Duration : 3 Hours   **[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.

1	Attempt any FOUR	**[20]**
a	Explain different Facets of data.	**5**
b	Explain different data transformation techniques with an example.	**5**
c	Differentiate between Linear regression and logistic regression.	**5**
d	Compare and contrast different types of NoSQL databases.	**5**
e	Define random sampling, systematic sampling and stratified sampling.	**5**

2 a What is probability? Explain different probability types? A new project **[10]** assignment order is received by an IT company and the authority wants to assign the projects according to the salary package. There are 60 employees without training and 45 employees who are trained with required professional courses. Now the task here is that you have to assign the job according to the salary package.

      i.   Find the probability of the employee that they have undergone professional training.
     ii.   Find the probability that an employee has attended professional training and also has a good salary package.
   iii.   Find the probability that an employee has a good salary package given that the employee has not undergone professional training.

Dataset for Employees				
Results		Training		Total
		Without Professional	With Professional	
Salary package obtained by employees	Poor salary	05	00	05
	Below-average	10	0	10
	Average salary	40	10	50
	Good salary	5	30	35
	Excellent salary	0	5	5
	Total	60	45	105

134B80B5B432F6F71E8805D5557DC0D0

b What is hadoop? Explain Hadoop ecosystem in detail. What are the Limitations of Hadoop.

3 a Elaborate and explain all the steps of Data Science Process. **[10]**

   b What is data science? Why is data science required? What is the role of data science in an autonomous car? **[10]**

4 A Explain linear regression algorithm. The average rainfall (in mm) for the last 5 years in the state of Maharashtra was recorded along with the umbrellas sold in that particular year. Calculate the Calculate the regression coefficients for the following dat. **[10]**

Rainfall in mm	121.2	152.6	98.4	171	85.6
Umbrella sold	52	72	40	100	34

   b What is confusion matrix? Explain TYPE I and TYPE II Error with an example. What is the difference between precision and recall? **[10]**

5 a Explain NoSQL databases. Explain CAP Theorem and BASE property of NoSQL. **[10]**

   b Explain Naïve Bayes Algorithm. Construct a Naïve Bayes classifier for the given data for the three attributes: long, sweet and yellow to predict what fruit is it. **[10]**

Type	Long	Not long	Sweet	Not sweet	Yellow	Not yellow	Total
Banana	400	100	350	150	450	50	500
Orange	0	300	150	150	300	0	300
Other	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000

6 a Explain in detail Customer Sentiment Analysis with a block diagram. **[10]**

   b Write short note on any two: **[10]**
      1. Customer Segmentation
      2. Recommendation System
      3. Visualization using Tableau

_____

134B80B5B432F6F71E8805D5557DC0D0