

Date: 5-03-2022

Registration number:

ST. JOSEPH’S COLLEGE (AUTONOMOUS), BENGALURU-27

MSc COMPUTER SCIENCE III SEMESTER

SEMESTER EXAMINATION: OCTOBER 2021

(Examination conducted in January-March 2022)

**CS 9118 – Data Warehousing and Data Mining**

Time- 2 ½ hrs Max Marks-70

This question paper contains **Two** printed pages

**Part A**

**Answer any seven questions from the following 7x10=70**

1. a)Explain the steps involved for the design and construction of data warehouses. (4 marks)

b) Explain the role of Meta data in a data warehouse. (3 marks)

c)Explain a 4 dimensional data cube. (3 marks)

1. a)It is always said that datawarehouse should be constructed separately, Justify . (5 marks)

(b) For a Supermarket Chain consider the following dimensions, namely Product, store, time, promotion. The schema contains a central fact table, sales facts with three measures unit\_sales, dollar\_sales and dollar\_cost. Design star schema for this application. (5 marks)

1. a)Describe the following OLAP operations using an example
   1. Slice
   2. Dice

iii) Rollup (6 marks)

b)Explain ROLAP Vs MOLAP (4 marks)

1. a)Describe Extraction, Transformation and Loading in datawarehousing. (5 marks)

b) What is data cube measures? Explain the categorization of measures. (5 marks)

1. a)Discuss about mining time-series and sequence data. (4 marks)

b) Partition the given data into 4 bins using Equi-depth binning method and perform smoothing according to the following methods.

Smoothing by bin mean

Smoothing by bin median

Smoothing by bin boundaries.

Data: 11,13,13,15,15,16,19,20,20,20,21,21,22,23,24,30,40,45,45,45,71,72,73,75.(6 marks)

6. a) Explain with an example how to build a decision tree using ID3 algorithm. (7 marks)

b} What is market basket analysis? (3marks)

1. a)Discuss the activities of data cleaning with the process associated with it. (5 marks)

b)Explain briefly hierarchical clustering techniques. (5 marks)

1. a)What are the various data pre processing techniques? (3 maks)

b)Demonstrate how Bayesian classification helps in predicting class

membership probabilities. (7 marks)

1. a)How does the k-means algorithm work? Explain with example. (5 marks)

b)What are the current trends in datamining? Explain any two application areas. (5 marks)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*