

Denis Gontcharov

Gontcharov Data Services BV
Proficient in Dutch, Russian, English, French and German
Henri Van Heurckstraat 9, bus 302 · 2000 Antwerp · Belgium
denis@gontcharov.be — <https://gontcharov.be> — <https://github.com/gontcharovd>

*Data engineer with a background in machine learning, a focus on the industrial sector and a passion for Linux.
My engineering education and experience enable me to understand software requirements on a technological level.*

Education

KU Leuven, Belgium 2016–2018

Master of Materials Engineering *burgerlijk ingenieur* (cum laude)
Industrial engineering exchange at Grenoble INP, France (one year)
Scientific research:

J. Everaerts, D. Gontcharov, B. Verlinden, M. Wevers, *The influence of load holds on the fatigue behaviour of drawn Ti-6Al-4V wires*. International Journal of Fatigue, 98, (2017), pp. 203-211.

KU Leuven, Belgium 2014–2016

Bachelor of Engineering

Experience

Freelance Data Engineer at Gontcharov Data Services BV in Antwerp, Belgium 09.2020 – Present

Development of big data pipelines for a machine learning recommendation engine

Role: data engineer

Sector: pharmaceuticals

Duration: 3 months (current project)

Technology: Apache Spark (PySpark), Python, Kedro, AWS, Argo, Kubernetes

Responsible for data engineering and interaction with business for the France country team on an EMEA-scale project of a multinational pharmaceutical company.

- Developed and maintained Spark data pipelines in Kedro.
- Interacted closely with data science and business teams to meet their data engineering requirements.
- Data profiling before ingestion into the data warehouse.

Development of scalable data pipelines for IoT-data

Role: data engineer

Sector: logistics

Duration: 5 months

Technology: Apache Airflow, Apache Kafka, Faust, Python, MySQL, Cassandra, Docker, ZABBIX, GCP

Laid the foundation of a batch-processing and data-streaming infrastructure for an early-stage startup:

- Deployed a Kafka cluster in Docker with three brokers and Kafka Connect connectors for Cassandra and MQTT.
- Developed a Python stream processing application in Faust (Robinhood) connected to the Kafka cluster.
- Configured ZABBIX on a remote server to monitor the Kafka cluster's JMX-metrics through a Java Gateway.
- Set up the client's infrastructure for Apache Airflow: Docker containers, Python packages for Airflow code, testing and CI/CD.
- Developed two Airflow DAGs with custom operators interfacing with the client's Cassandra production database, MySQL application database and Google Drive folders.

Data analysis and programming workshop for engineers of EIT RawMaterials

Role: instructor

Sector: academia

Duration: 1 month

Technology: R, RStudio Cloud, Zoom

Hosting of a two-day virtual workshop 20 students and young-professionals in the raw materials industry.

Data analysis and programming workshop for KU Leuven PhD-researchers

Role: instructor

Sector: academia

Duration: 1 month

Technology: Python, R, RStudio Cloud, Zoom

Hosting of a two-day virtual workshop for 20 PhD-researchers and professors.

Dashboard for gas-treatment time-series visualization

Role: data analyst

Sector: metallurgical industry

Duration: 3 months

Technology: R, Shiny, cron, VBA

Development of a data pipeline and visualization workflow to monitor the KPI's of three gas-filtering modules:

- Developed and automated R scripts to process a daily batch of raw data from text files.
- Developed a Shiny dashboard that can be customized by a user with no coding experience via an Excel sheet.

Junior Consultant Data Science at STATWORX GmbH in Frankfurt am Main, Germany

03.2020–08.2020

Tesseract OCR fine-tuning

Role: data scientist

Sector: accounting

Duration: 3 months

Technology: Tesseract OCR, Python, Docker

Development of an automated workflow to fine-tune the Tesseract OCR LSTM network to improve optical character recognition on German invoice images.

Machine learning proof of concept for NLP

Role: data scientist

Sector: retail

Duration: 2 months

Technology: TensorFlow (Keras), PyTorch, Python, GCP, Docker

Co-development of a proof of concept to extract and classify accounting information from German invoices:

- Experimented with multiple PyTorch transformer-models for NLP to classify invoice line items into three budget categories.
- Retrained the final model in TensorFlow and deployed it using Docker on Google Cloud Platform as a microservice.

Process Engineer at TRIMET Aluminium SE in Essen, Germany

09.2018–01.2020

Dashboard for metal production visualization

Role: process engineer - data analyst

Sector: metallurgical industry

Duration: 4 months

Technology: R, Shiny, SQL Server

Development of a dashboard to calculate the metal production of 360 units.

Dashboard for voltage and current time-series visualization

Role: process engineer - data analyst

Sector: metallurgical industry

Duration: 4 months

Technology: R, Shiny, SQL Server

Development of a dashboard to monitor electrical data and process control states of production units across two plants.

Anode-changing process control module

Role: process engineer - data analyst

Sector: metallurgical industry

Duration: 6 months

Technology: MATLAB, Simulink, Stateflow, SQL Server, SVN

Development of a software module with MATLAB, Simulink and Stateflow to control the anode-changing operation:

- Documented the software design and usage instructions into a technical specification.
- Wrote unit tests and integration tests and incorporated them into a CI/CD pipeline.
- Performed field tests of the software module on two different plants.

Process Engineer Intern at TRIMET Aluminium SE in Saint-Jean-de-Maurienne, France 02.2018–08.2018

Statistical analysis (design of experiments) of fluor emissions and electrical contacts

Role: data analyst

Sector: metallurgical industry

Duration: 6 months

Technology: R, Shiny, SQL Server, MATLAB

Introduced and established R as the default tool for data analysis at the plant.

Technical skills

- Scripting: Python (base, numpy, pandas), R, Bash
- DevOps & Cloud: Docker, AWS, GCP, Kubernetes
- Linux: Arch, Vim, Tmux
- Processing and orchestration: Airflow, Kafka, Spark
- Databases: SQL Server, PostgreSQL, MySQL, Cassandra, S3
- Machine learning: PyTorch, TensorFlow (Keras)
- CI/CD and Agile: Git, GitHub, GitLab, Jira

Voluntary work

EIT RawMaterials Alumni 08.2018–01.2020

- President of an international team of six board members and one full-time employed Alumni Manager
- Public speaker on the EIT RawMaterials Summit 2019 (500 people) and EIT Connect 2018 and 2019 (120 people)

Personal projects

Data pipeline and dashboard for compressor pressure sensor data

Duration: 3 months

Technology: Python, Apache Airflow, PostgreSQL, R, Shiny, Docker, GCP

Development of a data processing and visualization workflow to monitor pressure sensor data from an industrial compressor:

- Developed a data pipeline in Python and Apache Airflow that writes a daily batch of sensor data to an application database.
- Developed a dashboard showing pressure difference and sensor location.
- Deployed the project in Docker on an E2-instance virtual machine on Google Cloud Platform.

Computer vision competition

Duration: 4 months

Technology: Python, PyTorch

Developed a PyTorch model for the Steel Defect Detection competition on Kaggle whose test prediction finished in the top 23%.

Personal workstation for GPU-computing

Duration: 4 months

Technology: Arch Linux, computer hardware (GPU, CPU, memory, disk, power supply, cooling)

- Researched and assembled all hardware components.
- Installed and configured the Arch Linux operating system.