Muhammad Abdullah

Machine Learning Engineer | Computer Vision & Mobile ML Specialist

• +92 311 840 0589 • abdullah.muhammad.4315@gmail.com • Pakistan • LinkedIn • Github

Summary

Machine Learning Engineer with proven expertise in building and optimizing detection models for production deployment. Experienced in training object detection models, optimizing neural networks for mobile devices, and exporting models to Core ML format. Skilled in Python, real-time video analysis, and cross-functional collaboration with mobile developers. Comfortable with async-first workflows and delivering robust, scalable solutions that run efficiently on resource-constrained devices.

Skills

Computer Vision & Detection Models: Object Detection (YOLO, SSD, Faster R-CNN) · Real-Time Video Analysis · Image Preprocessing & Normalization · Edge Case Handling (Poor Lighting, Multiple Objects)

Model Optimization & Mobile Deployment: Model Quantization · Model Pruning · Core ML Export & Deployment · Mobile ML Optimization · Neural Engine Optimization (A14+) · Real-Time Inference Optimization · TensorFlow Lite · ONNX Model Conversion

Machine Learning: Supervised Learning · Unsupervised Learning · Recommendation Systems · NLP / NLU · Model Training & Validation · Hyperparameter Tuning · Model Evaluation & Metrics

AI & ML Frameworks: PyTorch · TensorFlow · Scikit-learn · Hugging Face · OpenCV

Backend Development: Python · FastAPI · Flask · REST APIs · AsyncIO · Microservices

Cloud & Infrastructure: AWS (Lambda, S3, EC2, SageMaker, CloudWatch) · GCP · Azure · Docker · CI/CD Pipelines

Databases & Tools: PostgreSQL · MySQL · Redis · DynamoDB · Vector Stores

Experience

Machine Learning Engineer — Jul 2025 - Present

Turing — United States, Remote

- Developed and deployed **object detection models** using **YOLO and Faster R-CNN** frameworks, optimizing for real-time inference on edge devices with **quantization and pruning techniques**, achieving **sub-200ms latency** while maintaining 85%+ accuracy.
- Exported trained PyTorch and TensorFlow models to **Core ML format** for iOS deployment, ensuring compatibility with **A14 Neural Engine** and enabling **30 FPS real-time performance** on target devices.
- Built **video analysis pipelines** for detecting **moving objects and skeletal keypoints** across sequential frames, implementing post-processing logic and confidence thresholds to handle edge cases in varying lighting conditions.
- Collaborated asynchronously with iOS developers via Slack and email to define **model input/output specifications**, recommend **preprocessing strategies** (image normalization, resizing), and troubleshoot integration issues.
- Created comprehensive **technical documentation** including model performance benchmarks, integration guides, and preprocessing requirements, reducing integration time for mobile teams by approximately **40%**.

AI Engineer — Jul 2024 - Jun 2025

Mindrift — Austria, Remote

- Trained and optimized **machine learning models** for supervised and unsupervised learning tasks, deploying via **FastAPI microservices** on **AWS/GCP** infrastructure to support **real-time inference pipelines**.
- Implemented **model optimization techniques** including **quantization and pruning** to reduce model size by **30-45%** while maintaining accuracy, enabling deployment on memory-constrained environments.
- Integrated trained models with **external APIs and tools**, building robust **data processing workflows** that handled **15-20 hours weekly** of automated task execution with minimal manual oversight.
- Maintained **scalable backend infrastructure** with **PostgreSQL** and **Redis** caching for **sub-100ms inference response times**, supporting production ML pipelines with high reliability.
- Collaborated with cross-functional teams using **async-first communication patterns** via Slack and email, providing timely technical guidance on model deployment and performance optimization.

Software Engineer Jan 2023 - Jun 2024

Asendia AI United States, Remote

- Developed **supervised and unsupervised machine learning models** in Python, integrated into **AWS/GCP production pipelines**, achieving **20-35% improvement** in model prediction accuracy and user engagement metrics.
- Built and optimized **data processing workflows** using **Pandas and NumPy**, preparing datasets for model training with proper **normalization, feature scaling, and augmentation techniques** to enhance model robustness.
- Designed backend services using **Flask and FastAPI** with **event-driven patterns**, connected to **PostgreSQL/MySQL databases** and **Redis caching**, ensuring **sub-200ms response times** and high availability.
- Participated in **Agile methodologies** including sprint planning and code reviews, collaborating with data scientists and frontend engineers to deliver **iterative ML features** on tight schedules.
- Implemented **continuous integration and deployment pipelines** to automate model training, validation, and deployment, reducing iteration cycles and supporting **scalable experimentation** across datasets.

Python Developer Feb 2020 - Dec 2022

Infostack Lahore, Pakistan

- Built **data processing pipelines** using **Python (Pandas, NumPy)** for cleaning, analysis, and transformation, supporting **data science workflows** that informed product decisions and improved operational efficiency by **40-60%**.
- Developed **robust backend APIs** with **Flask and FastAPI**, integrated with **PostgreSQL/MySQL** and **Redis caching**, supporting **scalable web applications** and reducing **endpoint latency by 25-35%**.
- Automated **repetitive data tasks** including ETL, file processing, and reporting, streamlining workflows and saving team members **10-15 hours weekly** with reliable Python automation scripts.
- Created **web scraping and data extraction pipelines** using **Python libraries (BeautifulSoup, Scrapy, Selenium)**, enabling **real-time data collection** that accelerated **reporting cycles by 40-60%** for business intelligence teams.
- Collaborated in **Agile environments**, participating in **stand-ups and task grooming** to ensure **timely delivery of backend features, data pipelines, and automation solutions**.

Projects

Product Searching and Recommendation System | Python, Keras, VGG16, ResNet, MongoDB, AWS, Angular

- Built an image-based product search and recommendation system using deep learning feature extraction and similarity matching.

Formula 1 Winning Driver Predictive Model | Python, Random Forest, Pandas, Matplotlib, AWS Lambda, S3, EC2, DynamoDB

- Developed a supervised learning model to predict Formula 1 race outcomes using historical and real-time data.

Conversational Voice Agent (Healthcare) | LiveKit, LangChain, Whisper STT, N8n, AWS Lambda, DynamoDB, Socket.io, PyTorch

- Built a **real-time healthcare conversational voice agent** to capture doctor-patient conversations, perform **low-latency speech-to-text**, extract medical entities, and generate **structured SOAP notes** with **EHR integration**.

Algorithmic Trading System (Real-Time Stream) | Python, AsyncIO, WebSockets, pandas, Binance WebSocket, NumPy

- Built a **fully autonomous, low-latency trading system** operating **24/7** using real-time candle streams and custom **Supertrend and EMA indicators**.

Education

University of the Punjab Lahore, Pakistan

Bachelor of Science in Computer Science 2016 - 2020

Certifications

IBM RAG and Agentic AI Specialization IBM (Coursera) – 2025

View Credential

Image Processing in Python DataCamp – 2023

View Credential

Supervised Machine Learning Coursera – 2022

View Credential