Q.6		Attempt any two:					
	i.	What are the advantages and disadvantages of using NoSQL databases?	**5**	3	5	3	3
	ii.	What is scikit-learn? Why is it popular in the field of machine learning?	**5**	4	5	3	4
	iii.	What are the different types of plots available in Matplotlib? Provide examples and typical use cases for at least two types.	**5**	5	5	3	3

******

Total No. of Questions: 6 *Total No. of Printed Pages:4*

Enrollment No.....

Faculty of Engineering / Science
End Sem Examination Dec 2024
CS3EL13 / BC3EL07 Data Science

Programme: B.Tech. Branch/Specialisation: CSE All / Computer Science

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

			Marks	**BL**	**CO**	**PO**	**PSO**
Q.1	i.	Machine learning is a subset of which of the following- (a) Artificial intelligence (b) Deep learning (c) Data learning (d) None of these	**1**	1	1	1	1
	ii.	Which of the following best describes the primary goal of data science? (a) To create databases for storing large amounts of data (b) To extract meaningful insights from data using various techniques (c) To build physical data warehouses (d) To write algorithms for web development	**1**	2	1	1	2
	iii.	What are the primary components of a data science workflow? (a) Data collection, data cleaning, data storage, data mining (b) Data collection, data cleaning, exploratory Data analysis, modeling, interpretation (c) Data entry, data analysis, data warehousing, data reporting (d) Data input, data management, data processing, data visualization	**1**	2	2	1	1

iv.	Which of the following is NOT a characteristic of the Normal distribution? (a) The distribution is symmetric about the mean (b) The mean, median, and mode are all equal (c) The shape of the distribution depends on the skewness of the data (d) The total area under the curve is equal to 1	**1**	3	2	2	2
v.	What is a Markov chain? (a) A type of linear regression model (b) A sequence of random variables where the future state depends only on the current state (c) A graph-based model for classification (d) A clustering algorithm for unsupervised learning	**1**	3	3	2	1
vi.	What is scikit-learn primarily used for? (a) Data visualization (b) Data manipulation (c) Machine learning (d) Database management	**1**	3	3	3	2
vii.	Which of the following is an example of a discrete random variable? (a) The height of students in a classroom (b) The time taken to run a marathon (c) The number of cars passing through a toll booth in an hour (d) The temperature in a city	**1**	3	4	3	3
viii.	What type of chart is best used for displaying the relationship between two continuous variables? (a) Box plot (b) Line chart (c) Scatter plot (d) Heatmap	**1**	3	4	1	3
ix.	Which function in Matplotlib is used to create a basic plot? (a) plt.plot() (b) plt.show() (c) plt.figure() (d) plt.bar()	**1**	3	5	2	3
x.	Which NoSQL database model is optimized for handling large volumes of unstructured data? (a) Column-family store (b) Document store (c) Key-value store (d) Graph database	**1**	3	5	3	3

Q.2	i.	What is data science?	**2**	1	1	1	1
	ii.	Differentiate between data analytics and data science.	**3**	1	1	2	1
	iii.	What is machine learning? Explain supervise learning.	**5**	1	1	2	2
OR	iv.	What are some common techniques for handling missing data in EDA?	**5**	1	1	2	
Q.3	i.	What are the key properties of the normal distribution?	**2**	2	2	2	1
	ii.	Define a random variable. Explain the difference between discrete and continuous random variables with examples.	**8**	2	2	1	2
OR	iii.	Define the normal distribution. Describe its probability density function (PDF).	**8**	3	2	2	2
Q.4	i.	Define Exploratory Data Analysis (EDA). Explain its importance in the data science workflow.	**3**	2	3	1	2
	ii.	Explain the concept of outliers. How can they be detected and treated in EDA?	**7**	3	3	1	2
OR	iii.	How can correlation analysis be useful in EDA? Describe how to interpret a correlation matrix.	**7**	3	3	2	3
Q.5	i.	Describe the different types of data visualization techniques. Provide examples of when to use each type.	**4**	3	4	3	3
	ii.	Compare and contrast bar charts and histograms. When should each be used.	**6**	3	4	2	3
OR	iii.	What are scatter plots used for in data analysis? Explain how they can reveal relationships between variables.	**6**	4	4	2	4

Marking Scheme
CS3EL13(T) Data Science

Q.1	i)	a. Artificial intelligence	**1**
	ii)	b. To extract meaningful insights from data using various techniques	**1**
	iii)	b. Data Collection, Data Cleaning, Exploratory Data Analysis, Modeling, Interpretation	**1**
	iv)	c. The shape of the distribution depends on the skewness of the data.	**1**
	v)	b. A sequence of random variables where the future state depends only on the current state	**1**
	vi)	c. Machine learning	**1**
	vii)	c. The number of cars passing through a toll booth in an hour	**1**
	viii)	c. Scatter Plot	**1**
	ix)	a. plt.plot()	**1**
	x)	b. Document Store	**1**
Q.2	i.	What is Data Science? Definition - 2 marks	**2**
	ii.	Differentiate between Data Analytics and Data Science Data Analytics and Data Science 1 mark per difference	**3**
	iii.	What is Machine learning? Explain supervise learning. Defination of ML 2 marks supervise learning 3 marks	**5**
OR	iv.	What are some common techniques for handling missing data in EDA? 5 marks	**5**
Q.3	i.	What are the key properties of the normal distribution? Three Properties- 2 marks	**2**
	ii.	Define a random variable- 4 marks Explain the difference between discrete and continuous random variables with examples-	**8**
OR	iii.	Define the normal distribution and describe its probability density function (PDF). Definition- 4 marks probability density function (PDF)- 4 marks	**8**
Q.4	i.	Define Exploratory Data Analysis (EDA) and explain its importance in the data science workflow. Definition 2 marks Importance- 1 mark	**3**
	ii.	Explain the concept of outliers. How can they be detected and treated in EDA? Definition of outlier- 3 marks How can they be detected and treated in EDA? 4 marks	**7**
OR	iii.	How can correlation analysis be useful in EDA? 3 marks Describe how to interpret a correlation matrix. 4 marks	**7**
Q.5	i.	Describe the different types of data visualization techniques. Provide examples of when to use each type. 4 marks	**4**
	ii.	Compare and contrast bar charts and histograms. When should each be used? Compare and contrast bar charts and histograms- 3 marks When should each be used? - 3 marks	**6**
OR	iii.	What are scatter plots used for in data analysis? Explain how they can reveal relationships between variables. What are scatter plots used for in data analysis? 3 marks Explain how they can reveal relationships between variables. 3 marks	**6**

Q.6 i. What are the advantages and disadvantages of using NoSQL databases? **5**

Adantages-2.5 marks

Disadvantages-2.5 marks

ii What is scikit-learn and why is it popular in the field of machine learning? **5**

Defination-2 marks

Advantages and functions- 3 marks

iii What are the different types of plots available in Matplotlib? Provide examples and typical use cases for at least two types. **5**

What are the different types of plots available in Matplotlib? 3 marks

Provide examples and typical use cases for at least two types. 2 marks

******