

1007**Code : 20CS51I**

Register Number	4	1	3	C	S	2	1	0	2	9
--------------------	---	---	---	---	---	---	---	---	---	---

V Semester Diploma Examination, December-2023**ARTIFICIAL INTELLIGENCE
AND MACHINE LEARNING****Time : 3 Hours |****| Max. Marks : 100****Instructions :** Answer any **one** full question from each section.**SECTION – I**

1. (a) Define AI and describe applications of AI in different domain. **10**
(b) Summarize any two cloud deployment models. **5**
(c) Write steps to create Repository in GitHub. **5**

2. (a) Explain how AI software development life cycle differs from traditional software development. **5**
(b) Differentiate between Supervised and Unsupervised Machine Learning. **5**
(c) Create a data frame with following data :

Ename	Department	Experience	Salary
Roshan	CS	10	45000
Amar	CS	20	50000
Ashwini	EC	15	30000
Lonith	EC	14	25000
Mohan	CS	9	10000
Pramod	EC	8	40000

- (i) Make a Pivot table that shows the average salary of each employee for each department. **10**
- (ii) Make a Pivot table that shows the sum and mean of the salaries of each employee.



SECTION – II

3. (a) How to handle missing values in the dataset ? Explain. 10
 (b) With suitable example explain map (), filter (), reduce () and Lambda () functions with Numpy. 10
4. (a) 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. 10
 (b) Describe Univariate and Multivariate Analysis. 10

SECTION – III

5. (a) How to handle Outliers in dataset ? Explain. 10
 (b) A company wants to study Iris dataset to make predictions. So perform the following operations : 10
 (i) Import libraries & read CSV file.
 (ii) Check for Missing Values.
 (iii) Count plot for the species.
 (iv) Scatter plot to compare petal length and petal width.
 (v) Visualize the distribution of any one column.
6. (a) Compare Classification algorithms with Clustering Algorithms. 5
 (b) Discuss different techniques of Cross validation. 5
 (c) The confusion matrix for a Machine Learning Model is given below. Evaluate. 10
 (i) Accuracy
 (ii) Precision
 (iii) Recall
 (iv) Specificity
 (v) F1-Score

		Actual	
		1	0
Predicted	1	45	8
	0	15	32

SECTION – IV

7. (a) Demonstrate simple Linear Regression considering a dataset that has two variables : 10
 “Rain fall rate” Independent variable
 “Crop yield” Dependent variable
(b) Compare overfitting with Underfitting. 5
(c) Discuss importance of dimensionality Reduction in Machine Learning. 5
8. (a) Build decision tree based model to perform the following operations on Breast cancer dataset : 10
 (i) Import libraries
 (ii) Perform preprocessing
 (iii) Split the dataset
 (iv) Find the Accuracy
 (v) Data prediction
(b) N grams are defined as the contribution of N keywords together consider the given sentence. 10
“Data Visualization is a way to express your data in a visual context so that patterns, correlations, trends between the data can be easily understood”.
(i) Generate bi-grams and tri-grams for the above sentence.

SECTION – V

9. (a) Demonstrate stemming and Lemmatization with suitable example. 10
(b) Explain briefly different stages involved in Machine Learning Operations (MLOPs) life cycle. 10
10. (a) With a neat diagram explain components of Docker. 10
(b) What are ethics in AI and why ethical practices should be followed while developing solutions using AI ? 10
-

Register No:

Code : 20CS51IT

V Semester Diploma Examination, May 2025
ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

Duration: 3 Hours

Max.Marks: 100

Instructions: Answer one full question from each Section.

SECTION-I

- 1.a Describe AI and its applications in various fields. 10 Marks
1.b Differentiate AI Software Development life cycle differs from traditional software development. 05 Marks
1.c List any 5 challenges associated with Machine Learning. 05 Marks
- 2.a Perform the following operations/write code snippet on Car manufacturing company dataset "auto-mpg.csv" given below using pandas. 10 Marks
- i) Read data from a file.
 - ii) Calculate mean value of "horsepower".
 - iii) Calculate Standard Deviation value of "acceleration".
 - iv) Get the number of cars manufactured in each year.

mpg	cylinders	displacement	horsepower	weight	acceleration	model	year	car name
18	8	307	130	3504	12	71		Chevrolet
15	8	350	165	3693	11.5	70		Skylark
18	8	318	150	3436	11	72		Plymouth
17	8	302	140	3449	10.5	70		Ford
14	8	455	225	4425	10	71		Pontiac
15	8	390	190	3850	8.5	70		Ambassador

- 2.b Explain how is AI software development life cycle different from traditional software development. 10 Marks

SECTION-II

- 3.a Handling missing values in a dataset is a crucial data pre-processing step, as missing data can lead to biased or incorrect results in your analysis or machine learning models. Elaborate on how missing values in the data sets can be handled. 10 Marks
3.b 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. 10 Marks

- 4.a Create two series as shown using pd. series() function. 10 Marks
Series A = [10, 20, 30, 40, 50]
Series B = [50, 60, 70, 80, 90]
- (i) Get the items not common to both.
 - (ii) Identify the largest element in the Series A.
 - (iii) Find the sum of Series B.
 - (iv) Calculate mean in the Series A.
 - (v) Find median in the given Series B.
- 4.b Explain the concepts of univariate and multivariate data types with suitable examples. 10 Marks

SECTION – III

- 5.a A machine learning model is built to detect whether a person has a disease (1) or does not have the disease (0). The following confusion matrix shows the model's performance: 10 Marks
5.b Explain Supervised and Unsupervised learning with examples. 5 Marks
5.c Compare overfitting with under-fitting. 5 Marks

- 6.a How to Choose the Right Number of Clusters in k-means clustering? Explain any one method. 5 Marks
6.b Compare "Classification algorithms" with "Clustering algorithm". 5 Marks
6.c Explain with examples: Scalars, Vectors, Matrices, Tensors and Gradients in Linear Algebra. 10 Marks

SECTION – IV

- 7.a N-grams are a type of linguistic model used in natural language processing (NLP) and computational linguistics. Consider the given sentence :"Artificial Intelligence is revolutionizing various sectors by enabling machines to perform tasks that typically require human intelligence." 10 Marks

- i. Generate bi grams for the above sentence
- ii. Generate tri-grams for the above sentence

- 7.b Explain how data exploration, pre-processing of data and splitting of data are performed on datasets. 10 Marks

- 8.a With examples demonstrate Stemming and Lemmatization normalization techniques. 10 Marks
8b. Explain any 2 techniques of cross validation used in Machine Learning. 5 Marks
8c. Brief explain different stages involved in the Machine Learning Operations (MLOps) lifecycle. 5 Marks

Section - V

- 9.a Demonstrate simple linear regression considering a dataset that has two variables:
"House price" (dependent variable)
"size of house" (independent variable) 10 Marks
9.b With a neat diagram explain components of Docker. 10 Marks
- 10.a Summarize any two cloud deployment models . 10 Marks
10.b Discuss any five ethical challenges in AI. 10 Marks

Reg.No:

Code: 20CS51IT

V Semester Diploma Examination, November /December 2024
ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

TIME: 3 HOURS

MAX MARKS: 100

Instructions:

- (i) Answer **one full question** from each section - I, II, III, IV and V.
- (ii) Each one full question carries **20 marks**.

SECTION - I

1. a) Artificial Intelligence (AI) is a promising state-of-the-art technology that provides intelligent solutions in every field today.
Justify your answer by describing AI and its applications in various fields. 10
- b) How AI Software Development life cycle differs from Traditional Software Development? 5
- c) Explain 5 V's of big data. 5

2. a) Perform the following operations on Car manufacturing company dataset "auto-mpg.csv" given below using pandas.

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin	car name
0	18	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15	8	350	165	3693	11.5	71	1	buick skylark 320
2	18	6	318	150	3436	11.0	70	1	plymouth satellite
3	16	4	304	150	3433	12.0	80	1	amc rebel sst
4	17	8	302	140	3449	10.5	70	1	ford torino

- a) Read data from an existing file 10
- b) Statistical details of dataset
- c) Get all cars with 8 cylinders
- d) Get the number of cars manufactured in each year

- b) Differentiate between Supervised machine learning and Unsupervised machine learning. 10

SECTION - II

3. a) How to handle missing values in the dataset? Explain. 10
- b) 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. 10

4. a) Explain Univariate and Multivariate analysis with example. 10
- b) Create two series as shown using pd.series()
Series A = [20, 30, 40, 50, 60]
Series B = [50, 60, 70, 80, 90]

- i) Get the items not common to both.
- ii) Identify the smallest and largest element in the Series A.
- iii) Find the sum of Series B.
- iv) Calculate mean in the Series A.
- v) Find median in the given Series B.

SECTION - III

5. a) A Machine learning model was built to classify spam emails as "spam"(1) or "not spam"(0). The confusion matrix for the model is as shown below. Evaluate accuracy, precision, recall, specificity and F1- Score. 10

		Actual	
		1	0
Predicted	1	140	10
	0	5	50

- b) Compare Overfitting with Underfitting. 5
 c) Explain any two techniques of Cross Validation used in Machine Learning. 5
 6. a) Demonstrate simple linear regression considering a dataset that has two variables. 10
 "Marks" (dependent variable)
 "Hours of study" (independent variable)
 b) Explain with examples: Scalars, Vectors, Matrices, Tensors & Eigen values in Linear Algebra. 10

SECTION - IV

7. a) How to Choose the Right Number of Clusters in k-means clustering? Explain any one method. 5
 b) Compare "Classification algorithms" with "Clustering algorithm". 5
 c) Explain Basic and Advanced Ensemble techniques. 10
 8. a) Explain various activation functions in Neural Networks. 10
 b) Explain how data exploration, pre-processing of data and splitting of data are performed on datasets. 10

SECTION - V

9. a) N-grams are defined as combination of N Keywords together.
 Consider the given sentence
 "Artificial Intelligence is a branch of computer science that focuses on creating intelligent machines capable of performing tasks that typically require human intelligence." 10
 (i) Generate bi grams for the above sentence.
 (ii) Generate tri-grams for the above sentence.
 b) Demonstrate Stemming and Lemmatization concepts with suitable examples. 10
 10. a) With a neat diagram explain components of Docker. 10
 b) Discuss any five ethical challenges in AI. 5
 c) Explain any two Cloud deployment Models. 5

7666**Code : 20CS51I**Register
Number

4	1	3	C	S	2	0	0	0	6
---	---	---	---	---	---	---	---	---	---

V Semester Diploma Examination, February/March-2023**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING****Time : 3 Hours |****[Max. Marks : 100****Instructions :** Answer **one** full question from each section.

6.

SECTION - I

1. (a) Summarize the challenges associated with Machine Learning. 5
(b) How AI Software Development life cycle differs from traditional software development ? 5
(c) Perform the following operations on Car manufacturing company dataset auto-mpg.csv given below using pandas. 10

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin	car name
0	18	8	307	130	3504	12.0	70	1	Chevrolet chevelle malibu
1	15	8	350	165	3693	11.5	71	1	buick skylark 320
2	18	6	318	150	3436	11.0	70	1	plymouth satellite
3	16	4	304	150	3433	12.0	80	1	amc rebel sst
4	17	8	302	140	3449	10.5	70	1	ford torino

- (i) Read data from an existing file
(ii) Statistical details of dataset
(iii) Get all cars with 8 cylinders
(iv) Get the number of cars manufactured in each year.

OR

20CS51I

2 of 4

2. (a) Differentiate between Supervised machine learning and Unsupervised machine learning. 5
- (b) Explore different sources of big data in machine learning. 5
- (c) Ramesh 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. 10

S. 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	6504
8	9	8965
9	10	7689

- (i) Perform an appropriate operation to add 1000 steps to all the observations
- (ii) Find out the days on which he walked more than 7000 steps

SECTION – II

3. (a) How to handle the missing values in the dataset ? Explain. 10
- (b) A dataset is given to you for creating a machine learning model. What are the steps followed before using the data for training the model ? Elaborate each step.

10

OR

4. (a) A company wants to study iris dataset to make predictions. However, the data gathered is not clean for analysis. The company requests you to write a python code to perform the following operations for data driven competitive advantage (Assume dataset with missing values)
- Check for missing values
- Replace missing values with mean value

10

Age Annual Income (₹) Spending Score (1-100)

100	100000	100000
100	100000	100000
100	100000	100000
100	100000	100000
100	100000	100000
100	100000	100000
100	100000	100000
100	100000	100000

Find and explain relevant metrics from above summary

SECTION - III

- (a) A Machine learning model was built to classify patient as COVID (1) or not (0). The confusion matrix for the model is as shown below. Evaluate accuracy, precision, recall, specificity and F1 score. 10

Actual		
0	1	0
0	100	100
1	100	100

- (b) Explain overfitting with underfitting.
 (c) How to choose the Right Number of Clusters in K-means clustering? Explain any one method. 8

OR

- (d) Cluster the following eight points (with (x_1, y_1) representing locations) into three clusters.
 Points are: A(2, 10), A(2, 3), A(5, 3), A(6, 7), A(7, 3), A(8, 8), A(9, 3), A(10, 9).
 Initial cluster centres are: A(2, 10), A(5, 8) and A(7, 3).
 The distance function between two points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined as

$$d(a, b) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Use K-Means Algorithm to find the three cluster centres after the first iteration.

[Turn over]

- (b) Statistical summary of Credit card dataset is as follows :

10

	Age	Annual Income (k\$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000
mean	38.850000	60.560000	50.200000
std	13.969007	26.264721	25.823522
min	18.000000	15.000000	1.000000
25%	28.750000	41.500000	34.750000
50%	36.000000	61.500000	50.000000
75%	49.000000	78.000000	73.000000
max	70.000000	137.000000	99.000000

Analyse and explain statistical metrics from above summary

SECTION - III

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

10

		Actual	
		1	0
Predicted	1	397	103
	0	126	142

5

- (b) Compare overfitting with underfitting.
 (c) How to Choose the Right Number of Clusters in k-means clustering ? Explain any one method.

5

OR

6. (a) Cluster the following eight points (with (x, y) representing locations) into three clusters :

10

A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)

Initial cluster centres are : A1(2, 10), A4(5, 8) and A7(1, 2).

The distance function between two points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined as

$$P(a, b) = |x_2 - x_1| + |y_2 - y_1|$$

Use K-Means Algorithm to find the three cluster centres after the first iteration

[Turn over]

- (b) Compare classification algorithms with clustering algorithm. 5
(c) K-means clustering with Euclidean distance suffer from the curse of dimensionality. Is the statement true and why ? 5

SECTION – IV

7. (a) N-grams are defined as the combination of N keywords together. Consider the given sentence : “The greatest glory in living lies not in never falling, but in rising every time we fall”. 10
(i) Generate bi grams for the above sentence
(ii) Generate tri-grams for the above sentence
(b) Discuss importance of dimensionality reduction in machine learning. 5
(c) Summarise different strategies of production deployment 5

OR

8. (a) Demonstrate stemming and Lemmatization concepts with suitable examples. 10
(b) Discuss different techniques of cross validation. 5
(c) What are MLOps ? Brief different stages that are involved in the MLOps lifecycles. 5

SECTION – V

9. (a) With a neat diagram explain components of Docker. 10
(b) Demonstrate Simple Linear Regression considering a dataset that has two variables : salary (dependent variable) and experience (Independent variable). 10

OR

10. (a) Summarize any two cloud deployment models. 10
(b) Discuss any five ethical challenges in AI. 10
-

Makeup Examination – Sept. 2023
V Semester Diploma Examination

ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

(Exam Date / Time: 22nd Sep. 2023 / 2.00 PM)

Time: 3 Hours

Max. Marks: 100

- Instructions:**
- (1) Answer one full question from each section.
 - (2) One full question carries 20 marks.

SECTION – I

- | | |
|---|----------|
| 1.a Describe AI and its applications in various fields. | 10 Marks |
| 1.b How AI Software Development life cycle differs from traditional software development. Explain. | 05 Marks |
| 1.c Summarize the challenges associated with Machine Learning. | 05 Marks |
| 2.a Perform the following operations/write code snippet on Car manufacturing company dataset "auto-mpg.csv" given below using pandas. | 10 Marks |
| i) Read data from a file. | |
| ii) Calculate mean value of "horsepower". | |
| iii) Calculate Standard Deviation value of "acceleration". | |
| iv) Get the number of cars manufactured in each year. | |

mpg	cylinders	displacement	horsepower	weight	acceleration	model year	car name
18	8	307	130	3504	12	71	Chevrolet
15	8	350	165	3693	11.5	70	Skylark
18	8	318	150	3436	11	72	Plymouth
17	8	302	140	3449	10.5	70	Ford
14	8	455	225	4425	10	71	Pontiac
15	8	390	190	3850	8.5	70	Ambassador

- 2.b Explain how is AI software development life cycle different from traditional software development? 10 Marks

SECTION – II

- | | |
|---|----------|
| 3.a Handling missing values in a dataset is a crucial data pre-processing step, as missing data can lead to biased or incorrect results in your analysis or machine learning models. Elaborate on how missing values in the data sets can be handled. | 10 Marks |
| 3.b 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. | 10 Marks |
| 4.a Create two series as shown using pd. series() function.
Series A = [20, 30, 40, 50, 60] Series B = [50, 60, 70, 80, 90] | 10 Marks |
| (i) Get the items not common to both. | 10 Marks |
| (ii) Identify the smallest and largest element in the Series A. | |
| (iii) Find the sum of Series B. | |
| (iv) Calculate mean in the Series A. | |
| (v) Find median in the given Series B. | |

a Referring to the number of variables or features in a dataset and the focus of analysis. Explain univariate & multivariate data types with examples. 10 Marks

SECTION – III

a A Machine learning model was built to classify spam emails as "spam"(1) or "not spam"(0). The confusion matrix for the model is as shown below. Evaluate accuracy, precision, recall, specificity and F1-Score. 10 Marks

		Actual	
		1	0
Predicted	1	140	10
	0	5	50

b Explain Supervised and Unsupervised learning with examples. 5 Marks

c Compare overfitting with under-fitting. 5 Marks

a How to Choose the Right Number of Clusters in k-means clustering? Explain any one method. 5 Marks

b Compare "Classification algorithms" with "Clustering algorithm". 5 Marks

c Explain with examples: Scalars, Vectors, Matrices, Tensors and Gradients in Linear Algebra.. 10 Marks

SECTION – IV

a N-grams are a type of linguistic model used in natural language processing (NLP) and computational linguistics. 10 Marks

Consider the given sentence:

"Artificial Intelligence is a branch of computer science that focuses on creating intelligent machines capable of performing tasks that typically require human intelligence."

i. Generate bi-grams for the above sentence ii. Generate tri-grams for the above sentence

b Explain how data exploration, pre-processing of data and splitting of data are performed on datasets. 10 Marks

a With examples demonstrate Stemming and Lemmatization normalization techniques. 10 Marks

b. Explain any 2 techniques of cross validation used in Machine Learning. 5 Marks

c. Brief explain different stages involved in the Machine Learning Operations (MLOps) lifecycle. 5 Marks

SECTION – V

a Demonstrate simple linear regression considering a dataset that has two variables: "Marks" (dependent variable) and "Hours of study" (independent variable) 10 Marks

b With a neat diagram explain components of Docker. 10 Marks

c.a Summarize any two cloud deployment models . 10 Marks

c.b Discuss any five ethical challenges in AI. 10 Marks

- 4.a Referring to the number of variables or features in a dataset and the focus of analysis. Explain univariate & multivariate data types with examples. 10 Marks

SECTION - III

- 5.a A Machine learning model was built to classify spam emails as "spam"(1) or "not spam"(0). The confusion matrix for the model is as shown below. Evaluate accuracy, precision, recall, specificity and F1-Score. 10 Marks

		Actual	
		1	0
Predicted	1	140	10
	0	5	50

- 5.b Explain Supervised and Unsupervised learning with examples. 5 Marks
- 5.c Compare overfitting with under-fitting. 5 Marks
- 6.a How to Choose the Right Number of Clusters in k-means clustering? Explain any one method. 5 Marks
- 6.b Compare "Classification algorithms" with "Clustering algorithm". 5 Marks
- 6.c Explain with examples: Scalars, Vectors, Matrices, Tensors and Gradients in Linear Algebra.. 10 Marks

SECTION - IV

- 7.a N-grams are a type of linguistic model used in natural language processing (NLP) and computational linguistics. 10 Marks
- Consider the given sentence:
"Artificial Intelligence is a branch of computer science that focuses on creating intelligent machines capable of performing tasks that typically require human intelligence."
- i. Generate bi-grams for the above sentence ii. Generate tri-grams for the above sentence
- 7.b Explain how data exploration, pre-processing of data and splitting of data are performed on datasets. 10 Marks
- 8.a With examples demonstrate Stemming and Lemmatization normalization techniques. 10 Marks
- 8.b. Explain any 2 techniques of cross validation used in Machine Learning. 5 Marks
- 8.c. Brief explain different stages involved in the Machine Learning Operations (MLOps) lifecycle. 5 Marks

SECTION - V

- 9.a Demonstrate simple linear regression considering a dataset that has two variables: "Marks" (dependent variable) and "Hours of study" (independent variable) 10 Marks
- 9.b With a neat diagram explain components of Docker. 10 Marks
- 10.a Summarize any two cloud deployment models . 10 Marks
- 10.b Discuss any five ethical challenges in AI. 10 Marks

1074

8

Code : 20CS51I

Register
Number

--	--	--	--	--	--	--	--	--	--

V Semester Diploma Examination, June/July-2023

ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

Duration : 3 Hours]

[Max. Marks : 100

Instructions : Answer **one** full question from each Section.

SECTION – I

1. (a) Artificial Intelligence (AI) is a promising state-of-the-art technology that provides intelligent solutions in every field today. Justify your answer by describing AI and its applications in various fields. **10**
(b) Write steps to create repository in GitHub and add file. **10**

2. (a) For the following scenarios you are required to build a predictive model. Which machine learning technique/algorithm can be applied/best suited for stated problems ? **10**
Justify your recommendations.
 - (i) Predicting the food delivery time.
 - (ii) Predicting whether the transaction is fraudulent.
 - (iii) Predicting the credit limit of a credit card application.
 - (iv) Predicting natural disaster.
(b) How is AI software development life cycle different from traditional software development ? Explain. **10**

SECTION – II

3. (a) How to handle missing values in the data set ? Explain. **10**
(b) Perform the following operations on car manufacturing company dataset auto-mpg.csv. Write code given below in Pand/numpy. **10**
(i) Read data from auto-mpg.csv



- (ii) Give code to get all cars with 8 cylinders.
 (iii) Get the number of Cars manufactured in each year.

	mpg	Cylinder	Displace- ment	HP	Weight	Accele- ration	Model year	Origin	Car- name
0	18	4	300	130	3504	12.0	70	1	Honda
1	15	8	325	165	3695	11.5	71	1	Nexon
2	18	8	318	160	3440	11.0	70	1	Ford
3	16	6	305	160	3450	12.0	75	1	Indica
4	17	8	307	150	3449	10.5	80	1	Swift

4. (a) Explain univariate & multivariate data types with examples. 10
 (b) Create two series as shown using pd. series() function.

Series – A = [10, 20, 30, 40, 50]

Series – B = [40, 50, 60, 70, 80]

- (i) Get the items not common to both.
 (ii) Identify the smallest and largest element in the Series A.
 (iii) Find the sum of Series B.
 (iv) Calculate average in the Series A.
 (v) Find median in the given Series B.

10

SECTION – III

5. (a) Assume that Iris dataset is given and write the code. 10
 (i) Print first 5 records.
 (ii) Print the size of the data for given dataset.
 (iii) Use Scatter plot to compare petal length and petal width.
 (iv) Check for missing values.
 (v) Print the summary of the dataset. 10
- (b) Explain supervised and unsupervised learning with examples. 10

6. (a) A dataset is given to you for creating a machine learning model. What are the steps followed before using the data for training the model ? Elaborate each followed step. 10
- (b) For the given dataset perform the following operations : 10
- Check Statistical info. of the data set.
 - Plot a line chart/plot showing total profit on y-axis and number column on x-axis.
 - Find the missing values.
 - Find the sum of total profit.
 - Find the max value from drawing sheets column.

Number	Pencil	Text Books	Drawing Sheets	Total Units	Profits
1	300	250	100	800	8000
2	350	350	200	1000	9500
3	400	400	200	1320	10256
4	500	420	250	1510	12000
5	520	500	300	2000	18000

SECTION - IV

7. (a) The confusion matrix for a machine learning model is given below. Evaluate. 10

		Actual	
		1	0
Predicted	1	45	8
	0	15	32

- Accuracy
- Precision
- Recall
- Specificity
- F1-Score

[Turn over

20CS51I

3 of 4

6. (a) A dataset is given to you for creating a machine learning model. What are the steps followed before using the data for training the model ? Elaborate each followed step. 10
- (b) For the given dataset perform the following operations : 10
- Check Statistical info. of the data set.
 - Plot a line chart/plot showing total profit on y-axis and number column on x-axis.
 - Find the missing values.
 - Find the sum of total profit.
 - Find the max value from drawing sheets column.

Number	Pencil	Text Books	Drawing Sheets	Total Units	Profits
1	300	250	100	800	8000
2	350	350	200	1000	9500
3	400	400	200	1320	10256
4	500	420	250	1510	12000
5	520	500	300	2000	18000

SECTION – IV

7. (a) The confusion matrix for a machine learning model is given below. Evaluate. 10

		Actual	
		1	0
Predicted	1	45	8
	0	15	32

- Accuracy
- Precision
- Recall
- Specificity
- F1-Score

[Turn over

- (b) (i) List and briefly explain various activation function in Neural Networks. 5
(ii) Explain neural network Architecture. 5
8. (a) Cluster the following eight points (with (x, y) representing locations) into three clusters : A1(2, 10); A2 (2, 5), A3 (8, 4) A4 (5, 8), A5 (7, 5), A6 (6, 4), A7 (1, 2), A8 (4, 9).
Initial cluster centers are :
A1 (2, 10), A4 (5, 8) and A7 (1, 2).
The distance function between two points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined on :
- $$p(a, b) = |x_2 - x_1| + |y_2 - y_1|$$
- use K-means Algorithm to find the three clusters after the first iteration. 10
- (b) Explain different basic and advanced ensemble learning techniques. 10
- SECTION – V**
9. (a) N-grams are defined as combination of N keywords together. Consider the given sentence :
“Machine Learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions relying on patterns and inference instead. It is seen as a subject of artificial intelligence”.
(i) Generate unigrams for the above sentence. 5
(ii) Generate trigrams for the above sentence. 5
(b) Summarize public and private cloud deployment models. 10
10. (a) What are ethics in AI and why ethical practices should be followed while developing solutions using AI ? 10
(b) Demonstrate simple linear regression considering a dataset that has two variables :
Salary (dependent variable) and experience (independent variable) 10

- (b) (i) List and briefly explain various activation function in Neural Networks. 5
(ii) Explain neural network Architecture. 5
8. (a) Cluster the following eight points (with (x, y) representing locations) into three clusters : A1(2, 10); A2 (2, 5), A3 (8, 4) A4 (5, 8), A5 (7, 5), A6 (6, 4), A7 (1, 2), A8 (4, 9).
Initial cluster centers are :
A1 (2, 10), A4 (5, 8) and A7 (1, 2).
The distance function between two points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined on :
 $p(a, b) = |x_2 - x_1| + |y_2 - y_1|$
use K-means Algorithm to find the three clusters after the first iteration.
- (b) Explain different basic and advanced ensemble learning techniques. 10

SECTION – V

9. (a) N-grams are defined as combination of N keywords together. Consider the given sentence :
“Machine Learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions relying on patterns and inference instead. It is seen as a subject of artificial intelligence”.
(i) Generate unigrams for the above sentence. 5
(ii) Generate trigrams for the above sentence. 5
(b) Summarize public and private cloud deployment models. 10
10. (a) What are ethics in AI and why ethical practices should be followed while developing solutions using AI ? 10
(b) Demonstrate simple linear regression considering a dataset that has two variables :
Salary (dependent variable) and experience (independent variable) 10