ML Ops Assignment-1 Report

# REPO LINK : <https://github.com/SobanSageer/mlops-assignment-1>

# 1. Problem Statement

The purpose of this ML Ops assignment is to demonstrate the end-to-end machine learning workflow including dataset selection, model training, experiment tracking with MLflow, and model registration.  
  
- Dataset: Iris dataset (classification problem).  
- Task: Predict the species of iris flowers (setosa, versicolor, virginica) based on their physical measurements.  
- Importance: This showcases how ML Ops practices help maintain reproducibility, scalability, and monitoring in ML projects.

# 2. Dataset Description

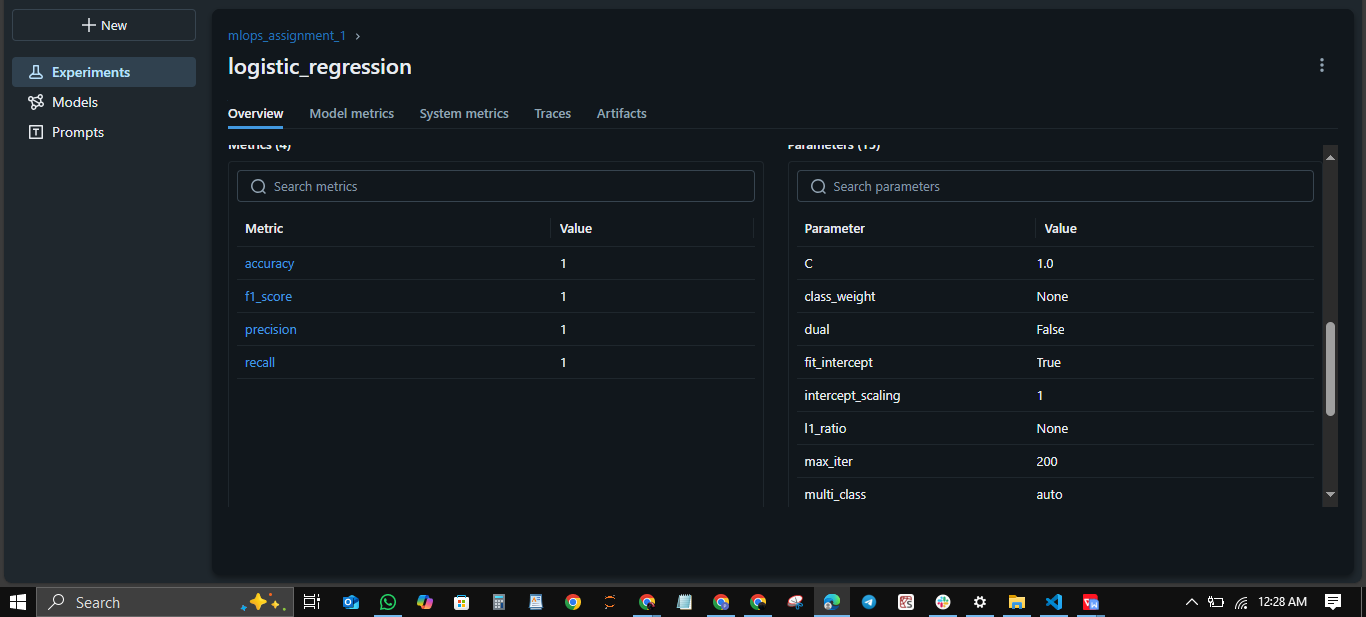
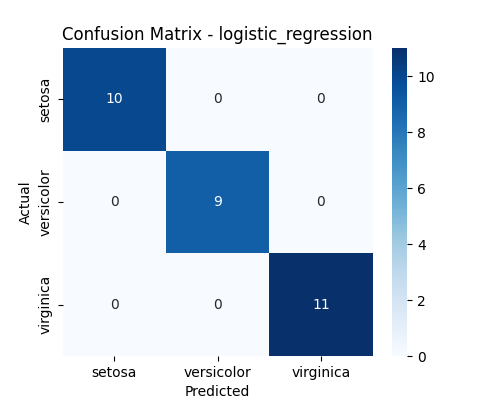
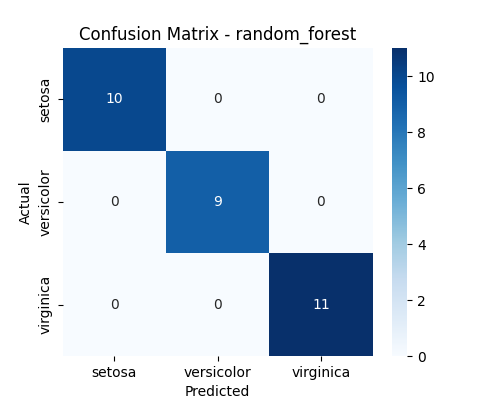
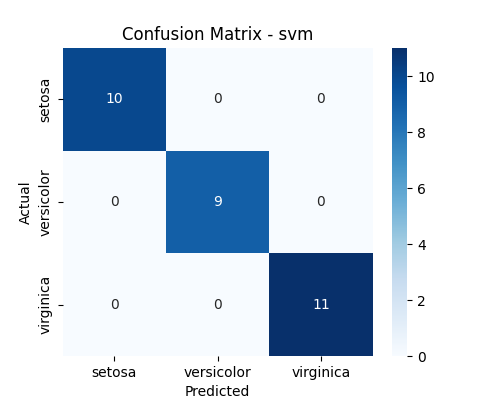
- Source: sklearn.datasets.load\_iris  
- Number of records: 150 samples  
- Features: 4 numerical features (sepal length, sepal width, petal length, petal width)  
- Target: Species of Iris (3 classes)

# 3. Model Selection & Comparison

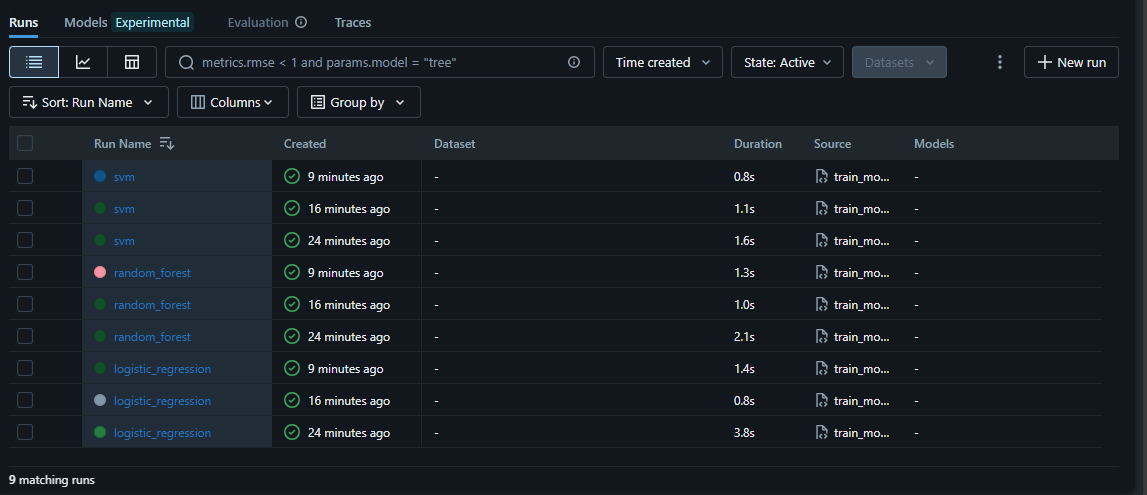
We tested the following models:  
1. Logistic Regression – baseline linear model.  
2. Random Forest Classifier – ensemble learning approach.  
3. Support Vector Machine (SVM) – linear kernel.

Performance Metrics:

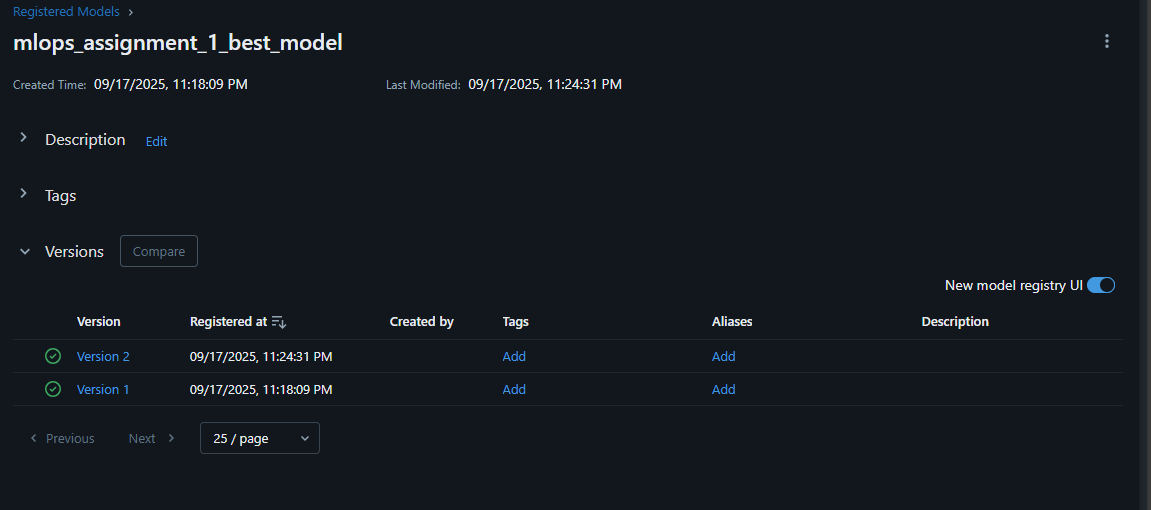
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-Score |
| Logistic Regression | 1.00 | 1.00 | 1.00 | 1.00 |
| Random Forest | 1.00 | 1.00 | 1.00 | 1.00 |
| SVM | 1.00 | 1.00 | 1.00 | 1.00 |

  
  
Final Model Selection: All models achieved perfect performance on this dataset. Logistic Regression was chosen as the final registered model for simplicity.

# 4. MLflow Logging

Each model was logged using MLflow with:  
- Parameters (e.g., n\_estimators, kernel, C)  
- Metrics (accuracy, precision, recall, F1)  
- Artifacts (saved model files, confusion matrices)  
  


# 5. Model Registration

The best-performing model (Logistic Regression) was registered in the MLflow Model Registry.  
  
- Model Name: mlops\_assignment\_best\_model  
- Version: 1 and 2  
- Stage: Production  
  


# 6. Instructions to Run the Code

1. Clone the GitHub repository:  
 git clone <https://github.com/SobanSageer/mlops-assignment-1>.git

cd mlops-assignment-1  
2. Create a virtual environment and install dependencies:  
 python -m venv venv  
 source venv/bin/activate # On Windows: venv\Scripts\activate  
 pip install -r requirements.txt  
  
3. Run the training script:  
 python train.py  
  
4. Start the MLflow UI to view experiment runs:  
 mlflow ui  
 Open in browser: http://127.0.0.1:5000  
  
5. Register the best model from the MLflow UI.

# 7. GitHub Submission

- All code, logs, and documentation have been pushed to GitHub.  
- The README.md contains setup instructions and references to MLflow screenshots.