

Applied Scaling & Classification Techniques in Political Science

Sixth Assignment



Deadline: 27 January 2023



- Use `uk_train.csv` and `uk_test2.csv`. They are a random sample of tweets discussing about UK politics during the 2014 EP elctions
- As an outcome focus on the «polite» variable: a dummy variable equals to 1 if a tweet uses a «polite» language; 0 an «impolite» language
- Use as your training-set the tweets included in `uk_train.csv`
- Given that we are talking about short texts (i.e., tweets) I would trim the texts very carefully. For example, by deciding to keep only those features that appear in at least 2 documents

Deadline: 27 January 2023



- Apply the ML algorithms discussed in class to classify the tweets included in `uk_test.csv`
- Comment the results you get from each algorithm
- Compare the results you get

EXTRA: consider as your Y also the «Sentiment» variable (with 3 categories: positive, negative and neutral)