DBA5106 Group Project

Predictive Modeling to Identify Diabetic Patients

A0231956Y / Gino Martelli Tiu
A0231905L / Susan Koruthu
A0231857Y / Widya Gani Salim
A0231930N / Xhoni Shollaj

Meeting Agenda

- **Introduction**
  - Problem Statement
  - Proposed Solution and Metrics
- **Data Analysis**
  - Datasets Overview
  - Exploratory Data Analysis (EDA)
- **Model Selection**
  - Model Performance
  - Results and Expected Benefits
- **Implementation**
  - Business Implication and Implementation
  - Next Steps

NUS National University of Singapore

The true costs of diabetes

7th

Leading Cause of Death

Annual Cost

US$327 Billion

1 in 4

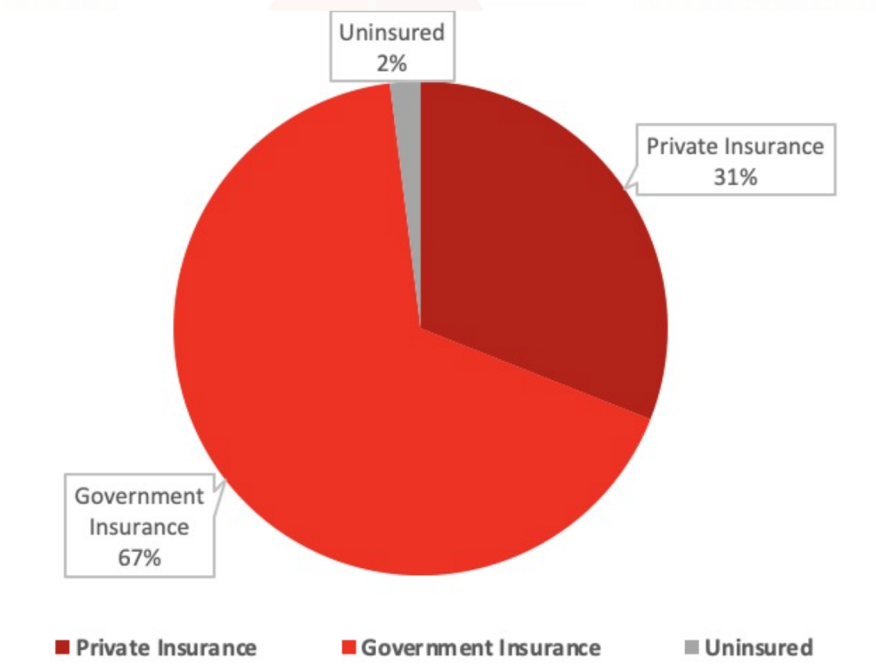
Healthcare Dollar Expenditure

Who covers these costs? We do.

NUS National University of Singapore

A costly problem

Distribution of Diabetes Treatment Claims Coverage



Annual cost (per individual)

Up to $18,652

Existing classification method detects only 30% of diabetics cases.

Average premium charged

$7,470

Annual loss (per individual)

$11,182

What can we do to manage this risk?

NUS National University of Singapore

Our proposed solution

A **classification model** with 2 key objectives:

(1) Compute the probability of diabetes diagnosis given lifestyle & medical test variables
(2) Identify high risk individuals for monitoring and intervention

Medical and personal data are readily available to us, let's take advantage.

NUS National University of Singapore

Our proposed performance metric

$$\begin{aligned} &Expected\ Value\ of\ Diabetes\ Expenses\ Claims\ = \\ &P(Diabetic) \times [\ P(Predicted\ Diabetic \mid Diabetic)\ \times \$16,752\ +\ P(Predicted\ NonDiabetic \mid Diabetic) \times \$18,652] \\ &+\ P(NonDiabetic) \times [\ P(Predicted\ NonDiabetic \mid NonDiabetic) \times \$0\ +\ P(Predicted\ Diabetic \mid NonDiabetic) \times \$200] \end{aligned}$$

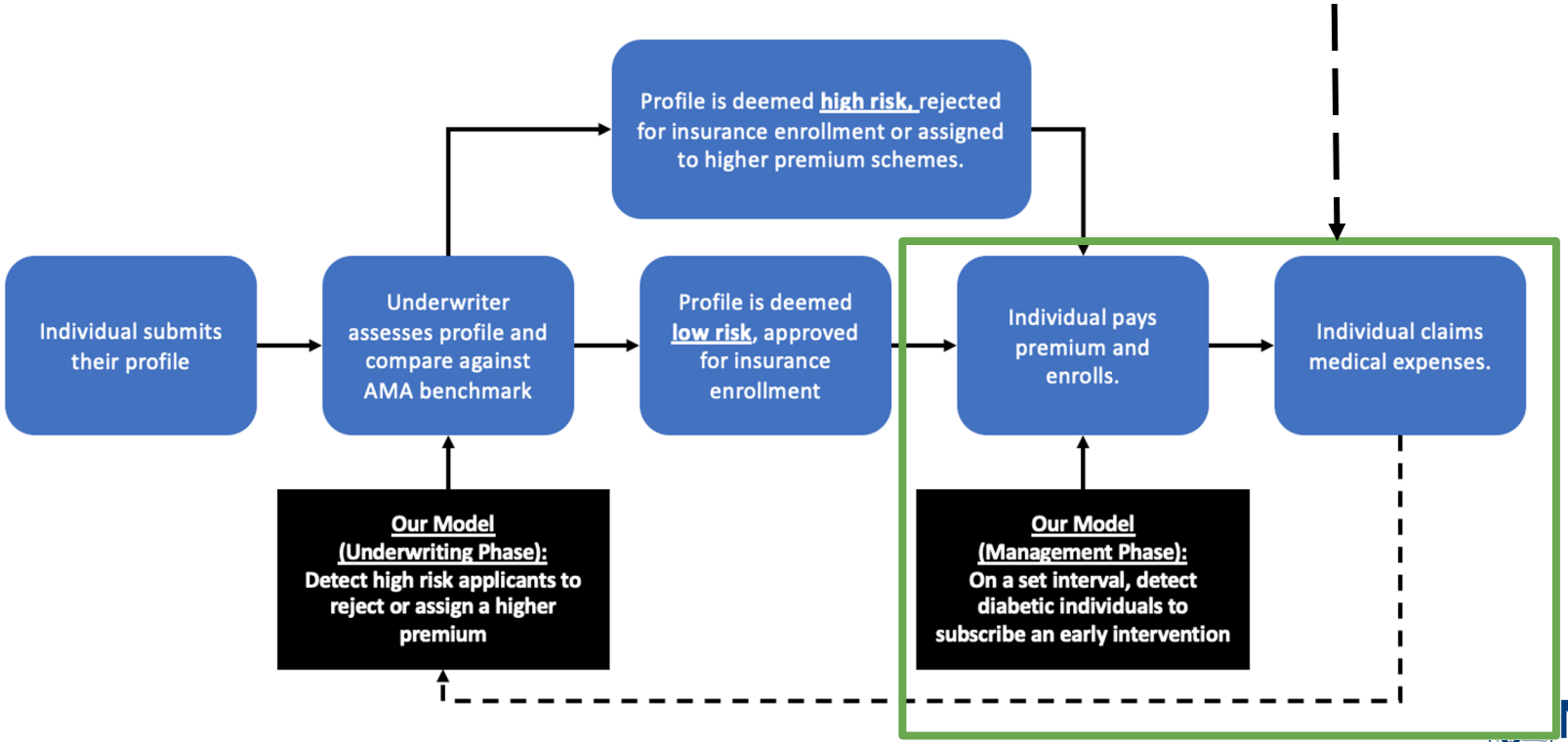
Probability	Case	Cost	Remark
P(Predicted Diabetic | Diabetic)	True Positive	$16,572	Base cost of diabetes treatment.
P(Predicted NonDiabetic| Diabetic)	False Negative	$18,652	Base cost + cost associated with complications due to the lack of early treatment ($1,900).
P(Predicted NonDiabetic | NonDiabetic)	True Negative	$0	No cost.
P(Predicted Diabetic | NonDiabetic)	False Positive	$200	Cost for diabetes screening test.

Proposed metric measures <u>cost saving</u> as compared to current classification method.

NUS National University of Singapore

Integration with our operations

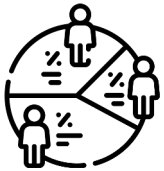
Trial run should focus on our existing customer first.



NUS National University of Singapore

Our existing dataset

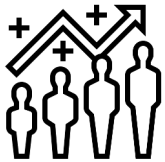
4	10,175	83
Tables	Observations	Original Features



Demographics

1. Gender
2. Age Group
3. Race
4. Education
5. Household income

...



Physical Exam

1. Body Mass Index
2. Blood pressure
3. Height
4. Weight
5. Grip strength

...

Blood Chemistry

1. Glycohemoglobin
2. Potassium
3. Lead
4. Cadmium
5. Insulin

...

Diabetic (y/n)

Identifying key features

Notable Key Features Engineered & Relabeled:

- BMI Group
- Age Group
- Socioeconomic Status
- Marital Status
- **High Risk**

High Risk
(based on AMA Guidelines)

Body Mass Index (BMI) ≥ 25

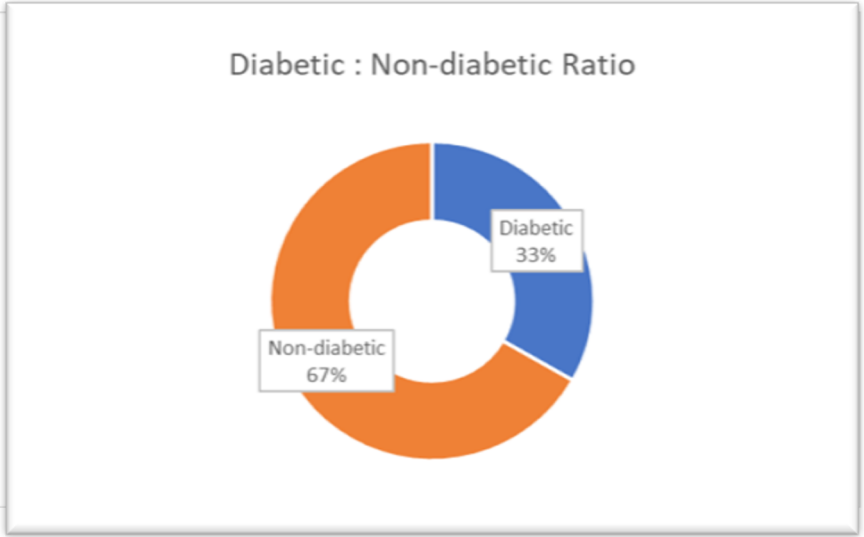
Glycohemoglobin (%) ≥ 5.7

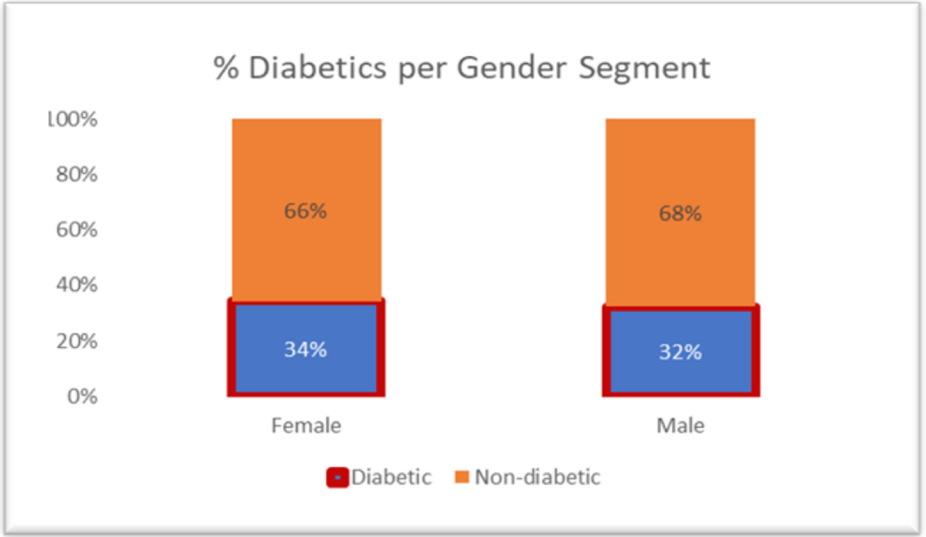
Oral Glucose Tolerance Test Level (mg/dL) ≥ 140**

NUS National University of Singapore

Exploratory data analysis

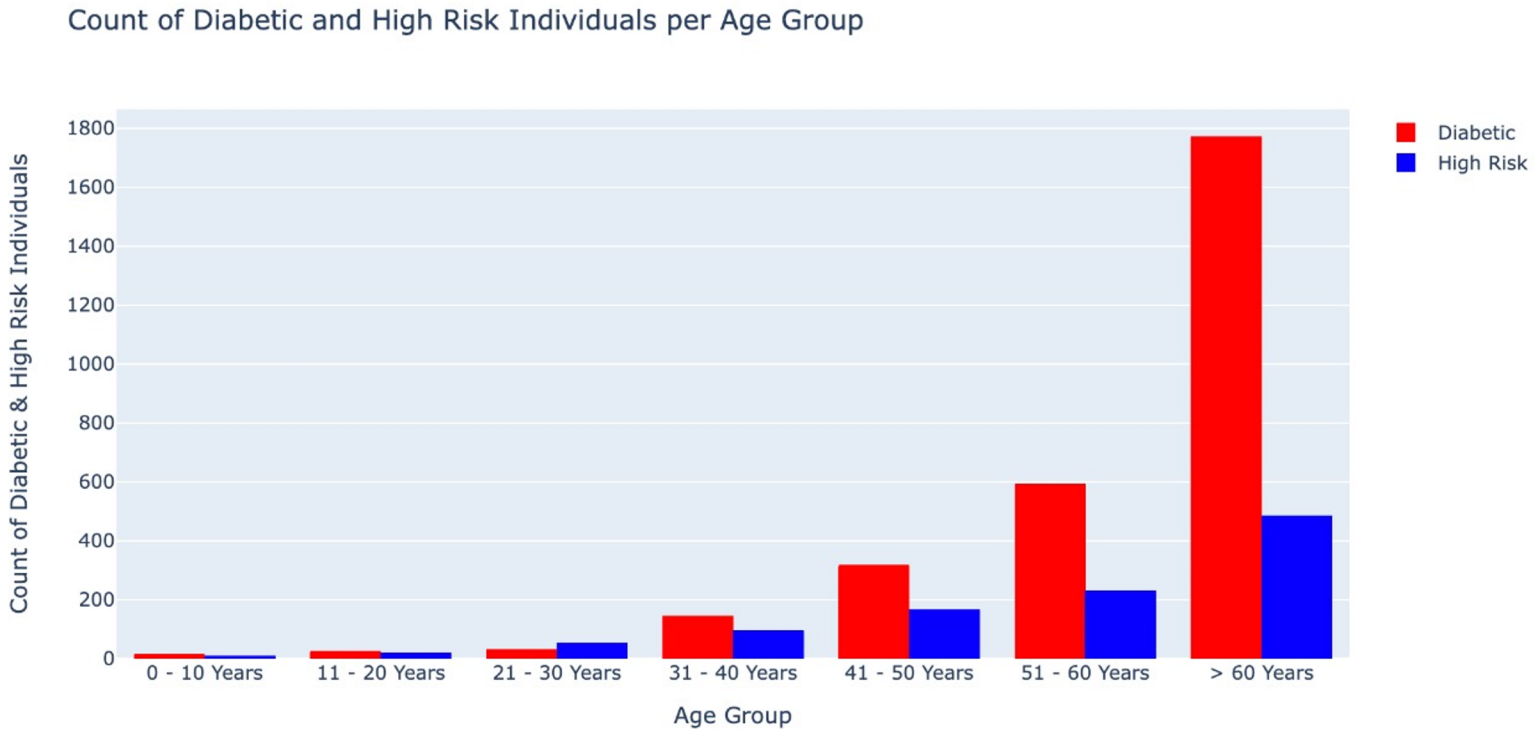
Focus is to find relevant features that could affect predisposition to diabetes.





NUS National University of Singapore

Exploratory data analysis



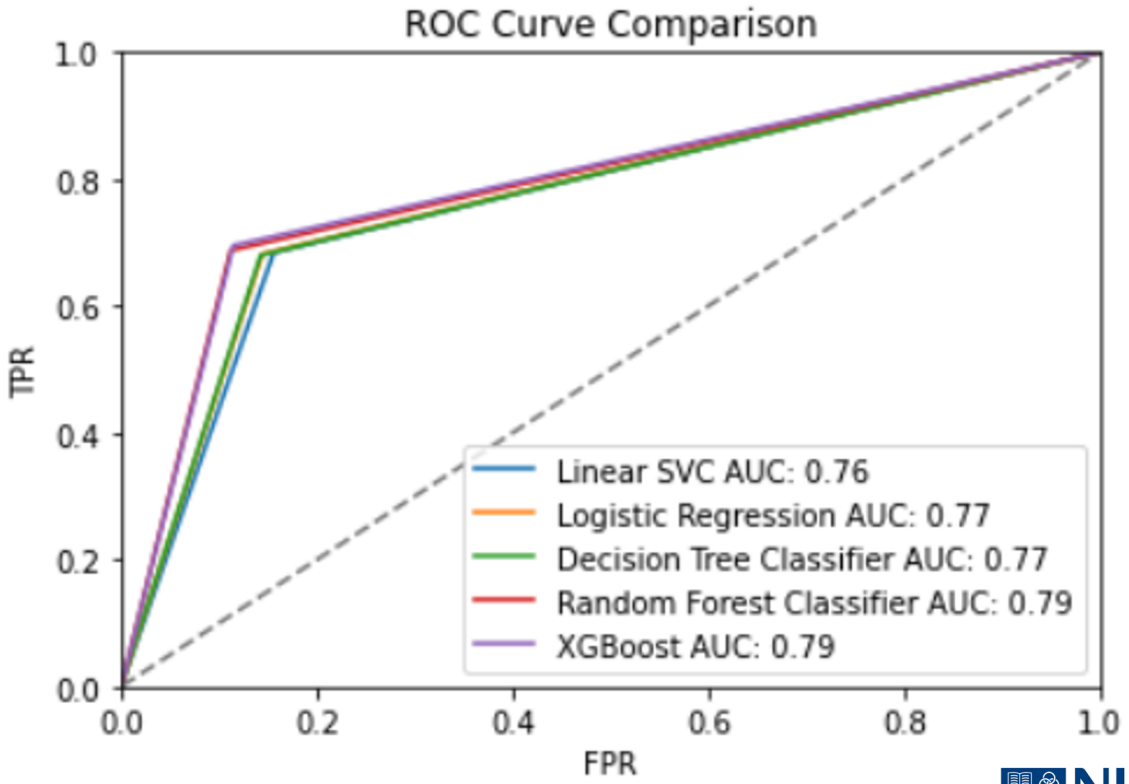
How can we use these insights to <u>manage our risks</u>?

NUS National University of Singapore

Selected model based on performance

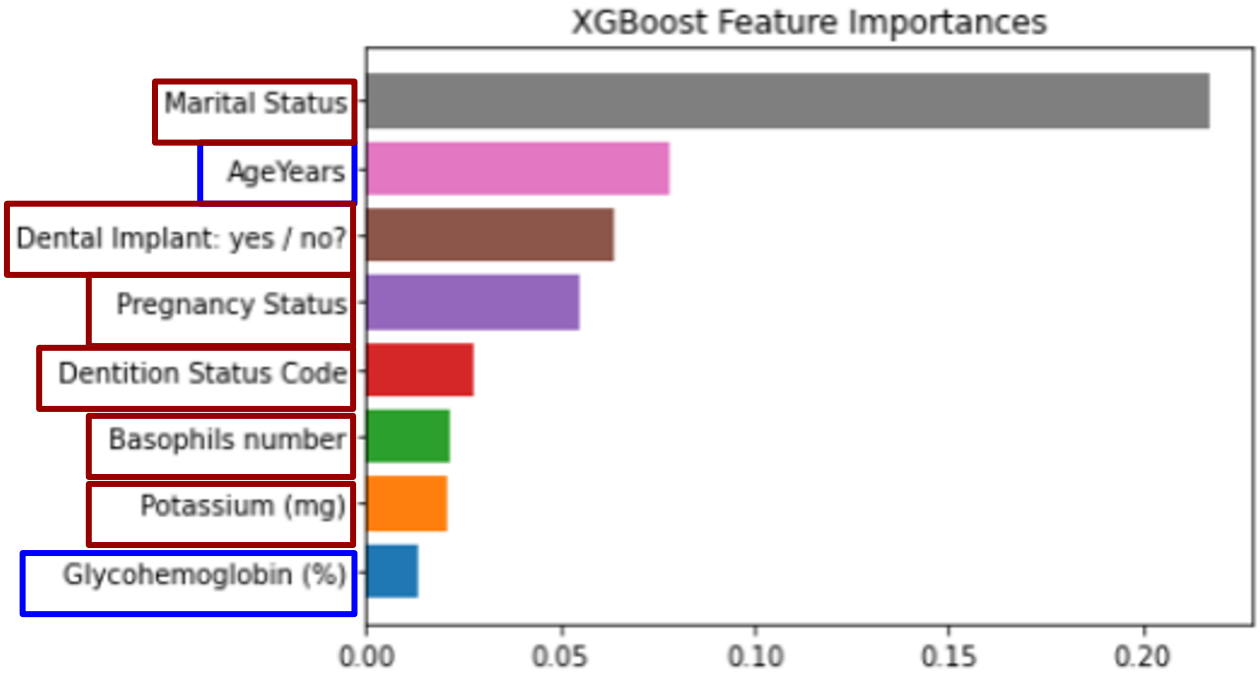
5
Models Compared

XG Boost performed best



NUS National University of Singapore

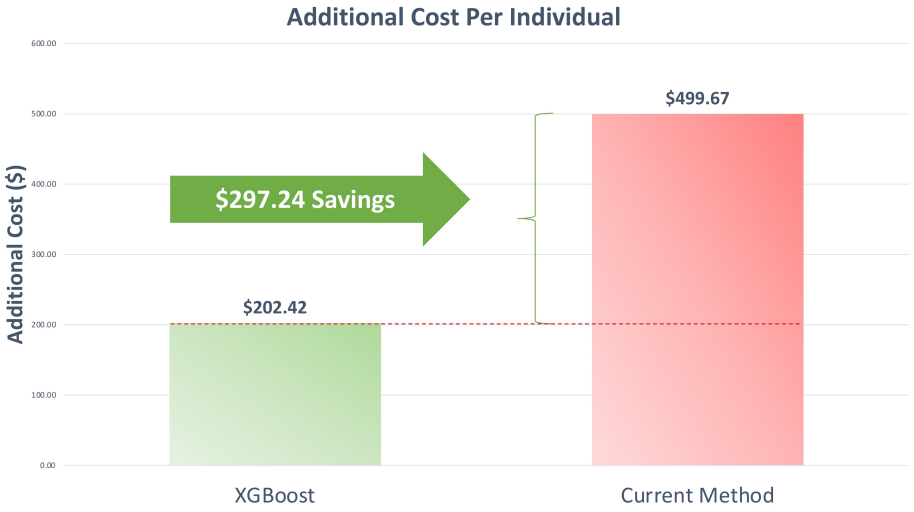
Interesting insights from top features

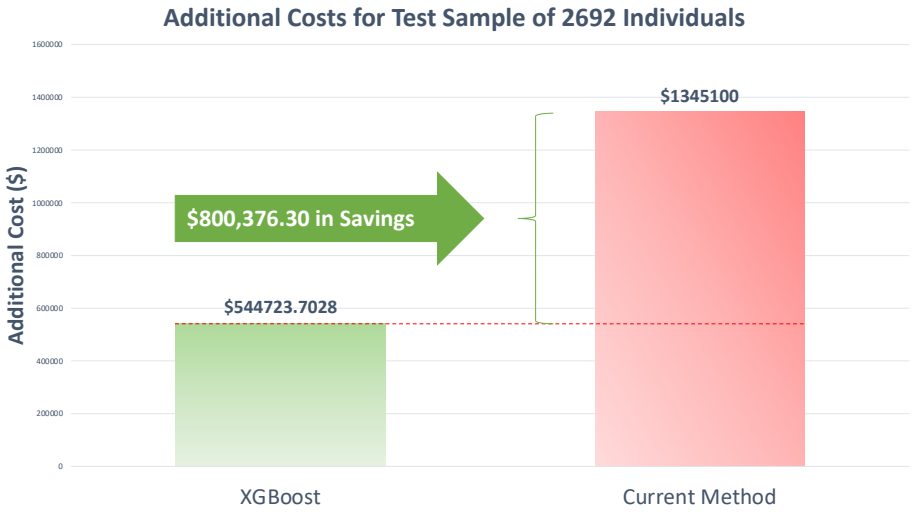


Extracted model features improve on the current detection capability and allow early intervention and better risk management.

NUS National University of Singapore

Our results and their expected benefit





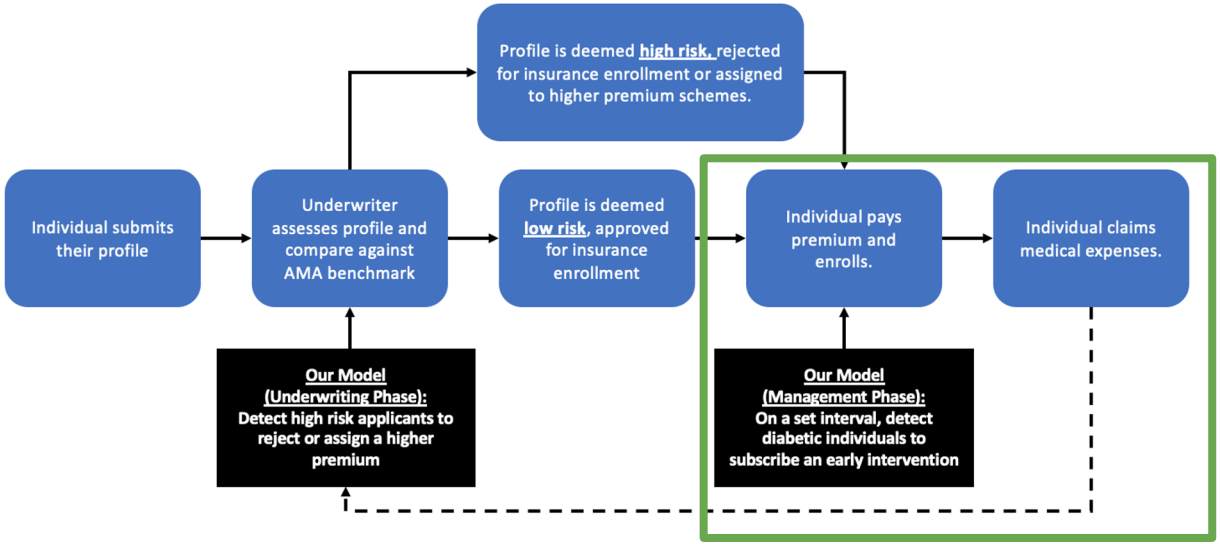
Potential cost saving is 59.49% from early and improved detection vs current industry standard…

NUS National University of Singapore

Implementation & Next Steps

Three steps to up to 59.49% cost-saving:

- Model deployment in management phase.
- Incentive or penalty scheme for those detected as diabetic.
- Track cost-savings as a performance metrics.



NUS National University of Singapore

Future Considerations

Phase 1 (Now)

- Deployment to Management Phase
- Compute Probability of Positive Diagnosis
- Identify High Risk Individuals

Performance & Business Impact Reviews

Phase 2

- Target Pre-Diabetic Cases
- Deployment to Underwriting Phase
- Provide Medical Benchmarks
- Connect Patients with Merchant Partners

The model is scalable and its cost-saving impact can be easily measured.

NUS National University of Singapore

Thank You

NUS National University of Singapore

Appendix

References:

Dataset: https://wwwn.cdc.gov/nchs/nhanes/
Model Libraries: https://scikit-learn.org/stable/user_guide.html
Facts on Diabetes(1): https://www.who.int/news-room/fact-sheets/detail/diabetes
Diabetes costs (2): https://www.diabetes.org/resources/statistics/cost-diabetes
CDCP Statistics (3): https://www.cdc.gov/diabetes/library/features/diabetes-stat-report.html
NUS MSBA 5106 Lectures