

Hall Ticket Number:

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

Code No. : 251 O

VASAVI COLLEGE OF ENGINEERING (Autonomous), HYDERABAD
M.C.A. II Year II-Sem (Old) Examinations, April-2018

Data Mining

Time: 3 hours

Max. Marks: 70

Note: Answer ALL questions in Part-A and any FIVE from Part-B

Part-A (10 × 2 = 20 Marks)

1. What is the significance of outlier analysis?
2. List the various techniques commonly used for data transformation.
3. List the various functions performed by back-end tools and utilities in Data Warehouse.
4. How are Concept hierarchies useful in OLAP?
5. Mention the benefits of Market Basket analysis.
6. Explain in short the Hash –based technique proposed to improve the efficiency of Apriori-based mining.
7. How is tree pruning useful in Decision tree induction?
8. Illustrate in short how accuracy of a classifier is estimated using k-fold cross validation.
9. List out the major tasks involved in Clustering Evaluation.
10. Write briefly about Statistical Modeling of Networks.

Part-B (5 × 10 = 50 Marks)
(All bits carry equal marks)

11. a) Describe any three challenges of Data Mining with respect to data mining methodology.
b) Describe the steps involved in Data Mining when viewed as a process of Knowledge Discovery.
12. a) Describe in detail a typical three-tier Data warehousing Architecture.
b) Explain the various schemas commonly used for modeling Multidimensional data.
13. a) A database has five transactions. Let min up = 60% and min conf = 80%.

TID	Items brought
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}

Find all frequent items using Apriori frequent pattern mining algorithm.

- b) Discuss briefly about Pattern Evaluation methods.
14. a) Why is naïve Bayesian Classification called “naïve”? Briefly outline the major ideas of naïve Bayesian Classification.
b) Discuss in detail the Learning phase of Bayesian Belief Network.

- 15. a) Explain the k-Means Clustering Algorithm with an example.
b) Briefly describe the different approaches behind the Statistical-based Outlier detection.
- 16. a) Explain in detail the various Data Reduction Techniques.
b) Compare the OLAP and OLTP Systems.
- 17. Answer any *two* of the following:
 - a) Explain in detail the procedure of Mining Multilevel Association rules.
 - b) Explain the Backpropagation Classification Algorithm.
 - c) How do you perform the Similarity Search on Time Series Data? Explain.
