Jakub Wisniewski

+48 535 304 552 | jakwisn@gmail.com | linkedin.com/in/jakwisn | github.com/jakwisn | jakwisn.github.io

Data Scientist with experience in machine learning, data analysis, and AI-driven solutions. Proven expertise in developing machine learning systems, deploying models to production, and delivering actionable insights through advanced analytics. Skilled in collaborating with cross-functional teams to design and implement impactful AI solutions. Pursuing a Master's in AI while successfully delivering data science solutions at SEB.

Technical Skills and Languages

Programming: Python, R, SQL, JavaScript, HTML/CSS
Frameworks and Libraries: Numpy, Pandas, PyTorch, Tensorflow, FastAPI, Plotly, Scikit-learn, Langchain, OpenCV, Mlflow, Weights&Biases, Dash, Selenium, Dalex (developer), fairmodels (lead developer)
Cloud Platforms: GCP (Cloud Run, Cloud Storage Buckets, BigQuery, Artifact Registry, Vertex AI)
Tools and Platforms: Docker, Git, DVC, Linux, HPC, Slurm, Tableau, PowerBI
Databases: MySQL, MongoDB, Neo4j, SQLite
Languages: English (Proficient C2), Polish (Native)

Experience

Data Scientist Feb. 2022 – Present
SEB *Copenhagen, Denmark*

- Developed internal Python package for processing payment data, integrating SQL and BigQuery, Pandas, Numpy, and Scikit-learn for processing, and Neo4j Database.
- Built containerized applications using Docker and FastAPI, hosted via GCP Cloud Run.
- Applied advanced ML techniques (e.g., autoencoders, XGBoost, RandomForest) for anomaly detection and multiclass classification.
- Coded a LLM Chain-of-Thought workflow enabling automated financial report section summary. Demonstrated the value of Generative AI in automating financial advisory processes within the team.
- Trained a Prophet time-series forecasting model, tracked experiments using MLflow, and integrated resulting models into the website via CI/CD pipelines.
- Automated data scraping using Selenium and BeautifulSoup, uploading results to SharePoint. Cutting down data acquisition time by 90%.
- Created an automated client project template using Cookiecutter that streamlined workflows and unified distributed tools which cut down the time for project delivery.
- Created interactive data visualizations (Plotly, Highcharts, Tableau) for key insights.
- Built a MongoDB database with over 7M documents for financial report analysis, accessible via a FastAPI backend. Data was tracked with DVC on every stage of processing pipeline.
- Conducted client presentations in front of treasury teams focusing on payment structure and inefficiencies using network exploration tools developed by our team.
- **Key Business Achievements:**
  - Established new and unique client offerings, including treasury-focused payment analysis through a self-hosted payment network exploration tool, interactive liquidity analysis, and debt review via custom dashboards. The tools developed by the team and me were used in over 20 client presentations.
  - Collaborated with cross-functional teams to integrate new data sources on GCP, enhancing data accessibility and analysis capabilities.
  - Drove the adoption of Generative AI by integrating the ChatGPT API for automated financial analysis, streamlining reporting, analytical processes, and financial modeling.

Research Software Engineer March. 2020 – Feb 2022
MI2 DataLab *Warsaw, Poland*

- Developed and maintained open-source Python and R packages for Explainable AI (XAI) and fairness analysis (e.g., fairmodels, DALEX).
- Created and managed a large-scale (500GB) DICOM image database, ensuring accessibility for researchers and radiologists.
- Trained deep learning models using PyTorch on a multi-GPU Slurm cluster.
- Designed a custom text annotation tool in Dash for efficient collaboration with radiologists.
- Published scientific articles, blogs, and designed landing pages for the software.
- Conducted presentations (a regular talk and workshops on useR! 2021) and presented team's work on poster session in CVPR workshop.
- **Key Role Achievements:**
  - Created model agnostic tool for measuring fairness metrics in Machine Learning models. R package gathered 80+ stars on GitHub. Later I integrated it into a Python Dalex package (over 1.4K stars). I published the work in a peer-reviewed journal.

    - Collaborated with a cross-disciplinary team of engineers, researchers, and radiologists while creating tools, training data databases, and models on a cluster. Our annotation tool speed-up time spent on annotating proprietary hospital data multiple times.
    - Created and managed a blog enabling the whole team to share their insights. To this day it gathered almost 400 followers.

Education

Technical University of Denmark (DTU) Denmark, Copenhagen
Master in Artificial Intelligence *Feb. 2023 - ongoing (estimated end Feb. 2025)*

- Thesis: Navigational support in obstetric ultrasound using AI
    - Collaboration with Pioneer Center for Artificial Intelligence and Copenhagen Academy for Medical Education and Simulation.
    - Tools used: Python, Docker, WebSockets, Dash, PyTorch, OpenCV

Warsaw University of Technology Poland, Warsaw
Bachelor of Science in Data Science *Oct. 2018 – Feb. 2022*

- Thesis: AI regulations database with the analytical user interface module
    - Collaboration with MI2 DataLab under the guidance of Przemysław Biecek.
    - Tools used: Python, PyTorch, hyper-optimizing BERT model for definition extraction, Huggingface, Docker.

Leadership and other activities

- President of Data Science Club - Led AI-based projects in archaeological scans, showcasing project management and cross-disciplinary collaboration skills. Took part in Kaggle competitions, hosted talks with industry leaders as well as hackathons and events promoting data science in the academic community.
- Co-organized Hackathon4Ukraine where participants had a chance to create valuable solutions to help Ukrainian people. We managed to raise 33 000 DKK for Polish Humanitarian Action.
- Conducted a commercial Python course tailored to the needs of PhD biology students. The course covered fundamental concepts of the Python language, data manipulation, visualization, and machine learning.
- Presented and won the award for the best poster at the MLinPL conference in 2020.
- Player and member of Lyngby Tennis Club, currently improving my Danish Tennis Federation raking.