Experiment Tracking with MLflow

Objective: Set up and use MLflow to track machine learning experiments, compare model performances, and manage experiment metadata.

Step:

1. Created and cloned the repository: Meh39/MLOps-Assignemnts.

2. Initialize the project structure as follows

```
├── data/
├── notebooks/
├── scripts/
├── mlruns/ # MLflow logs and experiment metadata
├── requirements.txt
├── train.py
├── README.md
```

Where train.py is the main training script
And mlruns directory stores MLflow logs and experiment metadata

3. Setup MLflow by installing the necessary packages by creating a requirements.txt file.

```
pandas
seaborn
scikit-learn
mlflow
```

4. Installed them in the virtual environment:

```
pip3 install -r requirements.txt
```

5. Trained two models: Linear Regression and Random Forest on the seaborn 'titanic' dataset

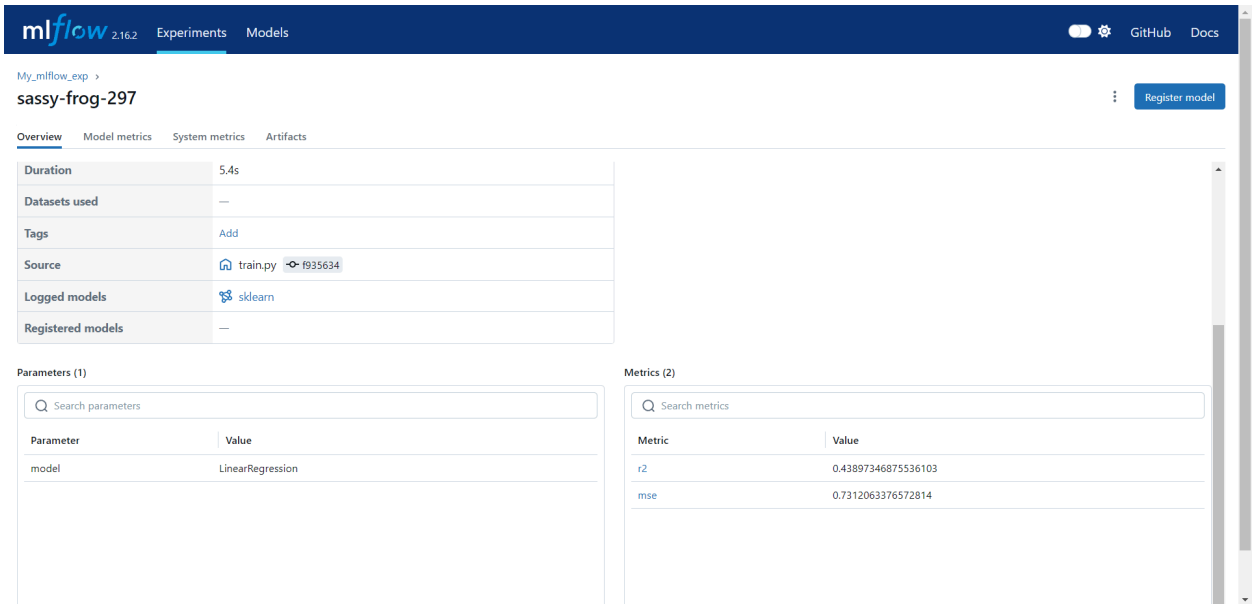
6. Ran Experiments and Tracked them with MLflow. Launched the MLflow UI in local environment using the command:

```
mlflow ui
```

7. Ran the training script with the command:

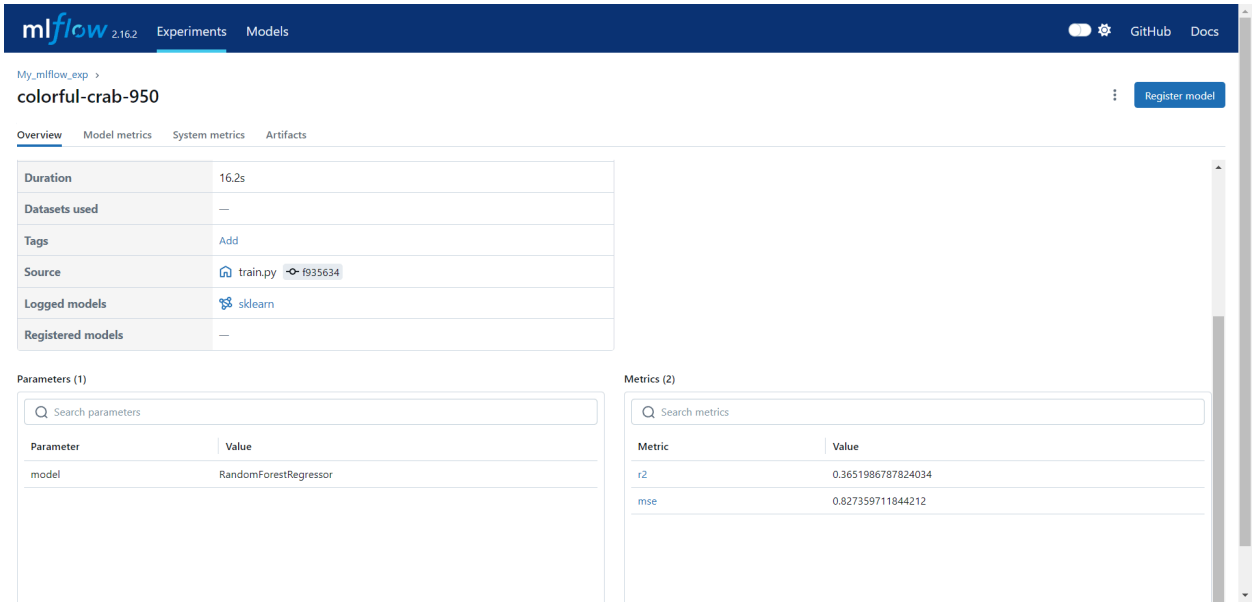
```
python train.py
```

And went to http://127.0.0.1:5000 to see the logs of different experiments in the MLflow UI



Linear Regression

Metric	**Value**
R2	0.43897346875536103
Mse	0.7312063376572814



RandomForestRegressor

Metric	Value
R2	0.3651986787824034
Mse	0.827359711844212

8. Compared the models by using the MLflow UI to compare the mse (Mean Squared Error) of the Linear_Regression and Random_Forest models.