

“Please check whether you have got the right question paper.”

- i) Q.1 and Q.6 are compulsory .Solve any two questions from the remaining ones in each section
 ii) Assume suitable data if necessary and state it clearly.

SECTION-A

- Q.1 a) Describe the snowflake scheme. 03
 b) Define the terms in brief – ROLAP, MOLAP, and HOLAP. 03
 c) What are the typical OLAP operations that can be performed on a data cube? 04
- Q.2 a) Explain a block diagram explain the steps to achieve Business Intelligence using data mining . 08
 b) Describe the differences between the various approaches for the integration (coupling)of a data mining 07
 system with a database or data warehouse system .
- Q.3 a) What is a KDD process ? What are its components if seen from machine Learning and Statistics ? 08
 b) What are Major issues in data mining ? 07
- Q.4 a) What is data dissimilarity ?Given two objects represented by the tuples (22,1,42,10) and (20,0, 36,8): 08
 i) Compute the Euclidean distance between the two objects .
 ii) Compute the Manhattan distance between the two objects
 iii) Compute the Minkowski distance between the two objects,using $q=3$.
 b) In real-world data , tuples with missing values for some attributes are a common occurrence . Describe 07
 various methods for handing this problem.
- Q.5 a) What are the properties of standard deviation ? How mean ,mode and standard deviation help in handing 07
 data dissimilarity?
 b) Write short notes on : (Any two) 08
 i) Data cleaning
 ii) ETL tools
 iii) Indexing schemes .

SECTION-B

- Q.6 a) What is Association rule mining ? 05
 b) Define and describe any five metrics used for measuring classifier performance 05
- Q.7 a) A database has five transactions. Find all frequent itemsets using Apriori algorithm by taking min 10
 support count as 2

TID	Items bought
T100	M,O,N,K,E,Y
T200	D,O,N,K,E,Y
T300	M,A,K,E
T400	M,U,C,K,Y
T500	C,O,O,K,I,E

Generate association rules by considering the min confidence as 80% .

- b) What is market –basket analysis ? Which measures of interestingness are used there ? 05

- Q.8 a) Describe k- means algorithm in brief ?What are its limitations? 08
b) How does k-medoids algorithm cluster in data ?What are its advantages? 07
- Q.9 a) Take any suitable database with 8 tuples. Describe the working of Decision tree classifier using ID3 algorithm .What is the role of Information Entropy and Information Gain? 08
b) What is the process of extracting rules from the Decision Tree? 07
- Q.10 a) What is the process of Intelligence creation and use in BI ?What is the role of BI governance. 07
b) Write short notes on (any two) 08
i) Use of Dashboards
ii) Rule based classifier
iii) Classification and Prediction.