Denis Semkin

Saarbrücken, Germany ✉ semkind001@gmail.com ☐ +4915119489367 in/denis-semk

SUMMARY

Master's student in Bioinformatics with a strong interest in using computational methods to solve complex biological problems. Experienced in analyzing datasets, ensuring data quality, and managing metadata. Skilled in Python programming and data visualization tools. I am looking to contribute to research and development projects. Dedicated to advancing bioinformatics through teamwork and cutting-edge technologies.

EXPERIENCE

Student research assistant

Saarland University AES   **July 2023 – Present, Saarbrücken 66123, Germany**

· Contributed to the development and implementation of a PV-Powered Battery System Modeling project, enhancing the efficiency and reliability of renewable energy systems.
· Conducted comprehensive data mining and processing using Python, resulting in improved data accuracy and actionable insights.
· Developed and maintained Python scripts for data management, analysis, and visualization.
· Provided robust code for modelling and optimization, significantly improving system performance and predictive accuracy.
· Managed and optimized databases using MongoDB and InfluxDB, ensuring efficient data storage and retrieval processes.
· Utilized Unix/Bash CronJob for automating model processes, increasing operational efficiency and reducing manual intervention.

PROJECT

MolFormer Fine Tuning

Saarland University, Spoken Language Systems · January 2025 – March 2025
· **Fine-tuned** a pre-trained **chemical language model** on a **lipophilicity regression task** using SMILES data and **Hugging Face** tools.
· Developed an end-to-end workflow for **SMILES** tokenization, dynamic padding, and integrating a regression head.
· Enhanced the model via unsupervised pre-training with a **Masked Language Modeling** objective.
· Applied **influence functions** with **LiSSA** to select high-impact external data for augmentation.
· Evaluated parameter-efficient strategies (**BitFit, LoRA, iA3**), identifying **iA3 with stratified sampling** as the most effective approach.

Molecular Docking

Saarland University, Volkamer ″Data Driven Drug Design″ Lab · October 2024 – February 2025
· Executed **GPCR docking** simulations with **GNINA** to investigate receptor-ligand interactions.
· Compared **flexible protein-flexible ligand** versus rigid protein-flexible ligand docking methodologies.
· Utilized **PyMol** for advanced 3D visualization of docking poses and interaction mapping.
· Analyzed the impact of **protein flexibility on docking accuracy** and binding affinity predictions.

Epigenetic data processing and analysis

Saarland University, Integrative Cellular Biology and Bioinformatics · August 2024 –August 2024
· Gained hands-on experience in **NGS-based epigenome analysis**, including RNA-seq, WG bisulfite-seq, ATAC-seq, and ChIP-seq.
· Conducted bioinformatic analyses using **UNIX shell** and **R** for data manipulation and interpretation.
· Performed **integrative multi-modal analysis** (clustering, differential analysis, segmentation, imputation) to extract meaningful insights.
· Collaborated in a team to **analyze epigenetic data** and **presented findings**, showcasing independent and teamwork skills.

EDUCATION

Master's degree in bioinformatics

Saarland University · Saarbrücken 66123, Germany · 2025

Bachelor's degree in biotechnology

Lomonosov Moscow State University · Moscow 119991, Russia · 2018 – 2022

SKILLS

Python: NumPy, Pandas, Polars, MatPlotLib, Seaborn, Scikit Learn, PyTorch, PvLib, PyOmo, RdKit, Smina, GNINA

R: Tidyverse, Bioconductor

Other: PyMol, Homology Modelling, Molecular Docking, PDB, Uniprot, ChEMBL, KLIFS, Gromacs, Bash, InfluxDB, MongoDB, Latex

Languages: English (C1), German (A2)