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: JANUARY 2024	**SESSION:** 1
DURATION: 3 HOURS	**MARKS:** 70

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

THIS EXAMINATION PAPER CONSISTS OF 9 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.

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 category, also called supervised machine learning methods in which the data is split on two parts.**

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 of 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. **A graph that uses vertical bars to represent data is called a ____.**

A. Bar graph
B. Line graph
C. Scatterplot
D. All the mentioned above

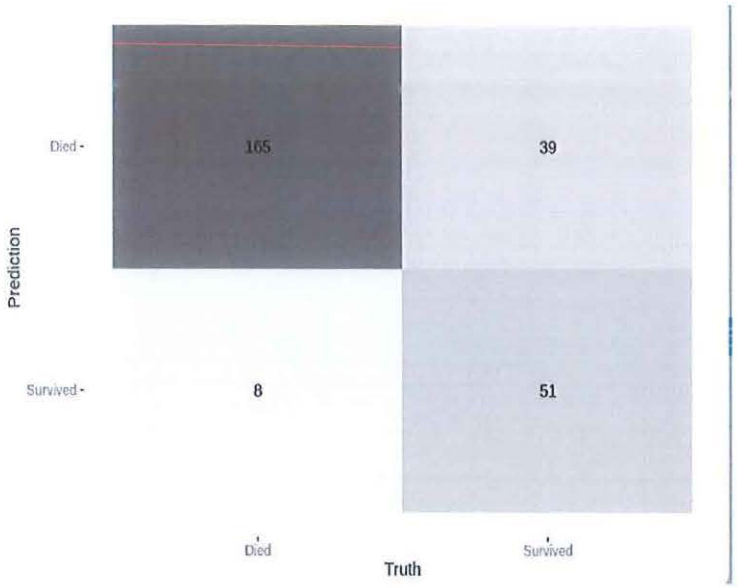
8. **Data Analytics uses ___ to get insights from data.**

a) Statistical figures
b) Numerical aspects
c) Statistical methods
d) None of the mentioned above

9. **Least Square Method uses ___.**

a) Linear polynomial
b) Linear regression
c) Linear sequence
d) None of the mentioned above

10. **Take a look at the confusion matrix above containing 263 observations. What is the accuracy of the predictions?**



A. The accuracy is equal to (165 + 51)/263 (82.1%).
B. The accuracy is equal to (165 + 8)/263 (65.8%).
C. The accuracy is equal to (51)/263 (19.4%).
D. The accuracy is equal to (39 + 8)/263 (17.9%)

11. What is Machine learning?

a) The autonomous acquisition of knowledge using manual programs.
b) The selective acquisition of knowledge using manual programs.
c) The autonomous acquisition of knowledge using computer programs.
d) The selective acquisition of knowledge using computer program.

12. Machine Learning is a field of AI consisting of learning algorithms that_____

a. At executing some task
b. Over time with experience
c. Improve their performance.
d. All mentioned above.

13. **Which of the following is not a supervised learning?**

a. PCA

b. Naive Bayesian

c. Linear Regression

d. Decision Tree

14. **Machine Learning technique that helps in detecting the outliers in data.**

a) Clustering

b) Classification

c) Anomaly Detection

d) All the above

15. **Which answer best describes standard deviation?**

a) Standard deviation is a measure of the spread of a dataset.

b) Standard deviation indicates how much individual values vary from the mean.

c) Standard deviation helps scientists summarize how much variation there is in a dataset or population.

d) All the above

16. **If the mean score for two different datasets is the same, the standard deviation will necessarily be the same.**

a) True

b) False

17. **If an experiment is repeated correctly several times, it should yield**

a) a distribution of measurements around some central value.

b) a single value that is obtained each and every time.

c) widely and randomly varying results.

d) Unpredictable results

18. **In Python, what is the result of the following operation '1'+'2'**

a. '2'

b. '3'

c. 3

d. '12'

19. **In Python, if you executed name = 'Lizz', what would be the output of print(name [0:2])?**

    a. Lizz
    b. L
    c. LI
    d. Liz

20. **What is the output of the following lines of code:**

```
x=1
if(x!=1):
print('Hello')
else:
print('Hi')
print('Mike')
```

    a) Mike
    b) Hello Mike
    c) The Mike
    d) Hi Mike

PART 2: STRUCTURED QUESTIONS

ANSWER ALL QUESTIONS

Questions 1

1. Explain the difference between the following term [10]
   - **a)** Supervised and Unsupervised machine learning.
   - **b)** Training and Test data sets
   - **c)** Logistic and Polynomial Regression
   - **d)** Tuple and List
   - **e)** Variance and Standard Deviation

Question 2

a) A class contains 39 children. The following children were chosen at random, and their weight were recorded in cm: 38, 51, 46, 79, and 57. Calculate their weight' standard deviation. [6]

b) Why Is Standard Deviation Often Used More Than Variance? [2]

Question 3

1. Explain the output of the following codes written in python programming language. [10]

a)
```
a = 2
b = 330
print("A") if a > b else print("B")
```

b)
```
Gemuse = ["apple", "banana", "cherry"]
print(len(Gemuse))
```

c)
```
Gemuse1 = ("apple", "banana", "cherry")
print(type(Gemuse1))
```

d)
```
import pandas as pd
df = pd.read_csv('data.csv')
df.fillna(130, inplace = True)
```

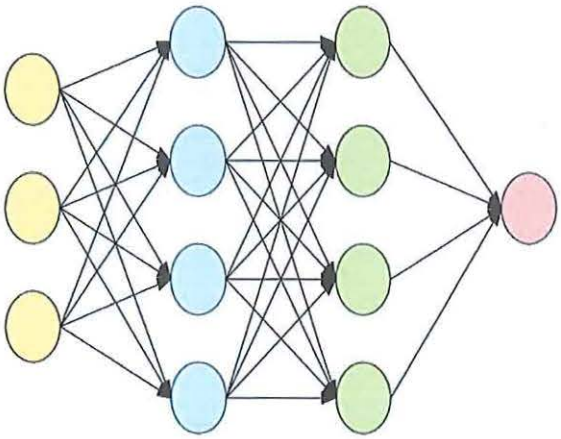
e) i = 1
while i < 6:
print(i)
i += 1

PART 3: APPLICATION OF MACHINE LEARNING

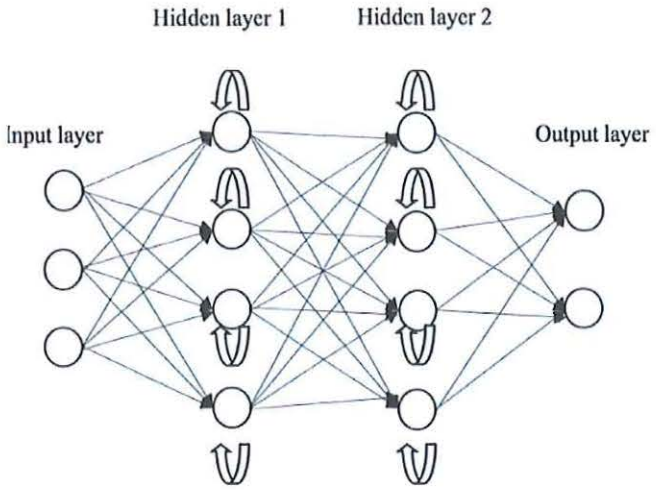
Question 4

a) Identify and explain the types of neural network algorithm presented in the pictures bellow [4]

A



B



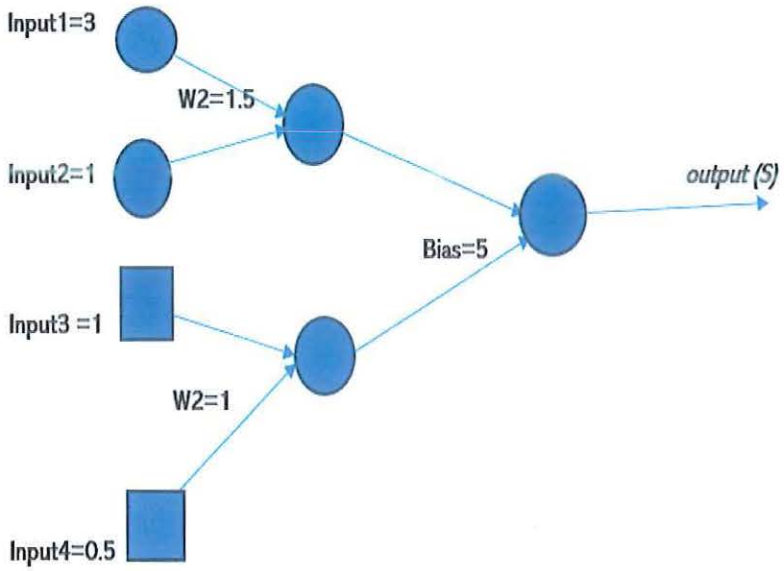
a) Look at the following diagram of Neural Network (NN). Given $input\ 1$ and $input\ 2$ that are independent of $Input\ 2$ and $Input\ 3$. The output is donated by S, and the bias is 5 in both cases .

i) **Calculate the Activation Function given the threshold below and state what will be the output**. Show your work. [7]

Threshold

$$\mathbf{0} = \boldsymbol{S} > \mathbf{10}$$

$$\mathbf{1} = \boldsymbol{S} \leq \mathbf{10}$$



PART 4: DATA PROTECTION

Question 5

Under the GDPR, organisations must meet six data protection principles whenever they process personal data. Explain the principles of the General Data Protection Regulation (GDPR) [10]

END OF QUESTION PAPER