FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

DEPARTMENT OF INFORMATION & MEDIA TECHNOLOGY

FIRST SEMESTER 2018/2019 EXAMINATION

COURSE CODE:	IMT 311
COURSE TITLE:	DATA MINING I
CREDIT UNITS:	2
TIME ALLOWED:	2 Hours
COURSE LECTURER(S):	I. O. ALABI & Y. M. SANI
NUMBER OF QUESTIONS:	5 ESSAY QUESTIONS
NUMBER OF PAGES:	3 (INCLUDING THIS PAGE)

INSTRUCTIONS

- Answer One (1) and any other Two (2) questions
- Do **not** use red pen
- Please use a clear handwriting
- This exam is closed book, closed notes, closed laptop and closed cell phone

Question 1: Compulsory (30 marks)

(a) Define data mining? **(3 marks)**

(b) Describe the steps involved in data mining when viewed as a process of knowledge Discovery? **(7 marks)**

(c) How is a data warehouse different from a database? How are they similar? **(5 marks)**

(d) Data warehouses often adopt a three-tier architecture? Explain this concept with the aid of a diagram. **(5 marks)**

(e) What are the major challenges of mining a huge amount of data (e.g., billions of tuples) in comparison with mining a small amount of data (e.g., data set of a few hundred tuple)? **(5 marks)**

(f) Data have qualities if they satisfy the requirements of the intended use. Mention and explain five (5) **(5 marks)**

Question 2 (15 marks)

(a) Explain the following concepts:

i. Machine learning
ii. Supervised learning
iii. Unsupervised learning **(2 marks each)**
iv. Semi-supervised learning
v. Active learning

(b) What kind of data can be mined? Mention and explain five (5)? **(1 mark each)**

Question 3 (15 marks)

(a) Giving the following data (in increasing order) for the attribute age: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Use smoothing by bin means to smooth these data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data? **(12 marks)**

(b) What are the other methods for data smoothing? **(3 marks)**

Question 4 (15 marks)

(a) Imagine that you need to analyze *All_Electronics* sales and customer data. You note that many tuples have no recorded value for several attributes such as customer *income*. How can you go about filling in the missing values for this attribute? Describe various methods for handling this problem. **(5 marks)**

(b) What are the major steps involved in data preprocessing? Enumerate them. (5 marks)

(c) Discuss the following *data mining functionalities*:

i. Characterization
ii. Discrimination **(1 mark each)**
iii. Association
iv. Clustering
v. Outlier analysis

Question 5 (15 marks)

(a) What is Business Intelligence? (3 marks)

(b) With the aid of a diagram, describe the main components of a Business Intelligence System? (7 marks)

(c) Discuss all the ethical issues associated with Business intelligence? (5 marks)