Enrollment No......................................

Faculty of Engineering
End Sem Examination Dec 2024
IT3ED07 Data Science

Programme: B.Tech.   Branch/Specialisation: IT

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	PO	CO	PSO
Q.1	i.	The different types of machine learning are-	1	01	01	01	
		(a) Supervised					
		(b) Unsupervised					
		(c) Reinforcement					
		(d) All of these					
	ii.	Identify the key data science skills among the following-	1	01	01	01	
		(a) Data visualization					
		(b) Machine learning					
		(c) Statistics					
		(d) All of these					
	iii.	What will be the probability of an impossible event?	1	03	02	02	
		(a) 0 (b) 1					
		(c) Infinity (d) None of these					
	iv.	A card is drawn from a pack of 52 cards. What is the probability of getting a queen card?	1	03	02	02	
		(a) 1/26 (b) 1/52					
		(c) 3/13 (d) 1/13					
	v.	What is the purpose of a correlation matrix in EDA?	1	01	01,3	01	
		(a) To visualize the distribution of data					
		(b) To display relationships between variables					
		(c) To identify missing values					
		(d) None of these					

P.T.O.

vi.	Which of the following gave rise to need of graphs in data analysis? (a) Data visualization (b) Communicating results (c) Decision making (d) All of these	**1**	01	01,3	01,02	
vii.	The most popular data visualization library in python is- (a) matinfolib (b) matplotlib (c) pip (d) matpiplib	**1**	01	01,4	02	
viii.	Which of the following is NOT a common visual element used in data visualization? (a) Bar chart (b) Line chart (c) Data table (d) Pie chart	**1**	01	02,4	01,02	
ix.	Which of the following is NOT a type of NoSQL database? (a) Document (b) Column-family (c) Relational (d) Key-value	**1**	01	02,5	02	
x.	Which function in Matplotlib is used to create a simple plot? (a) plot.show() (b) plt.plot() (c) plt.display() (d) plt.graph()	**1**	01	02,5	01,02	

Q.2	i.	Differentiate data mining and data science.	**2**	02	01	02
	ii.	Discuss big data with example.	**3**	03	01	02
	iii.	Explain different data science techniques.	**5**	02	01	02
OR	iv.	What are the different main components of data science?	**5**	02	01	02

Q.3	i.	Explain conditional probability and random variables with examples.	**4**	02	02	02
	ii.	Discuss normal distribution and its characteristics.	**6**	02	02	02
OR	iii.	Discuss Marcov chains in details with its applications.	**6**	02	02	02

Q.4	i.	Discuss EDA tools.	**3**	01	03	02,12
	ii.	How does exploratory data analysis works?	**7**	02	03	02,12
OR	iii.	Discuss data science process through case study.	**7**	02	03	02,12

Q.5		Attempt any two:				
	i.	What are basic tools for data visualization?	**5**	01	04	02,12
	ii.	Discuss basic principles of data visualization.	**5**	02	04	02,12
	iii.	Write down python code for bar chart and line chart for visualization.	**5**	02	04	02,12

Q.6		Attempt any two:				
	i.	Discuss python libraries SciPy and sci-kitlearn with utility.	**5**	01	05	02,12
	ii.	Discuss CAP theorem.	**5**	01	05	02,12
	iii.	Explain different types of No-sql databases.	**5**	01	05	02,12

*****

Q.1 i)   **(d) All of the above**   **1**

    ii)   **(d) all of the above**   **1**

    iii)   **(a) 0**   **1**

    iv)   **(d) 1/13**   **1**

    v)   **(b)To display relationships between variables**   **1**

    vi)   **(d) All of the mentioned**   **1**

    vii)   **(b) matplotlib**   **1**

    viii)   **(c) Data table**   **1**

    ix)   **(c) Relational**   **1**

    x)   **(b) plt.plot()**   **1**

Q.2 i.   Differentiate data mining and data science.   **2**
        1 mark for each difference

    ii.   Discuss big data with example.   **3**
        Definition 1.5
        Example 1.5

    iii.   Explain different data science techniques.   **5**
        1 mark for each

OR iv.   What are difference main components of data science.   **5**
        1 mark for each

Q.3 i.   Explain conditional probability and random variables with examples.   **4**
        2 marks for ech

    ii.   Discuss normal distribution and its characteristics.   **6**
        Definition3
        Char. 3

OR iii.   Discuss Marcov chains in details with its applications.   **6**
        Definition 3
        Applications 3

Q.4 i.   Discuss EDA tools.   **3**
        1 tool - 1 mark

    ii.   How does exploratory data analysis works?   **7**

OR iii.   Discuss data science process through case study.   **7**

Q.5 i.   Attempt any two:

    ii.   What are basic tools for data visualization?   **5**
        Each tool – 1 mark

OR iii.   Discuss basic principles of data visualization.   **5**
        0.5 mark for 1 principle
        Write down python code for bar chart and line chart for visualization.   **5**
        Bar chart – 2.5 marks
        Line chart – 2.5 marks

Q.6

    i.   Attempt any two:

    ii.   SciPy -2.5 marks   **5**
        sci-kitlearn -2.5 marks

    iii.   CAP theorem -2 marks   **5**
        Explanation C,A,P -1 marks each

    iv.   Explanation No-sql- 1 mark   **5**
        Types of No-sql databases -1 marks for each = 4 marks

*****