Taolan Jiaohaer, M.S.

Hands-on experience in machine learning with Python and data analysis with SQL

(858) 568-1191 | taolanjiaohaer@gmail.com | La JoIla, CA

EDUCATION

University of California, San Diego | San Diego, CA *Aug.2022-Jun.2024*
M.S. in Computational Social Science, GPA: 3.7/4.0

University of California, Berkeley | Berkeley, CA *Jan.2021-May.2021*
Exchange Undergraduate Student

Nanjing University | Nanjing, China *Sep.2018-Jun.2022*
B.A. in Political Science

SKILLS

Data Science Libraries: NumPy, Pandas, Scikit-Learn, TensorFlow, Pytorch, KenLM, OpenFST, Keras
Statistical/Machine Learning Models: Generalized Linear Model, Time Series Analysis, Logistic Regression, Decision Tree, Random Forest, K-means, Gradient Boosting, Support Vector Machine, Neural Network,
Tools: Python, SQL, R, Tableau, Stata, Bash, Git, QGIS, ArcGIS, Visual Studio
Certifications: IBM Databases and SQL for Data Science with Python.

PROFESSIONAL EXPERIENCE

Database Administrator and Visualization Specialist | MACRO Lab UCSD *Sep. 2023-Present*
• Creating, designing, and developing the **MySQL** database server for UCSD Macro Lab; using **PyTorch** to perform statistical analysis of the media data in the database.
• Building, monitoring, and maintaining **Tableau dashboards**. Making data analysis and visualization dashboards using lab data to conduct research on consolidation and financialization research in the media industry. Visualization Example.

Computational Linguist AI Internship | Sensory Inc *Oct.2022-May.2023*
• Using **Python** to build the Natural Language Processing(**NLP**) speech recognition architecture to transcribe spoken language into text using FST toolkits(OpenFST and KenLM).
•Deploying and Implementing Weighted Finite State Transducer(wFSTs) system to model phonetic patterns and language rules to accurately convert spoken utterances into written form with **Python**.

PROJECTS

Media Consolidation and Financialization Research Project | MACRO Lab UCSD *Spring 2024*
• Collecting, pre-processing, and cleaning the data from LSEG workspace as well as lab confidential data; converting data from Google Sheets format to **MySQL** databases; Conducting statistical analysis with the data including correlation, regression with machine learning models such as linear regression, non-linear regression with neural network with **Python** and **Tableau**. • Project page

Predicting Mice Behavior from Neural Activity Using Encoder-Decoder Transformer. *Spring 2023*
• Using **PyTorch** to model and predict the behavior of mice in a free-roam environment using a transformer architecture. Three models were tested, each to perform binary classification, multi-class classification, and sequence-to-sequence translation on neural activity and behavioral data collected from mice performing lever-held-down tasks.

Diabetes Prediction Using Machine Learning Algorithms *Winter 2023*
• Train multiple machine learning algorithms such as logistic regression, SVM, and decision trees to perform diabetes prediction using handcrafted features based on people's characteristics (e.g. age, gender, etc.) using **PyTorch** and **Skikit-Learn**. • Project page

Wave Height Prediction with Dimensionality Reduction *Fall 2022*
•Using **PyTorch** and publicly available data from NOAA to predict potential surf zone wave height in Hawaii North Shoreline with an Ensemble of Neural Networks and Regression algorithms.