SVKM's NMIMS
MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT & ENGINEERING

Programme: B.Tech (Computer)   Year: IV   Semester: VII

Academic Year: 2019-20

Subject: Data Warehousing and Mining

Date: 08 November 2019

Marks: 70
Time: 2.00 pm - 5.00 pm
Durations: 3 (hrs)
No. of Pages: __02__

Final Examination (2019-20)/ Re-Examination (2018-19)

Instructions: Candidates should read carefully the instructions printed on the question paper and on the cover of the Answer Book, which is provided for their use.
1) Question **No. 1** is compulsory.
2) Out of remaining **6** questions, attempt any **4** questions.
3) **In all 5 questions to be attempted.**
4) All questions carry equal marks.
5) **Answer to each new question to be started on a fresh page.**
6) Figures in brackets on the right hand side indicate full marks.
7) Assume suitable data if necessary.

Q.1	a)	Compare Agglomerative and Divisive clustering methods.	[3]	14
	b)	Compare Data warehouse with Data Marts.	[4]	
	c)	Why ER model is not suitable for Data Warehouse ?	[3]	
	d)	Explain any one method of partitional clustering with suitable example.	[4]	
Q.2	a)	Explain bottom up architecture of a data warehouse with neat diagram and with its advantages.	[7]	14
	b)	In real world data, tuples with missing values and noisy data are common occurrence. Describe various methods for handling this problem.	[7]	
Q.3	a)	What are aggregate fact tables? With suitable examples, explain multi-way aggregate fact tables.	[7]	14
	b)	Explain naïve Bayes algorithm for classification with its steps. Also list any three advantages of Naïve Bayes algorithm over KNN.	[7]	
Q.4	a)	What do you mean by Data transformation in ETL? Explain tasks involved in data transformation.	[7]	14
	b)	Use complete linkage algorithm to find the clusters from the following dataset. Also draw the dendogram for same.	[7]	

X	4	8	15	24	24
Y	4	4	8	4	12

| Q.5 | a) | What is OLAP? With neat diagrams explain any three OLAP models in detail. | [7] | 14 |
| | b) | Explain any six metrics for evaluating classifier performance using the Confusion matrix. | [7] | |

| Q.6 | a) | What is data mining? Discuss major issues in data mining. | [7] | 14 |
| | b) | A database has five transactions. Let minimum support count=2, and minimum confidence is 80%. Find all frequent item sets using Apriori algorithm. | [7] | |

TID	Items
T1	I1,I2,I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

Q.7	a)	Metadata acts as nerve center of a data warehouse. Justify.	[4]	14
	b)	Explain Clustering for large datasets using DBSCAN algorithm	[4]	
	c)	In brief, discuss Trends in data mining.	[3]	
	d)	Explain with an example type1, type2 and type3 changes in dimension table.	[3]	