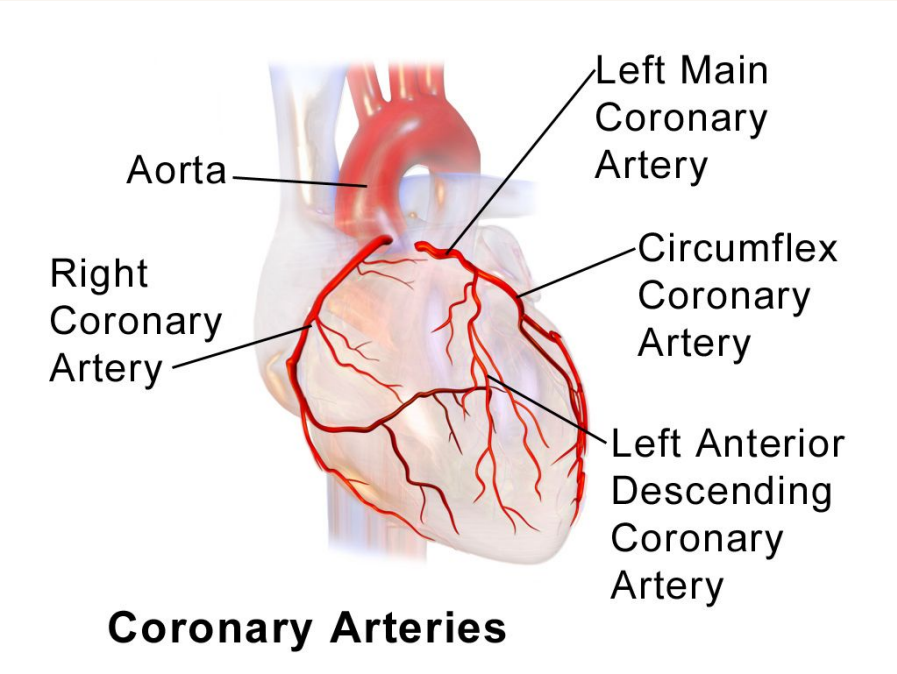
Coronary Heart Disease (CHD)

Group 4: Sisi Zhang, Harish Ram, Zoey Porter

Table of Contents



Background

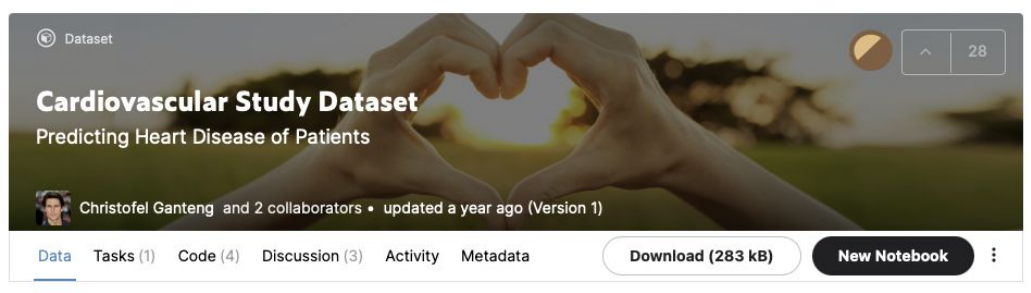
12 million deaths occur worldwide every year due to heart diseases

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")

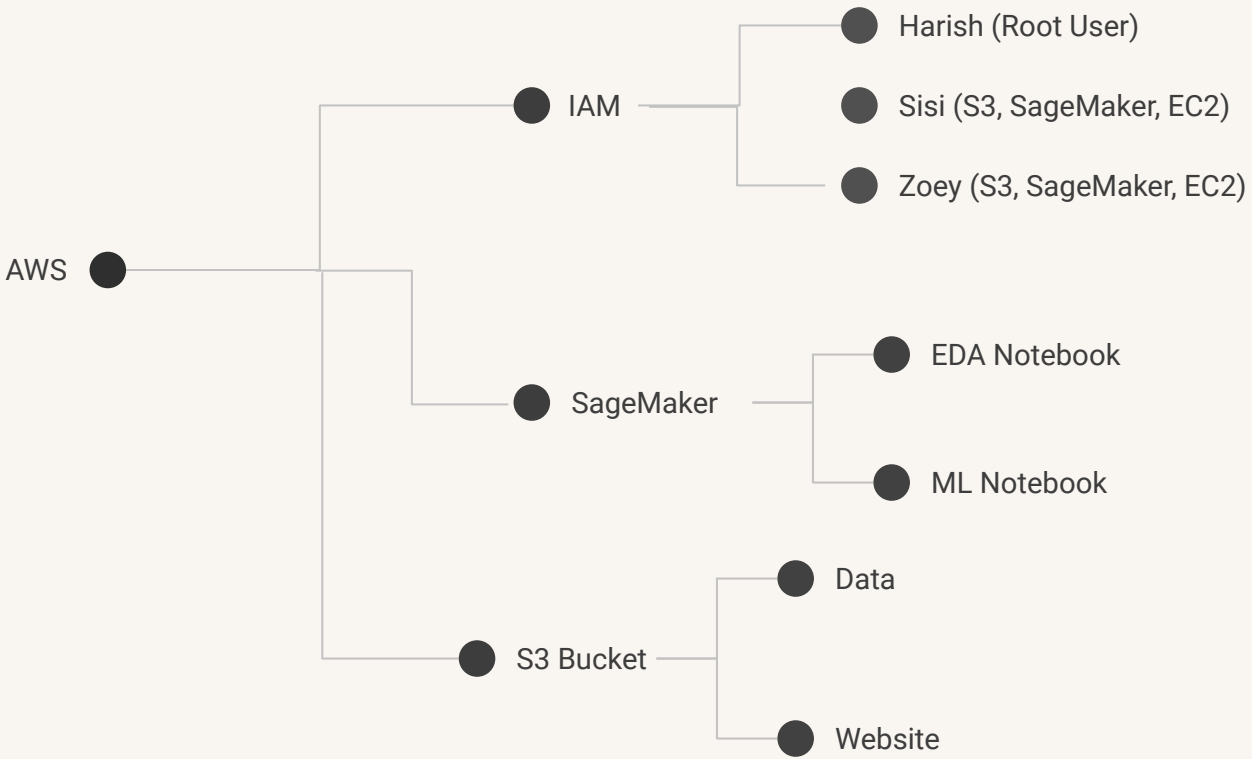
Demographic
- Sex
- Age

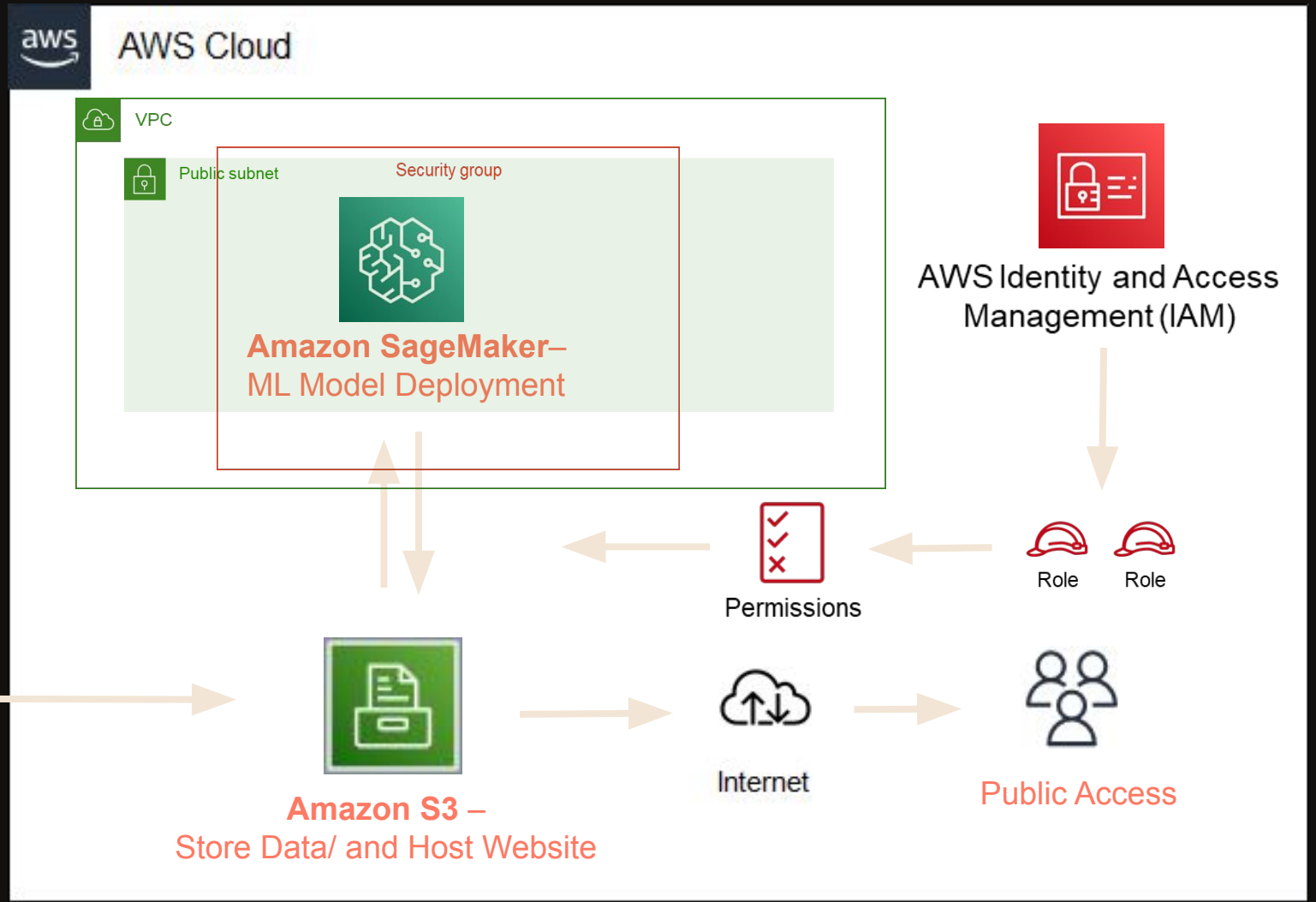
Behavioral
- Smoking
- Cigarettes per day

Medical history
- Blood pressure med
- Stroke
- Hypertensive
- Diabetes

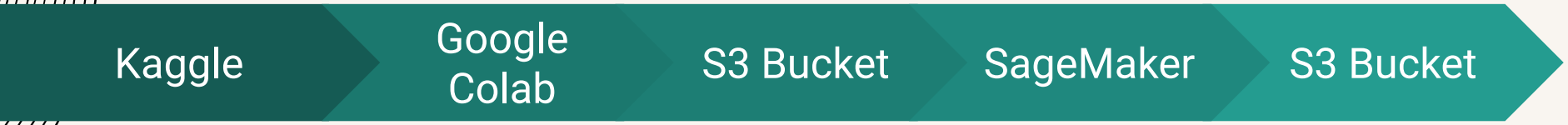
Medical current
- Cholesterol level
- Blood pressure
- BMI
- Heart rate

Implemented features




AWS Cloud
VPC
Public subnet
Security group
Amazon SageMaker–
ML Model Deployment
AWS Identity and Access Management (IAM)
Raw Data - Stored locally
Permissions
Role
Role
Google CoLab – Preprocessing and Modeling Training
Amazon S3 –
Store Data/ and Host Website
Internet
Public Access

Data Flow



- Download Data from Kaggle

- EDA Analysis
- Data Cleaning
- Model Building

- Store Data
- Store Notebook

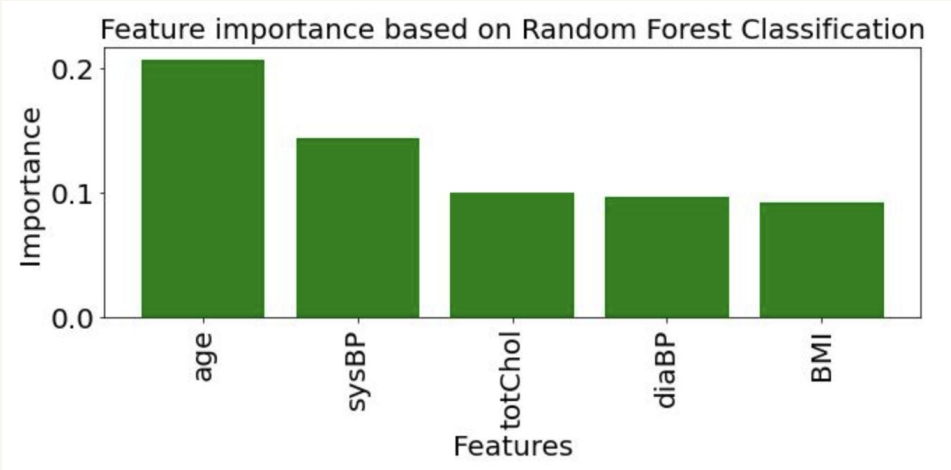
- Fine-tuning
- Generate Results

- 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