COLLEGE OF BUSINESS, PEACE, LEADERSHIP AND GOVERNANCE

CIS 406: DATA MINING AND DATA WAREHOUSING

END OF SECOND SEMESTER EXAMINATIONS

MAY/JUNE 2020

LECTURER: Mr Timothy Makambwa

DURATION: 48 HRS

INSTRUCTIONS

You are required to answer questions as instructed in each section

Start **each** question on a new page in your answer booklet

Answer *One* question from this Examination

Credit will be awarded for logical, systematic and neat presentations

QUESTION ONE

a) Explain how clustering can be used in the following areas
i) Marketing   (4 marks)
ii) Land use   (4 marks)
iii)Insurance   (4 marks)
b)Outline the steps of genetic learning algorithm   (4 marks)
c) Discuss how a decision tree is used as an algorithm in data mining
                                                               (6 marks)
d)Using an example explain the term Information gain in the context
of data mining   (4 marks)
e) Explain the basic steps used in neural networks back propagation
learning algorithm   (6 marks)
f)Using an example elaborate on the syntax of MDX   (6 marks)
g)Define the term neural network   (4 marks)
h)Discuss the following as used in knowledge discovery
i) Data Selection   (4 marks)
ii) Cleaning   (4 marks)
iii)Enrichment   (4 marks)

QUESTION TWO
   a)

DBMS schemas for decision support

 The basic concepts of dimensional modelling are: facts, dimensions
and measures. A fact is a collection of related data items,
consisting of measures and context data. It typically represents
business items or business transactions. A dimension is a collection
of data that describe one business dimension. Dimensions determine
the contextual background for the facts; they are the parameters over
which we want to perform OLAP. A measure is a numeric attribute of a
fact, representing the performance or behaviour of the business
relative to the dimensions.
Considering Relational context, there are three basic schemas that
are used in dimensional
Modelling:

1. **Star schema**

2. **Snowflake schema**

3. **Fact constellation schema**

Explain these Three basic schemas in detail   (25 marks)

b)

OLTP vs OLAP

OLTP stands for On Line Transaction Processing and is a data modelling approach typically used to facilitate and manage usual business applications. Most of applications you see and use are OLTP based. OLTP technology used to perform updates on operational or transactional systems (e.g., point of sale systems)
OLAP stands for On Line Analytic Processing and is an approach to answer multidimensional queries. OLAP was conceived for Management Information Systems and Decision Support Systems. OLAP technology used to perform complex analysis of the data in a data warehouse. Differentiate the two.   (25 marks)

QUESTION THREE

a) What motivated data mining? Why is it important?

The major reason that data mining has attracted a great deal of attention in information industry in recent years is due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from business management, production control, and market analysis, to engineering design and science exploration.

Explain detail Architecture of a typical data mining system/Major Components.   (25 marks)

b)

Data mining functionalities/Data mining tasks: what kinds of patterns can be mined?

Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks. In general, data mining tasks can be classified into two categories:

• Descriptive

• Predictive

Descriptive mining tasks characterize the general properties of the data in the database. Predictive mining tasks perform inference on the current data in order to make predictions.

Explain these in detail (25 marks)

END OF EXAMINATION