Can Karakoc

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

Recent UC Berkeley graduate with dual B.A. in Computer Science and Data Science, specialising in Computational Biology. Experienced in bioinformatics, machine learning, and biological dataset analysis. Seeking Graduate Programme/Research Associate roles in computational biology, bioinformatics, or biomedical data science.

Education

University of California, Berkeley

B.A. in Computer Science and B.A. in Data Science – Domain Emphasis in *Computational Biology Methods*

August 2021 - May 2025

GPA: 3.8/4.0 (First Class Equivalent)

Relevant Coursework: Machine Learning in Computational Biology; Introduction to Computational Molecular and Cell Biology; Designing, Visualizing, and Understanding Deep Neural Networks; Introduction to Machine Learning; Data, Inference & Decisions; Data Engineering; Optimization Models in Engineering; Introduction to Artificial Intelligence; Probability for Data Science; Machine Structures; Data Structures.

Technical Skills

Programming: Python, Java, C, SQL, Spark

Bioinformatics: Biopython, BLAST, JBrowse, AlphaFold

ML & Data: PyTorch, TensorFlow, Scikit-learn, Bambi, Pandas

Visualisation: R, Tableau, Matplotlib, Seaborn, Plotly

Domain knowledge: RNA Sequence Analysis, Protein Structure Modelling, Gene Annotations

Industry & Research Experience

Computational Plant Biology Research Assistant | UC Berkeley Mishler Lab (January - July 2022, Berkeley, CA, USA)

- 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.
- Presented research findings at the 6th North American Congress for Conservation Biology in Reno, NV [Abstract].

Data Analysis Intern | Veloxity Inc. (May - August 2023, Istanbul, Turkey)

A data-driven company that utilizes proprietary mobile footprint data to create user habit definitions and predict future behavior patterns.

- 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 to improve the accuracy of distinct consumer behavior prediction.

Website Content & Design Manager | Blum Center, UC Berkeley (June 2024 - August 2025, Berkeley, CA, USA

- Design and implement web pages for institutional webpages, increasing user engagement for current and prospective users [Sample Work].
- Collaborate with stakeholders to align website design with organizational goals and branding principles.
- Manage website content to ensure optimal functionality, accessibility, and a seamless user experience for both internal and external partners.

Projects

Single-Cell RNA-seq Analysis |GitHub | Persona Projectl | August - September 2025

- 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 | Paper, Static Website & GitHub UC Berkeley BIOENG C131 | August - December 2024

- Created a JBrowse genome browser enabling streamlined analysis of Lentivirus family genes, improving accessibility for cross-protein comparisons and 3D structural insights.
- Integrated gene annotations, synteny views for multiple sequence alignments, and AlphaFold-based 3D protein structure visualizations.

Energy Production and Carbon Emissions | Github | UC Berkeley DATA 102 | August - December 2024

- Investigated two research questions and presented findings about policy implications for renewable energy adoption and carbon emission reduction strategies:
  (1) 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 – achieved the highest likelihood in GLM with 5-fold Cross-Validation.
  (2) Modeled the causal relationship between state electricity prices and carbon emissions using Stabilized Inverse Propensity Weighing with bootstrapped propensity scores – estimated an average treatment effect, which indicated a significant reduction in emissions with higher electricity prices.

Leadership

Teaching Assistant | Human-centered Design DeCal (Fall 2023 - Spring 2023)

- Facilitated a student-led course on human-centered design within the Department of Design Innovation, focusing on design methodologies (user research, synthesis, ideation, prototyping) and design tools like Figma to over 60 UC Berkeley students.

Design Consultant | Berkeley Innovation (Spring, 2023 - Fall 2024)

- Consulted for client projects to enhance user experience, gaining hands-on practice in the Human-centered design process.