

PREDICTING HEART DISEASE

Applications of Machine Learning
Alexander Krasniewski

SEPTEMBER 2019

AGENDA

HEART DISEASE OVERVIEW

PROBLEM STATEMENT

TRENDS

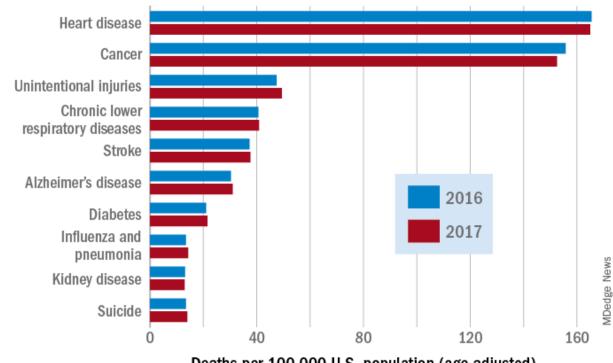
MODELS

CONCLUSIONS AND NEXT STEPS

STATISICS

- Leading cause of death for men and women
- About 630,000
 people die from
 heart disease every
 year (1 in 4 deaths)

Ten leading causes of death, 2016 and 2017



Deaths per 100,000 U.S. population (age adjusted)

Note: Based on data from the National Vital Statistics System.

Source: National Center for Health Statistics

TYPES OF HEART DISEASE

- Congenital artery disease (most common)
- Vascular disease
- Heart attack
- Heart failure

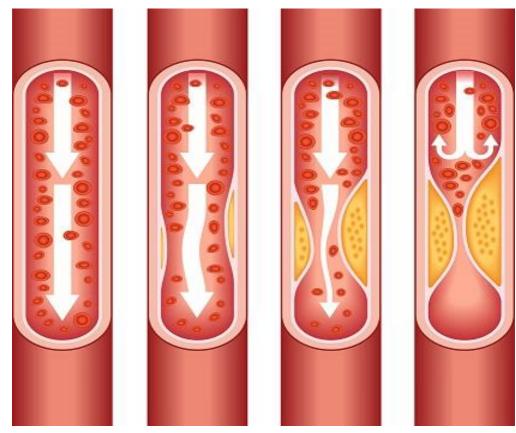


Figure 1. Plaque buildup
CDC Fast Stats: Heart Disease
Retrieved from https://www.cdc.gov/heartdisease/facts.htm

SPENDING

2015 Spending:

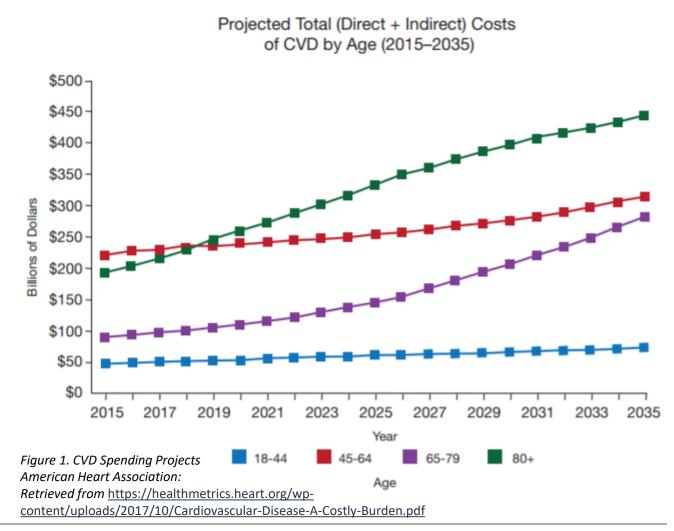
\$318 billion (direct)

\$237 billion (indirect)

2035 Projections:

\$749 billion (direct)

\$368 billion (indirect)



PROBLEM STATEMENTS

PROBLEM STATMENTS

Statements:

- 1. Can we predict heart disease accurately based on the variables from this data?
- 2. What are the trends of heart disease within our data?

Long Term Goals:

- Can we reduce heart disease with machine learning?
- 2. If we can predict heart disease with ML, could we decrease spending?



TRENDS

DATA

1025 observations and 14 variables

Age	1.0	-0.1	-0.1	0.3	0.2	0.1	-0.1	-0.4	0.1	0.2	-0.2	0.3	0.1	-0.2
Sex	-0.1	1.0	-0.0	-0.1	-0.2	0.0	-0.1	-0.0	0.1	0.1	-0.0	0.1	0.2	-0.3
Chest Pain Type	-0.1	-0.0	1.0	0.0	-0.1	0.1	0.0	0.3	-0.4	-0.2	0.1	-0.2	-0.2	0.4
Resting Blood Pressure	0.3	-0.1	0.0	1.0	0.1	0.2	-0.1	-0.0	0.1	0.2	-0.1	0.1	0.1	-0.1
Cholesterol Level	0.2	-0.2	-0.1	0.1	1.0	0.0	-0.1	-0.0	0.1	0.1	-0.0	0.1	0.1	-0.1
Fasting Blood Sugar	0.1	0.0	0.1	0.2	0.0	1.0	-0.1	-0.0	0.0	0.0	-0.1	0.1	-0.0	-0.0
Resting ECG Results	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	1.0	0.0	-0.1	-0.1	0.1	-0.1	-0.0	0.1
Max Heart Rate Achieved	-0.4	-0.0	0.3	-0.0	-0.0	-0.0	0.0	1.0	-0.4	-0.3	0.4	-0.2	-0.1	0.4
Exercise Induced Angina	0.1	0.1	-0.4	0.1	0.1	0.0	-0.1	-0.4	1.0	0.3	-0.3	0.1	0.2	-0.4
Old Peak	0.2	0.1	-0.2	0.2	0.1	0.0	-0.1	-0.3	0.3	1.0	-0.6	0.2	0.2	-0.4
Slope	-0.2	-0.0	0.1	-0.1	-0.0	-0.1	0.1	0.4	-0.3	-0.6	1.0	-0.1	-0.1	0.3
CA	0.3	0.1	-0.2	0.1	0.1	0.1	-0.1	-0.2	0.1	0.2	-0.1	1.0	0.1	-0.4
Thalassemia	0.1	0.2	-0.2	0.1	0.1	-0.0	-0.0	-0.1	0.2	0.2	-0.1	0.1	1.0	-0.3
Heart Disease	-0.2	-0.3	0.4	-0.1	-0.1	-0.0	0.1	0.4	-0.4	-0.4	0.3	-0.4	-0.3	1.0
	Age	Sex	Chest Pain Type	Resting Blood Pressure	Cholesterol Level	Fasting Blood Sugar	Resting ECG Results	Max Heart Rate Achieved	Exercise Induced Angina	Old Peak	Slope	CA	Thalassemia	Heart Disease

-0.6 -0.3 -0.0

-0.9

MODELS

LOGISTIC REGRESSION

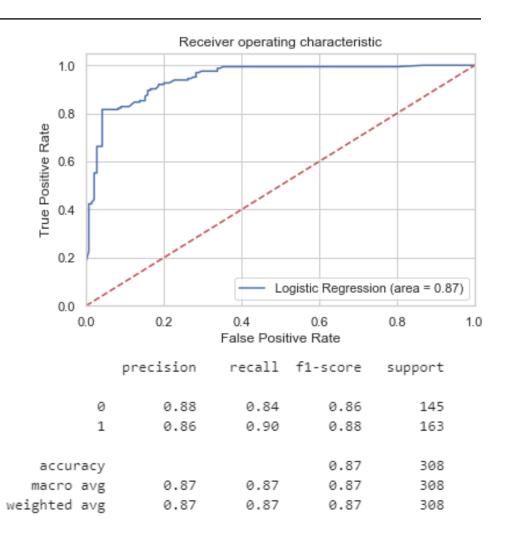
Baseline: 51% accuracy

Features:

Sex, Chest pain type, Resting blood pressure, Resting ECG results, exercised induced angina, number of major vessels colors, thalassemia

Results:

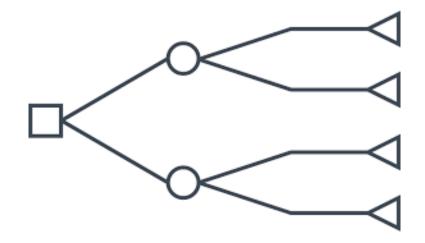
Precision score of 87% Accuracy of 87%



DECISION TREE

Baseline Model: 51% accuracy

Features: All Features



Results:

Precision score of 94%

Accuracy of 94 %

	precision	recall	f1-score	support
0	0.94	0.94	0.94	258
1	0.94	0.94	0.94	255
accuracy			0.94	513
macro avg	0.94	0.94	0.94	513
weighted avg	0.94	0.94	0.94	513

K-NEAREST-NEIGHBORS

Baseline: 51% accuracy

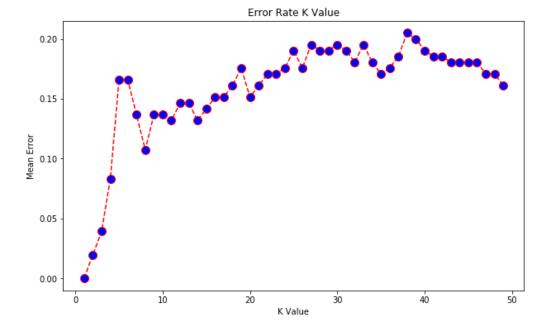
Results:

Precision score of 89%

Accuracy of 89 %

0.89 -		Λ					score
0.88 -		/ \					
0.87 -	/	\		\	\wedge		
0.86 -	/				$/ \setminus$		
0.85 -						\	
0.84 -							
0.83 -							
	6	8	10	12 k	14	16	18

	precision	recall	f1-score	support
0	0.92	0.87	0.90	109
1	0.86	0.92	0.89	96
accuracy			0.89	205
macro avg	0.89	0.89	0.89	205
weighted avg	0.89	0.89	0.89	205



K-NEAREST-NEIGHBORS

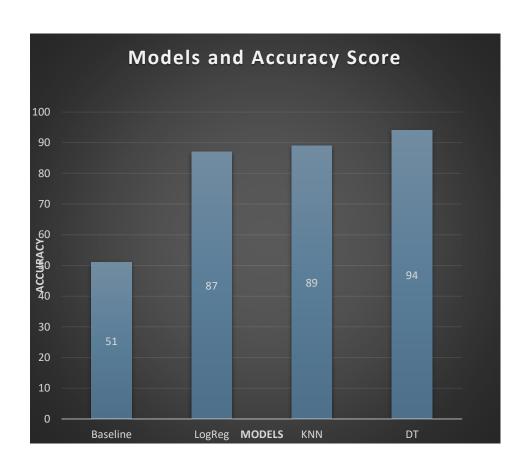
PREDICTION

```
new_patient = pd.DataFrame({
 'Age'
                           : 24,
 'Sex'
 'Chest Pain Type'
                           : 0,
 'Resting Blood Pressure' : 200,
 'Cholesterol'
                           : 150,
 'Fasting Blood Sugar'
                           : 1,
 'Resting ECG'
                           : 0,
 'Max Heart Rate'
                           : 185,
 'Exercise Induced Angina': 1,
 'oldpeak'
 'slope'
                           : 0,
 'ca'
                           : 0,
 'thalassemia'
                           : 1
}, index = [0])
new_patient
                                                                       Max
                                                                                  Exercise
               Chest Resting Blood
                                                   Fasting Resting
                                   Cholesterol
  Age Sex
                                                                      Heart
                                                                                  Induced oldpeak slope ca thalassemia
                                              Blood Sugar
            Pain Type
                           Pressure
                                                                       Rate
                                                                                   Angina
                                                                        185
                                                                                                      0 0
print(knn.predict proba(new patient).round(2))
print(knn.predict(new patient))
[[0.58 0.42]]
[0]
```

No Heart Disease!

CONCLUSIONS

CONCLUSIONS



NEXT STEPS

- Feature Engineering
- Fine tuning parameters
- Looking at the complete data set of 74 variables

SOURCES

- Devitt, Michael. "CDC Data Show U.S. Life Expectancy Continues to Decline." *AAFP Home*, AAFP, 10 Dec. 2018, www.aafp.org/news/health-of-the-public/20181210lifeexpectdrop.html.
- "Heart Disease Fact Sheet | Data & Statistics | DHDSP | CDC." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 2017, www.cdc.gov/dhdsp/data statistics/fact sheets/fs heart disease.htm.
- "CARDIOVASCULAR DISEASE: A COSTLY BURDEN FOR AMERICA. PROJECTIONS THROGH 2035" American Heart Association https://healthmetrics.heart.org/wp-content/uploads/2017/10/Cardiovascular-Disease-A-Costly-Burden.pdf

THANK YOU, NEXT