# Lucia Vargas

**Boston, MA** · lucia.vargas@example.com · (617) 555-0172 · linkedin.com/in/example-lucia-vargas

## My Key Qualifications

- Prepares training data daily at Clarify Health using **Python, SQL, and pandas**, covering feature generation, labeling, and data quality checks for care operations models.
- Lifted validation **ROC-AUC from 0.71 to 0.78** training and evaluating logistic regression and XGBoost classifiers with cross-validation and hyperparameter tuning.
- Helps maintain production ML pipelines with **Docker, FastAPI, and AWS** (S3, ECS, SageMaker), working under senior engineers as the JD's level expects.
- Reduced failed scheduled runs **from 5 to 2 per month** by adding MLflow experiment tracking, GitHub Actions checks, and CloudWatch alerts.
- **PySpark** on large appointment datasets: patient no-show forecasting capstone that raised F1 from 0.58 to 0.69 with XGBoost and MLflow.
- Healthcare data throughout: ML work on member outreach and scheduling analytics at Clarify Health, plus a clinical message triage classifier that handled **4,800 messages at 87% accuracy**.
- NLP exposure via relevant coursework and a **text classification** prototype for a nonprofit health hotline, built with scikit-learn and pandas.

## Experience

### **Junior Machine Learning Engineer** — **Clarify Health Solutions**, healthcare analytics software company

May 2025 – Present · 1 yr 6 mos

- Supports senior engineers and data scientists in preparing training data, training supervised models, and maintaining ML pipelines used in care operations analytics.
- Built a reusable feature-generation workflow for member outreach propensity models using Python, SQL, and pandas, cutting training dataset preparation from 6 hours to 2 hours for weekly model refreshes.
- Trained and evaluated baseline classification models for appointment outreach prioritization, testing logistic regression and XGBoost with cross-validation and hyperparameter tuning to improve validation ROC-AUC from 0.71 to 0.78.
- Added experiment tracking and pipeline reliability checks by logging runs in MLflow and adding GitHub Actions checks plus CloudWatch alerts, reducing failed scheduled runs from 5 per month to 2.

## Education

- **Northeastern University** — Bachelor of Science in Data Science · 2021 – 2025 · Relevant coursework: Machine Learning, Database Design, Applied Statistics, Natural Language Processing, Distributed Data Systems

Sample resume made with Specific Resume · specificresume.com

# Projects

- **Patient No-Show Forecasting Capstone**
  No-show prediction workflow for a university-affiliated clinic operations team, using historical appointment records to support scheduling outreach. Raised F1 score from 0.58 to 0.69 on the final held-out set.
  **Tech:** PySpark, XGBoost, Python, MLflow
- **Clinical Message Triage Class Project**
  Text classification prototype for a nonprofit health hotline routing inbound written questions into service categories for staff review. Classified 4,800 messages with 87% accuracy in the final evaluation.
  **Tech:** Python, scikit-learn, pandas, Jupyter

# Skills

- **ML & Data:** Python, SQL, pandas, NumPy, scikit-learn, XGBoost, PySpark, Feature Engineering
- **ML Ops & Deployment:** Docker, FastAPI, MLflow, GitHub Actions, CloudWatch, Git, Jupyter
- **Cloud:** AWS (S3, ECS, SageMaker)

Sample resume made with Specific Resume · specificresume.com