

BestHack 2021.Data

Команда RedLink



Михайлов
Баир



Сурков
Вячеслав



Гапонов
Максим



Тищенко
Евгений



Задание

ДАНО:

Aggregates – месячные отчеты о клиенте

Transaction – месячные отчеты о тратах клиента о категориях

ЦЕЛЬ:

- Провести разведочный анализ
- Придумать метрику качества
- Построить модель склонности к потребительскому кредиту

Склонность

ИДЕЯ

Для каждого клиента выбрать отчеты за последовательные 12 месяцев и разбить на 2 периода по 6

- Для каждого клиента выбирается отрезок отчетов за последовательные 12 месяцев. Делим на два периода по 6 месяцев
- Настоящее и прошлое: информация о клиенте обобщается по первому периоду, по ней определяется склонность
- Будущее: по второму периоду выясняется, взял ли клиент кредит
- Если в будущем клиент возьмет кредит, то у него высокая склонность, иначе низкая.

Разведочный анализ

Разведочный анализ

Метрика

- Важно не упустить потенциального потребителя!
- Показать рекламу незаинтересованному не настолько критично
- Склонность – вещественное число от 0 до 1
- Модель предсказала склонность y_{pred} , а правильный ответ – y_{test} .
- Метрика сильнее штрафует за $|y_{\text{test}} - y_{\text{pred}}|$, если $y_{\text{test}} = 1$, и слабее, если $y_{\text{test}} = 0$

Модель

Особенности:

- Небольшое количество клиентов (~10000)
- Много пропусков
- Большое количество атрибутов

Хорошо подойдут решающие деревья и леса.

Результаты: TODO