

PEIYUAN LEE

peiyuanlee@g.ucla.edu • github.com/peiyuanlee • +1 (408) 296-9996 • linkedin.com/peiyuanlee

EDUCATION

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Los Angeles, CA

Bachelor of Science, Statistics & Data Science, Minor in Data Science Engineering

Expected June 2025

- **GPA/ Major GPA:** 3.70/ 3.85
- **Awards:** Dean's Honors List, ASA DataFest Finalist
- **Coursework:** Neural Networks, Machine Learning, Linear Algebra, Data Mining, Experimental Design and Analysis, Database Systems, Algorithms and Complexities, Probability Theory, Bayesian Statistics, Causal Inference

EXPERIENCE

Data Science Intern

Kuala Lumpur, Malaysia

Pantas Softwares

Jul 2023 – Aug 2023

- Developed and integrated Optical Character Recognition (OCR) feature within company's website using **Django, AWS S3, JavaScript**, and **Bootstrap**
- Achieved ~98% OCR accuracy by leveraging **AWS Services** and iteratively testing and fine-tuning the OCR algorithm
- Collaborated in **code reviews** and coordinated with cross-functional teams to design, test, and deploy new features

Data Engineering Assistant

Los Angeles, CA

Backus Lab at UCLA

Jun 2023 – Present

- Built a comprehensive website for protein database lookup and merging using **Django, Javascript** and **HTML&CSS**
- Streamlined deployment with **Docker** for reproducibility and consistency and **integrated RESTful APIs** to enable seamless communication between the front-end and back-end of the protein database website
- Automated database updates with optimized **Python** scripts, improving efficiency and data reliability for research needs
- Contributed and submitted manuscript to Nature Computational Science (reference: *NATCOMPUTSCI-24-2940*)

NLP Research Assistant

Los Angeles, CA

Amazon X UCLA

Nov 2022 – Mar 2023

- Investigated the computational linguistics of short stories using **LLMs** under Amazon Researcher Haw Shiuan Chang
- Implemented heuristics to through **sentiment analysis** and **topic modeling** using **Gensim, GPT-3** and **Spacy** libraries
- Led NLP tasks using **HuggingFace** and fine-tuned pre-trained BERT model using **PyTorch** and **Trainer API**

Data Analytics Intern

Kuala Lumpur, Malaysia

Securities Commission Malaysia

Jun 2022 – Aug 2022

- Collected six articles and produced comprehensive **literature review** for text extraction algorithms
- Prepared and cleaned training textual data for supervised machine learning models using **pandas, SQL, NLTK** and **Excel**
- Automated labeling of company announcements using **Python** scripts, resulting in a **80%** decrease in labeling time

PROJECTS

Classifying Electroencephalogram (EEG) Signals

- Performed data augmentation techniques such as noise injection, subsampling and trimming to artificially increase dataset size
- Built a hybrid CNN and RNN architecture using **TensorFlow** to classify EEG signals, achieving an accuracy of ~ 72.14%
- Optimized model performance by fine-tuning hyperparameters with **KerasTuner**, enhancing prediction accuracy by ~10%

Analyzing Toxic Course Combinations for Student Performance

- Designed and implemented an ETL pipeline to preprocess course enrollment and grade data using **SQL** and **pandas**
- Feature-engineered a course combination binary matrix to analyze the impact of course pairings on student performance
- Conducted **statistical tests (t-tests, ANOVA)** and built analytics dashboards, translating complex statistical findings into actionable insights and recommendations using **Tableau, SciPy, statsmodels**
- Identified high-risk course combinations using random forests and logistic regression feature importance scores

Spotify Genre Predictions

- Performed **EDA, preprocessed data** and built **data visualizations** on Spotify data with 114000+ rows and 20 features
- **Engineered features** to address class imbalance, including adjusting prediction probability and applying SMOTE techniques
- Benchmarked algorithms like random forests, clustering analysis and PCA to achieve classification accuracy of ~70%

SKILLS

- **Languages:** Python, C++, R, SQL, HTML/CSS, JavaScript
- **Framework:** Tensorflow, Keras, Django, NumPy, Pandas, Matplotlib, PyTorch, Bootstrap, JQuery, Selenium, Scikit-learn
- **Tools:** Git/Github, JIRA, Docker, AWS services, Microsoft Excel, Word, Powerpoint, Access, Tableau, PowerBI

EXTRACURRICULAR ACTIVITIES

Founder & Entrepreneur at Miss Miso

- Designed and created unique jewelry pieces, overseeing the entire production process from concept to final product
- Achieved profit margins of 90% of the revenue through cost optimization and strategic pricing

Data Blogger at DataRes

- **Built compelling dashboards** and performed **statistical analysis** on a variety of datasets using **Tableau, Python** and **R**
- Partnered with cross-functional teams to create data-driven articles, enhancing engagement with impactful visualizations