

Ngoc (Alice) Hua

DATA SCIENTIST

626.677.1998



alicehua11@berkeley.edu



linkedin.com/in/alicehua/



github.com/alicehua11



alicehuacal.com



SKILLS

Python / R / SQL / AWS / PySpark /
Docker / Kubernetes / Databricks / Git /
GCP / GIS / ArcGIS / Docker / Tableau /
HTML / CSS / Carto / Mapbox

EDUCATION

Master of Information & Data Science
UC Berkeley
Anticipated 2021
GPA: 4.0

Bachelor of Arts, Geography
Magna Cum Laude, Phi Beta Kappa
UC Berkeley, 2020

Associate of Arts,
Social Sciences & Social Behavior
Certificate, Automotive Technology
Rio Hondo College, 2018
GPA: 3.8

SELECTED COURSEWORK

Applied Machine Learning
Statistics for Data Science
Fundamental of Data Engineering
Machine Learning at Scale
Research Design
Privacy Engineering
Experiments and Causal Inference
Deep Learning in the Cloud and at the Edge

CURRENT COURSEWORK

Capstone

LANGUAGES

English & Vietnamese

OTHER INFO

Aikido & Olympic weightlifting
Geographic Information System (GIS)
enthusiast

PROJECTS

Wildlife Object Detection for Conservation

2021

- Worked with WildTrack NGO to develop a POC and pipeline for wildlife detection using their raw drone footages. Trained YOLOv5 model using AWS P3dn.24xlarge instance with 96vCPUs and 8 GPUs on a Pytorch container
- Performed quality control on labeling, used various image augmentation techniques including GAN and a novel tiling solution for small object detection problem
- Deployed model on Jetson NX with Docker container on a drone video stream and sent detected frames with >50% confidence score to Flask webapp via MQTT to avoid false positives and missed detections

Flight Delay Prediction for ML at Scale

2021

- Built an end-to-end pipeline for binary classification with imbalanced dataset of air traffic and weather data. Pipeline ran on Databricks with r4.xlarge cluster
- Performed ETL, established baseline model with Logistic Regression, validated final model using Gradient Boosted Trees, performed timeseries k-fold cross validation and grid search CV to find best parameters and avoid data leakage problems

Regression Study of the Spread of Covid-19

2020

- Used classical linear model of OLS to estimate the causal relationship between adults from 19 to 34 and the number of Covid-19 cases in 50 US States
- Assessed the assumptions, estimated coefficients and confidence intervals. Established magnitude and direction to show that more young adults indeed result in higher Covid-19 cases

Machine Learning for Online News Prediction

2020

- Scaled out Python webscraping for raw data of 2020 news articles from Forbes
- Ran multiple ML algorithms with reproduced features and accessed the errors in predicting a continuous outcome

PySpark for Business Intelligence

2020

- Used Hadoop and Spark to build a data pipeline, from Docker cluster, consume messages from Kafka to Spark for data transformation (i.e. using flatMap() to unroll a nested json) and used PySpark SQL for querying

Descriptive Analysis: Bird Strikes & Vertipoint Analysis

2019

- Analyzed bird strikes trends from 1990 – 2018 against airline flights, temperature & migration datasets using R and LIDAR data
- Produced suitability analysis for vertipoint locations in using ArcGIS for the Civil and Environment Department at UC Berkeley

HIGHLIGHTED EXPERIENCE

FoodWare

Berkeley, CA | 2021

Software Engineer Intern

- Work on landing page. Tech stacks: JavaScript, React, Gatsby.js, Tailwind CSS, Node

CrowdStrike

Sunnyvale, CA | 2021

Data Engineer Intern

- Work on big data machine learning pipeline and infrastructure with senior data engineers

UC Berkeley School of Chemistry

Berkeley, CA | 2020

Data Analyst

- Discovered and collected large amounts of employment data of alumni using LinkedIn Sales Navigator to better engage UC Berkeley alumni and identify new donor prospects
- Analyzed IPOs, recent mergers, companies with unprecedented growth in recent years

UC Berkeley Urban Displacement Project

Berkeley, CA | 2018

Undergraduate Research Assistant

- Used GIS and data science skills to produce rigorous analysis on the nature of gentrification
- Developed interactive, online maps and website content from the results of data analysis
- Created Python course materials for Urban Data Analytics course on spatial/network analysis
- Worked with multiple researchers from various Universities. Example collaboration could be seen at <https://www.urbandisplacement.org/austin/austin-gentrification-and-displacement>