

Hall Ticket Number:

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

Code No.: 16605 O

VASAVI COLLEGE OF ENGINEERING (Autonomous), HYDERABAD
B.E.(IT) III Year II-Semester Old Examinations, May-2018

Data Warehousing and 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. List any two data reduction techniques.
2. Define Data Preprocessing.
3. Besides support and confidence, list and define any three other interesting measures for association rules.
4. How do you transform star schema into snowflake schema? Discuss with an example.
5. State why Associative classification is needed in Data mining?
6. How bagging and boosting are helpful in improving classifier and predictor accuracy?
7. Compute Euclidean Distance and Minkowski distance between the two objects represented by the tuples (22,1,42,10) and (20,0,36,8).
8. What is an outlier? How it is identified for the data points.
9. Define Time-series database.
10. Compare Text mining and World wide web mining.

Part-B (5 × 10=50 Marks)

- 11.a) Describe data mining functionalities in detail along with an example. [5]
- b) Illustrate the significance of Data Transformation with an example. [5]
- 12.a) Draw the architecture of a data warehouse and describe briefly the various components involved in it. [5]
- b) Write FP growth algorithm and explain with an example by constructing FP tree. [5]
- 13.a) Why is naïve Bayesian classification called “naïve”? Briefly outline the major ideas of naïve Bayesian classification. [6]
- b) Write the similarity measures for a binary variable in case of both symmetric and asymmetric cases. Also, explain dissimilarity between Categorical variables. [4]
- 14.a) Write k-means partitioning algorithm and explain with an example [5]
- b) Why is it that BIRCH encounters difficulties in finding clusters of arbitrary shape but OPTICS does not? How do you modify BIRCH to determine clusters of arbitrary shape? [5]
- 15.a) Explain Multimedia Data Mining along with an example. [4]
- b) Junk e-mail is one of the most annoying things in web-based business. Design an effective scheme that can be used to filter out junk email effectively. [6]

- 16.a) Explain Data Discretization and Concept Hierarchy Generation with an example. [5]
- b) Direct Hashing and Pruning (DHP) is an improvement over Apriori. Let {a,b,c} and {d,e,f} are two transactions and all 6 items are frequent. Then find the percentage improvement in the number of candidate 2-sets generated in DHP over Apriori. [5]
- 17. Answer any *two* of the following:
 - a) Why tree pruning is useful in decision tree induction? What is a drawback of using a separate set of tuples to evaluate pruning? [5]
 - b) Illustrate the significance of Hierarchical clustering with an example. [5]
 - c) i) What are the basic measures used for text retrieval? [5]
 - ii) How dimensionality reduction is performed using LSI on text data?

