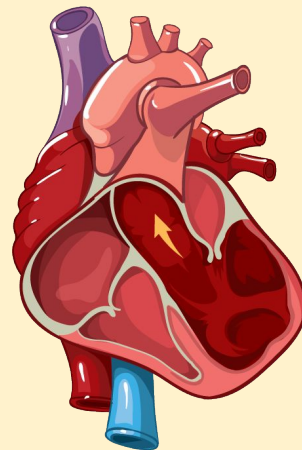
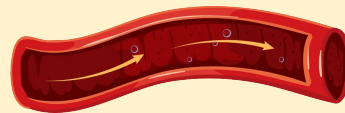


Promoting Cardiovascular Health for All

An ML approach

Yaire Souffront



01

Overview

What is Cardiovascular Disease (CVD)?

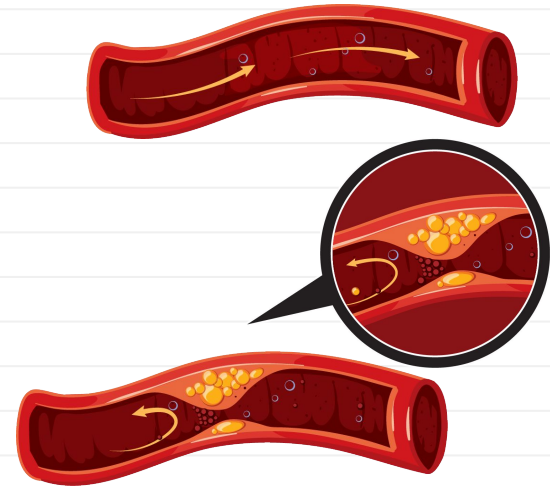
Causes

Can Vary
Plaque Buildup in Arteries
Infections, Aging

Risk Factors

High Blood Pressure
High Cholesterol
Lack of Physical Activity

Atherosclerosis



Sources: [Cleveland Clinic](#) , [WHO](#)

Purpose of this Project



Questionnaire: To determine whether or not someone is at risk of developing cardiovascular disease, only using their demographic information

Accessibility: participants only rely on demographic data (age, education level, weight, height, gender)

Learning: Application of data preprocessing, machine learning approaches, and model deployment

Why is this Important?

Problem

\$252.2B

- Plummeting Productivity
- Decrease in Age Mortality
- Silent Killer

Solution

- Use Machine Learning!
- Make informed decisions regarding diagnosis and treatments
- Competitive Advantage

Result

- Improved Quality of Life
- Early Detection
- Cost Reduction

Source: CDC

About the Data (NHANES)

Who

- US and Non US Citizens
- Participants chosen at random
- Must Live in US

What

- Study to assess health and nutritional status of participants
- Focused on Labs, Diet, Examination, & Demographic data

When

- NHANES program started in the 1960s
- Data was collected throughout the 2013-2014 year

Source: NHANES

A Data Science Approach



Data Collection & Cleaning

Finding Data from reliable source



EDA

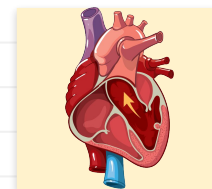
Finding Patterns + Insights



Model Building

Building & Optimizing Models

NHANES Datasets

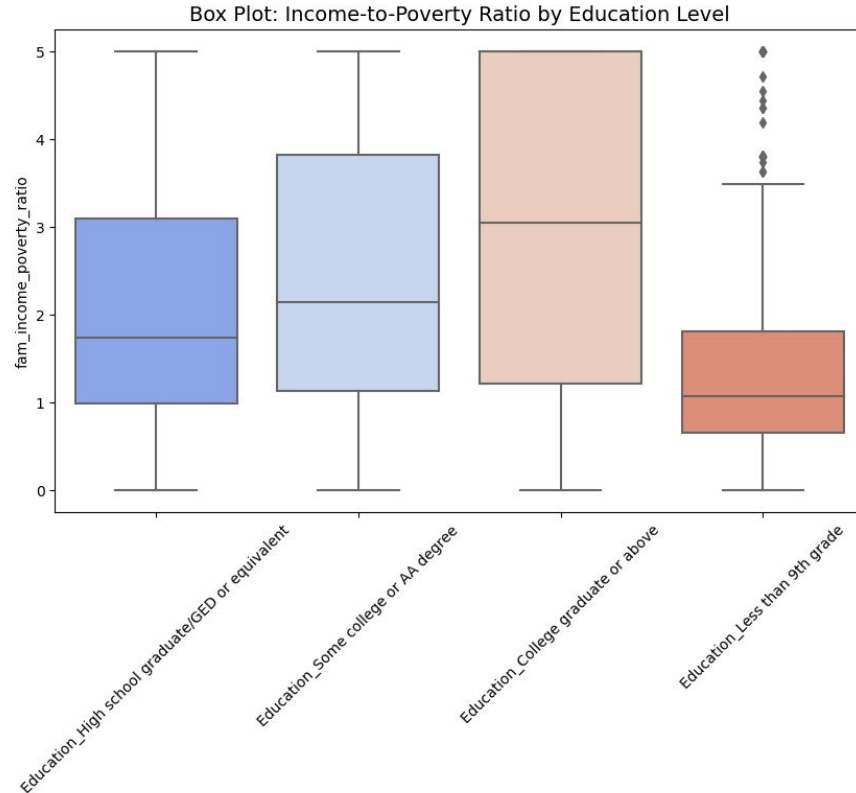


Demographic	Holds demographic data of participants, such as US citizenship, household income, gender, and age
Questionnaire	Contains basic level information of participant's medical history, physical activity, family-level information, and dietary behavior
Labs	includes cholesterol levels, blood glucose, and other metabolic indicators
Diet	Provides detailed information on participants' dietary intake
Examination	Contains physical examination of participants such as bmi, height, waist circumference, and blood pressure readings.
Medications	Contain prescription information, amount of medications used, and how frequent the medications are taken

02

Insights

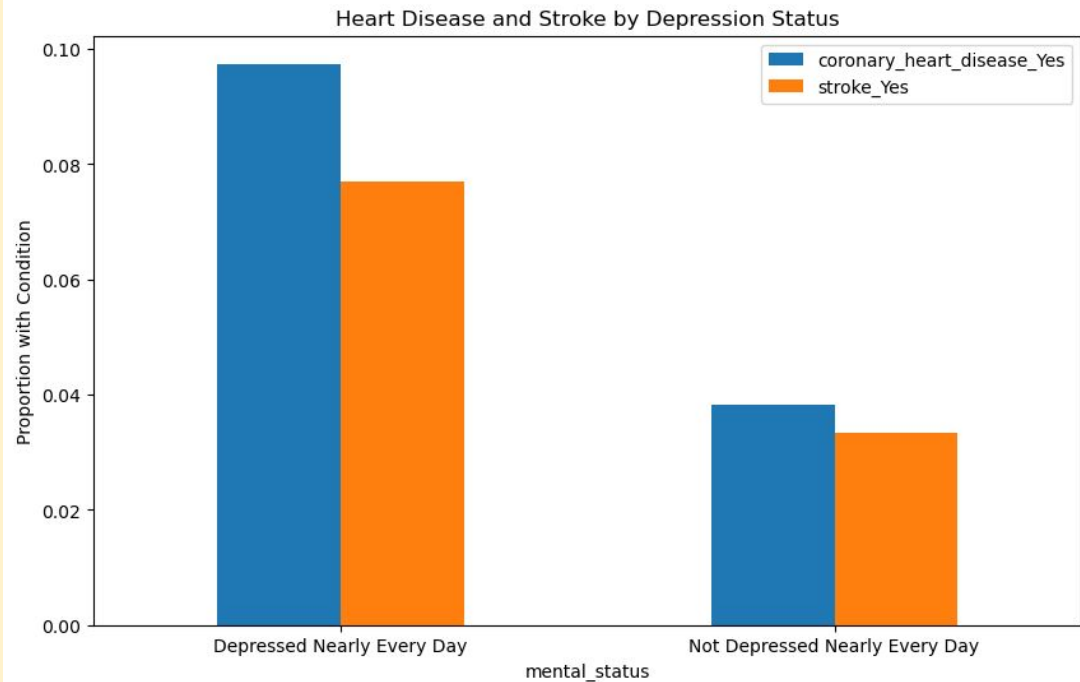
Education and Poverty Level Insights



Fam_income_poverty_ratio represents the ratio of family income to poverty

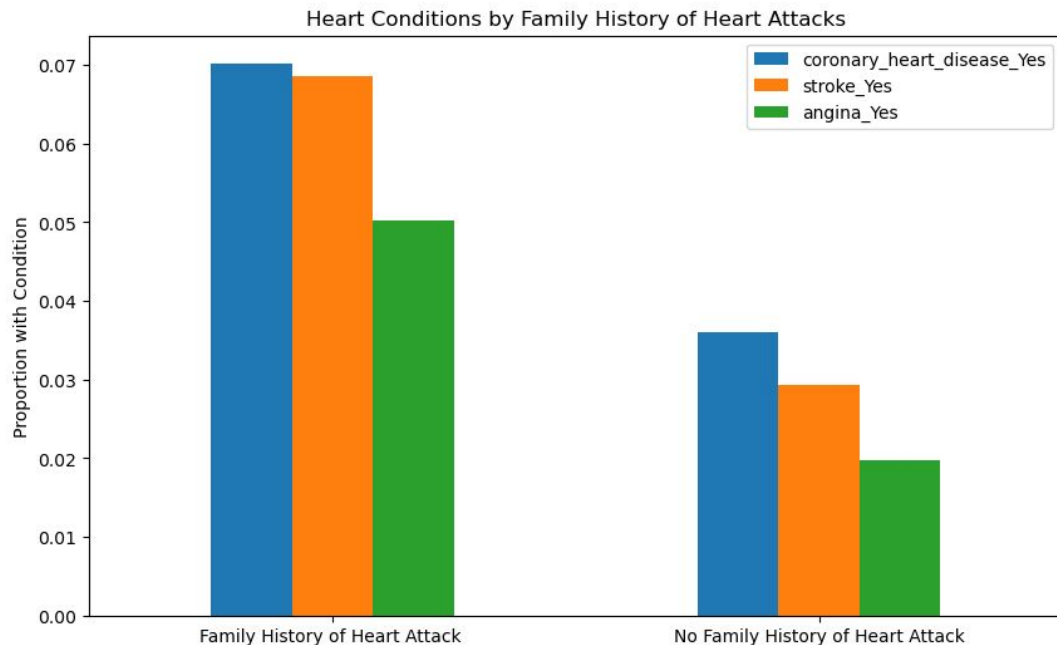
Those who have a **higher** education have a higher income above the poverty line.

CVD By Depression Status Insights



The proportion with having a stroke or being diagnosed with coronary heart disease more than **doubles** when a respondent reports that they are depressed nearly every day.

CVD Family History Insights

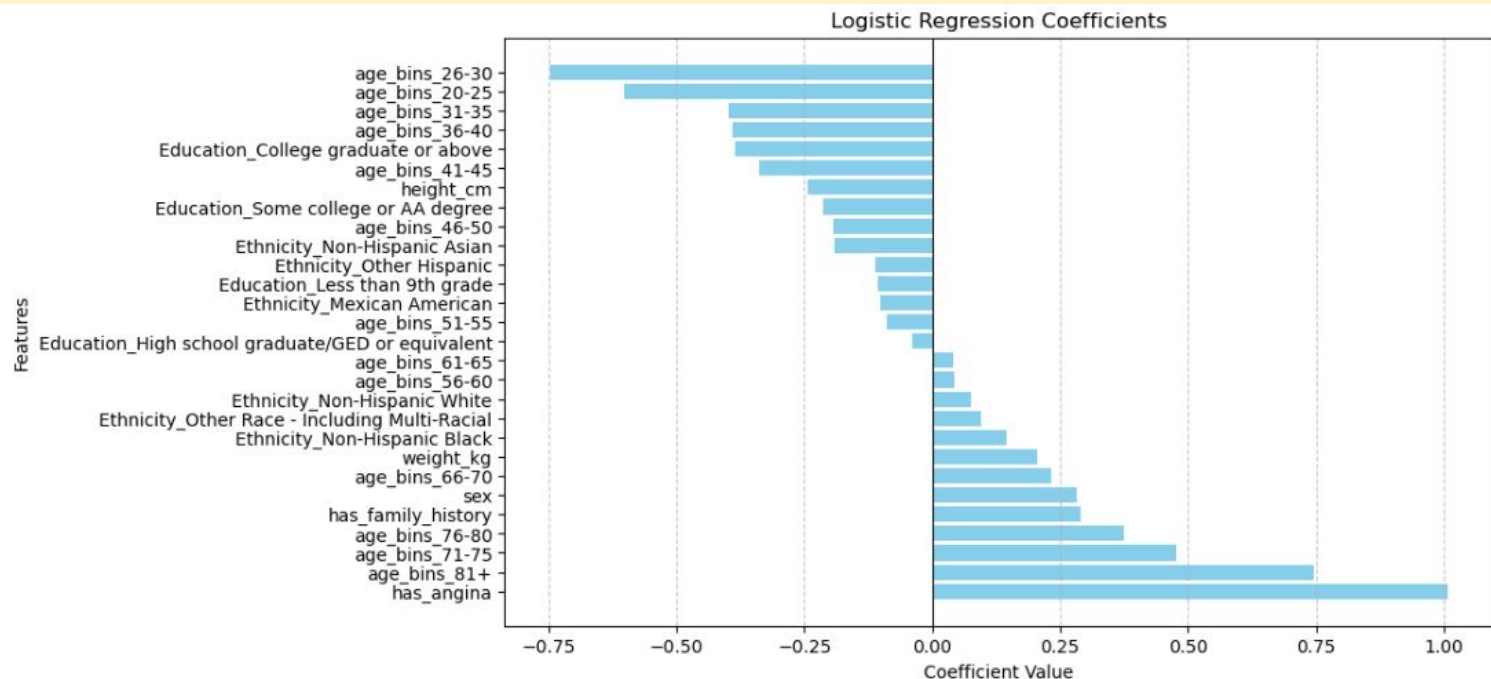


Respondents who have a family member younger than 50 who have suffered a heart attack are at an increased risk of developing cardiovascular disease, their risk **doubles** compared to those who have CVD with no family history.

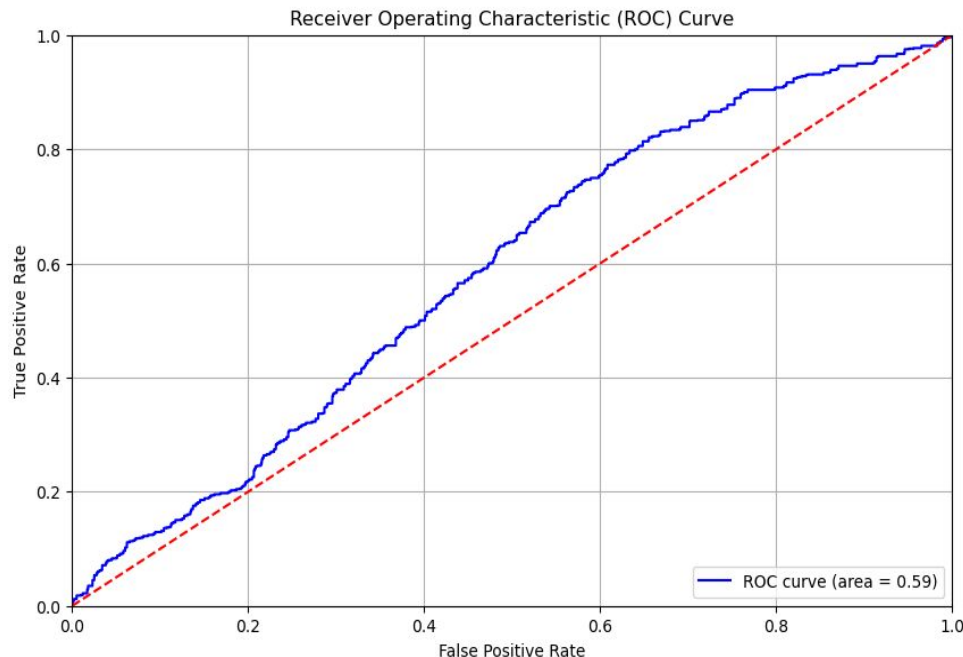
02

Models

Logistic Regression: Coefficients



Logistic Regression: Model Evaluation

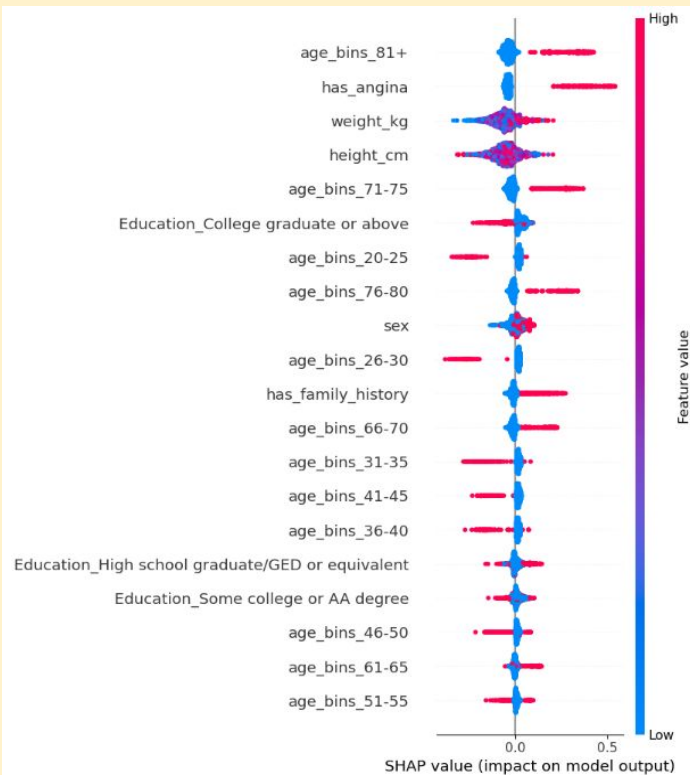


Optimal Threshold: 5.212798932325138e-12

Accuracy: 0.5826653306613226

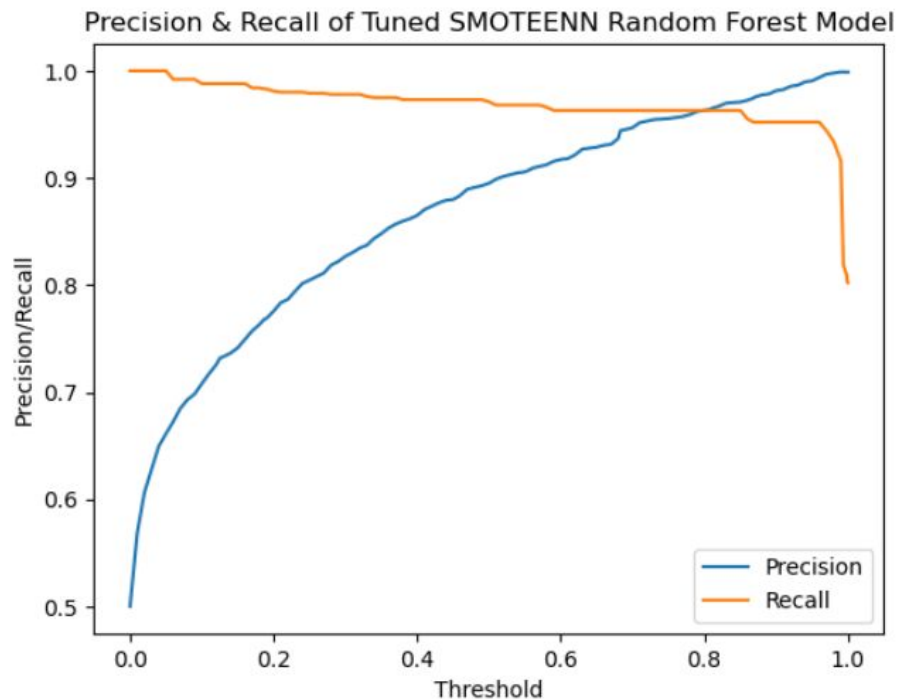
	precision	recall	f1-score	support
0.0	0.65	0.35	0.46	998
1.0	0.56	0.81	0.66	998
accuracy			0.58	1996
macro avg	0.61	0.58	0.56	1996
weighted avg	0.61	0.58	0.56	1996

Random Forest: Feature Importances



Respondent who belong in the older age bin, those who have chest pain symptoms, and those who are heavier are more likely to be at risk based on the high SHAP values of this model.

Random Forest: Model Evaluation



Confusion Matrix:

```
[[887 114]
 [ 29 972]]
```

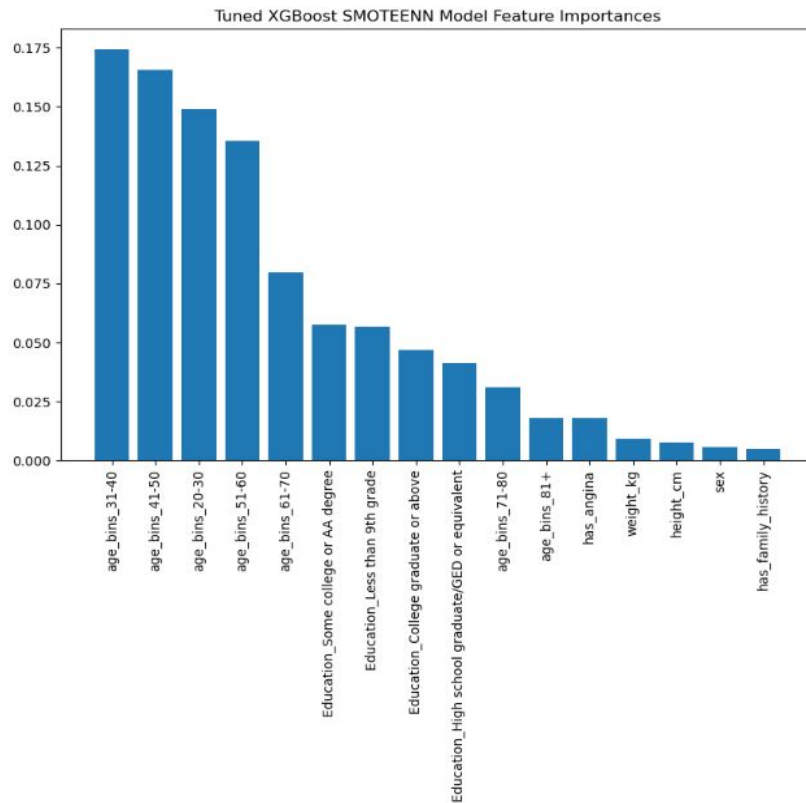
Classification Report with Threshold (0.5):

	precision	recall	f1-score	support
0	0.97	0.89	0.93	1001
1	0.90	0.97	0.93	1001
accuracy			0.93	2002
macro avg	0.93	0.93	0.93	2002
weighted avg	0.93	0.93	0.93	2002

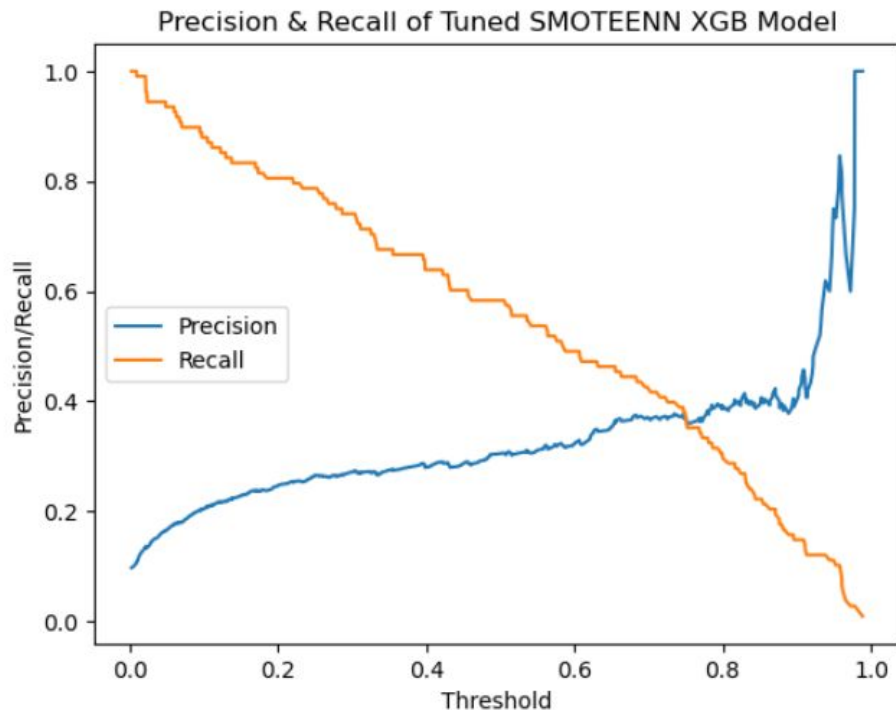
Accuracy Score:

0.9285714285714286

XGB Boost: Feature Importances



XGB Model: Model Evaluation



Classification Report with Default Threshold (0.5):

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.95	0.86	0.90	1002
---	------	------	------	------

1	0.31	0.58	0.40	108
---	------	------	------	-----

accuracy			0.83	1110
----------	--	--	------	------

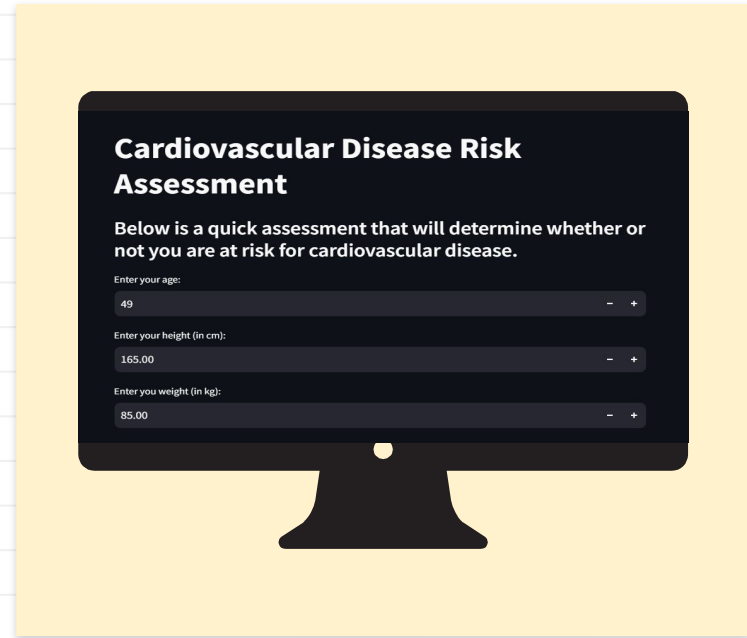
macro avg	0.63	0.72	0.65	1110
-----------	------	------	------	------

weighted avg	0.89	0.83	0.85	1110
--------------	------	------	------	------

Accuracy Score:

0.8306306306306306

Streamlit Demo!



Thanks!

CREDITS: This presentation template was created by [Slidesgo](#), and it includes icons by [Flaticon](#), infographics & images by [Freepik](#)

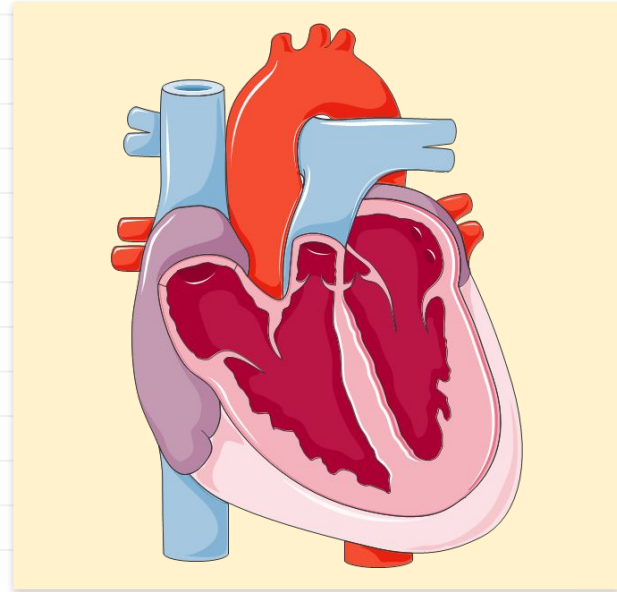


Illustration by Smart-Servier Medical Art