

HealthShield

Your Smart Doctor

Chou Chia Hsuan
Shih Hsin Yi



Table of contents

01

Introduction

Motivation and goal of
our work.

02

Data

Dataset and data
preprocessing.

03

Methods

Measured criteria and
fitted models.

04

Results

Performance of each
model.

05

Demo

Introduce cloud
architecture and
demonstrate our app
page.



01

Introduction

Motivation and goal of our work

Background of Diabetes

Top 5 death cause

in Taiwan (2024)

Increase 30%

adults aged 20-40 (2015-2020)



Motivation

01

Delayed Diagnosis

Diabetes is often diagnosed at a late stage.

02

Early Risk Prediction

Early risk prediction is needed to support prevention.

03

Cloud-based ML Solution

Develop a cloud-based machine learning system for diabetes risk prediction that can be accessed by more users.

Problem Definition

01

Input Data

Basic information, body measurements, medical history, lifestyle habits, blood pressure, and blood test.

02

Method

Machine learning models with SHAP-based interpretability.

03

Output

Personalized diabetes risk prediction and explanations. (HealthShield)

Novelty

Adds more feature categories

Combines lifestyle habits, blood pressure, and blood test data.

Allows “I don’t know” for selected variables

SHAP visual explanations

Provides individual-level and global feature importance visualizations.

End-to-end cloud-based system

A modular cloud architecture separates frontend, backend, and model inference services.



02 Data

Dataset and data preprocessing

Data Source

NHANES

Introduction

A nationally representative health survey by the U.S. NCHS.

Years

2007–2018

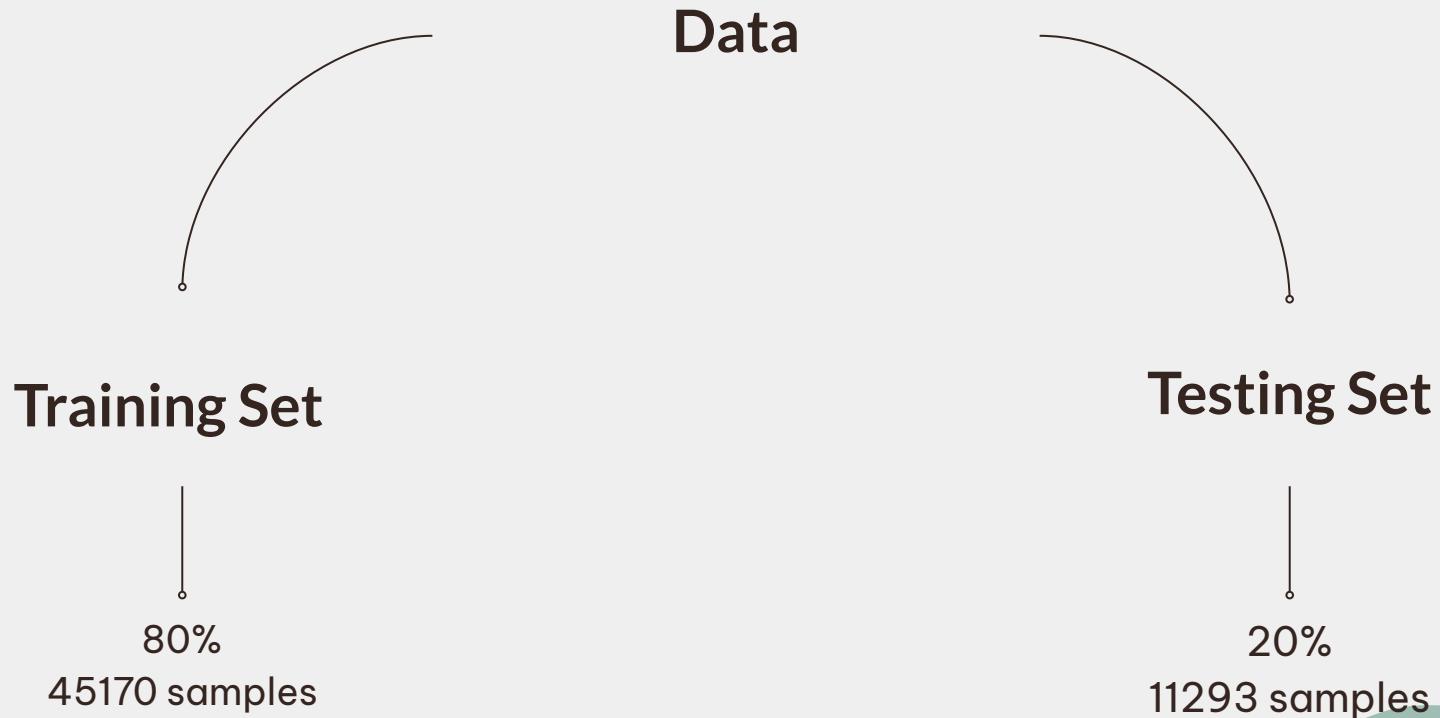
Samples

56,463 samples
22 features

Variables

Feature Categories	Variables Names
Target	Diabetes
Basic Information	Age, Gender
Body Measurements	Height, Weight, BMI, Waist
Family History & Lifestyle	family_diabetes, ever_smoked, alcohol_drinks, Sleep_Hours vigorous_activity, moderate_activity, general_health
Blood Pressure	systolic_1/2/3, diastolic_1/2/3
Blood Test Results	fasting_glucose, insulin, HbA1c, total_cholesterol, HDL, LDL, triglycerides

Data Splitting



Data Preprocessing - Missing Values

Variables	Method
Target, Age	Remove missing target or age = 0
height, weight, BMI	Use bmi formula if one is missing; otherwise median imputation
systolic_1/2/3, diastolic_1/2/3	Average into systolic_avg/diastolic_avg; otherwise median imputation.
waist, fasting glucose, insulin, HbA1c, cholesterol (total, HDL, LDL), triglycerides, general_health, sleep_hours, alcohol_drinks	Median imputation (alcohol_drinks age < 20, impute 0)
ever_smoked, vigorous_activity, moderate_activity, family_diabetes	Add “Missing” category. (ever_smoked age < 20, impute “No”)

Data Preprocessing -Scaling and Encoding

Min-Max normalization

age, height, weight, bmi, waist,
fasting_glucose, insulin, HbA1c,
total_cholesterol, HDL, LDL,
triglycerides, alcohol_drinks,
general_health, systolic_avg,
diastolic_avg, Sleep_Hours

One-hot encoding

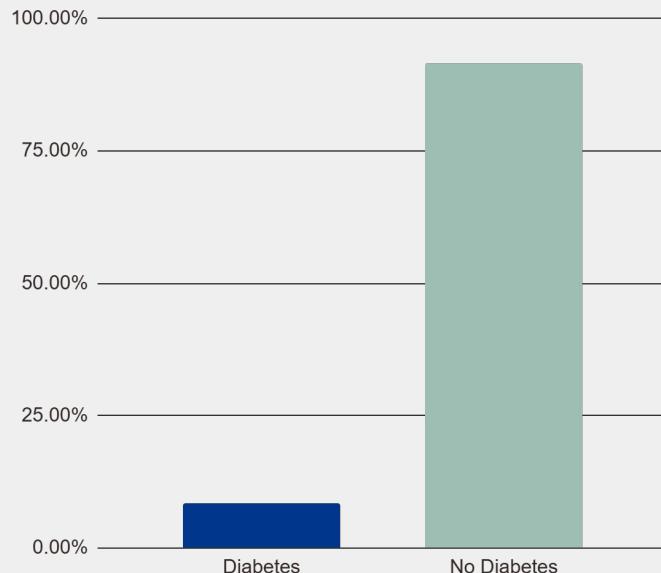
gender, ever_smoked, vigorous_activity,
moderate_activity, family_diabetes



03 Methods

Measured criteria and fitted models

Selection of Criteria



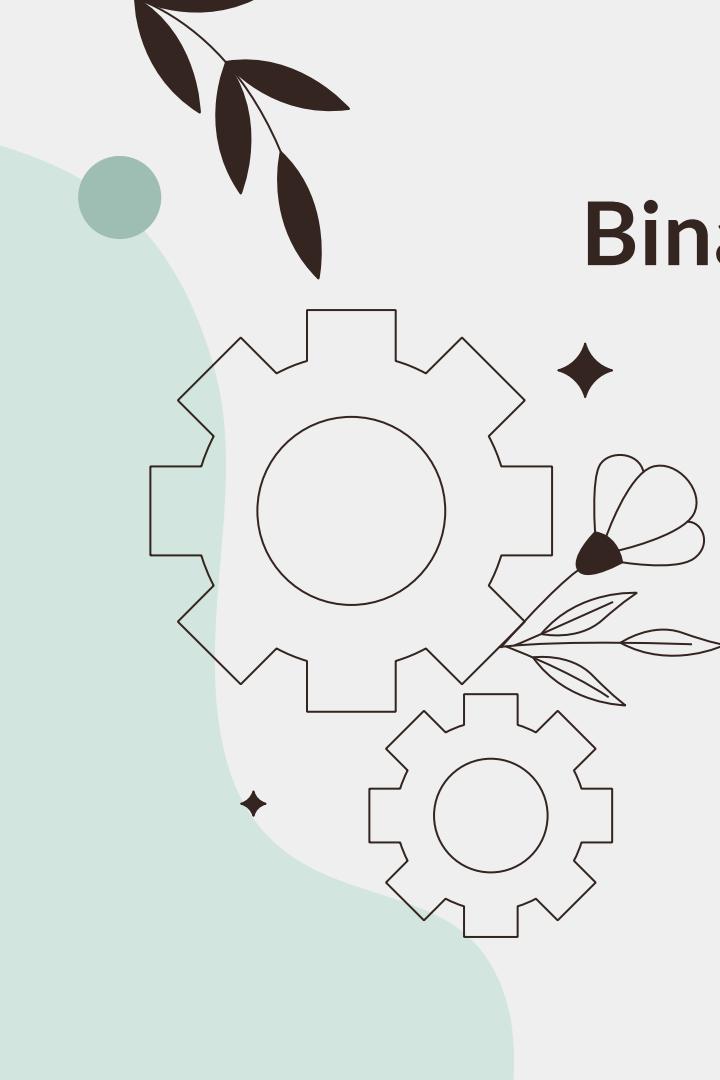
Diabetes
Diagnose to have diabetes

8.34%

No Diabetes
Not diagnose to have diabetes

91.66%

Imbalanced!
Adopt **Balanced Accuracy**



Fitted Models - Binary Classification Problem

Logistic Regression

Basic baseline that provides interpretable risk probabilities

SVM

Find the optimal boundary between healthy and at-risk patients

Random Forest

Ensemble of decision trees s.t. being more stable on imbalanced data

XGBoost

Iteratively corrects errors and offers built-in paras to maximize sensitivity for rare cases



04

Results

Performance of each model

Performance

Model	CV-Training				CV-Validation				Testing			
	BA	Sens	Spec	Acc	BA	Sens	Spec	Acc	BA	Sens	Spec	Acc
LR-L2	0.8894	0.8877	0.8910	0.8907	0.8881	0.8854	0.8909	0.8904	0.8954	0.9002	0.8906	0.8914
SVM	0.9000	0.9041	0.8959	0.8966	0.8915	0.8883	0.8948	0.8942	0.8945	0.8938	0.8952	0.8951
RF	0.9082	0.9183	0.8981	0.8998	0.8958	0.8949	0.8966	0.8965	0.8939	0.8970	0.8907	0.8913
XGB	0.9140	0.9203	0.9077	0.9087	0.9015	0.8970	0.9060	0.9052	0.9028	0.9023	0.9033	0.9032

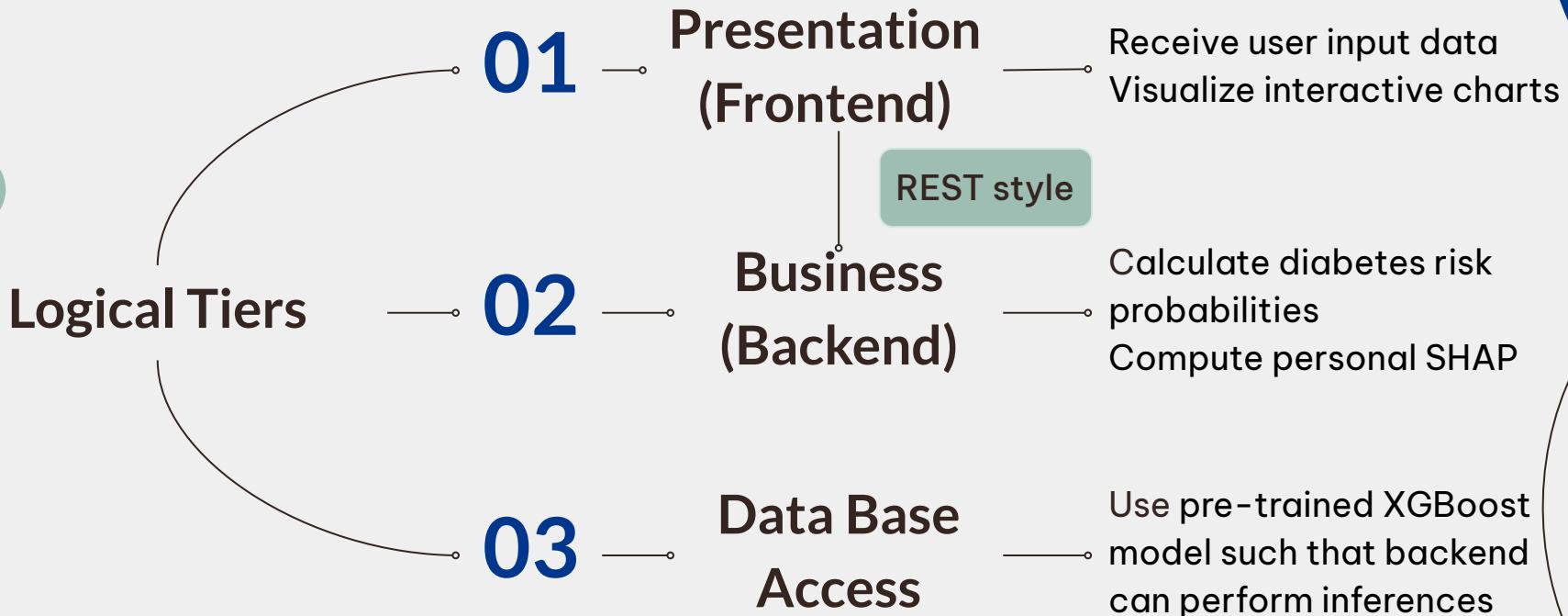


05 Demo

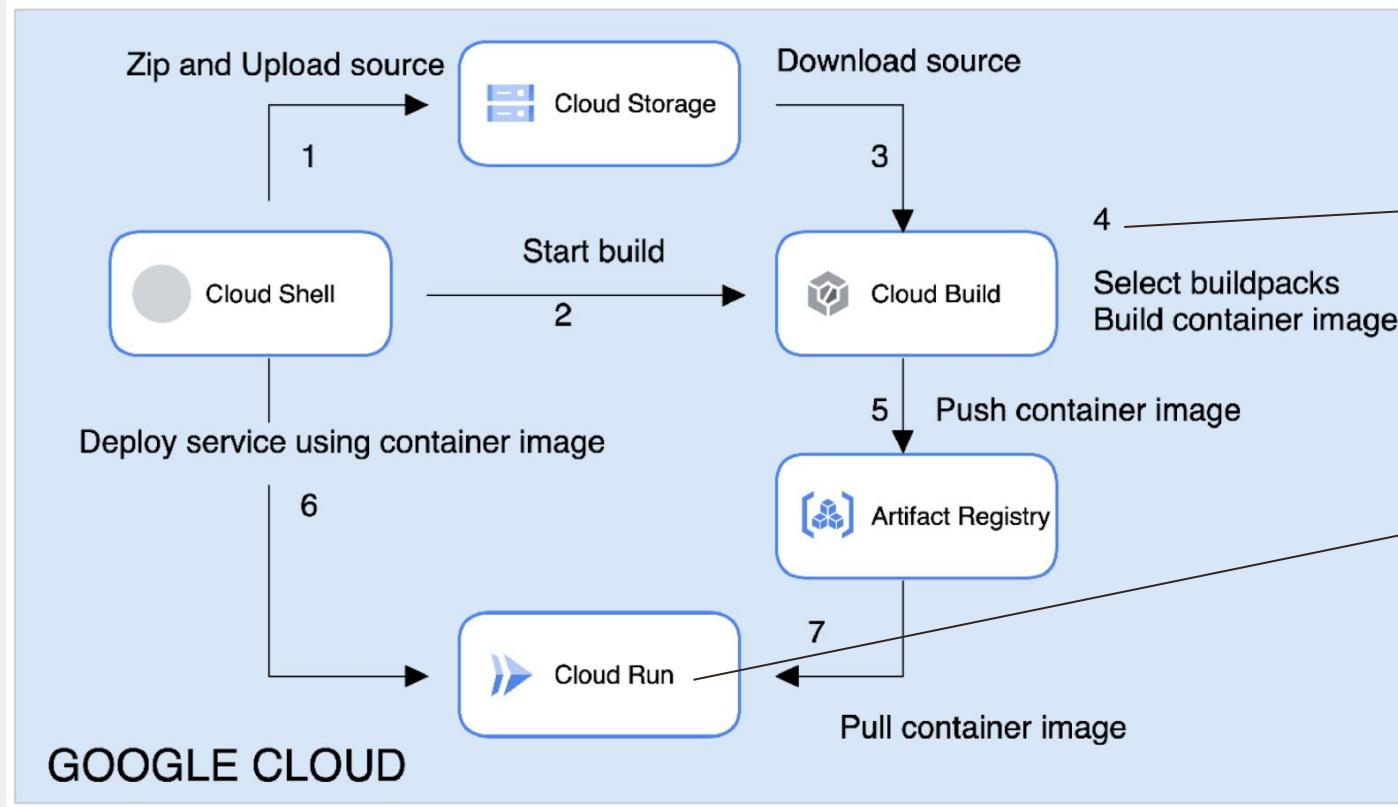
Introduce cloud architecture and demo web

HealthShield

3 Logical Tiers of Our Work



CI/CD Pipeline of Google Cloud Run



OS-level virtualization:
Dockerfile to define an isolated execution environment

PaaS:
focus on develop application codes

User Interface



User Data Input Page

Users can enter their personal health information required for diabetes risk prediction



Prediction Page

Present the diabetes risk prediction generated by the machine learning model.

User Data Input Page

Welcome to HealthShield

Know Your Diabetes Risk, Take Control of Your Health

Basic Information / 基本資料

Age: Gender / 性別: Choose an option

Body Measurements / 身體測量

Height / 身高 (cm): Weight / 體重 (kg): BMI / 身體質量指數:
 I don't know I don't know

Waist Circumference / 腰圍 (cm):
 I don't know

Blood Test Results / 血液檢查結果

Fasting Glucose / 空腹血糖 (mg/dL): Insulin / 胰島素 (μ U/mL): HbA1c / 糖化血色素 (%):
 I don't know I don't know I don't know

Total Cholesterol / 總膽固醇 (mg/dL): HDL Cholesterol / HDL 良固醇 (mg/dL): LDL Cholesterol / LDL 壞固醇 (mg/dL):
 I don't know I don't know I don't know

Triglycerides / 三酸甘油酯 (mg/dL):
 I don't know

Family History & Lifestyle / 家族病史 & 生活作息

Does a close relative have diabetes? / 您的近親是否患有糖尿病嗎？
Choose an option

Do you do moderate-intensity sports or fitness activities (e.g., brisk walking, swimming) weekly? / 您每週有從事中等強度運動或健身活動嗎(例如快走、游泳)？
Choose an option

What is your average alcoholic drinks per day? / 您平均每天飲用多少標準酒飲品？
 I don't know I don't know

How long do you sleep per night (hours)? / 您每晚睡眠時長 (小時) 是多久？
 I don't know

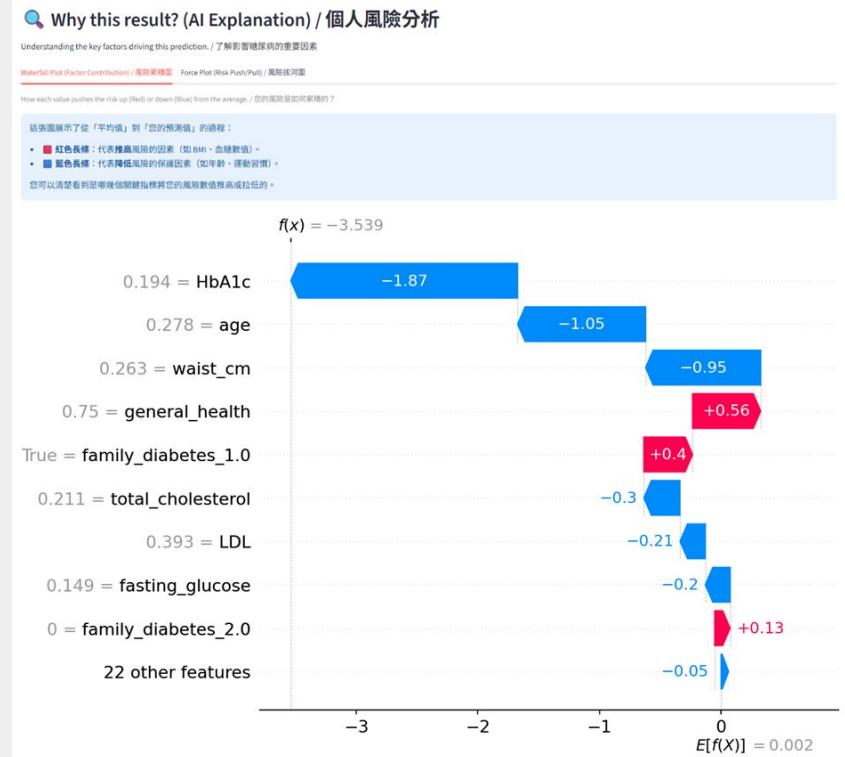
How is your self-reported health status? (1=Excellent, 5=Poor) / 您的自評健康狀況如何？(1=極佳, 5=差)
Choose an option

Blood Pressure / 血壓

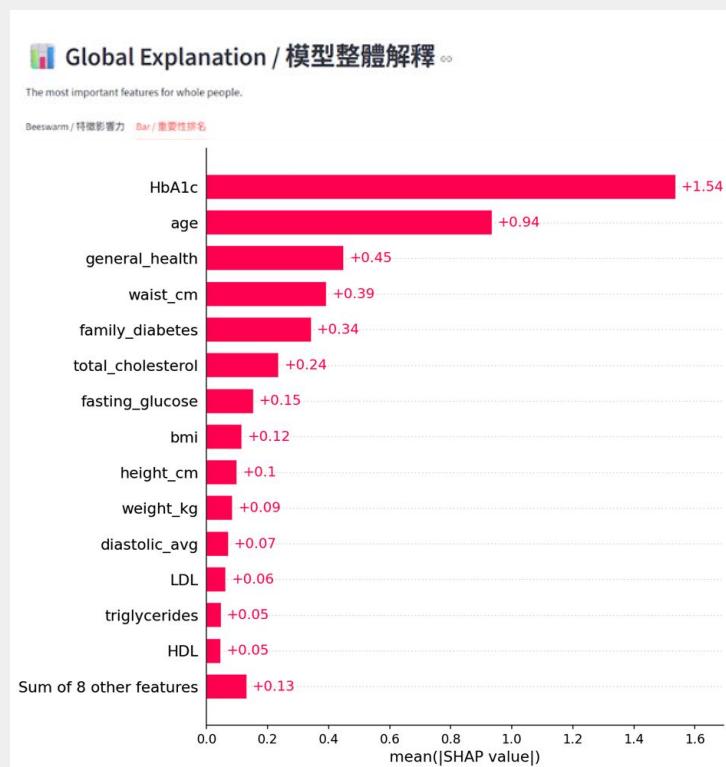
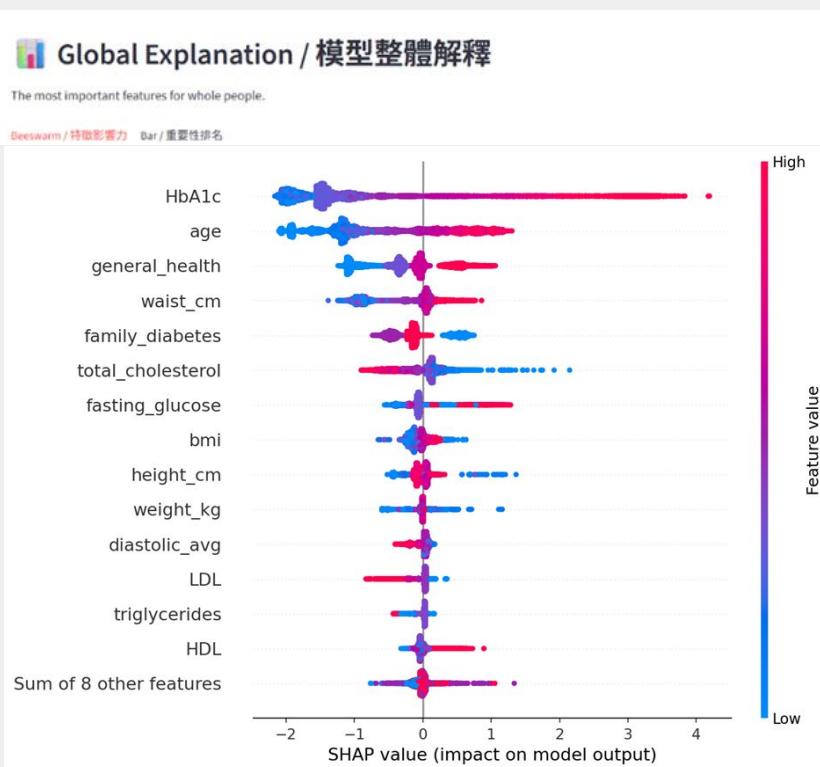
Systolic Blood Pressure / 收縮壓 (mmHg): Diastolic Blood Pressure / 舒張壓 (mmHg):
 I don't know I don't know

Get My Prediction → / 獲取我的預測結果 →

Prediction Page - Personal



Prediction Page - Global



**Thank you for
your listening!**