BE6-R4: DATA WAREHOUSING & DATA MINING

NOTE:

1. **Answer question 1 and any FOUR from questions 2 to 7.**
2. **Parts of the same question should be answered together and in the same sequence.**

Time: 3 Hours **Total Marks: 100**

1.

a) State differences between predictive modeling and descriptive modeling.

b) Explain the differences between discrete and continuous data and give three examples of each?

c) What is meant by hierarchical clustering? Explain.

d) What are the basic features and applications of Naïve Bayes algorithm?

e) What are the strengths and weaknesses of Decision trees?

f) What are the issues in data warehousing?

g) Compare and Contrast OLTP and Data Warehousing systems.

(7x4)

2.

a) Explain the architecture of Data. Also state the features of Data warehouse?

b) State why tree pruning is useful in decision tree induction? What are the approaches for tree pruning?

c) Explain three pre-processing steps required before feeding data into a data warehouse.

(6+6+6)

3.

a) Explain the K-means clustering algorithm.

b) Explain in detail the FASMI characteristics of OLAP systems?

(9+9)

4.

a) Explain the Kohonen's self-organizing Maps? State the applications of Neural Network?

b) Give the difference between hierarchical and non-hierarchical clustering? Consider eight data points with two dimensions x and y as candidate for agglomerative clustering. The data points are: P1(1,1), P2(6,7), P3(4,6), P4(5,7), P5(5,2), P6(2,3), P7(1,2), P8(3,1). Perform Agglomerative Clustering for the above points using Distance function and show all the steps involved.

(9+9)

5.

a) What are the steps involved in Data Transformation for making data suitable for Mining?

b) What is Data Cleaning? Explain the different data cleaning approaches?

c) How does snowflake schema overcome the disadvantages of star schema?

(6+6+6)

6.

a) What are support and confidence in association Rule mining? Write the Apriori algorithm for association rule mining.

b) Explain through example as how and why information gain is used to construct a decision tree?

c) Explain how temporal and spatial databases are different from the normal databases?

(7+6+5)

7. Write short notes on any **three** of the followings:

a) Multimedia Databases

b) Semantic Web

c) Collaborative Filtering

d) Hyperlink induced topic search (HITS) algorithm

(3x6)