Vladislav Konjushenko

29 Nov 2003 | Estonian | Copenhagen, Denmark

vlad.konjushenko@gmail.com | +4552671606 | LinkedIn | My Website Portfolio

Education

Aug 2025 – Jun 2027
Copenhagen, Denmark

IT University of Copenhagen, *Master of Data Science*

Planned curriculum: Algorithm Design, Advanced Applied Statistics, Data Wrangling & Visualization, Advanced Machine Learning, Algorithmic Fairness, Data Science in Production & Master's Thesis.

Aug 2022 – Jun 2025
Copenhagen, Denmark

IT University of Copenhagen, *Bachelor of Data Science (Graduated)*

Core courses: Machine Learning, NLP, Network Analysis, Applied Statistics, Data Visualization, Algorithms & Data Structures, Software Engineering, Security & Privacy, Business Foundations.

Professional Experience

Jul 2023 – Oct 2023
Tallinn, Estonia

IT Technician, *Haitek Group*

- Coordinated with project managers and negotiated with schools to schedule IT equipment installations.
- Installed and configured computers in schools across Tallinn for a major Atea AS project.
- Set up BIOS, Windows OS and software for peripherals.

Tallinn, Estonia

Web Developer

- Designed and developed four fully functional websites using WordPress and Tilda.
- Customized themes and plugins to meet client requirements.
- Implemented responsive layouts and front-end customization.
- Built an interactive online calculator using HTML, CSS, and JavaScript.
- Collaborated with clients to gather requirements and deliver solutions.

Skills

Data Science & Programming

- Programming Languages: Python, R, Go
- Data Analysis: pandas, NumPy, Jupyter Notebook, Google Colab
- Data Visualization: Matplotlib
- Machine Learning: scikit-learn, model evaluation
- Deep Learning: PyTorch, TensorFlow

Data Engineering & Cloud

- Big Data Processing: Apache Spark (PySpark)
- Databases & Warehousing: PostgreSQL, MS SQL Server
- Cloud Platforms: Microsoft Azure, Snowflake, AWS fundamentals
- ETL & Data Pipelines: Azure Data Factory, Databricks, Data Lake Storage
- Containerization & Deployment: Docker
- CI/CD & Automation: GitHub Actions

Data Visualization & BI

- Dashboards: Tableau, Power BI
- Network Visualization: Gephi
- Data Storytelling & visual analytics

Tools & Professional Practices

- Version Control: Git, GitHub, DVC
- Documentation: LaTeX (Overleaf)
- Office & Collaboration: Microsoft Office, Google Workspace
- Project Management: Jira

Systems & IT Foundations

- Operating Systems: Windows, iOS, Linux
- Mobile Support: iOS, Android
- Environment setup and troubleshooting

Web Development

- Front-End: HTML, CSS, JavaScript
- Back-End: Node.js, Express.js
- APIs & Authentication: REST APIs, JWT
- Security Fundamentals: password hashing (bcrypt), validation
- Databases: MongoDB
- Tools & Platforms: Postman, WordPress, Tilda, Figma

Languages

English: *C1* | **Danish:** *A2* | **Estonian:** *B2* | **Russian:** *Native* | **Ukranian:** *B1*

Bachelor Project

Developing Robust Neural Network Models for Network Intrusion Detection

Developed and evaluated neural network–based intrusion detection systems for detecting DDoS attacks in large-scale network traffic. The project compared a baseline autoencoder-based model with an enhanced architecture incorporating differentiable logic constraints in the training loss. Vehicle was used to define and enforce Differential Logic (DL) constraints during model training, and Marabou was applied as an SMT-based verifier to analyze model robustness and reduce false positives, while maintaining strong detection performance on highly imbalanced network data.

Data Science Projects

Public Transport & CO_2 Emissions

Analyzed public transport systems in Denmark, Estonia, Romania, and Germany to identify factors affecting usage and CO_2 emissions. Using interviews and online reviews, the project highlighted common challenges and proposed country-specific improvements focused on electric transport and infrastructure development to reduce emissions.

Clothing Classification Using Machine Learning Models

Built and compared models to classify clothing types from images using CNN, SVM, and Naive Bayes. Focused on data preprocessing, feature extraction, and hyperparameter tuning to improve accuracy. Worked collaboratively to analyze results.

Analysis of a Power Grid Network Robustness

This project focused on assessing the robustness of a power grid network under targeted attacks, random failures, and cascading failures. Conducted structured analyses to identify potential vulnerabilities and points of failure, and used Gephi to visualize network structure and resilience across different scenarios.

Data Visualization - Soviet Casualties in WWII

The project focused on designing a mountain-shaped timeline visualization of Soviet military losses during key WWII battles. Using Tableau, we developed an interactive dashboard that enabled viewers to quickly understand both the scale and timing of significant casualties, combining structured historical data analysis with clear and impactful visual storytelling.

Large-Scale Data Analysis and Machine Learning Lifecycle

Developed scalable data analysis pipelines using Apache Spark to process large datasets. Analyzed reviews and language patterns to study customer sentiment and business performance. Built machine learning models for rating prediction and wind power forecasting, applied feature engineering, and evaluated models using Accuracy and F1 score. Ensured reproducibility through MLflow, version control, and documented workflows.

Fine-Tuning BERT for Domain-Specific Named Entity Recognition

Our team fine-tuned pre-trained BERT model for Named Entity Recognition (NER) using data from J.R.R. Tolkien's Lord of the Rings universe. By adapting BERT with specialized datasets, we aimed to improve its accuracy in identifying domain-specific entities like characters, locations, and organizations. Training on both English Web Treebank and Lord of the Rings datasets led to significant performance improvements, demonstrating the value of incorporating domain-specific data.

Modular Data Science Pipeline with DVC & Dagger

Refactored a monolithic Python workflow into a modular, reproducible pipeline. Used DVC for data version control and Dagger (in Go) to orchestrate automated steps for data preprocessing, model training, and model selection. Integrated the pipeline with GitHub Actions for CI/CD, ensuring every run was traceable, consistent, and documented.

Energy Demand Forecasting and Surplus Analysis Across Nordic Regions

Analyzed short-term energy demand and supply conditions using historical electricity consumption, generation, weather, and calendar data collected via public APIs. The analysis focused on surplus periods and negative day-ahead price events, examining how renewable output, demand levels, and weather conditions drive price behavior across regions.

Reproducibility Study

Conducted a reproducibility study of the paper *"The language of opinion change on social media under the lens of communicative action"*. The project focused on verifying whether the original statistical findings and conclusions could be reproduced using the authors' released datasets and code. The study evaluated computational reproducibility, methodological clarity, and robustness of reported results.

Statistical Analysis of Housing Prices with Regularized Regression

Conducted a comprehensive statistical analysis of residential housing prices using the Ames Housing dataset. The project focused on identifying the most influential predictors of house prices and evaluating the stability and generalization of linear regression models in a high-dimensional setting. Regularization techniques were applied to address multicollinearity introduced by extensive categorical features.