Stojan **Jovanović, Ph.D.**

UK Tech Nation Exceptional Talent · Machine Learning Engineer · Analytics Engineer · Data Scientist

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 & analytics engineering, proficient at every layer of the modern data stack. Comfortable with both hands-on work, as well as leading a team of data professionals. 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 applied machine learning and analytics engineering.

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

Relevant Experience

Sensyne Health plc — *Oxford, UK*

Head of Machine Learning Engineering — *November 2021 - present*

- **Technical strategy role, focusing on ML and analytics engineering . Responsible for developing & maintaining the company's real-world-evidence data product. Leading a small team of ML and data engineers.**
- Led development of the company's real-world evidence clinical data platform, built on top of dbt, PostgreSQL and Kubernetes.
- Helped grow the ML Engineering and Data teams by designing & conducting 200+ pair-programming and system design interviews for the Data Engineer, ML Engineer and Data Scientist positions.
- Drove the adoption of tools and processes to improve our real-world evidence (RWE) data models reproducibility, provenance (lineage), and sustainability.
- **Tech stack: dbt / PostgreSQL / Kubernetes / Helm / ArgoCD / SQLAlchemy**

Sensyne Health plc — *Oxford, UK*

Lead Data Scientist — *May 2020 - November 2021*

- **Tech lead for all ML- & Analytics Engineering workstreams. Responsible for data product architecture across multiple projects simultaneously, working closely with teams of data scientists, ML engineers, software engineers and product managers.**
- Led technical development of the company's real-world evidence (RWE) data analytics SaaS product, throughout its full life cycle; from ideation & UX design, through data modelling & backend API development, to scalable cloud deployment.
- Developed a distributed ELT pipeline system, ingesting raw real-world electronic health record (EHR) data and producing standardised & analytics-ready datasets from multiple disparate NHS trust sources for use by internal research and engineering teams.
- Designed & implementated a distributed, end-to-end ML pipeline, running 400+ ML experiments a day across multiple CPU/GPU nodes. The pipeline ensured delivery on a commercial project worth 5 million pounds and orchestrated collaboration of 30 data scientists.
- **Tech stack: dbt / PostgreSQL / Kubernetes / Dask / Ray / Helm / ArgoCD / Slurm / DVC**

Sensyne Health plc — *Oxford, UK*

Senior Data Scientist — *April 2019 - May 2020*

- **Sole systems architect for the company's on-premise AI & data lab. Developed tooling, codes of conduct & best practices for reproducible and rapid data analytics, working with 4 R&D team leads.**
- 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
- Deployed & operationalized HPC facilities (9 node hybrid CPU/GPU system) in the company's AI lab. Served as the AI lab's SysAdmin.
- Designed a high-performance, conternized GPU-bound ML model training system, using Nvidia's Enroot technology. This system abstracted ML model training on complex hardware for our researchers & increased usage efficiency of our GPUs.
- **Tech stack: Modin / Dask / Pandas / PostgreSQL / Python / tidyverse / Ansible / Enroot**

Adverai ltd — *Stockholm, SE*

Machine Learning Engineer — *May 2018 - Mar 2019*

- **Sole ML engineering team member, responsible for introduction of AI 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: Python / 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 Institute 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*

Publications

2021	**Early risk assessment for COVID-19 patients from emergency department data using machine learning**, FS Heldt, MP Vizcaychipi, S Peacock, M Cinelli, L McLachlan, F Andreotti, S Jovanović, R Dürichen, N Lipunova, RA Fletcher, A Hancock, A McCarthy, RA Pointon, A Brown, J Eaton, R Liddi, L Mackillop, L Tarassenko, RT Khan	*Nature Sci Rep*
2020	**Deep Semi-Supervised Embedded Clustering (DSEC) for Stratification of Heart Failure**, O Carr, S Jovanovic, L Albergante, F Andreotti, R Dürichen, N Lipunova, J Baxter, RT Khan, B Irving	*arXiv preprint*
2020	**Prediction of the onset of cardiovascular diseases from electronic health records using multi-task gated recurrent units**, F Andreotti, FS Heldt, B Abu-Jamous, M Li, A Javer, O Carr, S Jovanovic, N Lipunova, B Irving, RT Khan, R Dürichen	*arXiv preprint*
2020	**Risk factors for clinical progression in patients with COVID-19: a retrospective study of electronic health record data in the United Kingdom**, RA Fletcher, T Matcham, M Tibúrcio, A Anisimovich, S Jovanović, L Albergante, N Lipunova, A Hancock, L Mackillop, L Tarassenko, A McCarthy, MP Vizcaychipi, RT Khan	*medRxiv preprint*
2016	**Interplay between graph topology and correlations of third order in spiking neuronal networks**, S Jovanović, S Rotter	*PLoS Comp Bio*
2015	**Cumulants of Hawkes point processes**, S Jovanović, J Hertz, S Rotter	*Physical Review E*