

Education

| | | |
|--|----------------------|--------------------------------|
| University of California, Berkeley | Current GPA: 3.7/4.0 | <i>Fall 2019 - Now</i> |
| <ul style="list-style-type: none">Statistics, BA – <i>STAT 154</i>: Modern Statistical Prediction and Machine Learning; <i>STAT 159</i>: Reproducible and Collaborative Statistical Data Science; <i>STAT 151A</i>: Linear Modelling: Theory and Applications; etc.Computer Science, BA – <i>CS 186</i>: Introduction to Database Systems; <i>CS 61C</i>: Machine Structures (A+); <i>CS 61B</i>: Data Structures and Programming Methodology; <i>DATA 100</i>: Principles & Techniques of Data Science; etc. | | |
| YKPao School, Shanghai, China | IB: 40/42 | <i>Fall 2007 – Spring 2019</i> |

Professional Experience

| | | |
|---|--------|----------------------------|
| DataCVG Shanghai | Intern | <i>May 2021 – Aug 2021</i> |
| <ul style="list-style-type: none">Performed extract-transform-load (ETL) using <i>Kettle</i> and <i>MySQL</i>. ► Combine outdated spreadsheets from two data sources (300+ spreadsheets) onto an updated data source.Constructed the architecture of destination spreadsheets by handpicking 3~5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 | <i>March 2021 – Aug 2021</i> |
| <ul style="list-style-type: none">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 <i>linear regression</i> with <i>stochastic regressors</i> and <i>feature mapping</i> (16 features). ► Collected data through 500+ in-game simulations, and initialized the model with naïve <i>feature map functions</i> and finalized the feature map function through <i>cross validation</i>. ► Achieved <i>prediction error</i> close to 0. ► Displayed the final model through web app using the <i>shinyapp</i> 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 <i>Python requests</i>, and overcame the protection barrier using <i>virtual headers</i>. ► Decoded the fetched <i>JSON</i> files to extract data of interest, then stored and updated data in <i>MySQL</i> database. ► Showcased final power rank on website^[2]. | | |
| Beijing PM2.5/PM10 Prediction | Project Leader | <i>Nov 2021 – Dec 2021</i> |
| <ul style="list-style-type: none">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 <i>EDA</i> through visualizing <i>univariate distribution</i>, <i>bivariate relation</i>, and <i>interaction significance</i>. ► Employed <i>AIC</i> with <i>backward stepwise selection</i> for <i>feature engineering</i>. ► Arrived with two models that was carefully diagnosed for model assumption and outliers.Tested models, achieving final <i>RMSE</i> of 0.6, and constructed <i>prediction intervals</i>. ► 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.

[1]: https://zhihao617.shinyapps.io/genshin_basic_damage_calculator

[2]: <https://www.youngmoe.com>