

USN

--	--	--	--	--	--	--	--	--	--

CMBIT LIBRARY
BANGALORE - 560 037

13MCA442

Fourth Semester MCA Degree Examination, June/July 2018

Data Warehousing and Data Mining

Time: 3 hrs.

Max. Marks:100

Note: Answer any FIVE full questions.

- 1 a. What is a data warehouse? Explain the implementation steps of data warehouse. (08 Marks)
- b. Explain the types of operations that can be performed on data cube. (08 Marks)
- c. Compare Operational Data Store (ODS) and Data Warehouse (DW). (04 Marks)
- 2 a. What is data mining? Explain the challenges that motivated the development of data mining. (08 Marks)
- b. With a neat diagram, explain the KDD (Knowledge Discovery in Database) process. (06 Marks)
- c. Explain different types of data mining task. (06 Marks)
- 3 a. What is data preprocessing? Explain any four types of data preprocessing techniques. (10 Marks)
- b. What is an attribute? Explain the types of attributes with example. (06 Marks)
- c. What is Euclidian distance? Find the Euclidian distance for four points P_1 , P_2 , P_3 and P_4 where x and y coordinates of four points are given below:

Point	x coordinate	y coordinate
P_1	0	2
P_2	2	0
P_3	3	1
P_4	5	1

(04 Marks)

- 4 a. Define the following:
 - i) Association rule
 - ii) Support of a rule
 - iii) Confidence of a rule
 (04 Marks)
- b. What is antimonotone property of support? Explain Apriori algorithm. (08 Marks)
- c. Explain FP growth algorithm. Construct FP tree for the following transaction data set:

TID	Items
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}
4	{a, d, e}
5	{a, b, c}
6	{a, b, c, d}
7	{a}
8	{a, b, c}
9	{a, b, d}
10	{b, c, e}

(08 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg. 42+8 = 50, will be treated as malpractice.

- 5 a. Explain measures of node impurity in decision tree induction. (05 Marks)
b. Explain sequential covering algorithm. (05 Marks)
c. State Bayes theorem. Explain Bayes classifier with example. (10 Marks)
- 6 a. How to estimate the accuracy of classification models? (10 Marks)
b. Write short notes on Bagging and Boosting. (06 Marks)
c. Explain multiclass problem. (04 Marks)
- 7 a. Explain the types of clusters. (08 Marks)
b. Write basic Agglomerative hierarchical clustering algorithm. (04 Marks)
c. What is DBSCAN algorithm? Explain core points, border points and noise points. (08 Marks)
- 8 a. Explain the statistical approach for outlier detection. (10 Marks)
b. Write relative density outlier score algorithm. (05 Marks)
c. Write short notes on proximity based outlier detection. (05 Marks)

* * * * *