

ИВАН КАТОРГИН

Data Scientist | Аналитик данных

 github.com/IvanKatorgin |  public.tableau.com/app/profile/ivan.katorgin


 Ижевск |  +7 (963) 455-19-23 |  @IvanKatorgin |  ivan.katorgin@gmail.com |  taplink.cc/ivankatorgin




О СЕБЕ

Data Scientist с более чем 20-летним опытом в нефтегазовой отрасли, где применял data-driven подходы для решения комплексных задач геологии и разработки месторождений. Специализируюсь на глубоком обучении (CV, NLP) и прогнозной аналитике. Разрабатываю решения для извлечения смысла из неструктурированных данных и создания интерактивных дашбордов. Имею опыт руководства командами и управления проектами. Открыт к сотрудничеству над проектами в области AI/ML, готов к релокации или удаленной работе

ЗНАНИЯ ЯЗЫКОВ

Русский  родной

Английский  upper intermediate

ПРОФЕССИОНАЛЬНЫЕ ОБЛАСТИ

Глубокое и машинное обучение (DL/ML), Компьютерное зрение (CV), Обработка естественного языка (NLP), Временные ряды (TS), BI-аналитика

DATA SCIENCE ПРОЕКТЫ

AI-диагностика патологий головного мозга по КТ с визуализацией решений *PyTorch, OpenCV, Grad-CAM*

- Разработал и обучил модели (ResNet18, EfficientNet) для классификации патологий по КТ-снимкам с точностью до 85%
- Реализовал механизмы внимания и визуализацию Grad-CAM для интерпретации решений модели
- Провел полный цикл: от анализа качества данных и аугментации до валидации и подготовки к клиническому применению

Многозадачная классификация атрибутов лица (пол, возраст, этнос) *TensorFlow, Keras, Transfer Learning, Pandas, NumPy, Scikit-learn*

- Спроектировал систему компьютерного зрения для одновременного решения 3 задач (классификация, регрессия) на общем энкодере
- Реализовал и сравнил 5 архитектур нейронных сетей. Лучшая модель достигла точности 92.3% для пола и 85.7% для этноса (MAE возраста: 6.2 лет)
- Применил механизмы внимания и многозадачное обучение для улучшения обобщающей способности модели

Генерация лиц по атрибутам с помощью GAN *TensorFlow, Keras*

- Разработал 5 архитектур GAN (Conditional GAN, DCGAN, DCGAN с Attention, гибридные модели) для генерации лиц по заданным атрибутам (пол, возраст, этнос)
- Создал функциональную систему генерации лиц с контролем атрибутов
- Проанализировал кластеризацию сгенерированных изображений в пространстве признаков

Сравнительный анализ механизмов внимания в Seq2Seq для машинного перевода *PyTorch, NLTK, Pandas, NumPy, Matplotlib, Seaborn*

- Реализовал и сравнил 3 механизма внимания (Scalar Dot-Product, Multi-Head, MLP-based) в архитектуре Encoder-Decoder
- Достиг BLEU-4 Score 0.85 на задаче англо-русского перевода
- Визуализировал матрицы внимания для анализа работы модели и провел абляционные исследования

Комплексное моделирование временных рядов: от ARIMA до машинного обучения *Statsmodels, Arch, Scikit-learn, Seaborn*

- Разработал универсальную методологию анализа и прогнозирования временных рядов

- Построил и сравнил множественные модели (ARIMA, SARIMAX, ARCH) для стационарных и нестационарных рядов, реализовал автоматический подбор параметров
- Достиг MAE 88 при прогнозировании на тестовых данных, используя как статистические методы, так и линейную регрессию на лагах

Прогнозирование кредитных дефолтов для микрофинансовой компании

Pandas, NumPy, Scikit-learn, XGBoost, Seaborn

- Разработал систему скоринга для прогнозирования дефолтов с accuracy 99.13% и AUC 0.99
- Провел полный цикл feature engineering, балансировку данных и сравнительный анализ 4 ML-моделей

Подробнее об этих и других DS-проектах:

github.com/IvanKatorgin

портфолио

КЛЮЧЕВЫЕ КОМПЕТЕНЦИИ

- Анализ данных в Python
- Feature engineering и предобработка данных
- CV: обработка изображений, сегментация, детекция
- NLP: работа с языковыми моделями, RAG-системы, анализ тональности
- TS: прогнозирование, аномалии
- DL: нейросетевые архитектуры
- Машинное обучение end-to-end
- Построение и комбинирование рекомендательных систем
- MLOps: развертывание и мониторинг моделей
- Generative AI: генерация текста, изображений
- Аналитика и BI: создание дашбордов и визуализация данных, A/B тесты

ТЕХНИЧЕСКИЙ СТЕК

- Программирование: Python
- Обработка и анализ данных: Pandas, NumPy, SciPy, Matplotlib, Seaborn, Plotly
- ML/DL: PyTorch, TensorFlow, Keras, Scikit-learn, XGBoost, LightGBM, Statsmodels
- NLP: NLTK, SpaCy, Transformers, BERT, Word2Vec, Natasha, Gensim
- CV: OpenCV, PIL, YOLO
- Рекомендательные системы: LightFM, Surprise
- БД: SQL, MySQL, PostgreSQL, MongoDB
- BI-инструменты, визуализация: Tableau, Power BI, Yandex DataLens
- Инструменты разработки: Git, Docker, Airflow, Hadoop, PySpark, Google Colab

ОПЫТ РАБОТЫ

01.2024 – наст. вр.

Учебные проекты Нетологии

Data Scientist

- Реализовал 15+ проектов полного цикла по направлениям ML, DL, NLP, CV
- Занял 1 место на воркшопе по аналитике данных Нетологии
- Финалист хакатона "ЭМкатон Экономики Москвы"
- Наставник студентов по курсу "Математика для Data Science"

10.2005 – наст. вр.

Нефтяная отрасль

Data Scientist / аналитик в нефтегазовой отрасли

Компании:

ЗАО «Ижевский нефтяной научный центр» (04.2018 – наст. вр.)

Главный специалист

ФАУ «Западно-Сибирский НИИ геологии и геофизики» (08.2016 – 12.2017)

Начальник отдела

ООО «Тюменский нефтяной научный центр» (02.2015-06.2016)

Менеджер департамента

ООО «РН-СахалинНИПИморнефть» (04.2007-02.2015)

Главный инженер проекта

ООО «Сервис Нафта» (10.2005-04.2007)

Разработчик

Основные обязанности и достижения:

- *Анализ данных и моделирование*: проводил анализ больших объемов промысловых и геологических данных, строил и калибровал геологические и гидродинамические модели
- *Проектирование разработки месторождений*: осуществлял оптимизацию разработки месторождений, прогнозирование добычи, расчет экономической эффективности, подготовку и защиту проектно-технологических документов в государственных органах
- *Оценка рисков*: применял вероятностные методы и статистическое моделирование для анализа неопределенностей и оценки рисков
- *Управление проектами и командами*: руководил мультидисциплинарными проектными командами (до 30+ человек), отвечал за планирование, исполнение и соблюдение сроков
- *Визуализация и презентация*: готовил комплексные дашборды, отчеты и презентации для руководства
- *Ключевые достижения*: за весь период работы подготовил и защитил в Государственных органах (ЦКР) 40+ проектно-технологических документов (портфолио: <https://drive.google.com/file/d/1K-yWzA7Fq6djo1b-Mq0pA3uxqzA1zxGn/view?usp=sharing>)

ОБРАЗОВАНИЕ

2025 **Нетология**

Data Scientist: с нуля до middle ([программа обучения](#))

2013 **РГУ Нефти и газа им. И.М. Губкина**

Экономика и управление на предприятиях нефтегазового комплекса

2006 **РГУ Нефти и газа им. И.М. Губкина**

Геология нефти и газа

Ключевые курсы:

- Data Science в медицине
- Основы Python: создаем Telegram-бота
- Tableau: анализ и визуализация данных
- Визуализация данных: от скучных графиков к интерактивным дашбордам