

C5-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) Differentiate between OLAP and OLTP.
- b) Differentiate between Classification and clustering.
- c) Define Composite Aggregates with an example.
- d) What are decision trees?
- e) Describe the types of situations that produce sparse or dense data cubes.
- f) What is Concept Hierarchy? Describe why Concept Hierarchies are useful in data mining?
- g) What is the use of Regression?

(7x4)**2.**

- a) What are the similarities and the differences between Star schema and snowflake schema? State their advantages and disadvantages.
- b) What is Data generalization? Discuss basic principle of Attribute Oriented Indication.
- c) Discuss distributive, algebraic and holistic measures.

(8+7+3)**3.**

- a) What is metadata? Explain.
- b) What are data marts? How they are different from traditional data warehouses?
- c) What is multidimensional data model? Explain.

(6+6+6)**4.**

- a) How multilevel association rules can be mined efficiently using concept hierarchy?
- b) What is the purpose of Apriori Algorithm?
- c) List out the OLAP operations in multidimensional data model.

(6+6+6)**5.**

- a) Compare the advantages and disadvantages of eager classification versus lazy classification.
- b) Briefly describe the classification processes using genetic algorithms, rough sets and fuzzy sets.

(9+9)**6.**

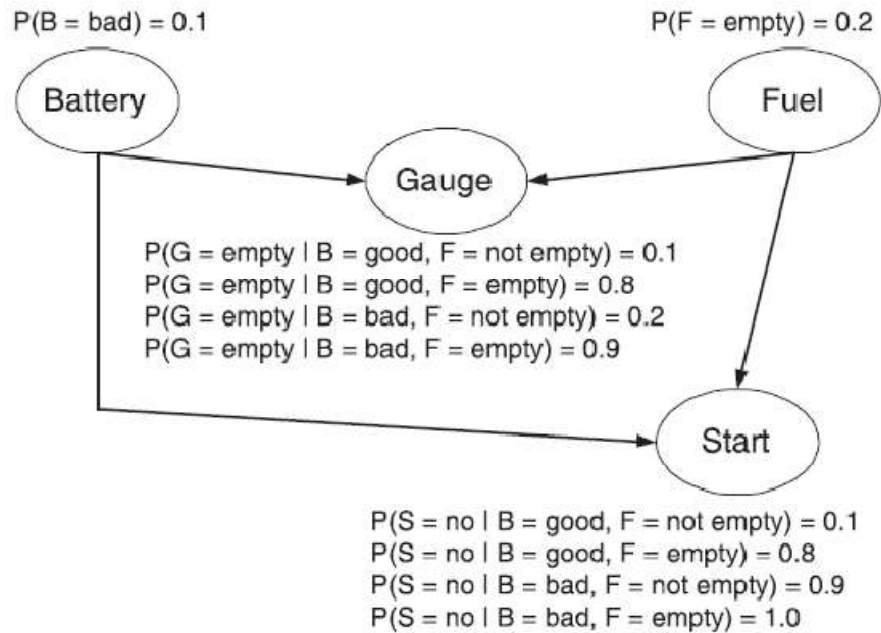
- a) Briefly describe the clustering methods. Give examples in each case.
- b) Briefly discuss the data smoothing techniques.

(12+6)

7.

a) Given then Bayesian network shown in below figure, compute the following probabilities:

- i) $P(B=\text{good}, F=\text{empty}, G=\text{empty}, S=\text{yes})$
- iii) $P(B=\text{bad}, F=\text{empty}, G=\text{not empty}, S=\text{no})$
- iii) Given that the battery is bad, compute the probability that car will start



- b) Write a short note on web usage mining.
- c) What is time series database?

(8+6+4)