Zachary Dawson

M.S. CS (ML), Georgia Tech — 2024–Present

zachwdawson@gmail.com | 508-446-3322 | zachwdawson.github.io

Summary

Machine-learning–focused data engineer completing M.S. CS (ML). Build and ship models + data pipelines on AWS/Spark. Open to U.S. relocation and will self-relocate.

Experience

Data Engineer (part-time), Learning Analytics Lab, Georgia Tech – Atlanta, GA   Aug 2025 – Present
- Develop and maintain an Azure-based learning-analytics platform: integrate new data sources, build PySpark ETL, contribute Terraform IaC and developer-experience improvements, and support dashboards/interactive visuals.

Senior Data Engineer, Global Atlantic Financial Group – Brighton, MA   Sept 2021 – May 2024
- Introduced the company's first ML model (cosine similarity + random forest) to detect duplicate contacts, improving accuracy by 10% vs. rules-based matching.
- Migrated 200+ Control-M jobs to AWS Glue with Terraform, enabling infra-as-code deployments and reducing manual scheduling.
- Built an automated test framework for database and stored procedure migrations (Python + CI), replacing manual checks and catching regressions pre-deploy.
- Integrated Salesforce Bulk API to read/write prospects, adding 5,000 new records to Salesforce and the enterprise data warehouse with auditability.
- Migrated annuity CRM data to the in-house warehouse to enable enterprise reporting at scale (700,000 policies; $40B in transactions).

Education

Georgia Institute of Technology | M.S. CS (ML) | GPA: 4.0   Aug 2024 – Expected May 2026
Relevant Coursework: Deep Learning, Machine Learning, Bayesian Statistics, Natural Language Processing
Northeastern University Honors Program | B.S. CS | GPA: 3.7   Sept 2017 – May 2021

Technical Skills

Languages: Python, SQL, Scala, Java, Bash
ML/Stats: PyTorch, scikit-learn, PyMC, Pandas, NumPy, SciPy
Data/Orchestration: Spark/PySpark, ETL pipelines
Cloud & IaC: AWS (S3, Glue, Redshift, Lambda), Azure (Synapse, Delta Tables), Terraform
Other: Git, Linux, MySQL, Jenkins, Jupyter

Projects

Bayesian Golf Handicapping   zachwdawson.github.io
- Developed a fairer golf handicapping system using Bayesian methods, reducing fairness gap between handicap groups from 10% (USGA) to 7% .
- Created random-walk and GLM models using PyMC5 to predict golfer's future score distributions based on previous scores.

Golf Swing Action Recognition   zachwdawson.github.io
- Improved frame-level golf swing labeling with 3D CNNs and custom loss functions, attaining frame-level PCE of 81.6% vs. 76.1% baseline.
- Produced model, loss functions, and training and evaluation loops using PyTorch for 1400 golf swing videos, totaling 219 hours of footage.