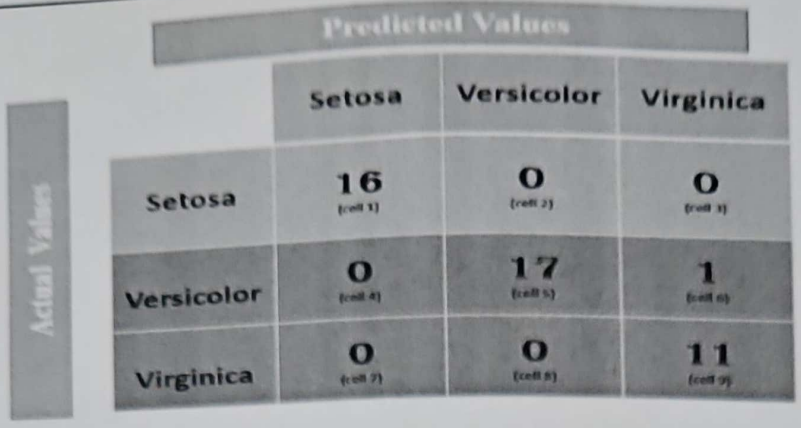
# National Forensics Sciences University, Goa Campus
## Mid-Sem Examination

**Program Name – M.Sc. Cyber Security** **Sem – I** **Date - 8/10/2024**
**Subject Name- Artificial Intelligence** **Subject Code - CTMSCS SI P4**
**Time- 1.5 hr** **Max. Marks- 50**
**Instructions - 1) Answer all questions. 2) Assume suitable data.**

| | | | |
|---|---|---|---|
| Q.1 | **Answer any 4** | | 20 marks |
| | I. | Mention the 5 Evaluation Metrics For Regression, state the equations and importance of each. | |
| | II. | Draw and Calculate the output of a perceptron, considering the input values (x1, x2, x3) as [ 0.5,0.7,0.2] and Weights (w1, w2, w3, b) as [0.4,0.3,-0.1,0.2]. | |
| | III. | Write a short on Hyper parameter tuning. Mention any 1 technique in detail. | |
| | IV. | Explain the types of learning in ML.<br>Describe reinforcement learning. | |
| | V. | Define Feature Extraction.<br>Give an example from cyber security.<br>How does Feature Extraction differ in ML & DL? | 5(=1+3+1) |
| Q.2 | **Attempt all.** | | 15 marks |
| | I. | Consider the scores of 5 students : 88, 58, 67, 79, 93.<br>Find mean, median, standard deviation and variance.<br>State the python library to find the mode of an array. | |
| | II. | Write a short note on Dendograms. | |
| | III. | Calculate the TP, TN, FP, FN for the Virginica class, of the given confusion matrix. | |

| | | |
|---|---|---|
| |  | |



| | | |
|---|---|---|
| Q.3 | **Answer any 1** | 7 Marks |
| 3.1 | Mention the 5 functions for Handling and Representing data.<br>State and elaborate the technique to apply all of them together on a project. | 7(= 5+2) |
| 3.2 | Give an example of Outlier in data.<br>Mention and elaborate two techniques for Outlier Detection in detail. | 7 (=1+3+3) |
| Q.4 | **Answer any 1** | 8 Marks |
| 4.1 | For the given dataset of children's health, implement 1 iteration of k-means considering the child 3 & child 4, as centroids.<br><br>Also, write down the equations for new centroids, such that a 3rd feature - Cholesterol level [x1 to x6] had been given.<br><br>Compare the K-means with DBSCAN. | 8 (=6+1+1) |
| 4.2 | Write an elaborate note on the Hierarchical Clustering and its types. | 8 (=4+4) |

| | Height ( in mm) | Weight ( in kg) |
|---|---|---|
| child 1 | 566 | 6 |
| child 2 | 674 | 4 |
| child 3 | **844** | **8** |
| child 4 | **429** | **3** |

# National Forensics Sciences University, Goa Campus

## TA-1 Examination

Program Name – M.Sc. Cyber Security | Sem – I | Date- 10 /09/2024

Subject Name- Artificial Intelligence | Subject Code- CTMSCS SI P4

Time- 45 minutes | Max. Marks- 25

Instructions - 1) Answer all questions. 2) Assume suitable data.

| | | |
|---|---|---|
| Q.1 | Match the pairs (1 mark each) | 5 marks |
| | Write down the combination of the alphabet letter with the roman number, such that the two relate to a certain ML skill.<br>a. SNV.<br>b. classification_report<br>c. Interactive data, Generate plots and Generate Code.<br>d. Metric for regression.<br>e. Criteria for ML based- Food Analysis product in USA<br>i. RPD<br>ii. Bias<br>iii. Precision, Recall, F1-score<br>iv. MS Co-pilot<br>v. Preprocessing of Spectral data. | |
| Q.2 | Answer as per the instruction:- | 5 marks |
| | I. State True/False : To obtain the titles of each of the features in the dataset, .col_heads() command in python has to be used. | 1 mark |
| | II. The python command which displays last 5 lines of the dataset is _____. | 1 mark |
| | III. Give an example to justify that the pd.Series() and np.arange() deliver different results for the same inputs as arguments. | 1 mark |
| | IV. An ML algorithm which predicts C02 emission with features - engine size and no.of cylinders is _____<br>a. univariate type<br>b. multi-variate type. | 1 mark |
| | V. State the use of .info() on a dataset, in python. | 1 mark |
| Q3 | Answer any 3 questions ( 3x5 marks each) | 15 Marks |
| | i. Write a short on .describe() function in python, with an example. Mention different visual tools in python. | 5 marks |
| | ii. State and describe the different types of learning in ML, with a neat diagram. | 5 marks |

P.T.O.

My marks 15

| | | |
|---|---|---|
| | iii. With the help of a neat diagram, justify : Increasing the precision of a ML model, decreases the recall, and vice-versa.<br><br>State and explain the metric : F1-score. | 5 (=3+2) marks |
| | iv. Calculate the TP, TN, FP, FN for the Versicolor class, of the given confusion matrix.<br><br>Mention the 2 functions of scikit-learn in Python for confusion matrix. | 5 (=4+1) marks |

| Actual Values \ Predicted Values | Setosa | Versicolor | Virginica |
|---|---|---|---|
| Setosa | 16 | 0 | 0 |
| Versicolor | 0 | 17 | 1 |
| Virginica | 0 | 0 | 11 |

TP = 17
FN = 1
FP = 0
TN = 27

Seat No.: _____ Enrolment No._____

# NATIONAL FORENSIC SCIENCES UNIVERSITY

## M.SC. Digital Forensics and Information Security

### Mid-Semester Exam

**Subject Code: CTMSCS SI P4** **Date: 17/03/2025**

**Subject Name: Artificial Intelligence**

**Time: 11:00 to 12:30 PM** **Total Marks: 50**

**Instructions:**

1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**
4. **Calculators are allowed.**

| Q.1 | | Attempt all. | 20 Marks |
|---|---|---|---|
| | a) | State and explain any 2 different cross-validation functions Python.<br>What is stratified CV? | 2+2 |
| | b) | Apply k-means for the dataset [2,3,4,6,12,14,15,16,21,23,25,30,31,35,38] , to find new **clusters.**<br>Given initial centroids : C1=2, C2=16, C3=38. | 2+2 |
| | c) | State the relevant step of machine learning at each of the comment, with the specified numerical order. | 2*5 |

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler

def data_preprocessing_pipeline(data):
    #1
    numeric_features = data.select_dtypes(include=['float', 'int']).columns
    categorical_features = data.select_dtypes(include=['object']).columns

    #2
    Unempty_data] = data[numeric_features].fillna(data[numeric_features].mean())

    #3
    for feature in numeric_features:
        Q1 = data[feature].quantile(0.25)
        Q3 = data[feature].quantile(0.75)
        IQR = Q3 - Q1
        lower_bound = Q1 - (1.5 * IQR)
        upper_bound = Q3 + (1.5 * IQR)
        data[feature] = np.where((data[feature] < lower_bound) | (data[feature] >
upper_bound),
                                 data[feature].mean(), data[feature])
```

```
# 4
scaler = StandardScaler()
scaled_data = scaler.fit_transform(data[numeric_features])
data[numeric_features] = scaler.transform(data[numeric_features])

# 5
data[categorical_features] =
data[categorical_features].fillna(data[categorical_features].mode().iloc[0])

return data
```

d) Draw & explain the regions of Bias-Variance trade-off plot in AI-ML. 2

| | | | |
|---|---|---|---|
| Q.2 | | **Attempt any four.** | **20 Marks** |
| | (a) | What is data remediation? Mentioning a library in sklearn, explain any 3 techniques used to handle the issue of missing values. | 2+3 |
| | (b) | Write a program in Python to obtain solution of set of two linear equations. | 5 |
| | (c) | Explain the various methods used to explore the relationship between two variables, and their implementation in Python. correlation, | 5 |
| | (d) | What are the stages of the data science- machine learning lifecycle, and how do they collectively assist in making data-driven decisions? | 5 |
| | (e) | Is it possible to improve the performance of a learning model? If so, explain the methods to enhance its performance. | 5 |
| Q.3 | | **Attempt any one.** | **10 marks** |
| | (a) | i. State the equation and importance of i) RMSE and ii) RPD.<br><br>ii. Mention the 2 functions of scikit-learn in Python for confusion matrix.<br><br>iii. Mention the FN for virginica class, of the given confusion matrix( Fig 1) | 4+2+2+2 |

| | Setosa | Versicolor | Virginica |
|---|---|---|---|
| Setosa | 21 | 0 | 0 |
| Versicolor | 0 | 24 | 3 |
| Virginica | 0 | 2 | 25 |

iv. From the given matrix, write meaning of each of 4 values( Fig. 2)

| | |
|---|---|
| 51 | 16 |
| 21 | 27 |

x-axis (non-event and event) and y-axis (non-event and event) are actual and predicted

| | | | |
|---|---|---|---|
| | (b) | i. Mention and explain 3 different encoding methods.<br><br>ii. Mention the commands for the TF-IDF pre-processing techniques in NLP.<br>Illustrate their functioning, using the following example:-<br>Sent 1: movie was good.<br>Sent 2: movie was bad.<br>Sent 3: movie was not good. | 6+4 |

**--- End of Paper---**

Seat No.: _____ Enrolment No._____

# NATIONAL FORENSIC SCIENCES UNIVERSITY

## M.Sc. Digital Forensics and Information Security

**Subject Code: CTMSDFIS SII P4** **Date: 10/02/2025**

**Subject Name: Artificial Intelligence**

**Time: 11:00 AM to 11:45 AM** **Total Marks: 25**

**Instructions:**

1. **Write down each question on a separate page.**
2. **Attempt all questions.**
3. **Make suitable assumptions wherever necessary.**
4. **Figures to the right indicate full marks.**

| 1. | Answer all | 10 Marks |
|---|---|---|

1. Which of the following is used to install external libraries in Python?
   a) install package
   b) pip install
   c) python install
   d) lib install

2. What is the correct way to create a dictionary in Python?
   a) dict = {1: "apple", 2: "banana"}
   b) dict = [1: "apple", 2: "banana"]
   c) dict = (1: "apple", 2: "banana")
   d) dict = {1, "apple", 2, "banana"}

3. Which of the following libraries is used for data visualization in Python?
   a) NumPy
   b) Pandas
   c) Matplotlib
   d) SciPy

4. What function is used to create a NumPy array?
   a) np.array()
   b) numpy.create()
   c) np.new_array()
   d) numpy.list_to_array()

5. What will be the output of the following Python code?

```
for i in range(2, 5)

print(i, end=" ")
```

   a) 2 3 4
   b) 2 3 4 5
   c) 1 2 3 4

d) 3 4 5

6. What is the dot product of two vectors **A = [1, 2]** and **B = [3, 4]**?
   a) 1*3 + 2*4 = 11
   b) 1*4 + 2*3 = 10
   c) 1+2+3+4 = 10
   d) 1*3 - 2*4 = -5

7. If the mean of a dataset is 50 and the standard deviation is 5, what is the variance?
   a) 5
   b) 25
   c) 50
   d) 100

8. Which of the following is **not** a measure of central tendency?
   a) Mean
   b) Median
   c) Mode
   d) Variance

9. In ML, which is the dependent variable?
   a) The input variable
   b) The output variable
   c) The constant term
   d) The error term

10. Which command is used to display the first 5 rows of a Pandas DataFrame df?

    a) df.top(5)

    b) df.above(5)

    c) df.head(5)

    d) df.first(5)

| 2. | **Attempt any three.** | **15 Marks** |
|---|---|---|
| a. | Write a short note on Python commands for the Handling and Representing Data. | 1*5 |
| b. | Given the data : failed Login Attempts, for a month of 4 weeks, [20,22,21,25,24],<br>Calculate all the statistics (mean, mode, median, standard deviation, variance) **per week,**<br>by writing a Python function Print_Attacks.<br>Write the results in an excel file. | 4+1 |
| c. | Using NumPy.<br>i) create a 1D array of numbers from 0 to 8,<br>ii) reshape it into a 3*3 array,<br>iii) create a DataFrame from the reshaped array using Pandas,<br>iii) plot a suitable graph of the DataFrame using Matplotlib.<br>iv) Save the figure. | 1*5 |
| d. | Explain the steps to set up a Python development environment using Anaconda.<br>Write a Python script that takes a user input number and prints "Even" or "Odd" accordingly. | 2+3 |

**--- End of Paper---**

Seat No.: _____ Enrolment No._____

# NATIONAL FORENSIC SCIENCES UNIVERSITY

**M.Sc. (DFIS) (Semester II)**

Semester End Examination –April- 2025

**Subject Code: CTMSDFIS24 SII P4** **Date:**29.04.2025

**Subject Name:** Artificial Intelligence

**Time:** 10:30 AM to 1:30 PM **Total Marks:** 100

**Instructions:**

1. **Write down each question on a separate page.**
2. **Attempt all questions.**
3. **Make suitable assumptions wherever necessary.**
4. **Figures to the right indicate full marks.**

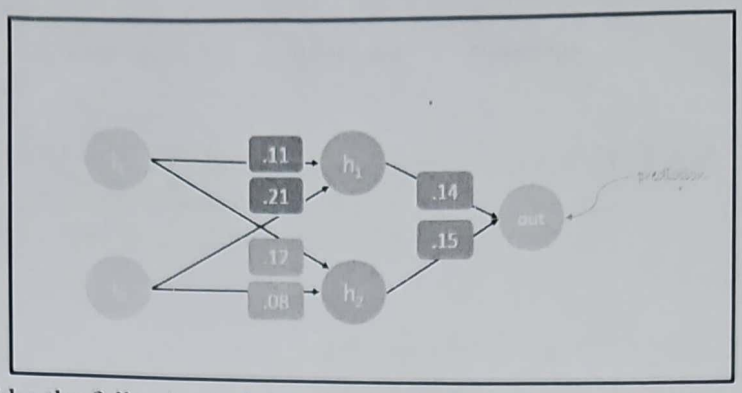
| | | Marks |
|---|---|---|
| **Q.1** | **Attempt any three.** | |
| (a) | Write a Python Script/Commands for following: create a 1D array of numbers from 0 to 5 using NumPy, ii) reshape it into a 3*2 array, iii) create a DataFrame from the reshaped array using Pandas, iv) Plot a bar graph of the DataFrame using Matplotlib. v) Save the figure. | 08 |
| | OR | |
| | Given the data of a laboratory having 5 systems for a month (of 4 weeks), failed Login Attempts, [20,20,24,32,28], write a Python script and User defind defined function Print_Attacks to calculate all the statistics (mean, mode, median, standard deviation, variance) for the lab per week. Write the results in an excel file. | |
| (b) | Consider the example below where the mass, y (grams) of a chemical is related to the time, x (seconds), for which the chemical reaction has been taking place according to the table: | 08 |

| Time, x (seconds) | 5 | 7 | 12 | 16 | 20 |
|---|---|---|---|---|---|
| Mass, y (grams) | 40 | 120 | 180 | 210 | 240 |

Find the equation of the regression line. Using resultant regression line predict the data points.

| | | |
|---|---|---|
| (c) | List out and briefly explain types of Clustering. Discuss the working of K-Means Clustering algorithm. | 08 |
| (d) | Bring out the difference between following:<br>1) Classification vs Prediction<br>2) Machine Learning vs Deep Learning | 08 |

| | | |
|---|---|---|
| **Q.2** | **Attempt any three.** | |
| (a) | Explain in brief the different types of layers in an ANN? Use perceptron algorithm and calculate value for hidden layer neurons ($h_1$, $h_2$). Consider inputs=[2, 3] and actual output=[1]. | 08 |




**(b)** Consider the following training data for the Naive Bayes Classifier **08**

| Type of family structure | Age group | Income status | Will they buy a car? |
|---|---|---|---|
| Nuclear | Young | Low | Yes |
| Extended | Old | Low | No |
| Childless | Middle-aged | Low | No |
| Childless | Young | Medium | Yes |
| Single Parent | Middle-aged | Medium | Yes |
| Childless | Young | Low | No |
| Nuclear | Old | High | Yes |
| Nuclear | Middle-aged | Medium | Yes |
| Extended | Middle-aged | High | Yes |
| Single Parent | Old | Low | No |

Given test inputs – {Single Parent, Young, and Low}, compute the probability of buying a car.

**(c)** Describe the types of data quality issues possible in real time dataset. Discuss different steps of Data Pre-Processing. **08**

**(d)** Write a note on following: **08**
(I)Overfitting (II) Perceptron (III) Image Segmentation (IV) Labelled vs Unlabeled Data

**Q.3** **Attempt any three.**

**(a)** The following confusion matrix shows data for a Binary-Class problem where classification model predicts whether a person loves Facebook or Instagram. Calculate Accuracy, TN rate, Precision, Recall, Error Rate, and FP Rate. **08**

| | | **Predicted** | |
|---|---|---|---|
| | | **Facebook** | **Instagram** |
| **Actual** | **Facebook** | 66 | 9 |
| | **Instagram** | 22 | 31 |

**(b)** Explain the working of a Convolutional Neural Network (CNN), describing its components and how it is used for image classification. Illustrate the process with an example. **08**

**(c)** Explain the key steps involved in the Natural Language Processing (NLP) pipeline, providing examples of their application. **08**

supervised, unsupervised and reinforcement

| | | | |
|---|---|---|---|
| | (d) | Discuss in detail the various types of Machine learning models. 5 | 08 |
| Q.4 | | **Both the questions are compulsory** | |
| | (a) | List out and discuss different static as well as dynamic features, which can be used for malware analysis using Machine Learning 3 | 07 |
| | (b) | What is activation function? Explain different types of activation function in detail? 5 | 07 |
| Q.5 | | **Both the questions are compulsory** | |
| | (a) | Define object tracking. List out and discuss types of object tracking. Explain object tracking algorithm with an example 3 | 07 |
| | (b) | Explain the architecture of LSTM with help of Diagram. What is significance of different types of gates in the working of LSTM? 4 | 07 |

**--- End of Paper---**