# Анализ поведения пользователей

## Лабораторный практикум

## Требования

* ОС Windows (либо ОС Linux) с установленным Python-интерпретатором и IDE (IDE опциональна, если вы собираетесь использовать Google Colab или Jupyter Notebook);
* Предварительно загруженные тренировочная и проверочная выборки данных;
* Подключение к сети Интернет;
* Предварительно установленные библиотеки для Python:
  + Pandas, numpy и matplotlib для работы с данными;
  + Scikit-learn, catboost, xgboost или подобные для работы с моделями машинного обучения.

## Постановка задачи

Требуется обучить какой-либо классификатор на данной тренировочной выборке. Позже, обученный классификатор будет необходимо проверить на проверочной выборке.

## Рекомендации по использованию инструментов для решения задачи

Для подготовки данных и разведывательного анализа использовать Pandas, Numpy и Matplotlib (можно использовать и другие библиотеки). По результатам анализа выбрать одну из моделей sklearn, catboost, xgboost и обучить ее на подготовленных ранее тренировочных данных. Для оценки качества обученной модели использовать метрику F1\_score из sklearn с параметром average = ‘weighted’.

## Ожидаемый результат

F-score на мультиклассовой классификации (использовать метрику F1\_score с параметром average=’weighted’ из библиотеки sklearn) выше 0.8

## Рекомендованная литература

1. <https://scikit-learn.org/stable/> - документация sklearn
2. <https://catboost.ai/en/docs/> - документация catboost
3. <https://towardsdatascience.com/a-guide-to-pandas-and-matplotlib-for-data-exploration-56fad95f951c> - базовый гайд по использованию Pandas и Matplotlib для анализа данных
4. <https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html> - гайд по выбору правильной модели исходя из данных и задачи