Stojan **Jovanović, Ph.D.**

UK Tech Nation Exceptional Talent · Full-Stack Data Scientist · ML Engineer · Analytics Engineer

Oxfordshire, United Kingdom

(+44)74 1316 7439 | stojanjo@pm.me | stojanjo.net | stojan211287 | https://www.linkedin.com/in/stojanjovanovic211287/

A professional with 6+ years of experience in machine learning, data science & data engineering, comfortable with every part of the modern analytics pipeline. Experienced data & ML production systems architect. Strong communicator, able to present abstract concepts in an accessible way to all organisational levels and diverse stakeholders . Experienced technical lead. So far, worked professionally in finance, advertising and healthcare.

Skills

Experienced data professional, specializing in analytics engineering, reproducible data analysis and applied machine learning.

Data science and data analysis	SQL, pandas, numpy, tidyverse
Distributed & high-performance analytics	Dask, Ray, Spark, Slurm, Modin, H2O.ai, Hadoop
Machine Learning	Scikit-Learn, Pytorch Lightning, Tensorflow, Keras, Pytorch
Data & Analytics Engineering	dbt, PostgreSQL, Prefect, DVC, MS SQL Server
DevOps & IaC	Docker, Kubernetes, Ansible, Helm, ArgoCD, Terraform
Programming languages	Python, Scala, R, shell (bash), Golang, C, CUDA
Software engineering	git, CI/CD (GitlabCI, CircleCI), Heroku, API development, ELK Stack
Web technologies	FastAPI, Flask, HTML, css/sass, JavaScript
Visualisation and reporting	Seaborn, Matplotlib, Streamlit, Shiny, Plotly Dash, ggplot2, lightdash
Cloud platforms	Microsoft Azure & AzureML, Google Cloud Platfrom

Relevant Experience

Sensyne Health plc — *Oxford, UK*

Lead Data Scientist — *May 2020 - present*

- **Technical lead for all ML, Data & Analytics Engineering, as well as DataOps workflows. Overseeing multiple projects simultaneously, while consulting with teams of data scientists, ML researchers, engineers, SysAdmins and project managers.**
- Led migration of MLOps processes & ELT pipelines from on-prem infrastructure to the Azure cloud platform. Co-led development of the company's data & analytics platform in the cloud, built on top of Github, CircleCI, Kubernetes, ArgoCD, PostgreSQL and AzureML.
- Helped grow the ML Engineering and Data teams by conducting 200+ pair-programming and system design interviews for the Data Engineer, ML Engineer and Data Scientist positions.
- Built an automated, distributed computing system for concurrently running 200+ end-to-end machine learning experiments, with simultaneous GPU & CPU workflows, data version control & extensive experiment result logging.
- Spearheaded deployment & operationalization of HPC facilities (9 node hybrid CPU/GPU system, WekaFS, InfiniBand) in the company's AI lab. Designed the HPC analytics & containerization software stacks. Served as the system's interim SysAdmin.
- Led a team of 4 data scientists to find clinical predictors of Covid-19 disease severity by analysing real-world EHR data, shared by an NHS partner trust during an ongoing pandemic. The research resulted in 2 publications.
- Drove the adoption of tools and processes to improve our data models reproducibility, provenance (lineage), and sustainability.
- **Tech stack: dbt / PostgreSQL / Kubernetes / Helm / ArgoCD / Dask / Ray / AzureML / Slurm / DVC / Enroot**

Sensyne Health plc — *Oxford, UK*

Machine Learning Architect — *Jan 2020 - May 2020*

- **Sole systems architect for the company's new AI lab. Developed tooling, codes of conduct & best practices for reproducible and rapid analytics, working with 4 other team leads.**
- Led development of a distributed ELT pipeline system, ingesting raw data and producing standardised & analytics-ready datasets from two disparate NHS trust sources for use by DL research teams. Increased team efficency by making it easy to re-use data transformation across multiple ML experiments.
- Designed & deployed central DB services, storing sensitive patient data from NHS trust partners, as well as AI lab's remote work infrastructure at Covid19 pandemic onset.
- Further developed tools and infrastructure to scale, support and improve the data & ML teams' tasks, from data transformation, through model development to distributed model training.
- Developed internal policies regarding data and ML best practices, remote working and analytics lab resource use.
- **Tech stack: Modin / Pandas / PostgreSQL / Python / TensorFlow / Dask / R / Parquet**

Sensyne Health plc — *Oxford, UK*

Senior Machine Learning Researcher — *Apr 2019 - Jan 2020*

- **One of 6 initial hires to the company's R&D team, was reponsible for setting up DataOps, data/ML engineering processes, infrastructure & tooling as we scaled out.**
- Led design and implementation of complex, end-to-end ML pipeline enabling delivery on a commercial project worth 5 million pounds. Coordinated collaboration of 4 teams of 30 data scientists.
- Built the company's AI lab infrastructure from scratch - from one shared server to a secure network of 50+ GPU-enabled workstations with air-gapped central GitLab services, private Docker registry, private *nix package repositories & secure remote storage
- Designed & wrote a modular ML pipeline system API with the purpose of enabling researchers to quickly prototype new AI model architectures. The API in question is still used by ML research teams.
- Developed & patented ML methodology for unsupervised stratification of patients based on structured EHR data, obtained through the company's NHS trust partners.
- **Tech stack: Python / Keras / Pandas / Dask / Nvidia-Docker / Prefect / Ansible / GCP**

Adverai ltd *Stockholm, SE*

Machine Learning Engineer *May 2018 - Mar 2019*

- **Sole ML engineering team member, responsible for introduction of ML capabilities into the company's core software products.**
- Designed and engineered Adverai's entire ML microservice stack, including model serving infrastructure, HTTP/gRPC model prediction APIs, as well as a CI/CD system for automated ML model training & deployment.
- Introduced online time-series prediction and anomaly detection functionality to the company's core software product.
- Mentored two junior data scientists in ML model deployment automation & large-scale data processing.
- **Tech stack: Pytho / TensorFlow (Serving) / gRPC / Docker / GitLab / PostgreSQL / GraphQL**

Qliro AB *Stockholm, SE*

Data Scientist *Apr 2017 - May 2018*

- **Data analytics team member, responsible for novel credit risk model development, reporting process automation & large-scale feature engineering.**
- Designed and built a system for scheduled re-training and scoring of credit risk models on top of Hadoop, Spark and H2O.ai, automating and significantly speeding up manual credit reserve reporting processes.
- Helped migrate ETL systems from RDBMS to the Spark / Hadoop ecosystem, as data volumes become unwieldy for processing inside our data warehouse.
- Worked closely with the data engineering team to operationalize continuous deployment pipelines for proprietary credit risk models, cutting average ML model deploy time to 1 week, from 2 months.
- **Tech stack: R / ELK stack / Scikit-Learn / Scala / Spark / H2O.ai / SQL Server**

Kavli Institute for Systems Neuroscience *Trondheim, NO*

Data Scientist *Apr 2016 - Apr 2017*

- **Sole data scientist in a biomedical experimental lab. Consulted on all matters of computational science & engineering, working with various teams of researchers.**
- Developed a novel, latent-variable RNN network architecture to predict time series of electrical activity of rat neurons. Publication still being prepared.
- **Tech stack: Torque / TensorFlow / Pandas / Scikit-Learn / Scikit-Image / Jupyter**

Education

KTH Royal Insitute of Technology & Albert-Ludwigs University *Stockholm, SE & Freiburg, DE*

PhD in Computer Science & Computational Neuroscience *Oct 2012 - Mar 2016*

Thesis: Correlations of Higher Order in Networks of Spiking Neurons. Supervised by Prof. Stefan Rotter & Prof. John Hertz
Joint degree, obtained as part of the Erasmus Mundus Joint PhD Programme

- Developed algorithms and analytics pipelines to simulate the behaviour of networks of neurons in the cortex, in Python, Cython and C. These pipelines simulated 100,000 interacting cortical neurons for days at a time, producing data that the subsequent analytics pipeline would analyse. Both pipelines were deployed on state-of-the-art supercomputers.
- Obtained a novel mathematical result, enabling prediction of future cortical activity from knowledge of functional connectivity of neurons in the cortex.

University of Belgrade *Belgrade, RS*

MSc in Applied Mathematics & Operations Research *Oct 2010 - Sep 2011*

- Principle subjects studied: queueing theory, graph theory, stochastic calculus, financial engineering

University of Belgrade *Belgrade, RS*

BSc in Statistics, Actuarial & Financial Mathematics *Oct 2006 - Sep 2010*

- Principal subjects studied: multivariate calculus, statistics, linear algebra, actuarial science, optimization & differential equations

Talks and workshops

Jan 2017 **Institute for Advanced Study**, The Hawkes Process as a Model of Cortical Networks *Princeton, NJ*

Aug 2016 **Los Alamos National Laboratory**, Temporal Evolution of Grid Cell Learning *Los Alamos, NM*