

■ AI/ML Study Roadmap — Focused on Training, Model, Inference, Quantization

■ Phase 1: Core ML Foundations (2–3 weeks)

Goal: Understand what happens in an ML pipeline.

Topics to Learn

- What is Machine Learning?
- Types: Supervised, Unsupervised, Reinforcement
- Dataset basics: features, labels, training/testing split

Mathematics for ML

- Linear algebra, Probability, Calculus basics

Python for ML

- numpy, pandas, matplotlib

Mini Exercises: Load a CSV, plot graph, compute mean

Resources: Google ML Crash Course, Python Machine Learning (Raschka)

■ Phase 2: Training & Model Fundamentals (3–4 weeks)

Goal: Learn how models are trained and stored.

- Forward & backward propagation
- Loss, optimization, overfitting
- Model representation (.pkl, .h5, .onnx)
- Frameworks: scikit-learn, TensorFlow, PyTorch

Hands-on: Train Linear Regression, save model, train image classifier

Goal: Understand how training creates and saves a model file.

■ Phase 3: Inference (2–3 weeks)

Goal: Learn how trained models make predictions.

- Difference between training & inference
- Inference pipeline (preprocess → predict → postprocess)
- Real-time vs batch inference

Hands-on: Load model, make predictions, measure latency

Goal: Understand real-time inference on devices.

■ Phase 4: Quantization & Optimization (3–4 weeks)

Goal: Make models smaller and faster.

- Floating point vs integer (FP32 → INT8)
- Post-training quantization, QAT
- Convert to .tflite, compare inference time

Goal: Optimize for embedded hardware like ESP32, STM32.

■ Phase 5: Hardware-Aware Learning (Optional Advanced)

Goal: Connect theory to practice.

- Edge AI / TinyML
- Frameworks: TensorFlow Lite Micro, Edge Impulse
- Hardware inference (ESP32, STM32, Raspberry Pi)

■ Tools to Learn During This Roadmap

Category	Tools / Libraries
Data & Math	numpy, pandas, matplotlib, seaborn
Machine Learning	scikit-learn
Deep Learning	TensorFlow, PyTorch
Optimization	TensorFlow Lite, ONNX, OpenVINO
Hardware	Edge Impulse, Arduino ML, TFLite Micro

■ Recommended Timeline

Phase	Focus	Duration
1	Core ML Foundations	2–3 weeks
2	Training & Model	3–4 weeks
3	Inference	2–3 weeks
4	Quantization	3–4 weeks
5	Hardware Practice	Optional