Árpád Goretity

Data Scientist, Bioinformatician

arpad@goretity.com
+36 30 183 1717
Budapest, Hungary
h2co3
H2CO3
goretity.com

WORK EXPERIENCE

Lead Bioinformatics Scientist

Medipredict Ltd.

Dec 2019 - Present

- Designed and implemented pipelines for analysis of gut microbiome (strain composition), blood metabolomics, and dietary data using Python, Scikit-learn, Rust, and Large Language Models
- Trained, verified, and deployed IBD and diabetes prediction models using Scikit-learn and Docker
- Obtained a patent on a IBD prediction model based on microbial functional gene compositions
- Created production-ready network visualization using Seaborn, GraphViz, and igraph
- Advised and mentored 2 junior team members on the fundamentals of statistics and programming
- Co-operated with research scientists to normalize architecture of internal knowledge bases
- Created ETL pipelines and API client scripts for accessing and preprocessing public scientific databases (KEGG, NCBI Taxonomy, RefSeq)
- Automated document generation tasks using Typst
- Held an internal seminar series on Python and data analytics best practices

Data Platform Engineer

PrivátDoktor (S+H Portfolio PLC)

Sept 2020 - Present

- Trained and deployed ML models for a medical decision support system: analyzed ECG, PPG, blood glucose, and blood pressure data using SciPy
- Improved performance of existing ECG signal processing pipeline with Rust and WASM
- Designed normalized database for training data using SQLAlchemy, SQLite, and PostgreSQL

Graduate Student Researcher

PPCU, Faculty of Information Technology

Sept 2020 - Aug 2022

- Systematically examined the complexity of modern query languages such as SQL and ORM DSLs
- Designed and partly implemented a strongly-typed ORM DSL supporting hierarchies and sum types

EDUCATION

MSc in Data Science

Università degli Studi di Padova, Dip. di Matematica

Sept 2018 - July 2020

- Thesis: Towards Personalized Disease Risk Prediction from Metagenome Analysis of the Microbiome
- Qualification: 110 / 110, Cum Laude

BSc in Bionic Engineering

Pázmány Péter Catholic University (PPCU FITB)

Sept 2013 - Dec 2016

- Thesis: Design and FPGA Implementation of a Protein Structure Comparison Method Based on Alignment of Backbone Conformations
- Qualification: 5 / 5 Excellent

SKILLS

Python | Rust | C | C++ | JavaScript, TypeScript | Statistics | Data analysis | Bioinformatics | SQLite | Postgres | MongoDB | Unix shell | Git | Docker | LaTeX | Typst | Teaching | Public speaking | English (C2) | French (C1) | Italian | Hungarian

OPEN SOURCE PROJECTS

NanoSQL

Databases, ORMs

- Strongly-typed, very lightweight data mapper for SQLite and Rust

Hindmarsh-Rose

Simulations, Neuroscience

- High-performance C++ implementation of the Hindmarsh-Rose neuron model

Avocado

Databases, ORMs

- Ergonomic, convenient MongoDB wrapper