



**PAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY**

FACULTY OF COMPUTING AND INFORMATICS

DEPARTMENT OF COMPUTER SCIENCE

| | |
|--|-----------------------------|
| QUALIFICATION: BACHELOR OF COMPUTER SCIENCE | |
| QUALIFICATION CODE: 07BACS | LEVEL: 7 |
| COURSE: DATA AND WEB MINING | COURSE CODE: DWM710S |
| DATE: JULY 2022 | SESSION: 1 |
| DURATION: 3 HOURS | MARKS: 100 |

| | |
|--|------------------------|
| SECOND OPPORTUNITY/SUPPLEMENTARY EXAMINATION QUESTION PAPER | |
| EXAMINER: | MR GEREON KOCH KAPUIRE |
| MODERATOR: | MRS JOY ANURIOHA |

THIS QUESTION PAPER CONSISTS OF 3 PAGES
(Excluding this front page)

INSTRUCTIONS

1. Answer all the questions.
2. When writing take the following into account: *The style should inform than impress, it should be formal, paragraphs set out according to ideas or issues and the paragraphs flowing in a logical order. Information provided should be brief and accurate.*
3. Please, ensure that your writing is legible, neat and presentable.

Question 1

[4 Marks]

During a workshop, a Senior Data Expert says, "People do not have time to look at data". Would you agree with that or not and why?

Question 2

[7 Marks]

What are the various areas where data mining is drawn? Please give seven examples.

Question 3

[4 Marks]

There are a number of data mining functionalities. Explain the difference between descriptive and predictive.

Question 4

[4 Marks]

Classification models are used to predict the class label of objects for which the class label is unknown. How is the derived model presented?

Question 5

[6 Marks]

Are all patterns of interest to potential users? What makes a pattern interesting?

Question 6

[6 Marks]

Explain the difference between Ordinal Attributes and Numeric Attributes. Give an example for each.

Question 7

[2 Marks]

By illustration with a practical example, please explain, "Data mining turns a large collection of data into knowledge".

Question 8

[3 Marks]

As a Senior Data Miner for a Marketing Company, you are tasked to visualize data. You have decided to use a Scatter Plot. How would you explain your decision as to why you would use that?

Question 9

[10 Marks]

Data in the real world is dirty: Lots of potentially incorrect data, e.g., instrument being faulty, human or computer error, transmission error, etc. Describe various methods for handling missing data.

Question 10

[6 Marks]

What are the best qualities to assess the quality of data?

Question 11

[2 Marks]

What would you discover when performing Frequent Itemset Mining?

Question 12

[4 Marks]

Explain the difference between Data Cleaning and Data Integration.

Question 13

[6 Marks]

Outliers are often discarded as noise. However, one person's garbage could be another's treasure. For example, exceptions in credit card transactions can help us detect the fraudulent use of credit cards. Using fraudulence detection as an example, propose two methods, which can be used to detect outliers and discuss which one is more reliable.

Question 14

[2 Marks]

Operational databases store huge amounts of data, and you may wonder, "*Why not perform online analytical processing directly on such databases instead of spending additional time and resources to construct a separate data warehouse?*". What would the reason be for this?

Question 15

[5 Mark]

You are an experienced Data Miner who applies the disciplines of Data Science and Philosophy together in tackling your daily duties. A junior Data Miner wants to understand the concept of Data Science to which you are applying. How or what would you explain to the junior Data Miner?

Question 16

[3 Marks]

You just joined a Data Mining Consulting Company as an Intern for five (5) months. During the last week of your internship program, your supervisor asks you to explain the concept of attribute construction during preprocessing. Please provide an example with the explanation.

Question 17

[12 Marks]

You are employed at Tech Mining Consultants as a Mining Engineer. Part of your duties is to teach Data Mining Interns data mining concepts. During a meeting, one intern wants you to list the most important data mining techniques. What are the different techniques used for data mining? Please list and explain them.

Question 18

[14 Marks]

Classification is the process of finding a model that describes and distinguishes data classes or concepts. The classification model can be presented in various forms. Transform the below IF-THEN rules into a Neural Network.

age(X, "youth") AND income(X, "high") -> class(X, "A")

age(X, "youth") AND income(X, "low") -> class(X, "B")

age(X, "middle_aged") -> class(X, "C")

age(X, "senior") -> class(X, "C")

<<<<<End of Exam Paper>>>>>