Lance Penaflor

850-686-2389 | penaflor.lance11@gmail.com | LinkedIn | Portfolio | GitHub

Education

University of West Florida Pensacola, FL

Bachelor of Science in Computer Science; Minor in Statistics; Certificate in Data Science *May 2027*

Experience

Machine Learning & Statistics Research Assistant *UWF CSDA Lab* Oct. 2024 – Present

- Built cancer classification pipeline on mass spectrometry data (**75% accuracy, 80% specificity**) using SVM, Random Forest, and XGBoost.
- Reduced feature space by **99.8%** with wavelet transforms + PCA, cutting training time and improving interpretability.
- Improved minority-class recall by **1̃5%** via SMOTE oversampling and Grid Search CV tuning. .
- Delivered model benchmark reports across **accuracy, precision, recall, and F1**, informing selection of optimal algorithm.

Data Science Educator *University of West Florida* Aug. 2024 – Present

- Founded a **35+ member Python/data science community**, lowering barriers to coding and Data Science.
- Created **10+ interactive Colab notebooks** for Python, statistics, and ML instruction.
- Led **weekly coding sessions (10+ attendees)** that improved student confidence with applied ML.

Undergraduate Teaching Assistant *University of West Florida* Jan. 2025 – Present

- Supported **120+ students** across C++, Algorithms, and Statistics courses, boosting performance through targeted guidance.
- Replaced calculator workflows with **Python labs**, improving reproducibility and aligning with industry practices.
- Co-Produced accessible course materials meeting **100% compliance** with disability mandates.

Full Stack Software Engineer *Andrews Institute* Sep. 2025 – Jun. 2026

- Designed **SQL schema** on **Azure SQL Database** for structured athlete injury data, user authentication, and access control.
- Configured **Azure Blob Storage** for unstructured video files linked to athlete reports, enabling efficient storage and retrieval at scale.
- Developed **R Shiny** application with **4 modules**: secure **authentication**, **injury questionnaires**, **dashboards** for injury insights, and **editable athlete records**.
- Implemented **role-based access control** in the cloud: coaches/teachers limited to their athletes, with admins granted global visibility.

Projects

Data Science Learning Website | *React, Node.js, TypeScript, Express*

- Deployed **browser-based coding + AI tutor platform** with **50+ users** using **React** and **Express**.
- Led a **3-person dev team** to maintain tutorials, ensure stability, and ship updates via Vercel and GitHub.
- Live: https://pads-murex.vercel.app/

Wavelet Perturbations on LLM Embeddings (In Progress) | *Python, HuggingFace, PyTorch, NumPy, PyWavelet*

- Designed wavelet-based embedding perturbations preserving semantics in GPT-2 outputs, advancing **robustness**.
- Compared logits and distributions to quantify differences, contributing to **LLM interpretability and safety**.
- GitHub: https://github.com/lancemelon/llm_perturbation

Music & Fashion Analysis (In Progress) | *Python, Pandas, Scikit-learn, ETL*

- Built ETL pipeline on **3 Kaggle datasets**, cleaning duplicates, removing nulls, and filtering.
- Integrated **LastFM API** to enrich Spotify data with genre tags for all unique songs.
- Applied **K-means clustering** to group songs into genre clusters, laying foundation for downstream trend analysis.

Technical Skills

Programming: Python, SQL, R, C++, JavaScript, TypeScript, Java
ML & Data: NumPy, Pandas, Scikit-learn, PyTorch, TensorFlow, MatPlotLib
Web: React, Node.js, Express, Docker, Git, Agile Development, REST API