Enrollment No......................................

Faculty of Management Studies
End Sem (Odd) Examination Dec-2019
MS3EG11 Data Mining and Business Intelligence

Programme: BBA   Branch/Specialisation: DM

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.   _____ is designed to overcome any limitations placed on the **1**
           data warehouse.
           (a) Operational database
           (b) Relational database
           (c) Multidimensional database
           (d) Data repository

     ii.   Data warehouse architecture is based on _____   **1**
           (a) DBMS   (b) RDBMS   (c) MySQL   (d) Sybase

     iii. The active data warehouse architecture includes which of the **1**
           following?
           (a) At least one data mart
           (b) Data extraction from numerous internal and external sources
           (c) Near real-time updates
           (d) All of these

     iv.   MOLAP stands for: **1**
           (a) Multidimensional Operational Analytic Processing
           (b) Multidimensional Online Analytical Processing
           (c) Mining Online Analytical Program
           (d) Mining Operational Analytical Processing

     v.   Attribute subset selection is a technique of   **1**
           (a) Data Reduction   (b) Data Transformation
           (c) Data Cleaning   (d) Data Integration

vi. Data transformation includes _____ **1**
(a) A process to change data from a detailed level to a summary level
(b) A process to change data from a summary level to a detailed level.
(c) Joining data from one source into various sources of data
(d) None of these

vii. Point out the wrong statement: **1**
(a) k-means clustering is a method of vector quantization
(b) k-means clustering partition n observations into k clusters
(c) k-nearest neighbour is same as k-means
(d) None of these

viii. Data can be updated in _____ environment. **1**
(a) Data Mining　　　(b) Data Warehouse
(c) Operational　　　(d) Informational

ix. The technology area associated with the CRM is **1**
(a) Specification　　　(b) Generalization
(c) Personalization　　　(d) Summarization

x. Which of the following areas are affected by BI? **1**
(a) Revenue　　(b) CRM　　(c) Sales　　(d) All of these

Q.2 i. What are Data Marts? Explain the types of Data Marts. **4**
ii. Briefly explain different types of Databases used in Data Mining from where data can be extracted. **6**
OR iii. Explain the typical architecture of Data Mining Process. **6**

Q.3 i. What is ETL? **2**
ii. Explain the Star and Snowflake schema of Data Warehouse. **3**
iii. Describe any five OLAP operations with example. **5**
OR iv. Differentiate between OLAP and OLTP. **5**

Q.4 i. What is Outlier Detection? How it can be used for marketing? **3**
ii. Explain data mining as a step process of knowledge discovery. Mention the Functionalities of Data mining. **7**
OR iii. What is Pre-Processing? Explain steps involved in Pre-Processing. **7**

Q.5 i. Explain decision tree induction with the example. **4**
ii. What is Classification and Clustering? Explain various Clustering techniques. **6**
OR iii. How prediction can be done using the Naïve Bayes Classifier. **6**

Q.6 Attempt any two:
i. Explain the term "Market Basket Analysis". How it can be used in Marketing? **5**
ii. Distinguish between Data Mining, Business Analytics & Business Intelligence. **5**
iii. Explain the use of Data Mining in e-commerce. **5**

*****

Marking Scheme
MS3EG11 Data Mining and Business Intelligence

Q.1 i. _____ is designed to overcome any limitations placed on the data warehouse. **1**
(c) Multidimensional database

ii. Data warehouse architecture is based on _____ **1**
(b) RDBMS

iii. The active data warehouse architecture includes which of the following? **1**
(d) All of these

iv. MOLAP stands for: **1**
(b) Multidimensional Online Analytical Processing

v. Attribute subset selection is a technique of **1**
(a) Data Reduction

vi. Data transformation includes _____ **1**
(a) A process to change data from a detailed level to a summary level

vii. Point out the wrong statement: **1**
(c) k-nearest neighbour is same as k-means

viii. Data can be updated in _____ environment. **1**
(c) Operational

ix. The technology area associated with the CRM is **1**
(c) Personalization

x. Which of the following areas are affected by BI? **1**
(b) CRM

Q.2	i.	What are Data Marts	1 mark	**4**
		Types of Data Marts.	3 marks	
	ii.	Different types of Databases (5 or 6 types)	6 marks	**6**
OR	iii.	Typical architecture of Data Mining Process.	6 marks	**6**
Q.3	i.	Define ETL	2 marks	**2**
	ii.	Star and Snowflake schema of Data Warehouse	3 marks	**3**
	iii.	Five OLAP operations with example.	5 marks	**5**
OR	iv.	5 Differentiate between OLAP and OLTP	5 marks	**5**
Q.4	i.	Outlier Detection	1 mark	**3**
		It can be used for marketing	2 marks	
	ii.	Data mining as a step process and Mention the Functionalities of		**7**

		Data mining.(7 Points)	(1 mark*7)	
OR	iii.	Pre-Processing	2 marks	**7**
		Steps involved in Pre-Processing.	5 marks	
Q.5	i.	Decision tree induction	2 marks	**4**
		Example.	2 marks	
	ii.	Classification	1 mark	**6**
		Clustering	1 mark	
		Various Clustering techniques.	4 mark	
OR	iii.	Prediction can be done using the Naïve Bayes Classifier.		**6**
			6 marks	
Q.6		Attempt any two:		
	i.	Market Basket Analysis	3 marks	**5**
		Used in Marketing	2 marks	
	ii.	Distinguish between Data Mining	2 marks	**5**
		And Business Analytics	1.5 marks	
		And Business Intelligence.	1.5 marks	
	iii.	Use of Data Mining in e-commerce	5 marks	**5**

*****