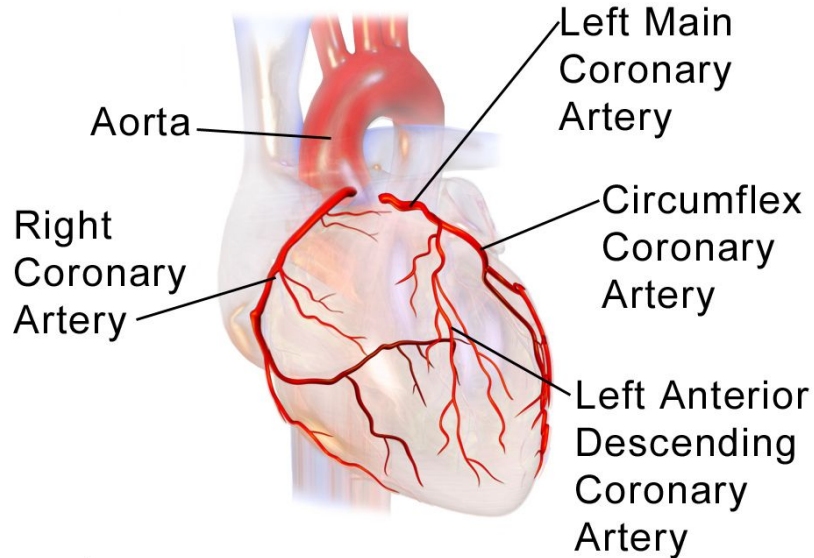


Coronary Heart Disease (CHD)

Group 4: Sisi Zhang and Harish Ram

Table of Contents

1. Background
2. Data Source
3. Implemented features
4. Logical Architecture
5. Data flow
6. Result
7. AWS Demo



Coronary Arteries

Background

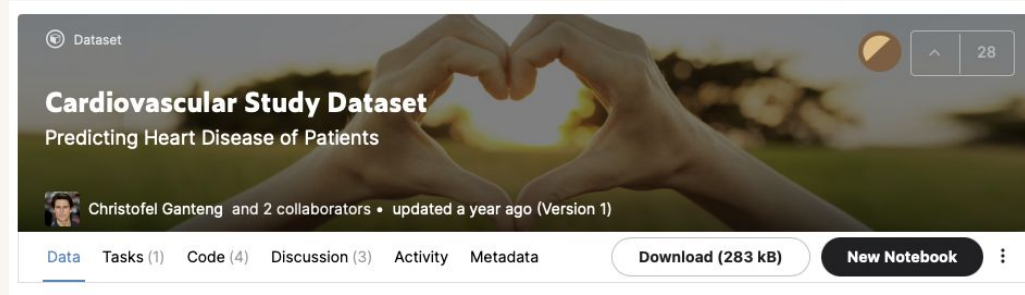
12 million deaths occur worldwide
every year due to heart diseases

659,000 people in the United
States die from heart disease each year

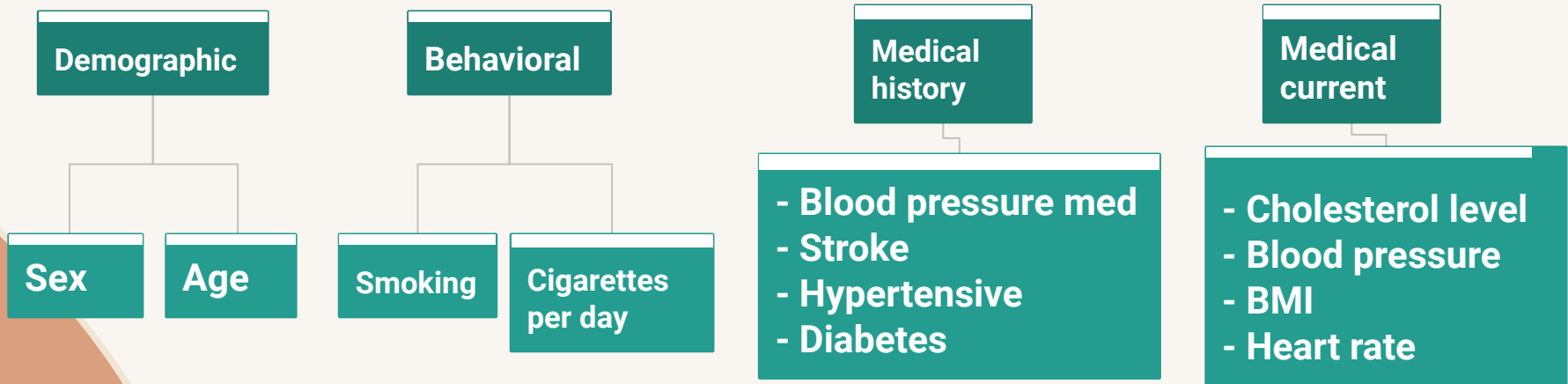
One person dies every **36 seconds** in
the United States from cardiovascular
disease

1 in every 4 death

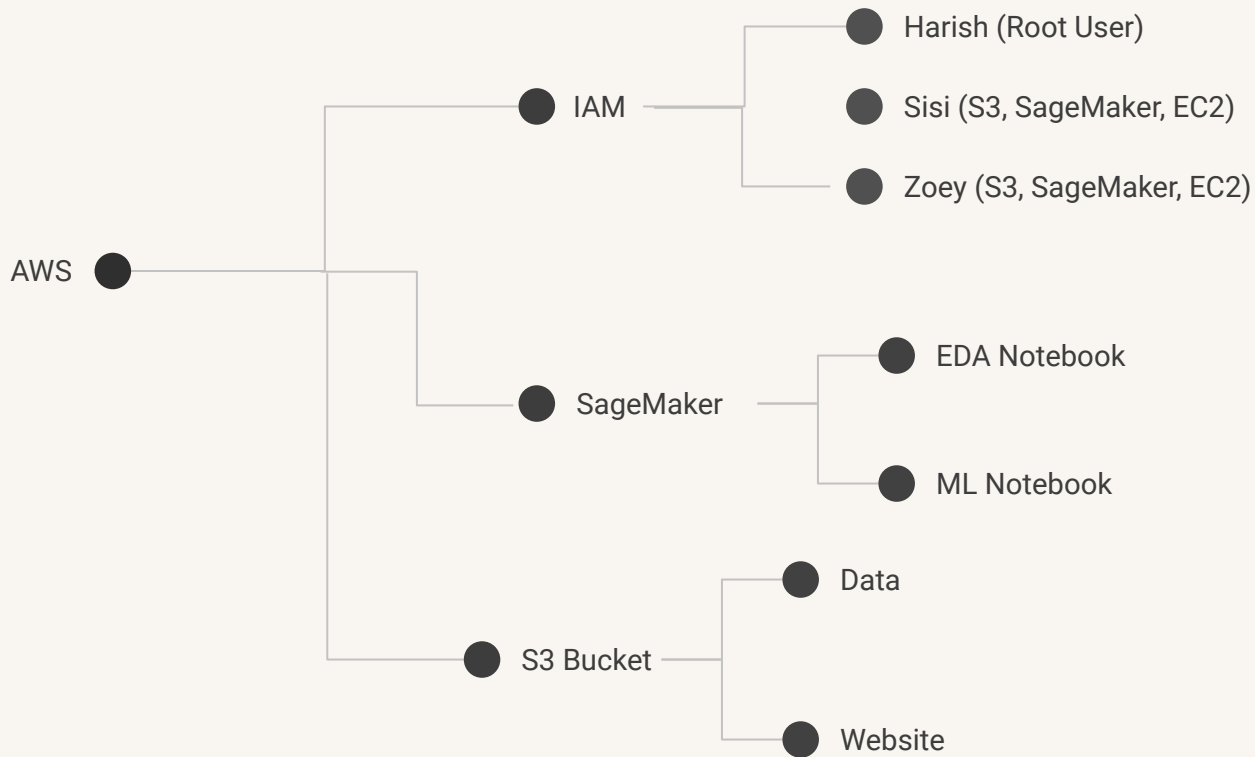
Data sources



- From an ongoing cardiovascular study
- The dataset provides patient's information and each attribute is a potential risk factor
- 10 year risk of coronary heart disease CHD (binary: "1", means "Yes", "0" means "No")



Implemented features





AWS Cloud



AWS Identity and Access Management (IAM)



Raw Data -
Stored
locally



Google CoLab –
Preprocessing and
Modeling Training



Amazon S3 –
Store Data/ and Host Website



Data Flow

Kaggle

- Download Data from Kaggle

Google Colab

- EDA Analysis
- Data Cleaning
- Model Building

S3 Bucket

- Store Data
- Store Notebook

SageMaker

- Fine-tuning
- Generate Results

S3 Bucket

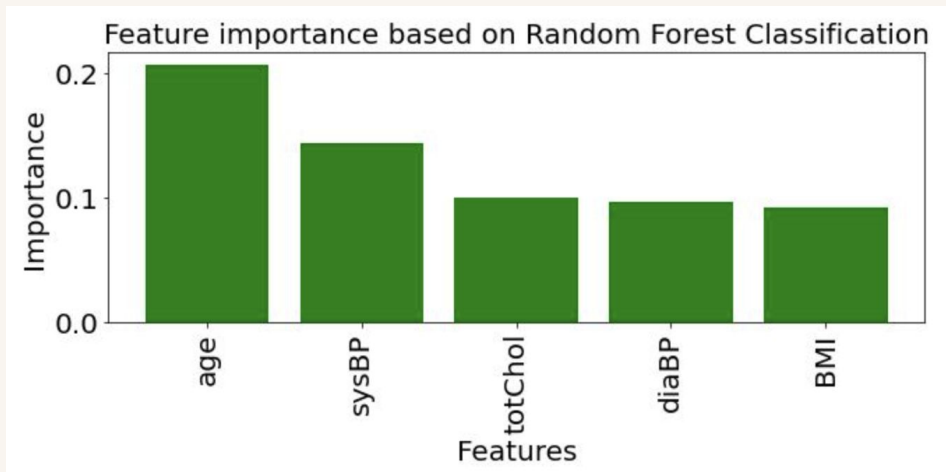
- Store graphs
- Store model result
- Host website

Demo

Result

- Improve the model accuracy score from 63.37% to 81.98% by:
 - Upsampling to balanced our dataset
 - Applying class weights to our model's f1 score

- Important features:



- Visualized the data to get more insight
 - <http://www.123456789.gwu.edu.s3-website-us-east-1.amazonaws.com>