

Roll No.....



**Plot No. 2, Knowledge Park-III, Greater Noida (U.P.) –201306
POST GRADUATE DIPLOMA IN MANAGEMENT (2024-26)
MID TERM QUIZ EXAMINATION (TERM -V)**

Subject Name:Text and Sentiment Analytics

Time: 00.30 hrs

Sub. Code: PGIT 52

Max Marks: 20

Note:

- 1. Writing anything except Roll Number on Quiz paper will be deemed as an act of indulging in unfair means and action shall be taken as per rules.**
- 2. There is no negative marking for wrong answers.**
- 3. Tick marks the correct answer.**

Attempt all questions. All questions are compulsory.

40×0.5 = 20 Marks

- 1. What is the primary purpose of Natural Language Processing (NLP)?**
 - (a) To convert numerical data into visual representations.
 - (b) To enable computers to understand, analyze, and process human languages.
 - (c) To generate binary code for machine execution.
 - (d) To analyze the grammar of computer programming languages.

Ans: b.

- 2. Which NLP application is used to generate captions for video content in real-time? (a)**
 - Language Translation
 - (b) Sentiment Analysis
 - (c) Autogenerated Captions
 - (d) Text Classification

Ans: d.

- 3. What is the main task of Semantic Analysis in NLP?**
 - (a) Dividing text into sentences and words.
 - (b) Checking the grammar of the sentences.
 - (c) Assessing the meaningfulness of the input text.
 - (d) Analyzing the relationship between successive sentences.

Ans: a.

- 4. What is a "corpus" in the context of text processing?**
 - (a) A single document.
 - (b) A special character in a sentence.
 - (c) A collection of written text from multiple documents.
 - (d) A word that has been stemmed.

Ans: c

- 5. The process of dividing sentences into words, numbers, and special characters is called: (a)**
 - Segmentation
 - (b) Tokenization

- (c) Stemming
- (d) Normalization

Ans: b

6. Which of the following is a key characteristic of "stop words"?
- (a) They add the most value to the corpus.
 - (b) They have the lowest occurrence in the text.
 - (c) They are frequently occurring words that do not add significant value.
 - (d) They are always meaningful and essential for context.

Ans: c

7. What is the difference between stemming and lemmatization?
- (a) Stemming is slower than lemmatization.
 - (b) Stemming produces a meaningful root word, while lemmatization does not.
 - (c) Stemming removes affixes, and the resulting word may not be meaningful, while lemmatization ensures the root word is meaningful.
 - (d) They are the same process with different names.

Ans: c

8. In the Bag of Words model, what does the "bag" symbolize?
- (a) That the sequence of sentences is preserved.
 - (b) That the unique words are stored in a bag-like structure.
 - (c) That the sequence of words or tokens does not matter.
 - (d) That only the most important words are kept.

Ans: d.

9. What does TFIDF stand for?
- (a) Term Frequency-Integrated Document Function
 - (b) Text Feature-Inverse Document Frequency
 - (c) Term Frequency-Inverse Document Frequency
 - (d) Total Frequency-In-Document Function

Ans: c

10. Which of the following is a TFIDF application?
- (a) Image Recognition
 - (b) Document Classification
 - (c) Voice Synthesis
 - (d) Chatbot Development

Ans: b

11. The conversion of a word like "healed" to its base form "heal" is an example of which text processing technique?
- (a) Tokenization
 - (b) Lemmatization or Stemming
 - (c) Sentence Segmentation
 - (d) Stop word removal

Ans: b

12. What is the key concept illustrated by the sentence "His face turned red after he found out that he took the wrong bag"?
- (a) The importance of grammar.
 - (b) The need for tokenization.
 - (c) The significance of context in understanding meaning.
 - (d) The process of stemming.

Ans: c

13. What is the core idea of a Bag of Words model?
- (a) Counting the number of unique words in a document.
 - (b) Creating a document vector for each document.

- (c) Preserving the grammatical structure of sentences.
- (d) Representing text as a collection of words, disregarding their order.

Ans: d.

14. Which step of text normalization ensures that "Hello," "hello," and "HELLO" are treated as the same word?

- (a) Tokenization
- (b) Stop word removal
- (c) Stemming
- (d) Converting text to a common case

Ans: d.

15. A word with a high TFIDF value is likely to have:

- (a) High term frequency and high document frequency.
- (b) Low term frequency and low document frequency.
- (c) High term frequency and low document frequency.
- (d) Low term frequency and high document frequency.

Ans: a

16. A document vector is created in which step of the Bag of Words algorithm? (a)

- Text Processing
- (b) Creating a Dictionary
- (c) Creating document vectors
- (d) TFIDF calculation

Ans: d

17. The process of converting the word "studying" to its base form "study" is an example of which technique?

- (a) Stemming
- (b) Lemmatization
- (c) Tokenization
- (d) Both (a) and (b)

Ans: d.

18. What is the primary function of keyword extraction?

- (a) To translate a text into different languages.
- (b) To classify a document into a category.
- (c) To automatically identify and extract the most important words or phrases from a text. (d)
- To check the grammatical structure of a sentence.

Ans: c

19. What is a "lemma" in the context of lemmatization?

- (a) A word with no meaning.
- (b) A word that has not been normalized.
- (c) A meaningful word that is the root of another word.
- (d) A word that is a stop word.

Ans: c

20. What is the main advantage of lemmatization over stemming?

- (a) It is faster.
- (b) It is less complex.
- (c) It always produces meaningful words.
- (d) It is better for identifying stop words.

Ans: c

21. _____ device helps to communicate with humans and abilities to make humans lives easier.

- a. Google Assistant
- b. Cortana
- c. Siri

d. All of the above

Ans: d. All of the above

22. _____ by collecting data from various reliable and authentic sources. a. Data Acquisition

b. Database

c. Data Mining

d. None of the above

Ans: a. Data Acquisition

23. Once the textual data has been collected, it needs to be processed and cleaned so that an easier version can be sent to the machine. This is known as _____.

a. Data Acquisition

b. Data Exploration

c. Data Mining

d. None of the above

Ans: b. Data Exploration

24. Once the text has been normalized, it is then fed to an NLP based AI model. Note that in NLP, modelling requires data pre-processing only after which the data is fed to the machine. a. Data Acquisition

b. Modelling

c. Data Mining

d. None of the above

Ans: b. Modelling

25. _____ helps in cleaning up the textual data in such a way that it comes down to a level where its complexity is lower than the actual data.

a. Speech Normalization

b. Text Normalization

c. Visual Normalization

d. None of the above

Ans: b. Text Normalization

26. _____ the whole corpus is divided into sentences. Each sentence is taken as a different data so now the whole corpus gets reduced to sentences.

a. Sentence Normalization

b. Sentence Segmentation

c. Sentence Tokenization

d. All of the above

Ans: b. Sentence Segmentation

27. _____ are the words which occur very frequently in the corpus but do not add any value to it.

a. Tokens

b. Words

c. Stopwords

d. None of the above

Ans: c. Stopwords

28. Stopwords are the words which occur very frequently in the corpus but do not add any value to it. for example_____.

a. Grammatical words

b. Simple words

c. Complex words

d. All of the above

Ans: a. Grammatical words

29. Stemming and lemmatization both are _____ processes.

a. Same process

b. Alternative process

c. Other process

d. All of the above

Ans: b. Alternative process

30. _____ identify each document in the corpus, find out how many times the word from the

unique list of words has occurred.

- a. Text Normalization
- b. Create Dictionary
- c. Document Vectors
- d. All of the above

Ans: c. Document Vectors

31. TFIDF stands for _____.

- a. Team Frequency and Inverse Document Frequency
- b. Term Frequency and Inverse Document Frequency
- c. Top Frequency and Inverse Document Frequency
- d. Table Frequency and Inverse Document Frequency

Ans: b. Term Frequency and Inverse Document Frequency

32. One of the applications of Natural Language Processing is relevant when used to provide an overview of a news item or blog post, while avoiding redundancy from multiple sources and maximizing the diversity of content obtained. Identify the application from the following a. Sentiment Analysis

- b. Virtual Assistants
- c. Text classification
- d. Automatic Summarization

Ans: d. Automatic Summarization

33. Companies use _____ application of NLP , to identify opinions and feelings/emotions online to help them understand what customers think about their products and services. a. Sentiment Analysis

- b. Automatic Summarization
- c. Text classification
- d. Virtual Assistants

Ans: a. Sentiment Analysis

34. Under _____, every word, number and special character is considered separately and each of them is now a separate token.

- a. Sentence Segmentation
- b. Removing Stopwords, Special Characters and Numbers
- c. Converting text to a common case
- d. Tokenisation

Ans: d. Tokenisation

35. During Text Normalisation, which step will come after removing Stopwords, Special Characters and Numbers.

- a. Converting text to a common case.
- b. Stemming
- c. Lemmatization
- d. Tokenisation

Ans: a. Converting text to a common case.

36. During Text Normalisation, when we convert the whole text into a similar case, we prefer

-
- a. Upper Case
 - b. Lower Case
 - c. Title Case
 - d. Mixed Case

Ans: b. Lower Case

Q37. Which of the following is not an example of smart bot?

- a. Siri
- b. Google Assistant
- c. Cortana
- d. Bots which are deployed in the customer care section and answer the basic queries.

Ans: d. Bots which are deployed in the customer care section and answer the basic queries.

38 Assertion (A): The Bag of Words model is a method for text representation that retains the grammatical structure of sentences.

Reason (R): The model disregards the sequence and grammar of words, focusing only on their

frequency.

- a) Both A and R are true, and R is the correct explanation of A.
- b) Both A and R are true, but R is not the correct explanation of A
- c) c) A is true, but R is false.
- d) d) A is false, but R is true.

Ans: d) A is false, but R is true.

39. Assertion (A): The process of breaking a large body of text into smaller units like words and sentences is called Tokenization.

Reason (R): Tokenization is a key step in Text Normalization, which simplifies textual data.

- a) Both A and R are true, and R is the correct explanation of A.
- b) Both A and R are true, but R is not the correct explanation of A
- c) c) A is true, but R is false.
- d) d) A is false, but R is true.

Ans: b) Both A and R are true, but R is not the correct explanation of A.

40. Assertion (A): Chatbots use NLP to simulate human conversation.

Reason (R): NLP enables computers to understand and process human language, which is essential for a chatbot to function effectively.

- a) Both A and R are true, and R is the correct explanation of A.
- b) Both A and R are true, but R is not the correct explanation of A
- c) c) A is true, but R is false.
- d) d) A is false, but R is true.

Ans:a) Both A and R are true, and R is the correct explanation of A.