Jovan Veljanoski

Data Science Lead

Netherlands

Contact: via LinkedIn

in/jovanvel/

JovanVeljanoski

jovan31

Jovan31

jovan.veljanoski

AI & Data Science Lead with over 11 years of experience solving complex problems with data-driven approaches. Currently leading a data science team in the e-commerce domain. Combines hands-on technical work with team management, operational process design and advice to senior management on AI/ML strategy. Previously a machine learning specialist at a Google Cloud Premier Partner. Created the *DevOps Prediction Engine* . Former core member of the statistical data validation team for the European Space Agency *Gaia* mission.

Experience

2023.02 – present
Amsterdam, Netherlands

Lead Data Scientist

Tiqets.com

- Directs and coordinates the AI and machine learning efforts in the company, advising the COO and Head of Product on strategy, feasibility, priorities, staffing and delivery.
- Manages a team of three data scientists (two senior and one mid-level), including recruitment, project allocation, performance reviews, regular feedback and professional development.
- Leads the technical delivery of an AI-assisted product onboarding and maintenance system, responsible for its core processing logic. It converts product data from various partner APIs into fully configured products within the Tiqets schema in minutes, with only manual approval required before publication. It has already been used to publish hundreds of products.
- Coordinates work across Operations, Editorial, Product and Engineering teams to adapt existing manual practices into AI-assisted processes.
- Established a company-wide testing and quality process for the AI components, including human-reviewed datasets, model and prompt evaluation, monitoring, quality thresholds and a structured review workflow.
- Extended the product ranking system to work with arbitrary entities like product groups and packages.
- Built and deployed a recommender system that suggests *alternative* products when a desired product is temporarily unavailable.
- Established a process for bulk creation and curation of product copy via large language models, ensuring consistent style in line with Tiqets' editorial guidelines. The process was used to edit and publish copy for more than 2500 products.
- Created a set of data applications that aggregate and summarize customer feedback from a variety of sources (public reviews, customer support interactions), allowing operational teams to quickly identify and address customer pain points.

Tech stack: Python - dspy, numpy, openai, pandas, scipy, streamlit; Airflow; Databricks; Docker, Langfuse, LiteLLM, SQL.

2021.11 – 2023.02
Amsterdam, Netherlands

Senior Data Scientist

Tiqets.com

- Improved the multi-armed-bandit product-ranking system, leading to a gross-margin increase of more than 10%, via changes to the optimization objective and feature set.

- Developed a company-wide platform for evaluating A/B tests and an internal Bayesian significance-testing library used to support product decisions.
- Created monitoring and explainability tools for the production product ranking system.

Tech stack: Python - mlflow, numpy, pandas, pymc, scikit-learn, scipy, streamlit; Airflow; Databricks; Docker; SQL.

2020.06 – 2021.10
Utrecht, Netherlands

Senior Data Scientist / Machine Learning Specialist
Cloud Technology Solutions

- Developed and deployed a computer vision system to classify over 100 factory components. Extracted images from videos, and performed data preprocessing, augmentation and quality control. Worked with the client to refine the use case, gather data and validate the results.
- Built and deployed a system that optimised greenhouse climate settings to balance production and operating cost according to the client's chosen objective, including data-source consolidation, backend development, and data and model validation.

Tech stack: GCP - AI Platform, App Engine, AutoML, BigQuery, Cloud Functions, Vision Edge; Airflow; Python - numpy, openCV, optuna, pandas, PIL, streamlit.

2018.09 – 2020.05
Hilversum, Netherlands

Senior Data Scientist
XebiaLabs (now Digital.ai)

- Created the *DevOps Prediction Engine*, a plug-in for the XebiaLabs release orchestration platform, which provides explainable predictions regarding the outcome of a DevOps release pipeline. Engineered meaningful features and developed custom transformers to clean and encode the data, together with a custom domain-informed estimator based on the k-nearest-neighbours algorithm.
- Wrote and filed a patent application for the DevOps Prediction Engine.
- Built customer-facing analytics web application prototypes, gathered feedback from customers and worked with the strategy team on the roadmap for future product features.

Tech stack: Python - flask, ipyvuetify, ipywidgets, pandas, plotly, scikit-learn, seaborn, vaex, voila; Docker.

2017.09 – 2023.01
Groningen, Netherlands

Co-founder & core contributor (part-time)
vaex.io

- Co-founded Vaex, an open-source Python library for working with datasets larger than available memory. Led the product roadmap and API design.
- Contributed to feature development and testing.
- Wrote articles and presented demos at companies and conferences.

2014.09 – 2018.09
Groningen, Netherlands

Post-doctoral Research Assistant
Kapteyn Astronomical Institute, University of Groningen

- Core member of the ESA Gaia statistical validation team, developing tests for large datasets (more than 1.6 billion samples and 100 features) to detect patterns, outliers and inconsistencies. Wrote qualitative and quantitative reports, and presented findings to an international group of stakeholders.
- Led a team of 4 scientists in developing a pipeline for detecting extremely rare events in large, imbalanced datasets. It achieved a precision of at least 90% on unseen data.
- Developed a pipeline that queries, cleans and reformats star catalogue data, and runs a series of clustering algorithms to detect and isolate patterns of interest. With this method I discovered nine new structures in the Milky Way.

Tech stack: Python - emcee, matplotlib, numpy, pandas, scikit-image, scikit-learn, scipy, vaex, xgboost; SQL; LaTeX.

Independent projects

MyLexilingo

MyLexilingo.com is a language learning platform that I designed and built completely independently.

- Facilitates the learning and practice of over 70 languages via flashcards, listening, reading and speaking exercises.
- Empowers language learners and teachers to create and curate their personalized study material.
- Enables language teachers to assign interactive homework to their students.
- Leverages AI models for dynamic content generation at the appropriate level for each student.

Education

2010.09 – 2014.05
Edinburgh, UK

Ph.D. in Astrophysics

University of Edinburgh

- Developed a pipeline to clean and process noisy image data. Implemented a Python algorithm to extract key measurements from the image data. Fitted a Bayesian model to the measurements using the Markov chain Monte Carlo framework.
- Learned how to coordinate work within a large international collaboration and communicate with different teams exchanging data products and results.
- Developed strong project management skills by structuring the work into a set of short-term (weekly or monthly checkpoints in software development, research or data analysis), medium-term (completion of publications) and long-term goals (completion of the thesis).
- Developed the ability to work independently on challenging problems over extended periods and to prioritise tasks across multiple parallel projects.
- Produced high-impact publications. Won the Robert Cormack Research Prize awarded by the Royal Society of Edinburgh for the best peer-reviewed paper of 2014/15 with a postgraduate researcher as lead author.

2006.09 – 2010.05
Edinburgh, UK

Bachelor of Physics with Honours in Astrophysics

University of Edinburgh

- First Class Honours.

Skills

AI & Data Science

Generative AI and LLM evaluation, recommender and ranking systems, experimentation, machine learning, cluster analysis, Bayesian and frequentist statistics

Development & Delivery

Python, SQL, Databricks, Airflow, Docker, AWS, GCP, MLflow, production monitoring

Leadership & Collaboration

AI and machine-learning strategy, project scoping, feasibility assessment and prioritization, stakeholder management, team management, recruitment and executive communication

Communication and outreach

2017.12 – 2018.03

Data Science competition

Natural Language Processing

Led a team of 4 data scientists competing in a natural language processing (multi-label classification) challenge organized by Jigsaw (a Google affiliate) and hosted on Kaggle.com. My team ranked in the top 5% of 4551 teams competing worldwide by stacking a combination of convolutional, recurrent and attention-based neural networks following careful preprocessing and encoding of the data with several word-vector models.

Public speaking

Presented modern data science techniques at several international conferences *(PyData Global 2022*, *GOTO Copenhagen 2021*, *EuroSciPy 2019*, *PyData London 2019*, *EuroPython 2019*, *PyGrunn 2019*, *PyParis 2018)*

Publications

- Published 32 peer-reviewed scientific papers, including 7 as a lead author, with more than 1500 citations in total.
- Published a number of data science blog posts describing modern data science techniques.

Languages

English
Full professional proficiency

Dutch
B1+ level

Macedonian
Native