IV/IV B.Tech. DEGREE EXAMINATIONS, NOVEMBER- 2019
First Semester
COMPUTER SCIENCE ENGINEERING
DATA ENGINEERING

Time: Three Hours **Maximum marks:60**

Answer Question No.1 Compulsory	**6X2=12 M**
Answer ONE Question from each Unit	**4X12=48 M**

1. a) Mention different data transformation techniques

   b) Brief on Entropy and Gain

   c) Fact and Dimension tables

   d) What is Jaccard coefficient?

   e) Compare and contrast classification and clustering

   f) Define support and confidence

UNIT-I

2. a) Classify and explain kinds of data that can be used for data mining.

   b) Explain various data mining tasks giving example for each.

(OR)

3. a) What is the need for preprocessing? List various preprocessing techniques.

   b) Explain the design of a data warehouse and steps in its implementation.

UNIT-II

4. a) Define sampling. Mention the different types of approaches for sampling and explain each in detail.

   b) Explain the concept of feature subset selection and its importance.

(OR)

5. Enumerate on the FP-Tree representations and with an example, explain the frequent item set generation in FP-Growth algorithm.

P.T.O

UNIT-III

6.   What is cluster analysis? Explain additional issues of K-means. Describe K-means as an optimization problem.

(OR)

7.   a)   What are outliers? How are they detected? Explain any one technique in detail.

    b)   Describe the center based approach for DBSCAN.

UNIT-IV

8.   a)   Mention the characteristics of Bayesian Belief Networks. Explain.

    b)   Explain spatial data cube construction and spatial OLAP.

(OR)

9.   a)   How does latent semantic indexing reduce the size of the term frequency matrix? Explain.

    b)   Describe the construction of a multilayered web information base.

    c)   What kinds of association can be mined in multimedia data? Explain in brief.

IV/IV B.Tech. (Supple) DEGREE EXAMINATIONS, JUNE- 2019
First Semester
COMPUTER SCIENCE ENGINEERING
DATA ENGINEERING

Time: Three Hours   **Maximum marks:60**

Answer Question No.1 Compulsory   **6X2=12 M**

Answer ONE Question from each Unit   **4X12=48 M**

1. a) Define data warehouse

    b) Rule-based hierarchies with suitable example.

    c) Metadata

    d) Mention measure of classifier accuracy

    e) Define outlier

    f) What is a sequence database?

UNIT-I

2. a) What are the major issues in Data Mining? Explain.

    b) What is meant by Data quality? Discuss.

(OR)

3. a) What is data mining? List and describe the motivating challenges of data mining.

    b) What is an attribure? Explain different types of attributes with examples.

UNIT-II

4. a) With an example, explain the FP-growth algorithm.

    b) Write short notes on discretization and binarization.

(OR)

5. a) What are the similarities and dissimilarities between simple attributes? Explain.

    b) Discuss about computational complexity of frequent itemset generation.

P.T.O

UNIT-III

6.   a)   Discuss in detail about agglomerative hierarchical clustering.

       b)   Explain clustering using representatives algorithm with example.

(OR)

7.   a)   Give an overview of cluster analysis.

       b)   Write the DBSCAN algorithm and explain how to the select DBSCAN Parameters.

UNIT-IV

8.   a)   Describe the general approach to solving a classification problem with a suitable example.

       b)   Discuss the back propagation algorithm for neural network-based classification of data.

(OR)

9.   Write short notes on the following

       a)   Web usage mining

       b)   Search engines

       c)   Multimedia databases Vs Transactional databases.