**DEDAN KIMATHI UNIVERSITY OF TECHNOLOGY**

University Examinations 2017/2018

**FIRST** YEAR **SEMESTER** II EXAMINATIONFOR THE DEGREE OF **BACHELOR OF SCIENCEIN BUSINESS INFORMATION TECHNOLOGY**

**CBT 2301: DATA MINING**

**DATE: DECEMBER 2017** **TIME: 2 HOURS**

***Instructions: Answer Question 1 and Any Other Two.***

## Question One (30 Marks)

Use the following table, a hypothetical customer relationship management (CRM) database for a company that makes computer parts

| Name | Publicly Held? | Market Cap. ($Mil) | Employees | Sales Channel | Customer since | Year | Units Sold | Profit |
|---|---|---|---|---|---|---|---|---|
| Dell | Y | 75000 | 37200 | Direct | 1997 | 2002 | 120000 | $1.2M |
| Dell | Y | 88000 | 39100 | Direct | 1997 | 2003 | 109000 | 1.1M |
| Gateway | Y | 2300 | 11000 | Direct | 1995 | 2001 | 60000 | .9M |
| Gateway | Y | 1400 | 11500 | Direct | 1995 | 2002 | 70000 | 1.0M |
| Gateway | Y | 1700 | 9600 | Retailers | 1995 | 2003 | 65000 | 1.9M |
| Compaq | Y | 10000 | 50000 | Retailers | 1993 | 2002 | 30000 | 1.4M |
| Hewlett-Packard | Y | 35000 | 95000 | Retailers | 1994 | 2002 | 80000 | 2.2M |
| Hewlett-Packard | Y | 60000 | 141000 | Retailers | 1994 | 2003 | 100000 | 2.5M |
| MA Micro | N | ? | 80 | Direct | 1995 | 2003 | 400 | 3500 |

The database above contains information on the companies we supply our products to, both general information about them (such as their market capitalization - the total value of their outstanding stock, the way they sell their products); and information about our sales to them and the profit we get from those sales. Our company's marketing department wants to use this information to better target their campaigns, thus increasing the sales of our company's products and the profit earned. The table you are shown is not complete - you can assume there are a lot more attributes both describing the company, and describing our sales to the company . The information shown above will be sufficient for you to answer the questions below

a) Data, Information and Knowledge all form important ingredients of decision. By differentiating between information and knowledge describe the process of achieving knowledge [4 Marks]

b) By defining the term Classification describe how this data mining method will be applied in direct marketing [4 Marks]

c) The people in marketing would like a better understanding of their different customers. They want to know what distinguishes customers – what are the key attributes that make a customer unique. Discuss two data mining technique to use to answer such question. [6 Marks]

d) The company has said "we have the data in a data warehouse, all you need to do is install a data mining tool and run it. Discuss what need to be done in order to successfully accomplish this task [4 Marks]

e) In 2002, Compaq and Hewlett-Packard merged (i.e. Hewlett-Packard bought Compaq). This distorts the data. Discuss TWO ways oh how to handle this incident [4 Marks]

f) Apart from Marketing discuss THREE other applications of Data Mining [6 Marks]

g) Distinguish between data warehouse and operational database [2 marks]

**Question Two (20 Marks)**

A marketing manager of Safaricom realizes that a large number of customers are discontinuing their service, leaving her company in favor of some competing provider. As can be imagined, low customer loyalty, also known as customer *attrition or churn, is a critical factor for many* companies operating in service industries. Suppose that the marketing manager can rely on a budget adequate to pursue a customer retention campaign aimed at 2000 individuals out of a total customer base of 16 million people. Hence, the question naturally arises of how she should go about choosing those customers to be contacted so as to optimize the effectiveness of the campaign. In other words, how can the probability that each single customer will discontinue the service be estimated so as to target the best group of customers and thus reduce churning and maximize customer retention? By knowing these probabilities, the target group can be chosen as the 2000 people having the highest churn likelihood among the customers of high business value

a) Discuss the term business Intelligence and its purpose in the Safaricom case above [4 Marks]

b) Construct a business intelligence system for the above case clearly showing the main components of the systems and the role of each of those components [16 Marks]

**Question Three (20 Marks)**

a) With the aid of a well labeled diagram, explain the data warehouse three-tier architecture [10 marks]

b) Use the following table as your training set to answer below question

| age | income | student | credit_rating | buys_computer |
|---|---|---|---|---|
| <=30 | high | no | fair | no |
| <=30 | high | no | excellent | no |
| 31…40 | high | no | fair | yes |
| >40 | medium | no | fair | yes |
| >40 | low | yes | fair | yes |
| >40 | low | yes | excellent | no |
| 31…40 | low | yes | excellent | yes |
| <=30 | medium | no | fair | no |
| <=30 | low | yes | fair | yes |
| >40 | medium | yes | fair | yes |
| <=30 | medium | yes | excellent | yes |
| 31…40 | medium | no | excellent | yes |
| 31…40 | high | yes | fair | yes |
| >40 | medium | no | excellent | no |

Construct a decision tree to find out who "buys_computer" [10 Marks]

**Question Four (20 Marks)**

Nyeri cooperative is owned by more than 30,000 farmer-members, it purchases, manufactures, and processes feed, seed, fertilizer and farm supplies. Thanks to strong customer loyalty and very high brand recognition among agricultural professionals, Nyeri Cooperative serves more than 200 retail locations in 5 counties and sells products to farmers and rural Kenya customers.
In order to maintain and extend its success, Nyeri Cooperative recognized that it needed to continually optimize its marketing efforts, especially its very high value direct marketing activities. However, the company faced some significant challenges to this goal. Not only did Nyeri Cooperative lack a way to bring together customer and marketing data from multiple sources across the company for analysis, but also it had no consistent or sophisticated tools to drive marketing analytics, instead relying on generic productivity tools, such as Microsoft Excel.

a) Discuss the main business problem faced by Nyeri Cooperative [4 Marks]
b) Discuss how predictive analytics can be applied in the Nyeri Cooperative case [8 Marks]
c) Explain the problems that can be solved by the optimization techniques if employed by Nyeri Cooperative [8 Marks]