**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 500абонентов из второго сегмента, потом 200из 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[[1]](#footnote-1).
2. Дополнительно каждый сегмент из файла contest\_segments.csv оценивается по метрике lift[[2]](#footnote-2).
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.

1. [Sklearn.metrics.f1\_score documentation](http://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html#sklearn.metrics.f1_score) [↑](#footnote-ref-1)
2. [Семинары по машинному обучению: раздел 1.2.3](http://www.machinelearning.ru/wiki/images/1/1c/Sem06_metrics.pdf) [↑](#footnote-ref-2)