Kenneth N. Reid, Ph.D.

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

Summary

Senior Data Scientist with expertise in generative AI, machine learning, and the entire data science pipeline. Experienced in leading cross-functional teams, technical oversight, and applying advanced data science techniques to solve complex problems across various domains and data types.

Technical Skills

- **Data Science:**
  - Machine Learning, GenAI, Visualization, Statistical Analysis, Data Collection & Cleaning
- **Code:**
  - Python, R, SQL, Java
- **Big Data:**
  - Spark, Hadoop
- **Visualization:**
  - Tableau, Seaborn, Matplotlib
- **Cloud:**
  - AWS, GCP
- **Version Control:**
  - Git, SVN
- **Other:**
  - Docker, LaTeX, Salesforce

Data Science Projects

Large Language Model for Library Meta-data Generation and Support

- Software including a fine-tuned LLM for meta-data generation tasks, including:
  - Licensing guidance for README files
  - Filetype matchups of uploaded data
  - Sensitive material checks
- Tech stack: Python, PyTorch, Transformers, GCS and HPC SLURM for training

LLM for Data Science Resource Assistance

- Developed a custom LLM to help researchers find data science resources at UM
- Implemented NLP techniques to find appropriate content from across UM internal sites
- Tech stack: Python, PyTorch, Hugging Face Transformers, AWS

Genomic Prediction Models

- Created variuos 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

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

- Lead data science projects across various domains, managing datasets up to a petabyte in size
- Develop and implement advanced ML models, solving previously unsolved problems
- Using network analysis, provided evidence for institute improving research collaboration, thus securing funding
- Aided in securing millions of dollars in funding through collaborative research proposals and industry partnering
- Organize and teach AI bootcamps, training hundreds of faculty members and industry professionals

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 in advanced data science techniques
- 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)

- Collaborated with British Telecommunications Plc. to optimize 25,000 engineers schedules, working onsite in the research and software development team
- Developed ML models using multi-spectral satellite data for malaria prediction via detection of small water bodies

Education

- **Ph.D. in Computing Science**, University of Stirling (05/2015 - 07/2019)
- **B.Sc. with Honours in Computing Science**, University of Stirling (08/2009 - 05/2013)

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.