

# Erdem Karaköylü

University Park, MD 20782 | 202-805-5246 | [erdemk@protonmail.com](mailto:erdemk@protonmail.com)  
[linkedin.com/in/erdem-karaköylü-49487455](https://www.linkedin.com/in/erdem-karaköylü-49487455) | <https://erdemkarakoylu.github.io>

## Data Scientist

Data science and machine learning expert specializing in Bayesian modeling, retrieval-augmented generation (RAG), and predictive analytics to enhance decision-making and operational efficiency. Skilled in designing and deploying large language models (LLMs), image classifiers, and Monte Carlo simulations to improve data accuracy and automation. Adept at integrating advanced AI-driven solutions to optimize processes in research, defense, and environmental applications.

## Skills

Machine Learning | Bayesian Modeling | Data Analysis | Data Visualization | Monte Carlo Simulation | Large Language Models (LLMs) | Retrieval Augmented Generation (RAG) | A/B Testing | Climate Data Processing | Satellite Oceanography

## Technical Skills

Git | Matplotlib | Numpy | Pandas | PyMC | Pytest | Pytorch | Scikit-learn | Scipy | Seaborn | SQL | XArray | XGBoost

## Experience

### Freelance Data Scientist

February 2025

Researched and developed predictive models for synoptic observational Oceanography and theoretical Marine Ecology.

- Built **XGBoost models** to predict phytoplankton community composition, improving predictive accuracy compared to baseline methods
- Developed **Bayesian models** to retrieve parameters of Ordinary Differential Equations describing marine ecological processes, providing insights into system dynamics.
- Currently developing a **Bayesian Additive Regression Trees (BART) model** to predict Marine Inherent Optical Properties from satellite data, potentially bypassing atmospheric correction challenges.
- Currently developing **Hierarchical Bayesian Regression models** to predict marine chlorophyll a concentration from satellite data.

### Research Innovations Inc., Alexandria, VA

November 2019 - January 2025

#### Data Scientist

Analyzed complex datasets to develop predictive models and data-driven solutions for DOD and DOJ projects initiatives.

- Developed and packaged a **Retrieval Augmented Generation (RAG)** system to enhance decision-making for military planners, improving information retrieval efficiency.
- Designed a **Bayesian A/B testing** study to identify optimal feature selection for the RAG system, refining its predictive capabilities.
- Built an **image classifier with an active learning data annotation loop**, improving model accuracy and reducing manual labeling efforts.
- Trained and deployed a **Large Language Model** for **Directed Sentiment Analysis**, enhancing sentiment detection in targeted applications.

### NASA Goddard Ocean Biology Processing Group and SAIC, Greenbelt, MD

October 2013 - November 2019

#### Data Analyst and Machine Learning Researcher

Developed machine learning models and analyzed ocean biology data to support NASA's research and satellite data processing.

- Designed Bayesian models to predict satellite oceanography products, improving data accuracy and forecasting reliability.
- Analyzed and visualized climate data for research papers and reports, enabling clearer insights and actionable conclusions.
- Implemented Monte Carlo simulations to quantify uncertainty in satellite data, improving error estimation and model robustness.
- Advocated for Bayesian methodologies, leading knowledge-sharing initiatives to promote probabilistic modeling best practices.

### **Additional Experience**

**Associate Research Scientist** | University of Maryland Center for Environmental Science), Solomons, MD

**Postdoctoral Researcher** | Scripps Institution of Oceanography, University of California, La Jolla, CA

### **Education**

**Doctor of Philosophy (Ph.D.) – Biological Oceanography and Marine Ecology**

Scripps Institution of Oceanography, University of California, La Jolla, CA

**Bachelor of Science (B.Sc.) – Oceanography** | Florida Institute of Technology, Melbourne, FL