

Practice exercises 4

You will explore different aspects of text data preprocessing, evaluating a model's performance, including both quantitative metrics and the model's robustness against adversarial attacks. This document is an overview of the practice exercises. Detailed instructions and explanations are provided in the accompanying Jupyter notebook.

Text data manipulation and preprocessing

- 4.1** Read text data from a given file and explore different reading functions.
- 4.2** Append new text data into an existing file.
- 4.3** Preprocess small chunks of text data using different techniques such as stopwords removal, stemming, and lemmatization.
- 4.4** Generate a visual representation of words that accounts for their frequency in the text.

Metrics

- 4.5** Get familiar with the metrics selection process based on the studied problem.
- 4.6** Get familiar with the BLEU metric for text. This exercise is designed to prepare you to think critically about the metrics you will use when building your own models, for example, for your mini-project.

Adversarial examples

- 4.7** Generate adversarial examples using the Fast Gradient Sign Method (FGSM). This exercise aims to demonstrate the vulnerability of image classification models, such as ResNet50, to imperceptible image perturbations.