NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATICS

DEPARTMENT OF INFORMATICS

QUALIFICATIONS: Bachelor of Computer Science; Bachelor of Informatics	
QUALIFICATION CODE: 07BCMS; 07BAIT	**LEVEL:** 6
COURSE CODE: DTA621S	**COURSE:** Data Analytics
DATE: November 2023	**SESSION:** 1
DURATION: 3 Hours	**MARKS:** 70

FIRST OPPORTUNITY EXAMINATION QUESTION PAPER	
EXAMINERS:	Mrs Ruusa Ipinge
MODERATOR(S):	Dr Jacob Ongala

THIS EXAMINATION PAPER CONSISTS OF 10 PAGES
(INCLUDING THIS FRONT PAGE)

INSTRUCTIONS FOR THE EXAMINER/MODERATOR

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.

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

Answer all questions. Select ONLY ONE BEST ANSWER to each question.

1. **___ is a type of usupervised machine learning methods where lost data point are assigned to the nearest group.**

a) Classification
b) Clustering
c) Data mining
d) None of the mentioned above

2. **An advantage of using computer programs for qualitative data is that they ___.**

A. Can reduce time required to analyse data.
B. Help in storing and organizing data.
C. Make many procedures available that are rarely done by hand due to time constraints.
D. All the mentioned above

3. **Logistic regression is used to find the probability of event = Success and event = ____.**

a) Failure
b) Success
c) Both A and B
d) None of the mentioned above

4. **This is the process of reorganising data and cleaning data by removing redundant and unstructured data and making the data look similar across all records**

a) Smoothing
b) Data aggregation
c) Discretization
d) Normalisation

5. **This is the type of research that It answers key questions such as "how many, "what" and "why".**

a) Quantitative
b) Qualitative
c) Nominal
d) Category

6. **___ 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

7. **This is the type of research that It answers key questions such as "how many, "how much" and "how often".**

a) Quantitative
b) Qualitative
c) Nominal
d) Category

8. **This is not an example of continuous data:**

a) The amount of time required to complete a project.
b) The weight of children.
c) The square footage of a two-bedroom house.
d) The number of injections or vaccine you received in your lie.

9. **Which statements is true about ordinal data?**

a) You cannot do arithmetic with ordinal numbers because they only show sequence.
b) Ordinal variables are considered as "in between" qualitative and quantitative data
c) The ordinal data is qualitative data for which the values are ordered.
d) All the mentioned above

10. **What is a hypothesis?**

A. A statement that the researcher wants to test through the data collected in a study.
B. Research questions the results will answer.
C. A theory that underpins the study
D. A statistical method for calculating the extent to which the results could have happened by chance.

11. **Amongst which of the following is / are the applications of Linear Regression,**

A. Biological
B. Behavioural
C. Social sciences
D. All the mentioned about

12. **refers to the ability to turn your data useful for business.**

A. Value
B. Variety
C. Velocity
D. None of the mentioned above

13. **To glean insights from the data, many analysts and data scientists rely on ___.**

A. Data mining
B. Data visualization
C. Data warehouse
D. All of the mentioned above

14. In Shayla's math class, she asks eight people out of the forty people in the class what grade they earned on the last exam. The data she collected is shown below. What is the sample mean for this sample?

Test scores: 89, 75, 61, 82, 95, 76, 83, 91

a) 81.5

b) 84.2

c) 78.3

d) 90.6

15. Which of the following is true about the sample standard deviation?

a. It is equal to the square root of the variance.

b. It is equal to the square root of the sample mean.

c. It is equal to the variance squared.

d. It is equal to the sample mean squared.

16. 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 labelled 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.

17. Why naive Bayes is 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.

18. What is one reason not to use the same data for both your training set and your testing set?

a) You will almost certainly underfit the model.
b) You will pick the wrong algorithm.
c) You might not have enough data for both.
d) You will almost certainly overfit the model.

19. Out of the data gathered by your digital analytics provider, which of the following categories of data are of a personal nature.

a) IP addresses only

b) Cookies Only

c) IP addresses, cookies, name of the site consulted and time of page consultation.

d) Name of the Service Provider

20. How does the GDPR define "Personal Data"?

a) Your personal bank details and postal address

b) Any information relating to an identified or identifiable natural person.

c) Any information relating to an identified or identifiable natural person.

d) None of the above

PART 2: STRUCTURED QUESTIONS

ANSWER ALL QUESTIONS

Questions 1

1. Explain the difference between the following term [10]
   - **a)** Machine learning and Artificial Intelligence
   - **b)** Normal Distribution and Uniform Distribution
   - **c)** Linear and Multiple Regression
   - **d)** Underfitting and Overfitting
   - **e)** Variance and Standard Deviation

Question 2

a) A class contains 50 children. The following children were chosen at random, and their weight were recorded in cm: 25, 26, 27, 30, and 32. Calculate the variance of their age. Show your work [5]

b) What is r2, what does it measure? [2]

Question 3

1. Explain the output of the following python codes [10]

a)

```
x = 3+5j
y = 5j

print(type(x)
print(type(y)
```

b)

```
a = 33
b = 200
if b > a:
  print("b is greater than a")
```

c)

```
fruits = ("apple", "banana", "cherry")
print(type(fruits))
```

d)

```
set1 = {"a", "b" , "c"}
 set2 = {1, 2, 3}
 set3 = set1.union(set2)
 print(set3)
```

e)

```
def my_function(*kids):
  print("The youngest child is " + kids[2])

my_function("Emil", "Tobias", "Linus")
```

PART 3: APPLICATION OF MACHINE LEARNING

Question 4

a) sing the table below, calculate the centroids X points, given that X is value between point A and B, and it need to be assigned to the correct cluster. Use the threshold set to indicate whether the output of X belong to cluster A or B [6]

Threshold

A= X>5

B=X< 5

A	B	Centroids difference(X)	A or B
7	1		
6	3.5		
8	7		
10	5		
9	2		
-1	-8		
7	8		

b) Assume the scientists predict that 350 test samples contain the genetic variant, and 150 samples don't. If they determine the actual number of samples containing the variant is 305, the actual number of samples without the variant is 195. These values become the "true" values in the matrix and the scientists enter the data in the table:

Predicted without the Variant	Predicted with the Variant
TP=200	FP=150
FN=105	TN=45

c) Using the following confusion matrix. Calculate the following and interpreter with really example of what the results mean. Show your work [9]

i. Recall rate.
ii. Accuracy
iii. Specificality

PART 4: DATA PROTECTION

Question 5

a) Explain the 4 fundamental rights of the General Data Protection Regulation? [8]

END OF QUESTION PAPER