

# Machine learning and classification

# Outcomes

Explain the core aspects of Machine Learning

Apply classification to a set of data to create a decision tree

Write a machine learning algorithm

# Machine Learning

“construction and study of systems that can learn from data”

Can be seen as building blocks to make computers learn to behave more intelligently

There are various *techniques* with various *implementations*.



# Question?

What instances of machine learning  
have you come across?

What types are they?



# Use-Cases

- Spam Email Detection
- Machine Translation (Language Translation)
- Image Search (Similarity)
- Clustering (KMeans) : Amazon Recommendations
- Classification : Google News

# Use-Cases (contd.)

- Text Summarization - Google News
- Rating a Review/Comment: Yelp
- Fraud detection : Credit card Providers
- Decision Making : e.g. Bank/Insurance sector
- Sentiment Analysis
- Speech Understanding – iPhone with Siri
- Face Detection – Facebook's Photo tagging

# Not Spam

Delete all spam messages now (messages that have been in Spam more than 30 days will be automatically deleted)			
<input type="checkbox"/>			HDFC Bank
LOAN upto Rs 25 lac - Disbursal in 2 days - open in fresh tab -- If you do not want to receive any more newsletters, please click here			9:40 pm
<input type="checkbox"/>			iEntry
Welcome iEntry Member - Ultimate Guide To Assessing			9:23 pm
<input type="checkbox"/>			New-Zealand-Jobs.067L
Come to New Zealand to find a great job and settle here (Search for all Jobs from diffe... - Search for all Jobs from different kinds of industries Find a Job in Enchanting			8:18 pm
<input type="checkbox"/>			CarSizzler
Assured Free Luxurious Ride worth Rs.300 with Uber Cabs - Home Home Buy New Car Buy New Car Sell Car Sell Car Tech Tics Tip & Tale Facebook 41727 others			6:05 pm
<input type="checkbox"/>			Supermarket Promotion
Enjoy Rs.1700 voucher valid at any supermarket! - If you are unable to view this mailer Click here HOW TO CONTACT US? BY EMAIL: support@savethedeals.in			4:51 pm
<input type="checkbox"/>			Entireweb Newsletter
Hire an SEO the Right Way -- 6 Tips You Must Remember for Life - Unsubscribe me View web version Become a fan on Facebook Follow us on Twitter September 5th, 21			1:24 pm
<input type="checkbox"/>			Max Bupa
A policy that understands your family's medical need - open in fresh tab -- If you do not want to receive any more newsletters, please			11:08 am
<input type="checkbox"/>			Scoop.it
Your Scoop.it Daily Summary - How to Maximize Your LinkedIn Publishing Exposure   SME a... - Scoop.it Facebook Twitter G+ H			9:30 am
<input type="checkbox"/>			standard charterer Bank
Instant approval on your Credit Card			7:27 am
<input type="checkbox"/>			CAR TRADE
Sell your car at no cost at all - If you are having trouble viewing this email,view web version   View this message in your mobile			Sep 4
<input type="checkbox"/>			Uday
VPS Web Hosting Services Provider - Dear Sir, I am Uday Sharma, Business development executive. We are providing quality VPS hosting for			Sep 4
<input type="checkbox"/>			Mark Regan, SPN
How to Find Your Most Valuable Keywords [Free Guide] - This is a SiteProNews/ExactSeek Webmaster Exclusive Mailing! To drop your subscription, use the link			Sep 4
<input type="checkbox"/>			HDFC Bank
LOAN upto Rs 25 lac - Disbursal in 2 days - open in fresh tab You have received this mailer from Shop@Best on behalf of HDFC Bank because you			Sep 4
<input type="checkbox"/>			CAR TRADE
Sell your car at no cost at all - open in fresh tab -- If you do not want to receive any further newsletters, please click here			Sep 4
<input type="checkbox"/>			ICICI Bank
Home Loan Interest Rate starting from 10.15%*. Get Instant Approval! - open in fresh tab -- If you do not want to receive any more newsletters, please Click Here			Sep 4
<input type="checkbox"/>			calculateyourwealth
It's good when your bank helps you manage your wealth and fulfill your ambitions - Calculate Now Dreams you wish to realize in your lifetime require enough wealth. C			Sep 4
<input type="checkbox"/>			Angel Broking
Get Low Brokerage - Free Demat & Trading Account - open in fresh tab -- If you do not want to receive any further newsletters, please click here			Sep 4
<input type="checkbox"/>			Bankbazaar
7 Minute Instant Online Approval for your PESONAL LOAN - Now get instant online Personal Loan approval in 7 minutes by BankBazaar.com from leading Banks in			Sep 4
<input type="checkbox"/>			Jayde
Welcome To The Jayde Newsletter! - Welcome To WebProNews Welcome To The Jayde Newsletter! Before we begin, make sure to add			Sep 4
<input type="checkbox"/>			ineedhits noreply
[ineedhits] Your ineedhits Account and Password - ACCOUNT CREATION Account ID : A1588368 Dear Rah, Welcome to ineedhits. Yo			Sep 4
<input type="checkbox"/>			Rekha
Mobility Apps for Your Business - While we look at the span of last 20 years, we could broadly look at two distinct eras, - Life in			Sep 4
<input type="checkbox"/>			SlideShare Newsletter
Top Tips From the World Champions of PowerPoint - View online version Remember to display images Meet the PowerPoint World Champs Top Tips From the			Sep 4
<input type="checkbox"/>			Dilshad Pathan
Feeling Hesitate to Discuss personal Health Queries - My Life Care Follow Us on facebook twitter linkedin Google+ Feeling Hesitate to Discuss personal			Sep 4
<input type="checkbox"/>			Vaishu
TAKE YOUR PICK. Register in SimplyMarry - TAKE YOUR PICK. Register in SimplyMarry -- Regards Vaishu			Sep 4

# NER (Named Entity Recognition)

## Stanford Named Entity Tagger

Classifier:

Output Format:

Preserve Spacing:

Please enter your text here:

When Mike Brannigan was 18 months old, he was diagnosed with autism. At the time, his doctors said he would likely need a special school and a group home. His mom, Edie, admits she thought he'd "never be able to function in the world." Fast-forward several years. Brannigan is now 17, and is a senior at Northport High School, a public school in Long Island, New York. He's doing well academically, he has friends -- and he also happens to be one of the best young athletes in the country. Continue Reading...

When **Mike Brannigan** was 18 months old, he was diagnosed with autism. At the time, his doctors said he would likely need a special school and a group home. His mom, **Edie**, admits she thought he'd "never be able to function in the world." Fast-forward several years. Brannigan is now 17, and is a senior at **Northport High School**, a public school in **Long Island, New York**. He's doing well academically, he has friends -- and he also happens to be one of the best young athletes in the country. Continue Reading...

Potential tags:

**LOCATION**

**TIME**

**PERSON**

**ORGANIZATION**

**MONEY**

**PERCENT**

**DATE**



# Similar/Duplicate Images

About 81 results (0.70 seconds)



Image size:  
250 × 321

No other sizes of this image found.

Best guess for this image: *taj mahal*

Visually similar images

Report images



**Check this :**

LIRE (Lucene Image REtrieval) library -  
<https://code.google.com/p/lire/>

Credit: <https://www.google.co.in/>

## Remember

### Features ?

(Feature Extraction)

Can be :

- Width
- Height
- Contrast
- Brightness
- Position
- Hue
- Colors

# Recommendations

**More Items to Consider**

**You looked at**

**You might also consider**

**JavaScript: The Good Parts** Paperback by Douglas Crockford  
\$29.99 **\$19.79**

**JavaScript: The Definitive Guide** Paperback by David Flanagan  
\$49.99 **\$31.49**

**CSS: The Missing Manual** Paperback by David McFarland  
\$34.99 **\$23.09**

**Learning jQuery 1.3** Paperback by Chaffer, Kar  
\$39.99 **\$35.99**

[Find similar items](#)

**Related to Items You've Viewed**

**You looked at**

**You might also consider**

**Forms that Work: Designing Web Forms for Usability** Paperback by Bruce Tognarelli  
\$24.99 **\$19.99**

**Don't Make Me Think: A Common Sense Approach to Web Usability** Paperback by Steve Krug  
\$19.99 **\$14.99**

**Letting Go of the Words: Writing Web Content that Sings** Paperback by Linda Ward Beech  
\$24.99 **\$19.99**

**Designing Web Interfaces: Principles for Creating Great User Experiences** Paperback by Steve Krug  
\$24.99 **\$19.99**

**Today's Recommendations For You**

Here's a daily sample of items recommended for you. Click here to [see all recommendations](#)

**Even Faster Web Sites: Performance Tuning for Developers** (Paperback) by Steve Souders  
★★★★★ (7) **\$23.10**  
[Fix this recommendation](#)

**Simply JavaScript** (Paperback) by Kevin Yank  
★★★★☆ (19) **\$26.37**  
[Fix this recommendation](#)

**The Art & Science of JavaScript** (Paperback) by John Resig  
★★★★★ (5) **\$24.99**  
[Fix this recommendation](#)

**Any Category** Algorithms Boxed Sets Business & Culture Java  
Graphic Design Microsoft Networking Networks, Protocols & APIs New SQL

# Terminology

## Features

The number of features or distinct traits that can be used to describe each item in a quantitative manner.

## Samples

A sample is an item to process (e.g. classify). It can be a document, a picture, a sound, a video, a row in database or CSV file, or whatever you can describe with a fixed set of quantitative traits.

## Feature vector

is an  $n$ -dimensional vector of numerical features that represent some object.

## Feature extraction

Preparation of feature vector

Transforms the data in the high-dimensional space to a space of fewer dimensions.

## Training/Evolution set

Set of data to discover potentially predictive relationships.



# Learning (Training)



Features:

1. Color: **Radish/Red**
  2. Type : **Fruit**
  3. Shape
- etc...



Features:

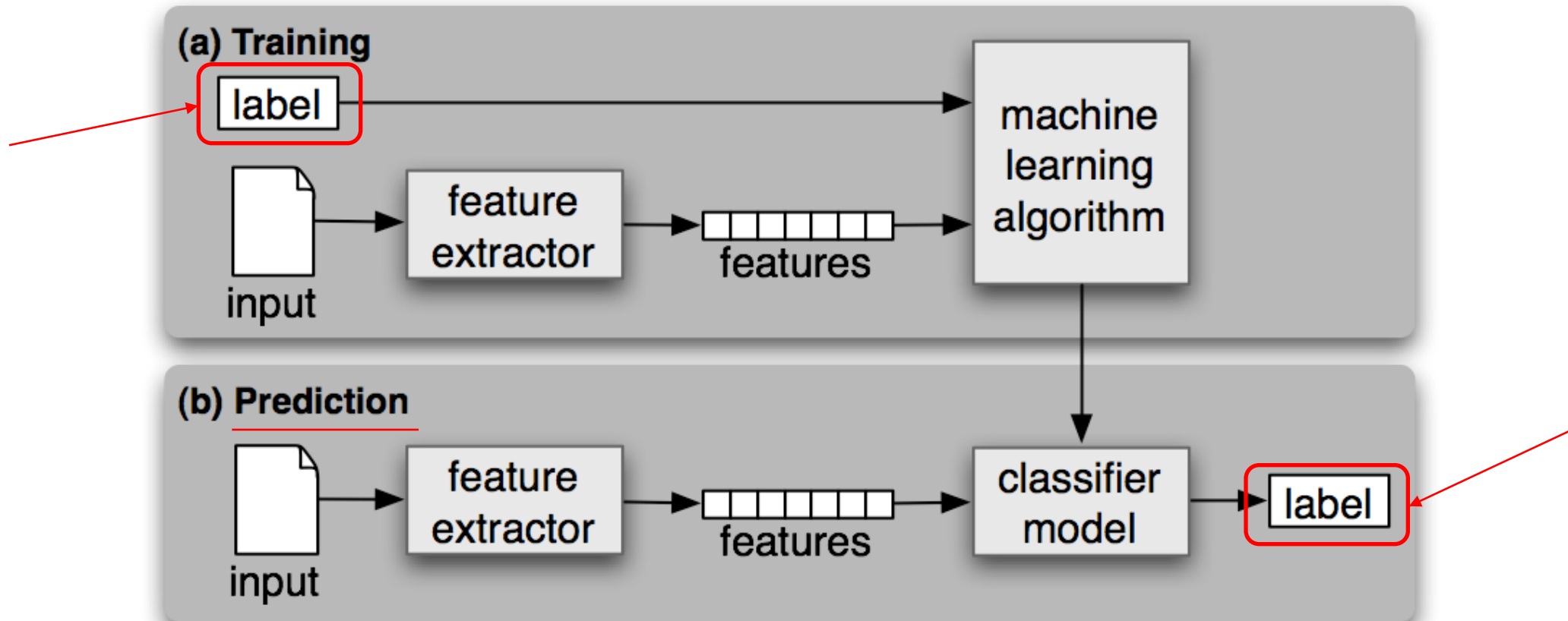
1. Sky Blue
  2. **Logo**
  3. Shape
- etc...



Features:

1. **Yellow**
  2. **Fruit**
  3. Shape
- etc...

# Workflow



# Property locations?

Each table has a set of “Top Trump” training-set cards.

Build a decision tree to sort them into “New York” and “San Francisco”.

You cannot use the target to sort them.



# Categories

Supervised Learning

Unsupervised Learning

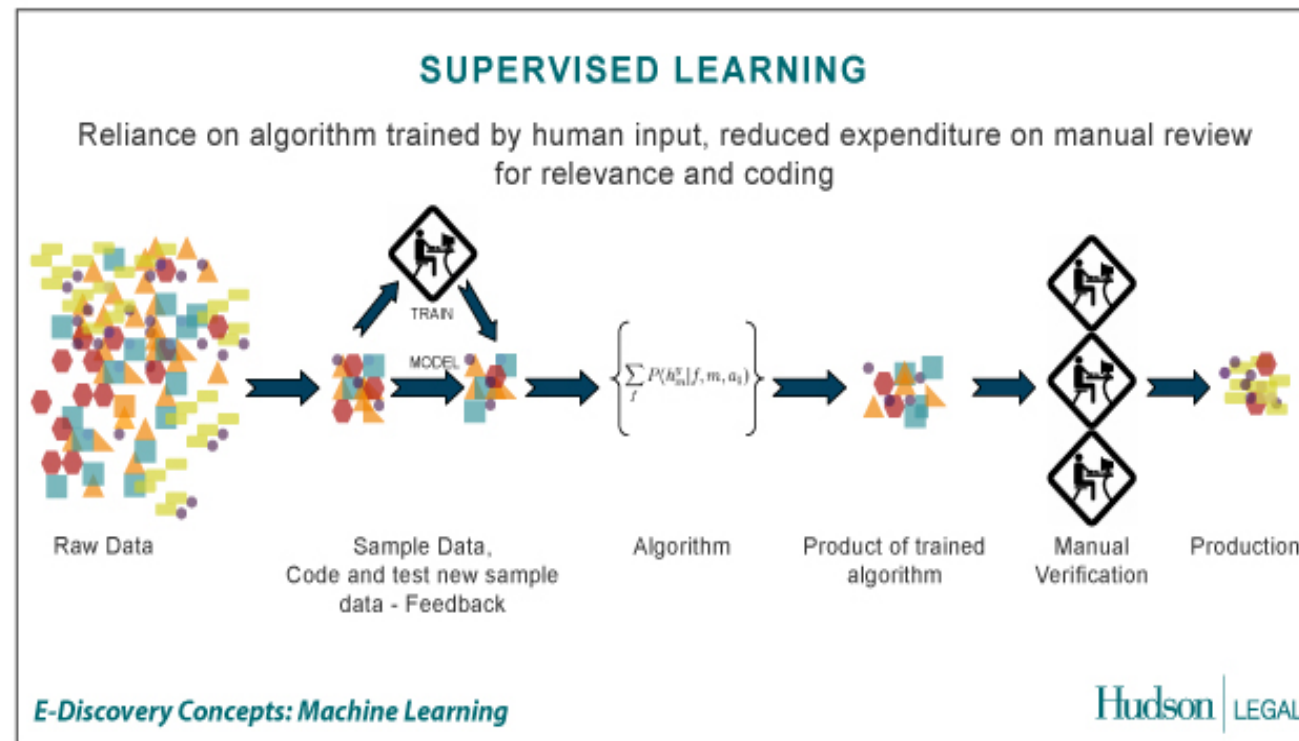
Semi-Supervised Learning

Reinforcement Learning



# Supervised Learning

The correct labels of the training data are known



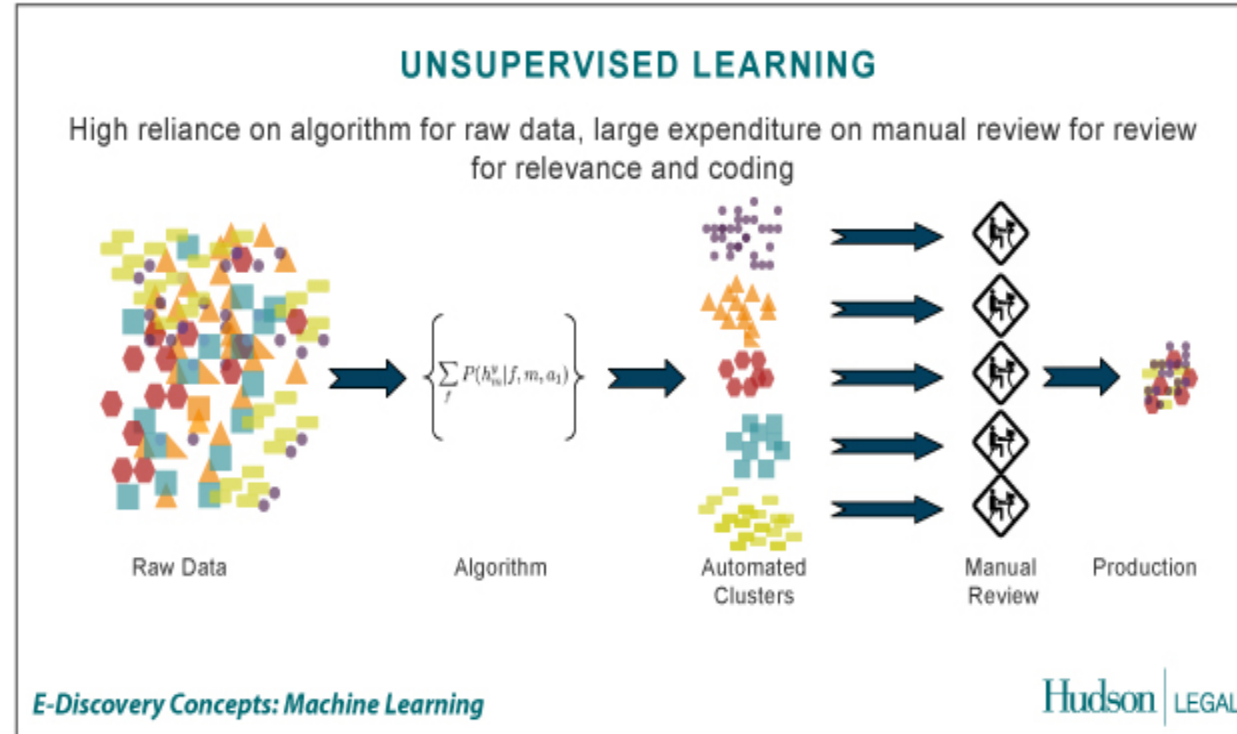
Credit: <http://us.hudson.com/legal/blog/postid/513/predictive-analytics-artificial-intelligence-science-fiction-e-discovery-truth>

Slide ideas thanks to @rahuldausa



# Unsupervised Learning

The correct labels of the training data are not known. Cluster discovery and pattern generation.



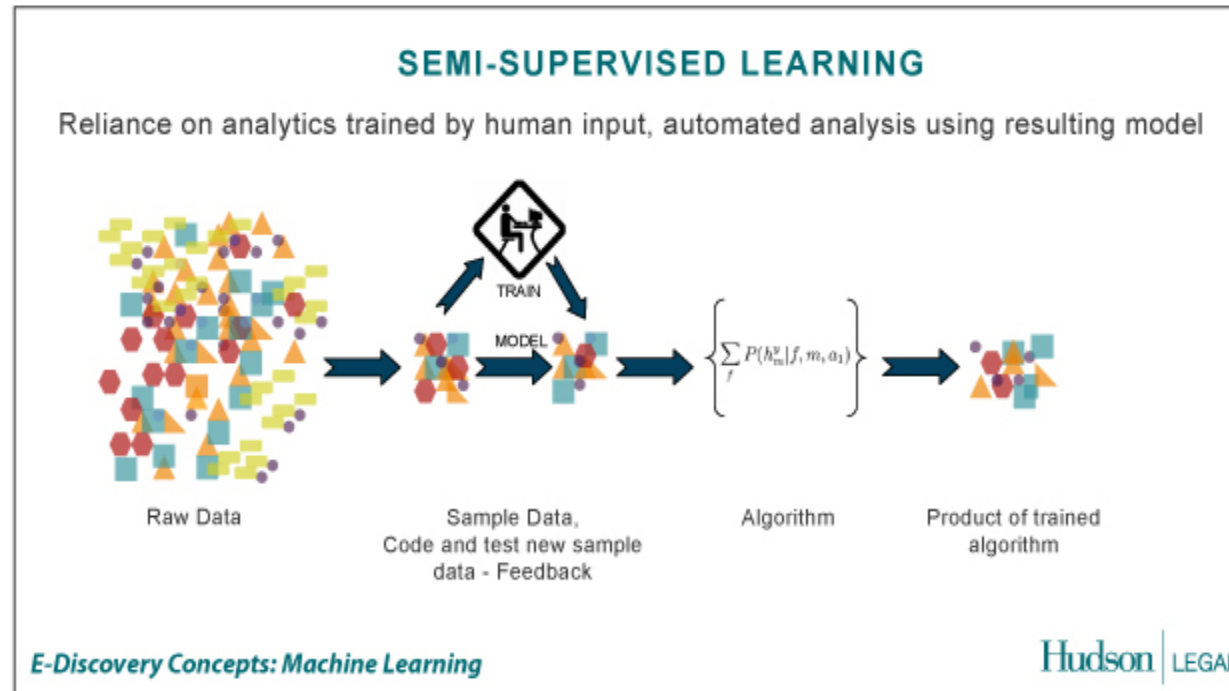
Credit: <http://us.hudson.com/legal/blog/postid/513/predictive-analytics-artificial-intelligence-science-fiction-e-discovery-truth>

Slide ideas thanks to @rahuldausa

# Semi-Supervised Learning

A Mix of Supervised and Unsupervised learning.

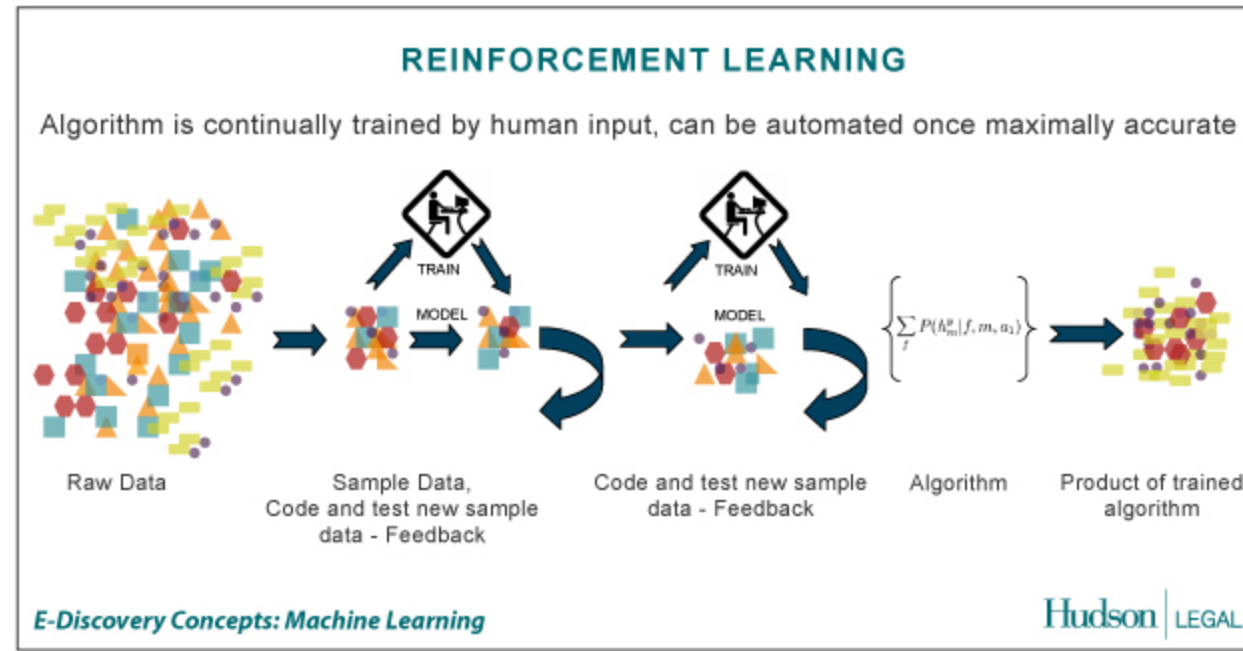
Some labels are known, but majority are not.



Credit: <http://us.hudson.com/legal/blog/postid/513/predictive-analytics-artificial-intelligence-science-fiction-e-discovery-truth>

# Reinforcement Learning

Software agent learns based on feedback and/or reward.



# Techniques

## **classification**

predict class from observations

## **clustering**

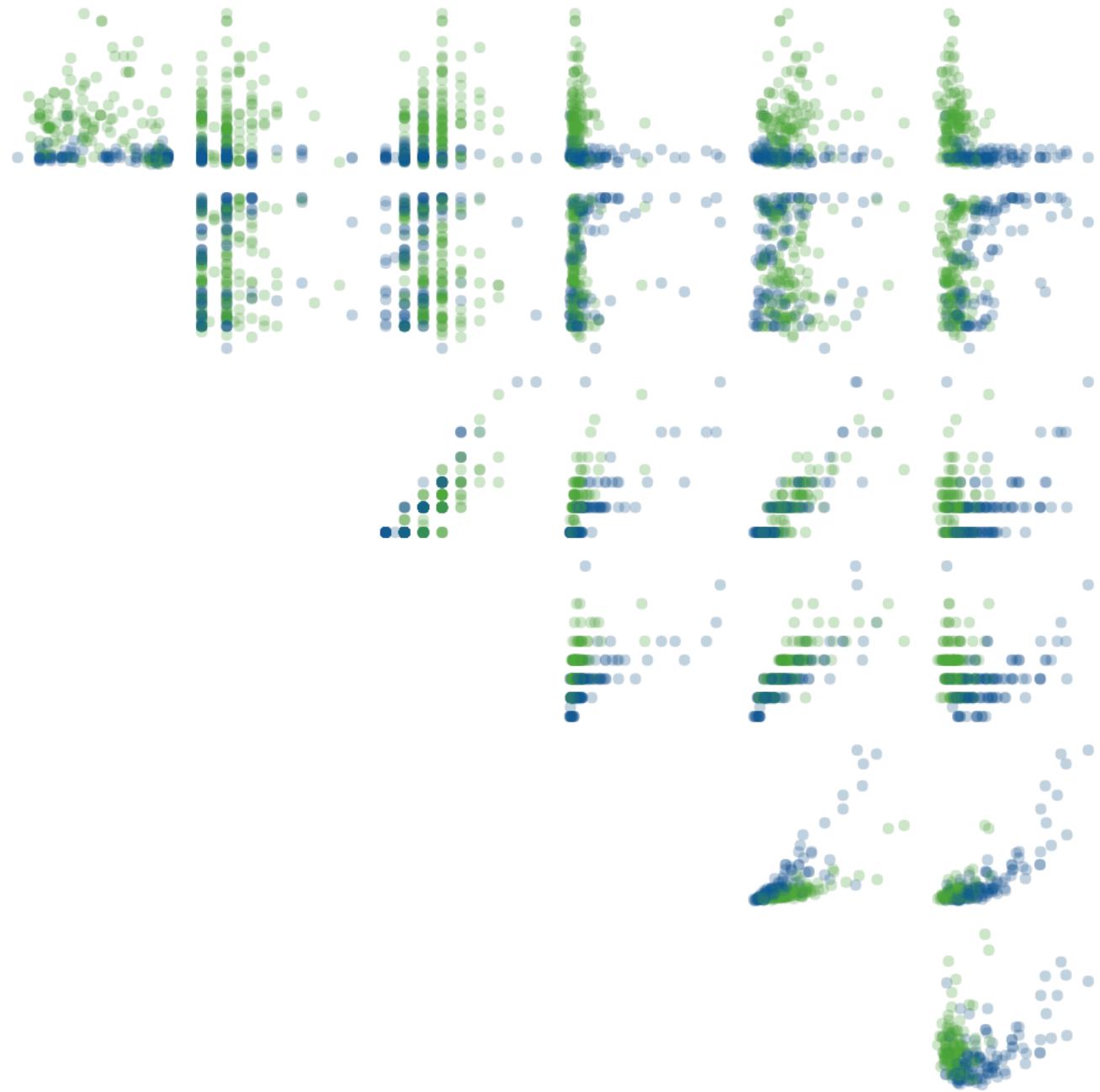
group observations into “meaningful” groups

## **regression (prediction):**

predict value from observations



r2d3.us



# Frameworks/Tools

R

Weka

Carrot2

Gate

OpenNLP

LingPipe

Stanford NLP

Mallet – Topic Modelling

Gensim – Topic Modelling (Python)

Apache Mahout

MLib – Apache Spark

scikit-learn - Python

LIBSVM : Support Vector Machines

and many more...

