

BE Sem 8, Choice based, 'c' Scheme, R2019-20, May 2025

Date- 23/5/2025  
Branch- EXTC

Duration: 3hrs

[Max Marks:80]

- 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 Discuss various text preprocessing techniques and their relevance in text mining ? 5
  - b Explain the HMM model in detail and state its limitations. 5
  - c Define web spamming and differentiate between content spam and link spam. 5
  - d Explain the concepts of "influence" and "homophily" in the context of social networks. 5
  - e Define opinion mining and explain the problems of opinion mining. 5
- 2 a Discuss various text preprocessing techniques and their relevance in text mining and text extraction tasks? Justify the answer with examples. [10]
- b What is opinion spam? Explain any one opinion spam detection technique in detail. [10]
- 3 a Explain with block diagram classical recommendation algorithms for social media . [10]
- b Discuss the techniques used for information extraction from unstructured text and their limitations. [10]
- 4 a What is N-gram modeling and its application in text mining? Explain unigrams, bigrams, and trigrams with examples. [10]
- b Explain the concepts of session analysis and visitor analysis in web usage mining. What types of insights can be gained from each of these analyses? [10]
- 5 a Differentiate between a standard web search engine and a meta-search engine. Explain the working of meta search engines. [10]
- b Explain decision tree classifier in detail. [10]
- 6 a Explain the concept of Latent Semantic Indexing (LSI). How does it attempt to overcome the limitations of traditional keyword-based search? [10]
- b Explain the concept of Named Entity Recognition (NER). Discuss the different categories of named entities and the challenges involved in accurately identifying them in text. [10]

\*\*\*\*\*