Total No. of Pages 3 2

25–

IVth SEMESTER

Roll No. .....

END SEMESTER EXAMINATION **MBA**

MGT-43 *May-2019*

Time: 3:00 Hours **Business Intelligence**

Max. Marks: 60

NOTE: All Questions compulsory. Attempt any five questions. Assume Suitable missing data , if any

Q1. Consider the following dataset, where each record represents the weather condition and the class attribute shows whether people generally play sports in that weather condition or not. Construct decision tree classifier using ID3 algorithm. Also, write various rule. [10+2]

Outlook	Temperature(°F)	Humidity (%)	Windy	Class
Sunny	75	70	True	Play
Sunny	80	90	True	Don't Play
Sunny	85	85	False	Don't Play
Sunny	72	95	False	Play
Sunny	69	70	False	Play
Overcast	72	90	True	Play
Overcast	83	78	False	Play
Overcast	64	65	True	Play
Overcast	81	75	False	Don't Play
Rain	71	80	True	Don't Play
Rain	65	70	True	Don't Play
Rain	75	80	False	Play
Rain	68	80	False	Play
Rain	70	96	False	Play

Q2. (i) Data warehousing is the only viable means to resolve the information crisis and to provide strategic information. List two reasons to support this assertion and explain them. [6]

(ii) What are the three major types of metadata in a data warehouse? Briefly mention the purpose of each type. Also, give examples supporting each type of meta data.[6]

Q3. (i) Why is data integration required in a data warehouse, more so there than in an operational application? Also, explain why data in data warehouse requires time element. [6]

(ii) You are the data analyst on the project team of an insurance company. You are building a data warehouse for a business intelligence system of the company. List the possible data sources from which you will bring the data into your data warehouse. State your assumptions. [6]

Q4. (i) BigBook, Inc. is a large book distributor with domestic and international distribution channels. The company orders from publishers and distributes publications to all the leading booksellers. Initially, you want to build a data warehouse to analyze shipments that are made from the company's many stores. Determine the metrics or facts and the business dimensions. Prepare an information package diagram. [8]
(ii) Explain the classification process with the help of a schematic diagram. [4]

Q5. (i) Consider the following data set consisting of the marks the of two subjects of seven individuals. Use k-means clustering two clusters the following dataset. [6]

Record	Subject 1	Subject 2
R1	1.0	1.0
R2	2.0	1.5
R3	4.0	3.0
R4	7.0	5.0
R5	5.0	3.5
R6	5.0	4.5
R7	4.5	3.5

(ii) Explain any three advantages of the STAR schema. Also describe the composition of the primary keys for the dimension and fact tables. [6]

Q6.(i) What do you understand by association rule mining? Apply Apriori algorithm to the following dataset to find strong association rules. Assume min_sup=25% and min_conf = 80%. [6]

Trans_ID	Item_ID
1	A,B,C
2	B,D
3	B,C
4	A,B,D
5	A,C
6	B,C
7	A,C
8	A,B,C,E
9	A,B,C

(ii) Differentiate between OLAP and OLTP. [6]

Q7.(i) Why is the entity-relationship modeling technique not suitable for the data warehouse? How is dimensional modeling different? [6]
(ii)Explain how do you compute the dissimilarity between objects described by the following type of variables: (a) Numerical variables (b) Asymmetric binary variables (c) Symmetric binary variables (d) Ratio scaled variables (e) Ordinal variables. [6]

-END-