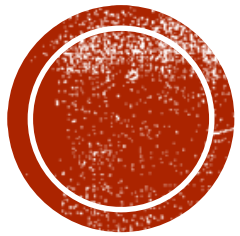

ASSIGNMENT

Text Classification Dataset

FIND A TEXT CLASSIFICATION DATASET

- Find a "standardized" text classification dataset
- Study the dataset
- 10 minute presentation
  - Description of the task
  - Features, statistics
  - Some data exploration on your chosen dataset
  - Why you recommend using this dataset
- Presentation slides due 23:59 Thursday 2nd of May
- Presentation on Friday 3rd of May
- Research proposal…???

SOME PLACES FOR TEXT CLASSIFICATION DATASETS

- UCI ML Repository: https://archive.ics.uci.edu/ml/datasets.html?format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=table
- Kaggle Text Data: https://www.kaggle.com/datasets?sortBy=hottest&group=public&page=1&pageSize=20&size=all&filetype=all&license=all&tagids=14104
- http://disi.unitn.it/moschitti/corpora.htm
- https://gengo.ai/datasets/the-best-25-datasets-for-natural-language-processing/?utm_campaign=c&utm_medium=quora&utm_source=community
- https://dataturks.com/projects/trending?type=TEXT_CLASSIFICATION

OTHER RESOURCES FOR NLP

- https://martin-thoma.com/nlp-reuters/
- How to solve 90% of NLP problems: a step-by-step guide
- The Essential NLP Guide for data scientists (with codes for top 10 common NLP tasks)
- Uncovering Hidden Trends in AirBnB Reviews
- Write SMS-spam detector with Scikit-learn
- Natural Language Processing with Python
- Text Classification Algorithms: A Survey (Github)