Balkan Yildirim

732-779-5455| balkanyildirim@gmail.com

Education

University of Wisconsin-Madison — Madison, Wisconsin
Intended Degrees: Bachelor of Science: Data Science, Mathematics — May 2028
GPA: 3.7/4.00

Relevant Coursework

Data Science Modeling, Machine Learning, Big Data Systems, Linear Algebra, Multivariable Calculus, Multivariate Analysis

Professional Experience

Data Science Intern — June 2026 - Present
University of Wisconsin, Office of Sustainability

- Built a reproducible GHG emissions data pipeline using Python, Jupyter, DuckDB, pandas, and OneDrive/SharePoint to replace manual Excel-based sustainability reporting workflows.
- Validated and loaded 995 raw SIMAP emissions records into DuckDB, preserving source-file metadata, load timestamps, scope values, emissions totals, and data quality checks.
- Created pilot dashboard-ready emissions tables by fiscal year, scope, and source category, incorporating source mappings, supervisor-reviewed exceptions, and exportable CSV/Excel outputs.

Undergraduate Research Assistant — May 2026 - Present
University of Wisconsin-Madison Ecology Laboratory

- Built ML pipelines mapping Landsat satellite spectral embeddings to airborne-derived vegetation traits across 290K+ pixel-year observations, bridging spaceborne and airborne remote sensing data streams
- Benchmarked uncertainty-aware training strategies under leave-one-year-out temporal validation, diagnosing how label-uncertainty weighting shifts model target distributions across 8 biophysical traits
- Tuned gradient-boosted models (XGBoost) with early stopping and per-trait depth optimization, improving normalized prediction error and identifying spectral-signal limits for weakly-observable traits

Undergraduate Research Assistant — October 2025 - April 2026
University of Wisconsin-Madison Physics Laboratory

- Analyze Cosmic Microwave Background (ACT DR6) and DESI galaxy survey maps using HEALPix spherical harmonic methods to probe how matter moves and clusters across the large-scale Universe
- Construct galaxy overdensity fields and compute masked/unmasked angular power spectra and CMB–galaxy cross-correlations to isolate physical cosmological signals from observational effects
- Develop statistical pipelines in Python (NumPy, Healpy, PyMaster) to support reconstruction of cosmic velocity fields and improve understanding of structure formation over cosmic time

Projects

Statistical & Machine Learning Climate Analysis (Independent Research) — January 2026 - February 2026

- Estimated global temperature trends using NASA GISTEMP and NOAA GlobalTemp datasets, applying OLS regression with bootstrap resampling to quantify uncertainty and sensitivity to start-year selection
- Evaluated nonlinear time-series models (Random Forest, Gradient Boosting) using lagged and rolling features; Gradient Boosting reduced NOAA test error from OLS RMSE = 0.565 to 0.496 and MAE = 0.504 to 0.405
- Demonstrated robustness of long-term warming trends across datasets and model classes, with ML-derived implied trends consistent with linear estimates despite modest gains in short-term predictive accuracy

Deforestation Detection from Satellite Imagery — November 2025 - January 2026

- Developed a geospatial pipeline using Sentinel-2 satellite imagery and Google Earth Engine to detect forest loss in the Brazilian Amazon via cloud masking, annual compositing, and NDVI-based change detection.
- Generated patch-level labeled datasets for supervised learning and ongoing work includes training and evaluating machine learning models (logistic regression, random forests, CNNs) and validating results with Global Forest data.

Skills and Interests

Skills: Java, Python, R, SQL, Pandas, TensorFlow, MATLAB, NumPy, Seaborn, Anaconda, Healpy, LoggerPro
Languages: English (Native), Turkish (Fluent), Spanish (Proficient)
Interests: Reading, Movies, Exercising, Triathlons, Cello, Scientific Applications