Machine Learning for Data Science (CS4786)
Lecture 26

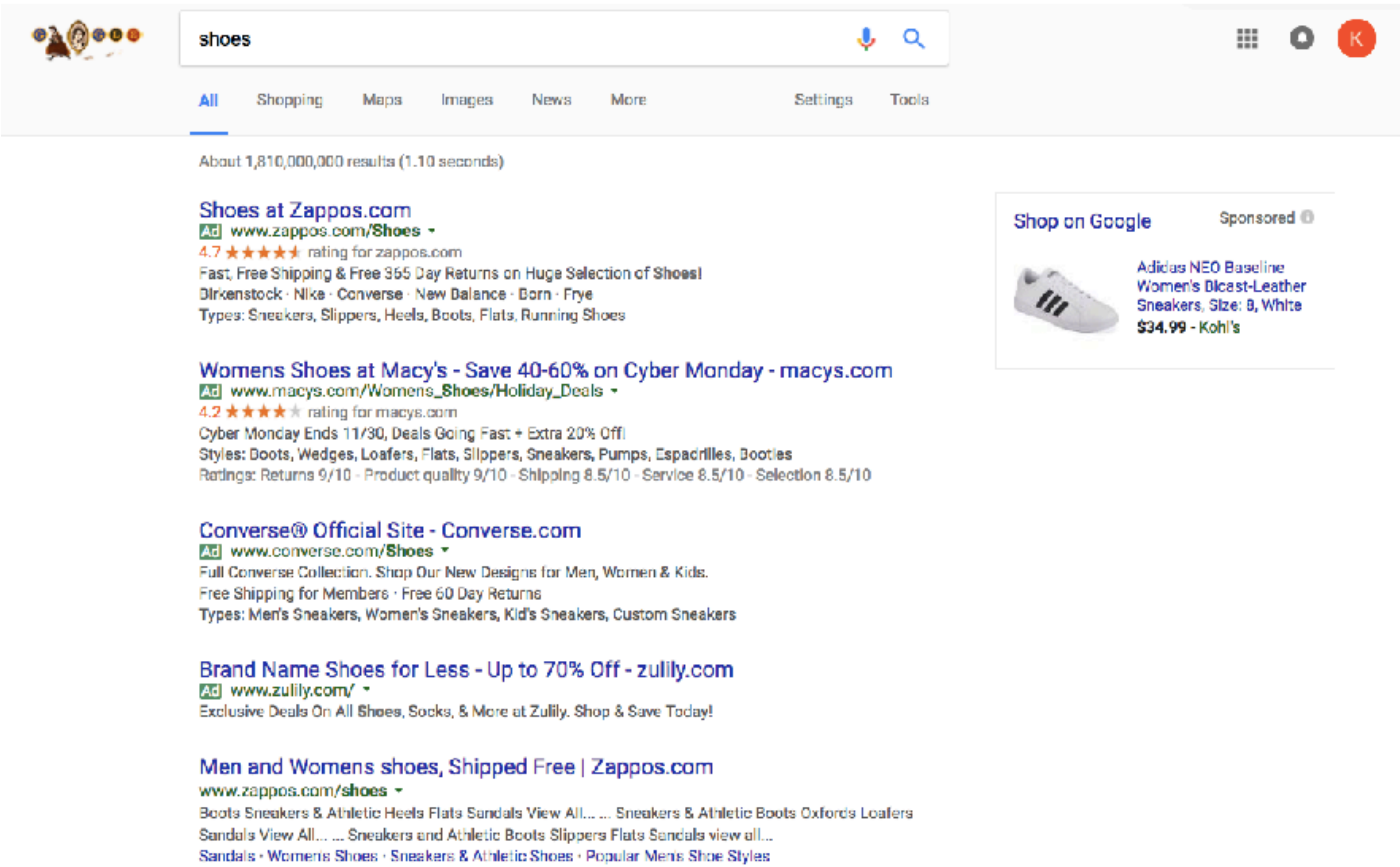
Fairness, Transparency and other Moral Issues in Machine Learning

Course Webpage :
http://www.cs.cornell.edu/Courses/cs4786/2016fa/

Announcements

- Survey 2, just over 80%
- Make sure you fill out the course eval
  - (If the participation on this is above 90% I will still drop worst assignment)

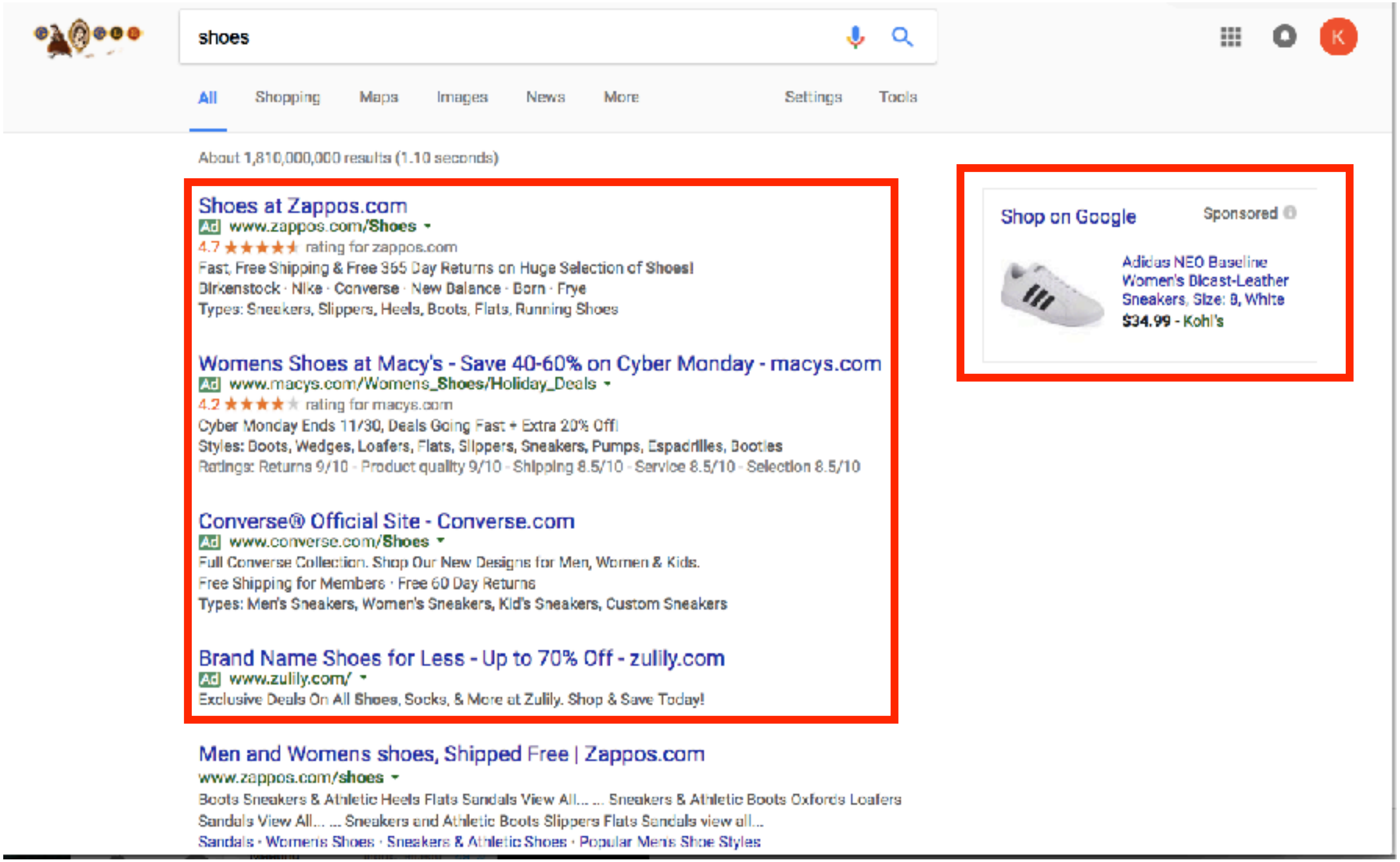
ML DREAM



shoes
All
Shopping
Maps
Images
News
More
Settings
Tools
About 1,810,000,000 results (1.10 seconds)
Shoes at Zappos.com
Ad www.zappos.com/Shoes
4.7 ★★★★★ rating for zappos.com
Fast, Free Shipping & Free 365 Day Returns on Huge Selection of Shoes!
Birkenstock · Nike · Converse · New Balance · Born · Frye
Types: Sneakers, Slippers, Heels, Boots, Flats, Running Shoes
Womens Shoes at Macy's - Save 40-60% on Cyber Monday - macys.com
Ad www.macys.com/Womens_Shoes/Holiday_Deals
4.2 ★★★★★ rating for macys.com
Cyber Monday Ends 11/30, Deals Going Fast + Extra 20% Off!
Styles: Boots, Wedges, Loafers, Flats, Slippers, Sneakers, Pumps, Espadrilles, Booties
Ratings: Returns 9/10 · Product quality 9/10 · Shipping 8.5/10 · Service 8.5/10 · Selection 8.5/10
Converse® Official Site - Converse.com
Ad www.converse.com/Shoes
Full Converse Collection. Shop Our New Designs for Men, Women & Kids.
Free Shipping for Members · Free 60 Day Returns
Types: Men's Sneakers, Women's Sneakers, Kid's Sneakers, Custom Sneakers
Brand Name Shoes for Less - Up to 70% Off - zulily.com
Ad www.zulily.com/
Exclusive Deals On All Shoes, Socks, & More at Zulily. Shop & Save Today!
Men and Womens shoes, Shipped Free | Zappos.com
www.zappos.com/shoes
Boots Sneakers & Athletic Heels Flats Sandals View All... ... Sneakers & Athletic Boots Oxfords Loafers
Sandals View All... ... Sneakers and Athletic Boots Slippers Flats Sandals view all...
Sandals · Women's Shoes · Sneakers & Athletic Shoes · Popular Men's Shoe Styles
Shop on Google
Sponsored
Adidas NEO Baseline Women's Blcast-Leather Sneakers, Size: 8, White
$34.99 - Kohl's

ML Dream

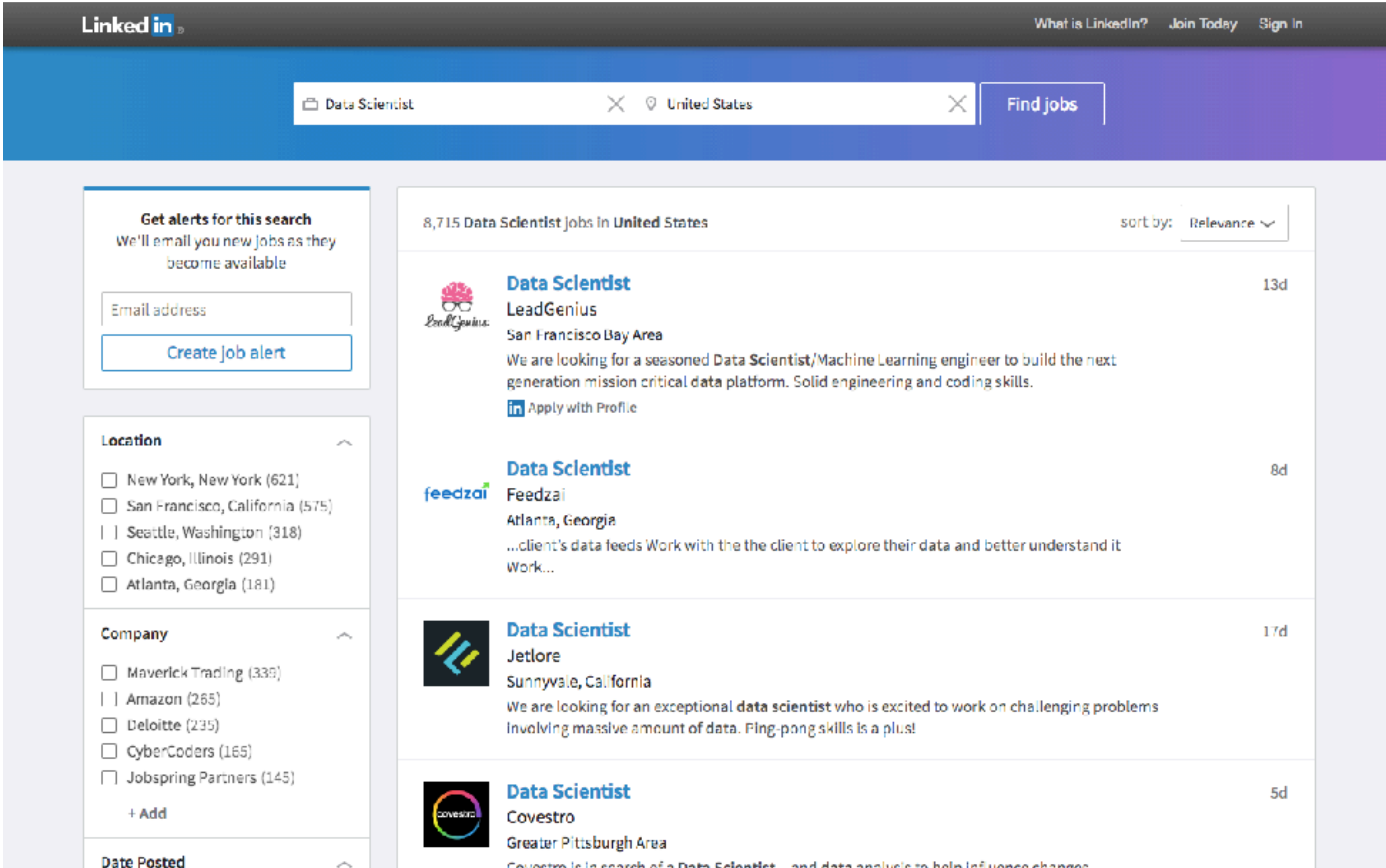
What product you would like…



What news would you prefer to read…

ML Dream

Find the best job for you…



ML Dream

- For every user predict: Ads, products, news, ...
- Have tons of data to learn this task well
- Have right models that can learn from all this data

ML Dream

- For every user predict: Ads, products, news, …
- Have tons of data to learn this task well
- Have right models that can learn from all this data

With Big Data comes Bigger Responsibilities …

Is ML Fair, Impartial?

Google

Women less likely to be shown ads for high-paid jobs on Google, study shows

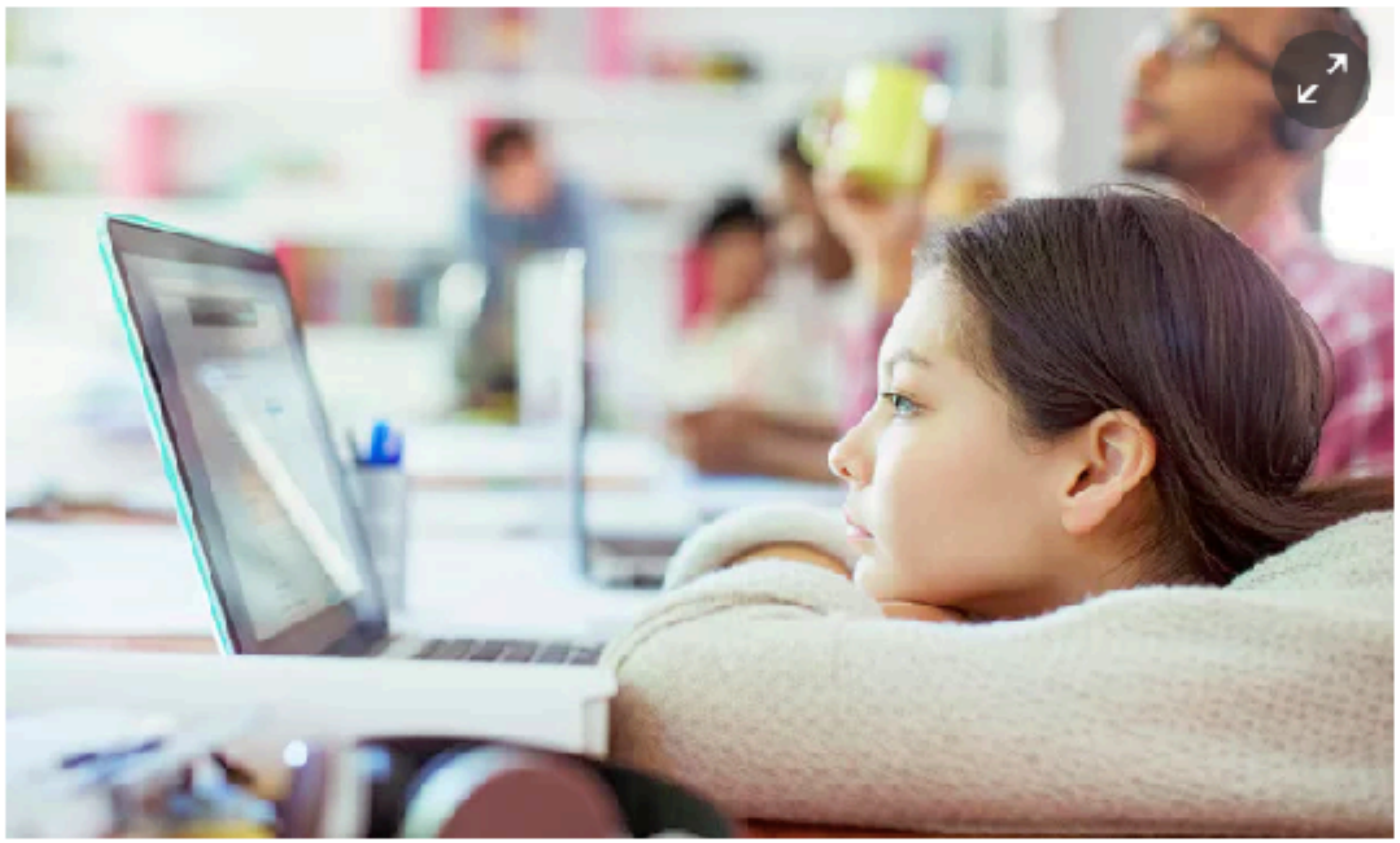
Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs

Samuel Gibbs

Wednesday 8 July 2015
06.29 EDT

This article is 1 year old

1120 140



One experiment showed that Google displayed adverts for a career coaching service for executive jobs 1,852 times to the male group and only 318 times to the female group. Photograph: Alamy

Female job seekers are much less likely to be shown adverts on Google for highly paid jobs than men, researchers have found.

Is ML Fair, Impartial?

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

WHY NOW?

Why Now?

Loads of data collected everywhere!

Why Now?

galvanize

Startup.ML

coursera

METIS

datascience@berkeley

The Data Incubator

Why Now?

galvanize

Startup.ML

coursera

datascience@berkeley

The Data Incubator

METIS

Raise in number of Data Scientists!

Why is ML Unfair?

the algorithms in themselves are neutral. "This program had absolutely nothing to do with race... but multi-variable equations,"

Why is ML Unfair?

- Data collection, labeling etc. can have unintentional biases
  - We learn from past data, historic biases
- Data in itself nor algorithms explicitly know of social inequities

Fairness Through Blindness?

- Ignore all protected attributes.

  Eg. Don't look at race, gender etc.

Fairness Through Blindness?

- Ignore all protected attributes.
  Eg. Don't look at race, gender etc.

- Problem: You don't need to look to be able to predict

Fairness Through Blindness?

- Ignore all protected attributes.

  Eg. Don't look at race, gender etc.

- Problem: You don't need to look to be able to predict

  Eg. User visits "www.artofmanliness.com"

  ...highly likely to be male

Eg. Real Vs Fake Names

- Biases are often not intentional . . .

Eg. Real Vs Fake Names

- Biases are often not intentional …
- Most training examples standard white American names: James, John, Robert, Jennifer, Michael, …
- Ethnic names often unique, much fewer training examples

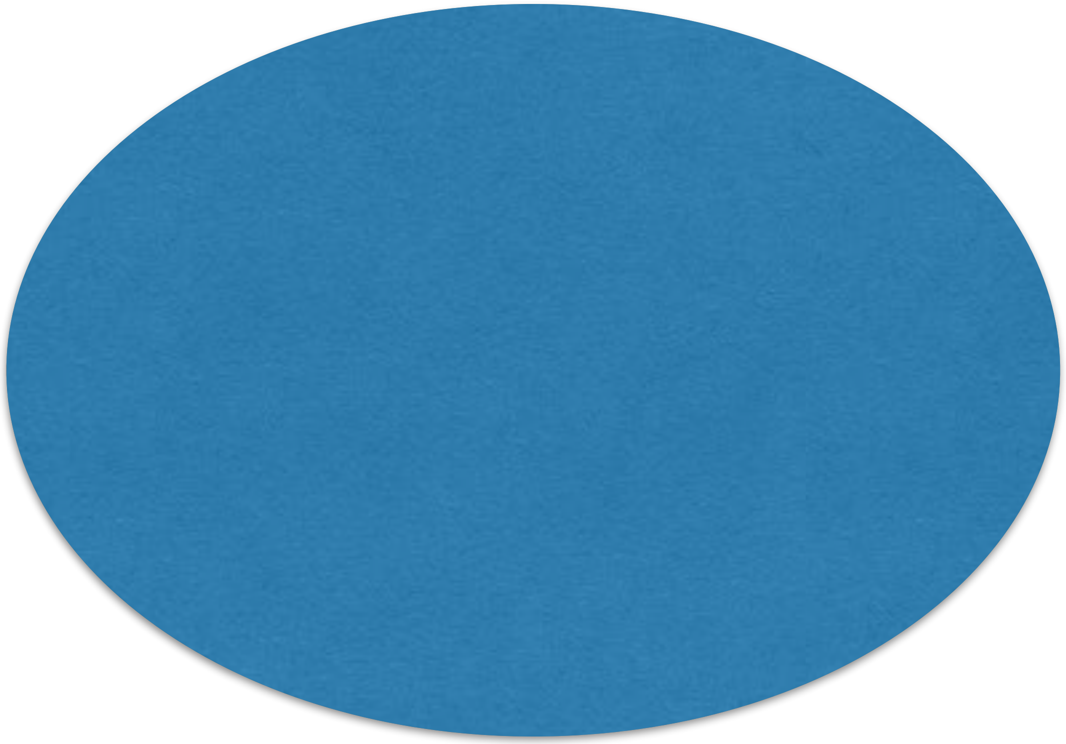
Eg. Real Vs Fake Names

- Biases are often not intentional ...
- Most training examples standard white American names: James, John, Robert, Jennifer, Michael, ...
- Ethnic names often unique, much fewer training examples

Most ML models aim for accuracy for the majority at the expense of mistakes on the smaller protected class

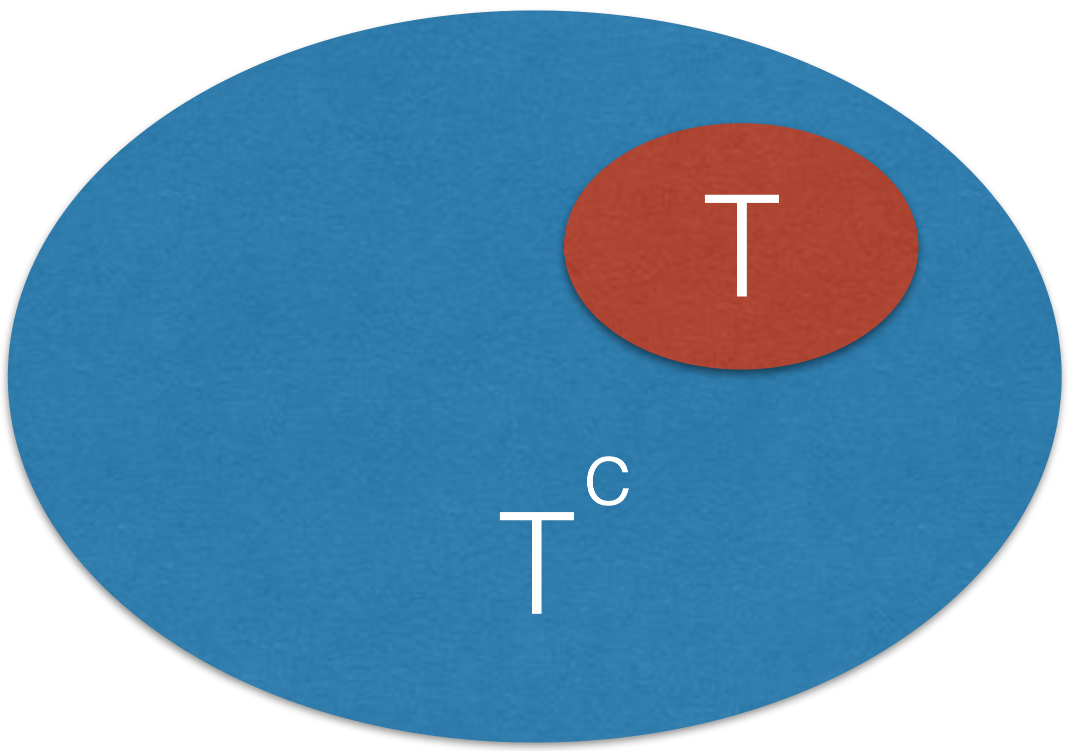
Fairness Through Awareness

Fairness Through Awareness



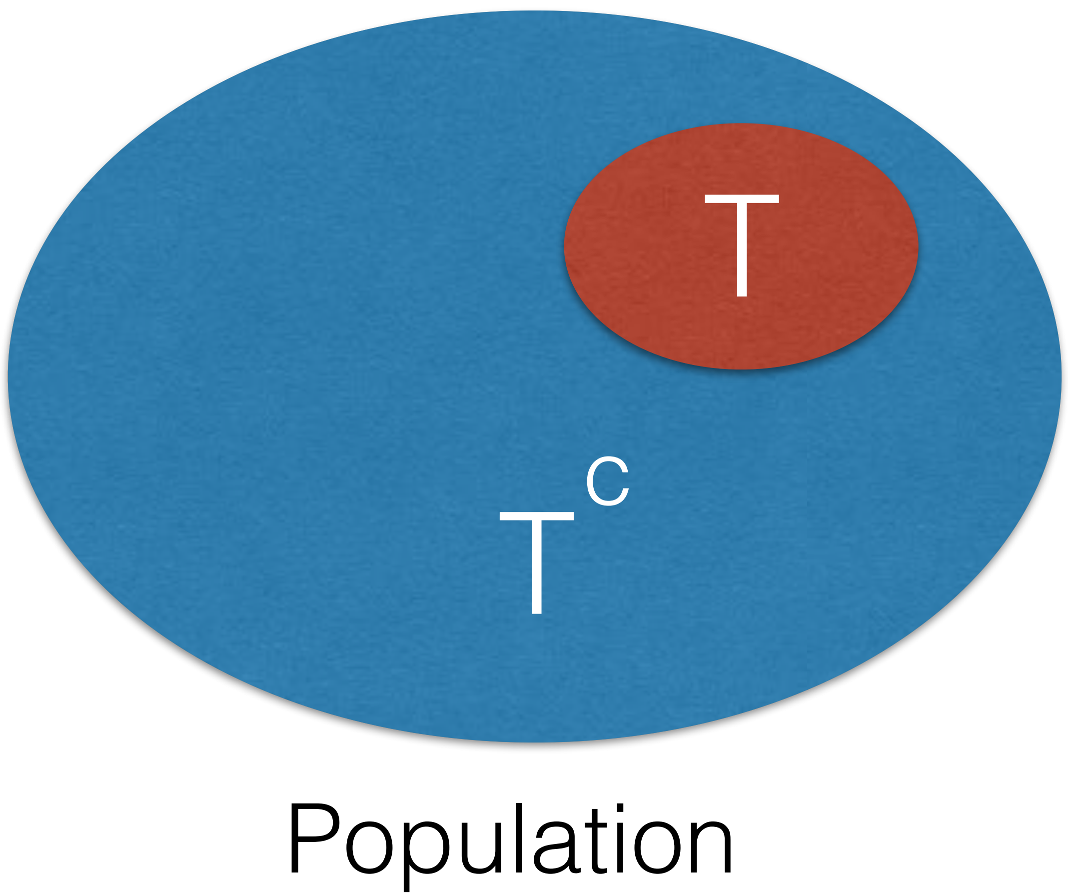
Population

Fairness Through Awareness



Population

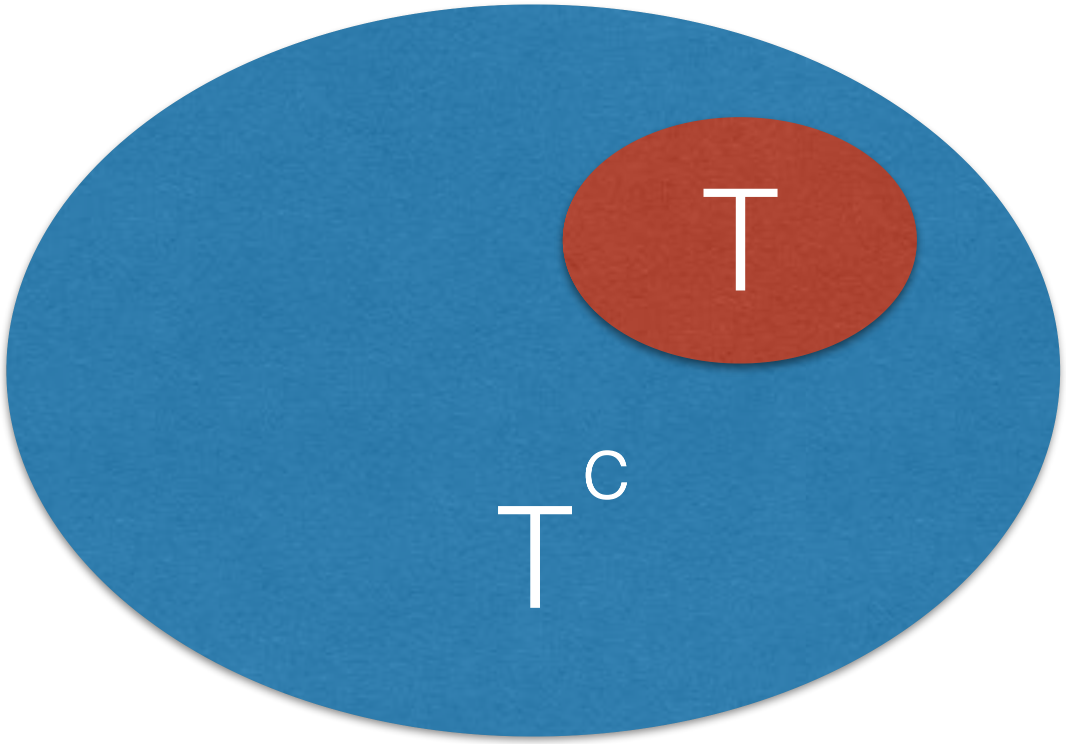
Fairness Through Awareness



T : Protected subset

T^c : Rest of the population

Fairness Through Awareness



Population

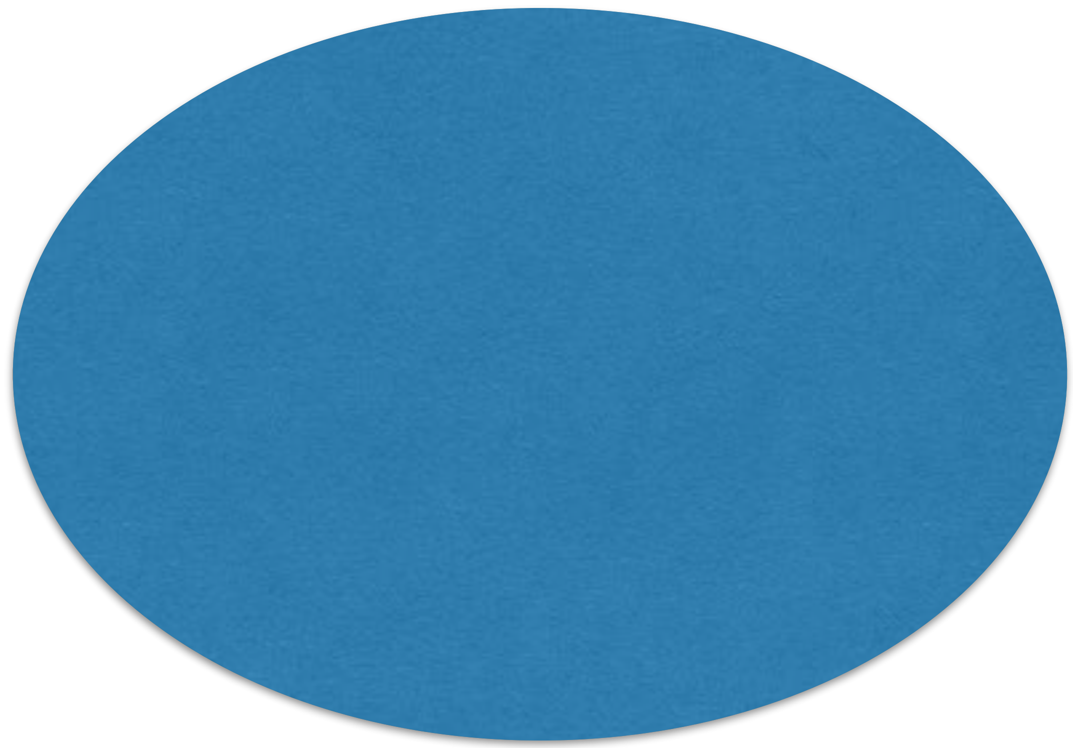
T : Protected subset

T^c : Rest of the population

$$P(\text{Outcome}|T) \approx P(\text{Outcome}|T^c)$$

Eg. Fraction of people shown high paying jobs in T and in T^c is equal

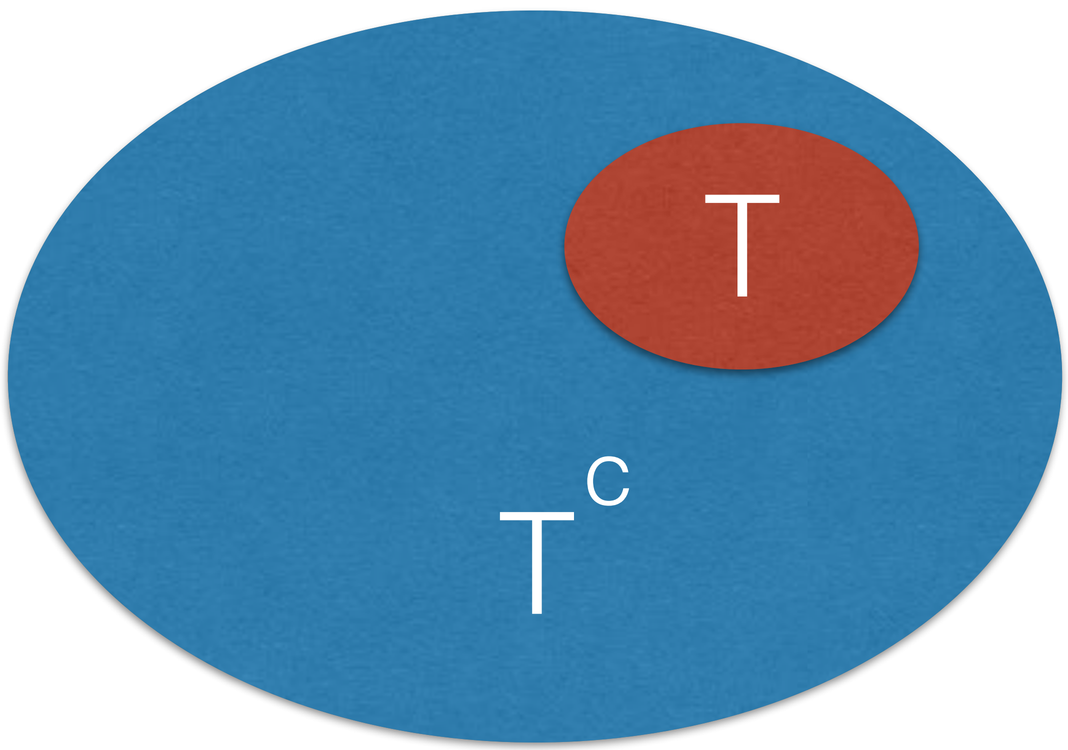
Fairness Through Awareness



Population

Eg. Fraction of people shown high paying jobs in T and in T^c is equal

Fairness Through Awareness



Population

Eg. Fraction of people shown high paying jobs in T and in T^c is equal

Eg. Fair K-means Clustering (very naive)

Eg. Fair K-means Clustering (very naive)

$$\text{Objective} = \sum_{j=1}^{K} \sum_{t \in C_j} \left\| \mathbf{x}_t - \mathbf{r}_j \right\|_2^2$$

$$\text{where } \mathbf{r}_j = \frac{1}{|C_j|} \sum_{\mathbf{x}_t \in C_j} \mathbf{x}_t$$

EG. FAIR K-MEANS CLUSTERING (very naive)

$$\text{Objective} = \sum_{j=1}^{K} \sum_{t \in C_j} \left\| \mathbf{x}_t - \mathbf{r}_j \right\|_2^2$$

$$\text{where } \mathbf{r}_j = \frac{1}{|C_j|} \sum_{\mathbf{x}_t \in C_j} \mathbf{x}_t$$

Fairness constraints: $\forall j \in [K], \quad \sum_{t: c_t = j} \mathbf{1}_{x_t \in T} = \sum_{t: c_t = j} \mathbf{1}_{x_t \notin T}$

Eg. Fair K-means Clustering (very naive)

$$\text{Objective} = \sum_{j=1}^{K} \sum_{t \in C_j} \left\| \mathbf{x}_t - \mathbf{r}_j \right\|_2^2$$

$$\text{where } \mathbf{r}_j = \frac{1}{|C_j|} \sum_{\mathbf{x}_t \in C_j} \mathbf{x}_t$$

$$\text{Fairness constraints:} \quad \forall j \in [K], \quad \sum_{t : c_t = j} \mathbf{1}_{x_t \in T} = \sum_{t : c_t = j} \mathbf{1}_{x_t \notin T}$$

Number of protected in cluster j = Number of unprotected in cluster j

Fair Classification

A view from a mile above:

Fair Classification

A view from a mile above:

Minimize Classification objective
(or whatever other surrogate loss you use usually)

Fair Classification

A view from a mile above:

Minimize Classification objective
(or whatever other surrogate loss you use usually)

Added Constraint: subject to proportion of labels in each class being same for protected and unprotected population

Fair Classification

Fair Classification

- Is this good enough?

Fair Classification

- Is this good enough?
    - Say there is this algorithm to select people to invite to apply for this **exclusive, credit card with high annual fee**

Fair Classification

- Is this good enough?
  - Say there is this algorithm to select people to invite to apply for this **exclusive, credit card with high annual fee**
  - One simple way to satisfy the fairness constraint:

Fair Classification

- Is this good enough?
  - Say there is this algorithm to select people to invite to apply for this **exclusive, credit card with high annual fee**
  - One simple way to satisfy the fairness constraint:
    - Make offer to higher income people in the unprotected class

Fair Classification

- Is this good enough?
  - Say there is this algorithm to select people to invite to apply for this **exclusive, credit card with high annual fee**
  - One simple way to satisfy the fairness constraint:
    - Make offer to higher income people in the unprotected class
    - Make offer to lower income people in protected class (in same proportion)

FAIR CLASSIFICATION

- Is this good enough?
  - Say there is this algorithm to select people to invite to apply for this **exclusive, credit card with high annual fee**
  - One simple way to satisfy the fairness constraint:
    - Make offer to higher income people in the unprotected class
    - Make offer to lower income people in protected class (in same proportion)

NOT REALLY FAIR!

Fair Classification

A view from a mile above:

Minimize Classification objective
(or whatever other surrogate loss you use usually)

FAIR CLASSIFICATION

A view from a mile above:

Minimize Classification objective
(or whatever other surrogate loss you use usually)

Added Constraint: subject to proportion of labels for **similar instances** in each class being same for protected and unprotected population

Its not always about fairness
There are other issues too!

What news would user prefer to read…

ML Dream

- Just because we can predict, should we?

ML DREAM

- Just because we can predict, should we?
- Say we have a fair, unbiased algorithm for prediction

ML Dream

- Just because we can predict, should we?
- Say we have a fair, unbiased algorithm for prediction
- Can there be other issues?

Extremizing Effect of ML

Extremizing Effect of ML

Extremizing Effect of ML

User 1

User 2

Extremizing Effect of ML

User 1

Apples are extremely rich in important antioxidants, flavanoids, and dietary fiber. The phytonutrients and antioxidants in **apples** may help reduce the risk of developing cancer, hypertension, diabetes, and heart

User 2

For fewer calories per fruit, **oranges** have higher levels of Vitamin C, folate, potassium, and protein.

Extremizing Effect of ML

User 1

Apples are extremely rich in important antioxidants, flavanoids, and dietary fiber. The phytonutrients and antioxidants in **apples** may help reduce the risk of developing cancer, hypertension, diabetes, and heart

User 2

For fewer calories per fruit, **oranges** have higher levels of Vitamin C, folate, potassium, and protein.

EXTREMIZING EFFECT OF ML

User 1

Apples are extremely rich in important antioxidants, flavanoids, and dietary fiber. The phytonutrients and antioxidants in **apples** may help reduce the risk of developing cancer, hypertension, diabetes, and heart

User 2

For fewer calories per fruit, **oranges** have higher levels of Vitamin C, folate, potassium, and protein.

EXTREMIZING EFFECT OF ML

Vs

User 1

Apples are extremely rich in important antioxidants, flavanoids, and dietary fiber. The phytonutrients and antioxidants in **apples** may help reduce the risk of developing cancer, hypertension, diabetes, and heart

User 2

For fewer calories per fruit, **oranges** have higher levels of Vitamin C, folate, potassium, and protein.

Extremizing Effect of ML

Extremizing Effect of ML

- Possible fixes:

Extremizing Effect of ML

- Possible fixes:
  - Good contents is the best fix

Extremizing Effect of ML

- Possible fixes:
  - Good contents is the best fix
  - Mix up articles: provide both sides

Extremizing Effect of ML

- Possible fixes:
  - Good contents is the best fix
  - Mix up articles: provide both sides
    - what if there are multiple sides and worse yet you can’t identify the sides

EXTREMIZING EFFECT OF ML

- Possible fixes:
  - Good contents is the best fix
  - Mix up articles: provide both sides
    - what if there are multiple sides and worse yet you can’t identify the sides
  - Mix up user profiles from time to time (or have a canonical user and every user is mix of individual and the canonical user)

Extremizing Effect of ML

- Possible fixes:
  - Good contents is the best fix
  - Mix up articles: provide both sides
    - what if there are multiple sides and worse yet you can't identify the sides
  - Mix up user profiles from time to time (or have a canonical user and every user is mix of individual and the canonical user)
  - This is a completely open topic …

Extremizing Effect of ML

- Possible fixes:
  - Good contents is the best fix
  - Mix up articles: provide both sides
    - what if there are multiple sides and worse yet you can’t identify the sides
  - Mix up user profiles from time to time (or have a canonical user and every user is mix of individual and the canonical user)
  - This is a completely open topic …

But a very important one …

Transparency in ML

- Another issue: ML methods are complex and we don't understand semantic meaning

Transparency in ML

- Another issue: ML methods are complex and we don't understand semantic meaning
- We need transparency of method for accountability

Transparency in ML

- Another issue: ML methods are complex and we don't understand semantic meaning
- We need transparency of method for accountability
- Transparency via interpretability.
    - Provide explanation for each decision
    - What makes an instance a negative instance according to the algorithm

data transparency lab
DTL2016
conference
November 16-19th
Columbia University
New York City

Sponsors

Google
MIT Connection Science
the technology of innovation
mozilla

Microsoft
DATA TRANSPARENCY LAB
NEW YORK UNIVERSITY INFORMATION LAW INSTITUTE
NYU Technology Law and Policy Clinic
Telefónica

Time for General Questions