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

**Name Enrollment No**

**Jaypee Institute of Information Technology**

**T3 Examination, Even 2019**

**B.Tech-IV Year**

**Course Title: Data and Web Mining Maximum Time: 2 Hrs.**

**Course Code: 15B1NCI635 Maximum Marks: 35 marks**

**Note: Attempt all questions**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | I1 | I2 | I3 | I4 | I5 | I6 | I7 | I8 | I9 |
| U1 | 4 | 5 |  | 5 | 1 | **??** | 3 | 2 | 1 |
| U2 |  | 3 | 4 | 3 | 1 | 2 | 1 |  |  |
| U3 | 2 |  | 1 | 3 |  | 4 | 5 | 3 |  |

**Q1 [CO6][Marks 4+4]** Given below is the utility matrix representing the ratings on a 1-5 star scale of nine items, I1 to I9, by three users U1, U2, U3. Compute the following from the given data:

1. What can be the missing predicted rating for item 6 (I6) in recommendation system.
2. If we consider I1 to I9 in above matrix as terms and U1 to U3 represents three documents. Therefore, according to the matrix document U1 contains item I1 4 times i.e. term frequency. Now which document will be retrieved using TF-IDF scoring method if we search the query containing following terms: “I2, I4, I8, I9, I8”.

|  |  |
| --- | --- |
| TID | Items |
| 1 | P,Q,V |
| 2 | Q,R,S |
| 3 | P,R,S,T |
| 4 | P,S,T,T |
| 5 | P,Q,R |
| 6 | P,Q,R,S |
| 7 | Q |
| 8 | Q,R,T |
| 9 | V |

**Q2 [CO5] [Marks 5+2+3]** A database has 9 transactions. Let min support = 30% & min confidence = 65%.

1. List the frequent k-itemset for the largest k using FP-growth (FP tree) and
2. Compare the results of part (a) using Apriori algorithm.
3. Which items X and Y can be placed on the same shelf, so that buyers of one item would be prompted to buy the other?

**Q3: [CO6] [Marks 3+3]** If we can get 1,000 pages pointing to our home page, but only have one link leaving from homepage to one of the incoming page.

1. What will be the page rank of all pages according to random surfer model? Damping factor = 0.15, iterations = 3.
2. What will be the page rank of all pages according to HITS? Iterations = 2

|  |  |  |
| --- | --- | --- |
| Student id | Quiz 1 | Quiz 2 |
| 1 | 8 | 6 |
| 2 | 6 | 5 |
| 3 | 5 | 6 |
| 4 | 4 | 5 |
| 5 | 9 | 8 |
| 6 | 7 | 9 |

**Q4: [CO3+CO4] [Marks 2+5+2+2] a)** Suppose we retrieve 11 documents and every odd numbered retrieved document is relevant. Find out the precision and recall of information retrieval system.

**b)** A group of six students whose performance based on marks obtained in quiz1 and quiz2 is given in table. Now this group is required to be sent in two teams for inter college competition. Extract the best two team members. Apply the appropriate clustering technique.

**c)** Prove that in DBSCAN for a fixed *MinPts* value and 2 neighbourhood thresholds, a cluster with respect to e1and *Minpts* must be subsets of a cluster C’ with respect to and *Minpts*.

**d)** Nine actual values of target variable in the training file is as follows: [0,0,1,1,0,1,0,1,1].

What is the entropy of the target variable?