

# 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 (Data Science & Analytics)** | 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.
- Conducted **statistical data 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 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.
- Built a **recommender system ranking pipeline** with XGBoost to generate relevance scores, delivering top recommendations with **Precision@5 = 0.88, Recall@5 = 0.86, and AUC = 0.89**; further boosted performance by ensembling with NCF.
- 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, scikit-learn, Surprise, XGBoost, Neural Collaborative Filtering, PyTorch, Hugging Face, NLTK, VADER, Streamlit

**Real Estate Investment Valuation App** | End-to-End Regression Pipeline | [GitHub Repo](#)

09/25 - Current

- Processed the Zillow Kaggle dataset with **EDA, missing value handling, and engineered 200+ domain-driven features** to enhance predictive accuracy.
- Trained and tuned regression models (**Linear, Ridge, Random Forest, LightGBM**) with cross-validation; achieved **best performance with Random Forest ( $R^2 = 0.874$ , MAE = 10.9k, RMSE = 271k, 85% within 10k of actual value)**.
- Integrated **SHAP explainability** (beeswarm, dependence, waterfall) and **Folium geospatial visualizations**; preparing deployment of an interactive **Streamlit app on AWS**.
- Tech Stack:** Python, Pandas, scikit-learn, Random Forest, LightGBM, SHAP, Folium, AWS

## TECHNICAL SKILLS:

- Languages:** Python, SQL, TypeScript, JavaScript
- Machine Learning & AI:** Regression, Classification, Ensemble Methods, Boosting, Neural Networks, Recommender Systems, NLP (BERT), scikit-learn, TensorFlow, PyTorch, Prompt Engineering (Cursor, Copilot)
- Data Engineering & Analytics:** ETL/ELT, PySpark (Spark), Hadoop, Snowflake, dbt, NoSQL (Firestore), Flask, Tableau, Power BI, Big Query, Matplotlib, Seaborn
- Cloud & Deployment:** Google Cloud (Cloud Run, Firebase Hosting), AWS, Azure, Hugging Face Spaces, CI/CD, Docker, Git, Kubernetes