**NLP - Summarization Workshop**

**Objective**

To practice using TFIDF as a tool for text featurization

**Exercise**

You are given a text file called “space\_invaders.txt”. Your task is to produce a 10-line summary of the text.

The algorithm is rather simple. First, compute the TFIDF value of each word in the text file by **treating each sentence as a document** and **the entire text file as the corpus**. Then, sum the **TFIDF** **values** of **each sentence**. Finally, selects sentences that have the **top 10 TFIDF sums** and print them **in the order that they appear in the text file**.

For pre-processing, remember to **remove stop-words** from your corpus and **stem each word** in the text.

Your summary might look like this:

Text, letter

Description automatically generated