



**PAMIBIA UNIVERSITY  
OF SCIENCE AND TECHNOLOGY**

**FACULTY OF COMPUTING AND INFORMATICS**

**DEPARTMENT OF INFORMATICS**

<b>QUALIFICATIONS:</b> BACHELOR OF INFORMATICS HONOURS (BUSINESS, WEB)	
<b>QUALIFICATION CODE:</b> 08BIFB; 08BIHW	<b>LEVEL:</b> 8
<b>COURSE CODE:</b> DSA821S	<b>COURSE:</b> DATA SCIENCE AND ANALYTICS
<b>DATE:</b> JANUARY 2024	<b>SESSION:</b> 1
<b>DURATION:</b> 2 HOURS	<b>MARKS:</b> 70

<b>SUPPLEMENTARY/SECOND OPPORTUNITY EXAMINATION QUESTION PAPER</b>	
<b>EXAMINERS:</b>	Mr Eliazer Mbaeva
<b>MODERATOR(S):</b>	Mrs Emilia Shikeenga

**THIS EXAMINATION PAPER CONSISTS OF 4 PAGES  
(INCLUDING THIS FRONT PAGE)**

**INSTRUCTIONS FOR THE CANDIDATE**

1. Answer all questions.
2. When writing, consider the following: The style should be to inform rather than impress.
3. Information should be brief and accurate.
4. Please ensure that your writing is legible, neat and presentable.

## SECTION A

---

### Question 1

Choose the correct answer from each of the options given.

(1 x 5 = 5 marks)

- 1.1. Data has been collected on visitors' viewing habits at a bank's website. Which technique is used to identify pages commonly viewed during the same visit to the website?
  - a. Clustering
  - b. Classification
  - c. Association Rules
  - d. Regression
  
- 1.2. Which of the following step is performed by data scientist after acquiring the data?
  - a. Data Integration
  - b. Data Replication
  - c. Data Cleansing
  - d. All the above
  
- 1.3. Which of the following is one of the key data science skills?
  - a. Machine Learning
  - b. Statistics
  - c. Data Visualization
  - d. All the above
  
- 1.4. Which of the following is the process of evaluating data through analytical and statistical tools?
  - a. Data Mining
  - b. Data Exploration
  - c. Data Analysis
  - d. Data Visualization
  
- 1.5. Which type of data is generated by POS terminal in a busy supermarket each day?
  - a. Source
  - b. Processed
  - c. Synchronized
  - d. All the above

## Question 2

State whether the following statements are **True** or **False**.

(1 x 5 = 5 marks)

- 2.1 A model only works with dependent variables.
- 2.2 Dependent variable represents an outcome.
- 2.3 Specifying a regression model involves selecting a dependent variable\_only.
- 2.4 The type of dependent variable determines the type of regression, linear or logistic.
- 2.5 Modelling step creates a representative model inferred from the data.

## SECTION B

---

### Question 1

1.1. Define the following concepts:

(10 marks)

- a. Data Science
- b. Data Mining
- c. Clustering
- d. Classification
- e. Machine Learning

1.2. State and describe three (3) main components of Hadoop.

(6 marks)

1.3. State the difference between supervised and un-supervised learning.

(4 marks)

1.4. State and define the 5 Vs of Big Data.

(10 marks)

**SECTION C**

**Question 1**

1.1 Given the data table below, find out the linear regression equation. For better understanding of your work, you may need to create subsequent table for your calculations. (10 marks)

<b>X</b>	2	3	5	8
<b>Y</b>	3	6	5	12

Figure 1: Data Table

1.2 Draw a scatterplot and regression line for your work. (5 marks)

1.3 A statistical model is described by the following equation and is used to predict numerical values:

$$Y = \sum(\text{weight} * \text{input}) + \text{bias}$$

Name the above model and briefly describe how it works. (5 marks)

1.4 Suppose that you are on campus, leaving the library after having been preparing for the next day's exam for data science and you were approached by an inspiring second year student X who has a great deal of interest in everything data science. In your conversation with X, arises the concept of Deep Learning and X is keen to know what that is. Explain Deep Learning to X, indicate how it works. (5 marks)

1.5 Consider the diagram below, explain the concept it represents, and all possible stages involved in the process. (5 marks)

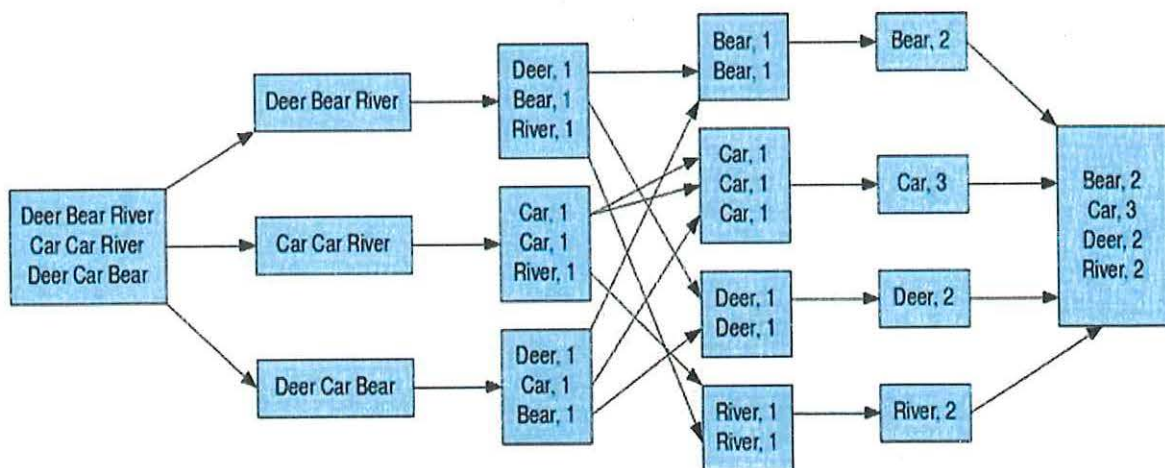


Figure 2: Words Diagram