

GREGORY SCHREITER

(507) 990-3557 | schreitergregory@gmail.com | gregschreiter.com | [GitHub](#) | Open to relocation

SUMMARY

Data scientist and analytics professional with experience building machine learning models, validating AI outputs, developing ETL workflows, and analyzing healthcare-related datasets using Python, SQL, R, TensorFlow, and scikit-learn. M.S. in Analytics from Georgia Tech with strong foundations in statistical modeling, Bayesian inference, regression, model evaluation, data preprocessing, and communicating technical findings to diverse audiences.

EXPERIENCE

AI Trainer | DataAnnotation.tech | April 2024 – Present

- Evaluated LLM, AI, and machine learning outputs for accuracy, reasoning quality, instruction adherence, and reliability
- Reviewed AI-generated Python code, simulations, data analysis, and chatbot responses to identify errors and improve output quality
- Performed QA on information extraction, summarization, and content generation workflows while providing structured model feedback
- Identified recurring model failure patterns and documented feedback to support continuous improvement of AI systems

Instructional Associate | Georgia Institute of Technology | Jan 2021 – May 2024

- Supported hundreds of data science students per year in Bayesian statistical inference methods
- Held weekly office hours covering Bayesian linear/logistic regression, Bayes theorem, and statistical modeling concepts
- Provided instruction on inference, Gibbs/Metropolis sampling algorithms, and MCMC simulation in BUGS software
- Communicated complex statistical and quantitative concepts clearly to students with varied technical backgrounds

Previous Experience | Healthcare, Education, and Student Support | 2011 – 2021

- Supported individuals with disabilities in a patient-centered environment while maintaining reliability, confidentiality, and documentation standards
- Certified in first aid, HIPAA, and universal health precautions
- Provided classroom, behavioral, and math instruction support in K-12 settings
- Tracked and visualized student progress data to support individualized learning plans and instructional decisions

SKILLS

Languages: Python, SQL, R, JavaScript

Machine Learning & Statistics: scikit-learn, TensorFlow, regression, classification, decision trees, neural networks, Bayesian inference, MCMC, model evaluation, hypothesis testing

Data Analysis & Engineering: pandas, NumPy, SQLAlchemy, pandera, glue, AWS (S3, PySpark, Terraform, ETL, Great Expectations), data cleaning, feature engineering, data validation, data preprocessing

Databases & Visualization: PostgreSQL, NoSQL, Tableau, Matplotlib, Excel/VBA

Tools: GitHub, LaTeX

PROJECTS

Diabetes Prediction Using Machine Learning | [GitHub](#)

- Built SVM, decision tree, and deep learning models to predict diabetes with up to 97% accuracy
- Analyzed patient health indicators including age, BMI, glucose levels, and demographic variables using Python, scikit-learn, and TensorFlow
- Applied data preprocessing, feature analysis, model training, and model evaluation techniques to a healthcare prediction problem

Medical Image Classification with Convolutional Neural Networks (CNNs) | [GitHub](#)

- Built TensorFlow CNN models to classify chest CT scans for cancer detection
- Evaluated custom CNNs, transfer learning, and ensemble methods to increase performance
- Improved validation accuracy to 97.5% with an ensemble model, outperforming individual pretrained models at 84.4%, 75.9%, and 50.6% and a base CNN model at 29.2%

Exploring Heart Disease Indicators: A Data Engineering Project | [GitHub](#)

- Built an ETL pipeline using pandas, pandera, SQLAlchemy, and PostgreSQL to clean and transform healthcare and population datasets
- Designed a relational database schema and ERD to support SQL and Python-based analysis
- Performed data quality checks, validation, preprocessing, and documentation to prepare data for reliable analysis

EDUCATION

Master of Science in Analytics, 4.0 GPA

Georgia Institute of Technology | Atlanta, GA

[Certificate, Data Visualization and Analytics Bootcamp](#)

University of Minnesota | Minneapolis, MN

Bachelor of Science in Mathematics, Economics and Statistics

University of Minnesota | Minneapolis, MN

[Machine Learning by Stanford University & DeepLearning.AI on Coursera](#) (August 7, 2026)

[Data Engineering by DeepLearning.AI & Amazon Web Services on Coursera](#) (August 22, 2026)

[TensorFlow Developer Professional Certificate by DeepLearning.ai](#) (September 4, 2026)

[Deep Learning Specialization by Stanford University & DeepLearning.ai](#) (October 10, 2026)