



ABES Engineering College, Ghaziabad

B. Tech Odd Semester Make-Up Test

Printed Pages: 2
Session: 2022-2023

Course Code: KDS501

Roll No:

Course Name: Introduction to Data Analytics and Visualization

Date of Exam: 27-Dec-2022

Maximum Marks:100

Time: 1:30-4:30

Instructions:

1. Attempt All sections.
2. If require any missing data, then choose suitably.

Q. No.	Question	Marks	CO	KL	PI																																														
Attempt All Questions		Total Marks: 10*10= 100																																																	
1a)	Cloud computing is a big shift from the traditional way businesses thinking about IT resources. Explain Public Cloud and Private cloud with their major five characteristics.	10	CO1	K2	2.2.3																																														
1b)	Data empowers to make decision informed, justify the statement and explain the various methods of primary and secondary data collection in detail.	10	CO1	K2	1.3.1																																														
2a)	<p>Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry. Solve this equation $P(S, D, A, \neg B, \neg E) = P(S A) * P(D A) * P(A \neg B \wedge \neg E) * P(\neg B) * P(\neg E)$ by probabilities given in below tables. Explain the Bayesian Belief network respectively.</p> <div><div><table><tr><td>T</td><td>0.002</td></tr><tr><td>F</td><td>0.998</td></tr></table></div><div><div>Burglary</div><div>B</div></div><div><div>Earthquake</div><div>E</div></div><div><table><tr><td>T</td><td>0.001</td></tr><tr><td>F</td><td>0.999</td></tr></table></div><div><div>Alarm</div><div>A</div></div><div><table><tr><td>B</td><td>E</td><td>P(A=T)</td><td>P(A=F)</td></tr><tr><td>T</td><td>T</td><td>0.94</td><td>0.06</td></tr><tr><td>T</td><td>F</td><td>0.95</td><td>0.04</td></tr><tr><td>F</td><td>T</td><td>0.69</td><td>0.69</td></tr><tr><td>F</td><td>F</td><td>0.999</td><td>0.999</td></tr></table></div><div><div>David Calls</div><div>D</div></div><div><div>Sophia calls</div><div>S</div></div><div><table><tr><td>A</td><td>P(D=T)</td><td>P(D=F)</td></tr><tr><td>T</td><td>0.91</td><td>0.09</td></tr><tr><td>F</td><td>0.05</td><td>0.95</td></tr></table></div><div><table><tr><td>A</td><td>P(S=T)</td><td>P(S=F)</td></tr><tr><td>T</td><td>0.75</td><td>0.25</td></tr><tr><td>F</td><td>0.02</td><td>0.98</td></tr></table></div></div> <td>10</td> <td>CO2</td> <td>K3</td> <td>1.3.1</td>	T	0.002	F	0.998	T	0.001	F	0.999	B	E	P(A=T)	P(A=F)	T	T	0.94	0.06	T	F	0.95	0.04	F	T	0.69	0.69	F	F	0.999	0.999	A	P(D=T)	P(D=F)	T	0.91	0.09	F	0.05	0.95	A	P(S=T)	P(S=F)	T	0.75	0.25	F	0.02	0.98	10	CO2	K3	1.3.1
T	0.002																																																		
F	0.998																																																		
T	0.001																																																		
F	0.999																																																		
B	E	P(A=T)	P(A=F)																																																
T	T	0.94	0.06																																																
T	F	0.95	0.04																																																
F	T	0.69	0.69																																																
F	F	0.999	0.999																																																
A	P(D=T)	P(D=F)																																																	
T	0.91	0.09																																																	
F	0.05	0.95																																																	
A	P(S=T)	P(S=F)																																																	
T	0.75	0.25																																																	
F	0.02	0.98																																																	

2b)	Write the steps involved in Principal Component Analysis (PCA). Consider the two dimensional patterns: (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8). Compute the principal component using PCA Algorithm.	10	CO2	K3	1.3.1																				
3a)	Elaborate the Decaying Window algorithm with graphical representation. Consider a sequence of Twitter tags below: {FIFA, IPL, FIFA, IPL, IPL, IPL, FIFA} Let each element weight is 1, and constant c is 0.1. Find the most trending element in the given stream.	10	CO3	K3	1.3.1																				
3 b)	Determine the distinct element in the stream using the Flajolet Martin algorithm. Stream: 4, 2, 5 ,9, 1, 6, 3, 7 Hash function, $h(x) = 3x + 1 \bmod 32$	10	CO3	K3	1.3.1																				
4a)	For the following given transaction data set, generate rules using the Apriori algorithm. Consider the values as Support=22% and Confidence= 70%. <table><tr><th>Transaction-ID</th><th>Item Purchased</th></tr><tr><td>1</td><td>I1, I2, I5</td></tr><tr><td>2</td><td>I2, I4</td></tr><tr><td>3</td><td>I2, I3</td></tr><tr><td>4</td><td>I1, I2, I4</td></tr><tr><td>5</td><td>I1, I3</td></tr><tr><td>6</td><td>I2, I3</td></tr><tr><td>7</td><td>I1, I3</td></tr><tr><td>8</td><td>I1, I2, I3, I5</td></tr><tr><td>9</td><td>I1, I2, I3</td></tr></table>	Transaction-ID	Item Purchased	1	I1, I2, I5	2	I2, I4	3	I2, I3	4	I1, I2, I4	5	I1, I3	6	I2, I3	7	I1, I3	8	I1, I2, I3, I5	9	I1, I2, I3	10	CO4	K3	2.4.1
Transaction-ID	Item Purchased																								
1	I1, I2, I5																								
2	I2, I4																								
3	I2, I3																								
4	I1, I2, I4																								
5	I1, I3																								
6	I2, I3																								
7	I1, I3																								
8	I1, I2, I3, I5																								
9	I1, I2, I3																								
4b)	Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9) Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2). The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as- $P(a, b) = x2 - x1 + y2 - y1 $ Use K-Means Algorithm to find the three cluster centers after the second iteration.	10	CO4	K3	2.4.1																				
5a)	Visualization is needed to present the facts available in the unstructured datasets. Explain, in brief, the most challenging issues that occur during effective data visualization in real life.	10	CO5	K2	2.4.4																				
5b)	Explain the various graphical techniques for visualizing the data facts and effective decision making.	10	CO5	K2	2.4.4																				

CO Course Outcomes mapped with respective question

KL Bloom's knowledge Level (K1, K2, K3, K4, K5, K6)

K1- Remember, K2- Understand, K3-Apply, K4- Analyze, K5: Evaluate, K6- Create