

David Garudo

Austin, TX – garudodav@gmail.com – <https://www.linkedin.com/in/GarudoDav> – 512-927-9020 – US Citizen

EDUCATION:

Texas A&M University, B.S. Data Engineering, College Station, TX

May 2028

Relevant Coursework — completed or expected by May 2027: Uncertainty Modeling, Quantitative Models for Statistical and Machine Learning, Statistical Learning and Decisions, Optimization of Analytics, Decision and Risk Analysis, Principles of Data Science, Computational Data Science, Probability and Statistics, Linear Algebra, Data Structures and Algorithms, Database Systems

SKILLS

Quantitative Methods: Statistical Modeling, Probability & Statistics, Predictive Modeling, Regression, PCA, Feature Engineering, Model Selection & Evaluation, Optimization, Anomaly Detection, Imbalanced Classification

Programming & Libraries: Python, SQL, R, C++, Java, NumPy, Pandas, SciPy, scikit-learn, PyTorch, TensorFlow, Matplotlib

Tools & Systems: Linux, Git, GitHub, Jupyter, Google Cloud Platform (GCP), Distributed Computing, ETL Pipelines

QUANTITATIVE RESEARCH PROJECTS:

Quantitative Sports Modeling — Texas A&M University

Jan 2026 - May 2026

- Built a Python research pipeline using NBA game data from 1996–2026 and engineered normalized player and team impact factors to support comparisons across changing pace and scoring environments.
- Applied statistical smoothing to reduce small-sample distortion, investigated anomalous model outputs, and iterated on factor definitions to improve stability and interpretability across teams and seasons.

Sparse Linear Image Modeling — Texas A&M University

Jan 2026 - May 2026

- Selected as 1 of 8 ADSC semester project leads after pitching an original **computer vision** and **machine learning** approach to sparse image reconstruction; recruited and led a 6-member team through research, predictive modeling, model development, and final poster presentation.
- Performed feature engineering and Principal Component Analysis (PCA) on **10M+** observations derived from a computer vision pipeline, benchmarking Logistic Regression, Random Forest, SVM, KNN, and MLP models for model selection and predictive modeling, **improving recall by ~22% and achieving ~97% recall** on imbalanced classification tasks.
- Developed optimized stroke-generation and scoring algorithms using **NumPy**, **SciPy**, and **PyTorch**, enabling efficient sparse image reconstruction and supporting downstream predictive modeling within a computer vision framework.

WORK EXPERIENCE:

Ascension

Austin, TX

Data Analyst Intern, Dell Children's Medical Center

June 2026 - August 2026

- Engineered analytic datasets across **25K+** pediatric encounters using PHIS, Cerner, and enterprise warehouse data, identifying 1,934 instances of resource overutilization and **\$405K** in standardized cost-saving opportunities.
- Built and deployed automated **ETL pipelines** integrating Cerner and enterprise data warehouse sources into **Power BI** and **Tableau** reporting layers using **SQL** and **Python**, supporting **9+** clinical quality initiatives and enabling data-driven monitoring of pathway performance and operational KPIs.

Lonestar NDE

Austin, TX

Data Analyst Intern

June 2024 - August 2024

- Built **Python** pipelines to ingest, analyze, and visualize ultrasonic, eddy-current, and resonance data collected by a robotic system inspecting the carbon-fiber outer structures of rockets, reducing manual review cycle times by approximately **80%**.
- Developed real-time **machine-learning** and deep-learning anomaly-detection models using **Python** and **PyTorch** to help identify hidden defects, weak bonds, and vulnerable areas in rocket structures before launch.
- Integrated **automated analysis and reporting capabilities** into the production inspection suite using ReportLab, OpenCV, and PyVista, improving the consistency and scalability of engineering review.

CERN — European Organization for Nuclear Research

Geneva, Switzerland

Python Software Development Intern

June 2023 – July 2023

- Engineered a **Python**-based infrastructure profiling and telemetry framework that extracted computer architecture, CPU, GPU, memory, OS, and system architecture metadata from distributed Linux compute environments supporting large-scale high-energy physics simulations; validated deployment on CERN compute nodes, including **NVIDIA A100 HPC systems**.
- Contributed the profiling module to the open-source MadGraph5 codebase on GitHub and integrated automated profiling into simulation workflows, **improving processing efficiency by ~25%** and supporting scalable execution across distributed computing clusters.

LEADERSHIP AND ACTIVITIES FEST Officer | Aggie Data Science Club | Texas A&M National Scholar Ambassadors | Climbing Team