

# Lance Penafior

850-686-2389 | penafior.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 **15%** 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 Data Engineer *Andrews Institute*

Sep. 2025 – Dec. 2025

- Designing **PostgreSQL database schema** for athlete injury data, user authentication, and access control.
- Developing **R Shiny** application with **4 modules**: secure **authentication**, **data entry** questionnaires, **dashboards** for injury insights, and **editable athlete records**.
- Implementing **role-based access control**: coaches/teachers restricted to their athletes, admin with global visibility.
- Integrating **database + Shiny app** into hospital workflow to enable **real-time data storage, monitoring, and visualization**.

## PROJECTS

---

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

- Deployed **browser-based coding + AI tutor platform** with **50+ users**, reducing onboarding by **40%**.
- Led a **3-person dev team** to maintain tutorials, ensure stability, and ship updates via Vercel.
- 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 research**.
- Compared logits and distributions to quantify differences, contributing to **LLM interpretability and safety**.
- GitHub: <https://github.com/lancemelon/llm-perturbation>

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

- Built ETL pipeline on **3 Kaggle datasets**, cleaning duplicates, removing nulls, and filtering by overlapping countries and timeframe.
- 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, C++, JavaScript, TypeScript, Java

**ML & Data:** NumPy, Pandas, Scikit-learn, PyTorch, TensorFlow, Oracle, Matplotlib, Statistics, Probability, ML

**Web:** React, Node.js, Express, Docker, Firebase, Vercel, Git, Agile Development, REST API