



# Мурзинов Леонид

Data Engineer

## КОНТАКТЫ

- +995 599 515 383
- leonid.murzinov@gmail.com
- <https://t.me/MurzinovLeonid>
- Тбилиси, Грузия

## ОБРАЗОВАНИЕ

### Бакалавриат

Санкт-Петербургский  
Государственный Университет  
2016 - 2020

### Магистратура

Санкт-Петербургский  
Государственный Университет  
Экономики и Финансов  
2020 - 2022

### Прохождение обучения

BEEINTERNS Data Engineering от  
Билайн  
2022

## SKILLS

- Python
- SQL
- Airflow
- Docker
- Hadoop
- NoSQL

## ОПЫТ РАБОТЫ

Data Engineer

ноя 2022 - н.в.

### NtechLab

#### Обеспечил переход на новую систему хранения и обработки данных

- Участвовал в реализации инструментов для миграции данных
- Произвел перенос данных, разметки и мета информации в новую систему работы с данными
- Реализовал библиотеку, осуществляющую выгрузку, кеширование, получение лейблов данных, что позволило создать новые инструменты аналитики при тестировании и обучении моделей, получение дополнительной статистической информации по каждому датасету
- Участвовал в проектировании и реализации системы хранения лейблов для данных в PostgreSQL

Стек: Python, PostgreSQL, FastAPI, Ceph

#### Реализовал на основе новой системы хранения данных пайплайн для запуска тестирования моделей

- Реализовал и поддерживал автоматизированную систему получения лейблов с имплементацией решения с использованием нейросетей
- Участвовал в переходе на новый фреймворк для запуска тестирования моделей
- Участвовал в реализации системы визуализации ошибок в тестовых данных
- Участвовал в реализации инструментов по выгрузке и нарезки видео
- Участвовал в автоматизации процессов первичного сбора и обработки данных

Стек: Python, Airflow, Docker, Luigi, FFmpeg, Superset, Streamlit

Data Engineer

июн 2022 - окт 2022

### АгроТерра

#### С нуля разработал сервис обработки данных

- Извлекал, фильтровал и агрегировал данные из нескольких DWH (PostgreSQL, MS SQL), с использованием SQL скриптов.
- Написал модуль по нахождению ошибок и несоответствий в данных, используя библиотеки pandas, numpy.
- Написал даги для организации регулярной проверки данных с помощью Airflow.
- Организовал все приложение в виде Docker контейнеров, что облегчило тестирование и разворачивание на сервере.
- Создал веб-приложение на Flask, которое позволило, пользователям ответственным за данные, корректировать данные, вносить комментарии по ошибке, заводить тикеты, рассылать сообщения.

Стек: Python, Airflow, Docker, Flask, HTML, CSS

#### Организация ETL процессов в проекте по моделированию оптимального распределения техники между агрохозяйствами

- Извлекал, фильтровал, агрегировал данные разной природы и из разных источников (данные из реляционных и нереляционных бд, изображения)
- Написал модуль по очистке и трансформации данных для оптимизатора, решающего целочисленные линейные задачи
- Создал бд и модуль для загрузки данных в нужном формате

Стек: Python, SQL, Pyomo

## Nexus

### Участие в проекте по детекции и распознаванию гос номеров и типов автомобилей

- Занимался скрапингом данных при помощи BeautifulSoup, Selenium, API
- Занимался разметкой и преобразованием изображений для нейронной сети, используя различные методы и библиотеки
- Разворачивал на серверах с использованием Docker нейронных сетей, запускал, тестировал.
- Оптимизировал метрики качества нс

**Стек: Python, YOLOv4, Pillow, Pandas, Selenium, BeautifulSoup, Docker**

### Написание сервиса по детекции и распознаванию лиц людей

- Написал инфраструктуру для работы с библиотекой по распознаванию лиц Insightface. Работа с получением данных по API, с веб-части и т.д
- Написание веб-приложения для добавления изображений на Flask
- Занимался определением метрик по качеству распознавания и улучшению их
- Подготавливал датасеты для тренировки нс

## ДОПОЛНИТЕЛЬНЫЕ СВЕДЕНИЯ

### Выход в финальную стадию в хакатоне ISS Hack «Детекция оружия»

- Являлся капитаном команды, нашел членов команды с различными навыками
- Занимался обработкой данных
- Использовал вместе с командой методы машинного обучения и статистики

### Прохождение обучения "BEEINTERNS Data Engineering"

В обучение входили следующие модули:

- SQL
- Linux/Bash
- CI/CD
- SQL
- Hadoop
- Scala
- Spark
- Airflow

**Знаком с HDFS, Hive, Pig**

**Знаком с NoSQL бд: HBase, Cassandra, MongoDB**