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

Faculty of Engineering

End Sem (Even) Examination May-2022

OE00072 Introduction to Data Science

Programme: B.Tech.   Branch/Specialisation: All

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.

Q.1 i.   _____is not one of the key data science skills.   **1**
      (a) Statistics   (b) Machine learning
      (c) Data visualization   (d) Software tester

   ii.   Which of the following is the correct statement?   **1**
      (a) Preprocessed data is original source of data
      (b) Raw data is original source of data
      (c) Raw data is the data obtained after processing steps
      (d) None of these

   iii. Let a pair of fair dice be thrown. What is the probability that the sum   **1**
      of scores will be 8?
      (a) 4/36   (b) 5/36   (c) 1/36   (d) 6/36

   iv.   Which is the first step of EDA?   **1**
      (a) Data wrangling   (b) Data visualization
      (c) Both (a) & (b)   (d) None of these

   v.   ML is a field of AI consisting of learning algorithms that?   **1**
      (a) Improve their performance
      (b) At executing some task
      (c) Over time with experience
      (d) All of these

   vi.   Which of the following is a widely used and effective machine   **1**
      learning algorithm based on the idea of bagging?
      (a) Decision tree   (b) Random forest
      (c) Regression   (d) Classification

                                                                    P.T.O.

vii. _____ is basic data structure of pandas can be think of SQL table or a spreadsheet data representation. **1**

    (a) Dataframe   (b) Series   (c) List   (d) Ndarray

viii. From matplotlib_____ module is used for plotting various plots. **1**

    (a) Scilearn   (b) Pyplot   (c) Scilab   (d) Matlab

ix. What is the work of Data Architect? **1**

    (a) Utilize large data sets to gather information that meets their company's needs.

    (b) Work with businesses to determine the best usage of the information yielded from data.

    (c) Build data solutions that are optimized for performance and design applications.

    (d) All of these

x. Which of the following are the Data Sources in data science? **1**

    (a) Structured   (b) UnStructured

    (c) Both (a) and (b)   (d) None of these

Q.2 i. Define Data Science with an example. **2**

    ii. Explain basic framework of data science. **3**

    iii. How many skills of data scientist may be observed? What are the role of data analyst and machine learning expert in data science? Explain it. **5**

OR iv. How data science is important in today's business world? **5**

Q.3 i. Explain sample spaces and events. **2**

    ii. How the conditional probability in statistics can be calculated? Explain with an example. **3**

    iii. Give the difference between descriptive and predictive statistics analysis. Also give name of primary tools used in descriptive and predictive analytics. **5**

OR iv. Define EDA. What are the main steps should be used while doing EDA? Explain with an example. **5**

Q.4 i. Define machine learning with an example. **2**

    ii. Explain feature selection and feature extraction method for dimensionality reduction. **3**

    iii. Distinguish between supervised learning and reinforcement learning. Illustrate with an example. **5**

OR iv. List out any five applications of machine learning. **5**

Q.5 i. What are the primary components of Data Science? Explain in detail. **4**

    ii. What are the document oriented NoSQL and key-value pair NoSQL? Also explain how they are differing than SQL. **6**

OR iii. How many python libraries may be used for data science project? Also write their usage in different stage of data science life cycle. **6**

Q.6   Attempt any two:

    i. What are the challenges from which a data scientist has to go through in any data science project? Also give the solutions for them. **5**

    ii. State the five different stages involved in a data science project. **5**

    iii. Explain the flow of data in any data science project. Also explain the life cycle of a data science project. **5**

******

Marking Scheme - OE00072 Introduction to Data Science

Q.1 i. _____is not one of the key data science skills. **1**
(d) Software tester

ii. Which of the following is the correct statement? **1**
(b) Raw data is original source of data

iii. Let a pair of fair dice be thrown. What is the probability that the sum **1** of scores will be 8?
(b) 5/36

iv. Which is the first step of EDA? **1**
(a) Data wrangling

v. ML is a field of AI consisting of learning algorithms that? **1**
(d) All of these

vi. Which of the following is a widely used and effective machine **1** learning algorithm based on the idea of bagging?
(b) Random forest

vii. _____ is basic data structure of pandas can be think of SQL table or **1** a spreadsheet data representation.
(a) Dataframe

viii. From matplotlib_____ module is used for plotting various **1** plots.
(b) Pyplot

ix. What is the work of Data Architect? **1**
(c) Build data solutions that are optimized for performance and design applications.

x. Which of the following are the Data Sources in data science? **1**
(c) Both (a) and (b)

Q.2	i.	Define Data Science	1 mark	**2**
		An example	1 mark	
	ii.	Explanation	2 marks	**3**
		Block Diagram	1 mark	
	iii.	Skills of data scientist	2 marks	**5**
		Role of data analyst	1.5 marks	
		Role of machine learning	1.5 marks	
OR	iv.	Any five points one mark for each	(1*5) marks	**5**

Q.3	i.	Sample spaces	1 mark	**2**
		Events	1 mark	
	ii.	Explanation	1 mark	**3**
		Formula	1 mark	
		Example	1 mark	
	iii.	Any 3 difference	(1*3) marks	**5**
		Primary tools used in descriptive	2 marks	
OR	iv.	Define EDA	2 marks	**5**
		Main steps	2 marks	
		Explain with an example.	1 mark	

Q.4	i.	Define machine learning	1 mark	**2**
		Example	1 mark	
	ii.	Feature selection	1.5 marks	**3**
		Feature extraction	1.5 marks	
	iii.	Any 3 distinguish points	(1*3) marks	**5**
		Any 2 example	(1*2) marks	
OR	iv.	Any five applications one mark	(1*5) marks	**5**

Q.5	i.	Any 4 primary components	(1*4) marks	**4**
	ii.	Document oriented NoSQL	1.5 marks	**6**
		Key-value pair NoSQL	1.5 marks	
		Any 3 difference	(1*3) marks	
OR	iii.	Python libraries name	2 marks	**6**
		Any 4 Usage in different stage	(1*4) marks	

Q.6		Attempt any two:		
	i.	Challenges	2.5 marks	**5**
		Solutions for	2.5 marks	
	ii.	For each stages involved	(1*5) marks	**5**
	iii.	Explanation	3 marks	**5**
		Life cycle of a data science	2 marks	

******