Unit 9 - Week 8

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   *1 point*
- ○ Cleaning Text
- ○ Tokenization
- ○ Stemming or lemmatization
- ○ All of the above

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   *1 point*
- ○ Text database
- ○ Data Collection
- ○ Corpus
- ○ Data Store

No, the answer is incorrect.
Score: 0
Accepted Answers:
Corpus

3) The primary assumption that sentence tokenization has already performed is necessary in   *1 point*
- ○ Treebank Word Tokenizer
- ○ Word Tokenizer
- ○ Punctuation Sentence Tokenizer
- ○ None of the above

No, the answer is incorrect.
Score: 0
Accepted Answers:
Treebank Word Tokenizer

4) Pattern = r'[^a-zA-Z0-9]' extract from text   *0 points*
- ○ Characters
- ○ Words
- ○ Alphanumeric Characters
- ○ None of the above

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?   *1 point*
- ○ no
- ○ nor
- ○ not
- ○ All of the above

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 presentation is called -   *1 point*
- ○ Vector space model
- ○ Term space model
- ○ Both a and b
- ○ None of the above

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?   *1 point*
- ○ Frequency count of terms
- ○ Part of Speech Tag
- ○ Dependency Grammar
- ○ All of these

No, the answer is incorrect.
Score: 0
Accepted Answers:
All of these

9) What is the right order for text classification model components?   *1 point*
1. Text cleaning
2. Text annotation
3. Model tuning
4. Text to predictors

- ○ 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