National Forensic Sciences University
School of Cyber Security and Digital Forensics
Course Name: M.Tech AI_DS (Batch: 2024-25)
Semester - 1 Exam: TA - I (SEP – 2024)

Subject Code: CTMTAIDS SI P4 **Time: 2:30 PM to 3:15 PM**

Subject Name: Fundamentals of Data Science and Machine Learning **Date: 10-sep-2024**

Q1. Multiple Choice Question. (Attempt All) **[05 Marks]**

1. Which statistical measure is used to quantify the spread of data around the mean?

   A. Mode
   B. Median
   C. Standard Deviation
   D. Range

2. In Pandas, what is the primary function of the dropna() method?

   A. Fills missing values with a specified value.
   B. Removes rows or columns containing missing values.
   C. Converts missing values to a specific data type.
   D. Replaces missing values with interpolated values.

3. Given a Pandas DataFrame df with a missing value in column 'A' at index 2, which interpolation method would be most suitable if you believe the missing value is likely to be close to the average of the values at indices 1 and 3?

   A. linear
   B. time
   C. nearest
   D. quadratic

4. In a confusion matrix for a binary classification problem, what does the diagonal represent?

   A. The number of correct predictions.
   B. The number of incorrect predictions.
   C. The total number of instances.
   D. The number of instances in the positive class.

5. A distribution with a long tail to the right is said to be:

   A. Skewed left
   B. Skewed right
   C. Symmetric
   D. Bimodal

Q2. Answer the following questions in short. (Attempt any 4) **[8 Marks]**

1. How can you detect outliers using a box plot? What is lower and upper extreme formula using IQR?
2. Why preprocessing of data performed?
3. Write any 2 Data Reduction method.
4. What is use of kurtosis? explain its types with Diagram.
5. Provide step included in Data Science Process
6. Difference between Normalization and Standardization.

Q3. Answer the following questions in Detail. (Attempt any 3) **[12 Marks]**

1. Write a python function to perform following transformation (final value can be float)

1 feet = 30.48 cm

1 inch = 2.54 cm

Name	Height
Ram	5'.11"
Shyam	6' 1"
Sita	5' 5"

=>

Name	Height
Ram	180
Shyam	185
Sita	165

2. Write a pandas code for creation of Dataframe as per following values.

Name	Height	Age	Weight	Salary
Ram	180	32	82	60K
Shyam	185	37	78	70K
Sita	165	30	58	50K

3. Provide 3 ways to handled null value with supportive python code.
4. What is Confusion Matrix, recall rate?
   A model outputs 3 TP, 4 TN, 2 FP, and 1 FN.
   Calculate the Accuracy, precision, FPR and Recall in Percentage.

5x30 + 2x11

150 + 22

150/172

Seat No:_____ Enrolment No: 240103007003

National Forensic Sciences University
School of Cyber Security and Digital Forensics
Course Name: M.Tech AI_DS (Batch: 2024-25)
Semester - 1 Exam: MID (OCT – 2024)

Subject Code: CTMTAIDS SI P4 **Time: 11:00 AM to 12:30 PM**

Subject Name: Fundamentals of Data Science and Machine Learning **Date: 10-Oct-2024**

Instruction:

1. **Do all necessary assumptions.**
2. **Support your answer with proper diagram**
3. **Make a cross line on the unused part of your answer sheet.**
4. **New question on new page.**

Q1. True/False (Attempt All) **[10 Marks]**

1. Range is a statistical measure used to quantify the spread of data around the mean.
2. A distribution with a long tail to the right is said to be Skewed right.
3. Machine learning algorithms can learn from data and improve their performance over time.
4. ROC curves are used to evaluate the performance of classification models.
5. EDA cannot be an iterative process.
6. A distribution with a kurtosis of 0 is considered mesokurtic, meaning it has tails similar to a normal distribution.
7. A Type 2 error is also known as a false negative.
8. A Confidence Interval = 1 + significance level (alpha).
9. As Threshold value increases, the number of true negatives will decrease.
10. A sigmoid function is utilized in logistic regression to model the relationship between the independent variables and the binary dependent variable.

Q2. Answer the following questions. (Attempt any 4) **[20 Marks]**

1. Difference between Normalization and Standardization with equations.
2. Explain any 2 classification techniques in detail.
3. Difference between Regression and Classification with example.
4. What is the use of ROC? explain it with a Diagram.
5. Provide diagram of real-world Data Science Process and elaborate.
6. What is the Confusion Matrix?
   A model outputs 13 TP, 14 TN, 2 FP, and 1 FN.
   Calculate the Accuracy, Sensitivity, Specificity, FPR and Precision.

Seat No:_____ Enrolment No:_____

Q3. Answer the following questions. (Attempt any 2) **[20 Marks]**

1. Write a python function to perform following transformation

Name	Weight	Salary
Ram	82.1	60,000
Shyam	78.3	70-000
Sita	58.2	50:000

Name	Weight	Salary
Ram	82	60,000
Shyam	78	70,000
Sita	58	50,000

2. Write a code where csv file provided as per following values. Now use linear regression to predict about salary from Experience=120, Age=40, IQ = 90.

Name	Experience	Age	IQ	Salary
Ram	5	32	82	60000
Shyam	6.5	37	78	70000
Sita	6	30	58	50000

3. Suppose we are now given a new feature vector:

   Weather condition: Rain, Road condition: good, Traffic condition: normal, Engine problem: no. The task is to predict if an accident will happen?

   Which classification will you apply here? why? how?

SNo.	Weather condition	Road condition	Traffic condition	Engine problem	Accident
1	Rain	bad	high	no	yes
2	snow	average	normal	yes	yes
3	clear	bad	light	no	no
4	clear	good	light	yes	yes
5	snow	good	normal	no	no
6	rain	average	light	no	no
7	rain	good	normal	no	no
8	snow	bad	high	no	yes
9	clear	good	high	yes	no
10	clear	bad	high	yes	yes