

MediScan

Сервис для ранней диагностики
злокачественных новообразований кожи,
определяющий тип новообразования по
фотографии

Выполнено:

Алексей Калягин, Мария Герасимова, Дмитрий Саперов,
Студенты 1-го курса магистратуры МОВС, ФКН, НИУ ВШЭ



Команда



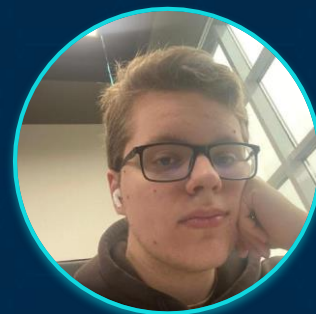
**Алексей
Калягин**



**Мария
Герасимова**



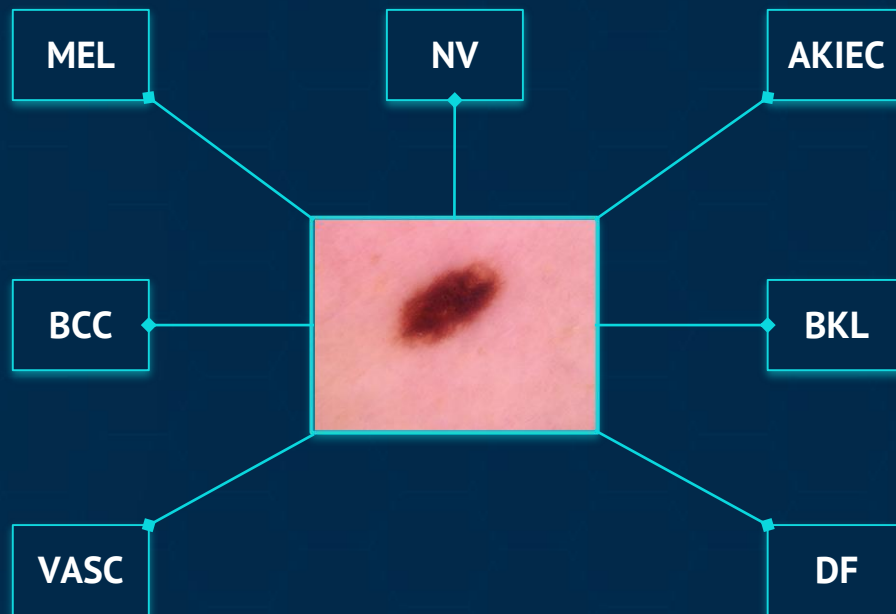
**Дмитрий
Саперов**



**Михаил
Никифоров**
Куратор

Описание задачи

Создание веб-сервиса с
возможностью определения
типа новообразования кожи
по фотографии.



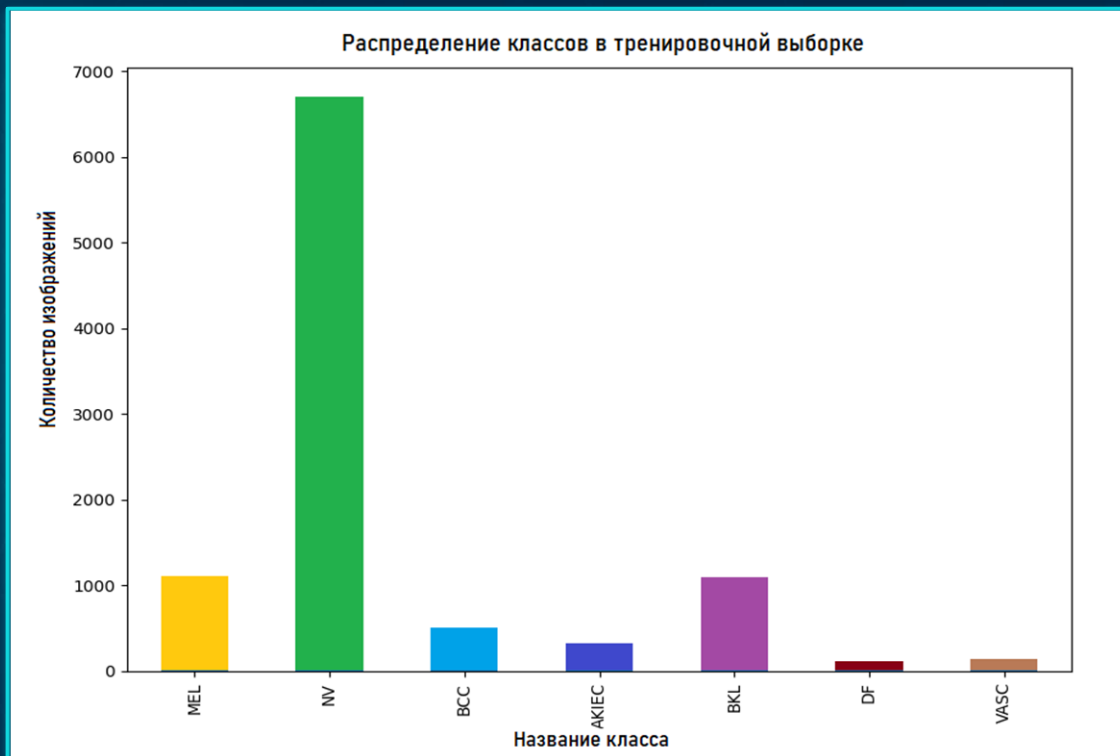
Данные

Датасет - ISIC2018 Task 3

- Тренировочная выборка – 10 015 изображений
- Тестовая выборка – 1 512 изображений
- Валидационная выборка – 193 изображения

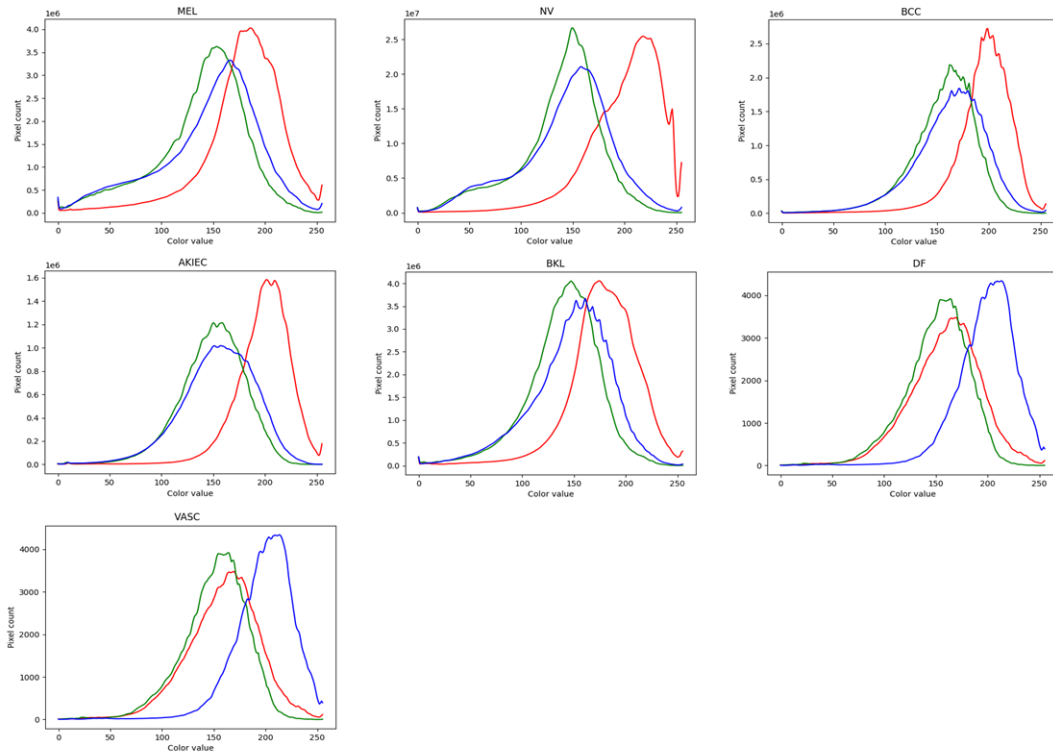
	MEL	NV	BCC	AKIEC	BKL	DF	VASC
Train	1113	<u>6705</u>	514	327	1099	115	142
Test	171	<u>909</u>	93	43	217	44	35
Val	21	<u>123</u>	15	8	22	1	3

Данные

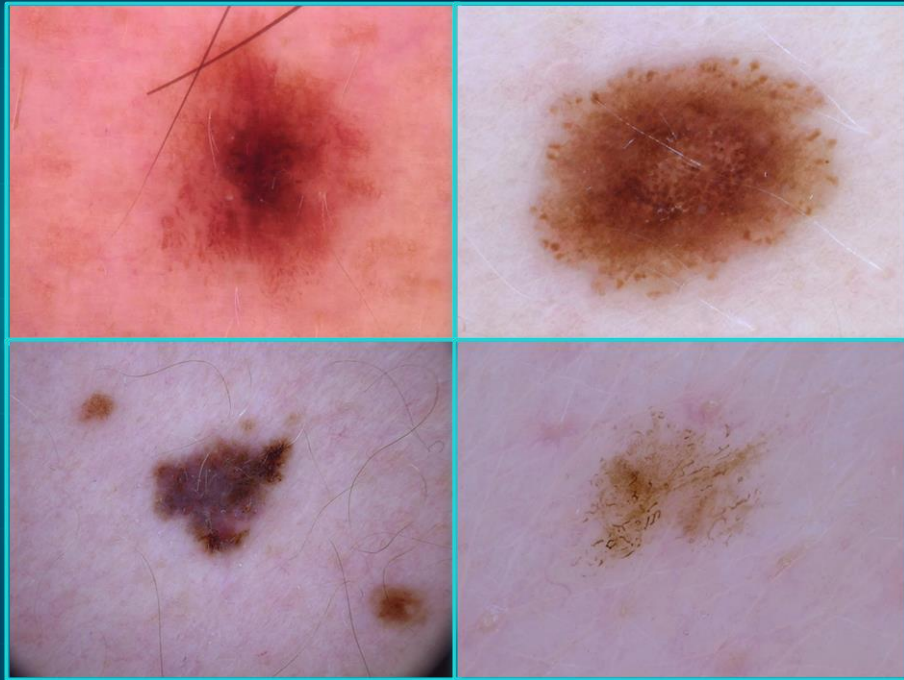


Данные

Распределение значений пикселей в изображениях



Данные



Изображения
центрированы, но на
многих присутствуют
артефакты в виде волос.

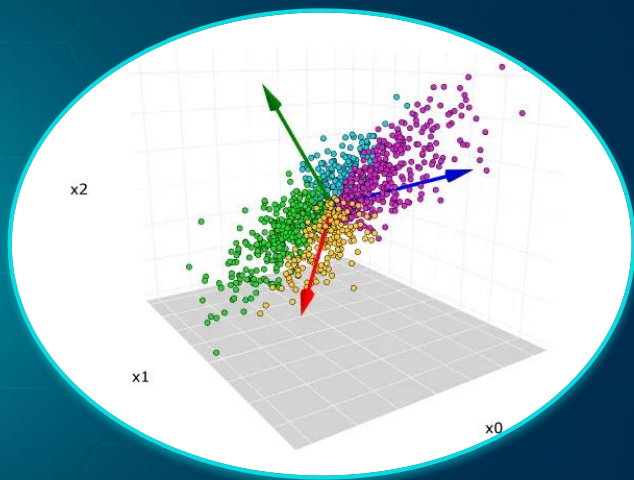
Обработка изображений

450x600

Размер стандартного изображения в датасете

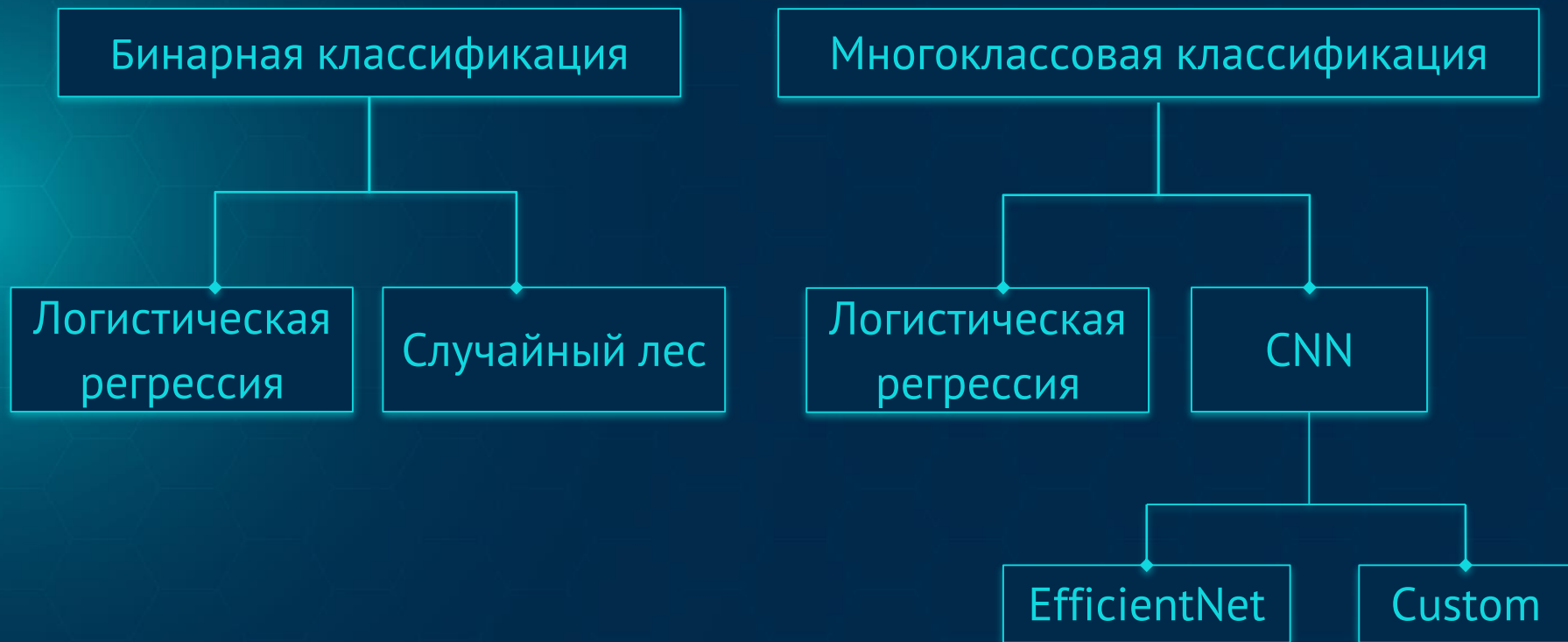
От 270k до 810k

Исходное количество признаков

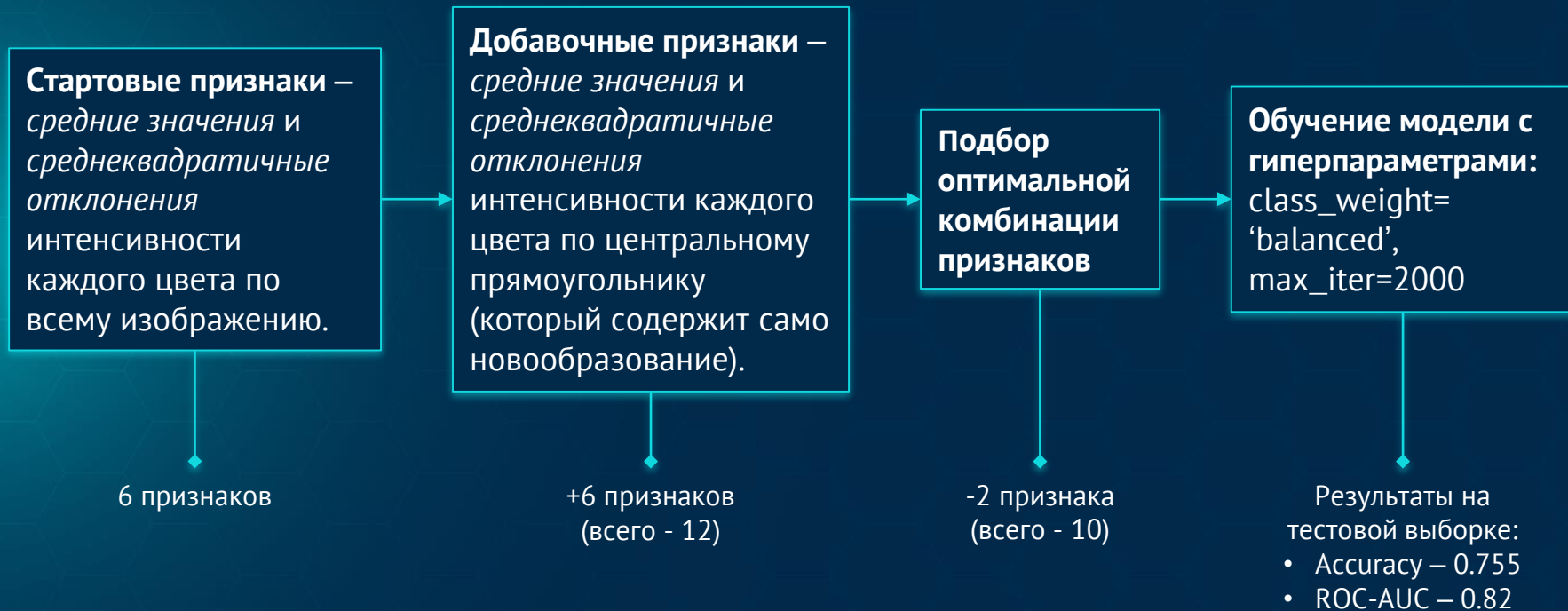


В классических моделях машинного обучения для уменьшения размерности был применен **PCA**

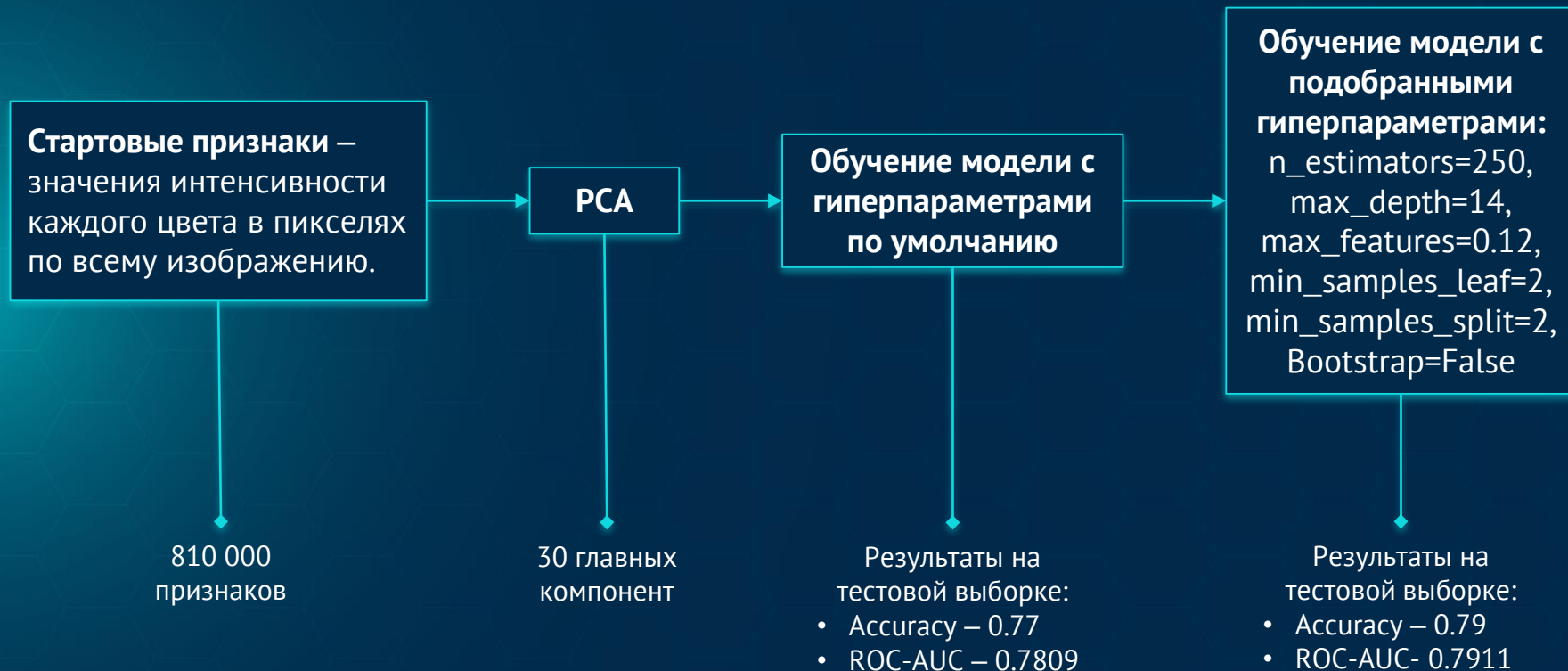
Рассмотренные подходы



Бинарная классификация - логистическая регрессия

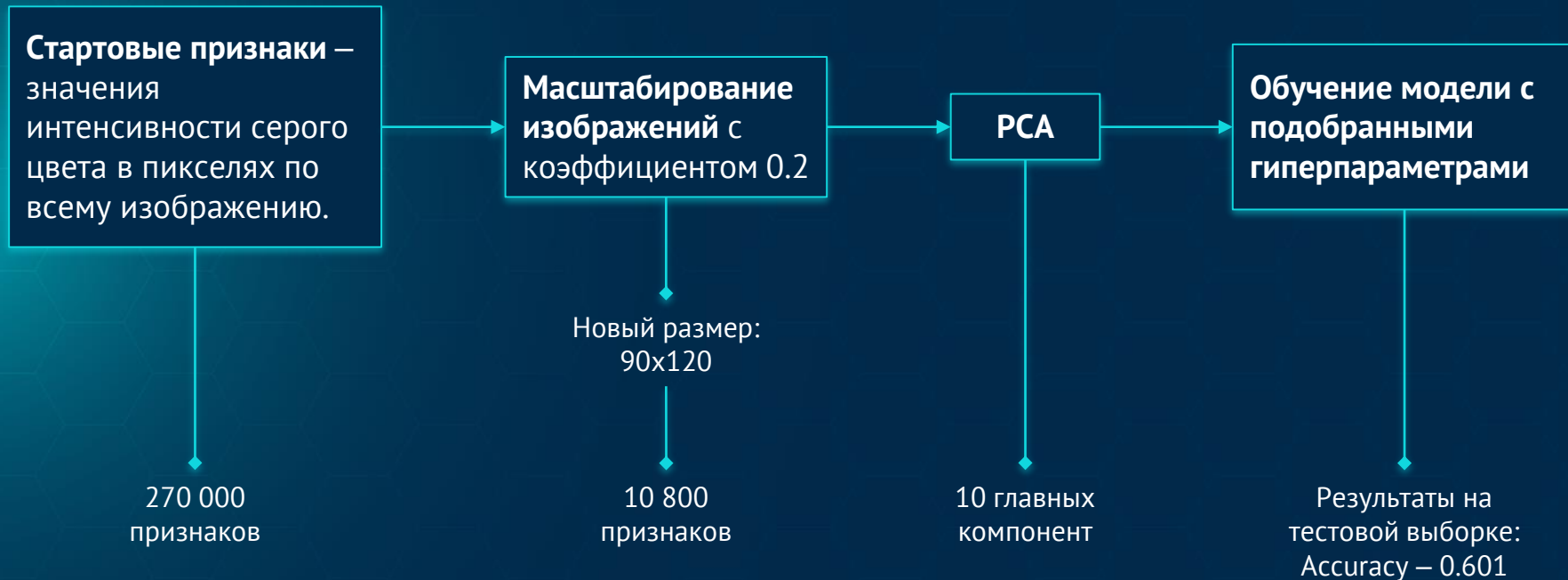


Бинарная классификация - случайный лес



Многоклассовая классификация — логистическая регрессия

В качестве алгоритма оптимизации использовался **SGD**



Custom convolutional neural network

3

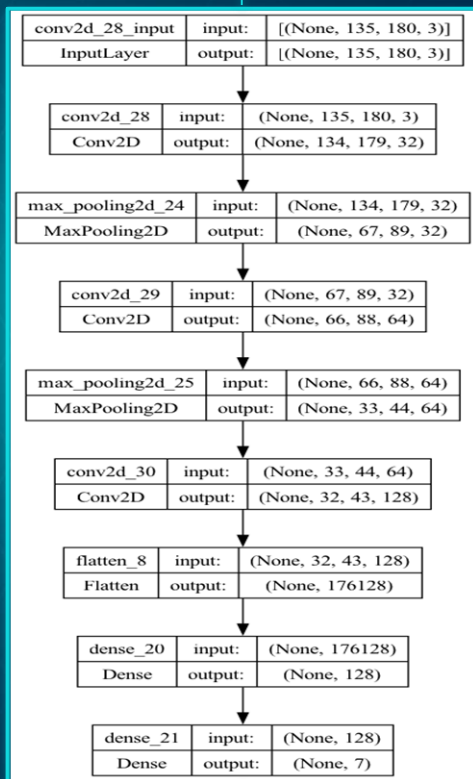
Количество сверток

2

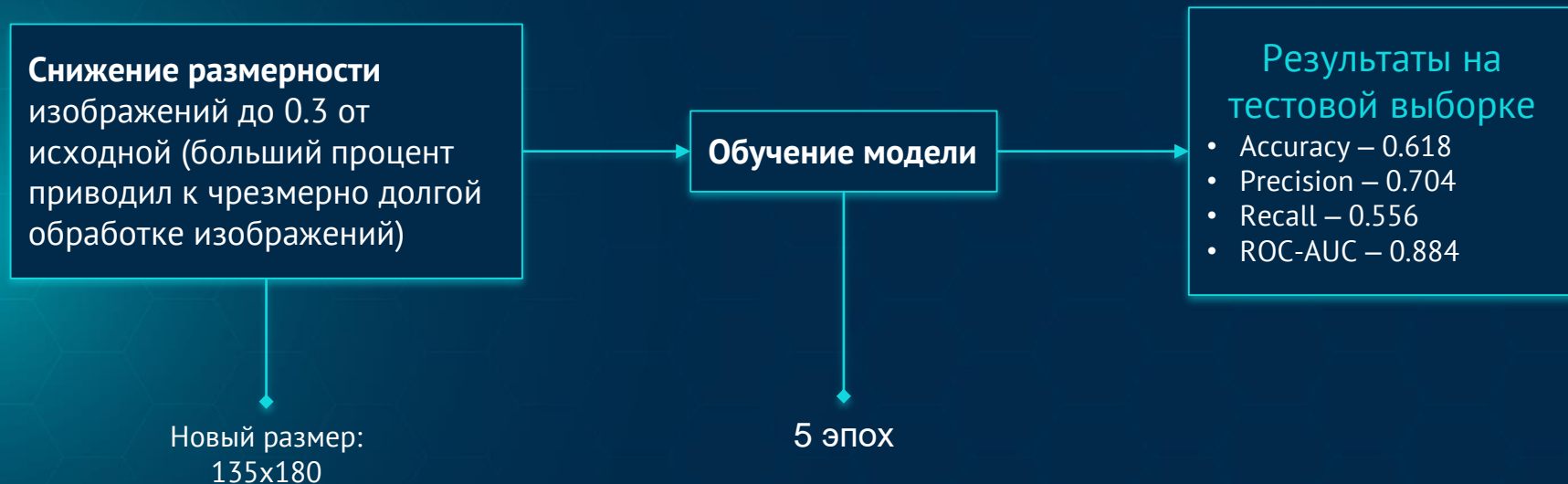
Количество полносвязных слоев

22,600,000

Общее количество параметров

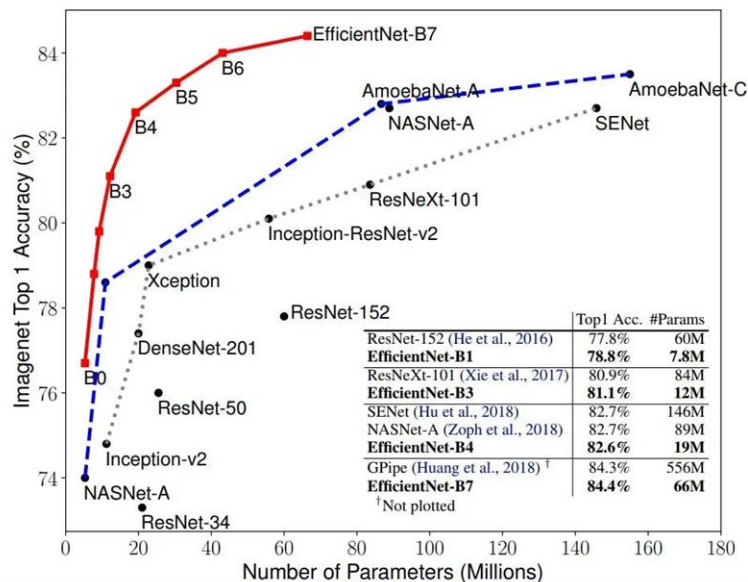


Custom convolutional neural network



EfficientNet

EfficientNet – семейство CNN, в котором используется метод **составного масштабирования**, когда модели равномерно масштабируются по глубине/ширине/разрешению на заданный коэффициент. Такой подход обеспечивает наилучший баланс между точностью модели и вычислительными затратами.



EfficientNet B0

237

Общее количество слоев

18

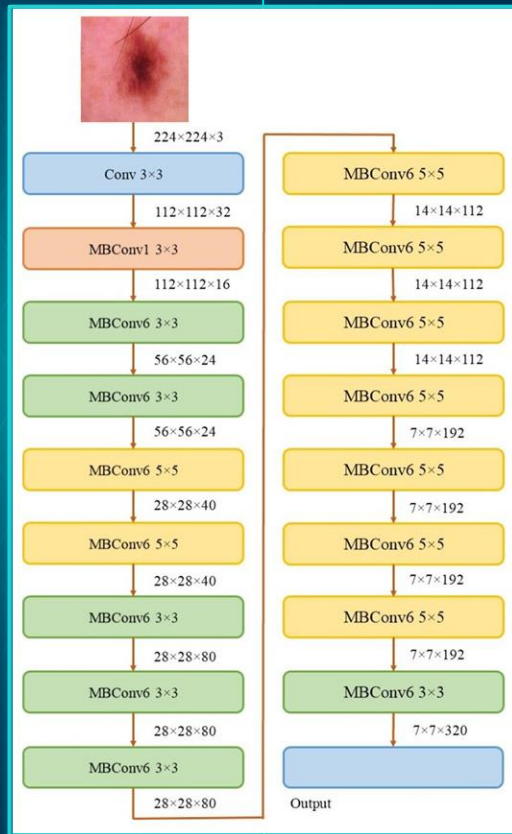
Количество сверточных слоев

224x224

Разрешение входного изображения

5,300,000

Общее количество параметров

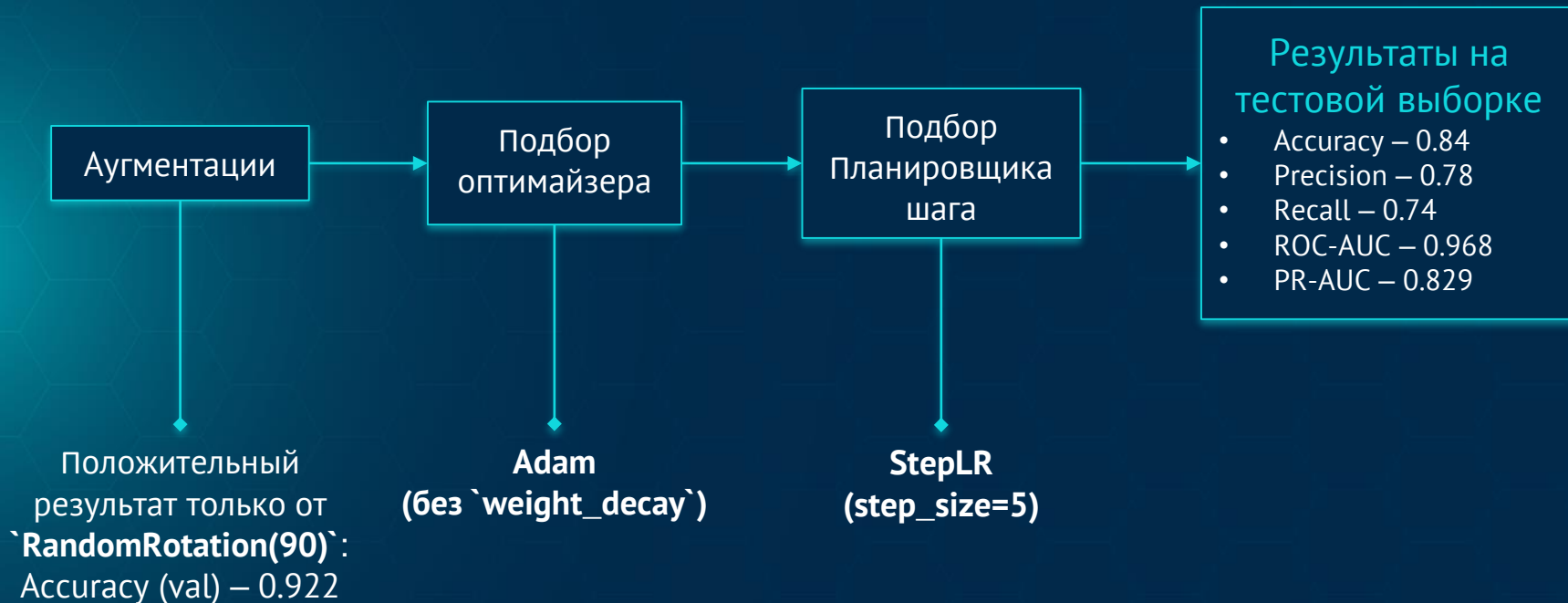


EfficientNet B0. Эксперименты

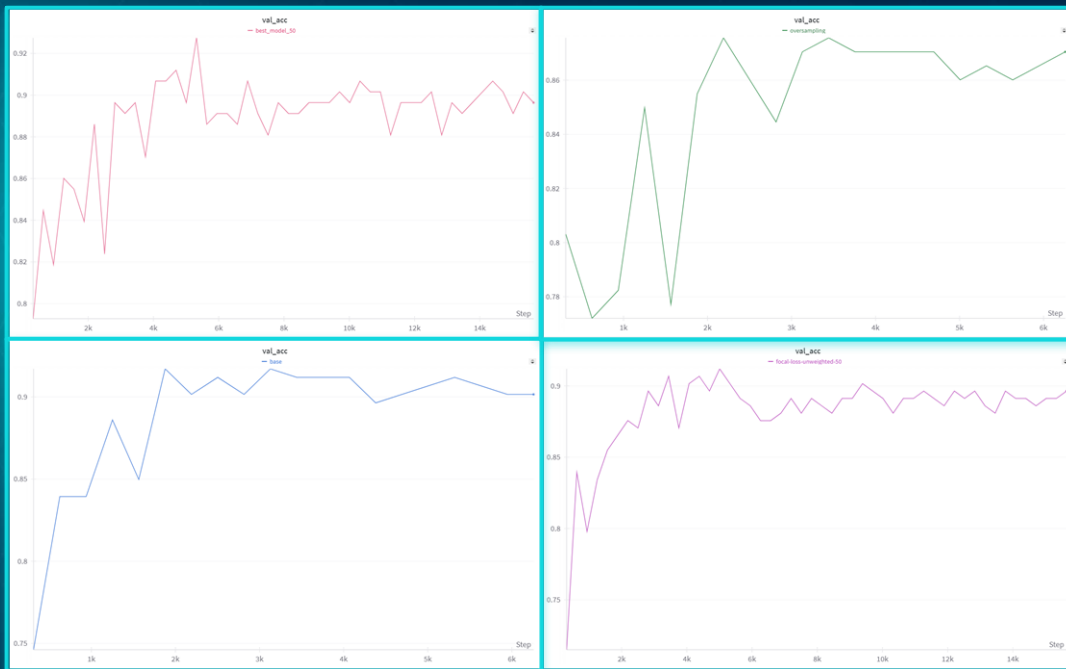
Работа с дисбалансом классов



EfficientNet B0. Эксперименты



EfficientNet B0. Эксперименты



Для визуальной оценки результатов проводимых экспериментов использовалось логирование значений метрики качества и функции потерь в сервис **wandb.ai**

EfficientNet B4

474

Общее количество слоев

380x380

Разрешение входного изображения

19,400,000

Общее количество параметров

Результаты на тестовой выборке

- Accuracy – 0.864
- Precision – 0.81
- Recall – 0.75
- ROC-AUC – 0.973
- PR-AUC – 0.854

Данная модель продемонстрировала наилучшие показатели качества среди рассмотренных и была выбрана в качестве **лучшей модели**.

Telegram бот

Основа — фреймворк **aiogram**.

/help

Выводит список доступных команд

/predict

Выдает наиболее вероятный тип новообразования на изображении, используя модель логистической регрессии

/predictmel

Выдает вероятность того, что на загруженном изображении меланомы

/predictcnn

Выдает наиболее вероятный тип новообразования на изображении, используя CNN

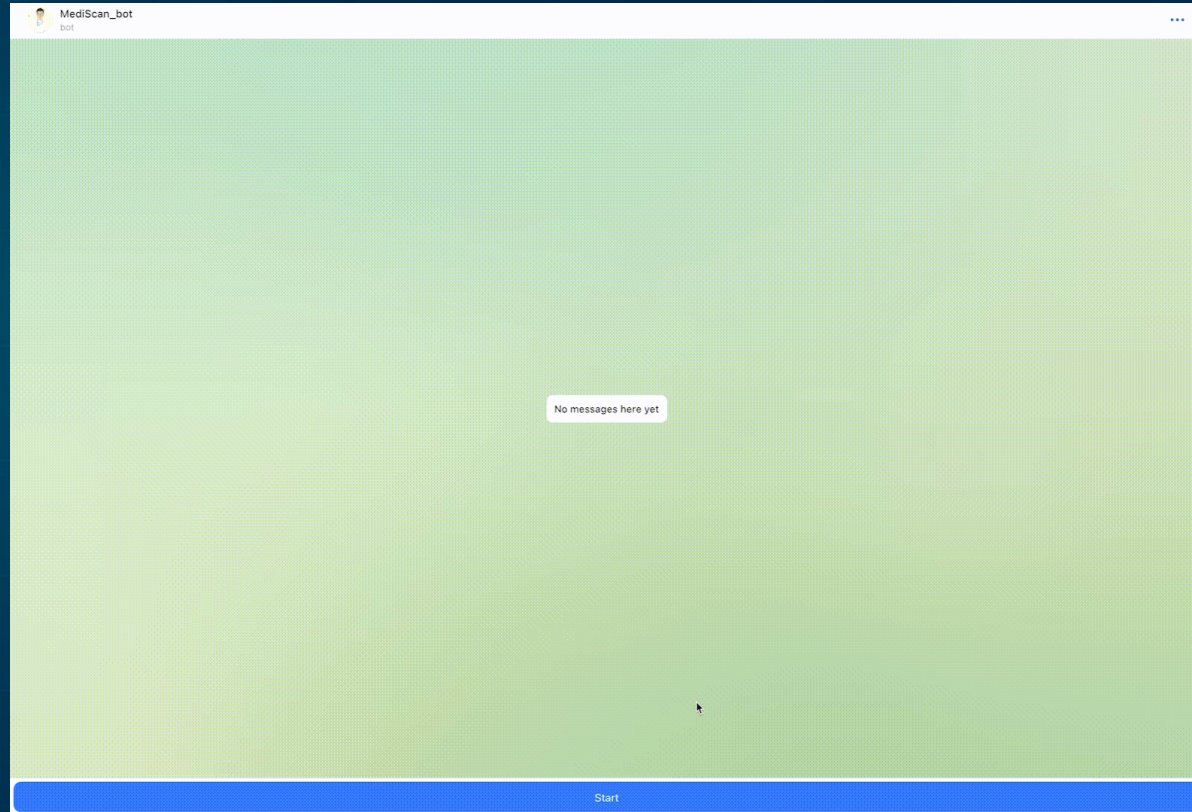
/rate

Предлагает оценить бота по шкале от 1 до 5

/showrating

Выводит накопившуюся усредненную оценку

Telegram bot



API

Основа – фреймворк **FastAPI**.

/ping

Проверяет доступность
сервиса

/predict

Выдает наиболее вероятный
тип новообразования на
изображении, используя
модель логистической
регрессии

/predict_cnn

Выдает наиболее вероятный
тип новообразования на
изображении, используя
CNN

/rate

Предлагает оценить бота по
шкале от 1 до 5

/current_rating

Выводит накопившуюся
усредненную оценку

/history

Выдает информацию о
предыдущих обращениях
к API

Хранение пользовательских данных

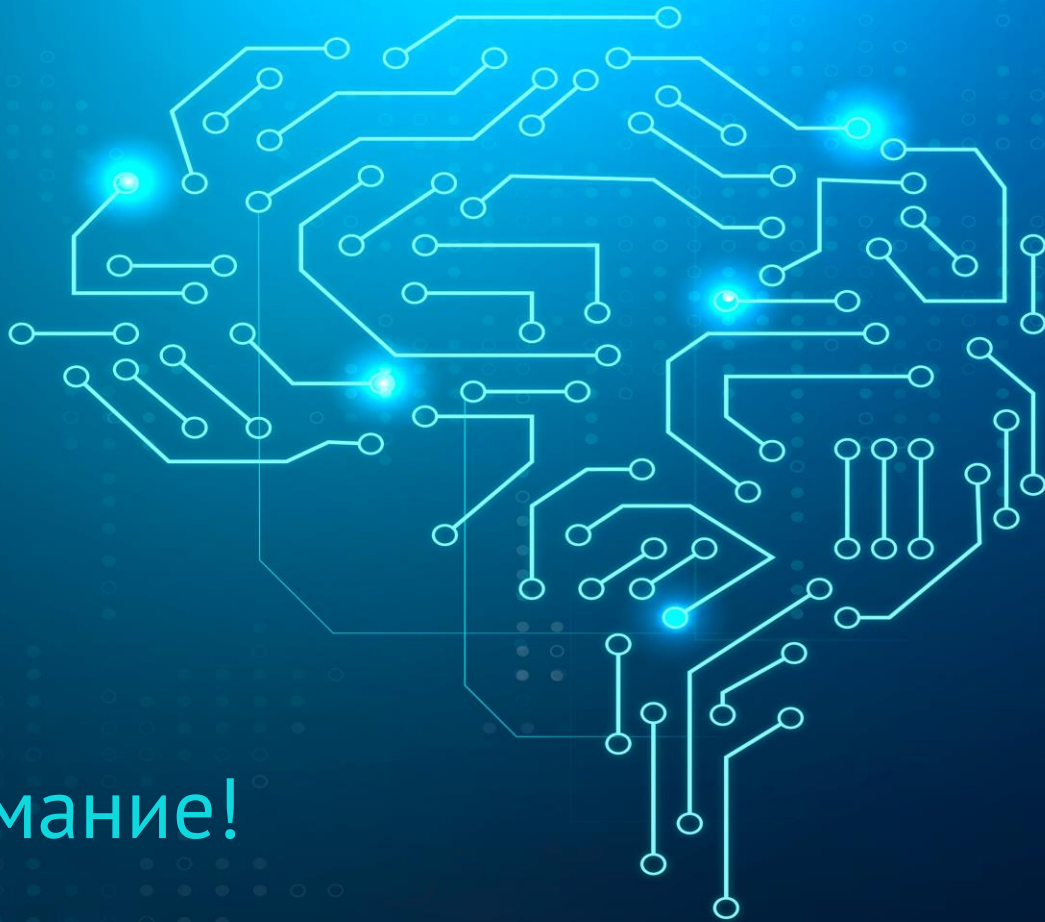


Развертывание



Бот и API-приложение запускаются
внутри отдельных Docker-контейнеров.
Их взаимодействие обеспечивается
посредством docker-compose

В качестве хостинга используется
облачная платформа **DigitalOcean**



Спасибо за внимание!