Assignment 2: Experiment Tracking and Model Lifecycle Management with MLflow

Overview

This assignment focuses on understanding and applying **MLflow** as a system for managing the machine learning lifecycle. You will design, track, compare, package, and register machine learning models using MLflow, emphasizing **reproducibility, governance, and best practices** rather than raw model performance.

Total Points: 100

Problem Context

You are working as an ML engineer on a small team building a predictive model (e.g., classification or regression). Multiple model variants will be trained with different hyperparameters. Your task is not only to train models, but also to ensure that experiments are traceable, reproducible, and deployable.

You may choose **any standard dataset** (e.g., Iris, Wine, Breast Cancer, California Housing), but your workflow must be fully managed using MLflow. Go to Kaggle to choose a dataset.

Part A: Experiment Design and Setup (10 points)

Tasks

- Create an MLflow experiment with a meaningful name.
- Clearly state the modeling task and dataset used.
- Define at least **three tunable hyperparameters**.

Deliverables

- Short written explanation (¡5 sentences)
- Screenshot of MLflow experiment page

Rubric

Clear problem definition and dataset choice	4
Well-defined hyperparameters	3
Proper MLflow experiment setup	3
Total	10

Part B: MLflow Tracking (25 points)

Tasks

For **at least 5 runs**:

- Log all hyperparameters using MLflow
- Log at least two evaluation metrics
- Log at least one artifact (e.g., model file, plot, confusion matrix)
- Use tags to describe run purpose (e.g., baseline, tuned etc)

Concept Focus

This section evaluates your understanding of:

- Runs vs experiments
- Parameters vs metrics
- Artifact management

Rubric

Correct parameter logging	7
Meaningful metric logging	7
Artifact logging	5
Effective use of tags	6
Total	25

Part C: Experiment Comparison and Analysis (15 points)

Tasks

- Compare all runs using the MLflow UI
- Identify the best-performing model
- Explain *why* this model is preferred (not just metrics)

Deliverables

- Screenshot of comparison table
- Written analysis (1 paragraph, ¡10 sentences)

Rubric

Correct identification of best run	5
Justification using metrics and reasoning	7
Clear comparison discussion	3
Total	15

Part D: Model Packaging and Signatures (20 points)

Tasks

- Log the best model using MLflow Models
- Include a model signature (input and output schema)
- Provide an input example

Concept Focus

This section assesses your understanding of:

- Model portability
- Training–serving skew
- Schema enforcement

Rubric

Model logged correctly	8
Valid model signature	7
Input example included	5
Total	20

Part E: Model Registry and Lifecycle Management (20 points)

Tasks

- Register the best model in the MLflow Model Registry
- Create at least two versions of the model
- Transition one model to **Staging** and one to **Production**
- Add comments explaining each transition

Rubric

Model registered correctly	6
Versioning demonstrated	6
Stage transitions performed	5
Clear governance rationale	3
Total	20

Part F: Best Practices Reflection (10 points)

Tasks

Write a short reflection (¡10 sentences) answering:

- What problems does MLflow solve in real ML systems?

- What mistakes could still occur even when using MLflow?
- How would MLflow integrate with tools like DVC or CI/CD pipelines?

Rubric

Conceptual depth	4
Accuracy of explanations	4
Clarity and organization	2
Total	10

Submission Instructions

- one zip file containing all the codes and model artifacts
- a separate pdf file named firstname_lastname.pdf with your written answers and screenshots of different stages of the work.
- Make sure to be able to show any stage of work if I ask you to. Failure to do so will result in heavy penalty.

Academic Integrity

You may discuss concepts, but all submitted work must be your own. Direct copying of code or reports is not permitted. You are expected to be able to answer any questions I may ask after you submit the assignment. I reserve the right the penalize in any way I like if I do not see your answers matching your assignments.