

Total No. of Pages: 2

USME, DTU East Campus

2nd Semester MBA (Business Analytics)

END SEMESTER EXAMINATION- MAY 2019

PAPER CODE: MB 204

TITLE OF PAPER- Data Warehousing and Data Mining

Time: 3:00 Hours

Max. Marks: 60

Note: Write your Roll no. on the top of this question paper.

Marks are indicated against each question. Parts of a question must be answered together.

Q1. Answer **all** the following questions:

[5* 4 marks = 20 marks]

- Describe the characteristics of different categories of clustering algorithms.
- Differentiate between OLAP and OLTP systems.
- Explain the architecture of a data mining system with the help of a schematic diagram.
- What is principal component analysis? Explain the steps involved.
- Explain the representation of text documents, weighting scheme and retrieval of documents using vector-space model with an example.

Q2. Attempt **any four** out of the following:

[4* 4 marks = 16 marks]

- Describe ROLAP and MOLAP models.
- Define data cube. Explain the operations of data cube with suitable examples.
- Explain the different types of data transformation methods with example.
- What is data discretization? Give two methods of concept hierarchy generation for numeric data.
- Describe in detail the application of data mining in retail or finance sector by taking two scenario cases.

Q3. Attempt **any three** out of the following questions:

[3* 6 marks = 18 marks]

- Suppose that a data warehouse for Big University consists of the following four dimensions: student, course, semester, and instructor, and two measures count and avg grade. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg grade measure stores the actual course grade of the student. At higher conceptual levels, avg grade stores the average grade for the given combination.
 - Draw a snowflake schema diagram for the data warehouse.
 - Starting with the base cuboid [student; course; semester; instructor], what specific OLAP operations (e.g., roll-up from semester to year) should one perform in order to list the average grade of CS courses for each Big University student.
- Describe the k-means clustering algorithm in detail and give its limitations.
- Explain different types of data mining functionalities with examples.

d) Find the clusters from given euclidean distance matrix of six points by using single link hierarchical clustering algorithm and draw the dendrogram.

	P1	P2	P3	P4	P5	P6
P1	0	0.24	0.22	0.37	0.34	0.23
P2		0	0.15	0.20	0.14	0.25
P3			0	0.15	0.28	0.11
P4				0	0.29	0.22
P5					0	0.39
P6						0

Q4. Answer any three out of the following questions:

[3* 2 marks = 6 marks]

- a) What is data warehouse?
- b) Discuss the issues to be considered during data integration.
- c) Describe binning with an example.
- d) Explain the types of sampling with an example.
