THI GIANG

San Jose, CA | (408) 398-6537 | Email | Portfolio | Linkedin

EDUCATION

Master of Science, Statistics (GPA: 3.87/4.0) May 2023
San Jose State University, San Jose, CA
Coursework: Mathematical Statistics, Regression Methods, Statistical & Machine Learning, R, Python-SQL

Associate of Arts, Mathematics, (GPA: 3.93/4.0) Dec. 2020
San Jose City College, San Jose, CA

Bachelor of Arts, Finance and Banking (GPA: 3.59/4.0) Jan. 2015
Can Tho University, Can Tho, VietNam

TECHNICAL SKILLS

Languages: Python, SQL, R, Matlab.

Libraries/Tools: NumPy, Pandas, SciPy, SciKit-Learn, TensorFlow, Keras, Seaborn, Matplotlib, ggplot2, Git, GitHub, Latex, Jupyter Notebook, Google Colab, Visual Studio Code, Microsoft Office,.

PROJECTS

Sentiment Analysis on Yelp Review Dataset (link) May 23-Current
- Collected 20,000+ restaurants' reviews through Yelp Fusion API using Python libraries.
- Removed stop-words and vectorized the review data using Bag-of-Words and TF-IDF for classification.

Building an AskMe Chatbot using Streamlit and OpenAI API (link) May 23
- Created an AskMe Chatbot similar to Chat GPT that can take in and answer any questions using from users.
- Used openai.Completion to generate a response for a given prompt and then used Streamlit session state and Streamlit message to get users' input questions, store and display the chat history.

Predicting lung diseases using ML models (link) Sep. 22 - Dec 22
- Transformed & manipulated 12,000 Chest X-ray images to pixel data & converted multilabel to a single label.
- Performed classification using Neural networks, Support vector machines (SVM), and Bayes classifier.

CAMCOS - Rollup Pricing Models Jan. 22 - June 22
- Participated in a group of student researchers to work on improving the EIP-1559 pricing mechanism on Ethereum network which was sponsored directly by Ethereum Foundation.
- Studied pricing models that incentivize the operators (miners) to participate in layer 2, a solution to scaling up transactions in layer 1 or the base Ethereum chain.

Visualized & Reduced Dimensions on Music Genre Dataset (link) May 22
- Analyzed 50,000 songs, explored the data distributions for 18 features, and handled missing data.
- Implemented dimension reduction techniques like PCA, LDA, and MDS in Matlab and successfully reduced the high dimensional dataset for better visualization and classification.

Predicting Rental Cost in Brazil (link) Nov. 21 – Dec. 21
- Cleansed, visualized and analyzed the data using ggplot2 in R to understand the important variables.
- Utilized information criteria, AIC and BIC, to select the best polynomial regression model. The model predicted the rental cost in Brazil with an accuracy of 91.5% and a mean square error of 0.008.

WORK EXPERIENCES

Precalculus Facilitator | Mathematics & Statistics Department, SJSU Aug. 21 – Dec. 21
- Weekly led 20 to 25 college students to work on precalculus practice problems.
- Communicated with the instructor to keep track of students' progress.