NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATICS

DEPARTMENT OF INFORMATICS

QUALIFICATIONS: Bachelor of Informatics Honours	
QUALIFICATION CODE: 08BIFH, 08BIHB	**LEVEL: 8**
COURSE CODE: DSA821S	**COURSE:** Data Science and Analytics
DATE: November 2025	**SESSION: 1**
DURATION: 3 Hours	**MARKS: 100**

FIRST OPPORTUNITY EXAMINATION QUESTION PAPER	
EXAMINERS:	Prof. Stephen Fashoto
MODERATOR(S):	Ms. Emilia Shikeenga

THIS EXAMINATION PAPER CONSISTS OF 4 PAGES

(INCLUDING THIS FRONT PAGE)

INSTRUCTIONS FOR THE CANDIDATE

1. Answer any four QUESTIONS.
2. When writing, take into account: The style should inform than impress, it should be formal, in third person, paragraphs set out according to ideas or issues, and the paragraphs flowing in a logical order.
3. Information should be brief and accurate.

Please ensure that your writing is legible, neat and presentable

QUESTION ONE **[25marks]**

a) Python environment is in two modes. Name them 2marks

b) What is another name for K-means clustering? 1mark

c) Differentiate between inter-cluster and Intra-cluster with the support of a diagram 3marks

d) List three limitations of K-means clustering 3marks

e) List two methods for choosing optimal value of K in K-means Clustering 2marks

f) List four clustering evaluation metrics 4marks

g) Given the data point in the table below, initialize the k-means clustering algorithm with two cluster centers c1 =(2,10) and c2=(5,8) using Manhattan distance. What are the values of c1 and c2 after one iteration of k-means clustering? What are the values of c1, and c2 after the second iteration of k-means clustering? 10marks

Manhattan distance formula d(x,y)=$\sum|x_i - y_i|$

Point	Coordinates
X1	(2,10)
X2	(2,5)
X3	(8,4)
X4	(5,8)
X5	(7,5)
X6	(6,4)
X7	(1,2)
X8	(4,9)

QUESTION TWO **[25Marks]**

a) List and explain three challenges of supervised learning models 6marks

b) Consider the binary classification problem in the Table below to calculate the following

i)Label the confusion matrix table appropriately first 1mark

ii) Accuracy 2marks

ii)Precision 2marks

iii)Recall 2marks

iv) F1-score 2marks

v) specificity 2marks

vi) Interpret the results based on the findings on precision and recall from the calculations 2marks

	Predicted:spam	Predicted:Not spam
Actual:spam	75	30
Actual:Not spam	15	110

c) What will happen if you deploy an AI model without evaluating its performance with known test set data? Support your answer with only three reasons 6marks

QUESTION THREE **[25Marks]**

a) Differentiate between the following

i) Overfitting and underfitting 2marks

ii) Supervised and unsupervised learning 2marks

b) I would like you to perform 5-fold cross-validation on any 10 data points 6marks

c) Write short notes on the steps involved in CRISP-DM 7marks

d) Assuming gender is the target variable in the Table below. what will be its implication when you carry out an exploratory data analysis on it and explain three ways it can be resolved from data quality perspective? 8marks

Student_number	Programme	Age	Religion	gender
001	Informatics	23	Christianity	Male
002	Computer science	21	Muslim	Male
003	Cybersecurity	32	Christianity	Male
004	Informatics	30	Christianity	Female
005	Informatics	22	Christianity	Male
006	Software engineering	25	Christianity	Male

007	Informatics	26	Muslim	Male
008	Computer science	27	Christianity	Male
009	Informatics	22	Muslim	Female
010	informatics	23	Christianity	Male

QUESTION FOUR **[25Marks]**

a) Explain three key skills required in Data Science with the support of a venn diagram 9marks

b) Write short note on any five challenges of Data Science 5marks

c) List and explain any five data quality problems that can affect classification model performance 5marks

d) Explain any three issues of data science methodologies 6marks

QUESTION FIVE **[25Marks]**

a) A database has five transactions as shown in the Table below, Apply Apriori algorithm on the Transaction data using 40% minimum support threshold and 60% of minimum confidence threshold. You are expected to stop at 3-itemset. 9marks

Transaction ID	Items
T1	B,M
T2	B,D,E,G
T3	C,D,G,M
T4	B,D,G,M
T5	B,C,D,M

b) Write out the output of using TransactionEncoder() on the Table above 5marks

c) Proof that the association rules below are commutative or not 1mark

i) {B,D}→{G} 2marks

ii) {G}→{B,D} 2marks

d) Write short note on the following

i) Lift 2marks

ii) Conviction 2marks

e) Differentiate between itemset and frequent itemset 2marks