



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 2019	SESSION: 1
DURATION: 3 HOURS	MARKS: 70

SECOND OPPORTUNITY SUPPLEMENTARY EXAMINATION QUESTION PAPER	
EXAMINER	MR GEREON KOCH KAPUIRE
MODERATOR	MRS JOY ANURIOHA

THIS QUESTION PAPER CONSISTS OF 4 PAGES
(Including this front page)

INSTRUCTIONS

1. Answer all questions.
2. When writing take the following into account: The style should inform than impress, it should be formal, in third person, 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

[8 Marks]

There are preprocessing steps to make the data more suitable for Data Mining. Often, the raw data must be processed in order to make it suitable for analysis.

- a) Give one objective and goal to support the statement.
- b) Provide two examples to support the statement.

Question 2

[4 Marks]

There are many types of data sets, and as the field of data mining develops and matures. A greater variety of data sets become available for analysis. Some objects have spatial attributes, such as positions or areas, as well as other types of attributes. Provide two important practical instances of spatial data.

Question 3

[4 Marks]

A graph can sometimes be a convenient and powerful representation for data. We consider one specific case: *the graph captures relationships among data objects*. Please provide a practical example to describe the case.

Question 4

[3 Marks]

By means of a practical example, please explain spatial auto-correlation.

Question 5

[4 Marks]

There are several strategies for dealing with missing data, each of which may be appropriate in certain circumstances. One strategy is that many data mining approaches can be modified to ignore missing values. Please illustrate this with an example.

Question 6

[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")

Question 7

[4 Marks]

Many people receive duplicate mailings because they appear in a database multiple times under slightly different names. To detect and eliminate such duplicate, to main issues must be addressed. Explain these issues.

Question 8

[4 Marks]

Ideally, data sets are accompanied by documentation that describes different aspects of the data; the quality of the documentation can either aid or hinder the subsequent analysis. Provide a practical example to explain the statement.

Question 9

[4 Marks]

Data pre-processing consist of several different strategies, such as aggregation and sampling. Using practical examples, please explain the difference between dimensionality reduction and the curse of dimensionality.

Question 10

[3 Marks]

Data visualization is the display of information in a graphic or tabular format. The arrangement and selection of items within the visual display is crucial. Explain the difference between these two crucial processes.

Question 11

[4 Marks]

A simple way to visualize the value of a dimension is to use a pixel where the color of the pixel reflects the dimension's value. Given the scenario: *AllElectronics maintains a customer information table which consists of four dimensions, income, credit-limit, transactions-volume, and age*. Using pixel visualization, what can you observe?

Question 12

[4 Marks]

During a press conference, a Data Analyst says, "*Every enterprise benefit from collecting and analyzing its data*". Would you agree with that or not and why? Explain why using a practical example.

Question 13

[2 Marks]

As a Data Miner for a Health Company, you are tasked to visualize data. You have decided to use a Quantile-Quantile plot. How would you explain your decision as to why you would use that?

Question 14

[4 Marks]

By illustration with a practical example, please explain the concept of Market basket analysis.

Question 15

[4 Marks]

Using practical examples, please explain the difference between Cluster analysis and

Evolution analysis.

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