Section-I

1.a what do you understand by the Artificial intelligence ? Describe how AI is finding its place in every sector? 10M

1.b How AI software development differs from traditional software development 5M

1.c what is Big-data? Brief out the four V's of Big data 5M

2.a. Differentiate between Supervised and Unsupervised machine learning algorithms . 5M

2.b Discuss how machine learning is changing travel industry today. 5M

2.c For the following scenario you are required to build machine learning model. Which machine learning algorithm can be applied for stated problem. 10M

a) Predicting the loan amount limit for the loan applicant.
b) Predicting cricket scores.
c) Predicting whether the transaction is fraudulent or not
d) Predicting the wine quality.

Section-II

3.a consider the given tip dataset and perform the following operation .write code given below in NumPy/pandas. 10M

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

i. find out the sum of the total_bill of each day
ii. Create a list containing tip values greater than '5' using 'filter' function
iii. Find the sum of the tip using 'reduce' function

3.b perform the following operations on the iris dataset 10M

i. read the data from an existing file
ii. statistical summary of the dataset
iii. scatter plot to analyse "sepal length" and " sepal width"
iv. check for missing values
v. replace the missing values with Mean

4.a How to handle missing values in the dataset 10M

4.b Accuracy and performance of the machine learning algorithm is depending on the quality of the data. Exploratory Data analysis helps in getting insights of dataset and data pre-processing is required task to clean and prepare the dataset for learning model. 10M

consider the dataset of lung cancer and the columns of the dataset are as follows

	Name	Surname	Age	Smokes	AreaQ	Alkhol	Result	Gender	location
0	John	Wick	35	3	5	4.0	1	F	United states
1	John	Constantine	27	20	2	5.0	1	M	florida
2	Camela	Anderson	30	0	5	2.0	0	M	canada

i. Understanding of dataset using info and describe functions.
ii. Perform univariate, bi-variate and multivariate analysis on the data.
iii. Find the outlier in the dataset using visualization technique and remove the outlier using IQR.

Section-III

5.a The confusion matrix for a machine learning model is given below. Evaluate the following 10M

i. Accuracy
ii. Precision
iii. Recall
iv. Specificity
v. F1.-Score

		Actual	
		1	0
Predicted	1	91	22
	0	24	139

5.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. 10M

6.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 centres are: A1(2, 10), A4(5, 8) and A7(1, 2). The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as P(a, b) = |x2 – x1| + |y2 – y1|.

Use K-Means Algorithm to find the three cluster centres after the first iteration 10M

6.b Demonstrate the use of activation function used in neural networks. List and explain activation functions in detail. 10M

Section-IV

7.a (i)Discuss importance of dimensionality reduction in machine learning. 5M

(ii) Compare overfitting with underfitting 5M

7.b Build DecisionTreeClassifier model for iris classification dataset and evaluate using classification metrics 10M

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

8.a (i) Compare classification algorithms with clustering algorithm. 5M

(ii) what are ensemble techniques ? explain advanced ensemble techniques in detail? 5M

8.b N-grams are defined as the combination of N keywords together. Consider the given sentence: “life is like riding bicycle.to keep balance, you must keep moving. If you are moving forward together, then success takes care of itself” 10M

a. Generate bi grams for the above sentence

b. Generate tri-grams for the above sentence

Section-V

9.a what is Cross validation ? Discuss different techniques of cross validation 10M

9.b With a neat diagram explain components of Docker . 10M

10.a Build simple linear regression machine learning model.Consider the dataset having two variables CIE(dependent variable) and SEE(target variable) and evaluate the model using performance metrics. 10M

10.b Summarize any two cloud deployment models. 10M