BE6-R4: DATA WAREHOUSE 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) Explain the CRISP-DM process model of data mining?
b) Explain the difference between Nominal, ordinal, ratio variables with example.
c) Is it true that data mining is a step in the Knowledge discovery process? Justify your answer by considering an example for it?
d) Explain the importance of the Extraction, Transformation and Loading process in data warehouse data?
e) What are the components of data warehouse? Explain each in brief?
f) What are Meta data? Explain its type in brief?
g) What is Hypothesis Testing? How it is generated? Explain with an Example.

(7x4)

2.

a) What is the difference between On Line Transaction Processing systems (OLTP) and Data Warehousing Systems? Explain with an example.
b) Draw and explain the hierarchy of data which helps users to build a data warehouse.
c) Explain the database design methodology for the data warehouse?

(6+6+6)

3.

a) What are the Association Rules? Describe the common types of rules produced by Association Rules?
b) How to build an Association Rule? Give the steps or concerns while building an Association Rule?
c) What are Dissociation Rules? How are they generated and explain its importance in business world?

(6+6+6)

4.

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

Name	Gender	Height (In Meters)	Output
A	FEMALE	1.6	SHORT
B	MALE	2	TALL
C	FEMALE	1.9	MEDIUM
D	FEMALE	1.88	MEDIUM
E	FEMALE	1.7	SHORT
F	MALE	1.85	MEDIUM
G	FEMALE	1.6	SHORT
H	MALE	1.7	SHORT
I	MALE	2.2	TALL
J	MALE	2.1	TALL
K	FEMALE	1.8	MEDIUM
L	MALE	1.95	MEDIUM
M	FEMALE	1.9	MEDIUM
N	FEMALE	1.8	MEDIUM
O	FEMALE	1.75	MEDIUM

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

b) What are distances based algorithms for classification? State K Nearest Neighbors algorithm with its drawback?

(12+6)

5.

a) Consider the following sets of items are given to form clusters:
{2,4,10,12,3,20,30,11,25} with K=2..
Using K-Mean Clustering algorithm find the clusters values?

b) How the clustering is different in case of large databases? Explain in detail?

(12+6)

6.

a) Give an overview of the IBM data mining tools explaining the basic layout and model which is Supported by the tool.

b) What are the ways to model the temporal events? Explain each of them in detail?

c) Explain Spatial Rules? Also explain how spatial classification, clustering and association is done in spatial mining?

(6+6+6)

7.

a) Using an example and explain the process of Selection, crossover and mutation in genetics?

b) What are Feed Forward Neural Networks? How are they beneficial in Artificial Neural Networks?

c) How to choose an efficient training data in case of Artificial Neural Networks? Explain what are Self-Organizing Maps (SOMs)?

(6+6+6)