Mark Andreev, Senior Software Engineer

London, United Kingdom

Phone: +4407458900595

Email: mark.andreev@gmail.com

Github: https://github.com/mrk-andreev   LinkedIn: https://www.linkedin.com/in/mrk-andreev/

Education

September 2016 - June 2018. Master's degree

Lomonosov Moscow State University, Master of Applied Mathematics and Informatics.

Thesis: "NLP in macroeconomics monitoring".

September 2012 - June 2016. Bachelor's degrees

Moscow Power Engineering Institute (National Research University). Mathematical modelling.
Thesis: "Face recognition".

Experience

Senior Software Engineer, **Sep 2022 - present**

Conundrum.AI, // London, England, United Kingdom

Development of Machine Learning Platform for Industrial AI

- Implement low level optimization for Python SDK that helps integration with platform | Decrease RAM usage for 30%
- Implement low level optimization for feature store on top of Kafka & Clickhouse (projections, application level query planner) | Increase one of the most popular queries 3x times
- Create performance regression tests (Gatling) | Cover 80% of queries that eliminates major performance degradations
- Create performance optimizations for kafka subscription proxy (java 21 virtual threads, shared subscription) | Decrease CPU load 5x times
- Implement security improvements for Platform (audit, L4 network policies, L7 network filter) | Apply security IS requirements at network level
- Cover platform's services with health performance metrics (Prometheus, Grafana, Alerts) | Decrease issue investigation time 3x times

Senior Software Engineer, **Sep 2019 - Aug 2022**

Conundrum.AI, // Moscow, Russia

Development of Machine Learning Platform for Industrial AI

- Migrate feature store to Kafka & Clickhouse (Column OLAP DB) | Increase query speed 15x times
- Create low level connectors for Industrial Data Exchange formats (MQTT, OPC UA, Historian) | Decrease CPU load to exchange server 3x times
- Migrate model serving runtime to Kubernetes (KubeAPI, Helm)
- Deploy platform to AKS (Azure Cloud) & K3s (on premisses, no internet)

Middle Software engineer, **Nov 2017 - Sep 2019**

Conundrum.AI, // Moscow, Russia

Development of Machine learning & Data pipelines for Industrial AI

- Create feature store for sensor's time series data (Java, Spring, PostgreSQL, TimescaleDB)
- Create model serving runtime server (Python, Processes)

- Create incident management service (Java, Spring, State machines)
- Create ETL based on S3, SQS, S3 SFTP

Junior Machine learning engineer, **May 2017 - Oct 2017**

Conundrum.AI, // Moscow, Russia

Development of Machine Learning Solutions as consulting projects

- Airline data clusterization. Create approach for data splitting for offline AB tests
- Telecom data chron. Offline chron scoring based on telecom data activity
- Web data gender detection. Offline gender detection based on web activity
- Mobile data geo analysis. Create reports about geo activity based on mobile location data
- Time Series data for Industrial data. Create data pipeline for failure prediction
- Create ETL based on S3, SQS, S3 SFTP

Machine learning engineer, **Oct 2016 - May 2017**

Big Data Indicators · Internship, // Moscow, Russia

Development of end to end data mining pipeline for Economic Indicators

- Create data collection & processing pipeline
- Use topic models for discover trends
- Create sentiment analysis models for trend prediction

Development stack

- **Java**. Spring: MVC, Data, AMQP, Kafka, Security, State machine, Apache Camel, Vert.x, GraalVM
- **Python**. Pandas, Scikit-learn, Matplotlib, XGBoost, LightGBM, PyTorch, Keras, Catboost, Tornado, Flask, FastAPI.
- PostgreSQL, Clickhouse, Kafka, MongoDB, RabbitMQ, Redis, Keycloak, Prometheus, Docker, Kubernetes, Helm, Airflow, Spark, Cassandra, Hadoop, Linux, Terraform, Ansible, MLFlow
- **Cloud**. **AWS**: EC2, S3, RDS, CloudFront, SQS, SNS, Lambda, Batch, IAM; **Azure**: VM, Blob.

Contribution to Open Source.

- Apache Airflow. Add ipc_mode for DockerOperator. AIRFLOW-27553
- Apache Ignite. Implemented target encoding preprocessor. IGNITE-13713
- Apache Ignite. Implemented Yandex Catboost inference integration. IGNITE-13714
- Apache Ignite. Implemented new distances (BrayCurtis, Canberra, JensenShannon etc). IGNITE-13386
- Apache Camel. Fix Azure Blob Storage and Azure Blob Queue interaction. CAMEL-16092
- Keycloak. Fix null username in ldap. KEYCLOAK-19743
- Keycloak. Fix NPE if the user does not exist in PolicyEvaluationRequest. KEYLOCK-16239
- Tornado Swagger. Swagger API Documentation builder for tornado server.

Conferences / Public speech

Dec, 2022. Speech "Kafka 101**"** in MLSystemdesign_2022_ods Meetup.

Feb, 2019. Speech "ML in production" in FunTech Meetup

May 2018. Volunteer Data Scientist at EnduringNet (founded by Ser-Huang Poon, prof Manchester University)

July 2017. Big Data approach to measure inflation expectations: the case of the Russian economy (IFABS 2017 Oxford Conference), Goloshchapova, I., & Andreev M.

May 2017. Measuring inflation expectations of the Russian population with the help of machine learning (Voprosy Ekonomiki), Goloshchapova, I., & Andreev M.