# Проект для «Викишоп»

Поздравляем! Вы прошли курс в тренажёре. Самое время проверить знания и решить новую задачу машинного обучения. Выполнять работу будете самостоятельно.

Когда закончите, отправьте её на проверку ревьюеру: он пришлёт комментарии в течение суток. После этого нужно доработать проект и пройти повторную проверку.

Скорее всего, вы доработаете кейс по комментариям ещё несколько раз. Это нормально.

Проект завершён, когда ревьюер одобрит все доработки.

## Описание проекта

Интернет-магазин «Викишоп» запускает новый сервис. Теперь пользователи могут редактировать и дополнять описания товаров, как в вики-сообществах. То есть клиенты предлагают свои правки и комментируют изменения других. Магазину нужен инструмент, который будет искать токсичные комментарии и отправлять их на модерацию.

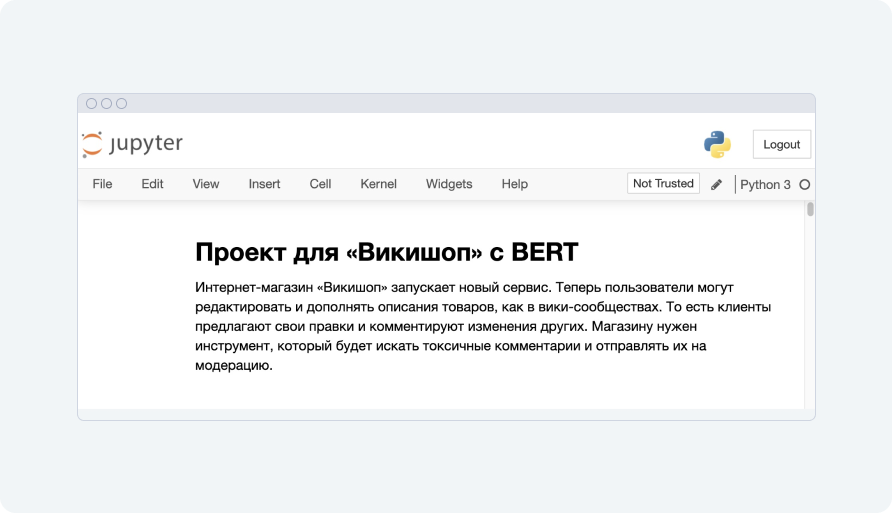
Обучите модель классифицировать комментарии на позитивные и негативные. В вашем распоряжении набор данных с разметкой о токсичности правок.

Постройте модель со значением метрики качества *F1* не меньше 0.75.

### Инструкция для проекта

Решить задачу можно как с помощью BERT, так и без этой нейронки. Если хотите попробовать BERT —

* Выполните проект локально. В тренажере тетрадь Jupyter ограничена 4 ГБ оперативной памяти — для проекта с BERT этого может не хватить.
* Упомяните BERT в заголовке проекта в первой ячейке:



Выполнить проект без BERT можно локально или в нашем тренажёре.

В любом случае алгоритм решения выглядит так:

* Загрузите и подготовьте данные.
* Обучите разные модели.
* Сделайте выводы.

### Описание данных

Данные находятся в файле /datasets/toxic\_comments.csv. [Скачать датасет](https://code.s3.yandex.net/datasets/toxic_comments.csv).

Столбец *text* в нём содержит текст комментария, а *toxic* — целевой признак.