Data Science Project Lifecycle with MLflow

---

1 Data Science Project Lifecycle (with MLflow Integration)

1. **Problem Definition**
   - Identify the business problem.
   - Define success metrics.
2. **Data Collection & Preprocessing**
   - Gather data from multiple sources.
   - Handle missing values, outliers, feature engineering, etc.
   - **MLflow:** Use MLflow Tracking to log dataset versions.
3. **Exploratory Data Analysis (EDA)**
   - Visualizing patterns & correlations.
   - Feature selection and hypothesis formulation.
   - **MLflow:** Log data insights and transformations.
4. **Model Training & Experimentation**
   - Train multiple models (Linear Regression, Random Forest, etc.).
   - Hyperparameter tuning using GridSearch, RandomizedSearch.
   - **MLflow:** Track different model experiments with `mlflow.log_param()` and `mlflow.log_metric()`.
5. **Model Evaluation**
   - Evaluate performance using metrics (RMSE, R^2, Precision, Recall).
   - Compare models to find the best one.
   - **MLflow:** Log evaluation metrics using `mlflow.log_metric()`.
6. **Model Deployment**
   - Deploy the best model using Flask, FastAPI, or a cloud platform.
   - **MLflow:** Use MLflow Model Registry for version control and deployment.
7. **Model Monitoring & Maintenance**
   - Continuously track model performance over time.
   - **MLflow:** Use MLflow Model Serving to track real-time performance.

---

2 MLflow Commands for Running the Project

```
# Install MLflow if not already installed
pip install mlflow

# Initialize MLflow Tracking
mlflow run .

# Set an experiment in MLflow
mlflow.set_experiment("House Price Prediction")

# Run an MLflow tracking UI (to visualize experiments)
```

```
mlflow ui

# Run MLflow in a specific directory (if needed)
mlflow run <path_to_project>

# Register the best model
mlflow.register_model("runs:/<run_id>/model", "House_Price_Model")

# Serve the MLflow model
mlflow models serve -m models:/House_Price_Model/Production -p 5001
```
