holmusk

Healthcare Data Challenge

Analysis of Drivers of the Cost of Care

By Sebastin Goh

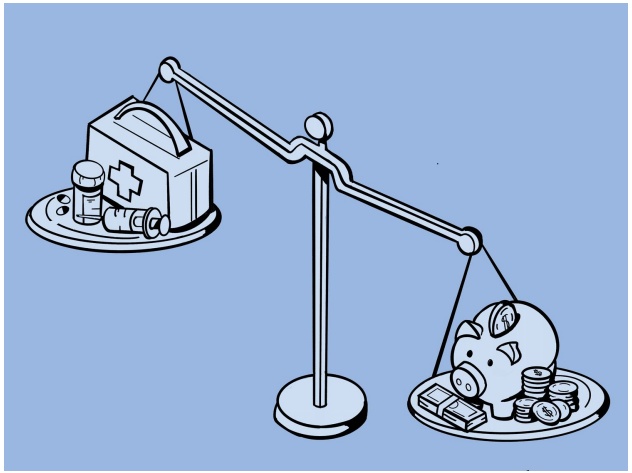
Table of contents

01

Introduction

Challenge Context
& Workflow

Background



Rising Cost of Care

- Trillions dollars annually spent to tackle healthcare challenges
- Current lifestyle trends will result in increased risk factors and costs

Need for Better Value

- Increase in healthcare spending alone does not lead to better outcomes
- Need for better understanding for sustainable future outcome per unit

(Fock & How, 2014)

Healthcare Data Challenge

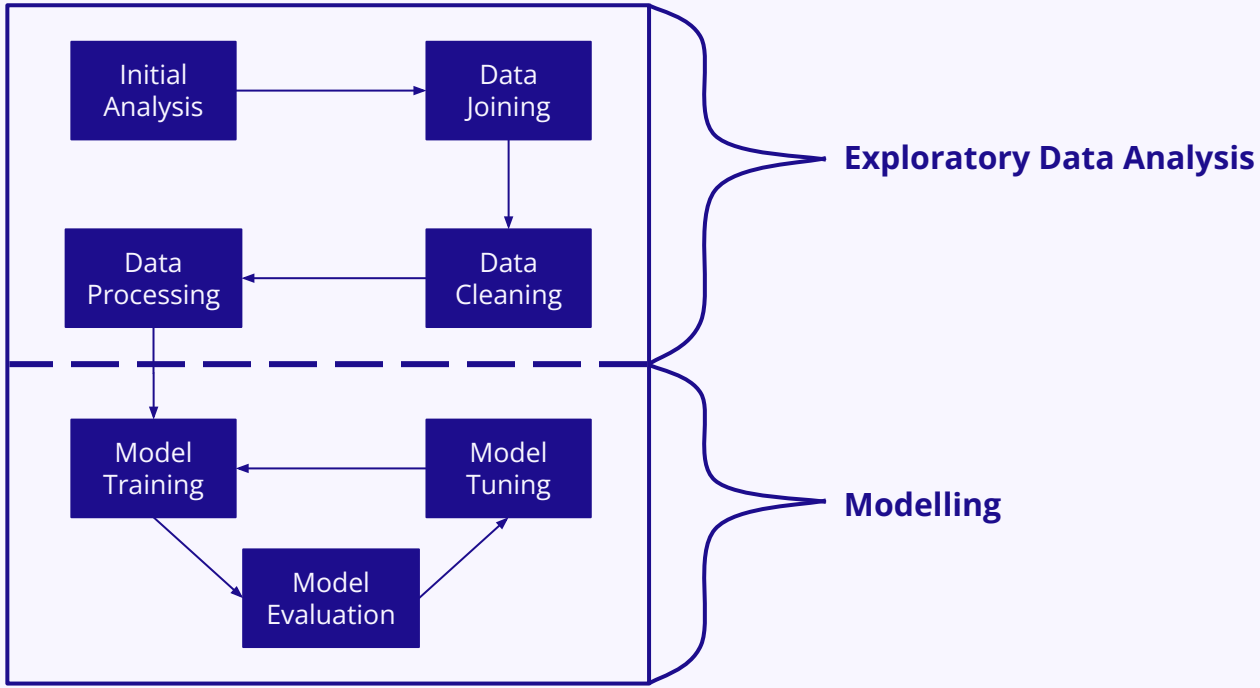
Task

- Analyze clinical & financial data of patients hospitalized for a certain condition
- Find insights about the drivers of cost of care

Deliverables

- Code (GitHub repository)
- Slides describing your approach, results, insights and case recommendations

Workflow

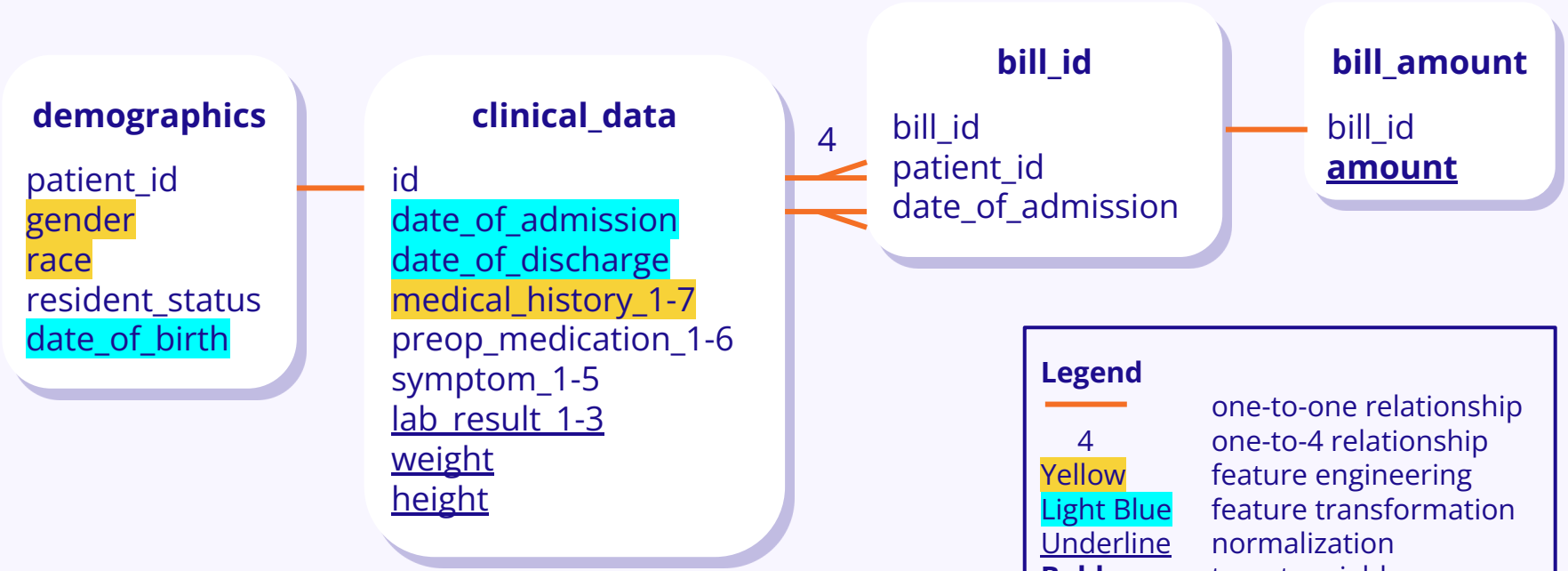

Initial Analysis
Data Joining
Data Processing
Data Cleaning
Model Training
Model Tuning
Model Evaluation
Exploratory Data Analysis
Modelling

02

Exploratory Data Analysis

Initial Analysis & Processing

Initial Analysis



Variable Processing

MinMax Normalisation

Feature Engineering & Dummy Encoding

Name	Type
length_of_visit	Continuous
age	Continuous
weight	Continuous
height	Continuous
lab_result_1-3	Continuous
amount	Continuous
medical_history_1-7	Binary
preop_medication_1-6	Binary
symptom_1-5	Binary
gender	Nominal
race	Nominal
resident_status	Nominal

03

Modelling

Model Training
& Evaluation

Model Training & Tuning

Regression Models

- Linear Regressor
- KNN XGB Regressor
- Random Forest Regressor
- Gradient Boosting Regressor
- Support Vector Regressor
- Cat Boost Regressor
- LightGBM Regressor

Tuning Methods

- Data subset comparison
  - ‘df’ (3400) vs ‘Nonull’ (2898)
- Hyperparameter Tuning
  - CV GridSearch for best parameters

Legend
Green — best performing
Bold — further analysis

Model Evaluation

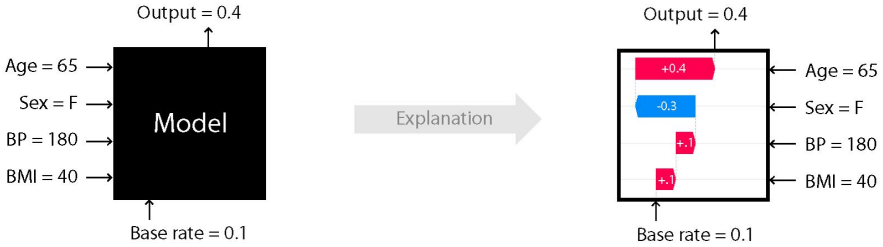
Regression Models	R2	MSE	RMSE
Linear Regressor	0.9399	0.0006	0.0242
KNN XGB Regressor	0.9564	0.0004	0.0206
Random Forest Regressor	0.8843	0.0011	0.0336
Gradient Boosting Regressor	0.9649	0.0003	0.0185
Support Vector Regressor	0.6050	0.0039	0.0621
Cat Boost Regressor	0.9891	0.0001	0.0103
LightGBM Regressor	0.9754	0.0002	0.0155
Cat Boost Regressor (Tuned)	0.9942	0.0000	0.0075

04

Insights

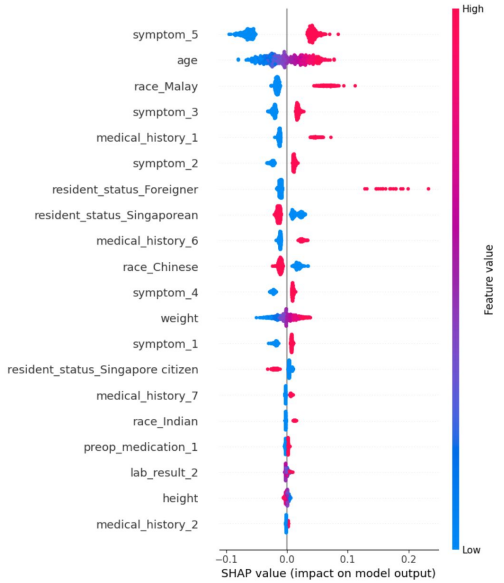
SHAP Analysis & Recommendations

SHapley Additive exPlanations



- Game theoretic approach to explain model output
- Optimal credit allocation with local explanations using classic Shapley values from game theory

(Lundberg, 2018)



SHAP Summary Plot of Cat Boost Regressor

SHAP Analysis

Key drivers contributing to cost of care

- Symptom_5
- **Age**
- **Malay**
- Symptom_3
- **Foreigner**

Older patients have historically been proven to be more likely to experience high cost of care due to:
- Age-related physiological changes
- Increased chance of underlying conditions

(Wastesson, 2018)

Possible differences in socioeconomic conditions might cause higher cost of care due to:
- Higher multi-morbidity
- Increased chance of underlying conditions

(Junxing et al., 2022)

The prevalence of medical tourism could cause higher cost of care due to:
- Specializations in advanced costlier treatments
- Overtreatment by practitioners

(Pocock & Phua, 2011)
(WTW, 2022)

Recommendation

Factor	False	True	Score
Presence of Symptom_5	0	1	1
Elderly (>65 years old)	0	1	0
Malay	0	1	1
Presence of Symptom_3	0	1	1
Foreigner	0	1	0
		Total Score	3

High Cost of Care Risk Assessment Matrix

- Created using top 5 patient variables correlated to a higher cost of care

- Simple matrix to quantify risk of higher cost of care for a patient undergoing hospitalisation for stated condition

- **Score (0-2)** Minor risk
- **Score (3-5)** Major risk

Recommendation

Potential Applications

- Programmed as a reminder system for practitioners before treatment
- Risk Assessment Card provided to practitioners for reference during treatment

Limitations

- No quantitative research done on risk thresholds
- No one size fits all: *Variance from patient to patient*

05

Conclusion

Limitations &
Future Developments

Limitations

1. Lack of Insights from Subject Matter Experts

Second opinion is invaluable for future steps

2. Limited Dataset

Current state of dataset prevents reliable and generalisable insights due to limited size and scope

3. Lack of data context

Anonymity of features prevent potential insights from further drill-down analysis

Future Developments

1. **Analysis Improvements**
   - Further transformation of non-normal variables
   - K-fold validation for models

2. **Subject Matter Expertise**
   - Inclusion of relevant patient information related to condition
   - Proper quantitative research to create accurate risk assessment matrix

Thank you!

Do you have any questions?

References

WTW. (2022, October 26). Healthcare benefit costs in Asia Pacific projected to jump by 10.2% in 2023. Healthcare benefit costs in Asia Pacific projected to jump by 10.2% in 2023, WTW survey finds. https://www.wtwco.com/en-sg/news/2022/10/healthcare-benefit-costs-in-asia-pacific-projected-to-jump-by-10-2-percent-in-2023

Junxing, C., Huynh, V. A., Lamoureux, E., Tham, K. W., & Finkelstein, E. A. (2022). Economic burden of excess weight among older adults in Singapore: A cross-sectional study. BMJ Open, 12(9). https://doi.org/10.1136/bmjopen-2022-064357

Wastesson, J. W., Morin, L., Tan, E. C. K., & Johnell, K. (2018). An update on the clinical consequences of polypharmacy in older adults: A narrative review. Expert Opinion on Drug Safety, 17(12), 1185–1196. https://doi.org/10.1080/14740338.2018.1546841

How, C., & Fock, K. (2014). Healthcare in Singapore: The present and future. Singapore Medical Journal, 55(3). https://doi.org/10.11622/smedj.2014027

Pocock, N. S., & Phua, K. H. (2011). Medical tourism and policy implications for health systems: A conceptual framework from a comparative study of Thailand, Singapore and Malaysia. Globalization and Health, 7(1), 12. https://doi.org/10.1186/1744-8603-7-12

Lundberg, S. (n.d.). Welcome to the shap documentation. Welcome to the SHAP documentation - SHAP latest documentation. https://shap.readthedocs.io/en/latest/index.html