

THI GIANG

San Jose, CA | (408) 398-6537 | thiyounguni22@gmail.com
<https://thi-giang.wixsite.com/portfolio> | <https://www.linkedin.com/in/thi-giang/>

EDUCATION

Master of Science, Statistics (GPA: 3.85/4.0)

Expected May 2023

San Jose State University, San Jose, CA

Coursework: Statistical & Machine Learning, Python-SQL, Data Visualization,
Mathematical Statistics, Regression Methods, R Programming

Associate of Arts, Mathematics, (GPA: 3.93/4.0)

Dec. 2020

San Jose City College, San Jose, CA

TECHNICAL SKILLS

Languages: Python, SQL, R, Matlab.

Libraries/Tools: NumPy, Pandas, SciPy, Matplotlib, SciKit-Learn, TensorFlow, ggplot2, Seaborn
Visual Studio Code, Microsoft Office, Github

PROJECTS

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 and determined metrics such as test error, accuracy rate, confusion matrix, F1 score, ROC, and AUC.
- Performed classification using multiple classifiers like Neural networks, Support vector machines (SVM), and Bayes classifier.
- Achieved the highest accuracy of 90.8% with SVM, followed by the neural network with an accuracy of 90.6%. Classification trees and logistic regression were more computationally intensive for high-feature data.

Visualized & Reduced Dimensions on Music Genre Dataset ([link](#))

May 22

- Analyzed 50,000 songs exploring the data distributions and evaluating dimension reduction methods for the 18 features.
- Utilized imputation to improve the data quality by imputing 4980 missing values in the dataset using 3 nearest neighbors.
- Implemented dimension reduction techniques like PCA, LDA, and MDS in Matlab and successfully reduced the high dimensional dataset for better visualization and classification.

Wordle Solver ([link](#))

May 22

- Reimplemented the Wordle game by collaborating with a diverse team & built a solver to play the game.
- The solver uses the feedback to enhance its future guesses and chooses the best words based on the relative frequency of use of the English language
- Utilized a lexicon of possible words to optimize the program to win at a rate of 98.68% and in 3.8 guesses on average.

Predicting Rental Cost in Brazil ([link](#))

Nov 21 – Dec 21

- Cleansed and visualized 10,962 observations with 13 features using ggplot2 in R to understand the relationship between 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.