UPES End Semester Examination, May 2024

Course: Cognitive Analytics

Program: B TECH(CSE+AI&ML-H/NH)

Course Code: CSBA3009

Semester: 6

Time : 03 hrs.

Max. Marks: 100

Instructions:

Section A shall have 5 Questions of 4 marks each. All the questions are compulsory.

Section B shall have 4 Questions of 10 marks each, out of which 3 Questions shall be compulsory and 1 Questions shall have internal choice to attempt any one.

Section C shall have 2 Questions of 20 marks each, out of which 1 Question shall be compulsory and 1 Question shall have internal choice to attempt any one.

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Explain the importance of data and data visualization in decision making, with a focus on machine learning.	4	CO1
Q2	Developing microservices for a specific cloud platform can lead to vendor lock-in. Explain the different approaches that the developers can use to design and implement cloud microservices in order to promote portability across different cloud providers like AWS, Azure, and IBM.	4	CO3
Q3	Discuss the various ways in which we can use entropy to evaluate the effectiveness of a cognitive intervention or training program.	4	CO2
Q4	It has been determined that 5% of drivers checked at a road stop show traces of alcohol and 10% of drivers checked do not wear seat belts. In addition, it has been observed that the two infractions are independent from one another. If an officer stops five drivers at random, calculate the probability that exactly three of the drivers have committed any one of the two offenses. Also calculate the probability that at least one of the drivers checked has committed at least one of the two offenses.	4	CO1
Q5	The ethical implications of using cognitive services need careful consideration. Discuss potential biases that might exist in some cognitive services, particularly those related to tasks like facial recognition or sentiment analysis. Explain how can developers mitigate these biases to ensure responsible and ethical use of cognitive services.	4	CO2

SECTION B
(4Qx10M= 40 Marks)

Q6	Describe the basics of statistics and its application in everyday life, through the lens of machine learning.	10	CO2
Q7	A study investigates the number of times participants press a button during a cognitive task. It follows a poisson distribution with a mean of 5 presses. Calculate the probability of a participant pressing the button exactly 3 times.	10	CO1
Q8	You have been given a dataset containing measurements in three dimensions for a set of objects. Perform PCA to reduce the dataset to 2 dimensions and list the new 2D coordinates of each object. **Dataset for 3D Measurements** Object Dimension 1 Dimension 2 Dimension 3 A 1 2 3 B 2 3 4 C 3 4 5 D 4 5 6	10	CO3
Q9	Consider the following user-item interaction matrix, where each cell represents the rating (on a scale of 1-5) a user has given to an item. Empty cells indicate unrated items. Using a simple collaborative filtering approach, predict the rating User 1 would give to Item 3. **User-Item Interaction Matrix** User/Item Item 1 Item 2 Item 3 Item 4 User 1 5 3 2 User 2 3 4 User 3 2 3 5 User 4 2 1 3 OR Assume a dataset with two features that perfectly correlate with each other. Explain how dimensionality reduction could be applied to this dataset and also find the expected outcome. **Correlated Features Dataset** <table><tr><td>Sample</td><td>Feature 1</td><td>Feature 2</td></tr><tr><td>1</td><td>10</td><td>20</td></tr><tr><td>2</td><td>20</td><td>40</td></tr><tr><td>3</td><td>30</td><td>60</td></tr><tr><td>4</td><td>40</td><td>80</td></tr></table>	10	CO3
SECTION-C **(2Qx20M=40 Marks)**			
Q10	Compare the performance of two different data mining techniques (e.g., K-means vs. DBSCAN) on a specific cognitive analysis task based on evaluation metrics. Discuss the factors influencing your choice of technique.	20	CO2

Q11	Explain any 5 popular R packages for machine learning along with their usage and applications. (4 marks each) OR In decision tree, the major challenge is attribute selection. Gini index is one of the methods used to decide the optimal split from a root node and subsequent splits. The following table consists of 14 rows and 4 columns. The table depicts that heart disease depends on factors like high blood pressure, high cholesterol and FBS(fasting blood sugar). Note that the original values are converted into 1 and 0 which depict numeric classification. Calculate the Gini index for high blood pressure, high cholestrol and FBS. Also Find the factor based on which the decision tree will be built.	20	CO2

SNO	HighBps	HighChol	Fbs	target
1	1	1	1	1
2	1	1	0	1
3	1	1	0	1
4	0	1	0	0
5	0	1	0	0
6	1	0	0	0
7	1	1	0	1
8	0	1	0	0
9	1	0	1	1
10	1	0	0	0
11	1	1	0	1
12	1	1	0	1
13	1	1	0	1
14	0	1	0	0