



Иван Каторгин Data Scientist, аналитик



КОНТАКТЫ



Ижевск (готов к переезду)



+7 (963) 455-19-23



Katorgin@inbox.ru



@IvanKatorgin



Сайт-визитка

ПОРТФОЛИО



[GitHub](#)

[Tableau](#)

НАВЫКИ

- ▶ Python
- ▶ SQL
- ▶ Big Data
- ▶ Machine learning
- ▶ Git
- ▶ Airflow
- ▶ Hadoop
- ▶ Tableau
- ▶ Power BI
- ▶ Yandex DataLens

Ключевые компетенции

- ✓ Работа с базами данных, написание SQL-запросов;
- ✓ PostgreSQL (CTE, views, window functions, subqueries, joins и т.д.);
- ✓ NoSQL (Mongo DB);
- ✓ Анализ данных в Python (Pandas, NumPy, MPL, SciPy, Pyspark, Seaborn, Scikit-learn, XGBoost, PyTorch, TensorFlow, Keras, LightBGM, NLTK, OpenCV и др. библиотеки);
- ✓ Feature engineering и предобработка данных;
- ✓ Выбор и реализация алгоритма под задачу;
- ✓ Построение ML-моделей;
- ✓ Создание нейросетей;
- ✓ Работа с текстовыми данными, изображениями и временными рядами;
- ✓ Построение и комбинирование рекомендательных систем;
- ✓ Планирование разработки проектов Data Science;
- ✓ Построение дашбордов и визуализаций в Tableau, Power BI, Yandex DataLens.

Опыт работы

01.2024 — наст. вр.

Учебные проекты в Нетологии
направление «Аналитика», курс «Data Scientist: продвинутая траектория»
(программа курса [по ссылке](#))

- ✓ Учебные проекты ([GitHub](#)):

Проект: Работа с атрибутами и построение модели на основе ансамблирования
Задачи: Классификации наличия болезни сердца у пациентов по показателям здоровья

Используемые технологии: Python (библиотеки pandas, numpy, MPL, seaborn, sklearn)

Результаты: Обучена модель определения наличия заболеваний и достигнуты показатели accuracy более 86%;

Проект: Нейронные сети

Задачи: Построение и обучение сверточной нейронной сети

Используемые технологии: Python (библиотеки tensorflow, keras, numpy, MPL, sklearn)

Результаты: построена и обучена большая сверточная нейронная сеть, которая дала на тестовой выборке accuracy больше 80%;

Проект: Гибридные рекомендательные системы

Задачи: Построение модели анализа предпочтений пользователей онлайн-кинотеатра

Используемые технологии: Python (библиотеки pandas, numpy, MPL, sklearn, lightfm, surprise)

Результаты: Обучена модель и получен список конкретных фильмов, рекомендованных пользователю;

Проект: Временные ряды. Нестационарный ряд

Задачи: Построение модели ARIMA

Используемые технологии: Python (библиотеки arch, pandas, numpy, MPL, seaborn, scipy, sklearn, statsmodels)

Результаты: Построены модели ARMA, ARIMA, ARCH, SARIMAX, а также



спрогнозирован график классическим ML, MAE которого составила 88.

- ✓ Участие в хакатонах и воркшопах (подробнее [по ссылке](#)):

Хакатон: Воркшоп по аналитике данных в Нетологии

Задачи: Исследование и решение кейса от компании DataNewton

Используемые технологии: Python (библиотеки pandas, numpy, MPL, seaborn, scipy, sklearn)

Результаты: Первое место (провел A/B-тесты и проверил несколько гипотез, построил модель линейной регрессии, предсказывающую выручку компании в зависимости от количества сотрудников);

Хакатон: Хакатон по направлениям BI, системного и бизнес анализа ([ЭМкатон Экономики Москвы](#))

Задачи: Решение тестовых заданий [Экономики Москвы](#)

Используемые технологии: Excel, Power BI

Результаты: Вышел в финал, но т.к. финал был офлайн, не смог прилететь из-за работы.

- ✓ В процессе обучения в Нетологии стал аспирантом-наставником на курсе «Математика для Data Science». Основные задачи: помощь студентам в выполнении домашних заданий, ответы на возникающие вопросы студентов.

10.2005 – наст. вр.

Работа в нефтяной отрасли, в области проектирования и моделирования разработки нефтяных и газовых месторождений в различных регионах и компаниях, на различных должностях (подробнее [по ссылке](#)).

Основные обязанности:

- ✓ Расчет прогнозных технологических показателей разработки;
- ✓ Анализ разработки месторождений;
- ✓ Подготовка предложений и мероприятий по оптимизации системы разработки;
- ✓ Подготовка программ исследовательских работ;
- ✓ Подготовка проектно-технологических документов на разработку месторождений нефти и газа;
- ✓ Защита проектных документов в государственных структурах.

Основные достижения (подробнее [по ссылке](#)):

- ✓ Подготовил и защитил в государственных структурах более 20 проектных документов на разработку месторождений нефти и газа и более 20 проектных документов на разработку месторождений подземных вод;
- ✓ Проектировал разработку месторождений с различными геолого-физическими характеристиками (нефть, газ, конденсат, разломная тектоника, газовые шапки, высоковязкие нефти, трудноизвлекаемые запасы, терригенные и карбонатные коллектора), с различными системами и стадиями разработки, по запасам от очень мелких до крупных и уникальных.

Образование



Нетология, 01.2024 - наст. вр (окончание обучения – 08.2025)

направление «Аналитика», курс «Data Scientist: продвинутая траектория» (программа курса [по ссылке](#))



Российский Государственный Университет нефти и газа им. И.М. Губкина, Москва, 2013

Экономика и управление на предприятиях нефтегазового комплекса



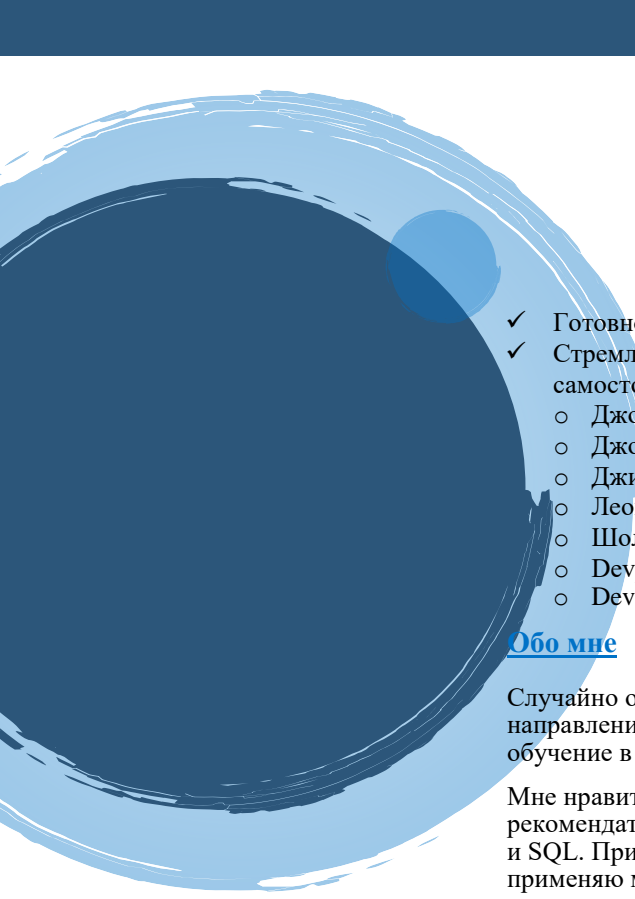
Российский Государственный Университет нефти и газа им. И.М. Губкина, Москва, 2006

Геология нефти и газа, горный инженер

(подробнее про образование: [тут](#) и [тут](#))

Дополнительная информация

- ✓ Иностранный язык: английский (B2 — upper intermediate);

- 
- ✓ Готовность к редким командировкам или переезду;
 - ✓ Стремление к профессиональному развитию. Помимо учебных материалов, самостоятельно изучил литературу по Data Science:
 - Джоэл Грас «Data Science. Наука о данных с нуля»;
 - Джон Форман «Много цифр. Анализ больших данных при помощи Excel»;
 - Джин Желязны «Говори на языке диаграмм»;
 - Леонард Апелцин «Data Science в действии»;
 - Шолле Франсуа «Глубокое обучение на Python»;
 - Devpractice Team «Pandas. Работа с данными»;
 - Devpractice Team «Библиотека Matplotlib» и др. учебные материалы.

Обо мне

Случайно открыв для себя Data Science и BI-аналитику в 2023 году, увлекся этим направлением и решил полностью изменить мою профессиональную жизнь, начав обучение в Нетологии.

Мне нравится работать с данными, проверять гипотезы, строить ML-модели и рекомендательные системы. В своей практической деятельности использую Python и SQL. При проведении A/B-тестов, а также на различных этапах анализа данных применяю математическую статистику. Знаком с нейросетями, люблю строить интерактивные дашборды в Tableau, либо в Power BI.

Ищу работу в удаленном, либо гибридном формате в качестве Data Scientist-a, либо аналитика в компаниях, которые занимаются развитием и внедрением машинного обучения и искусственного интеллекта.

