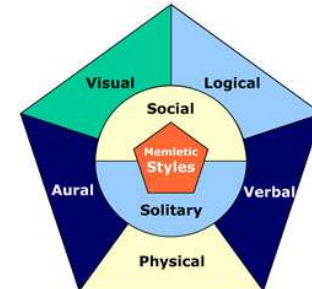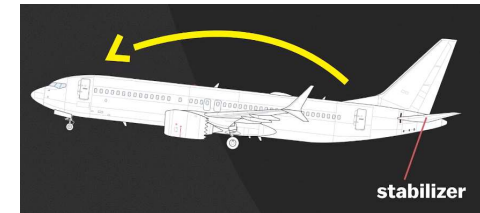# Teaching Practical Data Science via an Interactive Web-Based Tool

(or "Why Physicians Need to Understand Models")

John Kang

University of Rochester
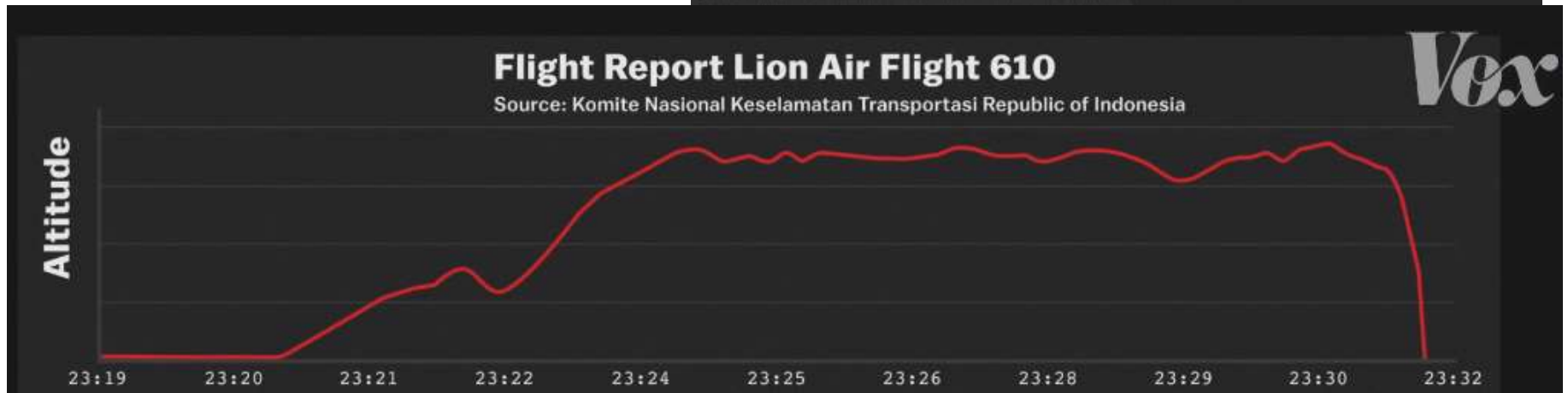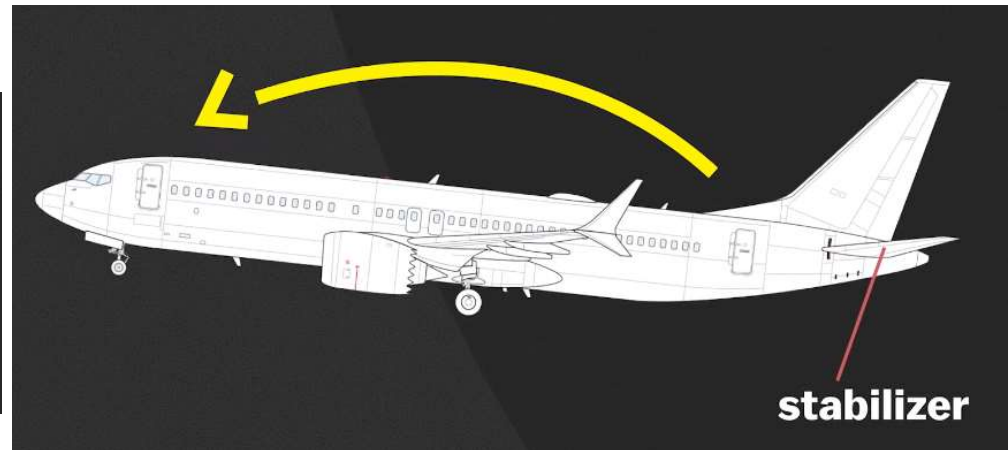
ROECSG 2019

john_kang@urmc.rochester.edu

Outline

- Why AI needs to be trustworthy?

- How can physicians improve this trustworthiness?

- How do we teach physicians the fundamentals of trustworthy models?

How trustworthy is the AI "black box"?

MCAS — **M**aneuvering **C**haracteristics **A**ugmentation **S**ystem

stabilizer

Flight Report Lion Air Flight 610
Source: Komite Nasional Keselamatan Transportasi Republic of Indonesia

Vox

"The real reason Boeing's new plane crashed twice", Vox. https://www.youtube.com/watch?v=H2tuKiiznsY

Is Rad Onc [...] physicians

Boeing

#737MAX
#OSH16

where there has been reference
to "self-certification."

Rep. Garret Graves (R-La.)

SUBSCRIBE

FAA chief grilled again over Boeing's 737 Max crisis

25,447 views

👍 191   👎 30   ➔ SHARE   ☰＋ SAVE   ...

wp Washington Post ✓
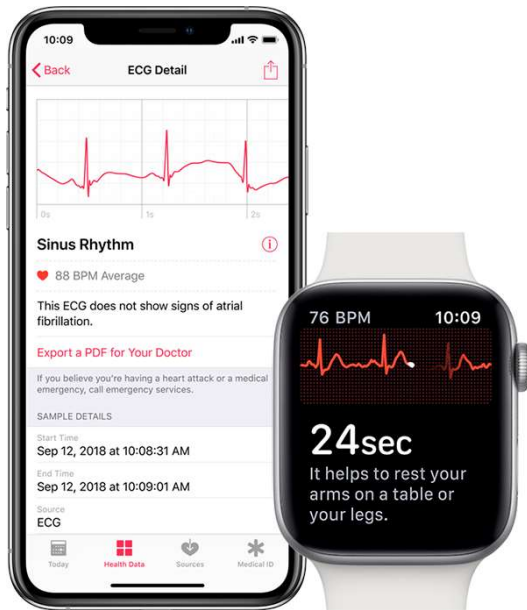Published on May 15, 2019

SUBSCRIBED 670K 🔔

Acting FAA administrator Daniel K. Elwell appeared before a House committee on May 15
following revelations that pilots complained to Boeing about the 737 Max months before the
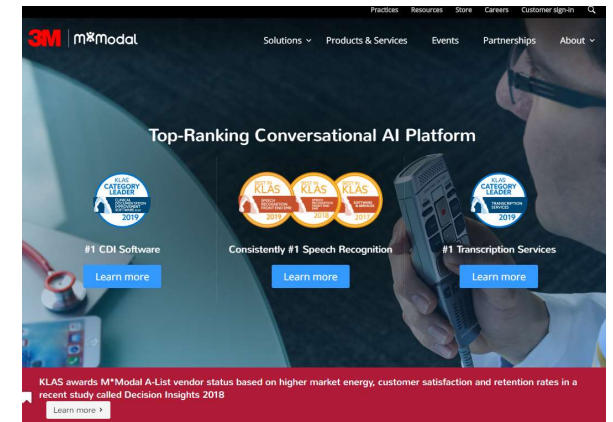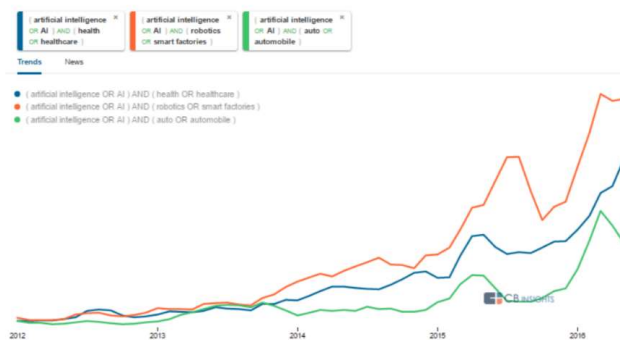Ethiopian Airlines crash. Read more: https://wapo.st/2JHB1A3. Subscribe to The
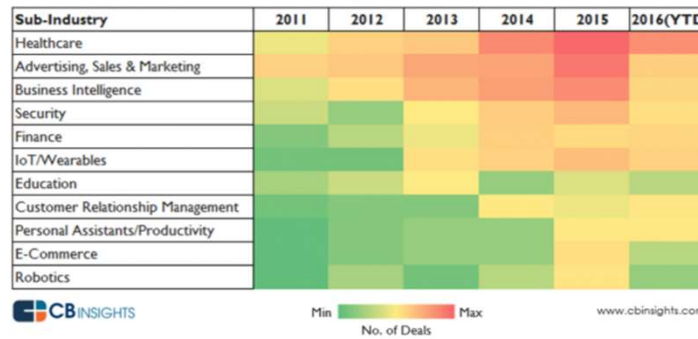
SHOW MORE

AI in medicine

Good performance ≠ trustworthy
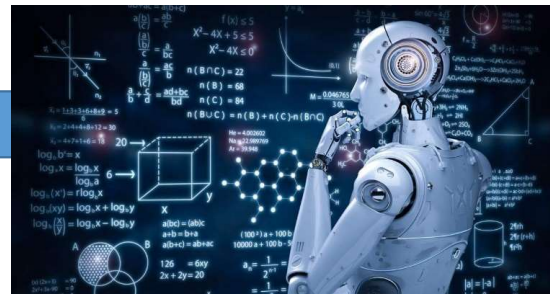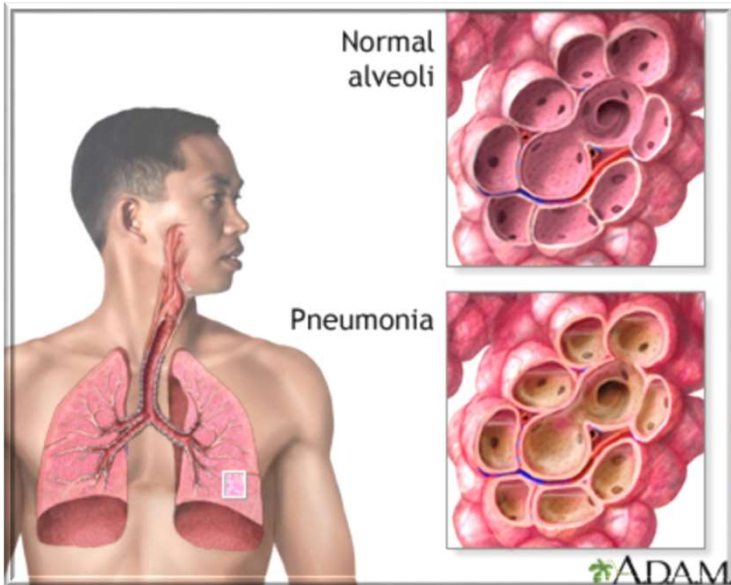
Making intelligence intelligible with Dr. Rich Caruana. Microsoft Research. May 2018.
https://www.microsoft.com/en-us/research/blog/making-intelligence-intelligible-dr-rich-caruana/.

How do we engage trainees in building trustworthy models

- Create awareness and proper conduct in AI
- Implement practical curriculum
- Create public resources
- Generate accelerated learning & funding opportunities

What makes a great educational tool?
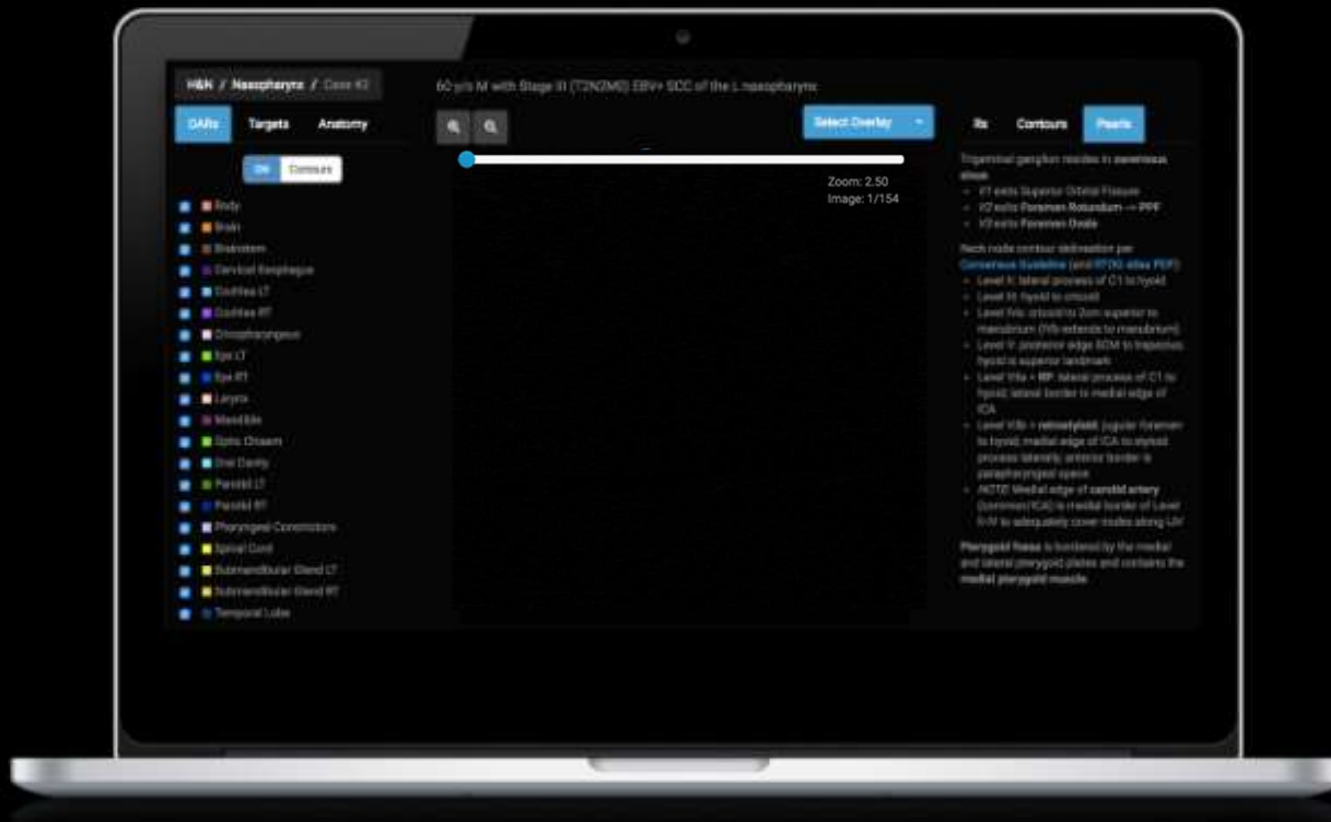
eContour

Interactive contouring guidelines for the busy radiation oncology professional

View Cases

Verbal Learning   Kinesthetic Learning   Visuospatial Learning

Verbal

Kinesthetic

Visuospatial

Runs in web browser

https://github.com/john-kang/DataScienceEducation/blob/master/StatsTeacher%20-%20Screening%20(separate%20figures).ipynb

Generate balanced clusters of training data of 100 patients

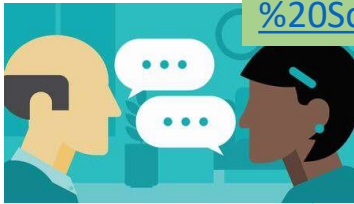**Q1**:How many samples to generate for each class (cancer, healthy)?

A1: n_samples = ___

```python
[356]: axis_names = ['protein A', 'protein B']
       class_names = ['cancer', 'healthy']
       cluster_std = 0.75
       random_state = 3

       X_D2, y_D2 = make_blobs(n_samples = [50,50], n_features = 2, centers = [[0.5,0.5], [2,2]],
                               cluster_std = cluster_std, random_state = random_state)
       y_D2 = y_D2 % 2
       plt.figure()
       plt.title('Simulated protein A and B co-expression for healthy (yellow) and cancer (purple) patients')
       plt.scatter(X_D2[:,0], X_D2[:,1], c=y_D2, marker= 'o', s=50)
       plt.xlabel(axis_names[0])
       plt.ylabel(axis_names[1])
       plt.show()
```

Simulated protein A and B co-expression for healthy (yellow) and cancer (purple) patients

Without the data, can we replicate the idea?

Generate balanced clusters of training data of 100 patients

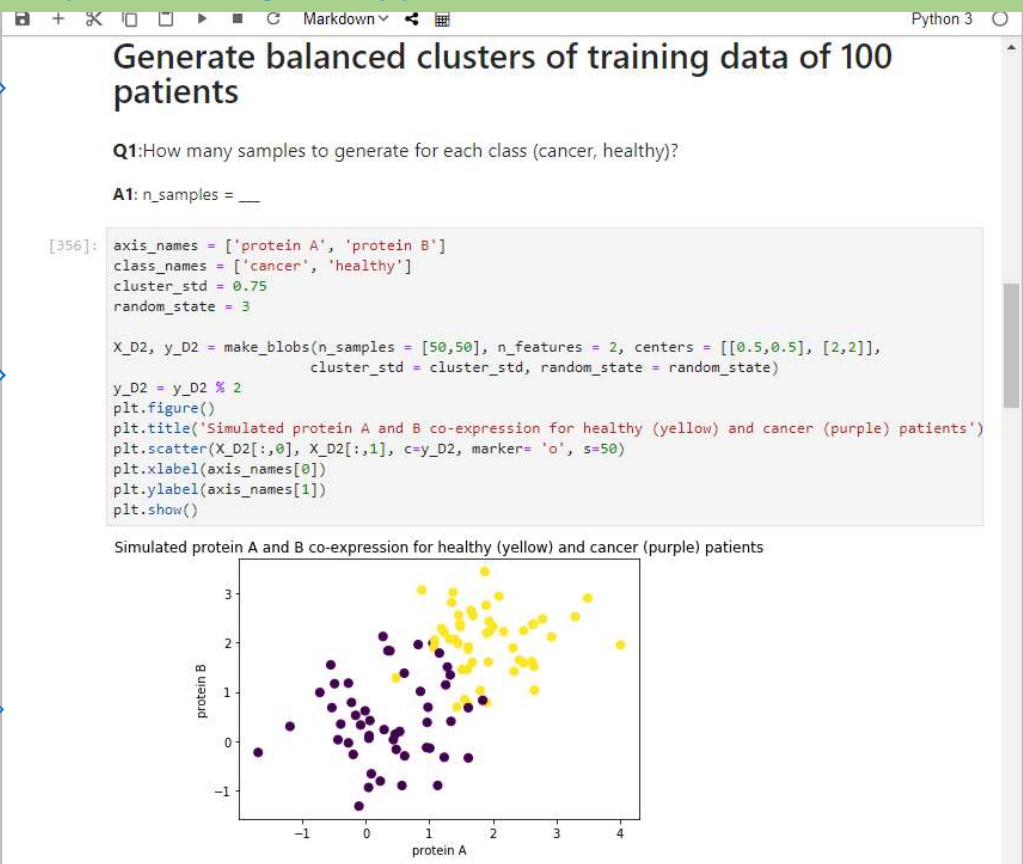Tr... al... training set of 100 patients

What happens if we vary k?

Tuning kNN clustering

How do evaluate our kNN model?

Training set (75%) p...

Test set (25%) set performance:

Generate an independent validation set of

Thank you for your attention

- How do you plan to gauge the trustworthiness of a model?