Berlin G

berlinmivi@gmail.com | linkedin.com/in/berlin-g | github.com/berlinmivi

PROFILE

Machine Learning Engineer with 3+ years of hands-on experience building and productionizing ML models across NLP, forecasting, and optimization use cases. Expertise in Python, classical ML algorithms, and deploying models via APIs, Docker, and AWS in product companies.

KEY SKILLS

Machine Learning
Linear Regression, Logistic Regression, Decision Trees, Gradient Boosting (LightGBM), Time-Series Forecasting, Recommendation Systems, Price Optimization, Feature Engineering

NLP & AI
Text Generation, Text Classification, Embeddings, Similarity Models, User Profiling, Feature Extraction, Transformers, Langchain, RAG Systems

Python & Libraries
Scikit-learn, TensorFlow, Pandas, NumPy, Matplotlib, Prophet, Pulp, LightGBM, SHAP

Productionization
FastAPI, Docker, CI/CD, REST APIs, AWS (Bedrock, Lambda, EC2, S3), Nginx, MLflow, Airflow

Data & Infrastructure
MySQL, PostgreSQL, Redis, Qdrant, ChromaDB, Git, GitHub

EDUCATION

IIT Bombay
M.Tech. Aerospace Engineering
CGPA: 8.01/10 | 2020-2022
Key Course: Introduction to Machine Learning
TA: Mentored students in ML algorithms

MIT Chennai
B.E. Aeronautical Engineering
CGPA: 8.8/10 | 2015-2019
Key Course: Computing Techniques

CERTIFICATIONS

• Introduction to Statistics - Stanford
• Credit Risk Modeling in Python
• Supply Chain Analytics - Rutgers
• ML with Python

ACHIEVEMENTS

• AIR 81 GATE Aerospace 2020

WORK EXPERIENCE

Althi Solutions | ML Engineer [Jul 2024 - Present]
RAG-Based Conversational System
• Engineered real-time NLP assistant using **FastAPI, AWS Bedrock & Qdrant** with **text generation** and **embeddings**; Implemented **semantic chunking, similarity models & feature extraction from unstructured text**.
• Productionized with **Docker on AWS EC2**, SNS-triggered webhook for updates, and **S3 knowledge base**; Integrated Grafana logging with **REST API** interface.

AML Compliance Assistant
• Built **fraud detection and risk assessment** agent using **Langgraph & AWS Bedrock**; Integrated MySQL for **transaction pattern analysis, anomaly detection & risk profiling**.
• Productionized with **FastAPI, Docker, Nginx** load balancing, and **CI/CD pipeline on AWS**; Implemented ModSecurity firewall and Redis authentication.

CloudWatch Triage BOT
• Developed **text classification system** using **Langchain React Agent**; Applied **NLP for pattern recognition, feature extraction from unstructured logs & automated incident categorization**.
• Productionized with **Docker** and Ollama; Built Streamlit interface for conversational log access.

Ikigai | ML Engineer [Jul 2022 - Jun 2024]
Inventory Optimization - Taager
• Implemented **time-series forecasting using Prophet** across 20K+ SKUs; Applied **feature engineering** for seasonal variations, promotions & geographic behavior.
• Designed **price optimization and procurement recommendation system** using **Pulp linear programming** with **gradient-based optimization**; Achieved **10% GMV increase & 9M AED savings**.

Episodic Forecasting - Spirit Halloween
• Built forecasting system for 10K+ SKUs using **gradient boosting (LightGBM), decision trees, and Prophet**; Applied **PCA** and **feature engineering** with time-series cross-validation.
• Achieved **WMAPE below 23% and $4M savings**; Utilized **SHAP values** for model interpretability and feature identification.

Time-Series Forecasting Q&A Model
• Developed RAG system using **OpenAI API** with **embeddings, similarity models & text generation**; Built knowledge base using **ChromaDB vector database** for **feature extraction from unstructured text**.
• Productionized with **Docker** and Streamlit UI; Implemented **REST API** interface for portable deployments.

PROJECTS

Response Assistant
• Fine-tuned **DeepSeek-R1 using LORA** for **text generation with user profiling**; Implemented **feature extraction from unstructured text** to mimic communication style and tone.

Airflow ETL Pipeline
• Designed end-to-end **ETL pipeline with Airflow orchestration**; Extracted via **REST API**, transformed with **Pandas**, loaded into **PostgreSQL**; Demonstrated **CI/CD** best practices.