

Report 2 for Text Detoxification task

The model chosen for the task is Encoder-Decoder model with Dot Product Attention mechanism.

`2.0-final-solution.ipynb` implements this model along with all improves described in Report 1.

It also adds **Serializing** and **Usage**

Its structure:

1. Data downloading and preprocessing

- a. Improved preprocessing part from `0.0-loading-and-preprocessing-data.ipynb`
- b. Constituting dataset (input dataset contains sentences that have greater tox score among the two) with filtering out long sentences
- c. Fitting and applying tokenizer
- d. Padding
- e. Train-test split (80/20)
- f. Forming batches

2. Model implementation

- a. Encoder implementation
- b. Dot Product Attention layer implementation
- c. Decoder implementation

3. Training

- a. Training and validation function implementations
- b. Code to run them

4. Serializing

5. Usage