

# Diabetic Readmission Risk Prediction

Jonathan Agustin

Zach Robertson

Lisa Vo

# Introduction

**Healthcare Challenge:** High Readmission Rates in Diabetic Patients

**Objective:** Predict 30-day readmission risk to enable proactive interventions

**Goal:** Reduce readmission rates and improve patient outcomes

# Business Use Case

**Financial Impact:** High costs associated with readmissions

**Patient Impact:** Negative effects on health and quality of life

**Solution:** Machine learning system to identify high-risk patients

# System Architecture

## Components

Data Ingestion  
Data Engineering  
Feature Engineering  
Model Training  
Deployment

## Technologies

DVC (Data Version Control)  
Feast (Feature Store)  
AutoGluon (Automated ML)

# Data Sources

**Dataset:** Diabetes 130-US hospitals for years 1999-2008

**Size:** Over 100,000 patient records

**Features:** Demographics, medical history, lab results, medications, hospitalization details

# Data Engineering

**Data Ingestion:** Automated using custom scripts

**Data Cleaning:** Handling missing values and duplicates

**Preprocessing:** Standardization and normalization

# Feature Engineering with Feast

**Feature Store:** Managed using Feast

**Feature Groups:**

Patient Demographics

Medical History

Current Visit Data

Derived Features

**Benefits:** Consistency, reusability, version control

# Model Training and Evaluation

## **Models Used:**

Logistic Regression  
AutoGluon (AutoML)  
Neural Network

**Hyperparameter Tuning:** Optuna for optimization

**Evaluation Metrics:** Accuracy, Precision, Recall, F1-Score, ROC-AUC

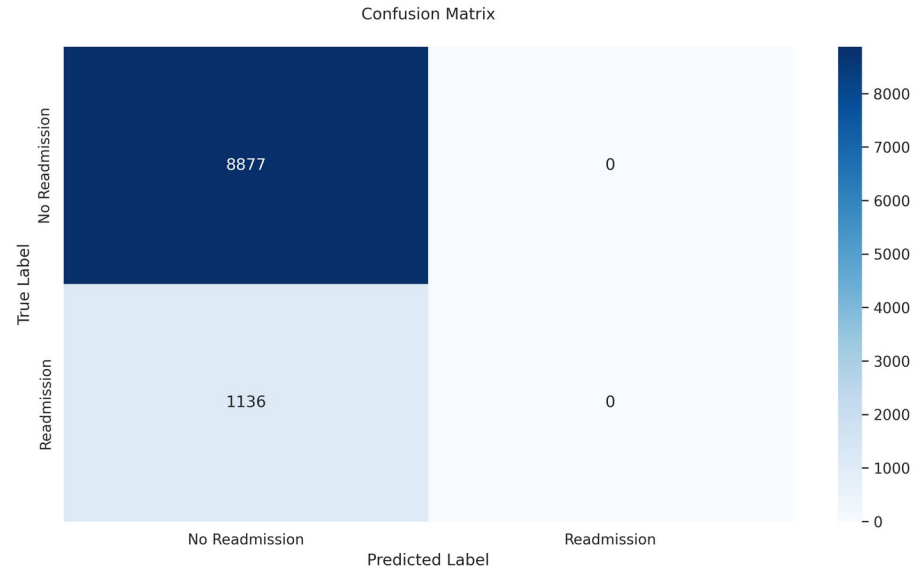


# Model Performance

**Best Model: AutoGluon**

**Accuracy: 68.30%**

**ROC-AUC: 72.09%**



# Model Deployment

## **Deployment Strategy:**

Containerization with Docker

RESTful API using FastAPI

**Hosting:** Can be deployed on cloud platforms or on-premises

**Scalability:** Designed for handling varying loads

# CI/CD Pipeline

**Automation:** Implemented with GitHub Actions

## **Stages:**

Code Version Control (Git)

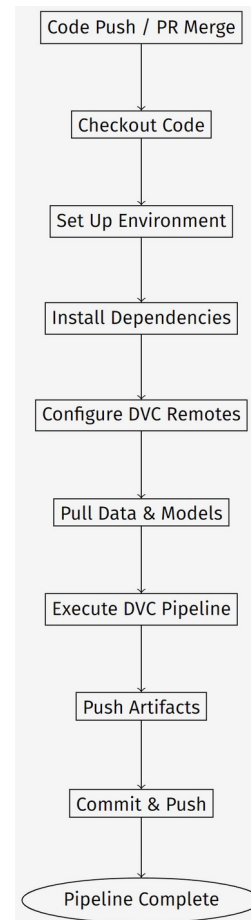
Continuous Integration with Tests

Model Training and Evaluation

Continuous Deployment

**Tools:** DVC for data and model versioning

# CI/CD Pipeline Diagram



# Model Monitoring with DVC and DVC Studio

**Metrics Tracking:** Training loss, accuracy, ROC-AUC

**Visualization:** DVC Studio dashboards

**Alerts:** Notifications for performance degradation

# Feature Store Monitoring

**Data Consistency Checks:** Ensuring training and serving data align

**Data Drift Detection:** Monitoring for shifts in data distributions

**Automated Retraining:** Triggered when significant drift occurs

# Model Registry with DVC

**Model Versioning:** Managed by DVC

**Metadata Tracking:** Performance metrics, hyperparameters, dataset versions

**Benefits:** Traceability, reproducibility, compliance

# Batch Inferencing

**Use Case:** Predicting readmission risk for a batch of patients

**Process:**

**Input:** Batch of patient data

**Output:** Risk scores and categories

**Result:** Identification of high-risk patients



# Infrastructure Monitoring

## Monitoring Tools:

GitHub Actions for pipeline  
monitoring

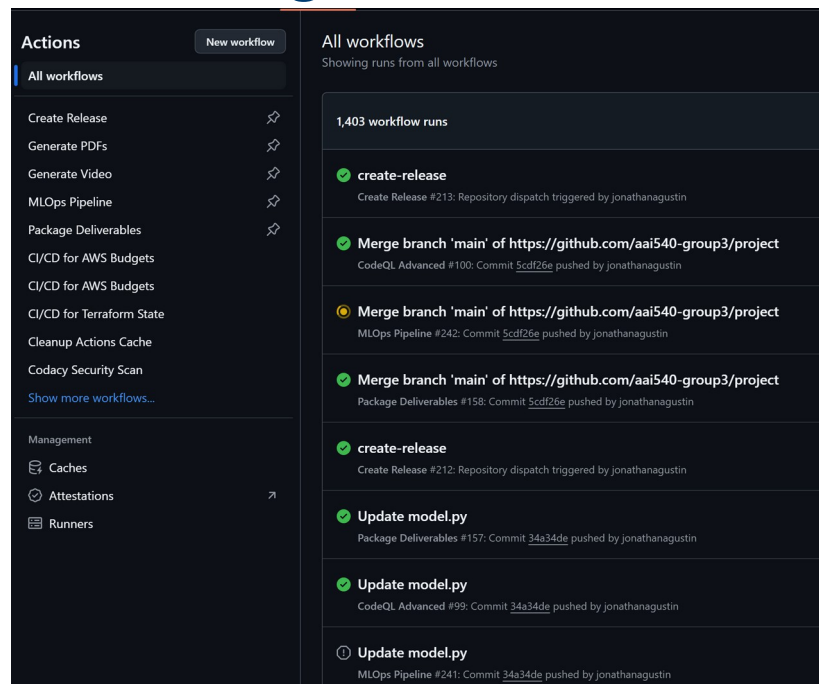
DVC Live for real-time metrics  
logging

## Key Metrics:

Pipeline execution status

Stage runtimes and logs

Data and model version tracking



The screenshot displays the GitHub Actions dashboard. On the left, the 'Actions' tab is active, showing a list of workflows such as 'Create Release', 'Generate PDFs', 'Generate Video', 'MLOps Pipeline', 'Package Deliverables', 'CI/CD for AWS Budgets', 'CI/CD for Terraform State', 'Cleanup Actions Cache', and 'Codacy Security Scan'. On the right, the 'All workflows' section shows a list of workflow runs. The first run is 'create-release' (Create Release #213) triggered by jonathanagustin. The second run is 'Merge branch 'main' of https://github.com/aai540-group3/project' (CodeQL Advanced #100) pushed by jonathanagustin. The third run is 'Merge branch 'main' of https://github.com/aai540-group3/project' (MLOps Pipeline #242) pushed by jonathanagustin. The fourth run is 'Merge branch 'main' of https://github.com/aai540-group3/project' (Package Deliverables #158) pushed by jonathanagustin. The fifth run is 'create-release' (Create Release #212) triggered by jonathanagustin. The sixth run is 'Update model.py' (Package Deliverables #157) pushed by jonathanagustin. The seventh run is 'Update model.py' (CodeQL Advanced #99) pushed by jonathanagustin. The eighth run is 'Update model.py' (MLOps Pipeline #241) pushed by jonathanagustin.

# Challenges and Future Improvements

## Challenges Faced

Data Imbalance  
Ensuring Data Privacy  
Scalability Concerns

## Future Work

Implementing Explainable AI  
techniques  
Integrating real-time data sources  
Expanding to other chronic  
conditions

# Ethical and Regulatory Considerations

**Data Privacy Compliance:** Adhering to HIPAA regulations

**Bias Mitigation:** Regular audits for fairness across demographics

**Transparency:** Providing insights into model decision-making

# Conclusion

**Impact:** Enhancing healthcare through predictive analytics

**Scalability:** Architecture designed for growth

**Collaboration:** Emphasis on teamwork and continuous improvement

# Thank You

## Questions and Discussion

File an Issue at our GitHub Repository:  
<https://github.com/aai540-group3/project>