Build report for PMLDL Assigment 1

(by Ksenia Shchekina k.shchekina@innopolis.university)

Baseline: Dictionary based

To solve the assignment model will paraphrase toxic text to normal text, so it should be text-to-text transform. There are a lot of pretrained models, but for complete this task I should have clear understanding all inside processes. So I cannot use too complicated model.

Hypothesis 1: Custom BERT will work for this task

BERT is a good model, clear enough even for basic level of knowledge. Also it performs text-to-text translation and already was used for paraphrasing tasks. It needs some additional things, but basically it will show good enough results.

Hypothesis 2: Also algorithm needs sets of toxic words etc.

At this task we can work in distinction from the overall meaning. Hence, simple list of not-allowed words and list of needs-to-replace words can be counted as a solution for this task.

Results: CondBERT

After exploring this task and also some information by David Dale, I decided to use CondBERT. For training I will just mask toxic words and use BERT to fill it by using words, that will be in allowed list.