Joe Bochnik

Nanuet, NY | (845) 499-0610 | joebochnik5@gmail.com | linkedin.com/in/joebochnik | joebochnik.github.io

Professional Summary

Machine Learning Engineer with 3+ years building production ML systems, from predictive models to LLM-powered applications. Experience deploying Azure AI services (Form Recognizer, Azure OpenAI) and developing end-to-end ML pipelines in both modern and legacy environments. M.S. in Data Science with thesis research applying ensemble methods, feature engineering, and probability calibration to rare-event prediction. Seeking a Data Scientist, AI Engineer, or ML Engineer role to build scalable AI solutions.

Education

CUNY School of Professional Studies – New York, NY

Master of Science in Data Science (Dec 2025) | GPA: 4.0

University at Buffalo – Buffalo, NY

Bachelor of Science in Computer Science (May 2020)
Magna Cum Laude | GPA 3.7

Technical Skills

Machine Learning: Ensemble Methods (XGBoost, CatBoost, LightGBM), Neural Networks, Supervised Learning, Feature Engineering, Probability Calibration, Model Evaluation, Walk-Forward Validation
Libraries / Frameworks: pandas, NumPy, scikit-learn, StatsModels, matplotlib, seaborn, Plotly
AI / NLP: LLM Integration, Prompt Engineering, Azure OpenAI, Azure Form Recognizer, OCR Processing
Languages: Python, SQL, R, JavaScript, ColdFusion, HTML/CSS
Data Engineering: ETL Pipelines, REST APIs, SQL Schema Design, Data Cleaning, Docker, Git
Tools/Platforms: Azure (OpenAI, Form Recognizer), VS Code, JupyterLab, GitHub, Google Maps API

Professional Experience

Roman & Associates - Uniondale, NY

Software Engineer / ML Engineer | Feb 2022 – Present

- Developed and deployed an ML-powered case-intake automation system using Azure Form Recognizer and custom classification models, reducing processing time by 50% and improving data extraction accuracy.
- Built a secure enterprise GPT platform on Azure OpenAI with SQL-backed conversation logging, enabling auditable AI-assisted workflows while maintaining data compliance.
- Created a geolocation analytics algorithm leveraging Google Maps API + cell-signal data to track task completion and generate delay alerts to managers.
- Led team migration to Git-based version control, establishing workflows and training colleagues.

Selected Projects

MLB Home-Run Prediction System | GitHub

- Built an ensemble ML pipeline comparing 5 algorithms (CatBoost, XGBoost, LightGBM, Random Forest, Neural Network) to predict rare-event outcomes across 43,000+ player-games.
- Engineered 50 features from player performance, weather data, pitcher matchups, and sportsbook odds using a 21-day rolling window.
- Implemented walk-forward validation with daily retraining and isotonic regression calibration to prevent temporal data leakage.
- Found that model profitability emerged only under restrictive high-edge thresholds, indicating that, while aggregate performance was unprofitable, selective confidence-based filtering yielded localized positive ROI.

Client Revenue & Engagement Trend Analysis Platform

- Built a data pipeline analyzing 3+ years of client billing and case activity to identify revenue trends and early engagement risk indicators.
- Integrated LLM summarization to generate executive-ready insights from complex SQL aggregations.
- Delivered actionable recommendations that informed resource allocation and client retention strategy.