

BE SEM VIII (choice based) Summer 2020

Duration: 3hrs

[Max Marks: 80]

2115/20

N.B. : (1) Question No 1 is Compulsory.

(2) Attempt any three questions out of the remaining five.

(3) All questions carry equal marks.

(4) Assume suitable data, if required and state it clearly.

- 1 Attempt any FOUR [20]
 - a What is the significance of Feature Selection in text classification? Name two common methods. 5
 - b Why Stop word removal is necessary in Natural Language Processing? Explain the process of stop word removal. 5
 - c Explain the concept of Homophily and its impact on social media influence. 5
 - d What is Inverted Index and its importance in web search. 5
 - e Explain the challenges of mining social media data. 5
- 2 a Explain Web Usage Mining and its phases. Discuss the challenges involved in identifying unique users and sessions from raw server logs. [10]
 - b Why is index compression necessary in Web Mining? Explain with neat diagram components of web search engine. [10]
- 3 a Compare Classical Recommendation Algorithms with modern algorithms that use Social Context. [10]
 - b What is NER? Explain different types of entities recognized from text data. Identify the entities in the following sentence "Narendra Modi visited the DRDO headquarters last Tuesday." [10]
- 4 a Define text clustering and explain various text clustering techniques [10]
 - b Discuss the detailed process of Opinion Lexicon Expansion using both dictionary and corpus-based techniques. [10]
- 5 a Explain the process of performing Behavior Analytics on social media platforms to predict user trends. [10]
 - b Compare and contrast Bayesian Networks and Conditional Random Fields (CRF) for text modeling. [10]
- 6 a Explain the Supervised approach to Document Sentiment Classification. [10]
 - b Explain opinion spam detection techniques in detail. [10]
