

HEART DISEASE CLASSIFICATION

(Members)

- 1.Anawat Pradabsuk 124039 (Leader)
- 2.Ritik Sareen 123844
- 3.Tanzil Al Sabah 123845
- 4.Chatpol Phongyen 123992
- 5.Munthitra Thadthapong 124022
- 6.Md Shafi Ud Doula 124047



CONTENT

- Introduction
- Related Industry
- Problem
- Solution
 - Rationale
 - Mockups
 - Diagrams
- Evaluation



INTRODUCTION

Background

- Cardiovascular diseases cause of death worldwide 17.9 M.
- Diagnosis tools have limitation (costly)

Motivation

- The potential of machine learning for heart disease detection is immense.
- The goal: early detection, timely interventions, and a significant reduction in the global impact



RELATED INDUSTRY



Health Monitoring Wearables
can notify users of irregular
heart rhythms.

The image features the logo for 'Cardiologs® AI Serving Cardiology' at the top left. Below it are three screenshots of a software interface. The first screenshot shows a table for 'Uploads' with columns for File Name, Patient Name, Date of Birth, and Labels. A file named 'orgP_uploaded_8.ecg' is listed. The second screenshot shows an ECG waveform with various analysis tools and data points. The third screenshot shows a detailed report with multiple ECG strips and associated data. Large blue numbers 1, 2, and 3 are overlaid on each screenshot respectively, with the text 'Upload the signal', 'Analyze the ECG guided by AI', and 'Generate the report' positioned below them.

Offers an ECG analysis platform powered by AI to detect arrhythmias and other conditions.

RELATED INDUSTRY

Risk Factor Calculator: 30 Year Risk Factors

We acknowledge Mr. Aaron Vaneps and the Mayo Clinic Cardiovascular Health Clinic who provided the interactive risk calculator.

Sex: Male Female

Systolic BP: 140

Age: 29

Diabetes:

Smoker:

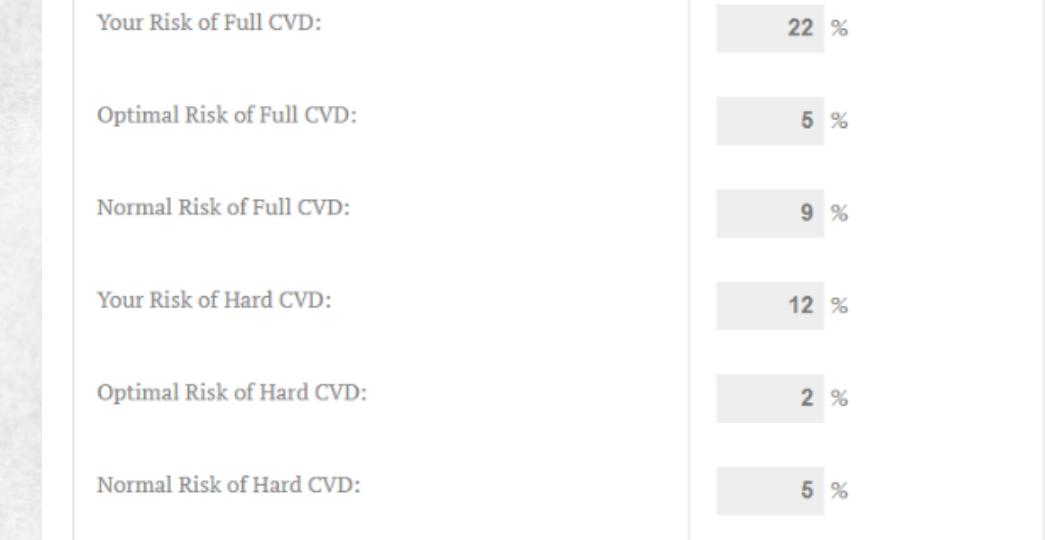
Treated Hypertension:

Total Cholesterol: 180

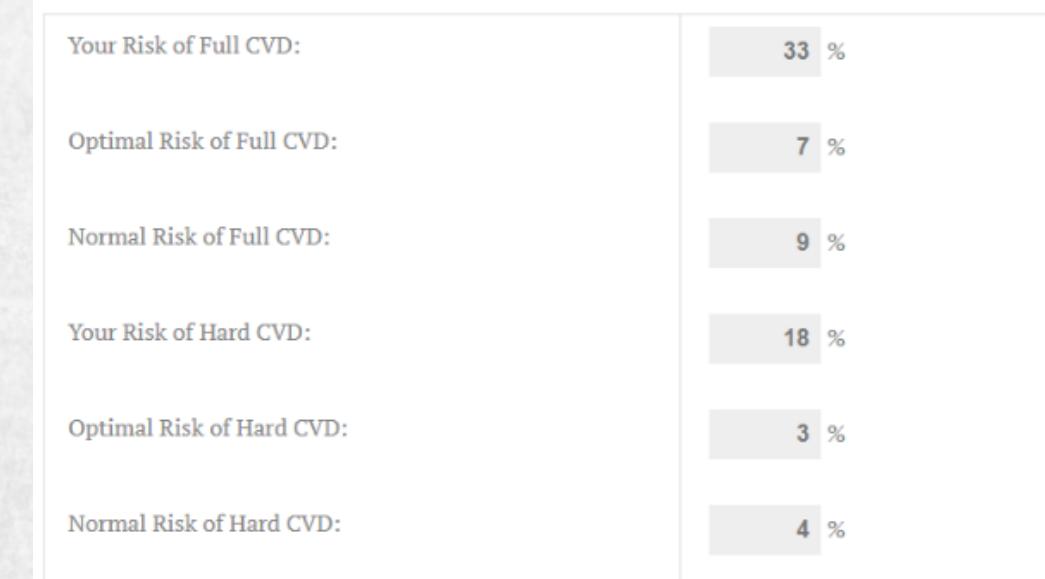
HDL Cholesterol: 66

BMI: 25

Lipids-Based Results



BMI-Based Results



Framingham Heart Study: A long-term, ongoing cardiovascular cohort study that has been a source for various ML projects due to its comprehensive datasets.

PROBLEM

- Heart rate and variability are not always linked directly to clinical outcomes.
- ECG Analysis by AI is not sufficient information.
- The Framingham Heart Study do not really reflect the global diversity of heart disease risk factors and manifestations.
- Some of the data might not adjust with modern lifestyle factors and risks associated with heart diseases todays.

RATIONALE

USE

UCI Heart
Disease Data
Set

CREATE

Modern
Supervised
machine
learning

PREDICT

Patient's risk of
heart disease

HELP

Detect and
diagnose heart
disease early



VALUES PROPOSITION



Early Detection



Accuracy Classification



User friendly interface



Data-Driven Insights



TARGET



Patients



**Cardiologists and Healthcare
Professionals**

Requirement Elicitation

- To classify the likelihood of heart disease based on a machine learning model.
- Add a user-friendly interface for simplicity of usage.
- Offer explanations for model predictions.
- Update the model frequently to increase precision.

Feasibility

- Developing machine learning within a two-month time frame.
- The dataset is well-documented and suitable for model training
- The availability of machine learning libraries and frameworks simplifies the development process.
- Maintaining data privacy and security may require additional time and resources.

RISKS



Data privacy



User Adopter



Model Accuracy



Ethical Considerations



Data Quality

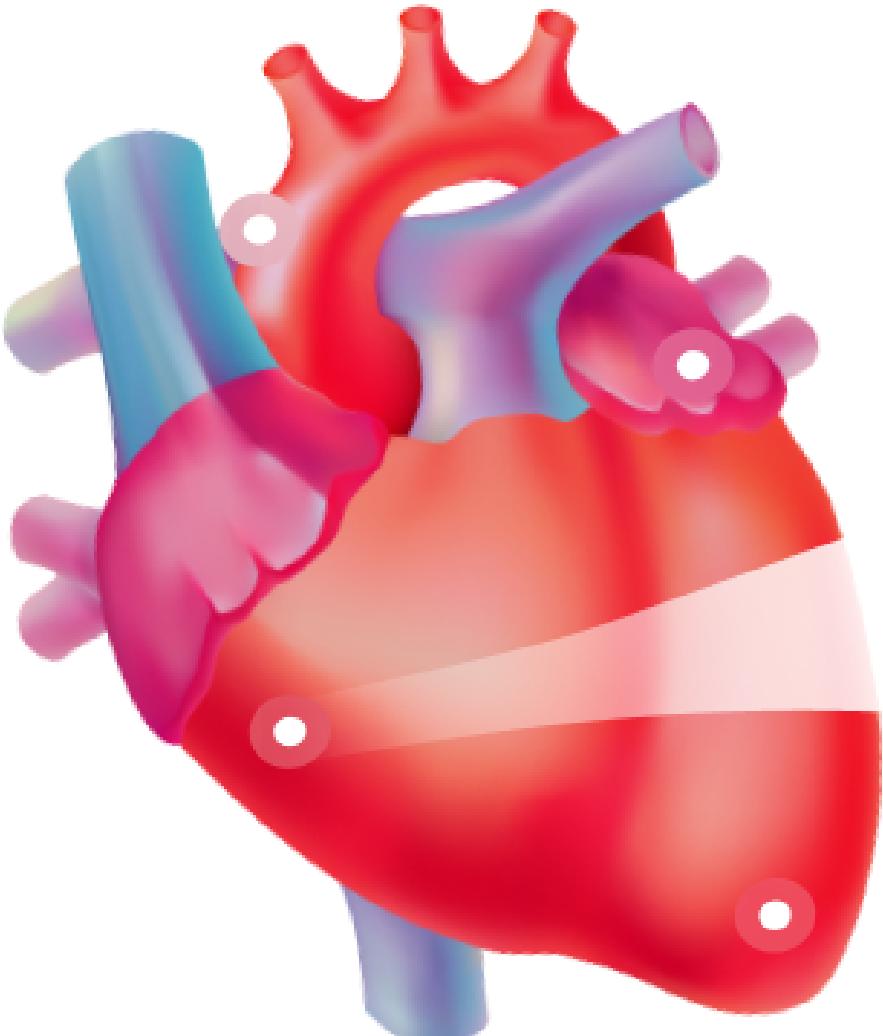
WEBSITE MOCKUPS



Desktop Application

Dashboard

Heart disease prediction



Motivation

Heart disease is a leading cause of death worldwide, making the accurate and timely diagnosis of utmost importance. To address this critical need, a predictive model has been developed using machine learning techniques. This model utilizes a comprehensive dataset of various patient attributes, including demographic information, medical history, and diagnostic test results.

Through its predictive capabilities, this model offers healthcare professionals a valuable tool for early detection and risk assessment of heart disease. Its accuracy and efficiency have the potential to assist medical practitioners in making informed decisions and providing timely interventions, ultimately contributing to improved patient outcomes and overall public health.

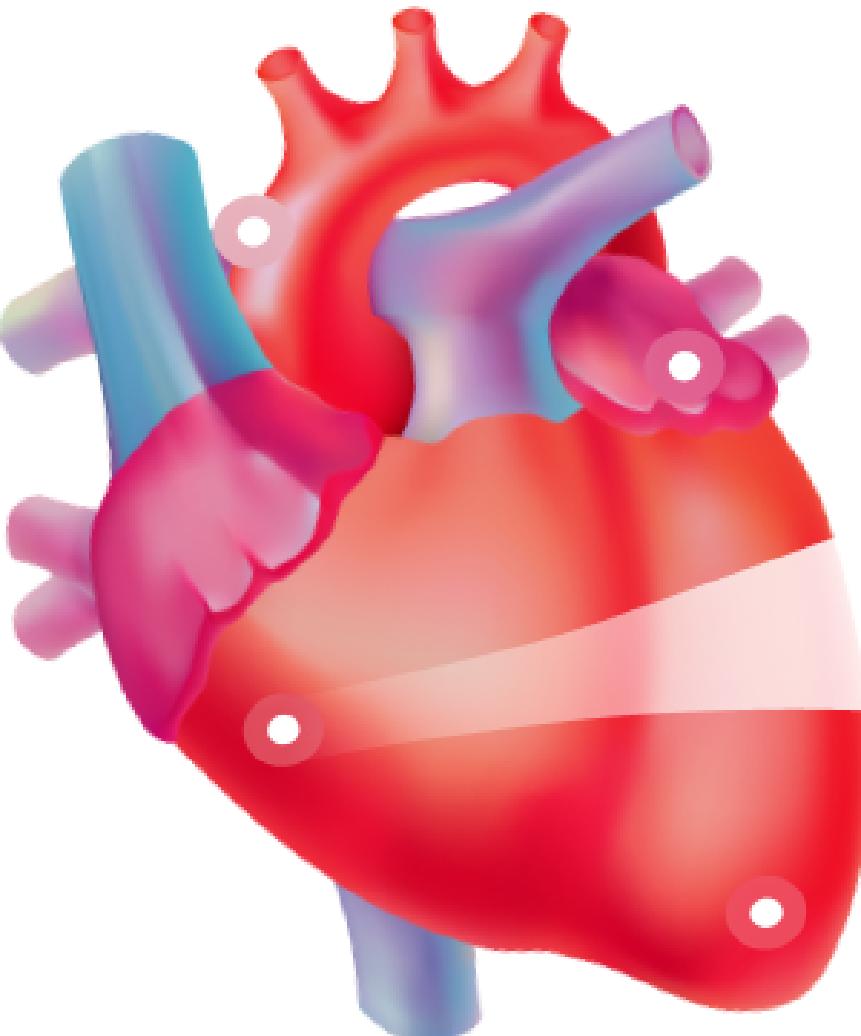
[Get Started](#)



Desktop Application

Dashboard

Patient Information



Age

Sex

Chest Pain Type

Resting Blood Pressure

(mm Hg)

Serum cholesterol

(mg/dl)

fasting blood sugar > 120 mg/dl

resting electrocardiographic

maximum heart rate achieved

(BPM)

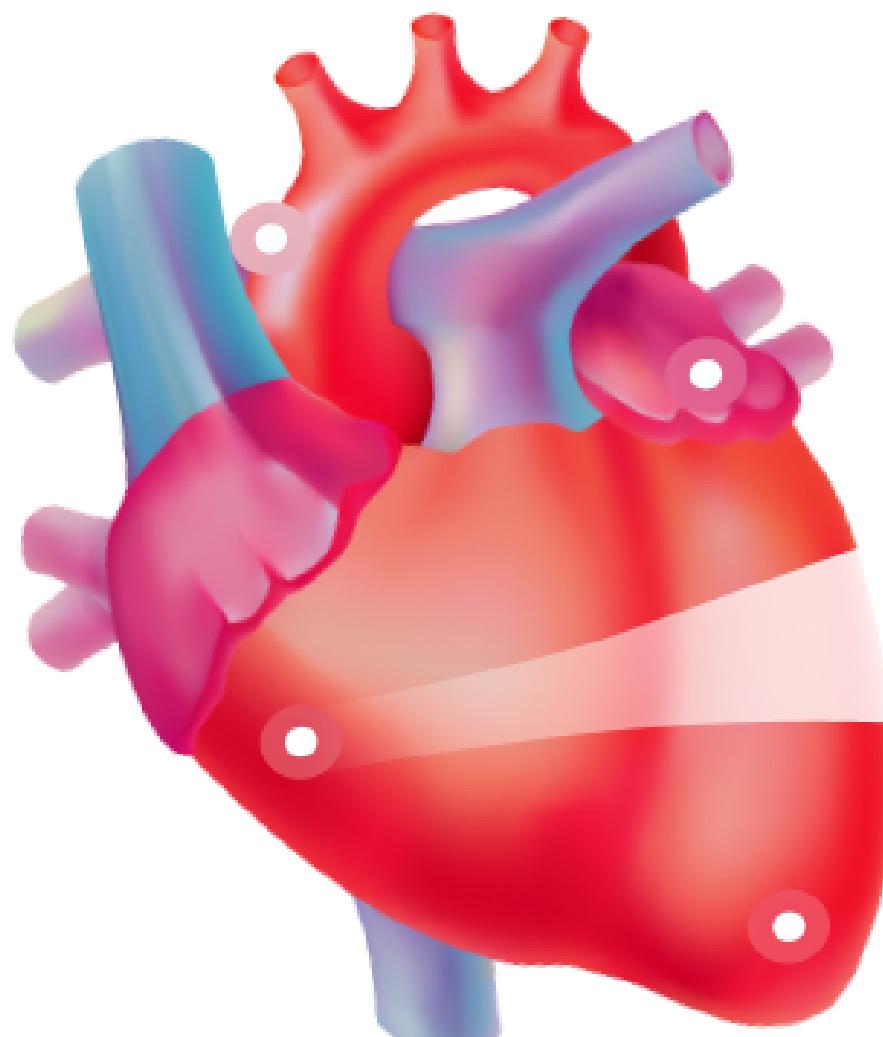
Next



Desktop Application

Dashboard

Patient Information



exercise induced angina

ST depression induced by exercise relative to rest

slope of the peak exercise ST segment

number of major vessels

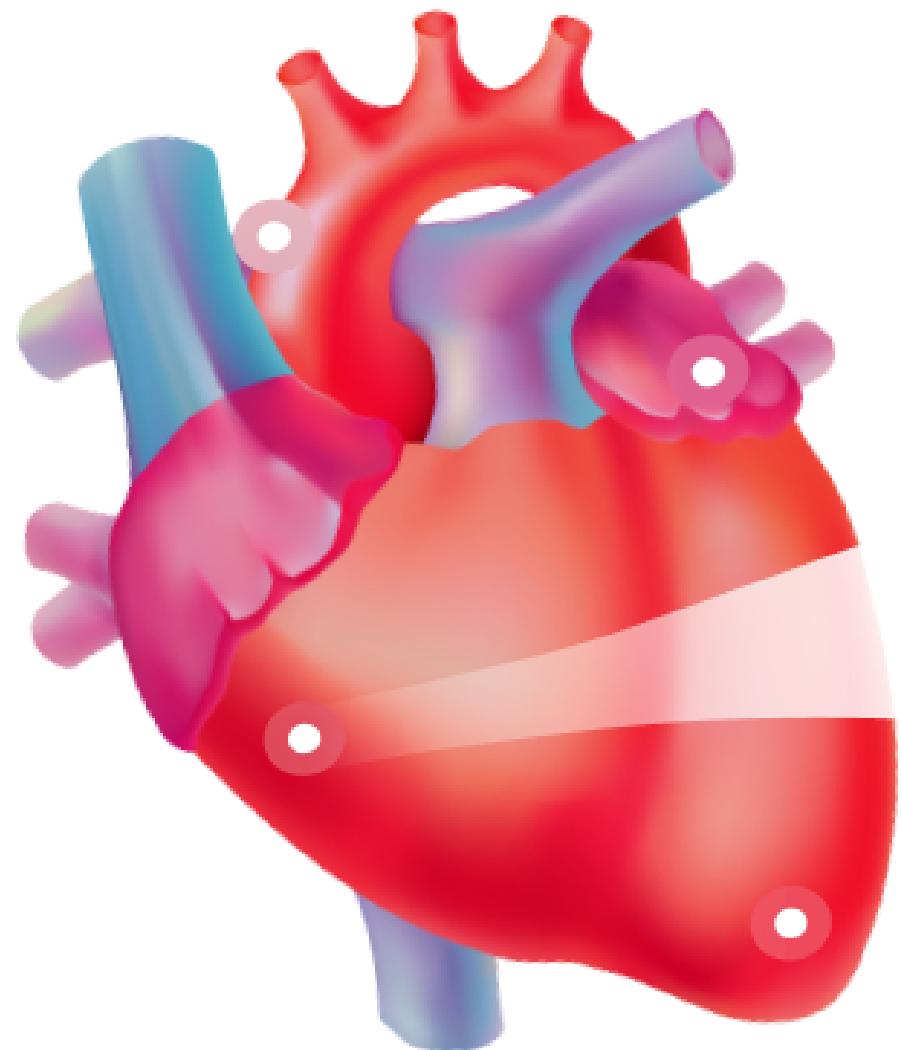
thal

Submit →

Desktop Application

Dashboard

Diagnostic Results

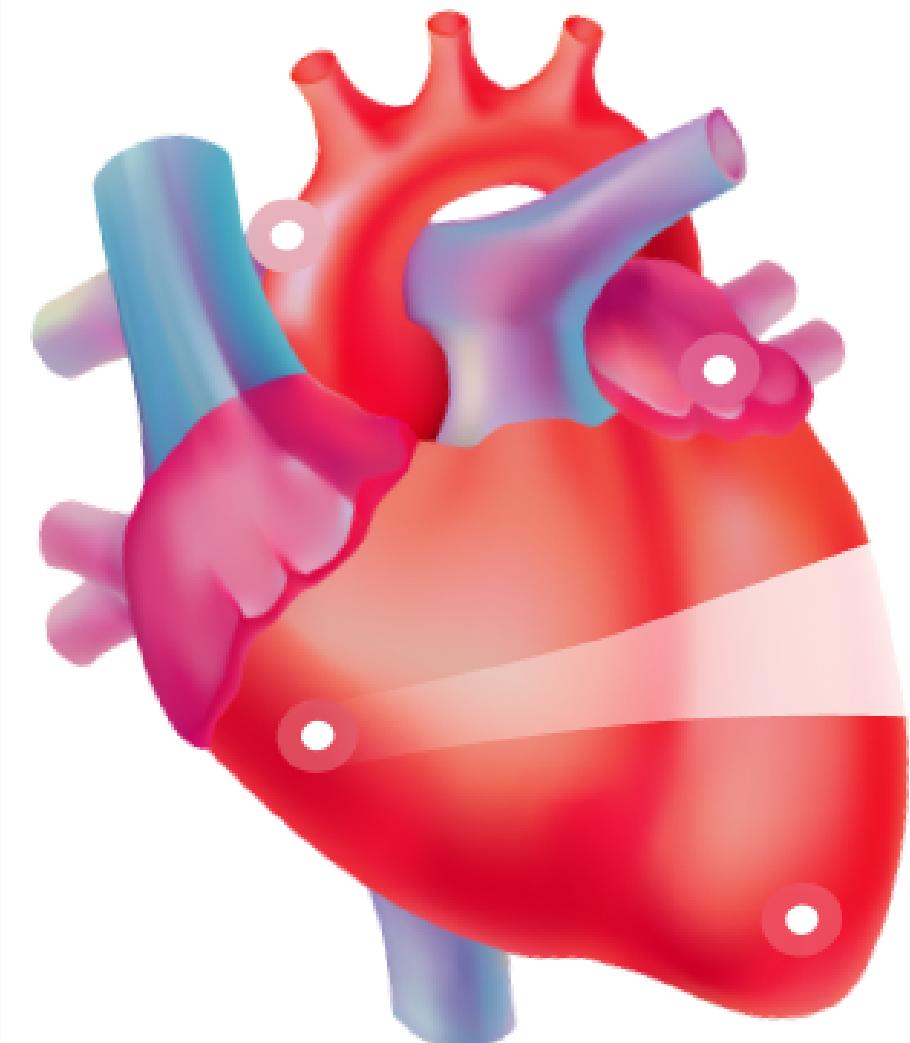


You Are
At Risk!

Desktop Application

Dashboard

Diagnostic Results



You Are
Safe!

Mobile Application

The image displays three screenshots of a mobile application interface for heart disease prediction.

Screenshot 1 (Left): The title "Heart Disease Prediction" is displayed above a subtitle: "Is a machine learning model which predicts the chances of having a heart decease". Below this is a large button labeled "Get Started" with a right-pointing arrow. In the top right corner, there is a "Skip" button.

Screenshot 2 (Middle): The title "Patient Information" is at the top. Below it are several input fields with dropdown arrows:

- Age
- Sex
- Chest Pain Type
- Resting Blood Pressure (mm Hg)
- Serum cholesterol (mg/dl)
- Fasting blood sugar > 120 mg/dl
- Resting electrocardiographic
- Maximum heart rate achieved (BPM)

Screenshot 3 (Right): The title "Patient Information" is at the top. Below it are several input fields with dropdown arrows:

- Exercise induced angina
- ST depression induced by exercise relative to rest
- Slope of the peak exercise ST segment
- Number of major vessels
- thal

A large blue "Submit" button with a right-pointing arrow is located at the bottom right.

Mobile Application

Model Prediction Results

You Are
Safe!

Model Prediction Results

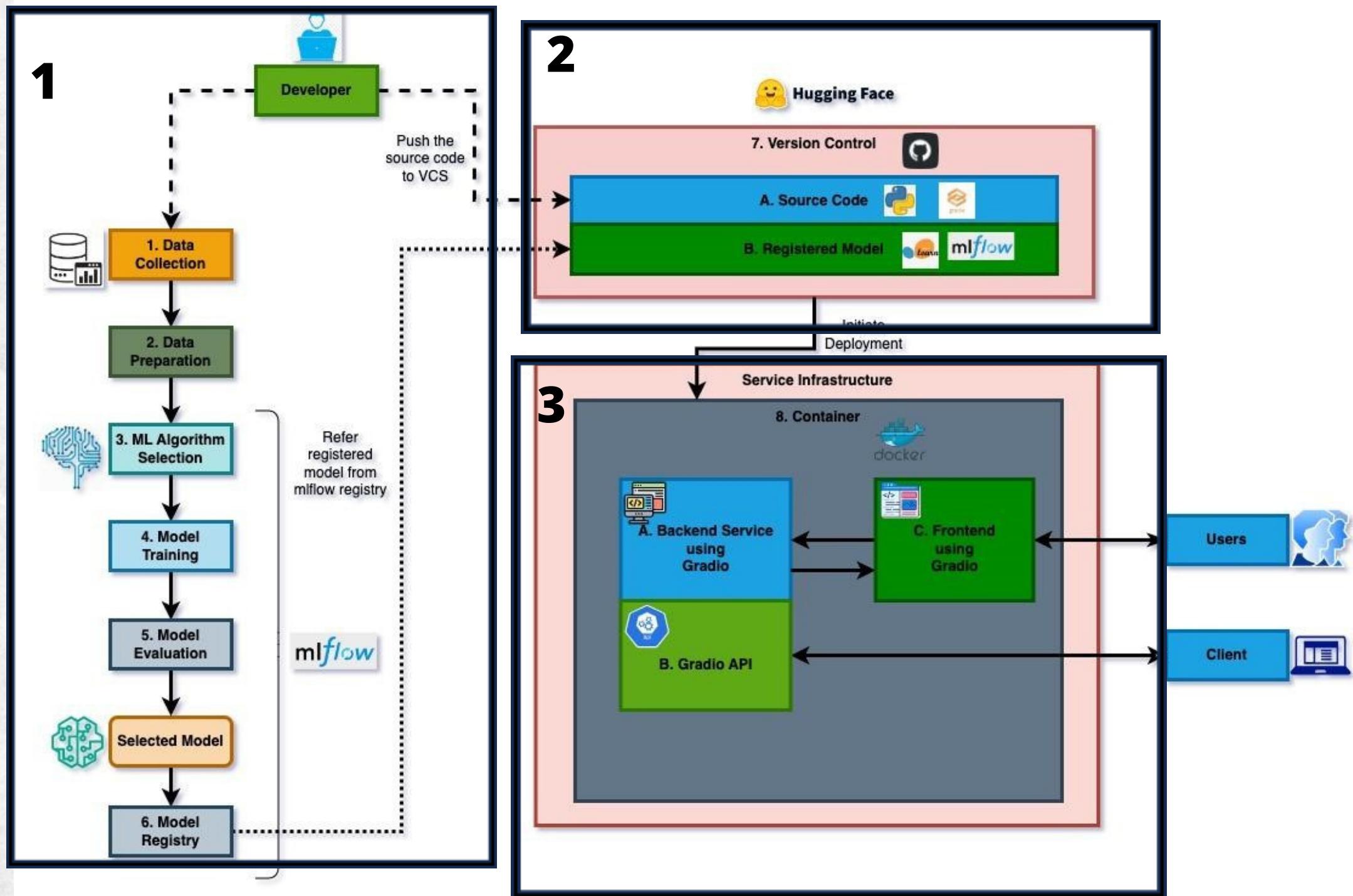
You Are
At Risk!

DIAGRAMS

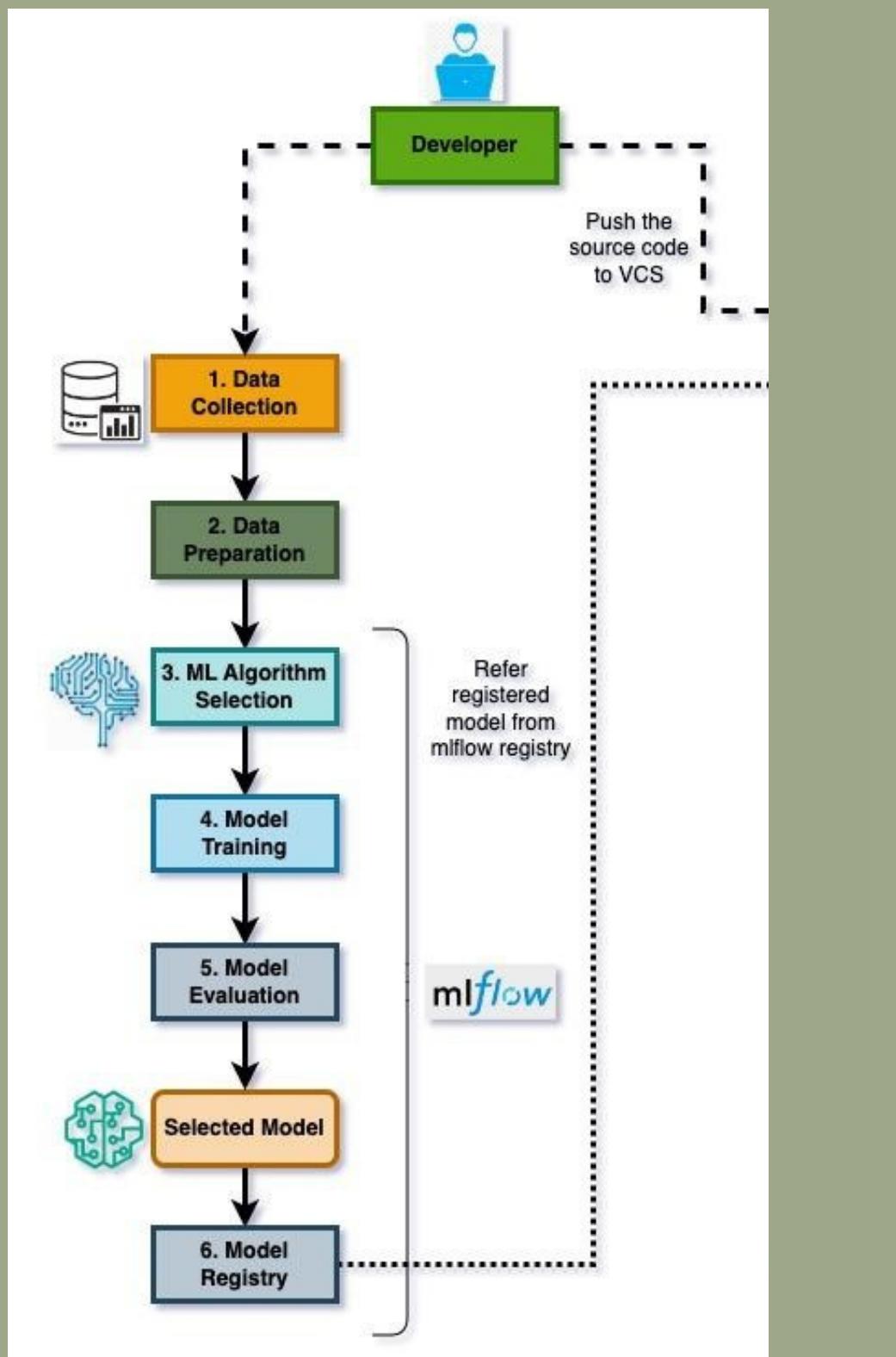
- 1.Training Process
- 2.Version Controlling
- 3.Deployment Process

Infrastructure Description

Task	Platform	Specification
Model training	Linux Machine	<ul style="list-style-type: none"> • vCPU = 16 • Memory = 64 GB
Model deployment	Huggingface Spaces	<ul style="list-style-type: none"> • vCPU = 2 • Memory = 16 GB

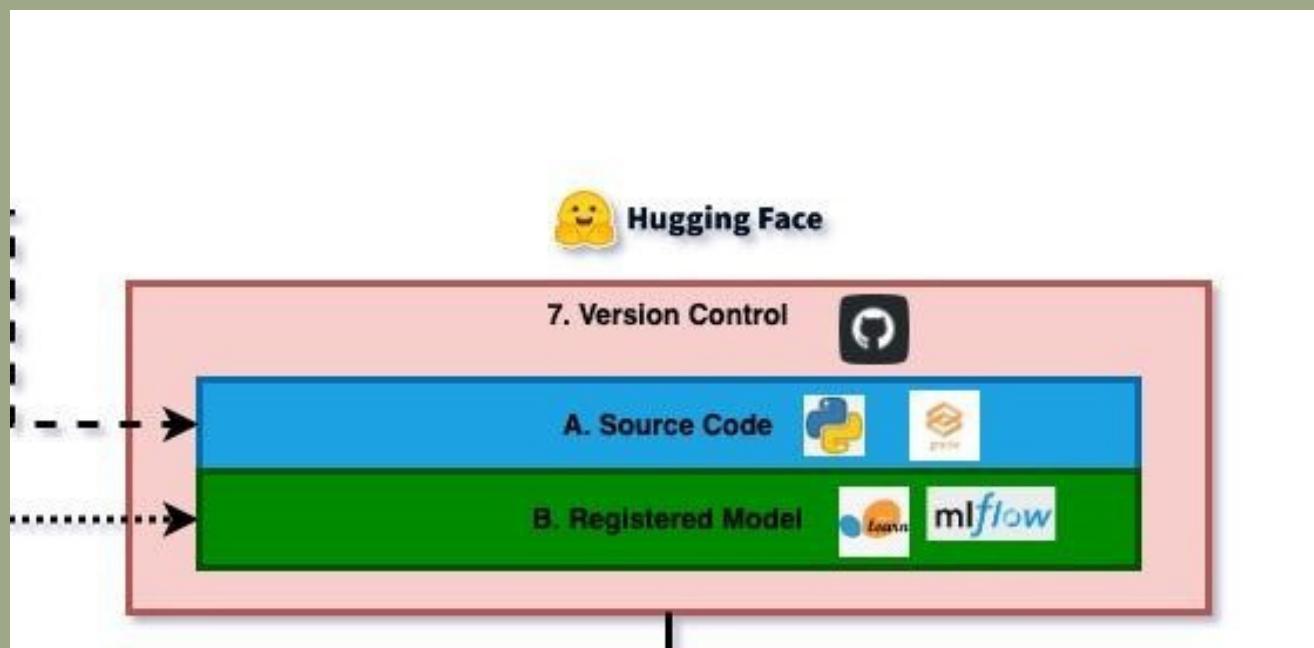


TRAINING PROCESS



- Use Heart Disease Data Set from UCI data repository
- Explore and use into train test split 80:20
- Preprocess train and test dataset.
- Cross Validation on multiple algorithms(k-fold: 10)
- Compare results and select the best performing algorithm
- Perform grid search to find the best set of hyperparameters
- Train the model using the settings we obtained from the previous steps then compute the performance using the test set
- Everything done inside MLFlow environment

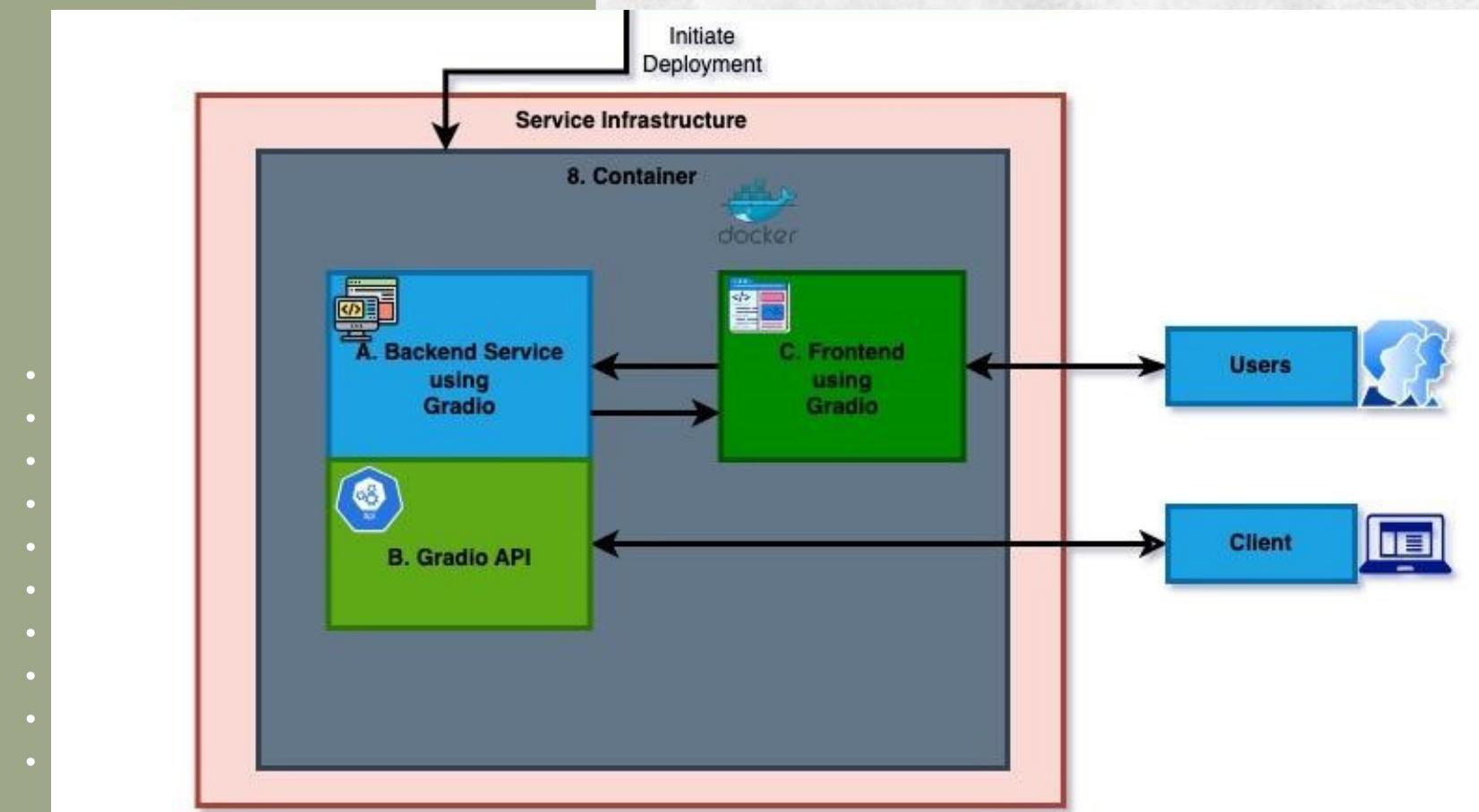
VERSION CONTROLLING



Creating a space for deployment in huggingface, a repository is automatically created for version control for both the model & source code of the application. Huggingface space maintains source versioning using the popular VCS git. So, working with HF Space for version control is the same as using github like VCS.

DEPLOYMENT PROCESS

Using Gradio



Gradio

- An open-source framework simplifying the creation of machine learning-based applications, requiring minimal GUI development.
- Deploying both the backend and frontend elements of these applications.

Gradio

```
Repository Root
|
└─ SRC
    ├ App.py
    └ requirements.txt
```

Gradio

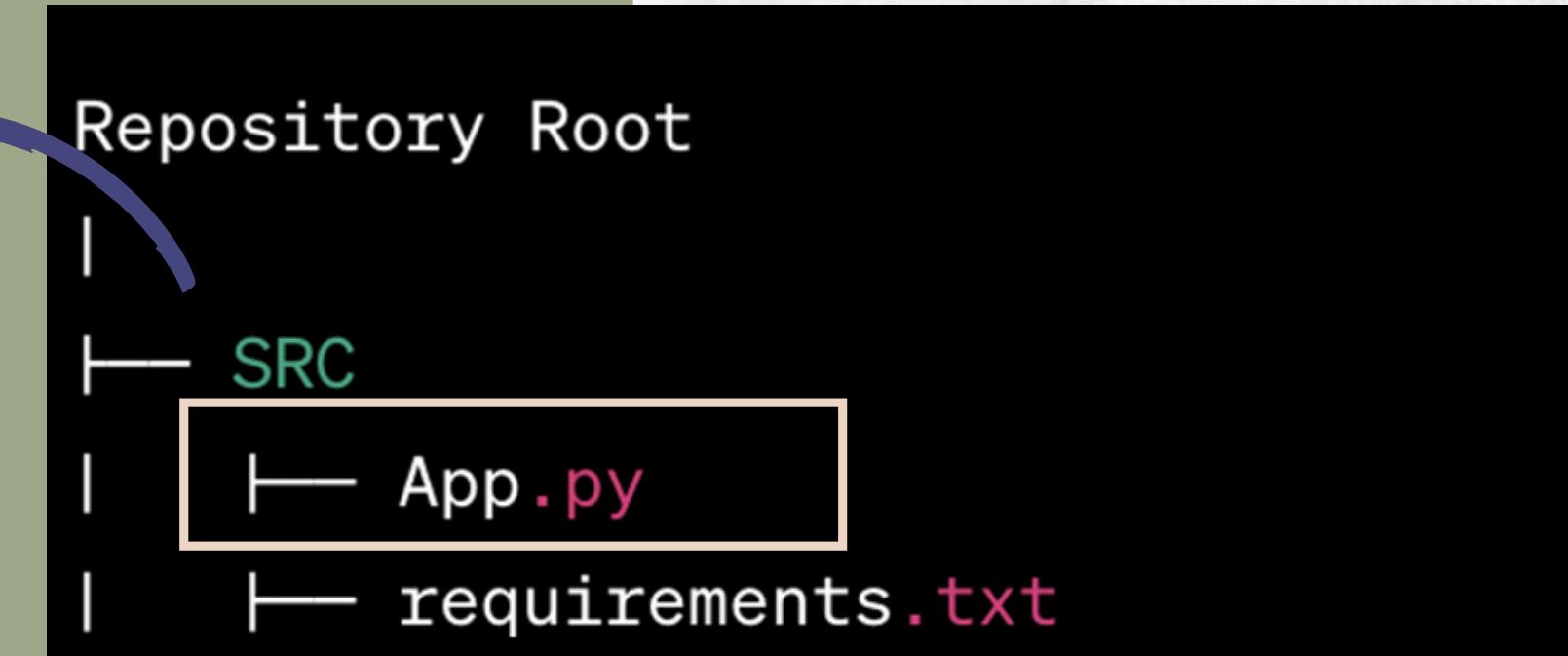
App.py

Backend

- Loads the scikit-learn model from MLflow Model Registry.
- Defines the inference function.
- Specifies input and output validation rules.
- Handles preprocessing and postprocessing of input and output.

Frontend

- Defines GUI components using predefined classes.
- Renders the web interface using a Gunicorn server.



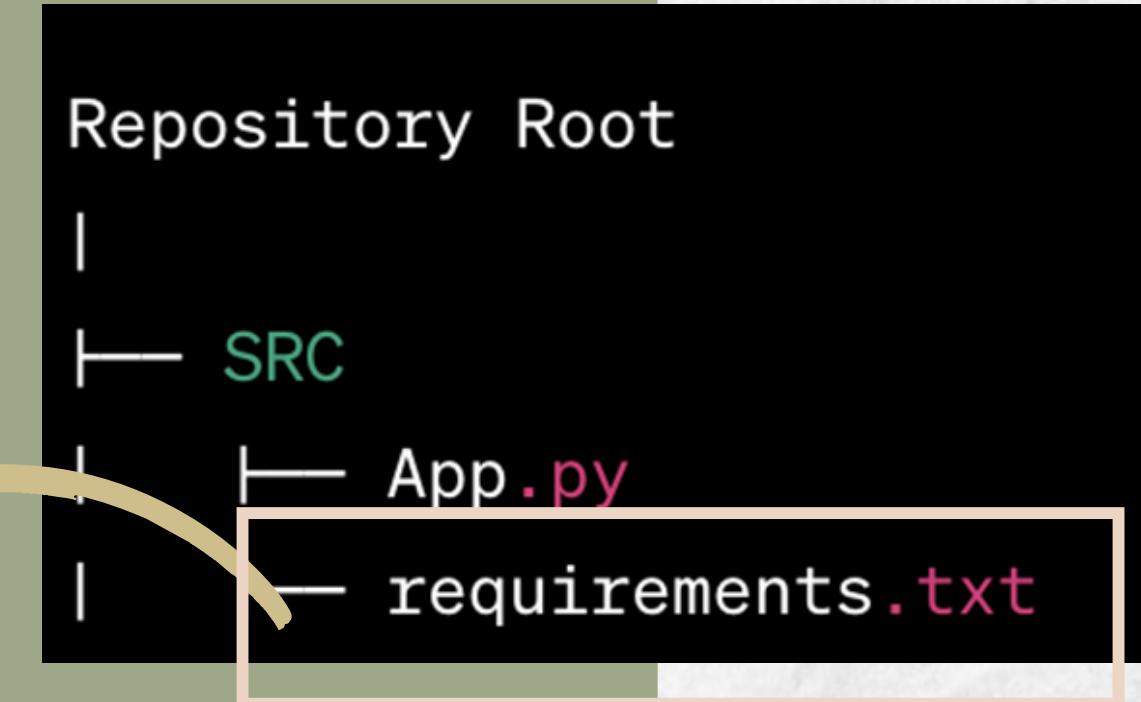
Gradio

requirements.txt

Including a list of libraries required to run the application, such as scikit-learn, numpy, mlflow and others.

MLflow Model Registry

This is the scikit-learn model that has been trained in the previous steps and registered in the MLflow Model Registry and is prepared for deployment after rigorous testing.





DEPLOYMENT SUMMARY

Code Push to vcs

Docker Image Creation

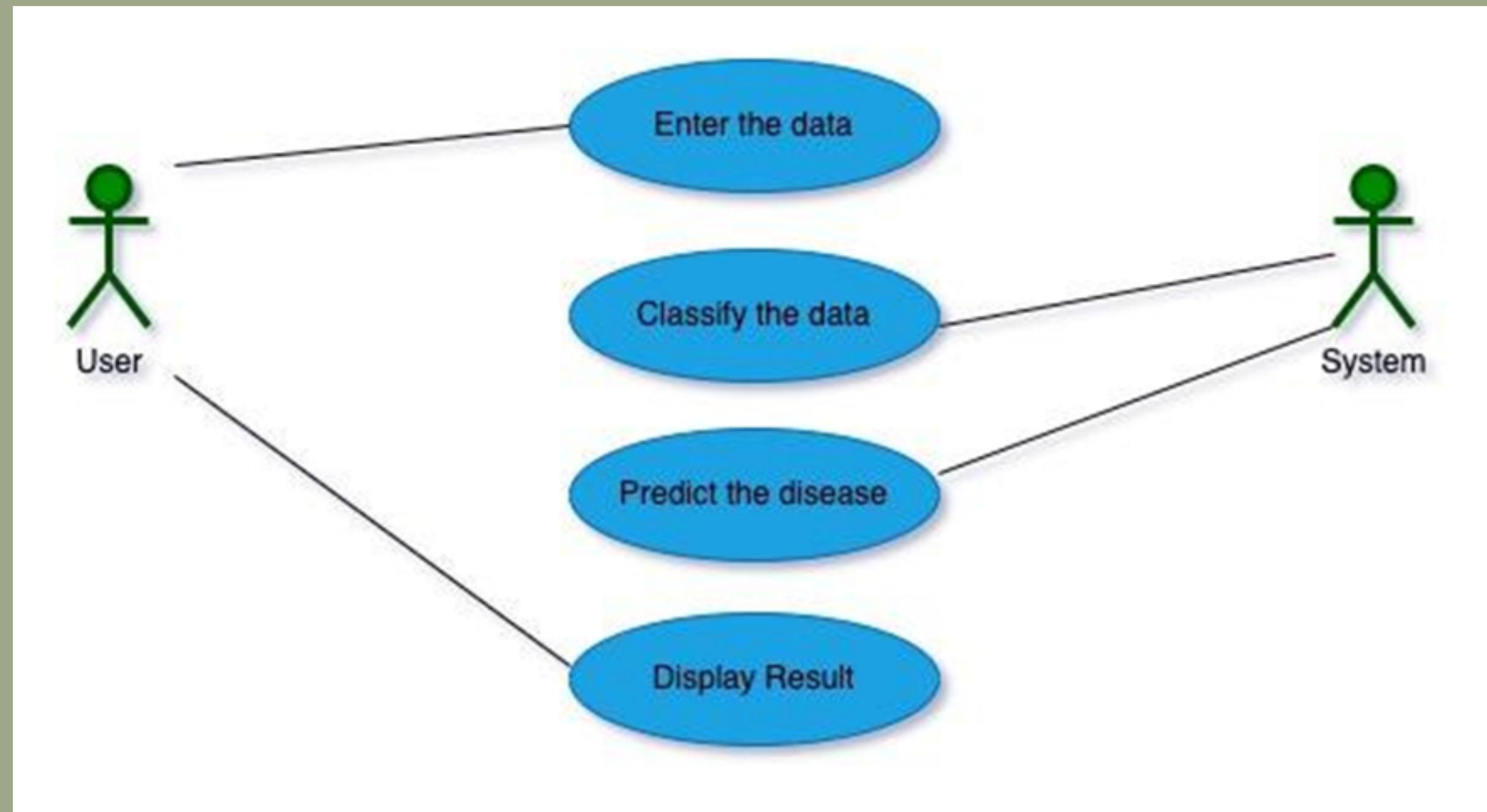
Image Build Success

Container Initialization

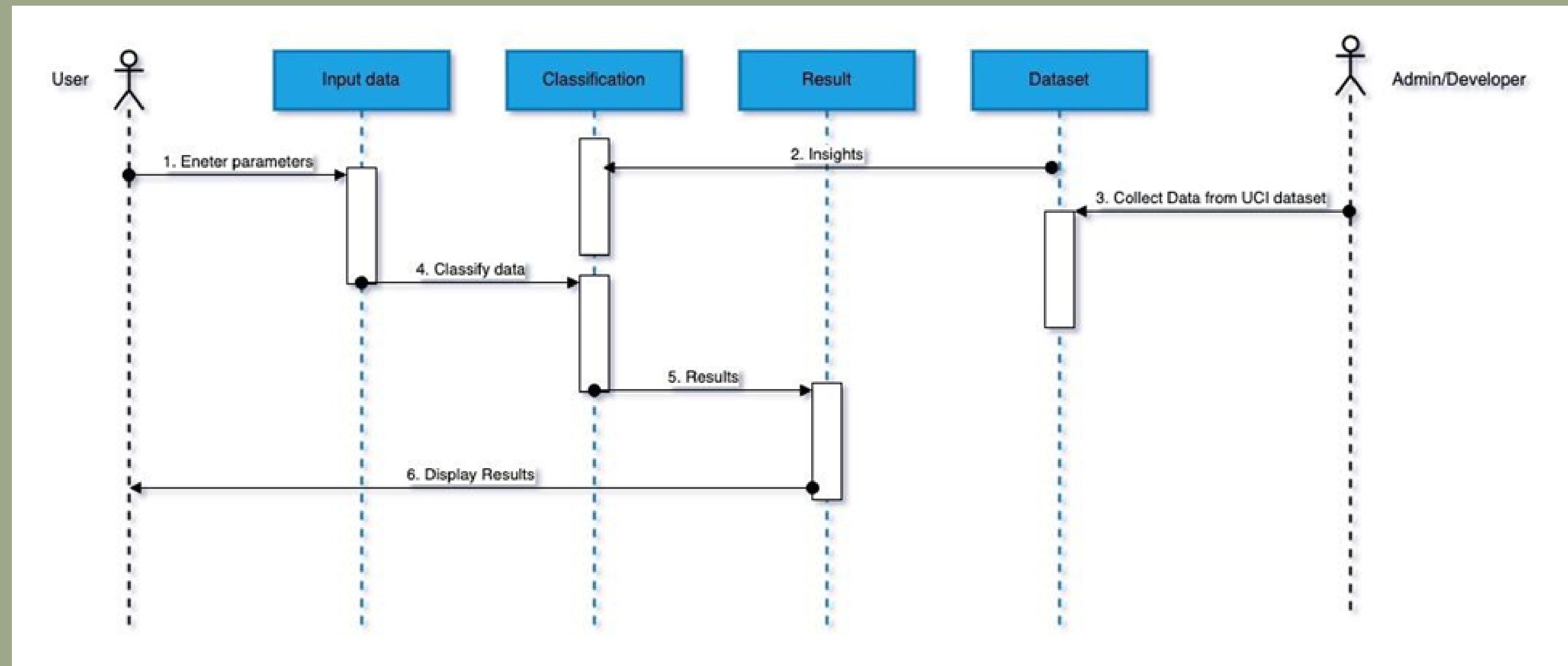
Service Availability

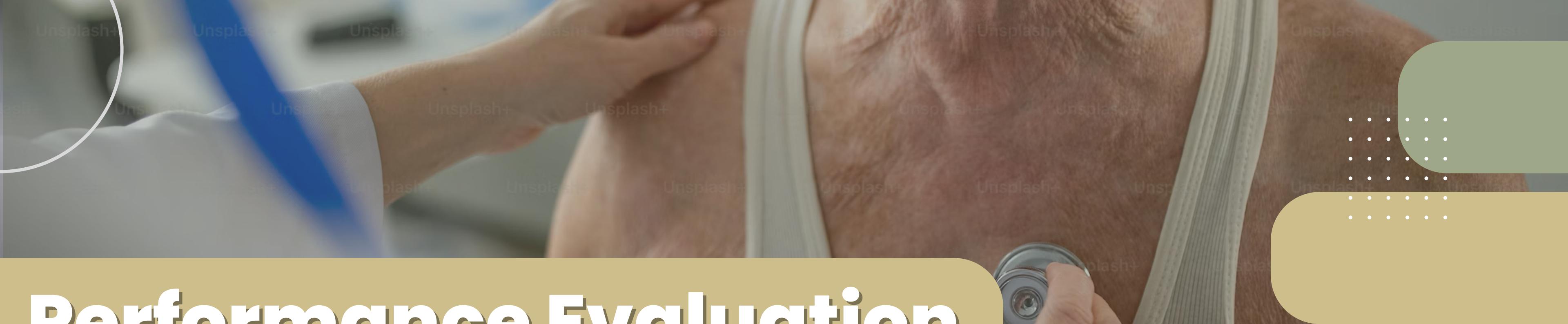
User Interaction

Use Case



Sequence Diagram





Performance Evaluation

Accuracy

Utilized various metrics such as confusion matrices, ROC curves, and precision-recall curves to assess the API's ability to correctly classify.

Precision

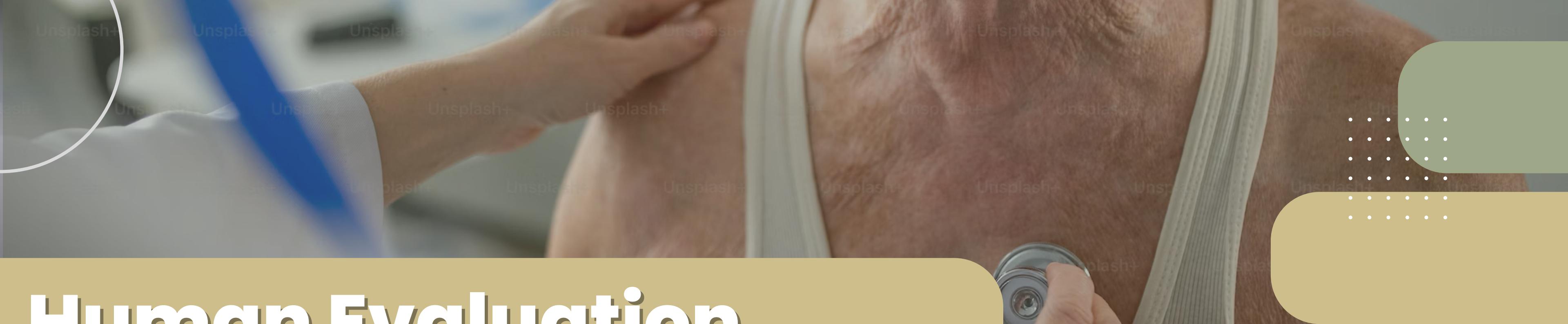
Emphasizing the reliability and quality of positive predictions by assessing the ratio of accurately identified true positives among all predictions categorized as positive.

Recall

Employed sensitivity analysis to determine the API's proficiency in identifying all relevant data in dataset.

F1 Score

Computed a unified metric, considering both precision and recall, to gauge the model's overall performance, particularly important when both precision and recall carry equal importance.



Human Evaluation

Satisfaction of End-Users

Utilized data-driven feedback mechanisms such as surveys and direct interactions to collect comprehensive feedback from end-users and developers, focusing on user experience and practical performance

Preference and User-Friendliness

Employed quantitative analysis to evaluate documentation and integration ease, specifically from a data science perspective. This ensured user-friendliness for data-centric operations and model integrations.

THANKYOU

