**POSSESION OF MOBILE IN EXAM IS UFM PRACTICE**

**Name Enrollment No**

**Jaypee Institute of Information Technology**

**T1 Examination, 2019**

**B.Tech -III Year**

**Course code: 16B1NCI635 Max. Marks: 20**

**Course Name: Data and Web Mining Max. Time: 1Hr**

**Note: Attempt all Questions:**

**Q1:** [CO1] **[Marks 5]** A genetics engineer was attempting to cross a tiger and a cheetah. She predicted a phenotypic outcome of the traits she was observing to be in the following ratio 4 stripes only: 3 spots only: 9 both stripes and spots. When the cross was performed and she counted the individuals she found 50 with stripes only, 41 with spots only and 85 with both. According to the Chi-square test, did she get the predicted outcome? (For degree of freedom 1, 2 and 3 at 0.05 significance level chi-square values are: 3.841, 5.991 and 7.815)

**Q2: [CO3] [Marks 4+2]** Consider a data having attributes Color, Height and Width, and the class can be either yes or no.

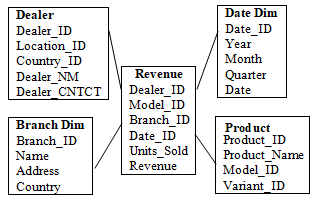
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Example | Color | Height | Width | Class |
| A | Red | Short | Thin | NO |
| B | Blue | Tall | Fat | YES |
| C | Green | Short | Fat | NO |
| D | Green | Tall | Thin | YES |
| E | Blue | Short | Thin | NO |

1. Design a decision tree based on ID3.
2. For the decision tree classifier, verify the results for the tuples (Green , Short, Thin,Yes) (Blue, short, fat, yes) and Determine the classifier accuracy.

**Q3:[CO2] [Marks 3]** Consider a schema of retail chain Bigmart. Perform the OLAP operations considering the base cuboid as (dealer ,year, branch, product).

a) List the revenue generated by all products in the year 2018.

b) List the total sales done by each dealers in January 2018.



|  |  |
| --- | --- |
| D1 | Computer electronic device computations. |
| D2 | Computer electronic device logical arithmetic computations. |
| D3 | CPU processing logical arithmetic computations. |

**Q4: [CO1] [3+1 marks]** Suppose we query an IR system for the query "Computer logical computations". The database collection consists of three pre-processed documents (D = 3) shown below:

1. Which document will be fetched using term-frequency document vectors incidence matrix?
2. Compute supremum distance between document D1 and D2.

|  |
| --- |
| 27, 26, 4, 21, 24, 34, 7, 15, 28, 29, 21, 9. |

**Q5:[CO1][2 Marks]** Consider the data for attribute “weight”:

1. Partition the data using equal width partitioning. . Number of intervals = 3.
2. Normalize the weight = 34 using min-max method (min=1, max=2)