Derek Lankeaux - Resume

Derek Lankeaux

Research Engineer | Machine Learning | Applied Statistics

dl1413@rit.edu | GitHub: dl1413 | Rochester, NY

Summary

Research-focused applied statistician with expertise in machine learning, ensemble methods, and large language models. Proven track record of developing production-ready ML pipelines achieving state-of-the-art performance on complex classification tasks. Strong foundation in Bayesian statistics, experimental design, and responsible AI practices. Passionate about applying rigorous statistical methods to solve real-world problems in healthcare, education, and AI systems.

Education

Master of Science in Applied Statistics
Rochester Institute of Technology | *Expected 2026*

- Coursework: Machine Learning, Bayesian Statistics, Experimental Design, Deep Learning, NLP
- Focus: Ensemble Methods, Probabilistic Programming, MLOps

Technical Skills

Programming Languages: Python, R, SQL

Machine Learning & AI: - Ensemble Methods: Random Forest, XGBoost, LightGBM, AdaBoost, Gradient Boosting, Stacking, Voting - Deep Learning: PyTorch, TensorFlow, Transformers, BERT - NLP & LLMs: GPT-4, Claude, Llama, LangChain, Hugging Face, Prompt Engineering - Explainable AI (XAI): SHAP, LIME, Model Interpretability

Statistical Analysis: - Bayesian Hierarchical Modeling: PyMC, ArviZ, MCMC Sampling - Classical Statistics: Hypothesis Testing, Regression, ANOVA, Cross-Validation - Inter-Rater Reliability: Krippendorff’s Alpha, Cohen’s Kappa

Data Engineering & MLOps: - Data Processing: Pandas, NumPy, Polars, Scikit-learn - Experiment Tracking: MLflow, Model Registry - API Development: FastAPI, REST APIs - Cloud & Deployment: Docker, CI/CD

Standards Compliance: IEEE 2830-2025 (Transparent ML), ISO/IEC 23894:2025 (AI Risk Management), EU AI Act

---

Research Projects

Ensemble Machine Learning for Breast Cancer Classification

January 2026

Developed and benchmarked eight state-of-the-art ensemble learning algorithms for binary classification of breast tumors using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset.

Key Achievements: - Achieved **99.12% accuracy**, **100% precision**, and **98.59% recall** with AdaBoost classifier, exceeding human inter-observer agreement rates (85-95%) - Implemented comprehensive preprocessing pipeline: VIF analysis for multicollinearity detection, SMOTE for class imbalance, RFE for feature selection (30 → 15 features) - Validated model robustness with 10-fold stratified cross-validation (98.46% ± 1.12%) - Achieved ROC-AUC of 0.9987 demonstrating near-perfect discrimination - Implemented SHAP-based explainability for clinical interpretability and IEEE 2830-2025 compliance

Technologies: Python, Scikit-learn, XGBoost, LightGBM, SMOTE, SHAP, MLflow, FastAPI

LLM Ensemble with Bayesian Hierarchical Modeling for Bias Detection

January 2026

Designed a novel computational framework for detecting and quantifying political bias in educational textbooks using an ensemble of frontier Large Language Models combined with Bayesian hierarchical modeling.

Key Achievements: - Processed **67,500 bias ratings** across **4,500 textbook passages** using GPT-4, Claude-3, and Llama-3 ensemble - Achieved excellent inter-rater reliability (Krippendorff's $\alpha = 0.84$) among LLMs, validating ensemble approach - Demonstrated statistically significant publisher-level bias differences (Friedman $\chi^2 = 42.73$, $p < 0.001$) - Implemented full Bayesian uncertainty quantification with 95% Highest Density Intervals - Built production-ready API processing pipeline with robust error handling and rate limiting

Technologies: Python, PyMC, ArviZ, OpenAI API, Anthropic API, LangChain, MLflow, MCMC Sampling

---

Technical Competencies

Domain	Expertise Level
Ensemble Learning Methods	Expert
Bayesian Statistical Modeling	Advanced
Large Language Models & Prompt Engineering	Advanced
Medical/Clinical ML Applications	Proficient
MLOps & Model Deployment	Proficient
Explainable AI (XAI)	Advanced
Statistical Hypothesis Testing	Expert

Publications & Reports

- **Breast Cancer Classification: Technical Analysis Report** - Comprehensive evaluation of ensemble methods for medical diagnostics
- **LLM Ensemble Textbook Bias Detection: Technical Analysis Report** - Novel framework for scalable content analysis

---

Professional Interests

- **Healthcare AI:** Computer-aided diagnosis, medical imaging analysis, clinical decision support systems
- **Responsible AI:** Fairness auditing, bias detection, model transparency, AI governance
- **NLP Research:** LLM evaluation, prompt engineering, multi-model ensemble systems
- **Statistical Methodology:** Bayesian inference, causal inference, uncertainty quantification

---

References

Available upon request