Simon Mokrov | Data Scientist

Amsterdam, the Netherlands

CONTACTS

mokrovsimon@gmail.com | +31 (06) 579-875-47 | linkedin.com/in/sem-wett

simon.mokrov@hva.nl | +31 (06) 579 875 47 | sem_wett

github.com/semwett0301 | gitlab.com/semen.mokrov | English | Russian | Dutch (A1)

ABOUT ME

Data Scientist with a robust foundation of **3+ years in software engineering**, bringing extensive experience in building scalable, high-performance systems across the full stack. Skilled in programming, data modeling, and statistical analysis, with hands-on expertise in Python, machine learning workflows, and visualization.

Professional background spans enterprise retail, EdTech, and banking software, where I delivered impactful features, optimized performance, and enhanced system scalability. Holding a Master's in Digital Driven Business (AUAS, Amsterdam), I specialized in data-driven development, machine learning, and applied statistics.

APPLIED PROJECTS

1. **Machine Learning Pipelines** — Performed full-cycle ML workflows including exploratory data analysis (EDA), feature selection and engineering, and model development using Decision Trees, Random Forest, Linear Regression, ElasticNet, and XGBoost. Applied cross-validation and hyperparameter tuning to improve generalization and performance.
2. **Deep Learning** — Developed and trained Multilayer Perceptron (MLP) and Recurrent Neural Network (RNN) models as part of AI systems coursework.
3. **Statistical Analysis & Clustering** — Conducted statistical EDA, applied feature engineering, and implemented clustering algorithms (KMeans, DBSCAN, HDBSCAN) to uncover data structure. Performed hypothesis testing and statistical validation to ensure robustness.
4. **Recommender Systems** — Built multiple recommendation engines (SVD++, item-based and user-based collaborative filtering, SLIM) and evaluated them across five key metric categories (accuracy, diversity, novelty, serendipity, coverage). Deployed systems with custom backend and frontend components.
5. **Causal Inference** — Designed and applied multi-level regression linear models to extract causal insights from large-scale scraped datasets, identifying key factors influencing observed outcomes.
6. **Web Scraping & Data Engineering** — Collected and stored a large-scale dataset (over 1M movie records) using Scrapy and Selenium.

RESEARCH PROJECTS

MultiEval: Bias-Mitigated, Cost-Efficient LLM-as-a-Judge Framework *In submission (2025)*

Co-authored with a Senior Researcher from **Amazon Research**, this forthcoming paper presents **MultiEval**, a novel evaluation framework addressing core limitations of LLM-as-a-Judge systems—namely *position bias*, *length bias*, *and high computational cost*.

The framework integrates two modular components:

Calibrated Judge: mitigates position bias using randomized prompt swapping and leverages a cascade of fine-tuned and high-capacity LLMs for cost reduction.

Calibrated Scorer: applies a post-hoc logistic regression model to correct for length bias in pairwise comparison results.

MultiEval was evaluated on **Vicuna** and **AlpacaEval** datasets, showing improved fairness, reduced verbosity-related discrepancies, and lower evaluation costs compared to existing frameworks. This work contributes a **scalable and bias-resistant evaluation pipeline** for future LLM benchmarking research.

SKILLS

Data Science & ML: Python, SQL, Pandas, NumPy, SciPy, scikit-learn, XGBoost, CatBoost, Decision Trees, Random Forest, ElasticNet, PyTorch, MLP, RNNs, Transformers, KMeans, DBSCAN, HDBSCAN, SVD++, User/Item-based CF, SLIM

Data Engineering & MLOps: ETL pipelines, MLflow, Docker, Git, CI/CD, AWS/GCP

Visualization: Matplotlib, Seaborn, Plotly, Tableau

Software Engineering: JavaScript, TypeScript, Java, Kotlin, C, React, Redux (TanStack/RTK), Vue.js, Next.js, NestJS, Spring Boot, FastAPI, gRPC, Kafka, RabbitMQ, PostgreSQL, MongoDB, Redis

SOFTWARE DEVELOPING EXPERIENCE (> 3 YEARS)

1. **Frontend Developer** *March 2024 – Present*

   X5 Tech — *Largest retail in Russia*

   Developed and supported subscription services ("Package", "Abonement") in a corporate, cross-functional environment.

   Tech stack: React, TanStack Query, pnpm, Axios, Chart.js, Python & FastAPI (reading)

- Increased customer retention by 12% through frontend improvements.
- Reduced release time-to-market by 15% within 6 months.
- Accelerated delivery of planned features by 10% via Scrum practices.
- Reduced production issues by 18% after refactoring codebase.

2. **Backend Developer** *July 2024 – Feb 2025*

ITMO University *— Banking software for Gazprombank*

Enhanced internal systems for bank guarantees, optimizing backend workflows and enterprise data processing.

Tech stack: Java, Kotlin, Spring Boot, Gradle, PostgreSQL, Kafka, SOAP, Docker

- Reduced average response latency by 12% via endpoint refactoring.
- Improved PostgreSQL query execution by up to 20%.
- Cut report build time from 90s to 60s.
- Lowered server load by 20% after migrating to Kafka.

3. **Full Stack Developer** *Nov 2023 – Apr 2024*

Tune IT *— EdTech solutions*

Developed and maintained the Skillfactory educational platform and the Polytech introduction system for abiturients, focusing on scalability, usability, and performance.

Tech stack: React, TypeScript, RTK Query, SCSS, Styled Components, Spring Cloud, Java, Kotlin

- Resolved 30+ product issues, improving stability.
- Reduced design-to-development cycle by 25% with custom UI components.
- Cut service latency by 13% through modularization with Spring Cloud.

4. **Full Stack Developer** *Mar 2023 – Nov 2023*

MagicGophers *— Social apps for VK (largest social network in Russia)*

Built VK mini-apps for third-party clients, combining frontend and backend development.

Tech stack: React, VK UI, VK Bridge, VK Router, VK API, NestJS, PostgreSQL, Redis

- Delivered apps adopted by 10,000+ VK users.
- Increased conversion rates by up to 20% via UI/UX improvements.
- Reduced bug reports by 30% within 6 months.
- Optimized DB queries, reducing latency by 25%.

5. **Frontend Developer** *July 2022 – Mar 2023*

ITMO University *— EdTech (remote exam monitoring)*

Redesigned frontend for ITMOproctor, improving scalability and user experience.

Tech stack: React, Redux, Vue.js, JavaScript, WebRTC, WebSockets, Axios

- Reduced time-to-market for new features by 20% through modular architecture.
- Decreased load times by 10% with optimized routing, state, and sockets.
- Enabled secure streaming for 5,000+ concurrent exams using WebRTC.

EDUCATION

Master (MSc) **AUAS (HvA)**
Amsterdam, the Netherlands *2024 – 2025*

Major: Digital Driven Business

Coursework:

- Statistical testing
- Machine learning
- Clustering
- Deep learning
- Scrapy & Selenium
- RecSys
- Causal inference
- MLR models
- Research skills

Bachelor (BSc – Exchange semester) **Ozyegin University**
Istanbul, Turkey *2023 – 2023*

Major: Computer Science (exchange)

Coursework:

- Testing and analyzis
- Parallel Computing
- Advanced Databases
- Computer Networks

Bachelor (BSc) **ITMO University**
Saint-Petersburg, Russia *2020 – 2024*

Major: Software Engineering (System and Applied Software)

Coursework:

- Machine Learning Fundamentals
- Artificial Intelligence Systems
- Cloud Infrastructure
- Linear algebra
- Probability theory
- Statistics
- Calculus
- Computer Vision
- Algorithms and Data Structures
- Databases and Data Management
- Operating Systems and Low-level Programming