



Radosław Szostak

Data Scientist



radekszostak.github.io
radoslawszostak@gmail.com

+48 794 625 183
Krakow, Poland

Overview

Certified data scientist and published researcher skilled in Python. Four years of professional experience in academic research specializing in the analysis and modeling of various types of data. Author of numerous software solutions addressing complex problems. Experience in deployment to production environments. Oriented towards sharing knowledge and problem-solving.

Experience

Personal projects

Portfolio assistant. Q&A chatbot using vector store knowledge base and keeping chat history context.

Prosumer prediction. Forecasting of energy production and consumption for photovoltaics prosumers.

Doctoral student Oct 2019 – Present
AGH University of Krakow (4 yrs)

Using Deep Learning and drones for estimation of river water level. An unconventional regression approach utilizing CNNs. Featured at prestigious AI conference NeurIPS 2021.

Calibration of drone thermal images using gradient descent. A custom application of the PyTorch library, in which many small models were optimized simultaneously to align positions and values of aerial thermal images.

Time series forecasting of groundwater levels. An approach that integrates autoregression with the incorporation of exogenous satellite data for predicting groundwater levels.

Semantic segmentation of rivers on satellite imagery. Standard segmentation using CNNs. Semiautomated labelling using OpenCV. Developed in two versions using PyTorch and TensorFlow.

Geospatial segmentation based on user prompt. GUI tool for semi-supervised aerial image clustering based on small amount of user prompted data.

Interpersonal skills proven as a deputy chairman of the departmental doctoral student government, university teacher and member of an international research group.

Tutoring Sep 2020 – Jun 2023
Self-employment (2 yrs 10 mos)

Development of a **web application** with an assignment database, which was utilized for online physics tutoring. The application successfully operated on the **production server**.

Industrial machines software developer Aug 2016 – May 2019
Averna (2 yrs 10 mos)

Experience in **commercial** projects, focusing on the development of reliable software solutions for production lines in automotive companies such as BMW and Tesla. Acquired knowledge in SQL, DBMS, and software architectures.

Skills

General

• deep learning • classical machine learning • large language models • convolutional neural networks • data cleaning • exploratory data analysis • feature engineering • data visualization • cross validation • computer vision • statistical analysis • hypothesis testing • gradient boosting algorithms • Bayesian optimization • hyperparameter search • multithreading • big data • cloud computing • object oriented programming • databases

Python

• Pandas / Polars • NumPy / Xarray • Scikit • PyTorch / TensorFlow • Langchain • StatsModels • OpenCV • Matplotlib / Seaborn • Plotly / Dash • Streamlit • GeoPandas / rasterio / GDAL • Django

Server

• Linux • Docker • Slurm • AWS • Postgres • MS SQL Server

Certificates

Data Science Professional Certificate
IBM

Aug 2023

Extensive certification program including **CRISP-DM** methodology, stakeholder **communication**, **Python** and Jupyter based software tools, **reporting** standards, interactive **dashboard** development, **data mining** and **machine learning** methods for unsupervised clustering and supervised classification and regression.

Fundamentals of Accelerated Computing with CUDA C/C++
NVIDIA

Jun 2022

Low level GPU programming in C++.

Education

Technical Physics
AGH University of Kraków

2013 – 2018
(5 yrs)

Engineering and Master's degree including courses in **statistics**, **mathematics**, several **programming** languages (Python, C, C++, Java), **databases** and **English**.

Interests

• Kaggle competitions • cooking • mathematics / physics • ecology • psychology • economy • rock music