Credit Default Predictor - Local MLOps Project

Welcome to the **Credit Default MLOps Lab**!
This project demonstrates a full local MLOps pipeline using MLflow and Streamlit.

Project Objectives

- Install MLflow and set up a local tracking server
- Train two ML models (April and May datasets)
- Track model experiments with MLflow
- Serve models with MLflow Models Server
- Test predictions with a Streamlit UI

Project Structure

```
api/
  data/
    raw/
      april_credit_data.csv
      may_credit_data.csv
training/
  train.py
ui/
  streamlit_app.py
Makefile
requirements.txt
README.md
```

Quick Setup

1. Clone the repository:

```
git clone https://github.com/iportilla/mlops-credit-default.git
cd mlops-credit-default
```

2. Install Python 3.11 (recommend using pyenv):

```
pyenv install 3.11.4
pyenv global 3.11.4
```

3. Install required Python packages:

```
pip install -r requirements.txt
pip install virtualenv
```

✅ `virtualenv` is required for MLflow model serving.

4. Create local mlruns folder manually:

```
mkdir -p ~/mlruns
```

5. Start MLflow Tracking Server:

```
mlflow server   --backend-store-uri sqlite:///mlflow.db   --default-
artifact-root ~/mlruns   --host 0.0.0.0   --port 5000
```

✅ This initializes `mlflow.db` and creates the required tables (experiments, runs, etc.).

6. Verify Setup:

- Confirm `~/mlruns` exists:

```
ls ~/mlruns
```

- Confirm `mlflow.db` exists:

```
ls mlflow.db
```

✅ If both exist, you're ready to train and serve models.

Commands Cheat Sheet

Task	Command
Start MLflow Tracking Server	`mlflow server --backend-store-uri sqlite:///mlflow.db --default-artifact-root ~/mlruns --host 0.0.0.0 --port 5000`
Train April Model	`python training/train.py api/data/raw/april_credit_data.csv`

Task	Command
Train May Model	`python training/train.py api/data/raw/may_credit_data.csv`
Serve Model Manually	`mlflow models serve -m /home/ubuntu/mlruns/1/<run_id>/artifacts/credit_defaults_model_ -p 5001`
Launch Streamlit App	`streamlit run ui/streamlit_app.py`

✅ When serving, make sure you point to the `credit_defaults_model_` folder which contains:

- `MLmodel`
- `model.pkl`
- `requirements.txt`
- `conda.yaml`

Example path after training:

```
/home/ubuntu/mlruns/1/<run_id>/artifacts/credit_defaults_model_
```

Architecture Overview

```
Local Machine
   ↓
MLflow Tracking Server (localhost:5000)
   ↓
MLflow Model Server (localhost:5001)
   ↓
Streamlit Frontend (localhost:8501)
```

Learning Outcomes

- Model training and logging
- Model versioning and experiment tracking
- Serving models automatically
- UI connection for predictions

License

This project is licensed under the MIT License.

✨ Happy Predicting with MLOps!