Rain Lin

(631) 829-1448 | Email | Portfolio |LinkedIn | GitHub | New York, NY 10009

EDUCATION:

Boston University | *Master of Science in Data Science* | GPA: 4.0 **Expected: 05/26**

University of North Carolina at Charlotte | *Bachelor of Science in Psychology* | GPA:3.8 **Completed: 12/24**

EXPERIENCE:

Software Engineer, Data & Infrastructure (Contract)| Vivi Swim School – *Remote* 07/25 - Current

- Built serverless ETL pipeline using Gmail API, regex, and Firestore, processing **100+ emails/day** with **>99% parsing accuracy** and **reducing manual workload by 80%.**
- Developed **full-stack appointment platform** (React + Firebase) handling **50+ daily queries under 2s latency** with role-based filtering, enabling efficient scheduling for coaches/admins.
- Deployed ML/infra services on **Google Cloud Run with Firebase CI/CD, cutting latency by 30%** and improving reliability of production systems.
- Published **cross-platform mobile app** (React Native + Firebase Functions) to **Apple App Store**, improving user stability and engagement through continuous feature updates and bug fixes.

Research Assistant | Levens Emotion and Cognition Lab – Charlotte, NC 01/24 - 12/24

- **Preprocessed and encoded 10k+ unstructured behavioral data entries** from Twitter, transformed **categorical/text responses into numeric values,** and **analyzed** them based on the codebook to improve data quality and support research analyses.
- Conducted **correlation analysis** and interactive **Excel visualizations** to support Multilevel Structural Equation Modeling (MSEM) and **causal inference**, identifying key emotional predictors.
- Applied **A/B testing, feature engineering, and behavioral data analysis** in weekly research discussions with PI and PhD students, contributing to experimental design improvements.

PROJECTS:

Amazon Electronics Recommender System | End-to-End Hybrid ML Pipeline | GitHub Repo | Live Demo 07/25 - 08/25

- Engineered product candidates using **SVD (collaborative filtering)** and **BERT content similarity**, and constructed feature vectors with prediction scores, sentiment analysis, and metadata to enrich ranking inputs.
- Ranked products with **XGBoost** to predict relevance scores, delivering top-N recommendations, achieving **15% higher accuracy (Precision@5 = 0.88, Recall@5 = 0.86, AUC = 0.89),** and compared with **Neural Collaborative Filtering** for ensembling.
- **Deployed** an interactive **Streamlit app** with **full MLOps** on **Hugging Face Spaces**, **reducing latency by 20%** and enabling users to explore electronic product recommendations in real time.
- **Tech Stack:** Python, Pandas, scikit-learn, Surprise, XGBoost, NCF, PyTorch, Hugging Face, NLTK, VADER, Streamlit

Zillow Property Value Prediction | End-to-End Regression Pipeline | GitHub Repo 09/25 - Current

- Processed Zillow Kaggle dataset with **exploratory data analysis (EDA), missing value handling, and feature engineering.**
- Trained and compared various regression models **(Linear Regression, Ridge Regression, Random Forest, LightGBM)** with cross-validation, achieving top performance with Random Forest **($R^2 \approx 0.979$, MAE ≈ $5,266, RMSE ≈ $87,717).**
- Currently performing **hyperparameter tuning** with Randomized Search and Grid Search, **modularizing the codebase**, and preparing for deployment with **SHAP** and **Amazon Web Services.**
- Tech Stack: Python, Pandas, scikit-learn (RandomizedSearchCV, GridSearchCV), XGBoost, LightGBM, SHAP, Folium, AWS

TECHNICAL SKILLS:

- **Languages:** Python, SQL, TypeScript, JavaScript
- **Machine Learning & AI:** Regression, Classification, Clustering, Ensemble Methods, Boosting Methods, Neural Networks, Recommender Systems, NLP (BERT), A/B Testing, scikit-learn, TensorFlow, PyTorch
- **Data Engineering & Analytics:** ETL/ELT, Spark, Firebase, Flask, Tableau, Power BI, Matplotlib, Seaborn
- **Cloud & Deployment:** Google Cloud (Cloud Run, Firebase Hosting), AWS, Azure, Hugging Face Spaces, CI/CD, Docker, Git/GitHub