Machine Learning

Machine Learning
CSx824/ECEx242
Bert Huang
Virginia Tech

25 Best Jobs in America for 2015

Glassdoor Team | January 27, 2015

Whether you want a new job, or just want to make sure you already have a great job, Glassdoor has released its inaugural report highlighting the **25 Best Jobs in America for 2015**.

This new report identifies the 25 best jobs based on each job's overall Glassdoor Job Score*. The Glassdoor Job Score is determined using three key factors – earning potential based on average annual base salary, career opportunities rating and number of job openings. The jobs that made this list stand out across all three categories. Check out the complete results:

1. **Physician Assistant – Glassdoor Job Score: 4.8**

- Number of Job Openings: 45,484
- Average Base Salary: $111,376
- Career Opportunities Rating: 3.5

2. **Software Engineer - Glassdoor Job Score: 4.6**

- Number of Job Openings: 104,828
- Average Base Salary: $98,074
- Career Opportunities Rating: 3.3

3. **Business Development Manager - Glassdoor Job Score: 4.6**

- Number of Job Openings: 11,616

Most Popular

Glassdoor Employees' Choice Award Winners Revealed: Google's Larry Page #1 Highest Rated CEO in 2015
Glassdoor Team | June 9, 2015

Top Oddball Interview Questions for 2015
Glassdoor Team | March 17, 2015

25 Best Jobs in America for 2015
Glassdoor Team | January 27, 2015

- Career Opportunities Rating: 3.4

8. **Product Manager - Glassdoor Job Score: 4.5**

- Number of Job Openings: 10,294
- Average Base Salary: $113,363
- Career Opportunities Rating: 3.3

9. **Data Scientist - Glassdoor Job Score: 4.4**

- Number of Job Openings: 3,449
- Average Base Salary: $104,476
- Career Opportunities Rating: 3.8

10. **Sales Manager - Glassdoor Job Score: 4.4**

- Number of Job Openings: 26,193
- Average Base Salary: $76,556
- Career Opportunities Rating: 3.3

11. **Solutions Architect - Glassdoor Job Score: 4.4**

Help Wanted: Black Belts in Data

Starting salaries for data scientists have gone north of $200,000

by Rodrigo Orihuela and Dina Bass
from **Bloomberg Businessweek**

June 4, 2015 – 1:07 PM EDT *Updated on* June 4, 2015 – 2:00 PM EDT

Photographer: Getty Images

A new species of techie is in demand these days—not only in Silicon Valley, but also in company headquarters around the world. "Data scientists are the new superheroes," says Pascal Clement, the head of Amadeus Travel Intelligence in Madrid. The description isn't exactly hyperbolic: The qualifications for the job include the strength to tunnel through mountains of information and the vision to discern patterns where

http://www.bloomberg.com/news/articles/2015-06-04/help-wanted-black-belts-in-data

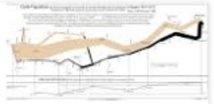
Google what is a data scientist

Bert

Web News Images Videos Shopping More Search tools

About 34,400,000 results (0.48 seconds)

Data Science is the extraction of knowledge from large volumes of **data** that are structured or unstructured, which is a continuation of the field **data** mining and predictive analytics, also known as knowledge discovery and **data** mining (KDD).



Data science - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/**Data_science** Wikipedia

Feedback

IBM - What is a Data Scientist? – Bringing big data to the ...
www.ibm.com/software/**data**/infosphere/**data-scientist**/ IBM
About **data scientists**. Rising alongside the relatively new technology of big **data** is the new job title **data scientist**. While not tied exclusively to big **data** projects, the **data scientist** role does complement them because of the increased breadth and depth of **data** being examined, as compared to traditional roles.

8 Skills You Need to Be a Data Scientist - Udacity - Climb ...
blog.udacity.com/2014/11/**data-science**-job-skills.html Udacity
Nov 7, 2014 - Many resources out there may lead you to believe that becoming a **data scientist** requires comprehensive mastery of a number of fields, such as software development, **data** munging, databases, statistics, machine learning and **data** visualization. Don't worry. In my experience as a **data scientist**, that's not the case.

How do I become a data scientist? An evaluation of 3 ...
datascopeanalytics.com/.../how-do-i-become-a-**data-scientist**-an-evaluatio...
Aug 4, 2014 - One of the most frequent questions we hear, right behind "so, what exactly is a **data scientist**" or "what makes a great **data scientist**", is "how do I ...

Data science - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/**Data_science** Wikipedia
Data Science is the extraction of knowledge from large volumes of data that are structured or unstructured, which is a continuation of the field data mining and ...

In the news

Luring The Curious Into The Data Science Kitchen
Forbes - 2 days ago
Data scientists are like the master chefs of the business world. Sequestered in their glass ...

How Data Science Shaped This Teen&Counseling&By&Text Service
Co.Exist - 22 hours ago

More news for what is a data scientist

Ads

Data Scientists
advance.bu.edu/Statistics
Accelerated, 1-Yr MS in Statistical Practice at BU. Download Brochure.

What is a Data Scientist?
www.sas.com/
A New Breed of Analytical Expert. Who Are They & What Do They Do?

Who Is A Data Scientist
www.edureka.co/**Data-Science**-Training
(800) 275-9730
Instructor Led Live Online Training
24x7 Support. Register Now!

Become a Data Scientist
www.galvanize.com/UNH
Complete a 12 months Masters program in San Francisco. Apply now

Data Scientist Jobs
www.fdmgroup.com/apply
Launch Your **Data** Analysis Career
Apply Now For New Roles In DC

See your ad here »

Data science

From Wikipedia, the free encyclopedia

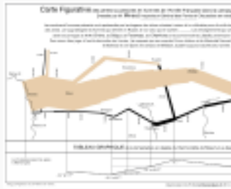
This article's **tone or style may not reflect the encyclopedic tone used on Wikipedia**. See Wikipedia's guide to writing better articles for suggestions. *(February 2014)*

Data Science is the extraction of knowledge from large volumes of data that are structured or unstructured,[1][2] which is a continuation of the field data mining and predictive analytics, also known as knowledge discovery and data mining (KDD). "Unstructured data" can include emails, videos, photos, social media, and other user-generated content. **Data Scientists** are qualified people with strength and patience to tunnel through mountains of information and the technical skills in writing algorithms to extract insights from these troves of information.

Overview [edit]

Data science employs techniques and theories drawn from many fields within the broad areas of nanotechnologies, physics, robotics, mathematics, statistics, information theory and information technology, including signal processing, probability models, machine learning, statistical learning, data mining, database, data engineering, pattern recognition and learning, visualization, predictive analytics, uncertainty modeling, data warehousing, data compression, computer programming, and high performance computing. Methods that scale to Big Data are of particular interest in data science, although

Part of a series o
Data visual



Major dimensions
Exploratory data analysis
Interactive data v
Descriptive statistics • In
Statistical graph
Data analysis • I
Data scie

Thought leaders
John W. Tukey • E

Information graphic types
Line chart • B
Histogram • Sc
Boxplot • Pare
Pie chart • Are
Control chart • I
Stem-and-leaf displa
Small multiple •
Table

Related Topics

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Languages
Deutsch

What is Data Science?

- Roughly defined: how one can use data to answer questions or guide decisions
- Research areas:
  - Computer Science
    - Machine learning 
    - Data mining
  - Statistics

Machine Learning

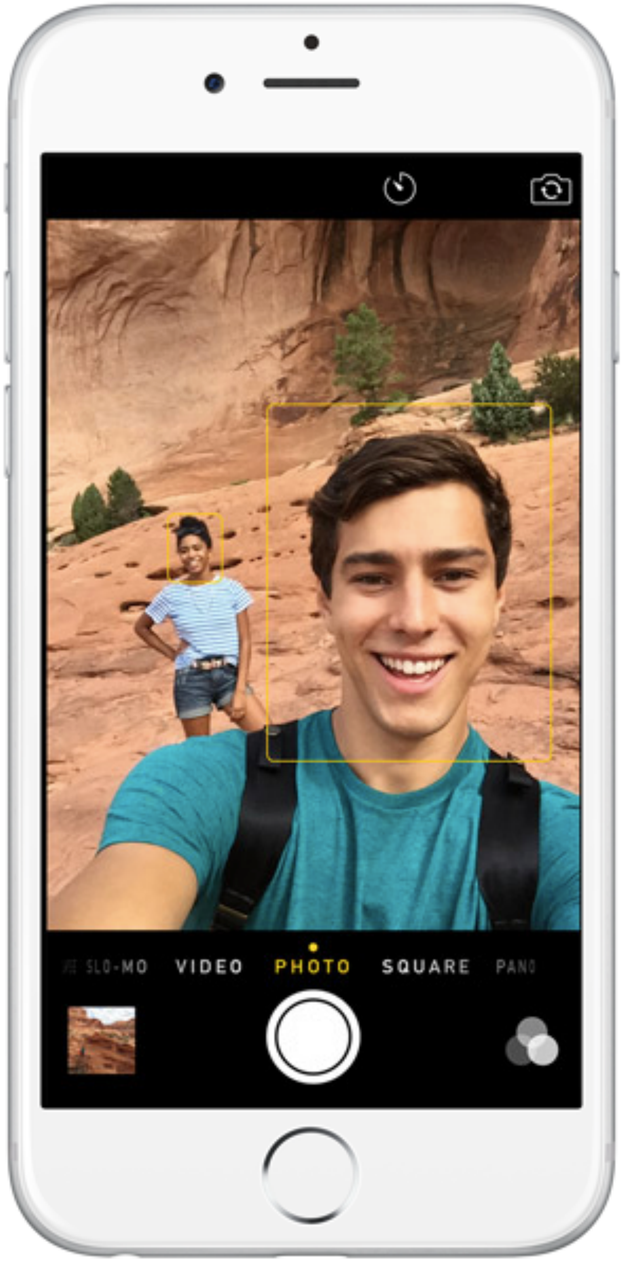
- Learning: improving with experience at some task
  - Improve over **task**
  - with respect to some **performance measure**
  - based on some **experience**
- Writing computer programs that write computer programs

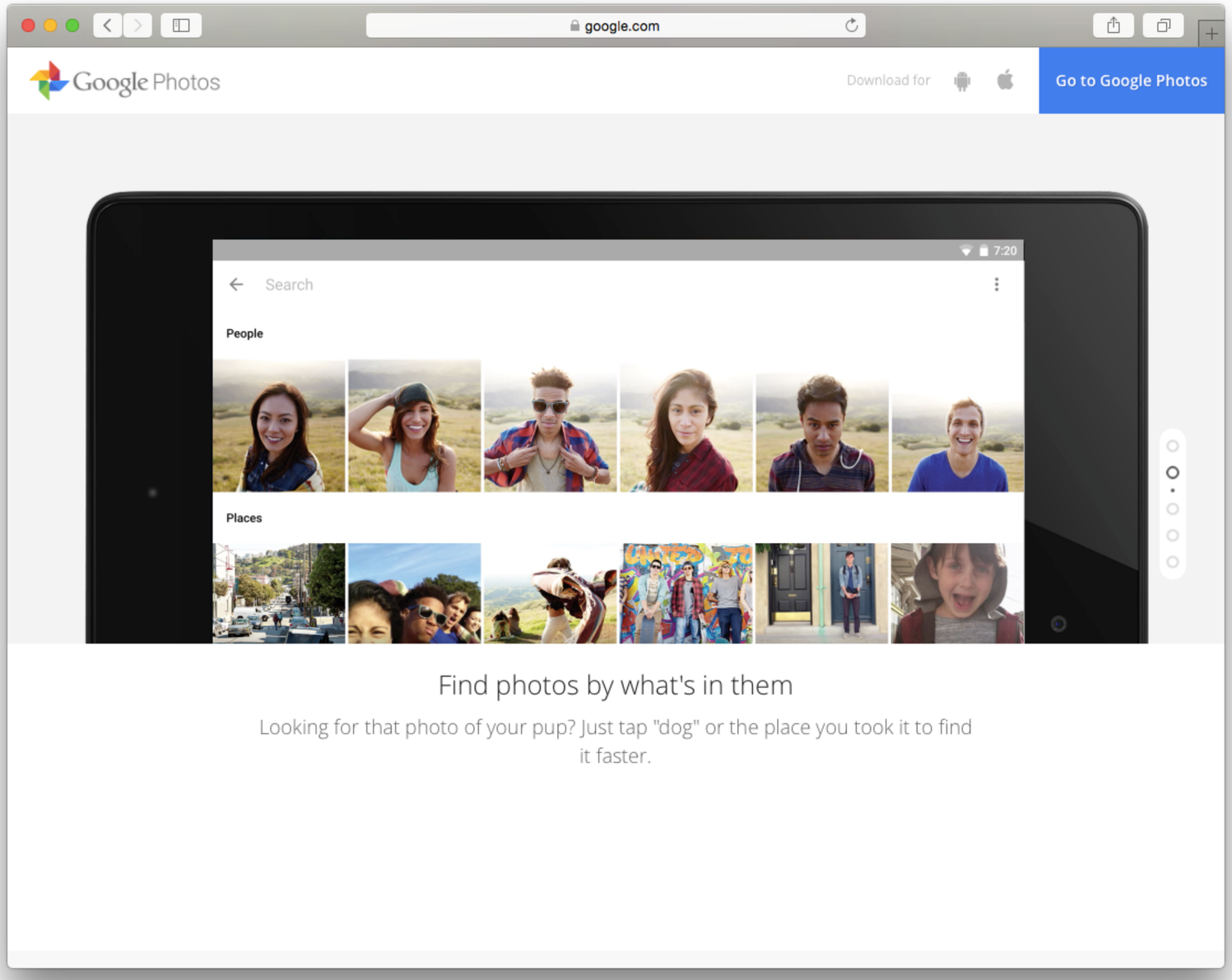
Learning definition by Tom Mitchell

Rest of Video

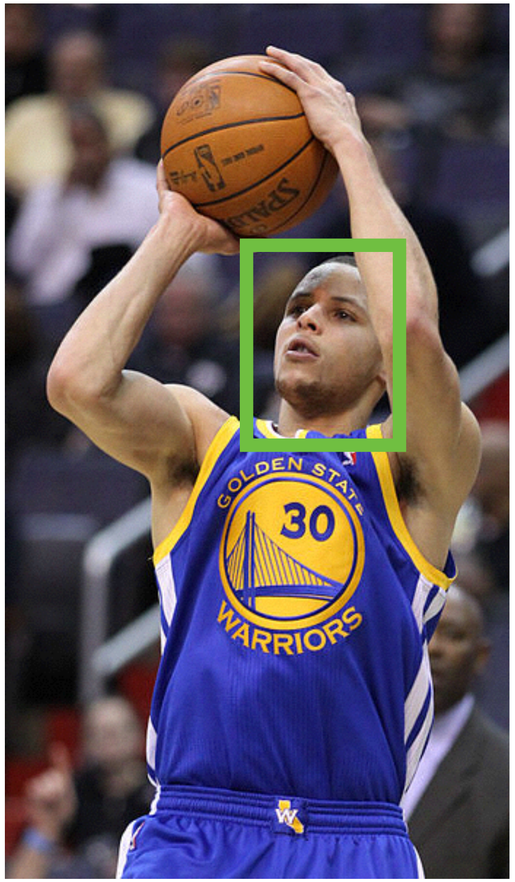
- Three machine learning success stories
- One cautionary tale
- Machine learning research

Machine Learning Story 1
Face Detection & Recognition


SLO-MO
VIDEO
PHOTO
SQUARE
PANO


google.com
Google Photos
Download for
Go to Google Photos
7:20
Search
People
Places
Find photos by what's in them
Looking for that photo of your pup? Just tap "dog" or the place you took it to find it faster.

What Does a Human Face Look Like?



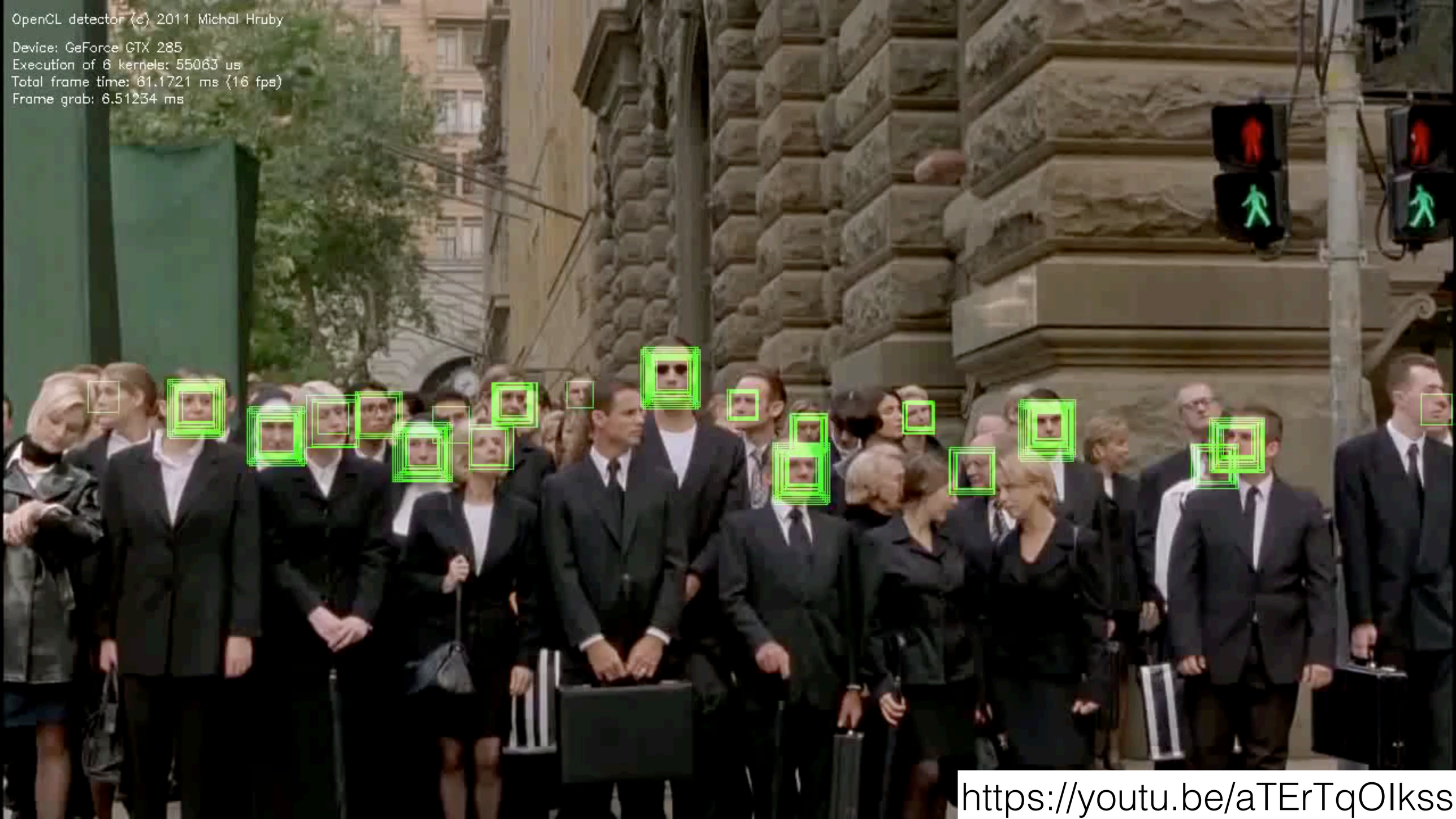
Apple II image from wikipedia.com.
Eyes added digitally.

Apple II image from wikipedia.com.
Eyes added digitally.

if pixel153 > 128 & pixel154 > 128 &
pixel155 > 128 & pixel156 < 64 &
sqrt(pixel157) < 82 &
log(pixel1132 * pixel1133) > 1 ….
then image is a face*

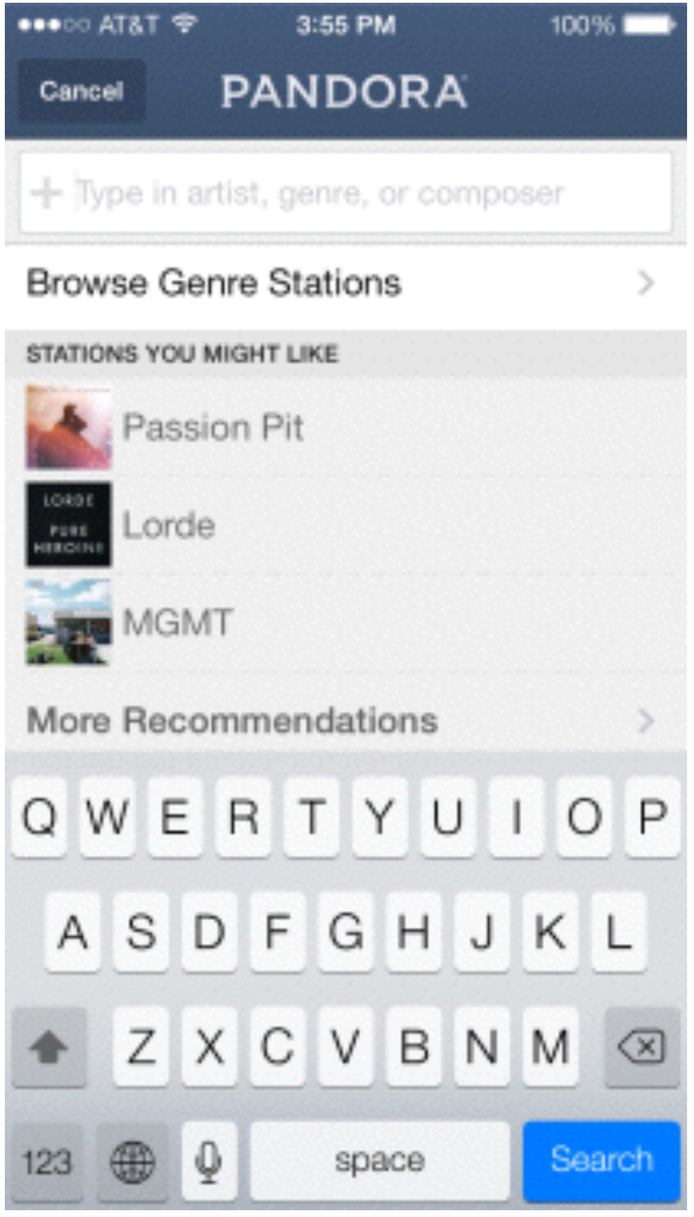
* (not a real face recognition program)

Apple II image from wikipedia.com.
Eyes added digitally.

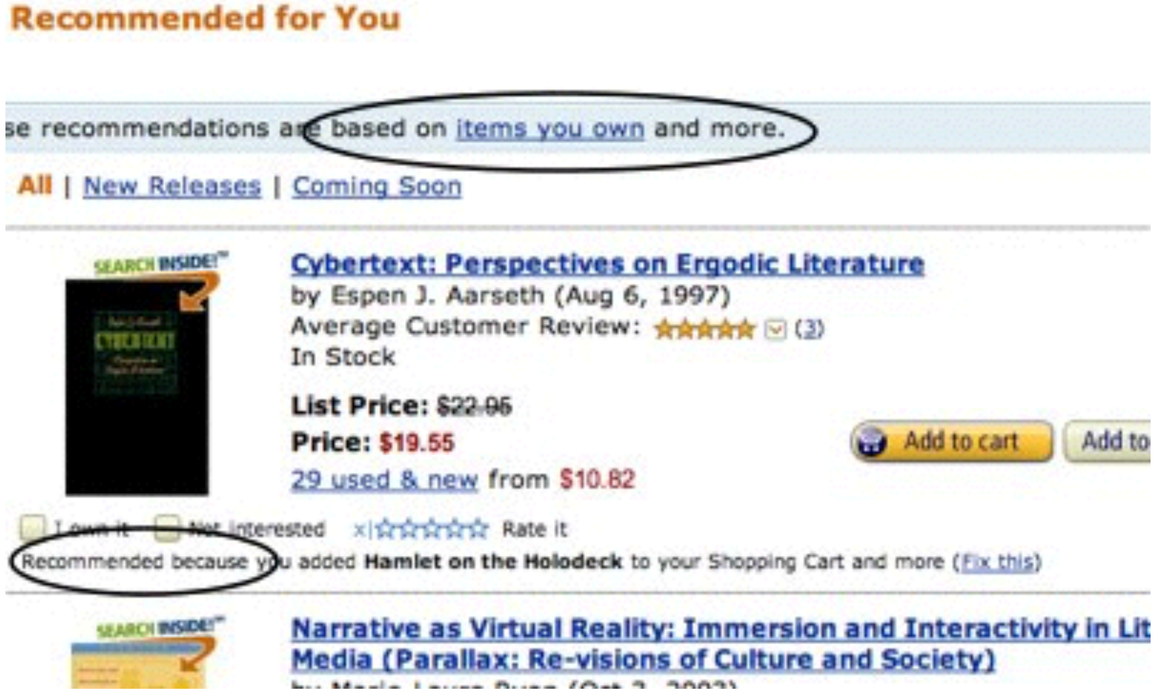

OpenCL detector (c) 2011 Michal Hruby
Device: GeForce GTX 285
Execution of 6 kernels: 55063 us
Total frame time: 61.1721 ms (16 fps)
Frame grab: 6.51234 ms
https://youtu.be/aTErTqOlkss

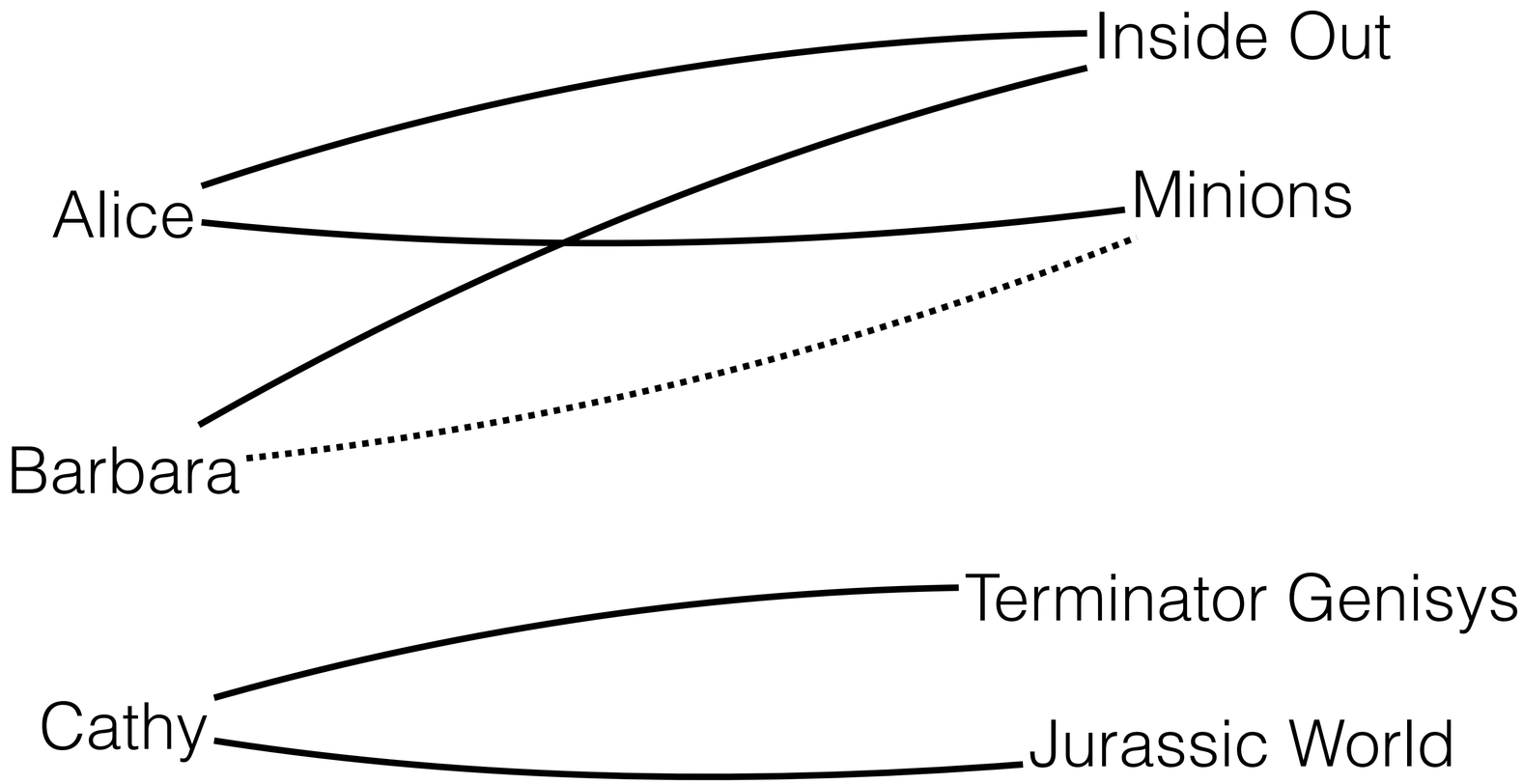
Machine Learning Story 2
Recommender Systems


NETFLIX
2009
DATE 09.21.09
PAY TO THE ORDER OF BellKor's Pragmatic Chaos
$1,000,000 00
AMOUNT ONE MILLION
00/100
FOR The Netflix Prize
Reed Hastings
The Netflix Prize


AT&T
3:55 PM
100%
Cancel
PANDORA
Type in artist, genre, or composer
Browse Genre Stations
STATIONS YOU MIGHT LIKE
Passion Pit
Lorde
MGMT
More Recommendations
Q W E R T Y U I O P
A S D F G H J K L
Z X C V B N M
123
space
Search


People You May Know
see all
Jim M
Add as Friend
Erin Elizabeth K
Add as Friend
Josh S
Add as Friend


Recommended for You
se recommendations are based on items you own and more.
All | New Releases | Coming Soon
SEARCH INSIDE!
Cybertext: Perspectives on Ergodic Literature
by Espen J. Aarseth (Aug 6, 1997)
Average Customer Review: (3)
In Stock
List Price: $22.95
Price: $19.55
29 used & new from $10.82
Add to cart
I own it
Not interested
Rate it
Recommended because you added Hamlet on the Holodeck to your Shopping Cart and more (Fix this)
SEARCH INSIDE!
Narrative as Virtual Reality: Immersion and Interactivity in Lit
Media (Parallax: Re-visions of Culture and Society)


Inside Out
Alice
Minions
Barbara
Terminator Genisys
Cathy
Jurassic World

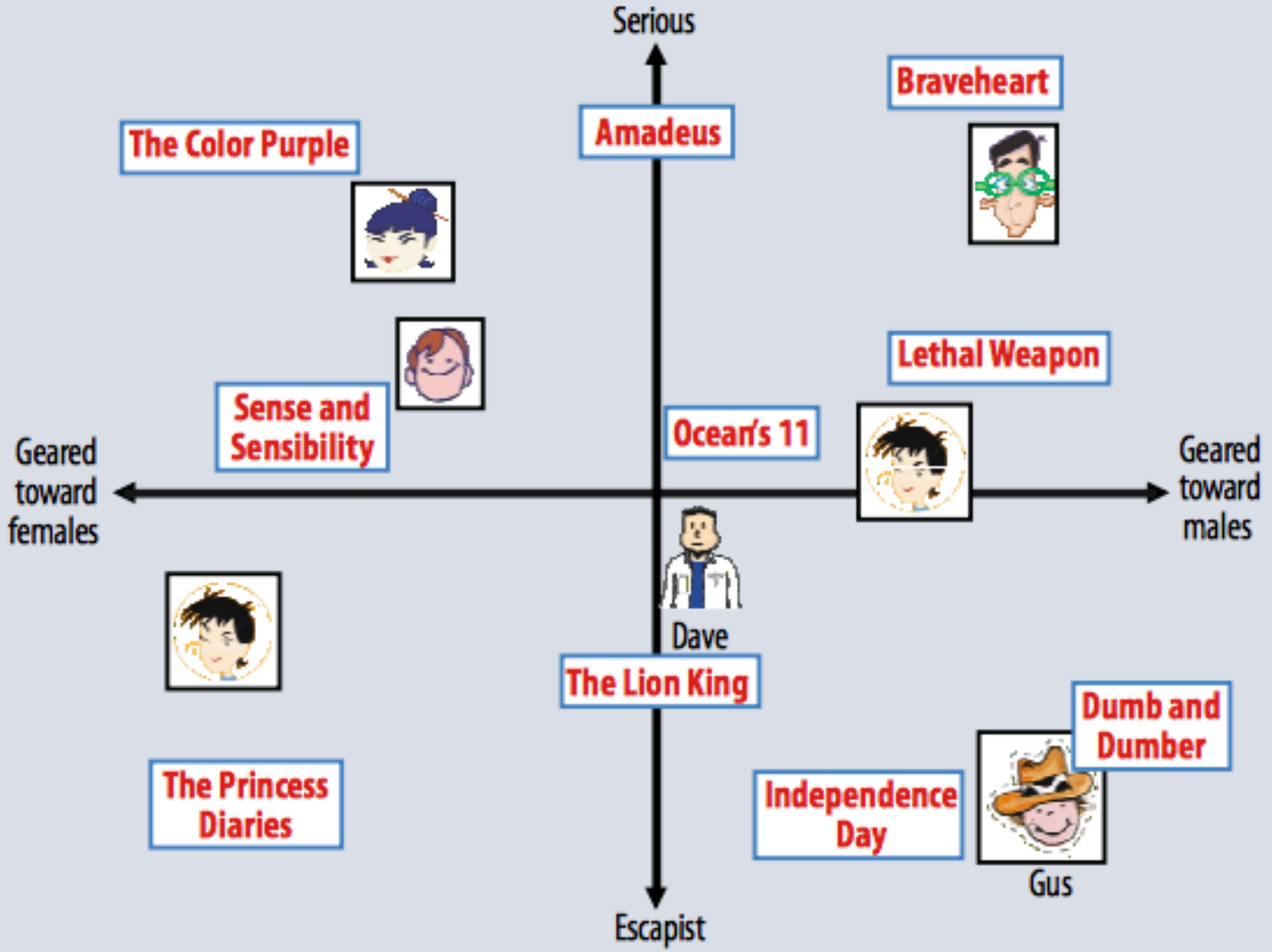


Figure from Koren, Bell, Volinksy, IEEE Computer, 2009

Machine Learning Story 3
EMBERS: Predicting the News with Social Media Data

EMBERS

Early Model-Based Event Recognition with Surrogates

newsweek.com

The EMBERS Project Can Predict the Future With Twitter | arxiv.org/pdf/1402.7035.pdf

US Edition | Social | Search for News | Try Newsweek

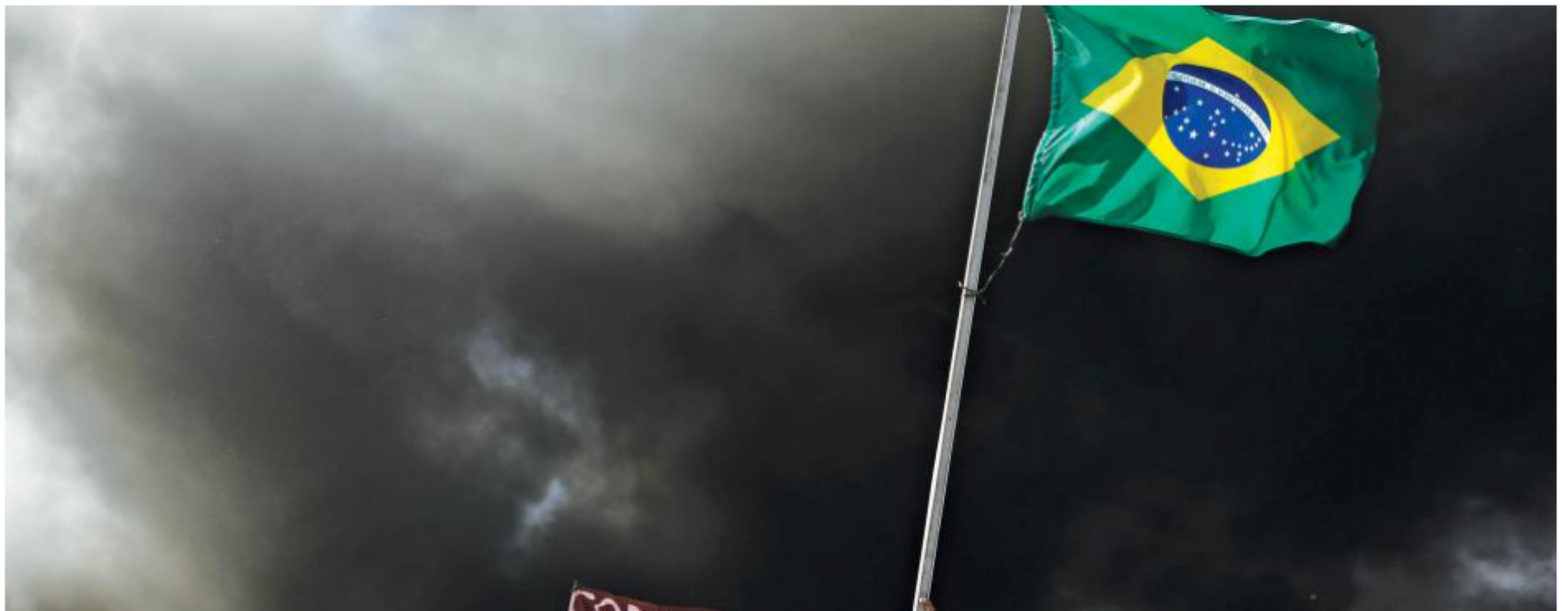
Newsweek

U.S. WORLD BUSINESS TECH & SCIENCE CULTURE SPORTS CAREER NEWS FEED THIS WEEK'S EDITION

TECH & SCIENCE

The EMBERS Project Can Predict the Future With Twitter

BY LEAH MCGRATH GOODMAN / MARCH 7, 2015 12:15 PM EST




```
{
"interaction": {
        "author": {
            …
            "username": ”TwitterUser”
        },
        "content": "Sindicato Unificado de Trabajadores se reunirá mañana!!! http://t.co/none",
        "created_at": "Tue, 14 Jan 2013 00:00:00 +0000",
        "geo": {
            "latitude": -36.7374525,
            "longitude": -71.0352945
        },
        …
}
```


BASIS RLP Enrichment performs natural language processing


```
"BasisEnrichment": {
        "enrichmentProcess": "RLP Java API v7.6.0",
        "entities": [
  {
     "expr": "Sindicato Unificado de Trabajadores",
     "neType": "ORGANIZATION",
     "offset": "134:138"
  }, {
     "expr": "15 de enero del 2013",
     "neType": "TEMPORAL:DATE",
     "offset": "402:407"
  },
```


TIMEN system (extended) for normalizing date expression


```
"eventSemantics": {
        "datetimes": [
 {
     "date": "2012-12-26",
     "id": 0,
     "phrase": "mañana",
     "type": “implicit"
 }, {
     "date": "2013-01-15",
     "id": 1,
     "phrase": "el 15 de enero del 2013",
     "type": "explicit"
 }, {
```

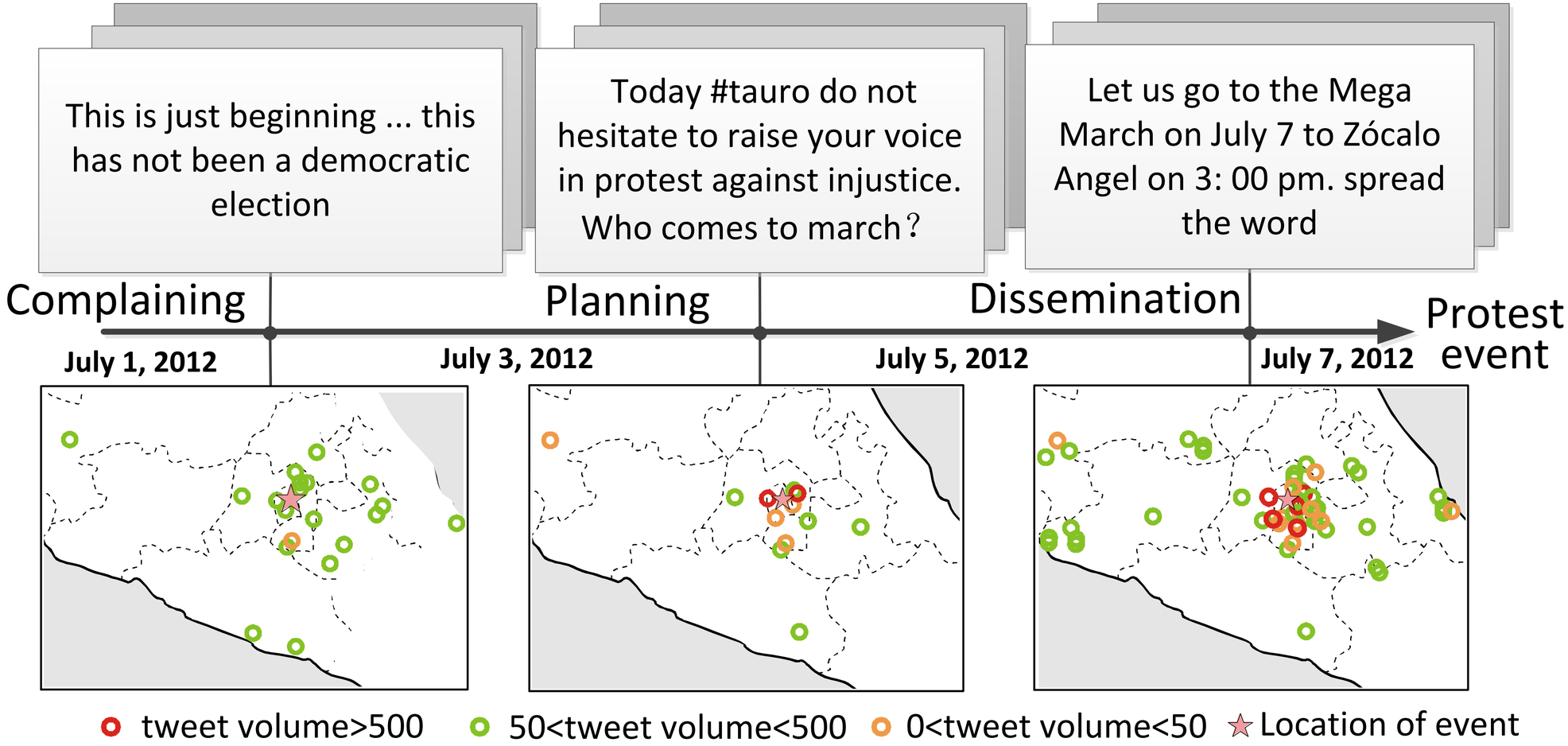


Figure from Zhao, Chen, Lu, Ramakrishnan, SDM 2015

Three Stories

- Face detection and recognition
- Recommendation
- News event prediction

- But not every use of machine learning is successful…

ASA Excellence in Statistical Reporting Award

The formula that killed Wall Street

Wall Street in the mid-1980s turned to the quants – brainy financial engineers – to invent new ways to boost profits. They and their managers, though laziness and greed, built a huge financial bubble on foundations that they did not understand. It was a recipe for disaster. The journalist **Felix Salmon** won the American Statistical Association's Excellence in Statistical Reporting Award for 2010. We reprint his article, first published as the cover story of *Wired* magazine, because it brilliantly conveys complex statistical concepts

A formula in statistics, misunderstood and misused, has devastated the global economy

In the years before 2008, it was hardly unth
that a math wizard like David X. Li might s
earn a Nobel Prize. After all, financial econo
even Wall Street quants – have received the
in economics before, and Li's work on measur
has had more impact, more quickly, than p
Nobel Prize-winning contributions to the field.
though, as dazed bankers, politicians, regulato
investors survey the wreckage of the biggest fi
meltdown since the Great Depression, Li is p
thankful he still has a job in finance at all. N
his achievement should be dismissed. He too
toriously tough nut – determining correlation,
seemingly disparate events are related – and

$$\Pr[T_A < 1, T_B < 1] = \Phi_2(\Phi^{-1}(F_A(1)), \Phi^{-1}(F_B(1)), \gamma)$$

The formula that killed so many pension plans: David X. Li's Gaussian copula, as first published in 2000. Investors exploited it as a quick – and fatally flawed – way to assess risk.

Probability

Specifically, this is a joint default probability – the likelihood that any two members of the pool (A and B) will both default. It's what investors are looking for, and the rest of the formula provides the answer.

Survival times

The amount of time between now and when A and B can be expected to default. Li took the idea from a concept in actuarial science that charts what happens to someone's life expectancy when their spouse dies.

Equality

A dangerously precise concept, since it leaves no room for error. Clean equations help both quants and their managers forget that the real world contains a surprising amount of uncertainty, fuzziness, and precariousness.

Copula

This couples (hence the Latinate term copula) the individual probabilities associated with A and B to come up with a single number. Errors here massively increase the risk of the whole equation blowing up.

Distribution functions

The probabilities of how long A and B are likely to survive. Since these are not certainties, they can be dangerous: Small miscalculations may leave you facing much more risk than the formula indicates.

Gamma

The all-powerful correlation parameter, which reduces correlation to a single constant – something that should be highly improbable, if not impossible. This is the magic number that made Li's copula function irresistible.

Machine Learning Research

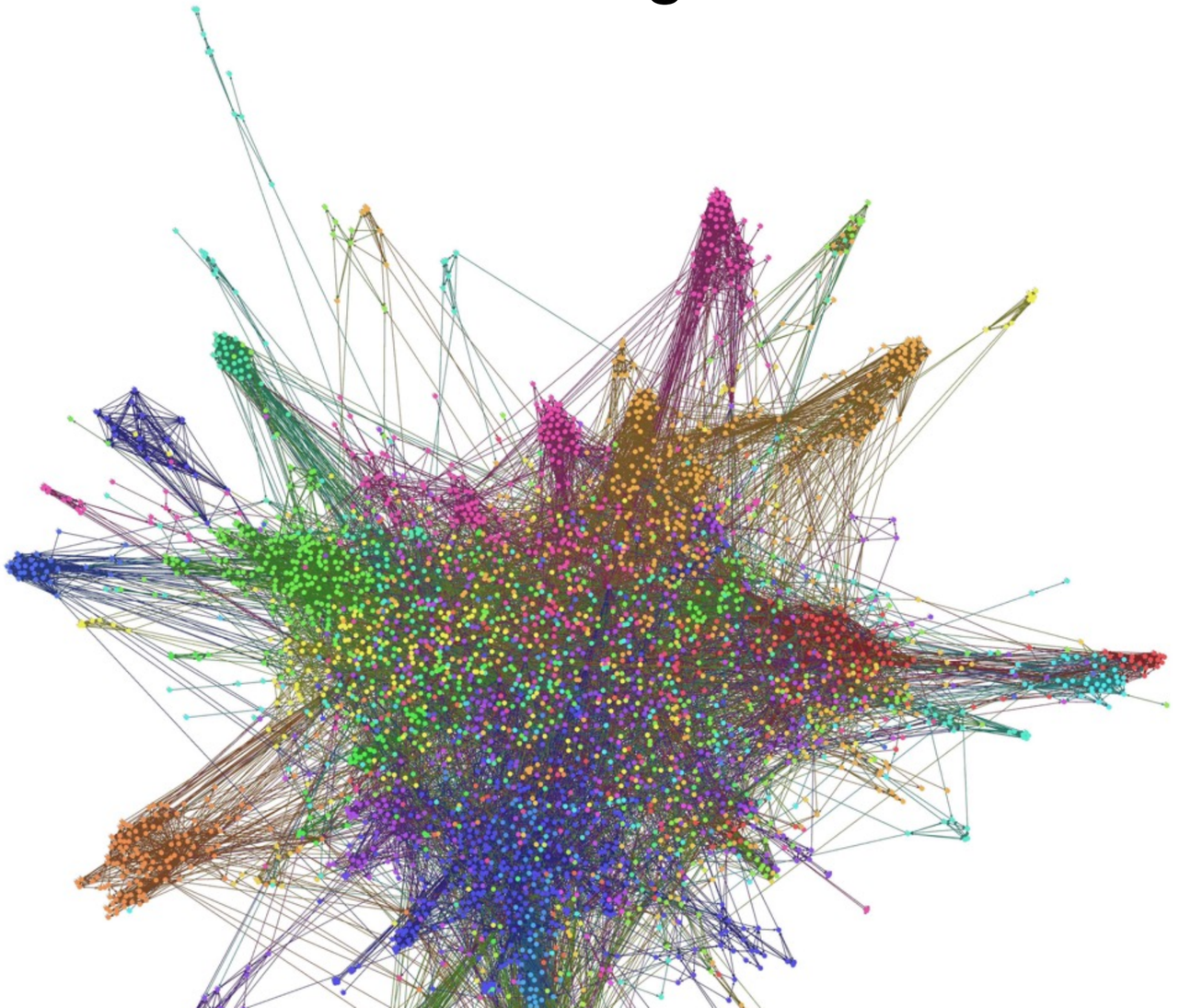
Research Questions

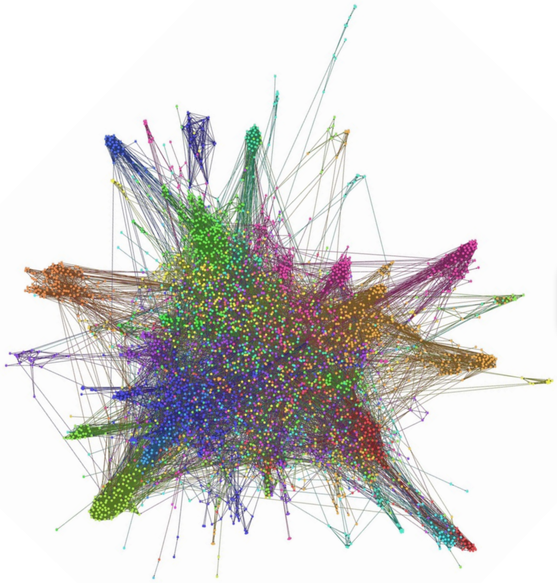
- How do we know if the models we learn are good? Or fair?
- How do we learn from lots of data **efficiently**?
- How do we learn about more complex ideas like natural language, social networks, or biology?

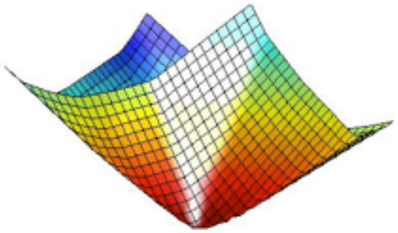

huffingtonpost.com
Culture
Life Lessons
iOS app
Android app
More
Desktop Alerts
Log in
Create Account
March 16, 2015
THE HUFFINGTON POST
Edition: U.S.
Search The Huffington Post
Like 4.8m
Follow
Follow 3.6M
FRONT PAGE
POLITICS
BUSINESS
Black Voices • Gay Voices • Sports • Crime • Science
HUFFPOST LIVE
NYPD Tried To Change Wikipedia
ELECTIO
BIBI V
foxnews.com
Fox News
Fox Business
Fox News Latino
Fox News Go
Fox News Radio
Fox Nation
Fox News Insider
Login
FOX NEWS .com
Fair & Balanced
March 16, 2015 -- Updated at 9:35 PM ET
Search
Listen to Fox News Radio Live
ON AIR NOW
9pet
The Kelly File
WATCH LIVE TV
10pet
Hannity
Hosted by Sean Hannity
On Air Personalities
Home
Video
Politics
U.S.
Opinion
Entertainment
Tech
Science
Health
Travel
Lifestyle
World
Sports
On Air
WH acknowledges Petraeus advising
49°
New York, NY
Detailed Forecast


?
@janedoe1: thanks, Obama
?
@johnsmith3: haha @janedoe1
huffingtonpost.com
Culture Life Lessons iOS app Android app More
Desktop Alerts Log in Create Account
March 16, 2015
THE HUFFINGTON POST
Edition: U.S.
Search The Huffington Post
Like 4.8m Follow Follow 3.6M
FRONT PAGE POLITICS BUSINESS ENTERTAINMENT TECH MEDIA WORLDPOST HEALTHY LIVING COMEDY HUFFPOST LIVE ALL SECTIONS
Black Voices • Gay Voices • Sports • Crime • Science • Religion
HUFFPOST LIVE NYPD Tried To Change Wikipedia
foxnews.com
Fox News Fox Business Fox News Latino Fox News Go Fox News Radio Fox Nation Fox News Insider
Login
March 16, 2015 -- Updated at 9:35 PM ET
Search
ON AIR NOW
9pet The Kelly File
WATCH LIVE TV
Subscribe
On Air Personalities
10pet Hannity
Hosted by Sean Hannity
FOX NEWS .com
Fair & Balanced
Listen to Fox News Radio Live
Home Video Politics U.S. Opinion Entertainment Tech Science Health Travel Lifestyle World Sports On Air
ELECTION EVE SURPRISE:
BIBI VEERS RIGHT-ER
WH acknowledges Petraeus advising
49°
New York, NY
Detailed Forecast

Many applications have complex structure
that traditional machine learning tools can’t handle.
We need new algorithms and new theory.







0.2: Major(*A*, *M*) ∧ Roommates(*A*, *B*) → Major(*B*, *M*)
2.1: Major(*A*, *M*) ∧ Friends(*A*, *B*) → Major(*B*, *M*)
1.4: Studies(*A*, *C*) ∧ Dept(*C*, *D*) → Major(*A*, D)

Apple II image from wikipedia.com.
Eyes added digitally.

Course Goals

- Primer on the science and research of machine learning
- Understand the mathematics and theory behind methods
- Flipped format
  - Watch lectures online
  - Solve problems, discussion, work on homework in class sessions