

PAN AFRICA CHRISTIAN UNIVERSITY

BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY

END OF SEMESTER EXAMINATION

DEPARTMENT: COMPUTING AND INFORMATION TECHNOLOGY

COURSE CODE: BSIT 307

COURSE TITLE: DATAMINING

CAMPUS: ROYSAMBU

EXAM DATE: THURSDAY, XXXX, DECEMBER, 2024

TIME:

INSTRUCTIONS

- This exam script has **TWO (2)** sections.
- Read all questions carefully before attempting.
- Answer All questions in Section **A** and any other Three questions in Section **B**.
- Write only your **student number** on the answer booklet provided.

SECTION A

(Answer ALL questions in this section)

Question 1:

- a) Explain the relationship between Data Mining and KDD **[3 Marks]**
- a) Pac University wishes to implement a data warehouse; however, they plan to roll it out from one school at a time. If it works well the management will be onboarded. Recommend the design approach suitable for their needs. Justify your choice **[4 Marks]**
- b) Using real life examples discuss any three known challenges in Data Mining **[3 Marks]**

SECTION B

(Answer any THREE (3) questions in this section)

Question 2:

- a) Explain two application areas for Association Rule mining in a supermarket **[4 Marks]**
- b) Highlight any two advantages of decision tree in data mining and any two limitations of decision tree mining **[4 Marks]**
- c) Explain web context mining and how it can be utilized in real life **[2 Marks]**

Question 3:

- a) Explain the following in the context of data mining. In each case, support your answer with an example, task or strategy
 - i. Data Preprocessing **[3 Marks]**
 - ii. Data Reduction **[3 Marks]**
- b) Differentiate between classification and regression as paradigms of data mining

[4 Marks]

Question 4:

During a data warehouse design, we perform data modelling. At this stage the data warehouse schema is defined. Explain schema and hence discuss three major types of schemas. Support your answer with simple diagrams **[10 Marks]**

Question 5:

“A Data Warehouse is a subject oriented, integrated, nonvolatile, and time variant collection of data in support of management’s decisions.” The four keywords, subject-oriented, integrated, time-variant, and nonvolatile, are defining features that distinguish data warehouses from other data repository systems, such as relational database systems, transaction processing systems, and file systems.

- a) Briefly explain these defining features with regard to data, nature of data, and usage **[9 Marks]**
- b) Describe multidimensional data as applied to data warehousing **[1 Mark]**

Question 6:

The applications of data mining in telecommunication industries can be grouped into three areas namely: Fraud detection, marketing/customer profiling and network fault isolation. Define data mining and briefly discuss these areas using known examples

[10 Marks]

EOF