

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VI - THEORY EXAMINATION (2025 - 2026)

Subject: Social 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:-

- (1.1) Tokenization in NLP refers to:(CO1,K1) 1
- (a) Removing punctuation from text
 - (b) Splitting text into individual words or sentences
 - (c) Converting words to their root form
 - (d) Encoding text into numbers
- (1.2) A unigram is an N-gram where N equals:(CO1,K1) 1
- (a) 0
 - (b) 1
 - (c) 2
 - (d) 3
- (1.3) Web content mining focuses on extracting useful information from:(CO2,K1) 1
- (a) Server access logs
 - (b) The actual content of web pages such as text and images
 - (c) Social media follower counts
 - (d) Domain registration records
- (1.4) The component of a search engine that traverses and downloads web pages automatically is called a:(CO2,K2) 1
- (a) Indexer
 - (b) Ranker
 - (c) Web crawler
 - (d) Query parser

- (1.5) Among which of the following metrics can be used to measure the size of a social network:(CO3,K2) 1
- (a) Number of users
 - (b) Number of connections between users
 - (c) Number of interactions between users
 - (d) All
- (1.6) Information diffusion in social media analytics is:(CO3,K1) 1
- (a) The spread of information from one social media platform to another
 - (b) The spread of information within a social media network
 - (c) The analysis of social media data to predict user behavior
 - (d) The mathematical formula for analyzing social media data
- (1.7) Which library is mainly used for numerical computation in Python?(CO4,K1) 1
- (a) NumPy
 - (b) Matplotlib
 - (c) Pandas
 - (d) NLTK
- (1.8) Which method groups similar text documents without labels?(CO4,K1) 1
- (a) Classification
 - (b) Clustering
 - (c) Ranking
 - (d) Parsing
- (1.9) Seasonal trend occurs due to:(CO5,K1) 1
- (a) Random errors
 - (b) Periodic patterns
 - (c) Hardware failure
 - (d) Data loss
- (1.10) Trend analysis is widely used in:(CO5,K2) 1
- (a) Gaming
 - (b) Finance
 - (c) Typing
 - (d) Printing

2. Attempt all parts:-

- (2.1) Write the significance of TF-IDF in text mining.(CO1,K2) 2
- (2.2) Define web mining and list its three main categories.(CO2,K1) 2
- (2.3) Explain why node is essential in social graph.(CO3,K2) 2
- (2.4) State the applications of text summarization in social media analytics.(CO4,K2) 2
- (2.5) Define trend analysis.(CO5,K1) 2

SECTION-B

30

Answer any one of the following:-

- (3.1) Explain tokenization, stemming, and stop word removal with examples. Why are 6

	these steps important in text preprocessing?(CO1,K2)	
(3.2)	Compare and contrast Text Clustering and Text Classification. When would you choose one over the other?(CO1,K2)	6
	Answer any one of the following:-	
(3.3)	Explain the working of a web crawler and the challenges faced during large-scale web crawling.(CO2,K2)	6
(3.4)	Given posting lists: P=[2,4,6,8,10], Q=[1,4,7,8,9], R=[2,4,5,8,11]. Solve: (a) P AND Q (b) Q OR R (c) P AND NOT Q (d) P AND Q AND R.(CO2,K3)	6
	Answer any one of the following:-	
(3.5)	Describe social network analysis model in social media analytics.(CO3,K2)	6
(3.6)	Explain how can influence and homophily be leveraged in marketing and advertising.(CO3,K2)	6
	Answer any one of the following:-	
(3.7)	Explain N-gram frequency count and its application in phrase identification.(CO4,K3)	6
(3.8)	Describe the major challenges involved in text summarization.(CO4,K2)	6
	Answer any one of the following:-	
(3.9)	Compare linear and non-linear trend with examples.(CO5,K2)	6
(3.10)	Write a note on emerging trends in text mining.(CO5,K2)	6
	SECTION-C	50
	4. Answer any <u>one</u> of the following:-	
(4.1)	Discuss the complete preprocessing pipeline used in text mining and sentiment analysis with suitable examples.(CO1,K2)	10
(4.2)	Design a complete sentiment mining system for analysing customer feedback. Discuss data collection, preprocessing, modelling, and output with architecture.(CO1,K4)	10
	5. Answer any <u>one</u> of the following:-	
(5.1)	Describe Social Graphs in detail including Interaction, Latent, and Following graphs with structure, properties, and applications.(CO2,K2)	10
(5.2)	Explain the role of machine learning in extracting knowledge from web data. Given a dataset of user clicks (10 positive, 5 negative), compute accuracy, precision, and recall for a classifier with predictions: TP=8, FP=2, FN=2, TN=3.(CO2,K2)	10
	6. Answer any <u>one</u> of the following:-	
(6.1)	Analyze the major challenges in social media mining such as data volume, velocity, variety, and veracity. Suggest possible solutions to overcome these challenges.(CO3,K4)	10
(6.2)	Compare different types of social graphs and analyze which type is most suitable for modeling friendship networks.(CO3,K3)	10
	7. Answer any <u>one</u> of the following:-	
(7.1)	Explain text summarization in detail and discuss its applications in text analytics.(CO4,K2)	10
(7.2)	Discuss the role of text extraction, classification, and clustering in building	10

summarization systems.(CO4,K3)

8. Answer any one of the following:-

- | | | |
|-------|---|----|
| (8.1) | Discuss the different categories of trend analysis with appropriate graphical illustrations.(CO5,K2) | 10 |
| (8.2) | Explain the significance of social media analytics in shaping digital marketing strategies with examples.(CO5,K4) | 10 |

REG_JAN_JUNE_2026