

# CONTINOUS INTEGRATION FOR ML MODEL

---

Charan Sai	AP21110011621
Manjari S	AP21110011620
Sai Charan Sake	AP21110011619
Ramanandhareddy	AP21110011622

# INTRODUCTION

---

Continuous Integration and Continuous Deployment (CI/CD) pipelines play a vital role in optimizing the development, testing, and deployment processes of machine learning (ML) models. In the landscape of ML, where models undergo frequent updates and enhancements, a robust CI/CD pipeline is essential for preserving model performance, ensuring reproducibility, and fostering collaboration among data scientists and developers.

This pipeline typically consists of multiple stages, starting with version control to track and manage changes to the codebase, including model architecture, data preprocessing, and training scripts. Automated testing follows, encompassing unit tests, integration tests, and validation checks to verify the model's functionality and adherence to predefined criteria, such as data consistency and model performance.

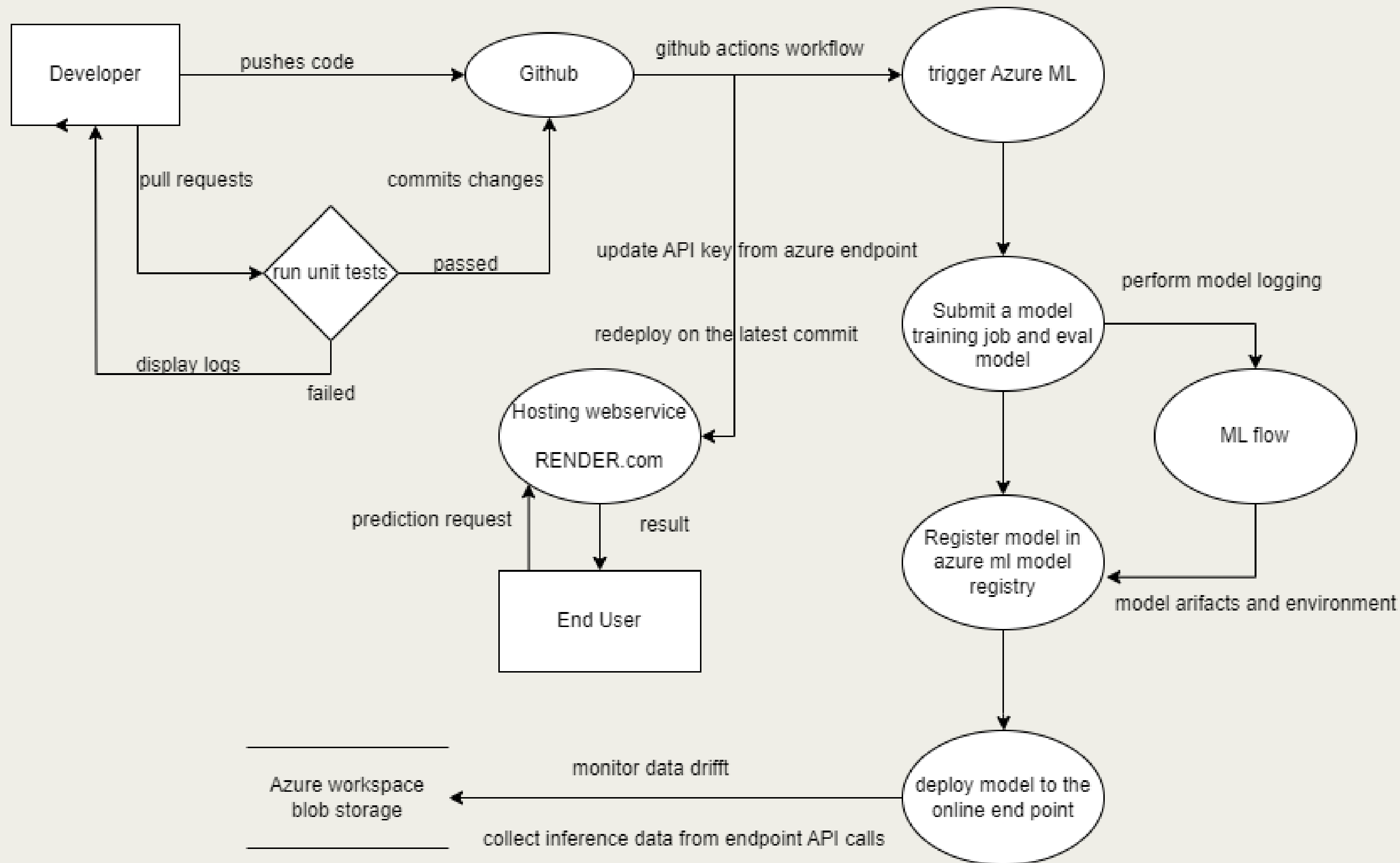
# PROJECT OVERVIEW

---

CI/CD pipeline for ML models involves continuous integration and deployment. Continuous integration automatically builds and tests the model every time the codebase or data sources are changed. Continuous deployment deploys the validated and tested model to the production environment automatically, ensuring rapid iteration and minimizing downtime.

The proposed system aims to implement a Continuous Integration and Continuous Deployment (CI/CD) pipeline for predicting diabetes using Logistic Regression Machine Learning (ML) model using GitHub, GitHub Actions, Azure ML, and MLflow for model logging in Azure, and for the UI streamlit is used.

# PROJECT WORKFLOW



# ROLES

---

**Developer:** A data scientist initially develops a model by analyzing and engineering features from available data, then passes it to an ML engineer. The ML engineer converts the model into a Python script, creates a GitHub Actions workflow for CI/CD, and pushes necessary files to the GitHub repository.

**GitHub:** GitHub facilitates CI/CD pipelines for ML models through automated workflows triggered by code changes. It organizes workflows for documentation, model training, validation, and deployment under `.github/workflows`, ensuring reproducibility and scalability.

**Azure ML:** Azure ML supports scalable model training, registration, and deployment. Models are registered in the Azure ML Model Registry, packaged into containers, and deployed to the appropriate environment. MLflow integration enables seamless transition from training to deployment.

# ROLES

---

**MLflow:** MLflow tracks experiments, logs parameters, metrics, and artifacts, facilitating model comparison and deployment. It logs models in a standardized format with dependencies, ensuring reproducibility and compatibility across environments.

**Render:** Render.com hosts the ML model's front end and manages API endpoints. It provides a platform for user-friendly interaction with the model, facilitates endpoint updates upon model changes, and securely manages environment secrets like API keys.

# GITHUB ACTIONS WORKFLOW

lolugucharansai / MLops-exercise

Type / to search

>

+

+

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

← Train Model

updated the consume -endpoint #71

Re-run all jobs

...

Summary

Jobs

development

production

Run details

Usage

Workflow file

Triggered via push last week

lolugucharansai pushed · 6a1afb1 · main

Status

Success

Total duration

17m 42s

Artifacts

1

manual-trigger-actions.yml

on: push

development9m 4s

production8m 17s

Annotations

5 warnings

development

node.js 16 actions are deprecated. Please update the following actions to use node.js 18: lukechilds/deploy-action-repository



# AZURE ML MODEL REGISTRY

Azure AI | Machine Learning Studio

Notebooks

Automated ML

Designer

Prompt flow

Assets

Data

Jobs

Components

Pipelines

Environments

**Models**

Endpoints

Manage

Compute

Monitoring

Data Labeling

Default Directory > aml-diabetes-dev > Models

Model List

+ Register

Refresh

Delete

Archive

Deploy

...

Show latest versions only

Include archived

Search

Filter

Columns

Name	☆	Versi... ↓	Type	Source	Experiment
diabetes_model		30	MLFLOW	This workspace	diabetes-data-example
diabetes_model		29	MLFLOW	This workspace	diabetes-data-example
diabetes_model		28	MLFLOW	This workspace	diabetes-data-example
diabetes_model		27	MLFLOW	This workspace	diabetes-data-example
diabetes_model		26	MLFLOW	This workspace	diabetes-data-example
diabetes_model		25	MLFLOW	This workspace	diabetes-data-example
diabetes_model		24	MLFLOW	This workspace	diabetes-data-example

25 / Page 16



# MONITORING THE DEPLOYED MODEL

Azure AI | Machine Learning Studio

Notebooks

Automated ML

Designer

Prompt flow

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Monitoring

Data Labeling

Linked Services

Default Directory > aml-diabetes-dev > Monitoring > diabetesdeployment1mon

diabetesdeployment1mon

Overview

Notification history

Job history

Refresh

Add signal

Previous run

Latest run

Trigger manually

Details

Model

diabetes\_model

Endpoint

diabetesend

Deployment

diabetesdeployment1

Model task type

Classification

Created on

May 5, 2024 4:30 PM

Created by

charan sai

Schedule

Next run

May 6, 2024 6:31 PM

# RENDER WEB SERVICE DEPLOYMENT ON LATEST COMMIT

WEB SERVICE

Mlops-exercise

Python 3

Free

Upgrade your instance →

Connect

Manual Deploy

lolugucharansai / Mlops-exercise

main

https://mlops-exercise.onrender.com

Events

Logs

Disks

Environment

Shell

Previews

Jobs

Metrics

Scaling

Your free instance will spin down with inactivity, which can delay requests by 50 seconds or more. [Upgrade now.](#)

Deploy live for 17c588e: Update on\_push\_workflow.yml

May 5, 2024 at 4:07 PM

Deploy started for 17c588e: Update on\_push\_workflow.yml

New commit via Auto-Deploy

May 5, 2024 at 4:05 PM

Deploy live for c056c69: updated the file

April 30, 2024 at 8:24 PM

Rollback



# CONCLUSION

---

In conclusion, the integration of Machine Learning (ML) models into Continuous Integration and Continuous Deployment (CI/CD) pipelines represents a significant advancement in the field of software engineering and AI. By automating the testing, validation, and deployment of ML models, organizations can accelerate the delivery of AI-powered solutions, improve collaboration between data scientists and developers, and ensure the reliability and scalability of their ML applications.

Throughout this project, we have explored the concepts, challenges, and best practices associated with implementing CI/CD pipelines for ML models. We have discussed the advantages of this approach, including faster iteration cycles, enhanced reproducibility, and reduced deployment errors, as well as the limitations and challenges that need to be addressed.

# Thank you!

---