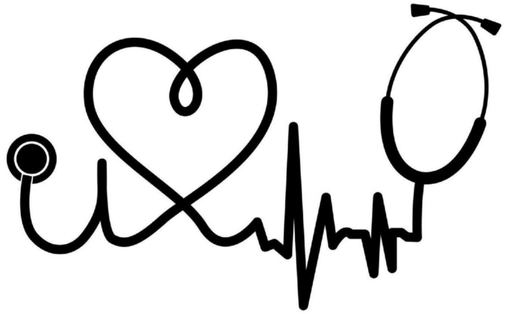
Heart Disease Patients Classification

with a Web App

METIS Data Science and Machine Learning Bootcamp

by Krystian Krystkowiak, 2022

Introduction



- **Early heart disease detection**
- Desired by medical institutions, insurances, medical apps, fitness/nutritionists, **health-conscious individuals**

- GOAL: **Raise awareness and flag potential issues** through **initial questionnaire** using **web app**

Methodology

METIS®

- Data from the **Behavioral Risk Factor Surveillance System**, an annual survey that gathers information on the health status of U.S. residents
- Data from **319,000 individuals**, including **19 factors** (medical and lifestyle)
- **HeartDisease (target)**, BMI, Smoking, AlcoholDrinking, Stroke, PhysicalHealth, MentalHealth, DiffWalking, Sex, AgeCategory, Race, Diabetic, PhysicalActivity, GenHealth, SleepTime, Asthma, KidneyDisease, SkinCancer

Pipeline

DATA INGESTION

- Behavioural Risk Factor Surveillance System questionnaire data
- download from Kaggle
- 300k+ rows of data

kaggle

TESTING/ROBUSTNESS

- Manual, annual data update.

DATA STORAGE

- Store data as a relational database with SQLAlchemy and SQL commands

SQL

PROCESSING

- Clean data using Python, NumPy and pandas
- Aggregate data with pandas
- Build predicting model

DEPLOYMENT

- Visualize the data and create a web application
- Deploy application to the web with Github and Streamlit

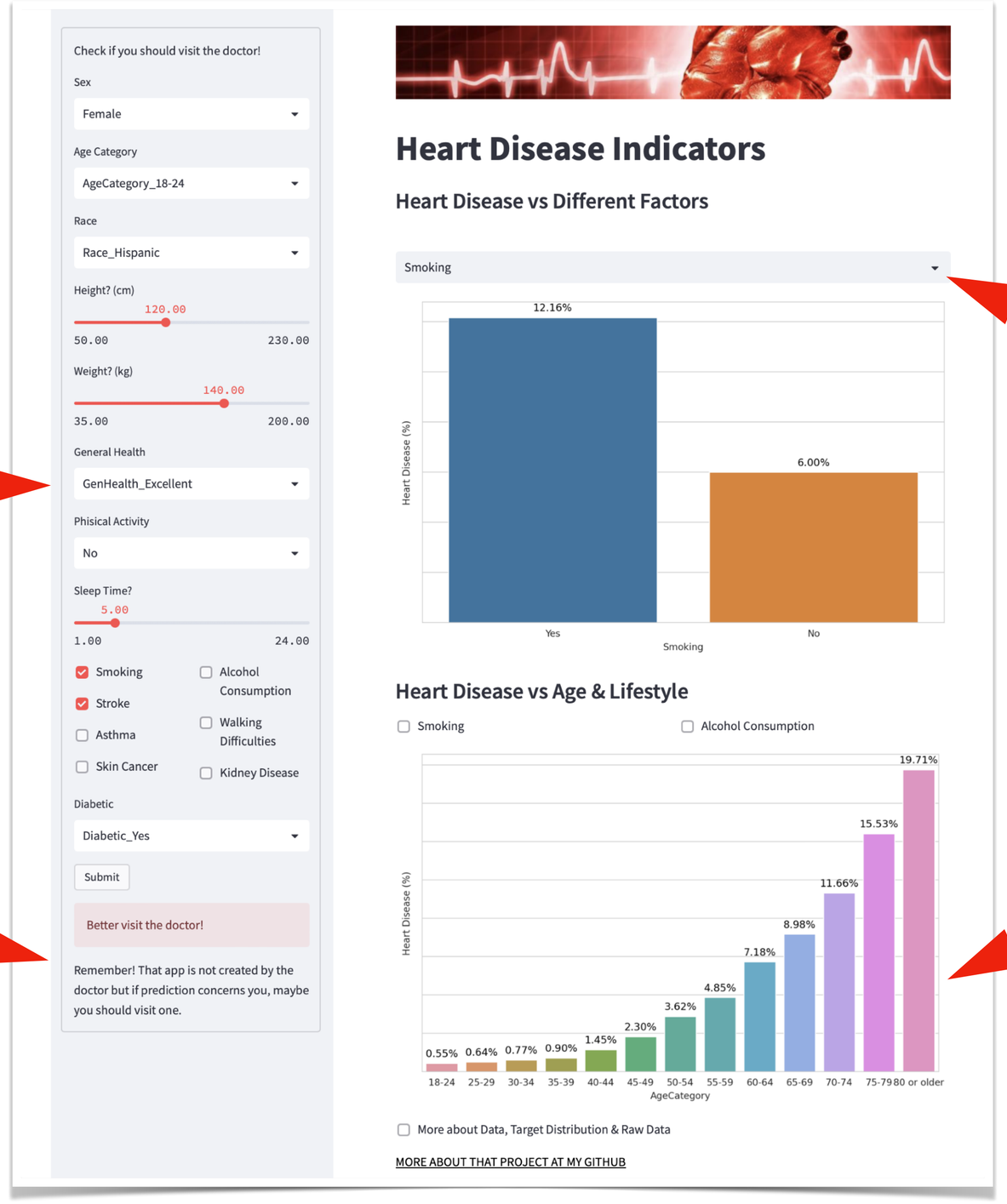
GitHub

Streamlit

Web App

pre-trained model
Naive Bayes
(Recall = 0.90)
(Accuracy = 0.81)

preprocessing for
compatibility
between
form and model data

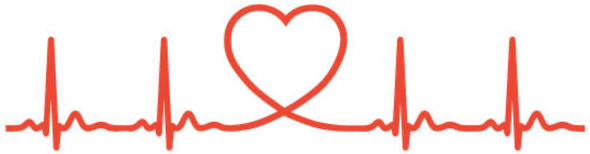


user-friendly
streamlit's
components

adaptable
seaborn
visualizations

METIS®

Future Work



METIS®

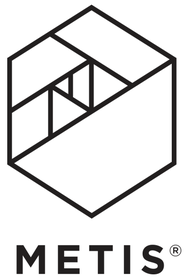
- **Automated data ingestion** process from BRFSS using **cron jobs**
- **AWS Cloud SQL** for improved data storage and management
- **Deep learning models** for improved performance
- **Enhance** the experience and accessibility (**Flask**?)

METIS®

Thank you!

Questions?

Appendix



- Original dataset: https://www.cdc.gov/brfss/
- Dataset: www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease (initially cleaned by Kamil Pytlak at Kaggle)
- Web App: https://krystkowiakk-heart-disease-pa-streamlit-appstreamlit-app-cjy0xv.streamlit.app/