**The LNM Institute of Information Technology**

**Department of Computer Science and Engineering**

|  |
| --- |
| **Information Retrieval (IR)**  **Make up Exam** |

**Time: 3 Hour Date: 02/08/2017** **Max. Marks: 4**0

**Instructions**: 1) Look through the whole exam and answer the questions that you find easiest first.

2) If necessary, you may make assumptions that are reasonable, and if you do make an

assumption, state it clearly.

3) You may use a calculator.

Q1. Explain the Naive Bayes assumption. Train the classifier by using the table below and classify the test sample X = (Gear = Automatic , Color=Yellow, Origin=Domestic , Type=SUV) as Stolen “Yes” or “No”. **[2 + 5]**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sample | Color | Type | Origin | Stolen? |
| 1 | Red | Sports | Domestic | Yes |
| 2 | Red | Sports | Domestic | No |
| 3 | Red | Sports | Domestic | Yes |
| 4 | Yellow | Sports | Domestic | No |
| 5 | Yellow | Sports | Imported | Yes |
| 6 | Yellow | SUV | Imported | No |
| 7 | Yellow | SUV | Imported | Yes |
| 8 | Yellow | SUV | Domestic | No |
| 9 | Red | SUV | Imported | No |
| 10 | Red | Sports | Imported | Yes |

**Q2.** What is NCUT Weighting? Compare NCUT weighting with TF-IDF. **[4+3]**

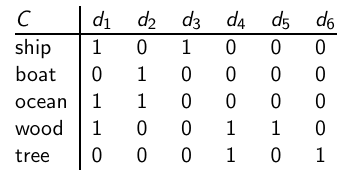
**Q3.** What are the classification performance evaluation parameters used in Information Retrieval? Discuss with an appropriate example **[5]**

**Q5**. Derive the confusion matrix for the following clusters and calculate the Rand Index and F-Measure.

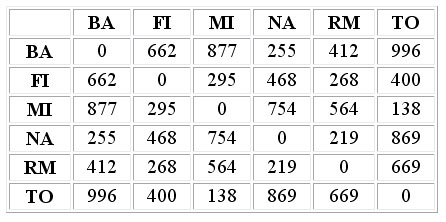
Cluster1= **{4, 4, 6, 6, 6, 8}** Cluster2= **{6, 6, 6, 6, 8}** Cluster3 = **{4, 4, 8 , 8, 8, 8}**

**[2+3+3]**

**Q6 What is SVD? Using Latent Semantic Analysis reduce the dimension of the following document matrix [2+6]**

****

**Q7. Apply the Hierarchical clustering algorithm to cluster the following data** **[5]**



{Best of Luck}