IVAN KATORGIN

Data Scientist | Data Analyst

github.com/IvanKatorgin | public.tableau.com/app/profile/ivan.katorgin
Izhevsk | +7 (963) 455-19-23 | @IvanKatorgin | ivan.katorgin@gmail.com | taplink.cc/ivankatorgin

ABOUT ME

Data Scientist with over 20 years of experience in the oil and gas industry, where I have applied data-driven approaches to solve complex geology and field development problems. I specialize in deep learning (CV, NLP) and predictive analytics. I develop solutions for extracting meaning from unstructured data and creating interactive dashboards. I have experience in team leadership and project management. I am open to collaborating on AI/ML projects and am willing to relocate or work remotely

LANGUAGES

Russian — native

English — upper intermediate

PROFESSIONAL DOMAINS

Deep Learning & Machine Learning (DL/ML), Computer Vision (CV), Natural Language Processing (NLP), Time Series (TS), BI Analytics

DATA SCIENCE PROJECTS

AI Diagnostics of Brain Pathologies from CT Scans with Decision Visualization
PyTorch, OpenCV, Grad-CAM
- Developed and trained models (ResNet18, EfficientNet) for pathology classification from CT scans with up to 85% accuracy
- Implemented attention mechanisms and Grad-CAM visualization for model decision interpretation
- Executed full project lifecycle: from data quality analysis and augmentation to validation and preparation for clinical application

Multi-Task Facial Attribute Classification (gender, age, ethnicity)
TensorFlow, Keras, Transfer Learning, Pandas, NumPy, Scikit-learn
- Designed a computer vision system for simultaneously solving 3 tasks (classification, regression) on a shared encoder
- Implemented and compared 5 neural network architectures. The best model achieved 92.3% accuracy for gender and 85.7% for ethnicity (Age MAE: 6.2 years)
- Applied attention mechanisms and multi-task learning to improve model generalization

Attribute-Guided Face Generation using GANs
TensorFlow, Keras
- Developed 5 GAN architectures (Conditional GAN, DCGAN, DCGAN with Attention, hybrid models) for generating faces based on specified attributes (gender, age, ethnicity)
- Created a functional face generation system with attribute control
- Analyzed clustering of generated images in the feature space

Comparative Analysis of Attention Mechanisms in Seq2Seq Architecture for Machine Translation
PyTorch, NLTK, Pandas, NumPy, Matplotlib, Seaborn
- Implemented and compared 3 attention mechanisms (Scalar Dot-Product, Multi-Head, MLP-based) in an Encoder-Decoder architecture
- Achieved BLEU-4 Score of 0.85 on English-Russian translation task
- Visualized attention matrices to analyze model performance and conducted ablation studies

Comprehensive Time Series Modeling: From ARIMA to Machine Learning
Statsmodels, Arch, Scikit-learn, Seaborn
- Developed a universal methodology for time series analysis and forecasting
- Built and compared multiple models (ARIMA, SARIMAX, ARCH) for stationary and non-stationary series, implemented automatic parameter selection

- Achieved MAE of 88 when forecasting test data using both statistical methods and linear regression on lags

Credit Default Prediction for a Microfinance Company
Pandas, NumPy, Scikit-learn, XGBoost, Seaborn
- Developed a scoring system for predicting defaults with 99.13% accuracy and 0.99 AUC
- Executed full cycle of feature engineering, data balancing, and comparative analysis of 4 ML models

More about these and other DS projects:
github.com/IvanKatorgin
portfolio

KEY COMPETENCIES

- Data Analysis in Python
- Feature engineering and data preprocessing
- CV: image processing, segmentation, detection
- NLP: large language models, RAG systems, sentiment analysis
- TS: forecasting, anomaly detection
- DL: neural network architectures
- End-to-end machine learning
- Building and combining recommender systems
- MLOps: model deployment and monitoring
- Generative AI: text generation, image generation
- Analytics and BI: creating dashboards and data visualization, A/B testing

TECHNICAL STACK

- Programming: Python
- Data Processing & analysis: 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
- Recommender systems: LightFM, Surprise
- Databases: SQL, MySQL, PostgreSQL, MongoDB
- BI tools, visualization: Tableau, Power BI, Yandex DataLens
- Development tools: Git, Docker, Airflow, Hadoop, PySpark, Google Colab

WORK EXPERIENCE

01.2024 – present
Training projects in Netology
Data Scientist
- Implemented 15+ end-to-end projects in ML, DL, NLP, CV domains
- Won 1st place at the Netology data analytics workshop
- Finalist at the "Emkaton of Moscow Economics" hackathon
- Mentor for the "Mathematics for Data Science" course

10.2005 – present
Oil and Gas Industry
Data Scientist / Analyst in the Oil and Gas Industry
Companies:
Izhevsk Petroleum Research Center (Rosneft Oil Company) (04.2018 – present)
Chief specialist

West Siberian Research Institute of Geology and Geophysics (08.2016 – 12.2017)
Department Head
Tyumen Petroleum Research Center (Rosneft Oil Company) (02.2015-06.2016)
Department Manager
RN-SakhalinNIPImorneft (Rosneft Oil Company) (04.2007-02.2015)
Chief Project Engineer
Service Nafta (10.2005-04.2007)
Reservoir Engineer

Main responsibilities and achievements:

- *Data analysis and modeling:* analyzed big data of field and geological data, built and calibrated geological and hydrodynamic models
- *Field development design:* optimized field development, predicted production, calculated economic efficiency, prepared and defended project technological documentation in State Authorities
- *Risk assessment:* applied probabilistic methods and statistical modeling for uncertainty analysis and risk assessment
- *Project and team management:* led multidisciplinary project teams (up to 30+ people), responsible for planning, execution, and meeting deadlines
- *Visualization and presentation:* prepared comprehensive dashboards, reports, and presentations for management
- *Main achievements:* prepared and defended in State Authorities 40+ project technological documentation throughout the career (portfolio: https://drive.google.com/file/d/1t2SvHT716eYRkbQu9WM8r2VoO_HuA-Wc/view?usp=sharing)

EDUCATION

2025 **Netology**
Data Scientist: Advanced track (course program)

2013 **Gubkin Russian State University of Oil & Gas**
Economics and Management at Oil & Gas Enterprises

2006 **Gubkin Russian State University of Oil & Gas**
Petroleum Geology

Key courses:

- Data Science in Medicine
- Python Basics: Creating a Telegram Bot
- Tableau: Data Analysis and Visualization
- Data Visualization: From Boring Charts to Interactive Dashboards