# Healthguard Insurance

Presented by Elijah Jacob

# PROJECT OUTLINE

## TOPICS TO COVER

**Objective**

**Initial Assumptions**

**Predictive Modeling**

**Insight Analysis**

**Impact Analysis**

**Recommendation and Next Steps**

# OBJECTIVE

BUILD A **PREDICTIVE MODEL** FOR THE DIAGNOSIS OF **CHRONIC ILLNESS**

# WHY?

**ACCURATE PREDICTION** OF CHRONIC ILLNESS CAN ENABLE US TO **PRICE INSURANCE PREMIUMS** WELL LEADING TO A **MAXIMIZED SHAREHOLDER AND CUSTOMER VALUE**

# Initial Data Assumptions

**CHRONIC CONDITIONS**

1 - Diagnosed
2 - Diagnosed during pregnancy
3 - Not Diagnosed

**TRAIN/TEST SPLIT**

No systematic differences between test and train data.

# PREDICTIVE MODELING

**Data Cleaning**

Filtering outliers

Data Standardization

**Exploratory Data Analysis**

Trend analysis

Correlation-matrix plot

**Feature Selection**

Feature importance

Recursive Feature Elimination

**Model Implementation**

Cross Validation

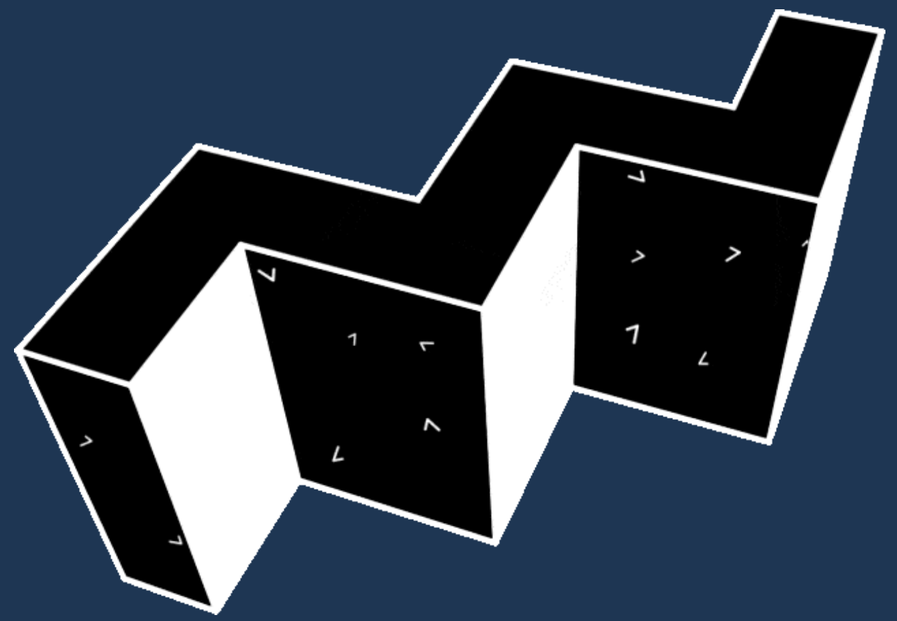
Hyperparameter Tuning

# SUMMARY OF MODELING PROCESS

First pass achieved **82.4% accuracy** with XGBoost as recommended by paper, then:

- Chose only the first 20 most important features to prevent overfitting of model
- Used GridSearchCV for hyperparameter tuning and cross-validation
- Used Smote to balance training data

Based on the papers in the appendix, I found that for the data I was working with the **XGBoost** model performed the best



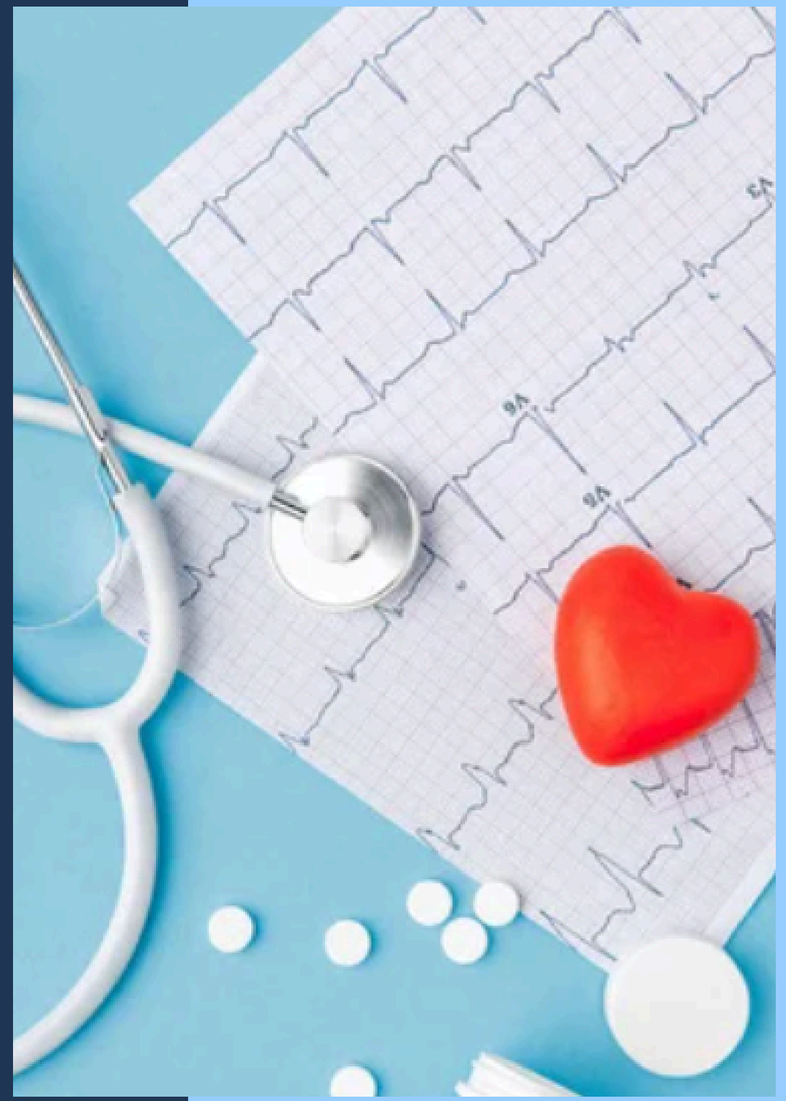
# 92%

MODEL ACCURACY IN PREDICTING CHRONIC ILLNESS ACHIEVED

Optimized with SMOTE, hyperparameter tuning and cross-validation

## Major Risk factors as seen from the data include:

| Feature | Importance |
|---|---|
| General Health | 0.162977 |
| Sex | 0.122104 |
| BMI Category | 0.09288 |
| Age | 0.06398 |
| Kidney Disease Status | 0.04905 |
| Last Routine Check Up | 0.04268 |



# MAXIMIZING CUSTOMER VALUE

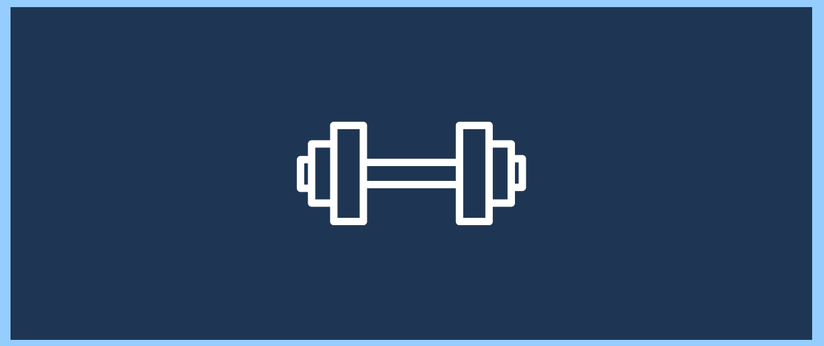
**Maximizing coverage** based on requirements:

Personalized coverage based on improving general health:

1) **Age** and **Gender**
2) Health metrics (eg. **BMI**, patient history)
3) **Exercise** and lifestyle choices (eg. smoking)

Benefits include:

- Potentially **lower premiums** due to lower risk
- **Wider coverage** based on lifestyle



# MAXIMIZING SHAREHOLDER VALUE

Ensuring insurance is priced at a level where **expected returns are consistently profitable**.

**Strategy:**

- **Tiered Pricing**: Adjust premiums based on risk level of key factors.
- **Risk Reduction**: Competitive pricing for low-risk individuals potentially being attractive to more customers.

# Recommendations and Next Steps

Balancing our shareholder and customer objectives

## PREVENTATIVE CARE INCENTIVES

Incentivizing preventative care and **check-ups** would result in more patient data as well as **early detection of disease**.

## EXERCISE INCENTIVES

Incentivizing exercise as the **trend between chronic illness and lack of exercise** is prevalent. This may look like a gym membership rebate.

## WHY?

### BETTER DATA

Through this process, **patient data is collected**.
Better data will enable better predictions which will enable us to price insurance better

### BETTER HEALTH

Incentivizing better health is a better outcome for both parties. **Customers are happier** and the insurance company makes **better margins** with lower payouts

# APPENDIX


Park, D.J., Park, M.W., Lee, H., et al. (2021). Development of machine learning model for diagnostic disease prediction based on laboratory tests. Scientific Reports, 11, 7567. https://doi.org/10.1038/s41598-021-87171-5

Lee, C., Jo, B., Woo, H., Im, Y., Park, R. W., & Park, C. H. (2024). Chronic disease prediction using the common data model: Development study. Artificial Intelligence in Medicine, 1(1), e41030. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11041444/pdf/ai_v1i1e41030.pdf

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357. https://doi.org/10.1613/jair.953