Steven Haworth

262-949-8102
0218steven@gmail.com
GitHub
LinkedIn
Personal Website

SKILLS

- **Languages (Ordered by Expertise)** - Python, SQL, R, CQL, Powershell, MATLAB, Julia, C++
- **Frameworks** - NumPy, Pandas, SK-Learn, TensorFlow, PyTorch, SciPy, Spark, Apache Kafka, MNE, NetworkX, AWS, FastAPI, Docker, Tableau, Matplotlib, Seaborn, Google AI Studio, Git, GitHub, GitLab

EXPERIENCE

UWHealth Hospital & Clinics — Madison, WI
ML Research Assistant — *May 2024 - July 2025*

- Applied Bayesian Network modeling and graph-based feature extraction to EEG data for mental state classification; achieved 0.88 ROC-AUC using supervised learning on network-derived embeddings
- Revealed meaningful manifold geometry in a previously inseparable feature space using manifold learning and hyperparameter optimization, leading to robust sleep stage differentiation (Precision: 82%, F1: 76%)
- Forecasted 72-hour seizure recurrence using EEG features and Cox regression, visualizing survival probabilities and identifying interpretable risk factors; results accepted for presentation at international conference

Washington University-St. Louis, Beth Israel Deaconess, Harvard Medical — Partially Remote
Graduate Researcher — *Sep 2024 – Present*

- Collaborated with Beth Israel Deaconess and WUSTL to forecast seizure risk using Convolution Neural Networks, LLMs, and fundamental machine learning methods (tree-based classifiers, linear/kernel methods)
- Queried and joined up to 30 TB of EEG and metadata from AWS S3 using Apache Spark; sampled a confounding variable controlled subject population to ensure fair and accurate seizure risk modeling
- Tuned a CNN-based embedding pipeline in PyTorch to convert long-form EEG signals into compact feature vectors, supporting temporal modeling of seizure recurrence
- Engaged in biweekly cross-institutional code reviews and research presentations to improve model quality, interpretability and share recent improvements to model architectures

PROJECTS

Transformer-Based EEG Topic Miner *GitHub Link*

- Encoded abstracts into dense embeddings and built a FAISS-based semantic search engine to support fast, scalable EEG topic retrieval across 10,000+ biomedical papers
- Containerized the full NLP pipeline with Docker and YAML-based configuration; integrated CI/CD hooks to detect model drift to retrain on new incoming literature
- Fine-tuned SciBERT using HuggingFace Transformers with cross-entropy loss and the AdamW optimizer on weakly labeled PubMed abstracts; achieving 97% accuracy on a hold-out set

GPU-Boosted Music Recommendation System *GitHub Link*

- Designed and deployed a Two-Tower deep learning recommender system using NVIDIA Merlin on the Million Song Dataset; optimized for implicit feedback and Top-N retrieval
- Engineered a full Spark-based ETL pipeline to preprocess 1M+ user–song interactions, outputting parquet files for GPU-accelerated training with NVTabular
- Integrated CI/CD workflows with GitHub Actions to automate testing, feature validation, and retraining upon model drift detection.

Synthetic Time Series with Diffusion Models *GitHub Link*

- Developed a denoising diffusion probabilistic model (DDPM) to generate realistic synthetic S&P 500 price series; achieved distributional precision in kurtosis, volatility, and autocorrelation
- Tuned a deep Bayesian network with attention-based architecture to stabilize training and match empirical statistics; improved standard deviation similarity from 1.12 to 0.58
- Benchmarked model performance against GAN and ARIMA baselines; synthetic data preserved structural trends and volatility regimes in holdout evaluations

EDUCATION

University of Wisconsin-Madison — **May 2025**
Bachelors in Data Science, Dean's List Student

University of Wisconsin-Madison — **Expected May 2027**
Master of Science in Data Science, Highly Qualified Candidate Scholarship