Rohit Paradkar
Experience: 4.5 years
Education: B. Tech

SUMMARY:

- AI/ML Engineer with 4.5 years of hands-on experience in developing real-time machine learning systems using Python, Apache Kafka, and modern streaming architectures.
- Proficient in designing cloud-native, scalable solutions on AWS, Azure, and GCP, ensuring high availability and low-latency processing for IoT and financial applications.
- Strong expertise in Apache Kafka (Kafka Streams, Kafka Connect, Schema Registry) and Azure IoT Hub for building event-driven and streaming data pipelines.
- Skilled in Python for data science, machine learning, scripting, and backend development, with additional experience in SQL, JavaScript, and R.
- Experienced with ML frameworks including PyTorch, TensorFlow, Scikit-learn, Keras, Hugging Face Transformers, LangChain, and CrewAI for advanced AI/LLM-based applications.
- Applied MLOps practices using MLflow, Airflow, and Kubernetes for model monitoring, retraining, and automated CI/CD pipelines with Docker.
- Designed and developed microservices and production-grade APIs using FastAPI, Flask, and WebSockets for integrating ML models into distributed systems.
- Adept in working with PostgreSQL, MongoDB, Redis, InfluxDB, and vector databases (Pinecone, Chroma) for data storage, retrieval, and real-time analytics.
- Well-versed in scalable system architecture, with a focus on modular design, security, observability, and fault-tolerant deployment strategies.
- Delivered end-to-end AI/ML solutions across domains such as financial analytics, industrial IoT, and autonomous agent-based systems with CrewAI.
- Demonstrated ability to collaborate cross-functionally with software engineering, DevOps, and product teams to ensure successful AI system integration and performance.

TECHNICAL SKILLS:

- **Languages:** Python, JavaScript, SQL, R, HTML, CSS
- **Streaming & Messaging:** Apache Kafka, Kafka Streams, Kafka Connect, Apache Flink, Schema Registry
- **Backend & APIs:** Node.js, Flask, FastAPI, RESTful APIs, WebSockets
- **Databases:** MongoDB, PostgreSQL, Vector DBs (Pinecone, Chroma), InfluxDB
- **Streaming & IoT:** Apache Kafka, Kafka Streams, Azure IoT Hub, Event-driven architecture
- **Cloud & Infra:** AWS (EC2, Lightsail, Sagemaker), Azure, Docker, Kubernetes, CI/CD
- **ML/AI:** PyTorch, TensorFlow, Keras, Scikit-learn, Hugging Face, LangChain
- **MLOps:** MLflow, Airflow, Model Monitoring, Versioning, Automated Retraining
- **System Design:** Scalable Architecture, Security, High Availability, Monitoring & Alerting

EXPERIENCE:

Company: Techno soft Engineering
Designation: AI/ML Developer
Duration: Sep 2023– Present

RESPONSIBILITIES:

- Architected and deployed real-time machine learning pipelines for streaming financial and IoT sensor data using Apache Kafka, Azure IoT Hub, and Python, ensuring low-latency and scalable processing.
- Built enterprise-grade AI systems leveraging TensorFlow, Scikit-learn, and PyTorch, increasing deployment speed by 50% and achieving 99.9% platform resiliency across hybrid cloud environments (Azure, AWS).
- Designed and implemented end-to-end ML workflows using MLflow for model tracking and versioning, and automated deployment using Docker, Kubernetes, and CI/CD pipelines, reducing go-to-production time by half.
- Developed scalable, production-ready microservices and APIs using FastAPI and Flask, integrating trained ML models with PostgreSQL, MongoDB, and Redis for efficient data handling and real-time predictions.
- Collaborated with cross-functional engineering teams to ensure seamless backend integration, high availability, and observability across microservice-based AI platforms.
- Built interactive data exploration and monitoring dashboards to visualize streaming analytics and model performance, enabling stakeholders to make informed decisions from real-time data insights.

Company: Exposys Data Labs
Designation: AI/ ML Intern
Duration: May 2022 – Jun 2022

RESPONSIBILITIES:

- Performed financial data analysis using Python and SQL, identifying market trends and actionable investment patterns across high-volume datasets stored in PostgreSQL and MongoDB.
- Developed and deployed forecasting models using Scikit-learn and TensorFlow, delivering real-time investment insights through dashboards hosted on cloud platforms like AWS and Azure.
- Designed streaming data pipelines with Apache Kafka for real-time market data ingestion and transformation, enabling near-instant model inference and alerting.
- Automated end-to-end reporting workflows using Dockerized Python scripts and scheduled CI/CD jobs, reducing manual reporting efforts by over 40%.
- Integrated Redis for real-time caching and feature retrieval to enhance model responsiveness and support interactive dashboard updates.
- Deployed containerized ML services using Docker and orchestrated on Kubernetes, ensuring scalable and reliable performance in production environments.

Company: Entuple Technologies
Designation: AI/ML Intern
Duration: Apr 2021– Jul 2021

RESPONSIBILITIES:

- Performed large-scale data analysis and time-series forecasting using Python, applying ML algorithms from Scikit-learn and TensorFlow to uncover actionable trends.
- Built robust data pipelines using Pandas, NumPy, and Apache Kafka for real-time ingestion and preprocessing of structured and unstructured datasets.
- Prepared and transformed raw data into optimized, ML-ready formats stored in PostgreSQL and MongoDB, accelerating model training and deployment cycles.
- Leveraged Docker to containerize data processing workflows and deployed scalable training environments on AWS and Azure cloud infrastructure.
- Integrated Redis for feature caching to improve training efficiency and supported deployment of ML models within Kubernetes-orchestrated environments.

PROJECTS:

Smart Safety Hat– IoT Streaming Platform
Technologies used: Azure IoT, Kafka, FastAPI, Python, Mongodb, Node

RESPONSIBILITIES:

- Designed and deployed a distributed IoT platform that streamed real-time sensor data from multiple smart safety helmets using Azure IoT Hub, Apache Kafka, and Python.
- Configured Kafka partitioning and consumer groups to enable scalable, parallel, and fault-tolerant processing of high-frequency sensor data across multiple nodes.
- Developed and deployed real-time anomaly detection models using Scikit-learn and TensorFlow, integrated via FastAPI and served over WebSocket APIs for instant alerting.
- Optimized system performance to achieve sub-second latency for critical alerts, leveraging Redis for low-latency data access and caching.
- Stored processed telemetry data in MongoDB and PostgreSQL for downstream analytics and compliance reporting.
- Containerized microservices using Docker and deployed the solution on Kubernetes (via Azure AKS or GCP GKE) for high availability and horizontal scaling in production environments.

Financial Data Analytics Platform
Technologies used: Kafka, AWS, Python, Flask

RESPONSIBILITIES:

- Developed a real-time data streaming pipeline using Apache Kafka and Python to ingest, process, and transform high-frequency financial market feeds.
- Built and deployed online ML inference services using Scikit-learn and TensorFlow, delivering personalized financial recommendations based on user behavior and market conditions.
- Utilized PostgreSQL and MongoDB for structured storage and historical analysis, with Redis for real-time feature caching and fast data retrieval during inference.
- Containerized the full application stack using Docker and orchestrated deployment on AWS with CI/CD pipelines, ensuring seamless updates and continuous delivery.
- Ensured system scalability and reliability by integrating with Kubernetes (EKS/GKE/AKS), supporting horizontal scaling and fault-tolerant microservices for production environments.

Financial Agent System
Technologies used: CrewAI, Python, Kafka, LangChain

RESPONSIBILITIES:

- Designed and developed an autonomous multi-agent financial intelligence system using CrewAI, orchestrating specialized agents (data analyst, strategy, alert) to process real-time financial data streams.
- Leveraged Apache Kafka for event-driven architecture, enabling asynchronous, scalable communication between agents and supporting real-time investment decision-making.
- Integrated LLM-powered agents using LangChain, Hugging Face, and Python, enabling contextual reasoning for signal generation, risk scoring, and personalized financial insights.
- Engineered real-time feature pipelines and personalized recommendation logic using Scikit-learn and TensorFlow, backed by Redis for low-latency data access and PostgreSQL/MongoDB for persistent storage.
- Deployed the system as microservices containerized with Docker and orchestrated using Kubernetes on AWS/GCP, ensuring high availability, scalability, and fault tolerance in production.
- Enabled CI/CD integration and cloud-native deployment strategies for continuous delivery, model updates, and real-time system monitoring on Azure and AWS.

CERTIFICATIONS:

- Google Data Analytics Professional Certificate (2022): SQL, Tableau, R, Data Analysis
- Multi AI Agent Systems with CrewAI (2025): AI Agents, GenAI Applications, RAG, Task Automation
- Professional Development: Active participant in AI/ML, Agetic AI, DevOps