

Time: 3 hrs.

Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.

2. *M*: Marks, *L*: Bloom's level, *C*: Course outcomes.

Module – 1			M	L	C												
Q.1	a.	What is a Data Warehouse? With a neat diagram, explain the 3-tier architecture of data warehouse.	10	L2	CO1												
	b.	Explain star schema, snowflake schema and fact constellation schema for data warehouse with a neat diagram.	10	L2	CO1												
OR																	
Q.2	a.	What are Data Warehouse Models? Explain the different types of data mart.	10	L2	CO1												
	b.	Discuss about OLAP operations on multidimensional data with example.	10	L2	CO1												
Module – 2																	
Q.3	a.	Explain the process of Knowledge Discovery in Databases (KDD) with a neat diagram.	10	L2	CO2												
	b.	Discuss various ways of handling missing values and noisy data during data cleaning.	10	L2	CO2												
OR																	
Q.4	a.	What is Data Pre-processing? Explain any two pre-processing steps in detail.	10	L2	CO2												
	b.	Explain normalization by min-max score, z-score, and decimal scaling with an example.	10	L2	CO2												
Module – 3																	
Q.5	a.	Explain the significance of data generalization and summarization based on characterization.	10	L2	CO2												
	b.	Explain Aprori algorithm for frequent item set mining. For given transaction data, generate frequent itemset and identify valid association rules with minimum support as 60% and minimum confidence as 75%. <table><tr><th>TID</th><th>ITEMS</th></tr><tr><td>1</td><td>Bread, Milk</td></tr><tr><td>2</td><td>Bread, Chocolate, Pepsi, Eggs</td></tr><tr><td>3</td><td>Milk, Chocolate, Pepsi, Coke</td></tr><tr><td>4</td><td>Bread, Milk, Chocolate, Pepsi</td></tr><tr><td>5</td><td>Bread, Milk, Chocolate, Coke</td></tr></table>	TID	ITEMS	1	Bread, Milk	2	Bread, Chocolate, Pepsi, Eggs	3	Milk, Chocolate, Pepsi, Coke	4	Bread, Milk, Chocolate, Pepsi	5	Bread, Milk, Chocolate, Coke	10	L3	CO2
TID	ITEMS																
1	Bread, Milk																
2	Bread, Chocolate, Pepsi, Eggs																
3	Milk, Chocolate, Pepsi, Coke																
4	Bread, Milk, Chocolate, Pepsi																
5	Bread, Milk, Chocolate, Coke																

1 of 2

OR

Q.6	a.	Explain Associative Rule Mining. What do you mean by incremental ARM?	10	L2	CO2
	b.	Construct the FP-Tree for the following data set (Given minimum support is 2):	10	L3	CO2

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	B C
8	A B C
9	A B D
10	B C E

Module – 4

Q.7	a.	Explain Decision Tree Induction classification method with an example.	10	L2	CO3
	b.	Write a note on Naïve Bayes Classifier.	10	L2	CO3

OR

Q.8	a.	Explain linear and logistic regression prediction methods with example.	10	L2	CO5
	b.	Write a note on data mining tools DB miner, WEKA and DTREG.	10	L2	CO5

Module – 5

Q.9	a.	Explain data mining for business applications balanced scorecard and fraud detection.	10	L2	CO4
	b.	Explain the state of the practice in analytics role of data scientists.	10	L2	CO4

OR

Q.10	a.	Explain the main phases of data analytics life cycle in detail.	10	L2	CO4
	b.	Discuss the key stakeholders of analytics project.	10	L2	CO4
