AI Engineering Internship Assignment: Model Deployment with FastAPI

Topic Chosen: Deploying Machine Learning Models with FastAPI

This assignment will test your ability to:

- Learn a technical AI engineering concept (model deployment).
- Teach it back to your peers using a clear, well-structured presentation.
- Demonstrate deep technical understanding through a full code walkthrough.
- Solve a real-world challenge and communicate your solution effectively.

1. Learning Resources

Use the following curated references to master the topic:

1. Official Documentation

- **FastAPI Official Docs:**
  https://fastapi.tiangolo.com
  The authoritative guide for FastAPI, including tutorials, deployment, and advanced features1.
- **FastAPI Docs (DevDocs mirror):**
  https://devdocs.io/fastapi/
  Fast, searchable documentation for FastAPI2.

2. Beginner-Friendly Tutorials

- **DataCamp FastAPI Tutorial:**
  https://www.datacamp.com/tutorial/introduction-fastapi-tutorial
  Step-by-step introduction to FastAPI for building APIs, with plenty of code examples3.
- **GeeksforGeeks FastAPI Introduction:**
  https://www.geeksforgeeks.org/python/fastapi-introduction/
  Covers FastAPI basics, setup, and a simple example API4.
- **FastAPI ReadTheDocs:**
  https://fastapi-tutorial.readthedocs.io
  Community-maintained FastAPI documentation and guides5.

3. Model Deployment with FastAPI (Blogs & Guides)

- **Deploying ML Models with FastAPI and Docker (Dev.to):**
  https://dev.to/code_jedi/machine-learning-model-deployment-with-fastapi-and-docker-llo
  Detailed walkthrough for deploying a scikit-learn model as an API, including Dockerization6.
- **Founding Minds: FastAPI & Docker for ML Deployment:**
  https://www.foundingminds.com/deploying-ml-models-with-fastapi-and-docker/
  Explains deployment best practices, AWS integration, and production tips7.
- **GeeksforGeeks: Deploying ML Models as API using FastAPI:**
  https://www.geeksforgeeks.org/machine-learning/deploying-ml-models-as-api-using-fastapi/
  Step-by-step guide for deploying a model as an API, with code and explanations8.
- **TestDriven.io: Deploying and Hosting a ML Model with FastAPI:**
  https://testdriven.io/blog/fastapi-machine-learning/
  Covers deploying a stock prediction model with FastAPI and Heroku, including Docker and CI/CD9.
- **Uptrace: How to Use FastAPI [Detailed Python Guide]:**
  https://uptrace.dev/blog/python-fastapi
  In-depth FastAPI guide with advanced features and deployment advice10.

4. Video Tutorials

- **Deploy ML Models as APIs with FastAPI (Full YouTube Tutorial):**
  https://www.youtube.com/watch?v=0sOvCWFmrtA
  Comprehensive, hands-on video for deploying ML models with FastAPI11.
- **FastAPI Tutorial: Build a REST API in 15 Minutes:**
  https://www.youtube.com/watch?v=iWS9ogMPOI0
  Quick, practical intro to building APIs with FastAPI12.
- **Deploy ML Models with FastAPI, Docker, and Heroku:**
  https://www.youtube.com/watch?v=h5wLuVDr0oc
  End-to-end deployment, including Dockerization and cloud deployment11.
- **FastAPI Deployment Tutorials Playlist:**
  https://www.youtube.com/playlist?list=PLZoTAELRMXVPgsojPOHF9i0u2L83-m9P7
  Multiple videos on API creation and ML model deployment13.
- **How to Deploy ML Solutions with FastAPI, Docker, & AWS:**
  https://www.youtube.com/watch?v=pJ_nCklQ65w
  Full-stack deployment, including cloud integration and best practices14.

5. Advanced/Production Deployment

- **FastAPI in Containers (Docker):**
  https://fastapi.tiangolo.com/deployment/docker/
  Official guide to containerizing FastAPI apps for production15.

- **Northflank: Step-by-Step ML Model Deployment:**
  https://northflank.com/blog/how-to-deploy-machine-learning-models-step-by-step-guide-to-ml-model-deployment-in-production
  Covers containerization, CI/CD, and infrastructure setup16.

6. IDE/Development Environment

- **VS Code FastAPI Tutorial:**
  https://code.visualstudio.com/docs/python/tutorial-fastapi
  How to build, debug, and run FastAPI apps in Visual Studio Code17.

7. Real-World Use Case Example

- **Deploying ML Models with FastAPI and Azure:**
  https://datastud.dev/posts/ml-fastapi/
  Walks through deploying a model with FastAPI on Azure, including project structure and CI/CD basics18.

2. Assignment Instructions

A. Learn & Summarize

- Study the above resources to understand:
  - What is FastAPI and why is it used for model deployment?
  - How do you build, test, and deploy an API for a machine learning model?
  - Best practices for productionizing ML models.

B. Teach the Concept

- Prepare a **PPT (8–12 slides)** that:
  - Explains FastAPI and its relevance in AI engineering.
  - Illustrates the steps to deploy a machine learning model as an API.
  - Uses diagrams, code snippets, and real-world analogies.
  - Is tailored for new joiners with basic Python/ML background.

C. Code Walkthrough

- Choose a simple ML model (e.g., Iris classifier, sentiment analysis).
- Build and save the model in Python (using scikit-learn or similar).
- Develop a FastAPI application to serve predictions.
- **Full Code Walkthrough:**
  - Explain each section: model loading, API endpoints, request/response structure, error handling, and testing.

    - Discuss the math behind the model (e.g., how logistic regression or decision trees work).
    - Highlight deployment considerations (e.g., Docker, scaling, monitoring).

D. Real-World Challenge

Scenario:
A client wants to automate flower species identification from petal/sepal measurements (Iris dataset). They need a REST API that receives measurements and returns the predicted species.

- Build the solution end-to-end:
    - Train a model on the Iris dataset.
    - Deploy it as a FastAPI service.
    - Demonstrate the API with sample requests.

E. Video Submission

- Record a **video [NO time constraint]** :
    - Teach the FastAPI deployment concept using your PPT.
    - Walk through your code in detail, explaining logic, math, and design choices.
    - Demo your API solving the real-world challenge.
- Upload the video to Google Drive (or similar) and share a link with **download or read access so that we could evaluate it.**

3. Submission Checklist in the drive

- PPT file (Google Slides or PDF)
- Well-commented code (GitHub or zip)
- Video link (Google Drive with download/read access)
- Short summary (2–3 sentences) of your project and learnings [.docx]

4. Evaluation Criteria

- Clarity and depth of concept explanation
- Quality and accuracy of code walkthrough (including math and technical reasoning)
- Effectiveness of teaching (PPT and video)
- Relevance and completeness of real-world solution
- Communication skills and ability to simplify technical content

Tip:
Teaching a concept simply is the best test of your own understanding. Focus on clarity, structure, and real-world relevance.