

CEFERINO JASON MALABED

Data Scientist

San Francisco, CA • +1 (707) 301-9900 • cmalabed0125@gmail.com • linkedin.com/in/ceferinojmalabed • github.com/Ceferinojm

SUMMARY

Data Scientist with an M.S. in Data Science and AI, experienced in engineering autonomous LLM agents, scalable predictive models, and distributed PySpark data pipelines. Proven track record applying ML architectures (PyTorch, XGBoost, Isolation Forests) to production systems and communicating data insights to executive stakeholders.

EXPERIENCE

Data Scientist, Intern — American Civil Liberties Union of Northern California

Oct 2025 - Jun 2026

San Francisco, CA

- Engineered and deployed an autonomous, LLM-driven AI agent (Python, PyTorch) to ingest and analyze police body camera footage in real time, mapping telemetry against civil rights protocols to flag non-compliance patterns.
- Unified disparate public data silos by ingesting and cross-referencing 1M+ Automated License Plate Reader (ALPR) records with emergency call logs using PySpark and SQL.
- Implemented unsupervised anomaly detection frameworks (Isolation Forests) to benchmark compliance and equity across diverse demographics.
- Translated complex statistical anomalies and predictive pipeline findings into executive-ready data visualizations and technical frameworks to support public-sector legal action and city policy updates.

On-Site Manager / Production Assistant — S.R.G. Designs

Dec 2022 - Jul 2025

San Francisco, CA

- Served as primary technical point of contact for high-profile clients and enterprise VIP stakeholders, conducting discovery to resolve real-time operational issues in fast-paced environments.
- Streamlined complex vendor workflows and physical logistics using AutoCAD to generate precise structural and spatial layouts, increasing operational efficiency.

PROJECTS

Spiffy

Apr 2026 - May 2026

- Built a scalable web application predicting apparel resale metrics via an XGBoost regression pipeline (500,000 listings, \$30.90 MAE) integrated with a PyTorch CNN for asset classification.
- Integrated a PyTorch convolutional neural network (CNN) for automated multi-class visual asset classification to streamline listing ingestion.

Agentic Workout Planner and Coach

Oct 2025 - Present

- Designed a personalized multi-agent workflow application using LLM reasoning, prompt engineering, and context-aware orchestration to evaluate user goals against dynamic environments.
- Developed an optimization algorithm querying the Google Calendar API to ingest schedule data and execute autonomous background planning within open blocks.

EDUCATION

M.S. in Data Science and Artificial Intelligence — University of San Francisco

Jul 2025 - Jul 2026

- Relevant Coursework: Machine Learning, Deep Learning, Data Acquisition, Distributed Computing, Time Series Analysis, A/B Testing, Relational Databases, MLOps, Generative AI.

B.S. in Software Engineering — Chapman University

Aug 2014 - Dec 2017

TECHNICAL SKILLS

- Languages & Core Frameworks: Python, SQL, PySpark, Bash, Scikit-Learn, PyTorch, TensorFlow, XGBoost, Pandas, NumPy
- Machine Learning & AI: LLM Agents, Multi-Agent Orchestration, Deep Learning, NLP, Time Series Analysis, Anomaly Detection (Isolation Forests), A/B Testing, Computer Vision (CNNs)
- Data Engineering & Cloud: PySpark, Distributed Computing, ETL Pipelines, Google Cloud Platform (GCP), REST APIs, SQL Databases
- MLOps & Developer Tools: Docker, Git, CI/CD Pipelines, Model Deployment, Matplotlib, Plotly