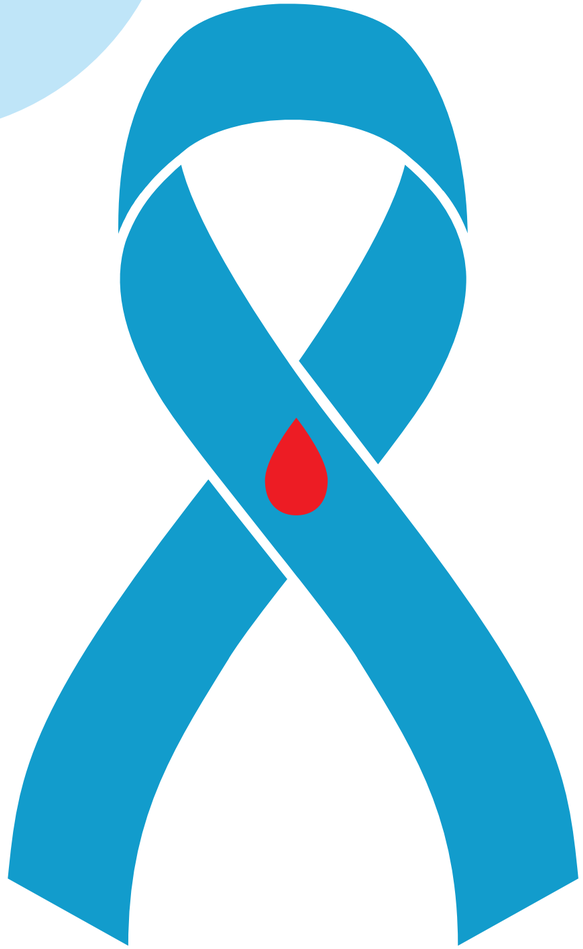
BC2406

DIABETES PREDICTION MODEL



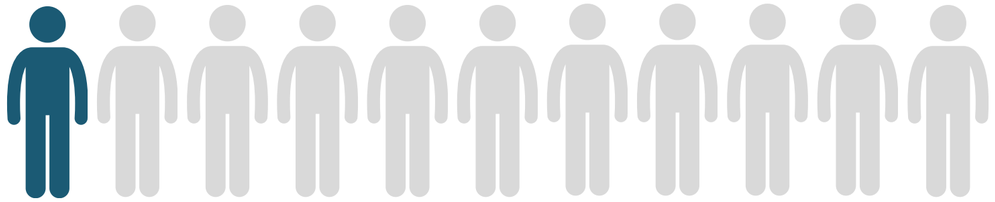
Group 7

ALOYSIOUS LAW JIA JIAN (U2420162J)
BRYAN LIM HOW MENG (U2420547L)
DE SOUZA ALYSSA ANNE (U2210429D)
NG YONG WEI (U2440621H)

ALOYSIOUS LAW JIA JIAN

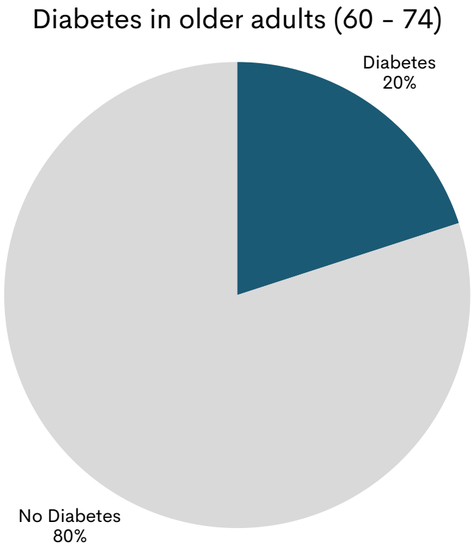
Diabetes in Singapore

SOCIAL IMPACT



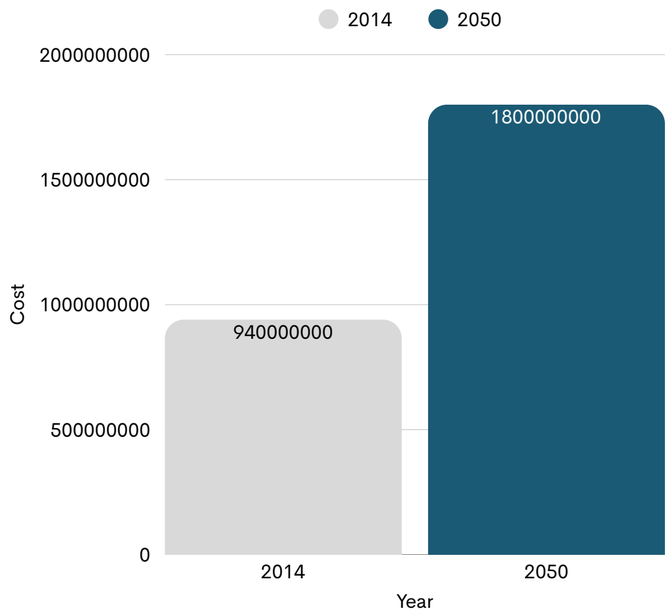
1 in 11

suffer from diabetes in Singapore



Among older Singaporeans aged 60 - 74, **20% are diagnosed with diabetes**

FINANCIAL IMPACT



The financial impact of diabetes, including medical costs and lost productivity, was projected to increase from over $940 million in 2014 to **$1.8 billion by 2050**.

ALOYSIOUS LAW JIA JIAN

Diabetes in Singapore

THE PROBLEM

Many Singaporeans are unaware of their personal risk of developing diabetes

Thus, no preventive measures are taken to lower their risk of developing diabetes

ALOYSIOUS LAW JIA JIAN

Dataset Overview

- Source: **CDC Diabetes Health Indicators**
  → 253,680 survey responses
  21 features related to chronic disease and lifestyle.
- Target variable: **Diabetes_binary**
  → "0" = Non-diabetic; "1" = Prediabetic/Diabetic

Balanced Subset Used:

70,692 responses, **50-50 class split**
→ Ensures fair model training

diabetes_binary_5050split_health_indicators_BRFSS2015.csv

Diabetes_binary	HighBP	HighChol	CholCheck	BMI	Age	Smoker	HvyAlcoholConsump	DiffWalk	GenHlth
0	1	0	1	26	4	0	0	0	3
0	1	1	1	26	12	1	0	0	3
0	0	0	1	26	13	0	0	0	1
0	1	1	1	28	11	1	0	0	3
0	0	0	1	29	8	1	0	0	2

ALOYSIOUS LAW JIA JIAN

Variable Analysis

1. Correlation analysis (heatmap)

- Top correlated features:
  General Health (r=0.294), **HighBP** (r=0.263), **BMI** (r=0.218)
- Heatmaps confirmed **no multicollinearity** (all r < 0.6).

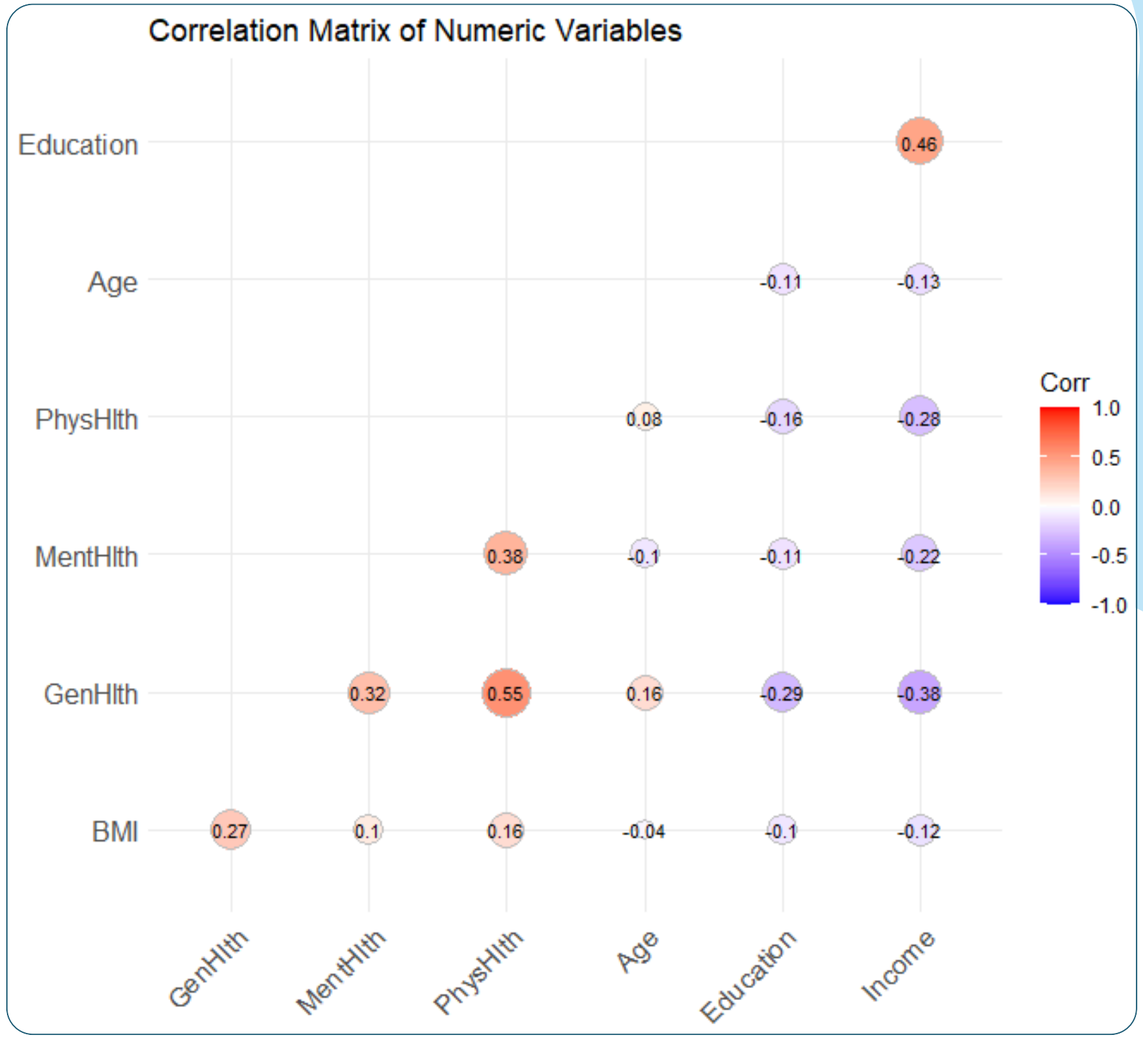
2. Statistical tests

a) Chi-square test:

- Used on catgorical variables
- Strongest categorical predictors → **HighBP, DiffWalk, and HighChol**

b) T-test:

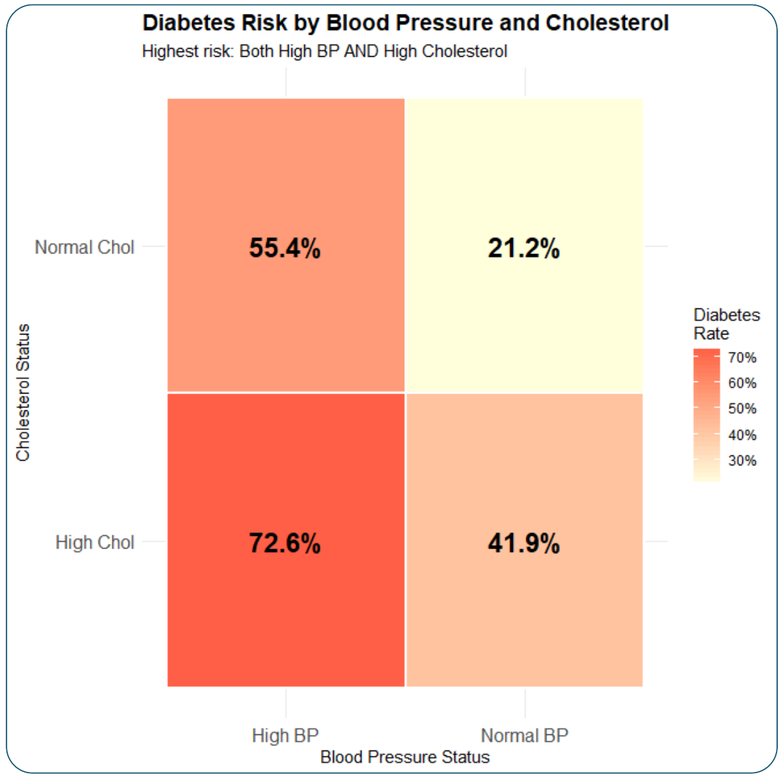
- Used on continuous variables
- Factors with **p < 0.001** = highly significant
  → 7 significant continuous predictors

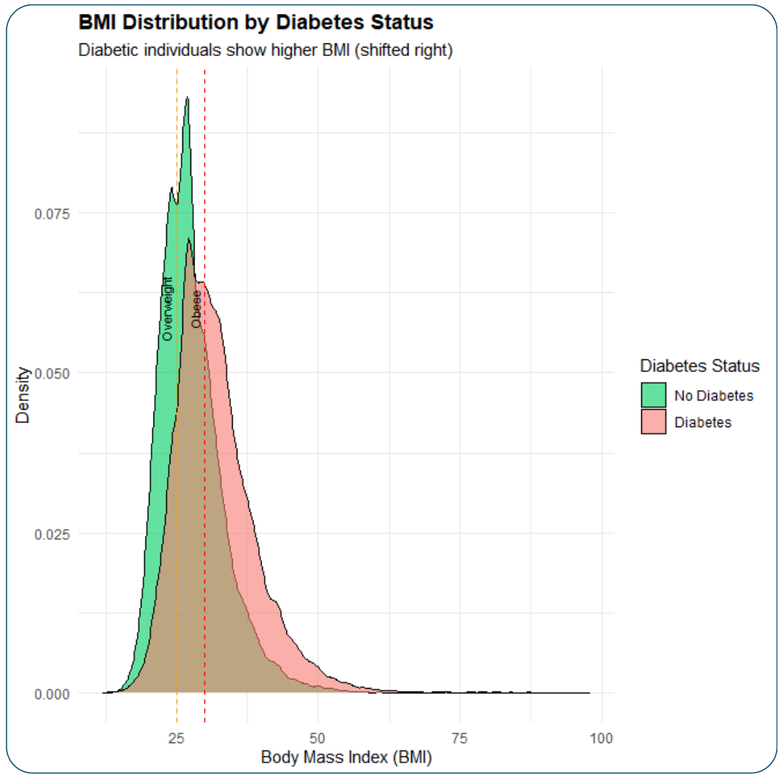


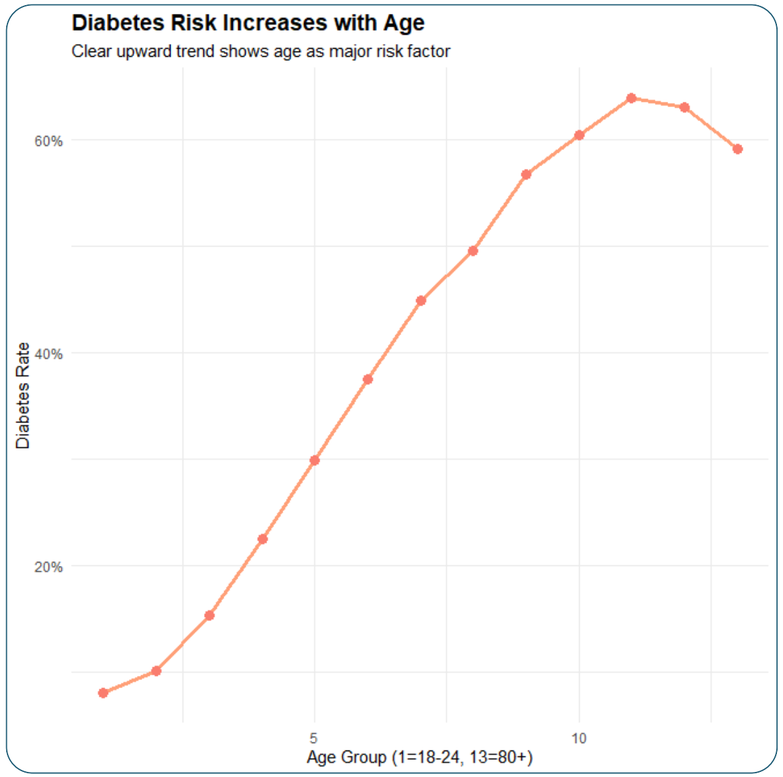
NG YONG WEI

c) Visualisations for top predictors:

- Strong diabetes risk patterns with BMI, Age, Blood Pressure, and Cholesterol







- Highest risk: **High BP + High Cholesterol**
- Normal levels → lowest risk
- Shows compounding effects of cardiovascular factors

- Diabetic group has **higher BMI** overall
- Most fall into Obese (BMI ≥ 30)
- Reinforces link between weight and diabetes risk

- Risk **increases steadily** with age
- **Slight dip after 75–80** due to survival bias & underdiagnosis

NG YONG WEI

Predictive Models

Logistic Regression Model

The problem we faced...

For every 100 prediabetic/diabetic patients:

Model with 0.5 threshold

Missing **23% of prediabetic/diabetic cases** is unacceptable in healthcare!

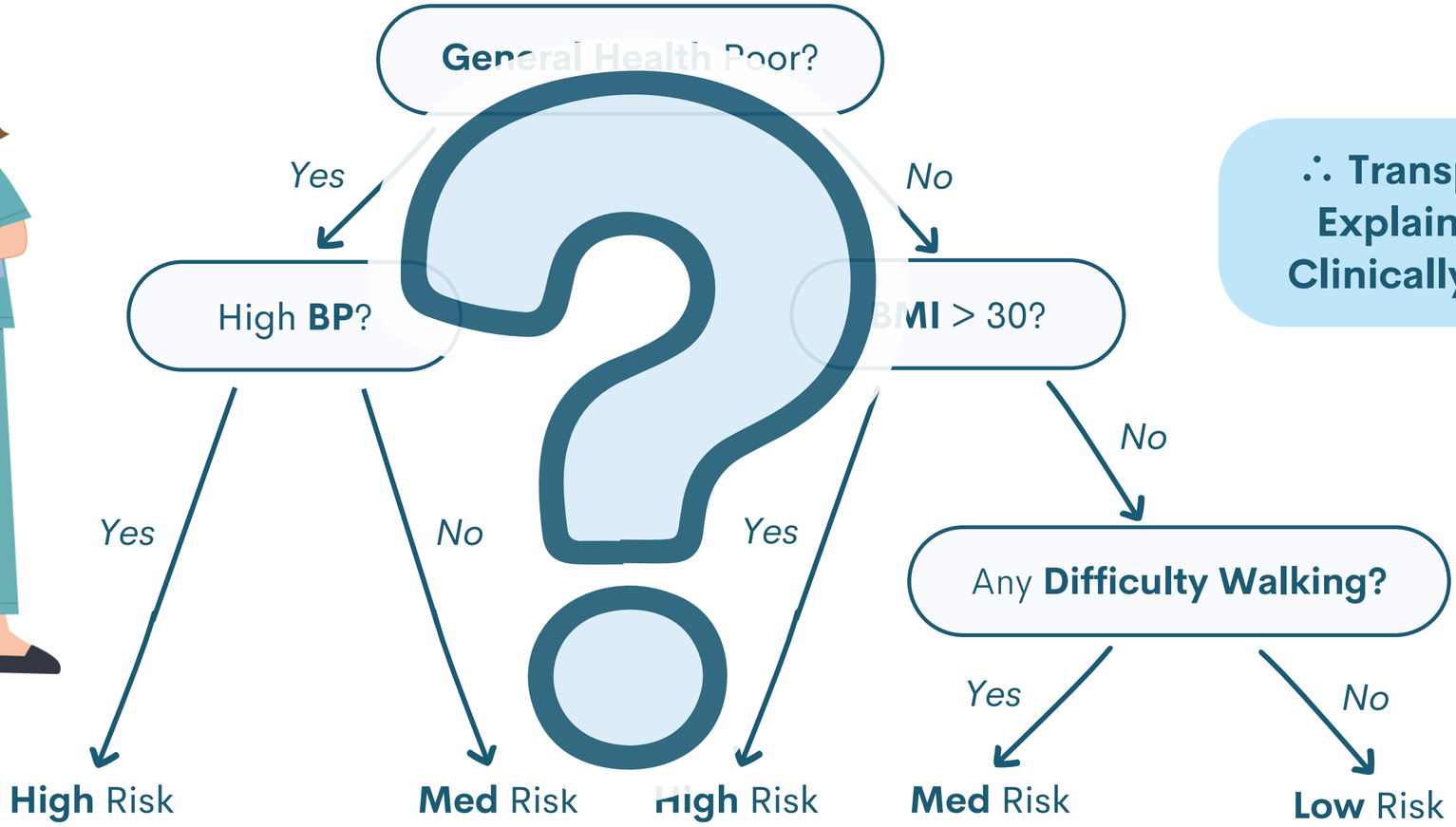
	Before (Threshold 0.5)	After (Threshold 0.35)
Missed Cases	2488	*1151* ↓
Detection Rate	76.5%	*89.2%* ↑

∴ **12.7% increase** in succesful detection
46.3% reduction in missed cases

NG YONG WEI

CART Model

→ Automated Clinical Reasoning



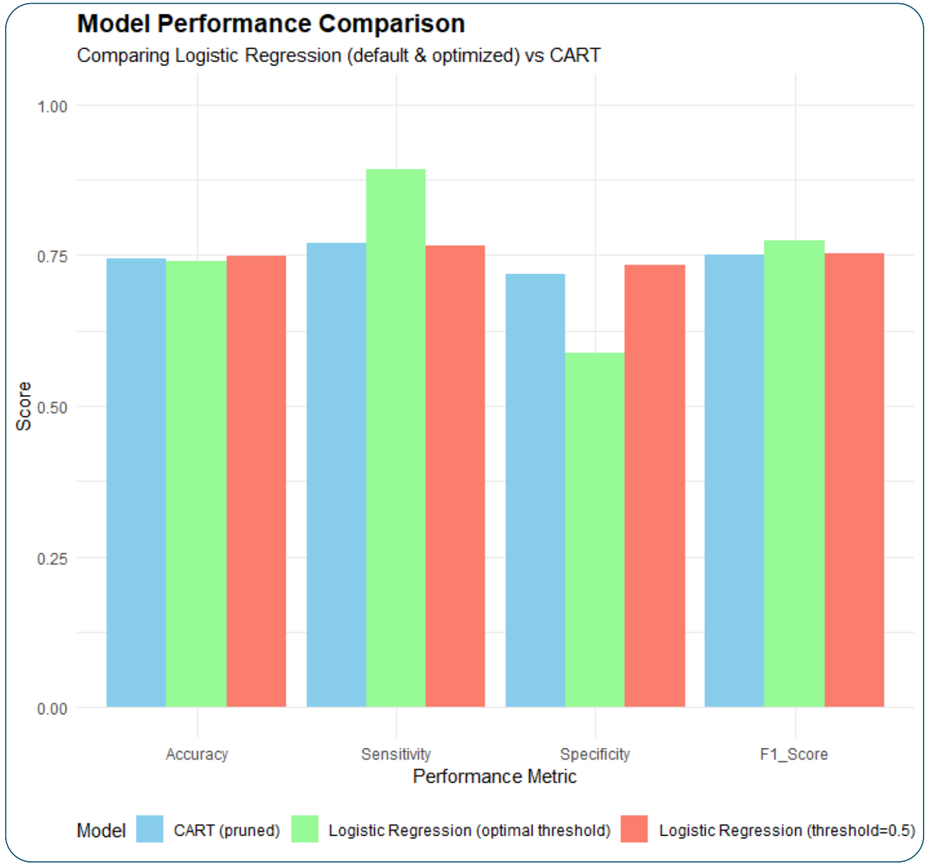
∴ Transparent, Explainable & Clinically useful!

NG YONG WEI

Models Comparison

Model	Accuracy	Sensitivity	F1 Score*	False Negatives
Logistic Regression (threshold 0.5)	***0.7491***	0.7654	0.7531	2488
Logistic Regression (threshold 0.35)	0.7401	***0.8915***	***0.7742***	***1151***
CART	0.7444	0.7700	0.7508	2439

* F1 Score = Mean of Precision and Sensitivity



∴ The Optimised Logistic Regression model has the highest F1 score → **optimal balance,** best AUC score → **strong discriminative ability**

RISK SCORING

- Model coefficients converted into a 0–100 risk score
- Based on all 21 variables from the dataset
- Higher score → higher likelihood of diabetes
- Simple, interpretable, and ready for integration into LifeSG

Risk Scoring Formula:

$$\text{Risk Score} = \frac{1}{1 + e^{-(\beta_0 + \sum_i \beta_i x_i)}} \times 100$$

- Low Risk (0–40)
- Moderate (40–70)
- High Risk (70–100)

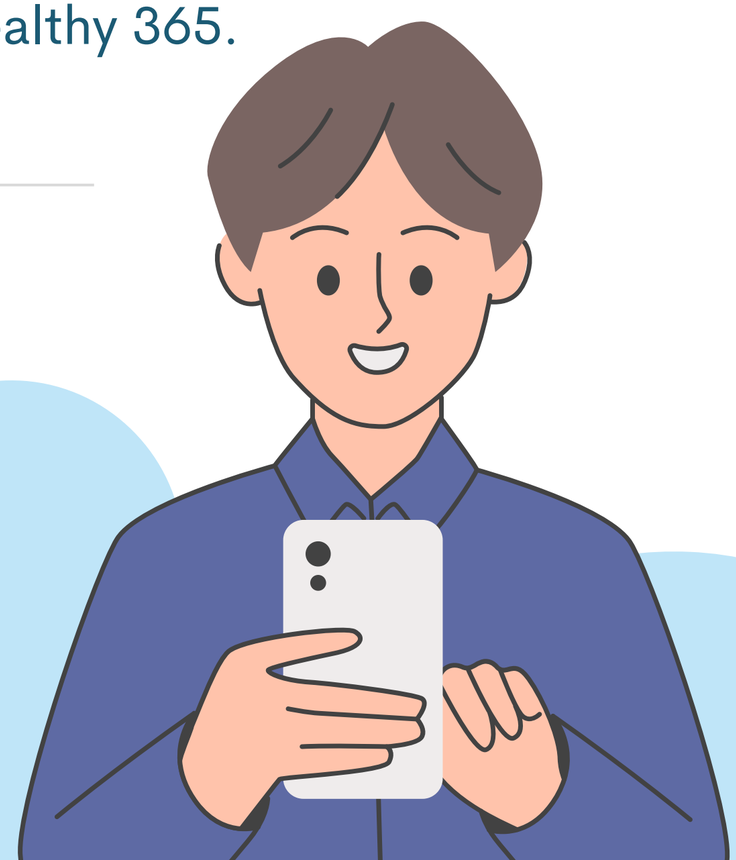
BRYAN LIM HOW MENG

Implementation Concept

Embedded under LifeSG's Health & Wellness section.

Users can input or sync data from HealthHub / Healthy 365.

- Instant feedback: personalised risk score + recommendations.
- Shared (securely) with family doctors for follow-up.



BRYAN LIM HOW MENG

Implementation Concept

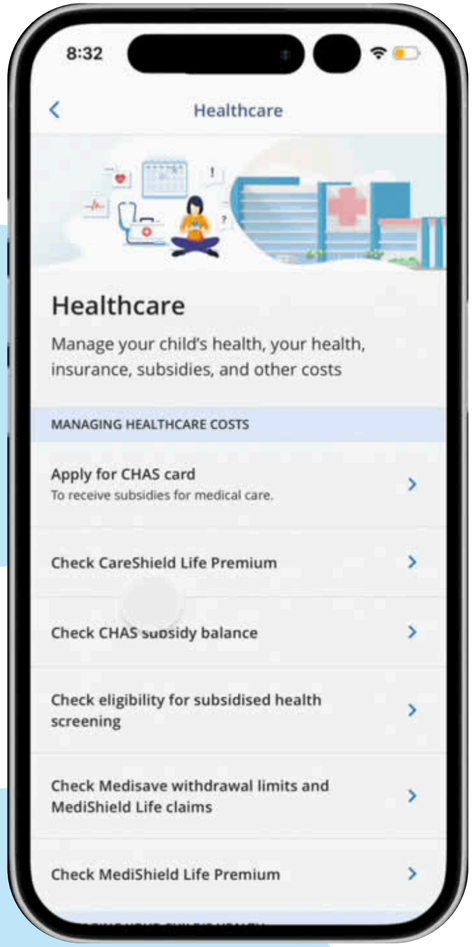
- Encourages early detection and health ownership.
- Provides aggregated, anonymised data to identify high-risk areas.
- Enables targeted national campaigns and resource planning.
- Supports Healthier SG's shift from reactive care to proactive prevention.

DE SOUZA ALYSSA ANNE

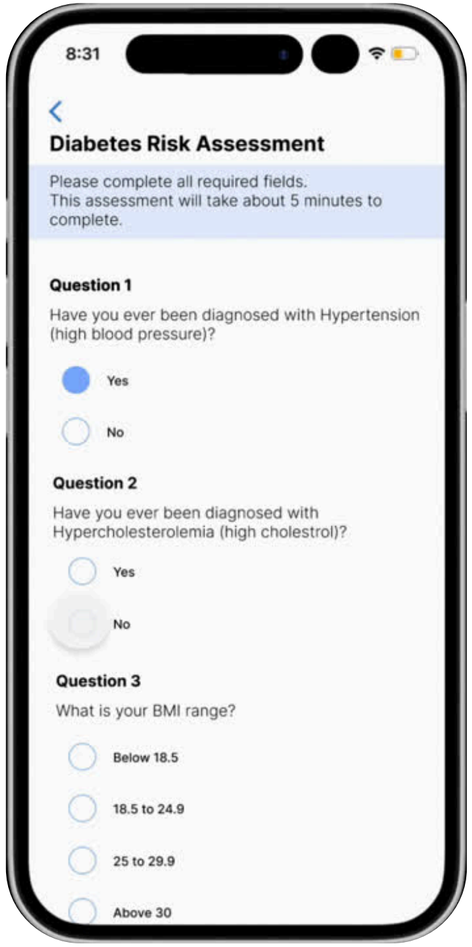
Implementation Output

USER FLOW DEMONSTRATION

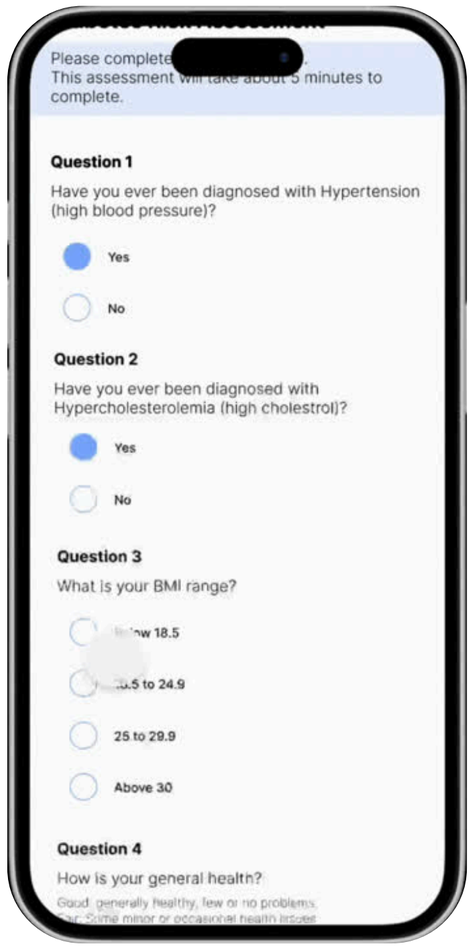
LOW RISK



MEDIUM RISK



HIGH RISK



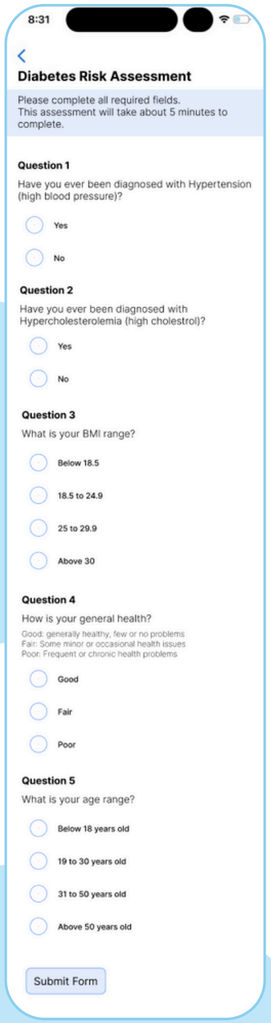
Users will access the Diabetes Risk Assessment through the "Healthcare" section on LifeSG.

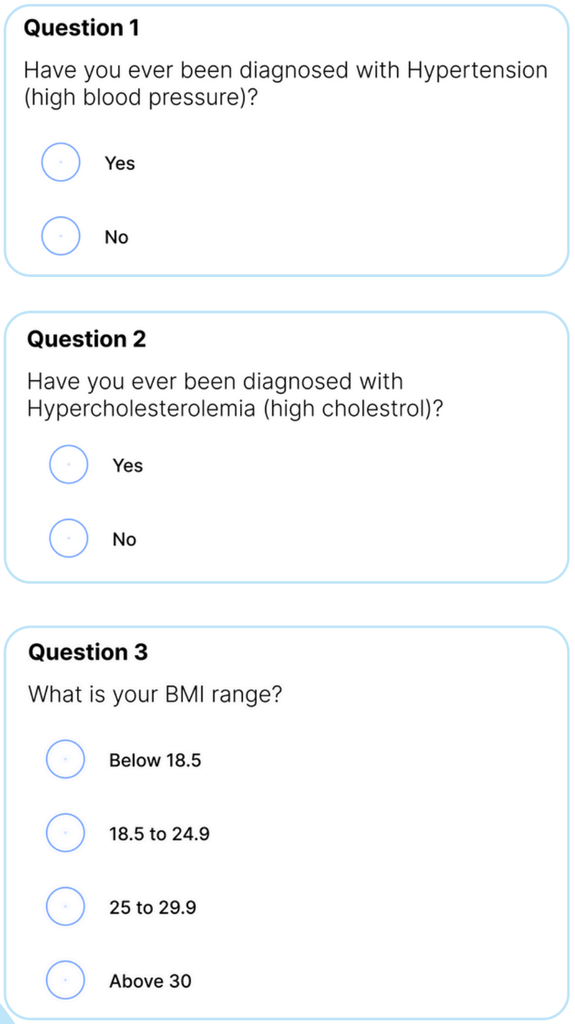
The risk assessment is in the form of a short questionnaire. Once completed, users will be grouped into "Low", "Medium" and "High" diabetes risk based on their personal risk score.

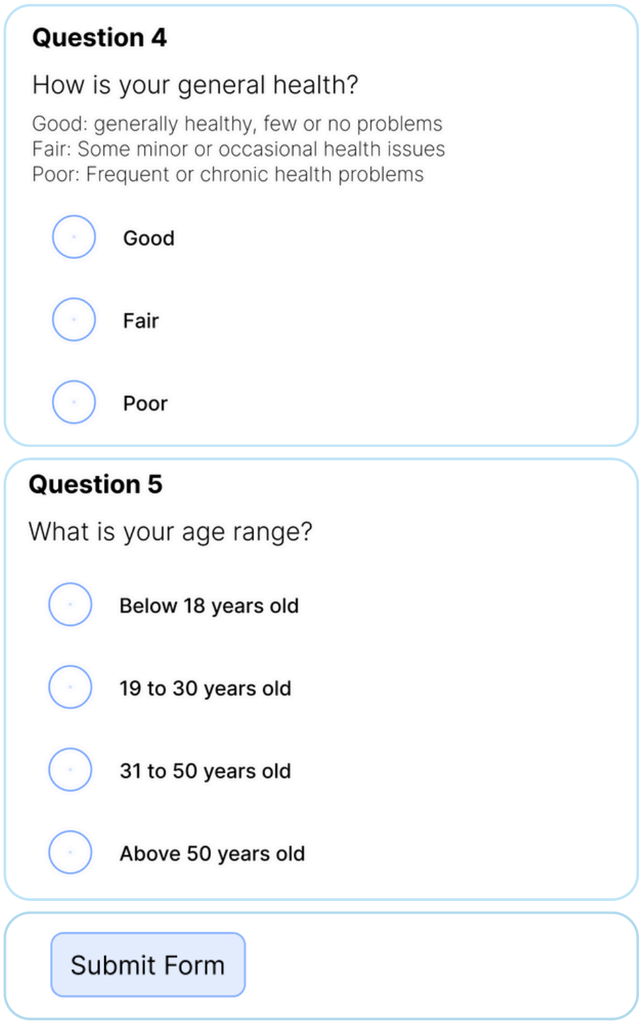
DE SOUZA ALYSSA ANNE

Implementation Output

RISK ASSESSMENT - QUESTIONNAIRE





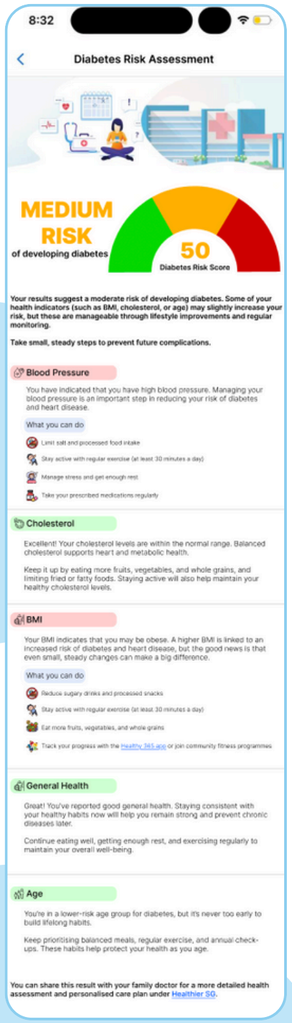


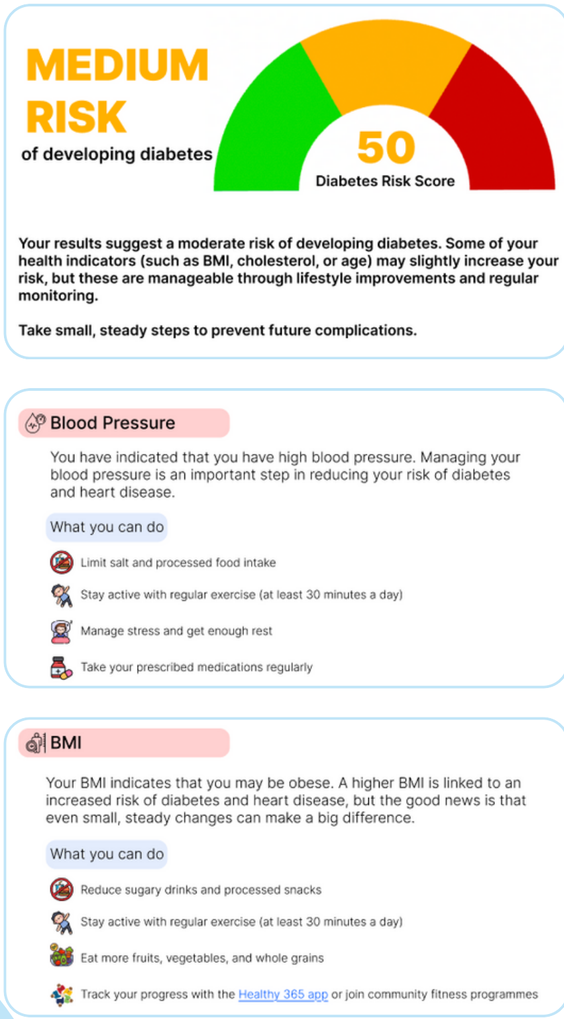
- **User-friendly UI/UX**: 5 simple MCQ questions to select from, with additional clarification for questions that need them
- **Data handling**: All responses are securely stored and synced with HealthHub/Healthy 365 upon user consent

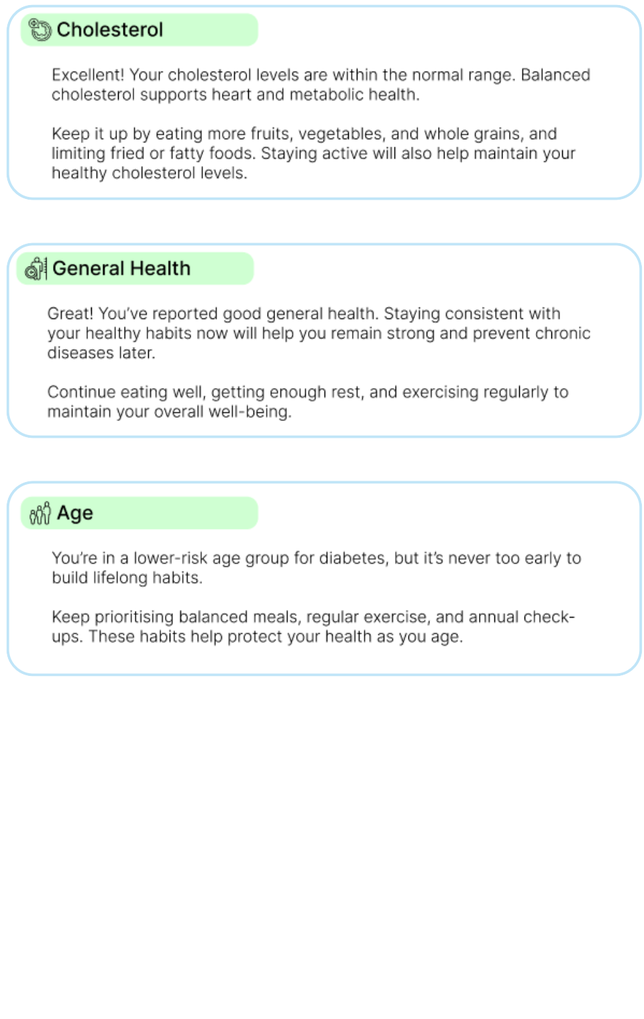
DE SOUZA ALYSSA ANNE

Implementation Output

RISK ASSESSMENT - RESULTS







Risk score is displayed using a traffic light system, and each result is paired with targeted, Healthier SG-aligned advice.

Data captured

↓

Sent to model

↓

Risk score generated

DE SOUZA ALYSSA ANNE

Conclusion

KEY TAKEAWAYS

Predictive Diabetes Risk Model: empowers Singaporeans to understand and manage their health proactively and independently

Integration with LifeSG: allowing preventive healthcare into a personal, accessible, and data-driven experience

Actionable Insights: individuals can be more independent with their health as data insights are translated into simple and actionable insights

HEALTHIER SG ALIGNMENT

The early detection and preventive action encourage engagement with national health initiatives and enables data-driven policy-making through population-level insights.

BROADER SOCIAL IMPACT

Shifts Singapore's healthcare focus from reactive treatment to proactive prevention, and contributes to building a healthier and more sustainable nation.