

# Build report for PMLDL Assignment 1

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

## Introduction

For solving this task I want to use CondBERT algorithm.

## Data analysis

Initial filtered.tsv dataset is too heavy for fast training, so I decided to cut it for two small datasets: train set with 50000 rows and test set with 10000. I'm sure it will give a clear representation of model efficiency.

## Model Specification

Also for training CondBert model needs another binary classification model. This model will classify, is predicted text toxic or is not. Linear regression will good for that purpose

## Training Process & Evaluation

A preprocess dataset and extract texts. Unfortunately, I did not have time for training and evaluation.

## Results

Now I have preprocessed datasets and plan of future work.