|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |

PRESIDENCY UNIVERSITY BENGALURU

SCHOOL OF ENGINEERING

MAKE-UP EXAMINATION JULY 2024

|  |  |
| --- | --- |
| **Semester :IV** | **Date :01 JULY 2024** |
| **Course Code : CSA2007** | **Time : 9:30AM-12:30PM** |
| **Course Name : DATA MINING** | **Max Marks :100** |
| **Program :BCA** | **Weightage :50%** |

**Instructions:**

1. *Read all questions carefully and answer accordingly.*
2. *Question paper consists of 3 parts.*
3. *Scientific and non-programmable calculator are permitted.*
4. *Do not write any information on the question paper other than Roll Number.*

|  |  |  |  |
| --- | --- | --- | --- |
| **PART A** | | | |
| **ANSWER ANY 5 QUESTIONS 5Q X 4M=20M** | | | |
| 1 | Define Data Mining and mention four alternative names of data mining. | (CO 1) | [Knowledge] |
|  | | | |
| 2 | Data mining is a multi disciplinary field. Justify. | (CO 1) | [Knowledge] |
|  | | | |
| 3 | Describe the types of attributes with an example. | (CO 2) | [Knowledge] |
|  | | | |
| 4 | Find the similarity between given data points by using Euclidean distance.  Product X: (50,4.5) (Price, Rating)  Product Y: (45,4.8) | (CO 2) | [Knowledge] |
|  | | | |
| 5 | State the concept of Support and Confidence in Association Rule Mining. | (CO 3) | [Knowledge] |
|  | | | |
| 6 | Outline the differences between supervised and unsupervised learning. | (CO 4) | [Knowledge] |
|  |  |  |  |
| 7 | Recite any four key challenges in anomaly detection? | (CO 5) | [Knowledge] |
|  | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **PART B** | | | |
| **ANSWER ANY 5 QUESTIONS 5Q X 10M=50M** | | | |
| 8 | Illustrate the steps involved in knowledge discovery process with suitable diagram. | (CO 1) | [Comprehension] |
|  | | | |
| 9 | Discuss in detail about data cleaning and data integration methods in preprocessing with suitable example. | (CO 2) | [Comprehension] |
|  | | | |
| 10 | Compute Min-max and decimal scaling normalization from the following data:  300,150,600,450 and 750 | (CO 2) | [Comprehension] |
|  | | | |
| 11 | Built the FP-Growth tree from the given table: Let assume minsupport:3   |  |  | | --- | --- | | TID | ITEMS | | T100 | f,a,c,d,g,i,m,p | | T200 | a,b,c,f,l,m,o | | T300 | b,f,h,j,o | | T400 | b,c,k,s,p | | T500 | a,f,c,e,l,p,m,n | | (CO 3) | [Comprehension] |
|  | | | |
| 12 | Use KNN classification for predicting the class for a new sample D when X1 = 70 and X2 = 80 and K = 3 from the given data:   |  |  |  | | --- | --- | --- | | X1 | X2 | D | | 40 | 50 | Bad | | 95 | 75 | Good | | 60 | 70 | Good | | 45 | 35 | Bad | | 80 | 85 | Good | | (CO 4) | [Comprehension] |
|  | | | |
| 13 | Suppose we have a dataset of vehicles classified as either "Car" or "Motorcycle" based on features like Engine Type, Fuel Efficiency, and Top Speed.    Predict the class of a new vehicle with the following features using Naive Bayesian Classification:  (Engine Type: Electric, Fuel Efficiency: Low, Top Speed: Medium)=> Class:? | (CO 4) | [Comprehension] |
|  |  |  |  |
| 14 | Explain the following: i) Global Outliers ii) Local Outliers iii) Point Outliers iv) Contextual Outlier v) Collective Outliers | (CO 5) | [Comprehension] |
|  | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **PART C** | | | |
| **ANSWER ANY 2 QUESTIONS 2Q X 15M=30M** | | | |
| 14 | Consider the following database consisting of 10 transactions:   |  |  | | --- | --- | | Transaction ID | Items | | T101 | I1,I2, I5 | | T102 | I2, I4 | | T103 | I2, I3 | | T104 | I1,I2,I4 | | T105 | I1, I3 | | T106 | I2, I3 | | T107 | I1,I3 | | T108 | I1, I2, I3, I5 | | T109 | I1, I2, I3 |   Use Apriori Algorithm to find the strong association rules.  Assume minimum support count= 2 and minimum confidence= 70%. | (CO 3) | [Application] |
|  | | | |
| 15 | The given data set consists 14 customers historical data that records if they have purchased computer depending on age, income, student, credit rating. Apply CART algorithm to find out which attribute can be selected as root node for constructing decision tree.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Age | Income | Student | Credit\_Rating | Buys\_Computer | | Youth | High | No | Fair | No | | Youth | High | No | Excellent | No | | Middle\_Aged | High | No | Fair | Yes | | Senior | Medium | No | Fair | Yes | | Senior | Low | Yes | Fair | Yes | | Senior | Low | Yes | Excellent | No | | Middle\_Aged | Low | Yes | Excellent | Yes | | Youth | Medium | No | Fair | No | | Youth | Low | Yes | Fair | Yes | | Senior | Medium | Yes | Fair | Yes | | Youth | Medium | Yes | Excellent | Yes | | Middle\_Aged | Medium | No | Excellent | Yes | | Middle\_Aged | High | Yes | Fair | Yes | | Senior | Medium | No | Excellent | No | | (CO 4) | [Application] |
|  | | | |
| 16 | a. Illustrate in detail about text analysis with an example.  b. Demonstrate various types and applications of web mining. | (CO 5) | [Application] |
|  | | | |
|  | | | |