**Hiscox Claims Prediction**

**Marcus Trinder**

**Overview**

This repository contains a machine learning solution designed to predict the likelihood of claims for insurance applications. The project includes functionalities for data collection, preprocessing, model training, evaluation, and deployment using AWS SageMaker. A CI/CD pipeline is also set up using GitHub Actions to ensure seamless integration and deployment.

Detailed user guide and implementation steps provided in ‘readme.md’.

**Key Features**

1. **Data Processing**:
   * Collects and preprocesses data from a simulated SQL database.
   * Handles missing values and feature transformations.
2. **Model Training and Evaluation**:
   * Trains an XGBoost classifier to predict claim likelihood.
   * Evaluates the model using various metrics such as accuracy, F1 score, ROC AUC, and more.
3. **Model Deployment**:
   * Deploys the trained model to AWS SageMaker for scalable and reliable inference.
   * Includes deployment scripts to automate the model training and deployment process.
4. **CI/CD Pipeline**:
   * GitHub Actions workflow for continuous integration and continuous deployment.
   * Automated testing and deployment on every push to the main or version branches.

**Project Structure**

hiscox-claims-prediction/

├── .github/

│ └── workflows/

│ └── ci-cd.yaml

├── data/

├── models/

├── notebooks/

├── src/

│ ├── \_\_init\_\_.py

│ ├── data\_processing.py

│ ├── model\_training.py

│ ├── model\_evaluation.py

│ ├── predict.py

├── tests/

├── scripts/

│ ├── train\_model.py

│ ├── evaluate\_model.py

│ ├── predict.py

├── deployment/

│ ├── train\_model.py

│ ├── deploy\_model.py

│ ├── sg\_create\_group.py

│ ├── sg\_register\_model.py

├── requirements.txt

├── setup.py

├── README.md

**Setup Instructions**

1. **AWS Configuration**:
   * Configure AWS CLI and set up an IAM role for SageMaker.
   * Upload training data to S3.
2. **Installation**:
   * Clone the repository and install the required Python packages.
3. **Usage**:
   * Use provided scripts for data processing, model training, evaluation, and predictions.
4. **CI/CD Pipeline**:
   * GitHub Actions workflow for automated testing and deployment.
5. **Model Versioning**:
   * Manage model versions using SageMaker Model Registry.

**Remarks**

I have thoroughly enjoyed working on this task and look forward to feedback and future communication.