BE6-R4: DATA WAREHOUSING AND 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) What is Data Scrubbing? Why it is important while building a data warehouse?
b) Which schema is best suitable for data mart and data warehouse and why?
c) Give some advantages of OLAP systems.
d) Differentiate between Discrete attribute and Continuous attributes.
e) How is prediction different from classification?
f) Explain four views which must be considered during the design of a data warehouse.
g) Define meta data and explain meta data repository.

(7x4)

2.

a) What are Neural Networks? Explain multilayer feed-forward network with diagram.
b) List out the advantages and disadvantages of snowflake schema.
c) Write a short note on web mining.

(6+6+6)

3.

a) Explain regression models.
b) Compare OLTP and OLAP systems.
c) What is OLAP server architecture like?

(6+6+6)

4.

a) Write a short note on cluster analysis.
b) Discuss various ways of handling missing values during data cleaning.
c) Explain the KDD (Knowledge Discovery Process). What are the major issues in data mining (DM)?

(6+6+6)

5.

a) A data-warehouse for a university consists of four dimensions – student, course, semester and instructor. Two measure are maintained – count and average-grade. Average grade is the average grade for a course, semester, instructor at the lowest level; count is the number of students. Draw a star schema for the data-warehouse.
b) Write short note on the following:
 i) Interquartile Range
 ii) Five- number summary

(10+8)

6.

a) How crossover and mutation is performed in Genetic Algorithm? Explain with example.
b) What is noise? Explain data smoothing methods as noise removal technique to divide given data into bins of size 3 by bin partition (equal frequency), by bin means, by bin median and by bin boundaries. **Consider the data: 10 , 2 , 19 , 18 , 20 , 18 , 25 , 28 , 22.**
c) What is Hypo Thesis? How is it tested?

(8+6+4)

7.

a) Why naïve Bayes classification is called naïve? Briefly outline the major ideas of naïve Bayes classification.

b) Find all frequent item sets in following transactional database using Apriori (minimum support is 40%). Also, write down steps used in each pass.

TID	A	B	C	D	E
T_1	1	1	1	0	0
T_2	1	1	1	1	1
T_3	1	0	1	1	0
T_4	1	0	1	1	1
T_5	1	1	1	1	0

(8+10)