



Sberbank Data Science Contest

Андрей Черток,
Head of R&D, Data Science, Sberbank

12 ноября 2016 г., DI Telegraph, г. Москва

Цифры



- ▶ 3181 регистраций;
- ▶ 718 активных пользователей;
 - ▶ больше, чем Kaggle от российских компаний;
- ▶ 11789 сабмитов;
- ▶ 1 запрос в секунду (50 запросов в секунду в последний день);
- ▶ 30 Гб данных с решениями.

Данные



► transactions.csv (~7 млн транзакций)

customer_id	tr_datetime	mcc_code	tr_type	amount	term_id
91472082	70 13:31:50	742	1110	-23986.38	
87695373	228 15:08:33	6010	7070	15721.41	888990
39026145	364 00:00:00	5499	1110	-2560.34	J145389

► tr_mcc_codes.csv (184 категории трат)

mcc_code	mcc_description
742	Ветеринарные услуги
4121	Лимузины и такси

► tr_types.csv (156 типов трат)

tr_type	tr_description
2322	Платеж с карты через АТМ
2411	Перевод с карты на счет др. лица в одном тер. банке через АТМ

Sberbank Data Science Contest



А. Предсказание пола клиента



Можно ли по истории транзакций восстановить демографический портрет клиента, включая его пол?

Сложность: низкая
Целевая метрика: AUC

Б. Предсказание суммарных трат



Как предсказать суммарные траты всех клиентов в следующем месяце на такие категории, как еда или связь?

Сложность: средняя
Целевая метрика: RMSLE

В. Предсказание трат клиентов



Можно ли предсказать, сколько каждый клиент потратит в следующем месяце денег, в разрезе категорий затрат?

Сложность: высокая
Целевая метрика: RMSLE

Идеи



- ▶ Сегментация клиентов
- ▶ Персонализация предложений;
- ▶ Next Best Offer;
- ▶ Предсказание надёжности клиента («устойчивость» профиля) - скоринг
- ▶ Антифрод - признаки мошеннических транзакций.

Факторы успеха



- ▶ **Политическая воля, \$**
 - ▶ Вадим Кулик, Александр Ведяхин, Максим Ерёменко, Максим Коновалихин
- ▶ **Данные**
 - ▶ Дмитрий Берестнев, Александр Карев
- ▶ **Проработка задач**
 - ▶ Александр Фонарёв, Алексей Кузьмин, Михаил Гавриков
- ▶ **Экспертиза, трансляция в сообщества**
 - ▶ Алексей Натёкин
- ▶ **Платформа**
 - ▶ РентаТим
- ▶ **Оргвопросы**
 - ▶ Анастасия Елисеева, Ксения Тотьянина, Ирина Шахназарова, Сергей Лоханов, Кирилл Сарайкин
- ▶ **Внимание к деталям**
- ▶ **Обратная связь и быстрая реакция, забота об участниках**

Институт соревнований



- ▶ Связи с data science-сообществом, нетворкинг
 - ▶ выступление на конференциях, ml-тренировках;
 - ▶ статьи;
 - ▶ R&D-разработки;
 - ▶ НИ;
- ▶ Развитие DS и AI в России;
- ▶ Повышение лояльности к бренду;
- ▶ Есть ли прямое business-value;
- ▶ В чём отличие от Kaggle?
- ▶ Контесты по AI

ОТЗЫВЫ



- ▶ **skipjack:** Спасибо организаторам! Конкурс очень интересный. Хотелось бы, чтобы в будущем было повторение подобного конкурса.
- ▶ **atikhonov:** Присоединяюсь к поздравлениям! Было интересно и полезно. Спасибо! P.S. верьте кросс-валидации ;)
- ▶ **IgorDoynikov2016:** Присоединяюсь к поздравлениям! Очень профессионально сделано. Особенная ценность для студентов - попробовать себя на реальных банковских данных, это огромнейшая возможность и ценность. Было бы такое лет 10 назад :)
- ▶ **RinatKurgan:** Так же присоединяюсь. UI-функций немного, но самое главное работают они идеально на все сто! А это главное! Спасибо и огромный респект!
- ▶ **Bernadsky2016:** Огромное спасибо организаторам и большие поздравления победителям! :) Было интересно. Единственное, лично мне было тяжело одновременно думать о трех задачах. Хотя они все из одной области, но все-таки требуют много времени на эксперименты. Успел плотно поработать только над первой задачей. На мой взгляд, было бы логично проводить соревнование поэтапно. Но это лично мое мнение. А так все очень достойно!

Контакты



- ▶ платформа contest.sdsj.ru
- ▶ <https://vk.com/sdsjru>
- ▶ <https://www.facebook.com/groups/sdsjru/>
- ▶ МОЯ ПОЧТА: avchertok.sbt@sberbank.ru



Спасибо
за крутое соревнование!