Cornelia Ilin

CONTACT
INFORMATION

3416 Plateau Dr
Belmont, CA 94002

(608) 338-4844
cornelia.ilin@berkeley.edu
http://corneliailin.github.io

Citizenship: U.S. permanent resident

PROFILE

I am a Faculty in Data Science at UC Berkeley, School of Information, where I teach Applied Machine Learning (ML) and Capstone classes to MS-level students. My research sits at the intersection of health and the environment. In my research, I combine **ML** with **geospatial** and **causal inference** methods. I draw from more than 13 years of experience across academia, industry, and consulting, with highlights including:

- Highly rated Data Science lecturer. Presentation and writing skills (I attend conferences and write papers for journal publications). Collaboration skills (direct interaction with faculty, postdocs, and graduate students).
- Computer languages include Python, GIS, SQL, Bash shell; version control with Git(Hub)
- Experience with big data analytics: analyzed data using Hadoop/HDFS and Dask frameworks on CloudLab.us clusters and Google Cloud Platform
- Experience with packages such as TensorFlow2, SHAP for explainable AI, Scikit-Learn, Statsmodels, Geopandas, OSMnx, Rasterio
- Experience with IRB applications and working with **HIPAA protected health data**, including claims and electronic health records (e.g., diagnosis, lab values, clinician notes, images).
- Two years of **industry and consulting experience** with contributions to high-profile litigation cases in the healthcare industry

RESEARCH
AFFILIATIONS

UC-Berkeley, Global Policy Lab   2023 - present

- The Aerial History Project: the objective is to convert 1.6 million historical aerial photos taken between the 1940s and 1990s into a data set comparable to modern satellite imagery. The final product will provide input to various downstream tasks, such as understanding how climatic shifts have mobilized populations.
- Data: The aerial photos were collected by the British Directorate of Overseas Surveys while mapping what was formerly known as the British Empire. It covers large parts of Africa, the Far East, and the islands of the Caribbean Sea, Atlantic, and Pacific Oceans.
- Methods: computer vision algorithms for stitching and georeferencing of mosaics, and predictions

UW-Madison, Environmental Research Group   2020 - present

- Project 1: Understand the effects of wildfire exposure in-utero and first years of life on the likelihood of respiratory, cancer, and diabetes conditions in one's emergency room or inpatient visit later in life. A version of this project won the School of Information's **MIDS Hal Varian** award in the Fall of 2022.
- Project 2: Develop Ped-BERT, a state-of-the-art deep learning model that accurately predicts the likelihood of 100+ conditions in a pediatric patient's next medical visit.
- Data: emergency room and inpatient visits, vital statistics, wind and wildfire data

- Methods: record linkage, causal analysis with instrumental variables, bidirectional encoder representations from transformers (BERT)

PROFESSIONAL
EXPERIENCE

UC-Berkeley, School of Information, **MIDS Lecturer**   2021 - present

- DataSci 207: Applied Machine Learning (course coordinator). Topics include linear regression, logistic regression, decision trees, random forests, unsupervised learning (clustering and dimensionality reduction), deep neural networks (FNN, CNN, RNN/LSTM, Transformers), embeddings, transfer learning, model interpretability and fairness
- DataSci 210: Capstone. A project-based course fusing core data science and soft skills learned throughout the MIDS program.

Stanford University, RegLab, **Research Scientist**   2021 - 2022

- Project (in partnership with the EPA): explore how machine learning can be used to protect human health with a focus on environmental justice and health outcomes in California.
- Data: Satellite imagery to detect Concentrated Animal Feeding Operations (CAFOs), atmospheric data (wind patterns from NASA's MERRA-2 product), water pollutant discharge monitoring data (ICIS-NPDES), hospital/ER data (CDPH), census tract data (Geolytics Neighborhood Change Database)
- Methods: computer vision algorithms for time series satellite images classification to detect building construction and expansion, causal inference with instrumental variables

UC-Berkeley, School of Information, **Postdoctoral Fellow**   2020 - 2021

- Advisor: Joshua Blumenstock, Ph.D.
- Project 1: Provide real-time feedback for managing the spread of COVID-19 at local, national, and global scale. Focus on the impact of non-pharmaceutical policies on human mobility, and the usefulness of cellphone data in predicting the spread of the pandemic.
- Project 2 (in partnership with the CDC): Empirically estimate the impact of changes in non-pharmaceutical policy interventions, mobility, and other avoidance behaviors on growth rate of COVID-19 cases, COVID-19 deaths, and economic output for all countries in the world where GDP data is available.
- Data: COVID-19 cases and deaths (John Hopkins CSSE), human mobility data (Google, Facebook, SafeGraph, InfoGroup), non-pharmaceutical policy interventions (CDC and other sources), quarterly GDP data.
- Methods: Causal inference, predictive analysis.

UW-Madison, Department of Applied Economics, **Faculty Associate**   2018 - 2020

- Teaching: Object Oriented Programming and Data Analytics with Python; Practicum for Applied Economists
- Topics: Data types, functions, classes, exceptions, IO files, data visualization, descriptive statistics, causal inference, cloud computing (incl. Bash), GIS with Python.
- Assistant Program Director: M.S. in Quantitive and Applied Economics program.

Analysis Group, Inc., Menlo Park, CA, **Associate Economist**   2017 - 2018

- Litigation consulting: Contributed to several high-profile litigation cases in the healthcare industry (e.g., Des Roches, et al. v. Blue Shield and Magellan).
- Research: Contributed to manuscripts and posters documenting the effectiveness of leptin replacement therapy in treating lipodystrophy.

- Data: Claims (e.g., mental health, substance abuse), quasi-experimental and surveys.
- Methods: Surveys, discrete choice analysis (multinomial logistic and hierarchical Bayesian regressions), matching algorithms, Cox hazard models.

UW-Madison, Department of Applied Economics, **Research Assistant**   2012 - 2017

University of Zürich, Department of Economics, **Research Assistant**   2011

EPFL, Department of Computer Science, **Research Assistant**   2010

DHL European Headquarters, Belgium, **Intern**   2009

EDUCATION		
	UW-Madison, Ph.D. in Applied Economics	2012 - 2017
	University of Lausanne, Switzerland, M.S. in Economics	2009 - 2011
	Academy of Economic Studies of Bucharest, Romania, B.S. in Economics	2004 - 2008

JOURNAL PUBLICATIONS

Global Health and Economic Impacts of Behavior Change During the COVID-19 Pandemic (*under review, Nature*, with J. Tseng, K.C. Coy, A.C. Ewing, T. Chong, S.M. Marks, I. Bolliger, N.M. Gonzalez, K. Bell, A.J. Hakim, S. Hsiang, 2023)

Public Mobility Data Enables COVID-19 Forecasting and Management at Local and Global Scales (*Nature - Scientific Reports, volume 11, article number: 13531*, with S. Annan-Phan, X.H. Tai, S. Mehra, S. Hsiang, J. Blumenstock, 2021).

Competition, Price Dispersion and Capacity Constraints: The Case of the U.S. Corn Seed Industry, *European Review of Agricultural Economics, 2021* (with G. Shi).

in preparation:
Ped-BERT: Early Detection of Disease for Pediatric Care (2022)

The Role of Birth and Contemporaneous Pollution Exposure on Health Outcomes. Evidence from California (with D. Phaneuf, 2020).

MANUSCRIPTS AND POSTERS

Longitudinal Matching. A Method for Generating Comparable Samples of Treatment and Treatment-Naive Patients with Progressive Conditions (Analysis Group Inc., 2018).

Effect of Leptin Replacement Therapy on Survival and Disease Progression in Generalized and Partial Lipodystrophy (study funded by Aegerion Pharmaceuticals Inc., 2018).

Patient Quality of Life and Benefits of Leptin Replacement Therapy in Generalized and Partial Lipodystrophy (study funded by Aegerion Pharmaceuticals Inc., 2018).

OTHER TEACHING EXPERIENCE

UW-Madison:
TA, World Hunger and Malnutrition: Spring 2017
TA, Applied Econometric Analysis I: Fall 2016
TA, Applied Microeconomic Theory: Fall 2014
Lecturer, Math Camp for Incoming M.S. and Ph.D. Students: Summer 2014

FELLOWSHIPS, SCHOLARSHIPS AND GRANTS

Science of ADRD Workshop (competitive), University of Southern California, 2022
Research Grant, American Bar Association, Section of Antitrust Law, 2016

Ph.D. Summer Program (competitive), Edgeworth Economics, Washington, DC, 2016
Kenneth and Pauline Parsons Graduate Fellowship Fund, UW-Madison, 2016
Best Paper Presentation Award, UW-Madison, 2016
SASC Graduate Funds, University of Lausanne, 2010 - 2011
Hessen Summer School (competitive), Goethe University of Frankfurt am Main, Germany, 2008
WU Summer School (competitive), Vienna University of Economics and Business, Austria, 2007
Excellency in Research Award, Academy of Economic Studies of Bucharest, Romania, 2007

SEMINAR AND CONFERENCE PRESENTATIONS

Stanford Maternal and Child Health Research Institute Symposium, 2021, 2022
Association of Environmental and Resource Economics, 2020
UW-Madison, Healthcare Group seminar, 2016, 2019
University of Connecticut, 2017
European Association for Research in IO (Rising Stars section), Lisbon, Portugal, 2016
AAEA Meetings, Boston, Massachusetts, 2016

PROFESSIONAL ACTIVITIES

Reviewer for Nature Scientific Reports, 2022 - present
Reviewer for the American Public Health Association (APHA), 2019 - present
Social Chair, THC Club of AAE Department, UW-Madison, 2015 - 2016
Seminar Organizer, THC Club of AAE Department, UW-Madison, 2014 - 2015

LANGUAGE SKILLS

Romanian (native), English (fluent), French (basic)