

Tomasz Tkaczyk Software, Data & AI Engineer

tom.tkaczyk11@gmail.com - +49 15253390797 - Berlin, DE - [linkedin.com/in/tomtkaczyk](https://www.linkedin.com/in/tomtkaczyk) - [tomasztkaczyk.com](https://www.tomasztkaczyk.com)

AI, Data, Software Engineer with 8 YoE in developing and delivering data-driven products and applications. Proven track record in machine learning, NLP, cloud computing, and ETL pipeline development. Strong communicator with a product-oriented mindset, capable of engaging with clients and participating in all stages of development.

SKILLS

Programming: Python, SQL, JavaScript, Golang, Bash, Docker, Linux, Kubernetes
ML: Kubeflow, MLFlow, sklearn, Pytorch, PySpark, Spacy, Deep Learning, NLP, Transformers, Langchain, LLamaIndex
Data Analysis: Data Mining, Web Scraping, A/B Testing, Statistical Analysis, Streamlit, Dash, Plotly
Cloud: GCP (GCS, GKE, CloudRun, PubSub, BigQuery, VertexAI), AWS (S3, EC2, SageMaker, DynamoDB)
Storage: BigQuery, Postgres, ElasticSearch, Pinecone, ChromaDB, Firestore, DynamoDB, MongoDB

WORK EXPERIENCE

Doit International

Software Engineer – Data & AI

Remote

08/2021 - Present

- Engineered and maintained business-critical ETL pipelines (SQL, Dataform) processing +1TB/day of cloud billing data powering our flagship SaaS - Flexsave
- Championed a cross-team project to build an internal FinOps platform for asset management and analytics; lowering daily operational overhead 10x, improved the profitability of Flexsave by ~25%
- Built and supported ML pipelines (Kubeflow, ARIMA, and XGBoost) forecasting cloud spend for +2000 customers
- Designed an automated decision-making system leveraging ML predictions which saved ~1 FTE of work and drastically reduced inefficiencies, thus increasing the profit margins by 20%
- Led and mentored a team of 5 engineers to develop and deploy a right-sizing tool in just 4 weeks, resulting in an estimated cost savings of \$300k per year and a promotion

Axel Springer NMT

Software & Data Engineer

Berlin, Germany

08/2020 - 08/2021

- Engineered a transformer-based (SentenceBERT) topic modeling pipeline (PyTorch, UMAP, HDBSCAN); which saved countless hours of manual work by automating a process of labeling +3k news articles per month
- Build +10 Slack bots monitoring the front page of BILD.de and delivering insights and alerts to the editorial team; saved 5h per week of manual checks and cut 90% of the reaction time in case of anomalies
- Led a cross-team project to develop a ML model to accurately predict click-through rates of news articles on BILD.de, resulting in a 15% increase in overall engagement metrics

IBM Research

Data Science Researcher

San Jose, California

03/2019 - 04/2020

- Engineered a cybersecurity anomaly detection ML pipeline for IBM's Recovery Orchestration using Random Forest and DNN; detecting +95% of intrusions when exposed to unseen ransomware strains
- Implemented secure ETL pipeline using malware Sandbox Cuckoo, VMware, and Go-lang. It fully automated the process of obtaining training data by infecting various VMs with over 200+ ransomware strains for analysis
- Collaborated with cross-functional teams to integrate the ML system into IBM's Recovery Orchestration product

Digitas Pixelpark

AI / Data Science Intern

Berlin, Germany

08/2016 - 02/2019

- Built and monitored over +20 ETL pipelines in Apache Airflow and Hadoop Map-Reduce processing millions of records every day for multiple dashboards; consistently excelling all required Data Quality metrics and SLAs
- Designed a PoC of dashboarding CMS using JavaScript & Python (D3.js, Flask, PySpark, React) which became a default communication tool with multiple clients
- Architected and implemented 10+ Tableau dashboards for Data-Driven marketing; used daily by marketing teams at Mercedes-Benz, McDonald, or BMW to monitor digital marketing KPIs

Telekom Innovation Laboratories

Student Researcher

Berlin, Germany

02/2016 - 08/2016

EDUCATION

Master of Science - Technical University Berlin

Computer Science & Media

Graduation: 2021

Thesis: Design of a embedding-based topic modeling system for news articles