Unit 9 - Week 8

urse outline		
urse outilite	Assignment 8	
to access the portal?	The due date for submitting this assignment has passed.	Due
k 1	As per our records you have not submitted this assignment.	
k 2	The process of transforming unstructured text involves following steps	
	Cleaning Text	
k 3	Tokenization	
k 4	Stemming or lemmatization	
	All of the above	
k 5	No, the answer is incorrect. Score: 0	
k 6	Accepted Answers: All of the above	
7	A large collection of text data in text mining is known as	
8		
,	Text database Data Collection	
kt Collection &	Corpus	
nsformation Part I	O Data Store	
t Collection & nsformation Part II	No, the answer is incorrect. Score: 0	
Mining & Modeling Part I	Accepted Answers: Corpus	
Mining & Modeling Part II	Corpus	
t Mining & Modeling Part III	3) The primary assumption that sentence tokenization has already performed is necessary in	
iz : Assignment 8	Treebank Word Tokenizer	
_	○ Word Tokenizer	
ution For Assignment 8	Punctuation Sentence Tokenizer	
NLOAD VIDEOS	None of the above	
2401/11/11/		
EEDBACK LINK	No, the answer is incorrect. Score: 0	
	Accepted Answers:	
	Treebank Word Tokenizer	
	4) Pattern = r'[^a-zA-Z0-9]' extract from text	
	Characters	
	Words	
	Alphanumeric Characters	
	None of the above	
	No, the answer is incorrect. Score: 0	
	Accepted Answers:	
	Alphanumeric Characters	

Assignment 8	
The due date for submitting this assignment has passed. As per our records you have not submitted this assignment.	Due on 2019-09-25, 23:59 IST.
1) The process of transforming unstructured text involves following steps Cleaning Text Tokenization Stemming or lemmatization All of the characters.	1 point
No, the answer is incorrect. Score: 0 Accepted Answers: All of the above	
2) A large collection of text data in text mining is known as Text database Data Collection Corpus	1 point
No, the answer is incorrect. Score: 0 Accepted Answers:	
3) The primary assumption that sentence tokenization has already performed is necessary in Treebank Word Tokenizer Word Tokenizer Punctuation Sentence Tokenizer	1 point
No, the answer is incorrect. Score: 0 Accepted Answers: Treebank Word Tokenizer	
4) Pattern = r'[^a-zA-Z0-9]' extract from text Characters Words Alphanumeric Characters	0 points
No, the answer is incorrect. Score: 0 Accepted Answers: Alphanumeric Characters 5) Process of converting the word to its base form or root form is called	1 point
Tokenization Lemmatization Stemming Both b and c No, the answer is incorrect.	
Score: 0 Accepted Answers: Both b and c	
 6) Which of the following are stopwords in the English language? no nor not All of the above 	1 point
No, the answer is incorrect. Score: 0 Accepted Answers: All of the above	
7) In the text mining, we transform and represent text document as numeric vectors of specific terms. This form of proverous pace model Both a and b None of the above	esentation is called - 1 point
No, the answer is incorrect. Score: 0 Accepted Answers: Both a and b	
8) Which of the following features can be used for accuracy improvement of a classification model? Frequency count of terms Part of Speech Tag Dependency Grammar All of these	1 point
No, the answer is incorrect. Score: 0 Accepted Answers: All of these	
9) What is the right order for text classification model components? 1. Text cleaning 2. Text annotation 3. Model tuning 4. Text to predictors	1 point
1,2,3,4 1,3,4,2 1,2,4,3 1,4,3,2	
No, the answer is incorrect. Score: 0 Accepted Answers: 1,2,4,3	
10) A document-term matrix for the tweets is given, where every tweet has been treated as one document. Which of the following is correct?	1 point
1.Removal of stopwords from the text will reduce the dimensionality of the matrix 2.Normalization of weights of the matrix will produce more relevant features	
Only 1 Only 2 Both 1 and 2 None of the above	
No, the answer is incorrect. Score: 0 Accepted Answers: Both 1 and 2	