Jared Klug

jared.a.klug@gmail.com ❖ (516) 512-4273 ❖ New York, NY ❖ jaredaklug.github.io

WORK EXPERIENCE

Memorial Sloan Kettering **Dec. 2022 – Present**
Data Analyst *New York, NY*

- Optimized data processing pipeline using parallel computing and cluster technology, reducing computational time by 50%, leading to faster clinical insight
- Implemented advanced statistical techniques, including a robust method to quantify surgeons' contribution to outcome variation. This data-driven approach led to a redesigned feedback system, resulting in a 25% increase in patient survey completion rate and the establishment of a targeted surgeon shadowing program.
- Maintained and expanded the team's R package, implementing specialized functions to improve coding efficiency and optimized validation workflows. Resulting in a 70% reduction in validation time and significantly enhancing the reliability and reproducibility of downstream analyses
- Developed and executed complex SQL queries to create robust patient cohorts, joining and preprocessing data from multiple clinical databases.

Weill Cornell Medicine **June 2021 – Dec. 2022**
Computational Research Technician *New York, NY*

- Engineered a scalable bioinformatics pipeline for high-throughput assembly of genomes, deploying it on a Linux-based HPC cluster using SLURM schedulers for parallel processing.
- Conducted genome-wide association studies on WGS data to identify genomic variations in fungal strains related to Inflammatory Bowel Disease using advanced bioinformatics tools.
- Leveraged data visualization libraries (e.g., ggplot2, ploty) and web framework (Shiny) in R to construct interactive dashboards showcasing differential analysis between IDB cohorts.
- Developed an in-house database using MySQL to efficiently manage and query bioinformatics data.

Massachusetts General Hospital **Sept. 2018 – May 2019**
Undergraduate Researcher *Charlestown, MA*

- Developed a machine learning model using Support Vector Machines and Principal Component Analysis in R to predict Alzheimer's risk, achieving 90%+ accuracy.
- Architected a scalable ETL pipeline in R, integrating API data retrieval and advanced data wrangling techniques to process large-scale hematological datasets.
- Optimized SVM hyperparameters through grid search and cross-validation, ensuring model generalizability and mitigating overfitting.

EDUCATION

Columbia University **May 2022**
Master of Science in Biostatistics *New York, NY*
GPA: 4.0/4.0

Boston University **May 2019**
Bachelor of Science in Biomedical Engineering *Boston, MA*
GPA: 3.4/4.0 *Cum Laude*

SKILLS & INTERESTS

- **Technical Skills:** R Programming; SQL; Python; Statistical Programming; Advanced Statistics; High-Performance Computing; ETL; Agile Methodologies; Research Collaboration; Git;
- **Interests/Hobbies:** 3D Printing & Design; Mechanical Keyboards; Pickleball; World of Warcraft; Overwatch