Credit Card Fraud Detection MLOps Project

Overview

This project focuses on building an end-to-end MLOps pipeline for credit card fraud detection. It involves data preprocessing, model selection, hyperparameter tuning, and model deployment using tools like scikit-learn and MLflow.

Repository

You can find the project repository here.

Project Structure

```
credit_fraud_detection/
■
■■■ data/
■   ■■■ default_of_credit_card_clients.csv
■
■■■ jupyter_notebooks_experiments/
■   ■■■ data_processing.ipynb
■
■■■ .gitignore
■■■ README.md
■■■ requirements.txt
```

Data

The dataset used in this project is the Credit Card Fraud Detection dataset. It contains information about credit card clients, including their payment history, bill statements, and whether they defaulted on their payments.

Setup

Clone the repository: `bash git clone https://github.com/vijaytakbhate2002/credit_fraud_detection_project_with_mlflow_dagshub_dvc.git cd credit_fraud_detection`

Install dependencies: `bash pip install -r requirements.txt`

Run the Jupyter notebook: `bash jupyter notebook jupyter_notebooks_experiments/data_processing.ipynb`

Model Training

The project uses two models for fraud detection: - **Logistic Regression**: Optimized using GridSearchCV for hyperparameter tuning. - **Random Forest**: Also optimized using GridSearchCV.

The best parameters for each model are logged using MLflow for reproducibility.

Evaluation

The models are evaluated using the following metrics: - Accuracy - Precision - Recall - F1 Score

MLflow Integration

MLflow is used to track experiments, log parameters, and metrics. You can view the experiments at:

```
http://localhost:5000/#/experiments/<experiment_id>
```

Future Work

- Model deployment and monitoring
- Integration with CI/CD pipelines
- Real-time fraud detection

Contributing

Contributions are welcome! Please open an issue or submit a pull request.

License