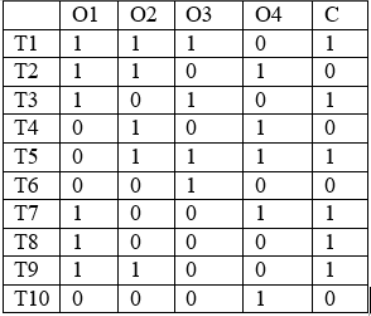
**BITS Pilani, Hyderabad Campus**

**Comprehensive ExaminationTest (Regular Exam)**

**Course No.: CSF415 Course Title: Data Mining Max. Marks: 70**

Section 1:

Q.1 Consider the testing data shown in below Table:



***What is the accuracy of the Classifier2 ? [1.5]***

***NOTE: classifier1: if O1+O2+O3+O4>=2 then C=1 else C=0***

***Classifier2: if O3+O4>=1 then C=1 else C=0***

***Classifier3: if O2==O1 then C=O2 elseif O1==0 then C=0 else C=1***

Q.2 The 1-itemset X in the original large 1-itemsets L1 is a loser in the updated database DB U db if and only af X.SupportUD > s \* (D + d) ***[1.5]***

***True***

***False***

Q.3 If confidence (Object1-->Object2)=confidence(Object2-->Object1), then support(Object1)== support(Object 2) ***[1.5]***

***True***

***False***

Q.4 The variance of n1=(a+b)/2 and n2=(a-b)/2 is a\*b/2 ***[1.5]***

***True***

***False***

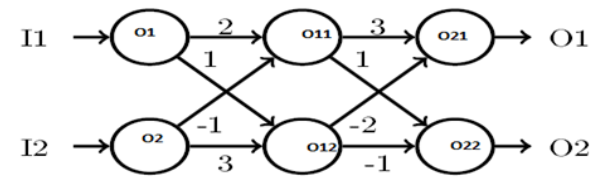
Q.5 The 1-itemset X in the original large is infrequent become infrequent in the updated database DBUdb if and only af X.Support\_db < s \* d ***[1.5]***

***True***

***False***

Section 2:

Q.1 Figure1 represents a feed-forward neural network:



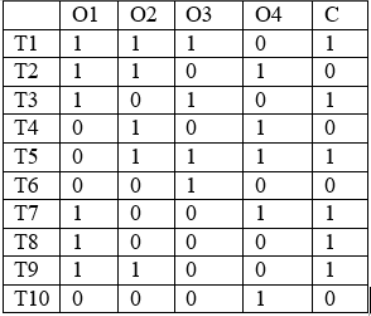
Each of the node uses the following activation function: (! is used to represent factorial of number like 3!=3\*2=6 )

**F(x)= x!>x3 if x>1**

**1 if x<=1**

**What is the output of neuron O1(final output) for input I1=2 and I2=4 *[2]***

Q.2 Consider the testing data shown in below Table:



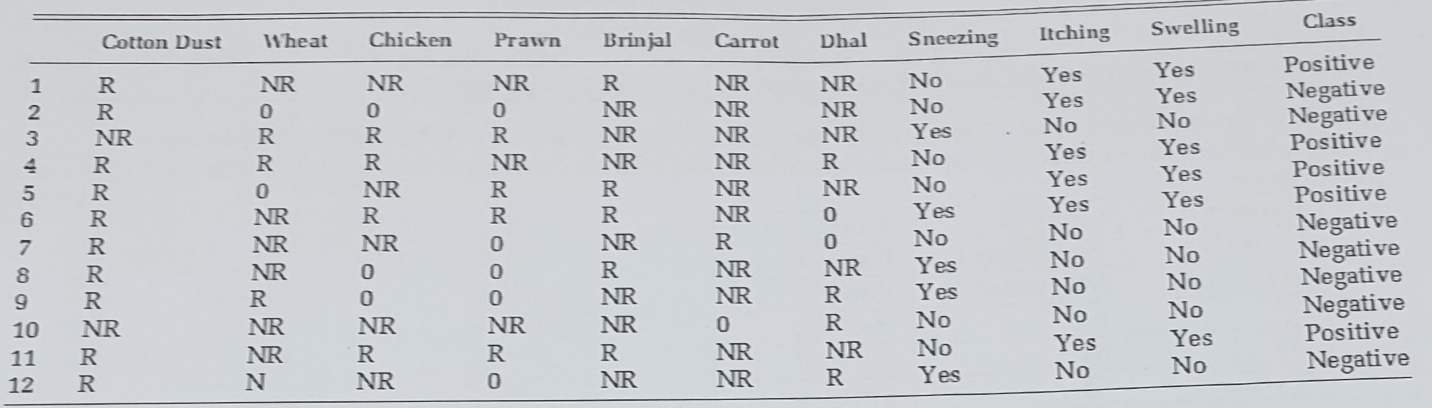
***Which of the following classifier has high value of F-Measure? [2]***

***classifier1: if O1+O2+O3+O4>=2 then C=1 else C=0***

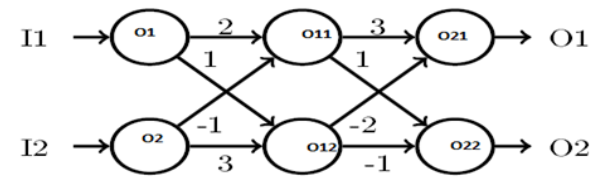
***Classifier2: if O3+O4>=1 then C=1 else C=0***

***Classifier3: if O2==O1 then C=O2 elseif O1==0 then C=0 else C=1***

Q.3 Consider the data shown in below Table. What is the value of similarity metric between data samples 1 and 2 ***[2]***



Q.4 Figure1 represents a feed-forward neural network:



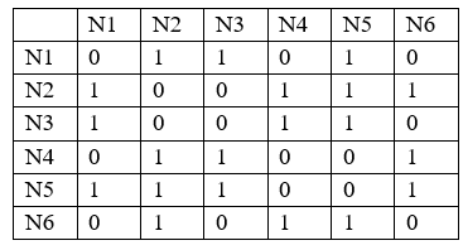
Each of the node uses the following activation function: (! is used to represent factorial of number like 3!=3\*2=6 )

**F(x)= x!>x3 if x>1**

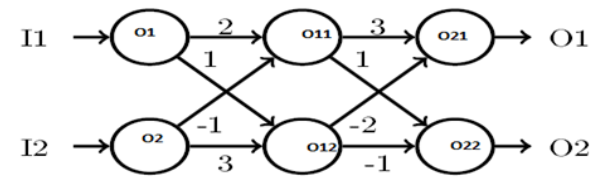
**1 if x<=1**

**What is the output of neuron O1(final output) for input I1=1 and I2=1 *[2]***

Q.5 For the undirected graph given below in the form of an adjacency matrix. What is the summation value of EBC scores for all edges if start node is N1? ***[2]***



Q.6 Figure1 represents a feed-forward neural network:



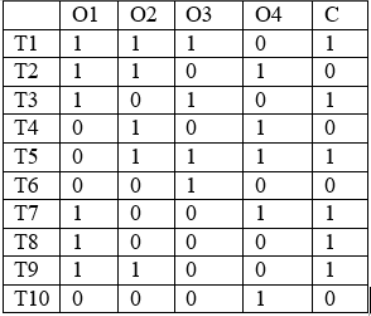
Each of the node uses the following activation function: (! is used to represent factorial of number like 3!=3\*2=6 )

**F(x)= x!>x3 if x>1**

**1 if x<=1**

**What is the output of neuron O11 for input I1=2 and I2=4 *[2]***

Q.7 Consider the testing data shown in below Table:



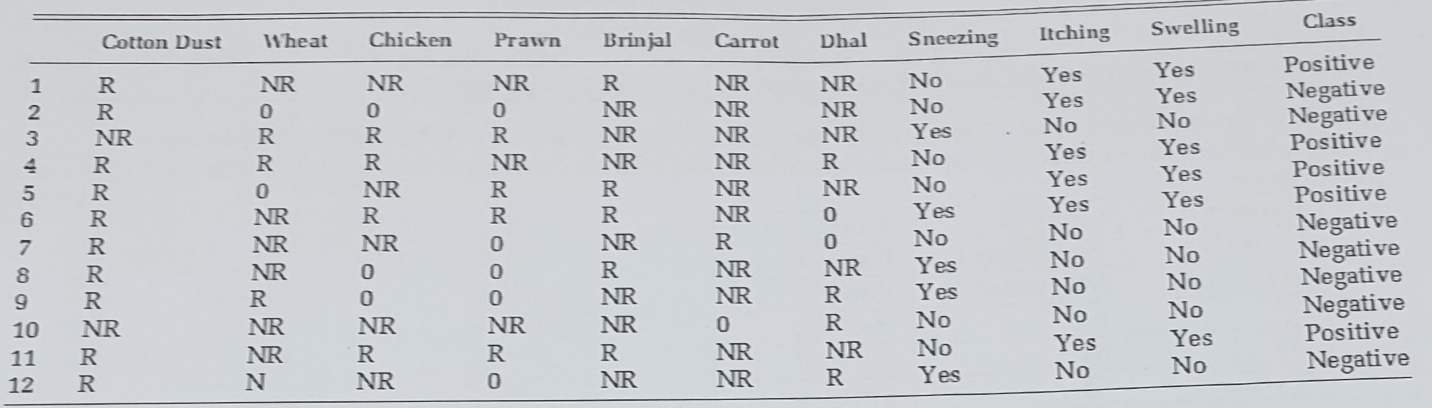
***What is the number of correctly classified samples of the model using major voting ensemble techniques? [2]***

***NOTE: classifier1: if O1+O2+O3+O4>=2 then C=1 else C=0***

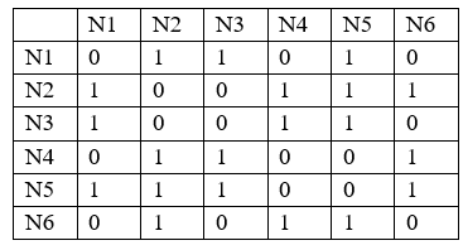
***Classifier2: if O3+O4>=1 then C=1 else C=0***

***Classifier3: if O2==O1 then C=O2 elseif O1==0 then C=0 else C=1***

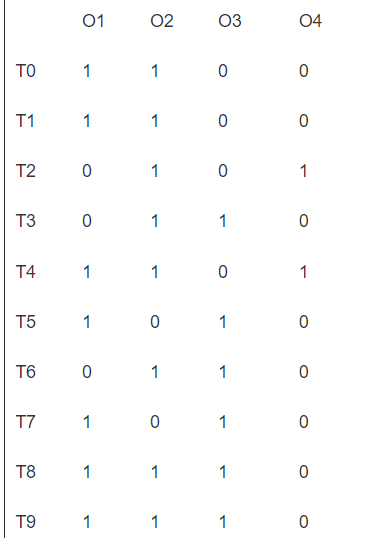
Q.8 Consider the data shown in below Table. What is the value of disimilarity metric between data samples 1 and 2 ***[2]***



Q.9 For the undirected graph given below in the form of an adjacency matrix. What is the value of EBC scores of (N1—N2) if start node is N1? ***[2]***



Q.10 Consider the transaction data shown in the below Table:



*What is number of scan required to find 2-frequent item-sets using FUP algorithm with min sup=30% if we will add four more transactions.* ***[2]***

*i.T10: 01,04*

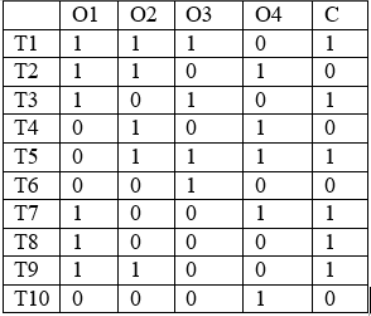
*ii.T11: 02,O4*

*iii.T12: O3*

*iv.T13: O1*

Section 3:

***Q1: Consider the training data shown in below Table:***



***The trained classifiers on above data after 1st, 2nd, and 3rd rounds are given below:***

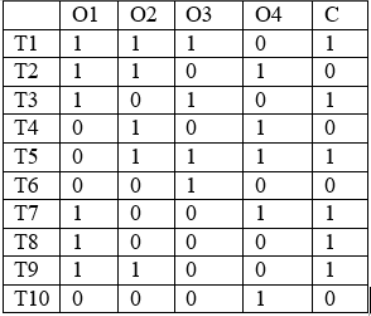
***1st Round: if O1+O2+O3+O4>=2 then C=1 else C=0***

***2nd Round: if O3+O4>=1 then C=1 else C=0***

***3rd Round: if O2==O1 then C=O2 elseif O1==0 then C=0 else C=1***

***What is the value of weight after 1st round classifier for T1? (consider log2) [4]***

***Q2: Consider the training data shown in below Table:***



***The trained classifiers on above data after 1st, 2nd, and 3rd rounds are given below:***

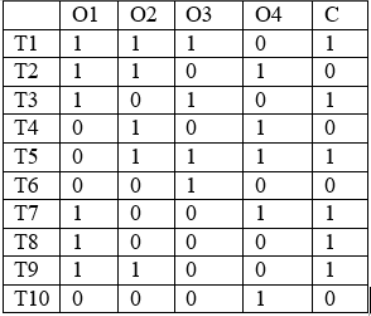
***1st Round: if O1+O2+O3+O4>=2 then C=1 else C=0***

***2nd Round: if O3+O4>=1 then C=1 else C=0***

***3rd Round: if O2==O1 then C=O2 elseif O1==0 then C=0 else C=1***

***What is the value of error rate for 1st round classifier? [4]***

***Q3: Consider the training data shown in below Table:***



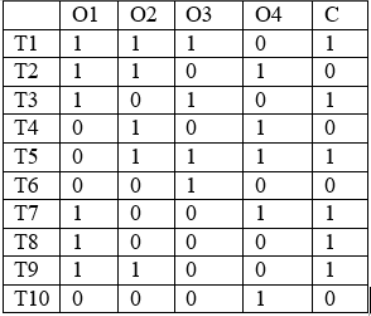
***Construct major voting ensemble technique using following desision tree:***

***1. Decision tree with O1 feature***

***2. Decision tree with O4 Feature***

***What is the number of correctly classified C==1 samples of above model ? (NOTE: consider 0 class for same probability)? [4]***

***Q4: Consider the training data shown in below Table:***



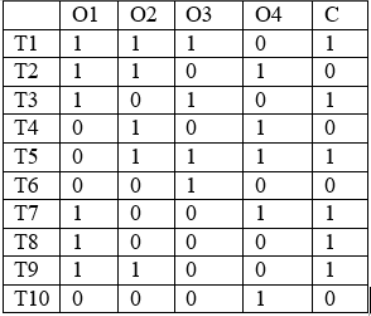
***What is the accuracy of the model using major voting ensemble techniques?***

***NOTE: classifier1: if O1+O2+O3+O4>=2 then C=1 else C=0***

***Classifier2: if O3+O4>=1 then C=1 else C=0***

***Classifier3: if O2==O1 then C=O2 elseif O1==0 then C=0 else C=1 [4]***

***Q5: Consider the training data shown in below Table:***



***The trained classifiers on above data after 1st, 2nd, and 3rd rounds are given below:***

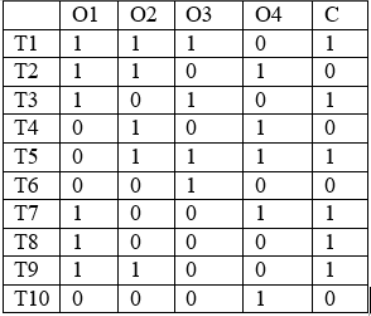
***1st Round: if O1+O2+O3+O4>=2 then C=1 else C=0***

***2nd Round: if O3+O4>=1 then C=1 else C=0***

***3rd Round: if O2==O1 then C=O2 elseif O1==0 then C=0 else C=1***

***What is the value of error rate for 2nd round classifier? [4]***

***Q6: Consider the training data shown in below Table:***



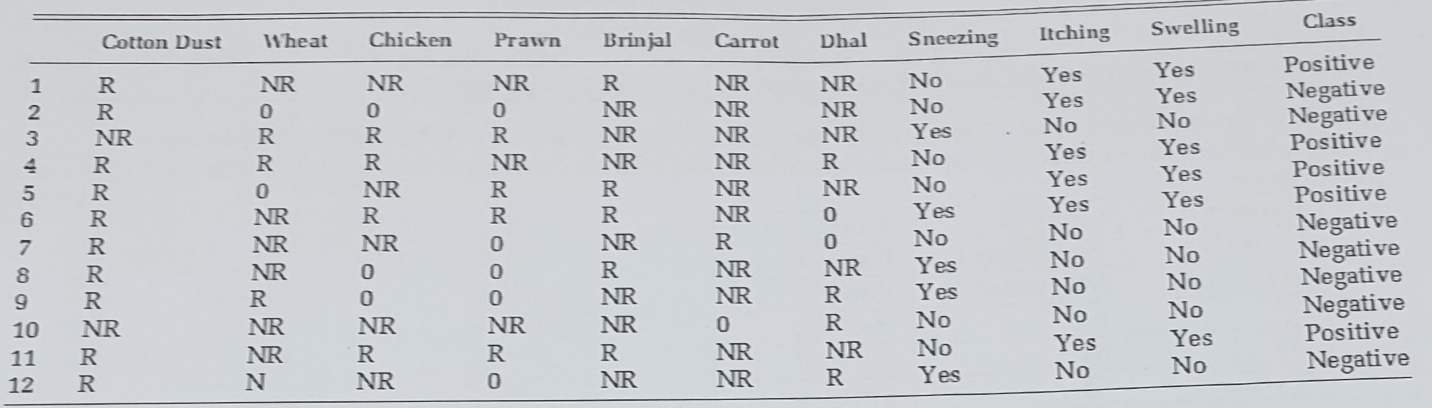
***Construct major voting ensemble technique using following desision tree:***

***1. Decision tree with O1 feature***

***2. Decision tree with O4 Feature***

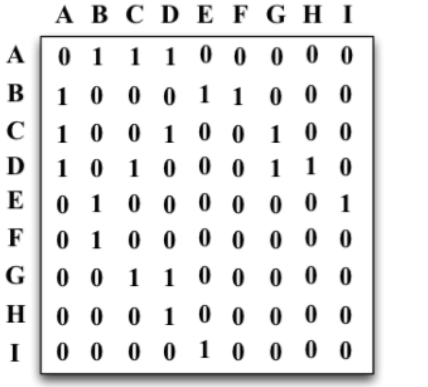
***What is the training accuracy value of above model ? (NOTE: consider 0 class for same probability)? [4]***

***Q.7 Consider the data shown in below Table. Construct model using 5 nearest neighbour algorithm and find the class for data sample 12. Consider 1-10 data sample for training. [4]***

******

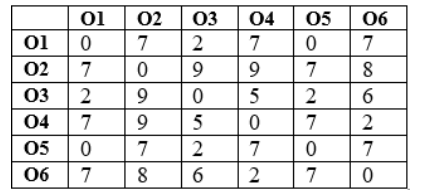
Section 4:

**Q1: For the undirected graph given below in the form of an adjacency matrix to all questions in this section:**



1. ***What is the value of the degree centrality for the node E?(ans must be in decial format)* [1.5]**
2. ***What is the value of Clustering Coefficient for node D? (ans must be in decial format)* [1.5]**
3. ***Which of the follwong node has the highest betweenness centrality?* [1.5]**
4. ***Which of the follwong node has the least betweenness centrality?* [1.5]**

**Q2: Consider the Similarity or Dissimilarity Matrix data shown in below Table to answer all question in this section.**



1. **What is the value of DIS3 (3-distance of an object) for object O4? [1.5]**
2. ***What is the value of 3-Reachability distance from O1 to O3?* [1.5]**
3. **What is the number of 3-distance neighbourhood for object O4? [1.5]**
4. ***What is the value of Local outlier factor (LOF3) for Object O1? (answer must be in decimal format)* [2]**
5. **What is the value of 3-Local reachability density (lrd3) for Object O1? [2]**