Erdem Karaköylü

University Park, MD | Citizenship: U.S. Citizen | Clearance: Top Secret

erdemk@protonmail.com | erdemkarakoylu.github.io | LinkedIn

Data Scientist | Bayesian Analysis & Machine Learning

- Data scientist specializing in Bayesian modeling, uncertainty-aware machine learning for decision optimization.
- Experienced in developing and evaluating probabilistic models in high-stakes domains including defense, remote sensing, and environmental prediction.
- Strong track record of applying principled statistics to improve system performance and experimental design.

Core Skills

Bayesian & Statistical Modeling Hierarchical Modeling (Regression & Classification) • Bayesian Additive Regression Trees (BART) • Probabilistic Programming • A/B Testing

Emerging Interests Causal Inference (incl. do-calculus in PyMC) • Bayesian Decision Theory

Technical Stack Python • PyMC • Scikit-learn • XGBoost • PyTorch • Pandas • Git • Matplotlib • Seaborn • Arviz • XArray • SQL

Experience & Research Highlights

Freelance Data Scientist — Marine Remote Sensing and Ecological Forecasting

- Developed Bayesian Additive Regression Tree (BART) and hierarchical models to estimate marine optical properties and chlorophyll concentrations from satellite radiance data (paper in prep).
- Built predictive XGBoost models to infer phytoplankton community structure, outperforming baseline approaches; preprint available at https://doi.org/10.20944/preprints202508.0184.v2.
- Used probabilistic ODE parameter estimation to analyze nonlinear dynamics in marine ecological systems.
- Published reproducible Bayesian modeling workflow guide for remote sensing model developers (preprint available at https://doi.org/10.31223/X54J1J.

Data Scientist — Research Innovations Inc. (Alexandria, VA)

- Contributed to the development of a Retrieval-Augmented Generation (RAG) system that improved information retrieval for military planners.
- Led Bayesian A/B testing to optimize system components and refine model selection for production environments.
- Built and iteratively refined an active-learning image classification pipeline to reduce manual annotation requirements.

- Supported targeted sentiment analysis using fine-tuned large language models for sensitive domains.

Machine Learning Researcher NASA Goddard Ocean Biology Processing Group / SAIC

- Developed Bayesian models to predict satellite-derived ocean color products, improving chlorophyll and particulate property estimates.
- Conducted Monte Carlo simulations to quantify uncertainty and error propagation in remote sensing reflectance (Rrs) data.
- Created climate data visualizations and analysis pipelines supporting scientific reports and satellite mission deliverables.
- Advocated for probabilistic approaches and led internal discussions on Bayesian methods for biogeophysical modeling.

Researcher UC San Diego / Scripps Institution of Oceanography

- Adapted a planar laser-induced fluorescence imaging system to quantify real-time feeding states in individual marine zooplankton.
- Captured high-resolution time series of gut pigment dynamics to infer behavioral state transitions (feeding, digestion, resting).
- Built an individual-based model linking physiological state to vertical foraging behavior under environmental constraints.
- Calibrated imaging measurements against chemical extraction to ensure accuracy and repeatability across individuals.
- Follow-up research conducted on temperature effects on feeding behavior and gut dynamics.

Education

Ph.D. Biological Oceanography Scripps Institution of Oceanography, UC San Diego

B.Sc. Oceanography Florida Institute of Technology

Languages

- English – Native/Trilingual
- French – Native/Trilingual
- Turkish – Native/Trilingual
- Spanish – Advanced