**Name Enrollment No**

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

**Jaypee Institute of Information Technology**

**Test 2 Examination, 2019**

**B.Tech -III Year**

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

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

**Note: Attempt all Questions:**

Q1:[CO3] **[Marks 3]** Below is a table showing how two human judges rated the relevance of a set of 12 documents to a particular information need (0 = nonrelevant, 1 = relevant). Let us assume that you’ve written an IR system that for this query returns the set of documents {4, 5, 6, 7, 8}.

1. Calculate precision, recall, and F1 of your system if a document is considered relevant only if the two judges agree.
2. Calculate precision, recall, and F1 of your system if a document is considered relevant if either judge thinks it is relevant.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Students | A | B | C | D | E | F | | Food | 2 | 6 | 4 | 9 | 8.5 | 1 | | Entertainment | 4 | 6 | 2 | 3 | 1 | 5 |   Q2: [CO4]**[Marks 4]** The daily expenditure on the food and the entertainment of six students are shown in given table. Use agglomerative hierarchical clustering based on longest distance to form the final customer stores. Plot the dendogram.   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | Pepperoni | Pineapple | PicledOnion | Liked | | A | True | True | True | False | | B | True | False | False | True | | C | False | True | True | False | | D | False | True | False | True | | E | True | False | False | True |   Q3**:[CO1][Marks 4+2+2]** The Pants Pizza Parlour sells pizzas with optional toppings: pepperoni, pineapple and pickled onion. Every day this week you have tried a pizza (A to E) and kept a record of which you liked:  a) Show how the naive Bayes classifier would classify {pepperoni = true, pineapple = true, pickledOnion= false}.  b) Are pineapple = true and pickledOnion = true conditionally independent given liked = false? Show your working.  c) Show how the 2NN classifier would classify {pepperoni = false, pineapple = false, pickledOnion= true}; |  |

Q4:[CO1] **[Marks 2]** Assume you apply DBSCAN to the same dataset, but the examples in the dataset are sorted differently. Will DBSCAN always return the same clustering for different orderings of the same dataset? Give reasons for your answer.

Q5:[CO1] **[Marks 3]**Try placing the following faces into groups using any of the criteria: glasses, gender, and smile? Which method would you apply to justify. Show the results.

|  |  |  |  |
| --- | --- | --- | --- |
| **Summary of the face characteristics** | | | |
| **case** | **sex** | **glasses** | **smile** |
| 1 | m | y | y |
| 2 | m | y | n |
| 3 | f | n | y |
| 4 | f | n | n |
| 5 | m | y | y |
| 6 | f | n | y |
| 7 | m | y | y |

AQ1 Suppose we are building a classifier that says whether a text is about sports or not. Our training data has 5 sentences:

|  |  |
| --- | --- |
| **Text** | **Tag** |
| “A great game” | Sports |
| “The election was over” | Not sports |
| “Very clean match” | Sports |
| “A clean but forgettable game” | Sports |
| “It was a close election” | Not sports |

Now, which tag does the sentence *A very close game* belong to?

<https://monkeylearn.com/blog/practical-explanation-naive-bayes-classifier/>

AQ2:

Now suppose that we have a new document that we don’t know the label of. What is the probability that a word in the document is wordtype 1?

<https://people.cs.umass.edu/~brenocon/inlp2015/midterm-review-sols.pdf>