

Module 1: Introduction to Text Mining and NLP

Case Study

edureka!

edureka!

© Brain4ce Education Solutions Pvt. Ltd.

Case Study

Write code snippets in python to perform the following tasks:

1. Import nltk module in python
2. List all the documents present in NLTK 'corpora'
3. Import 'twitter_samples' document
4. Print all the files present in 'twitter_samples' file
5. Download 'swadesh' corpora using download in nltk
6. Print all the files present in nltk Gutenberg corpora
7. Print the contents of 'shakespeare-macbeth.txt'
8. Save 'shakespeare-macbeth.txt' content's in a text file as 'Macbeth.txt'
9. Count the number of articles in file 'Macbeth.txt' (Articles are 'a', 'an' & 'the') and save the file as 'Macbeth-ArticlesRemoved.txt'
10. Remove all the punctuations from the file. (Punctuations are . , ' ; : # () etc.) and save as 'Macbeth-punctuationsRemoved.txt'
11. Download 'names' corpora using download in nltk
12. Print the total number of words present in 'names' corpora of nltk
13. For each alphabet print the frequency of names present in 'names' corpora
14. Plot a bar graph showing frequency of names for each alphabet
15. Add appropriate titles, X-axis labels and Y-axis labels to the above created bar graph
16. Save the bar graph as "FrequencyOfNamesOfEachAlphabet.png"