Invigilator's Signature and Date

Set-1

KIIT DEEMED TO BE UNIVERSITY

Spring End Semester Examination-2022

Roll No.	
Registration No.	
Name	
Date of Exam	

DATA ANALYTICS (IT3006)

6th Semester B.Tech (Open Elective-I / Minor-I)

SECTION-A

(Answer All Questions)

Time: 30 Minutes **Full Marks = 2 × 7 = 14 Marks**

Question No	Question	Write the correct option here.
Q.No:1	Consider 10 bit length bloom filter with 5 hash functions. What is the probability that slot 7 is hashed? a. 0.00001 b. 0.0001 c. 0.001 d. 0.01 e. 0.1	
Q.No:2	What is the correlation coefficient of below sample? Subject X Y 1 43 99 2 21 65 3 25 79 4 42 75 5 57 87 a. 0.636 b. 0.634 c. 0.631 d. 0.639 e. 0.681	
Q.No:3	Consider the transaction details where TID stands for transaction ID, G stands for grapes, A stands for apple, M stands for mango and O stands for orange. TID G A M O 1 1 1 1 1 2 1 0 1 1 3 0 0 1 1 4 0 1 0 0 5 1 1 1 1 6 1 1 0 1 What is support(G) and Confidence({G, A} => {M})? a. 0.666 and 0.667 b. 0.666 and 0.667 c. 0.666 and 0.666 c. 0.667 and 0.667 e. 1 and 1	

Q.No:4

What is Euclidean distance, Squared Euclidean distance and Manhattan distance between two data points p1(1,4) and p2(8,2)?

a. 7.28, 53, 9
b. 8.52, 64, 9
c. 6.51, 36, 7
d. 7.35, 50, 12
e. 1, 1, 1

Q.No:5

Consider the mark of the following students.

Roll No	Mark
1	10
2	9
3	8

What would be the values in the question mark places (i.e.,?1 and ?2) for the proximity matrix with the consideration of Euclidean distance?

Roll No	1	2	3
1	0	?1	
2		0	
3	?2		0

a. ?1 = 1 and ?2 = 2
b. ?1 = 2 and ?2 = 1
c. ?1 = 0 and ?2 = 0
d. ?1 = 1 and ?2 = 1
e. ?1 = 2 and ?2 = 2

Q.No:6

Consider 3 dataset i.e., A = {1, 5}, B = {9, 10} and C = 12, 15}. When the data to be plotted in Venn and Euler diagram, how many dataset would result into overlapping for each diagram?

a. 3, 0
b. 0, 3
c. 1, 3
d. 3, 1
e. 3, 3

Q.No:7

Consider three tables that capture details about analytics.

Table A

Row No	Analytics
1	Descriptive
2	Diagnostic
3	Predictive
4	Prescriptive

Table B

Row No	Human input to decision
1	Data
2	Insight
3	Decision

Which one of the following combination (i.e., Table A (Row No) – Table B (Row No) is correct?

a. Table A (1) - Table B (1)
b. Table A (2) - Table B (2)
c. Table A (3) - Table B (2)
d. Table A (4) - Table B (3)
e. All of the above are correct

****

Invigilator's Signature and Date

Set-1

KIIT DEEMED TO BE UNIVERSITY

Spring End Semester Examination-2022

DATA ANALYTICS (IT3006)

6th Semester B.Tech (Open Elective-I / Minor-I)

SECTION-B

(Answer Any Three Questions.)

Time: 1 Hour and 30 Minutes **Full Marks = 12 × 3 = 36 Marks**

Q.No:8 A. Consider the following dataset wherein YoE stands of Years of Experience (independent variable) and SAL represents salary (independent variable).

Row No	YoE	SAL
1	1	32323
2	1.1	45207
3	1.3	39751
4	2	43525
5	2.2	39891
6	2.7	56542
7	3	60150
8	3.2	54545
9	3.9	63218
10	4	55794
11	4.1	56081
12	4.2	57081

Demonstrate step-by-step procedure in establishing the linear regression model by determining the value of intercept, slope and random error. In addition, comment on the relation type between the variables.

B. Consider the sales dataset for a retailer which was observed from the year 2010 till 2021.

Year	Sales
2010	420.735
2011	392.943
2012	440.593
2013	450.037
2014	430.345
2015	471.033
2016	423.456
2017	458.989
2018	470.767
2019	420.368
2020	432.456
2021	487.409

Using Simple Exponential Smoothing, illustrate the step-by-step procedure to forecast the sale of 2022 and calculate Mean Square Error and Root Mean Square Error. Initialize the first forecast as the first observation and consider the smoothing constant as 0.255.