Ricky Martin

rickylenmartin@gmail.com | (661) 808-3055 | www.linkedin.com/in/rickylm/ | portfolio | New York, NY

PROFESSIONAL EXPERIENCE

DCCA *New York, NY (Remote)*
AI/ML Engineer *July 2023 - Present*
- Led full-stack development, deployment, and monitoring of a secure internal agentic RAG platform (Bedrock, FAISS, LangGraph, LangSmith) used by 80+ employees, supporting document QA, reasoning workflows, and tool-calling.
- Operated GPU-backed SageMaker endpoints for Llama-2/3+ models, powering three enterprise proof-of-concept applications I oversaw construction of: a Document summarizer, an internal secure SQL Agent, and an interactive report generation tool.
- Owned end-to-end development of a drug Price Forecasting API (XGBRegressor) predicting 2-8 quarters ahead with 4-11% WMAPE, enabling proactive pricing decisions and saving hundreds of analyst hours annually.
- Built distributed PySpark anomaly-detection pipelines over big datasets, identifying millions of dollars in potential fraud charges.

Ardent *New York, NY (Remote)*
Mid-Level Machine Learning Engineer *January 2022 - May 2023*
- Designed and maintained 20+ containerized ML training and serving pipelines (scikit-learn, PyTorch, AWS) for microservice health models with retraining, versioning, and drift detection, reducing system downtime by 25%.
- Devised document classification pipeline for highly imbalanced data using Textract OCR, TF-IDF, and XGBoost, achieving PRAUC of 0.91 through SMOTE and feature engineering.
- Partnered with the security team to implement end-to-end time-series anomaly detection (ARIMA, Kalman Filter, Prophet) on batched Splunk logs, reducing incident detection latency by 80%.

Redica Systems *Los Angeles, CA (Remote)*
Data Scientist *October 2020 - October 2021*
- Led improvements to a semi-supervised DBSCAN document clustering system using NLP preprocessing, TF-IDF, and feature engineering, increasing precision/recall by 10% and reducing manual labeling effort by 64+ hours per month.
- Collaborated with cross-functional teams to deliver KPI dashboards and data-driven impact analysis to clients and executives.
- Gave 30+ stakeholder review meetings translating technical outcomes into business value and delivery impact.

UCSD Bio Inspired Robotics and Design Lab *San Diego, CA*
Data Science (ML) Researcher *May 2018 - June 2021*
- Developed five distinct ML-on-edge prototypes, collecting large-scale sensor data and using GPU accelerated deep learning pipelines (Python, C++, CUDA, LSTM) to model touch sensation from force sensors embedded in artificial skin.
- Built GAN-based synthetic data generator for sensor signals, improving model generalization by introducing noisy data conditions.

PROJECTS

Low Latency Agentic AI on Edge Device
- Built agentic control system for robotic platform using open-source LLMs with tool-calling optimized for Raspberry Pi.

Monitored and CI/CD-Enabled Chatbot for Personal Website Link
- Built and Deployed AI Chatbot using FastAPI, LangGraph, and AWS Bedrock with containerized CI/CD pipeline..

LLM from scratch in Python Colab

TECHNICAL SKILLS

Languages: Python, SQL, C++
Frameworks: Numpy, Pandas, Scikit-Learn, Matplotlib, Seaborn, OpenCV, TensorFlow, Keras, PyTorch, FastAPI, Pydantic, Langchain, OpenAI, Langgraph, Transformers, HuggingFace, FAISS, Streamlit, boto3
Technologies: AWS, CICD, Docker, Excel, Git, Jupyter, JSON, SageMaker, Tableau, MLFlow, DVC, AWS CLI

EDUCATION

UC San Diego *Graduated June 2021*
B.S. in Cognitive Science (Machine Learning Specialization)

PUBLICATIONS

Classification of components of affective touch using rapidly manufacturable, soft, sensor skins - IEEE-RAS 2020
Towards classification of components of affective touch using rapidly manufacturable, soft, sensor skins - RoboTac 2019