

KENNETH N. REID, PH.D.

Ken@kenreid.co.uk · github.com/drkenreid

SUMMARY

Data Scientist with experience in generative AI, machine learning, and the entire data science pipeline. Have worked in multi-disciplinary teams, applying advanced data science techniques to solve complex problems across various domains and data types. Experienced educator of data science to multi-disciplinary faculty at the University of Michigan.

TECHNICAL SKILLS

- **Data Science:**
 - Machine Learning, GenAI, Visualization, Statistical Analysis, Data Collection & Cleaning
- **Code:**
 - Python, R, SQL, Java
- **Version Control:**
 - Git, SVN
- **Visualization:**
 - Tableau, Seaborn, Matplotlib
- **Other:**
 - Docker, \LaTeX

CURRENT & PAST DATA SCIENCE PROJECTS

Large Language Model for Library Meta-data Generation and Support

- Collaborating with team members to implement software for meta-data generation tasks, including:
 - Generation of READMEs, sourced from data and researcher input
 - Licensing guidance users, based on data
 - Filetype matchups of uploaded data
 - Sensitive material checks
- Tech stack: Python, AWS, LMStudio, AnythingLLM, LLM Agents

Data Science Resource Expertise

- Developing a custom LLM to help researchers find data science resources at UM
- Implemented NLP techniques to find appropriate content by scraping thousands of internal UM sites
- Co-authored and maintained guides for researchers on use of generative AI in research

Genomic Prediction Models

- Contributed to the creation of various machine learning models for genomic prediction, increasing accuracy by up to 15%
- Optimized hyperparameters using evolutionary algorithms, improving efficiency and results by several times over
- Feature selection and HDF5 files aided in computation efficiency improvements
- Tech stack: Python, TensorFlow, Scikit-learn, DEAP

Employee Scheduling Optimization

- Developed a hybrid metaheuristic algorithm for real-world employee scheduling
- Reduced scheduling conflicts by 26%, improved employee satisfaction scores by 20%, among other improvements
- Tech stack: Python, DEAP, Java, SVN, JavaScript, SQL, CPLEX Optimizer

RELEVANT WORK EXPERIENCE

Data Scientist, University of Michigan MIDAS (04/2023 - Present)

- Provided data science expertise across various domains and disciplines, and across a variety of data types
- Collaborated on the development of advanced ML models, solving novel problems with experts at UM
- Used network analysis to provide evidence for improving research collaborations, aiding in securing funding
- Organized and taught at AI bootcamps for faculty, staff and industry professionals in advanced data science
- Worked closely with the team lead to align projects with organizational goals

Research Associate, Michigan State University (11/2019 - 04/2023)

- Applied ML techniques to genomic data, improving prediction accuracy by 15%
- Optimized LLVM compiler pass sequences, massively reducing compile time for various common problems
- Mentored 2 Ph.D. and 5 M.Sc. students
- Published peer-reviewed papers in top-tier journals and conferences in various fields using data science techniques

Research Assistant, University of Stirling, UK (04/2019 - 10/2019)

- Worked closely with British Telecommunications Plc. to optimize 25,000 engineers schedules
- Developed ML models using multi-spectral satellite data for malaria prediction via detection of small water bodies
- Set the foundation of a new project in trajectory mining search optimization

EDUCATION

- **Ph.D. & B.Sc. with Honours in Computing Science**, University of Stirling (2019 & 2013)

SELECTED PUBLICATIONS

- Boyko, J., et al. (2023). An Interdisciplinary Outlook on Large Language Models for Scientific Research. *arXiv. Under Review in Science Advances*.
- Han, J., Gondro, C., Reid, K.N., Steibel, J.P. (2021). Heuristic hyperparameter optimization of deep learning models for genomic prediction. *G3 Genes | Genomes | Genetics*.
- Reid, K.N., et al. (2019). A Hybrid Metaheuristic Approach to a Real World Employee Scheduling Problem. *GECCO'19: The Genetic and Evolutionary Computation Conference 2019. ACM*.