



DEDAN KIMATHI UNIVERSITY OF TECHNOLOGY

University Examinations 2015/2016

**THIRD YEAR SUPPLEMENTARY/SPECIAL EXAMINATION FOR THE
DEGREE OF BACHELOR OF SCIENCE IN BUSINESS INFORMATION
TECHNOLOGY**

CBT 2301: DATA MINING

DATE: 21ST MARCH 2016

TIME: 2.00PM – 4.00PM

INSTRUCTIONS: Answer Question 1 and Any Other TWO Questions

Question 1

30 Marks

- a) Discuss the following data mining terms: (8 Marks)
- i. Machine Learning
 - ii. Web Data Mining
 - iii. Big Data
 - iv. Business Intelligence
- b) Compare and contrast neural networks from decision trees (3 Marks)
- c) Briefly the steps involved in the knowledge discovery process (7 Marks)
- d) Explain the mining techniques you can use to detect network intrusion by hackers and other malicious software. (6 Marks)
- e) Explain the following data mining tasks (6 Marks)
- i. Associations
 - ii. Outlier Analysis
 - iii. Cluster Analysis

Question 2

20 Marks

- a) Discuss the main reasons why data mining is an essential task in a organizations.(8 Marks)
- b) Discuss FIVE issues associated with data mining (5 Marks)

c) Outline the tools you can use to visualize patterns discovered during data mining. (5 Marks)

d) Differentiate between web structure mining and web usage mining. (2 Marks)

Question 3

20 Marks

- a) State FOUR sources of data for mining (4 Marks)
- b) Discuss the process of data pre-processing (8 Marks)
- c) Discuss Four legal and ethical issues associated with data mining (8 Marks)

Question 4

20 Marks

- a) Discuss FOUR kind of problems you can solve with data mining (8 Mark)
- b) Suppose the fraction of diploma students who smoke is 15% and the fraction of undergraduate students who smoke is 23%. If one-fifth of the college students are diploma students and the rest are undergraduates:
 - i) Using the Bayesian Theorem, determine the probability that a student who smokes is a graduate student? (6 Marks)
 - ii) From the information in part (i) above, is a randomly chosen college student more likely to be a diploma or undergraduate student? Give a reason(2 Marks)
- c) Discuss how the following measures assess the quality of text retrieved from documents. (6 Marks)
 - i) Precision
 - ii) Recall
 - iii) F-score

Question 5

20 Marks

- a) Describe how each of the following data mining techniques can be applied by a internet access provider. (10 Marks)
 - i) clustering
 - ii) classification
 - iii) association rule mining
 - iv) anomaly detection
- b) Describe the steps involved in discovering knowledge from databases. (10 Marks)