Total No. of Questions: 6 *Total No. of Printed Pages:3*

Enrollment No.....

Faculty of Engineering

End Sem (Odd) Examination Dec-2019

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. Which of the following process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evolution and knowledge presentation? **1**

(a) MDX process (b) ETL process

(c) KTL process (d) KDD process

ii. Extreme values that occur infrequently are called as _____. **1**

(a) Outliers (b) Rare values

(c) Dimensionality reduction (d) All of these

iii. If T consist of 500000 transactions, 20000 transaction contain bread, 30000 transaction contain jam, 10000 transaction contain both bread and jam. Then the support of bread and jam is ____. **1**

(a) 2% (b) 20% (c) 3% (d) 30%

iv. Converting data from different sources into a common format for processing is called as _____. **1**

(a) Selection (b) Preprocessing

(c) Transformation (d) Interpretation

v. Which of the following is a predictive model? **1**

(a) Clustering (b) Regression

(c) Summarization (d) Association rules

vi. In _____ the groups are not predefined. **1**

(a) Association rules (b) Summarization

(c) Clustering (d) Prediction

vii. In the ETL process, what must be updated first? **1**

(a) Dimensions (b) Fact table

(c) Indices (d) Surrogate Key

P.T.O.

viii. The star schema is composed of _____ fact table(s). **1**
(a) One (b) Two (c) Three (d) Four

ix. In web mining, _____ is used to find natural groupings of users, pages, etc **1**
(a) Classification (b) Associations
(c) Sequential analysis (d) Clustering

x. SD(CLARANS) and CLARANS Extention algorithms are used for: **1**
(a) Spatial Classification (b) Spatial Association Rule
(c) Spatial Clustering (d) None of these

Q.2 i. Write down the steps involved in a typical data mining process. **4**

ii. How is data mining influenced by fields like DBMS, statistics, Mathematics and Biology? **6**

OR iii. How is predictive mining different from descriptive mining? Briefly explain the data mining techniques used in both types of mining. **6**

Q.3 i. What are concept hierarchies? Give example and show how they are helpful in data mining? **4**

ii. Explain data transformation and data reduction with example. **6**

OR iii. Describe A- Priori Algorithm. Why is it used? Also write down its disadvantages. **6**

Q.4 i. Differentiate between partition clustering and hierarchcal clustering. **4**

ii. Explain the process of classification using Decision Tree induction algorithm. **6**

OR iii. Describe K- Means algorithm? Write down its applications. **6**

Q.5 i. “Data warehouse is different from database management system”. Do you agree? Justify your answer. **4**

ii. Propose a multi dimentional data model for an insurance company. There are many types of insurance including vehicle insurance, life insurance, fire insurance etc. Insurance company need to analyse: potential customers, fraud customers, liabilities etc. Suggest OLAP queries to retrieve useful information from the data model. **6**

OR iii. Describe data warehouse architecture. What are the hardware and software requirements for creating a data warehouse? **6**

Q.6 Attempt any two:

i. Throw light on the need of mining text data. Explain the procedure of text mining. **5**

ii. Write down the applications of spatial data mining. Also describe different data structures used for spatial mining. **5**

iii. What are different types of web mining? How can the web mining techniques help to accelerate current business scenario? **5**

******

Marking Scheme
CA5CO15 Data Warehousing and Mining

Q.1	i.	Which of the following process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evolution and knowledge presentation? (d) KDD process	**1**
	ii.	Extreme values that occur infrequently are called as _____. (a) Outliers	**1**
	iii.	If T consist of 500000 transactions, 20000 transaction contain bread, 30000 transaction contain jam, 10000 transaction contain both bread and jam. Then the support of bread and jam is ____. (a) 2%	**1**
	iv.	Converting data from different sources into a common format for processing is called as _____. (c) Transformation	**1**
	v.	Which of the following is a predictive model? (b) Regression	**1**
	vi.	In _____ the groups are not predefined. (c) Clustering	**1**
	vii.	In the ETL process, what must be updated first? (a) Dimensions	**1**
	viii.	The star schema is composed of _____ fact table(s). (a) One	**1**
	ix.	In web mining, _____ is used to find natural groupings of users, pages, etc (d) Clustering	**1**
	x.	SD(CLARANS) and CLARANS Extention algorithms are used for: (c) Spatial Clustering	**1**
Q.2	i.	Steps involved in a typical data mining process. 1 mark for each steps (steps can be more than 4) (1 mark * 4)	**4**
	ii.	Data mining influenced by fields like DBMS, statistics, Mathematics and Biology 2 marks for each field subject to maximum (2 marks *3)	**6**
OR	iii.	Predictive mining different from descriptive mining 2 marks Data mining techniques 4 marks	**6**
Q.3	i.	Definition concept hierarchies 1 marks Uses 3 marks	**4**
	ii.	Data transformation with example 3 marks Data reduction with example 3 marks	**6**
OR	iii.	Description A- Priori Algorithm 4 marks. Why is it used? 1 mark Its disadvantages 1 mark	**6**
Q.4	i.	Differentiate between partition clustering and hierarchcal clustering 1 mark for each difference (1 mark * 4)	**4**
	ii.	Process of classification using Decision Tree induction algorithm Theory 4 marks Algorithm 2 marks	**6**
OR	iii.	Description of K- Means algorithm 5 marks Its applications 1 mark	**6**
Q.5	i.	"Data warehouse is different from database management system". For agreement 1 mark For justification 3 marks	**4**
	ii.	Suggest OLAP queries to retrieve useful information from the data model. For describing concept 2 marks For OLAP operations of slice, dice, pivot drill down, rollup etc. 4 marks	**6**
OR	iii.	Data warehouse architecture Drawing the 3 tier architecture 3 marks Hardware and software requirements 3 marks	**6**
Q.6		Attempt any two:	
	i.	Need of mining text data 1 mark Procedure of text mining 4 marks	**5**
	ii.	Applications of spatial data mining 2 marks Data structures used for spatial mining 3 marks	**5**
	iii.	Types of web mining 2 marks Web mining techniques help to accelerate current business scenario 3 marks	**5**

******