LIFE AND DEATH DECISIONS:

TESTING DATA SCIENCE

WHY TESTING IS IMPORTANT

Elon Musk ✔
@elonmusk

Owner video of Autopilot steering to avoid collision with a truck[m.youtube.com/watch?feature=…](m.youtube.com/watch?feature=…)

12:34 AM - Apr 18, 2016

♡ 5,579   💬 2,496 people are talking about this

4 . 1

Blog   Videos   Press   Customer Storie

A Tragic Loss

The Tesla Team   •   June 30, 2016

Be the firs
news, eve

We learned yesterday evening that NHTSA is opening a preliminary evaluation into the performance of Autopilot during a recent fatal crash that occurred in a Model S. This is the first known fatality in just over 130 million miles where Autopilot was activated. Among all vehicles in the US, there is a fatality every 94 million miles. Worldwide, there is a fatality approximately every 60 million miles. It is important to emphasize that the NHTSA action is simply a preliminary evaluation to determine whether the system worked according to expectations.

example

Following our standard practice, Tesla informed NHTSA about the incident immediately after it occurred. What we know is that the vehicle was on a divided highway with

https://www.tesla.com/blog/tragic-loss

6

SINCE THEN

- 23 March 2018 - Tesla X crashed into a "poorly maintained crash attenuator"[1]
- 18 March 2018 - Uber Ford Fusion hit a pedestrian[2]

1. https://www.theguardian.com/technology/2018/mar car-crash-autopilot-mountain-view
2. https://www.theguardian.com/technology/2018/mar self-driving-car-kills-woman-arizona-tempe

CONTEXT

- Approximately 37,000 deaths per year source on US roads, where 93% of them can be attributed to human error source

How will driving change?

1. How fast will road travel grow?   ⓘ
1.0% annually ▾

2. When will AVs be introduced?
2025 ▾

3. How many years until full diffusion?   ⓘ
40 ▾

4. At the time of full adoption, what fraction of road travel be autonomous?
80% ▾

How will safety evolve?

Compared to today's drivers: ⓘ

5. How safe will non-AVs ultimately b...
just as safe (1.12) ▾

6. How safe will AVs be at introductio...
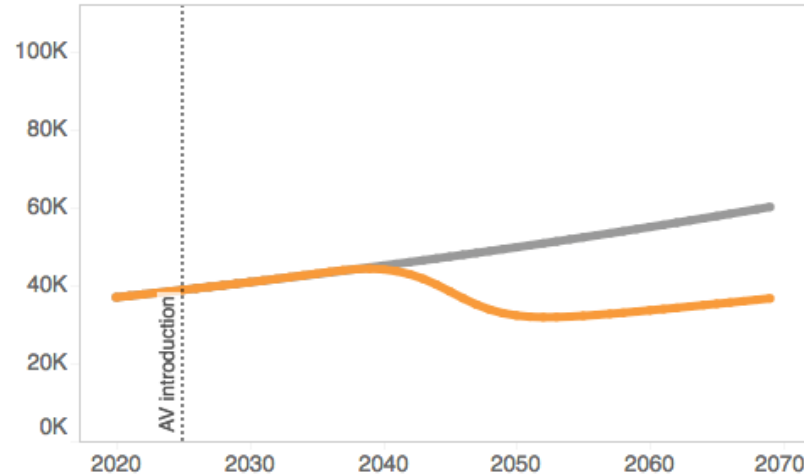half as safe (1.68) ▾

7. How safe will AVs ultimately be?
twice as safe (0.56) ▾

8. Will the change in the AV fatality rate occur quickly or slowly?   ⓘ
slowly ▾

Show: Annual Fatalities ▾

Future without AVs (gray)   Future with AVs (orange)
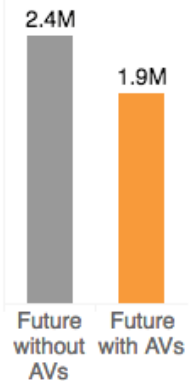
Total Fatalities

2.4M   1.9M

Future without AVs   Future with AVs

In this scenario, autonomous vehicles will be introduced into the marketplace in **2025**, when they are **half as safe** as current drivers. They will be fully adopted in **2065**, when they account for **80%** of miles traveled in the United States. Over this period, their safety **changes slowly**, so that by **2048** they are **twice as safe** as current drivers. Additionally, road travel will increase **1.0% annually** and non-AVs will be **just as safe** as current drivers by 2070.

As shown in the bar graph, a future without AVs would have **2.38 million** fatalities by 2070, while a future with AVs would have **1.87 million** fatalities. Over this time, AVs will have **saved 0.50 million** lives, compared to a future without AVs.

RAND Corporation via Tableau

SENSATIONALISM?

CORRECTIONAL OFFENDER MANAGEMENT PROFILING FOR ALTERNATIVE SANCTIONS (COMPAS)

2016

- Borden arrested for stealing an $80 childrens bike.
- Rated as **high** risk

--

- Prater arrested for stealing $86 worth of tools from Home Depot.
- Rated as **low** risk.
- Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica May 23, 2016
- Inspired by Cody Marie Wild via HackerNoon

Two years later:

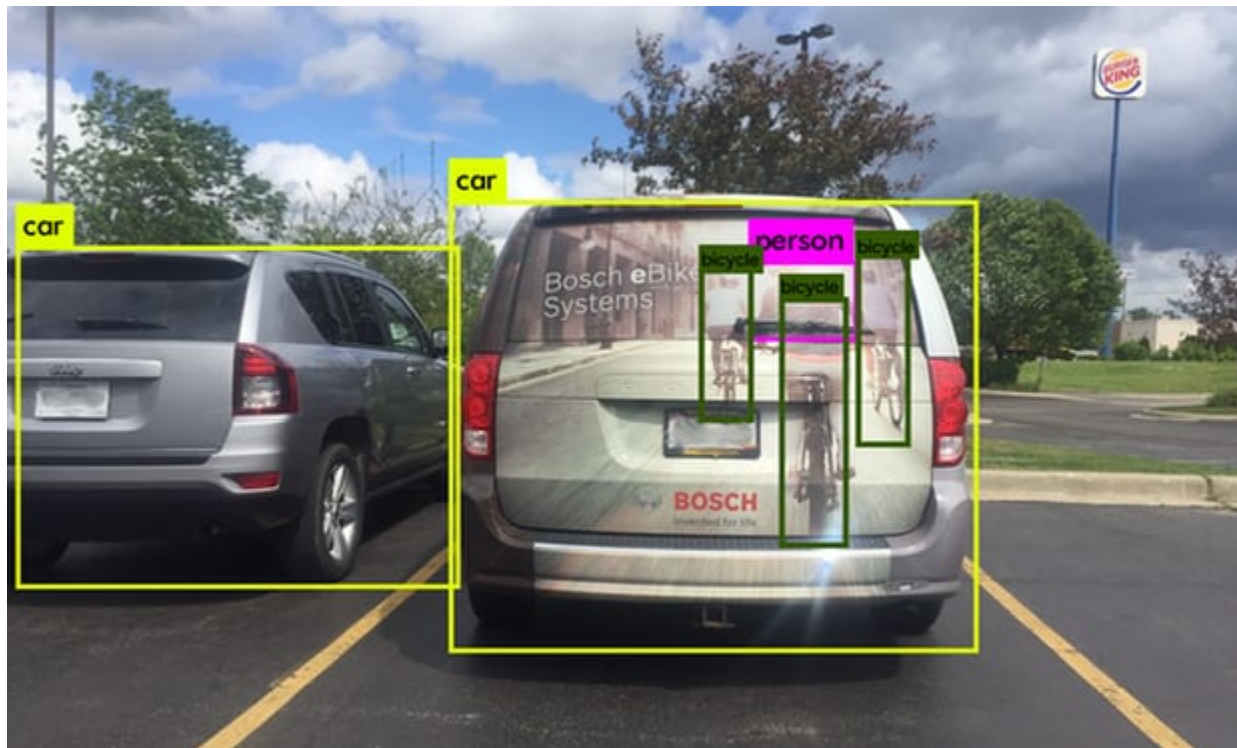- Borden was not charged with any other crimes.

--

- Prater is serving an eight-year prison term for stealing thousands of dollars' of electronics from a warehouse.

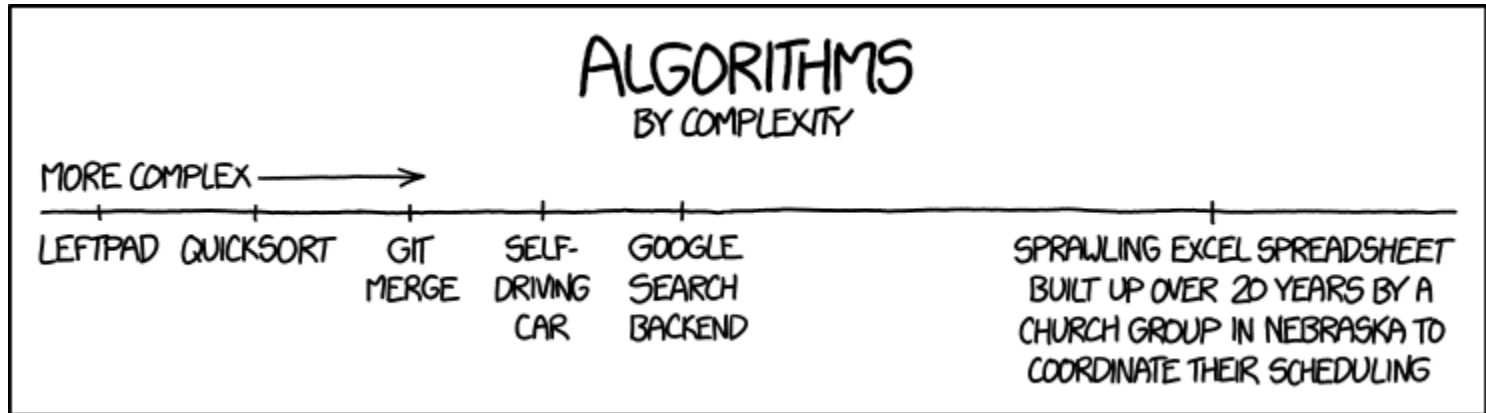- Borden was an **African American Female**

--

- Prater was a **Caucasian Male**
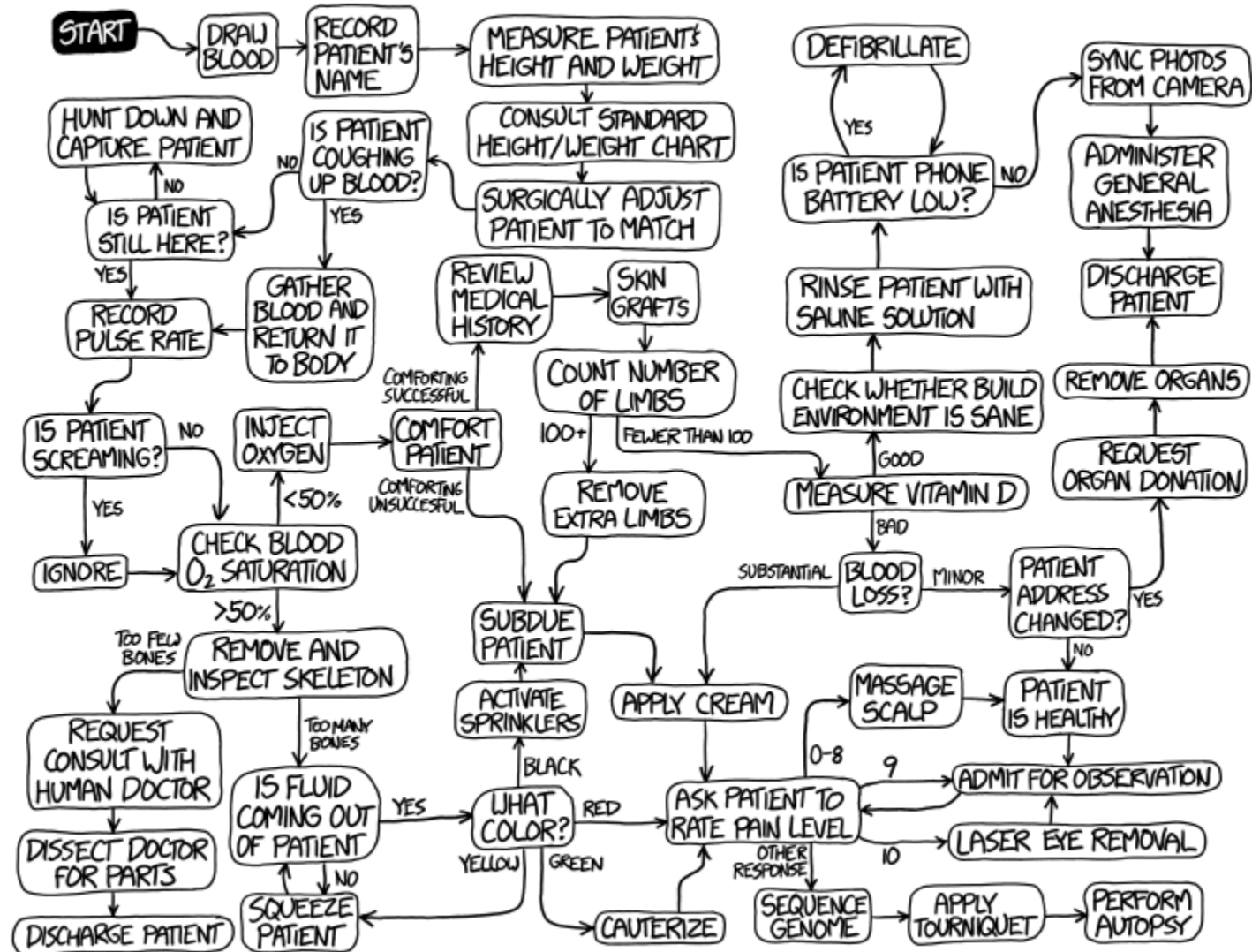
TESTING DURING TRAINING

The Guardian, Assume self-driving cars are a hacker's dream? Think again, 30 Aug 2017
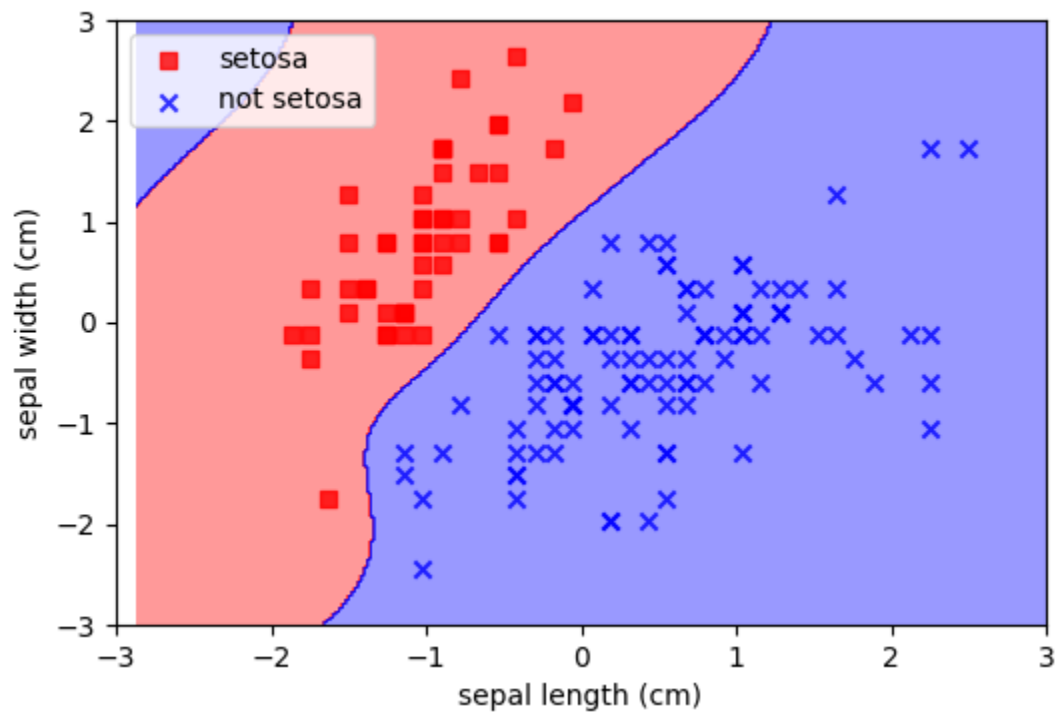
COMPLEXITY

XKCD - Algorithms

A GUIDE TO THE MEDICAL DIAGNOSTIC AND TREATMENT ALGORITHM USED BY IBM's WATSON COMPUTER SYSTEM

- XKCD - Watson Medical Algorithm
- Explanation of XKCD Watson Medical Algorithm

MORE DATA IS BETTER

METRICS

- Business metrics
- Technical metrics

TYPES OF PROBLEM

- Regression
- Clustering
- Classification

CONFUSION MATRIX

A confusion matrix is a table showing the degree of class confusion.

	actual positive	**actual negative**
predict true	true positive	false positive
predict false	false negative	true negative

UNBALANCED CLASSES

- Very common to have classes that are very rare.

Skews of $1 : 100$ are common in fraud detection, and skews of greater than $1 : 10^6$ have been reported in other applications.

- 1) Attenberg, J., & Provost, F. (2010). Why label when you can search?...

	positive	negative
yes	0	0
no	1	49

$$accuracy = \frac{\text{Number of correct decisions made}}{\text{Total number of decisions made}} =$$

After balancing classes:

	positive	negative
yes	0	0
no	25	25

$$accuracy = \frac{25}{50} = 50\%$$

NUMERICAL MODEL EVALUATION

- Every decision you make has some impact on the business.

We need to talk in the same language as the business. Profit/Loss.

EXPECTED VALUE

Marketing example:

- **true positive**: £50 - £9 = £41.
- **true negative**: Did not send, would not buy.
- **false positive**: Sent, did not buy: -£9
- **false negative**: Did not send, but would have bought. £0, but opportunity cost?

We can represent this cost benefit data, $b(\mathbf{x})$, in a matrix:

	p	n
Y	41	-9
N	0	0

VISUAL MODEL EVALUATION

- Numerical scores are necessary for optimisation.

 But visual representations are far more intuitive.
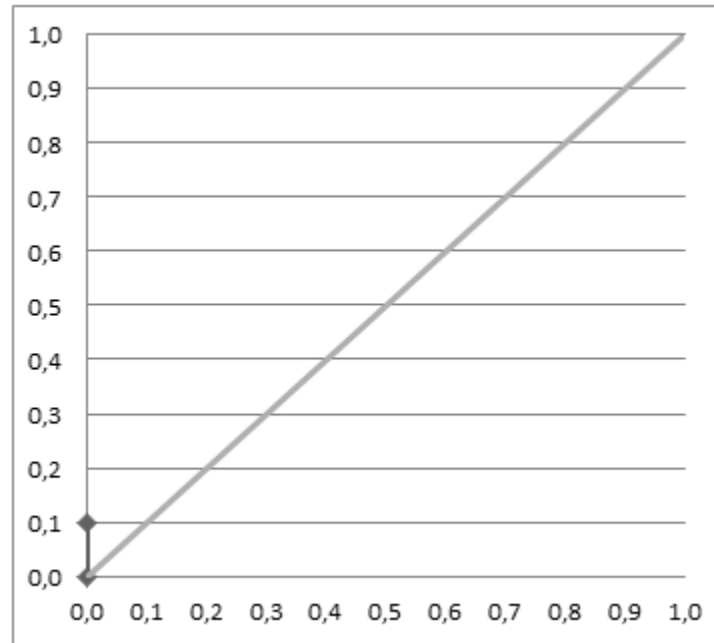
RANKING

- Most algorithms output some kind of score

For example:

- Probabilities
- Distances
- Errors

We can alter the *threshold* of that score to generate a new confusion matrix.
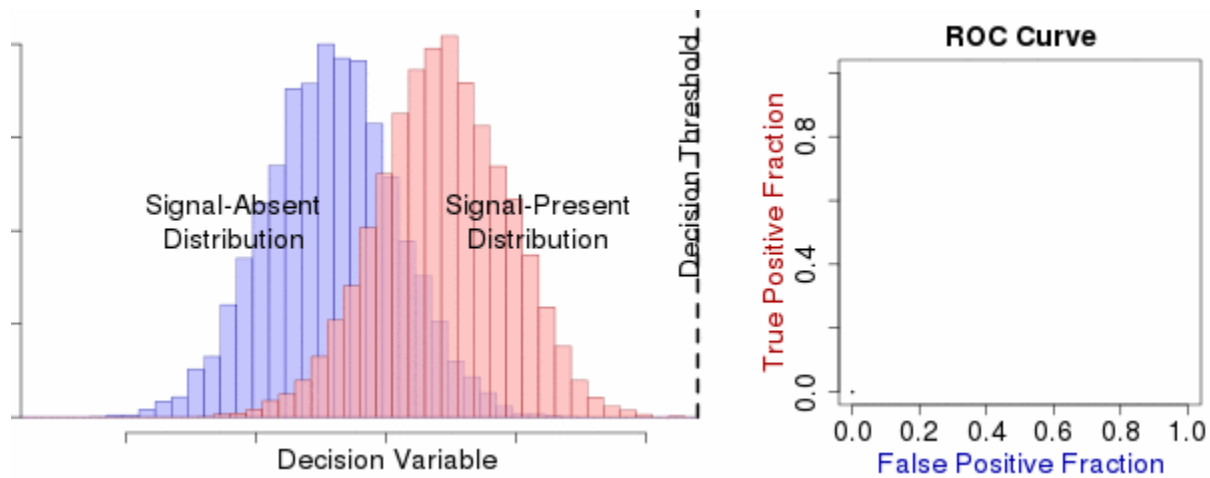
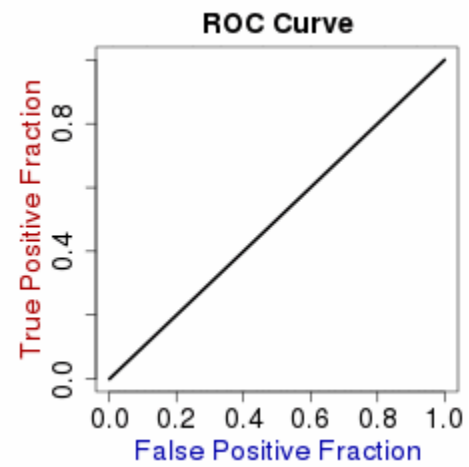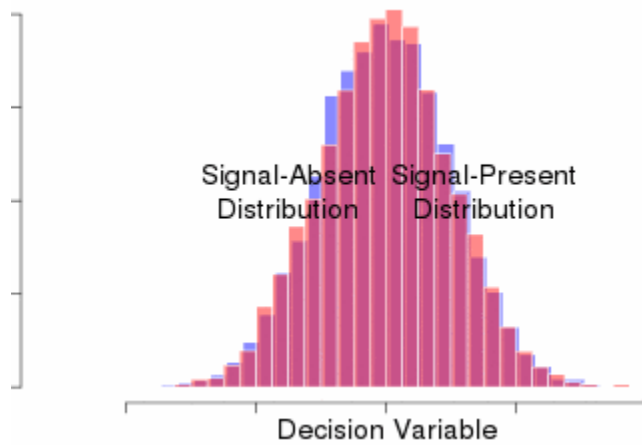| #	C	Score
1	P	0,9
2	P	0,8
3	N	0,7
4	P	0,6
5	P	0,55
6	P	0,54
7	N	0,53
8	N	0,52
9	P	0,51
10	N	0,505
11	P	0,4
12	N	0,39
13	P	0,38
14	N	0,37
15	N	0,36
16	N	0,35
17	P	0,34
18	N	0,33
19	P	0,3
20	N	0,1

http://mlwiki.org/index.php/ROC_Analysis

ROC PLOTS AND CURVES

- Receiver Operating Curves (ROC) plot the *false positive rate* vs. the *true positive rate*

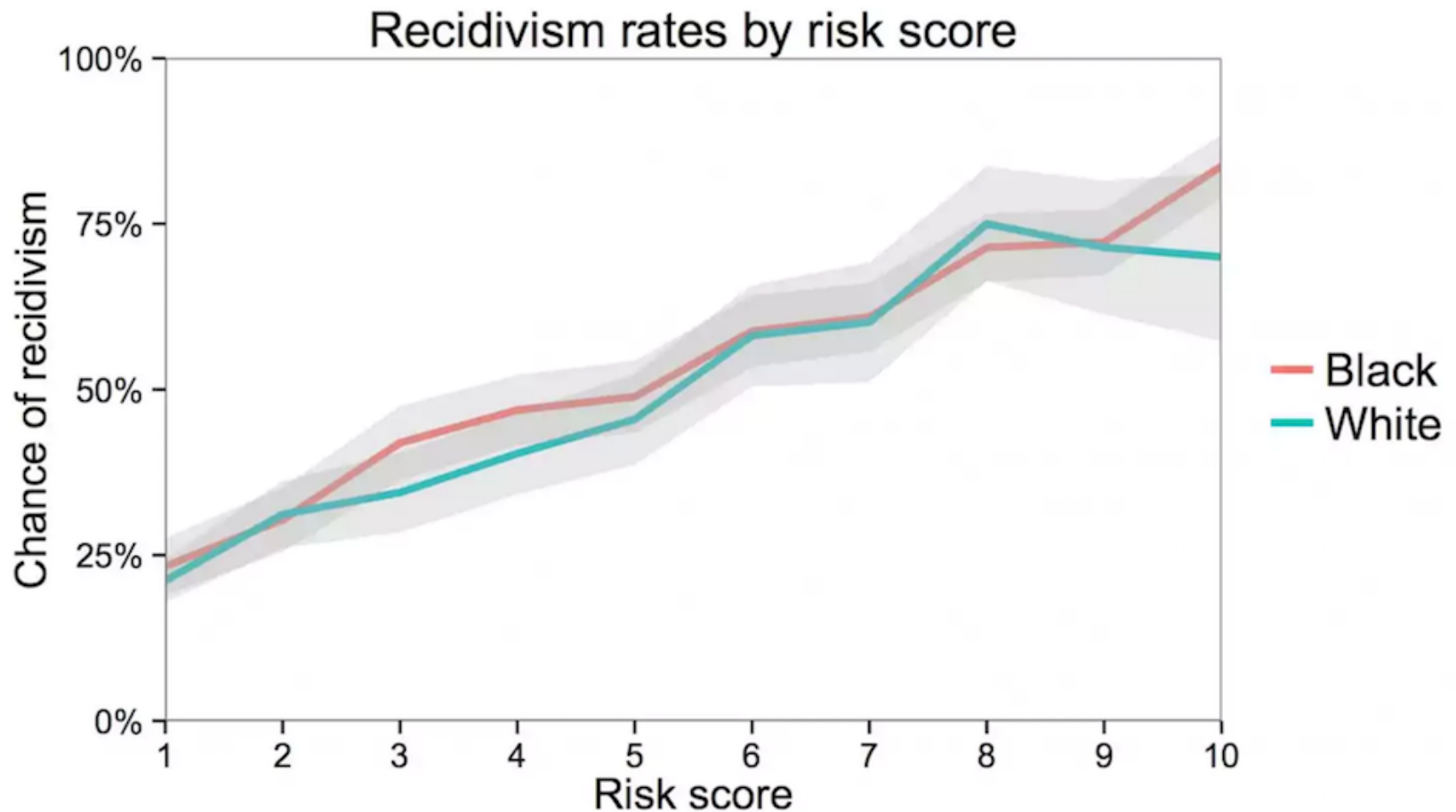http://davidgbrown.co/

33

Signal-Absent Distribution   Signal-Present Distribution

Decision Variable

ROC Curve

True Positive Fraction

False Positive Fraction

http://davidgbrown.co/

34

WARNING!

Many types of curves are not balanced!!

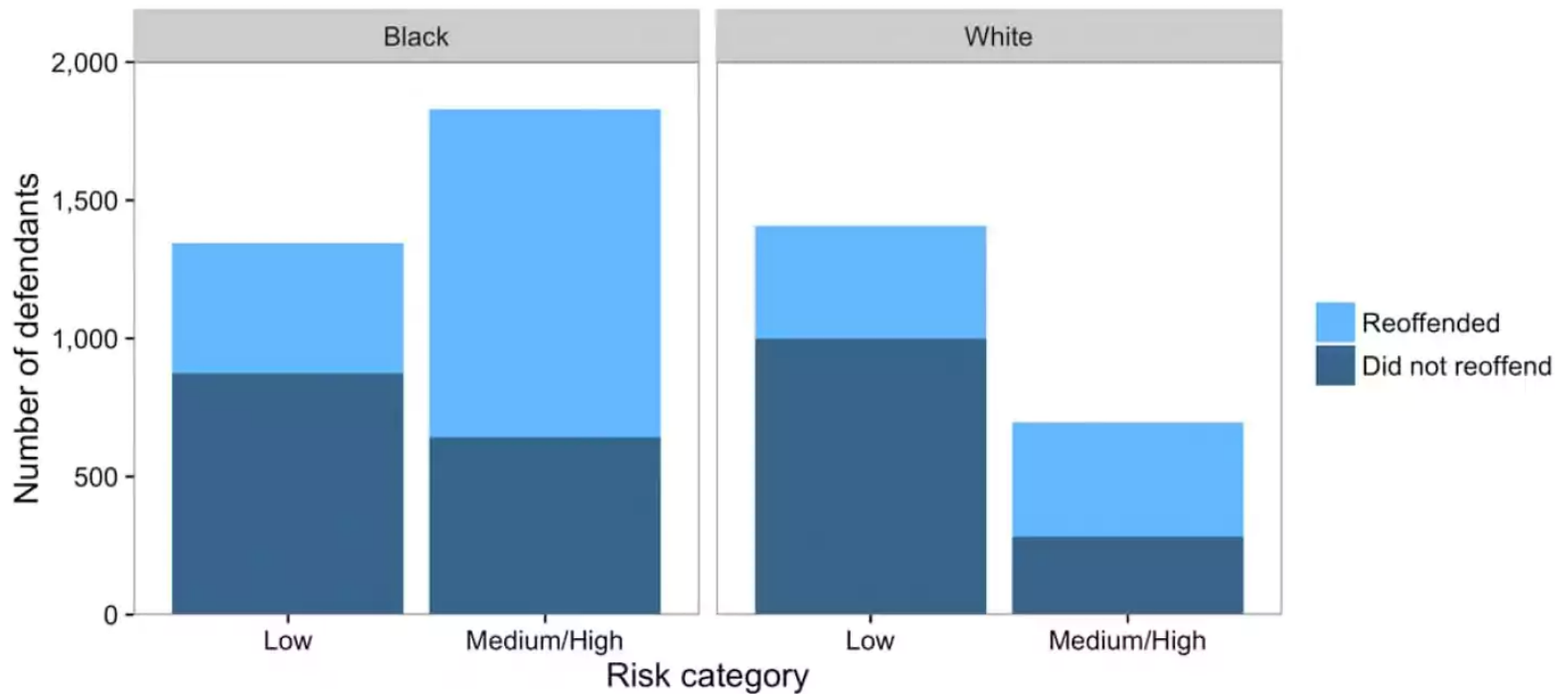# IS COMPAS FLAWED?

Recidivism rates by risk score

Sam Corbett-Davies, Emma Pierson, Avi Feller and Sharad Goel, Washington Post

ProPublica's definition of "fairness" focused on prisoners that did not reoffend.

	Caucasian	**African American**
Higher Risk, Didn't Re-Offend	23.5%	44.9%
Lower Risk, Did Re-Offend	47.7%	28.0%

ProPublica

Sam Corbett-Davies, Emma Pierson, Avi Feller and Sharad Goel, Washington Post

Caucasian defendants

- Accuracy: 0.67
- False positive rate: 0.23
  ProPublica Data

African American defendants

- Accuracy: 0.64
- False positive rate: 0.45
  ProPublica Data

- Accuracy is similar for each risk category

- There are twice as many African American reoffenders

...

- There are twice as many "false positives" for African Americans

- Balance groups over scores (calibration)
- Balance groups over "not-reoffend" (negative class balance, false positive rate)
- Balance groups over "reoffended" (positive class balance, false negative rate)

Choose 1

Inherent Trade-Offs in the Fair Determination of Risk Scores, Jon Kleinberg, Sendhil Mullainathan, Manish Raghavan

- Cost/Benefit Analysis and optimise for expected value?
  - Does parole work?
  - Should everyone be treated fairly?
  - Should offenders be treated fairly?
  - Is it more important to keep prospective criminals off the street?
- Underlying problems
  - Why is there so much skew in the base rate?

- More questions hidden in the data
  - Differences between Men and Women
  - Similar problems with people of Hispanic descent
  - So many confounders (violence, drugs, background, etc.)
- Sampling problems
  - Can't sample everyone. Are samples representative?
- Local effects
  - Is it reasonable to assume one model for the whole of the US?

We found that sometimes people's names or dates of birth were incorrectly entered in some records - which led to incorrect matches between an individual's COMPAS score and his or her criminal records. We attempted to determine how many records were affected. In a random sample of 400 cases, we found an error rate of 3.75 percent (CI: +/- 1.8 percent).

WHAT DOES THIS MEAN FOR ENGINEERS?

- Fallout

Anneliese Bruner
@AnnelieseMaria

The workings of COMPAS R kept secret but, used by judges to determine inmate parole, it may be biased vs minorities.businessinsider.com/killer-robots-…

9:49 AM - Oct 5, 2017

Forget killer robots — bias is the real AI danger

Google's AI chief isn't worried about intelligent killer robots. Instead, he's concerned about the danger that may be inside the machine-

businessinsider.com

♡   See Anneliese Bruner's other Tweets

47

- Be careful
- You are biased
- Your data are biased
- Ask others to break it
- Visualise all the things
- Don't rely on single metrics
- Communicate performance (expected value framework)
- Be upfront with weaknesses

Revert "yolo" ...
idcrosby committed on Jun 13   31160d3   <>

yolo
idcrosby committed on Jun 13 ✓   04e67dd   <>

- Stream or batch
- Regression
- Clustering
- Testing in production/monitoring
- Feedback
- Online/offline differences
- Deployment
- Provenance
- Snapshotting
- Testing types (model/data/regression/etc.)
- Get the easy things right

Data Science **Training, Consultancy, Development**

✦ **@DrPhilWinder**

DrPhilWinder

🌐 https://WinderResearch.com

phil@WinderResearch.com

🌐 **https://TrainingDataScience.com**

TRAINING DATA SCIENCE   Training   Workshops   Tips & Tricks   Help   Members

Free Data Science Training

Receive my free monthly data science training worksheets to help you learn data science. Use them to boost your project portfolio and to gain practical experience. Sign up now, cancel at any time!

| Email Address | Submit |

Why I Created Training Data Science

I'm tired of click-bait. I don't agree with the use of the word AI. I wanted to create a place where people could learn data science without the hype. This is for people that want to learn data

@DrPhilWinder | WinderResearch.com