



MACHAKOS UNIVERSITY

University Examinations for 2017/2018

SCHOOL OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTING AND INFORMATION TECHNOLOGY

**THIRD YEAR FIRST SEMESTER EXAMINATION FOR BACHELOR OF SCIENCE IN
INFORMATION TECHNOLOGY**

DATA MINING AND KNOWLEDGE DISCOVERY

DATE: 13/12/2017

TIME: 8.30-10.30 AM

INSTRUCTIONS: Answer Question ONE and any other TWO Questions

QUESTION ONE (30 MARKS)

- a) Differentiate between knowledge discovery and data mining. (2 marks)
- b) Explain the two “high-level” primary goals of data mining, in practice. (2 marks)
- c) Explain how data mining can be used in Telecommunication Industry and Biological Data Analysis. (4 marks)
- d) Explain two categories of functions involved in data mining. (3 marks)
- e) Differentiate between star schema and snowflake schema. (3 marks)
- f) Explain Association algorithm in Data mining. (3 marks)
- g) Describe three data smoothing techniques. (3 marks)
- h) Present an example where data mining is crucial to the success of a business. What data mining functions does this business need? Can they be performed alternatively by data query processing or simple statistical analysis? (4 marks)
- i) Based on your observation, describe another possible kind of knowledge that needs to be discovered by data mining methods but has not been listed in this chapter. Does it require a mining methodology that is quite different from those outlined in this chapter? (2 marks)

- j) Describe the differences between the following approaches for the integration of a data mining system with a database or data warehouse system: *no coupling*, *loose coupling*, *semi tight coupling*, and *tight coupling*. State which approach you think is the most popular, and why. (4 marks)

QUESTION TWO (20 MARKS)

- a) Describe the steps involved in data mining when viewed as a process of knowledge discovery. (8 marks)
- b) Describe how the goals of prediction and description can be achieved by using the primary data mining tasks: (6 marks)
- c) In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem. (6 marks)

QUESTION THREE (20 MARKS)

- a) Describes the three major issues data mining. (6 marks)
- b) Using a diagram discuss how Online Analytical Mining can be integrated with Online Analytical Processing with data mining and mining knowledge in multidimensional databases. (6 marks)
- c) Outline the major research challenges of data mining in one specific application domain, such as stream/sensor data analysis, spatiotemporal data analysis, or bioinformatics. (4 marks)
- d) Suppose your task as a software engineer at Machakos-University is to design a data mining system to examine the university course database, which contains the following information: the name, address, and status (e.g., undergraduate or graduate) of each student, the courses taken, and their cumulative grade point average (GPA). Describe the architecture you would choose. What is the purpose of each component of this architecture? (4 marks)

QUESTION FOUR (20 MARKS)

- a) Describe how data mining engine is very essential to the data mining system with its set of functional modules. (5 marks)
- b) Discuss five criteria on how a data mining system can be classified. (5 marks)

- c) Explain five features to look into while selecting a data mining system. (5 marks)
- d) Explain five trends in Data Mining and knowledge discovery. (5 marks)

QUESTION FIVE (20 MARKS)

- a) Describe Decision Tree Induction Algorithm as model for Data Mining. (4 marks)
- b) Discuss Bayesian Classification as model for Data Mining. (4 marks)
- c)
 - i Explain three applications of cluster Analysis. (3 marks)
 - ii Explain three Requirements of Clustering in Data Mining. (3 marks)
 - iii Explain three categories of Clustering methods. (3 marks)
- c) What is the difference between discrimination and classification? Between characterization and clustering? For each of these pairs of tasks, how are they similar? (3 marks)