

Эффективный мониторинг территорий: *MLOps для семантической сегментации спутниковых снимков*

Павел Кикин

Газпромнефть-региональные продажи

Руководитель центра

t.me/pavel_kikin



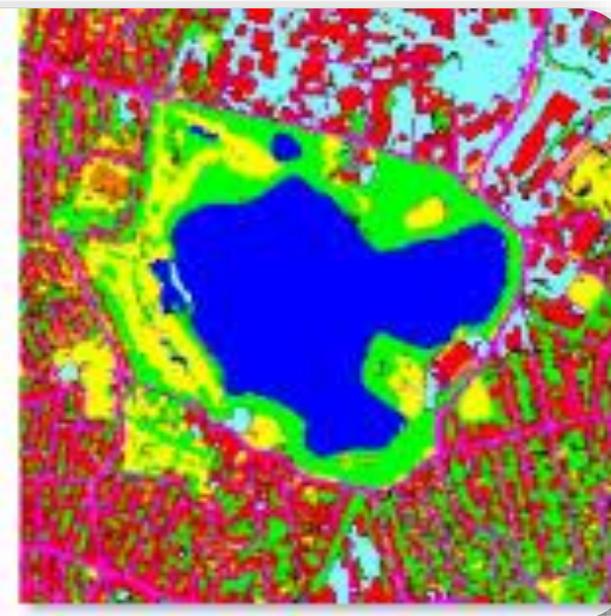
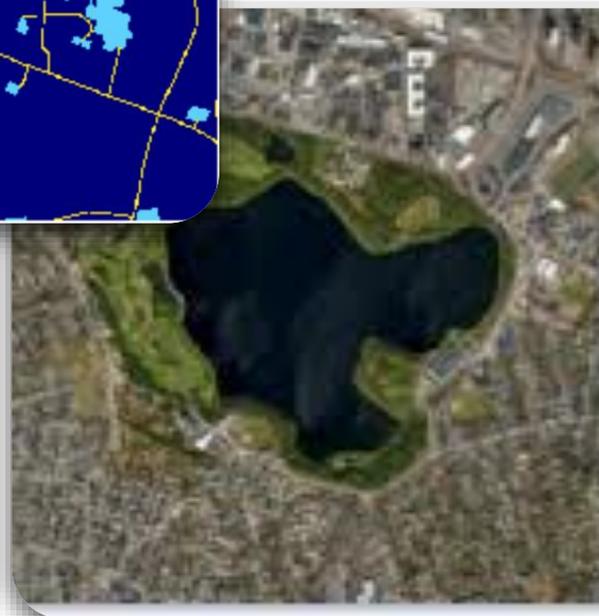
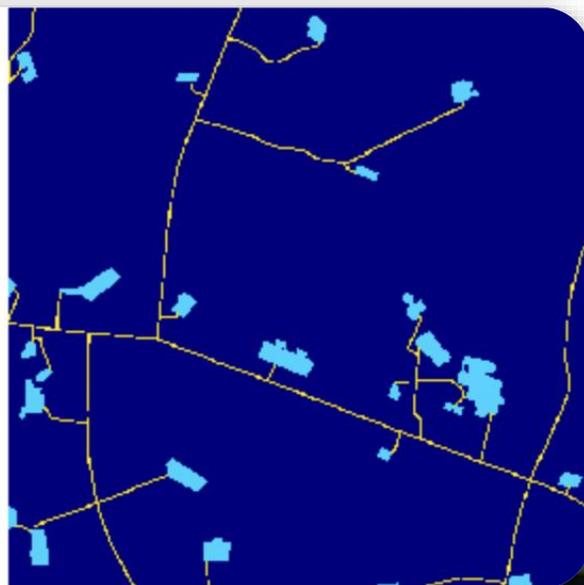
@PAVEL_KIKIN





Семантическая сегментация для мониторинга территорий

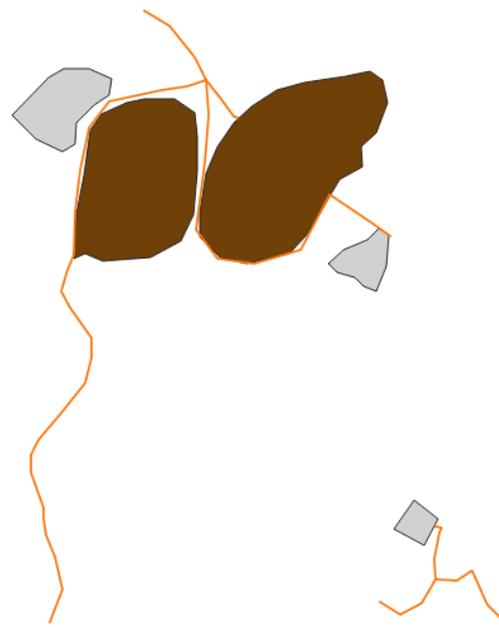
Пример:





Ключевые процессы для создания сервиса семантической сегментации

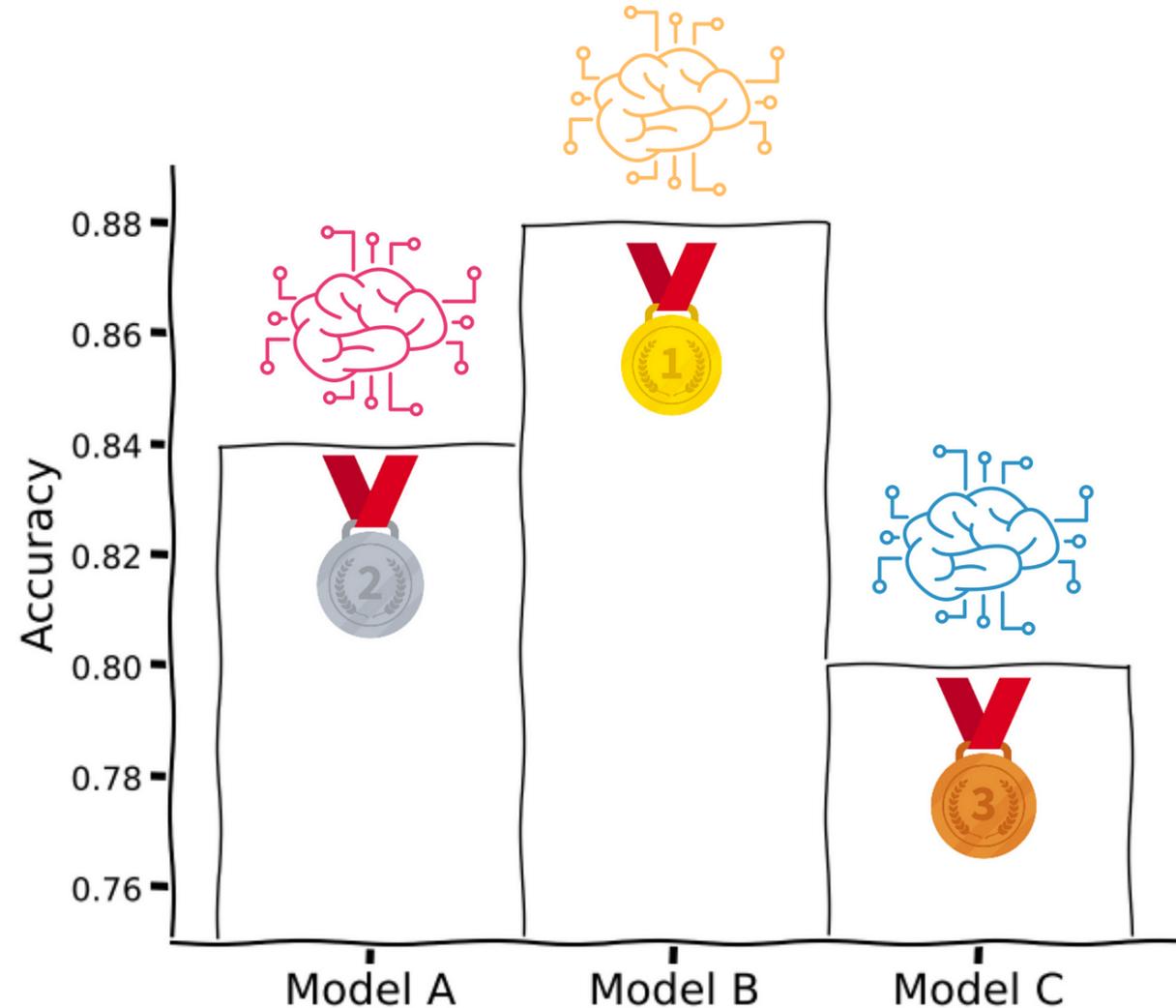
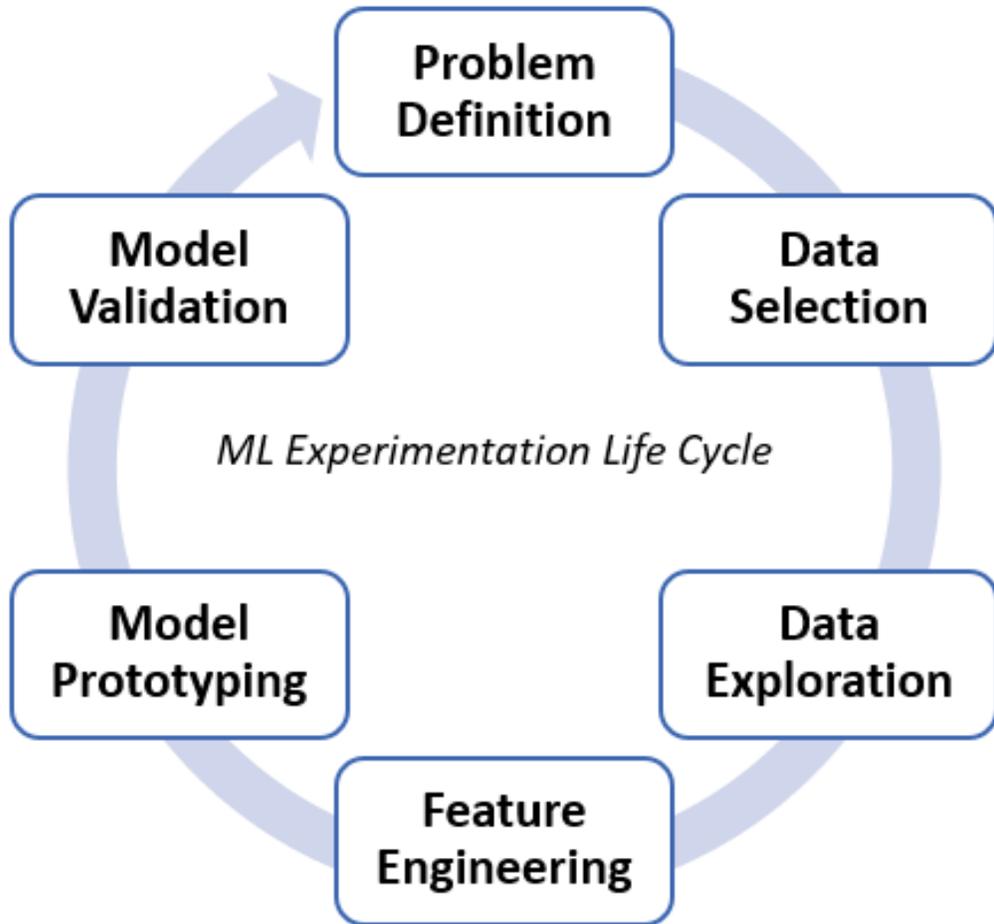
Разметка данных





Ключевые процессы для создания сервиса семантической сегментации

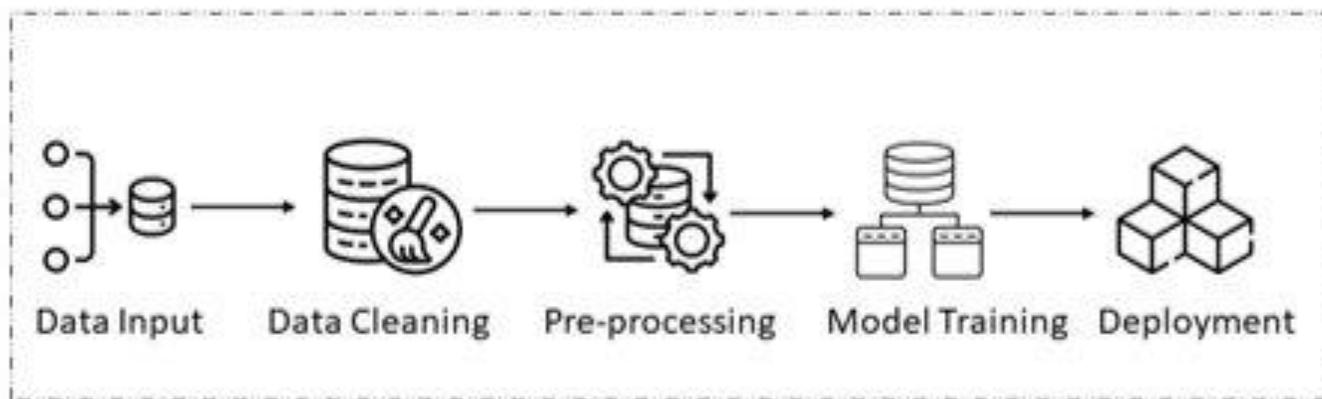
Проведение экспериментов





Ключевые процессы для создания сервиса семантической сегментации

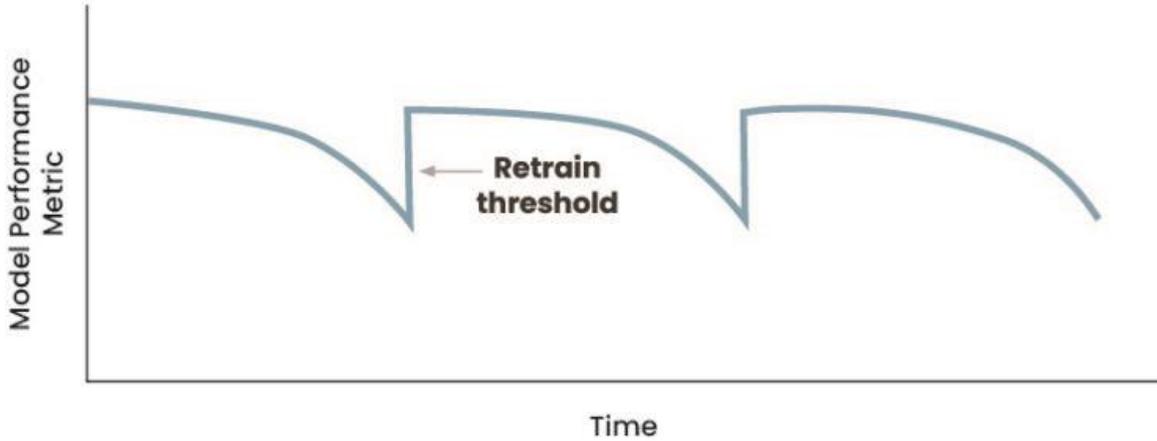
Обучение моделей





Ключевые процессы для создания сервиса семантической сегментации

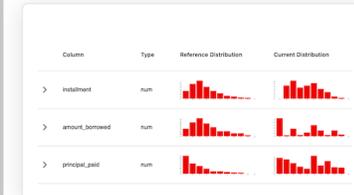
Устойчивое функционирование моделей



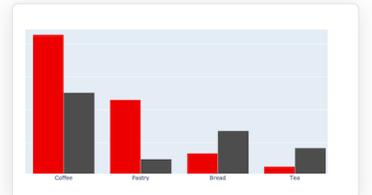
Data quality



Data drift



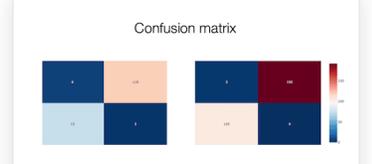
Prediction drift



Data integrity



Model quality





Требования к сервису

- Высокая производительность
- Воспроизводимость
- Устойчивость
- Актуальность
- Масштабируемость

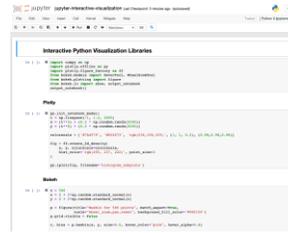


Monkey DS

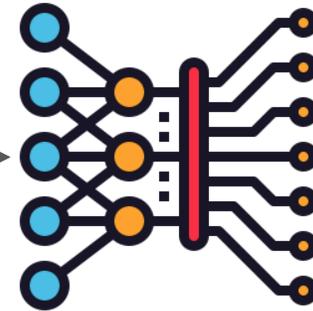
Ребят, файлы много весят, может на флешке передать?



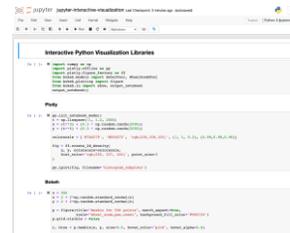
Конг, мне стало неудобно высекать метрики в камне!? А ты где их записываешь?



Я модель на гит залил



Ты их еше и записываешь?! Я в уме держу!





MLOps driven DS

- Грамотно спроектированная инфраструктура
- Продуманные процессы
- Систематизация
- Документирование
- Унификация



- Высокая производительность
- Воспроизводимость
- Устойчивость
- Актуальность
- Масштабируемость



Инфраструктура хранения данных

Требования

- Хранение больших объемов
- Геопривязка и картографические проекции
- Общий доступ с различных приложений

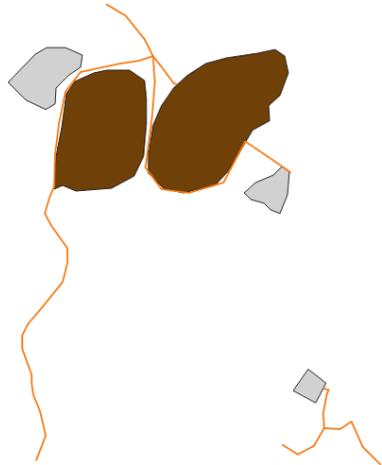


Инфраструктура хранения данных

Хранение векторных и табличных данных



PostgreSQL
PostGIS



SHP / GeoPackage

Метаданные

- пользователи
- статусы разметки
- параметры экспериментов
- размеры тайлов
- проекция



Инфраструктура хранения данных

Хранение растров



GeoTIFF



GeoTIFF



Инфраструктура хранения данных

Публикация геоданных через *https*



GeoServer

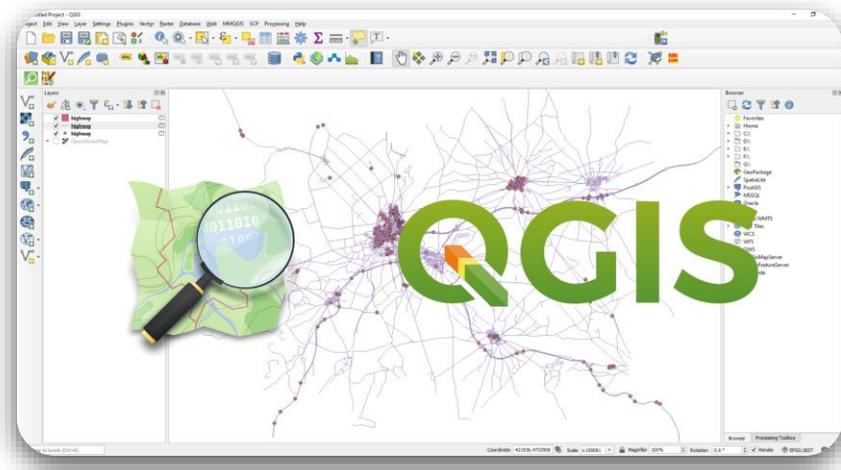


WMS/WFS



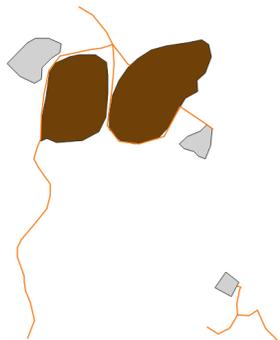


Инфраструктура хранения данных

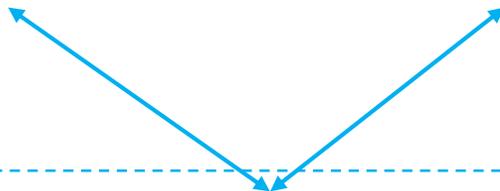


NGINX

WMS/WFS



GeoTIFF





Разметка *Требования*

- Контроль качества
- Эффективность:
 - Распараллеливаемость
 - Высокая скорость разметки
- Масштабируемый



Разметка

Функционал разметки

Sputnik_test

Выберите тип объекта для создания:

Промплощадки

Карьеры

Дороги

Строения

OK Отмена

Sputnik_test

Статус блока:
Не размечен

Назначенный исполнитель:
Фамилия ИО 1

Отправить на проверку

Загрузить автоматическую разметку

Загрузить OSM разметку



Разметка Процесс

GeoServer



1. Загрузка тайла

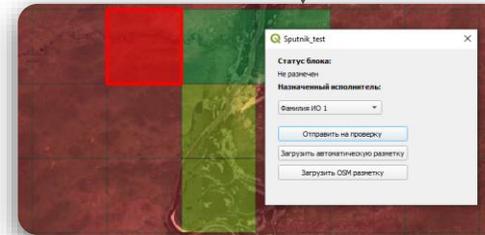
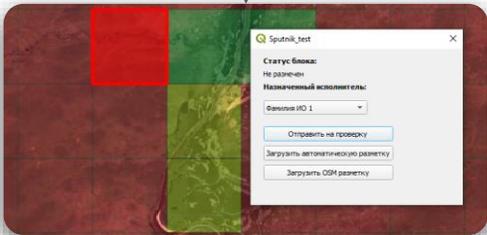
2. Разметка и
сбмит решения

3. Загрузка разметки на
сервер со статусом: test

4. Загрузка разметки
для валидации

4. Загрузка
корректной
разметки в
статусе: valid

5. Валидация и
корректировка





Разметка

Инфраструктура для пре-разметки и валидации

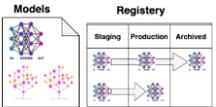


NGINX

Apache Airflow

ML-Service

mlflow



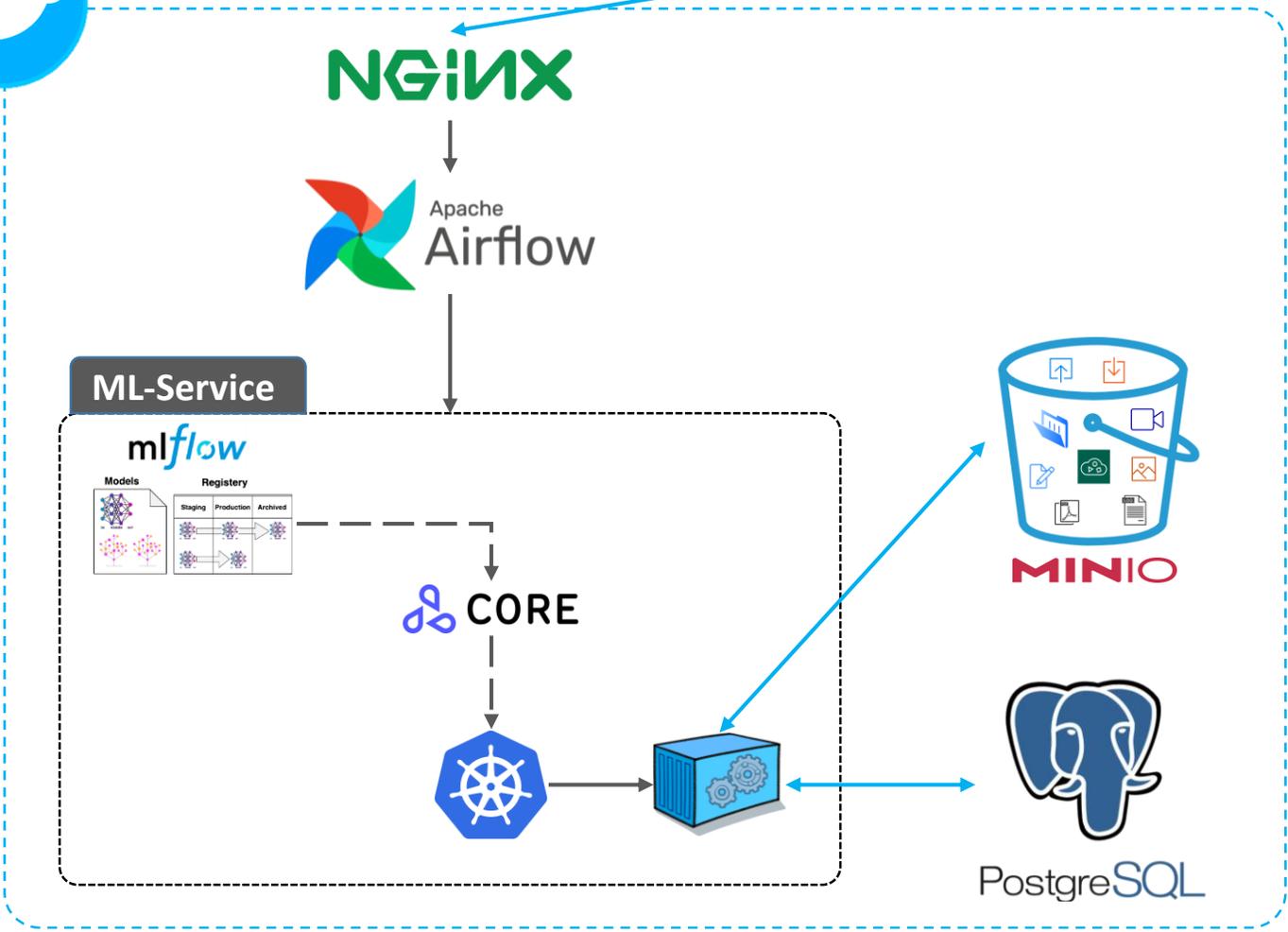
CORE



MINIO



PostgreSQL





Разметка

Дополнительные фичи

- Быстрое формирование и визуализация комбинаций каналов и индексов в QGIS (стандартными инструментами)
- Автоматическая активация лучшей растровой подложки при начале разметки



Проведение экспериментов

Требования

Задачи:

- Воспроизводимость
- Трекинг
- Пакетный запуск
- Сохранение и визуализация результатов сегментации (и метрик для каждого тайла с разбивкой по классам)



Проведение экспериментов

Инфраструктура

Источник данных
Оптимизатор
Набор моделей
Гиперпараметры
Комбинации каналов + индексы
Виды обработки данных



NGINX

mlflow



MINIO



PostgreSQL



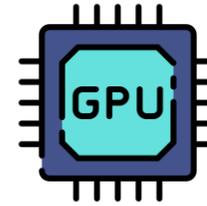
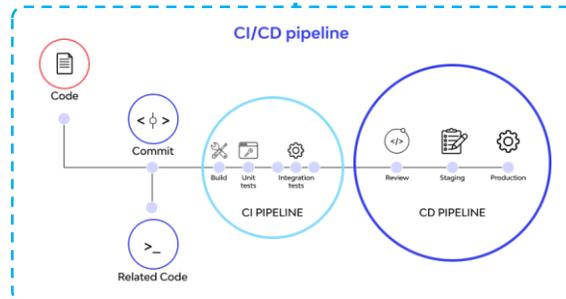
Apache
Airflow



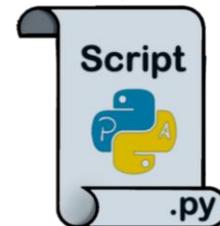
GitLab



nexus
repository



RUNNER



Script

.py

Эффективный мониторинг территорий: *MLOps для семантической сегментации спутниковых снимков*

Павел Кикин

Газпромнефть-региональные продажи

Руководитель центра

t.me/pavel_kikin

