

Applied Scaling & Classification Techniques in Political Science

Seventh Assignment



Deadline: 6 February 2023



Form 3 groups. I propose the following 3 groups (with no reason but the alphabetic index!)

1) ANDRIESSEN, Ivo Maarten; ASO, Mitsuhiro; JI, Xinru; KAMINAGA, Karen

2) LIM, Andrew; LIU, Yaru; NUKUI, Hikaru; MENG, Xiang

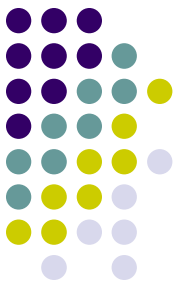
3) UNDERLAND, Jake; YASUTAKA, Seikai; ZHANG, Yuzhi; ZULUAGA, Katrina

Deadline: 6 February 2023



- Each group must run one single search on Twitter as you like and download between 5,000 and 10,000 tweets in any language you want
- Then define a set of categories for the tweets you have downloaded (it can be a 2-set categories such as positive/negative, or a 3-set categories such as positive/negative/neutral, or anything else you want)
- Discuss within the group the meaning of each class you have selected

Deadline: 6 February 2023



- Each member of the group must draw separately a training-set of around 300 tweets (if you have a 2-set categories) or 500 (if you have a 3-set categories, etc.). The remaining tweets will be part of the test-set. Ensure however that there is an overlap of 100 tweets in the training-sets extracted by each student in the group
- Check your inter-coder reliability on this set of 100 tweets. Report and comment your results
- Then run on the training-set the 2 ML algorithms we discussed in class
- Cross-validated your results and pick up the best algorithm
- Then classify the test-set

Deadline: 6 February 2023

- ✓ You can rearrange the groups as you like, or form a group with a lower number of persons
- ✓ You can also run this assignment by your own, but in this case skip you must of course skip the inter-coder reliability part

