**The LNM Institute of Information Technology**

**Department of Computer Science and Engineering**

|  |
| --- |
| **Information Retrieval and Web Search(IRWS)**  **End Semester Exam** |

**Time: 3 Hour Date: 29/04/2017** **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 = (age<=30, Income=medium, Student=yes , Credit\_rating=Fair) as buys\_computer “Yes” or “No”. **[2 + 5]**

|  |  |  |  |
| --- | --- | --- | --- |
| **income** | **student** | **Credit\_rating** | **Buys\_computer** |
| high | no | fair | No |
| high | no | excellent | No |
| high | no | fair | Yes |
| medium | no | fair | Yes |
| low | yes | fair | Yes |
| low | yes | excellent | No |
| low | yes | excellent | Yes |
| medium | no | fair | No |
| low | yes | fair | Yes |
| medium | yes | fair | Yes |
| medium | yes | excellent | Yes |
| medium | no | excellent | Yes |
| high | yes | fair | Yes |
| medium | no | excellent | No |

**Q2.** Write and compare various weighting methods used in Information Retrieval. **[5]**

**Q3.** What are the clustering performance evaluation parameters used in Information Retrieval?

Discuss with an appropriate example. **(Give the example other than in Q5)** **[5]**

**Q4.** Draw the Decision tree for the Car dataset given in the below table: **[6]**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

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

Cluster1= **{a, a, a, b, c, c}** Cluster2= **{b, b, b, b, c}** Cluster3 = **{a, a, b, c, c, c, c}**

**[2+2+2]**

**Q6.** Explain the role of Information Retrieval in Social Media. At lease take two popular social media for your discussion **(Only Research aspects) [6]**

**Q7.** Describe the opportunities and challenges in building XML based information retrieval system. Draw the DOM tree for the XML document given below. **[5]**

<play>

<author>Shakespeare</author>

<title>Macbeth</title>

<act number=“I”>

<scene number=“”vii”>

<title>Macbeth’s castle</title>

<verse>Will I with wine

…</verse>

</scene>

</act>

</play>

{Best of Luck}