3 marks
Q.5	i.	Data Warehouse definition 0.75 for each part of definition	3 (0.75 mark * 4)
	ii.	Diagram and explanation of Snowflake Difference	5 marks 2 marks
OR	iii.	Typical OLAP operations with diagram. Explanation Diagram	7 4 marks 3 marks
Q.6		Attempt any two:	
	i.	Write short note: 2.5 Marks for each note (a) Mining WWW (b) Spatial Mining	(2.5 marks * 2) 5
	ii.	Social impact of data mining. 1 Mark for each impact	5 (1 mark * 5)
	iii.	Applications of Data mining Applications	2.5 Marks 2.5 Marks 5
