

Яндекс  Лавка

Обеспечение качества работы ML-моделей

от обучения до внедрения в прод

Коля Смирнов
ML-инженер



Обо мне

- 01 ML-инженер в Яндекс Лавке
- 02 Старший преподаватель
ВШПИ ИКНК СПбПУ Петра Великого
- 03 Бэкендер, QA-автоматизатор,
scrum-мастер, тимлид



О чём поговорим

Сервис поиска Яндекс Лавки

01

О чём поговорим

Сервис поиска Яндекс Лавки

01

Поломки в обучении модели

02

О чём поговорим

Сервис поиска Яндекс Лавки

01

Поломки в обучении модели

02

Внедрение и применение модели в рантайме

03

О чём поговорим

Сервис поиска Яндекс Лавки

01

Поломки в обучении модели

02

Внедрение и применение модели в рантайме

03

Наблюдение и сбор данных

04

История одного бага



```
72 def extract_thresholds(recommender, shard, place_name):
73     threshold_names_and_param_names = [
74         (FT_DSSM_THRESHOLD, DSSM_CANDIDATES_THRESHOLD_PARAM),
75         (FT_BERT_DSSM_THRESHOLD, DSSM_CANDIDATES_THRESHOLD_PARAM),
76         (FT_BM25_THRESHOLD, BM25_CANDIDATES_THRESHOLD_PARAM),
```



```
72 def extract_thresholds(recommender, shard, place_name):
73     threshold_names_and_param_names = [
74         (FT_DSSM_THRESHOLD, DSSM_CANDIDATES_THRESHOLD_PARAM),
75         (FT_BERT_DSSM_THRESHOLD, DSSM_CANDIDATES_THRESHOLD_PARAM),
76         (FT_BM25_THRESHOLD, BM25_CANDIDATES_THRESHOLD_PARAM),
```


Кандидаты на запрос «сыр твёрдый»



-40%
167₽ 279₽
Сыр Ламбер
230 г · 🕒
В корзину



-40%
263₽ 439₽
Сыр твёрдый Пармезан
Ricrem 6 месяцев
200 г · 🕒
В корзину



-40%
233₽ 389₽
Сыр твёрдый Пармезан
Cheese Gallery
175 г · 🕒
В корзину



-40%
239₽ 399₽
Сыр сливочный Natura
300 г · 🕒
В корзину

DSSM score

0,74

0,71

0,66









0,59

✂ threshold=0.61



{1, 2, 3}

Кандидаты на запрос «сыр твёрдый»

	 <p>-40% 167₽ 279₽ Сыр Ламбер 230 г · ◯</p> <p>В корзину</p>	 <p>-40% 263₽ 439₽ Сыр твёрдый Пармезан Ricrem 6 месяцев 200 г · ◯</p> <p>В корзину</p>	 <p>-40% 233₽ 389₽ Сыр твёрдый Пармезан Cheese Gallery 175 г · ◯</p> <p>В корзину</p>	 <p>-40% 239₽ 399₽ Сыр сливочный Natura 300 г · ◯</p> <p>В корзину</p>	<p>✂ threshold=0.61</p> <p>↓</p> <p>{1, 2, 3}</p>
DSSM score	0,74	0,71	0,66	0,59	
	 <p>-40% 167₽ 279₽ Сыр Ламбер 230 г · ◯</p> <p>В корзину</p>	 <p>-40% 263₽ 439₽ Сыр твёрдый Пармезан Ricrem 6 месяцев 200 г · ◯</p> <p>В корзину</p>	 <p>-40% 455₽ 759₽ Сыр твёрдый из овечьего молока Рота-Агро 100 г · ◯</p> <p>В корзину</p>	 <p>-40% 251₽ 419₽ Сыр твёрдый Брест- Литовск 12 месяцев 150 г · ◯</p> <p>В корзину</p>	
BERT score	0,035	0,033	0,031	0,026	

Кандидаты на запрос «сыр твёрдый»

	1	2	3	4	Filter
					✂ threshold=0.61 ↓ {1, 2, 3}
DSSM score	0,74	0,71	0,66	0,59	
					✂ threshold=0.61 ↓ ∅
BERT score	0,035	0,033	0,031	0,026	

На что повлиял баг?

Обучение

на сломанных логах



Коля Смирнов

Обучение

на сломанных логах



В проде

меньше кандидатов



Обучение

на сломанных логах



В проде

меньше кандидатов



А/Б тест

невалиден,
поехали сроки





Хорошо, что нашли! (с)



Часть 01

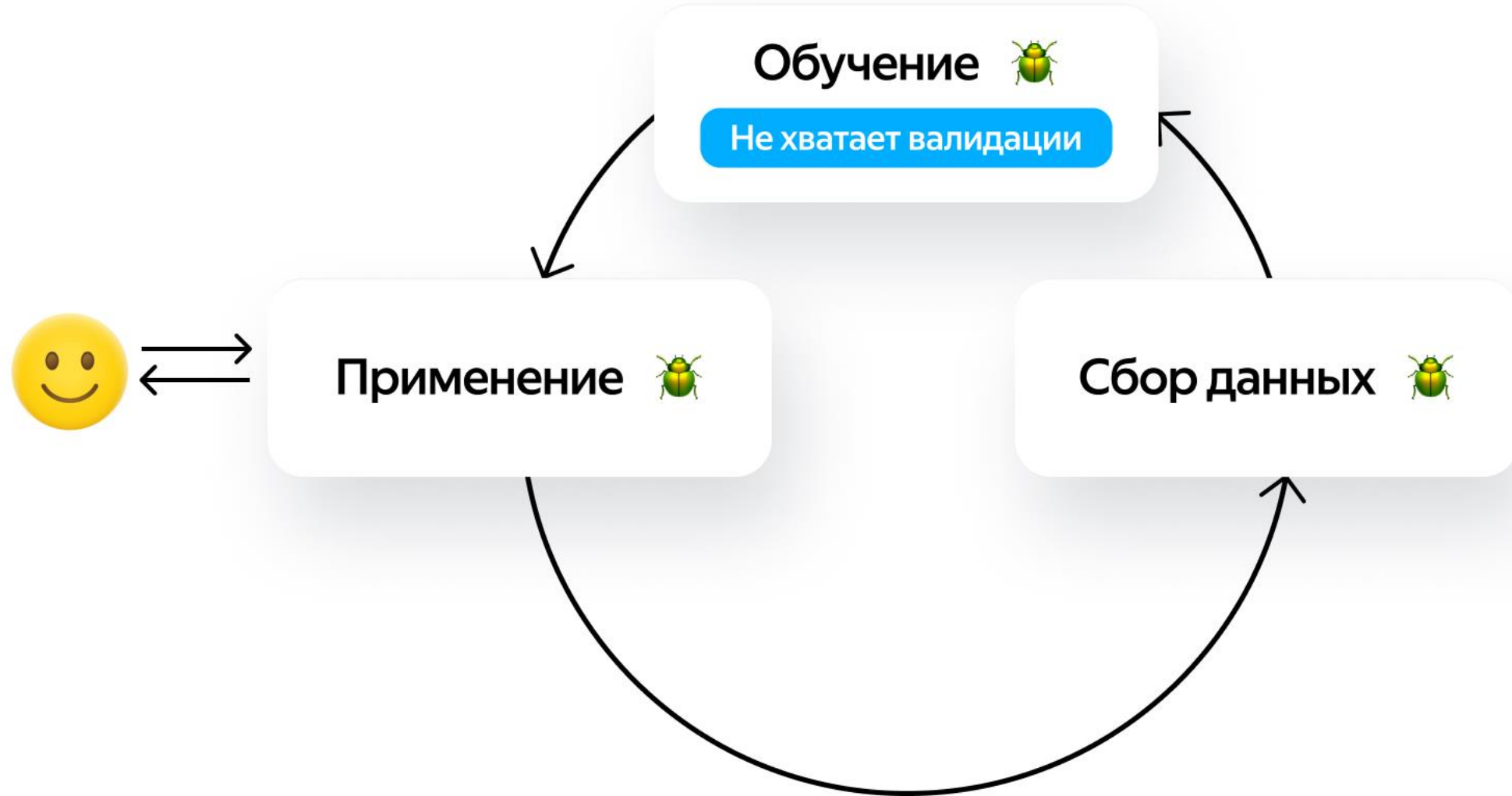
В каком месте системы можно отловить баг?

Сервис поиска Яндекс Лавки

Сервис поиска Яндекс Лавки



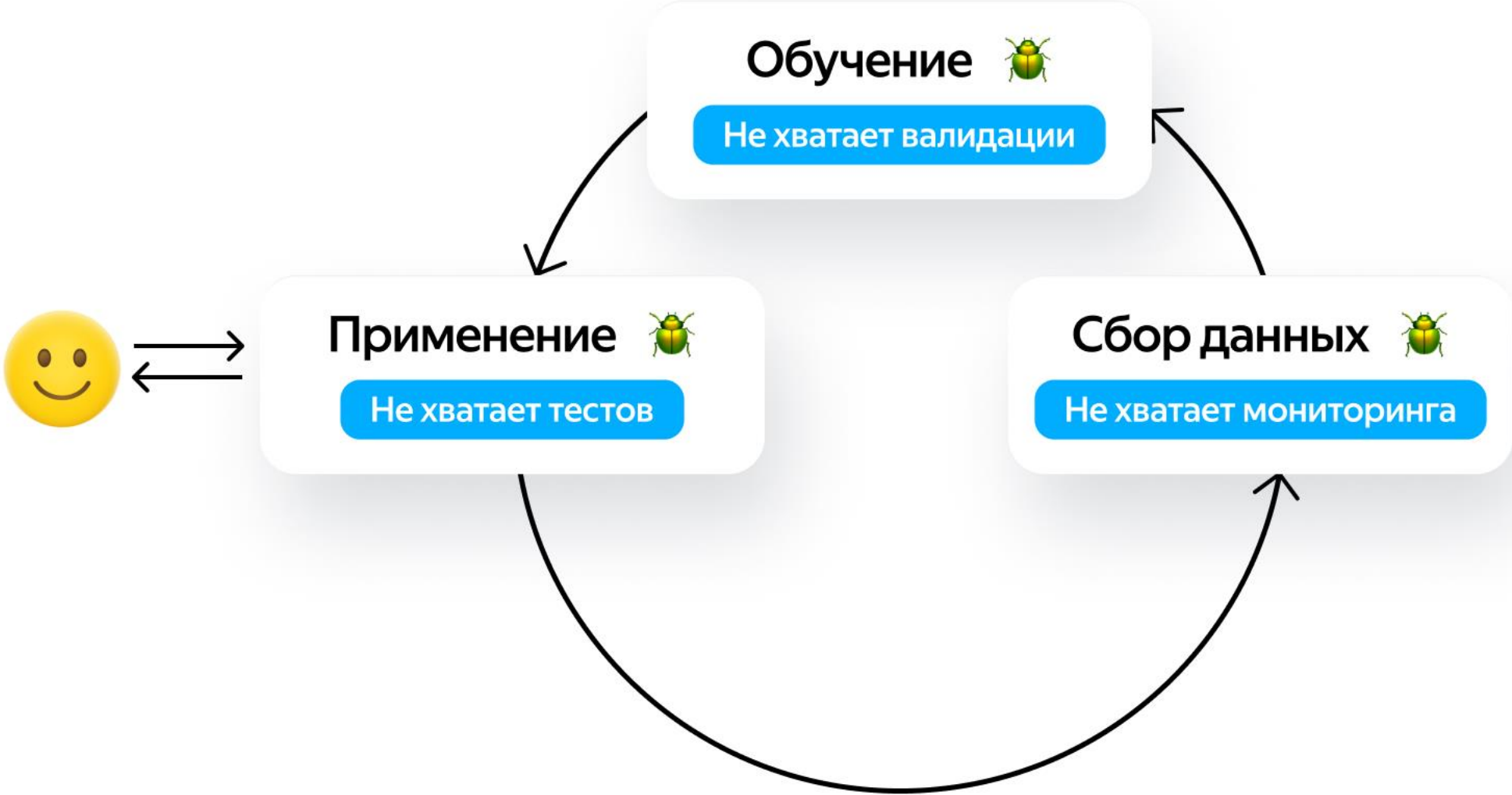
Сервис поиска Яндекс Лавки



Сервис поиска Яндекс Лавки

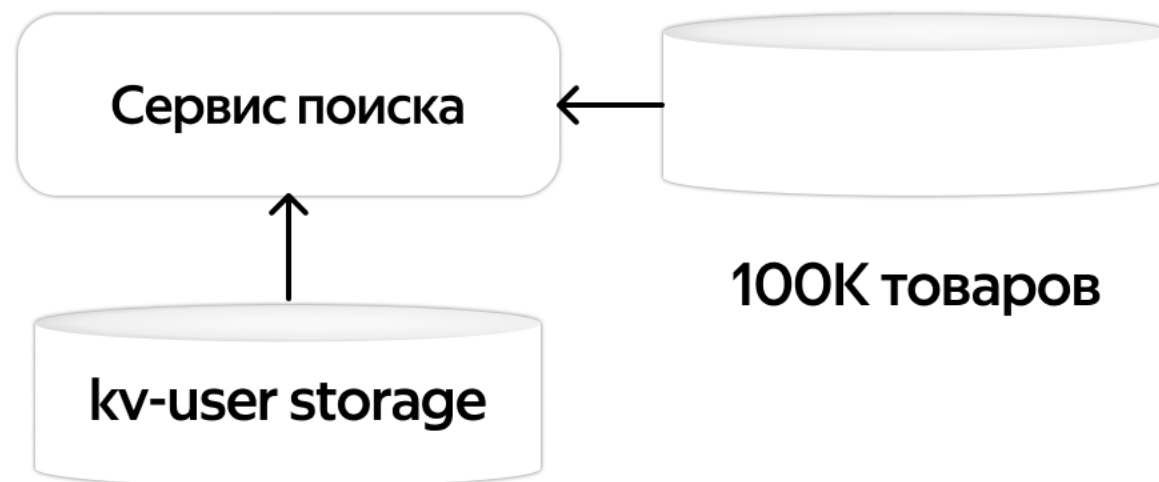


Сервис поиска Яндекс Лавки



Сервис поиска Яндекс Лавки

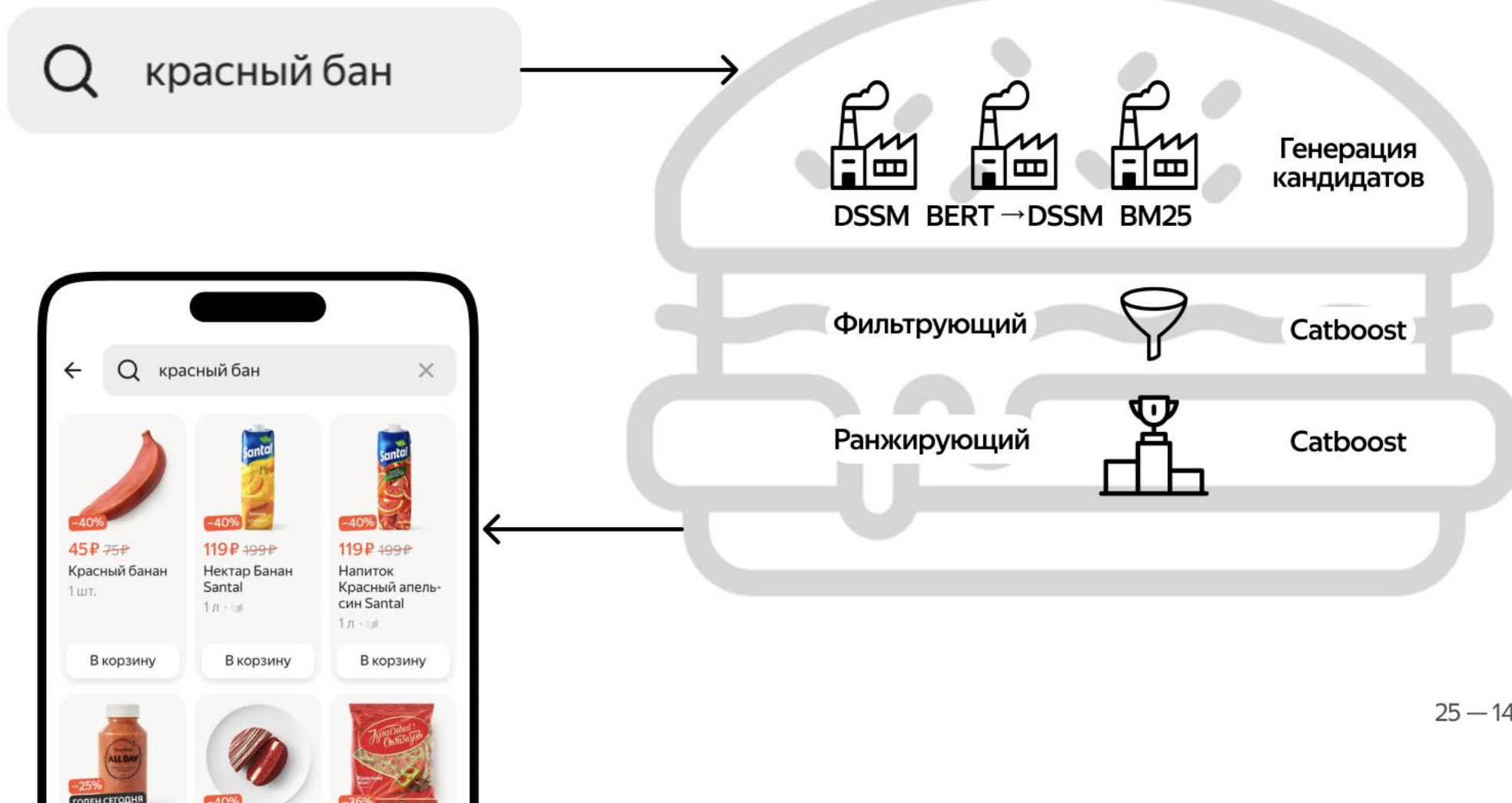
Условия задачи

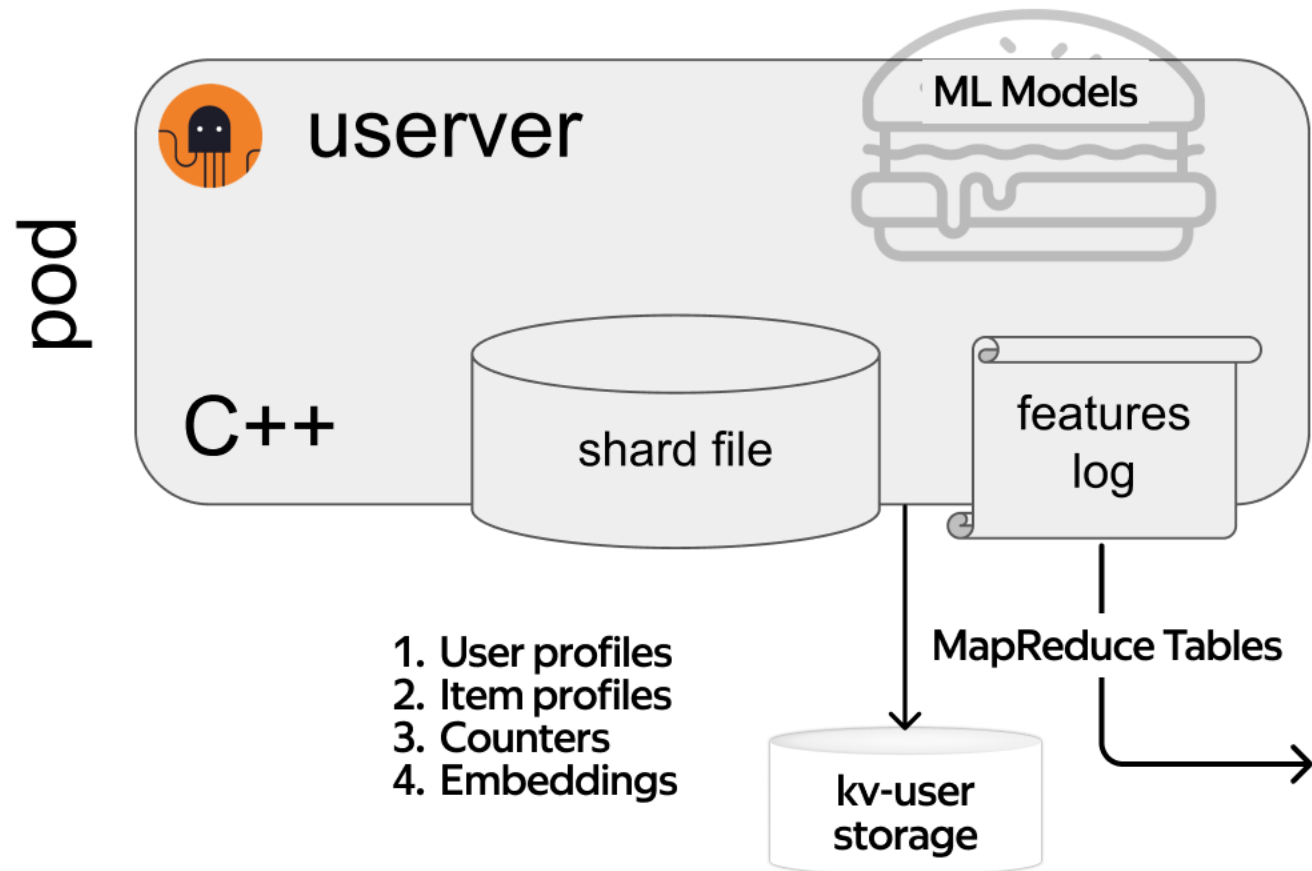


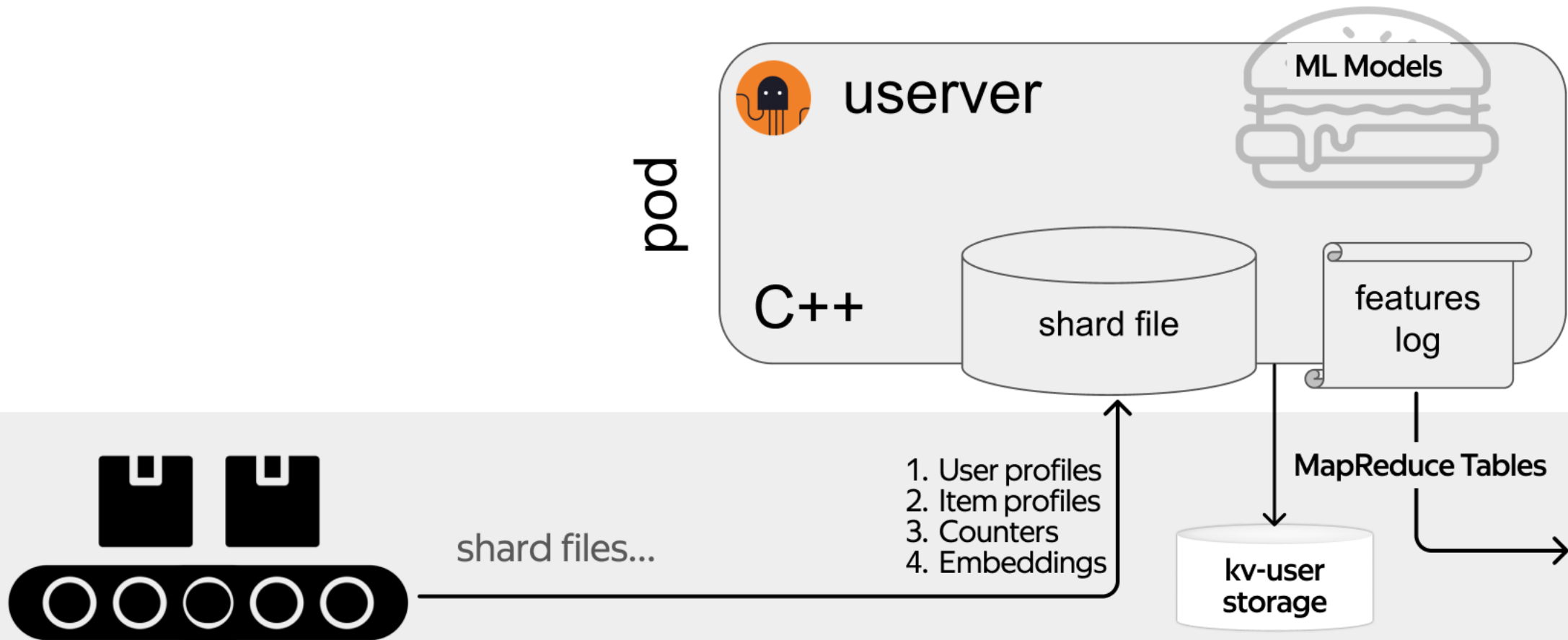
* префиксный поиск

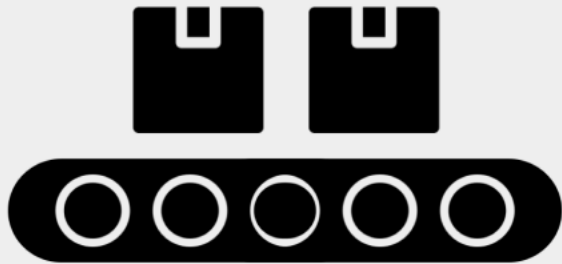
Сервис поиска Яндекс Лавки

Бургер моделей



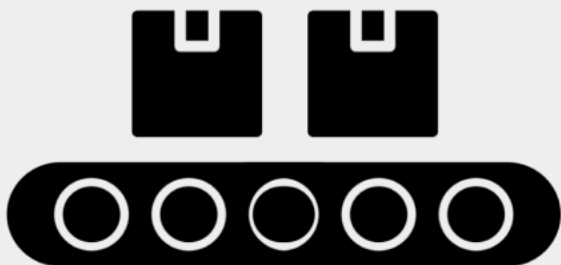
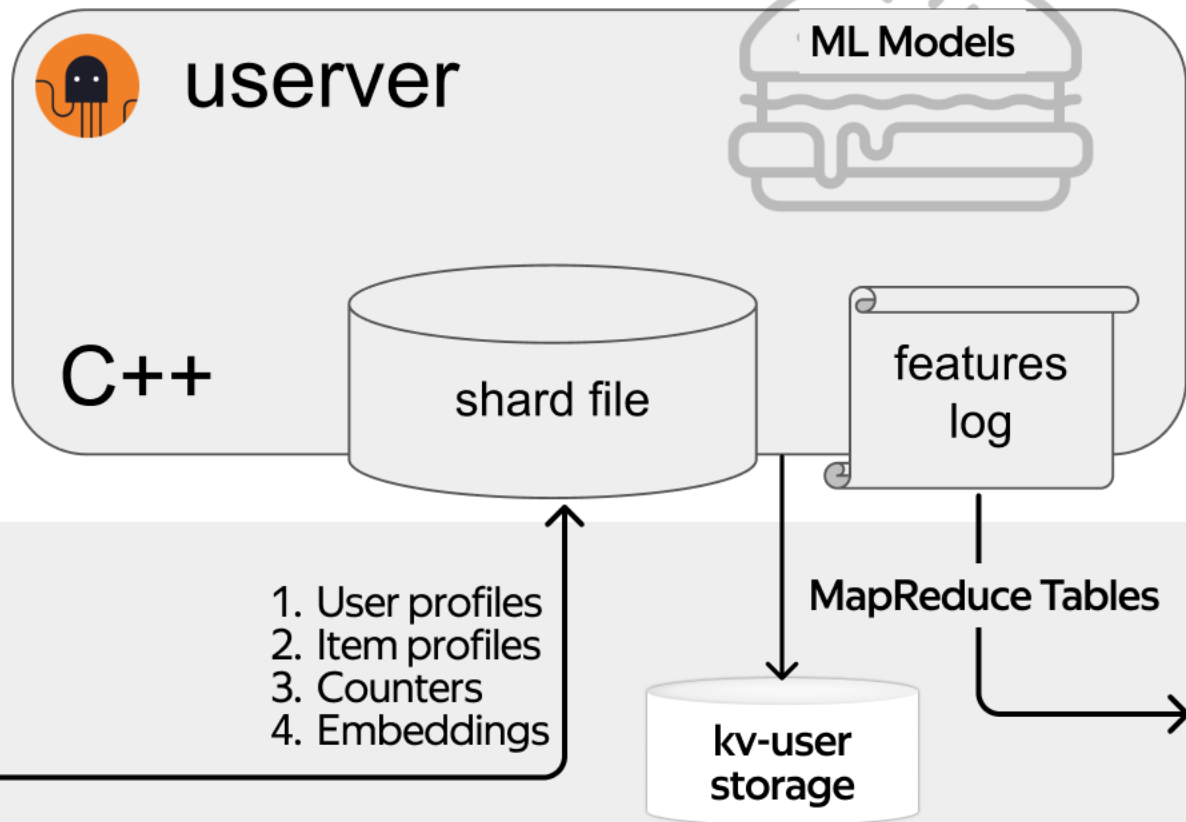




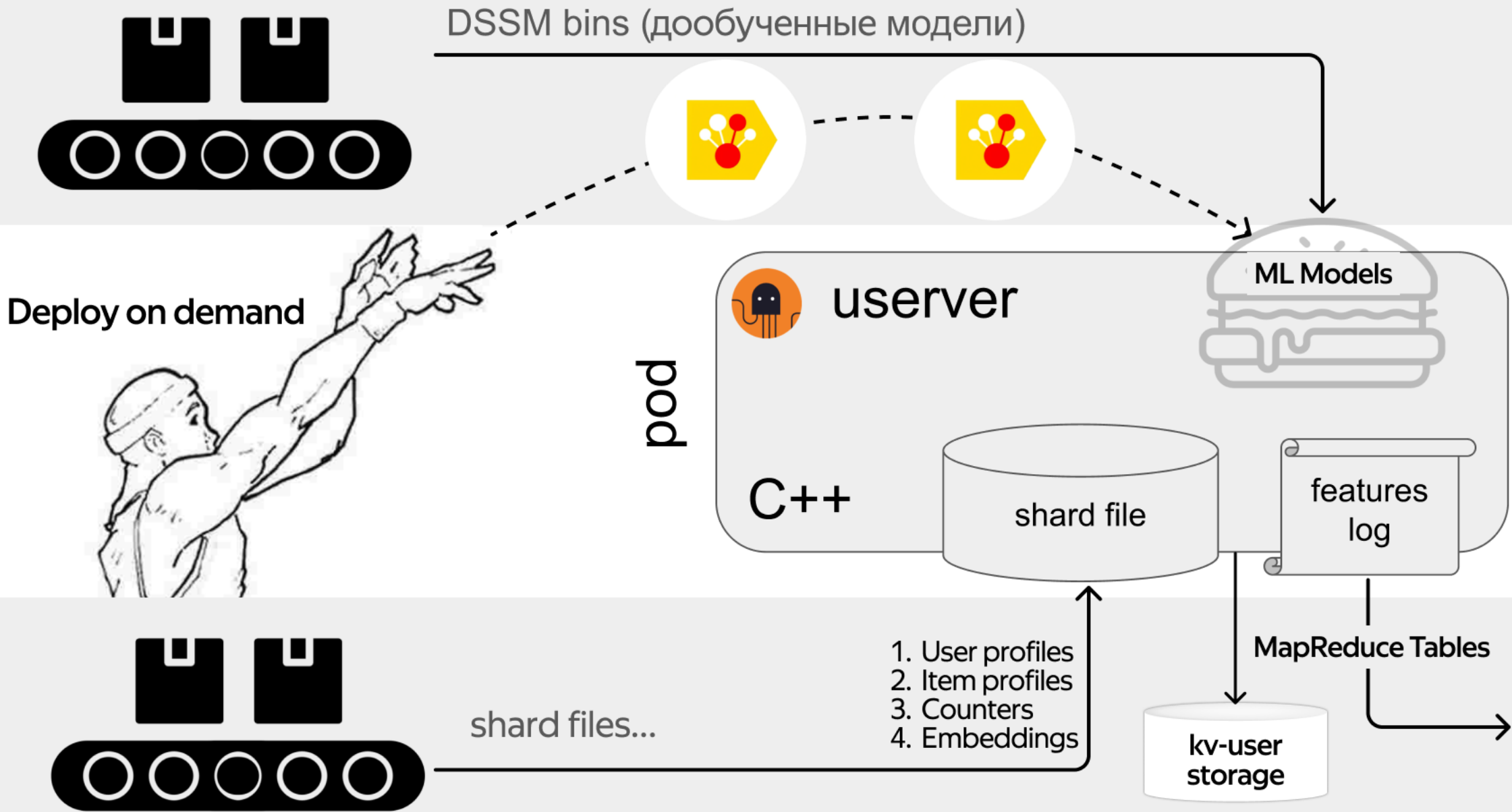


DSSM bins (дообученные модели)

pod



shard files...



Что может пойти не так?

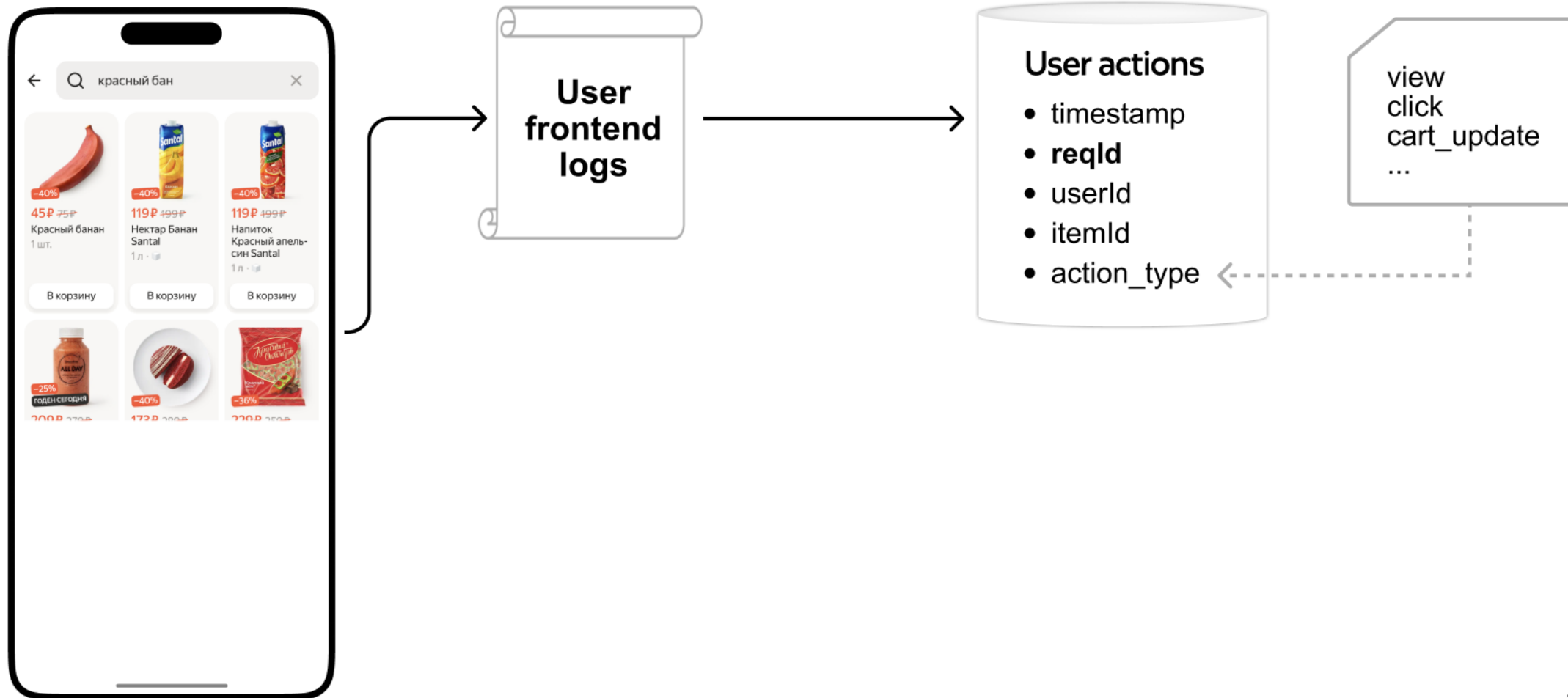
The screenshot shows the Yandex.Lavka mobile application interface. At the top, there is a browser address bar with the URL 'lavka.yandex.ru' and a search bar containing 'Колбаса'. The page header includes the Yandex.Lavka logo, a search bar with 'Колбаса', a location selector 'Пресненская н..., 10с1', a delivery time indicator '10-20 мин', a shopping cart icon labeled 'Корзина', and a login button 'Войти'. A left sidebar menu titled 'Каталог' lists various food categories: 'Придумано Яндекс Лавкой', 'Готовая еда', 'Овощной прилавок', 'Молочный прилавок', 'Булочная', 'Вода и напитки', 'Сладкое и снеки', 'Мясо, птица, рыба', and 'Заморозка'. The main content area displays a grid of food items, each with a 'В корзину' button above it. The items shown are: 1. BASKER WHEELS pancakes (36g) for 17 rubles (30% off from 25 rubles). 2. Karpatzho (100g) for 83 rubles (30% off from 119 rubles). 3. Farfalle Barilla pasta (400g) for 90 rubles (30% off from 129 rubles). 4. Pryanik Tulyjskiy (140g) for 59 rubles (25% off from 79 rubles). The bottom of the screen shows the beginning of another row of products, including a sausage and a cheese product.

Часть 02

Поломки в обучении модели

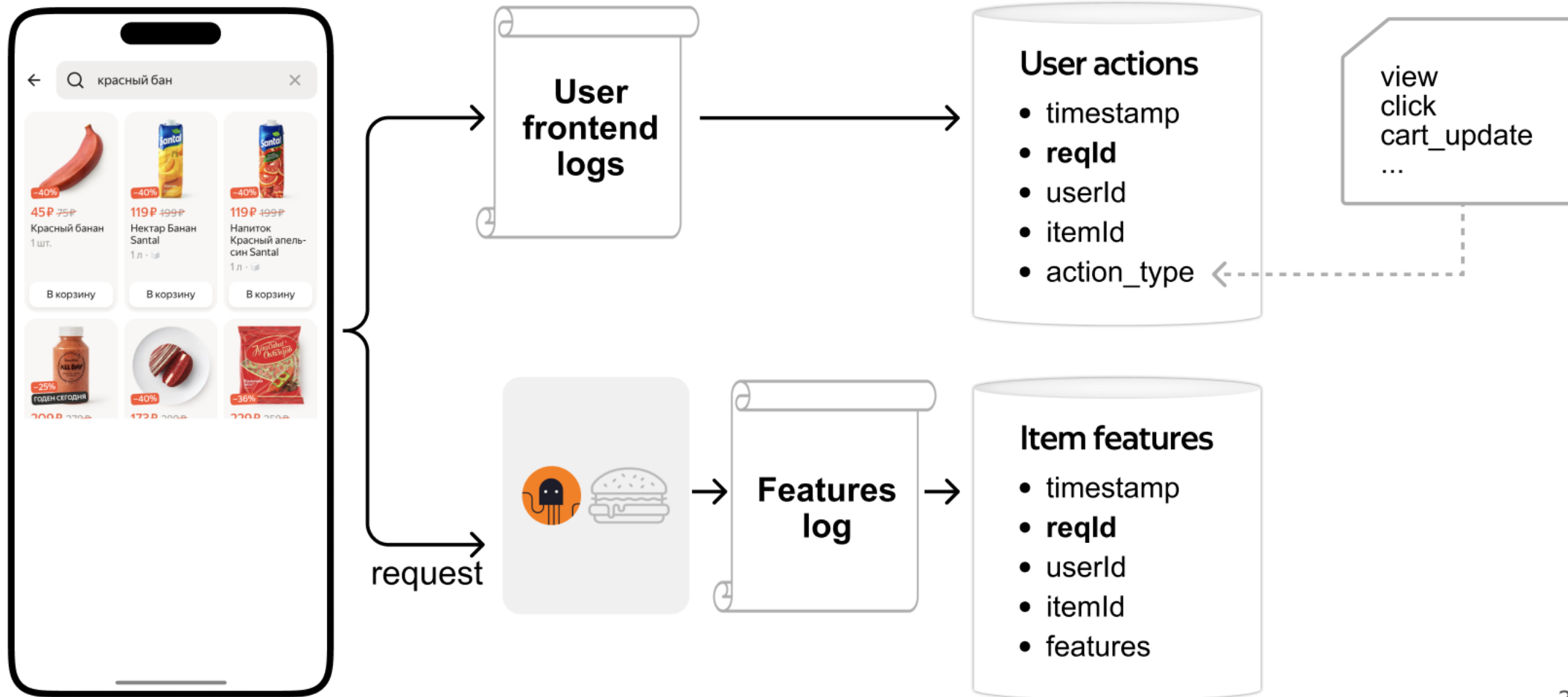
Поломки в обучении модели

Откуда берутся данные



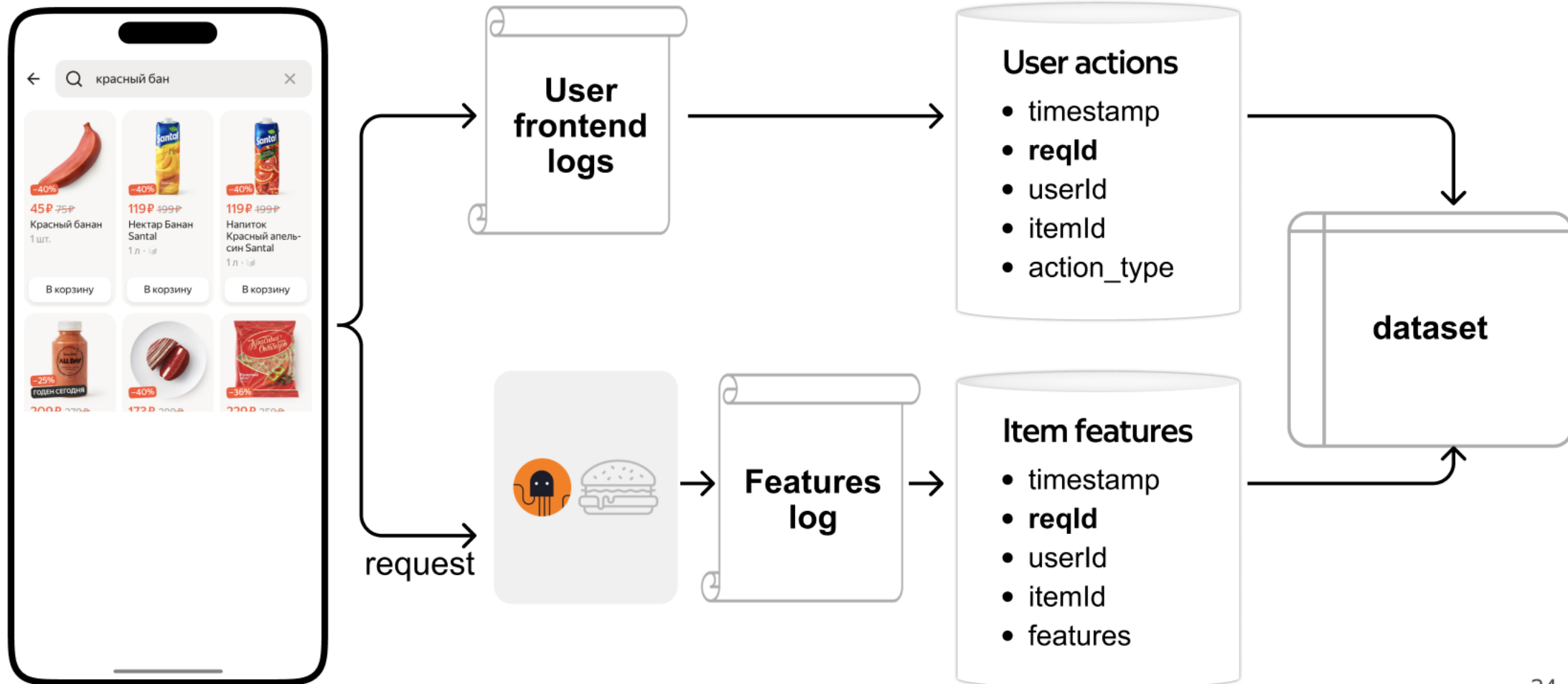
Поломки в обучении модели

Откуда берутся данные

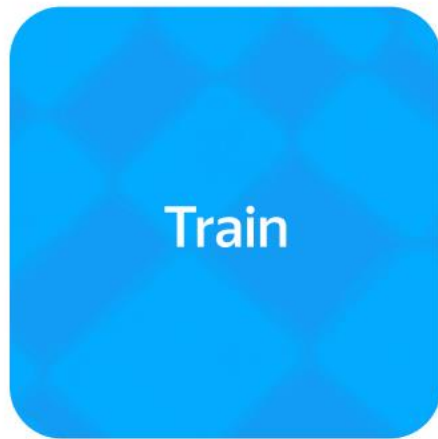


Поломки в обучении модели

Откуда берутся данные



Поломки в обучении модели



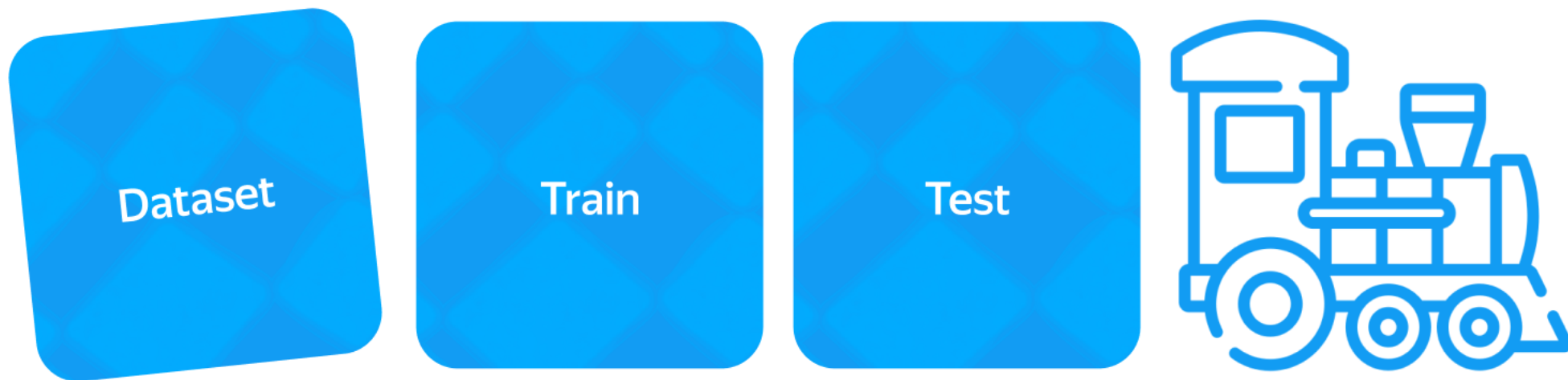
Поломки в обучении модели

01 Сломанные фичи (пустые, константные, меняющиеся во времени)



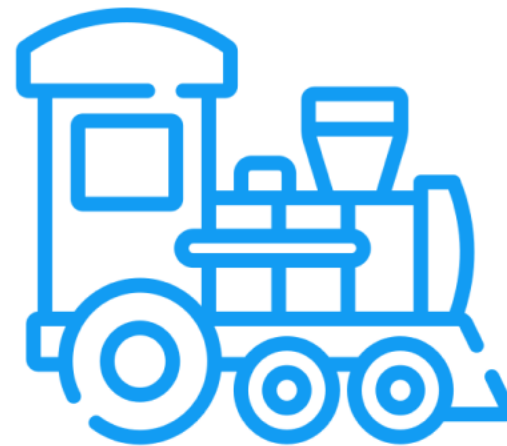
Поломки в обучении модели

- 01 Сломанные фичи (пустые, константные, меняющиеся во времени)
- 02 Расхождение способа обработки фичей в рантайме и обучении



Поломки в обучении модели

- 01 Сломанные фичи (пустые, константные, меняющиеся во времени)
- 02 Расхождение способа обработки фичей в рантайме и обучении
- 03 Неверные гиперпараметры модели



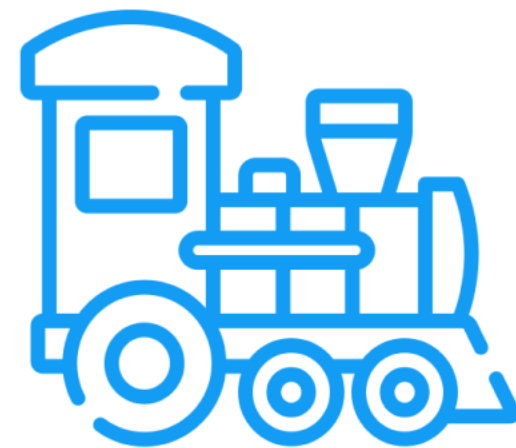
Поломки в обучении модели

- 01 Сломанные фичи (пустые, константные, меняющиеся во времени)
- 02 Расхождение способа обработки фичей в рантайме и обучении
- 03 Неверные гиперпараметры модели
- 04 Некорректное разбиение датасета



Поломки в обучении модели

- 01 Сломанные фичи (пустые, константные, меняющиеся во времени)
- 02 Расхождение способа обработки фичей в рантайме и обучении
- 03 Неверные гиперпараметры модели
- 04 Некорректное разбиение датасета
- 05 Сдвиг фичи в каскаде моделей



Поломки в обучении модели

Сломанные фичи: валидация пула

NewInGeneral_general
NumItemsInStock
Period_general
Price_w_vat_rub
ProbabilityRepeatOrder7Days
StmProduct_general
TimeOfDay
TimesOrdered30Days
TimesOrdered7Days
TimesOrdered_general
itemPageCarousel_ctr_7

Поломки в обучении модели

Сломанные фичи: валидация пула

	Index	Column type	Min value	Max value	Nonzero ratio	Non-default-min ratio	Unique values	Mode	Mode frequency	Mean	Std
NewInGeneral_general											
NumItemsInStock											
Period_general											
Price_w_vat_rub											
ProbabilityRepeatOrder7Days											
StmProduct_general											
TimeOfDay											
TimesOrdered30Days											
TimesOrdered7Days											
TimesOrdered_general											
itemPageCarousel_ctr_7											

Поломки в обучении модели

Сломанные фичи: валидация пула

NewInGeneral_general	Index	Column type	Min value	Max value	Nonzero ratio	Non-default-min ratio	Unique values	Mode	Mode frequency	Mean	Std
NumItemsInStock											
Period_general	22	Num	0	73.878	0.99284	1	95324	0	178954	5.4255	2.8843
Price_w_vat_rub	23	Num	0	34990	0.99087	1	41941	129	1042700	205.86	268.76
ProbabilityRepeatOrder7Days											
StmProduct_general											
TimeOfDay											
TimesOrdered30Days											
TimesOrdered7Days											
TimesOrdered_general											
itemPageCarousel_ctr_7											

Сломанные фичи: валидация пула

✔ Оценка значений фичей в собранном пуле

🤔 const or drift

NewInGeneral_general	Index	Column type	Min value	Max value	Nonzero ratio	Non-default-min ratio	Unique values	Mode	Mode frequency	Mean	Std
NumItemsInStock											
Period_general	22	Num	0	73.878	0.99284	1	95324	0	178954	5.4255	2.8843
Price_w_vat_rub	23	Num	0	34990	0.99087	1	41941	129	1042700	205.86	268.76
ProbabilityRepeatOrder7Days	24	Num	-1	0.91542	1	1	61513	-1	767283	0.21209	0.23851
StmProduct_general	25	Num	0	1	0.12373	1	2	0	23436771	0.12373	0.32927
TimeOfDay	26	Num	0	86399	1	1	85047	60787	314	47095	17001
TimesOrdered30Days	27	Num	-1	3.3346e+05	1	1	21370	-1	767312	17595	24410
TimesOrdered7Days	28	Num	-1	79455	0.99979	1	9357	-1	767305	4143.2	5881.6
TimesOrdered_general	29	Num	0	1.2126e+06	0.99284	1	31447	0	179073	41058	62541
itemPageCarousel_ctr_7	30	Num	-1	-1	0.97423	-1	18479	13862544	13862544	-0.43259	0.6035

Поломки в обучении модели

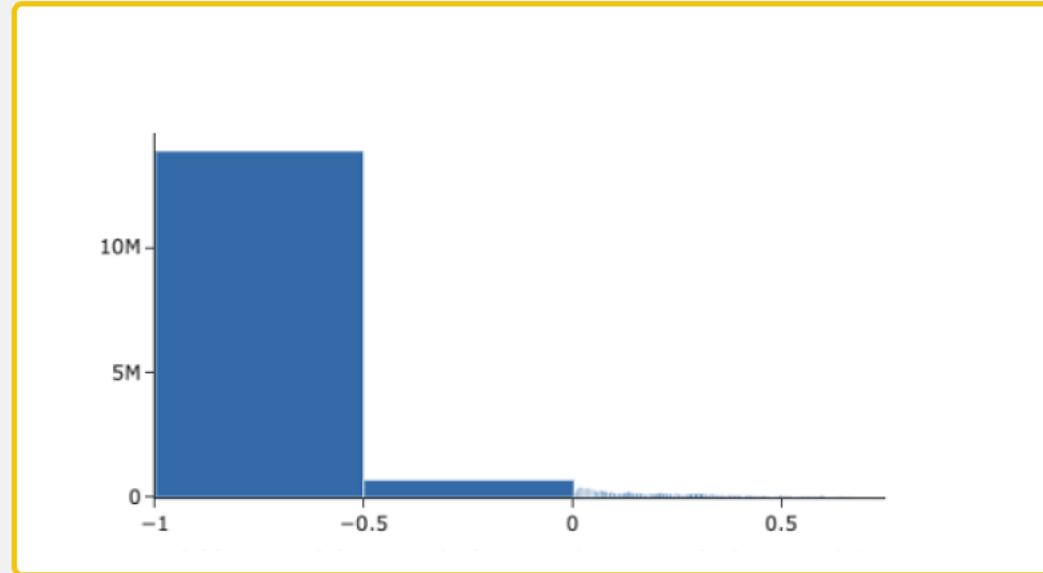
Сломанные фичи: валидация пула

✓ Гистограмма распределения фичи

✓ Изменение значений во времени

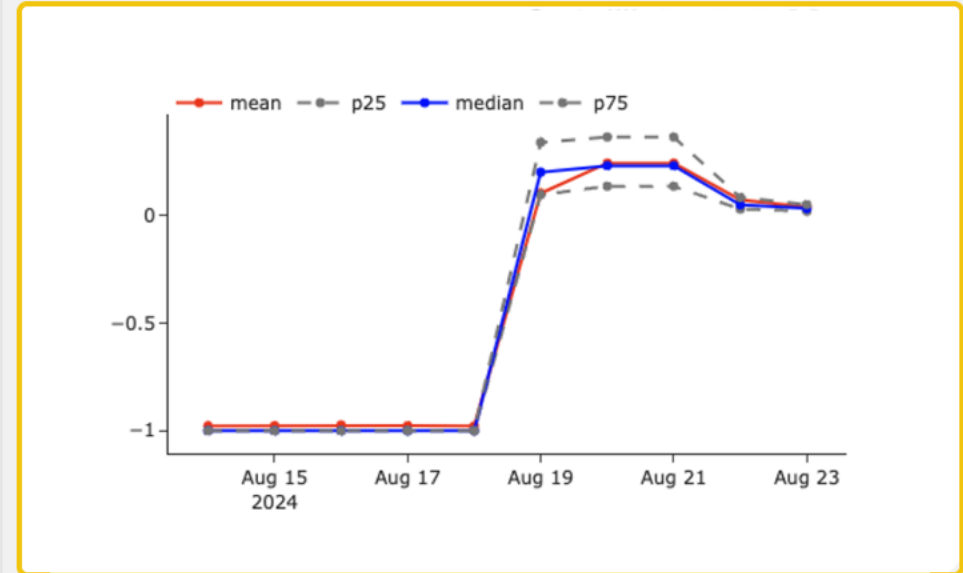
Histogram

▼ click to view



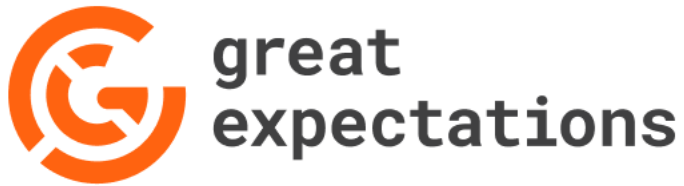
Time Series

▼ click to view



Поломки в обучении модели

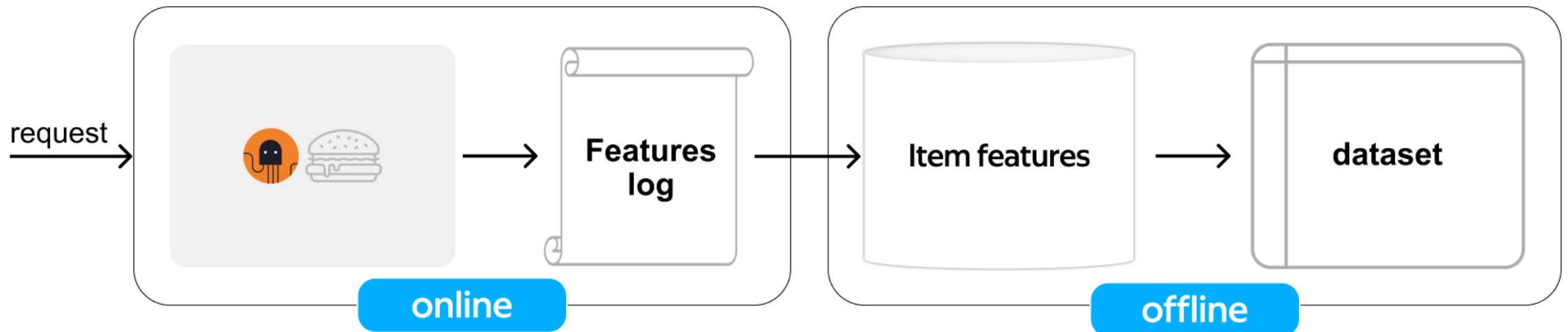
Data Quality: что предлагает рынок



Feature	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test	Drift Score
> target	num			Not Detected	Wasserstein distance (normed)	0.031996
> prediction	num			Not Detected	Wasserstein distance (normed)	0.013115
> AveBedrms	num			Not Detected	Wasserstein distance (normed)	0.055536
> AveRooms	num			Not Detected	Wasserstein distance (normed)	0.051092
> AveOccup	num			Not Detected	Wasserstein distance (normed)	0.048945
> MedInc	num			Not Detected	Wasserstein distance (normed)	0.038174
> Longitude	num			Not Detected	Wasserstein distance (normed)	0.031671

Поломки в обучении модели

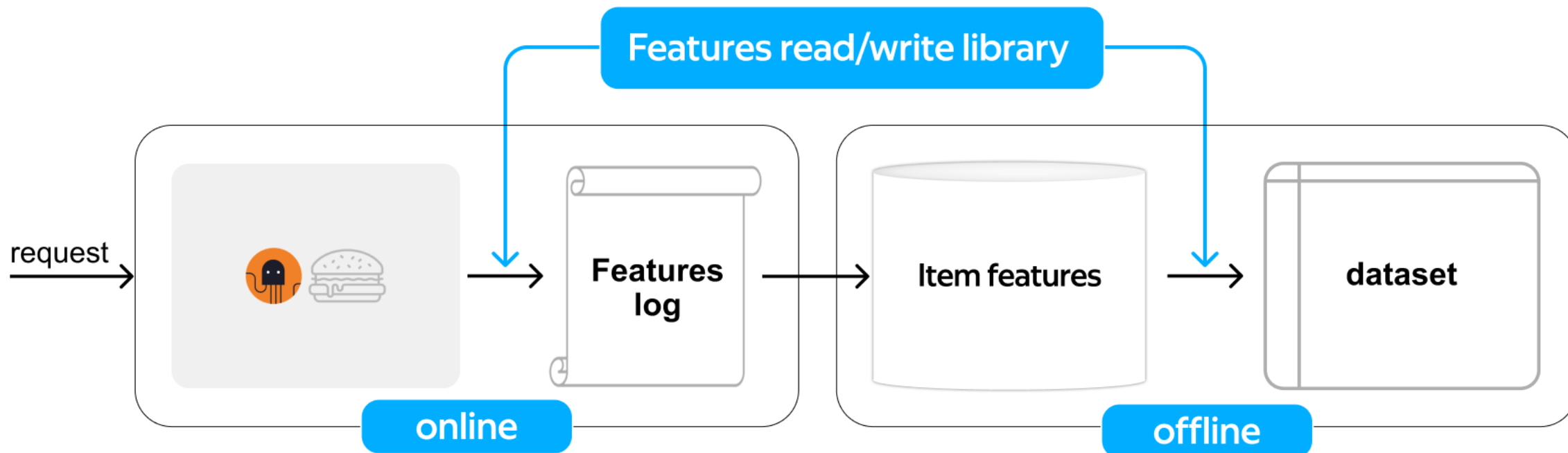
Расхождение способа обработки фичей



Поломки в обучении модели

Расхождение способа обработки фичей

✓ Один инструмент для работы с фичами C++, protobuf



Поломки в обучении модели

Подбор гиперпараметров

Применение оптимизатора на GridSearch, RandomSearch, BayesOpt

Поломки в обучении модели

Подбор гиперпараметров

✓ Применение оптимизатора на GridSearch, RandomSearch, BayesOpt



OPTUNA

Scikit-Optimize

Поломки в обучении модели

Подбор гиперпараметров

✓ Применение оптимизатора на GridSearch, RandomSearch, BayesOpt



O P T U N A Scikit-Optimize

✓ Ревью процесса обучения (code review)

Поломки в обучении модели

Подбор гиперпараметров

✓ Применение оптимизатора на GridSearch, RandomSearch, BayesOpt

🎯 OPTUNA Scikit-Optimize

✓ Ревью процесса обучения (code review)



Поломки в обучении модели

Разбиение датасета

✓ Исключение протекания “будущего” в обучающую выборку

✓ Учёт сезонности

Train

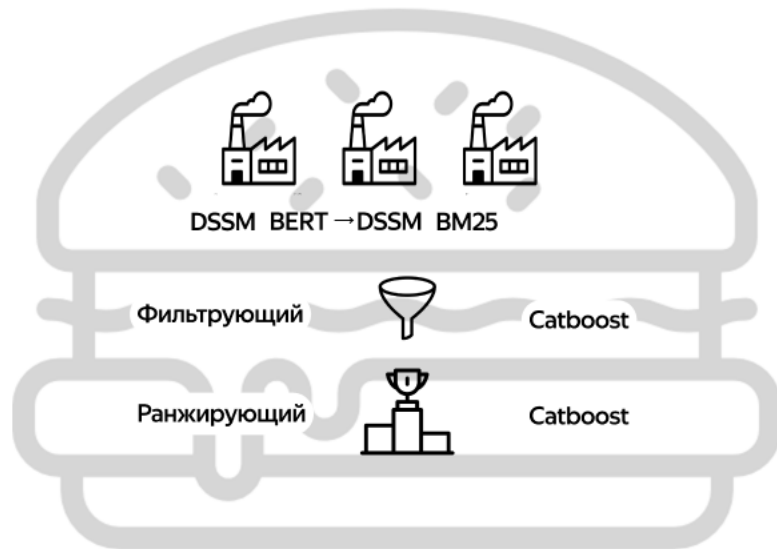
Validate

Test

Время

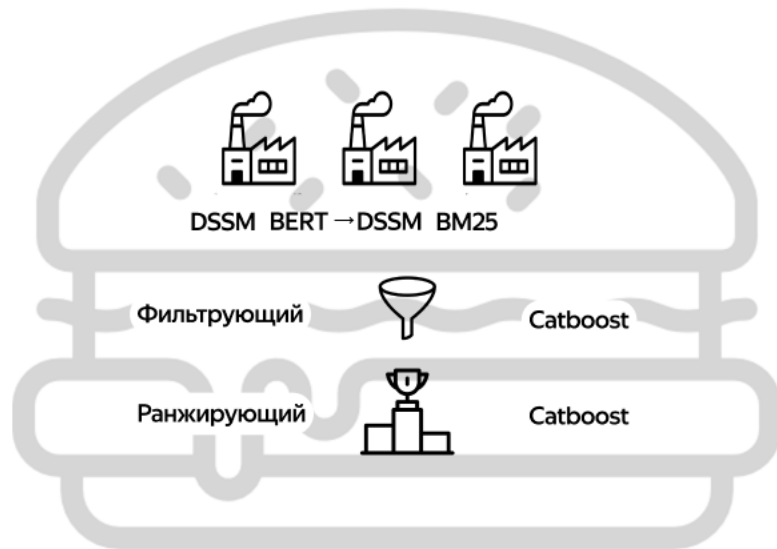
Сдвиг фиши в каскаде моделей

01 Проблема новых товаров



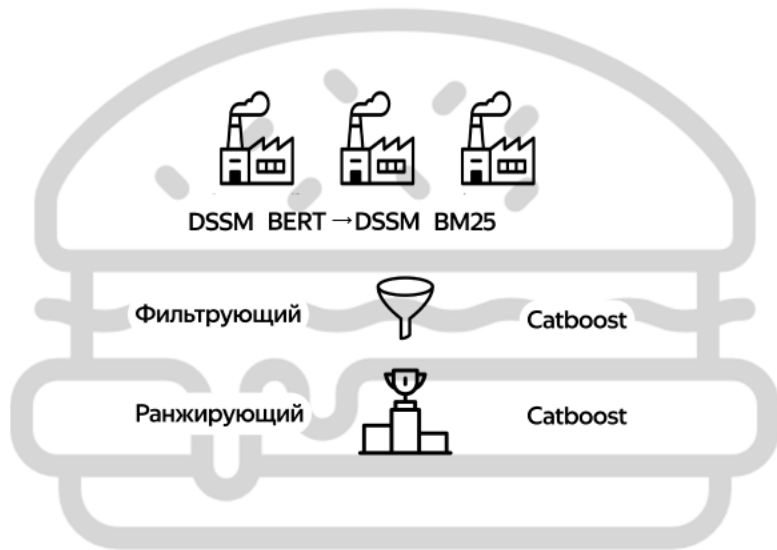
Сдвиг фиши в каскаде моделей

- 01 Проблема новых товаров
- 02 Влияние дообучения DSSM на score

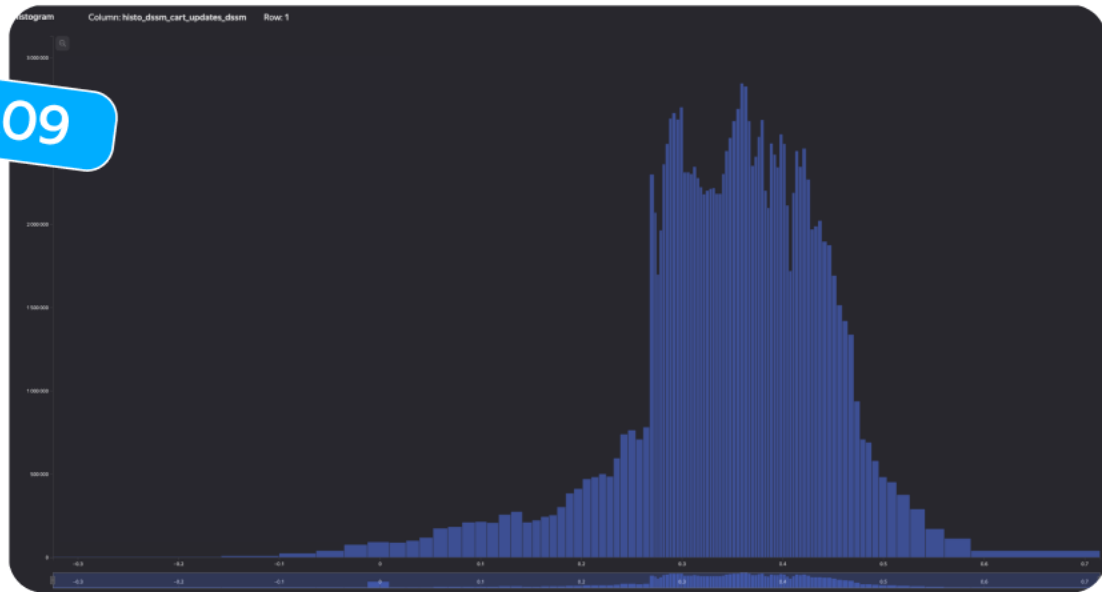


Сдвиг фиши в каскаде моделей

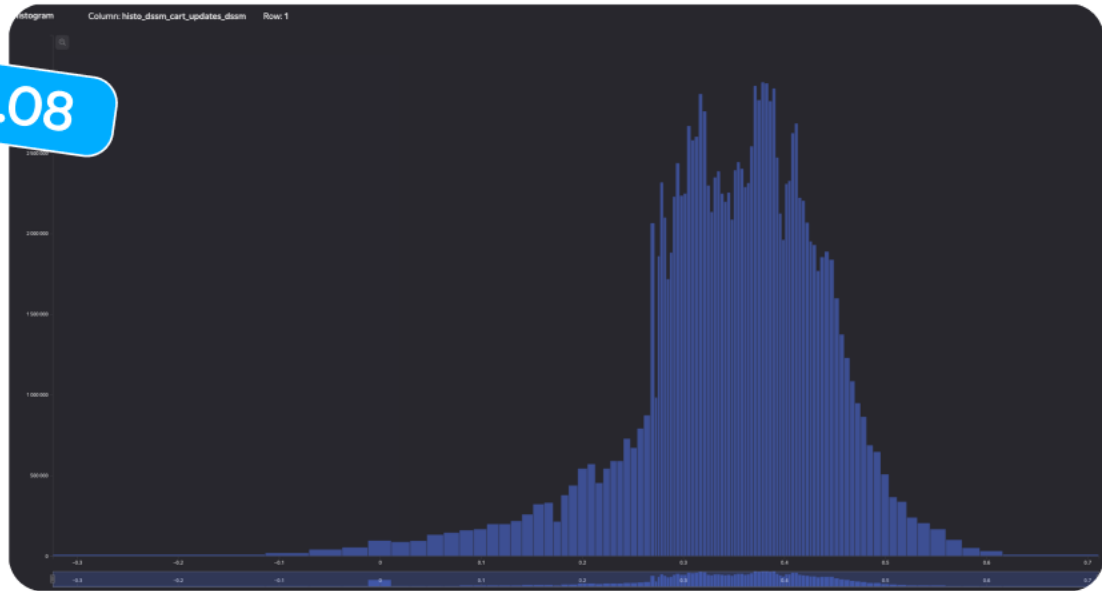
- 01 Проблема новых товаров
- 02 Влияние дообучения DSSM на score



11.09

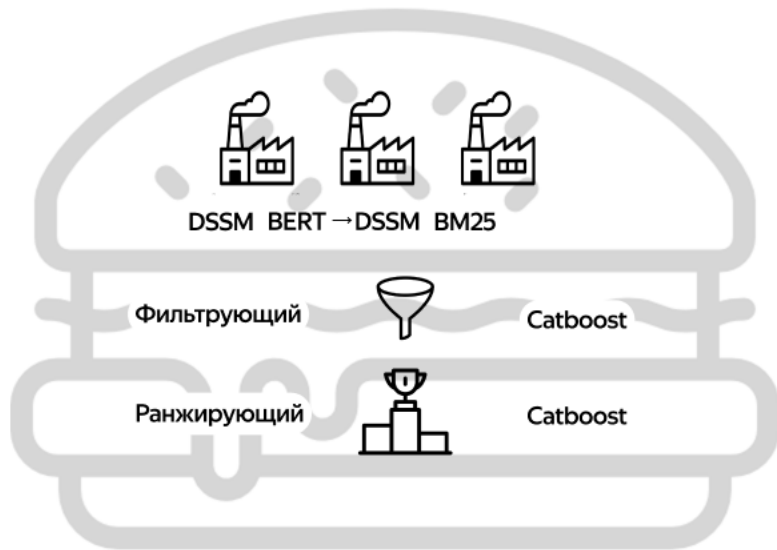


14.08



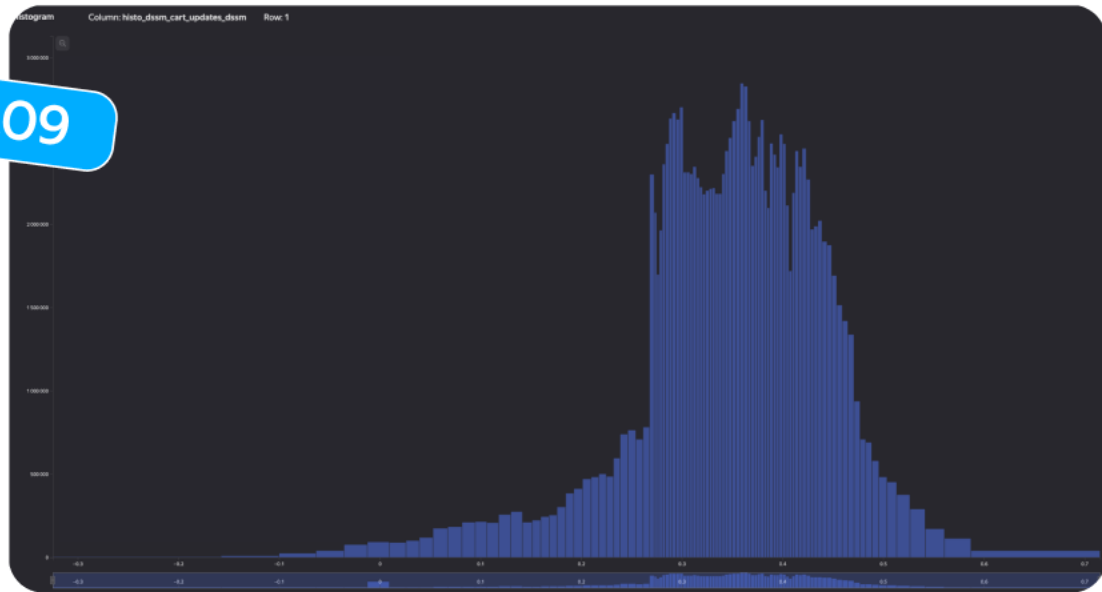
Сдвиг фичи в каскаде моделей

- 01 Проблема новых товаров
- 02 Влияние дообучения DSSM на score
- 03 Приёмка модели в конвейере

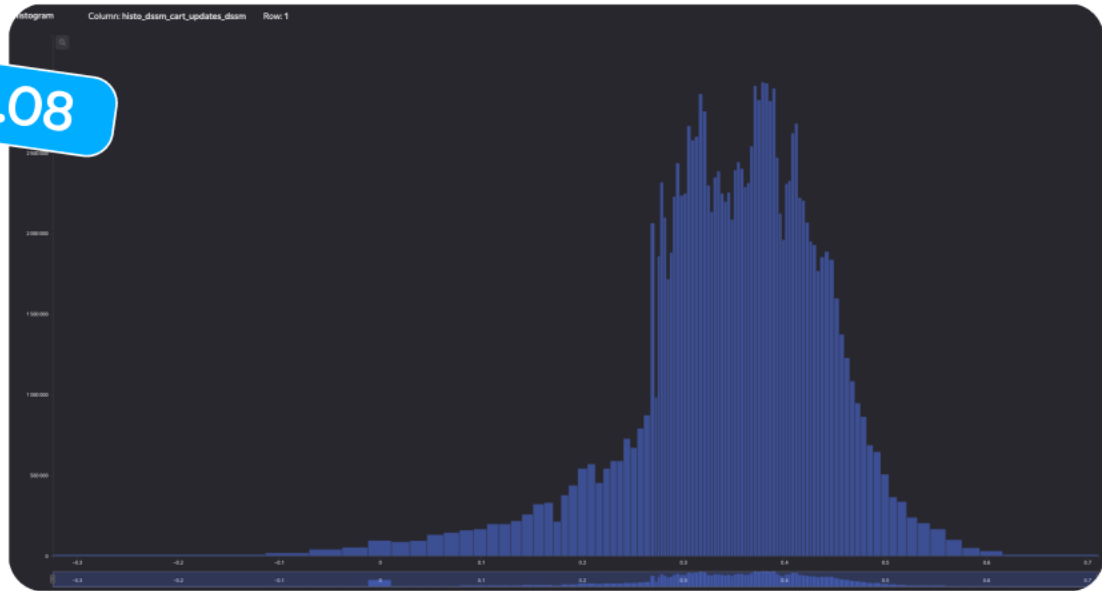


Коля Смирнов

11.09

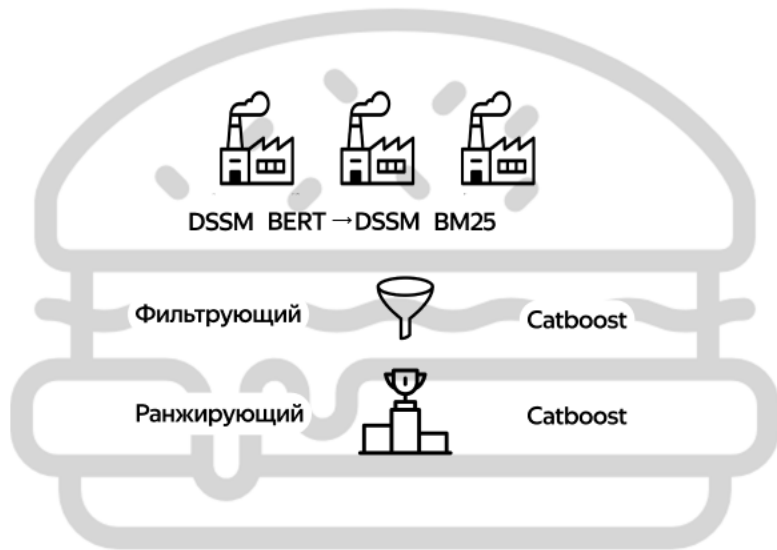


14.08

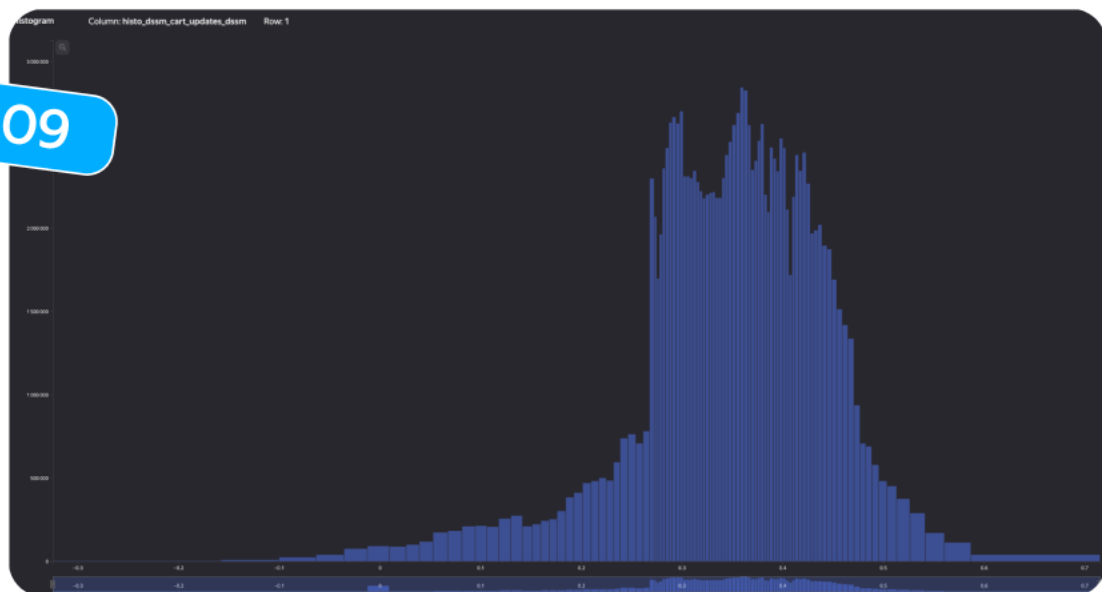


Сдвиг фиши в каскаде моделей

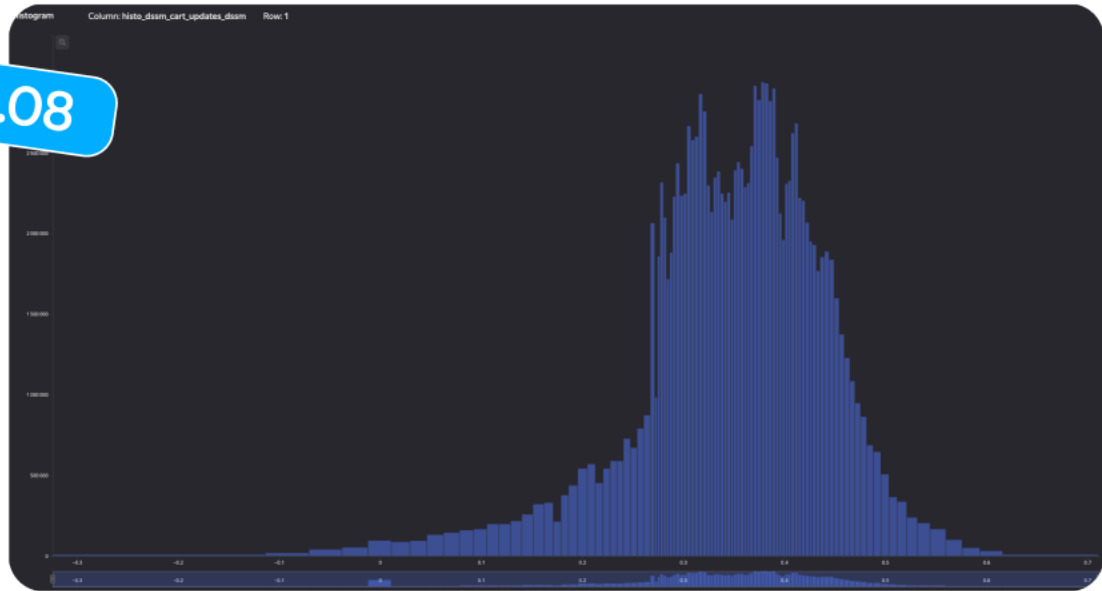
- 01 Проблема новых товаров
- 02 Влияние дообучения DSSM на score
- 03 Приёмка модели в конвейере
- 04 Обучение DSSM с нуля



11.09



14.08



Сдвиг фичи в каскаде моделей

Возможные решения

01 Скоринг старой и новой моделью → feature log

Сдвиг фичи в каскаде моделей

Возможные решения

- 01 Скоринг старой и новой моделью → feature log
- 02 Скоринг в A/B группе пользователей → feature log

Сдвиг фичи в каскаде моделей

Возможные решения

- 01 Скоринг старой и новой моделью → feature log
- 02 Скоринг в A/B группе пользователей → feature log

Обучение зависимых моделей



Сдвиг фичи в каскаде моделей

Возможные решения

- 01 Скоринг старой и новой моделью → feature log
- 02 Скоринг в A/B группе пользователей → feature log
- 03 Нормализация фичи

Обучение зависимых моделей



Обучение модели

Итоги

- + Отловили отсутствие фичей в пуле

Обучение модели

Итоги

- + Отловили отсутствие фичей в пуле
- + Подбор гиперпараметров дает ~3%% к качеству модели

Обучение модели

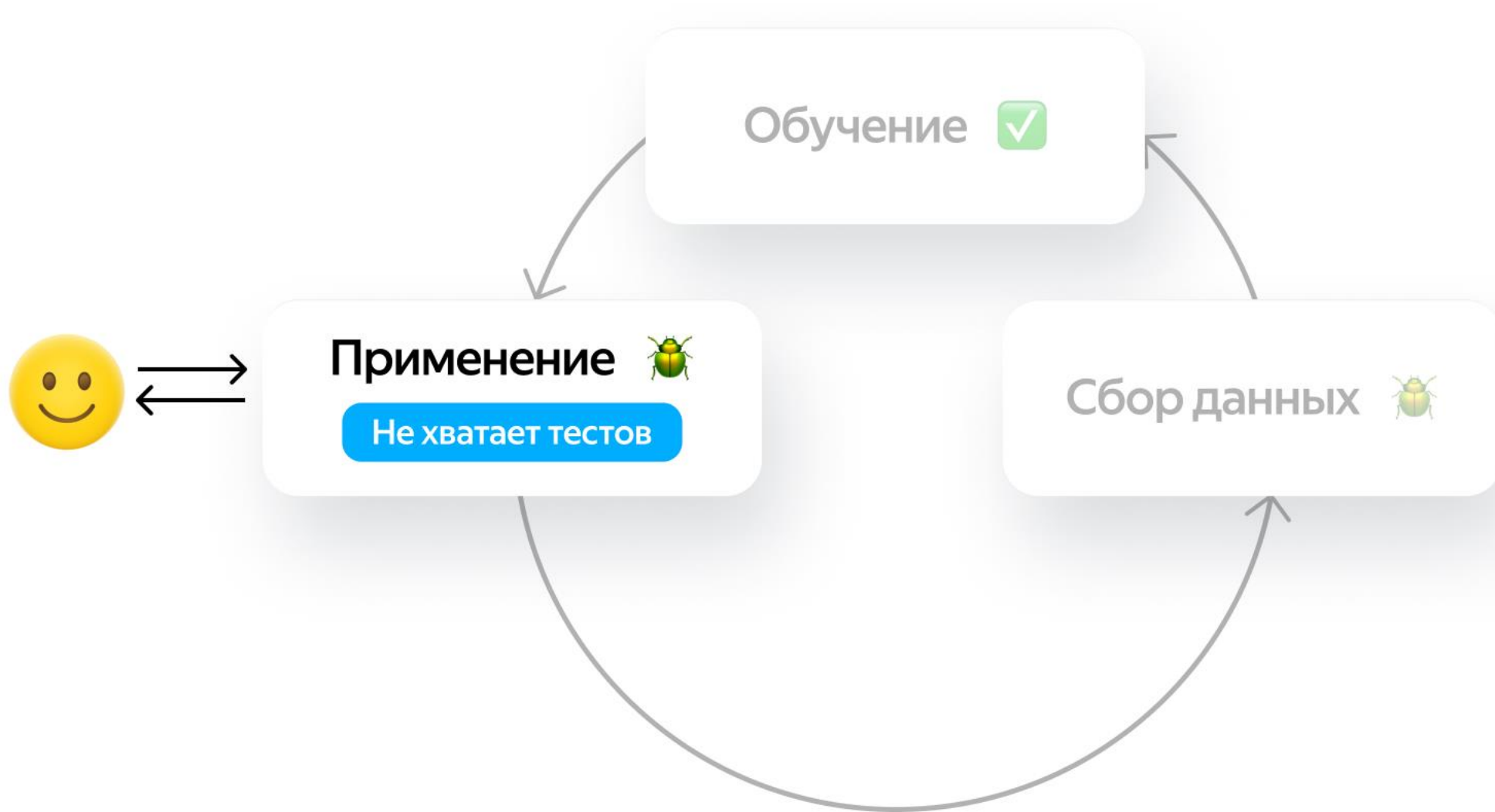
Итоги

- + Отловили отсутствие фичей в пуле
- + Подбор гиперпараметров дает ~3%% к качеству модели
- + Полечили DSSM, не сломав формулы

Обучение модели

Итоги

- + Отловили отсутствие фичей в пуле
- + Подбор гиперпараметров дает ~3%% к качеству модели
- + Полечили DSSM, не сломав формулы
- Проблема отсутствия кандгена осталась



Часть 03

Внедрение и применение модели в рантайме

Внедрение и применение модели

Интеграция модели в рантайм

Тестирование

Модульное

Интеграционное

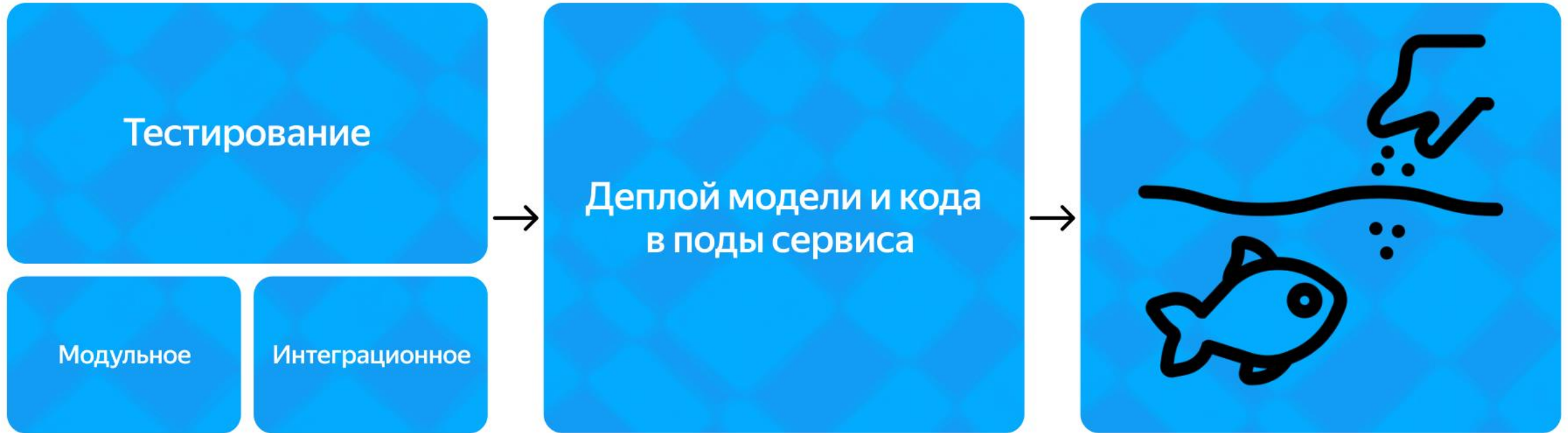
Внедрение и применение модели

Интеграция модели в рантайм

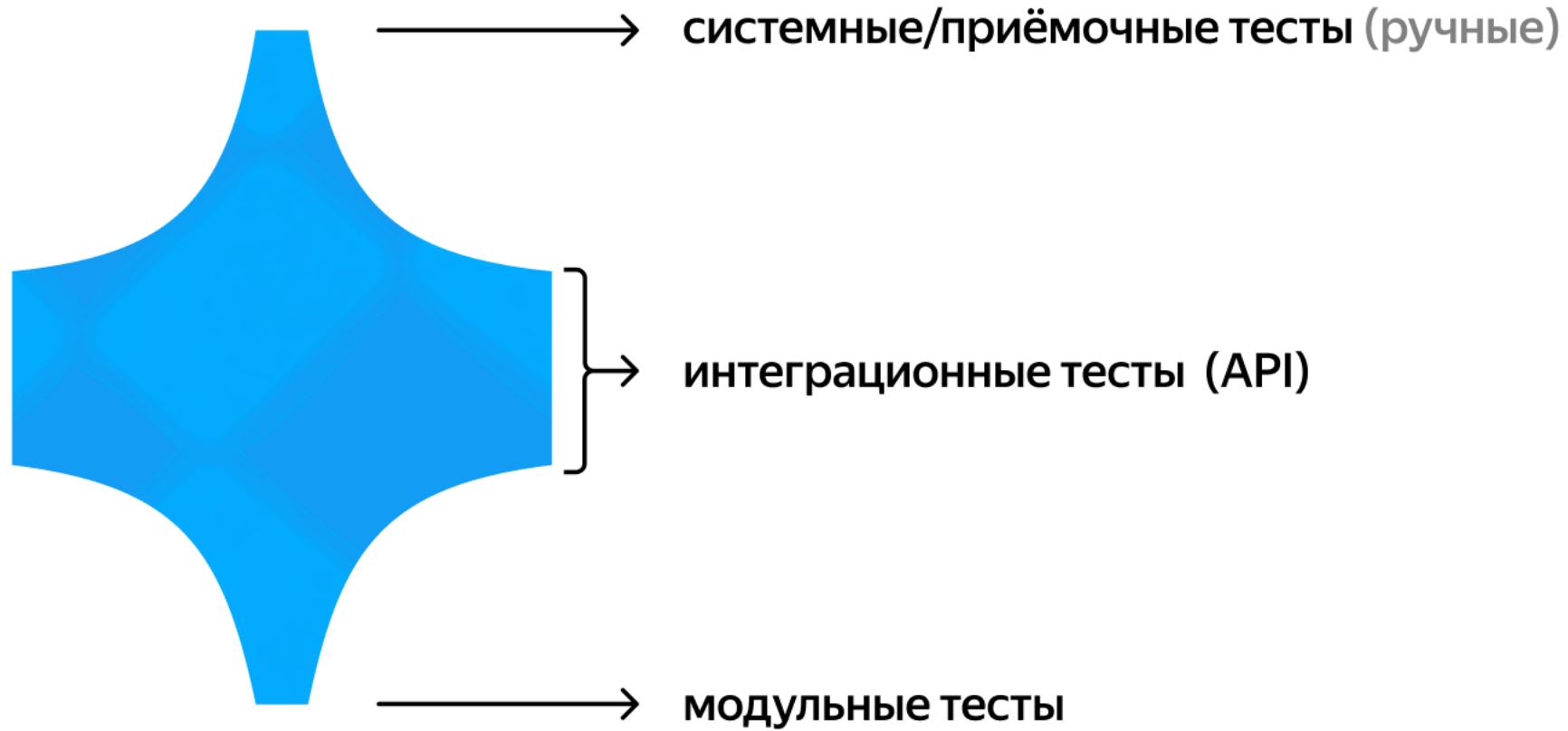


Внедрение и применение модели

Интеграция модели в рантайм



Внедрение и применение модели
Юла тестирования



Внедрение и применение модели

Особенности автотестов

- 01 Модульных тестов мало — фреймворк снижает сложность кода
- 02 Интеграционных много — больше покрытие и доверие
- 03 Необходимо подготовить тестовые данные (shard file)

Внедрение и применение модели

DJ Recommender: config

```
def search_program():
    recommender = Recommender()
    program = recommender.program(name="search")

    extract_dssm_query_embed = recommender.component(
        ExtractDssmQueryEmbed,
        'extract_dssm_query_embed',
        erf_input='search_query',
        input='query_embedding',
    )

    program.step(extract_dssm_query_embed)
```

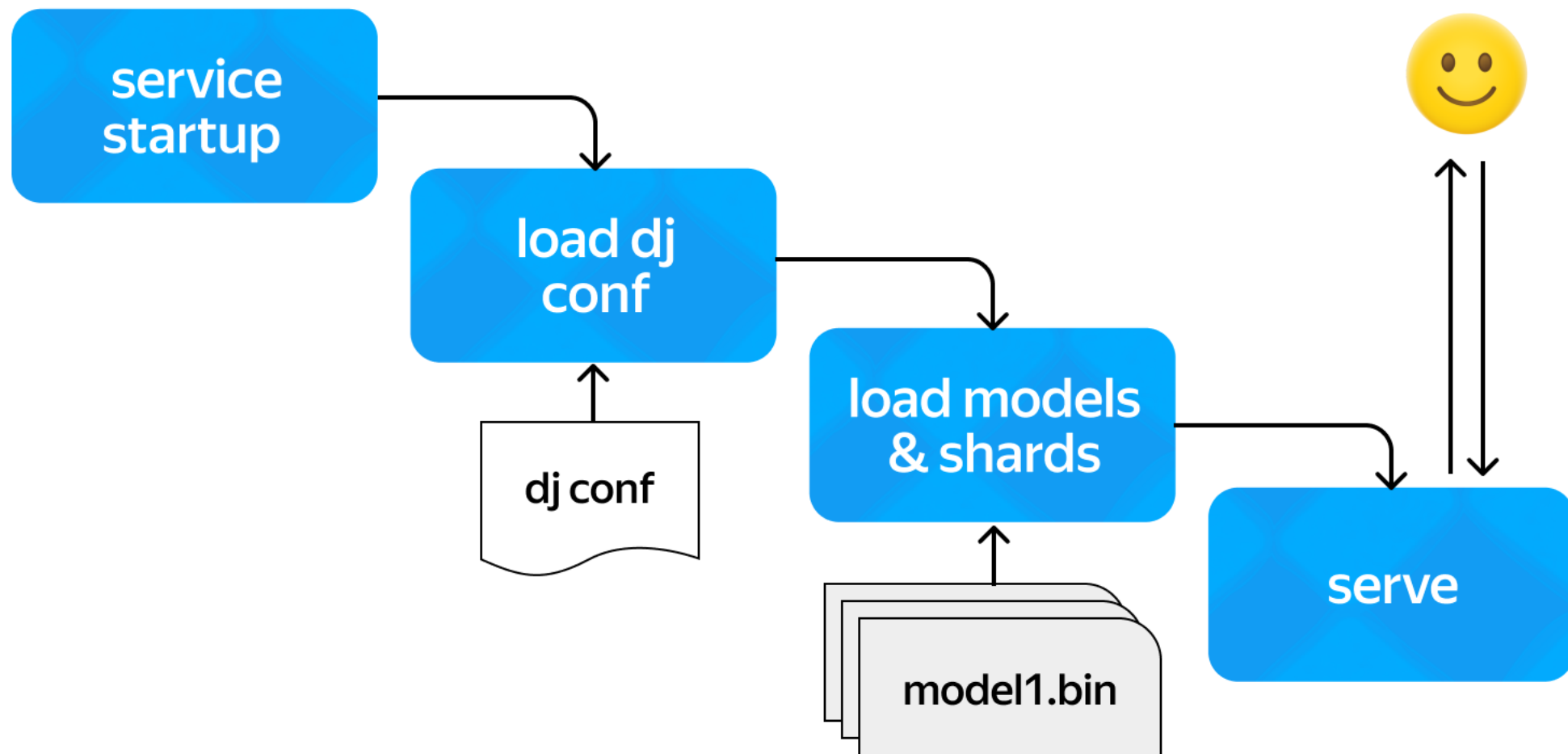
Внедрение и применение модели

DJ Recommender: config

```
shard = recommender.resource(Shard, 'shard.bin')
apply_dssm = recommender.component(
    ApplyModifyProfileRules,
    'apply_dssm_rules',
    shard,
)
program.step(apply_dssm)
# ...
program.step(extract_features)
program.step(rearrange_candidates)
program.step(log_component)
```

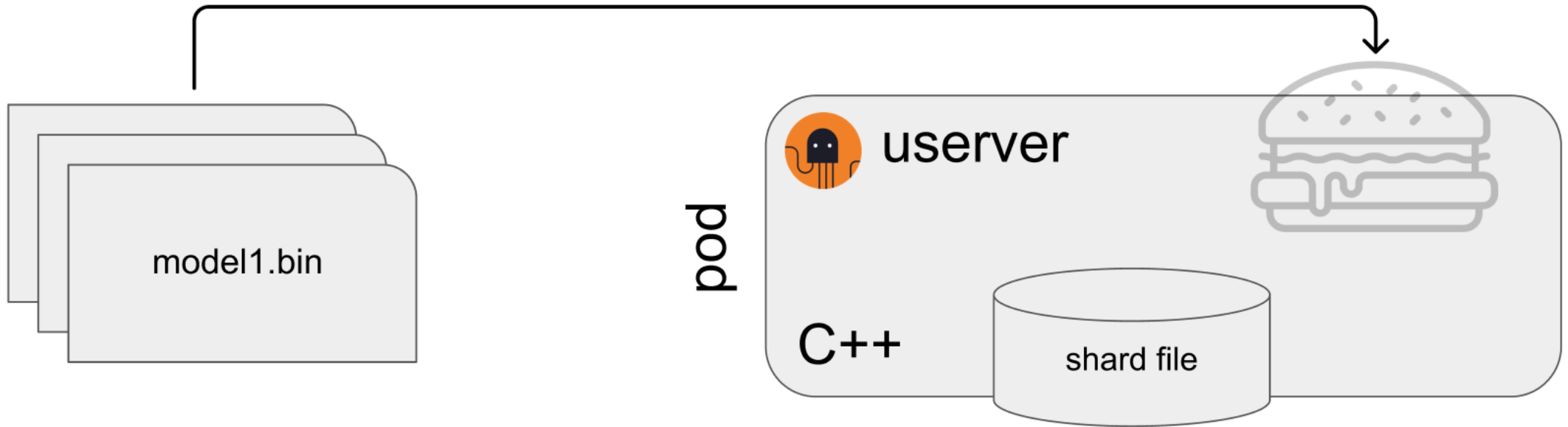
Внедрение и применение модели

DJ Recommender: runtime



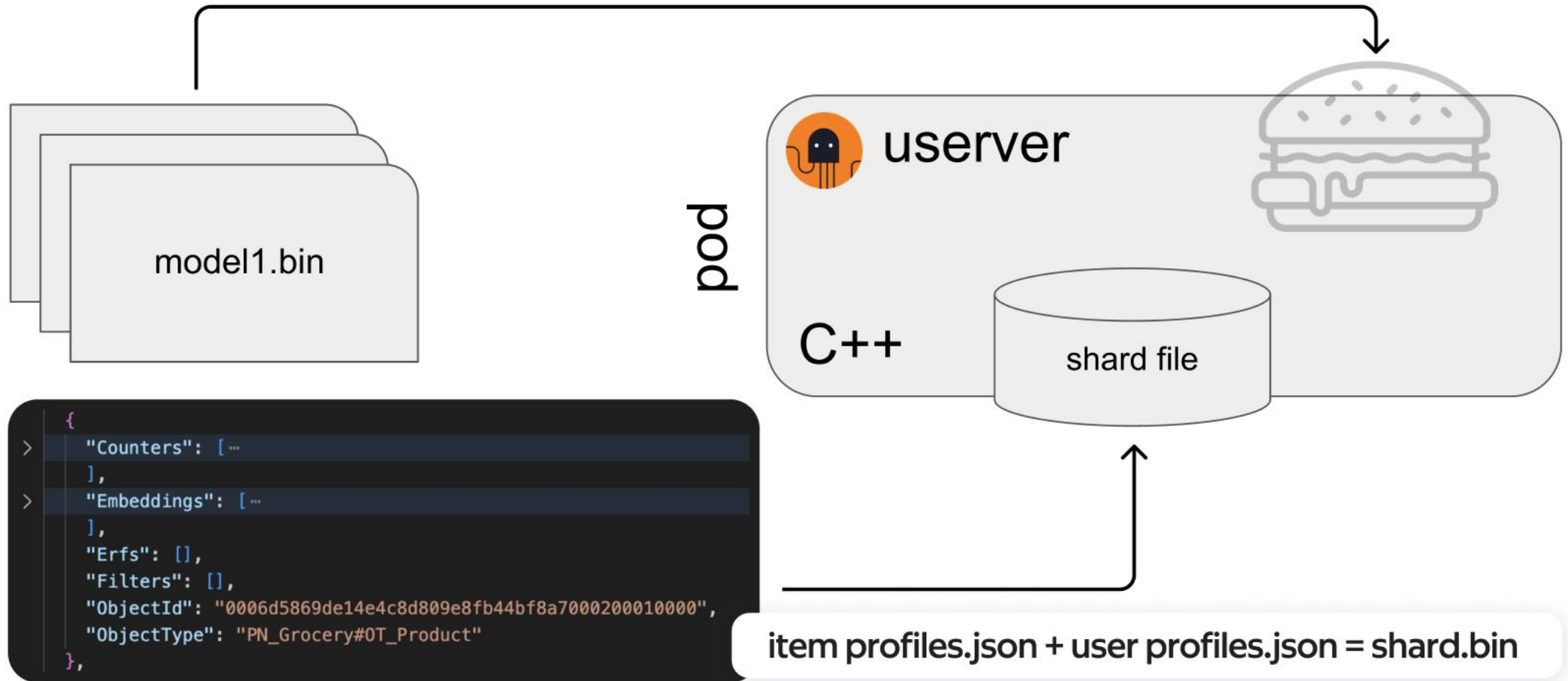
Внедрение и применение модели

Интеграция модели в рантайм



Внедрение и применение модели

Интеграция модели в рантайм



Внедрение и применение модели

Интеграционный API тест



```
@EXP_NO_BERT_CG
async def test_ml_search_no_bert(service_under_test):
    response = await service_under_test.post(
        '/grocery-search/v1/ml-search',
        data=json.dumps({'search_query': 'молоко'}))
    )
    expected_items = ['cart_update_dssm_08', 'cart_update_dssm_02']
    assert expected_items == response.json()['found_items']
```

Внедрение и применение модели

Когда пишем тесты

01 Внедрение новой модели

Внедрение и применение модели

Когда пишем тесты

- 01 Внедрение новой модели
- 02 Значительные изменения в модели

Внедрение и применение модели

Когда пишем тесты

- 01 Внедрение новой модели
- 02 Значительные изменения в модели
- 03 Добавление новых фичей

Внедрение и применение модели

Когда пишем тесты

- 01 Внедрение новой модели
- 02 Значительные изменения в модели
- 03 Добавление новых фичей
- 04 Изменение бизнес-логики процессинга

Внедрение и применение модели

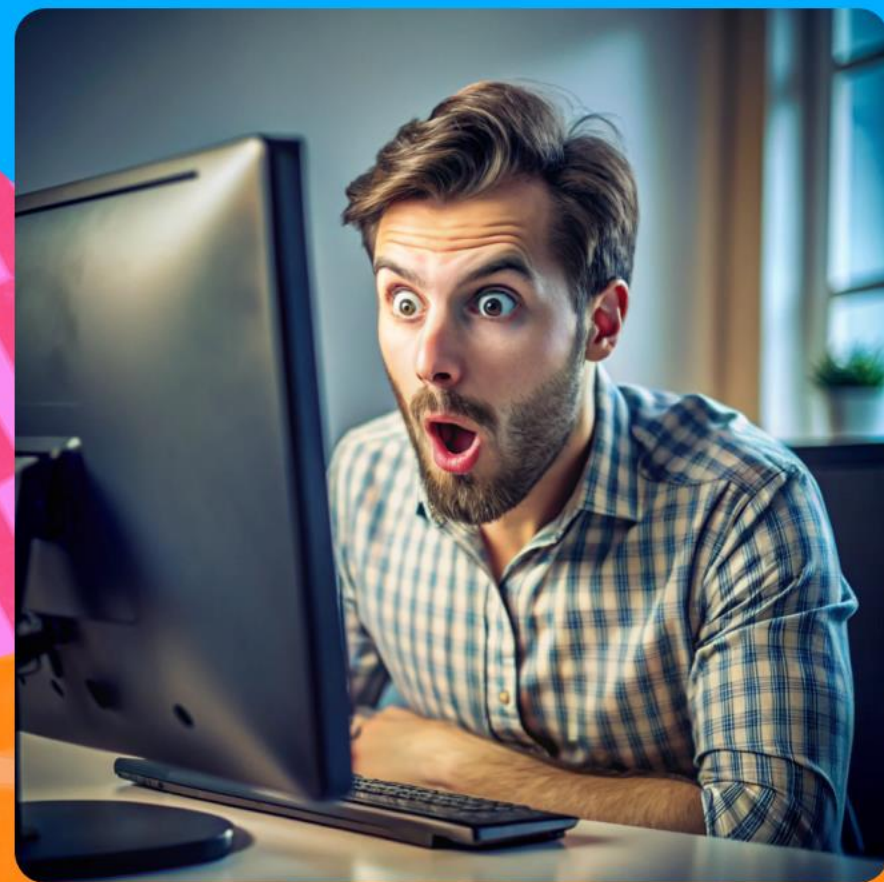
Когда пишем тесты

- 01 Внедрение новой модели
- 02 Значительные изменения в модели
- 03 Добавление новых фичей
- 04 Изменение бизнес-логики процессинга
- 05 Для уверенности в себе!



Системные тесты

- 01 Задеплоить модель в прод
- 02 Включить модель на команду разработки
- 03 Посмотреть глазами

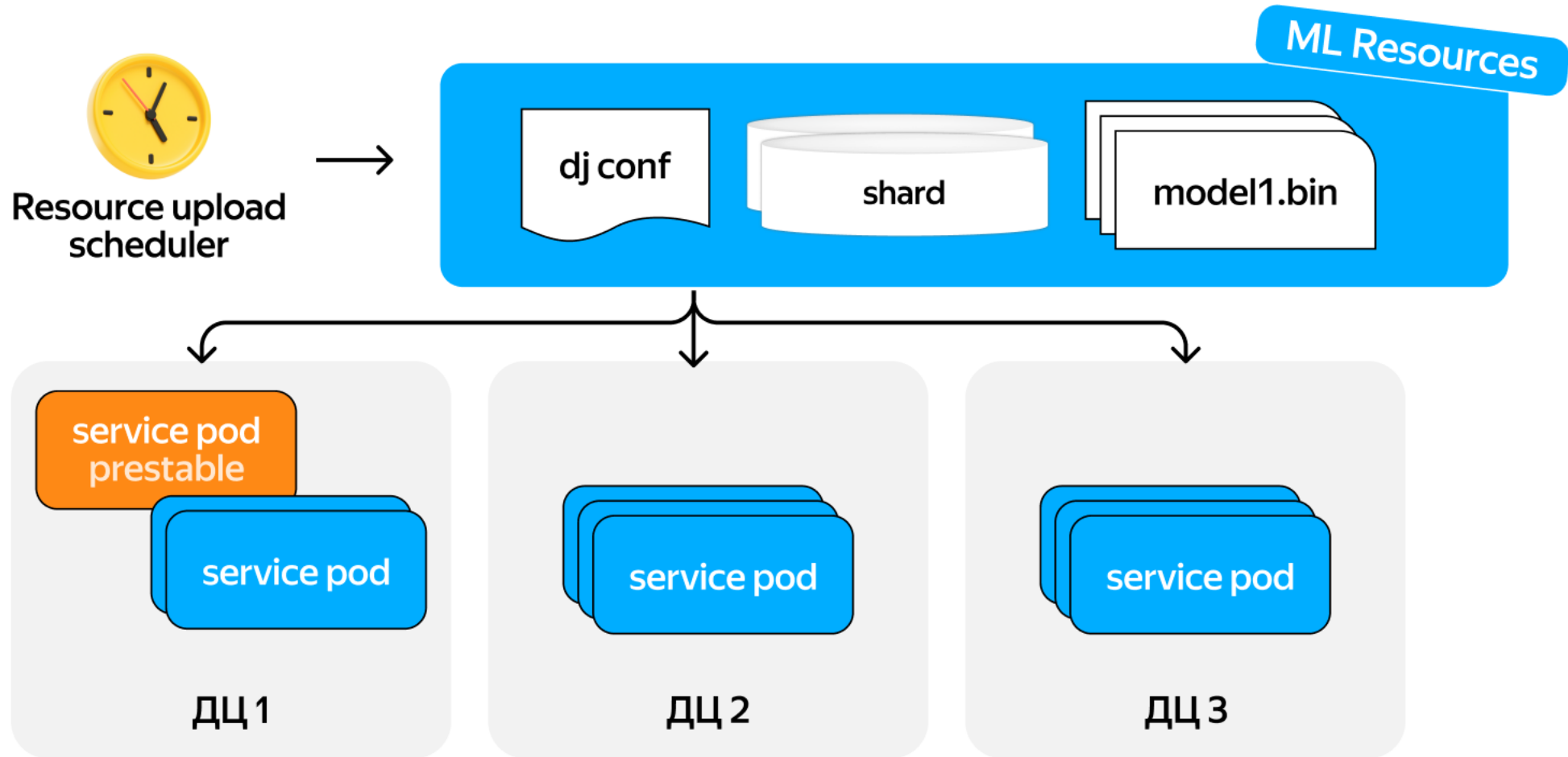


Внедрение и применение модели **Deployment**



Внедрение и применение модели

Deployment



Внедрение и применение модели

Особенности развёртывания

01 Prestable — первая фаза релиза кода сервиса

Внедрение и применение модели

Особенности развёртывания

- 01 Prestable — первая фаза релиза кода сервиса
- 02 ML Resource катается независимо от релиза сервиса
 - Конвейер моделей
 - Конвейер shard-файлов

Внедрение и применение модели

Особенности развёртывания

- 01 Prestable — первая фаза релиза кода сервиса
- 02 ML Resource катается независимо от релиза сервиса
 - Конвейер моделей
 - Конвейер shard-файлов
- 03 В случае отказа сервиса работает fallback (aka baseline)























“
Три движка для одной Лавки
как эволюционировала система поиска в сервисе
”



Внедрение и применение модели **Системные тесты**

«Помидоры черный принц» с разным порогом фильтр-формулы

45.		Прокладки Kotex Ultra Soft product_id: 69238206e external_id: 10004834
46.		Умный свет Яндекс Лампа product_id: d31ae04c7 external_id: 14019
47.		Маска для лица Aravia product_id: a30c5c6fd3 external_id: 10043823
48.		Шампунь и гель для душа product_id: 7ebd56a10 external_id: 10006489
49.		Умный свет Яндекс Лампа product_id: ab79159b6 external_id: 10015938
50.		
16.		Томаты маринованные product_id: ae23c5878 external_id: 10041228
17.		Томаты протёртые Pomi product_id: 083ebf673c external_id: 10037065
18.		Томаты маринованные product_id: f21c40a31b external_id: 1416
19.		Ассорти овощное 6 соевых соусов product_id: 7b1b1abfd7 external_id: 9740
20.		Томатная паста Pomi product_id: 50e2ff85b1 external_id: 2434
21.		
7.		Томаты черри сливовидные product_id: e2aa2580086 external_id: 10030942
8.		Помидоры Бычье сердце product_id: 8c47db06947 external_id: 10323
9.		Помидоры черри product_id: b6648ef47797 external_id: 6573
10.		Томаты черри маринованные product_id: ee556529e28 external_id: 10041229
11.		Ассорти помидоры и огурцы product_id: 6544f4da772 external_id: 12878
12.		
5.		Помидоры Бычье сердце Дагестан product_id: 5f6f719176cf4e5f9f3d5551293240 external_id: 2365
6.		Помидоры черри на ветке product_id: 3b323e490d994d2cbb5021adde74 external_id: 6576
7.		Томаты черри сливовидные product_id: e2aa258008694ce48f7556584f16f6 external_id: 10030942
8.		Помидоры Бычье сердце Азербайджан product_id: 8c47db06947a44b998d0d28376ac external_id: 10323
9.		Помидоры черри product_id: b6648ef477974120b894c9b1484f6 external_id: 6573
10.		

Внедрение и применение модели

DJ Viewer: режим дебага

Handler

Headers

Parameters

DjParams


Search query

Debug mode

Local viewer

[Recommend](#)

[Result](#) [Items](#) [Profiles](#) [Parameters](#) [Statements](#)



Улитки «Лавка 100» в панцире

product_id: 8ba8fc3b22804700be2baad4858f92b600

external_id: 10050985

Внедрение и применение модели

DJ Viewer: режим дебага

Handler:

Headers:

Parameters:

DjParams:

Search query:

Debug mode:

Local viewer:

[Recommend](#)

Result [Items](#) [Profiles](#) [Parameters](#) [Statements](#)



Улитки «Лавка 100» в панцире
product_id: 8ba8fc3b22804700be2baad4858f92b600
external_id: 10050985

	Dura...	Name	Item...	Item...	Item...	Diff
1	397 ms	search_with_promo	0	113	113	Show
2	390 ms	└─extract_features_and_ranking	0	113	113	Show
3	376 ms	└─log_search_result	0	113	113	Show
4	0 ms	└─dssm_applier_click_dssm	0	0	0	Show
5	0 ms	└─dssm_applier_dssm_model_version_0	0	0	0	Show
6	0 ms	└─dssm_applier_bert_distill_dssm	0	0	0	Show
7	0 ms	└─dssm_applier_bert_distill_dssm_model_versio...	0	0	0	Show
8	0 ms	└─bm25_tokenizer	0	0	0	Show
9	130 ms	└─master_cg_search	0	1152	1152	Show
10	25 ms	└─bm25_candidates_by_shard	0	600	600	Show
11	23 ms	└─default_bm25_candidates	0	600	600	Show

Внедрение и применение модели

DJ Viewer: режим дебага

Result Items Profiles Parameters Statements

```
[
  {
    "FloatFeatures": {
      "category_min_action_pmi_to_context": 0,
      "category_min_npmi_to_hist": 0,
      "category_min_pmi_to_hist": 0,
      "coef_b": 0.08271999657154083,
      "dssm_cart_updates_dssm": 0.49870139360427856,
      "dssm_threshold": 0.27000001072883606,
      "feed_conversion_30": 0.006191950291395187,
      "feed_conversion_7": 0.0038910505827516317,
      "feed_conversion_90": 0.0008611715747974813,
      "feed_ctr_30": 0.02683178521692753,
      "feed_ctr_7": 0.031128404662013054,
      "feed_ctr_90": 0.018482066690921783,
      "feed_view_30": 969,
      "feed_view_7": 257,
      "feed_view_90": 105670,
      "filter_fm1_score": 0.36848291754722595,
```

Result Items Profiles Parameters Statements

	Dura...	Name	Item...	Item...	Item...	Diff
1	397 ms	search_with_promo	0	113	113	Show
2	390 ms	└─extract_features_and_ranking	0	113	113	Show
3	376 ms	└─log_search_result	0	113	113	Show
4	0 ms	└─dssm_applier_click_dssm	0	0	0	Show
5	0 ms	└─dssm_applier_dssm_model_version_0	0	0	0	Show
6	0 ms	└─dssm_applier_bert_distill_dssm	0	0	0	Show
7	0 ms	└─dssm_applier_bert_distill_dssm_model_versio...	0	0	0	Show
8	0 ms	└─bm25_tokenizer	0	0	0	Show
9	130 ms	└─master_cg_search	0	1152	1152	Show
10	25 ms	└─bm25_candidates_by_shard	0	600	600	Show
11	23 ms	└─default_bm25_candidates	0	600	600	Show

Результаты внедрения

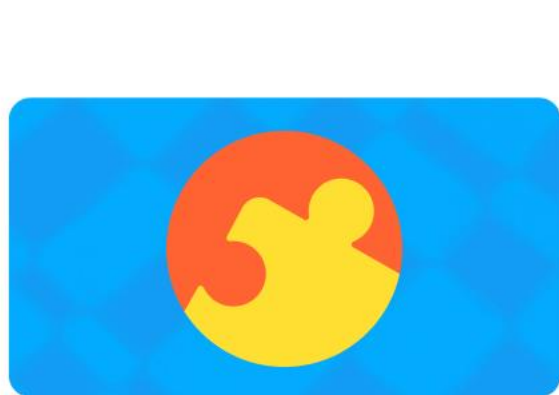
- 01 API тесты дают уверенность в деплое
- 02 Проблема отсутствия кандгена решена 🥳
- 03 Системные тесты
 - Адекватность модели
 - Отладка и настройка

Часть 04

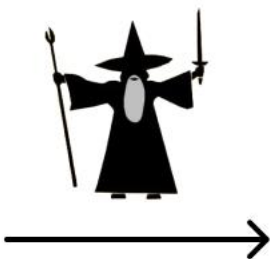
Наблюдение модели и сбор данных

Наблюдение и сбор данных

Наблюдение и анализ



Офлайн-оценка качества



A/B-тестирование




Непрерывный мониторинг

Офлайн-оценка качества поисковой выдачи

Оцените, насколько запрос и товар соответствуют

- rel+** В запросе есть и продукт (перец), и хотя бы одна его характеристика (острый)
- rel** В запросе есть продукт, но нет характеристик
- rel-** Есть противоречивые характеристики запроса и товара
- non-rel** Вообще не то

🔍 красный перец острый ✕



-40%

185₽ ~~309₽~~

Перец микс
500 г

В корзину

Наблюдение и сбор данных

Обстрел модели корзиной запросов



Наблюдение и сбор данных

Обстрел модели корзиной запросов



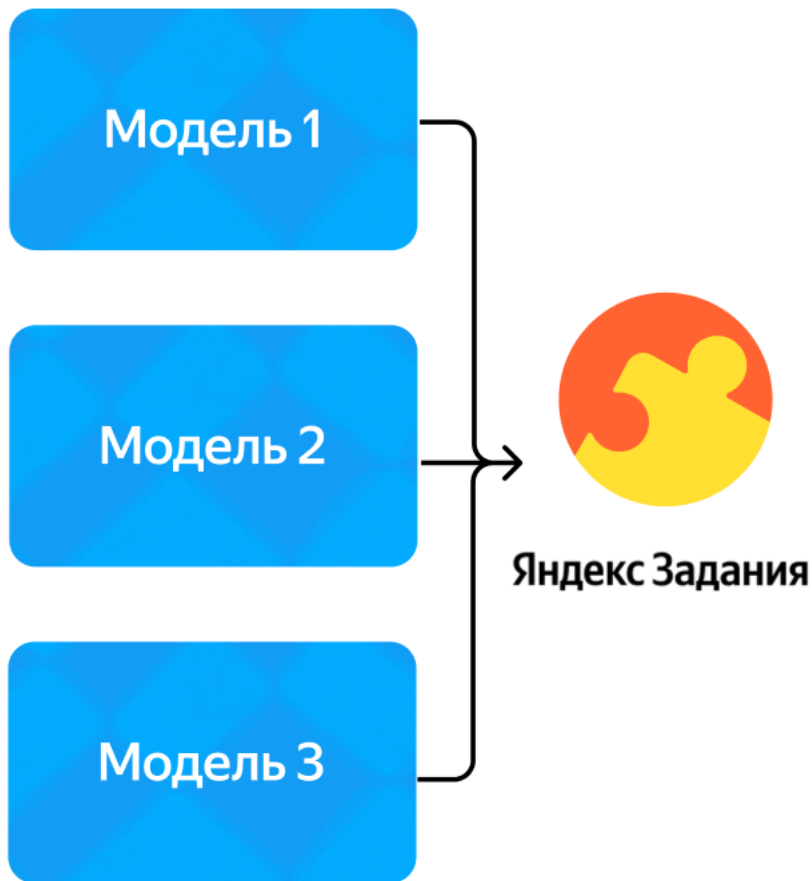
Модель 1

Модель 2

Модель 3

Наблюдение и сбор данных

Обстрел модели корзиной запросов



Наблюдение и сбор данных

Обстрел модели корзиной запросов



Модель 1

Модель 2

Модель 3



Яндекс Задания



Наблюдение и сбор данных

Офлайн-оценка качества поисковой выдачи

The screenshot displays a web interface for search engine evaluation. At the top, there are tabs for 'Calculations' (active) and 'Enrichments' (with a count of 4). A '+ Add enrichment' button is visible next to the 'Enrichments' tab. Below the tabs, there are three buttons: 'Expand all groups' (with a checkmark), 'Only significant metrics', and a search box labeled 'Search metric...'. The main content area is a list of metrics grouped into sections:

- gs_mean_serp_len**
 - gs_mean_serp_len
- gs_filtered_out_queries**
 - gs_filtered_out_queries
- nDCG**
 - gs_ndcg_at_5
 - gs_ndcg_at_9
 - gs_ndcg_at_17
 - gs_ndcg_at_40
 - gs_ndcg_at_80
 - gs_ndcg_total
- pfound**
 - gs_pfound_at_5
 - gs_pfound_at_9
 - gs_pfound_at_17
 - gs_pfound_at_40
 - gs_pfound_at_80
 - gs_pfound_total

Наблюдение и сбор данных

Офлайн-оценка качества поисковой выдачи

The screenshot displays a dashboard for search quality evaluation. At the top, there are several tabs for different filter levels: 'filter low', 'filter med', 'filter high', and 'prod'. Below these, a 'COMPONENT FILTER' dropdown is set to 'Wizard off'. The main section is titled 'Calculations' and includes a search bar for metrics. The left sidebar lists several metric groups:

- gs_mean_serp_len**
- gs_filtered_out_queries**
- nDCG**
- pfound**

Each group contains a list of specific metrics, such as 'gs_ndcg_at_5', 'gs_ndcg_at_9', 'gs_ndcg_at_17', 'gs_ndcg_at_40', 'gs_ndcg_at_80', and 'gs_ndcg_total' for nDCG. The 'pfound' group includes 'gs_pfound_at_5', 'gs_pfound_at_9', 'gs_pfound_at_17', 'gs_pfound_at_40', 'gs_pfound_at_80', and 'gs_pfound_total'. The right side of the dashboard is currently blank, suggesting a table of results is present but not visible.

Наблюдение и сбор данных

Офлайн-оценка качества поисковой выдачи

Open details Add serpset Edit serpsets

56785954 2024-08-30 16:34 filter low 56785920 2024-08-30 16:16 filter med 56785920 2024-08-30 16:31 filter high 56785817 2024-08-30 16:22 prod

COMPONENT FILTER Wizard off

Calculations Enrichments 4 Add enrichment

Expand all groups Only significant metrics Search metric...

	VALUE	DIFF	% DIFF	VALUE	DIFF	% DIFF	VALUE	DIFF	% DIFF	BASELINE
gs_mean_serp_len										
gs_mean_serp_len	45.61745	-22.745	-33.27%	34.99661	-33.2542	-48.72%	29.945392	-38.1638	-56.03%	68.109215
gs_filtered_out_queries										
gs_filtered_out_queries	0	0	0.00%	0	0	0.00%	0	0	0.00%	0
nDCG										
gs_ndcg_at_5	0.970703	0.0006	+0.06%	0.979455	0.0009	+0.10%	0.985731	0.0039	+0.40%	0.981782
gs_ndcg_at_9	0.968772	0.0018	+0.18%	0.977736	0.0024	+0.24%	0.984575	0.006	+0.61%	0.978621
gs_ndcg_at_17	0.966961	0.002	+0.21%	0.976598	0.0033	+0.34%	0.983801	0.0071	+0.73%	0.976684
gs_ndcg_at_40	0.965238	0.0026	+0.27%	0.975675	0.0047	+0.49%	0.983151	0.0088	+0.91%	0.974327
gs_ndcg_at_80	0.964516	0.0032	+0.33%	0.975077	0.0054	+0.56%	0.982662	0.0096	+0.99%	0.973052
gs_ndcg_total	0.964463	0.0033	+0.34%	0.975064	0.0055	+0.57%	0.98263	0.0097	+1.00%	0.972894
pfound										
gs_pfound_at_5	0.882857	0.0016	+0.19%	0.889693	-0.0003	-0.03%	0.888898	-0.0051	-0.57%	0.893988
gs_pfound_at_9	0.88519	0.0014	+0.16%	0.891677	-0.0009	-0.10%	0.890584	-0.006	-0.67%	0.896615
gs_pfound_at_17	0.885395	0.0014	+0.16%	0.891772	-0.001	-0.11%	0.890617	-0.0062	-0.69%	0.896805
gs_pfound_at_40	0.885436	0.0013	+0.15%	0.891781	-0.0011	-0.13%	0.890618	-0.0063	-0.70%	0.896916
gs_pfound_at_80	0.885443	0.0013	+0.14%	0.891783	-0.0012	-0.13%	0.890618	-0.0063	-0.71%	0.896964
gs_pfound_total	0.885444	0.0013	+0.14%	0.891783	-0.0012	-0.13%	0.890618	-0.0063	-0.71%	0.896964

Офлайн-оценка качества модели

01 Объективность

Офлайн-оценка качества модели

- 01 Объективность
- 02 Подбор параметров

Офлайн-оценка качества модели

- 01 Объективность
- 02 Подбор параметров
- 03 Оцифровка качества

Офлайн-оценка качества модели

- 01 Объективность
- 02 Подбор параметров
- 03 Оцифровка качества
- 04 Принятие решений о «размене»

A/B-тестирование перед запуском

Как реальные пользователи Яндекс Лавки оценят нововведение?

Просмотр эксперимента grocery_search_common

Подробная документация по работе с экспериментами 3.0 в [docs](#).

Аналитика и расчёты

Laba

New

Посмотреть историю

время последнего изменения: 11.10.20

Доля пользователей в группе

1/3

1/3

1/3

> 16. **russia_control** (russia_control)

> 17. **russia_test_old_click_dssm_embed** (russia_test_old_click_dssm_embed)

> 18. **russia_test_retrained_click_dssm_embed** (russia_test_retrained_click_dssm_embed)

Наблюдение и сбор данных

A/B-тестирование перед запуском

01 Графики сравнения продуктовых метрик (CTR, конверсия)

Контроль

Модель 1

Модель 2

Модель 3

	grocery_sea	grocery_s	Δ	Δ, %	grocery_s	Δ	Δ, %	grocery_s	Δ	Δ, %
GMV и дискавери метрики по поверхности 1ого уровня										
Поиск										
Поиск: GMV				4.99%	▲		5.46%			7.03%
Поиск: GMV Share, %				4.55%	▲		4.62%			5.90%
Поиск: Unique Items				2.74%	▲		2.94%			5.26%
Поиск: Unique Items Share, %				3.35%	▲		3.21%			5.09%
Поиск: Avg Item Price				2.72%	▲		2.90%			1.67%
Поиск: Front Margin				5.11%	▲		5.66%			7.29%
Поиск: Front Margin, %				0.13%	▲		0.20%			0.23%

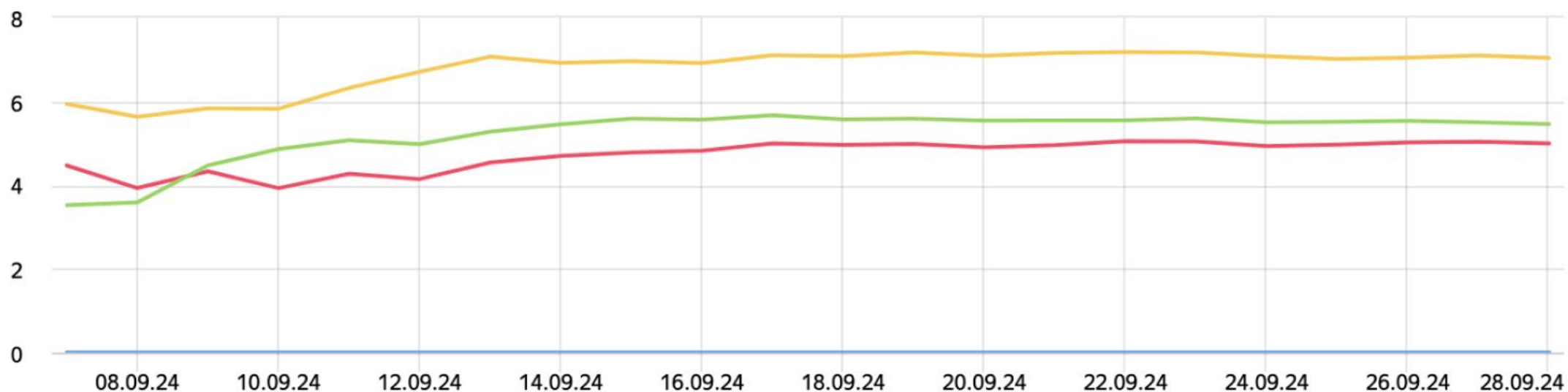
Наблюдение и сбор данных

A/B-тестирование перед запуском

← Поиск: GMV `surfaces.surface_lv1.search.gmv`

Значения Дифф Кумулятивный дифф Кумулятивный дифф в процентах

Кумулятивный дифф в процентах

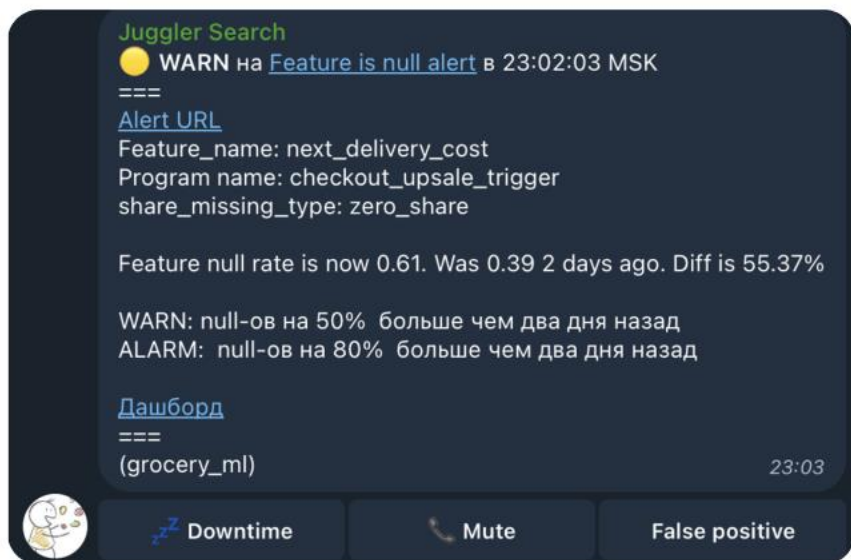


Инциденты и следствия



Инцидент

01 Перестали логировать фичу




Juggler Search
● WARN на [Feature is null alert](#) в 23:02:03 MSK
===
[Alert URL](#)
Feature_name: next_delivery_cost
Program name: checkout_upsale_trigger
share_missing_type: zero_share

Feature null rate is now 0.61. Was 0.39 2 days ago. Diff is 55.37%

WARN: null-ов на 50% больше чем два дня назад
ALARM: null-ов на 80% больше чем два дня назад

[Дашборд](#)
===
(grocery_ml) 23:03

 Downtime Mute False positive

Коля Смирнов

Действие

> Мониторинг доли нулевых/
дефолтных значений

Инцидент

- 01 Перестали логировать фичу
- 02 Фича поменяла значение

Действие

- Мониторинг доли нулевых/дефолтных значений
- Мониторинг распределения фичей

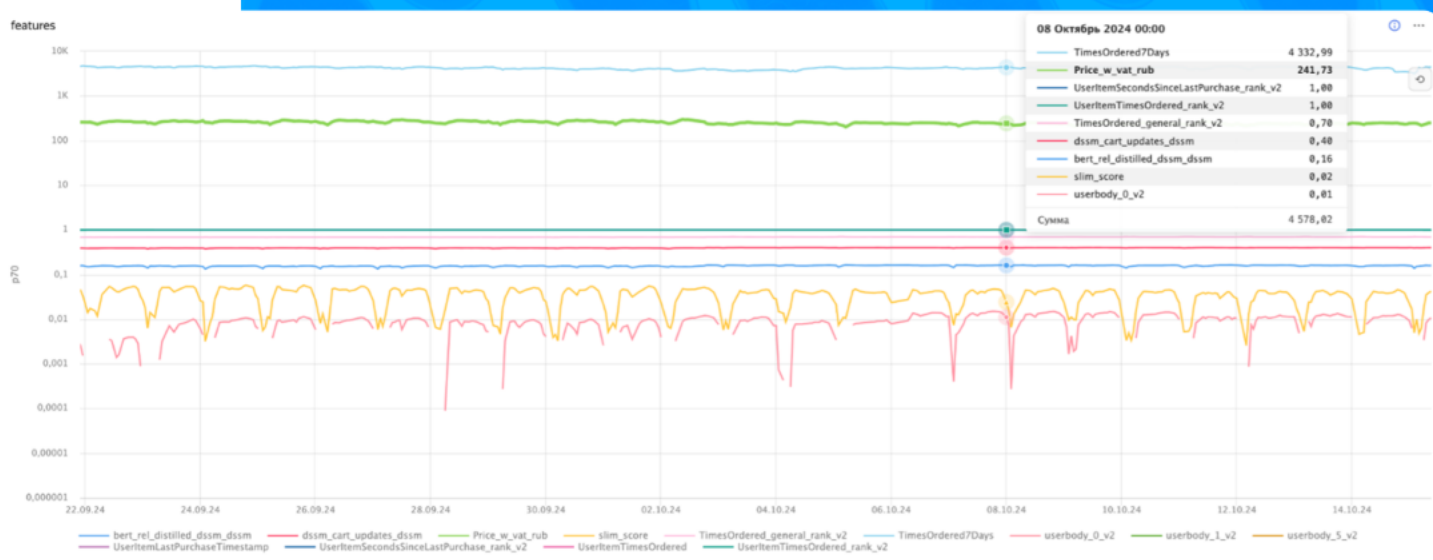
Juggler Search
● **WARN** на [Feature is null alert](#) в 23:02:03 MSK
===
[Alert URL](#)
Feature_name: next_delivery_cost
Program name: checkout_upsale_trigger
share_missing_type: zero_share

Feature null rate is now 0.61. Was 0.39 2 days ago. Diff is 55.37%

WARN: null-ов на 50% больше чем два дня назад
ALARM: null-ов на 80% больше чем два дня назад

[Дашборд](#)
===
(grocery_ml) 23:03

Downtime Mute False positive



Инцидент

- 01 Перестали логировать фичу
- 02 Фича поменяла значение
- 03 Заглохла выкатка ресурса

Действие

- Мониторинг доли нулевых/дефолтных значений
- Мониторинг распределения фичей
- Мониторинг возраста ресурсов

Juggler Search
● WARN на [Feature is](#)
===
[Alert URL](#)
Feature_name: next_del
Program name: checkou
share_missing_type: zer
Feature null rate is now (
WARN: null-ов на 50% больше чем два дня назад
ALARM: null-ов на 80% больше чем два дня назад
[Дашборд](#)
===
(grocery_ml) 23:03

Legend:
- slim_item_profiles_ready
- lavka_stats_profiles_ready
- search_profiles_ready
- final_item_profiles_ready
- noninc_item_profiles_ready
- bandits_profiles_ready

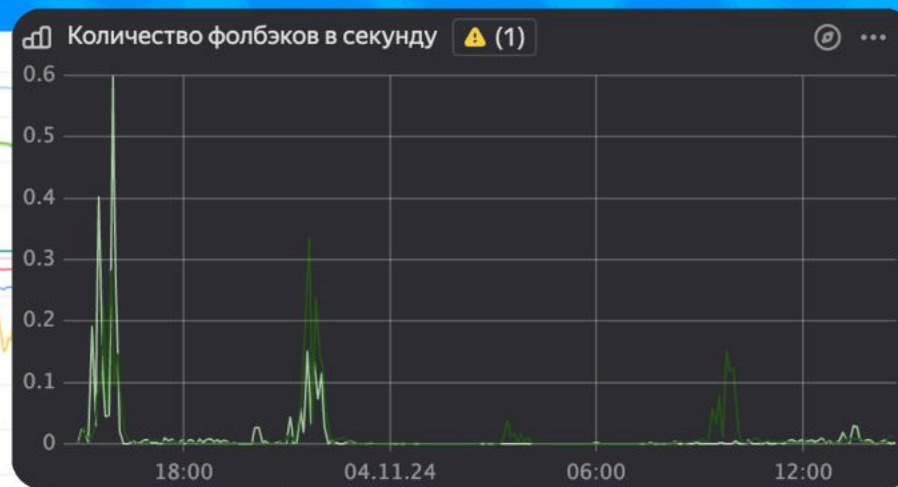
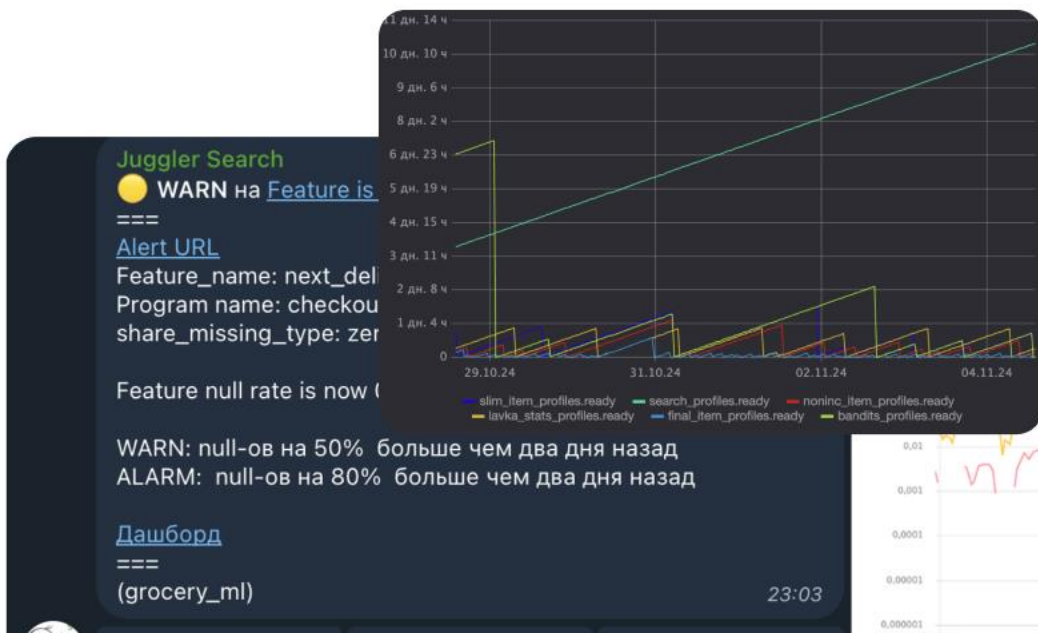


Инцидент

- 01 Перестали логировать фичу
- 02 Фича поменяла значение
- 03 Заглохла выкатка ресурса
- 04 Выкатка с 500-ми

Действие

- Мониторинг доли нулевых/дефолтных значений
- Мониторинг распределения фичей
- Мониторинг возраста ресурсов
- Мониторинг количества фолбеков

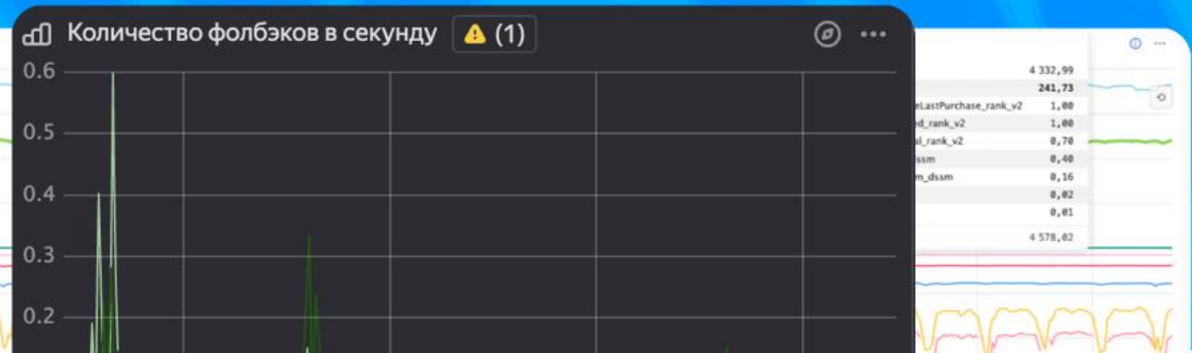
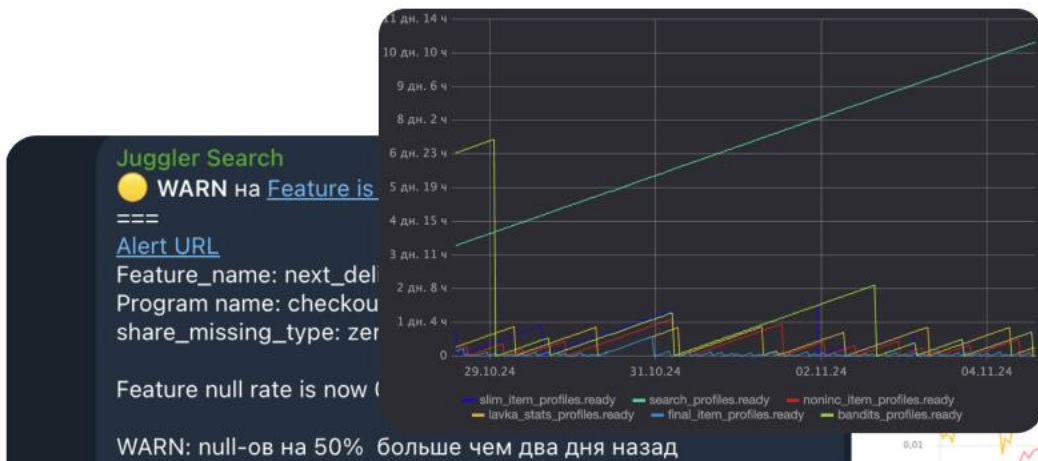


Инцидент

- 01 Перестали логировать фичу
- 02 Фича поменяла значение
- 03 Заглохла выкатка ресурса
- 04 Выкатка с 500-ми
- 05 Растёт доля пустых выдач

Действие

- Мониторинг доли нулевых/дефолтных значений
- Мониторинг распределения фичей
- Мониторинг возраста ресурсов
- Мониторинг количества фолбеков
- Мониторинг доли пустых выдач

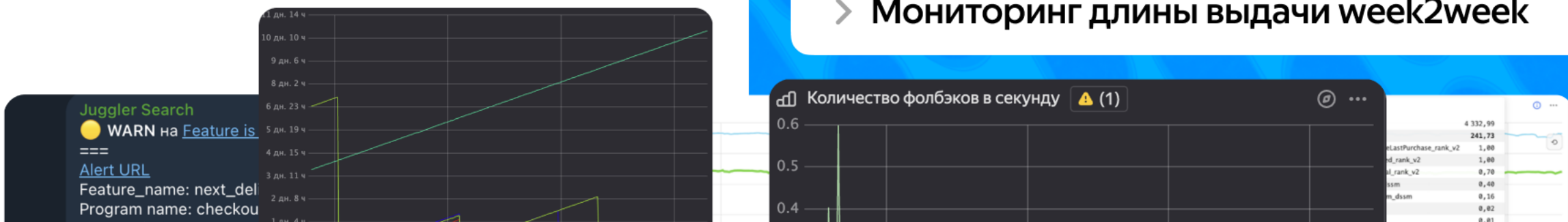


Инцидент

- 01 Перестали логировать фичу
- 02 Фича поменяла значение
- 03 Заглохла выкатка ресурса
- 04 Выкатка с 500-ми
- 05 Растёт доля пустых выдач
- 06 Сократилась длина выдачи

Действие

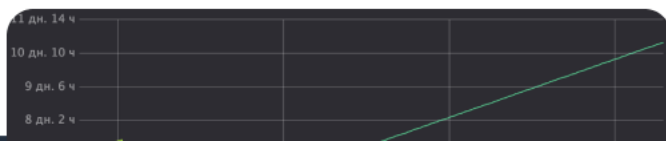
- > Мониторинг доли нулевых/дефолтных значений
- > Мониторинг распределения фичей
- > Мониторинг возраста ресурсов
- > Мониторинг количества фолбеков
- > Мониторинг доли пустых выдач
- > Мониторинг длины выдачи week2week



Инцидент

- 01 Перестали логировать фичу
- 02 Фича поменяла значение
- 03 Заглохла выкатка ресурса
- 04 Выкатка с 500-ми
- 05 Растёт доля пустых выдач
- 06 Сократилась длина выдачи
- 07 ...

Коля Смирнов



Действие

- > Мониторинг доли нулевых/дефолтных значений
- > Мониторинг распределения фичей
- > Мониторинг возраста ресурсов
- > Мониторинг количества фолбеков
- > Мониторинг доли пустых выдач
- > Мониторинг длины выдачи week2week
- > Мониторинг...

Инцидент

- 01 Перестали логировать фичу
- 02 Фича поменяла значение
- 03 Заглохла выкатка ресурса
- 04 Выкатка с 500-ми
- 05 Растёт доля пустых выдач
- 06 Сократилась длина выдачи
- 07 ...

Действие

- > Мониторинг доли нулевых/дефолтных значений
- > Мониторинг распределения фичей
- > Мониторинг возраста ресурсов
- > Мониторинг количества фолбеков
- > Мониторинг доли пустых выдач
- > Мониторинг длины выдачи week2week
- > Мониторинг... Мониторинг...

Инцидент

- 01 Перестали логировать фичу
- 02 Фича поменяла значение
- 03 Заглохла выкатка ресурса
- 04 Выкатка с 500-ми
- 05 Растёт доля пустых выдач
- 06 Сократилась длина выдачи
- 07 ...

Действие

- > Мониторинг доли нулевых/дефолтных значений
- > Мониторинг распределения фичей
- > Мониторинг возраста ресурсов
- > Мониторинг количества фолбеков
- > Мониторинг доли пустых выдач
- > Мониторинг длины выдачи week2week
- > Мониторинг... Мониторинг... Мониторинг...

Инцидент

- 01 Перестали логировать фичу
- 02 Фича поменяла значение
- 03 Заглохла выкатка ресурса
- 04 Выкатка с 500-ми
- 05 Растёт доля пустых выдач
- 06 Сократилась длина выдачи
- 07 ...

Действие

- > Мониторинг доли нулевых/дефолтных значений
- > Мониторинг распределения фичей
- > Мониторинг возраста ресурсов
- > Мониторинг количества фолбеков
- > Мониторинг доли пустых выдач
- > Мониторинг длины выдачи week2week
- > Мониторинг... Мониторинг... Мониторинг...
Мониторинг... Мониторинг... Мониторинг...

Наблюдение и сбор данных

Проблемы мониторинга

- 01 Горшочек, не вари!
- 02 Дежурства: 🔥👉🔥
- 03 False positive срабатывания
- 04 Влияние A/B

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

- Рантайм ошибки сервиса

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

- Рантайм ошибки сервиса
- Изменения поисковой выдачи

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

- Рантайм ошибки сервиса
- Изменения поисковой выдачи
 - % наполнения корзины из поиска

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

- Рантайм ошибки сервиса
- Изменения поисковой выдачи
 - % наполнения корзины из поиска
 - % пустых выдач

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

- Рантайм ошибки сервиса
- Изменения поисковой выдачи
 - % наполнения корзины из поиска
 - % пустых выдач
 - ...

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

- Рантайм ошибки сервиса
- Изменения поисковой выдачи
 - % наполнения корзины из поиска
 - % пустых выдач
 - ...

Офлайн-контур

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

- Рантайм ошибки сервиса
- Изменения поисковой выдачи
 - % наполнения корзины из поиска
 - % пустых выдач
 - ...

Офлайн-контур

- Дрейф офлайн-статистик (профили)

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

- Рантайм ошибки сервиса
- Изменения поисковой выдачи
 - % наполнения корзины из поиска
 - % пустых выдач
 - ...

Офлайн-контур

- Дрейф офлайн-статистик (профили)
- Возрасты ресурсов

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

- Рантайм ошибки сервиса
- Изменения поисковой выдачи
 - % наполнения корзины из поиска
 - % пустых выдач
 - ...

Офлайн-контур

- Дрейф офлайн-статистик (профили)
- Возрасты ресурсов
- Проблемы инфраструктуры (квоты)

Наблюдение и сбор данных

Каналы нотификаций

Онлайн-контур

- Рантайм ошибки сервиса
- Изменения поисковой выдачи
 - % наполнения корзины из поиска
 - % пустых выдач
 - ...

Офлайн-контур

- Дрейф офлайн-статистик (профили)
- Возрасты ресурсов
- Проблемы инфраструктуры (квоты)
- Падение офлайн-процессов

Результаты наблюдения

01 Офлайн-оценка: объективное сравнение

Результаты наблюдения

- 01 Офлайн-оценка: объективное сравнение
- 02 A/B: голосует пользователь

Результаты наблюдения

- 01 Офлайн-оценка: объективное сравнение
- 02 A/B: голосует пользователь
- 03 Мониторинг: всё, до чего дотянулись

Что унести с собой

01

В любом месте цепи Обучение–Внедрение–Наблюдение может быть провал 🧑

Что унести с собой

- 01 В любом месте цепи Обучение–Внедрение–Наблюдение может быть провал 🧑
- 02 Выстраиваем оборону

Что унести с собой

- 01 В любом месте цепи Обучение–Внедрение–Наблюдение может быть провал 🧑
- 02 Выстраиваем оборону
 - Валидацией и метриками в обучении

Что унести с собой

- 01 В любом месте цепи Обучение–Внедрение–Наблюдение может быть провал 🧑
- 02 Выстраиваем оборону
 - Валидацией и метриками в обучении
 - Интеграционными тестами и «посмотреть глазами»

Что унести с собой

- 01 В любом месте цепи Обучение–Внедрение–Наблюдение может быть провал 🧑
- 02 Выстраиваем оборону
 - Валидацией и метриками в обучении
 - Интеграционными тестами и «посмотреть глазами»
 - Постоянным мониторингом

Что унести с собой

- 01 В любом месте цепи Обучение–Внедрение–Наблюдение может быть провал 🧑
- 02 Выстраиваем оборону
 - Валидацией и метриками в обучении
 - Интеграционными тестами и «посмотреть глазами»
 - Постоянным мониторингом
- 03 И... ❤️



Яндекс Лавка

Спасибо!

LinkedIn



Telegram



Коля Смирнов
ML-инженер

