Module 9: ML Deployment

Data Science with Python – BSc Course

Why ML Deployment?

Netflix deploys thousands of ML models in production. Spotify's Discover Weekly serves 500 million personalized playlists. Yet 87% of ML projects never make it past the prototype stage. The gap between a working notebook and a production system is where most data science projects die.
Goldman Sachs' Marquee platform serves models via APIs. Stripe's ML fraud detection handles millions of API calls daily. BlackRock's Aladdin deploys risk models at scale. The message is clear: deployment is where data science creates business value.

A model in production is worth infinitely more than a model in a notebook

Why This Matters

- **Business value:** A model in a notebook earns nothing; deployed models serve thousands of users
- **Scalability:** Serve predictions to stakeholders without requiring Python knowledge
- **Reliability:** Monitoring, versioning, and rollback capabilities ensure production stability
- **Communication:** Dashboards make models accessible to non-technical stakeholders

Deployment is the bridge from data science to real-world impact

What You'll Learn

By the end of this module, you will be able to:

- Save and load trained models reliably using pickle and joblib
- Build REST APIs to serve predictions using FastAPI
- Create interactive dashboards for stakeholders using Streamlit
- Deploy ML applications to the cloud with Docker and CI/CD
- Implement model versioning and monitoring in production

From local prototype to cloud-deployed production system

Lesson Roadmap

Lesson	Topic	Focus
L41	Model Serialization	Pickle, joblib, model versioning
L42	REST APIs with FastAPI	Endpoints, request/response
L43	Streamlit Dashboards	Interactive web apps
L44	Cloud Deployment	Docker, CI/CD, cloud platforms

Each lesson takes you one step closer to production-ready systems

Key Concepts

- **Model Serialization & Versioning** – Save trained models and track versions over time
- **REST APIs with FastAPI** – Expose models as HTTP endpoints for programmatic access
- **Interactive Dashboards (Streamlit)** – Build web apps that stakeholders can use without code
- **Cloud Deployment & CI/CD** – Automate deployment and updates to cloud platforms

Deployment skills turn you from researcher into engineer

Finance Application

Scenario: Real-Time Risk Dashboard
Using the skills from this module, you will build a system that:

- Serializes a trained risk model (VaR or portfolio optimization)
- Exposes it via FastAPI so traders can query risk metrics programmatically
- Builds a Streamlit dashboard showing portfolio risk in real-time
- Deploys to the cloud so the risk team can access it from anywhere

This is exactly how banks deliver risk analytics to trading desks.

Transform a research model into a production risk system

Who Uses This?

- **Big Tech:** Netflix and Spotify deploy thousands of ML models serving billions of predictions
- **Investment Banks:** Goldman Sachs Marquee platform serves models via APIs to clients
- **Fintech:** Stripe's ML fraud detection serves millions of API calls daily with ¡100ms latency
- **Asset Management:** BlackRock Aladdin platform deploys risk models at scale for institutions

Deployment is where ML meets business value across all industries

What's Next: Module 10 – Capstone & Ethics

You have all the skills. Now bring them together in a real project — and learn why responsible AI matters in finance.
Module 10 covers end-to-end project work, presentation skills, and ML ethics: bias detection, fairness requirements, explainability, and regulatory compliance.
Prerequisite check: Can you build, evaluate, deploy, and monitor an ML model? If yes, you are ready to lead a data science project — and do it responsibly.

Technical skills plus ethical awareness make you a complete data scientist

Let's Begin!

First up: L41 – Model Serialization

Open your laptop and follow along.