

Xiaoyi(Alleria) Gu

323-716-8103 | xiaoyigu@usc.edu | <https://www.linkedin.com/in/alleria1809>

EDUCATION

University of Southern California, Los Angeles, CA, U.S

Jan 2022 - Dec 2023

Master of Science, Applied Data Science

3.90/4.0

Courses: Data Management, Machine Learning, Analysis of Algorithms, Knowledge Graph

Beijing Foreign Studies University, Beijing

Sep 2017 - Jun 2021

Bachelor of Management, Information Management and Information Systems

3.97/4.0

Courses: Data Mining, Statistics, Python Analytics, Database, Operations Research

Awards: National Scholarship, Merit Student, 2021 Beijing Outstanding Graduates

TECHNICAL SKILLS

- Python, PostgreSQL, MongoDB, MySQL, PySpark, MapReduce, Hadoop, Jupyter Notebook, HiveSQL, Linux OS, AWS

INTERSHIP EXPERIENCE

Data Scientist & Machine Learning Research Volunteer, HireBeat Inc., NY, U.S

May 2022 – Jul 2022

- Led the project to **construct a data dashboard** visualizing data metrics. Built the dashboard UI connected with PostgreSQL using Python and Streamlit, which empowered the users to screen and interact with the database with specific filters, reducing the manual cost by 30%.
- Built a pipeline for **classifying the candidates' resumes** to the job categories they applied for. Extracted resume text features with the **NLP(TF-IDF) method** and constructed classification models with KNN, SVM and Neural Network. Used the models to recommend targeted job categories to the applicants and improved the job application process.
- Performed a **clustering analysis** based on user profiles to compare premium and free trial account features and used the model for marketing promotion.

Data Analyst Intern, Innovation Institute, Beijing Qingbo Big Data Technology Co., Ltd., China

Jun 2021 - Sep 2021

- Constructed **Python Selenium crawlers** to obtain over 1,000 public opinion data from social media platforms based on hot issues, applied text mining methods to calculate word frequency and post sentiments, and visualized the results.
- Created the **market performance visualization** of new energy vehicle brands, based on the sales volume and share of voice data of 7 domestic and foreign new energy vehicles such as Tesla and Weilai.

Speech Recognition Intern, AI Group, iHuman Inc., Perfect World, China

Apr 2020 - May 2021

- Led the audio annotation and screened raw data input. Constructed a pipeline to **query records and aggregate large-scale data using SQL**, download audios from the server based on the records and update the database(Linux).
- Carried out **training set and test set analysis** with Python and visualized the results to improve software performance.

RESEARCH PROJECTS

University Information Web Portal | Team Course Work

Jan 2022 - May 2022

- Collaborated with team members to **establish a Website** visualizing reviews, rankings, safety index, traffic convenience and climate of different universities using **Python Flask**.
- Built Selenium crawlers to download data from multiple sources and connected the database to Google Firebase.
- Used the web portal to facilitate both domestic and international students to select the right study programs and reduce the applicants' search costs.

Review Sentiment Classifier | Independent Course Work

May 2022

- Built a **review sentiment classifier** to classify movie reviews into positive and negative sentiments. Cleaned the text corpus, tokenized sentences and applied **word embedding** as the text feature extractor.
- Trained different neural network models with Keras and compared the performances. Conducted parameter tuning which improved the performance by 10% compared to the baseline model.

Interdisciplinary Research in Genetics and Management

Dec 2020 – Dec 2021

- Processed more than 500,000 pieces of genetic feature data with **Python Pandas**, manually screened more than 70,000 sample labels, and applied **statistical regression and intermediary analysis** to build interdisciplinary models on individuals' genetic endowments and leadership.

- Performed **A/B test** to analyze how genetic endowments of learning ability, psychological status and health status would influence individuals' potential to be future leaders.
- Completed a paper winning Beijing Foreign Studies University Excellent Graduation Thesis for Undergraduates Award.

SARS-CoV2 Waste-Water Surveillance

Jul 2020 - Dec 2021

- Collaborated with a UC Berkeley team to build a website for waste-water surveillance. Independently integrated Python Flask framework, R script and C script, constructed a genetic analysis website demo and deployed it to the Ubuntu server, which enabled researchers to efficiently upload files and receive results in emails.