

# Mikhail Maximov

[linkedin.com/in/m-maximov/](https://www.linkedin.com/in/m-maximov/) | maksimov.xxxxxxx@xxxxx.com | +44 7767 5984XX | Berlin, Germany  
6+ YOE in back-end development of high-load SaaS, distributed web applications and MLOps

## EXPERIENCE

### Software Engineer II, Inference team

#### DataRobot

Mar 2019 – Aug 2022

- Developed a solution for preloading heavy ML models in low latency prediction servers, which resulted in a significant improvement in request processing time after the initial cold start. Specifically, the new preloading mechanism reduced the time to process requests for such models by 90%. It reduced amount of customers who experienced cold start delays from 3% to 0% which led to better service and a higher customer satisfaction rate
- Designed and implemented a Dual Inference environment project that enabled the separation of the API and compute components of the Predictions server. This resulted in savings of \$50,000 per year for the company
- Built and integrated a zero-downtime system of Prediction Cluster Reprovisioning. It allowed reducing the number of related production incidents from 3 incidents per year to 0 for the next 3 years and until now
- Achieved a high resiliency of the prediction servers even during database outages by introducing new ML models artifacts lifecycle founded on reliable storage, which allows avoiding hours of downtime (100k+ failed requests).

### Backed Engineer, AI API team

#### DataRobot

- Led the development of Request Logging and Activity Tracking Dashboard: Designed the system, coordinated the work of 4 other engineers in different time zone (Kyiv, Boston), and facilitated cross-functional collaboration by organizing and leading teamwide meetings, cross domains communication
- Worked on the development of API and SDKs for the company's new B2C product. Managed up to 5 developers in the team: reporting, conducting weekly 1x1s and providing guidance on career and goal planning
- Built performance monitoring service founded on APM for the team and generalized the solution for the company. Afterward, this solution was integrated by 3 other teams

### Software Engineer

#### DataArt

Jan 2017 – Mar 2019

- Built tracks and images assets management system integrated with 6 world-class audio streaming platforms
- Designed and implemented a high-load system for assets processing (transcoding and modification). The performance was increased by 40% for images and 35% for audio compared with the legacy system
- Led the implementation and automation of a new audio track transcoding, which replaced the legacy system and dramatically reduced the number of invalid track outputs from 0.1% to 0%

### Backend Developer

#### Design and Test Lab

Jun 2015 – Jan 2017

- Developed Risk Investment Calculator web application with a load of 300 generated reports per second in peak
- Designed and built Identity Management SaaS with billing system and industrial printer integrations
- Implemented automated testing and monitoring, which resulted in a 90% decrease in the number of bugs

## TECHNICAL SKILLS

- **Technologies:** AWS (S3, SQS, SNS, Lambda, DynamoDB, etc.), MongoDB, PostgreSQL, ELK, Redis, Kafka, Cassandra, RabbitMQ, GCP, Docker, Kubernetes, Snowflake, Bash, Jenkins, Terraform, Linux, Git, ffmpeg
- **Languages and related frameworks:** Python, Flask, Pandas, NumPy, Tornado, FastAPI, Java, JUnit, JS, React

## EDUCATION

### Kharkiv National University of Radio Electronics

SEP 2011 – May 2017

Bachelor of **Computer Engineering** (2015)

Kharkiv, Ukraine

Master's Degree in **Specialized Computer Systems** (2017)

GPA: 81.2/100

Certificates and Courses:

- **AWS Certified Solutions Architect** – Associate (2017)
- Cloud Computing Concepts, Part 1, Part 2 (Coursera)
- Cloud Computing Applications (Part 1-2): Cloud Systems and Infrastructure, Big Data (Coursera)
- Advanced Algorithms and Complexity (Coursera)