Natural Language Processing with Disaster Tweets

<https://www.kaggle.com/competitions/nlp-getting-started/data?select=test.csv>

Steps :

Cleaning

* Remove site links
* Remove special characters
* Remove words that only have a letter

Tokenisation

* Create 2 new columns: 1 to count nb ‘positive’ words, another for ‘negative’ words
* Top words for target 1 / for target 0: check their frequencies

Training the Data

* Try ridge classifier
* Test these models out: TFIDF, LSA, LSTM / RNNs, etc.

The metric for this competition is F1

lejuanjo@gmail.com