**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 – *STAT 154:* Modern Statistical Prediction and Machine Learning; *STAT 159:* Reproducible and Collaborative Statistical Data Science; *STAT 151A*: Linear Modelling: Theory and Applications; etc.
* Computer Science, BA – *CS 186:* Introduction to Database Systems; *CS 61C:* Machine Structures (A+); *CS 61B:* Data Structures and Programming Methodology; *DATA 100:* Principles & Techniques of Data Science; etc.

|  |  |  |
| --- | --- | --- |
| YKPao School, Shanghai, China | IB: 40/42 | *Fall 2007 – Spring 2019* |

**Professional Experience**

|  |  |  |
| --- | --- | --- |
| **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 ►Combine outdated spreadsheets from two data sources (300+ spreadsheets) onto an updated data source.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.

**Personal Experience**

|  |  |  |
| --- | --- | --- |
| **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”: Built 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. ►Showcased final power rank on website [2].

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

* Predicted the real-time measurement of air quality index PM2.5/PM10 in Beijing. ►Imported data (30,000 rows, 12 features) from Beijing Environmental Monitoring Center between 2013 and 2017.
* 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.
* Tested models, achieving final *RMSE* of 0.6, and constructed *prediction intervals*. ►Presented the result in a report with detailed description and diagrams.

**Skills**

**High level programming language**: Java (proficient), Python (proficient), R (proficient), C.

**Query language**: MySQL (proficient), SQL Server.

**Front end development language**: HTML, CSS, JavaScript.

**Tools**: AzureML, Git (proficient), Shell (proficient), Jupyter (proficient), Kettle, MS Office.

**Language Proficiency**:GRE: 334/340 (2022), TOEFL: 114/120 (2019).

**Awards 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.