



бак

Мы мисис 177 бачонок!!!

Our team



Даниил
Душенев

ML Engineer

@daniil_d_d



Данил
Кузнецов

ML Enginner



@sapf3ar



Назарий
Карпов

NLP Engineer

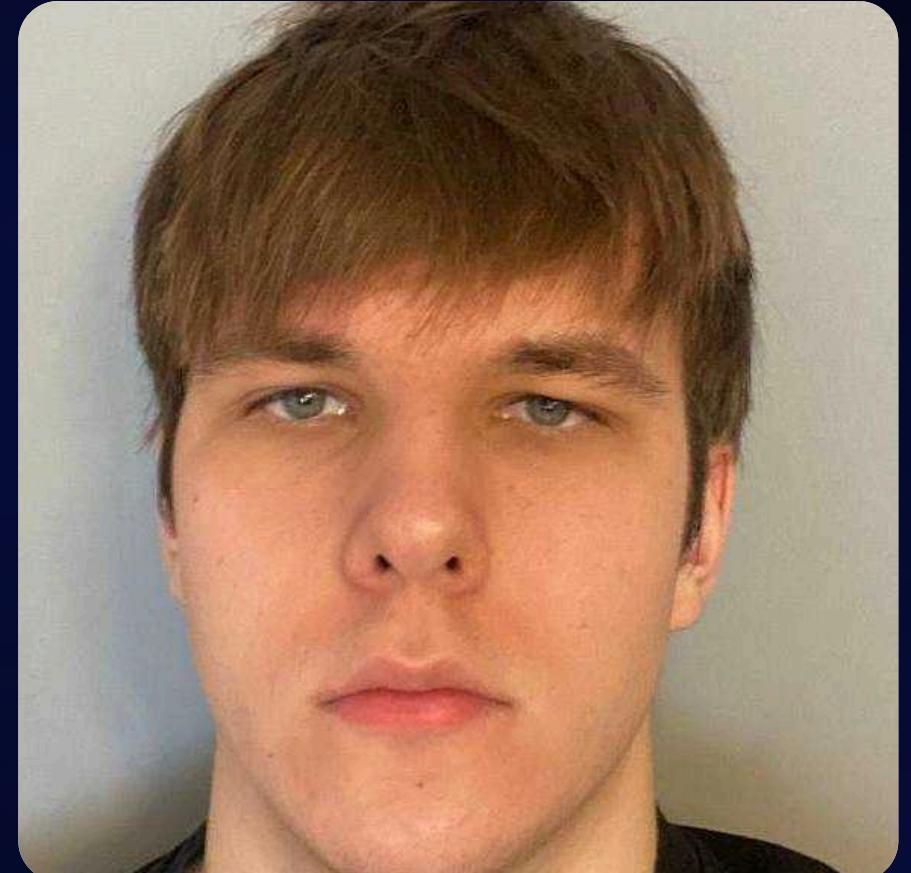
@rezvey



Максим
Хандусь

NLP Engineer

@duck_mood



Павел
Смирнов

CV Engineer

@pa8lusha

Шеф

Умный помощник для заказа еды

Почему еда?

Регулярная потребность

Большинство людей думают о том, что поесть,
как минимум 3 раза в день

Перегруженность

Слишком большой ассортимент,
множество фильтров, тяжело

User:
Настроил фильтры, хочу X

Подборки:
Вот тебе X', Y, A, X

Боль пользователя:
"Я просто хочу сказать,
что мне нужно, и получить это"

А что, если можно лучше?

- А что, если я мог найти дешевле?
- А что, если я мог найти доставку быстрее
- А что, если у меня где то был кешбек?
- А что, если я мог найти лучше по отзывам?



Время

Не хочу тратить десятки
минут на заказ

Эффективность

Один запрос - Один ответ

User:
Я хочу X

Шеф:
Вот тебе X

А куда лучше?

Учитывая историю, запрос,
актуальные предложения:
AI подберет лучший товар



Время

Даем ответ через секунды

01

Поиск товаров

02

Фильтрация

03

Персонализация

04

Интерпретация

Поиск похожих товаров на основе текстовой близости

RAG

rubert-tiny-turbo

Пользователь попросил:

- подешевле
- быстрее доставка

gemma3:4b

Учитываем:

- Историю пользователя
- Персональные кешбеки

Объясняем, почему получились именно эти товары

Закажи недорогую и вкусную пиццу под фильм, не дольше 30 минут

1. Отбор кандидатов

Ищем следующие товары:
Пицца

Продукт	Цена	Рейтинг	Время доставки (мин)	Заказов за месяц	Персональный кешбек
Маргарита	600р	4.96	30	2	10
Маргарита	550р	4.82	40	0	0
Мясная	1200р	4.25	30	2	0
Грибная	400р	4.77	40	0	5
Сицилийская	720р	4.82	50	0	0
Вегетерианска	800р	4.23	40	1	0
Морская	540р	4.52	30	1	0
Мясная	930р	4.75	20	0	0
Пицца 4 сыра	600р	4.87	40	0	0
Пицца 4 сыра	1300р	4.68	50	0	0

Закажи недорогую и вкусную пиццу под фильм, не дольше 30 минут

2. Фильтрация

Из запроса:

- Недорого
- <= 40 минут

Продукт	Цена	Рейтинг	Время доставки (мин)	Заказов за месяц	Персональный кешбек
Маргарита	600р	4.96		30	2 10
Маргарита	550р	4.82		40	0 0
Мясная	1200р	4.25		30	2 0
Грибная	400р	4.77		40	0 5
Сицилийская	720р	4.82		50	0 0
Вегетерианска	800р	4.23		40	1 0
Морская	540р	4.52		30	1 0
Мясная	930р	4.75		20	0 0
Пицца 4 сыра	600р	4.87		40	0 0
Пицца 4 сыра	1300р	4.68		50	0 0

Закажи недорогую и вкусную пиццу под фильм, не дольше 30 минут

3. Персонализация

Из суммаризации истории
диалога:

- не любит грибы
- не любит морепродукты

Продукт	Цена	Рейтинг	Время доставки (мин)	Заказов за месяц	Персональный кешбек
Маргарита	600р	4.96		30	2 10
Маргарита	550р	4.82		40	0 0
Мясная	1200р	4.25		30	2 0
Грибная	400р	4.77		40	0 5
Сицилийская	720р	4.82		50	0 0
Вегетерианска	800р	4.23		60	1 0
Морская	540р	4.52		30	4 0
Мясная	930р	4.75		20	0 0
Пицца 4 сыра	600р	4.87		40	0 0
Пицца 4 сыра	1300р	4.68		50	0 0

Закажи недорогую и вкусную пиццу под фильм, не дольше 30 минут

3. Персонализация

Из суммаризации истории диалога:

- заказывал из этого ресторана
- часто заказывает с кэшбэком

Продукт	Цена	Рейтинг	Время доставки (мин)	Заказов за месяц	Персональный кешбек
Маргарита	600р	4.96		40	2 10
Маргарита	550р	4.82		40	0 0
Мясная	1200р	4.25		30	2 0
Грибная	400р	4.77		40	0 5
Сицилийская	720р	4.82		50	0 0
Вегетерианска	800р	4.23		60	1 0
Морская	540р	4.52		30	1 0
Мясная	930р	4.75		20	0 0
Пицца 4 сыра	600р	4.87		40	0 0
Пицца 4 сыра	1300р	4.68		50	0 0

На основе чего интерпретируем решение пользователю:

- Количество заказов из этого ресторана
- Рейтинг блюда
- Скорость доставки
- Персональный Кешбек
- Стоимость

Например:

- Оценка этого блюда выше остальных - 4.95!
- Это самая быстрая доставка из доступных - всего 20 минут!
- Вы заказывали из этого ресторана уже 7 раз!

Демонстрация

Пользователь:

1

Снижение
когнитивной нагрузки

2

Снижение времени
совершения заказа

Банк:

1

Увеличение количества заказов
через приложение на X%

2

Повышение NPS на X%

Arenadata

Decision pen' is all you need



В области связи и интернет-трафика значительную проблему представляет пользователи, создающие **вредоносный трафик**. Такой трафик инициирует спам-атаки, фишинговые рассылки и прочие виды кибератак, что **негативно сказывается** на стабильности и безопасности систем связи.

Проблема становится еще серьезнее, когда оказывается, что взломанных юзеров чрезвычайно мало, а находить их **необходимо**

Нас интересует **скорость и точность**.
Ложных срабатываний должно быть **по минимуму**



Random Forest, LogReg

Высокий скор на валидации,
низкий на боевом сете

Детекция аномалий (autoencoder, z-score)

Хороший перфоманс был у z-score. В остальном модели переобучались или просто выдавали ужасные результаты

01

UpTx > DownTx

02

N > 7

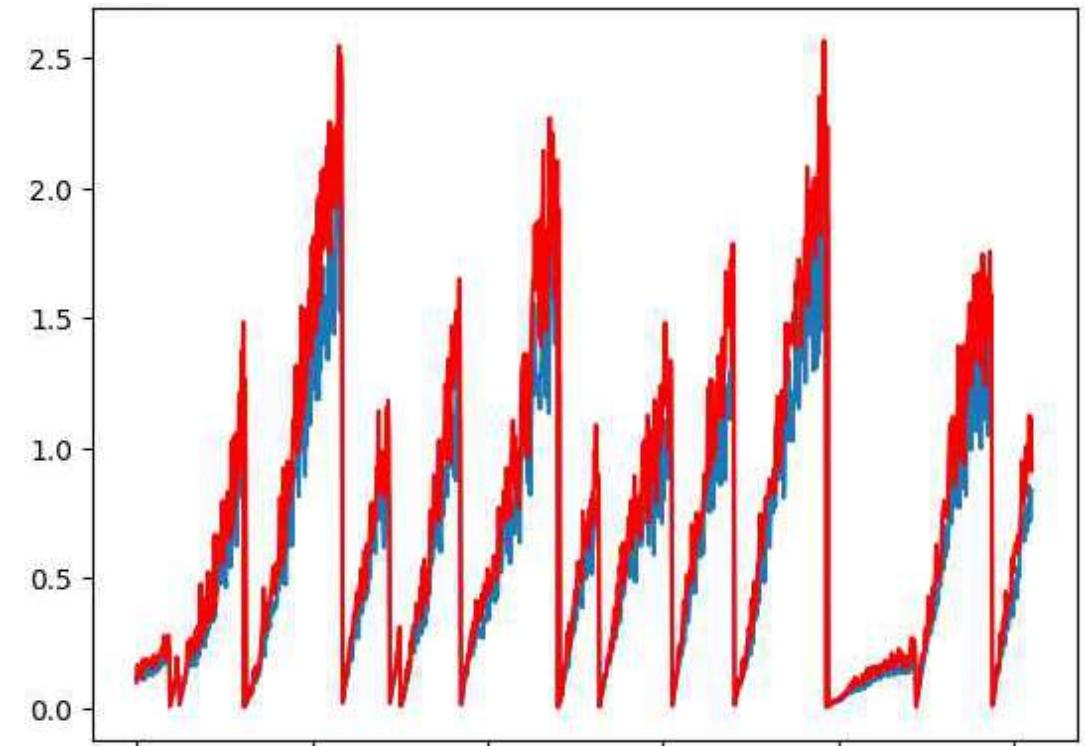
Для каждого юзера считаем
кол-во временных точек
где Uptx > Downtx

03



04

Profit!

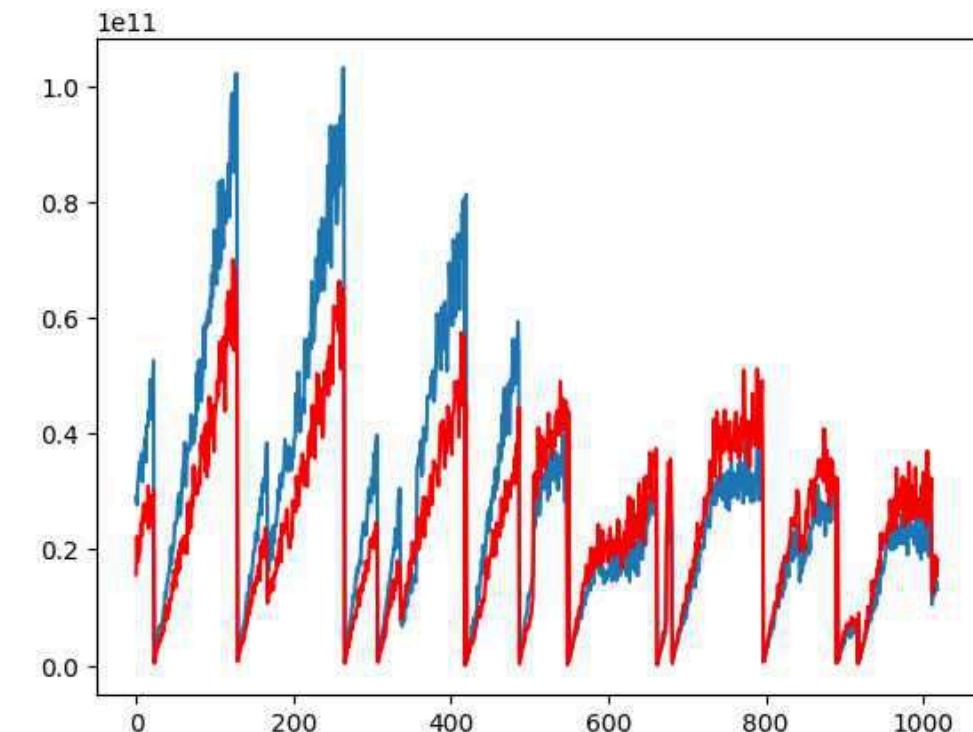


**Вот так выглядит здоровый
абонент**



Инсайты

Методом пристального взгляда была выдвинута гипотеза что такая ситуация с аномальным трафиком у всех взломанных юзеров, тогда было решено реализовать модель решающего ПНЯ



А вот так Взломанный

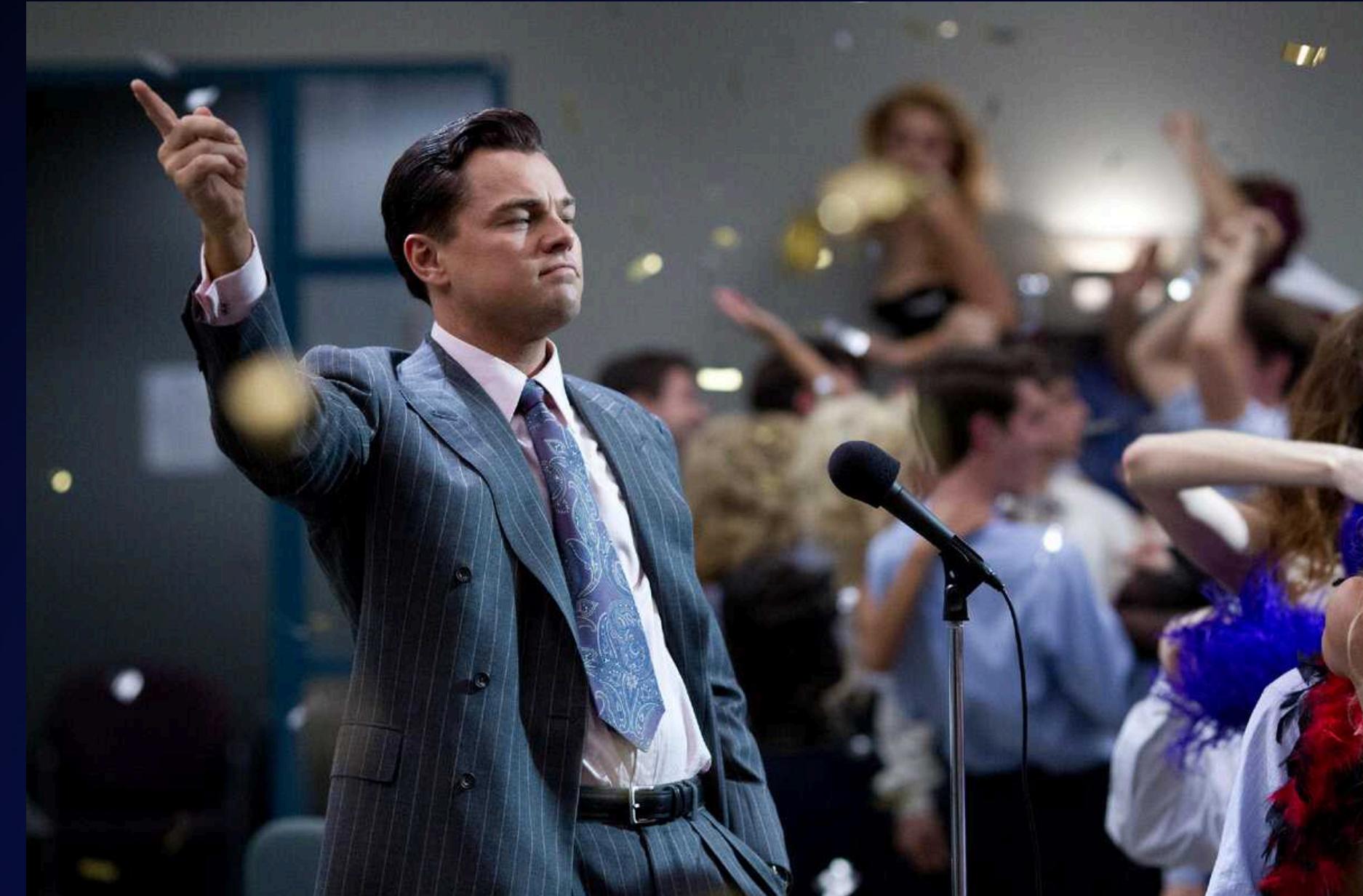
И в итоге...

99%

Вот настолько совпадают наши
предсказания с файлом RESULT!

Алгоритм в сумме отработал минут за 5(потребовался компьютер с 30гб оперативки)

Если соединить все витрины - получится тот же RESULT, что получается при прогоне полнового боевого датасета



Вывод



Быстрое решение

Качественное решение

Понятное решение



АВИЛОН

Мы мисис 177 бачонок!!!

Проблематика

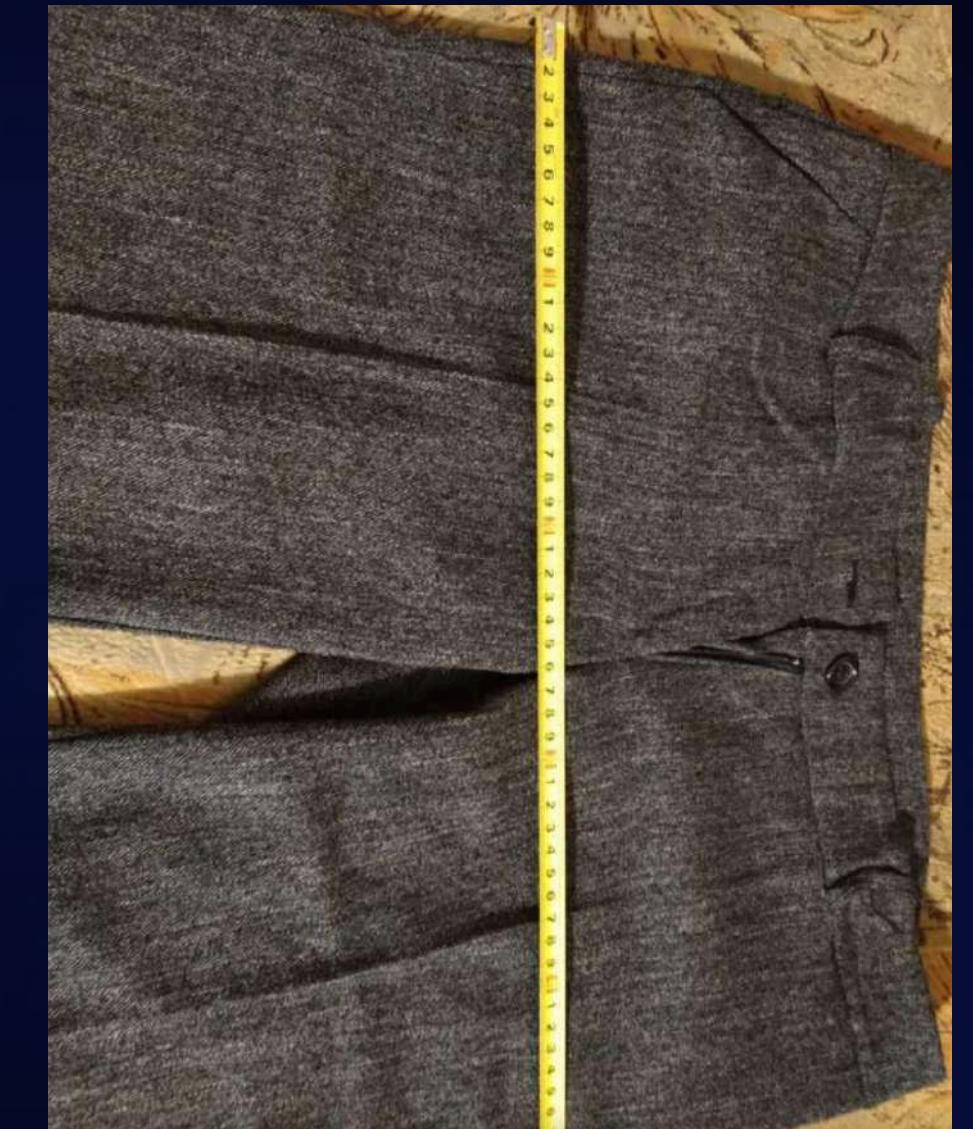
Хотим репрезентативные дескрипторы для пользовательских данных

Инференс должен быть быстрым

Данные

Специфика данных

1. Пользовательская разметка
2. Дисбаланс классов
3. Дубликаты
4. Мусор в данных



Фильтрация данных

- | | |
|------------------------------|---|
| 1. Пользовательская разметка | VLM разметка |
| 2. Дисбаланс классов | Кастомный лосс, учитывающий палитру |
| 3. Дубликаты | Кластеризация эмбеддингов |
| 4. Мусор в данных | Open-domain детекция, для подтверждения
нахождения объекта на фото |

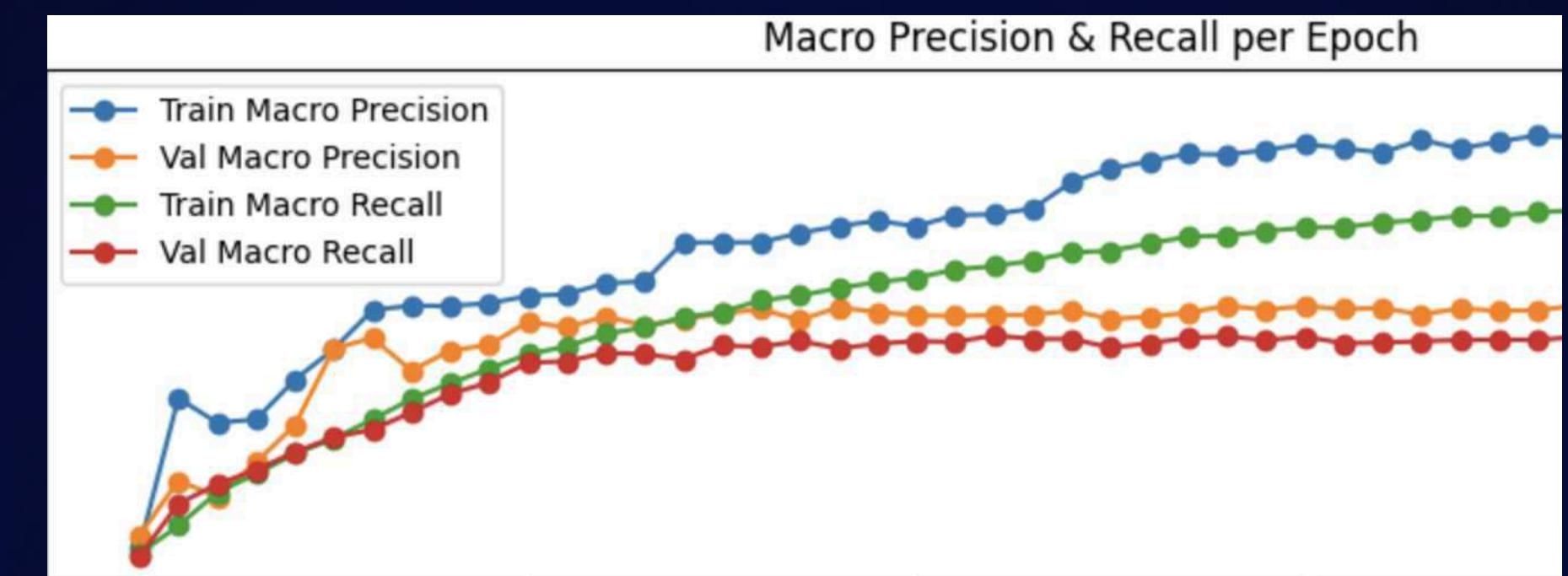
Обучили робастный классификатор
На основе zero-shot пайплайна
Использовали Yolo-World для Open-domain детекции
DINOv2 для создания репрезентативных эмбеддингов
И опах-гри для симуляция прод среды

Результат

Macro F1: 0.76

Инференс < 200ms

Оптимизация классификации
смежных цветов





МТС

Мы мисис 177 бачонок!!!

Нужно постоянно поддерживать
оптимальную утилизацию

Важна **доступность** сервиса к
новым аллокациям

Тяжело делать эффективно **миграции**

VM

+100

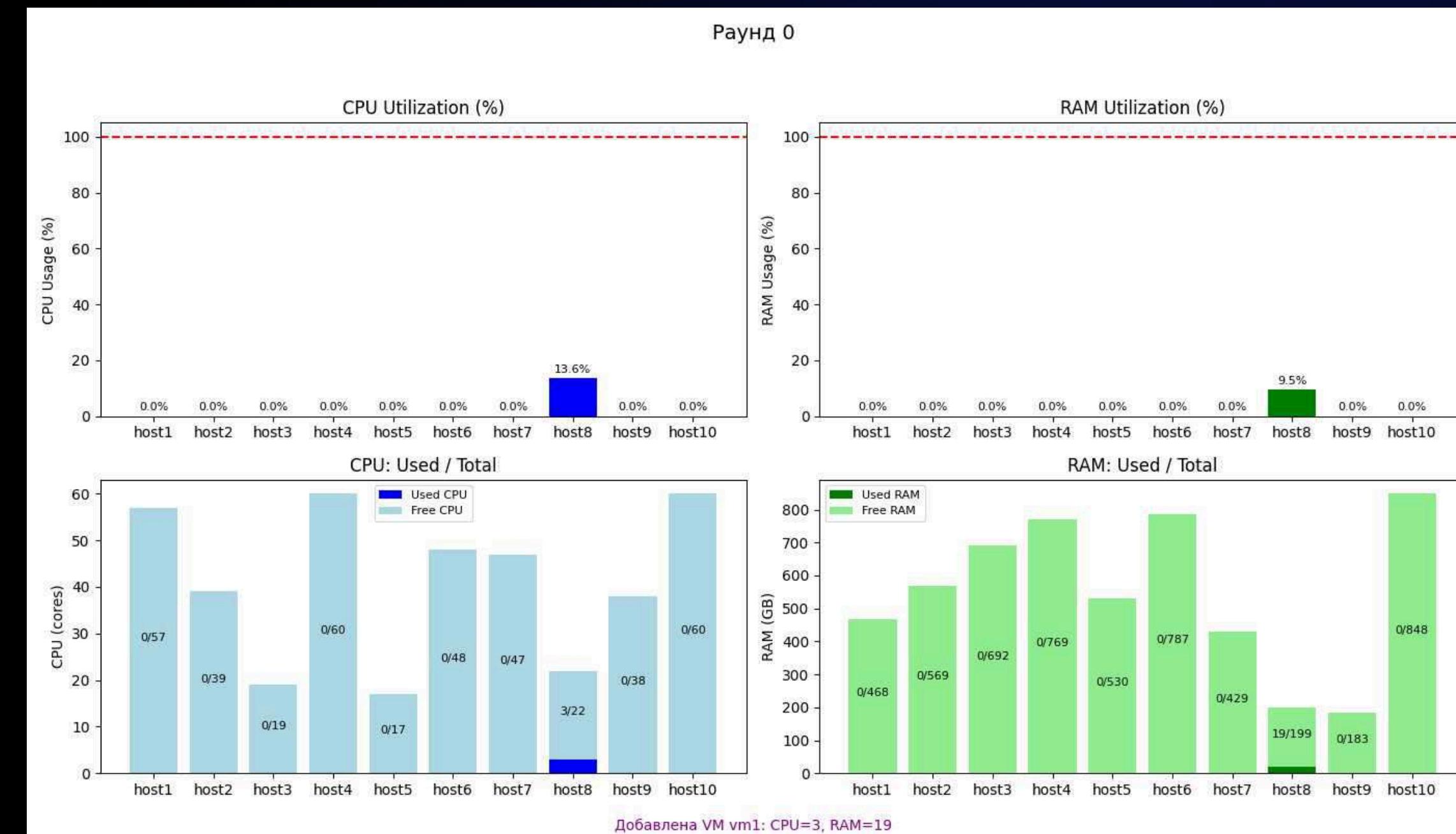
Жадный алгоритм основанный на
эвристиках

Плюсы:

- Легко внедрить
- Эффективен на небольших данных
- Уменьшает фрагментацию
- Работает в онлайне

Минусы:

- Субоптимален
- Не гарантирует соблюдение SLA в пиковых нагрузках



Genetic Algo

Генетический алгоритм

Популяция: 80 индивидов
(распределений ВМ).

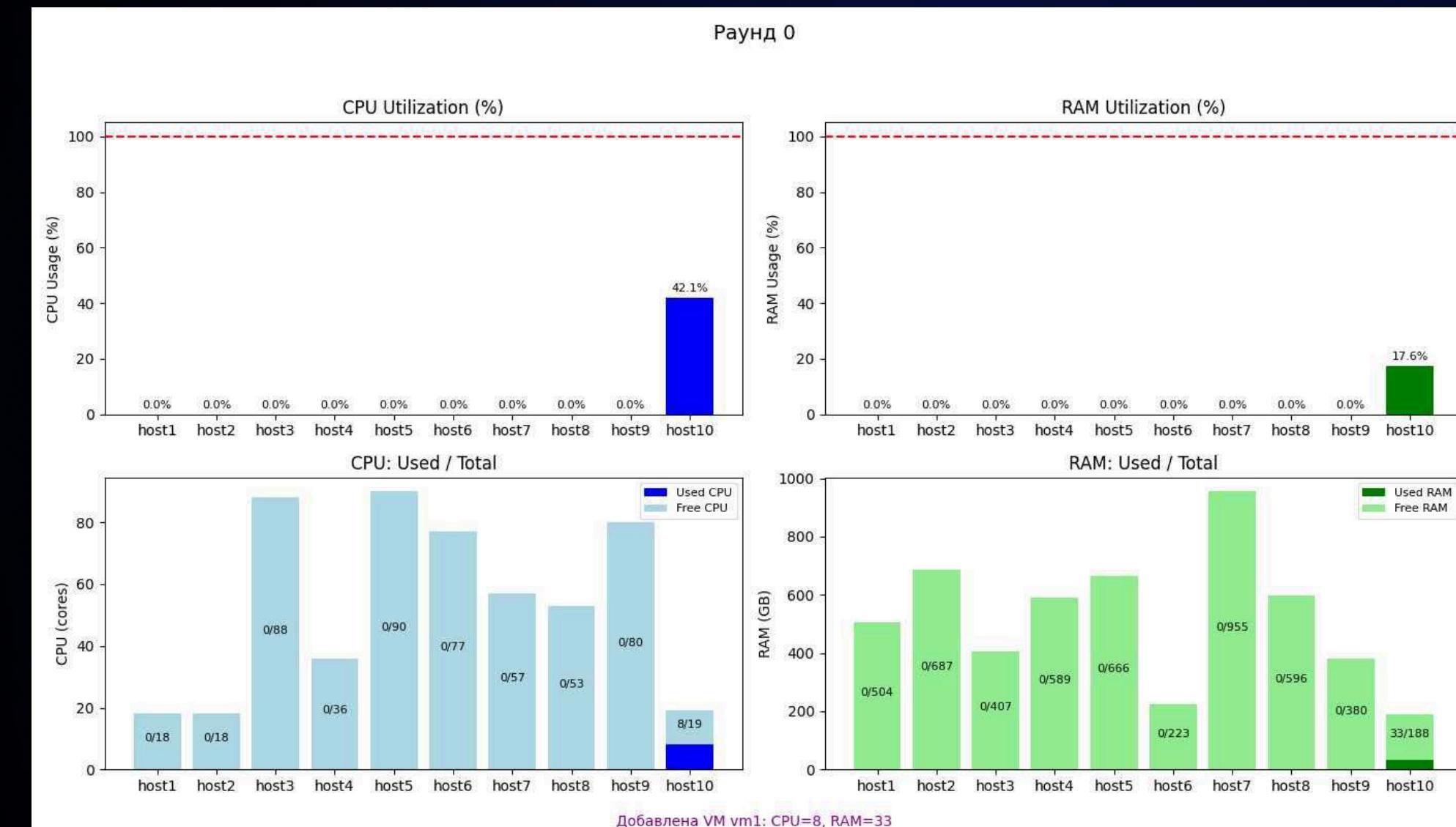
Поколения: 50 итераций
эволюции.

Плюсы:

- Более оптимальная оптимизация
- Балансировка нагрузки
- Энергоэффективность, поощряет выключение пустых хостов
- Можно комбинировать с BestFit

Минусы:

- Большая выч. сложность
- Субоптимальность
- Не всегда эффективно делает миграции



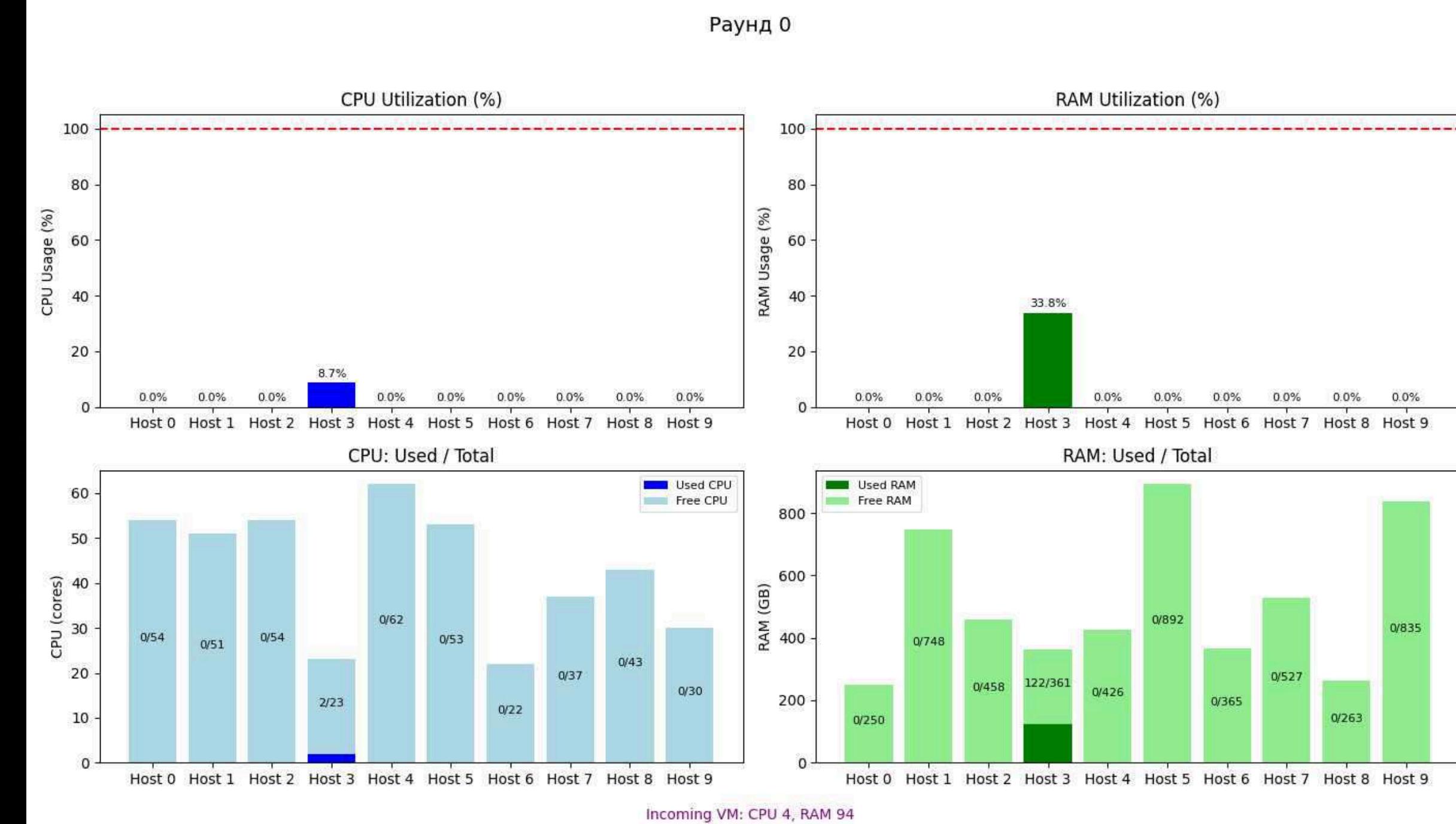
Алгоритм машинного обучения. На основе обучения с подкреплением

Плюсы:

- Адаптивность к динамическим условиям
- Оптимизация долгосрочных целей
- Работа со сложными состояниями
- Автоматизация без ручных правил

Минусы:

- Большая выч. сложность
- Сложность проектирования функции награды
- Нестабильность обучения



Что выбрать?

- Для крупных облаков с долгосрочными целями (например, энергоэффективность).
- Если нагрузка нестабильна и требует адаптивных решений.

RL

- В small-scale системах с простыми требованиями.
- Если нужны гарантированно быстрые решения (например, real-time SaaS).

**Best Fit, GA,
алгоритмы на эвристиках**



SBER

Сложность ручного
планирования при множестве
зависимостей

Конфликт целей: время vs ресурсы

Учет динамических ограничений

Необходимость баланса между
автоматизацией и гибкостью

Состояния (States)

- Нормализованные даты начала/окончания задач.
- Общая длительность проекта (в % от года).

Действия (Actions)

- Сдвиг задачи вперед/назад (Discrete Space: len(tasks)^2).

Награда (Reward)

- +N баллов за каждый сокращенный день проекта.
- -M баллов за нарушение зависимостей.

Модель

- PPO (Proximal Policy Optimization).

Обработка зависимостей:

- Граф задач строится через NetworkX.
- Рекурсивный сдвиг дочерних задач при изменении родительских.

Иерархия задач:

- Рекурсивный парсинг вложенных задач и подпроектов.

Учет календаря:

- Пропуск нерабочих дней при расчете сроков.

Использование обучения с подкреплением (RL):

- Адаптивность
- Оптимизация по нескольким параметрам

Обработка сложных зависимостей

Учет календаря

Масштабируемость:

- Поддержка вложенных задач и подпроектов.



спасибо

Мы мисис 177 бачонок!!!