You Are What You Eat:

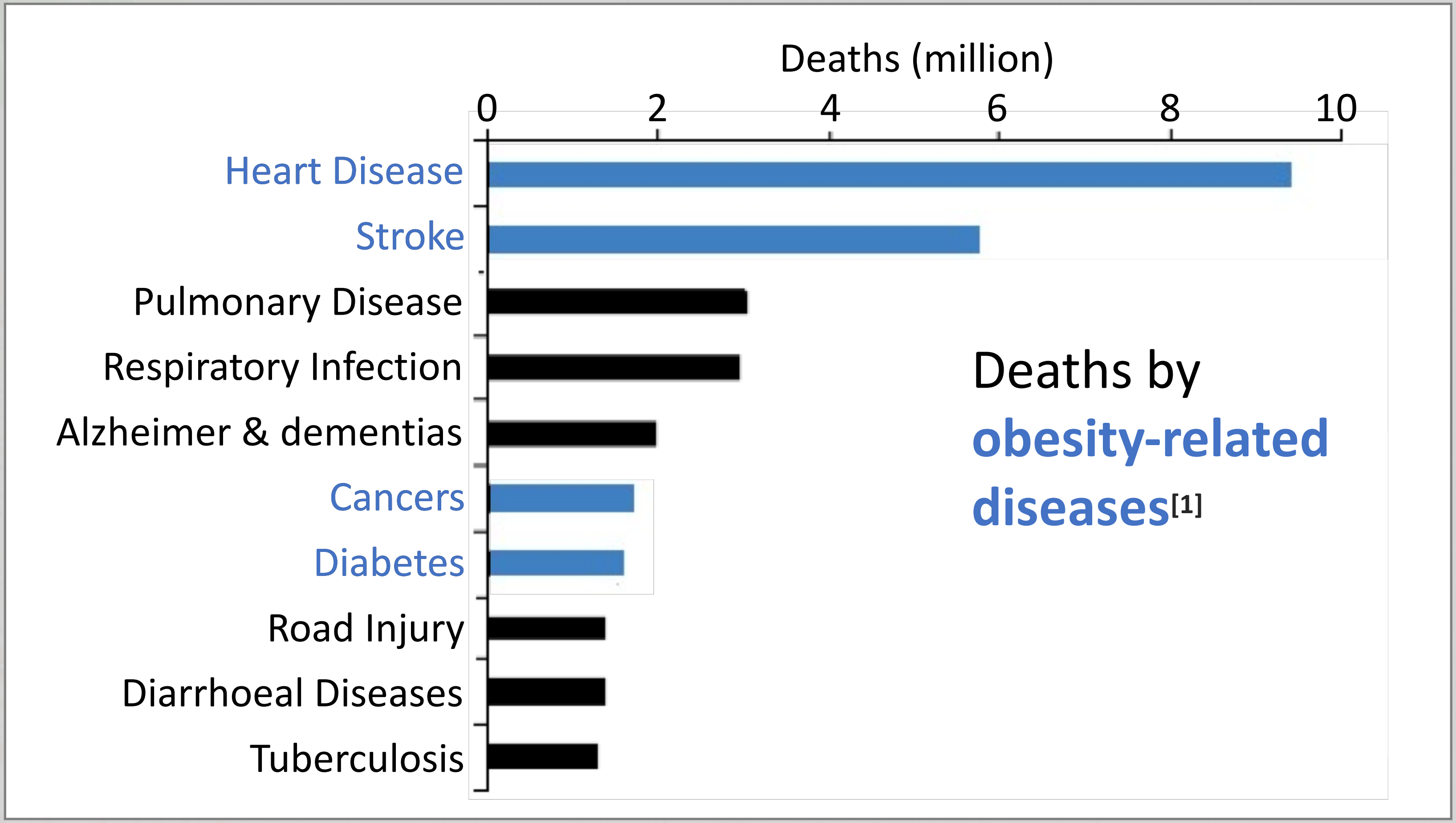
Obesity Classification Based on Weekly Habits

Project-2 @ **METIS**

Jhonsen Djajamuliadi

Obesity-Related Diseases

Top 10 Causes of Death Worldwide



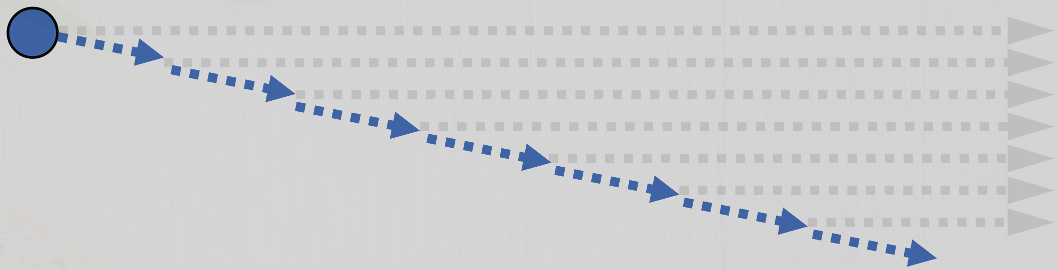
Global Health Estimates 2016, from WHO, 2018

[1] *Obesity related diseases* https://www.cdc.gov/obesity/adult/causes.html

Project Motivation & Theme

Small changes ∝ Big impact

today *future*



Charles Eugster, 97

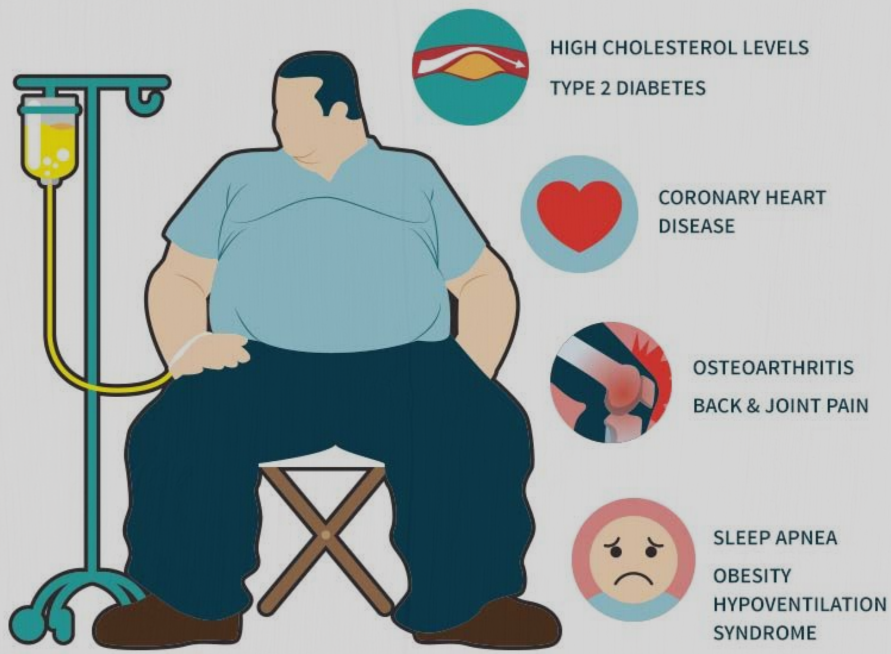
https://www.agebrilliantly.org/charles-eugster-champion-elderly-world-record/

Obese | **Not Obese**

Disease *Healthy*

Question:

What small changes can I make, today?



https://www.pinterest.com/pin/367606388318247552/

Data-Driven Investigation

American Time Use Survey (ATUS)
Eating & **Health Module** Microdata Files

BLS
BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR

6 years
(2006-2008, 2014-2016)

Feature Selection

37 Questions and Answers

BMI

"User-Friendly" Features

- Actionable, on a weekly basis
- Doable immediately
- No privacy concerns

Time_Eat
Time_Snack
Exercise_Freq
Fastfood_Freq
General_Health

Features of Importance

Obese

30

Not Obese

Target

Data-Driven Investigation

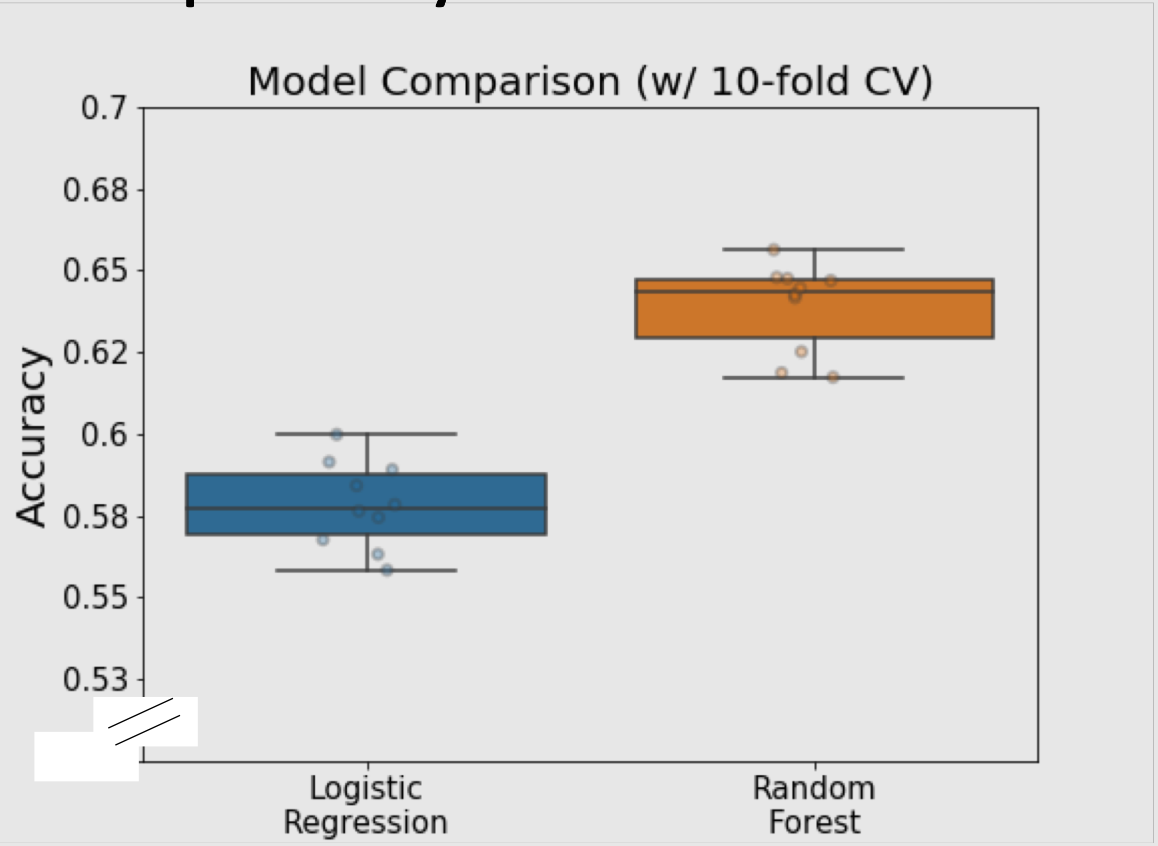
American Time Use Survey (ATUS)
Eating & **Health Module** Microdata Files

BLS
BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR

Feature Selection

Model Selection

Interpretability vs. **Predictive Power**



- Misclassifying a diseased-patient is costly

Data-Driven Investigation

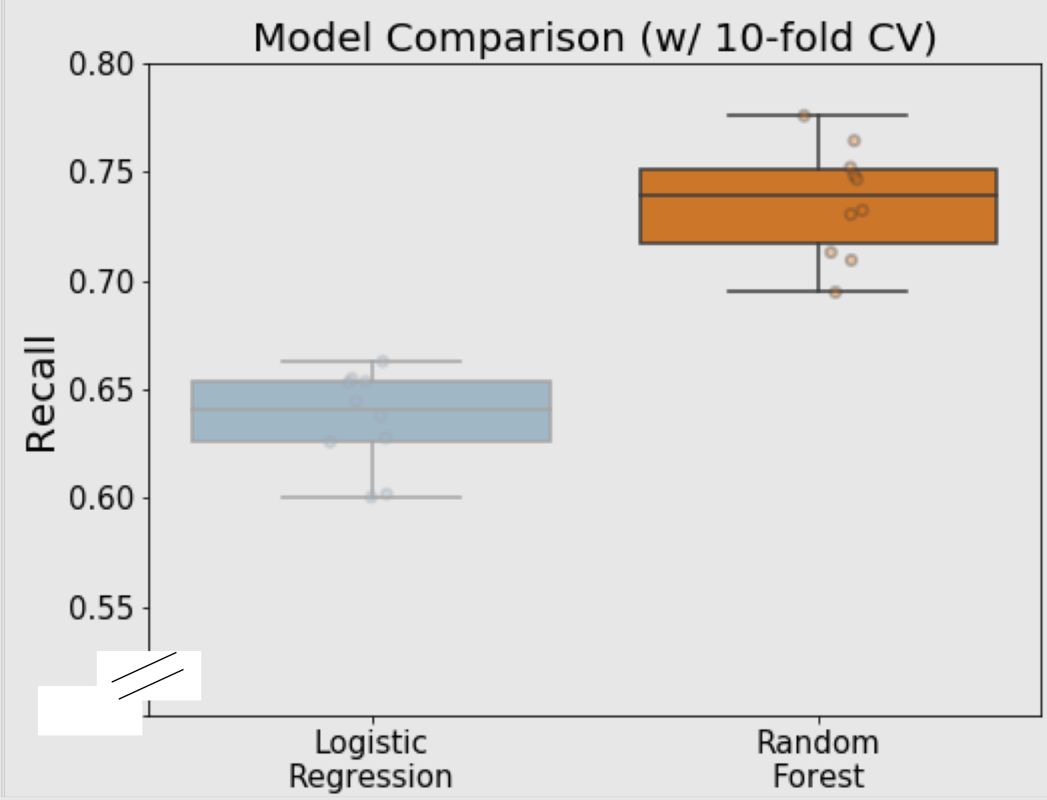
American Time Use Survey (ATUS)
Eating & Health Module Microdata Files

BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR

Feature Selection

Model Selection

Interpretability vs. **Predictive Power**



Data-Driven Investigation

American Time Use Survey (ATUS)
Eating & Health Module Microdata Files

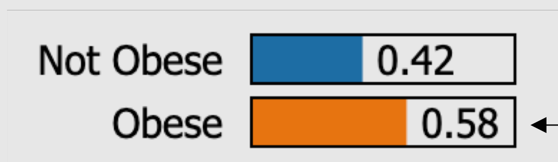
BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR

- Feature Selection
- Model Selection
- Evaluation
- Deployment

Interpretability & **Predictive Power**

Local
Interpretable
Model-agnostic
Explanations

Case# *1902*



TRUTH : Obese

Variation of Features

- $\uparrow$ *exerciseFreq* (0 $\rightarrow$ 1 a day)
  - P(Obese) = **0.51**
- $\downarrow$ *genHealth* (Good $\rightarrow$ Poor)
  - P(Obese) = **0.62**
- $\downarrow$ *genHealth*, $\uparrow$ *exerciseFreq*
  - P(Obese) = **0.57**

LIME references:
- https://github.com/marcotcr/lime
- https://arxiv.org/abs/1602.04938
- https://www.oreilly.com/learning/introduction-to-local-interpretable-model-agnostic-explanations-lime

Summary & Future Work

Project Challenges:
- Survey data (unreliable)
- BMI for obesity?

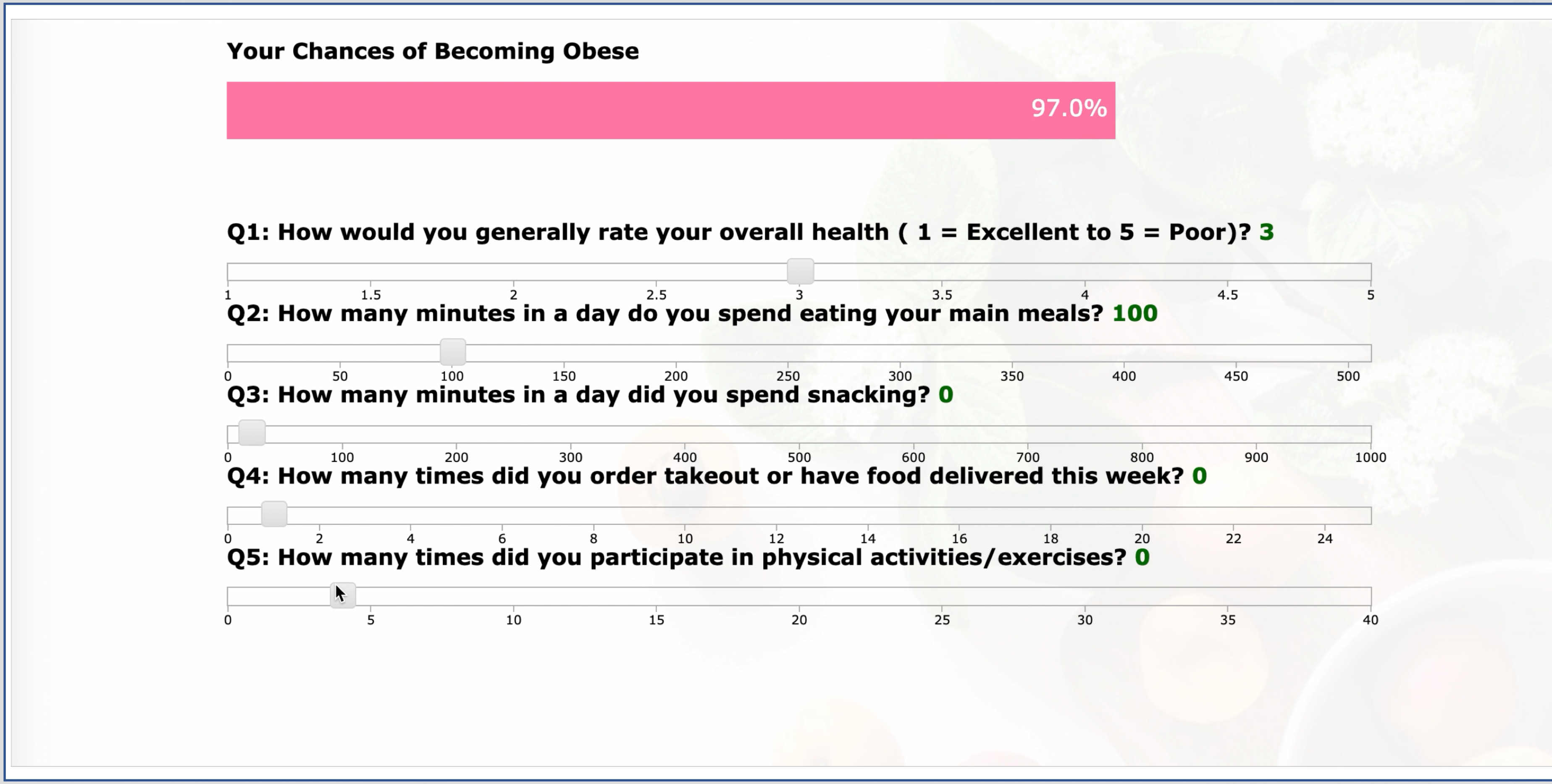
Completed:
- Model with predictive power
- Interpretability

Room for Improvement:
- Test set accuracy **~0.63**
- Test set recall **~0.69**

Future Work:
1. Retrain model with **more data**
2. Adding relevant **features**
3. Other dataset (non-survey)

Obesity Prediction Web App

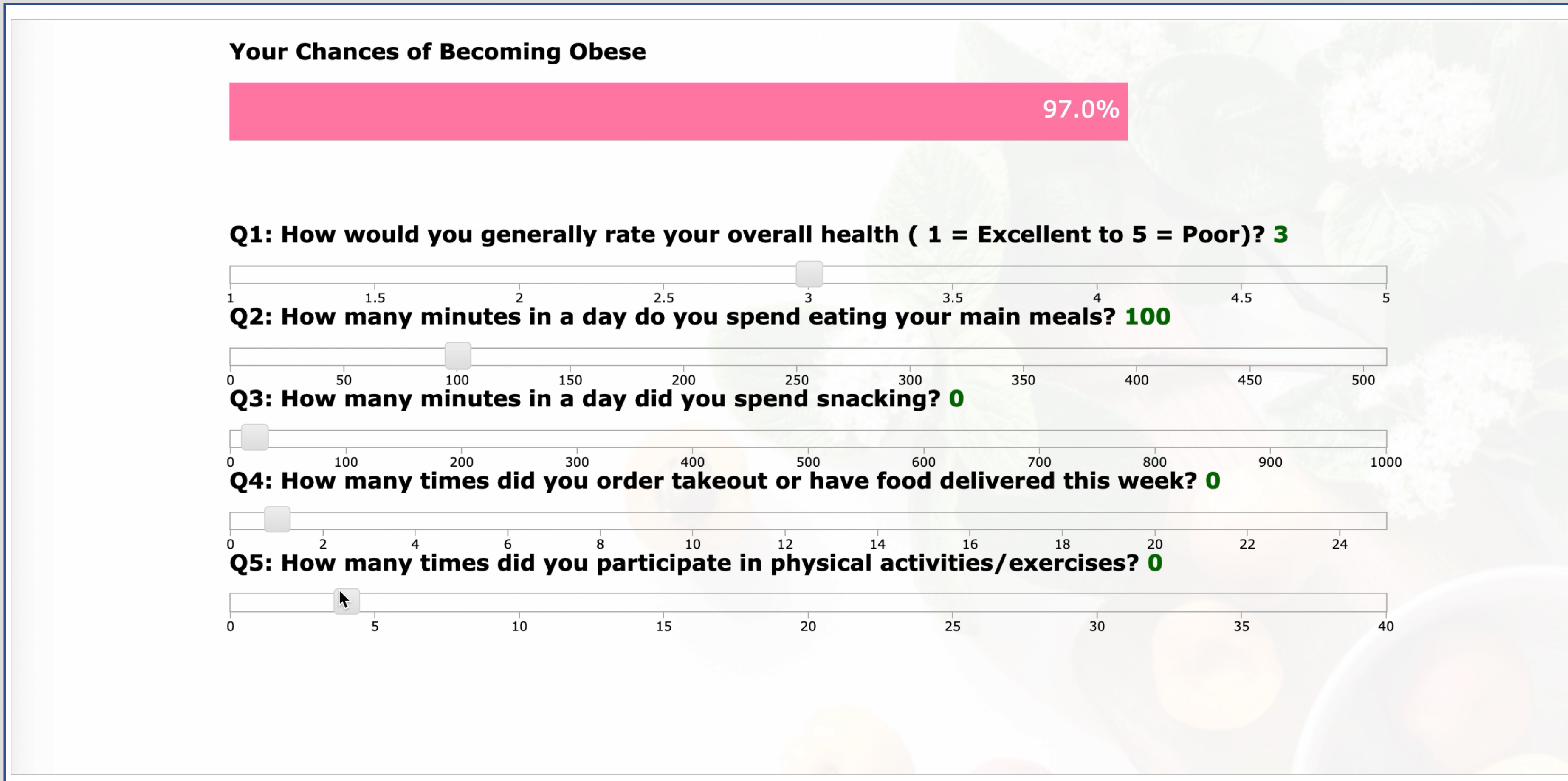
https://obesity-predictor.herokuapp.com/



+ + Flask →

Obesity Prediction Web App

https://obesity-predictor.herokuapp.com/



Question:

- *What small changes can I make, today?*

Hints:

- *Make adjustments to our weekly habits*

Thank You

Data Scientist | Chemist

jhonsen

jhonsen

jhonsen_dj

Project Description and Goal

Data Science
for Social Good

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

FDA

HSC Public Health Agency

World Health Organization

orangetheory FITNESS

24 HOUR FITNESS

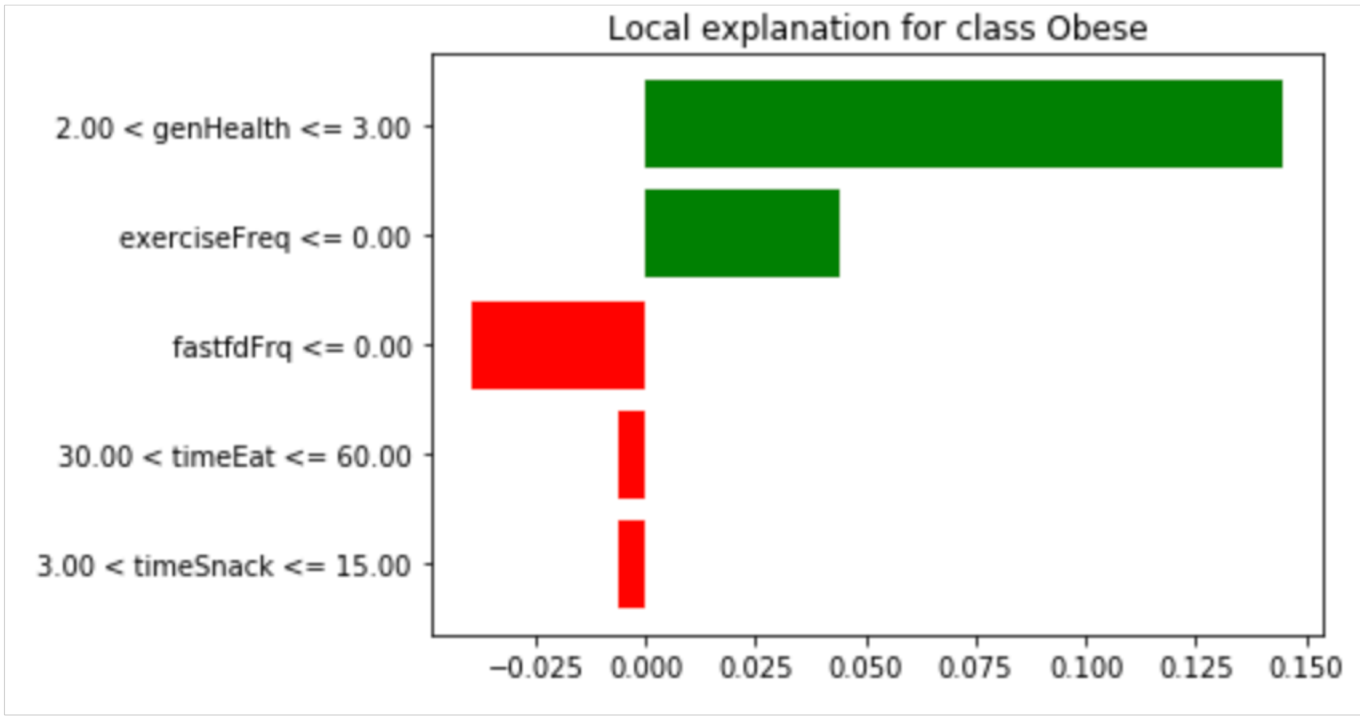
Deployment

Photos from https://www.pexels.com/

Bringing Interpretability to "Black-Box"

<u>Appendix to slide #7</u>

- The impact of each variable on P(Obese)



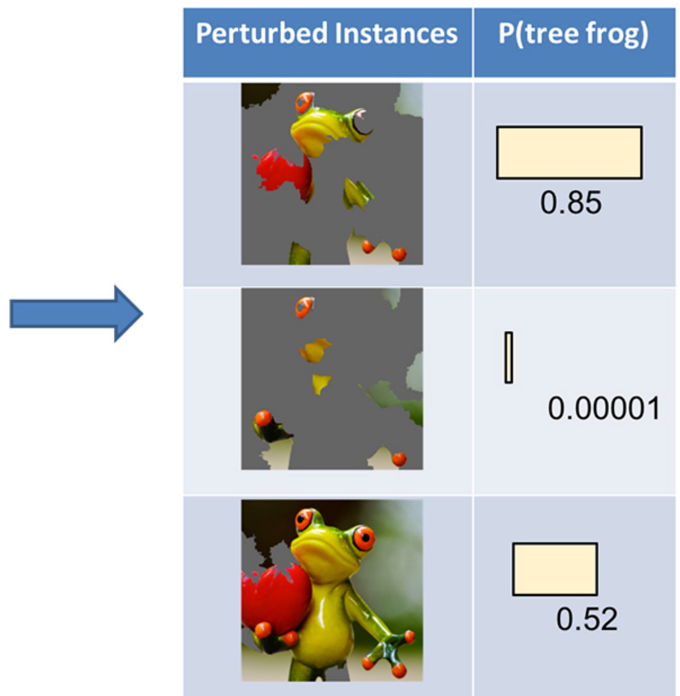
Not Obese | Obese

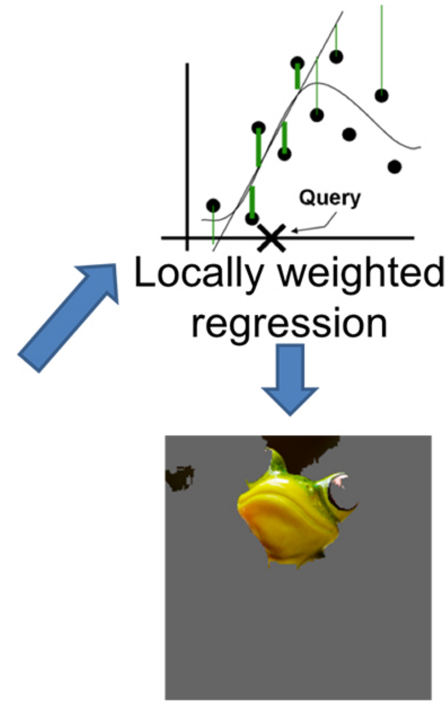
Bringing Interpretability to “Black-Box”

<u>Appendix to slide #7</u>

- Quick blurb about **LIME**

Original Image
P(tree frog) = 0.54





Explanation

Interpretable
Components

LIME references:

- https://github.com/marcotcr/lime
- https://arxiv.org/abs/1602.04938
- https://www.oreilly.com/learning/introduction-to-local-interpretable-model-agnostic-explanations-lime

Bringing Interpretability to “Black-Box”

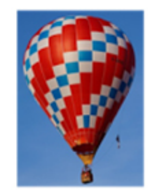
<u>Appendix to slide #7</u>

- Quick blurb about **LIME**

Local segments of tree-frog picture

P() = 0.54

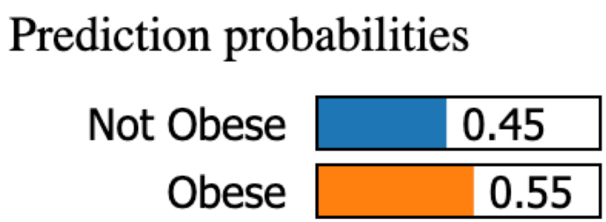
P() = 0.07

P() = 0.05

Strange *Feature-Target* Pair

Appendix to slide #8

```
row number:  1742
predicted: 1, truth: 1. Therefore, the classifier is CORRECT
```



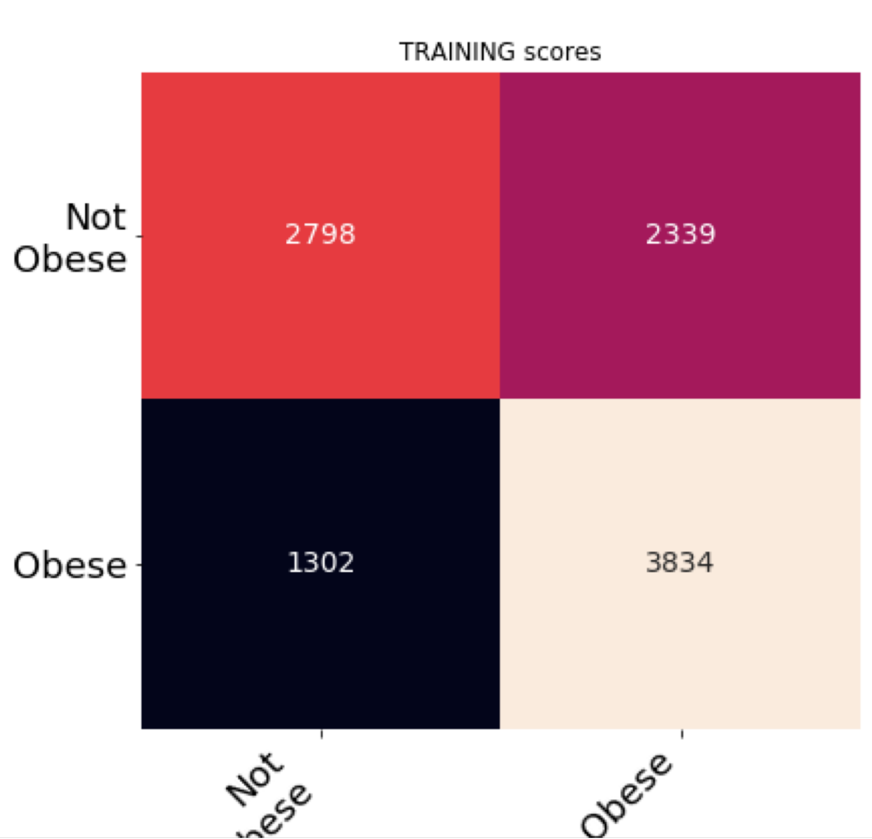
Feature	Value	
genHealth	2.00	(Very Good)
exerciseFreq	3.00	
timeEat	120.00	(2 hrs)
timeSnack	0.00	
fastfdFrq	1.00	

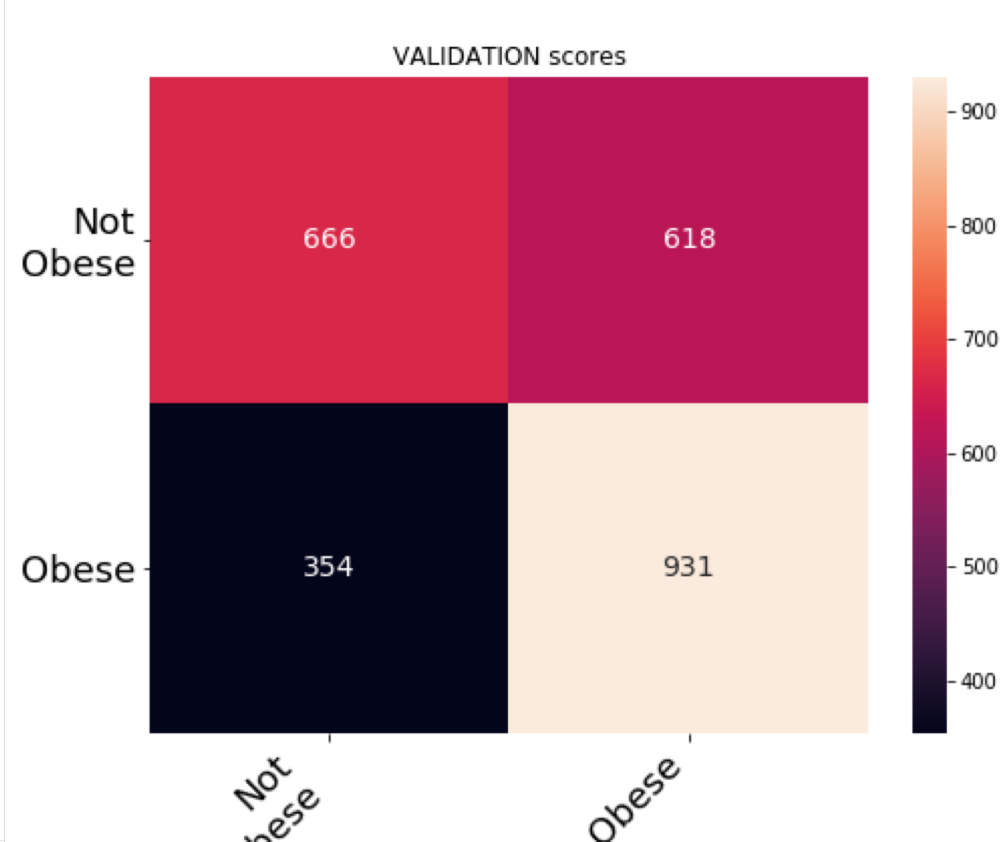
- *Isn't this person generally healthy?*

Confusion Matrix for Train-Val Set

Appendix to slide #6

- Performance of models based on Training Set





ROC Curve of Training-Val Set

<u>Appendix to slide #6</u>

- Performance of models based on Training Set
- **Random Forest** model