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

Faculty of Engineering
End Sem Examination May-2023
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. Assume suitable data if necessary. Notations and symbols have their usual meaning.

Q.1 i. Which of the following is correct skills for a Data Scientist? 1
        (a) Probability & Statistics   (b) Machine Learning / Deep Learning
        (c) Data Wrangling   (d) All of these

    ii. Which of the following are correct component for data science? 1
        (a) Data Engineering   (b) Advanced Computing
        (c) Domain expertise   (d) All of these

   iii. _____ answers the question "What has happened"? 1
        (a) Descriptive analytics   (b) Predictive analytics
        (c) Prescriptive analytics   (d) None of these

    iv. _____ are used when we want to visually examine the relationship 1
         between two quantitative variables.
        (a) Bar graph   (b) Scatterplot (c) Line graph (d) Pie chart

     v. What is true about Machine Learning? 1
        (a) Machine Learning (ML) is that field of computer science
        (b) ML is a type of artificial intelligence that extract patterns out of
            raw data by using an algorithm or method
        (c) The main focus of ML is to allow computer systems learn from
            experience without being explicitly programmed or human
            intervention.
        (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) Regression
        (c) Classification   (d) Random Forest

P.T.O.

vii. _____ is most important language for Data Science. **1**
(a) Java   (b) Ruby   (c) R   (d) None of these

viii. _____ phase of the data analytics lifecycle usually takes the longest **1** time.
(a) Phase 2: Data Preparation
(b) Phase 3: Model Planning
(c) Phase 4: Model Building
(d) Phase 5: Communicate Results

ix. _____ of the following is not a step-in data analysis. **1**
(a) Obtain the data   (b) Clean the data
(c) EDA   (d) None of these

x. 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

Q.2   i.   Explain Data Science with example.   **2**
     ii.   Differentiate between data analytics and data science.   **3**
     iii.   Explain basic framework of data science.   **5**
OR   iv.   What are the basic expertise required for data science?   **5**

Q.3   i.   Explain basic principles of data visualization.   **2**
     ii.   (a) How does exploratory data analysis differ from summary analysis?   **8**
         (b) What are the Descriptive, Predictive and Prescriptive statistics?
OR   iii.   Explain different EDA tool. Also compare them.   **8**

Q.4   i.   Differentiate between Data Science and Machine Learning.   **3**
     ii.   What is reinforcement learning? Explain in detailed concepts.   **7**
OR   iii.   What are the basic design issues and approaches to machine learning?   **7**

Q.5   i.   What are the primary components of Data Science?   **4**
     ii.   Explain Nosql and its features.   **6**
OR   iii.   Explain the role of data warehousing in data science.   **6**

Q.6   Attempt any two:
    i.   Explain the execution flow of data science project management.   **5**
    ii.   Explain users of Data Science and its hierarchy also give the overview   **5** of different Data Science techniques.
    iii.   What are the challenges and the scope of data science project   **5** management?

*****

Marking Scheme
OE00072 Introduction to Data Science

Q.1 i) c) Data Wrangling   1

ii) d) All of the above   1

iii) a)Descriptive analytics   1

iv) a)Bar graph   1

v) d)All of the above   1

vi) d) Random Forest   1

vii) c)R   1

viii) a)Phase 2: Data Preparation   1

ix) d)None of the above   1

x) c)build data solutions that are optimized for performance and design   applications.   1

Q.2 i. Explain Data Science with example?   2

Definition   -1 mark

Example   -1 mark

ii. Differentiate between data Analytics and data science?   3

Each point   -1 mark

iii. Explain basic framework of data science.   5

Explanation   3 marks

Diagram   2 marks

OR iv. What are the basic expertise required for data science?   5

Each point   -1 mark

Q.3 i. Explain basic principles of data visualization.   2

ii. a)How does exploratory data analysis differ from summary analysis?   -4 marks   8

b)What are the Descriptive, Predictive and Prescriptive statistics   -4 marks

OR iii. Explain different EDA tool and also compare them   8

Q.4 i. Differentiate between Data Science and Machine Learning?   3

Each point   -1 mark

ii. What is reinforcement learning .Explain in detailed concepts?   7

Definition   -3 marks

Concepts (Example)   4 marks

OR iii. What are the basic design issues and approaches to machine learning?   7

Design issues   -4 marks

Approaches   -3 marks

Q.5 i. What are the primary components of Data Science?   4

Each component   -1 mark

ii. Explain Nosql and its features?   6

Definition   -2 marks

Features   -4 marks

OR iii. Explain the role of data warehousing in data science?   6

Q.6   Attempt any two:

i. Explain the execution flow of data science project management?   5

ii. Explain users of Data Science and its hierarchy also give the overview of different Data Science techniques.   5

Explain users of Data Science   -2.5 marks

its hierarchy   -2.5 marks

iii. What are the challenges and the scope of data science project management.   5

Challenges   -2.5 marks

Scope   -2.5 marks

******