

VICTOR LAI

laivictor2718@gmail.com

github.com/laivictor • linkedin.com/in/victor-lai-3461ab19b/ • (737) 704-5901

EDUCATION

The University of Texas at Austin

May 2023

Master of Science, Business Analytics

The University of Texas at Austin

May 2022

Bachelor of Science, Electrical Engineering (Honors), Minor: Business

Overall GPA: 3.62/4.00

PROJECTS

ML/Data Analytics for Edge AI

2022

- Worked with both PyTorch and Tensorflow to build convolutional neural networks trained on the CIFAR 10 dataset submitted to the TACC's Maverick2 supercomputer via shell. The models were then converted to ONNX and TFLite for deployment on edge devices such as the Odroid MC1 and Raspberry Pi. Used data augmentation, pruning, and hyperparameter tuning to achieve certain goals in metrics such as accuracy, latency, and energy consumption.

PGPAIML - Post Graduate Program in Artificial Intelligence and Machine Learning

2021-2022

- Course projects portfolio: <https://eportfolio.greatlearning.in/victor-lai>
- Performed sentiment analysis on tweets using NLP techniques, with libraries including Spacy, Textblob, BeautifulSoup, NLTK, UnicodeData, and RE in Python
- Classified plant seedlings based on images using CNNs with models built in Tensorflow
- Identified whether or not bank customers would churn with ANNs in Tensorflow
- Did customer segmentation with silhouette visualization, clustering and pattern profiling
- Used cross validation, regularization, pipelines, hyperparameter tuning, and up/down sampling to predict whether customers would leave
- Used ensemble techniques including bagging and boosting to predict customers who would buy a newly introduced package

Data Science Lab

2021

- In Python, worked on a variety of data science problems, including data visualization, classifiers such as decision trees or random forests, both logistic and linear regressors, boosting and tuning parameters with XGBoost, and ensembling methods such as stacking, and bagging. Did convolutional neural networking and deep learning with PyTorch, FastAI, and ResNet, both with and without pretrained models. Other used libraries include Pandas, Seaborn, SciPy, SkLearn, and more.

Concurrent and Distributed Systems

2021

- In Java, worked on projects that involved processing data in parallel and multithreading, with techniques such as semaphores, locks, synchronization, atomic variables, and several mutex algorithms. Also worked with client server programming with both TCP and UDP sockets. Worked with MapReduce in Hadoop to make a text analyzer for entire books that runs within minutes.

App Scrape

2021

- Part of a larger project called PrivacyCheck by the UT Center of Identity, built a crawler in Python to scrape Apple's new privacy cards, one hot encoded based on the presence or absence of elements, and exported the resulting dataframe to csv files. Also scraped privacy policy links and their corresponding text. Worked with BeautifulSoup to scrape and Pandas to format.

Data Science Principles

2020

- Used NumPy and Matplotlib(Python) to solve various data science problems such as doing linear regressions and measuring training vs testing errors. Worked with MNIST dataset to recognize handwritten digits using K Nearest Neighbors.

Olympics

2020

- As part of a team, built a website that provides information about the Olympics, such as event records and notable athletes. Used data mining tools like Scrapy and BeautifulSoup to obtain information, and Flask to build template webpages. Mainly in HTML and Python.

BlogApp

2020

- Developed a space themed blog hosted online where users can view blog posts, sign in to create their own, search blogs based on title or content, use the fixed navigation bar at the top to switch between pages, and subscribe to a daily digest by email of new blog posts made. Hosted on Google Cloud Platform, with a mixture of Java, HTML, CSS, Cron, and more.

HONORS

Predicting Bad Debt Societies Hackathon - First Place

Dec 2021

- Built the model with the best mean precision score out of 87 participants predicting whether accounts would be unable to pay off debt, given past data about loan applicants.

NBFC Loan Default Hackathon - First Place

Oct 2021

- Built the model with the best accuracy out of 60 participants predicting whether a client would default on the loan payment or not, given past data about bank accounts.

TECHNICAL SKILLS

Programming Languages: Python, Linux, Java, HTML

Data Science Tasks: Convolutional/Deep Neural Networks, Model Deployment, Natural Language Processing, Clustering, Pipelines, Hyperparameter Tuning, EDA, Data Visualization, Data Preprocessing, Web Crawling, Scraping, Classification, Regression, Ensembling - Bagging, Boosting, Stacking, Parallel Processing, Multithreading

Natural Language Processing: Sentiment Analysis, Count/TFIDF Vectorization, Text Preprocessing, Tokenization, Stopword Filtering, Regex, Stemming, Lemmatization

Convolutional Neural Networks: Sequential Model Building, Convolution, Pooling, Activation Functions, Fully Connected Layers, Data Augmentation, Network Pruning, Model Deployment, Optimizing, Checkpoints

ADDITIONAL INFORMATION

Languages: Fluent in English, Conversational in Chinese, Basic Knowledge in Spanish

Interests: Swimming, reading, videos, duolingo, leetcode

Work Eligibility: Eligible to work in the US with no restrictions