INTERNATIONAL UNIVERSITY OF EAST AFRICA

'Learning to Succeed'

FACULTY OF SCIENCE AND TECHNOLOGY

END OF SEMESTER EXAMINATIONS - APRIL 2025

PROGRAMME: BSSE

YEAR/SEM: YEAR 4/SEMESTER 2

COURSE CODE: SWE4202

NAME: DATA MINING AND WAREHOUSING

DATE: 2025-04-14

TIME: 2:00-5:00PM

INSTRUCTIONS TO CANDIDATES:

1. Read the instructions very carefully
2. The time allowed for this examination is STRICTLY three hours
3. Read each question carefully before you attempt and allocate your time equally between all the Sections
4. Write clearly and legibly. Illegible handwriting cannot be marked
5. Number the questions you have attempted
6. Use of appropriate workplace examples to illustrate your answers will earn you bonus marks
7. Any examination malpractice detected will lead to automatic disqualification.

DO NOT WRITE ANYTHING ON THE QUESTION PAPER

Section A This section is compulsory

Question 1:

The architecture of data mining involves several key components that work together to extract valuable information from large datasets. These components include the user interface, knowledge base, pattern evaluation, data mining engine, and data warehouse. Answer the following questions based on your understanding of these components and their roles in data mining:

a. a. Describe the architecture of data mining. (6 marks)
b. b. Explain the role of the following components in the data mining architecture:

i. i. User Interface (6 marks)
ii. ii. Knowledge Base (6 marks)
iii. iii. Data Mining Engine (8 marks)
iv. iv. Pattern Evaluation (6 marks)
v. v. Data Warehouse (8 marks)

Section B Answer any three questions from this section

Question 1:

(a) Explain the **key steps** involved in the Apriori algorithm. (6 Marks)
(b) Define **support, confidence, and lift**, and explain their significance in association rule mining. (6 Marks)
(c) Given the following transactions, determine the **frequent itemsets** using a minimum support of 50%. (4 Marks)

Transaction ID	Items Purchased
1	Milk, Bread, Butter
2	2 Milk, Bread
3	Bread, Butter
4	Milk, Butter
5	Milk, Bread, Butter

d. Identify one **association rule** from the frequent itemsets and calculate its confidence. (4 Marks)

Question 2:

a) Define data mining and explain its significance in modern data analysis. (4 Marks)
b) Discuss the motivation behind data mining and identify two major challenges faced in its implementation. (4 Marks)
c) Explain the key tasks of data mining and provide an example of each. (4 Marks)
d) Differentiate between structured and unstructured data and explain two common types of datasets used in data mining. (4 Marks)
e) Describe two real-world applications of data mining and explain how they benefit organizations. (4 Marks)

Question 3:

a) Explain the concept of data quality and discuss three key characteristics of high-quality data. (4 Marks)
b) Describe the role of data preprocessing in data mining and explain why it is necessary. (4 Marks)
c) Briefly explain the following data preprocessing techniques and provide one example for each:
    i. Aggregation
    ii. Sampling
    iii. Dimensionality reduction
    iv. Feature selection (4 Marks)
d) Define discretization and transformation in data preprocessing, and explain how they help improve data mining efficiency. (4 Marks)
e) Explain how similarity and dissimilarity between data objects can be measured and describe one technique used for each. (4 Marks)

Question 4:

a. Explain the term data mining. (4 marks)
b. Explain the major steps involved in the data mining process. (6 marks)
c. Why do we need data mining? (6marks)
d. Explain the key properties of data mining (4marks)

Question 5:

a. Explain the six steps involved in data mining process (12 marks)
b. Explain the issues that are involved in data mining (8 marks)

Question 6:

A supermarket chain wants to analyze customer purchasing behavior using association rule mining. The transactions database contains records of items bought together by different customers.

a) Explain the concept of association rule mining and its importance in data mining and warehousing. (5 Marks)
b) Define the following key measures used in association rule mining:
    a. Support
    b. Confidence
    c. Lift (5 Marks)
c) Given the following transaction dataset:

Transaction ID	Items Purchased
T1	Bread, Butter, Milk
T2	Bread, Diapers, Beer
T3	Bread, Butter, Diapers
T4	Milk, Diapers, Beer
T5	Bread, Milk, Diapers, Beer

If the minimum support threshold is 40% and the minimum confidence threshold is 50%, identify whether the rule {Bread} â?? {Diapers} is a strong association rule. (5 Marks)

a. Describe one real-world application of association rule mining in a business environment. (5 Marks)