HEART FAILURE

PREDICTIVE ANALYSIS

Nithin Reddy Muduganti

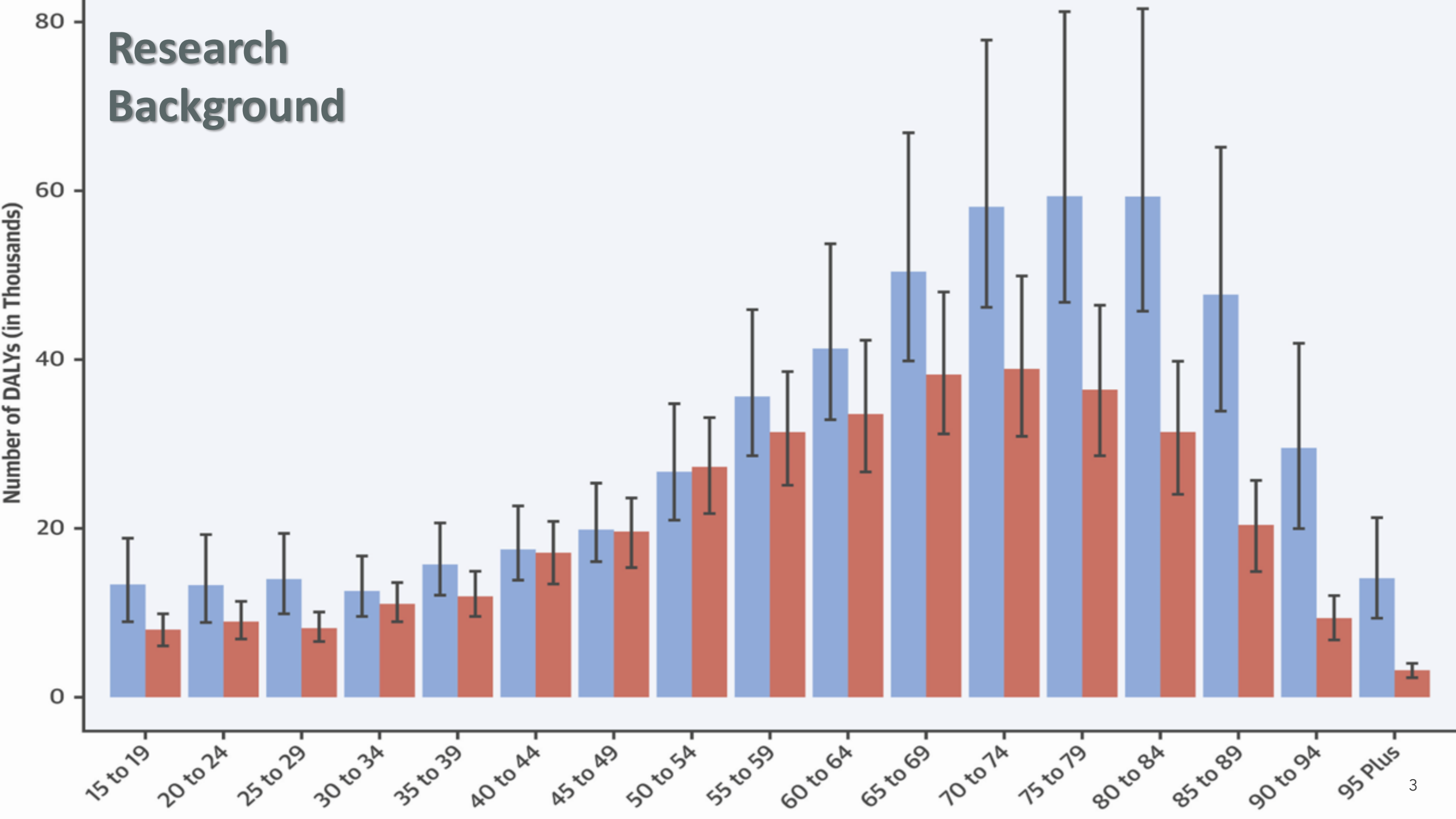Bhanu Nithin Yadav

Neehar Sai Gopisetty

Dilli Prasad Siragala

Ameer Basha Mittigar

Research Background

HEART DISEASE
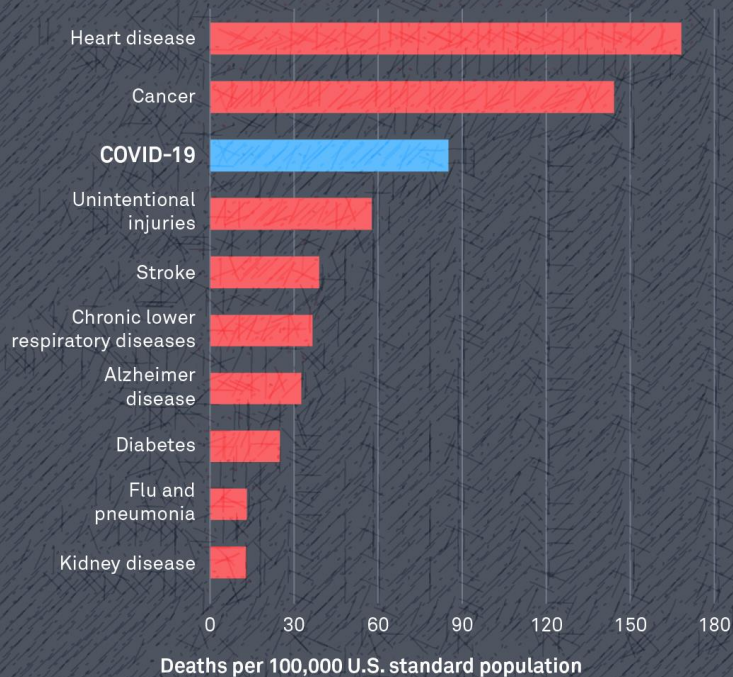
Leading causes of death in 2020

Cause	
Heart disease	
Cancer	
COVID-19	
Unintentional injuries	
Stroke	
Chronic lower respiratory diseases	
Alzheimer disease	
Diabetes	
Flu and pneumonia	
Kidney disease	

0   30   60   90   120   150   180

Deaths per 100,000 U.S. standard population

Source: Centers for Disease Control and Prevention

20%
Percentage of heart attacks that are silent

18.2M
Americans aged 20 and older with coronary heart disease

#1
Number one leading cause of death for men and women in the U.S.

647K
Number of Americans who die from heart disease each year

80%
Percentage of preventable cases of heart disease and stroke

VENTION

healthcentral

4

Business Problem

Heart failure (HF) occurs when the heart cannot pump enough blood to meet the needs of the body.

Common causes of heart failure include coronary artery disease, heart attack, high blood pressure, diabetes, excessive alcohol use, anemia, and cardiomyopathy.

Heart Failure Prediction

Physicians have long needed to identify, quantify, and interpret relationships among variables to improve patient care. Hence identifying or predicting the disease at the earliest is very important & has been the biggest challenge to avoid any unwanted casualties.
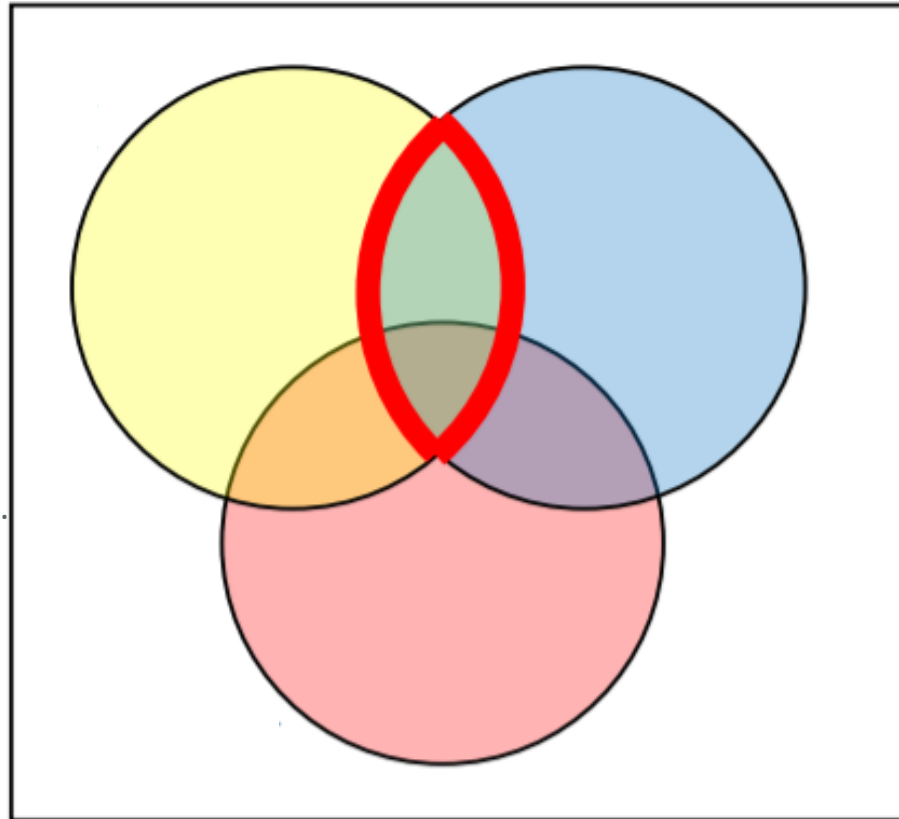
Historical Solutions

Machine learning can predict patients' survival from their data. **Most of the models developed for this purpose reach only modest accuracy, with limited interpretability from the predicting variables.**

Current Proposal

Build a new model that can be used to predict an individual's chance of heart failure more accurately and back the prediction with evidence.

Opportunities in BA

Forecasting & Planning

- Target first level healthcare centers.
- Capitalize on feedback.
- Boost the efficiency.

Reporting for Insights

- Trends & Events.
- Complex Processing.
- Evidence based management.

Optimization

- Pruning.
- Effective and efficient decision making.

Dataset Exploration

- ❖ The dataset contains the medical records of the heart failure patients collected from Faisalabad Institute of Cardiology and at the Allied Hospital in Faisalabad (Punjab, Pakistan), during April–December 2015.

- ❖ The patients consisted of 105 women and 194 men, and their ages range between 40 and 95 years old.

- ❖ All the patients had left ventricular systolic dysfunction and had previous heart failures that put them in classes III or IV of New York Heart Association (NYHA) classification of the stages of heart failure.

- ❖ Link - https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records https://drive.google.com/file/d/1FGeb5_xwQmwEEXT3AW Z9iuKh1H7DB_8S/view?usp=sharing

Attribute Information

Feature	Explanation	Measurement	Range
Age	Age of the patient	Years	[40,..., 95]
Anaemia	Decrease of red blood cells or hemoglobin	Boolean	0, 1
High blood pressure	If a patient has hypertension	Boolean	0, 1
Creatinine phosphokinase (CPK)	Level of the CPK enzyme in the blood	mcg/L	[23,..., 7861]
Diabetes	If the patient has diabetes	Boolean	0, 1
Ejection fraction	Percentage of blood leaving the heart at each contraction	Percentage	[14,..., 80]
Sex	Woman or man	Binary	0, 1
Platelets	Platelets in the blood	kiloplatelets/mL	[25.01,..., 850.00]
Serum creatinine	Level of creatinine in the blood	mg/dL	[0.50,..., 9.40]
Serum sodium	Level of sodium in the blood	mEq/L	[114,..., 148]
Smoking	If the patient smokes	Boolean	0, 1
Time	Follow-up period	Days	[4,...,285]
(TARGET) death event	If the patient died during the follow-up period	Boolean	0, 1

1.mcg/L: micrograms per liter. mL: microliter. mEq/L: milliequivalents per litre

Process Flow
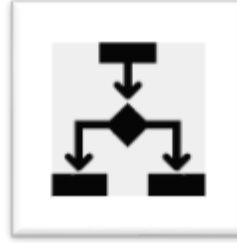
Load & explore the data.

Cleaning the data.

Standardizing & normalizing the data.

Splitting & transforming the data.

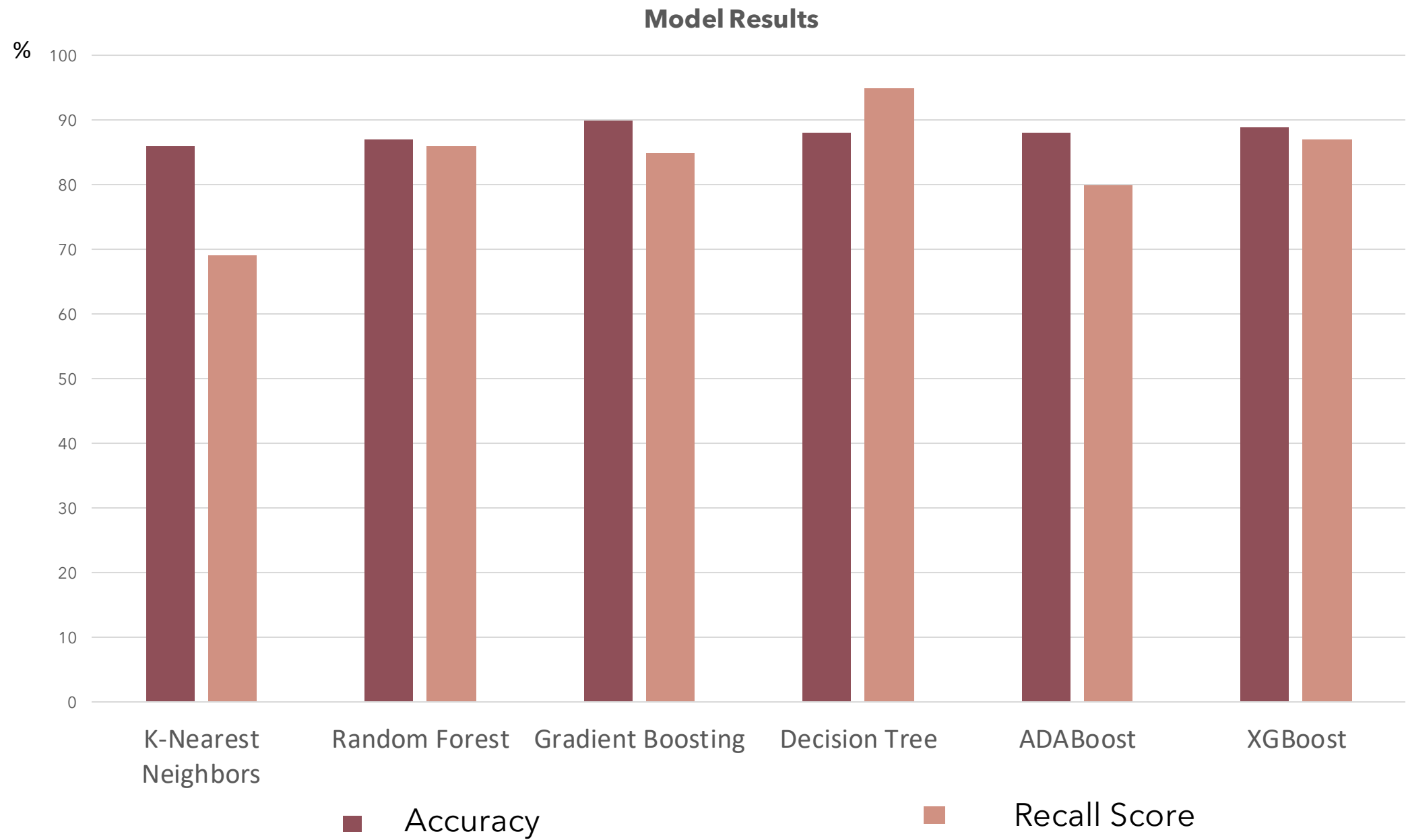Model Implementation

Model Visualization

Conclusions

Model Selection

Results

Model Results

Selection Of Predictors

Gradient Boosting:

The accuracy score achieved with this predictor is 90% which is more than that of the modest ML models existing for this case.

Decision Tree:

The recall score obtained with this predictor is 95% and is almost exact to use in identifying both positive and negative examples.
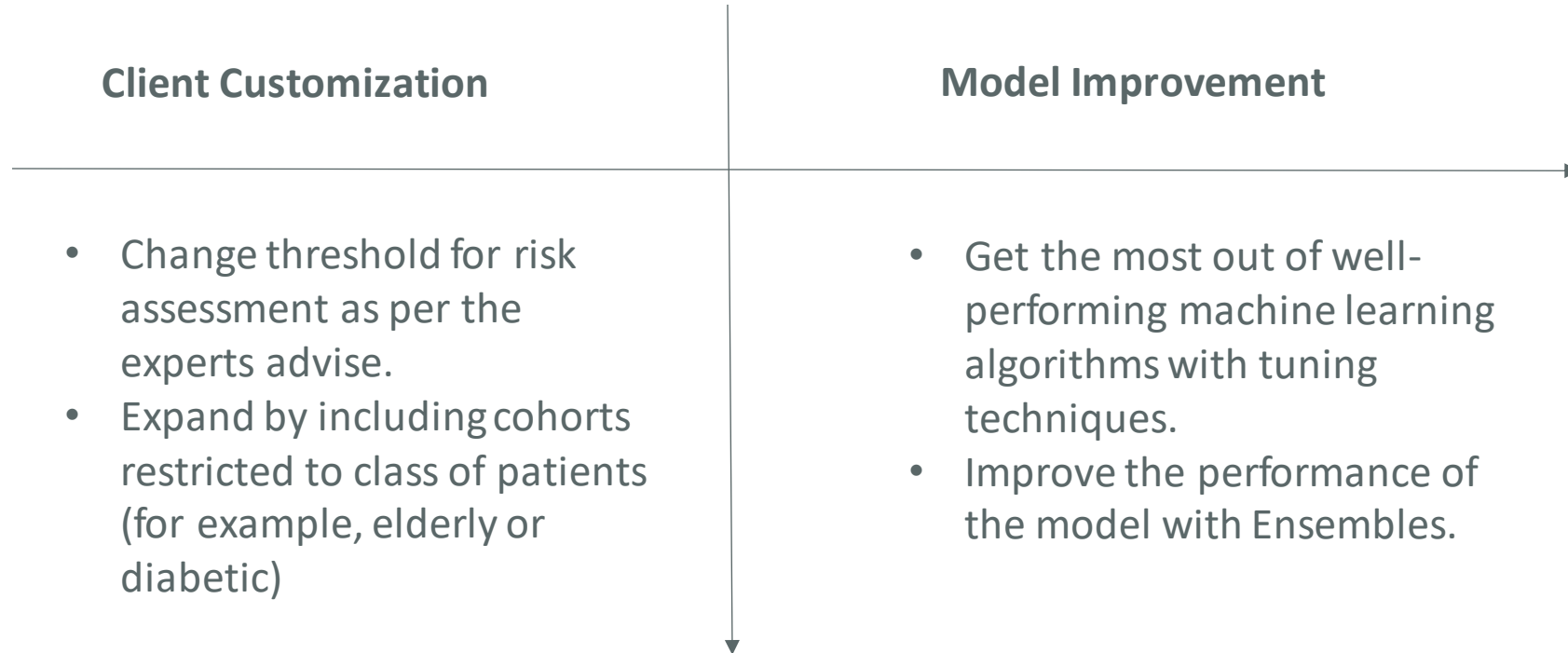
Model Optimization Scores

- **Accuracy scores:**

  - Gradient Boost - 90%

  - XGBoost – 89%

  - Decision Tree - 88%

  - ADABoost - 88%

  - Random Forest - 87%

  - K-NN model - 86%.

Further Refinements

Client Customization   **Model Improvement**

- Change threshold for risk assessment as per the experts advise.
- Expand by including cohorts restricted to class of patients (for example, elderly or diabetic)

- Get the most out of well-performing machine learning algorithms with tuning techniques.
- Improve the performance of the model with Ensembles.

❑ An addition to this development would be the pharmaceutical discoveries made for each stage of heart failure along with the current details of the individual.

Conclusions

- ✓ From all the prediction models we have performed, it is observed that gradient boost has been most effective for the training data that we have used and fetched accuracy of 90 percent with a recall score of 85, to reduce the bias error and to eliminate the death factor gradient boost has been found to be the best as our data set has many biased results. By analyzing these clinical records, we can predict the death event of an individual by various parameters taking into consideration which affects the heart. This analysis and data reports will be useful to the health organizations, hospitals and care centers etc. to predict and control the chances of death event by the various attributes mentioned and creating awareness among the people about food habits, food safety and lifestyle.
- ✓ benefits:

Benefits

1)Our model can be used by the first level healthcare centers to obtain the risk level of an individual with his/her clinical records and help them run a with more precautions by a step ahead.

2)We can also alert the audience to be watchful of the parameters and their levels that will push them close to the risky situation and guide them towards a healthy lifestyle.

3)The improved level of prediction capacity [Obtained from the better recall/Precision/F - scores and ROC curve] in our models will aim to nullify the error of predicting the Death Event.

References

- https://www.cdc.gov/nchs/products/databriefs/db425.htm
- https://www.statista.com/statistics/184515/deaths-by-heart-diseases-in-the-us-since-1950/
- https://www.healthcentral.com/condition/heart-disease

Thank You