

## Кейс:

Стоит задача сегментировать абонентов по трем категориям (0,1,2).

## Постановка задачи:

1. Необходимо построить модель, которая предскажет, к какому из трех сегментов относится каждый абонент.
2. Выбрать 500 абонентов из второго сегмента и 200 абонентов из третьего сегмента, для которых уверенность модели в ответе наибольшая.

## Форма представления результата

Работающая модель в .ipynb, которая принимает файл `contest_test.csv` из корневой папки и записывает в эту же папку два файла:

- `contest_answer.csv` — файл с предсказанными сегментами, в котором порядок абонентов такой же, как и в файле `contest_test.csv`. В файле должно быть два столбца (без индекса): `ID` с идентификаторами абонентов и `TARGET` с предсказанным сегментом. Каждому сегменту отвечает цифра 0, 1 или 2.
- `contest_segments.csv` — файл, в котором находится 500 абонентов из второго сегмента и 200 абонентов из третьего сегмента, для которых степень уверенности модели в ответе самая высокая. В файле два столбца (без индекса): `ID` и `TARGET`. Порядок абонентов в файле: сначала 500 абонентов из второго сегмента, потом 200 из третьего сегмента.

Файл с кодом `<фамилия латиницей>.ipynb` и два `.csv` с ответом следует поместить в архив с названием `<фамилия_имя>`.

Важно! В каждом из файлов должно быть только два столбца (без столбца индекса) в последовательности: `ID,TARGET`. Разделитель **запятая** (,).

## Критерии оценивания

1. Точность предсказания оцениваем по метрике `macro_f_score`<sup>1</sup>.
2. Дополнительно каждый сегмент из файла `contest_segments.csv` оценивается по метрике `lift`<sup>2</sup>.
3. Оценка кода: структура, последовательность действий, комментарии и выводы, чистота и понимание кода.

## Описание файлов

1. Обучающая выборка `contest_train.csv` состоит из следующих столбцов:
  - `ID` с идентификаторами абонентов.
  - `TARGET` с соответствующим абоненту сегментом.
  - `FEATURE_0...FEATURE_259` — данные абонента.
2. Тестовая выборка `contest_test.csv` состоит из столбца `ID` и следующими за ним столбцами `FEATURE_0...FEATURE_259`.

---

<sup>1</sup> [Sklearn.metrics.f1\\_score documentation](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html)

<sup>2</sup> [Семинары по машинному обучению: раздел 1.2.3](#)