



АО «Гринатом»

Тестовое задание для junior-специалистов по направлению Data Science

Уважаемый кандидат, в рамках тестового задания вам предлагается разработать веб-сервис для оценки комментариев (отзывов) к фильмам.

Мы предлагаем вам взять открытый набор данных, который содержит в себе отзывы о фильмах, а также соответствующие им оценки рейтинга. Рейтинг может служить ориентиром для построения модели классификации отзывов.

<https://ai.stanford.edu/~amaas/data/sentiment/>

Более подробное описание структуры файлов данных а также сами данные вы можете найти по ссылке: https://ai.stanford.edu/~amaas/data/sentiment/aclImdb_v1.tar.gz

Прежде чем вы приступите к разработке решения рекомендуем вам изучить статью: https://ai.stanford.edu/~amaas/papers/wvSent_acl2011.pdf

Для выполнения тестового задания вам необходимо:

1. Обучить модель на языке Python для классификации отзывов.
2. Разработать веб-сервис на базе фреймворка Django для ввода отзыва о фильме с автоматическим присвоением рейтинга (от 1 до 10) и статуса комментария (положительный или отрицательный).
3. Развернуть сервис в открытом доступе для оценки работоспособности прототипа.
4. Подготовить отчет о работе с оценкой точности полученного результата на тестовой выборке.
5. Отправить ответным письмом ссылку на прототип сервиса, ссылку на открытый репозиторий github с исходным кодом проекта, отчет о проделанной работе в формате pdf.

Рекомендованные бесплатные курсы для подготовки:

1. **Введение в машинное обучение.**
Константин Вячеславович Воронцов, Профессор
Факультет компьютерных наук НИУ ВШЭ, Школа анализа данных Яндекса
<https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie/>
2. **Deep Learning (осень 2019-2020): продвинутый поток**
Школа глубокого обучения МФТИ
<https://stepik.org/course/57457/>
3. **Нейронные сети и обработка текста**
Samsung Research Russia Open Education
<https://stepik.org/course/54098>
4. **Web-технологии**
Mail.ru Group
<https://stepik.org/course/154/>

