رواد مصر الرقمية

وزارة الاتصـــــالات
وتكنولوجيا المعلومات

Heart Disease Risk Prediction

Presented By : Mostafa Gamal Abdel-Fatah Fouda

Date : 15/5/2025

Project Idea

➢ Problem Definition:

Heart disease remains the leading cause of death globally. Many patients remain undiagnosed until severe symptoms appear. There is a crucial need for early, accessible, and affordable prediction tools that help identify individuals at risk before complications occur.

➢ Proposed Solution:

This project delivers a complete machine learning pipeline to predict the presence of heart disease based on structured patient data. The system is deployed as a user-friendly web application using Streamlit, allowing users to input clinical parameters and receive instant feedback.

➢ Objectives:

- Automate heart disease risk prediction using machine learning.
- Provide an interactive and accessible web interface.
- Improve early detection and support health decisions.

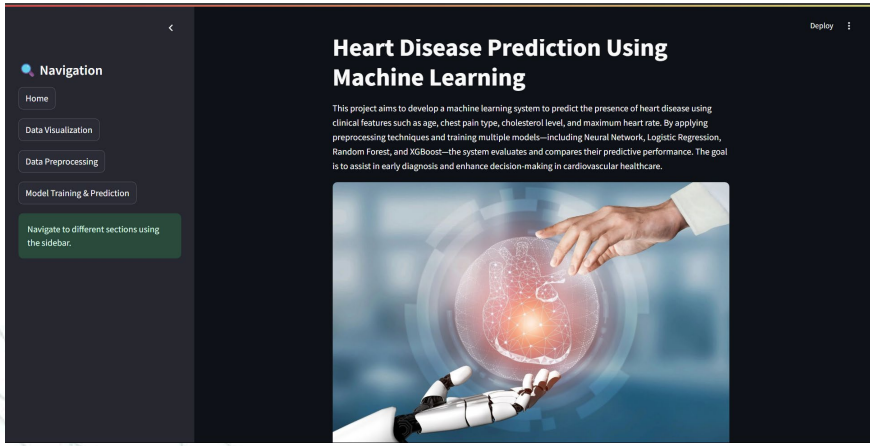
➢ Unique Value Proposition:

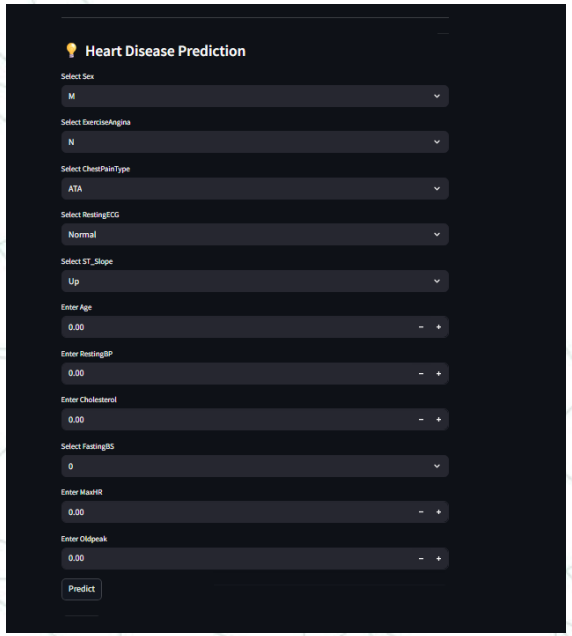
- **End-to-End Pipeline:** Covers all steps from data preprocessing to deployment.
- **Multi-Model Comparison:** Trained multiple models Random Forest, Logistic Regression, XGBoost and Neural Network.
- **Real-Time Prediction:** Immediate feedback via a live Web app.
- **Visual Insights:** Includes intuitive data visualizations for users and analysts.
- **Scalable & Extendable:** Can be enhanced with more features or applied to other diseases.

➢ Real-World Impact:

- Assists doctors with a quick second opinion.
- Empowers individuals to assess their risk anytime.
- Can be integrated into health screening tools or mobile apps in the future.

Project Wireframe





User Journey Overview:

1. Home Page – Project Overview

1. Brief introduction to the purpose of the app.
2. Explains the problem and the goal of the prediction system.
3. Navigation menu for moving between pages.

2. Preprocessing Page

1. Shows how raw data is cleaned and prepared.
2. Displays steps like: missing value handling, outlier removal, feature encoding & scaling.
1. Visual feedback for each preprocessing stage.

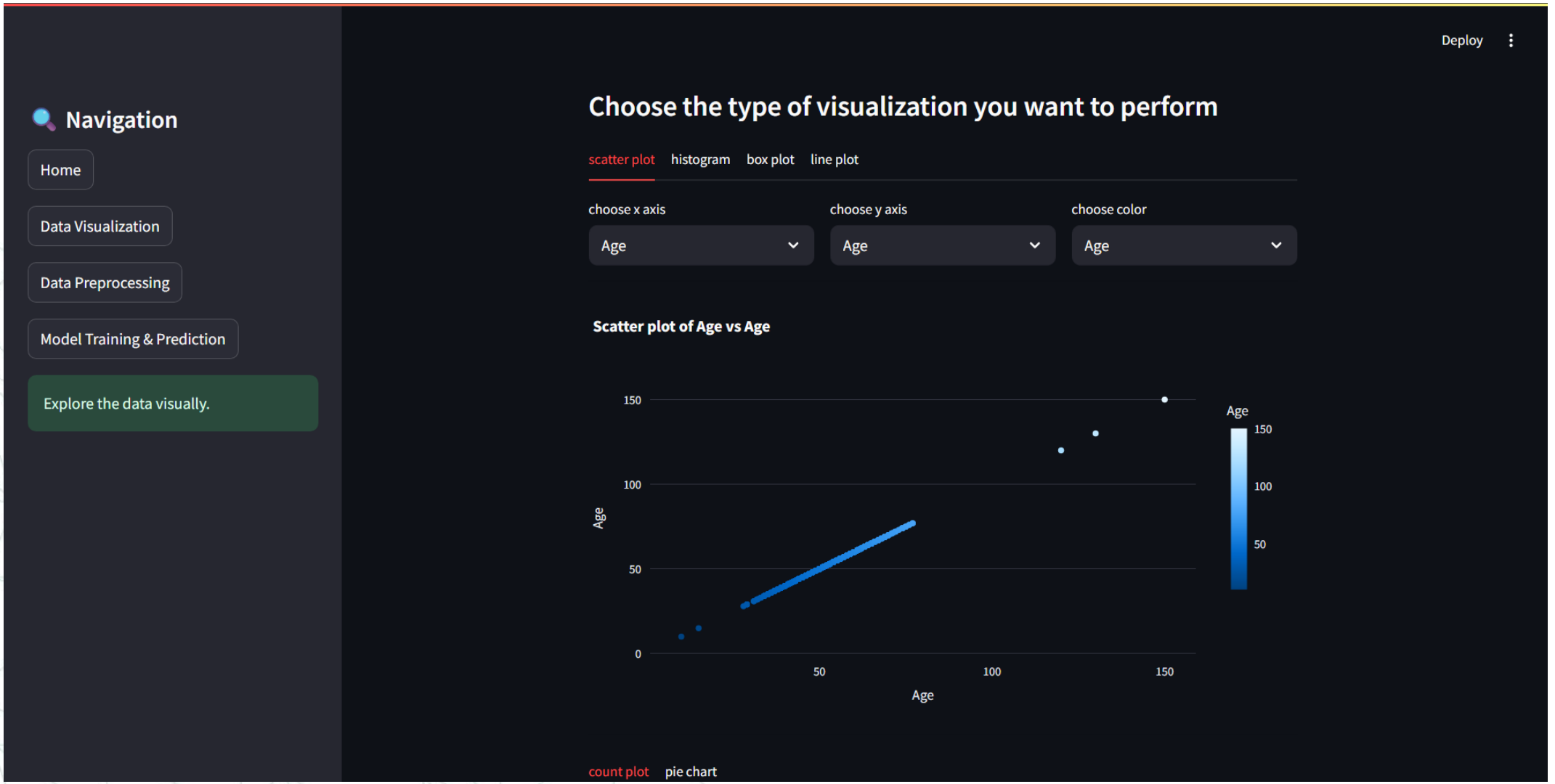
3. Visualizations Page

1. Exploratory Data Analysis (EDA) to understand feature distributions.
2. Charts like bar plots, pie charts, correlation heatmaps.
3. Helps users explore important risk indicators.

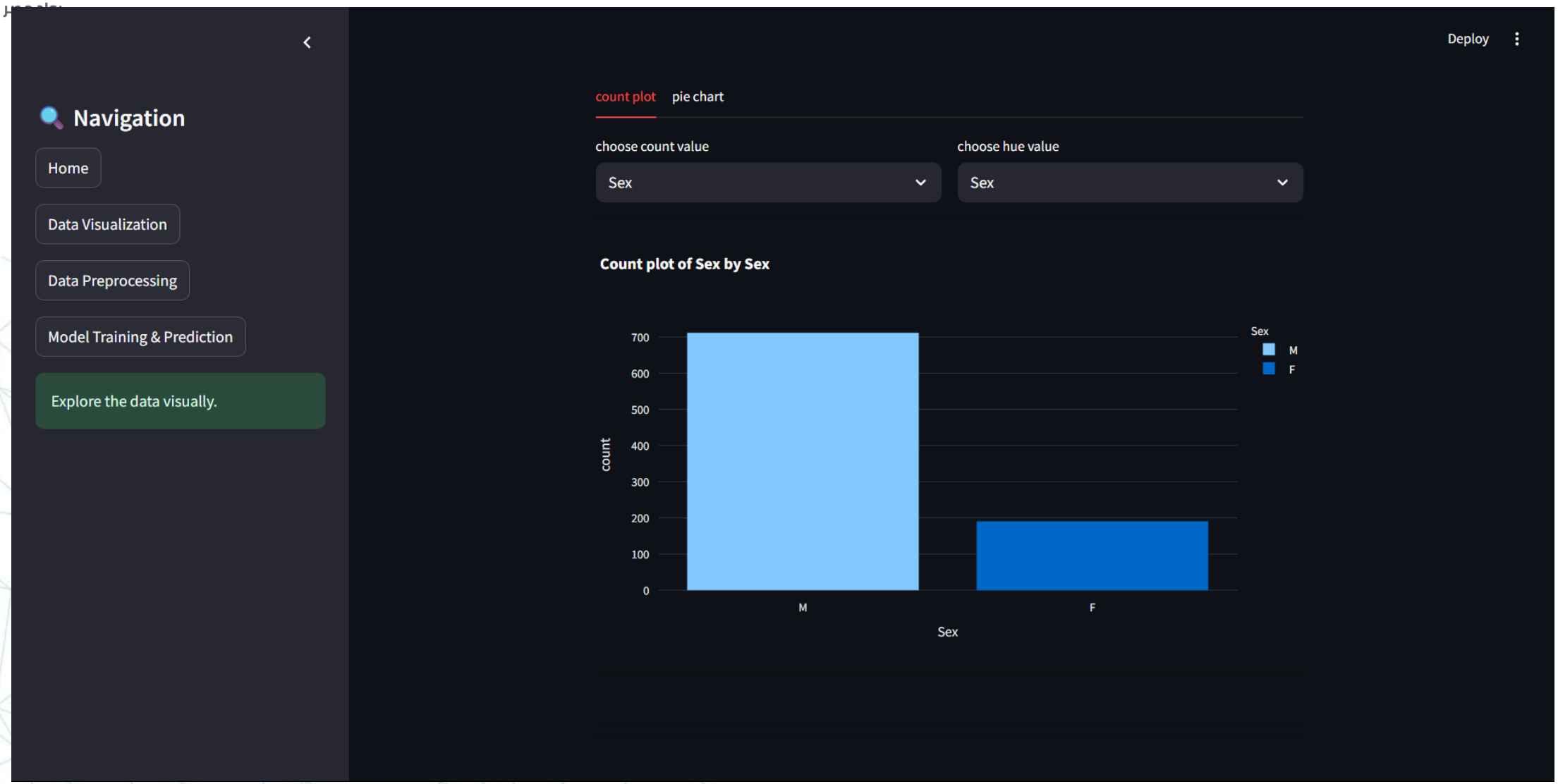
4. Modeling Page

1. Allows users to input medical data manually.
2. Choose from multiple models: Random Forest, Logistic Regression, XGBoost, Neural Network.
3. Click "Predict" to get real-time risk prediction and probability.
4. Shows evaluation metrics like accuracy, precision, recall.

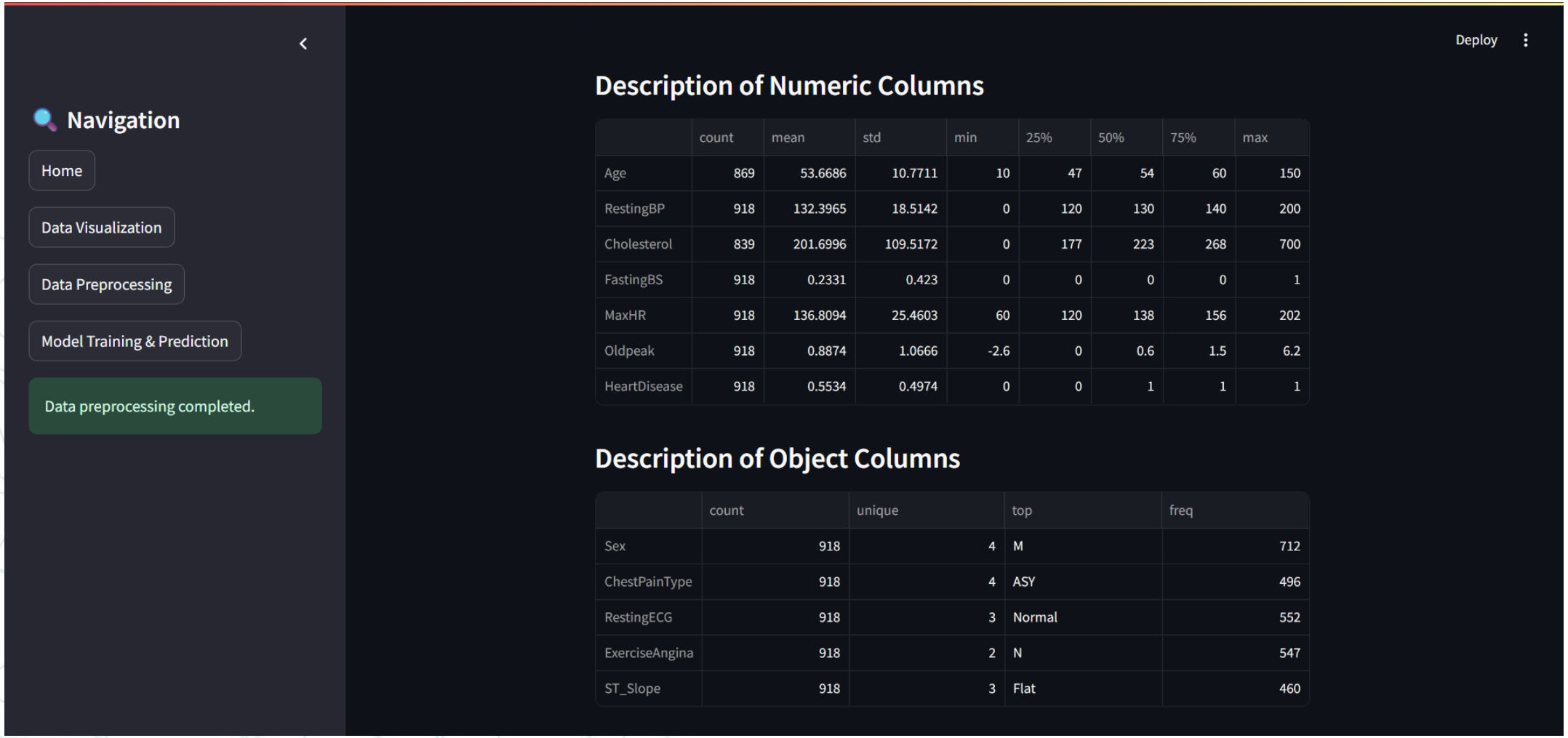
Project Wireframe



Project Wireframe



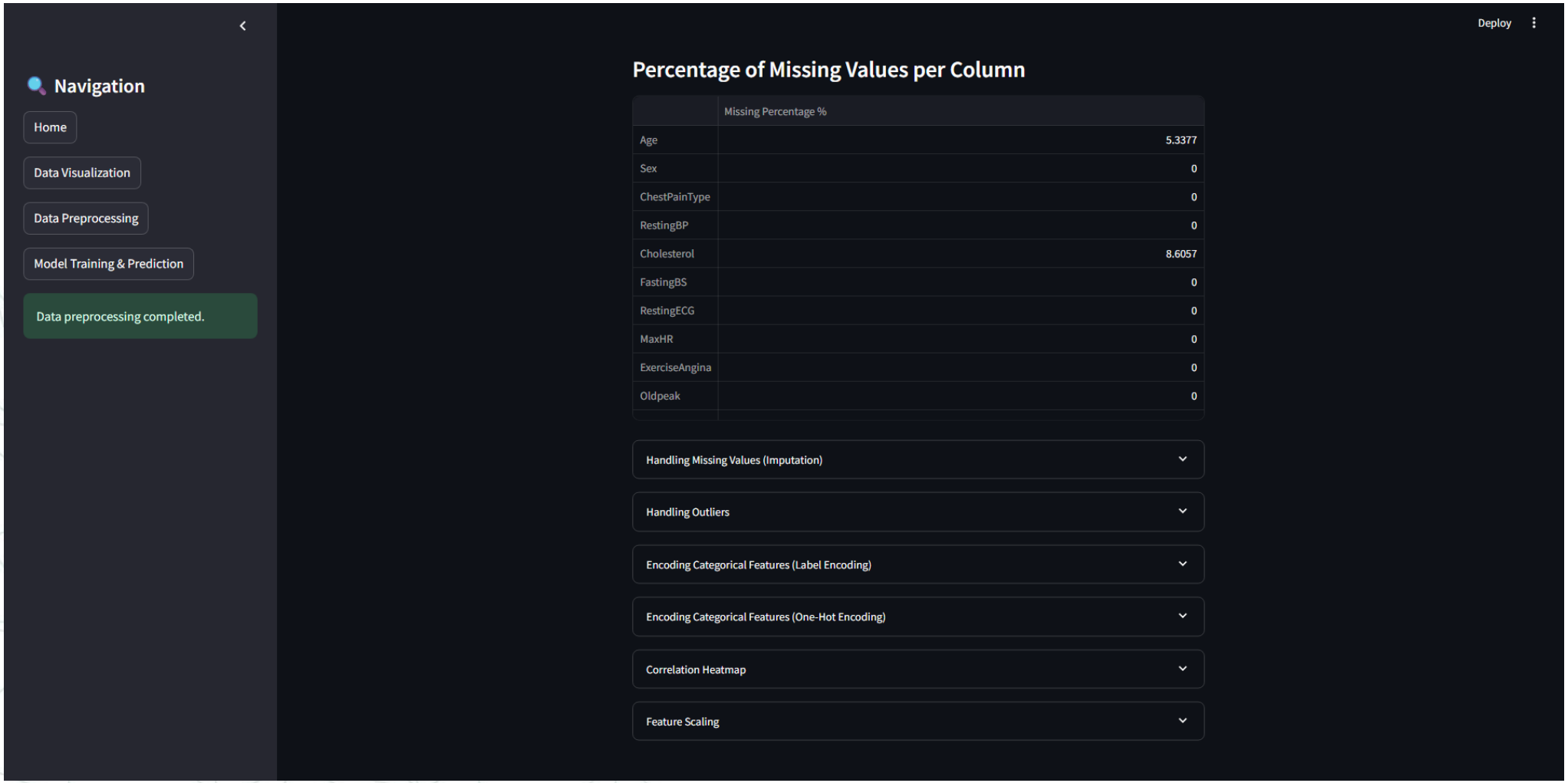
Project Wireframe



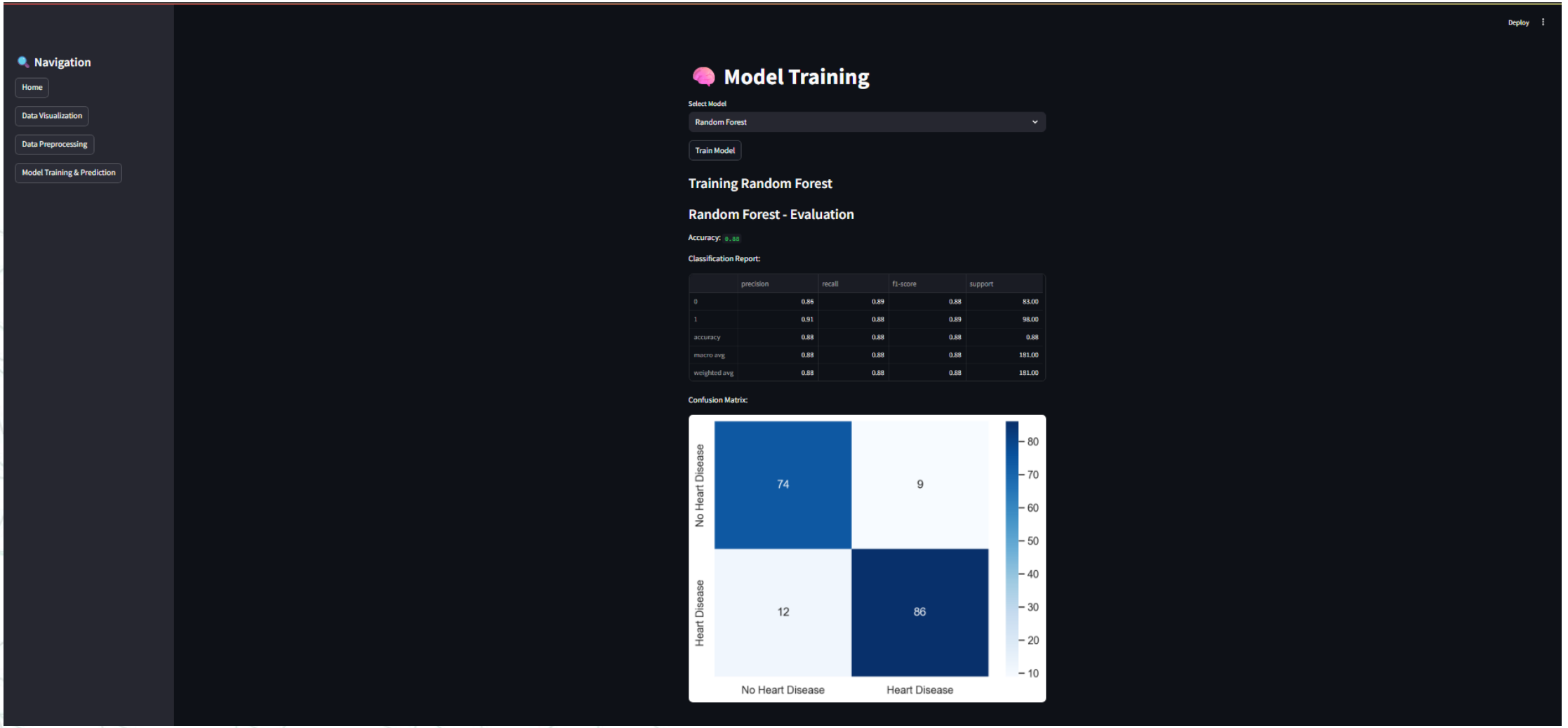
	count	mean	std	min	25%	50%	75%	max
Age	869	53.6686	10.7711	10	47	54	60	150
RestingBP	918	132.3965	18.5142	0	120	130	140	200
Cholesterol	839	201.6996	109.5172	0	177	223	268	700
FastingBS	918	0.2331	0.423	0	0	0	0	1
MaxHR	918	136.8094	25.4603	60	120	138	156	202
Oldpeak	918	0.8874	1.0666	-2.6	0	0.6	1.5	6.2
HeartDisease	918	0.5534	0.4974	0	0	1	1	1

	count	unique	top	freq
Sex	918	4	M	712
ChestPainType	918	4	ASY	496
RestingECG	918	3	Normal	552
ExerciseAngina	918	2	N	547
ST_Slope	918	3	Flat	460

Project Wireframe



Project Wireframe



End Users + Features

➢ Target Users:

- **Doctors & Medical Professionals**

  Use as a fast decision-support tool to identify patients at risk and prioritize further testing.

- **Health-Conscious Individuals**

  Allows non-experts to check their risk level based on simple medical inputs, promoting early awareness.

- **Healthcare Programs & Institutions**

  Useful for public health screening and can be integrated into digital awareness platforms.

➢ Key Features:

- Real-time heart disease prediction using clinical inputs.
- Multiple models supported: Random Forest, Logistic Regression, XGBoost, Neural Network.
- Clean and user-friendly interface built with Streamlit.
- Visual feedback (charts, probabilities) to enhance understanding.
- Fully deployed and accessible online.

➢ How It Helps:

- Speeds up early diagnosis and supports clinical decisions.
- Encourages individuals to monitor their health proactively.
- Scales well for public use and educational demonstrations

Data Structure

- **Data Source:**
  A structured CSV file containing labeled medical records of patients. The dataset includes various clinical measurements and diagnosis labels.
- **Data Type:**
  - Mixture of **numerical features** (e.g., Age, RestingBP, Cholesterol, MaxHR)
  - And **categorical features** (e.g., Sex, ChestPainType, ExerciseAngina)
- **Target Variable:**
  - **Heart Disease** (Binary Classification):
    0 : No heart disease
    1 : heart disease
- **Preprocessing Steps:**
  - **Handling Categorical Variables:**
    - Applied **Label Encoding** for binary categories (e.g., Sex, ExerciseAngina)
    - Used **One-Hot Encoding** for multi-class features (e.g., ChestPainType, RestingECG)
  - **Data Cleaning:**
    - Removed or treated unrealistic values (e.g., 0 in blood pressure or cholesterol)
    - Checked for duplicates and missing values
  - **Feature Scaling:**
    - Standardized numerical features to improve model performance and convergence

Programming Languages + Frameworks

- **Language:**

Python

- **Frameworks & Libraries:**

Streamlit

Scikit-learn

TensorFlow/Keras

XGBoost

- **Tools & Hosting:**

Pandas

NumPy

Matplotlib

Seaborn

Streamlit Cloud

plotly

Live Application + Test

Section	Details
Application Status	- Live and fully functional. - Publicly accessible via Streamlit Cloud - Four main pages: Home, Preprocessing, Visualizations, Modeling.
Deployment	- Hosted on Streamlit Cloud. - No installation required (browser-based). - GitHub integrated for version control and updates.
Testing – Unit	- Core functions tested individually (e.g., encoding, scaling, prediction logic). - Ensured each part behaves as expected.
Testing – Validation	- Train/Test split used to evaluate model. - Calculated accuracy, precision, recall on unseen data.
Testing – User (UX/UI)	- Interface tested with non-technical users. - Feedback applied to improve usability, layout, and navigation.
Outcome	- Reliable predictions with clear feedback. - Ready for educational use or real-time health risk checks.

Deliverables (Reports, etc.)

Deliverable	Description
Streamlit Web Application	Fully functional, interactive web app for real-time heart disease prediction.
Cleaned Dataset	Preprocessed and cleaned version of the original dataset, ready for analysis and modeling.
Source Code & Notebooks	All code files and Jupyter notebooks used in data processing, EDA, model training, and testing.
Documentation	Includes: • Data Dictionary • Preprocessing steps • Model explanation • User Guide
Deployment Link	Publicly accessible link to the deployed web app (hosted on Streamlit Cloud). https://heart-disease-predictor-lyz98qemywla9hy8fxz76y.streamlit.app/

Project Team + Roles

Name	Role	Key Responsibilities	Tools Used
Mostafa Gamal Abdel-Fatah Fouda	Data Scientist	Modeling, Deployment , Presentation	GitHub, Jupyter, Streamlit, Power point , Scikit-learn (sklearn), XGBoost , Python, TensorFlow/Keras, Streamlit Cloud
Mohamed Emad Ibrahim Mostafa	Data Scientist		
Lucas Alkomos Philopater Zakher Hanna	Data Scientist	EDA, Visualization, Documentations	Python, Seaborn, Matplotlib, Plotly, Jupyter, Microsoft Word, PowerBI
Youssef Alaa El Din Metwally Musallam	Data Scientist		
Maryam Gamal Ahmed Kamal Askar	Data Scientist	Data Cleaning, Data preprocessing, User testing	Python, Pandas, NumPy, Jupyter, Scikit-learn (sklearn), Seaborn
Marco Wael Issa Ibrahim	Data Scientist		

Thank You

Thank you for your attention and interest in this project!

Feel free to ask any questions or share your feedback.

Contact:

Email: mostafagamalf9@gmail.com

GitHub: https://github.com/MostafaGmalFouda

LinkedIn: linkedin.com/in/mostafa-gamal-fouda-2645212a4