



MACHAKOS UNIVERSITY

University Examinations for 2021/2022

SCHOOL OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTING AND INFORMATION TECHNOLOGY

fourth YEAR SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE (INFORMATION TECHNOLOGY) AND

BACHELOR OF (COMPUTER SCIENCE)

SIT408/SCO415: DATA MINING

DATE:

TIME:

INSTRUCTIONS: Answer Question ONE and other TWO Questions

QUESTION ONE (COMPULSORY) (30 MARKS)

- a) Highlight four key properties of data mining (4 marks)
- b) List two approaches in classifying data mining. (2 marks)
- c) Classify the tasks involved in data mining in the context of large e-commerce data warehouse. (6 marks)
- d) Imagine you have been given the role of mining specific data in an Electronic Health Record data warehouse. Demonstrate the six steps you would carry out to mine data about trends in non-infectious diseases in a certain region. (6 marks)
- e) Discuss three aspects, which differentiate the Classification and Prediction Methods in data mining (6 marks)
- f) Analyse Classification by Backpropagation as applied in data mining (6 marks)

QUESTION TWO (20 MARKS)

- a) While mining data, several factors impede such a process, sometimes making it difficult to achieve the objectives. Discuss five such factors. (10 marks)
- b) Enumerate the seven steps used in knowledge discovery process (7 marks)
- c) Highlight three aspects of a data warehouse. (3 marks)

QUESTION THREE (20 MARKS)

- a) Define OLAP (1 mark)
- b) Analyze three data warehouse models, indicating the areas of application for each. (9 marks)
- c) Discuss three analytical processes in OLAP (6 marks)
- d) Differentiate data mining and data ware housing (4 marks)

QUESTION FOUR (20 MARKS)

- a) Analyze four problems that data mining solves, (8 marks)
- b) Explain the decision tree algorithm (2 marks)
- c) Illustrate how the following algorithms are applied in data mining
 - i). Time series algorithm (3 marks)
 - ii). Association algorithm (3 marks)
 - iii). Sequence clustering algorithm (3 marks)
- d) Define data cube. (1 mark)

QUESTION FIVE (20 MARKS)

- a) Highlight three criteria for Frequent pattern mining classification. (3 marks)
- b) Illustrate four methods of moving data between servers and databases in sql server. (8 marks)
- c) Study the table below and answer the following questions.

Transaction ID	Milk	Bread	Butter	Water
1	1	1	0	0
2	1	1	1	0
3	0	0	1	0
4	0	0	0	1
5	1	1	0	1
6	1	0	1	1

- i) Calculate the Support of the item set {milk, bread, water} (3 marks)
- ii) Calculate the confidence of the rule {milk, Bread} => {Butter} (3 marks)
- iii) Calculate the lift of the rule (ii) above (3 marks)

