Andrew Nguyen

(408) 946-4400 | andrew.nguyen@gmail.com
https://www.linkedin.com/in/andrew-nguyen-122a90228/ | https://github.com/AndrwNgyn
Quick learner, problem solver, fluent in English and Vietnamese.

EDUCATION

University of California, Santa Barbara (UCSB) | B.S. Statistics and Data Science
GPA: 3.83/4.00
Coursework: Probability and Statistics, Linear Regression, Design of Experiments, Stochastic Processes, Statistical Machine Learning (Graduate Division), Time Series, Bayesian Data Analysis

SKILLS

Languages: Python, R, SQL
Familiar Applications: NumPy, Pandas, Scikit-learn, Matplotlib, TensorFlow/Keras, seaborn, Flask, ggplot, tidymodels, tseries

EXPERIENCE

Picarro Inc. - Automation and Test Intern

Santa Clara, California | JUNE 2022 - SEPTEMBER 2022

— Designed and improved automated testing applications using Python, pandas, and NumPy to reduce user-interface heavy work using wxPython, suitable for both Windows and Linux platforms
— Implemented SQL to record product test data into the company database and analyzed said data to provide prospects on how to improve production efficiency

Picarro Inc. - Data Analytics Intern

Santa Clara, California | JUNE 2023 - SEPTEMBER 2023

— Developed and tested a neural network model to detect instrument measurement anomalies (0.01% class imbalance) with a **93%** F1 score. Said model is automated using Slack chatbot hosted on AWS EC2 instance to notify Field Application Engineers of anomalies to quickly consult customers for reconfiguration of product instruments, completely alleviating manual labor by **100%.** Pipelines are also constructed to automatically feed measurement data into predictive models.
— Extracted performance data via PostgreSQL from Jira to performance metrics dashboard using Apache Superset to highlight team performance
— Executed MSSQL queries to run computational diagnostics and provide data to other engineers

PROJECTS

US Electricity Production Forecast: **MARCH 2023**

Time Series Analysis of US Electricity Production Types in R

- Forecast and analysis of US Electricity Production types using time series models in R. Production values for various production types (fossil fuels and renewable energy) are forecasted using SARIMA models.
- Independent research of US Electricity Production type and behavior, and gained insight into energy production trajectory for the future. The report can be viewed here.

Fetal Health Classifier MARCH 2023

- Classifying Fetal Health using Statistical Machine Learning Models in R
- Observed multiple models in R to determine which model would predict the fetal health given cardiotocography (CTG) data, and concluded with a boosted tree model with an accuracy of **97%**. The project can be viewed here

EXTRACURRICULARS

Santa Barbara Science Olympiad: AUGUST 2020 - JUNE 2022 - Santa Barbara, California

— Helped create tests and supervised events for invitational, regional, and state-level competitions
— Collaborated with multiple teams to formulate adaptations to online competitions

VOLUNTEER WORK

Circle K International: NOVEMBER 2021 - PRESENT - Santa Barbara, California

— Provided community service around Santa Barbara County, including aiding homeless shelters and cleaning the community
— *Public Relations Chair* - Liaison between other organizations on campus, updated club websites and other social media platforms, and advertised organization through campus and social media outlets, increasing attendance by **125%**