

# Santander Product Recommendation

## 8 / 1787

Александр Пономарчук  
Владимир Овсянников (sh1ngg)

# Задача

- Предсказать заказ новых услуг
- 24 услуги  
счета, депозиты, кредиты...
- Метрика MAP@7

# Данные

- Train: 2015-01 – 2016-05
- Test: 2016-06
- 24 продукта
- 22 фичи: возраст, пол, доход, провинция...
- 13.5M строк
- 0.55M продуктов добавлено

# Метирка

$\text{MAP}@7$  - ранжирующая метрика

Нужно предсказать упорядоченный список продуктов.

# Факторы

- Обычные фичи по-разному закодированные
- Продукты за предыдущие месяцы (до 5ти)
- Агрегаты по продуктам (min, max, std...)
- Повторяющийся доход как категория:
  - анонимизированные круглые числа?
  - бюджетники?
  - большие домохозяйства?

# Модель

- Мультиклассовая классификация
- Не нужно предсказывать вероятность, что продукт добавят, только отранжировать
- XGBoost, LightGBM
- Валидация: март и апрель 2016, LB

# Веса

- Что делать, если добавили больше одной услуги?  
Дублировать!
- Есть 2 клиента, 2 примера:  
1й клиент добавил услугу 0  
2й клиент — 1, 2
- Предсказываем для обоих  
0 —  $\text{MAP} = 0.5$   
1 или 2 —  $\text{MAP} = 0.25$
- $w = \exp(1/n)$  ;  $n * \log(w) = 1$

# 5 месяцев

- Train: 2015-01 — 2016-05
- Test: июнь 2016
- До июня 2015 есть 5 месяцев истории
- Июнь выделяется



# Мердж

- Сохраняли матрицу вероятностей:  
клиент \* продукт
- Среднее геометрическое (другие варианты)
- Учет сгора отдельного сабмита
- Жадный алгоритм, учитывающий корреляцию

# Не сработало\*

- Заполнить историю до 2015 года
- Заполнить пробелы по времени
- Ранжирование
- История больше 5 месяцев
- Какие-то факторы, статистики

\* Но помогло при мердже

# Не сделано

- Учет истории фичей (#1 учитывал)
- Использовать все данные (#1 использовал)
- Другие модели, не GB (#1...)
- Свертки по истории

Спасибо!

<https://github.com/yaxen>