

Problem Statement

Problem(s) Identified

Many Singaporeans are unaware of their diabetic status, creating a gap between perceived health and actual risk, therefore a dire need exists to promote early detection and prevention of diabetes in Singapore.

*To close the
gap(s)...*

Business Goals

Develop a practical tool to estimate an individual's diabetes risk using their health and lifestyle data through analytics. This tool should addresses the critical awareness gap by enabling early detection, supporting both public and clinical prevention efforts in line with the Healthier SG initiative.

CART Model Building

Maximum Tree

Pruning

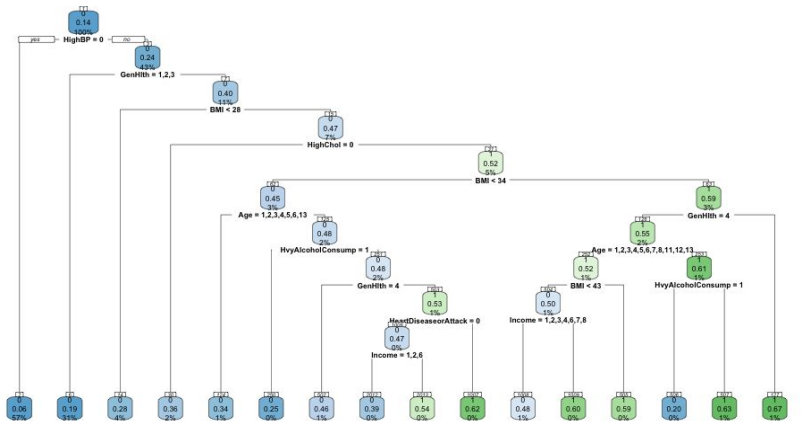
Final Tree

Diabetes.csv

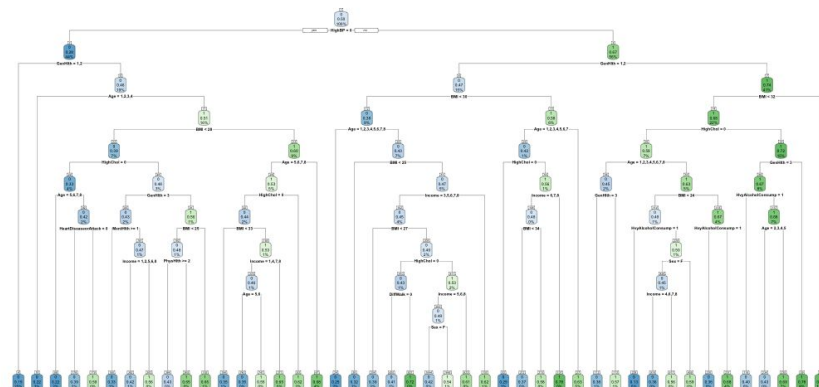
Diabetes5050.csv

Models are complicated/
extensive
→ **Most likely overfitted !**

Maximal Tree (diabetes.csv)



Maximal Tree (Diabetes5050.csv)



CART Model Building

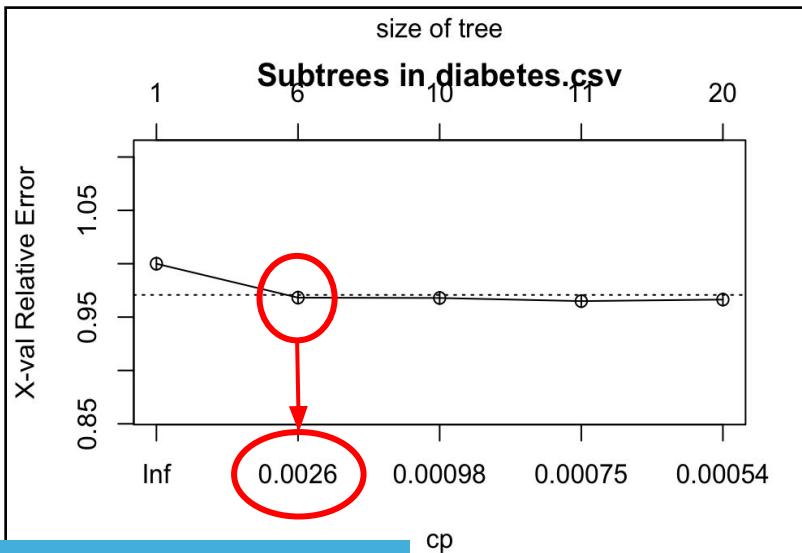
Maximum Tree

Pruning

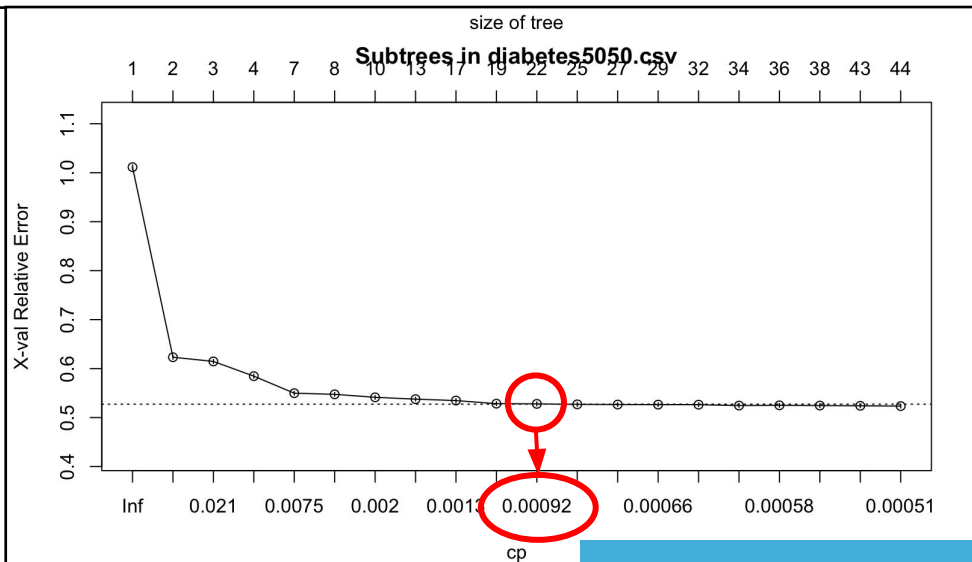
Final Tree

Diabetes.csv

Diabetes5050.csv



Pruning $cp = 0.0026$



Pruning $cp = 0.00092$

CART Model Building

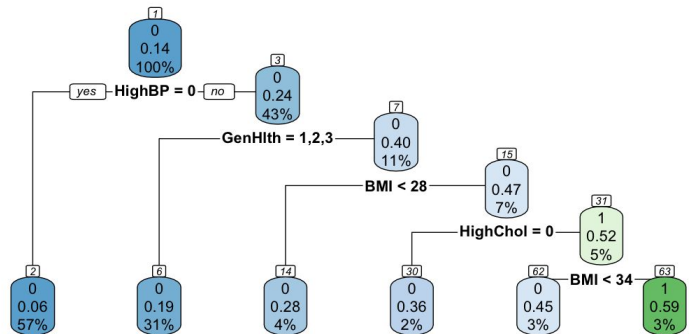
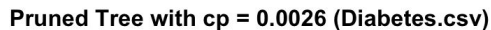
Maximum Tree

Pruning

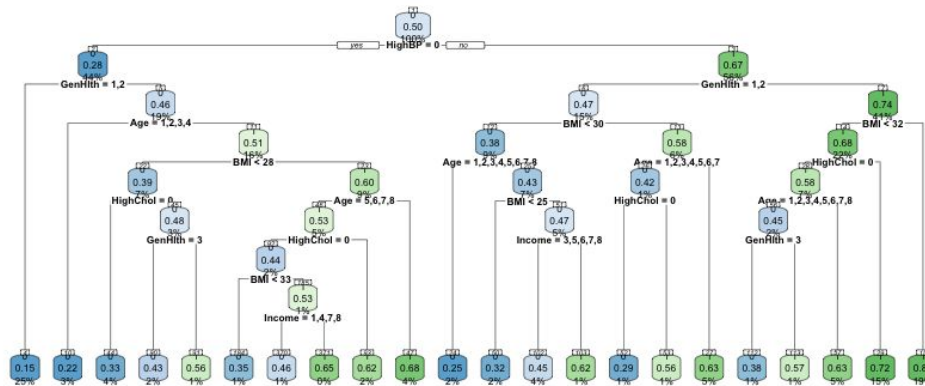
Final Tree

Diabetes.csv

Diabetes5050.csv



Pruned Tree (5050 dataset) with $cp = 0.00092$



CART Model Performance + Selection

Selected Model

Diabetes.csv

```
> confusion
```

Confusion Matrix and Statistics

	Reference	
Prediction	0	1
0	64676	824
1	9514	1090

Accuracy : 0.8642
95% CI : (0.8617, 0.8666)

No Information Rate : 0.9749
P-Value [Acc > NIR] : 1

High false -ve but
low false +ve



Bad in Medical context
(no risk when actually
have risk)

Higher accuracy
(86%)



Compare to NIR (97%)
⇒ model underperforms
significantly

Diabetes5050.csv

```
> confusion2
```

Confusion Matrix and Statistics

	Reference	
Prediction	0	1
0	7357	3247
1	2217	8387

Accuracy : 0.7424
95% CI : (0.7364, 0.7482)

No Information Rate : 0.5486
P-Value [Acc > NIR] : < 2.2e-16

Balanced false -ve
and +ve



Implications are less
serious (can always go
for other check-ups)

Lower accuracy
(74%)



Compare to NIR (54%)
⇒ Model outperforms
by a good margin

Proposed Solution – Prototype

Model used

Final Logistic Regression & CART Model

Shiny Package

Web page (ui) connected to ongoing R session (server)

Input – ui

Translation

Output – server

6 Significant Variables from CART model:

1. HighBP
2. HighChol
3. Age
4. BMI
5. General Health
6. Income

Collect all user data into a single data frame

→ new user input feeds into prediction models

Combined probabilities of both models

⇒

If > 0.5 = At Risk

If < 0.5 = Low Risk

Diabetes Risk Prediction Prototype App

Does the patient have High Blood Pressure? (1=Yes, 0=No)

0

Does the patient have High Cholesterol? (1=Yes, 0=No)

0

What is the patient's age group? (based on 13-level age category)

1

What is the patient's BMI?

25

Would the patient say that in general their health is: scale 1-5 1 = excellent 2 = very good 3 = good 4 = fair 5 = poor?

1

What is the patient's Income Category? (based on 8-level income scale)

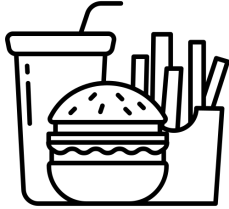
1

Predict Risk

Risk Prediction

Persona File - Mr Harold Tan [54 y/o Retiree]

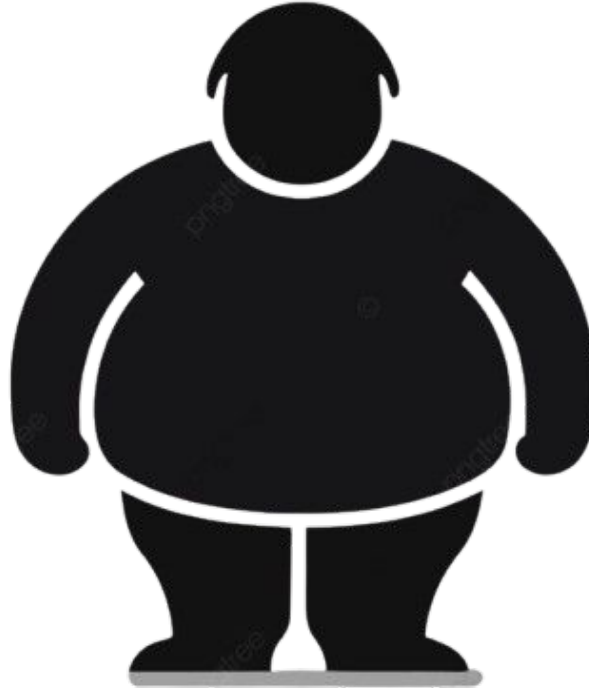
Unhealthy
diet



Smokes &
Drinks daily

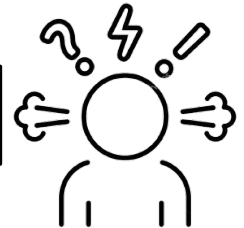


BMI = 35



Sedentary
Lifestyle

Easily Agitated
& High Stress



Prototype Test - Mr Harold Tan

Diabetes Risk Prediction Prototype App

Does the patient have High Blood Pressure? (1=Yes, 0=No)

1

Does the patient have High Cholesterol? (1=Yes, 0=No)

1

What is the patient's age group? (based on 13-level age category)

7

What is the patient's BMI?

35

Would the patient say that in general their health is: scale 1-5 1 = excellent 2 = very good 3 = good 4 = fair 5 = poor?

5

What is the patient's Income Category? (based on 8-level income scale)

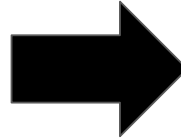
3

Predict Risk

69.3%

Risk Prediction

Probability: **69.3%** -> At Risk. Please arrange follow-ups.



Diabetes Risk Prediction Prototype App

Does the patient have High Blood Pressure? (1=Yes, 0=No)

1

Does the patient have High Cholesterol? (1=Yes, 0=No)

1

What is the patient's age group? (based on 13-level age category)

7

What is the patient's BMI?

30

Would the patient say that in general their health is: scale 1-5 1 = excellent 2 = very good 3 = good 4 = fair 5 = poor?

1

What is the patient's Income Category? (based on 8-level income scale)

3

Predict Risk

33.4%

Risk Prediction

Probability: **33.4%** -> Low Risk. Stay Vigilant.

The doctor would use this results and advise him about his high risk of developing diabetes. To reduce the risk levels, he should change his lifestyle habits such as weight management or improving lifestyle by taking care of his well-being.

If he does take his doctor's advice and improve his lifestyle habits, his risk of developing diabetes is greatly reduced. Stopping drinking and smoking could improve his High Blood Pressure and High Cholesterol, further reducing risk.