

Wenqian Zhao

+1(858)-888-6970 wez020@ucsd.edu

San Diego

<http://www.linkedin.com/in/wez020> <https://github.com/wenqian-zhao>

EDUCATION

University of California, San Diego 2019.09 - 2023.06

Data Science Bachelor La Jolla, CA

- GPA: 3.9 / 4.0
- Marshall College Honor Program

Saint George's School 2015.09 - 2019.06

High School Spokane, WA

- GPA: 3.86/4.0
- IB Diploma Programme candidates

SKILLS

- **Programming:** Python(PyTorch, Numpy, etc.), JavaScript(D3.js, G6.js), Java, Bash
- **Languages:** Chinese(Mandarin), English
- **Other:** Docker, Git, Linux, Latex

WORK EXPERIENCE

ByteDance Ltd. 2022.06 - 2022.09

Content Strategy Data Analyst Intern TikTok Music Shanghai, China

- Build up preprocessing and visualization pipeline for diffusion networks over 100k edges with SQL, Python, and JavaScript.
- Utilize web-scraper to collect musician account data and construct a natural language classification model on musician detection based on content creators' biography and video descriptions with domain expertise on music, NLP, and parametric/non-parametric models.

Inspur Group 2021.06 - 2021.09

Product Manager Cloud-Collab Jinan, Shandong, China

- Conduct a competitive study with coworkers in order to discover the customer's demands based on user satisfaction statistics and research.
- Make prototypes of the software's new features by Axure10 and communicate with back-end engineer.
- With back-end engineers and front-end engineers to investigate and reproduce user-reported bugs and assist them in fixing the bugs.

PROJECT EXPERIENCE

Weibo's Political Sentiment Reactions to China's New COVID-19 Policies 2022.11 - Present

Data Scientist

- Using existing Weibo (much like Chinese Twitter) posts scraper to gather people's posts regarding the new COVID-19 policies.
- Perform sentimental analysis and output a customized index that evaluates public opinion on the latest "Zero COVID" policy change.

Heterogeneous Casual Relationship with Honest Forest 2022.09 - Present

Data Analyst

- Utilized R and causalTree with an honest approach to determine the association between educational attainment and income level.
- Perform the procedure on gathered real-world data and simulated data in high-dimensional covariate space to determine the robustness of the model and the significance of its underlying assumptions.
- Utilized Docker Hub to save the custom image for future research into this topic.

User Intent Classification on Amazon Massive Dataset 2022.11 - Present

Data Scientist

- Utilized the pre-trained BERT to predict users' intent based on the Amazon Massive Intent dataset.

- Explored Transformer and PyTorch API to arrange customized dataloader, optimizers, and fine-tune the BERT with warmed-up cosine learning rate scheduler and reinitiated layers.
- Experimented with SupCon loss and compared its performance on an unsupervised clustering challenge to that of cross-entropy loss.
- Arrange all computations on the cloud.

Image-Caption Generator

2022.11 - 2022.12

Data Scientist

- Constructed an encoder-decoder architecture network with pre-trained ResNet-50 and LSTM based on PyTorch.
- Arrange the model training and predicting process on remote Linux machines through SSH.
- Trained, tuned, and evaluated the network with CIFAR-10 dataset and reached 0.7 BLEU-1 scores.

New York Airbnb Price Prediction

2022.01 - 2022.02

Data Analyst

La Jolla

- Cleansed, imputed, and engineered both numerical and text features in the 2019 NYC Airbnb price dataset.
- Fitted and fine-tuned CatBoost, LightGBM, and XgBoost through 5-fold cross-validation; stacked the results with fine-tuned weights.
- Awarded fourth place in the contest and presented my pipeline.

DataHacks2021 Champion

2021.04 - 2021.04

Data Analyst

- Served as team leader to join the UCSD DataHack and won first place in the contest with the highest classification accuracy.
- Given the cryptocurrency transaction dataset, perform data preprocessing including data imputation, over/downsampling, and dimension reduction by PCA.
- Fit and tune various models based on Sklearn's API and used a tuned random classifier as the final model.

Kaggle Competition - House Price Prediction

2020.08 - 2020.09

Data Analyst

- Cleansed the data, ran data imputation using a number of approaches such as polynomial fit, eliminated outliers that may have hampered the performance of my model, and performed dimensionality reduction using PCA.
- Fitted and tuned Xgboost regressor, and LightGBM regressor and achieved a final MSE of 0.126, which ranked top 30% in the contest.

OTHER EXPERIENCE

Triple C

2021.06 - Present

Data Engineer Kaggle Competition Department

San Diego

Participate in various Kaggle competitions that aim to promote data sensitivity and the ability of data processing.

Responsible for data cleaning, analysis, and visualization processes.