

# Постановка задачи

Создать сервис определения степени токсичности текста

## Требования

Необходимо разработать веб-сервис. Веб-сервис должен иметь метод (он же endpoint, POST или GET по желанию), принимающий на вход небольшой текст (см. примечание) и возвращающий оценку степени его токсичности (от 0 до 1). Оценка токсичности должна формироваться ML-моделью. Модель может быть как "коробочной", так и самописной (со всеми возможными комбинациями).

Примечание: предполагается, что входящий текст по характеристикам (длина, язык, употребляемые слова, etc) схож с текстами корпуса, использовавшегося при обучении модели.

## Данные

Для обучения собственной модели следует использовать любой открытый датасет, например из соревнования на Kaggle: <https://www.kaggle.com/c/jigsaw-multilingual-toxic-comment-classification/overview>

## Инструментарий

Язык программирования: Python.

Для ML-составляющей можно использовать любой удобный инструментарий/библиотеки, например: SK-learn, PyTorch, TensorFlow, Transformers, etc. Допускается так же использование предобученных моделей, например из <https://pytorch.org/hub/> или <https://www.tensorflow.org/hub>. Для создания веб-сервиса следует использовать Flask.

## Прочее

Код решения можно передать любым удобным способом, например с помощью ссылки на GitHub или обычным zip-архивом.

Необходимо предусмотреть инструмент, позволяющий проверить работоспособность созданного решения. Например:

- Можно сформировать инструкцию по локальному развертыванию (набор команд с пояснениями), а так же сопроводить решение скриптом или Web-UI (например на streamlit'e), позволяющим увидеть оценку токсичности произвольного текста
- Можно развернуть созданное решение на каком-либо внешнем ресурсе, и опять-таки сопроводить решение скриптом, позволяющим увидеть оценку токсичности произвольного текста
- <собственный вариант>