NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATICS

DEPARTMENT OF INFORMATICS

QUALIFICATION : BACHELOR OF INFORMATICS, BACHELOR OF COMPUTER SCIENCE	
QUALIFICATION CODE: 07BAIT, 07BCMS	**LEVEL:** 6
COURSE: DATA ANALYTICS	**COURSE CODE:** DTA621S
DATE: JANUARY 2023	**SESSION:** 1
DURATION: 3 HOURS	**MARKS:** 100

SECOND OPPORTUNITY/SUPPLEMENTARY EXAMINATION QUESTION PAPER	
EXAMINER(S)	MRS RUUSA IPINGE
MODERATOR:	Dr. JACOB ONGALA

THIS QUESTION PAPER CONSISTS OF 8 PAGES

(Including this front page)

INSTRUCTIONS

- Answer ALL questions in Question1, Question2, Question3, Question 4 and Question 5
- NUST examinations rules apply
- DO NOT open this examination cover until you are instructed to do so.
- DO NOT FORGET to write down your student number at the designated places in the examination page.

Question1: MULTIPLE QUESTIONS (20 MARKS MAXIMUM 1 MARK FOR EACH CORRECT ANSWER)

Answer all questions. Select ONLY ONE BEST ASWER to each questions.

1. The 'right to be forgotten' gives an individual the right to have their personal data, what?

   a) Read by everyone
   b) Amended
   c) Erased
   d) Restricted

2. Point out the correct statement
   a) Raw data is the data in excel sheet
   b) Raw data is the data obtained after pre-processing
   c) Raw data is the data before pre-processing
   d) None of the above

3. Are used when we want to visually examine the relationship between two quantitative variables.
   a. Bar graph
   b. Scatterplot
   c. Line graph
   d. Pie chart

4. A graph that uses vertical bars to represent data is called a _____.
   a) Bar graph
   b) Line graph
   c) Scatterplot

d) All of the mentioned above

5. **Data Analysis is a process of,**
   a) Inspecting data
   b) Data Cleaning
   c) Transforming of data
   d) All of the mentioned above

6. **What are the only grounds for individuals to have an absolute right to objection?**
   a) Legitimate interest
   b) Public Interest
   c) Direct Marketing
   d) None of the above

7. **The students divided into different groups according to their intelligence and gender will generate".**
   a) Quantitative
   b) Qualitative
   c) Continuous data
   d) Constant

8. **The mode of 2,9 and 7 is:**
   a) No mode exists.
   b) 2
   c) 9
   d) 6

9. **The data must be arranged before computing the**
   a) Mean
   b) Mode
   c) Median

d) All of the above

10. To predict a binary value. use ___

    a)   Logistic

    b)   Clustering

    c)   Classification

    d)   Dimensionality reduction

11. How do machine learning algorithms make more precise predictions?

    a) The algorithms are typically run more powerful servers.

    b) The algorithms are better at seeing patterns in the data.

    c) Machine learning servers can host larger databases.

    d) The algorithms can run on unstructured data

12. Machine learning is a subset of?

    a) Artificial Intelligence

    b) Deep Learning

    c) Artificial intelliget

    d) None of the above

13. In linear regression, we try to   the least square errors of the model to identify the line of the best fit.

    a) Minimise

    b) Maximise

    c) Change

    d) None of the above

14. What is the data type of print (type)(("Loide")?

    a) Float,

    b) String,

    c) Integer

    d) Boolean

15. Under the GDPR, how many main rights do individuals have?

    a) 6

    b) 8

    c) 9

    d) 7

16. Within how long must a data breach be reported from when it is first identified?

    a) 24

    b) 48

    c) 12

    d) 72

17. This is the process of programmatically identifying and grouping objects or ideas into pre-determined categories

    a) Classification

    b) Binary

    c) Sorting

    d) Variance

18. You want to identify global weather patterns that may have been affected by climate change. To do so, you want to use machine learning algorithms to find

patterns that would otherwise be imperceptible to a human meteorologist. What is the place to start?

  a) Find labeled data of sunny days so that the machine will learn to identify bad weather.
  b) Use unsupervised learning have the machine look for anomalies in a massive weather database.
  c) Create a training set of unusual patterns and ask the machine learning algorithms to classify them.
  d) Create a training set of normal weather and have the machine look for similar patterns.

19. **Why is naive Bayes called naive?**

  a) It naively assumes that you will have no data.
  b) It does not even try to create accurate predictions.
  c) It naively assumes that the predictors are independent from one another.
 d) It naively assumes that all the predictors depend on one another.

20. **Which statistical model and supervised machine learning algorithm uses independent variables to predict the values of a dependent variable?**

  a) Lasso regression.
  b) Multiple Regression.
  c) Logistic Regression
  d) Linear regression.

Question 2   **[8 Marks]**

Look at the following boxplots about Emma's performance in physics and chemistry and answer the following questions.
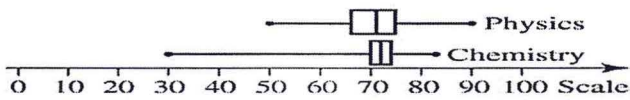
    (a)   State the median mark for each subject.   (2)

    (b)   Find the range of marks in each subject.   (2)

    (c)   Calculate the interquartile for Chemistry?   (3)

    (d)   Give one of the benefits of using a scatter plot dring ?   (1)

Question 3   **[32 Marks]**

    a) List and describe 4 different types of data distribution   (8)
    b) List and explain the Methods of Data Integration.   (8)
    c) Explain three methods of how you can clean noise data   (6)
    d) Using Examples, explain the process of Data Science   (10)

Question 4   **[20 Marks]**

    a) Explains under fitting, and how can you avoid It?   (6)
    b) Explain the difference between the data controller and data processor in GDPR?   (4)
    c) Describe the 4 methods of how you could evaluate the performance of machine learning (4)

d) Briefly Explain Neural Network. (2)

e) Explain the Support vector machine learning Algorithm (SVM). (2)

Question 5 [20 Marks]

4.1 Based on the Jupiter notebook, explain what the following command means

a) df.dtypes (2)

b) df.head (2)

c) df.drop duplicates() (2)

d) df.nsmallest(n, `values`) (2)

e) df.fillna(`values`) (2)

4.2 Write a command in Jupiter notebook that will allow you to perform the following tasks

a) Give the statistics of each columns (2)

b) df.iat (1, 2) (2)

c) Give the minimum value of each object (2)

d) To group object by columns (2)

e) Calculate the standard deviation of each object (2)

THE END OF EXAM