# Natural Language Processing in Python (Part 1)

Maximilian Spliethöver

April 29, 2020

**PADERBORN UNIVERSITY**

# Goal of this tutorial

- Short introdcution to NLP in Python

- Task: *Predicting the political ideology of users*

  - Corpus: *debate.org dataset [1, 2]*

- Main steps:

  - *Importing and inspecting the data (Part 2)*

  - *Preparing and preprocessing the data (Part 3)*

  - *Generating predictions and evaluating them (Part 4)*

# Assumptions

- Python **version 3.6.10**

- Python knowledge
    - Writing/running programs
    - Using `pip`
    - Using `pandas`, `scikit-learn` **and** `spacy` libraries

# Sources

**[1]** *Esin Durmus and Claire Cardie. 2019. A Corpus for Modeling User and Language Effects in Argumentation on Online Debating. In Proceedings of the 57th Conference of the Association for Computational Linguistics. Florence, Italy. Association for Computational Linguistics.*

**[2]** *http://www.cs.cornell.edu/~esindurmus/ddo.html*

# Next Part:
Data loading and exploring