LNMIIT/B.Tech./CSE/PE/2018-19/ODD/CSE/ET

LNMIIT

The LNM Institute of Information Technology

Department of Computer Science and Engineering

Ujjawal Kumar
17UCS171

Information Retrieval (IR)
End Semester Exam

Time: 3 Hour **Date: 10/12/2019** **Max. Marks: 40**

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. Draw the Decision tree for the data sample given in the below table: **[5 + 1]**

Below are the symptoms and diagnosis of the patients examined by the doctor

Patid#	Chills	Runny nose	Headache	Fever	Flue
P1	Y	Y	Mild	Y	Yes
P2	Y	N	Strong	N	No
P3	Y	N	Mild	Y	Yes
P4	N	Y	Strong	Y	Yes
P5	N	N	No	Y	Yes
P6	N	Y	Strong	N	No
P7	N	Y	Mild	Y	Yes
P8	N	N	No	N	No
P9	Y	Y	Mild	Y	Yes
P10	N	N	No	N	No

Using the drawn decision table with above training data Classify the test sample **X= {(Chills="Yes", Runny nose ="No" , Headache="Mild" and Fever = "Yes")} as Flue or Not.**

Q2. Consider the below documents and group the semantically related terms using complete-link hierarchical agglomerative clustering.

- **Doc1:** [Coach, play, referee, goal, game, goal, mistake, linesman, referee, win, coach, mistake] 12
- **Doc2:** [score, goal, match, World, Cup, quarter, final, win, goal, tournament, score, win] 12
- **Doc3:** [World, Cup, game, lose, win, tournament, match, goal, team, lose, semi, final, World, Cup, tournament, win, game, semi final] 19
- **Doc4:** [World, Cup, quarter, final, score, goal, score, beat, team, extra, time, referee, score, quarter, game, win, final] 17

Show all similarity calculations needed for clustering, and the final cluster hierarchy. Apply normalization and *tf-idf* weights with ***cosine similarity***. **[2+2+4]**

doc2 doc4 — 0.72

doc2 doc4 doc3 → 0.48

doc1 doc2 doc3 doc4 → 0.26

2
12
12
19
17
60

LNMIIT

Q3. Write short notes on any one with suitable example: [4]

a) Latent Semantic Indexing.
b) Recommender System.

Q4. Explain the role of Information Retrieval in Social Media. At lease take one popular social media for your discussion (Only Research aspects) [4]

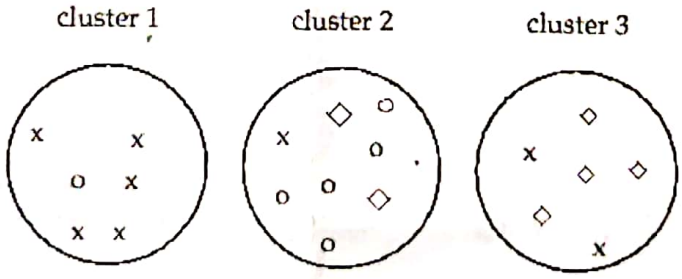
Q5. Consider the text document given below and suggest some **extractive text summarization model** to get the comprehensive summary.

[Prime Minister Narendra Modi needs to jump-start the economy to keep that 2014 campaign promise. Months into a second term, GDP growth of 4.5% is about half the pace ideal for its 1.3 billion people. Modi surprised markets in good ways, and bad, before. Here's how he might do so again.

A crisis of confidence is crippling the private sector, which is hobbled by $200 billion of bad debt, and isn't investing as demand weakens and businesses adjust to new rules on everything from sales tax to insolvency. If India Inc is now acutely aware of the costs of corruption, the pendulum of fear has swung too far. Industrialists once willing to back large infrastructure projects are now risk-averse.

One big problem is unpredictable policymaking that has seen companies abruptly stripped of key licences or whacked with costly new contractual terms. Prashant Ruia, whose Essar Group lost its steel asset through a new bankruptcy regime to ArcelorMittal, told the audience at Breakingviews' Predictions event in Mumbai on Dec. 6 that fixing regulatory risk is key to revival.] [6]

Q6. Consider the below outcome of clustering algorithm and calculate the **purity, NMI, Rand Index** and **F-Measure** to evaluate the cluster quality. [2+3+3+2+2]



{Best of Luck}