Zhihao (Johnson) Du
+1 (510) 833-4417
zhihao617@berkeley.edu

Education

University of California, Berkeley	Current GPA: 3.7/4.0	*Fall 2019 - Now*

- Statistics, BA – Modern Statistical Prediction and Machine Learning; Linear Modelling: Theory and Applications; Concepts in Computing with Data.
- Computer Science, BA – Machine Structures (A+); Discrete Mathematics and Probability Theory; Data Structures and Programming Methodology; Designing Information Devices and Systems II.

YKPao School, Shanghai	Final Grade: 40(+2)/42(+3)	*Fall 2007 – Summer 2019*

- International Baccalaureate bilingual diploma – Physics HL, Mathematics HL

Personal Experience

Beijing PM2.5/PM10 Prediction	Project Leader	*Nov 2021 – Dec 2021*

- Lead and coordinated group of 3, aimed to predict the real-time measurement of PM2.5/PM10in Beijing using imported data (30,000 rows, 12 features) from Beijing Environmental Monitoring Center between 2013 and 2017;
- We performed robust EDA through visualizing univariate distribution, bivariate relation, and interaction significance; employed AIC with backward stepwise selection for feature engineering; arrived with two models that was carefully diagnosed for model assumption and outliers.
- We tested our models, which achieved final RMSE of 0.6, and constructed prediction intervals. We presented the result in a report with detailed description and diagrams.

HOYO Lab	Project Leader	*March 2021 – Aug 2021*

- Self-initiated, interest-driven, data-centered project on the RPG game Genshin Impact:
- Subproject I "Damage Calculator": Predicted the game's underlying damage mechanism using linear regression with stochastic regressors and feature mapping (16 features); collected data through 500+ in-game simulations, and initialized the model with naïve feature map functions and finalized the feature map function through cross validation; achieved prediction error close to 0. Displayed the final model through web app using the shinyapp package in R [1], then utilized the model and published tutorial articles that received 400,000 views and 50,000+ likes.
- Subproject II "Celestial Database": Aimed to build an in-game character power rank based on character usage in combat challenge. With a group of 3, crawled 1000+ user profile in game associated mobile app using Python requests, and overcame the protection barrier using virtual headers. Decoded the fetched JSON files to extract data of interest, then stored and updated data in MySQL database for every 3 weeks; showcased final power rank on website [2].

DataCVG Shanghai	Intern	*May 2021 – Aug 2021*

- Performed extract-transform-load (ETL) using Kettle and MySQL; combine outdated spreadsheets from two data sources (300+ spreadsheets) onto an updated data source.
- Constructed the architecture of destination spreadsheets by handpicking 3~5 primary keys along with necessary metadata columns; reengineered incompatible columns through grouping rows, pivoting columns, and modifying column datatypes; loaded the final results through merging transformed spreadsheets by matching primary keys; debugged the final data through a series of rigorous diagnostics for any architectural failures.
- The ETL project had helped the client, Fosun Pharma, to upgrade onto the newest, higher performing database system.

Skills

High level programming languages: R, C, Python, Java; **Web development related language**: HTML, CSS, JavaScript; **Assembly language**: RISC-V; **Query language**: MySQL, SQL Server

Award and Honors

Anna Sohmen Pao Award: Awarded for being the top 3 students in class of 2019 (high school)

Hobbies

Piano – Performed on high school graduation ceremony; **Choir** – Participated in ISCMS XII and ISCMS XIII (International Schools Choral Music Society); **Violin –** High school orchestra 2nd violinist; **Soccer** –high school varsity team left defensive midfielder.

[1]: https://zhihao617.shinyapps.io/genshin_basic_damage_calculator
[2]: https://www.youngmoe.com