Intro to Machine Learning

Crowdsourcing and Human Computation
Lecture 12

Instructor: Chris Callison-Burch
TA: Ellie Pavlick

Website: crowdsourcing-class.org

What is Machine Learning?

- How can we build computer systems that automatically improve with experience and what are the fundamental laws that govern all learning processes?

–Tom Mitchell, CMU

The age of machine learning is now

- In the 90s we rediscovered probabilistic models and statistics and applied it to CS and machine learning
- We now have access to much better computing infrastructure
- We have so much data that we can barely store it, and it provides great opportunities for analysis

What can we do with machine learning?

- Find groups of related things via “clustering”. Used for recommendations by Amazon, Netflix, etc
- Are two items the same? Named entity disambiguation
- Classification: Is this email spam? What language is this web page written in? Whose face is shown in a picture?

Your Amazon.com › Recommended for You

(If you're not Chris Callison-Burch, click here.)

These recommendations are based on items you own and more.

view: **All** | New Releases | Coming Soon

More results

Just For Today

Browse Recommended

Recommendations

Amazon Instant Video
Amazon MP3 Store
Appliances
Appstore for Android
Arts, Crafts & Sewing
Automotive
Baby
Beauty
Books
Books on Kindle
Camera & Photo
Cell Phones & Accessories
Clothing & Accessories
Computers
Electronics
Grocery & Gourmet Food
Health & Personal Care
Home & Kitchen
Home Improvement
Industrial & Scientific
Jewelry
Kitchen & Dining
Magazine Subscriptions
Magazines on Kindle

1.

Carter's Keep Me Dry Waterproof Fitted Quilted Crib Pad, White
by Kids Line (December 11, 2009)
Average Customer Review: ★★★★½ (383)
In Stock

List Price: ~~$12.99~~
Price: $12.79
17 used & new from $11.02

Add to Cart | Add to Wish List

I own it | Not interested | Rate this item

Recommended because you added **Summer Infant Contoured Changing Pad Amazon Frustration F...** to your Shopping Cart and more (Fix this)

2.

Nosefrida The Snotsucker Nasal Aspirator
by FridaBaby (April 1, 2010)
Average Customer Review: ★★★★½ (1,859)
In Stock

List Price: ~~$15.99~~
Price: $14.78
43 new from $9.86

Add to Cart | Add to Wish List

I own it | Not interested | Rate this item

Recommended because you added **Summer Infant Infant Character Change Pad Cover, Safari S...** to your Shopping Cart and more (Fix this)

3.

Safety 1st Heavenly Dreams White Crib Mattress
by Dorel Home Products (December 11, 2010)
Average Customer Review: ★★★★½ (627)
In Stock

List Price: ~~$54.99~~

amazon Prime
Chris's Amazon.
Shop by Department
Search All
Your Amazon.com
Your Browsing History
Your Amazon.com › Recomme
(If you're not Chris Callison-Burch, click here.)
Just For Today
Browse Recommended
Recommendations
Amazon Instant Video
Amazon MP3 Store
Appliances
Appstore for Android
Arts, Crafts & Sewing
Automotive
Baby
Beauty
Books
Books on Kindle
Camera & Photo
Cell Phones & Accessories
Clothing & Accessories
Computers
Electronics
Grocery & Gourmet Food
Health & Personal Care
Home & Kitchen
Home Improvement
Industrial & Scientific
Jewelry
Kitchen & Dining
Magazine Subscriptions
Magazines on Kindle
HalloWeen Shop
Shop now
0 Cart
Wish List
These
view:
More results
1.
2.
3.
hopping Cart and more (Fix
h List
pping Cart and more (Fix this)
Price: $52.99
Add to Cart
Add to Wish List

Amazon.com: Why is this recommended for you?

amazon.com

Help | Close window

Recommended for You

Nosefrida The Snotsucker Nasal Aspirator
by FridaBaby (April 1, 2010)
In Stock
List Price: ~~$15.99~~
Price: $14.78
43 new from $9.86

Add to Cart | Add to Wish List

Rate this item
☒ ☆☆☆☆☆
☐ I own it
☐ Not interested

Because you purchased...

GE 51386 Metal Shade With Flower Design Incandescent Night Light

☒ ☆☆☆☆☆
☐ This was a gift
☐ Don't use for recommendations

Munchkin Arm & Hammer Diaper Pail Refill Bags, 30 Count
by Munchkin

☒ ☆☆☆☆☆
☐ This was a gift
☐ Don't use for recommendations

My Brest Friend Original Pillow, Bluebells
by Zenoff Products

☒ ☆☆☆☆☆
☐ This was a gift
☐ Don't use for recommendations

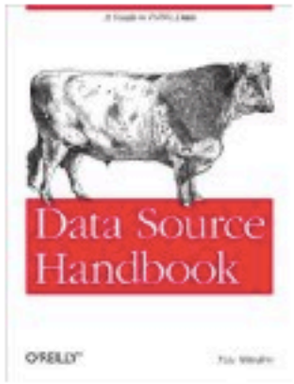
Because your Wish List includes...

WiFi Baby 2.0 (2013 Model) - iPhone, iPad, Android, Baby Monitor & Nanny Cam DVR. Video, Audio, Recording. Anywhere. Same Look, New Features (WFB2013)
by WiFi Baby

☒ ☆☆☆☆☆
☐ Don't use for recommendations

More Items to Consider

You viewed

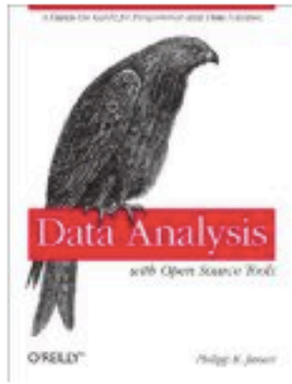


Data Source Handbook
› Pete Warden
Paperback
★★☆☆☆ (8)
~~$29.99~~ $20.78

Customers who viewed this also viewed

View or edit your browsing history ›

Big Data Glossary
› Pete Warden
Paperback
★★★☆☆ (5)
~~$19.99~~ $14.78



Data Analysis with Open Source Tools
› Philipp K. Janert
Paperback
★★★★☆ (35)
~~$39.99~~ $24.22

1. ASICS Women's GT-1000 Running Shoe
ASICS

2. ASICS Women's GEL-Noosa Tr Running Shoe
ASICS

3. ASICS Women's Gel-Nimbus 14 Running Shoe
ASICS

4. ASICS Women's Gel-Kayano 19 Running Shoe
ASICS

5. ASICS Women's GEL-Cumulus Running Shoe
ASICS

›See all Best Sellers in Women's Running

Related to Items You've Viewed

You viewed

Bose QuietComfort 20i Acoustic Noise...
★★★★½ (144)

Customers who viewed this also viewed

View or edit your browsing history ›

Bose IE2 Audio Headphones
★★★★½ (1,153)
$99.95

Bose® MIE2i Mobile Headset
★★★★☆ (632)
$129.95

Best Sellers

Gift Cards Store

Updated hourly

1. Amazon Gift Card - E-mail
Amazon
$50.00

2. Amazon Gift Card - E-mail - Hap Birthday (Candles)
Amazon
$50.00

3. Amazon Gift Card - E-mail - Tha You (Note)
Amazon
$50.00

4. Amazon Gift Card Upload Your Photo - Gift for You
Amazon
$50.00

5. Amazon.com Gift Cards - E-mail

More Items to Consider

You viewed

Customers who viewed this also viewed

View or edit your browsing history

1. ASICS Women's GT-1000 Running Shoe
ASICS

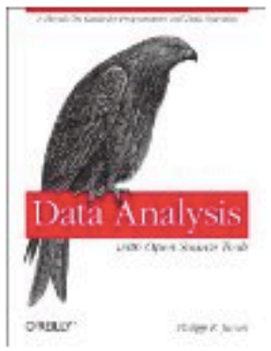
2. ASICS Women's GEL-Noosa T
Running Shoe
ASICS

3. ASICS Women's Gel-Nimbus 1
Running Shoe
ASICS

4. ASICS Women's Gel-Kayano 1

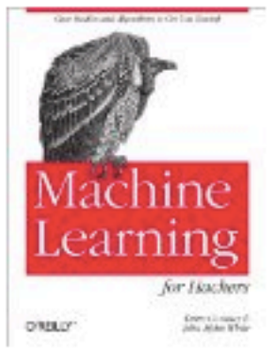
Continue Shopping: Customers Who Bought Items in Your Recent History Also Bought

Page 1 of 13

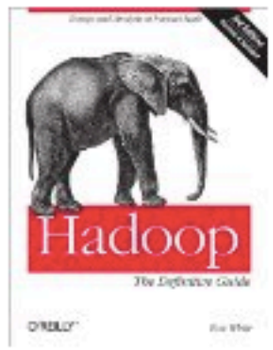


Data Analysis with Open Source Tools
› Philipp K. Janert
(35)
Paperback
$24.22 Prime

Python for Data Analysis
› Wes McKinney
(34)
Paperback
$24.68 Prime



Machine Learning for Hackers
› Drew Conway
(22)
Paperback
$33.66 Prime



Hadoop: The Definitive Guide
› Tom White
(36)
Paperback
$29.99 Prime

Bose QuietComfort 20i Acoustic Noise...
(144)

Bose IE2 Audio Headphones
(1,153)
$99.95

Bose® MIE2i Mobile Headset
(632)
$129.95

$50.00

3. Amazon Gift Card - E-mail - Tha You (Note)
Amazon
$50.00

4. Amazon Gift Card Upload Your Photo - Gift for You
Amazon
$50.00

5. Amazon.com Gift Cards - E-mail

Does Anne Hathaway News Drive Berkshire Hathaway's Stock?

ALEXIS C. MADRIGAL | MAR 18 2011, 10:50 AM ET

1
Like
483
Tweet
14
+1
258
Share
More

Given the awesome correlating powers of today's stock trading computers, the idea may not be as far-fetched as you think.

A couple weeks ago, Huffington Post blogger Dan Mirvish noted a funny trend: when Anne Hathaway was in the news, Warren Buffett's Berkshire Hathaway's shares went up. He pointed to six dates going back to 2008 to show the correlation. Mirvish then suggested a mechanism to explain the trend: "automated, robotic trading programming are picking up the same chatter on the Internet about 'Hathaway' as the IMDb's StarMeter, and they're applying it to the stock market."

The idea seems ridiculous. But the more I thought about the strange behavior of algorithmic trading systems and the news that Twitter sentiment analysis could be used by stock market analysts and the fact that many computer programs are simply looking for tradeable correlations, I really started to

VIDEO

Advice to a Younger Me: Michelle Peluso

The CEO of Gilt reflects on leadership and work-life balance

WRITERS

Molly Ball
Republicans Shut Down the Government for Nothing 11:17 AM ET

Julie Beck
Study: People Love to Cheat 11:10 AM ET

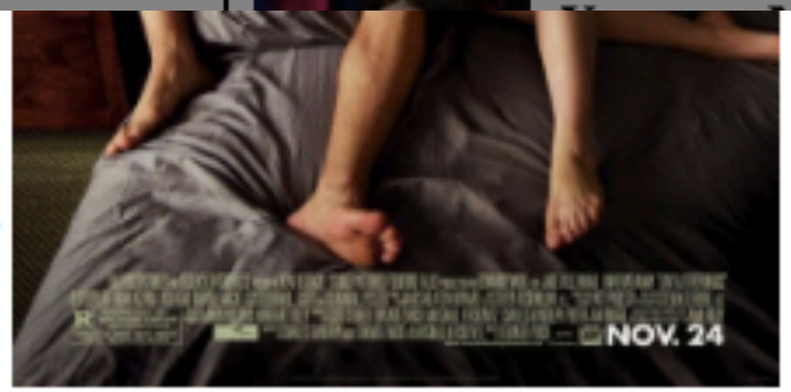
Matthew O'Brien
The Undeadline: Why Halloween Could (Seriously) Be Our Last Day to Save America From Default 10:56 AM ET

Kasia Cieplak-Mayr von Baldegg
The Highs (and Lows) of Starting Your Own Company 10:53 AM ET

Derek Thompson
American Politics Deserves to Be Downgraded 10:20 AM ET

Does Anne Hathaway News Drive Berkshire Hathaway's Stock?

ALEXIS C.



On the Friday before the Oscars, Berkshire shares rose a whopping 2.02%. And on the Monday just after the Academy Awards, they rose again, this time 2.94%. But it's not just an Oscar bounce, or something Warren Buffett may have said in the newspaper, or even necessarily something the company itself is doing (i.e. rumors afoot to buy Costco). Just look back at some other landmark dates in Anne Hathaway's still young career:

Oct. 3, 2008 - *Rachel Getting Married* opens: BRK.A up .44%
Jan. 5, 2009 - *Bride Wars* opens: BRK.A up 2.61%
Feb. 8, 2010 - *Valentine's Day* opens: BRK.A up 1.01%
March 5, 2010 - *Alice in Wonderland* opens: BRK.A up .74%
Nov. 24, 2010 - *Love and Other Drugs* opens: BRK.A up 1.62%
Nov. 29, 2010 - Anne announced as co-host of the Oscars: BRK.A up .25%

My guess is that all those automated, robotic trading programming are picking up the same chatter on the internet about "Hathaway" as the IMDb's StarMeter, and they're applying it to the stock market. Of course, this isn't necessarily bad news for the investor. After all, can you imagine what might have happened to Berkshire stock if Warren Buffett had appeared nude in *Love and Other Drugs* rather than Anne Hathway? Perhaps it's best if we don't think about it.

of algorithmic trading systems and the news that Twitter sentiment analysis could be used by stock market analysts and the fact that many computer programs are simply looking for tradeable correlations, I really started to

VIDEO
Advice to a
luso
ership
1
Like
483
Tweet
14
+1
258
Share
More
ent for
M ET
uld
merica
n
r Own
Derek Thompson
American Politics Deserves to Be Downgraded 10:20 AM ET

iPhoto

LIBRARY
- Events
- Photos
- Faces
- Places

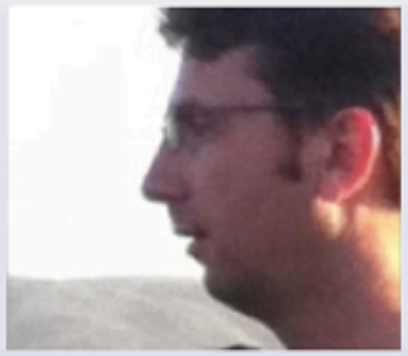
RECENT
- Oct 3, 2009
- Last 12 Months
- Last Import
- Flagged
- Trash

WEB
- Photo Stream
- callison-burch

DEVICES
- ccb's iPhone

EVENTS

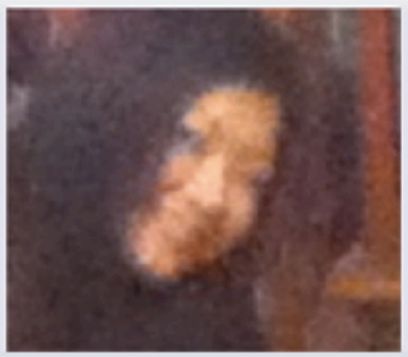
ALBUMS
- Last 12 Months
- 2nd floor bathroom
- 3029 Elm Ave


Alex Klementiev


Alexa Callison-Burch


Ann Irvine


Archie Burch

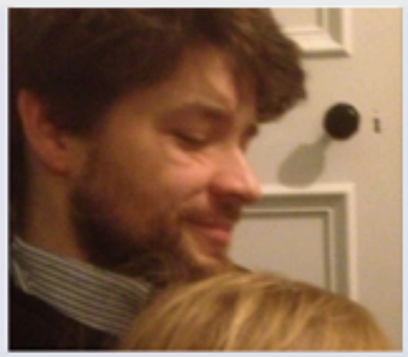

Charley Chan


Chris Callison-Burch


Chris Landers

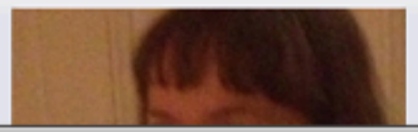

Colin Bannard

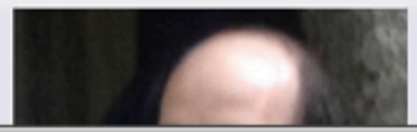

Danielle Matthews

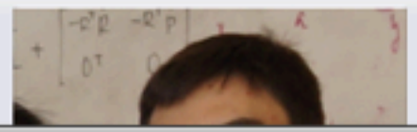

Dave Hawkey

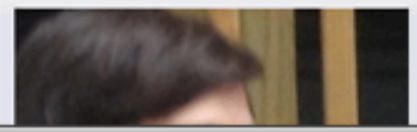

Dawn Mercurio


Diane Callison









Search Zoom Slideshow Find Faces

LIBRARY
- Events
- Photos
- Faces
- Places

RECENT
- Oct 3, 2009
- Last 12 Months
- Last Import
- Flagged
- Trash

WEB
- Photo Stream
- callison-burch

DEVICES
- ccb's iPhone

EVENTS

ALBUMS
- Last 12 Months
- 2nd floor bathroom
- 3029 Elm Ave


Chris Callison-Burch


Chris Landers


Colin Bannard


Courtney Napoles


Danielle Matthews

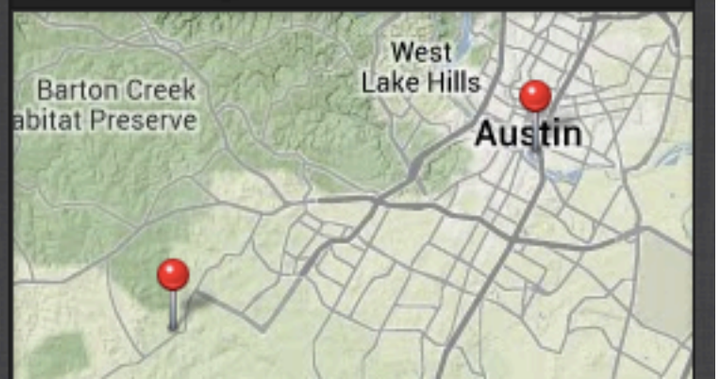

Dave Hawkey


Colin Bannard

Colin Bannard
bannard@mail.utexas.edu

12/6/12 - 12/8/12
3 photos
34 unconfirmed matches

Texas



iPhoto

Colin Bannard 37 photos Photos


Colin Bannard


Colin Bannard


Colin Bannard

Unconfirmed Faces


click to confirm


click to confirm


click to confirm

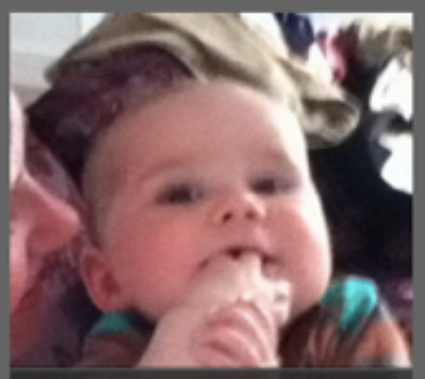

click to confirm


click to confirm

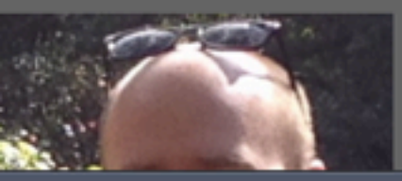

click to confirm

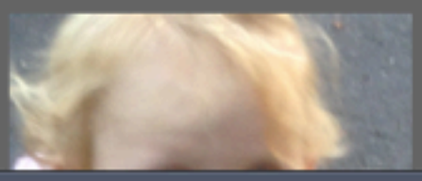

click to confirm

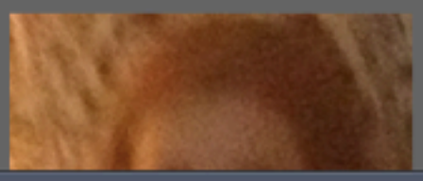

click to confirm

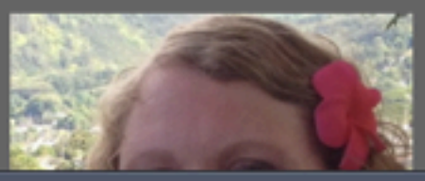

click to confirm

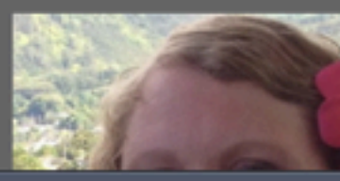

click to confirm





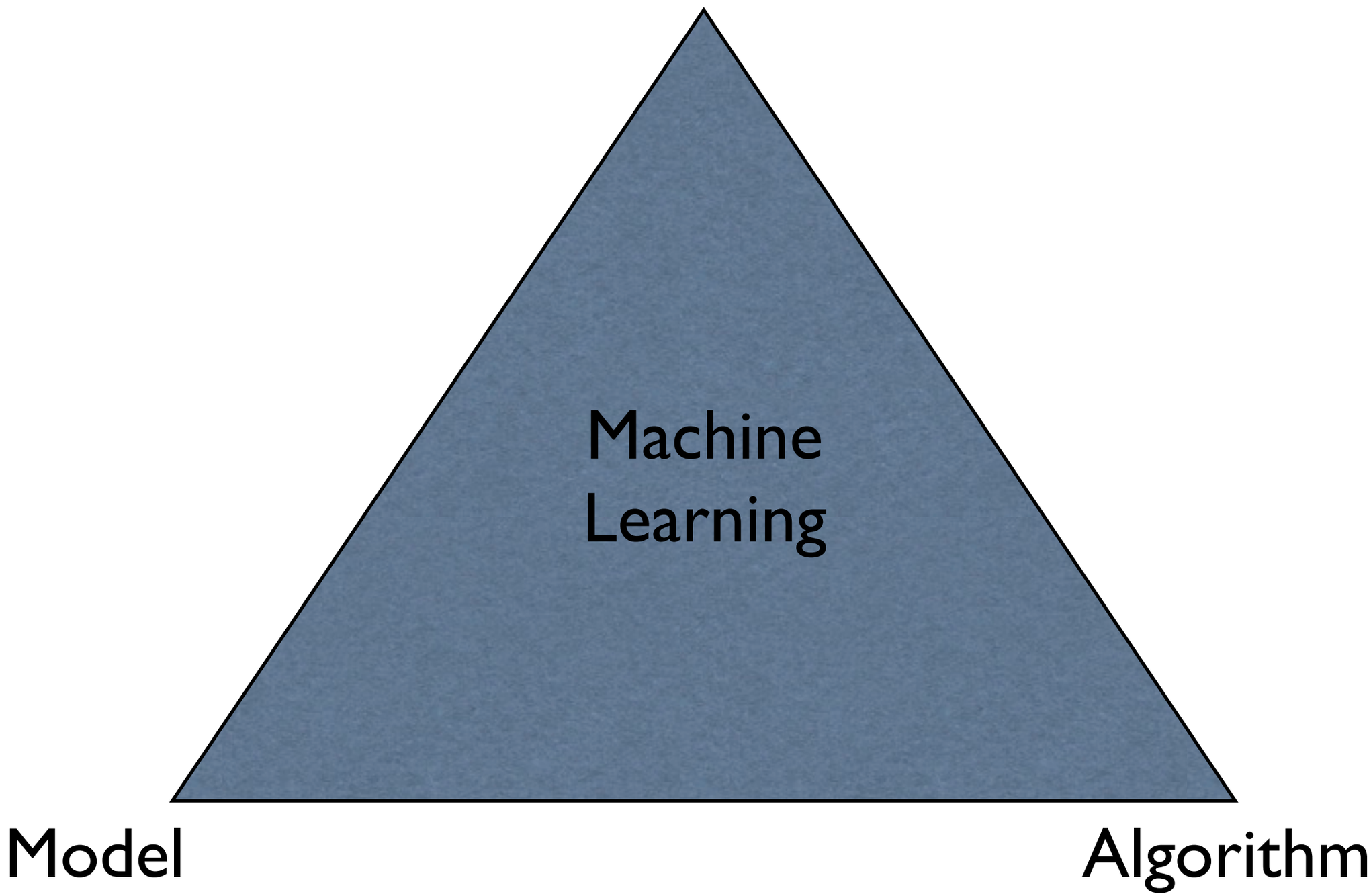






Click to confirm or Option-click to reject photos

Cancel


Data
Machine
Learning
Model
Algorithm

Supervised v. Unsupervised Learning

- In supervised learning you are starting with a labeled training set of data
- In unsupervised learning you don’t (yet) have that data

Kinds of data?

- Text and speech
- Images and video
- Geographic information
- Time series information
- Transaction data from customers
- Climate data
- Census data

Where does data come from?

- Some datasets are available for free: http://delicious.com/pskomoroch/ dataset
- Some is owned by companies
- Sometimes you can assemble it yourself
- Crowdsourcing!

Classification

- Classification is the assignment of a label to unlabeled input based on previously seen data
- Learn f(x)...
- that outputs a label ...
- along with a probability that that label is true

Example classification tasks

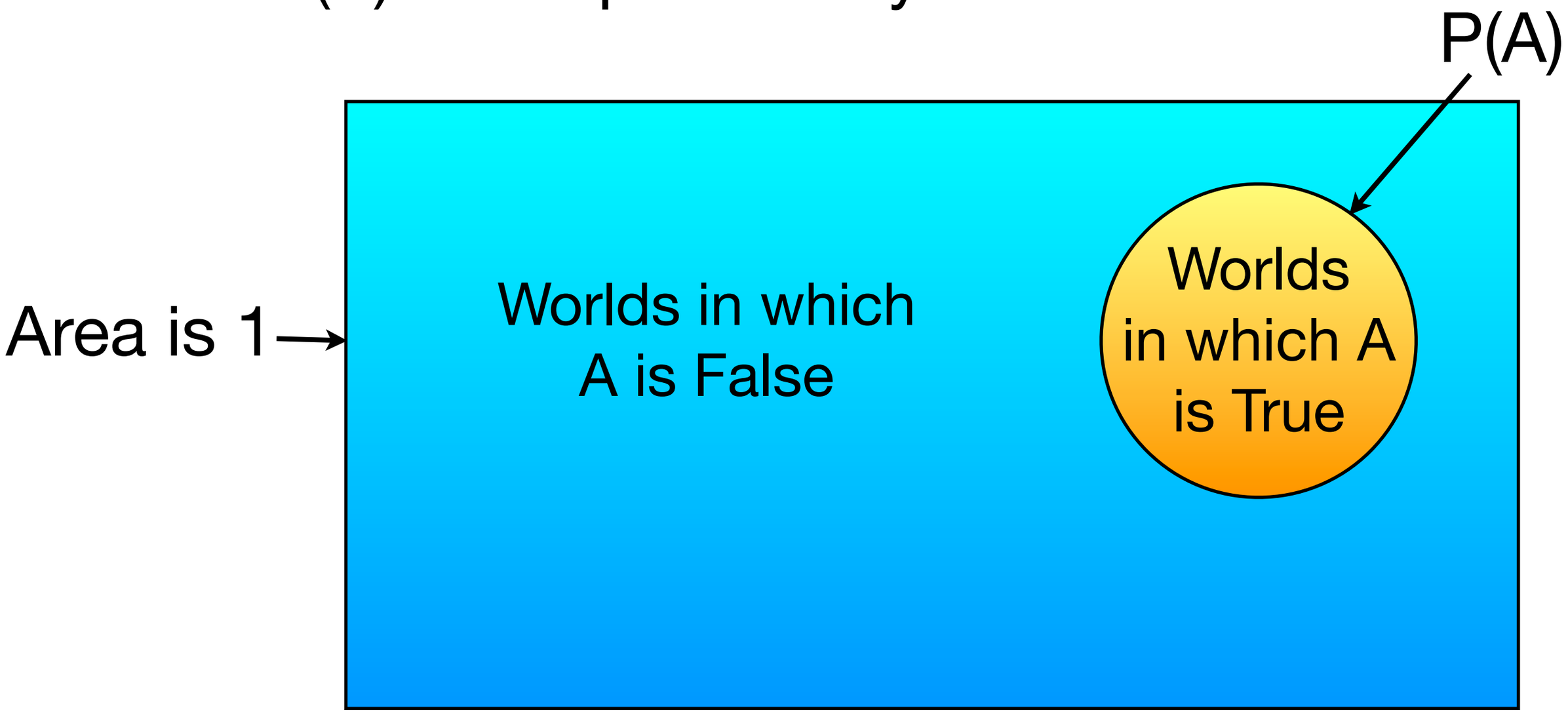
- Spam v. not Spam
- Face detection
- Language Identification

Naive Bayes

- First used for Spam filters in the 1990s
- Math is just counting, dividing and multiplying
- No calculus

Probability

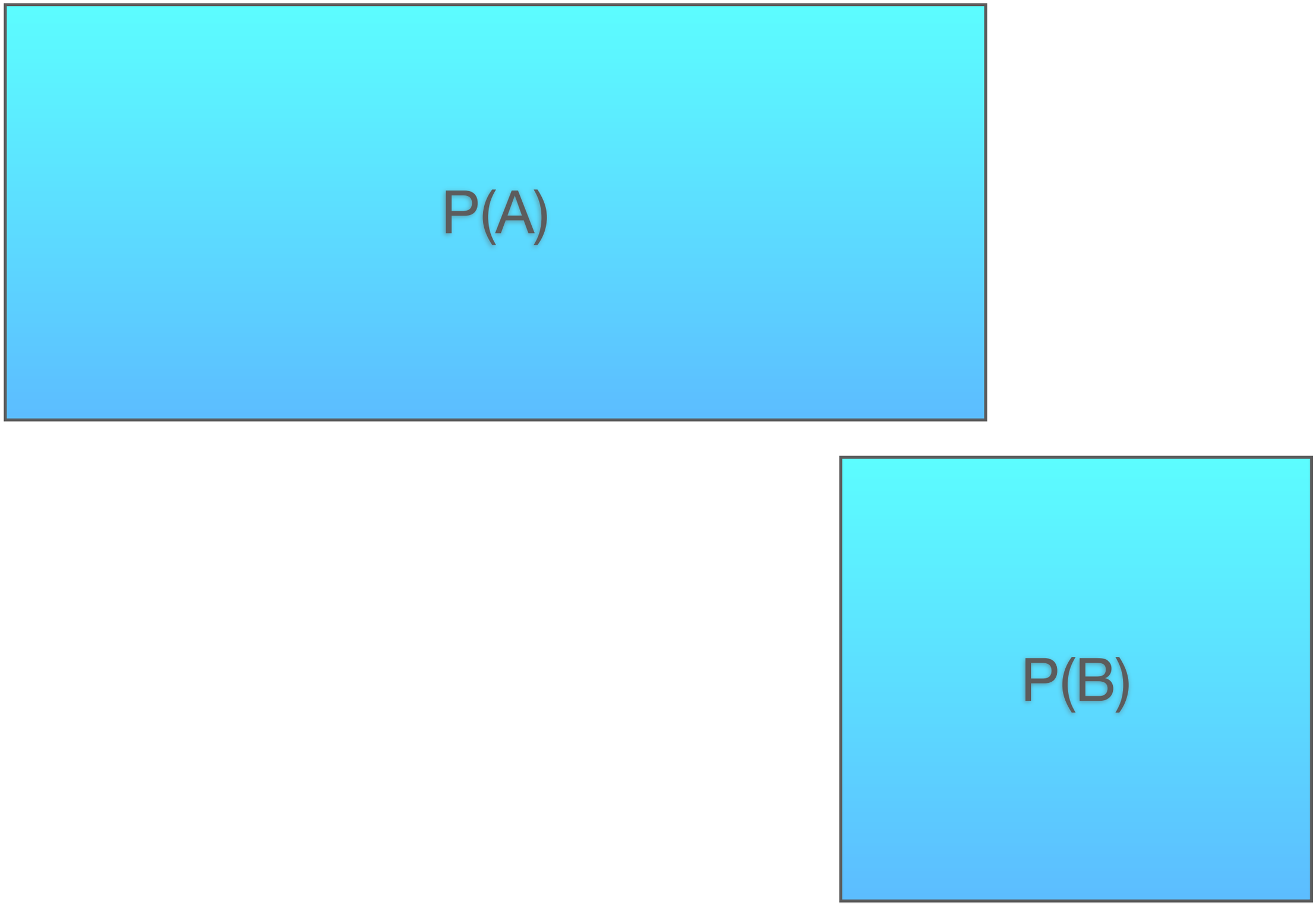
P(A) is the probability that A is true



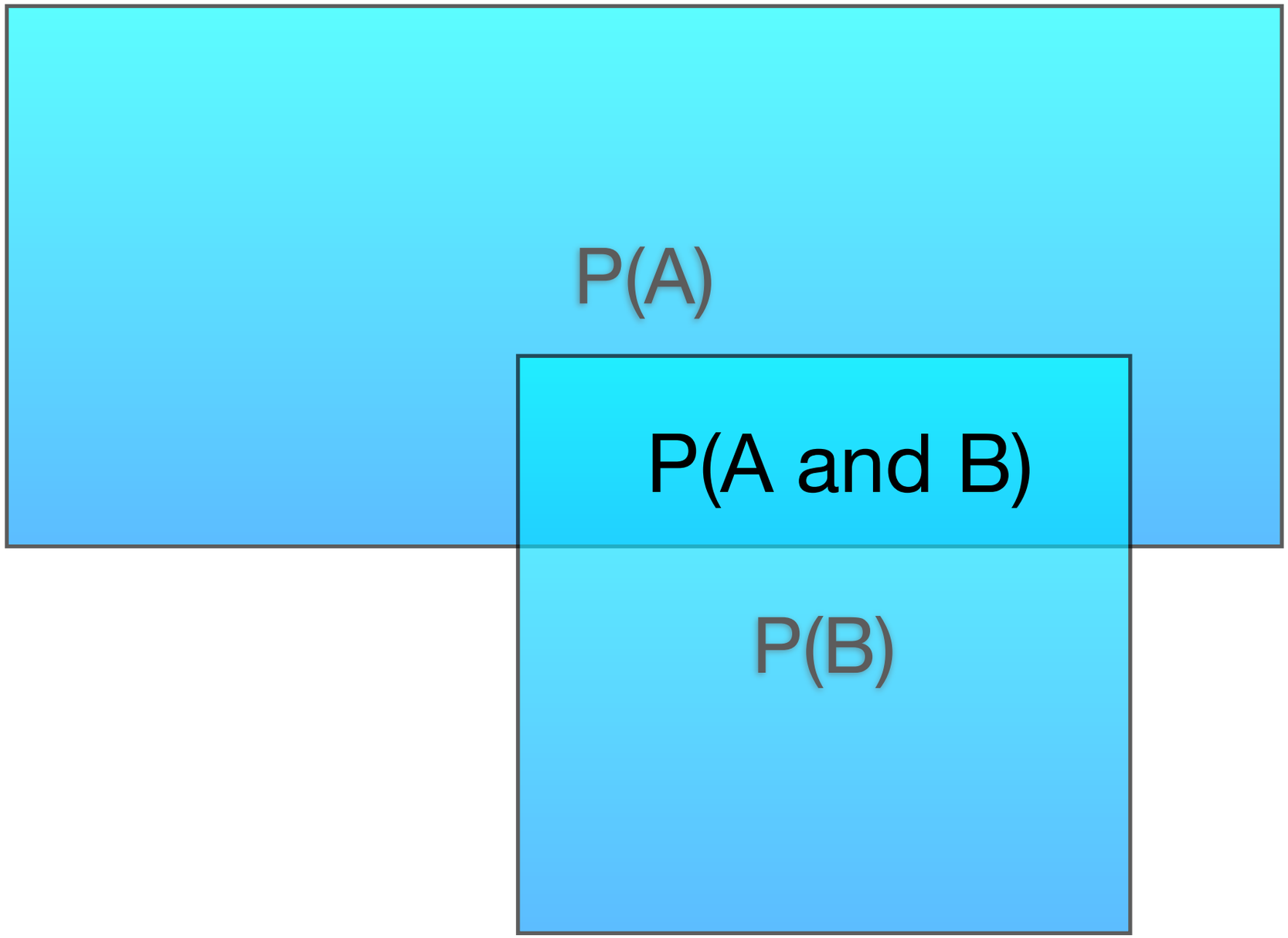
Axioms of Probability

- P(True) = 1
- P(False) = 0
- $0 \leq P(A) \leq 1$
- P(A or B) = P(A) + P(B) - P(A and B)

P(A or B) = P(A) + P(B) - P(A and B)



P(A or B) = P(A) + P(B) - P(A and B)



Bayes Law

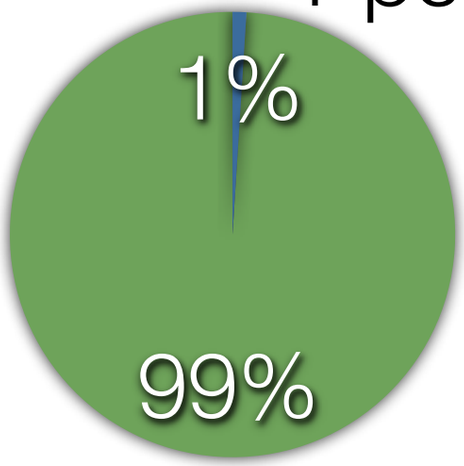
$$P(B \mid A) = P(A \mid B) * P(B) / P(A)$$

Example

- In a group of 10,000 people
- 1% of them have a rare disease
- There is a test that is 99% effective
  - 99% of sick patients test positive
  - 99% of healthy patients test negative
- Given a positive result, what is the probability that the patient are sick?

Sick population (100 people)

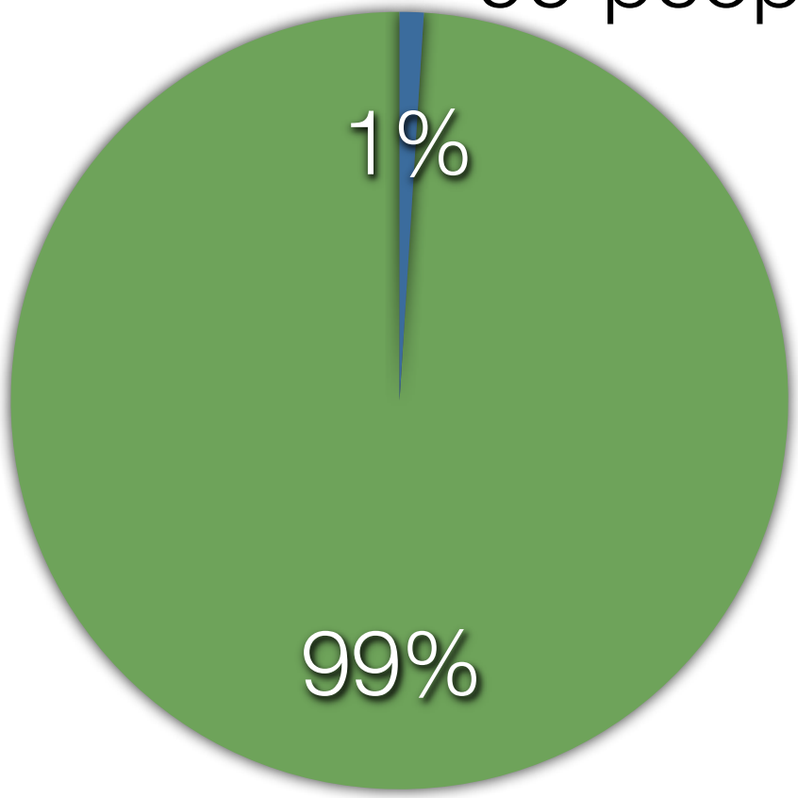
1 person



99 people

Healthy population (9900 people)

99 people



9801 people

False positives

True positives

False negative

True negatives

Disease diagnosis

- 99 sick patients test positive, 99 healthy patients test positive
- Give a positive test result, there is a 50% chance that a patient has the test result

Bayesian disease

$$p(\text{sick} \mid \text{test\_pos}) = \frac{\overbrace{p(\text{test\_pos} \mid \text{sick})}^{99/100} \; * \; \overbrace{p(\text{sick})}^{1/100}}{p(\text{test\_pos})} = \frac{99}{198} = \frac{1}{2}$$

$$99/10{,}000 + 99/10{,}000 = 198/10{,}000$$

(true positives + false positives)

We know the effectiveness of test (probability of testing positive given being sick), and the prior probability of being sick.

How do we use it for machine learning?

Naive Bayes Classifier

prior x likelihood

$$p(C|F_1,\ldots,F_n) = \frac{p(C)\ p(F_1,\ldots,F_n|C)}{\text{[redacted]}}.$$

posterior

evidence

$$p(C)\ p(F_1,\ldots,F_n|C) \ = \ p(C,F_1,\ldots,F_n)$$

Simplifying assumption

assume each feature is independent of the others

$$
\begin{aligned}
p(C|F_1,\ldots,F_n) &\propto p(C,F_1,\ldots,F_n) \\
&\propto p(C)\ p(F_1|C)\ p(F_2|C)\ p(F_3|C)\ \cdots \\
&\propto p(C)\prod_{i=1}^{n} p(F_i|C)\,.
\end{aligned}
$$

Text classification example

- NYTimes labels the type of each of its articles
- Has hand curated its categorization for the past 100 years
- Provides a Developer API
- You can get an API key from developer.nytimes.com

The New York Times

Developer Network BETA

API Documentation and Tools

You already know that NYTimes.com is an unparalleled source of news and information. But now it's a premier source of data, too — why just read the news when you can hack it?

- **Overview**
- APIs
- Keys
- Tools
- Forum
- Gallery
- API Console

Getting Started

The Times Developer Network is our API clearinghouse and community. Here's how to get started:

- Request an API key
- Read the API documentation, FAQ and Terms of Use
- Use the API Tool to experiment without writing code
- Browse the application gallery
- Connect with other developers in the forum

> To see your API keys and rate limits, visit the Keys page.

News

New York State Legislature API Deprecated

As of April 15, the New York State Legislature API is no longer supported. Learn more on the Open blog.

Article Search API: Full Service Restored

Jan. 30: After an outage over the weekend, the Article Search API is now working properly. Thanks for your patience.

API Logos and Branding Guide

Get Times API logos for your apps: review The New York Times Branding Guide and the Attribution Guidelines.

OPEN

A blog written by New York Times developers.

Go to Open Blog »

Blogroll

NOTABLE SITES

- Amazon Web Services
  The AWS blog
- BBC Backstage
  The BBC's developer network
- Daylife
- Google Code
- Mashable
- Programmable Web
- ReadWriteWeb
- Reuters Labs
- The Sunlight Foundation
- Yahoo Developer

OPEN-SOURCE SOFTWARE

- The Apache Software Foundation
- Django
- MySQL
- PHP
- Ruby on Rails

JOURNALISM & TECHNOLOGY

- Adrian Holovaty
  Creator of Django
- NICAR
  The National Institute for Computer-Assisted Reporting

Get the Silverlight kit for New York Times APIs.

Featured Applications and Projects

intro_web_data — emacs — 101×36

```
#!/usr/bin/env python
# encoding: utf-8
"""
nytimes_pull.py

Created by Hilary Mason on 2011-02-17.
Copyright (c) 2011 Hilary Mason. All rights reserved.
"""

import urllib
import json

def main(api_key, category, label):

    content = []
    for i in range(0,5):
        # print "http://api.nytimes.com/svc/search/v1/article?query=classifiers_facet:%s&api-key
offset=%s" % (category, api_key, i)
        h = urllib.urlopen("http://api.nytimes.com/svc/search/v1/article?query=classifiers_facet
api-key=%s&offset=%s" % (category, api_key, i))
        data = json.loads(h.read())
        for result in data['results']:
            content.append(result['body'])

    f = open(label, 'w')
    for line in content:
        f.write('%s\n' % line.encode('utf-8'))

    f.close()

if __name__ == '__main__':
    main("e561d1c482633983193 84f3e1343fe2c:19:68267067", "[Top/Features/Arts]","arts")
    main("e561d1c482633983193 84f3e1343fe2c:19:68267067", "[Top/News/Sports]","sports")
```

-uu-:---F1 **nytimes_pull.py**   All L1   (Python)----------

Loading python...done

Derian Hatcher scored midway through the third period and Antero Niittymaki stopped 38 shots, helpi
 the Philadelphia Flyers beat the Florida Panthers, 1-0, on Sunday for their third straight victory
It was the third career shutout and first of the season for Niittymaki, who has appeared in 111 gar
 over parts of four seasons with the Flyers.
The University of Hawaii's star-kissed run to the Sugar Bowl has so many improbable story lines tha
it practically reads like a movie script. The Warriors had four fourth-quarter comeback victories,
ey compete on a shoestring budget and their two best players, quarterback Colt Brennan and wide rec
ver Davone Bess, each spent time in jail before
Kerry Collins waited all season to prove his value in Tennessee. The Titans would not have been cel
rating Sunday night without him. Collins rescued his teammates in the most precarious of circumstar
s, leading them to three second-half field goals in relief of the injured Vince Young and helping T
nessee rally for a 16-10 victory over the
One of the risks the NFL Network faced by having CBS and NBC simulcast the Patriots-Giants game Sat
day night was the exposure to a wider world it would give Bryant Gumbel, the channel's play-by-play
nnouncer. The NFL Network reaches 43 million cable and satellite subscribers, but the simulcast inc
ased the viewing universe to 113 million TV
It almost sounds like a joke. Florida State's football team will be short-handed and vulnerable in
e Music City Bowl against Kentucky on Monday because the Seminoles will be missing players who were
uspended for cheating on, of all things, a music test. Ten to 15 years ago, at the height of its pr
inence in college football, Florida State was
Rutgers seemed refreshed after a nearly three-week layoff. Epiphanny Prince scored 20 points to lea
the No. 6 Scarlet Knights to a 70-34 victory against visiting Temple on Sunday. Rutgers had been on
for the last 18 days for exams, last playing Dec. 12 against Princeton. ''We had a chance to rechar
 ourselves after playing a brutal schedule
Whether slumped at his locker in a dress shirt and tie, or slumped on the bench in his blue warm-up
 Eddy Curry presented an apt portrait Sunday of all that has gone horribly wrong for the Knicks. Th
day began with Curry out of the starting lineup for the first time in two seasons, the latest despe
tion move by Coach Isiah Thomas to break his
For a split second in overtime Sunday night, Rangers center Chris Drury and forward Brendan Shanaha
spotted the same sliver of space that makes the difference between an inconsequential play and a ga
-winning goal. Drury snagged the puck at mid-ice after a Montreal turnover and saw Shanahan -- 38-y
r-old legs and all -- skating past the
This was at an N.B.A. Board of Governors meeting in New York not long ago, James L. Dolan facing hi
colleagues-in-ownership for the first time since the sexual harassment trial brought by Anucha Brow
 Sanders had slimed Madison Square Garden and by extension the league. According to a person who at

Google Prediction API

+1 907

What is the Google Prediction API?

Google's cloud-based machine learning tools can help analyze your data to add the following features to your applications:

- Customer sentiment analysis
- Spam detection
- Message routing decisions
- Upsell opportunity analysis
- Document and email classification
- Diagnostics
- Churn analysis
- Suspicious activity identification
- Recommendation systems
- ***And much more...***

RESTful API

Asynchronous cloud-based training, automatic model selection and tuning, and the ability to add training data on the fly.

Flexible Input

Numeric or text input that can output hundreds of discrete categories or continuous values.

Multi-Platform Support

Google App Engine, Apps Script (Google Docs), web & desktop apps, and the command line.

Open Source Code

- We will look at open source code in Python for doing Machine Learning
  - scikit-learn: classification, regression, clustering, etc.
  - Hilary Mason's video: Introduction to Machine Learning with Web Data
  - https://github.com/hmason/ml_class

```
#!/usr/bin/env python
# encoding: utf-8
"""
classify.py

Created by Hilary Mason on 2011-02-17.
Copyright (c) 2011 Hilary Mason. All rights reserved.
"""

import re, string

from nltk import FreqDist
from nltk.tokenize import word_tokenize
from nltk.stem.porter import PorterStemmer

class NaiveBayesClassifier(object):

    def __init__(self):
        self.feature_count = {}
        self.category_count = {}

    def probability(self, item, category):
        """
        probability: prob that an item is in a category
        """
        category_prob = self.get_category_count(category) / sum(self.category_count.values())
        return self.document_probability(item, category) * category_prob

    def document_probability(self, item, category):
        features = self.get_features(item)

        p = 1
        for feature in features:
            print "%s - %s - %s" % (feature, category, self.weighted_prob(feature, category))
-uuu:---F1  classify.py    Top L1     (Python)----------------------------------------
Mark set
```

```
        w_prob = ((weight*ap) + (totals * basic_prob)) / (weight + totals)
        return w_prob

    def train(self, item, category):
        features = self.get_features(item)

        for f in features:
            self.increment_feature(f, category)

        self.increment_cat(category)

if __name__ == '__main__':
    labels = ['arts', 'sports'] # these are the categories we want
    data = {}
    for label in labels:
        f = open(label, 'r')
        data[label] = f.readlines()
        # print len(data[label])
        f.close()

    nb = NaiveBayesClassifier()
    nb.train_from_data(data)
    print nb.probability("Early Friday afternoon, the lead negotiators for the N.B.A. and the playe
s union will hold a bargaining session in Beverly Hills — the latest attempt to break a 12-month st
lemate on a new labor deal.", 'arts')
    print nb.probability("Early Friday afternoon, the lead negotiators for the N.B.A. and the playe
s union will hold a bargaining session in Beverly Hills — the latest attempt to break a 12-month st
lemate on a new labor deal.", 'sports')
```

```
the - arts - 0.915670103093
a - arts - 0.618356164384
, - arts - 0.975051546392
. - arts - 0.154545454545
12-month - arts - 0.5
Beverly - arts - 0.193333333333
Early - arts - 0.5
Friday - arts - 0.132
Hills - arts - 0.5
N.B.A. - arts - 0.25
afternoon - arts - 0.17
and - arts - 0.7368
attempt - arts - 0.14
bargaining - arts - 0.5
break - arts - 0.25
deal - arts - 0.5
for - arts - 0.304444444444
hold - arts - 0.26
in - arts - 0.657837837838
labor - arts - 0.5
latest - arts - 0.25
lead - arts - 0.1
negotiators - arts - 0.5
new - arts - 0.155
on - arts - 0.460888888889
players - arts - 0.125
session - arts - 0.5
stalemate - arts - 0.5
to - arts - 0.598461538462
union - arts - 0.5
will - arts - 0.151666666667
— - arts - 0.5
6.1474279934e-16
```

```
the - sports - 0.994845360825
a - sports - 0.815616438356
, - sports - 0.935463917526
. - sports - 0.118181818182
12-month - sports - 0.5
Beverly - sports - 0.166666666667
Early - sports - 0.5
Friday - sports - 0.132
Hills - sports - 0.5
N.B.A. - sports - 0.26
afternoon - sports - 0.125
and - sports - 0.7368
attempt - sports - 0.155
bargaining - sports - 0.5
break - sports - 0.26
deal - sports - 0.5
for - sports - 0.578222222222
hold - sports - 0.25
in - sports - 0.795945945946
labor - sports - 0.5
latest - sports - 0.26
lead - sports - 0.164
negotiators - sports - 0.5
new - sports - 0.14
on - sports - 0.421777777778
players - sports - 0.17
session - sports - 0.5
stalemate - sports - 0.5
to - sports - 0.677230769231
union - sports - 0.5
will - sports - 0.133333333333
— - sports - 0.5
2.06771958824e-15
```