



Microsoft

# barcode



# Команда и экспертиза

## Хабр

★ Тutorial по uplift моделированию ★

★ scikit-uplift библиотека ★

★ User Guide scikit-uplift ★

★ 2 место Retail Hero uplift modeling соревнования ★



Максим Шевченко  
Data Science  
[Uplift, Geo]



Ирина Елисова  
Data Science  
[Uplift, RecSys]



Сергей Курочкин  
Javascript  
[React]



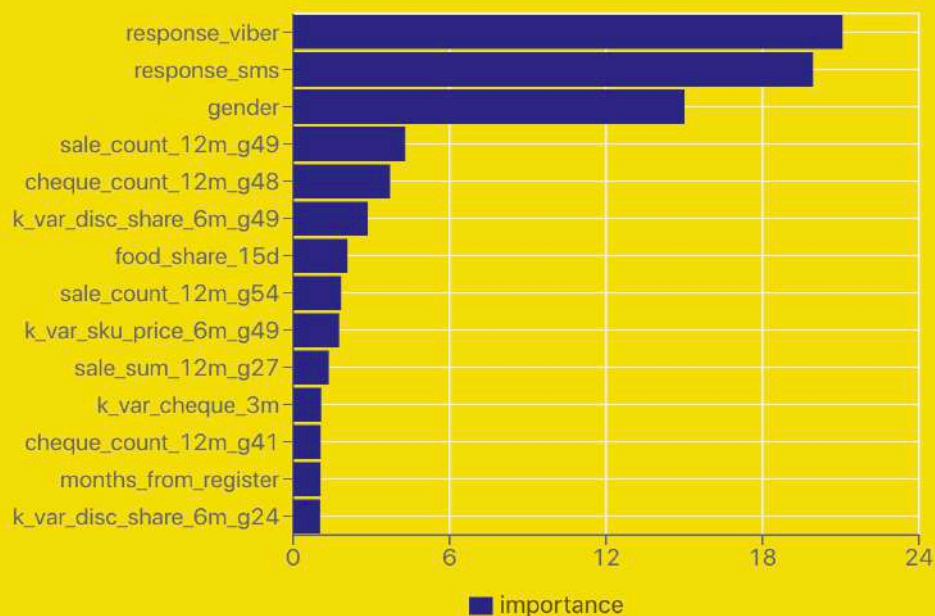
Кирилл Ликсаков  
Data Science  
[Uplift, FinTech]

# Инсайты

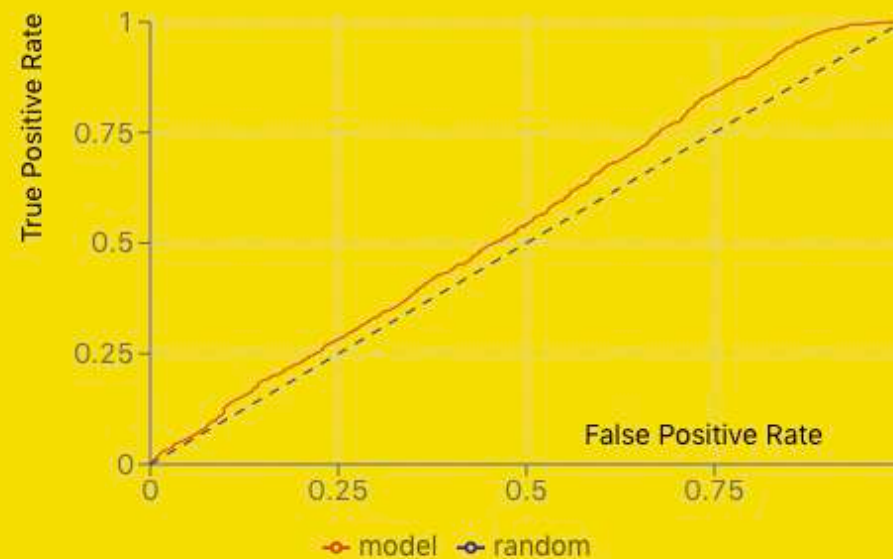
## Adversarial validation

Проверка разбиения на группы:  
Классификатор, отличающий  
контрольную группу от тестовой.

Adversarial Validation. Test vs Control



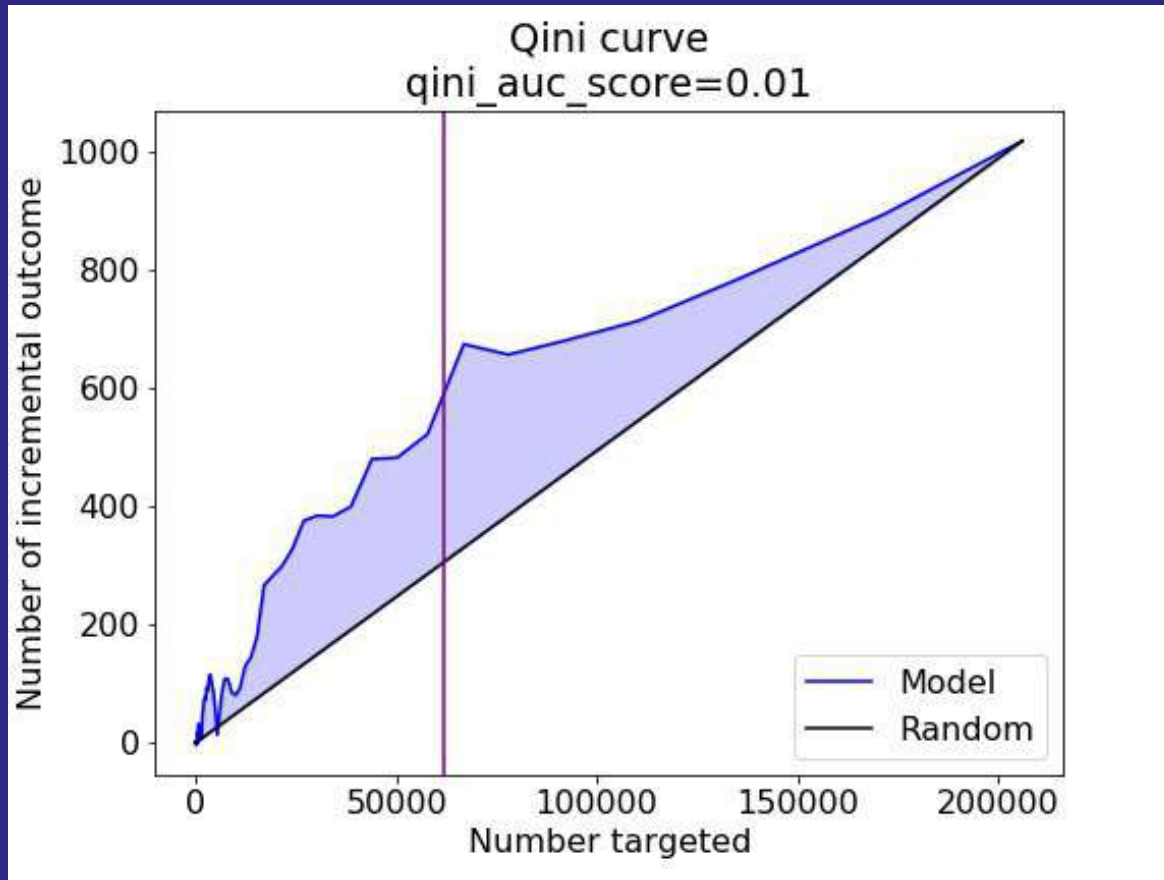
Adversarial Validation. Test vs Control



🤔 Признаки в топе Adversarial validation совпадают с признаками в топе uplift модели ClassTransformation

# Инсайты

Признак **cheque\_count\_12m\_g48** -> вместо предсказанного uplift



Такая эвристика работает **статистически  
значимо лучше** в сравнении со **случайным**  
алгоритмом

uplift@25 random:

**0.75%**

**vs**

uplift@25 mean:

**1.13%  $\pm$  0.33%**



# Генерация признаков

Для датасета на хакатоне


мы сгенерировали  
категориальные признаки

---

Категория товара  
топ 1, топ 2, ..., топ n

## Пример

index	cheque_count_6m_g20	cheque_count_6m_g21	cheque_count_6m_g25
0	0	12	9
1	1	0	2
2	6	0	9
3	2	1	11
4	0	1	2



index	top_g_cheque_count_6m
0	g21
1	g40
2	g25
3	g25
4	g40

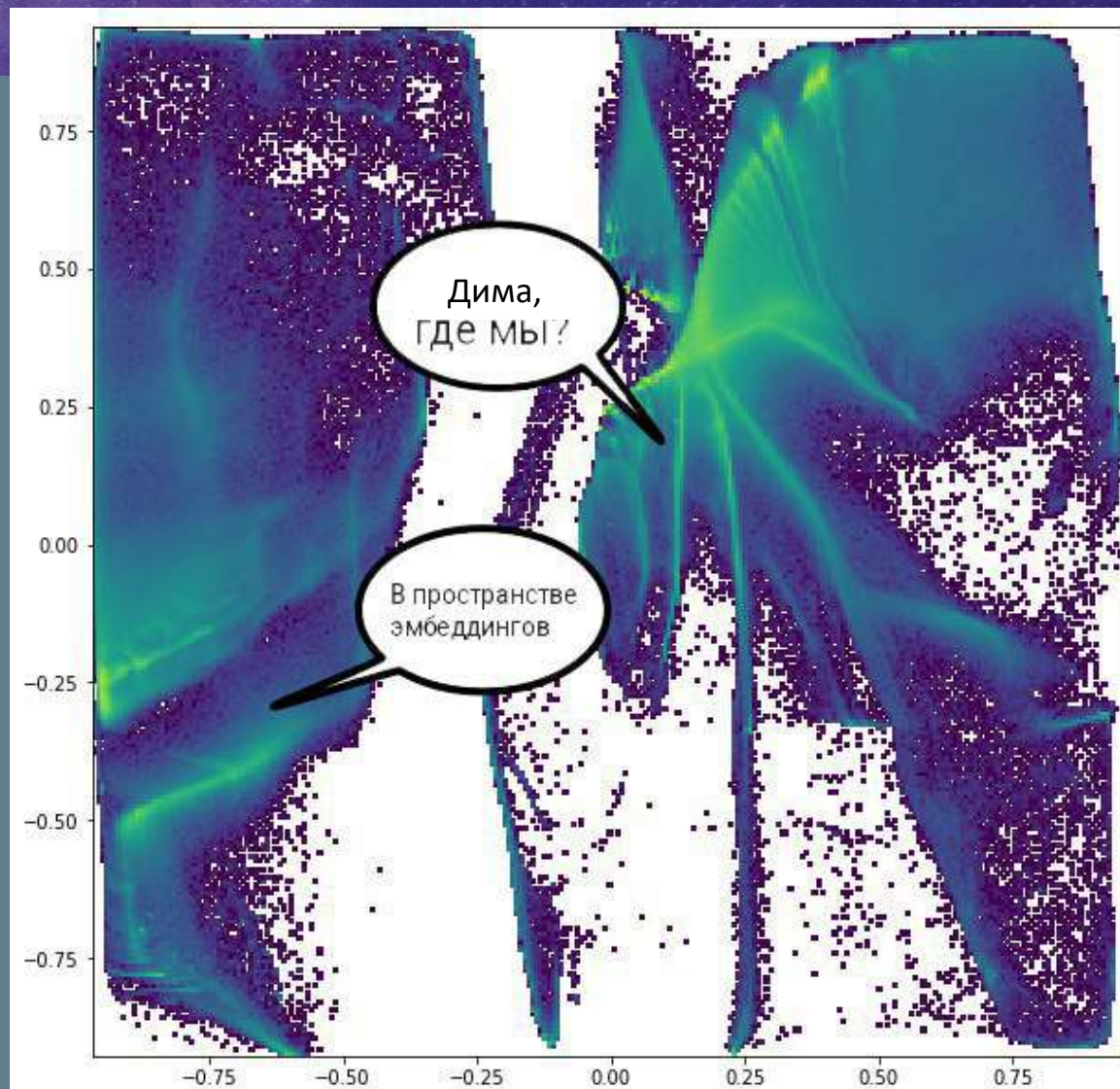


# Генерация признаков

Следующие шаги

Что можно было бы сделать с сырыми данными

- Эмбединги
- Категориальные фичи (любимый магазин, категория, топ продукт)
- Гео
- Графы
- Привязка покупок к календарю и времени суток
- ...



# Валидация

на  $N \geq 100$  разбиениях

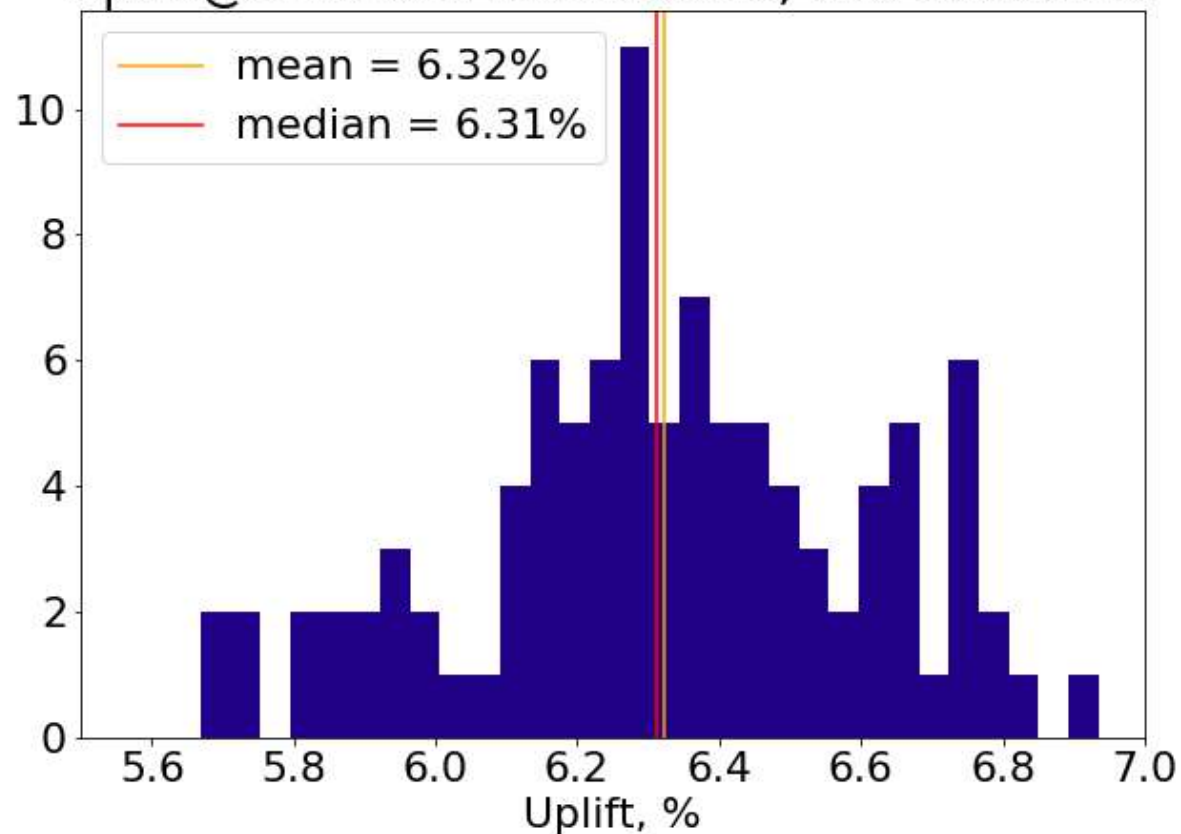
**uplift@k: std 0.4%**

большой шум

Контролируем  
переобучение  
в uplift моделях

Смотрим метрику не только  
на **validation**, но и на **train**

Uplift@25% test distribution, 100 iterations



# Microsoft Azure ML



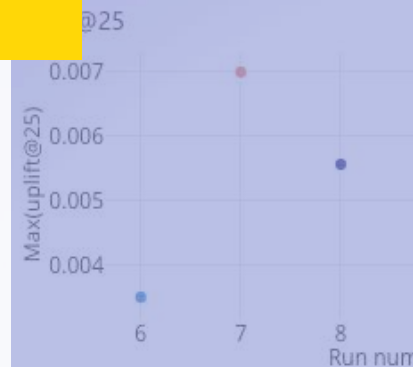
Логирование любых переменных

- Сравнение запусков с разными гиперпараметрами и выбор лучшего

- Хранение моделей

- Нужна возможность кастомизации графиков

(например, поворот осей или выбор переменной для оси)



Last(max\_depth)

5

5

5

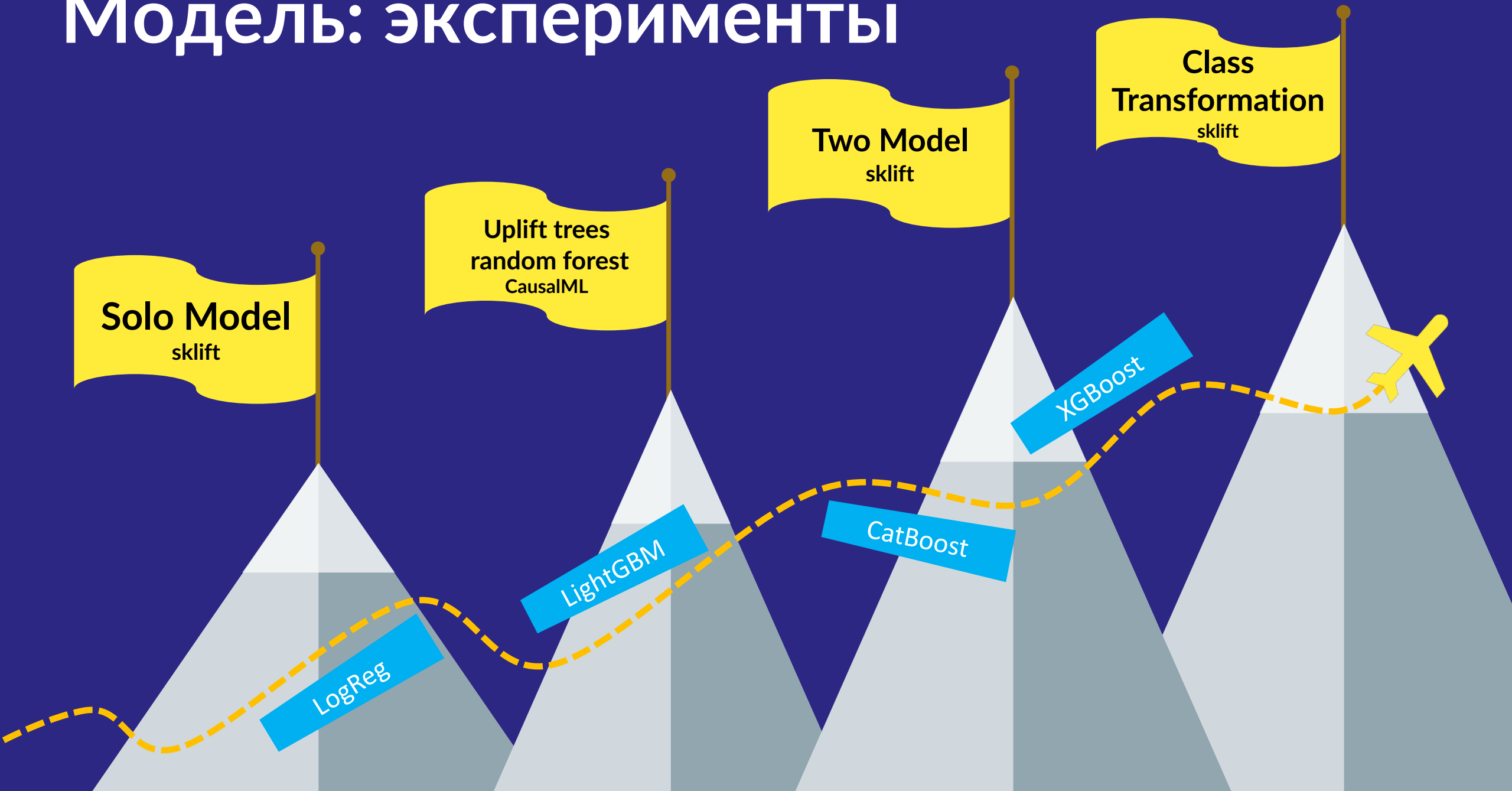
5

5

5

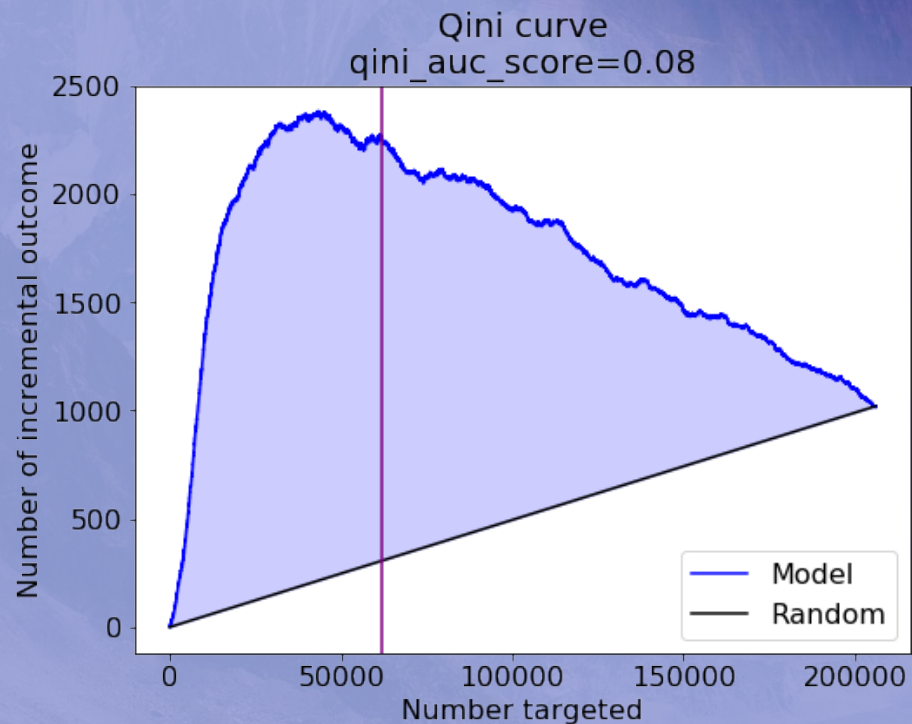


# Модель: эксперименты



# Class Transformation

+ классификатор



uplift@25

std  
0.28%

6.32%

std  
0.66%

uplift@5

25.07%

без  
балансировки  
классов  
качество  
оказалось  
лучше

	cardholder	response...	response...	k_var_da...	perdelta...	months_f...	stdev_da...	k_var...
<input type="radio"/>	16478978	0.7	0.167	0	0	9	0	0.262
<input type="radio"/>	15940845	1	0.75	NaN	0	1	NaN	0.514
<input checked="" type="radio"/>	15941038	0.867	0	0	0	19	0	0
<input type="radio"/>	16404535	0.889	0.125	0	0	26	0	0
<input type="radio"/>	15950619	1	0	0	0	55	0	0
<input type="radio"/>	16425048	1	0.6	0	0	22	0	NaN

15941038

Uplift: -0.426905

Shap

uplift = 2 \* (output\_value) - 1

Интерпретация прогнозов

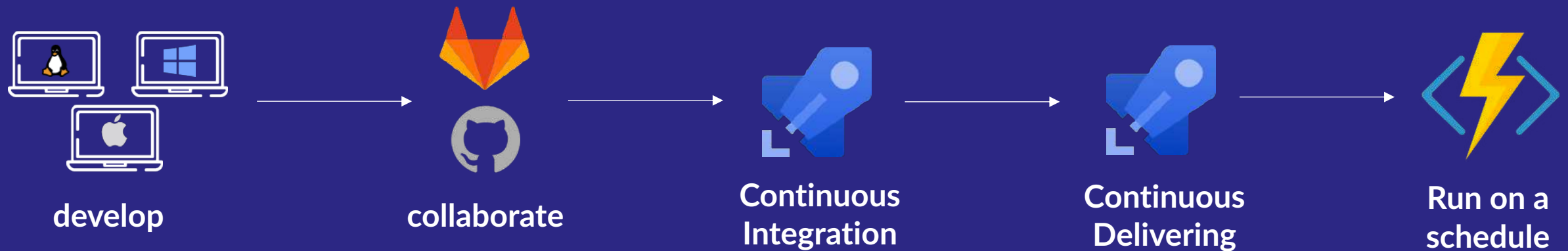
🤔





# Продуктивизация модели

Github/Gitlab + Azure DevOps + Azure MLOps + Azure Function



# Финансовая оценка

	Супермаркеты	Гипермаркеты
Доля клиентов в данных	0.101	0.899
Средний чек	800 р	1 400 р
Результат кампании в рублях	<b>1,1 М</b>	<b>17,5 М</b>

Разница в **CTR модель vs случайная рассылка**: \*  
 $16,57\% - 11,01\% = 5,56\%$  (клиенты, которых мы привели)

**Пилот на 250 000 клиентов (25% от 1 млн)**

$250\,000 * 5,56\% = 13\,900$  клиентов совершит целевое действие после рассылки

- Затраты на рассылку верхняя оценка =  $250\,000 * 2\text{ р} = 0,5\text{ М}$
- Расходы на предоставление скидки (336 рублей в ср.) =  $13\,900 * 336\text{ р} = 4,7\text{ М}$

**Итого**

$$17,5 + 1,1 - 0,5 - 4,7 =$$

**13,4 М**

\* Uplift = 6,32%  
CTR тестовая группа = 11,01 %  
CTR контрольная группа = 10,25 %  
CTR uplift модель группа = 6,32% + 10,25% = 16,57%



Команда

Опыт

Инсайты

Скоринг

## КОМАНДА BARCODE



Максим Шевченко

Data Science   Питч  
Презентация

Power Uplift'er. Соавтор цикла статей на хабре «Тutorial по Uplift моделированию». Разработчик библиотеки scikit-uplift для моделирования Uplift.



Ирина Елисова

Data Science   Презентация  
Управление командой

Data Scientist. Соавтор цикла статей на хабре «Тutorial по Uplift моделированию». Разработчик библиотеки scikit-uplift для моделирования Uplift.



Сергей Курочкин

Разработка   Frontend

Javascript разработчик. Ниндзя React приложений с опытом разработки коммерческих решений в сфере bigdata.



Кирилл Ликсаков

Data Science   Research

Математик. Танцор. Призер соревнований по Uplift моделированию от X5. Разработчик библиотеки scikit-uplift для моделирования Uplift.





