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 are the indexes supported in a data warehouse?
b) What are the Contextual information and its type in Data warehouse?
c) Compare and contrast OLTP and Data Warehousing Systems.
d) Formally define association rule mining problem.
e) Explain the SEMMA process Model of data mining?
f) What is bagging? How does it improve performance?
g) How does snowflake schema overcome the disadvantages of star schema?

(7x4)

2.

a) What are the characteristics of a data warehouse?
b) Draw and explain the Data warehouse Model? State the structure of data inside the data warehouse.
c) What are the steps involved in Data Transformation for making data suitable for Mining?

(6+6+6)

3.

a) Explain the various schema of the data warehouse?
b) Explain the data selection, cleaning, enrichment, coding and analysis of knowledge discovery process?

(9+9)

4.

a) Consider the following data set shown in below table where each record represents the weather condition and class attributes shows whether people generally play sports in that weather condition or not.

Outlook	Temperature (Fahrenheit)	Humidity (%)	Windy	Class
Sunny	75	70	True	Play
Sunny	80	90	True	Don't play
Sunny	85	85	False	Don't play
Sunny	72	95	False	Don't play
Sunny	69	70	False	Play
Overcast	72	90	True	Play
Overcast	83	78	False	Play
Overcast	64	65	True	Play
Overcast	81	75	False	Play
Rain	71	80	True	Don't play
Rain	65	70	True	Don't play
Rain	75	80	False	Play
Rain	68	80	False	Play
Rain	70	96	False	Play

Draw the decision tree for the above table using ID3 algorithm?

b) Write an algorithm for K-nearest neighbor classification.

(12+6)

5.

a) How crossover and mutation is performed in Genetic algorithms? Explain with example?

b) Consider the task of clustering people into two clusters based on their heights and weights given in table using K-Means Algorithms?

ID	NAME	HEIGHT(Inch)	WEIGHT(kg)
X1	Ram	64	60
X2	Shayam	60	61
X3	Gita	59	70
X4	Mohan	68	71

(9+9)

6.

a) Consider the dataset given below and perform the Apriori algorithm with a minimum support of 60%.

TID	ITEMS
1	A,C,D,F,G,I,M,P
2	A,B,C,F,L,M,O
3	B,F,H,J,O
4	B,C,K,P
5	A,C,E,F,L,M,N,P

b) Explain the FP-Growth Algorithm and show how to construct the FP-tree?

c) Explain the temporal data mining tasks and give data mining techniques involved in such analysis methods?

(7+6+5)

7. Write short notes on the followings:

a) Hypothesis testing

b) Linear and Non-linear regression

c) Termporal and spatial database models

(6+6+6)