DEDAN KIMATHI UNIVERSITY OF TECHNOLOGY

UNIVERSITY EXAMINATIONS 2016/2017

FOURTH YEAR **SEMESTER** 1 EXAMINATION FOR THE DEGREE OF **BACHELOR OF SCIENCE IN CRIMINOLOGY**

BSM 2419: DATA MINING

DATE: 22ND JUNE 2017   **TIME: 11.00-01.00 P.M**

Instructions: Answer Question 1 and Any Other Two.

Question One (30 Marks)
a)   Define the following terms as used in data mining and data warehousing:
   i) Data mining   **(1mark)**
   ii) Meta data   **(1 mark)**
   iii) Data Mart   **(1 mark)**
   iv) Dimension   **(1 mark)**
   v) Online Analytical Processing (OLAP)   **(1 mark)**
b) Outline **THREE** application areas of data mining   **(3 marks)**
c) i)   Explain **THREE** types of data warehouse application   **(6 marks)**

   ii) Distinguish between data warehouse and operational database   **(6 marks)**

d) Explain the **TWO** categories of functions involved in data mining **(4 marks)**

e) Describe the **THREE** data warehouse models   **(6 marks)**

Question Two (20 Marks)
a) i)   Distinguish between data warehouse and data warehousing   **(2 marks)**

   ii) Outline **FOUR** characteristics exhibited by a data warehouse that supports the
        management's decision making process.   **(4 marks)**

b) Explain **FOUR** reasons why clustering is required in data mining   **(4 marks)**

c) With the aid of a well labeled diagram, explain the data warehouse three-tier architecture
                                                                         (10 marks)
Question Three (20 Marks)
a) Explain **FIVE** advantages of Neural Networks   **(5 marks)**

b) With the aid of a well labeled diagram, explain the steps involved in the *Knowledge Discovery in Databases (KDD)* process **(15 marks)**

Question Four (20 Marks)

a) Outline **THREE** reasons why partitioning is necessary in a data warehouse **(3 marks)**
b) Explain **FIVE** criteria used in comparing the methods of Classification and Prediction
(5 marks)
c) Describe the **THREE** fundamental measures used in assessing the quality of text retrieved
(6 marks)

d) Data mining is soon becoming a trend and before it develops into a conventional, mature and trusted discipline, many pending issues still needs to be addressed. Describe the following pending issues in data mining: **(6 marks)**

   i) Mining methodology issues
   ii) Performance issues
   iii) Data source issues

Question Five (20 Marks)

a) Outline **FIVE** methods of clustering **(5 marks)**

b) With the aid of a suitable illustration explain rule-based classifier **(3 marks)**

c) Describe **FOUR** ways in which data mining systems can be categorized **(8 marks)**
d) Distinguish between *association analysis* and *outlier analysis* as used in data mining
(4 marks)