FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

DEPARTMENT OF INFORMATION AND MEDIA TECHNOLOGY

FIRST SEMESTER 2016/2017 EXAMINATION

COURSE CODE:	IMT 412
COURSE TITLE:	DATA MINING II
CREDIT UNITS:	2
TIME ALLOWED:	2 hours
COURSE LECTURER(S):	I. O. ALABI
NUMBER OF QUESTIONS:	4
NUMBER OF PAGES:	3 (INCLUDING THIS PAGE)

INSTRUCTIONS

- Answer all questions
- Do **not** use red pen
- Please use a clear handwriting
- This exam is closed book, closed notes, closed laptop and closed cell phone
- Please use non-programmable calculators only

1. a) Explain the following:
   i) Supervised learning ii) Data mining techniques iii) Data mart
   iv) Data clustering v) Market-basket transactions
   (*2 marks each*).

   b) What are the major data mining processes? (5 marks)

2. Consider the Point Of Sales (POS) transactions below.

   a. Convert the POS transactions into a binary form. What is the width of Transaction 6? *(2 marks)*

   b. State the Support count and Confidence metrics. What is their significance? *(2 marks)*

   d. Compute the Support count and the rule's count of:

   (i) {Butter, Bread},

   (ii) {Milk, Diapers}, and

   (iii){Beer, Milk, Diapers} itemsets. *(6 marks)*

   e. Compute the Confidence rule for:

   (i) {Milk, Diapers}→{Beer, Milk, Diapers};

   (ii) {Butter, Milk}→ {Milk, Diapers, Bread, Bread}. *(5 marks)*

Table 1: POS Transactions

Transaction ID	Items Bought
1	{Milk, Beer, Diapers}
2	{Bread, Butter, Milk}
3	{Milk, Diapers, Cookies}
4	{Bread, Butter, Cookies}
5	{Beer, Cookies, Diapers}
6	{Milk, Diapers, Bread, Butter}
7	{Bread, Butter, Diapers}
8	{Beer, Diapers}
9	{Milk, Diapers, Bread, Butter}
10	{Beer, Cookies}

3. a) Explain the following Decision tree test conditions, use sketches as necessary:

i) Binary attributes ii) Nominal attributes iii) Ordinal attributes
iv) Continuous attributes v) Target attributes *(2 marks each)*.

b) Sketch a lattice for all possible item-sets I = { a,b,c,d,e}. (2 *marks*).

c) State the *Apriori* principle, then use it to generate the frequent itemset for the POS transactions in Table 1. *(3 marks)*

4. a) What is a Confusion matrix? Use the matrix of table 2 to explain its content and hence the following:
i) True-Positive
ii) False Positive
iii) False Negative, and
iv) True Negative. *(10 marks)*

b) What are the metrics that can be computed from the matrix? *(5 marks)*

Table 2: Matrix A

		Predicted	
		Class 1	Class 2
Actual	Class 1	TP 67	FN 15
	Class 2	FP 20	TN 13