Register Number 183CS22014

Code: 20CS51IT

V Semester Diploma MakeUp Examination July 2025
Department of Computer Science & Engineering

Course: Artificial Intelligence & Machine Learning Max Marks: 100

Course Code: 20CS51IT Duration: 3 Hrs

Instruction to the Candidate: Answer one full question from each section.

Section-1

1a. Describe any 5 machine learning challenges. 5M

1b. Differentiate between supervised and unsupervised machine learning 5M

1c. Write steps to Create repository in GitHub and add file 10M

OR

2a. How Big data is different from the data stored in traditional databases? Elaborate 5M

2b. Explore different sources of big data in machine learning 5M

2c. A person decides to walk 10000 steps every day to combat the effect that lockdown has had on his body's agility, mobility, flexibility and strength. Consider the following data from fitness tracker over a period of 10 days 10M

Sl.No	Day	Steps
0	1	4335
1	2	9552
2	3	7332
3	4	4504
4	5	5335
5	6	7552
6	7	8332
7	8	8965
8	9	8965
9	10	7689

i) Write a program that returns the steps walked if the steps walked are more than 9000.

ii)Print an array containing steps walked in sorted order.

Section-2

3a. Explain how to handle the outliers in the dataset? 10M

3b. A dataset is given to you for creating machine learning model. What are the steps followed before using the data for training the model? Elaborate each step.

10M

OR

4a. Describe univariate, bivariate, and multivariate analysis with suitable examples 10M

4b. Elaborate statistical analysis for the given dataset.

10M

	Age	Study Hours (per week)	Test Score (out of 100)
Count	150	150	150
Mean	20.4	15.2	72.5
Std	2.8	4.5	10.6
Min	18	5	45
25%	19	12	65
50% (Median)	20	15	75
75%	22	18	80
Max	27	25	95

Analyse and explain statistical metrics from above summary

Section 3

5a. A Machine learning model was built to classify patient as cancer +ve(1) or – ve(0). The confusion matrix for the model is as shown below. Evaluate accuracy, precision, recall, Specificity and F1-Score 10M

5b. Compare classification algorithm with clustering algorithms 5M

5c. N-grams are defined as the combination of N keywords together. Consider the given sentence: "Big Data refers to extremely large and complex datasets that are difficult to process, manage, or analyze using traditional data processing tools." Generate bi-grams and tri-grams for the above sentence

5M

OR

6a. For the given data set perform the following operations:

i) Check statistical info of the data set

ii) Plot a line plot showing total profit on 'y' axis and number column on 'x' axis

iii) Find the missing values

iv) Find the sum of total profit

v) Find the max value from Drawing sheets column

10M

Number	Pencil	Textbooks	Drawing Sheet	Total units	Profit
1	300	250	100	700	80000
2	350	350	125	1075	9500
3	400	400	190	1320	10256
4	500	420	210	1510	12000
5	520	500	250	1516	15000

6b. Explain stemming and lemmatization with examples 5M

6c. Compare Bagging with Boosting. 5M

Section 4

7a. Write python code to explain map(), filter(), reduce(), lamba(),describe() 10M

7b. Discuss importance of dimensionality reduction in machine learning 5M

7c. Explain any 5 application of sentimental analysis 5M

OR

8a. Explain any two-activation function in neural network 10M

8b. Differentiate between forward propagation and back propagation 5M

8c. Explain different stages that are involved in the MLOps lifecycle 5M

Section 5

9a. With a neat diagram explain components of docker. 10M

9b. Demonstrate Simple Linear Regression considering a dataset that has two variables: salary (dependent variable) and experience (Independent variable). 10M

OR

10a. Summarize any two cloud deployment models 10M

10b. Write Ethical challenges in Artificial Intelligence 10M