Can Karakoc

cankarakoc@berkeley.edu | linkedin/can-karakoc | github/can-karakoc

About Me

Recent **UC Berkeley** graduate with a dual B.A. in Computer Science and Data Science, specializing in computational biology methods. Interested in data-driven research at the intersection of biology, machine learning, and human-centered design.

Education

University of California, Berkeley

B.A. in *Computer Science* & B.A. in *Data Science*

Domain Emphasis: Computational Biology Methods

GPA: 3.8/4.0

Technical Skills

Computational Tools

Python, Java, C, SQL, Pandas, PyTorch, TensorFlow, Scikit-learn, Bambi, Spark, HTML/CSS, JavaScript, React.

Bioinformatics Tools:

BioPython, JBrowse, BLAST, AlphaFold.

Data Visualization:

R, Tableau, Power BI, Matplotlib, Seaborn, Plotly.

Leadership

Teaching Assistant

HCD Decal | Aug 2023 - May 2024

- Facilitated a student-led course on human-centered design to 60+ UC Berkeley students.

Industry & Research Experience

Computational Plant Biology Research Assistant

Mishler Lab, UC Berkeley | Jan - July 2022 | Abstract

- Compiled dataset of 1,300+ alien flora species from 6+ databases, including characteristics such as habit, bioregion, and threatened species rating.
- Conducted chi-squared hypothesis testing using R among plant traits and invasiveness profile, and identified 14% of species as invasive.

Data Analysis Intern

Veloxity Inc. | May - Aug 2023

- Engineered predictive features from proprietary mobile footprint data to enhance user segmentation, identifying key behavioral patterns.
- Developed interactive data visualizations (Tableau) to differentiate consumer segments and communicate model insights to stakeholders.
- Refined user segmentation criteria by analyzing proprietary mobile footprint data and optimizing features.

Website Content & Design Manager

UC Berkeley Blum Center | June 2024 - Aug 2025 | Sample Work

- Develop and maintain web pages for the Blum Center and Master's of Development Engineering websites.

Projects

Single-Cell RNA-seq Analysis

Personal Project | Aug - Sept 2025 | GitHub

- Processed and clustered 3,000 PBMC scRNA-seq profiles (from 10X Genomics) with QC, normalization, HVG selection, and Leiden clustering.
- Annotated immune cell types using marker genes and trained classifiers (logistic regression, random forest, neural network) for cell-type prediction.
- Achieved strong accuracy and produced UMAP/confusion matrix visualizations to interpret results.

Lentivirus Gene Search Tool

BIOENG C131 | Aug - Dec 2025 | GitHub | Website | Paper

- Created a browser-based genome viewer (**JBrowse**) for Lentivirus family gene and protein exploration and comparison.
- Integrated gene annotations, synteny views for multiple sequence alignments (**Clustal Omega**), and 3D protein structure visualizations.

Energy Production and Carbon Emissions

DATA C102 | Aug - Dec 2024 | GitHub

Investigated two research questions and presented findings about policy implications for renewable energy adoption and carbon emission reduction strategies:

- Modeled state-level renewable energy production from state characteristics, e.g., size, temperature, political leaning, and electricity prices, using Gaussian GLM, Random Forest, and KNN.
- Modeled the causal relationship between state electricity prices and carbon emissions using Stabilized Inverse Propensity Weighing with bootstrapped propensity scores to estimate an average treatment effect.