Printed Page:- 04 Subject Code:- ACSBS0715

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VII - THEORY EXAMINATION (2024 - 2025)

Subject: Advanced Social, Text and Media Analytics

Time: 3 Hours Max. Marks: 100

General Instructions:

IMP: *Verify that you have received the question paper with the correct course, code, branch etc.*

1. This Question paper comprises of three Sections -A, B, & C. It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.

2. Maximum marks for each question are indicated on right -hand side of each question.

3. Illustrate your answers with neat sketches wherever necessary.

4. Assume suitable data if necessary.

5. Preferably, write the answers in sequential order.

6. No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.

SECTION-A 20

1. Attempt all parts:-

1-a. Select probabilistic model which is commonly used for information extraction in text mining?(CO1,K1) 1

(a) Naive Bayes

(b) Decision Trees

(c) Support Vector Machines

(d) K-Means

1-b. Identify the full form of "TF-IDF" in context of text mining? (CO1,K4) 1

(a) Text Frequency-Inverse Document Frequency

(b) Term Frequency-Inverse Data Format

(c) Total Frequency-Inverse Document Filter

(d) Token Frequency-Inverse Data Function

1-c. In the context of NLP, what does POS tagging stand for?(CO2, K4) 1

(a) Parts of Speech tagging

(b) Positional Syntax tagging

(c) Positive Sentiment tagging

(d) Personal Object Subject tagging

1-d. In content analysis, the term "coding" outline the feature. (CO2,K1) 1

(a) Creating a new software program

(b) Assigning categories or labels to textual data

(c) Decrypting encoded messages

(d) Analyzing numerical data

1-e. Select the correct statement that represents "viral post" correctly. (CO3,K2) 1

(a) A post that has been viewed by a large number of people

(b) A post that has been shared by a large number of people

(c) A post that has received a large number of likes

(d) A post that has been deleted by the social media platform

1-f. Select statement that describe hybrid recommendation system. (CO3,K1) 1

(a) A recommendation system that combines collaborative filtering and content-based filtering

(b) A recommendation system that recommends only popular items

(c) A recommendation system that recommends items based on demographic data

(d) A recommendation system that recommends items randomly

1-g. In social network analysis, what does a node represent? (CO3, K1) 1

(a) A computer server

(b) An individual or entity in the network

(c) A website URL

(d) A type of computer software

1-h. Which of the following is an example of a directed graph? (CO4,K1) 1

(a) Facebook friendship network

(b) Instagram follower network

(c) Twitter retweet network

(d) All of the above

1-i. In the context of link analysis, what does the term "anomaly" typically refer to? (CO5,K1) 1

(a) 2

(b) Predicted connections

(c) Unexpected patterns or irregularities

(d) Sentiment trends

1-j. In link prediction models, what does the term "triadic closure" refer to?(CO5,K1) 1

(a) 3

(b) Identifying influential nodes

(c) Analyzing emotional tone

(d) Predicting connections based on shared neighbors

2. Attempt all parts:-

2.a. Explain the concept of emotion analysis in text mining.(CO1,K2) 2

2.b. List some common text preprocessing techniques.(CO1,K1) 2

2.c. Which analytics tool stands out for tracking user behaviour? (CO3,K1) 2

2.d.	How are graphs and matrices used to represent social networks?(CO4,K2)	2
2.e.	Can link analysis techniques be applied to identify relationships and dependencies within a body of text?(CO5,K2)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Explain public and private data in social media analytic with real life example. (CO1,K2)	6
3-b.	Explain K-means clustering algorithm with example. (CO1,K2)	6
3-c.	Discuss the role of overfitting in predictive modeling? (CO2,K2)	6
3-d.	Explain, how does the length of a text affect sentiment prediction accuracy? (CO2,K2)	6
3.e.	Describe the purpose of clickstream data.(CO3,K2)	6
3.f.	Explain the importance of user-generated content in social media analytics.(CO4,K2)	6
3.g.	How does link analysis contribute to the understanding of network evolution over time? (CO5, K2)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Discuss the concept of named entity recognition (NER) and its role in extracting valuable information from texts. (CO1,K2)	10
4-b.	Explain, how are Conditional Random Fields (CRFs) utilized in information extraction, and what advantages do they offer over other models? (CO1,K2)	10

5. Answer any one of the following:-

5-a.	List the challenges faced in clustering short and noisy texts, such as tweets or product reviews and Discuss innovative approaches and algorithms designed to improve the accuracy of clustering for these specific types of data. (CO2,K2)	10
5-b.	Explain Latent Dirichlet Allocation (LDA) as a probabilistic model for topic detection. How does it work, and what are its limitations? (CO2,K2)	10

6. Answer any one of the following:-

6-a.	Describe the impact of A/B testing on website optimization, with its key components? (CO3, K2)	10
6-b.	Analyze the impact of ranking algorithms on web traffic and website visibility, discussing their methodologies and effects on search engine results.(CO3,K4)	10

7. Answer any one of the following:-

7-a.	Compare and contrast different types of graph structures, including directed, undirected, weighted, and bipartite graphs. Provide examples of social media data that can be represented using each type of graph.(CO4,K4)	10
7-b.	Describe the role of data visualization in social media analytics. Explain how visual representations such as charts, graphs, and heat maps can enhance the	10

understanding of social media data patterns. (CO4,K2)

8. Answer any <u>one</u> of the following:-

8-a.	Illustrate, how do multiplex and multilayer networks contribute to understanding complex relationships within evolving networks? (CO5,K4)	10
8-b.	In social media, Explain the role of natural language processing (NLP) in extracting information about identity from posts? (CO5,K2)	10