

Ваша задача – сделать систему рекомендации друзей в социальной сети, которая для данного пользователя выдает список пользователей, с кем пользователь вероятней всего подружится. В качестве исходных данных вам дан анонимизированный датасет пользователей, где для каждого пользователя указан список текущих друзей пользователя. Для каждой пары пользователей известно, сколько времени прошло с того момента когда они подружились и насколько активно взаимодействуют в соц. сети, обе величины измеряются в некоторых абстрактных величинах.

Формат датасета:

uid1, uid2, time, intensity

1) Придумайте, как переформулировать продуктовую задачу в Data Science задачу.

2) Сделайте два варианта работы алгоритма:

- через эвристику, не требующую машинного обучения
- через любой близкий вам вариант машинного обучения – обучение с учителем, рекомендательные системы, итд.
- эвристика будет вашим бейзлайн-алгоритмом. Докажите, что ваш ML алгоритм лучше решает поставленную задачу.

Данные три пункта необходимо оформить в виде Jupiter Notebook, с вашими комментариями, почему вы приняли то или иное решение в ходе выполнения задания.

4) Напишите утилиту (это может быть программа на python или docker контейнер, в зависимости от ваших предпочтений), которая будет эмулировать работу вашего алгоритма в production. Программа должна отвечать на запросы за адекватное время.

Программа интерактивно принимает на вход по одному идентификатору пользователя на каждой строке и для каждой строки из ввода возвращает 5 пользователей, с которыми данный пользователь вероятней всего подружится.

Пример:

python recommend\_friends.py

INPUT:

15

9

5

OUTPUT:

1 5 29 3 100

8414 725 6 814 41

15 399 592 9959 159

В итоге, решение должно содержать Jupiter Notebook с аналитической частью и исходный код программы, которая применяет полученный алгоритм.