(864) 617-7166
samvoisin@protonmail.com

Sam Voisin

SUMMARY STATEMENT

Driven and results-oriented Data Scientist with extensive experience in designing and implementing innovative data pipelines, models, and algorithms. Published climate science researcher. Seeking a challenging role in a forward thinking organizations where my experience in machine learning and climate science can contribute to impactful solutions.

SKILLS

- Expertise in probability, statistical inference, predictive modeling, and optimization
- Proficiency in Bayesian and classical modeling and inference
- Comprehensive knowledge of Python; experience with R, Julia, SQL
- Proficiency in deep learning frameworks, especially PyTorch
- Practical machine learning experience with a strong theoretical understanding of algorithm complexity
- Expertise in Natural Language Processing (NLP) and advanced text mining techniques
- Effective communication and presentation skills for sharing insights and results with both technical and non-technical stakeholders

PROFESSIONAL EXPERIENCE

Infinia ML, Durham, NC — *Data Scientist*

MARCH 2022 - OCTOBER 2023

- Developed generic and bespoke pipelines for analyzing unstructured documents
- Implemented solutions for a variety of NLP tasks such as NER, text classification, etc.
- Constructed and deployed scalable document processing pipelines capable of handling hundreds of thousands of documents per day
- Designed and implemented large language model (LLM) infrastructure

Geometric Data Analytics, Inc, Durham, NC — *Data Scientist*

JUNE 2020 - MARCH 2022

- Led development of a novel vortex tracking algorithm for oceanographic research (See publications)
- Directed experiment design and analysis for assessing new compression algorithm designed for edge computing (see publications)
- Developed a tiered system for modeling local, regional and global shipping data
- Performed ad hoc statistical analyses for clients including DARPA, NRL and AFRL
- Planned architecture and managed conversion of research code into deployable libraries

Duke University, Durham, NC — *Research Assistant*

APRIL 2019 - SEPTEMBER 2019

- Formulated research problems and procedures; Designed and executed experiments to meet research objectives
- Designed and constructed pipelines for data preprocessing and modeling
- Developed and optimized MCMC samplers for Bayesian hierarchical regression models

Ally Financial Services, Charlotte, NC — *Analyst*

JANUARY 2015 - JUNE 2018

- Analyzed exchange data and business metric relationships to mitigate risk
- Automated data gathering and processing for significant lead time and error reduction
- Acted as program lead and mentor for department internship program

EDUCATION

Duke University, Trinity College of Arts and Sciences — *Master of Statistical Science*

AUGUST 2018 - MAY 2020

- Master of Statistical Science; Focus on high-dimensional geometric and topological data analysis
- Thesis: *Graph Diffusion for Gesture Classification* focused on pre-processing signals to improve quality of downstream human-computer interfaces
- Expertise in machine learning algorithms, with a strong theoretical understanding of algorithm complexity

Clemson University, College of Business and Behavioral Science — *Bachelor of Science in Financial Management*

AUGUST 2010 - DECEMBER 2014

- Understand economics of businesses operating in a dynamic environment

AWARDS & PUBLICATIONS

G. Koplik et al., "Topological Simplification of Signals for Inference and Approximate Reconstruction," 2023 IEEE Aerospace Conference, Big Sky, MT, USA, 2023, pp. 1-11, doi: 10.1109/AERO55745.2023.10115654.

Voisin, S., Hineman, J., Polly, J. B., Koplik, G., Ball, K., Bendich, P., D'Addezio, J., Jacobs, G. A., & Özgökmen, T. (2022). Topological feature tracking for Submesoscale eddies. Geophysical Research Letters, 49(20). https://doi.org/10.1029/2022gl099416

First place at University of South Carolina Big Data Health Science Conference 2020 case study competition.