ALAN WONG

San Jose, CA | (408) 505-1595 | alan.j.wong@columbia.edu | linkedin.com/in/alanwong11/

EDUCATION

Columbia University New York, NY
MS in Data Science Expected Dec 2025
Relevant Coursework: Computer Systems and Databases, Applied Machine Learning, Statistical Inference and Modeling

University of California, Los Angeles (UCLA) Los Angeles, CA
BS in Statistics and Data Science, GPA: 3.8/4.0 Jun 2024
Relevant Coursework: Probability and Statistics, Exploratory Data Analysis and Visualization, Data Structures and Algorithms

EXPERIENCE

Stanford University - SHTEM Program Stanford, CA
Research Intern-Mentor May 2023 - Aug 2023
- Administered Fast Fourier Transform (FFT) to efficiently convert 10,000 audio files into graphs with known peak values
- Trained a convolutional neural network (CNN) spanning 1000 epochs that leveraged a CSV file of audio features
- Co-authored a paper regarding system for effective audio separation and instrument identification of 20 instruments

Code Ninjas San Jose, CA
Assistant Branch Manager/Coding Instructor Aug 2021 - Sep 2022
- Instructed students of varying ages and levels of expertise in Python, Javascript, and Lua, focusing on practicality
- Mentored youth in comprehensive curriculum for CAD, Minecraft Modeling, and Roblox Studio with hands-on projects
- Collaborated in teaching programming fundamentals to over 50 students, such as data types, loops, and functions

PROJECTS

Societal and Environmental Impacts of Air Traffic Mar 2024
- Enhanced two maps of geospatial data with QGIS to show how air pollution impacts communities surrounding airports
- Fine-tuned visualizations of air travel discrimination with Tableau to reflect social dynamics influencing travel decisions
- Refined website expanding upon two disproportionate effects of air traffic on marginalized communities and environment

Baseball Hall of Fame Prediction Dec 2023
- Performed exploratory data analysis using visualization techniques from Tableau to understand Hall of Fame voting trends
- Developed four predictive models with Random Forest and Boosted Tree (XGBoost) algorithms based on player statistics
- Achieved high prediction accuracies (~90%) in identifying potential inductees, demonstrating effectiveness of models

IMDb Sentiment Analysis Nov 2023
- Conducted sentiment analysis on 50,000 IMDb movie reviews, operating machine learning essentials for text classification
- Applied data pre-processing methods including text cleaning, TF-IDF vectorization, and PCA for dimensionality reduction
- Designed and compared four machine learning models for predictive accuracy and analyzed model performance metrics

Good Lawyer, Bad Lawyer - ASA Datafest @ UCLA Apr 2023
- Implemented sentiment analysis with NLTK across a dataset of 200,000 client-lawyer conversations, identifying key tones
- Created R and Tableau visualizations depicting sentiment distributions by state and analyzed attorney language trends
- Analyzed sentiment transitions in over eight types of cases to provide advice regarding intricate linguistic patterns

Differences in Autonomous and Human Driving - SHTEM @ Stanford Uni. Jun 2020 - Aug 2020
- Executed coordinate differencing and spline interpolation to understand lane-keeping behavior in autonomous vehicles
- Utilized Python and Jupyter Notebook to run detailed data analyses on Waymo and Lyft autonomous vehicle datasets
- Co-authored paper detailing three major driving patterns of autonomous vehicles for user practicality and efficiency

LANGUAGES AND TECHNOLOGIES

Languages: Python, R, SQL, Javascript, Java, C++, HTML/CSS
Operating Systems: macOS, Linux (Raspberry Pi OS Bullseye, Ubuntu), Windows
Communications: SSH, VNC, Tailscale, RDP
Technologies: PostgreSQL, Jupyter Notebook, Visual Studio Code, Tableau, AWS Cloud (EC2, S3), Git/Github, APIs, Express.js, Node.js, Postman, Dropbox, RStudio, Microsoft Office, Sublime Text