Bài thực hành cơ bản nhất

Dưới đây là một **bài thực hành MLflow cơ bản nhất**, cùng với **hướng dẫn cách xem giao diện MLflow UI**.

Mục tiêu:

- Hiểu cách ghi lại (log) các tham số, metric và mô hình bằng MLflow.
- Chạy MLflow UI để xem kết quả trực quan.

I. Cài đặt MLflow

```
python3 -m venv venv
source venv/bin/activate


pip install mlflow scikit-learn pandas
```

II. Tạo file mlflow_basic.py

```
# mlflow_basic.py


import mlflow
import mlflow.sklearn
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
from sklearn.datasets import load_diabetes
from sklearn.model_selection import train_test_split
import pandas as pd

# Load dataset
data = load_diabetes()
X = pd.DataFrame(data.data, columns=data.feature_names)
y = pd.Series(data.target)


# Train-test split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Bắt đầu một MLflow run
with mlflow.start_run():

    # Tham số mô hình
    n_estimators = 100
    max_depth = 4

    # Log parameters
    mlflow.log_param("n_estimators", n_estimators)
    mlflow.log_param("max_depth", max_depth)

    # Train model
    model = RandomForestRegressor(n_estimators=n_estimators, max_depth=max_depth)
    model.fit(X_train, y_train)

    # Predict & evaluate
    predictions = model.predict(X_test)
    rmse = mean_squared_error(y_test, predictions, squared=False)

    # Log metrics
    mlflow.log_metric("rmse", rmse)

    # Log mô hình
    mlflow.sklearn.log_model(model, "model")

    print(f"Done! RMSE: {rmse}")
```

III. Chạy file

```
python mlflow_basic.py
```

MLflow lưu trữ các kết quả trong thư mục mlruns (mặc định)

Nếu gặp cảnh báo

> warnings.warn(2025/04/23 00:10:40 WARNING mlflow.models.model: Model logged without a signature and input example. Please set `input_example` parameter when logging the model to auto infer the model signature.

Cập nhật đoạn `log_model()` như sau:

```
import numpy as np

# Log mô hình kèm input_example
mlflow.sklearn.log_model(
    model,
    artifact_path="model",
    input_example=X_test.iloc[:5],  # hoặc: X_test[:1]
    signature=mlflow.models.infer_signature(X_test, predictions)
)
```

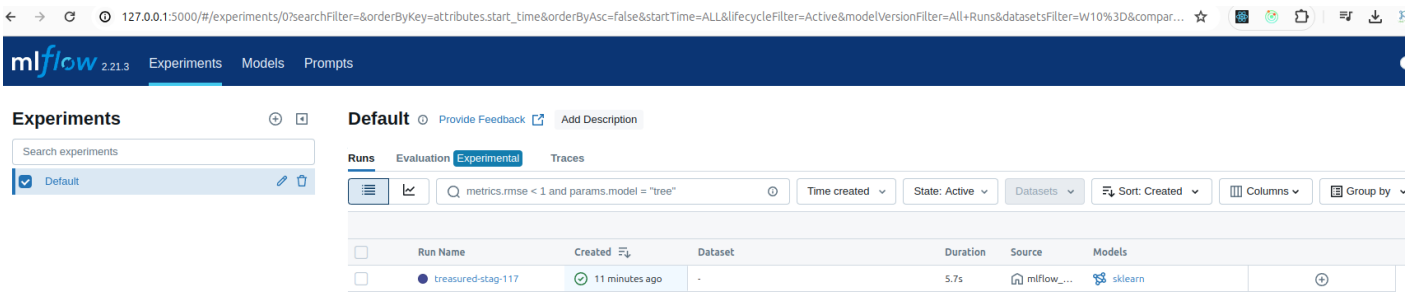
Giải thích:

- `input_example`: Một ví dụ dữ liệu đầu vào cho mô hình. MLflow dùng nó để minh họa cách input phải như thế nào.
- `signature`: MLflow sẽ tự động suy luận kiểu input/output của mô hình.

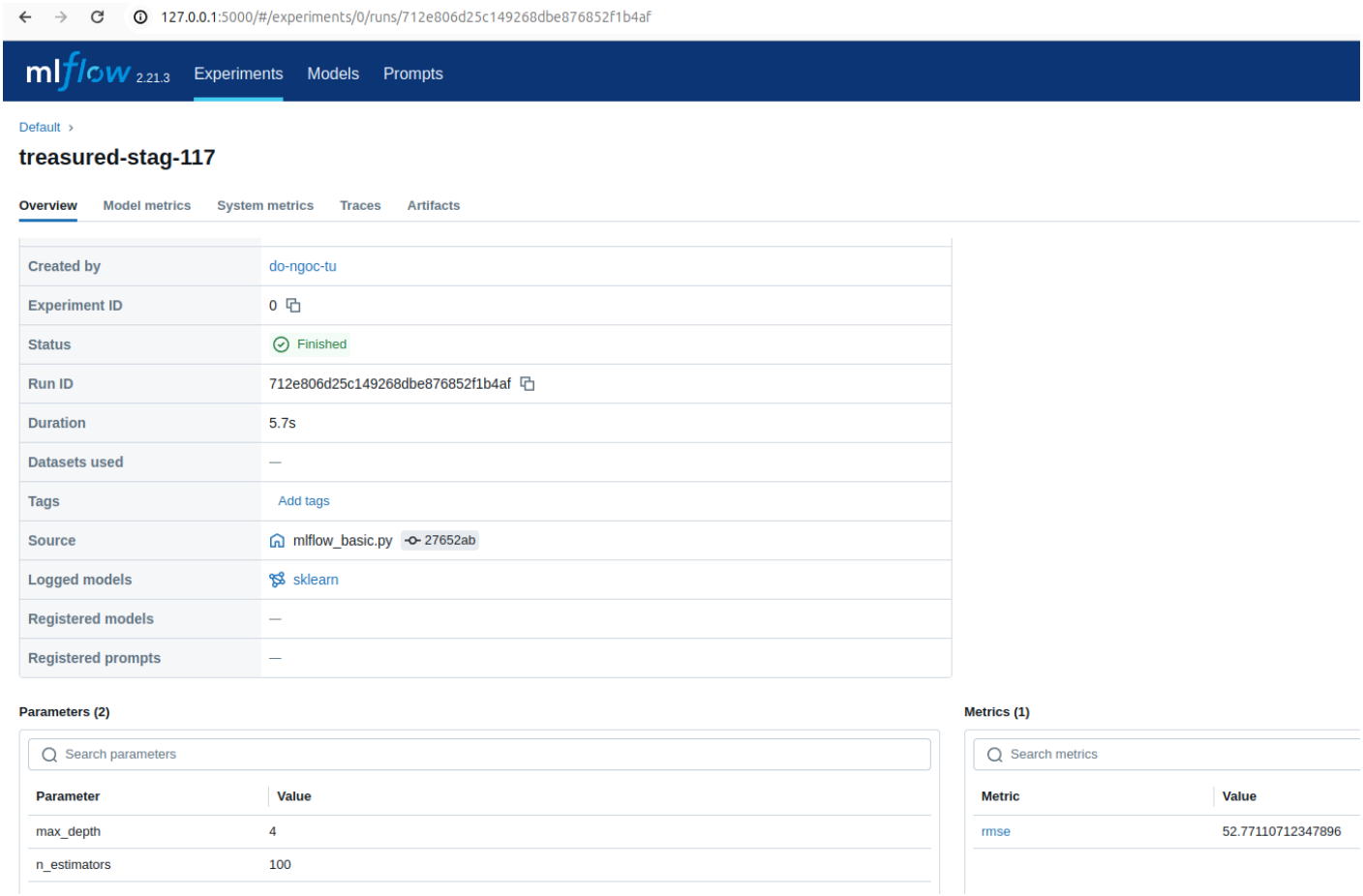
IV. Xem giao diện MLflow UI

```
mlflow ui
```

MLflow sẽ chạy trên `http://127.0.0.1:5000` (mặc định). Mở trình duyệt và truy cập vào địa chỉ đó.



Click vào **treasured-stag-117**, bạn sẽ thấy



V. Tổng quan: "Log" trong MLflow nghĩa là gì?

Trong MLflow, "log" nghĩa là **ghi lại và lưu trữ** các thông tin như:

Loại thông tin	Ví dụ	MLflow gọi là
Tham số (số lớp, số cây, learning rate, v.v.)	`n_estimators=100`	`log_param`
Kết quả đánh giá mô hình	`rmse=54.772`	`log_metric`
Mô hình đã huấn luyện	file `.pkl` hoặc `.joblib`	`log_model`

Ở ví dụ trên dùng `RandomForestRegressor`, mình sẽ ghi lại:

1. **Tham số (Parameters)**

```
mlflow.log_param("n_estimators", n_estimators)
mlflow.log_param("max_depth", max_depth)
```

Ghi lại cấu hình mô hình để sau này dễ tái hiện.

2. **Metric (hiệu suất mô hình)**

```
rmse = mean_squared_error(y_test, predictions, squared=False)
mlflow.log_metric("rmse", rmse)
```

Ghi lại giá trị RMSE để so sánh nhiều mô hình với nhau.

3. **Ghi lại mô hình đã huấn luyện**

```
mlflow.sklearn.log_model(model, "model")
```

MLflow sẽ lưu mô hình để sau này có thể load lại, dùng để deploy, hoặc tái huấn luyện.

Tất cả đặt trong 1 "Run"

MLflow cần phải có 1 "chạy thử nghiệm" (`run`) để lưu trữ thông tin:

```
with mlflow.start_run():
    # log_param()
    # log_metric()
    # log_model()
```

Tóm tắt chúng ta đã học

Bạn muốn...	**Dùng hàm...**
Ghi lại một tham số	`mlflow.log_param(name, value)`
Ghi lại một kết quả đánh giá	`mlflow.log_metric(name, value)`
Ghi lại mô hình đã huấn luyện	`mlflow.sklearn.log_model(model, "model")`
Bắt đầu một "chạy thử nghiệm"	`with mlflow.start_run():`

Tác giả: Đỗ Ngọc Tú

Công Ty Phần Mềm VHTSoft

Phiên bản #1
Được tạo 22 tháng 4 2025 17:13:54 bởi Đỗ Ngọc Tú
Được cập nhật 22 tháng 4 2025 17:40:11 bởi Đỗ Ngọc Tú