Zachary Lee Skidmore

Senior Quantitative Scientist

Data Science | Machine Learning | Statistical Modeling | Feature Engineering | Data Visualization | Software & Package Development

St. Louis, MO | GitHub | https://zlskidmore.github.io | ORCiD

Professional Summary:

Senior quantitative scientist and data analyst with 12+ years of experience applying hypothesis testing, statistical modeling, machine learning, and software development to complex, noisy, high-dimensional data across academic and industry settings. Proven track record of building robust data pipelines, developing and validating predictive models, and translating analytical results into clear, actionable insights. Experienced in end-to-end quantitative problem solving spanning experimental design, feature engineering, modeling, and interpretation. Author of 50+ peer-reviewed publications and patent co-inventor. Mentor and instructor in quantitative and computational methods.

Skill Summary *(See Page 2 for a complete list)*:

Education: University of Illinois - Chicago; Master of Engineering in Bioinformatics — May 2013

Work Experience:

Delfi Diagnostics — Remote

Position: Senior Quantitative Scientist, Translational Science, Data Science - Research — Aug 2022 - Oct 2025

2023 - 2025 Highlights

- Co-inventor for provisional patent leveraging high-dimensional features to classify samples. (WO2024173277A2)
- Increased classification performance by aggregating and transforming high-dimensional features using supervised and unsupervised learning approaches (e.g., generalized additive models, hierarchical clustering, PCA), improving signal-to-noise and model robustness.
- Built dimensionality reduction and projection workflows (PCA) to embed incoming data into learned feature spaces for downstream stratification tasks.
- Designed probabilistic models (binomial) to quantify the impact of experimental modifications and structural assumptions on feature behavior and model performance.
- Lead author and analyst for the application of a production level predictive model evaluated in a prospective cohort. PMID: (39433569)
- Maintained and iteratively improved a production level random forest model and validated performance using novel engineered features.
- Delivered interpretable analytical results to cross-functional teams and external partners, directly supporting multi-million-dollar research and development collaborations.
- Executed 15+ end-to-end applied data science projects, spanning experimental design, feature engineering, modeling, validation, interpretation, and delivery to internal and external stakeholders.

McDonnell Genome Institute, Washington University Department of Medicine — St. Louis, MO

Position: Senior Staff Scientist - Griffith Lab — Nov 2013 - Aug 2022

2013 - 2022 Highlights

- Designed, cross-validated, and deployed a machine learning model (e.g., XGBoost) that significantly improved classification accuracy and reduced manual classification effort in large, complex datasets.
- Refactored and optimized production analytical software, achieving up to a 20× improvement in performance and memory efficiency.
- Developed and maintained reusable data processing pipelines using workflow languages and containerization (CWL, Docker) for institute-wide use. (https://hub.docker.com/u/zlskidmore, bioinformatic, analysis-workflows)
- Led the development and public release of open-source analytical and visualization software (GenVisR), adopted by external users and cited in peer-reviewed research. PMIDs: (27288499, 34506690)
- Prototyped interactive visualization tools (e.g., Shiny applications) to enable exploratory analysis and stakeholder interpretation of high-dimensional data. (saturation-mutagenesis-viz)
- Trained and mentored scientists and engineers across disciplines in statistical modeling, data analysis, and reproducible computational practices.
- Created and maintained static websites (Jekyll) as a training and teaching aids.
- Applied statistical and quantitative analysis to large, heterogeneous cohorts, contributing to 50+ peer-reviewed research publications.

Technical Instruction & Public Speaking:

- Instructor, teaching assistant, and content author for domestic and international technical workshops focused on data science, statistical modeling, visualization, and best practices (2017–2025)
  - CRI, *Physalia-courses*, Evomics

Technical and Analytical Expertise	
Advanced (***): Extensive multi-year experience with deep understanding. Have mentored, taught, or critically evaluated the topic. **Proficient (**):** Multi-year practical experience. Demonstrated use in published research or public repositories. **Familiar (*):** Working knowledge through guided exposure or light use.	
Statistical Analysis	Machine Learning and Modeling
• Hypothesis Testing (Parametric/Non-Parametric)*** • Survival Analysis (Kaplan-Meier)*** • Batch Effect Detection/Correction** • Power Analysis** • Distribution-based Modeling (e.g., binomial, negative binomial, poisson)** • Multivariate Analysis **	• Supervised and Unsupervised Learning ** • Dimensionality Reduction (PCA) ** • Clustering (hierarchical, k-means if applicable) ** • Feature Engineering ** • Random Forest** • Extreme Gradient Boosting** • Generalized Linear Models** • Generalized Additive Models ** • Cross-Validation** • Model Evaluation (ROC, AUC, Precision-Recall)**
Data Cleaning & Visualization	Programming & Software***
• Data Cleaning & QC *** • Large-Scale Data Processing *** • Data Visualization *** • Interactive Dashboards / Tools **	• R (inc. ggplot2, data.table, S4, shiny) *** • UNIX *** • GNU Utils (incl. awk, sed, vim) *** • Docker (incl. Image building) *** • Git/GitHub *** • High Performance Computing (e.g. LSF, Slurm) *** • Amazon Web Services (incl. EC2, S3) *** • Confluence/JIRA ***
Programming & Software**	Programming & Software*
• Python • Common Workflow Language and Pipeline Implementation • Perl • Jekyll/Static Site Building • HTML/Web Development • Markdown	• SQL • CSS/SASS • D3.js