MATTHEW CHEN

mattc2371@gmail.com | (530) 379-6538 | https://github.com/Matt2371

EXPERIENCE

National Science Foundation Graduate Research Fellow | UC Davis | September 2022 - Present

- Designed novel RNN and LSTM deep learning architectures to accurately model a large sample of reservoir operations across the US using PyTorch. Devised a transfer learning scheme to finetune models to specific sites.
- Applied Principal Component Analysis (PCA) to analyze LSTM internal states, and correlated states to meaningful latent variables (such as reservoir storage) to reveal key temporal patterns with the input and output data and enhance model interpretability.
- Conducted permutation testing/randomization inference to identify scenarios where models exhibited varying performance.
- Leveraged logistic regression to predict climate change effects using historical environmental data, achieving high model performance across specificity, sensitivity, accuracy, and AUC metrics.
- Presented research findings to diverse audiences, including an international conference with 25,000+ attendees (American Geophysical Union, 2022). Received positive feedback from peers and experts.

Product Development Data Scientist (Long-term Intern) | Genentech, Inc. | June 2024 - June 2025

- Analyzed complex longitudinal datasets for four early phase clinical trials using R, SAS, and SQL to produce statistical deliverables directly relevant to a $700m acquisition
- Applied MMRM (Mixed Model with Repeated Measures) and implemented multiple covariance structures to assess treatment efficacy
- Spearheaded development of a modular R-shiny app to enable stakeholders across multi-functional teams (including Biostatistics, Data Management, and Clinical Science) to quickly explore insights as patient data is updated
- Developed data transformation pipelines and oversaw rigorous quality control, including unblinding workflows in a randomized control trial, to ensure data quality and integrity
- Accelerated team adoption of Git version control from legacy workflows by providing best practices and hands-on support

EDUCATION

PhD Student, Civil and Environmental Engineering | Expected 2026 | University of California, Davis
MS Statistics | December 2023 | University of California, Davis | GPA: 4.0
BS Civil Engineering | June 2022 | University of California, Davis | GPA: 4.0

PUBLICATIONS

- Chen, M., & Herman, J. (2024). Detection time for nonstationary reservoir system performance driven by climate and land use change. *Journal of Water Resources Planning and Management.*

SKILLS

- Python (pandas, numpy, matplotlib, scikit-learn, PyTorch, tensorflow, keras), R, SQL
- Statistical data analysis methodology, linear regression, experimental design, hypothesis testing. Static and interactive data visualization.
- Supervised and unsupervised machine learning, deep learning using PyTorch and tensorflow/keras, optimization theory and algorithms.
- Git/Github version control, UNIX shell scripting in cloud environment

PROJECTS

ML for Solving Differential Equations: Physics-Informed Neural Networks

- Leveraged autodifferentiation to implement custom loss functions that allow neural networks to solve differential equations in fluid mechanics. Obtained R^2 scores of 0.996+ across predicted velocity and pressure fields in a turbulent flow reconstruction problem, given only sparse training observations, i.e., 0.5% randomly chosen data samples in time and space.

Neural NLP: Yelp Review Text Classification

- Classified Yelp restaurant reviews into food or service-related categories by finetuning BERT using PyTorch and HuggingFace Transformers after manually labelling a dataset of 1000 reviews into one-hot categories. Compared results to a baseline bi-directional LSTM using pre-trained word2vec embeddings.

Causal Inference: Proposition 99 and Synthetic Control

- Estimated causal effects of Proposition 99 on lowering cigarette sales in California using the synthetic control method and randomization inference, controlling for several state-level covariates. Compared findings with Abadie et al., (2010).

RELATED COURSEWORK (GRADUATE LEVEL)

- Probability Theory, Mathematical Statistics, Regression Analysis, Computational Statistics, Statistical Machine Learning, Optimization for Big Data Analytics, Big Data and High-Performance Computing, Data and Web Technology, Causal Inference, Analysis of Variance

AWARDS

NSF Graduate Research Fellowship (2024-Present)

- Three-year fellowship awarded by the National Science Foundation for high research potential in STEM

Outstanding Masters Student Award in Statistics (2023)

- The Outstanding Masters Student Award is an annual award given to the best performing students in the Masters of Statistics program

Outstanding Senior Award (2022)

- Selected by a faculty committee as the top student for academic excellence and contributions to the larger learning community.