# Heart Attack Analysis

Justin Lee

# Overview

Analyst: Justin Lee

Stakeholder: Ministry of Health and Family Welfare (MoHFW)

Motive: Leveraging ML models to identify high-risk characteristics to heart attacks and allocate medical resources accordingly, which can be used for public awareness and public health initiatives targeting key risk factors.

सत्यमेव जयते

**Ministry of Health & Family Welfare**
**Government of India**

# Data Preparation

**Model Iteration Steps**

1. Baseline logistic regression model with raw data
2. Applied random undersampling to logistic regression model to account for class imbalance
3. Iterated on random undersampled logistic regression model for decision tree classifier

# Data Understanding

Baseline model accuracy: 50%

- 10,000 patient records
- 25 categorical features
- 0 null values
- Target class imbalance
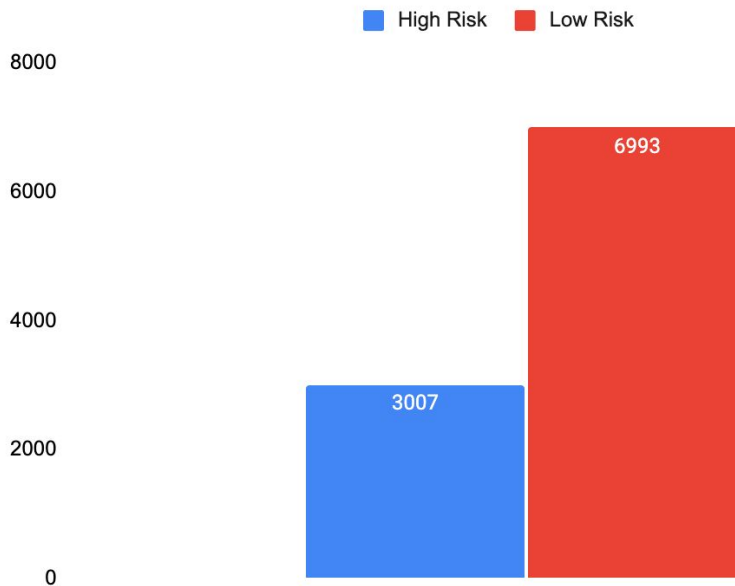
# Decision Tree Model Evaluation

- Classification report (target class)
  - Precision: 0.50
  - **Recall: 0.49**
  - F1-Score: 0.49
- All metrics resilient to class imbalance
- Recall is our north star metric to minimize the error of missing high risk patients
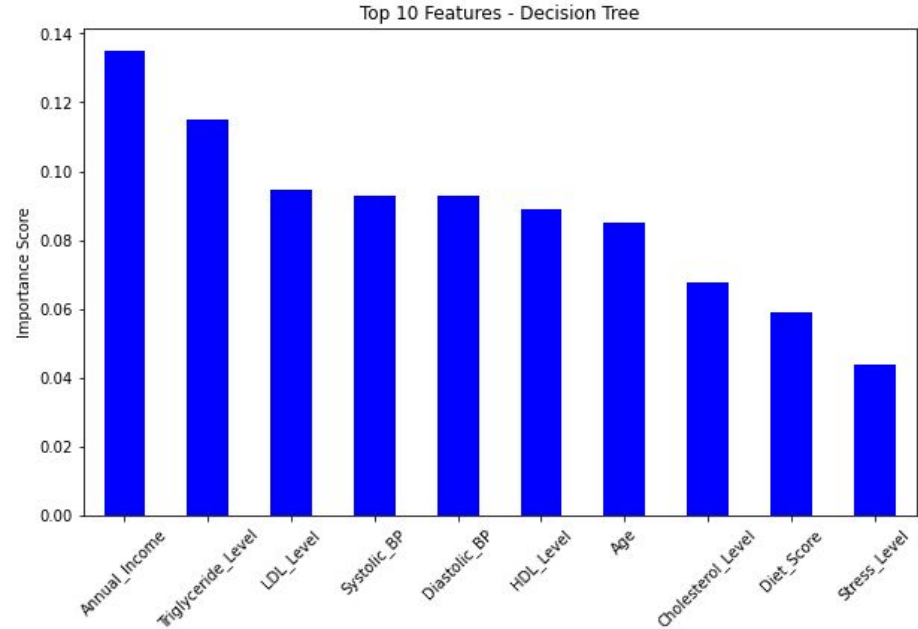
Confusion Matrix

| | |
|---|---|
| *TN: 456* | *FP: 447* |
| *FN: 463* | *TP: 439* |

# Decision Tree Feature Importance

Feature importance ranking to high heart attack risk based on Y-axis importance score:

1. Annual income
2. Triglyceride level
3. Low-density lipoprotein cholesterol (LDL) level
4. Systolic blood pressure
5. Diastolic blood pressure



*Top 10 Feature Analysis*

# Next Steps

1. Expand on data limitations of binary values
2. Sub-factors of annual income/socioeconomic level
   a. Health insurance type
   b. Coverage level
   c. Out-of-pocket costs
   d. Employment tier
3. Draft evidence-based policy interventions

# Contact Information

Justin Lee

Email: jlee8150@gmail.com

LinkedIn:
https://www.linkedin.com/in/justinlee2023/