

Part	sub/obj	Marks	Question	Answer Option 1	Answer Option 2	Answer Option 3	Answer Option 4	Correct Answer(A/B/C/D)	CO	Bloom's Taxonomy Level
A	obj	1	The Bag-of-Words approach_____	keeps word order, keeps word multiplicity	keeps word order, disregards word multiplicity	disregards word order, disregards word multiplicity	disregards word order, keeps word multiplicity	D	CO1	L3
A	obj	1	Machine Translation is that convert_____	Any human language to English	Human language to machine language	One human language to another	Machine language to human language	C	CO1	L3
A	obj	1	What kind of charts are used in visualizations of results?	Scatterplot	Barchart	Windowplot	Both A & B	D	CO1	L3
A	obj	1	The primary use of data cleaning is	Removing the noisy data	Correction of the data inconsistencies	Transformations for correcting the wrong data	All of the above	D	CO1	L3
A	obj	1	Which are python libraries used in NLP?	pandas	spacy	nlTK	All of above	D	CO1	L3
A	obj	1	What is the field of Natural Language Processing (NLP)?	Artificial Intelligence	Linguistics	Computer Science	All	D	CO1	L3
A	obj	1	_____ is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights.	Data mining	Text mining	File mining	Deep mining	B	CO1	L3
A	obj	1	Text mining is being used by large media companies, to clarify information and to provide readers with greater search experiences,	TRUE	FALSE			A	CO1	L3
A	obj	1	The initial steps concerned in the process of knowledge discovery is	Data Selection	Data Integration	Data Cleaning	Data Transformation	C	CO1	L3
A	obj	1	Typical text mining tasks include?	text categorization	text clustering	entity relation modeling	All of the above	D	CO1	L3
A	obj	1	KDD stands for	Knowledge Discovery Database	Knowledge Definition Data	Knowledge Data Discovery	Knowledge Data Definition	A	CO1	L3
A	obj	1	What is the right order for a text classification model components?	Text cleaning -> Text annotation -> Gradient descent -> Model tuning	Text cleaning -> Text annotation -> Text to predictors -> Gradient	Text cleaning -> Gradient descent -> Model tuning -> Text	Text cleaning -> Text annotation -> Model tuning -> Text to		CO1	L3
A	obj	1	What is the oUtput : from nltk.tokenize import sent_tokenize as ST text1 = "Hello everyone. Welcome to Javatpoint. We are	Hello everyone.', 'Welcome to Javatpoint.', 'We are studying NLP	Hello everyone.' 'Welcome to Javatpoint.' 'We are studying NLP Tutorial'			A	CO1	L3
A	obj	1	Which of the below sentence correct with respect to Text Mining & Text Analytics	Text Analytics and Text mining both are same	Text mining comes before text analytics	Text analytics used for preprocessing of text and text mining	None of the above	C	CO1	L3
A	obj	1	What is meant by TF (Term Frequency) ²	(Number of times term T appears in the particular row)	(Number of times term T appears in the particular row) / (number of terms in	(number of terms in that column) / (Number of times	(number of terms in that row) / (Number of times term T appears	B	CO1	L3
A	obj	1	Sentiment analysis also known as opinion mining	TRUE	FALSE			A	CO1	L3
A	obj	1	Text Mining can be used in	Detecting spam model	Predicting stock Movements	Both A and B	None of the above	C	CO1	L3
A	obj	1	Which of the following are the correct statements regarding similarities between text mining and data mining?	Analytics tasks are modeled using similar machine learning	Both are about learning from samples of recorded experiences or examples	Both are about finding valuable patterns in data	All	D	CO1	L3

Part	sub/obj	Marks	Question	Answer Option 1	Answer Option 2	Answer Option 3	Answer Option 4	Correct Answer(A/B/C/D)	CO	Bloom's Taxonomy Level
A	obj	1	Which of the following are typical text mining problems?	Document Classification	Clustering and Organizing Documents	Information Extraction	Information Retrieval	D	C01	L3
A	obj	1	Text Mining is a field related to	Computer Science	Artificial Intelligence	Natural Language Processing (NLP)	All of the above		C01	L3
A	obj	1	What are the possible features of a text corpus Count of word in a document Boolean feature – presence of word in a document. Vector notation of word Part of Speech Tag Basic Dependency	12345	1234	123	12	A	C02	L3
A	obj	1	Which step is the process of breaking down documents into smaller units of analysis?	Ngrams	Tokenization	Stopwords	Corpus	B	C02	L3
A	obj	1	Which is a model of measuring the incidence of known words?	A low weight in TF-IDF	A high weight in TF-IDF	A corpus	A bag of words	D	C02	L3
A	obj	1	Which is a high term frequency and low document frequency?	A low weight in TF-IDF	A high weight in TF-IDF	A corpus	A bag of words	B	C02	L3
A	obj	1	Which are common words usually removed in an NLP analysis?	NLTK	NLP-LIB	Scikit-Learn	PyNLP	A	C02	L3
A	obj	1	Which of the following Affective States does Sentiment Analysis mostly focus on?	Personality Traits	Mood	Emotion	Attitudes	D	C02	L3
A	obj	1	Which is a collection of documents?	Corpus	Ngrams	Tokenization	Stopwords	A	C02	L3
A	obj	1	Consider the following given sentences. Match the lexical relations between the first word (w1) to the second word (w2) i.e. w1 is a of w2. * Invention of the wheel is one of the	1-iii, 2-ii, 3-iv, 4-i	1-ii, 2-iii, 3-i, 4-iv	1-ii, 2-iii, 3-iv, 4-i	1-i, 2-ii, 3-iii, 4-iv	B	C02	L3
A	obj	1	_____ technique looks at the meaning of the word.	Stemming	Lemmatization	Stop word identification	Morphological Analysis	B	C02	L3
A	obj	1	In Sentiment analysis	List the topics that a document deals with	Assess the emotional content of a document	Compress a document as much as possible without	Given a question in natural language, provide an appropriate	B	C02	L3
A	obj	1	Tf-Idf Helps You To Establish?	Most Frequently Occurring Word In The Document	Most Important Word In The Document	Most Important Sentence In The Document	Most Frequently Occurring Sentence In The Document	B	C02	L3
A	obj	1	Which of the following is not true input for the NLP?	Image	Text	Types input	Speech	A	C02	L3
A	obj	1	NLP Stands for	Natural Language Protocol	Natural Lingual Protocol	Natural Lingual Processing	Natural Language Processing	D	C02	L3
A	obj	1	Text Mining is	Conceptual	Theoretical	Empirical	None of the above	C	C02	L3
A	obj	1	Predictive text analytics tasks include	Prediction	Classification	Clustering	All of the above	D	C02	L3
A	obj	1	In a text corpus of N documents, one document is randomly picked. The document contains a total of W words and the word 'mining' appears K times. What would be the cell value for the	K	1	W	0	B	C02	L3

Part	sub/obj	Marks	Question	Answer Option 1	Answer Option 2	Answer Option 3	Answer Option 4	Correct Answer(A/B/C/D)	CO	Bloom's Taxonomy Level
A	obj	1	Text analysis and NLP is an integral part of	Machine Learning	Database Systems	Artificial Intelligence	Operating Systems	C	CO2	L3
A	obj	1	Consider the following sentences:•Sentence 1: The brown dog is running.•Sentence 2: The black dog is in the black room.•Sentence 3: Running in the room is forbidden.how many	8	10	9	11	C	CO2	L3
A	obj	1	Which matrix counts the number of occurrences of each word in the document?	Sentence term matrix	Word term matrix	Text term matrix	Document term matrix	D	CO2	L3
A	obj	1	Which tokenizer is used to split the punctuations?	WordPunct	WordPunc	WordPunctuation	WordPun	A	CO2	L3