

Pamela Nguyen

254-350-3406 | cao.nn.p@gmail.com | Austin, TX

<https://github.com/pcn-physics> | <https://www.linkedin.com/in/pam-nguyen/>

TECHNICAL/COMPUTER SKILLS

Programming Languages: C++, Java, Python (NumPy, pandas, scikit-learn, matplotlib, PyTorch, SymPy, SciPy), Fortran

Tools and Platforms: Jupyter Notebooks, Bash/Unix/Linux, MATLAB, LaTeX, Google Cloud Platforms, Git

Databases: SQL, Postgres, BigQuery, NoSQL, Excel

Languages: Spanish (A1), French (A2), English (Native), Vietnamese (Social)

PROJECTS

Ensemble Classifiers Analysis on Credit Card Transactions, Austin TX

Nov 2021

- Partitioned the data for cross validation using a stratified k-fold and fitted scikit-learn algorithms: AdaBoost and Random Forests, on over 140,000+ confidential credit card transactions to predict fraudulent charges.
- Concluded that the random forest algorithm yielded a higher accuracy of determining fraudulent charges when calculating the TPR, PPV, FPR and plotting the ROC curves for each fold. Used: Python (scikit-learn, matplotlib)

RDBMS Hardware and Software Optimizations, Austin TX

May 2022

- Deployed, tested, and improved the runtime of a relational database management system (RDBMS) environment on a Compute Engine instance in GCP using Postgres 13. Used: Postgres, GCP, Git, bash
- Improved the latency of writing 8 million records (+20%), server upgrade cut time (+17%), and bandwidth (+4%) by adjusting the batch size, upgrading the bandwidth, and increasing the CPU and RAM.

Infectious Disease Simulation, Austin TX

Oct 2019 - Dec 2019

- Modeled and tracked disease propagation within a population from an infectious person statistically using an explicit SIR model in C++.
- Analyzed and plotted the effects contagion probabilities, population sizes, contact spreading probabilities, and inoculation percentages had on the contagiousness of the disease until the population has obtained herd immunity in Matlab. Used: C++, bash, Frontera (TACC), Matlab

RESEARCH & EXPERIENCE

UT Austin: Molecular Biosciences, Austin TX

January 2019 - April 2023

Undergraduate Researcher under Professor Elif Sarinay Cenik

- Generated and executed computational pipelines using Python (pandas, NumPy, seaborn, matplotlib), Bash, and R for parsing, visualization, statistical analysis, and other processing of data generated by NGS workflows.
- Produced and tested bash pipelines that converted raw sequencing data (FASTQ, FASTA) to binaries (BAM) in order to save storage and create readable variant data (VCF) for analysis. This pipeline was used to convert all sequencing data in the lab.
- Improved the efficiency (about 20 mins per 1TB of data) and the accuracy (+8%) of the sequencing pipelines by generating test data, adjusting pipeline parameters, and increasing CPU/RAM.
- Consistently communicates with peers, professors, and researchers on implementing modern research tactics and improving accuracy of results and data.
- **Senior Research Thesis**: Filtered, sorted, plotted and correlated the effects of rDNA copy number extremities on gene expression and genomic variation of millions of *C. elegans* records. Used: Python: pandas, NumPy, seaborn

EDUCATION

The University of Texas at Austin, Austin, TX

December 2022

Bachelor of Science in Mathematics, Computational Physics

- Certificate in Elements of Computing, Scientific Computation & Data Sciences
- **Relevant Coursework**: Numerical Methods I/II, Probability, Statistics, Matrices, Game Theory, High Performance Computing (HPC), Data Analytics, Databases, Data Structures and Algorithms, Software Design, Electronic Techniques, Quantum Mechanics, Statistical Mechanics, Electromagnetics, Modern Physics & Lab