Для проверки ваших решений вам предоставляется тестовый датасет. Он состоит из файлов НПА, аналогичных файлам в датасете для обучения.

Вам необходимо осуществить классификацию выделенного текста.

Предсказания вы записываете в csv файл (разделитель точка с запятой). Название файла должно совпадать с названием команды.

**Структура csv файла**

|  |  |
| --- | --- |
| file\_id | Название файла |
| id | Номер выделенного текста |
| class | Предсказанный класс |

Пример

Рассмотрим пример для файла “1”.Нам предоставляется такой текстИзображение выглядит как текст

Автоматически созданное описание

Как можно заметить, в нём 6 выделенных отрезков.

Допустим мы считаем, что 1 и 5 отрезок относятся к классу 38,   
2, 3, 4 к классу 23, а 6 к классу 17. Тогда наш csv файл будет выглядеть подобным образом.

|  |  |  |
| --- | --- | --- |
| file\_id | id | class |
| 1 | 1 | 38 |
| 1 | 2 | 23 |
| 1 | 3 | 23 |
| 1 | 4 | 23 |
| 1 | 5 | 38 |
| 1 | 6 | 17 |

В первой колонке стоят единицы, так как смотрим файл “1”, во второй числа от 1 до 6, потому что в файле 6 отрезков текста, а в третьем столбце наши предсказания.

Csv файл выкладывается на ваш github.

Метрика: accuracy