

# Sam Voisin

---

## SUMMARY STATEMENT

Skilled data scientist and software engineer with extensive experience in developing robust data pipelines, scalable architectures, and efficient algorithms. Proficient in integrating data science techniques into software solutions, with a strong foundation in statistical modeling, machine learning, and software design. Seeking a position to contribute to innovative data-driven projects.

## SKILLS

- Expertise in probability, statistical inference, and predictive modeling
- Advanced understanding of Python and R
- Proficiency in deep learning frameworks, especially PyTorch
- Expertise in Natural Language Processing (NLP) and advanced text mining techniques
- Familiar with RESTful API fundamentals; experience with FastAPI
- Practical experience in database querying and management using SQL, Neo4j, PySpark
- Competent in the ML development toolset including Linux, Docker, and Kubernetes
- Well-versed in software development and Agile methodology
- Effective communication skills with both technical and non-technical stakeholders

## PROFESSIONAL EXPERIENCE

### **Tradewind Data Science, Chicago, IL** — *Senior Data Scientist*

OCTOBER 2023 - CURRENT

- Executed efficient ETL operations on large-scale data clusters using SQL and PySpark
- Analyzed dynamic consumer trends and translated insights into clear actionable recommendations for clients, facilitating data-driven decision making
- Enhanced client revenue through advanced econometric modeling techniques to optimize pricing strategies and revenue generation

### **Infinia ML, Durham, NC** — *Data Scientist*

MARCH 2022 - OCTOBER 2023

- Developed adaptable and bespoke pipelines for analyzing unstructured documents
- Implemented NLP solutions for entity recognition and text classification
- Constructed and deployed scalable pipelines capable of processing hundreds of thousands of documents per day
- Developed architecture integrating LLMs into the data science tech stack

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

JUNE 2020 - MARCH 2022

- Developed novel multi-object tracking algorithm (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 pattern-of-life data

- Performed statistical analyses for clients including DARPA, NRL and AFRL
- Planned architecture and managed conversion of research code into deployable libraries

### **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* Signal processing to improve downstream prediction quality

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

AUGUST 2010 - DECEMBER 2014

- Strong working knowledge of finance and economics
- Understanding of business needs and constraints

## AWARDS & PUBLICATIONS

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>

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.

First place at U of SC Big Data Health Science Conference 2020 case study competition.

## OPEN-SOURCE PROJECTS

### **GestuReMote**

- Computer vision based system for controlling a personal computer
- <https://github.com/samvoisin/gesture-control>