



PREDICTING HEART DISEASE

Applications of Machine Learning
Alexander Krasniewski

SEPTEMBER 2019

AGENDA

HEART DISEASE OVERVIEW

PROBLEM STATEMENT

TRENDS

MODELS

CONCLUSIONS AND NEXT STEPS

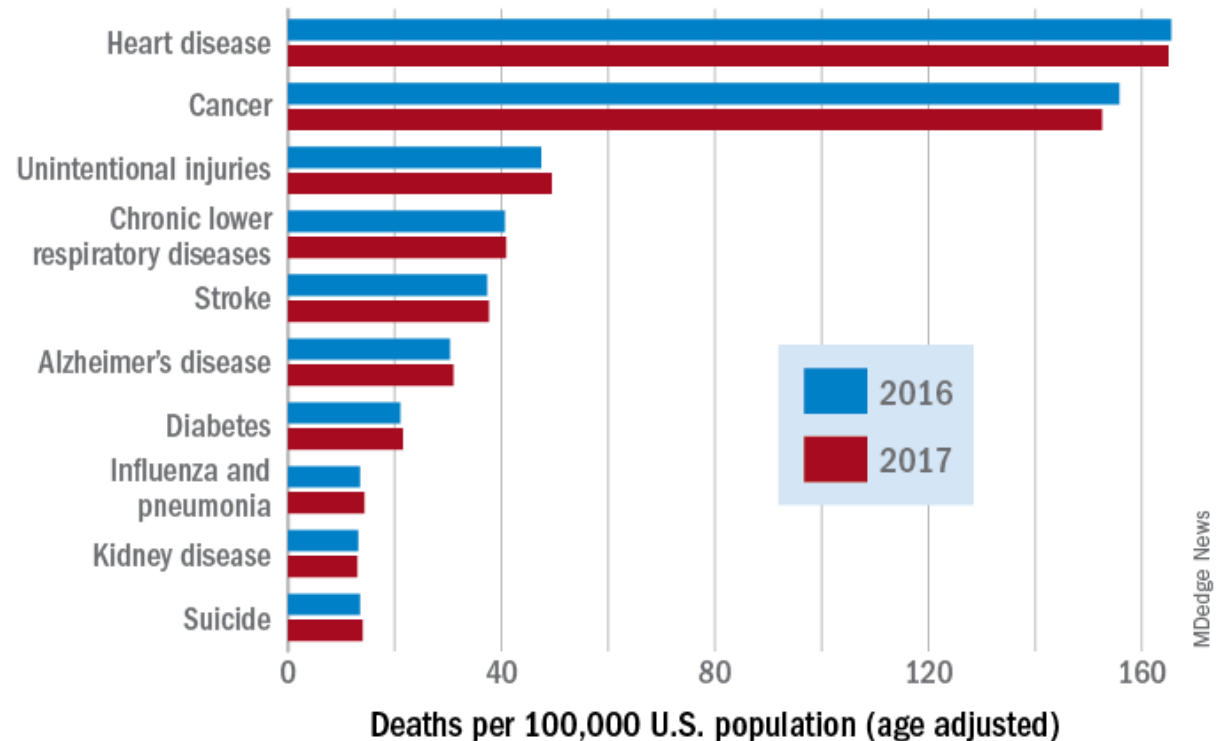
HEART DISEASE

HEART DISEASE

STATISTICS

- *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



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

Source: National Center for Health Statistics

HEART DISEASE

TYPES OF HEART DISEASE

- *Coronary artery disease*
- *Vascular disease*
- *Heart attack*
- *Heart failure*

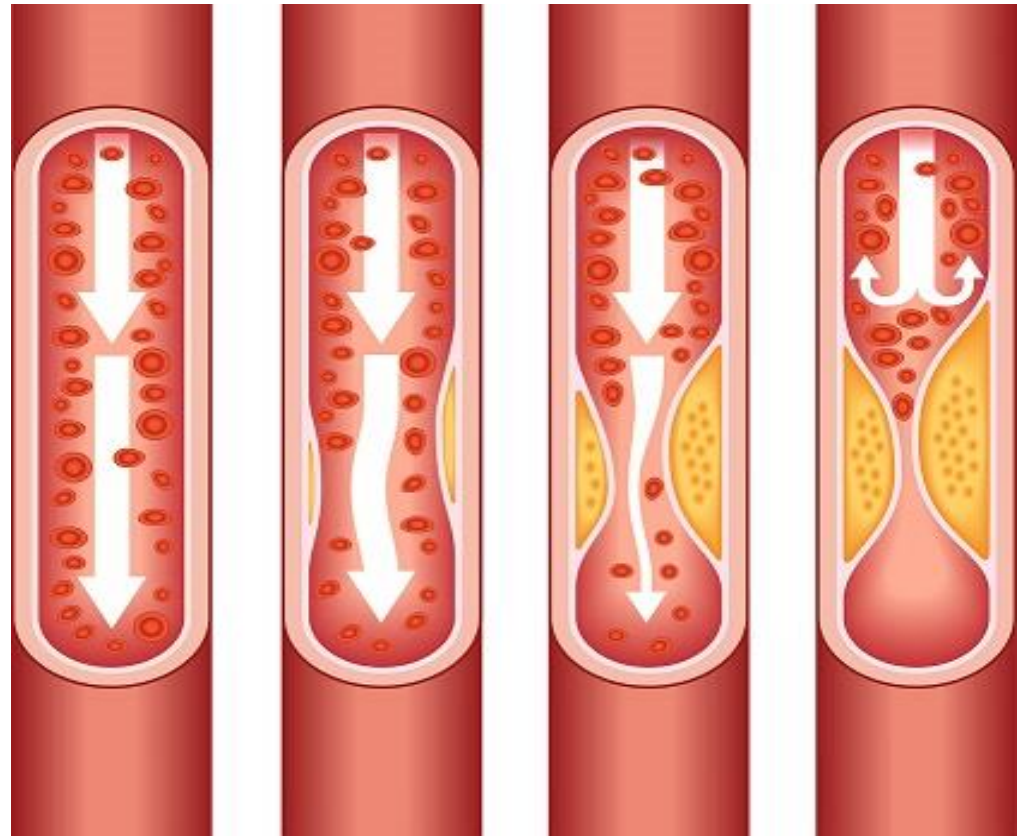


Figure 1. Plaque buildup

CDC Fast Stats: Heart Disease

Retrieved from <https://www.cdc.gov/heartdisease/facts.htm>

HEART DISEASE

SPENDING

2015 Spending:

\$318 billion (direct)

\$237 billion (indirect)

2035 Projections:

\$749 billion (direct)

\$368 billion (indirect)

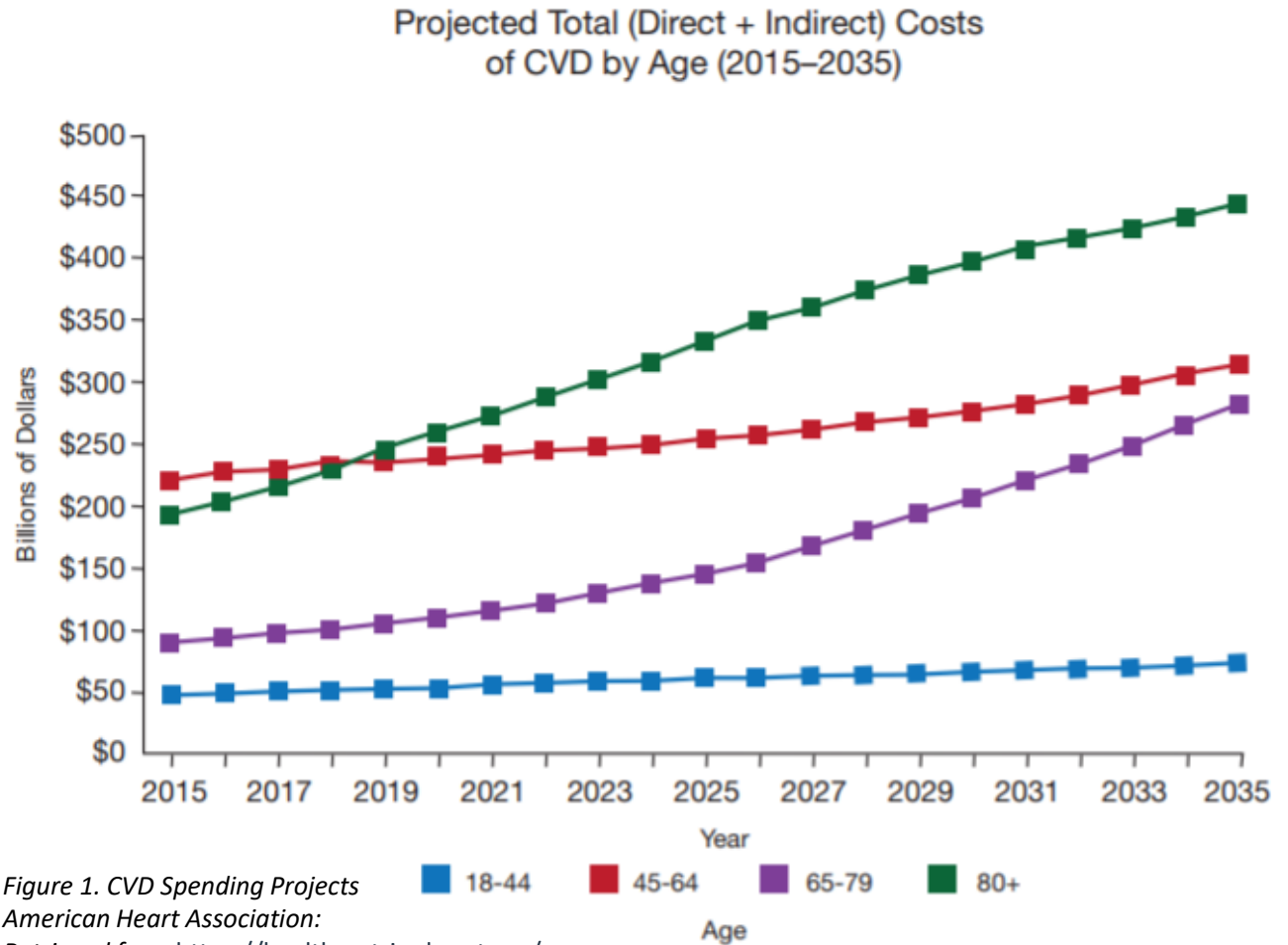


Figure 1. CVD Spending Projects

American Heart Association:

Retrieved from [https://healthmetrics.heart.org/wp-](https://healthmetrics.heart.org/wp-content/uploads/2017/10/Cardiovascular-Disease-A-Costly-Burden.pdf)

[content/uploads/2017/10/Cardiovascular-Disease-A-Costly-Burden.pdf](https://healthmetrics.heart.org/wp-content/uploads/2017/10/Cardiovascular-Disease-A-Costly-Burden.pdf)

PROBLEM STATEMENTS

PROBLEM STATEMENTS

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:

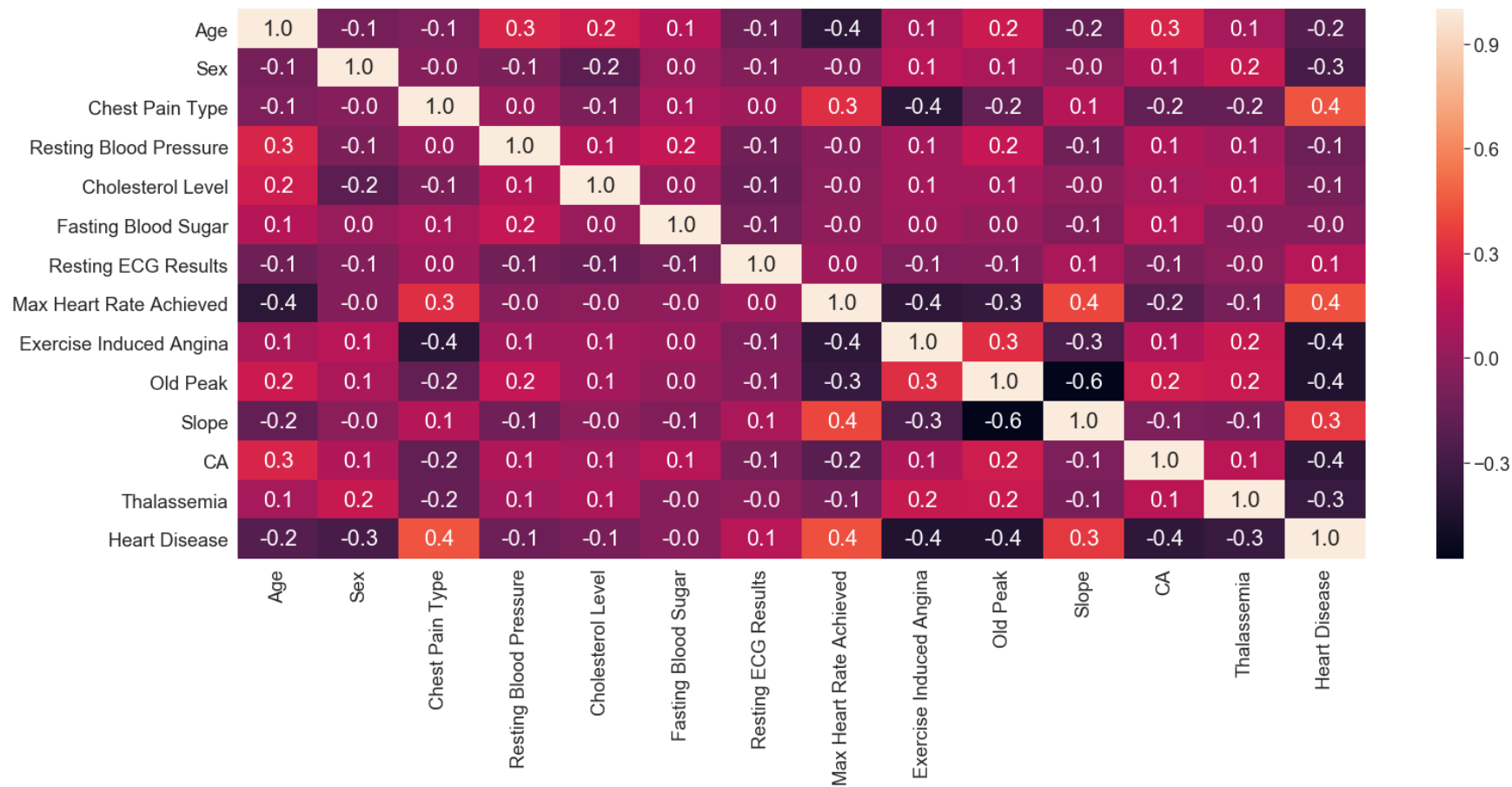
1. Can we reduce heart disease with machine learning?
2. If we can predict heart disease with ML, could we decrease spending?



TRENDS

DATA

1025 observations and 14 variables



MODELS

LOGISTIC REGRESSION

Baseline: 51% accuracy

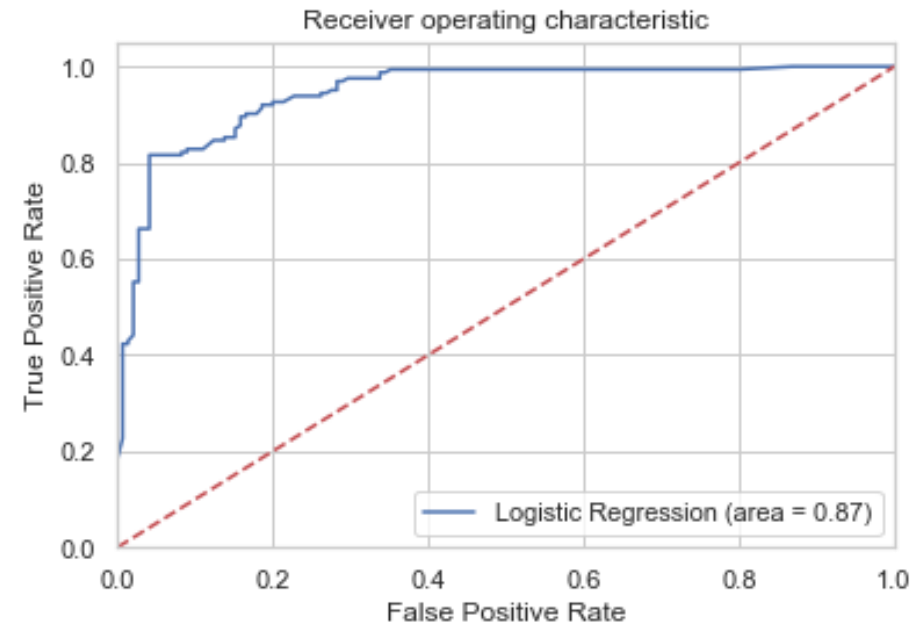
Features:

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%

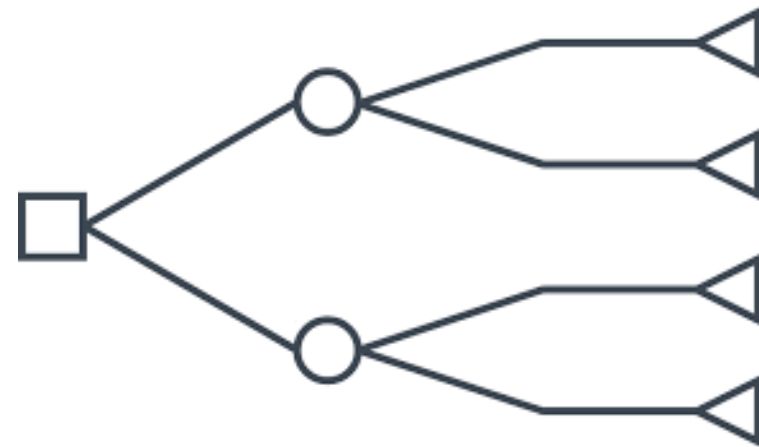


	precision	recall	f1-score	support
0	0.88	0.84	0.86	145
1	0.86	0.90	0.88	163
accuracy			0.87	308
macro avg	0.87	0.87	0.87	308
weighted avg	0.87	0.87	0.87	308

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

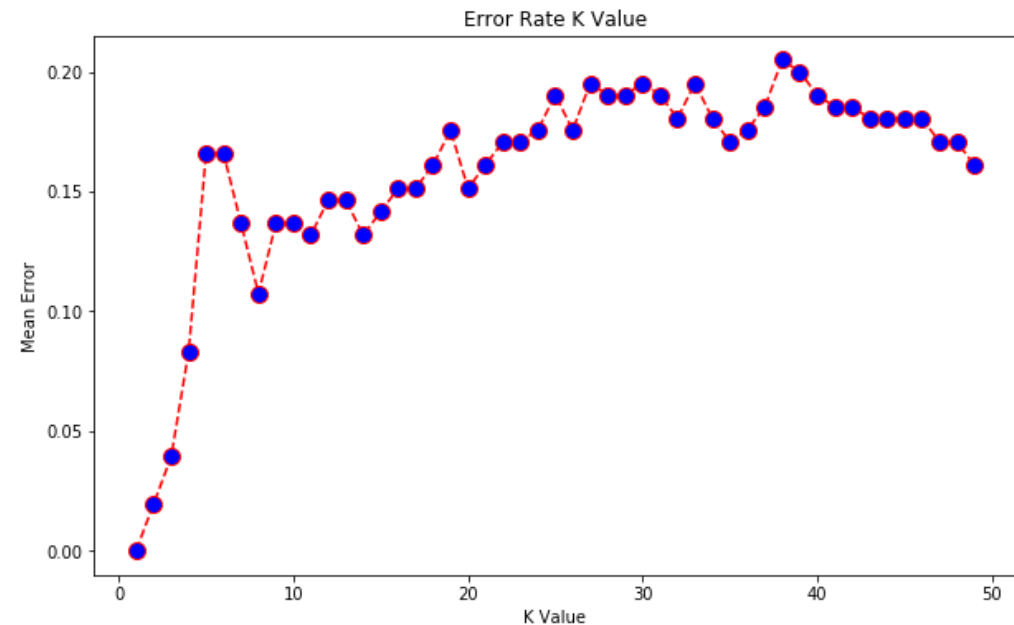
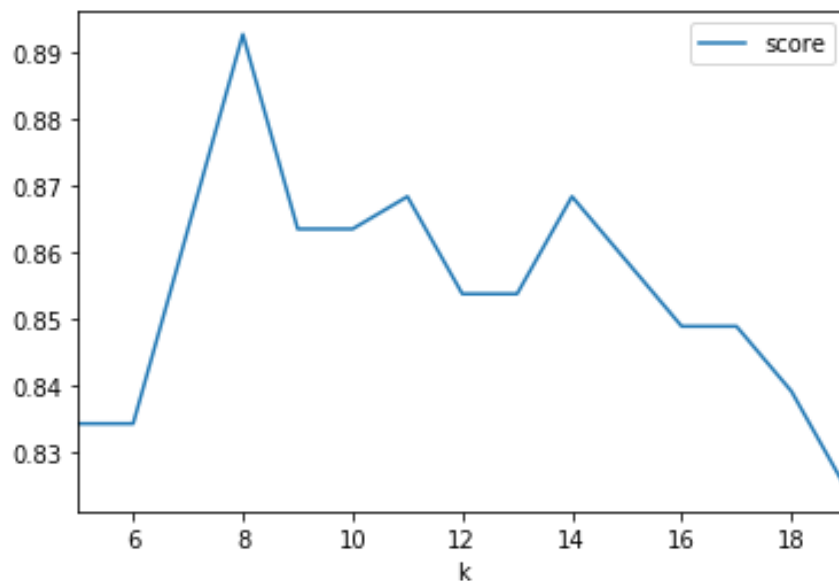
Baseline: 51% accuracy

Results:

Precision score of 89%

Accuracy of 89 %

	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'           : 1,
    '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'        : 0,
    'slope'          : 0,
    'ca'             : 0,
    'thalassemia'    : 1
}, index = [0])
```

new_patient

	Age	Sex	Chest Pain Type	Resting Blood Pressure	Cholesterol	Fasting Blood Sugar	Resting ECG	Max Heart Rate	Exercise Induced Angina	oldpeak	slope	ca	thalassemia
0	24	1	0	200	150	1	0	185	1	0	0	0	1

```
print(knn.predict_proba(new_patient).round(2))
print(knn.predict(new_patient))
```

```
[[0.58 0.42]]
[0]
```

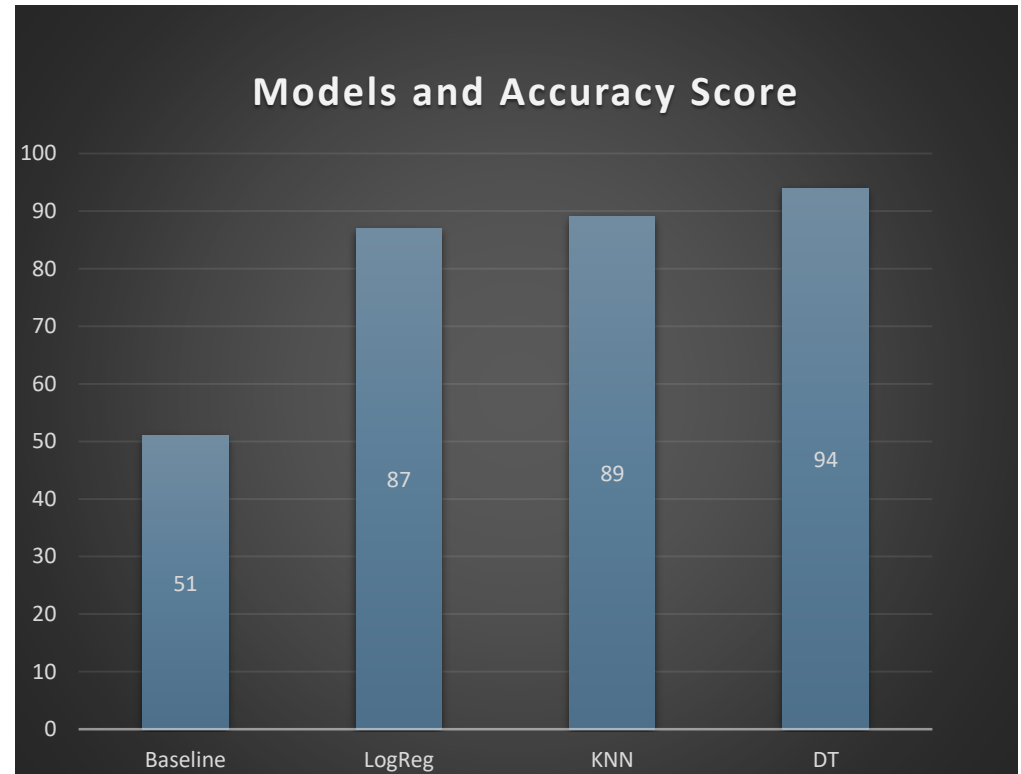
No Heart Disease!

CONCLUSIONS

CONCLUSIONS

NEXT STEPS

- *Feature Engineering*
- *Looking at the complete data set of 74 variables*
- *Can this be applied to bigger sets of data?*



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 THROUGH 2035" American Heart Association <https://healthmetrics.heart.org/wp-content/uploads/2017/10/Cardiovascular-Disease-A-Costly-Burden.pdf>

THANK YOU

*Alexander Krasniewski
Krasniewski_alexander@bah.com
240-888-7423*