# Assignment 4:

**Content warning:** this assignment contains an analysis of offensive language examples.

**Learning goals:**

- Applying and comparing different ensemble methods for hate speech detection
- Getting insights into current challenges in hate speech detection by performing an error analysis of the cross-domain ensemble results

**Steps of the assignment:**

- Develop three ensemble approaches
- Perform a quantitative error analysis (cf. assignments 2 and 3) both in the in-domain and cross-domain setups
- Perform a qualitative error analysis (cf. assignment 2) in the cross-domain setup focusing on the main challenges in hate speech detection
- Submit the notebook and write a report about your findings

**1. Data**

Same as in the assignment 3 (i.e., the OLIDv1 and HASOC datasets).

**2. Experimental setup**

Same as in the assignment 3 (i.e., the in-domain and cross-domain setups).

**3. Methods**

Combine 3 (or more) models from the assignment 3 within three different ensembling strategies. You can use different types of models (e.g., two transformer-based models and one machine learning). You can also use 3 transformer models.

Execute 3 ensembling strategies below:

- ➢ Hard majority voting
- ➢ Soft majority voting (note: not all classifiers output prediction probability, e.g., LinearSVC)
- ➢ Stacking ensemble: k-fold cross-validation + additional features + meta-model

**4. Analysis**

Focus on the ensemble approach that showed the best results.

- ➢ Perform a **quantitative analysis** of the ensemble results (both in-domain and cross-domain).
    - Is there an improvement over best-performing individual model(s) when using the ensemble approach?
    - Is it the case both for the in-domain and cross-domain setups?
    - Is there a drop in performance in the cross-domain setup?
- ➢ Discuss the reason for the obtained (improved) results when using the ensemble approach.
    - Do component models provide comparable results?
    - Do the component models produce uncorrelated predictions (use Pearson correlation coefficient)?
- ➢ Perform a **qualitative analysis** of the ensemble results in the cross-domain setup getting inspiration from the challenges discussed in (van Aken et al., 2018, section 6).
    - Focus on false negatives and false positives.
    - Provide your own ideas on how each type of error can be reduced.
    - Compare your finding from the error analysis to assignment 2.

**5. Write an academic report with introduction, related work, and conclusions… (you are allowed to copy what you described in previous assignments, e.g., data)**