AMPBA-Batch 24 Term-2
CT1 MLOPS Individual Assignment

This deliverable has <u>60% weightage</u> in the Consolidated Score Sheet.

<u>Due Date</u>: 14th Dec 2025, 11:55 PM

Assignment Submission Form must be **attached separately** with the name and PGID

<u>**General Instructions**</u>:

1. This is a **Individual assignment**. Please adhere to the details posted on LMS.
2. **Do NOT submit .zip files**.
3. Please note that both deliverables along with the video **code files you have used (.ipynb/collab) are mandatory** for evaluation.
4. Code files rendered/exported as pdfs will **strictly not be considered for evaluation**.
5. Any late submission will attract a penalty as mentioned in the course outline.
6. **Please adhere to the given instructions, otherwise, your submission will not be accepted, or a severe penalty will be applied**.
7. The Honor Code for this submission is **2N-b.**
8. Upload your submission to the section **"Individual Assignment"** submission Link on LMS.
9. **Email submissions are NOT allowed**. All the submissions must be made on LMS.
10. There is no penalty for early submissions!

Deliverables (Submit All of the Following):

1. **Video Demonstration (Max 3 mins)**
    - Covers model building, Docker setup, AWS deployment, and inference
    - Voice-over is mandatory
    - Submit on LMS
2. **Code Files**
    - Model code (.ipynb or .py), Dockerfile, API script (app.py), requirements.txt, saved model file
3. **ReadMe File**
    - Short summary of the problem, model, Docker steps, AWS setup, and how to test the endpoint
4. **Optional (Recommended):**
    - Screenshots of Docker, AWS instance, and API inference
    - Sample request/response JSON for testing

Mention all these screenshots, in the PDF report. PDF report should be very short with Steps you followed, screenshots not exceeding 2-3 pages.

ASSIGNMENT

MLOps Assignment: Model Deployment and Inference Using Docker and AWS

Assignment Title: End-to-End MLOps Workflow: Model Development, Containerization, Deployment, and Inference

Submission Type: Video Demonstration with Voice-over (Max 3 Minutes)

Tools Expected: Python, Docker, AWS (EC2 or other relevant services)

Business Problem Statement

Business Problem:

A leading bank runs marketing campaigns (like phone calls, emails, or messages) to offer financial products such as term deposits. However, reaching out to every potential client is expensive and time-consuming. Despite strong outreach, only a small percentage of customers actually accept the offer.

The core business challenge is:

"How can the bank improve the efficiency of its marketing promotions by identifying the right customers to target—those most likely to accept the offer—thereby reducing costs and increasing conversion rates?"

Data set: Provided on LMS.

LMS Upload Limit: Maximum limit of all files combined that you can upload is 200MB.'

Objective:

This assignment is designed to give students hands-on experience with the complete **MLOps pipeline**—from developing a machine learning model locally, to containerizing it using Docker, deploying it on a cloud platform (AWS), and finally performing inference using the deployed model.

You are a **business problem** and a **dataset**. Your task is to solve the business problem by building an appropriate ML model, and then go through the entire MLOps cycle as outlined below.

Assignment Tasks & Instructions:

Part 1: Model Development (Local Machine)

1. Understand the business problem and explore the given dataset.

2. Perform any necessary data cleaning, preprocessing, and EDA.

3. Train an appropriate Machine Learning model to solve the problem.

4. Evaluate the model's performance using appropriate metrics.

Part 2: Containerization Using Docker

5.   Package your code and model into a Docker container.

6.   Write a Dockerfile to define your container environment.

7.   Ensure that the container exposes a REST API (e.g., using Flask or FastAPI) that can accept input and return predictions.

Part 3: Deployment to AWS

8.   Launch an instance on AWS (preferably EC2).

9.   Transfer your Docker image/code to the AWS instance.

10. Run the Docker container on the cloud.

11. Make sure the deployed model is accessible via public IP for inference (test using Postman or curl).

Part 4: Inference and Final Testing

12. Send test inputs to the deployed model endpoint.

13. Collect and interpret the responses.

14. Verify that the results are consistent with your local predictions.

Video Demonstration

You are required to **record a video (maximum 3 minutes)** that captures the key steps of your assignment:

- Brief overview of the business problem and dataset.

- Model training and local testing.

- Docker container creation and running.

- Deployment on AWS.

- Inference using the live endpoint.

- **Voice-over is mandatory** explaining what you're doing in each step.

Note: Keep your video crisp and concise. Focus on key steps and proof of successful deployment and inference.