

**EXAMINATION PAPER**

**FACULTY: COMPUTER SCIENCE & MULTIMEDIA**

**COURSE: MASTER OF COMPUTER SCIENCE**

**YEAR/SEMESTER: FIRST / SECOND**

**MODULE TITLE & CODE: Data Warehousing and Big Data: DWB121**

**DATE:**

**TIME ALLOWED: 3 HOURS  START: 1:00 PM FINISH: 4:00 PM**

**Instruction to candidates**

1. This question paper has THREE (3) Sections.
2. Answer all **7** questions in Section A, VSAQ
3. Answer **7** questions out of 9 in Section B, SAQ.
4. Answer **2** questions out of 3 in Section C, LAQ.
5. No scripts or answer sheets are to be taken out of the Examination Hall.
6. Lengthy answers do not win more marks. Students are advised to write clear and concise answer with valid relevant examples

***Do not open this question paper until instructed***

*(Candidates are required to give their answers in their own words as far as practicable)*

**SECTION A**

**Very Short Answer Questions**

**Attempt all seven (7) questions [2 × 7 = 14]**

1. What are the characteristics of OLAP??
2. What is data aggregation?
3. Why can data be incomplete?
4. What is concept hierarchy specification?
5. What is the need of association rule mining?
6. What are the issues in classification and prediction?
7. List the basic measures for text retrieval.

**SECTION B**

**Short Answer Questions**

**Attempt only seven (7) questions out of eight (9) questions [7 × 8 = 56]**

Q.1. Why business analysis does is motivated for data mining? Does data mining generate and help to expand a business? Explain. [Unit 1 Introduction]

Q.2. Explain the way to construct a data warehouse system .[Unit 2 Data warehouse for Data Mining]

Q.3. “How are data actually stored in ROLAP and MOLAP architectures? Explain. [Unit 3 OLAP Technology for Data Mining]

Q.4. When computing a cube of high dimensionality, we encounter the inherent curse of the dimensionality problem. How can we deal with curse of dimension? Explain briefly [Unit 4 Tuning For Data Warehouse]

Q.5. Compare linear and logistic regression analysis with suitable figure and example. [Unit 5 Data Mining Technique]

Q.6. How can users specify the forms of presentation and visualization to be used in displaying the discovered patterns? [Unit 6 Data Mining Query Language]

Q.7. Association rule mining often generates a large number of rules. Discuss effective methods that can be used to reduce the number of rules generated while still preserving most of the interesting rules. [Unit 7 Association Rule]

Q.8. Explain data transformation and reduction technique for classification and prediction. [Unit 8 Cluster analysis, Classification and Prediction]

Q.9. Design the process for web usage mining. [Unit 9 Advance concept in data mining]

Section C

Long Answer Questions

Attempt any two (2) questions out of three(3) questions

(2\*15=30)

Q.1. Consider a franchise of retail stores having the business setup only in Nepal. The analysis requirements of the franchise include getting to know which items are purchased together by each individual consumer. They wish to know the sales figures in terms of sales amount in Rupees as well as quantity of the individual stores and also for the city, state and region in which they are located. They also wish to know how sales differ over different months, quarters and years; how sales figures change with the hour of the day – e.g., how sales of morning hours are different from sales of evening hours, etc.; how buying habits of male consumers are different from that of the female consumers; how buying habits of married consumers are different from that of the unmarried consumers; how buying habits of consumers vary with their native languages.

1. Design a star schema for such a data warehouse clearly identifying the fact table and dimension tables, their primary keys, and foreign keys.
2. Write one SQL statement that runs on your schema and returns the number of purchases made during the evening hours by the married customers and the unmarried customers in the month of May 2020. [Unit 2 Data warehouse for Data Mining]

Q.2. If your dataset contains missing value, discuss the basic analysis and the corresponding decisions you will take in the preprocessing phase of data mining process. Recommend a solution for data cleaning process and present with example. [Unit 5 Data mining Techniques ]

Q.3. Trace the results of using the Apriori algorithm on the grocery store example with support threshold s=33.34% and confidence threshold c=60%. Show the candidate and frequent itemsets for each database scan. Enumerate all the final frequent itemsets. Also indicate the association rules that are generated and highlight the strong ones, sort them by confidence. [Unit 7 Association Rules]

