ΠΑΜΙΒΙΑ UΠΙVERSITY
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: JUNE 2019	SESSION: 1
DURATION: 3 HOURS	MARKS: 70

FIRST OPPORTUNITY 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 [6 Marks]

During a press conference, a person interested in Data Mining hears you make the statement *"Data sets differ in a number of ways"*. The person asks you to explain the statement further. As an experienced data miner, you decide to explain to the person.

    a) Provide an example to show the importance of the statement above.

    b) The topic becomes very interesting to the person, that he further asks you to answer the question *"Can most data mining techniques tolerate some level of imperfection in the data?"*. Please provide reasons to defend your decision.

Question 2 [6 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. Sequential data can be thought of as an extension of record data, where each record has a time associated with it. Given the above, provide a practical situation and example, and what patterns can be found.

Question 3 [4 Marks]

A graph can sometimes be a convenient and powerful representation for data. We consider one specific case: *the data with objects themselves are represented as graphs*. Please provide a practical example to describe the case.

Question 4 [4 Marks]

By means of a practical example, please explain the process of data collection error.

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 sometimes missing data can be reliably estimated. Please illustrate this with an example.

Question 6 [9 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 Decision Tree.

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]

Data can contain inconsistent values. Once an inconsistency has been detected, it is sometimes possible to correct the data. Show using an instance when this is detected.

Question 8 [5 Marks]

Some data starts to age as soon as it has been collected. Do you agree with this statement and provide motivation for your decision?

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 aggregation and sampling.

Question 10 [2 Marks]

Data visualization is the display of information in a graphic or tabular format. Some data points can obscure other data points. For example, the shape of an object cannot be used to encode a characteristic of that object if there is only one pixel available to display it. What would you do in this situation?

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.* Can we analyze the correlation between income and the other attributes by visualization?

Question 12 [5 Marks]

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

Question 13 [4 Marks]

As a Data Miner for a Health 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 14 [4 Marks]

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

Question 15 [5 Marks]

Point-of-sale data collection (bar code scanners, radio frequency, and smart card technology) have allowed retailers to collect up-to-the-minute data about customer purchases at the checkout counters of their stores. What can the business do with the data?

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