Project proposal:

We want to create a program that gives an overview of the Spanish news of the day (articles that appear on El País), using sentiment analysis.

Different steps of the program:

1. Webscraping

scrape articles (titles + text, no images/image titles/comments…)

1. POS-tagging

We would omit functional words like prepositions, articles and conjunctions, so we only work with meaningful types of words: nouns, verbs and adjectives. It will be challenging to make sure also inflections of adjectives, all forms of the verb and so on are also included.

1. Check if the article contains good/bad news
   1. Adding a value to the word: matching words with
      1. Dictionary/list with positive words
      2. Dictionary/list with negative words

Maybe there are already some toolkits available online to split up words of a text into categories (positive and negative).

* 1. Check if 20 most common words are positive or negative
  2. If the word is not in the dictionary/list yet, add it, so it will be taken into account for future analysis.

1. Output
   1. 20 most common words in the text
   2. General value of the text (positive/negative)