

# MCQ ON TEXT MINING

Multiple-choice questions (MCQs) related to Text Mining in Predictive Modelling:

1. Text mining is the process of:

- a) Extracting information from images
- b) Analyzing structured data
- c) Analyzing unstructured text data
- d) Identifying relationships between variables

Correct Answer: c) Analyzing unstructured text data

2. Which of the following tasks can be performed using text mining in predictive modeling?

- a) Sentiment analysis
- b) Clustering
- c) Regression
- d) All of the above

Correct Answer: d) All of the above

3. In text mining, term frequency-inverse document frequency (TF-IDF) is used to:

- a) Remove stop words from the text
- b) Measure the importance of words in a document compared to a collection of documents
- c) Extract named entities from the text
- d) Convert text data into numerical vectors

Correct Answer: b) Measure the importance of words in a document compared to a collection of documents

4. Which of the following is an example of a text mining technique used to categorize documents into predefined topics?

- a) Sentiment analysis

b) Named entity recognition

c) Text classification

d) Word cloud generation

Correct Answer: c) Text classification

5. The bag-of-words (BoW) model in text mining represents a document as:

a) A collection of images

b) A sequence of sentences

c) A set of words disregarding grammar and word order

d) A structured table with rows and columns

Correct Answer: c) A set of words disregarding grammar and word order

6. Which of the following techniques can be used to handle the high dimensionality of text data in predictive modeling?

a) Principal Component Analysis (PCA)

b) Feature scaling

c) Decision trees

d) K-means clustering

Correct Answer: a) Principal Component Analysis (PCA)

7. In sentiment analysis, the goal is to:

a) Categorize documents into predefined topics

b) Extract named entities from the text

c) Determine the sentiment or emotion expressed in the text

d) Convert text data into numerical vectors

Correct Answer: c) Determine the sentiment or emotion expressed in the text

8. Which of the following algorithms can be used for text classification tasks in predictive modeling?

a) Decision tree

b) k-Nearest Neighbors (KNN)

c) Logistic Regression

d) All of the above

Correct Answer: d) All of the above

9. Named Entity Recognition (NER) is a text mining technique used to:

a) Identify important terms in the text

b) Extract features from text data

c) Recognize named entities such as names, locations, and dates in the text

d) Classify documents into predefined categories

Correct Answer: c) Recognize named entities such as names, locations, and dates in the text

10. Which of the following libraries is commonly used in Python for text mining tasks?

a) Pandas

b) NumPy

c) Scikit-learn

d) NLTK (Natural Language Toolkit)

Correct Answer: d) NLTK (Natural Language Toolkit)

11. Text mining can be used in predictive modeling to:

a) Analyze numerical data

b) Analyze images and videos

c) Analyze unstructured text data

d) Analyze data in a tabular format

Correct Answer: c) Analyze unstructured text data

12. Which of the following is a preprocessing step in text mining?

a) Feature selection

- b) Scaling the data
- c) Tokenization
- d) Model evaluation

Correct Answer: c) Tokenization

13. In text mining, tokenization refers to:

- a) Removing stop words from the text
- b) Converting text into numerical vectors
- c) Splitting the text into individual words or tokens
- d) Categorizing documents into predefined topics

Correct Answer: c) Splitting the text into individual words or tokens

14. Which of the following techniques can be used to handle the curse of dimensionality in text mining?

- a) Principal Component Analysis (PCA)
- b) Decision tree pruning
- c) Feature scaling
- d) Word cloud generation

Correct Answer: a) Principal Component Analysis (PCA)

15. The process of converting text data into numerical vectors is known as:

- a) Tokenization
- b) Bag-of-words representation
- c) Feature scaling
- d) Named Entity Recognition (NER)

Correct Answer: b) Bag-of-words representation

16. In text mining, the term "stop words" refers to:

- a) Words that carry little or no meaning and are often removed during preprocessing

- b) Words that have strong emotional sentiment
- c) Words that are misspelled in the text
- d) Words that are used to represent numbers in the text

Correct Answer: a) Words that carry little or no meaning and are often removed during preprocessing

17. Which of the following text mining techniques can be used to determine the most frequent words in a document?

- a) Sentiment analysis
- b) Word cloud generation
- c) Named Entity Recognition (NER)
- d) Text classification

Correct Answer: b) Word cloud generation

18. In text mining, the goal of topic modeling is to:

- a) Determine the sentiment of the text
- b) Categorize the text into predefined topics
- c) Identify important words in the text
- d) Convert text data into numerical vectors

Correct Answer: b) Categorize the text into predefined topics

19. Which of the following is an unsupervised text mining technique used to find similar words or documents in a collection?

- a) Word cloud generation
- b) Named Entity Recognition (NER)
- c) Clustering
- d) Sentiment analysis

Correct Answer: c) Clustering

20. In text mining, what does the term "stemming" refer to?

- a) Converting text into numerical vectors
- b) Removing stop words from the text
- c) Reducing words to their root or base form
- d) Tokenization of the text

Correct Answer: c) Reducing words to their root or base form

21. Which of the following techniques is used to handle text data that is highly skewed or imbalanced?

- a) Decision tree pruning
- b) Feature scaling
- c) Oversampling the minority class
- d) Principal Component Analysis (PCA)

Correct Answer: c) Oversampling the minority class

22. In text mining, the process of converting words to their root form is known as:

- a) Lemmatization
- b) Tokenization
- c) Bag-of-words representation
- d) Named Entity Recognition (NER)

Correct Answer: a) Lemmatization

23. Which of the following metrics is commonly used to evaluate the performance of a text

classification model?

- a) R-squared value
- b) F1 score
- c) Mean Absolute Error (MAE)
- d) Root Mean Squared Error (RMSE)

Correct Answer: b) F1 score

24. Which of the following statements about text mining is true?

- a) Text mining can only be used for analyzing numerical data.
- b) Text mining involves converting structured data into unstructured text data.
- c) Text mining techniques can be applied to analyze both structured and unstructured text data.
- d) Text mining is primarily used for generating word clouds.

Correct Answer: c) Text mining techniques can be applied to analyze both structured and unstructured text data.

25. Which of the following is NOT a common application of text mining in predictive modeling?

- a) Sentiment analysis of customer reviews
- b) Spam email detection
- c) Predicting stock market prices
- d) Topic modeling of news articles

Correct Answer: c) Predicting stock market prices

26. Text mining is most commonly used in which type of predictive modeling task?

- a) Image recognition
- b) Time series forecasting
- c) Natural language processing
- d) Financial modeling

Correct Answer: c) Natural language processing

27. In text mining, the term "n-grams" refers to:

- a) The number of documents in a collection
- b) The most frequent words in a document
- c) A sequence of n consecutive words in a text
- d) The number of clusters in a clustering algorithm

Correct Answer: c) A sequence of n consecutive words in a text

28. Which of the following text mining techniques can be used to identify the language of a text document?

- a) Named Entity Recognition (NER)
- b) Lemmatization
- c) Language detection
- d) Sentiment analysis

Correct Answer: c) Language detection

29. Which of the following is an example of an application of sentiment analysis in predictive modeling?

- a) Identifying topics in customer reviews
- b) Detecting spam emails
- c) Predicting stock prices
- d) Analyzing customer feedback for product improvement

Correct Answer: d) Analyzing customer feedback for product improvement

30. Which of the following is an example of a feature engineering technique commonly used in text mining?

- a) Feature scaling
- b) Principal Component Analysis (PCA)
- c) Word embedding
- d) Decision tree pruning

Correct Answer: c) Word embedding

31. The process of converting words into numerical vectors in text mining is known as:

- a) Lemmatization
- b) Tokenization
- c) Word cloud generation



d) Word embedding

Correct Answer: d) Word embedding

32. In text mining, the process of grouping similar words together based on their meaning is known as:

a) Lemmatization

b) Tokenization

c) Word cloud generation

d) Word clustering

Correct Answer: d) Word clustering

33. Which of the following techniques can be used to convert text data into numerical vectors while preserving the semantic meaning of words?

a) Bag-of-words (BoW) representation

b) Word cloud generation

c) Word embedding

d) Lemmatization

Correct Answer: c) Word embedding

34. Which of the following text mining techniques can be used to identify and extract entities such as names, locations, and dates from a text document?

a) Sentiment analysis

b) Named Entity Recognition (NER)

c) Text classification

d) Word cloud generation

Correct Answer: b) Named Entity Recognition (NER)

35. Which of the following is an example of a text mining application in the healthcare industry?

a) Identifying topics in customer reviews for a restaurant chain

b) Sentiment analysis of social media posts about a new movie release

- c) Detecting fraudulent activities in financial transactions
- d) Extracting medical terms and diagnoses from patient records

Correct Answer: d) Extracting medical terms and diagnoses from patient records

36. Which of the following text mining techniques can be used to identify the most frequently occurring words in a collection of documents?

- a) Named Entity Recognition (NER)
- b) Text classification
- c) Word cloud generation
- d) Sentiment analysis

Correct Answer: c) Word cloud generation

37. Which of the following is an example of a text mining technique used to identify the sentiment or emotion expressed in a text document?

- a) Named Entity Recognition (NER)
- b) Text classification
- c) Sentiment analysis
- d) Word cloud generation

Correct Answer: c) Sentiment analysis

38. Which of the following techniques is commonly used to preprocess text data in text mining?

- a) Feature scaling
- b) Principal Component Analysis (PCA)
- c) Tokenization
- d) Clustering

Correct Answer: c) Tokenization

39. In text mining, what does the term "n-grams" refer to?

- a) The number of documents in a collection

- b) The most frequent words in a document
- c) A sequence of n consecutive words in a text
- d) The number of clusters in a clustering algorithm

Correct Answer: c) A sequence of n consecutive words in a text

40. Which of the following text mining techniques can be used to identify the language of a text document?

- a) Named Entity Recognition (NER)
- b) Lemmatization
- c) Language detection
- d) Sentiment analysis

Correct Answer: c) Language detection

41. Which of the following is an example of an application of sentiment analysis in predictive modeling?

- a) Identifying topics in customer reviews
- b) Detecting spam emails
- c) Predicting stock prices
- d) Analyzing customer feedback for product improvement

Correct Answer: d) Analyzing customer feedback for product improvement

42. Which of the following is an example of a feature engineering technique commonly used in text mining?

- a) Feature scaling
- b) Principal Component Analysis (PCA)
- c) Word embedding
- d) Decision tree pruning

Correct Answer: c) Word embedding

43. The process of converting words into numerical vectors in text mining is known as:

- a) Lemmatization
- b) Tokenization
- c) Word cloud generation
- d) Word embedding

Correct Answer: d) Word embedding

44. In text mining, the process of grouping similar words together based on their meaning is known as:

- a) Lemmatization
- b) Tokenization
- c) Word cloud generation
- d) Word clustering

Correct Answer: d) Word clustering

45. Which of the following techniques can be used to convert text data into numerical vectors while preserving the semantic meaning of words?

- a) Bag-of-words (BoW) representation
- b) Word cloud generation
- c) Word embedding
- d) Lemmatization

Correct Answer: c) Word embedding

46. Which of the following text mining techniques can be used to identify and extract entities such as names, locations, and dates from a text document?

- a) Sentiment analysis
- b) Named Entity Recognition (NER)
- c) Text classification
- d) Word cloud generation

Correct Answer: b) Named Entity Recognition (NER)

47. Which of the following is an example of a text mining application in the healthcare industry?

- a) Identifying topics in customer reviews for a restaurant chain
- b) Sentiment analysis of social media posts about a new movie release
- c) Detecting fraudulent activities in financial transactions
- d) Extracting medical terms and diagnoses from patient records

Correct Answer: d) Extracting medical terms and diagnoses from patient records

48. Which of the following text mining techniques can be used to identify the most frequently occurring words in a collection of documents?

- a) Named Entity Recognition (NER)
- b) Text classification
- c) Word cloud generation
- d) Sentiment analysis

Correct Answer: c) Word cloud generation

49. Which of the following is an example of a text mining technique used to identify the sentiment or emotion expressed in a text document?

- a) Named Entity Recognition (NER)
- b) Text classification
- c) Sentiment analysis
- d) Word cloud generation

Correct Answer: c) Sentiment analysis

50. Which of the following techniques is commonly used to preprocess text data in text mining?

- a) Feature scaling
- b) Principal Component Analysis (PCA)
- c) Tokenization

d) Clustering

Correct Answer: c) Tokenization

51. What is the goal of sentiment analysis in text mining?

a) To determine the language of a text document

b) To identify the most frequent words in a document

c) To classify a text document into predefined categories

d) To determine the sentiment or emotion expressed in the text

Correct Answer: d) To determine the sentiment or emotion expressed in the text

52. Which of the following is an example of a supervised learning algorithm used in text mining for sentiment analysis?

a) k-Nearest Neighbors (KNN)

b) k-Means clustering

c) Decision tree

d) Principal Component Analysis (PCA)

Correct Answer: c) Decision tree

53. In text mining, which technique can be used to convert words into numerical vectors while considering their frequency and importance in a document?

a) Lemmatization

b) Word cloud generation

c) Term frequency-inverse document frequency (TF-IDF)

d) Tokenization

Correct Answer: c) Term frequency-inverse document frequency (TF-IDF)

54. In text mining, the process of identifying and extracting entities such as names, locations, and dates from a text document is known as:

a) Sentiment analysis

b) Named Entity Recognition (NER)

- c) Clustering
- d) Word cloud generation

Correct Answer: b) Named Entity Recognition (NER)

55. Which of the following statements about text mining and natural language processing is true?

- a) Text mining is a subset of natural language processing.
- b) Natural language processing and text mining are the same thing.
- c) Text mining is used to analyze structured data, while natural language processing is used to analyze unstructured text data.
- d) Text mining and natural language processing are two unrelated fields.

Correct Answer: a) Text mining is a subset of natural language processing.

56. Which of the following is an example of an unsupervised learning technique used in text mining for clustering similar documents?

- a) Support Vector Machine (SVM)
- b) Logistic Regression
- c) k-Means clustering
- d) Decision tree

Correct Answer: c) k-Means clustering

57. In text mining, which technique can be used to convert words into numerical vectors while considering their frequency and importance in a document?

- a) Lemmatization
- b) Word cloud generation
- c) Term frequency-inverse document frequency (TF-IDF)
- d) Tokenization

Correct Answer: c) Term frequency-inverse document frequency (TF-IDF)

58. In text mining, the process of identifying and extracting entities such as names, locations, and dates from a text document is known as:

- a) Sentiment analysis
- b) Named Entity Recognition (NER)
- c) Clustering
- d) Word cloud generation

Correct Answer: b) Named Entity Recognition (NER)

59. Which of the following statements about text mining and natural language processing is true?

- a) Text mining is a subset of natural language processing.
- b) Natural language processing and text mining are the same thing.
- c) Text mining is used to analyze structured data, while natural language processing is used to analyze unstructured text data.
- d) Text mining and natural language processing are two unrelated fields.

Correct Answer: a) Text mining is a subset of natural language processing.

60. Which of the following is an example of an unsupervised learning technique used in text mining for clustering similar documents?

- a) Support Vector Machine (SVM)
- b) Logistic Regression
- c) k-Means clustering
- d) Decision tree

Correct Answer: c) k-Means clustering