Zhihao (Johnson) Du

zhihao617@berkeley.edu | +1 (510) 833-4417

Links

Github

github.com/JohnsonJDDJ

Personal Website

zhihao.myxd.place
(for info on all projects)

Education

Bachelor of Arts
Statistics
UCBerkeley | 05/2023

Bachelor of Arts
Computer Science
UCBerkeley | 05/2023

Skills

Languages

Python (since 2017)
SQL (since 2019)
Java (since 2020)
R (since 2020)
C (since 2021)
HTML, CSS (since 2017)

Coursework

Machine Learning
Database Systems
Reproducible Data Science
General Linear Models
Data Structures
Machine Structures
Linear Algebra

Tools

Git (since 2020)
Jupyter (since 2020)
Microsoft Azure (since 2022)
DBeaver (Summer 2021)
Command Line
Microsoft Office

Experience

ETL Engineer Intern
DataCVG Shanghai | 05/2021 - 08/2021

Helped the client "FosunPharma" manage their database system with pharmaceutical data by performing extract-transform-load (ETL) on relational databases. The project involved the following processes:

- Designed the architecture for the target relational database.
- Wrote queries to combine tables from two relational data sources.
- Debugged architecture failures through long diagnostic process.

Research

Python ML Engineer
Project AEI - Koer A.I., Inc. | 01/2022 - Now

Responsible for the algorithm behind emotion discernment and early warning system for police aggression. Built a parallel CNN transformer using pytorch to discern emotion from human voice. The steps are:

- Trained the model using large emotional databases (RAVDESS, SAVEE).
- Preprocessed data through augmentation using librosa and pytorch.
- Training process carried through Azure ML cloud compute platform.

Projects

ZileanLeague
Machine Learning in Python | 06/2022 - Now

Predicted League of Legends match outcomes with game statistics before the 16 minute mark. Tuned a random forest classifier and a XGBoost classifier.

zilean
Python Package Development | 05/2022 - Now

Developed a python package "zilean" that can extract and process complex objects from the Riot Games API. It essentially builds a bridge between the API and the traditional python data science environment. Installable via pip, with full documentation and test coverage.

Forest Fire Prediction
Reproducible Data Science in Python | 04/2022 - 05/2022

A forest fire prediction project with a robust and consistent reproduciblility framework. Tools and technologies include: Makefile, myBinder, Jupyter Book, Github Pages, Github Actions, and unit tests for an installable python package.

HOYO Lab
Data Modelling in R | 03/2021 - 08/2021

Predicted Genshin Impact's damage calculation algorithm by building a basic linear model, then displayed the result using R ShinyApp. Data collected from in-game simulations. Used the model to wrote game tutorials receiving 400,000+ views.