Lan Dinh

(510)759-5926 github.com/lannd3217 linkedin.com/in/lan-dinh3217 ngocld32@berkeley.edu

Education

Georgia Institute of Technology Expected December 2027
Master of Science in Computer Science, Machine Learning Specialization *Atlanta, GA*

University of California, Berkeley December 2024
Bachelor of Arts in Data Science, Concentrated in Business and Industrial Analysis *Berkeley, CA*

- **GPA:** 3.86
- **Coursework:** Advanced Business Analysis, Inference and Decisions, Advanced Statistical Inference Seminar, Introduction to Machine Learning, Deep Neural Network, Natural Language Processing, Optimization Models in Engineering, Principles and Techniques of Data Science, Probability for Data Science.

Technical Skills

Languages: Python, SQL.
Frameworks & Libraries: Pandas, NumPy, Matplotlib, Scipy, Seaborn, Pytorch, Sklearn, Scikit-learn.
Tools: AWS, Tableau, Excel, Jupyter Notebook, Google Colab, Github.
Statistical Models/ Machine Learning Models: Linear Regression, Logistic Regression, Decision Trees, K-Nearest Neighbors.

Experiences

Rubicon Bakers LLC March 2025 – Present
Data Analyst *Richmond, CA*

- Develop **Tableau dashboards** to standardize sales data from 5 distributor portals, enabling **real-time product performance tracking** and **reducing manual reporting by 40%** to support data-driven sales strategies
- Clean and standardize **100,000+ rows of sales data**, improving reporting accuracy and **identifying underperforming products** to optimize promotions and inventory planning.

UC Berkeley Data Discovery –Oakland Natives Give Back Fund Inc. September 2024 – Present
Data Science Intern *Berkeley, CA*

- Engineer and optimize data pipelines by **cleaning, standardizing, and integrating diverse datasets** from Oakland Unified School District, ensuring high-quality, actionable data for predictive analysis.
- Design and deploy **Random Forest model** with **95% accuracy** by using **Sklearn, AWS** to predict students at risk of chronic absenteeism, driving data-driven insights and delivering a comprehensive report to inform strategic interventions.

UC Berkeley Data Science Undergraduate Studies January 2023 – December 2024
Data Science Modules Developer *Berkeley, CA*

- Developed data science assignments for **6 diverse subjects**, **benefiting 100+ students** by collecting, processing, and handling unstructured datasets using **Python and Pandas**.
- Designed **20+ Jupyter notebooks** containing **40+ interactive visualizations** using **matplotlib and seaborn**; conducted **exploratory data analysis (EDA)** to uncover data trends, informing data-driven decisions.

Robotics Cats, CITRIS And The Banatao Institute June 2023 – August 2023
Machine Learning Intern *Berkeley, CA*

- Trained a predictive model **over 500 images** that detected smoke, aiding nearby forest residents in **mitigating risks early and in a timely manner.**
- Used toolbox based on Pytorch, resulting in an approximately **45% increase in mIoU** as an accuracy measurement; laid groundwork for further development by subsequent data scientist.

Research & Teaching

UC Berkeley College of Computing, Data Science, and Society July 2024 – August 2024
Data Science Seminar Facilitator *Berkeley, CA*

- Guided group research project that analyzed educational disparities using Python, Pandas, and Google Colab, **identifying key factors** influencing African American college enrollment and **providing actionable insights** for educational equity.
- Conducted a thorough literature review and pulled **total of 4 datasets from the US Census and the National Center for Education Statistics.**

UC Berkeley College of Computing, Data Science, and Society August 2023 – December 2023
Data Tutor *Berkeley, CA*

- Led a weekly 2-hour lab for Foundations of Data Science course that provides academic support for **50+ students.**
- Course Content: Python Programming, Data Analysis, Data Visualization, Distributions, Probability, Designing Experiments, Hypothesis Testing, A/B Testing, Sampling, Regression, Classification.

Projects

Predicting Housing Prices in Cook County | *Python, Pandas, Sklearn, Seaborn*

- Built a **Linear Regression machine learning model** to predict housing prices using **500,000+ Cook County, Illinois records** that potentially assisted real estate stakeholders with data-driven insights.
- Achieved 200,000 RMSE on out-of-sample data, which translates to a **44% accuracy increase** over a random model.