Ali Khan

CSCE 4290/5290

9/2/2021

NLP ICE-1 Lab

Article Used

<https://text.npr.org/2021/06/26/1009487890/spacexs-new-rocket-factory-is-making-its-texas-neighbors-mad>

Output

[('SpaceX', 21), ('launch', 15), ('says', 12), ('Musk', 10), ('test', 10), ('rocket', 9), ('Starship', 9), ('plans', 8), ('Rio', 7), ('Grande', 7)]

# Github Link

<https://github.com/ayk314/NLP>

# GraphChart, line chart Description automatically generated

# Questions

1. **Why we use stopwords? Why stopwords are not necessary for NLP frequency distribution.**

We use stopwords to ignore the non important words. We don’t want common words like “is” and “the” to be analyzed. We don’t want the stopwords in the NLP frequency distribution because they are naturally the most common words in a given text; the frequency distribution would be skewed with them as the highest frequencies.

1. **Based on high frequency words what information you can extract from the graph?**

In general, the concepts that the article is about are also have the highest frequency. The article we used is about SpaceX and Elon Musk and those words are also some of the most frequently present in the text.

1. **Can you provide different visualization for frequency distribution? If yes, please perform. If no, why?**