Regd. No. | | | | | | | | | | |

VIGNAN'S
Foundation for Science, Technology & Research
(Deemed to be University)
-Estd. u/s 3 of UGC Act 1956

**Regulation: R13** **Code No: CS425/11**

IV B.Tech. I Semester Regular Examinations - December, 2018

# DATAWAREHOUSING AND DATA MINING

Time: **3** hours (IT) Max. Marks: **60**

## SECTION – A

**Answer all ten questions** **10×1M=10M**

1. _____ is a comparison of the general features of target class data objects with the general features of objects from one or a set of contrasting classes.
2. The 0-D cuboid which holds the highest level of summarization is called as _____.
3. _____ is a measure that must be computed on the entire data set as a whole.
4. Histogram partition the values for an attribute into disjoint ranges called _____.
5. _____ is a process that abstracts a large set of task relevant data in a database from a relatively low conceptual level to higher conceptual levels.
6. _____ constraints specify the type of knowledge to be mined.
7. _____ learner will construct a generalization model before receiving new tuples to classify.
8. _____ is based on the establishment of equivalent classes within the given training data.
9. Write a formula to calculate Euclidean distance.
10. _____ is a grid based multi resolution clustering technique in which spatial area is divided into rectangular cells.

## SECTION – B

**Answer all five questions** **5×2M= 10M**

11. Name the data mining tasks.
12. State the difference between data warehouse and data mart.
13. What is meant by sequential data mining?
14. Define the term association rules.
15. List the requirements of clustering.

## SECTION – C

**Answer all four questions** **4×5M = 20M**

16. Describe various data mining issues.

**(OR)**

17. Write short notes on architecture of data warehouse.

18. Write notes on data pre-processing.

**(OR)**

19. How will you measure the dispersion of data? Explain it.

20. Outline on data cleaning.

**(OR)**

21. Illustrate data transformation.

22. What are the various optimization techniques used for efficient computation of data cubes? Discuss about it.

**(OR)**

23. Describe frequent pattern mining.

## SECTION – D

**Answer all two questions** **2×10M= 20M**

24. Express an algorithm for inducing a decision tree from training tuples. Elaborate it.

**(OR)**

25. With example, explain SVM.

26. Discuss about density based methods.

**(OR)**

27. How will you cluster high dimensional data? Explain it.