DENIS VOLK

Full Stack Data Science Engineer, PhD in Math

the.denis.volk@gmail.com
+41 762713058
Zürich, Switzerland
denis-volk
dsvolk
Work permit

SUMMARY

Full-stack data scientist and ML engineer, highly skilled in modern generative AI, machine learning, MLOps, data analysis, mathematical modeling, data pipelines, big data, and cloud infrastructure. Expertise includes LLMs and AI chatbots, time series analysis and forecasting, geospatial data analysis, natural language processing, probabilistic modeling, data engineering, and team leading.

Crafting Complete Data, ML, and AI Solutions: From Idea to Production-Ready

SKILLS

Languages: Python, SQL, Scala, R, C++

GenAI: GPT, LLaMa, LangChain, RAG, Agents, Fine-tuning, Evals, Observability

ML topics: Time Series, NLP, Probabilistic ML, Graph ML, LLMs, ...

ML tools: Pandas, NumPy, Scikit-learn, XGBoost, PyTorch, TensorFlow, PyMC, ...

MLOps: Sagemaker, MLFlow, W&B, DVC, Prefect, Airflow, Grafana

DevOps: Bash, Docker, CI/CD, AWS, GCP

Databases: SQL, NoSQL, Vector DBs, Neo4j

EXPERIENCE (SELECTED)

2023 – 2024 **Founder, CTO** **OkGPT**

- Created OkGPT, an AI personal assistant messenger bot, with RAG over documents and web search
- Optimized the code to parallelize user queries' processing to increase the bot performance by magnitude.
- Integrated multiple APIs, including Telegram, OpenAI, Google, Redis, Amplitude, Datadog, and more.

`Python / SQL / ChatGPT / LangChain / RAG / Prefect / Docker / CI/CD / GCP`

2023 **AI Engineer** **Generative Tech Startup**

- Developed an MVP of an app to query enterprise data in natural language. Upon receiving a question in natural language about the data, the app outputs the answer as a plot or a small table.
- Engineered and fine-tuned the prompts to improve the quality and correctness of SQL code generation.
- Created an automatic annotator for the database columns and the final table.

`Python / ChatGPT / LangChain / SQL / NLP`

2021 – 2022 **Senior Data Scientist** **Perfect**

- Architected and directed the creation of a core Similarity engine to score candidates.
- Utilized pre-trained NLP deep neural networks to create semantic text embeddings, significantly improving the Similarity engine output results.
- Constructed a Big Data pipeline in Databricks and Spark to enrich input data and prepare features for ML.
- Prepared custom deep learning models to build richer embeddings, incorporating various data sources and metadata.

`Python / SQL / ETL / Spark / NLP / Deep Learning / Grafana / AWS`

2020 – 2021 **Senior Data Scientist** **Doorstead**

- Built an end-to-end machine learning solution predicting fair real-estate property prices, eliminating manual assessment and enabling quick customer responses.
- Designed and implemented an auto-refreshing ETL pipeline that injects, cleans, joins, and enriches new data from AWS S3 storage daily.
- Developed an interpretable machine learning model with Scikit-learn, CatBoost, Lifelines, FBProphet, FAISS, and SHAP, consisting of several submodels and satisfying business monotonicity constraints.
- Supervised other data science team members and coordinated with the engineering team.

`Python / SQL / ETL / CatBoost / Lifelines / FBProphet / FAISS / SHAP / AWS`

2017 – 2019 **Senior Data Scientist** **KPMG**

- Created a machine learning model that predicted revenues for a retail store chain based on store location, local demographic data, GIS features, seasonality, and other factors.
- Developed and deployed an interpretable machine learning model that scored B2B customers for payment default risks and provided explanations for the scores. The model massively reduced workload for weekly risks assessment.
- Built a probabilistic Bayesian machine learning model to predict which apartment buildings still under construction would fail to be commissioned in time. The model helped reduce the funds needed to hedge risks by two times.
- Developed and deployed NLP models to automatically label a vast body of housing contracts by contract type and extract contractor party names, address entities, and other attributes.

`Python / SQL / Machine Learning / NLP / Bayesian Models / MLFlow / DVC / Hadoop`

2010 – 2017 **4 Postdoc positions in Math and Neuroscience** **Academia**

- SISSA, KTH, Tor Vergata, HSE: authored 15+ papers in top academic journals

`Mathematics / Neuroscience / Research`

2004 – 2007 **Software Developer** **A4Vision**

- Implemented and tuned 3D surface reconstruction algorithms.
- Designed and implemented algorithms for face detection on a 2D image.
- Migrated the algorithmic core to an embedded platform.

`C++ / Machine Learning / Computer Vision / Embedded`

EDUCATION

2004 – 2010 **Ph.D. in mathematics** **Lomonosov Moscow State University**

1999 – 2004 **Combined B.S./M.S. in mathematics, Summa cum laude** **Lomonosov Moscow State University**

AWARDS

1998 **2nd place** **Russian National Mathematical Olympiad**

LANGUAGES

English - fluent, **Italian** - good, **German** - beginner, **Russian** - native

PERSONAL

In my free time, I enjoy

- Hiking and skiing with my wife and friends
- Reading sci-fi and pop science