Nodar Okroshiashvili

Email: n.okroshiashvili@gmail.com
LinkedIn: linkedin.com/in/nodar-okroshiashvili
Github: github.com/Okroshiashvili

Data Software Engineer with 7+ years of experience architecting production-grade data applications and intelligent systems across fintech, e-commerce, cybersecurity, and document intelligence domains. Expert in Python ecosystem, cloud platforms, and real-time data processing with proven track record delivering end-to-end solutions from data collection to user-facing APIs. Successfully built custom search engines, credit scoring systems, anomaly detection platforms, and document intelligence solutions. Combines deep technical expertise with business domain understanding to solve complex problems through scalable and reliable data applications. Active open source contributor with experience collaborating on community-driven projects.

WORK EXPERIENCE

Senior Data Software Engineer   Jul 2023 – present
EPAM Systems

- Digitized 10,000+ IEC/ISO standards by architecting end-to-end XML-to-RDF transformation pipeline, enabling searchable knowledge graphs for enterprise compliance teams
- Developed data validation framework using SHACL shapes ensuring data integrity before RDF ingestion
- Prevented 95% of data integrity issues in production by implementing dual-layer validation system comparing XML sources against RDF output and RDF output to API response
- Created data enrichment pipeline generating embeddings and implemented semantic matching and search
- Accelerated recruitment efficiency by building job description validation pipeline using LLM tools
- Built intelligent email routing system to classify emails and send to dedicated address
- Enhanced system reliability through cross-team collaboration on data flow architecture
- Mentoring EPAM learning campus students on production data engineering practices

Senior Python Developer - Data Applications   May 2023 – Jun 2025
Nexus

- Built credit scoring decision engine from scratch, automating loan approval process and reducing application time from days to hours
- Developed anti-fraud detection system preventing fraudulent loan applications in real-time
- Architected medallion-pattern data lake on AWS (S3, Glue, Athena, PySpark) transforming raw transactional data into analytics-ready datasets for business dashboards and reporting
- Built data processing pipeline for e-commerce platform, cleaning raw data to enable faceted search capabilities
- Mentoring junior developers and helping them to curve out their career path
- Onboarded and trained 3 new team members on data engineering best practices

Associate Professor   Sep 2019 – Sep 2024
Business and Technology University

- Head of bachelor program in CS/AI

Python Developer   Dec 2023 – May 2024
InvestSuite   *Georgia, Tbilisi*

- Implemented duplicate event detection system for robo advisor platform, preventing duplicate trading orders and ensuring data integrity
- Designed and implemented document upload API processing and validating financial document submission
- Delivered production-critical bug fixes for public/private APIs serving as core product offering
- Enhanced system reliability by implementing comprehensive testing framework and code quality standards

Data Software Engineer   Dec 2022 – May 2023
CyFox   *Georgia, Tbilisi*

- Built real-time anomaly detection system identifying suspicious hacker activity using Windows event logs
- Collaborated with cybersecurity researchers collecting ML training data based on MITRE ATT&CK framework

- Integrated MITRE matrix threat intelligence enhancing detection accuracy for known attack patterns and techniques

Data Software Engineer Jul 2021 – Apr 2023

[Bestlist](#) *Georgia, Tbilisi*
- Built custom search engine delivering ad-free, curated products results eliminating noise from traditional search
- Enhanced search accuracy to 95%+ by training custom NER models for product entity extraction from unstructured data
- Created comprehensive metadata collection system automatically gathering data to enrich search results
- Achieved 30x performance improvement by enhancing existing data collection and storing logic
- Deployed serverless architecture using AWS Lambda and Step Functions with MongoDB for scalable search processing

Data Scientist Oct 2021 – Aug 2022

[Intrro](#) *Georgia, Tbilisi*
- Built similarity matching engine comparing job descriptions and resumes
- Trained Siamese neural networks for deep similarity pattern identification and scoring
- Designed and implemented data enrichment pipeline collecting additional data to augment candidate information
- Enhanced data collection workflows optimizing transmission and processing for improved matching performance
- Collaborated with backend developers to integrate matching systems into production platform

Data Scientist Jan 2021 – Sep 2021

[Delivery Hero](#) *Germany, Berlin*
- Built anomaly detection system identifying exogenous events during marketing campaigns
- Created time series similarity matching system for unequal-length data using dynamic time warping algorithm
- Conducted cluster analysis to optimize campaign allocation strategies for A/B testing
- Enhanced A/B testing framework by integrating anomaly detection logic into campaign evaluation workflows

Data Scientist Aug 2019 – Sep 2021

[MaxinAI](#) *Georgia, Tbilisi*
- Built end-to-end Question-Answering system processing PDF files and automatically answering business questions
- Integrated multiple OCR services for high-accuracy handwritten document text extraction
- Fine-tuned and deployed RoBERTA model for domain-specific question-answering
- Created automated email workflow enabling PDF submission and answer delivery via email

Data Scientist Aug 2018 – Jul 2019

[Credo Bank](#) *Georgia, Tbilisi*
- Built customer segmentation pipeline analyzing behavior and needs for targeted marketing campaign preparation
- Implemented market basket analysis identifying new customer segments for untapped marketing opportunities
- Developed credit scoring model migrating legacy system to Python for improved performance and maintainability

Data Analyst Oct 2017 – Jul 2018

[Liberty Bank](#) *Georgia, Tbilisi*
- Managed ATM cash demand prediction software and coordinated fleet dispatch operations
- Built clustering algorithm to group ATMs by location and behavior for optimized cash distribution strategies
- Developed custom time series models (ARIMA, SARIMA, EWMA) to improve cash demand prediction accuracy

EDUCATION

Georgian Technical University Georgia, Tbilisi

PhD - Computational Linguistics *Dropout to pursue industry work*

International School of Economics at Tbilisi State University Georgia, Tbilisi

Master of Arts, Economics *Aug. 2014 – Aug 2016*

SKILLS

Languages & Core Technologies: Python, SQL, FastAPI, Pydantic
Data Processing: Pandas, Polars, PySpark, spaCy, NLTK, Pillow
Machine Learning: PyTorch, Keras, Scikit-Learn, NumPy, SciPy, XGBoost, Hugging Face
Databases: PostgreSQL, MongoDB, Redis, DuckDB, RDF/SHACL, Qdrant, Chroma, Milvus
Cloud & Infrastructure: AWS (S3, Lambda, Glue, Athena, EC2, SQS, SNS, Redshift), Azure Databricks
DevOps: AWS (SAM, Cloudformation), CI/CD pipelines (Github Actions, Gitlab CI/CD), Terraform
Search: OpenSearch, Elasticsearch, MongoDB Atlas
Tools: Airflow, dbt, Kafka, Weights & Biases, DVC, MLflow, Docker, Git