CHUKA  UNIVERSITY

UNIVERSITY EXAMINATIONS

FOURTH YEAR SECOND SEMESTER EXAMINATION FOR THE AWARD OF BACHELOR OF SCIENCE APPLIED COMPUTER SCIENCE

ACMP 452: DATA MINING AND KNOWLEDGE DISCOVERY IN DATABASES

STREAMS: BSC. APPLIED COMP.SCI **TIME: 2 HOURS**

DAY/DATE: THURSDAY 11/4/2019 **11.30 A.M. – 1.30 P.M.**

INSTRUCTIONS:

- Answer Question **ONE** and any other **TWO** questions.
- Diagrams should be used whenever they are relevant to support an answer.
- Sketch maps and diagrams may be used whenever they help to illustrate your answer
- Do not write anything on the question paper
- This is a **closed book exam**, No reference materials are allowed in the examination room
- There will be **No** use of mobile phones or any other unauthorized materials
- Write your answers legibly and use your time wisely

SECTION A

ANSWER ALL THE QUESTIONS IN THIS SECTION

QUESTION ONE [30 MARKS]

a) Describe the four characteristics of Big Data [4 Marks]

b) Differentiate between classification and regression as paradigms of data mining [4 Marks]

c) Explain the term cluster analysis and highlight two practical applications in modern systems [4 Marks]

d) Highlight any two advantages of decision tree in data mining and any two limitations of decision tree mining. [4 Marks]

e) Identity and explain the two classes of knowledge. [4 Marks]

f) Discuss any three benefits of data visualization tools in an organization. [6 Marks]

g) Explain the relationship between Data Mining and KDD [4 Marks]

SECTION B

ANSWER <u>ANY TWO</u> QUESTIONS FROM THIS SECTION

QUESTION TWO [20 MARKS]

a) Data pre-processing and conditioning is one of the key factors that determine whether a data mining project will be a success or not. For each of the following topics, describe the affect this issue can have on our data mining session and what techniques can we use to counter this problem.

i. Noisy data [2 Marks]
ii. Missing data [2 Marks]
iii. Data normalization and scaling [2 Marks]
iv. Data type conversion [2 Marks]
v. Attribute and instance selection [2 Marks]

b) Describe the three main components of calculating Bayesian probabilities [6 Marks]

c) Discuss two reasons why data mining is popular now than it was 20 years ago. [4 Marks]

QUESTION THREE [20 MARKS]

a) Contrast OLTP and OLAP based on the following features:

i. User and System Orientation [2 Marks]
ii. Data Contents [2 Marks]
iii. Database Design [2 Marks]
iv. View [2 Marks]
v. Access Patterns [2 Marks]

b) List three real application areas of Data Mining and highlight which data mining techniques can be applied for each application area [6 Marks]

c) With the aid of a suitable example present the structure of a Bayesian network then use it to explain the concept of Bayesian networks in data mining [4 Marks]

QUESTION FOUR [20 MARKS]

a) A dataset is given as follows:

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

a) Using Attributes Outlook, Temperature, Humidity and Wind and Classification PlayTennis, apply ID3 algorithm to develop a decision tree [14 Marks]

b) Create rules from the decision tree created in a) above [6 Marks]

QUESTION FIVE [20 MARKS]

Consider the market basket transactions shown in the following table. Assume that min_support=2% and min_confidence=70%. Assume also that the Apriori algorithm is used to discover strong association rules among transaction item.

Transaction ID	Items Bought
1	{Bread, Butter, Milk}
2	{Bread, Butter}
3	{Beer, Cookies, Diapers}
4	{Milk, Diapers, Bread, Butter}
5	{Beer, Diapers}

a) By following the Apriori Algorithm, generate strong association rules for the dataset [16 Marks]

b) Identify two application areas for Association Rule mining in a supermarket [4 Marks]
