

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 What is the role of stemming , lemmatization and stop word removal in text preprocessing? Briefly explain each technique. 5
- b Explain the HMM model in detail and state its limitations. 5
- c Discuss two common hiding techniques employed by spammers to conceal their activities from users and search engine crawlers. 5
- d Define link spamming and describe common techniques used in it. 5
- e How is behavior analytics used in the context of social media data? 5
- 2 a Explain the concept of an inverted index and Latent Semantic Indexing (LSI) in web mining. How do these techniques facilitate efficient web search? [10]
- b Explain how spammers manipulate website content and links to unfairly boost search rankings, and discuss methods to detect and prevent such web spam. [10]
- 3 a Describe different approaches to model and integrate social context into recommendation algorithms. Explain evaluation metrics used to assess recommendation system performance. [10]
- b Discuss the various relation extraction and NER extraction techniques from unstructured text. [10]
- 4 a Explain distance based clustering algorithm in detail. [10]
- b Explain web usage mining process in detail. How association and correlation analysis is performed in web usage mining? [10]
- 5 a Explain behavior analytics in the context of social media. [10]
- b Explain any two text classification algorithms [10]
- 6 a Discuss unsupervised methods of document sentiment classification. [10]
- b Explain text mining applications and challenges. [10]
