

Duration: 3hrs

[Max Marks: 80]

- N.B. :**
- (1) Question No 1 is Compulsory.
 - (2) Attempt any three questions out of the remaining five.
 - (3) All questions carry equal marks
 - (4) Assume suitable data, if required and state it clearly

- Q.1** Attempt any **FOUR** [20]
- a** Explain the different tools and skills required in data science. [5]
 - b** How to define research goals and create a project charter in the data science process? [5]
 - c** Explain physical architecture of HDFS and the characteristics of Big Data. [5]
 - d** Explain the different sampling techniques with an example. [5]
 - e** Explain entropy and information gain with an example. [5]
- Q.2**
- a** Elaborate and explain all the steps of the Data Science Process. [10]
 - b** Compare business intelligence and data science. How to translate business challenges into data science challenges? [10]
- Q.3**
- a** Discuss the Hadoop Ecosystem in detail. Explain its limitations. [10]
 - b** Explain Linear Regression, Logistic Regression and Multinomial Logistic Regression. [10]
- Q.4**
- a** Explain Bayes Theorem and how Naive Bayes Algorithm works. [10]
 - b** Explain the applications of machine learning in data science. [10]
- Q.5**
- a** Explain ACID the core principles of Relational databases, CAP Theorem and the BASE Principles of NoSQL databases. [10]
 - b** Let there is a disease that is present in 2% of the population. The disease can be detected by a nasal swab, and the test is 95% accurate for both positive and negative cases. If Mr. A was tested for the disease and received a positive test result. Consider this case study to explain confusion matrix, True positive, True negative, False Positive, False Negative. Draw confusion matrix. What is the probability that Mr. A actually has the disease? [Assume a population of 1000 people.] [10]
- Q.6**
- a** Explain recommendation system and customer segmentation in detail. [10]
 - b** Write short note on any two : [10]
 - 1. NoSQL Database Types
 - 2. Sentiment analysis
 - 3. Data visualization using Tableau