

POSSESSION OF MOBILES IN EXAM IS UFM PRACTICE.

Name _____

Enrollment No. _____

Jaypee Institute of Information Technology, Noida
B.Tech (5th Sem)

Course Title: Introduction to Big Data & Data Analytics **Maximum Time-1 hour**

Course Code: 20B12CS333

Maximum Marks: 20

| | |
|-----|---|
| CO1 | To demonstrate the fundamental concepts of growing field of big data analytics. |
| CO2 | To make use of tools required to manage and analyze big data like Hadoop, NoSql, MapReduce. |
| CO3 | To apply predictive model and advanced computing paradigms for big data analytics. |
| CO4 | To analyze the big data using intelligent & visualization techniques. |
| CO5 | To design and create predictive and mathematical model to solve complex real-worlds problems for decision making. |

Note: Attempt all questions.

Q1a. Imagine you are managing a manufacturing company that produces three different products. Discuss the use of optimization to determine the production quantity of each product to maximize profit while considering constraints such as raw materials availability, labour hours and production capacity.

Q1b. How would the problem change if you had to simultaneously decide on the production quantity, inventory levels and distribution logistics? Explain the application of simultaneous decision making.

[CO1(Remembering), 3+2=5 marks]

Q2. Consider the following table to treat the missing values (NaN) by the approach of:-

| Person | Weight | Height | Age |
|--------|--------|--------|-----|
| A | 70 | 175 | 28 |
| B | 65 | 160 | NaN |
| C | 50 | 155 | 22 |
| D | 80 | 180 | NaN |
| E | 58 | 165 | 25 |
| F | 68 | 170 | 24 |

- Mean Imputation and Median Imputation.
- KNN Imputation.

[CO3(Applying), 2+3=5 marks]

Q3. A small bakery wants to predict the daily sales based on the number of hours the bakery is open (X_1), and the number of items sold (X_2). The collected data is as follows:-

| Hours Open(X_1) | Items Sold(X_2) | Daily Sales(Y) |
|---------------------|---------------------|--------------------|
| 8 | 120 | 500 |
| 6 | 90 | 350 |
| 10 | 150 | 650 |
| 7 | 110 | 450 |
| 5 | 80 | 300 |

Compute a multiple linear regression model of the form:-

$$Y = \beta + \beta_0 \cdot X_1 + \beta_1 \cdot X_2$$

[CO4 (Analyzing), 5 marks]

Q4. The cost function for a linear regression model is given by:-

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (\theta_0 + \theta_1(x^i) - y^i)^2$$

- Find the derivative of $J(\theta_0, \theta_1)$ with respect to θ_0, θ_1 .
- Explain how these derivatives would be used in the gradient descent algorithm.
- Justify the statement "The cost function $J(\theta_0, \theta_1)$ is convex".

[CO4 (Analyzing), 2+2+1=5 marks]
