ZIXIANG TANG

Email: shawntang27@gmail.com| (203) 770-2776
LinkedIn: https://www.linkedin.com/in/zixiang-tang| Wellesley, MA

WORK & RESEARCH EXPERIENCE

Yale University Center for Biodiversity and Global Change, Deep Learning Software Engineer Dec. 2023 – Present

- Collaborated as a key member of one of the six finalist teams in the **XPRIZE** competition, competing for a **$10 million** prize purse
- Built an automatic **machine learning pipeline** to provide insights for the biodiversity (habitat and species) in the rainforest
- Created an automated pipeline to label training data, use the labeled data to train and fine-tune foundation deep learning models
- **Automated** the process of species identification, **deployed and integrated** the automated pipeline into current web application

Brave Career, Machine Learning Engineer (Contract) Apr. 2023 – Nov. 2023

- Built a job recommender app hosted on **AWS** to send emails to users with top matched job recommendations in a daily basis
- Built front end with **Streamlit**, web scraped job postings with **Apify,** and sent emails with **AWS SES** service
- Parsed job postings, extracted and summarized resumes, and created embeddings using **OpenAI GPT** models with **Langchain**
- Automated an **ETL** pipeline using **AWS Lambda** and **Eventbridge**, stored data in **Elasticsearch** and visualized by **Kibana**

Frederick National Laboratory for cancer research, Intern Sept. 2022 – Jan. 2023

- Designed an active learning framework to filter out fitted compounds, **reduced 40% time** in drug discovery process
- Built an automatic data preprocess pipeline for data cleaning and transformation, performed EDA on **10GB+ data**
- Implemented **Random Forest** and **Neural Networks** on docking scores and ranking using ATOM modeling pipeline (AMPL)
- Achieved **0.83 of f1 score** on classifying top 5% docking score; conducted hyperparameter tuning to obtain **12% improvement**

China Coal Technology Engineering Group, Data Scientist Jun. 2020 – Jul. 2021

- Developed a machine learning pipeline to detect damage on conveyor belt in mining industry, **improving 25% performance**
- Worked closely with the cross functional team to collect data, analyze requirements and perform data analysis
- Developed an **ETL** pipeline to maintain the **Cloud Database** and wrote complex **SQL** queries to generate statistical reports
- Built **Logistic Regression**, **XGBoost** and **Neural Network** models to predict on damages and fine-tuned these models
- Integrated the model into the system and deployed, saved **20% cost**, and improved the overall **throughput by 28%**

Meyers' Lab in UCSD, Research Assistant Oct. 2018 – Apr. 2020
Research in Active Defense Mechanisms of Thorny Catfish

- Performed **quantitative analysis** with Python to identify the mechanism of catfish defensive system and its mechanical properties
- Coauthored the research paper, **published in *Materials Today***: *Active defense mechanisms of thorny catfish*

SKILLS

- **Programming**: Python, Pandas, Numpy, Git, TensorFlow/Keras, PyTorch, SQL, NoSQL, Opencv, Langchain, Rust, R, Java
- **Machine Learning**: Scikit-learn, Deep Learning, NLP, Time Series, Predictive Modeling, Large Language Model
- **Data Analytics**: A/B Testing, Tableau, PowerBI, SAS, Google Analytics, Azure Synapse, Matplotlib, Seaborn, Plotly
- **Data Engineering**: AWS, Google Cloud Platform, Azure, BigQuery, Spark, Databricks, Terraform, Docker

EDUCATION

Columbia University **New York, NY**
Master of Science in Data Science | GPA: 3.8/4.0 **Feb. 2023**

Relevant Courses: Applied Machine Learning, Statistical Inference & Modeling, Natural Language Processing, EDA/Visualization

University of California San Diego **La Jolla, CA**
Bachelor of Science in Mechanical Engineering, Cum Laude | GPA: 3.8/4.0

Minor: Mathematics **Mar. 2020**

PROJECTS

Neural Network - Photo restoration – moire pattern removal with CNN TensorFlow/Keras Oct. 2022 – Dec. 2022

- Remove moire pattern (digital noise) to restore the textures and colors of the original photo
- Built a convolutional neural network based model using customized model, layers and loss function with **TensorFlow and Keras**
- Constructed moire texture removal blocks with multiple dense, convolutional layers and learnable band pass filters
- Successfully removed moire pattern, achieved **0.72 accuracy and 0.78 SSIM**, obtained 10% improvement by tuning the model

NLP - Image Captioning with LSTM Generators TensorFlow Feb.2022 - May. 2022

- Maintained **3GB+ pictures** and text from 3 different websites and extracted the features from the pictures
- Implemented **RNN, LSTM and BERT** to generate the captions for each image via **TensorFlow**, **reached bleu score of 0.32**
- Further optimized the models with beam search decoder to keep three highest probable choices, **improved score by 21%**