

# Heart Disease MLOps Project - Documentation Index

---

## Assignment Submission Structure

This documentation is organized according to the assignment requirements. Navigate through each section below:

---

### **1** Setup/Install Instructions

**Location:** [01\\_Setup\\_Installation/](#)

Essential guides for setting up the project environment and running the application:

- [START\\_HERE.md](#) - Quick start guide to get the project running
  - [LOCAL\\_DEPLOYMENT\\_GUIDE.md](#) - Complete local setup with virtual environment, dependencies, and model training
  - [EXECUTION\\_GUIDE.md](#) - Step-by-step execution instructions for the entire pipeline
  - [CLOUD\\_DEPLOYMENT\\_GUIDE.md](#) - Azure Container Apps deployment setup and configuration
- 

### **2** EDA and Modelling Choices

**Location:** [02\\_EDA\\_Modelling/](#)

Exploratory data analysis and model development documentation:

- [01\\_EDA.ipynb](#) - Jupyter notebook with complete exploratory data analysis, visualizations, and statistical summaries
- **Model Selection:** Logistic Regression, Random Forest, Gradient Boosting
- **Best Model:** Random Forest (88.52% accuracy, 0.95 ROC-AUC)
- **Feature Engineering:** StandardScaler preprocessing pipeline
- **Performance Metrics:** See FINAL\_SUBMISSION\_REPORT.md Section 4

#### Key Insights:

- 303 patient records, 13 features, 0 missing values
  - Balanced dataset (45.5% no disease, 54.5% disease)
  - Top predictors: chest pain type, ST depression, number of vessels
  - 5-fold cross-validation for robust evaluation
- 

### **3** Experiment Tracking Summary

**Location:** [03\\_Experiment\\_Tracking/](#)

MLflow experiment tracking and model versioning:

- [MODEL\\_STORAGE\\_INFO.md](#) - Model storage locations, versioning strategy, and cloud upload guide

## Experiment Tracking Details:

- **MLflow Version:** 2.5.0
- **Experiments Logged:** 3 models with all hyperparameters
- **Metrics Tracked:** Accuracy, Precision, Recall, F1-Score, ROC-AUC
- **Artifacts Stored:** Trained models (.pkl), confusion matrices, ROC curves, feature importance plots
- **UI Access:** <http://localhost:5000>
- **Model Registry:** Complete version control and lineage tracking

## Model Storage:

- Local: `models/best_model.pkl`, `models/preprocessor.pkl`
- MLflow: `mlruns/` directory with complete experiment history
- Production: Version-tagged models ready for deployment

---

## 4 Architecture Diagram

**Location:** [04\\_Architecture/](#)

System architecture and component design:

- **ARCHITECTURE.md** - Comprehensive system architecture documentation with component descriptions

## Architecture Overview:

- **API Layer:** FastAPI with 4 endpoints (/health, /predict, /predict/batch, /metrics)
- **Model Layer:** Random Forest classifier with preprocessing pipeline
- **Data Layer:** UCI Heart Disease dataset with feature engineering
- **Deployment:** Docker containers orchestrated by Kubernetes
- **Monitoring:** Prometheus + Grafana with 15+ metrics
- **CI/CD:** GitHub Actions with automated testing and deployment
- **Cloud:** Azure Container Apps with auto-scaling

## Key Components:

1. FastAPI REST API (Port 8000)
2. Prometheus Metrics Server (Port 9090)
3. Grafana Dashboard (Port 3000)
4. MLflow Tracking Server (Port 5000)
5. Kubernetes Cluster (3 replicas, LoadBalancer)

---

## 5 CI/CD and Deployment Workflow Screenshots

**Location:** [05\\_CI\\_CD\\_Deployment/](#)

Visual documentation and deployment guides:

- **GRAFANA\_SETUP\_GUIDE.md** - Grafana dashboard setup with monitoring panels

- [GRAFANA\\_GUIDE.md](#) - Dashboard usage and metric interpretation
- [DASHBOARD\\_SETUP\\_COMPLETE.md](#) - Monitoring stack configuration

### Screenshots Available in: [.. /screenshots/](#)

- GitHub Actions workflow runs
- MLflow experiment tracking UI
- Grafana monitoring dashboards
- API Swagger documentation
- Kubernetes deployment status
- Docker containers running
- Test coverage reports

### Video Demonstration: [.. /recorded\\_video\\_project\\_pipeline/](#)

### CI/CD Pipeline Stages:

1. **Lint & Test:** flake8, black, pylint + pytest (85%+ coverage)
  2. **Build:** Docker image creation with vulnerability scanning
  3. **Integration Tests:** API endpoint testing with health verification
  4. **Deploy:** Conditional deployment to staging/production
- 

## 6 Link to Code Repository

### **Location:** [06\\_Repository\\_Links/](#)

Access instructions and repository navigation:

- [ACCESS\\_INSTRUCTIONS.md](#) - How to access all deployed services and UIs
- [SAMPLE\\_DATA\\_GUIDE.md](#) - Test data usage and API testing guide
- [DATASET\\_LOCATION\\_UPDATE.md](#) - Dataset location and download information

### GitHub Repository: [MLOpsExperimentalLearning\\_Assignment\\_1\\_Group\\_81](#)

### Key Repository Links:

#### **Source Code:**

- [src/train.py](#) - Model training pipeline
- [src/app.py](#) - FastAPI application
- [src/preprocessing.py](#) - Data preprocessing

#### **CI/CD:**

- [.github/workflows/ci-cd.yml](#) - GitHub Actions pipeline

#### **Infrastructure:**

- [Dockerfile](#) - Container configuration
- [deployment/kubernetes/](#) - K8s manifests
- [deployment/prometheus/](#) - Prometheus config

- [deployment/grafana/](#) - Grafana dashboards

## Tests:

- [tests/](#) - Test suite (30+ tests, 85%+ coverage)

## Models & Data:

- [models/](#) - Trained models (best\_model.pkl, preprocessor.pkl)
- [data/processed/](#) - Processed dataset

## Documentation:

- [README.md](#) - Main project README
- 

## Main Submission Report

[FINAL\\_SUBMISSION\\_REPORT.md](#) - Comprehensive 10-page final report covering:

- Executive Summary
  - Complete MLOps methodology
  - Dataset analysis and EDA insights
  - Model development and evaluation (88.52% accuracy)
  - Experiment tracking with MLflow
  - API development and containerization
  - CI/CD pipeline implementation
  - Production deployment (Kubernetes + Azure)
  - Monitoring and observability (Prometheus + Grafana)
  - Performance analysis and results
  - Challenges, solutions, and lessons learned
  - Future enhancements
  - Complete reference links
- 

## Quick Navigation

Requirement	Primary Document	Supporting Files
<b>Setup Instructions</b>	LOCAL_DEPLOYMENT_GUIDE.md	START_HERE.md, EXECUTION_GUIDE.md
<b>EDA &amp; Modelling</b>	01_EDA.ipynb	FINAL_SUBMISSION_REPORT.md (Sec 2-4)
<b>Experiment Tracking</b>	MODEL_STORAGE_INFO.md	EXECUTION_GUIDE.md (MLflow section)
<b>Architecture</b>	ARCHITECTURE.md	FINAL_SUBMISSION_REPORT.md (Sec 8-9)
<b>CI/CD Screenshots</b>	screenshots/ folder	GRAFANA_SETUP_GUIDE.md, DASHBOARD_SETUP_COMPLETE.md

Requirement	Primary Document	Supporting Files
Repository Links	FINAL_SUBMISSION_REPORT.md	ACCESS_INSTRUCTIONS.md

## Project Statistics

- **Models Trained:** 3 (Logistic Regression, Random Forest, Gradient Boosting)
- **Best Accuracy:** 88.52% (Random Forest)
- **Test Coverage:** 85%+ (30+ tests)
- **API Endpoints:** 4 (/health, /predict, /predict/batch, /metrics)
- **Monitoring Metrics:** 15+ Prometheus metrics
- **Deployment:** 3 Kubernetes replicas with auto-scaling
- **API Performance:** <50ms average response time, 250 req/sec max throughput
- **Uptime:** 99.9% reliability

## Getting Started

1. Start with [START\\_HERE.md](#) for quick setup
2. Follow [LOCAL\\_DEPLOYMENT\\_GUIDE.md](#) for complete setup
3. Review [01\\_EDA.ipynb](#) for data insights
4. Check [ARCHITECTURE.md](#) for system design
5. Read [FINAL\\_SUBMISSION\\_REPORT.md](#) for comprehensive overview

## Team Members

Name	Student ID
GOBIND SAH	2024AA05643
VISHAL SINGH	2024AA05641
YASH VERMA	2024AA05640
AVISHI GUPTA	2024AA05055
ASIT SHUKLA	2023AC05956

**Date:** January 6, 2026

**Course:** MLOps - Experimental Learning (S1-25\_AIMLCZG523)

**Status:**  Final Submission Ready

**END OF INDEX**