

Stojan Jovanović, Ph.D.

UK TECH NATION EXCEPTIONAL TALENT · DATA SCIENTIST · ANALYTICS ENGINEER

Oxfordshire, United Kingdom

☐ Upon Request | ✉ stojanjo@pm.me | 🌐 stojanjo.net | 📠 stojan211287 | 🌐 <https://www.linkedin.com/in/stojanjovanovic211287/>

A professional with 6+ years of experience in data science & 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 team lead. So far, worked professionally in finance, advertising and healthcare.

Skills

Experienced data professional, specializing in applied data science and analytics engineering.

Data & Analytics Engineering	dbt, PostgreSQL, RedPanda, DVC, MS SQL Server
Data science and data analysis	SQL, pandas, numpy, tidyverse, SQLAlchemy
Distributed & high-performance analytics	Dask, Ray, Spark, SLURM, Modin
Visualisation and reporting	Seaborn, Matplotlib, Streamlit, ggplot2
DevOps & InfrastructureAsCode	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

- **Senior team leader, focusing on data and analytics engineering. Responsible for the company's data warehouse & DataOps lifecycle. Leading a small team of data and analytics engineers.**
- Led development of a serverless data ingestion & standardisation pipeline, cutting data-related cloud costs by 60%.
- Helped grow the ML Engineering and Data teams by designing & conducting 200+ pair-programming and system design interviews for the Data Engineer and ML Engineer positions.
- Drove the adoption of tools and processes to improve reproducibility, sustainability and ensure provenance of our data artefacts.
- **Tech stack: dbt / PostgreSQL / Kubernetes / Helm / ArgoCD**

Sensyne Health plc

Oxford, UK

LEAD DATA SCIENTIST

May 2020 - November 2021

- **Team leader & tech lead for all Data & 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.
- Developed a distributed ELT system, ingesting real-world electronic health records (EHR) and producing analytics-ready datasets, built from multiple disparate NHS trust sources for use by internal analysts and engineering teams.
- Designed & implemented an end-to-end ML pipeline, running 400+ ML experiments a day across multiple CPU/GPU nodes, ensuring delivery on a commercial project worth 5 million pounds.
- **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 R&D lab. Developed infrastructure, tooling, codes of conduct & best practices for reproducible and rapid data analytics, working with 4 R&D team leads.**
- Built the company's R&D lab from scratch - air-gapped network of 50+ GPU-enabled workstations with central GitLab services, private Docker registry, private *nix package repositories & secure remote storage
- Deployed & operationalized on-premise high-performance compute facilities (9 node hybrid CPU/GPU system).
- Developed a containerized, GPU-bound ML model training system, using Nvidia's Enroot technology, abstracting distributed ML model training & increasing 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

DATA SCIENTIST

Stockholm, SE

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 ecosystem, as data volumes made processing inside our data warehouse untenable.
- 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

DATA SCIENTIST

Trondheim, NO

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

MSC IN APPLIED MATHEMATICS & OPERATIONS RESEARCH

Belgrade, RS

Oct 2010 - Sep 2011

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

University of Belgrade

BSC IN STATISTICS, ACTUARIAL & FINANCIAL MATHEMATICS

Belgrade, RS

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