Ivan Katorgin
Data Scientist, analyst

CONTACTS

Izhevsk (ready to move)

+7 (963) 455-19-23

Katorgin@inbox.ru

@IvanKatorgin

Website

PORTFOLIO

GitHub

Tableau

SKILLS

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

Core Competencies

- ✓ Database management:
  - SQL query writing & optimization;
  - PostgreSQL: CTEs, views, window functions, complex joins, subqueries;
  - NoSQL: MongoDB;
- ✓ Python data stack:
  - Core: Pandas, NumPy, SciPy;
  - Visualization: Matplotlib, Seaborn;
  - ML: Scikit-learn, XGBoost, LightGBM;
  - Deep Learning: PyTorch, TensorFlow, Keras;
  - NLP/Computer Vision: NLTK, OpenCV;
  - Big Data: PySpark;
- ✓ Machine Learning Pipeline:
  - Feature engineering & data preprocessing;
  - Algorithm selection & implementation;
  - End-to-end model development;
  - Neural network architecture design;
- ✓ Domain Expertise:
  - Text data processing;
  - Image recognition;
  - Time series analysis;
- ✓ Production Systems:
  - Recommendation system development;
  - DS project lifecycle planning;
- ✓ Data Visualization:
  - Interactive dashboards (Tableau, Power BI);
  - Yandex DataLens.

Work Experience

01.2024 — present
Training projects in Netology
Analytics, course «Data Scientist: advanced track» (course program by link)

- ✓ Training projects (GitHub):

Project: Work with attributes and building a model based on ensembleing
Task: Classification of the presence of heart disease in patients according to health indicators
Used stack: Python (Pandas, Numpy, MPL, Seaborn, Sclearn)
Result: The model of determining the presence of diseases is trained and the Accuracy have been achieved more than 86%;

Project: Neural networks
Task: Building and training of a convincing neural network
Used stack: Python (Tensorflow, Keras, Numpy, MPL, Sklearn)
Result: a large convincing neural network has been built and trained, which gave more than 80%in the Accuracy test sample;

Project: Hybrid recommendation systems
Task: Building a model of analysis of user preferences of an online cinema

Used stack: Python (Pandas, Numpy, MPL, Sclearn, LightFM, Surprise)
Result: a model trained and a list of specific films recommended by the user was obtained;

Project: Temporary ranks. Non-stationary series
Task: Building the Arma model
Used stack: Python (Arch, Pandas, Numpy, MPL, Seaborn, Scipy, Sklearn, Statsmodels)
Result: Models ARMA, Arima, Arch, Sarimax were built, and the graph of the classic ML was predicted, MAE 88.

- ✓ Participation in hackathons and workshops (detailed **by link**):

Hackathon: Workshop on data analytics in Netology
Task: Cases Research and Solving from DataNewton
Used stack: Python (Pandas, Numpy, MPL, Seaborn, Scipy, Sklearn)
Result: first place (spent A/B tests and checked several hypotheses, built a linear regression model, predicting revenue of the company depending on the number of employees);

Hackathon: Hackathon of BI, System and Business Analysis (**Emkaton of the Moscow Economics**)
Task: Solving test tasks of the **Moscow economy**
Used stack: Excel, Power BI
Result: I went to the final, but because the final was offline, could not fly in due to work.

- ✓ During of learning in Netology, became a graduate student on the course "Mathematics for Data Science". The main tasks: assistance to students in doing homework, answers to students' questions.

10.2005 – present
Work in the oil industry, in the field of design and modeling of the development of oil and gas fields in various regions and companies, in various positions (detailed **by link**).

Main responsibilities:

- ✓ Calculation of forecast technological indicators of development;
- ✓ Analysis of field development;
- ✓ Preparation of proposals and measures to optimize the development system;
- ✓ Preparation of research programs;
- ✓ Preparation of design and technological documents for the development of oil and gas fields;
- ✓ Protection of design documents in state structures.

Basic results of work and achievements (detailed **by link**):

- ✓ Prepared and defended in government structures more than 20 design documents for the development of oil and gas fields and more than 20 design documents for the development of groundwater deposits;
- ✓ Designed the development of deposits with various geological and physical characteristics (oil, gas, condensate, fragmentary tectonics, gas caps, high -skinned oils, hard -to -reclaimed reserves, terrigenous and carbonate collectors), with various systems and stages of development, in reserves from very small to large and unique.

Education

Netology, 01.2024 - present (ending of training – 08.2025)
Analytics, course «Data Scientist: advanced track»
(course program **by link**)

Gubkin Russian State University of Oil and Gas, Moscow, 2013
Economics and Management at Oil and Gas Enterprises

Gubkin Russian State University of Oil and Gas, Moscow, 2006
Petroleum Geology

(more about education: **here** and **here**)

Additional Information

- ✓ Language: English (B2 — upper intermediate);
- ✓ Readiness for rare business trips or moving;
- ✓ Commitment to professional growth: Beyond formal coursework, completed extensive self-study including:
  - Joel Grus "Data Science from Scratch"
  - John Foreman "Data Smart: Using Data Science to Transform Information into Insight" (Excel-focused big data analysis)
  - Gene Zelazny "Say It With Charts" (data visualization)
  - Leonard Apeltsin "Data Science in Action"
  - François Chollet "Deep Learning with Python"
  - Devpractice Team "Pandas: Data Manipulation"
  - Devpractice Team "Matplotlib Library" and other educational materials

About Me

I discovered Data Science and BI analytics by chance in 2023 and became deeply passionate about the field, ultimately deciding to completely reshape my professional path.

What excites me most is working with data:

- Testing hypotheses and uncovering insights
- Building ML models and recommendation systems
- Creating interactive dashboards in Tableau/Power BI
- Applying mathematical statistics for A/B testing and data analysis

My toolkit includes Python and SQL, and I'm actively expanding my knowledge of neural networks.

I'm now seeking remote or hybrid roles as a:

- Data Scientist
- Analyst

in organizations that drive innovation through machine learning and AI implementation.