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 2025 | **SESSION:** 2 |
| **DURATION:** 2 HOURS | **MARKS:** 85 |

| **SUPPLEMENTARY/SECOND OPPORTUNITY EXAMINATION QUESTION PAPER** | |
|---|---|
| **EXAMINER(S)** | MRS RUUSA IPINGE |
| **MODERATOR:** | MR SEBASTIAN MUKUMBIRA |

**THIS QUESTION PAPER CONSISTS OF 11 PAGES**

(Excluding this front page)

**INSTRUCTIONS**

- Answer ALL questions in Part 1, Part 2 and Part 3,
- 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.

**PART 1: MULTIPLE CHOICE QUESTIONS (25 MARKS MAXIMUM 1 MARK FOR EACH CORRECT ANSWER)**

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

1. **This helps in ensuring that a model is generalizable to new data rather than just fitting the training data well.**

   a) Classification
   b) Clustering
   c) Data mining
   d) Cross validation

2. **This is the process of selecting a subset of relevant features (variables, predictors) from a larger set to improve model performance, reduce overfitting, and enhance interpretability. It's a crucial step in the machine learning pipeline, especially when dealing with high-dimensional data.**

   a) Feature Selection
   b) Generalisation
   c) Overfitting
   d) Underfitting

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. **It includes functions to visualize distributions, relationships, and categorical data, making it easy to create complex visualizations such as heatmaps, violin plots, and pair plots.**

   a) Matplotlib
   b) Pandas
   c) Seaborn
   d) Normalisation

5. **This is the type of research that it answers key questions such as , "what" and "why".**
   a) Quantitative
   b) Qualitative
   c) Nominal
   d) Category

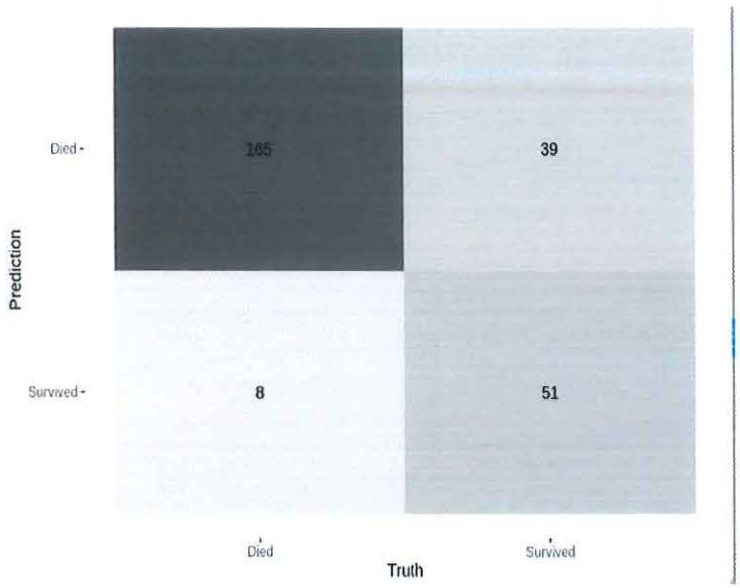
6. **This shows how a specific variable changes across time periods (especially with grouped or stacked plot).**
   a) Bar graph
   b) Scatterplot
   c) Line graph
   d) Pie chart

7. **This refers to the 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) Science tools
   b) Numerical aspects
   **c)** Statistical methods
   d) None of the mentioned above

9. **This approach is commonly employed in regression analysis to find the best-fitting line or model for a given set of data point.**
   a) Least Square Method
   b) Linear regression
   c) Linear sequence
   d) None of the mentioned above

10. **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. What is the primary distinction between Artificial Intelligence (AI) and Machine Learning (ML)?**

a) AI is solely about mimicking human behaviour, while ML is about programming machines to perform specific tasks.
b) AI encompasses a wide range of technologies that simulate human intelligence, whereas ML specifically focuses on algorithms that allow machines to learn from data.
c) AI and ML are interchangeable terms that refer to the same concept of creating intelligent machines.
d) AI requires large amounts of data to function, while ML does not depend on data.

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

a) PCA (Principal Component Analysis)
b) Naive Bayesian
c) Linear Regression
d) Decision Tree

**14. Which of the following Machine Learning technique 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. What is the primary goal of supervised learning?**

a) To find patterns in unlabelled data
b) To predict outcomes based on labelled data
c) To optimize a model without any data
d) To cluster similar items

**17. Which of the following algorithms is commonly used for classification tasks?**

a) Linear Regression
b) K-Means Clustering
c) Decision Trees
d) Principal Component Analysis

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. **How can you read a CSV file into a Pandas DataFrame?**

a) pd.read_table('file.csv')
b) pd.load_csv('file.csv')
c) pd.read_csv('file.csv')
d) pd.import_csv('file.csv')

21. **What method would you use to get a quick overview of a DataFrame's structure and data types?**

a) df.describe()
b) df.info()
c) df.head()
d) df.summary()

22. **How do you select a specific column in a DataFrame named df?**

a) df.column_name

b) df['column_name']

c) df.column_name()

d) df.get('column_name')

23. **What does DPIA stand for in the context of GDPR?**

a) Data Protection and Information Assessment

b) Data Processing Impact Analysis

c) Data Privacy and Incident Assessment

d) Data Protection Impact Assessment

24. **What role does a Data Protection Officer (DPO) play under the GDPR?**

a) Ensuring marketing compliance

b) Overseeing data protection compliance

c) Managing IT infrastructure

d) Handling customer support

25. **How soon should organizations report a data breach to the supervisory authority under the GDPR?**

a) Within 24 hours

b) Within 48 hours

c) Within 72 hours

d) Within one week

## PART 2: STRUCTURED QUESTIONS [60 MARKS]

**ANSWER ALL QUESTIONS**

### QUESTIONS 1

1. Explain the difference between the following terms [8]
   a) Supervised and Unsupervised machine learning.
   b) Logistic and Polynomial regression
   c) Tuple and list
   d) Variance and standard deviation

### QUESTION 2

**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's standard deviation. [6]

**QUESTION 3**

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

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

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

c)
```
fruits=("apple", "banana", "cherry")
mytuple=fruits* 2
print(mytuple)
```

d)
```
thislist=["apple", "banana", "cherry"]
del thislist[0]
          print(thislist)
```

e)
```
x= 41
if x> 10:
  print("Aboveten,")
  if x> 20:
    print("and also above 20!")
  else:
    print("but not above 20.")
```

## QUESTION 5

5. Explain the following pandas codes output [16]

a) filtered_df = df[df['Age'] > 28]
b) df.describe()
c) df.info()
d) df.head(10)
e) df.summary()
f) df['Salary'] = [50000, 60000, 70000]
g) df.to_csv('output.csv', index=False)
h) df = df.drop('Salary', axis=1)

## 6. QUESTION 6

a) 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**