Xiongjie (Jack) Dai

📱 217-305-0416 | @ xdai12@illinois.edu | 🔗 linkedin.com/in/xiongjie-dai | 🐙 github.com/Xiongjiedai

EDUCATION

University of Illinois Urbana-Champaign (UIUC) — Urbana-Champaign, Illinois, USA
Master of Science in Statistics; **GPA: 3.84/4.00** — *August 2021 – May 2023*

Jinan-Birmingham Joint Institute — Guangzhou, China
Economic Statistics Program (Full-time Dual Degree) **GPA: 3.93/4.25** — *September 2017 – June 2021*

Jinan University (JNU) — Guangzhou, China
Bachelor of Economics — *September 2017 – June 2021*

University of Birmingham (UoB) — Birmingham, England
Bachelor of Science in Applied Mathematics with Statistics (First-class degree) — *September 2017 – June 2021*

RELEVANT COURSEWORK

Machine Learning: Artificial Intelligence, Deep Learning, Natural Language Processing, Statistical Learning

Data Science: Data Science Programming Methods, Statistical Data Management, Statistical Consulting

SKILLS

Programming: Python, R, SQL

Technologies: Git, Shell, PyTorch, Keras, NumPy, Pandas, Matplotlib, ggplot2, Tidyverse, R Shiny app

PROJECTS

GPU Benchmarks on LLM Inference | *Git, Shell* — *December 2023*
- Used one of the most active open-source communities around **large language model** (LLM) inference, **llama.cpp**, to benchmark the **LLaMA** models' inference speed on different NVIDIA GPUs on RunPod and various Apple Silicon hardware. Provided solid data support for **localized** implementing LLM inference. Hit over 200 stars on Github repo.

Neural Machine Translation: RNN and Transformer Models | *Python, PyTorch* — *May 2023*
- Spearheaded a Neural Machine Translation project, innovating with PyTorch to develop high-accuracy **Recurrent Neural Network** (RNN) and **Transformer** models for Spanish to English translation.
- Demonstrated deep learning prowess by integrating Gated Recurrent Units and **attention** mechanisms, building both encoder and decoder of the model, achieving outstanding **BLEU-4** scores (RNN: **0.058**, Transformer: **0.059**).

Reinforcement Learning in Snake | *Python, Numpy* — *December 2022*
- Trained an **AI agent** using **reinforcement learning** to play a simple version of the Snake game, which consists of a snake agent, a food pellet, and walls around the environment.
- Operated agent in the environment by defining a Markov Decision Process containing states, actions, and rewards. Implemented the TD version of the **Q-learning** algorithm to maximize its reward by learning to get as many food pellets as possible without dying.

Sentiment Analysis for Amazon Review & **Drug Dataset** | *R markdown, Word2Vec* — *December 2022*
- Focused on **text** classification and sentiment analysis of two datasets, Amazon Review and Drug Review.
- Implemented four classic **word embeddings** and **Natural Language Processing** methods using R packages, including BoW, Word2Vec, GloVe, and fastText. Visualized most represent words by Word Cloud. Used **Naive Bayes** and **Random Forest** algorithm to classify the sentiment of the reviews.
- FastText performed best with **86.49%** accuracy for Amazon Review Dataset and **78.69%** for Drug dataset.

Explaining the Effects of Data Augmentation with CNN | *Python, PyTorch, Matplotlib* — *May 2022*
- Trained three different **Convolutional Neural Network** (CNN) models with three different seeds to build a cat-dog **image** classifier using the Oxford-IIIT Pet dataset after **data augmentation** techniques.
- Applied **regularization** techniques, such as dropout, L1, and L2 penalties, to avoid overfitting. Used binary cross entropy as the cost function and SGD as the optimizer. The best CNN model achieved the accuracy with **74.8%** under the data augmentation technique of flipping, as compared with **59.6%** from the original data.

Recommendation System for Movie | *R, Tidyverse, R Shiny app* — *December 2021*
- Constructed **recommendation system** based on 1M movie ratings by **user-based collaborative filtering** (UBCF) and **item-based collaborative filtering** (IBCF); designed and built an interactive **shiny app** based on the system.

WORK EXPERIENCE

Graduate Course Assistant — University of Illinois Urbana-Champaign
Fundamentals of Deep Learning, Statistical Learning — *September 2022 – December 2022*
- Effectively collaborated with faculty and assisted students, developing machine learning content and creative coding assessments in R and Python. My role in grading and feedback highlights my strong **self-directed** analytical skills, communication ability, and passion for contributing constructively to team-driven machine-learning projects.