Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Use Blackboard´s forum if a question may be relevant to other students, too. Email always both joeran.beel@scss.tcd and doug.leith@scss.tcd.ie. Give a meaningful subject, starting with "*[ML1819]*". No file attachments.

Week 01 (2): Introduction to Machine Learning

CS7CS4/CS4404 Machine Learning

v4 2018-09-22 16:06

Dr Joeran Beel
Assistant Professor in Intelligent Systems
Department of Computer Science and Statistics
Trinity College Dublin, Ireland

Dr Douglas Leith
Professor in Computer Systems
Department of Computer Science and Statistics
Trinity College Dublin, Ireland

Any questions?

https://kaiserhealthnews.files.wordpress.com/2017/02/khnoncall-2-2.jpg?w=1024

Update

- **Have a look at slides from previous lecture for literature recommendation, and marks from previous year**
- **Begin to find your team already now**
  - 3 Members
  - Ideally same module code
  - Ideally 1 native English speaker

Outline

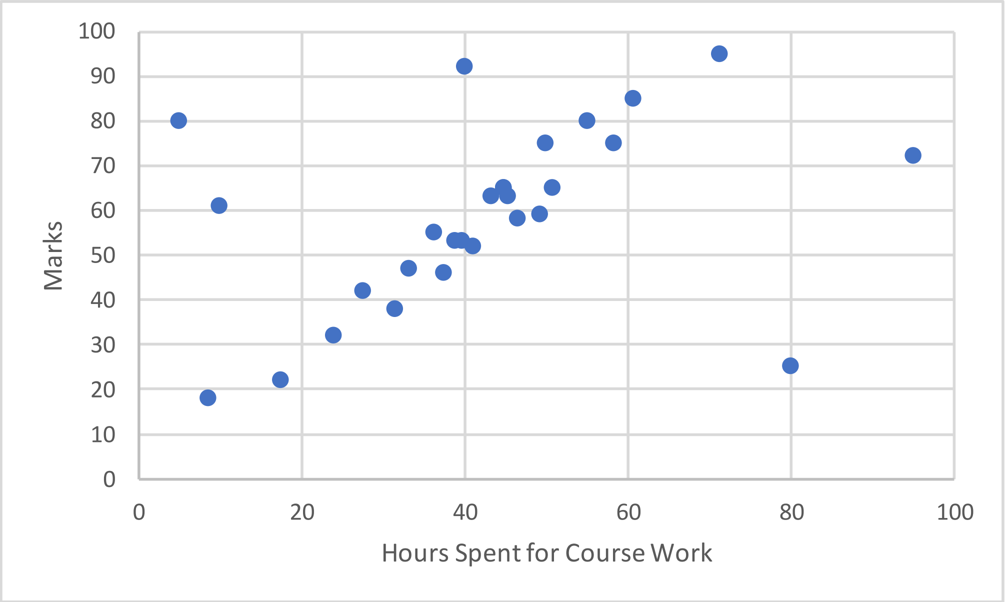
1. **Machine Learning Examples**
2. **Types of Machine Learning**
3. **Definition**
4. **Traditional Approaches to Problem Solving**
5. **Overview of the Machine Learning Pipeline**
6. **The Machine Learning Landscape**
7. **Strength and Weaknesses of Machine Learning**

Machine Learning Examples

Example 1: Regression

Predict Marks for Students

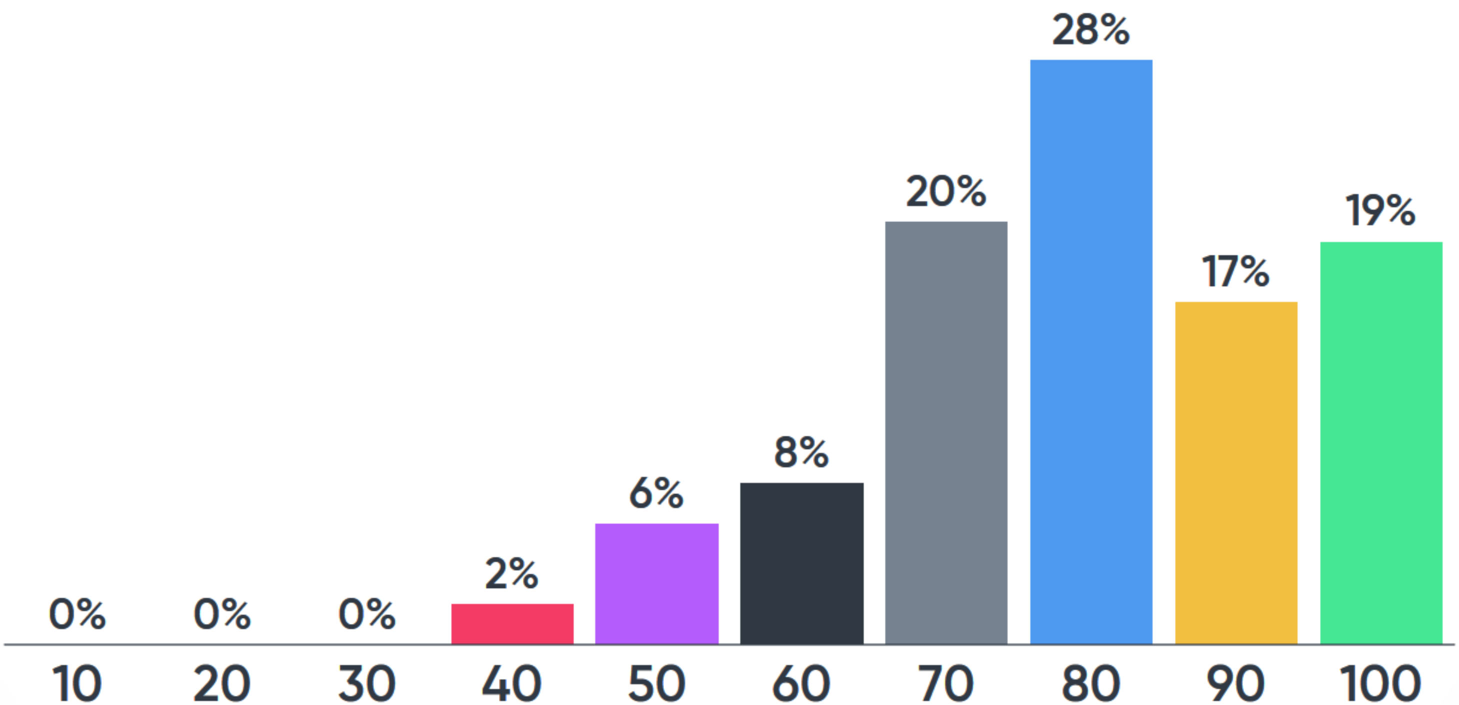
- **A student tells you that she spent 90 hours for the course work in the Machine Learning module. How many marks do you think she will receive?**
- **Without any other information difficult to answer**
- **Have a look at her fellow students**
  - x = Time spent
  - y = Achieved marks
- **Now, again: How many marks will she receive? 10, 20, 30 … 90, 100?**



Go to **www.menti.com** and use the code **20 20 72**

How many marks will a student receive who spends 90 hours?

Mentimeter

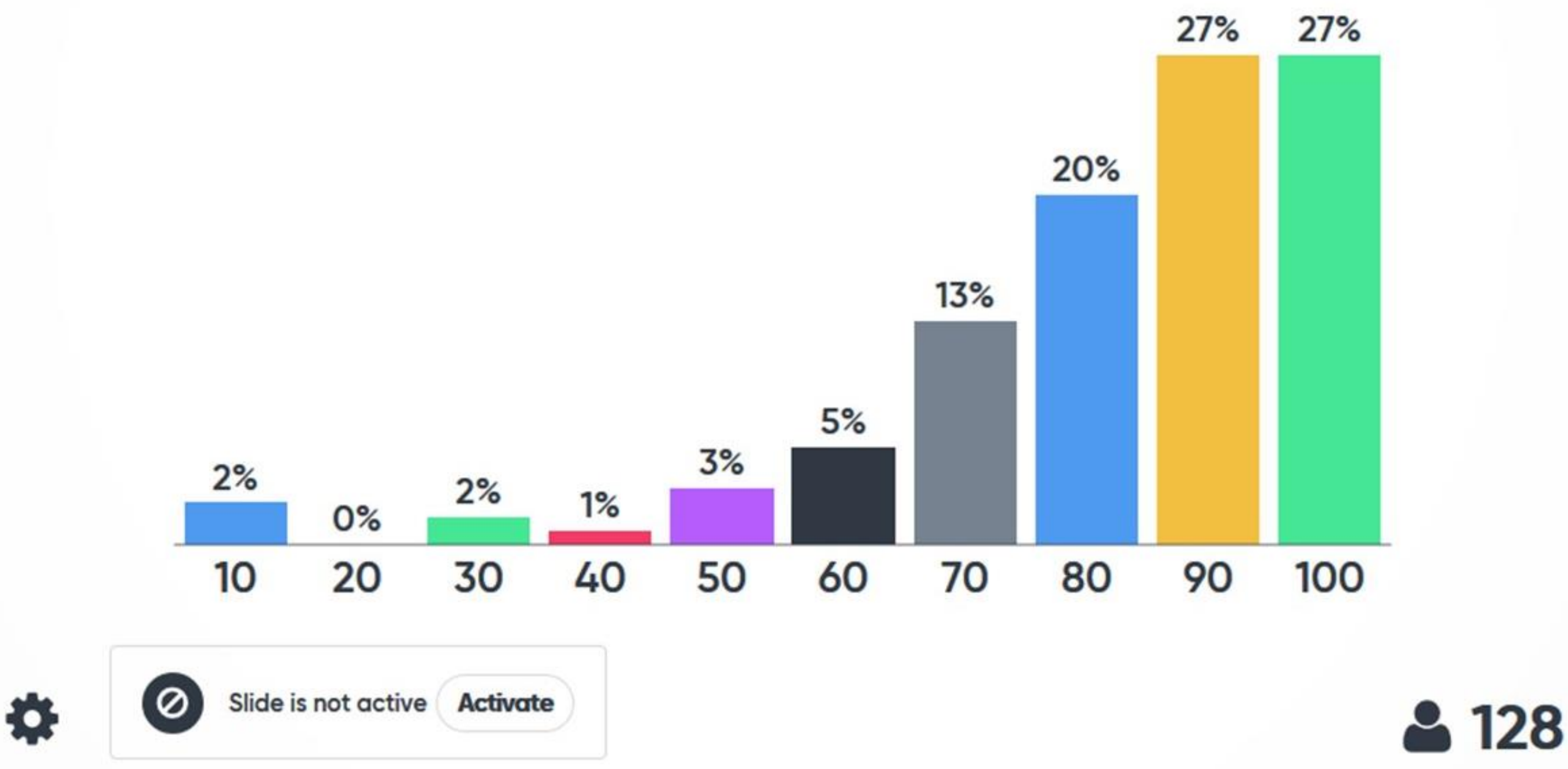


Slide is not active

Activate

103

Last year´s results

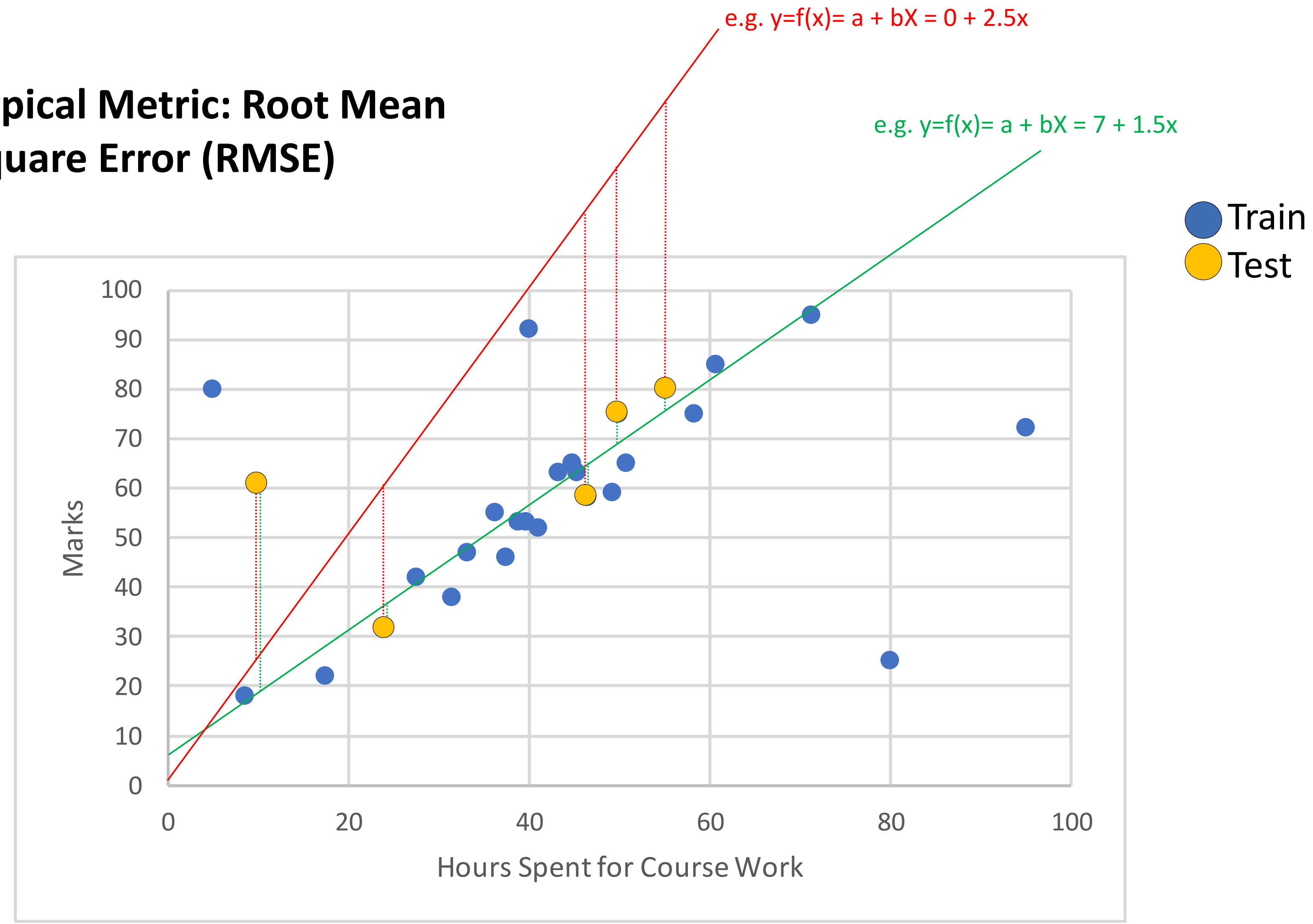


Example

- **Instance based: What marks did other students receive who invested 90 hours?**
  - Find similar student(s)
  - Calculate value for new student
  - „K-Nearest Neighbour“ algorithm
    - Neighbourhood size *k* is a hyperparameter
- **Model based: What function can approximate the existing data best?**
  - Find a function y=f(x)= a + bX
  - E.g. f(x)= 10 + 0.7x
  - Calculate y, given x
  - E.g. f(90)= 10 + 0.7*90 = 73
  - Linear regression
    - *a* and *b* are hyperparameters

Loss Function / Optimization

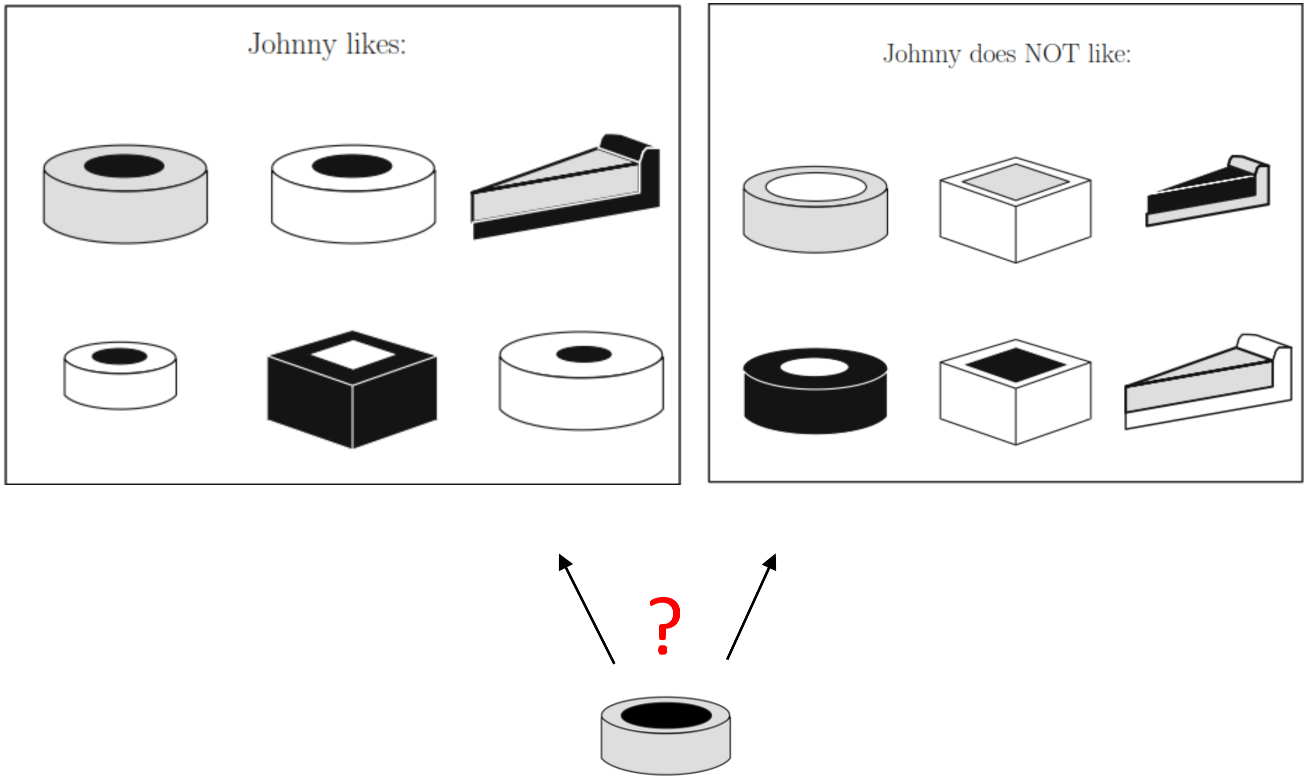
- **Typical Metric: Root Mean Square Error (RMSE)**



Example 2: Classification

Example

Will Johnny like or dislike the pie?

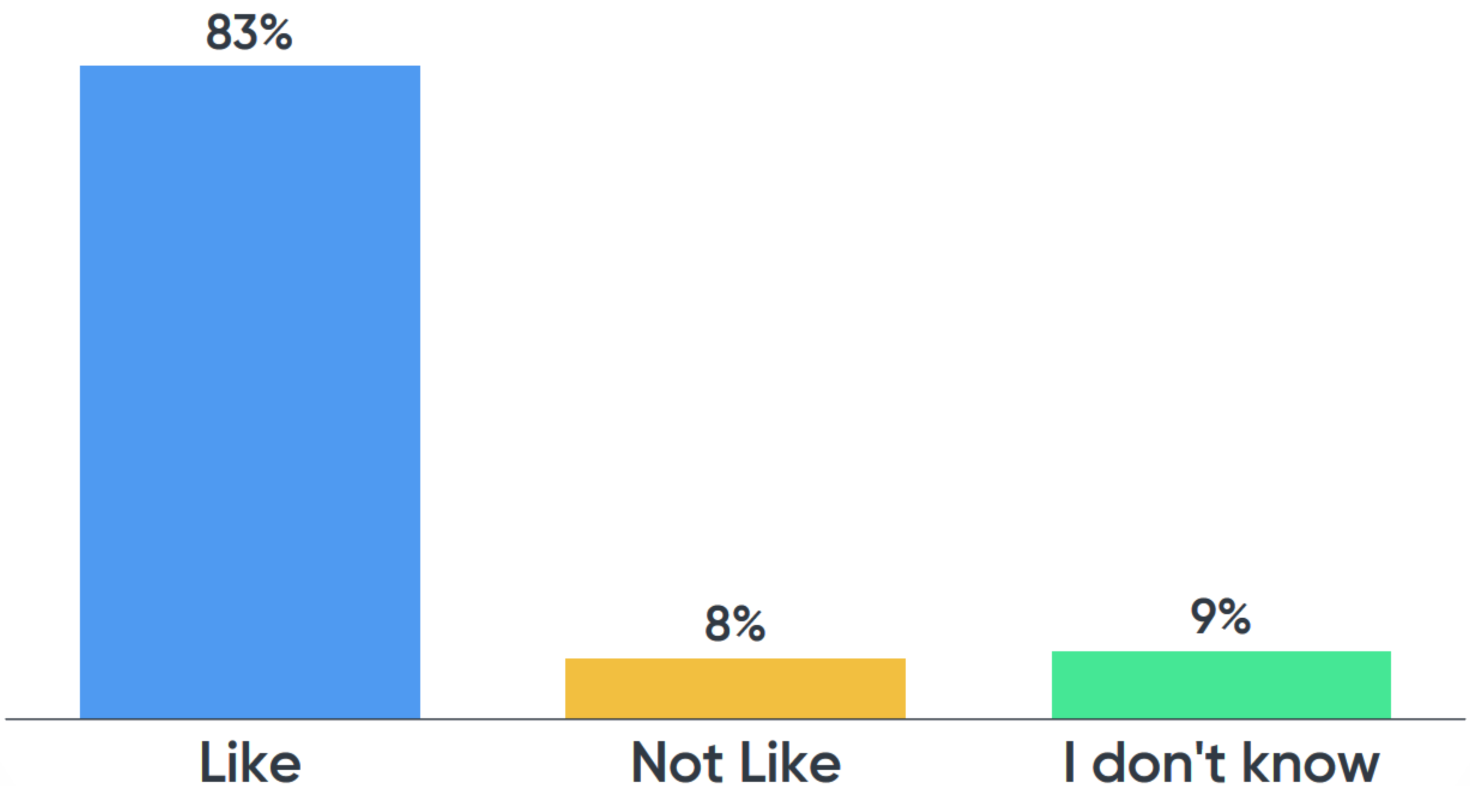


Miroslav Kubat, *An Introduction to Machine Learning* (Springer, 2015).

Go to **www.menti.com** and use the code **20 20 72**

Will Johnny like or dislike the pie?

Mentimeter

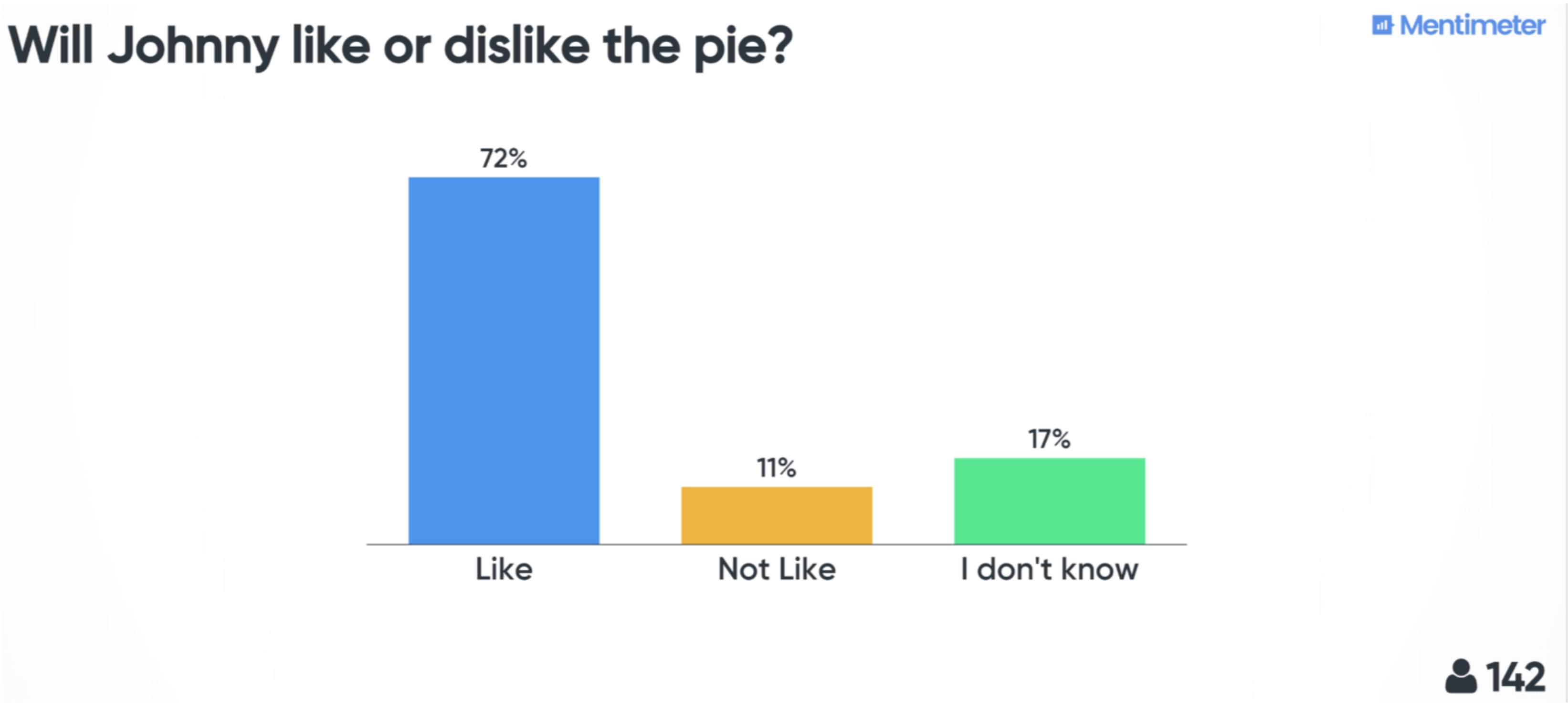


Slide is not active

Activate

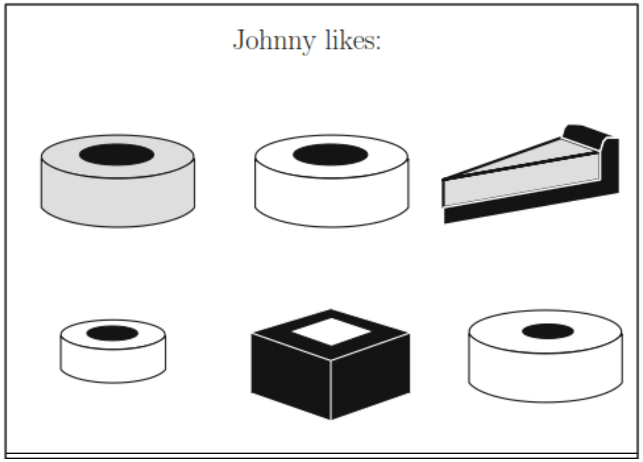
103

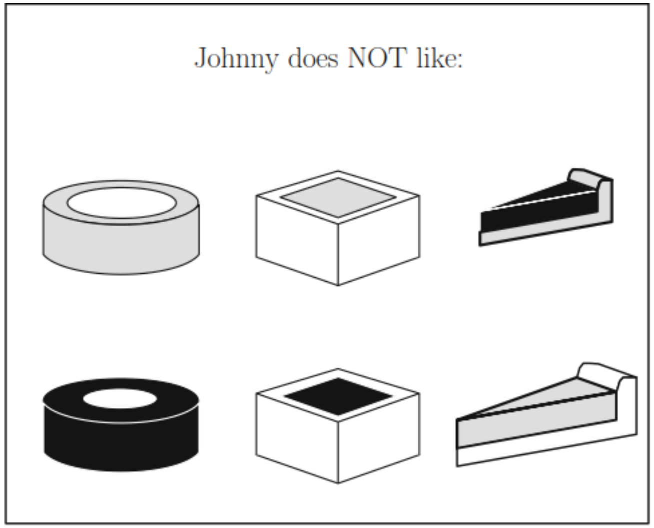
Last Year's Results



Features/Attributes

What are the features that distinguish the pies?

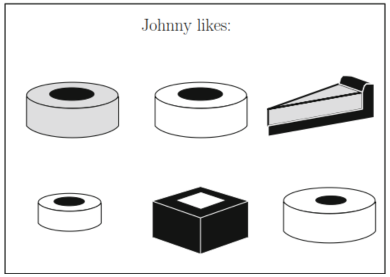


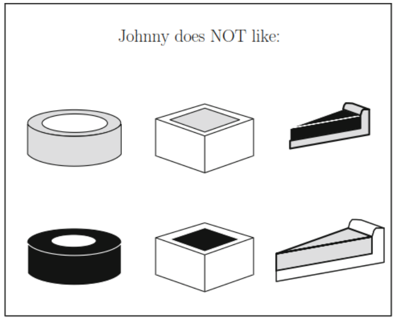


Miroslav Kubat, *An Introduction to Machine Learning* (Springer, 2015).

Representation

example	shape	crust		filling		class
		size	shade	size	shade	
ex1	circle	thick	gray	thick	dark	pos
ex2	circle	thick	white	thick	dark	pos
ex3	triangle	thick	dark	thick	gray	pos
ex4	circle	thin	white	thin	dark	pos
ex5	square	thick	dark	thin	white	pos
ex6	circle	thick	white	thin	dark	pos
ex7	circle	thick	gray	thick	white	neg
ex8	square	thick	white	thick	gray	neg
ex9	triangle	thin	gray	thin	dark	neg
ex10	circle	thick	dark	thick	white	neg
ex11	square	thick	white	thick	dark	neg
ex12	triangle	thick	white	thick	gray	neg

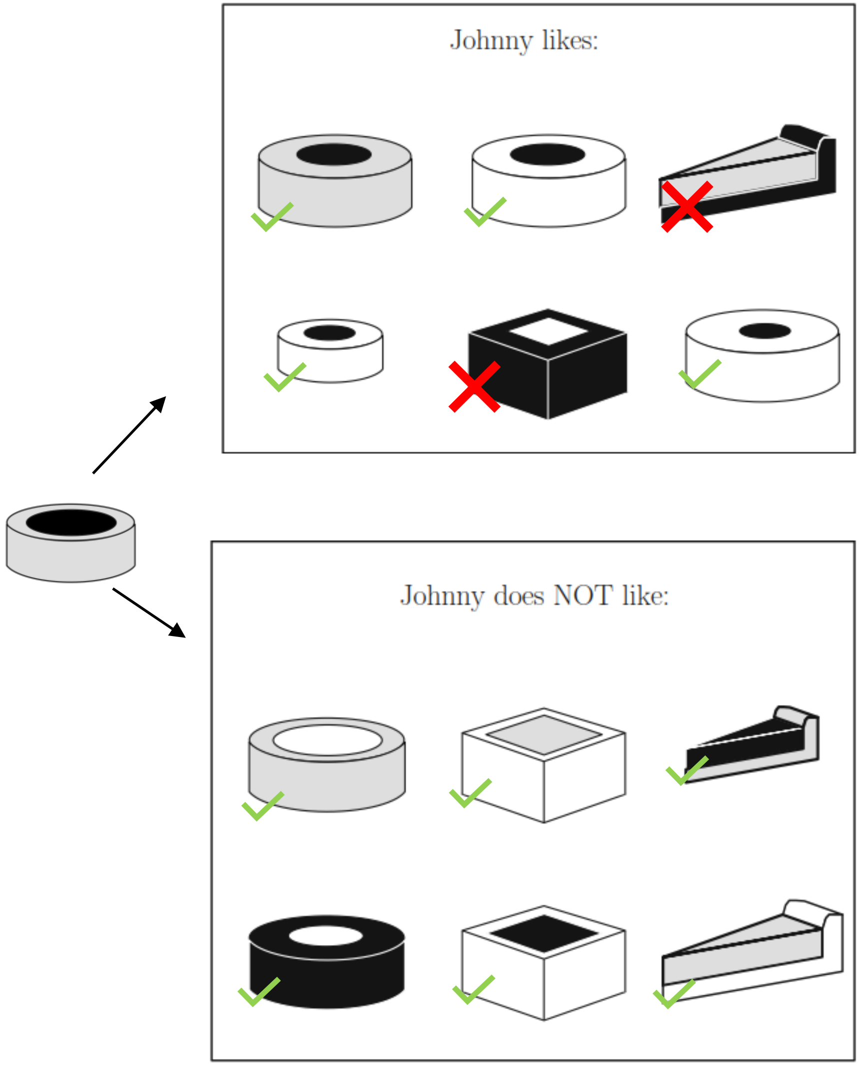




Miroslav Kubat, *An Introduction to Machine Learning* (Springer, 2015).

Rule Solution

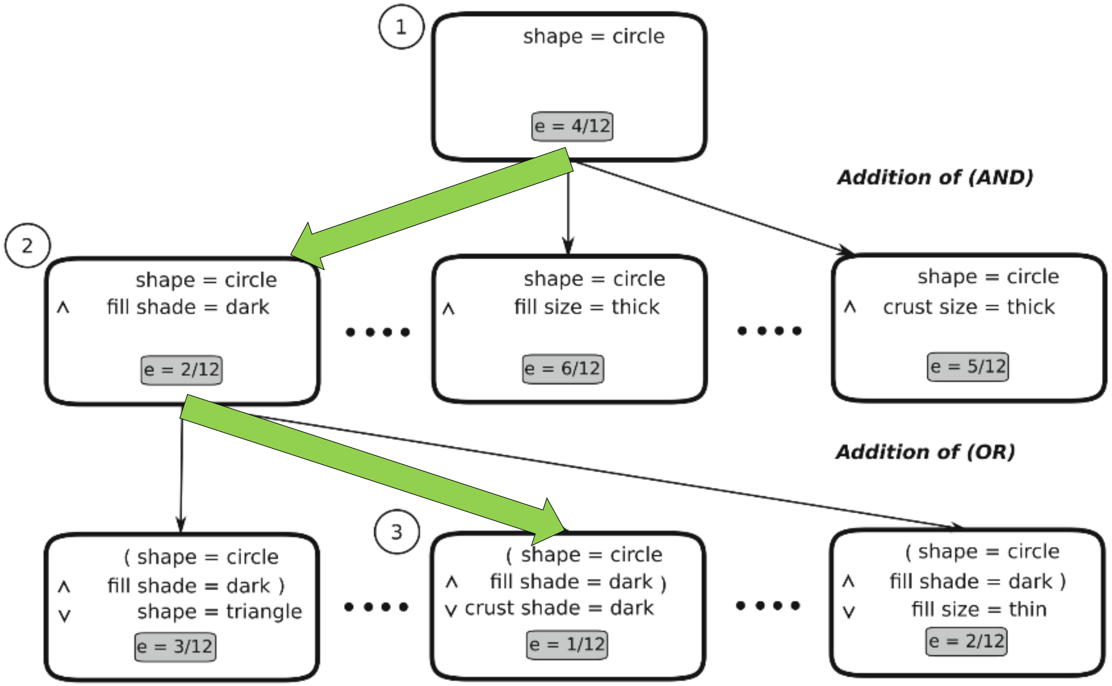
- **If [(shape=circle) AND (filling-shade=dark)] THEN “Like” ELSE “Not Like”**
  - Classifies all „Not Like“ correctly
  - Classifies 4 out of 6 „Like“ correctly
  - → Accuracy: 10 of 12 (83%)
- **What would you do for thousands of users (and new pies)?**

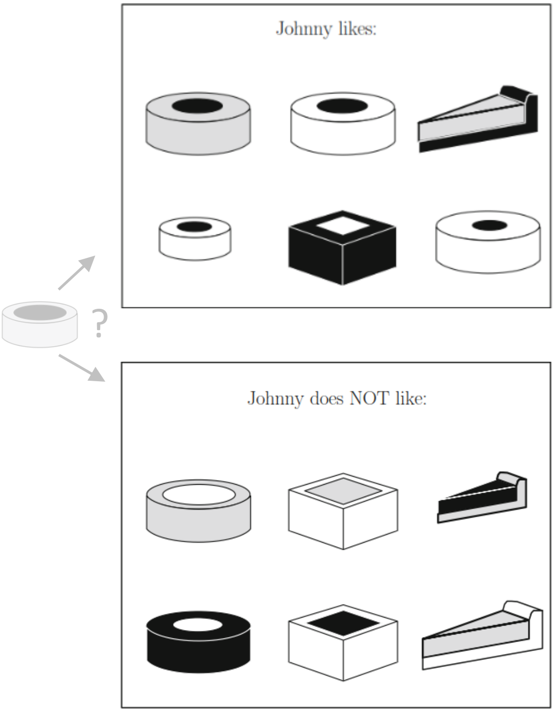


Miroslav Kubat, *An Introduction to Machine Learning* (Springer, 2015).

Machine Learning Solution (Hill Climbing)

- **Iterative algorithm**
- **Start with arbitrary solution**
- **Incremental improvements**
- **Repeat until no improvement is achieved**
- **Disadvantage: Finds only local optima & computing intensive**





Miroslav Kubat, *An Introduction to Machine Learning* (Springer, 2015).

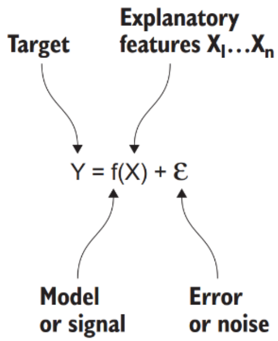
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Machine Learning Classifications

Supervised and Unsupervised Learning…

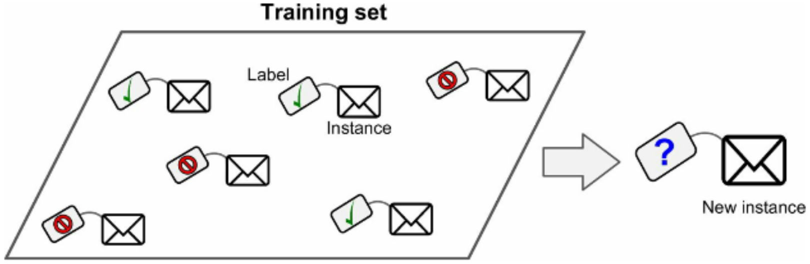
Supervised Learning

Henrik Brink, Joseph Richards, and Mark Fetherolf, *Real-world machine learning* (Manning Publications Co., 2016)



- **Typical Tasks**
  - Prediction/Regression (predict a number: the price of a house, income of a person, …)
  - Classification (predict a class: *Spam* or *Not Spam*?)
- **Typical Algorithms**
  - K-Nearest Neighbours
  - Regression
  - Support Vector Machines
  - Decision Trees and Random Forests

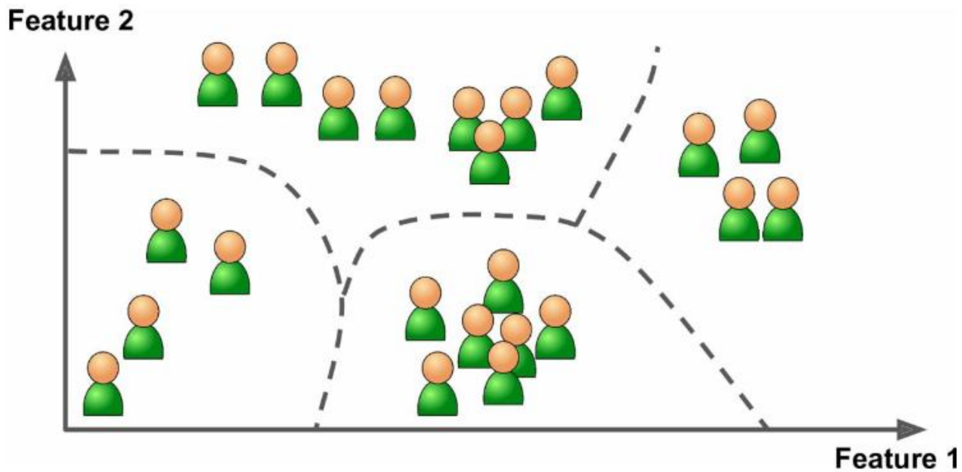
- **Training Input:**
  - Labelled data / training data
  - Instances/data points with known values for (independent) variables/features/attributes ***and*** known dependent variable/response/target/label
- **Production**
  - Input: new instance with known independent variables
  - Output: Predicted target/label



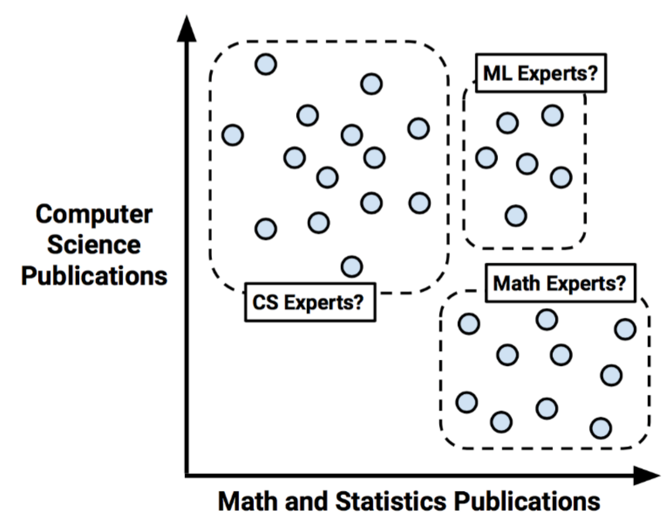
A. Géron, *Hands on Machine Learning with scikit-learn and Tensorflow*. O'Reilly Media, 2017.

Unsupervised Learning

- **Unlabelled data**
  - Work on the current data instead of predicting for future unknown instances
- **Typical task(s)**
  - Clustering (group instances into previously unknown groups)
  - Dimensionality Reduction
- **Algorithms**
  - Clustering
    - K-means
    - Hierarchical Cluster Analysis (HCA)
    - Expected Maximization
  - Dimensionality Reduction
    - Principal Component Analysis
    - Kernel PCA



A. Géron, *Hands on Machine Learning with scikit-learn and Tensorflow*. O'Reilly Media, 2017.



https://hub.packtpub.com/introduction-clustering-and-unsupervised-learning/

… and Semi Supervised and Reinforcement Learning…

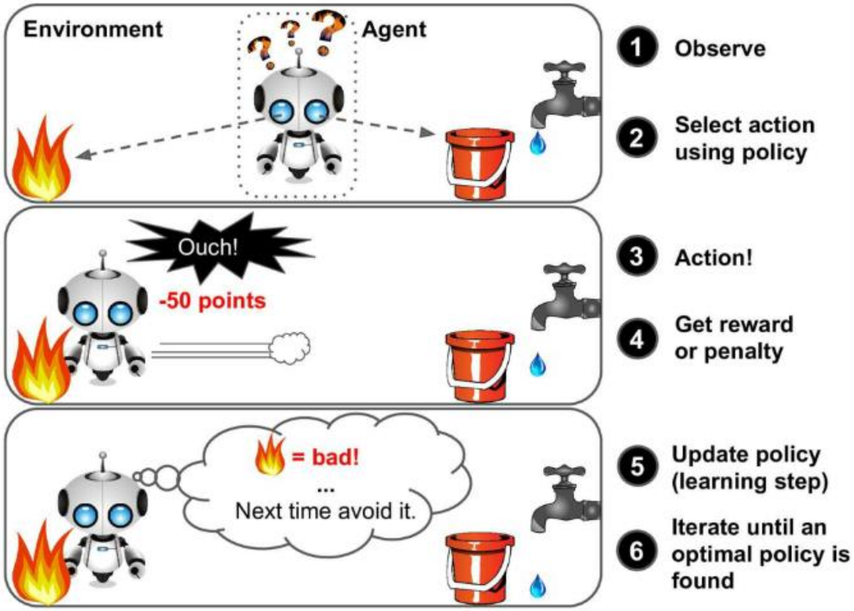
Semi-Supervised Learning

- **Some labelled but mostly unlabelled data**
- **For instance, Google Photos:**
  - Identification of the same persons (unsupervised)
  - Manually adding labels to clusters/persons (supervised)
- **Often based on Deep Belief Networks (DBN) / Restricted Boltzmann Machines (RBM)**

A. Géron, *Hands on Machine Learning with scikit-learn and Tensorflow*. O'Reilly Media, 2017.

Reinforcement Learning

- **An agent observes an environment, chooses between actions and gets rewards (or penalties) in return. The agent learns the best strategy („policy“) for what to do in a given situation.**
- **Often used for making robots to learn how to walk; also for DeepMind‘s AlphaGo**



A. Géron, *Hands on Machine Learning with scikit-learn and Tensorflow*. O’Reilly Media, 2017.

Example (Reinforcement learning)

- **https://www.youtube.com/watch?v=V1eYniJ0Rnk**

… and other views

Instance Based vs. Model Based

- **Instance**
  - No training
  - Algorithms
    - k-nearest neighbour
    - kernel machines
    - RBF networks
- **Model**
  - Model is learned through training data

Parametric vs. Nonparametric Models

- **Parametric**

– Based on the features (and certain weights), Y can be calculated

– E.g. (linear) regression

$$f(\mathbf{X}) = \beta_0 + \mathbf{X}_1 \times \beta_1 + \mathbf{X}_2 \times \beta_2 + \ldots$$

- **Nonparametric**

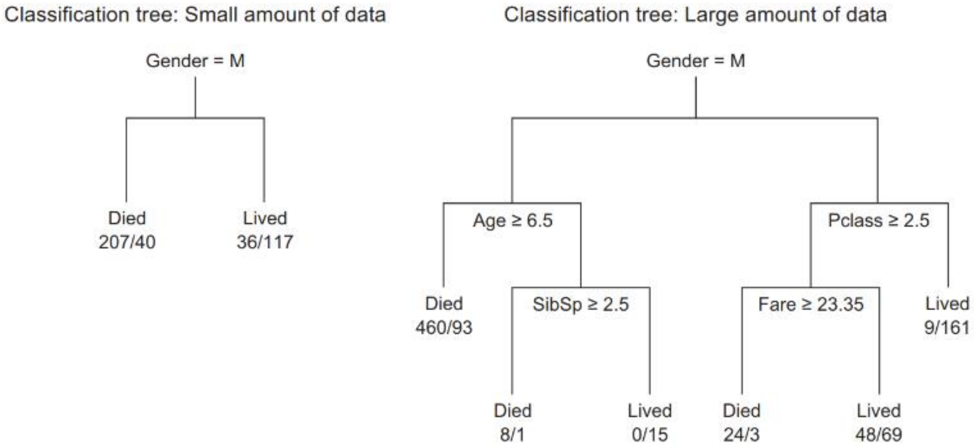


Figure 3.2 A decision tree is an example of a nonparametric ML algorithm, because its functional form isn't fixed. The tree model can grow in complexity with larger amounts of data to capture more-complicated patterns. In each terminal node of the tree, the ratio represents the number of training instances in that node that died versus lived.

Henrik Brink, Joseph Richards, and Mark Fetherolf, Real-world machine learning (Manning Publications Co., 2016).

The five tribes of machine learning

- The five tribes of machine learning, Webinar by Pedro Domingos https://learning.acm.org/webinar_pdfs/PedroDomingos_FTFML_WebinarSlides.pdf
- Pedro Domingos, *The master algorithm: How the quest for the ultimate learning machine will remake our world* (Basic Books, 2015), http://www.idi.ntnu.no/emner/tdt4173/papers/Domingos-SVM-NN-CBR.pdf.
- See also John Paul Mueller and Luca Massaron, Machine Learning for Dummies (John Wiley & Sons, 2016).

Tribe	Origins	Problem	Solution / Master Algorithm
Symbolists	Logic, Philosophy	Knowledge composition	Inverse deduction
Connectionists	Neuroscience	Credit Assignment	Backpropagation
Evolutionaries	Evolutionary Biology	Structure Discovery	Genetic programming
Bayesians	Statistics	Uncertainty	Probablistic Inference
Analogizers	Psychology	Similarity	Kernel Machines

Definition

Machine Learning Definitions

- ~~**„Machine Learning is the science (and art) of programming computers so they can *learn from data.*“**
A. Géron, Hands on Machine Learning with scikit-learn and Tensorflow. O'Reilly Media, 2017.~~
- ~~**„[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed“**
Arthur Samuel, 1959 in A. Géron~~
- **„A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E.**“ Tom Mitchell, 1997 in A. Géron

Ignores the eventual goal, i.e. solving a task.

Experience = data

Typical Tasks (from an application point of view)

1. **Prediction**
   1. House Prices
   2. Diseases
   3. Products a customer will like
   4. …
2. **Recognition**
   1. Faces
   2. Voices
   3. Gestures
   4. …
3. **Creation / Modification**
   1. Art (Paintings, Music, …)
   2. Videos (DeepFakes, …)
   3. Augmented Reality
   4. …
4. **Anything else?**

Typical Tasks (2)

Input A → Output B

What Machine Learning Can Do

A simple way to think about supervised learning.

INPUT A	RESPONSE B	APPLICATION
Picture	Are there human faces? (0 or 1)	Photo tagging
Loan application	Will they repay the loan? (0 or 1)	Loan approvals
Ad plus user information	Will user click on ad? (0 or 1)	Targeted online ads
Audio clip	Transcript of audio clip	Speech recognition
English sentence	French sentence	Language translation
Sensors from hard disk, plane engine, etc.	Is it about to fail?	Preventive maintenance
Car camera and other sensors	Position of other cars	Self-driving cars

SOURCE ANDREW NG

https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now

Typical Experiences

- **Data with features from which the task can be inferred**
  - Instance/Data Point
  - Attribute/Feature/(Independent)Variable
  - Target/Label/(Dependent) Variable

			Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	…	Target / Label
Dataset	Training Data	**Data Point *d1***	72263	27245	17275	43964	15567	56433	83212
		Data Point *d2*	84139	18072	87332	50153	9561	40396	15706
		Data Point *d3*	78266	71688	53857	93655	67871	75253	26774
		Data Point *d4*	36848	48316	47676	11324	76572	4863	87264
		Data Point *d5*	25388	32838	67949	33058	15464	66957	53946
		Data Point *d6*	16600	86143	67239	15401	70959	50995	80351
		Data Point *d7*	1660	60325	11626	44432	1323	92803	17490
		Data Point *d8*	18567	31649	5142	34048	85153	19318	50444
		Data Point *d9*	56760	72157	6882	22721	85580	30351	62197
		Data Point *d10*	31857	77264	43623	35524	36979	94473	8850
	Test Data	**Data Point *d11***	63372	13691	19661	33752	78327	54604	54750
		Data Point *d12*	93095	25329	72530	79730	14969	86179	86887
		Data Point *d13*	41598	91224	73105	5623	85319	31115	10267
		Data Point *d14*	5429	34300	41889	45034	70645	79225	13049
		Data Point *d15*	60513	84518	54860	99854	82111	38099	8633

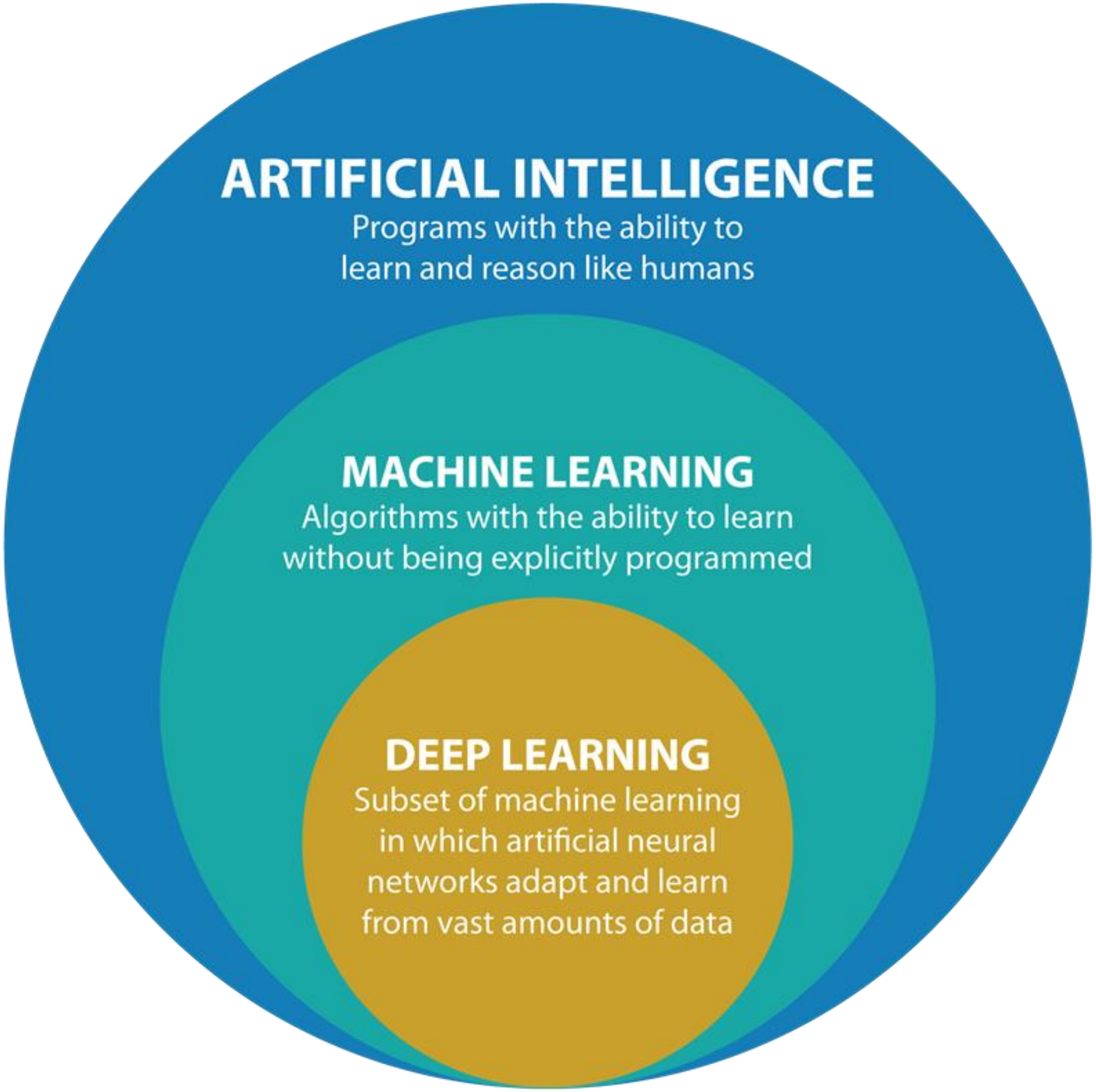
Production Data	**Data Point *p1***	38366	16911	40533	56296	73633	43034	?
	Data Point *p2*	30984	28986	23832	70127	21389	29660	?
	Data Point *p3*	98767	71808	2057	13791	41601	80384	?
	Data Point *p4*	79887	56619	99693	42540	25042	99478	?
	Data Point *p5*	67501	66324	78065	94571	93229	6776	?
	…	59989	7313	68544	64573	66802	33831	?

Build Machine-Learning Model(s)

Test the Model(s)

Apply the (best) Model

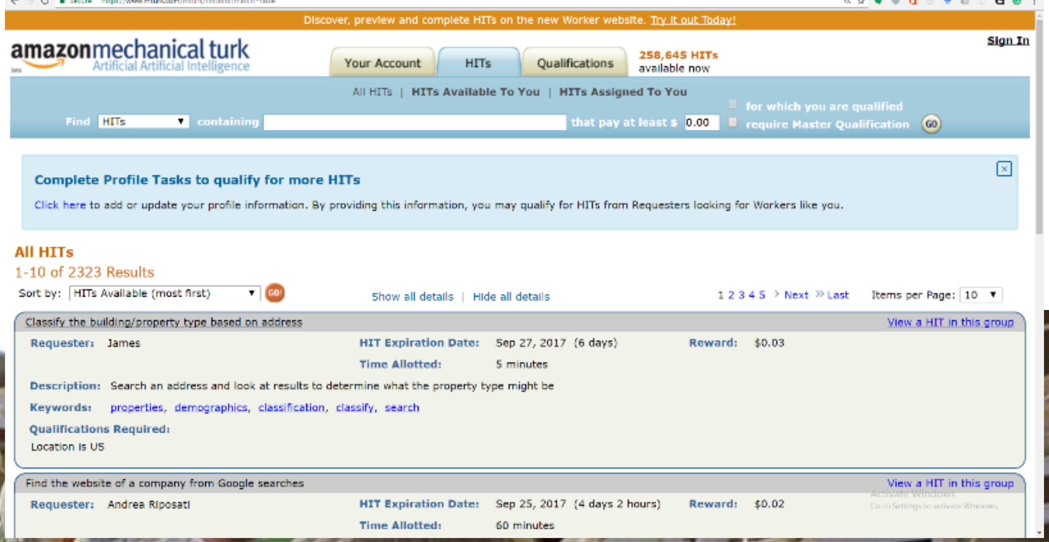
ML vs. AI vs. DL (vs. Data Science)

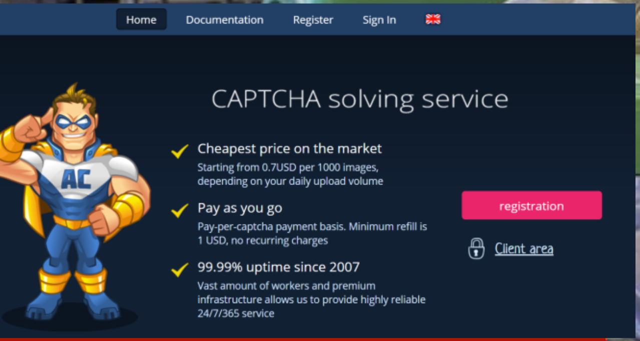


https://www.argility.com/wp-content/uploads/2018/04/image10.png

Traditional approaches to problem solving

Human Labour



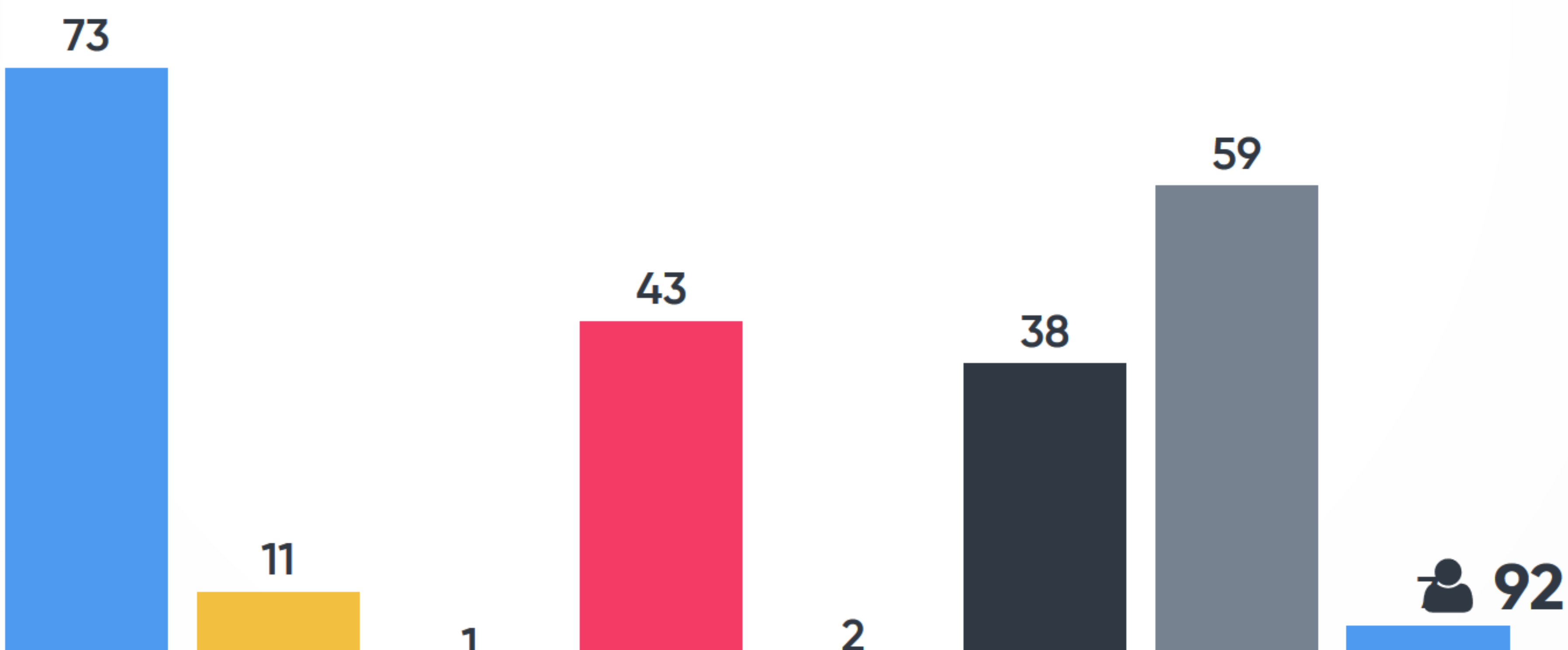


http://www.captcha.net/images/recaptcha-example.gif
http://motorcitymuckraker.com/wp-content/uploads/2014/04/Shinola_2315-701x468.jpg
https://anti-captcha.com/

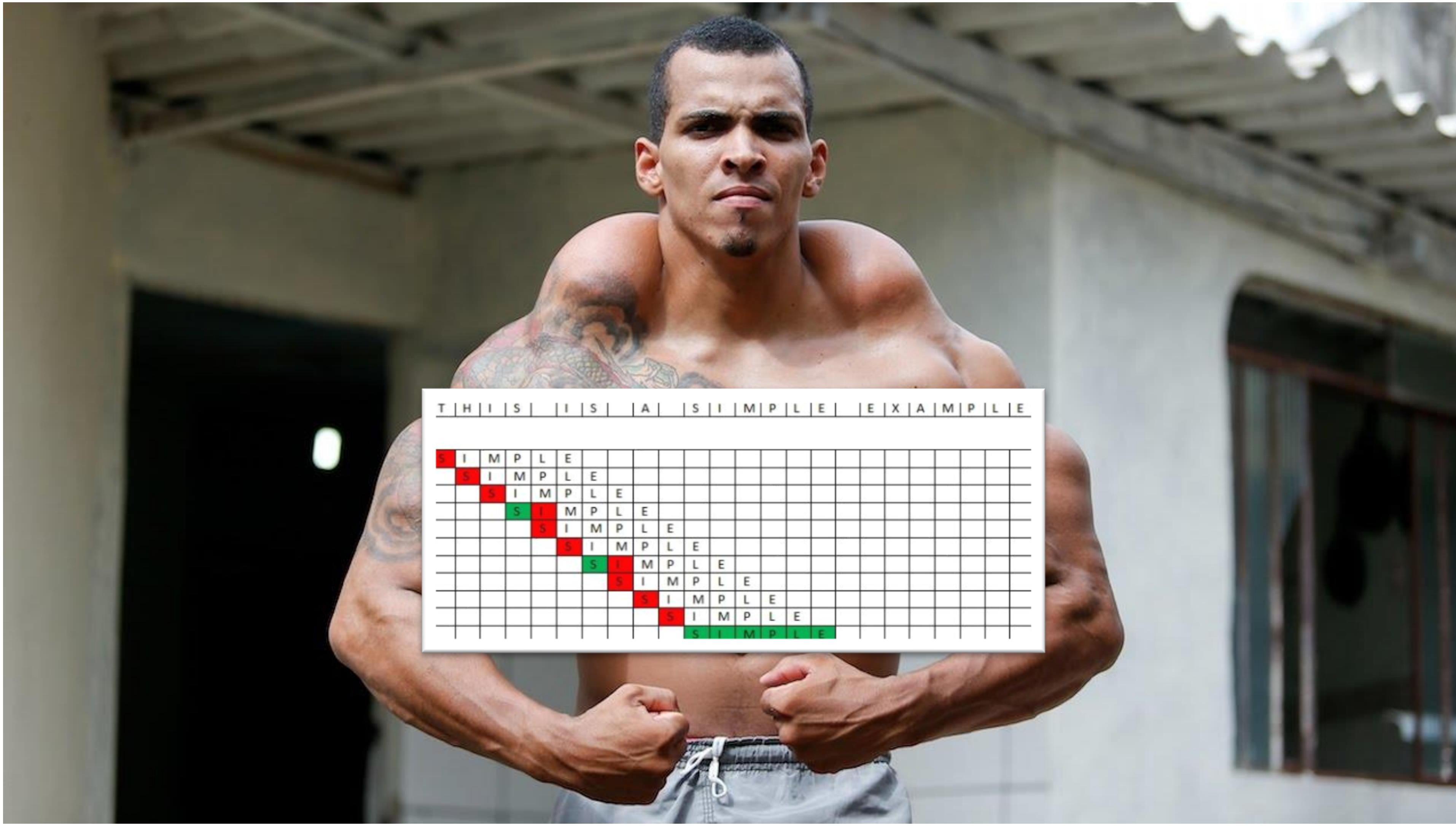
Go to **www.menti.com** and use the code **20 20 72**

Lecture Evaluation

Mentimeter



Bruteforce



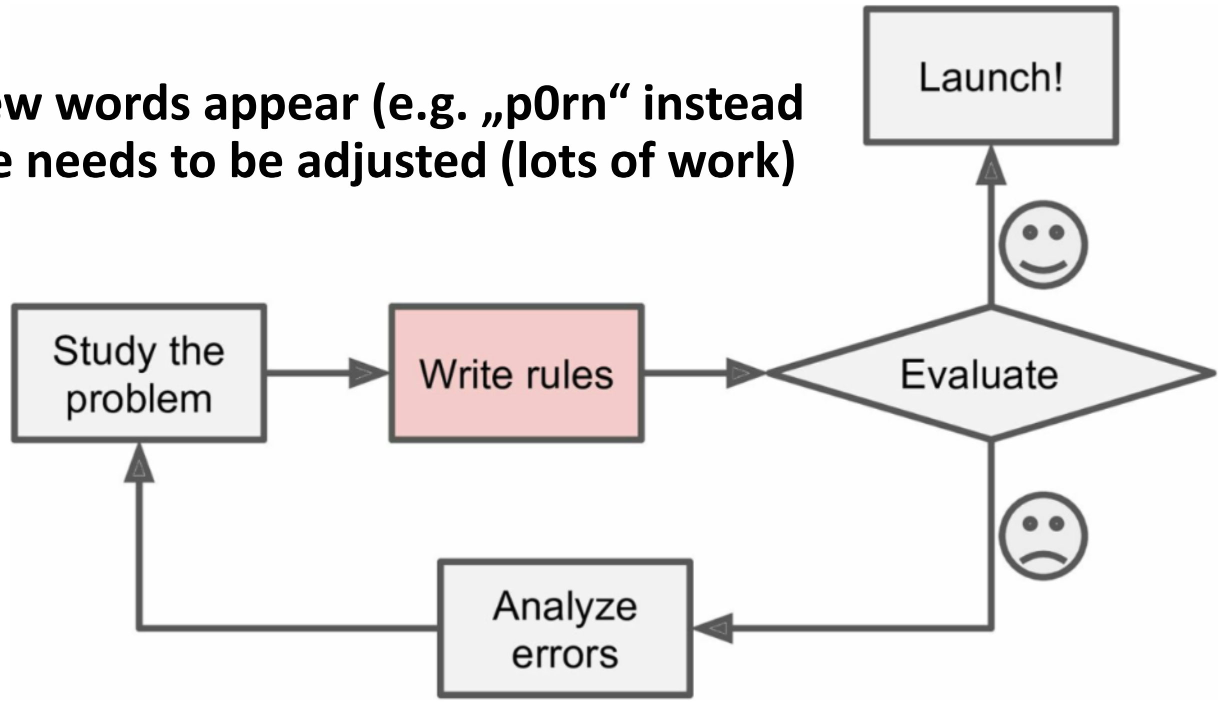
https://i.ytimg.com/vi/SobAPTAAX1s/maxresdefault.jpg
http://1.bp.blogspot.com/-YJDalyxz6XY/UFCYnd2_nBI/AAAAAAAAAB4/uewJpXgs9Mc/s1600/Brute+Force.jpg

Rules / Heuristics / Algorithms

https://mensconferenceuk.files.wordpress.com/2013/08/the-rule-book.jpg

Example for Rules: Spam Detection

- **Goal: Decide if an email is spam or not? (classification problem)**
- **Potential Rule: If an email contains the terms „porn, sex, viagra, …“ then it is spam. Maybe with wildcards, e.g. p?rn, or sex***
- **Problem: When new words appear (e.g. „p0rn“ instead of „porn“), the rule needs to be adjusted (lots of work)**

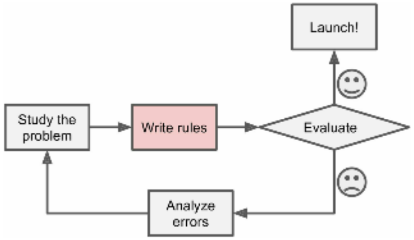


A. Géron, *Hands on Machine Learning with scikit-learn and Tensorflow*. O’Reilly Media, 2017.

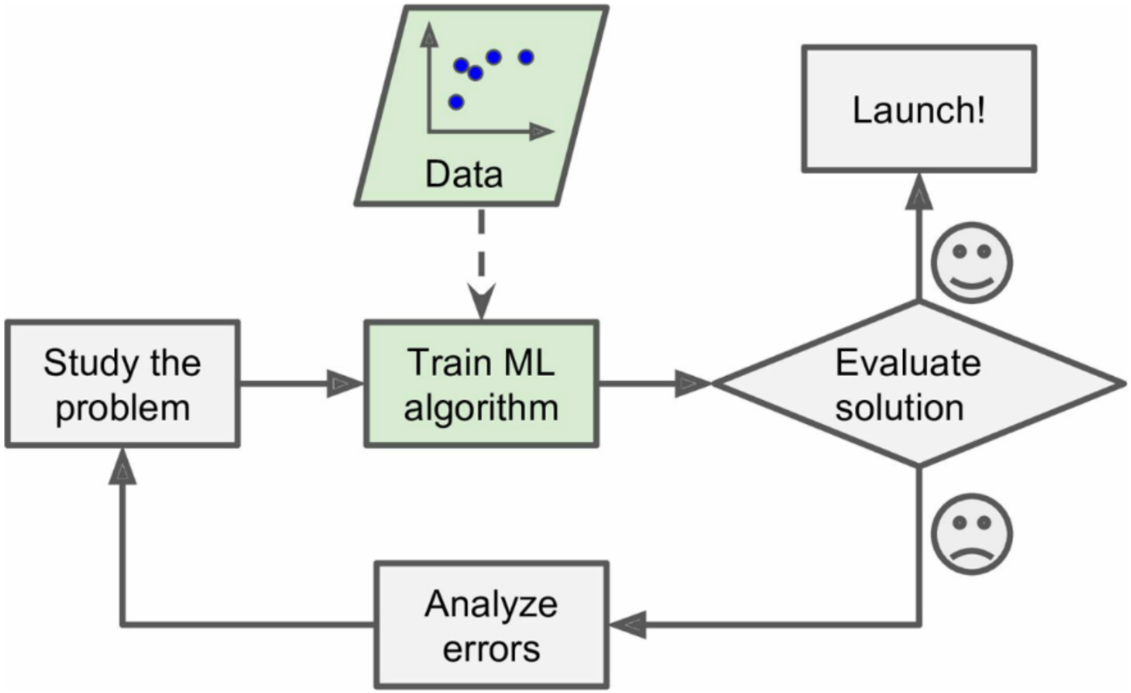
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

The Machine Learning Approach to Problem Solving

Machine-Learning Approach to Spam Detection



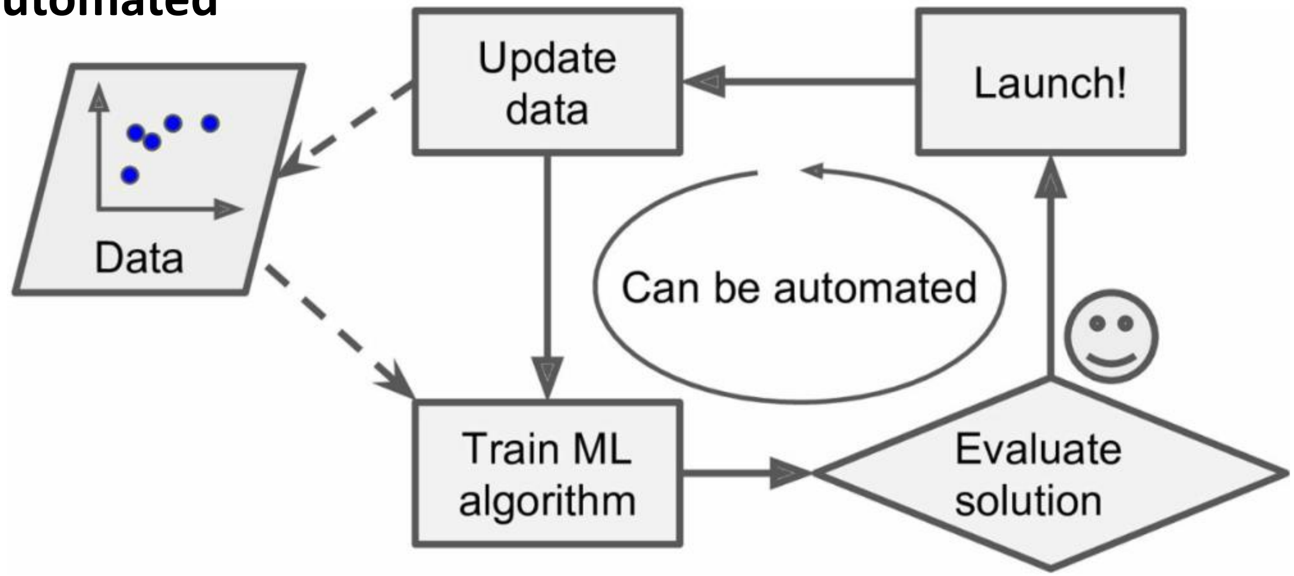
- „A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E.“
- T = Detect Spam Emails (flag new emails as spam/not-spam)
- P = Number of correctly classified (not-) spam emails
- E = Dataset with emails being classified as spam/not-spam ← the more data, the better P becomes



A. Géron, *Hands on Machine Learning with scikit-learn and Tensorflow*. O'Reilly Media, 2017.

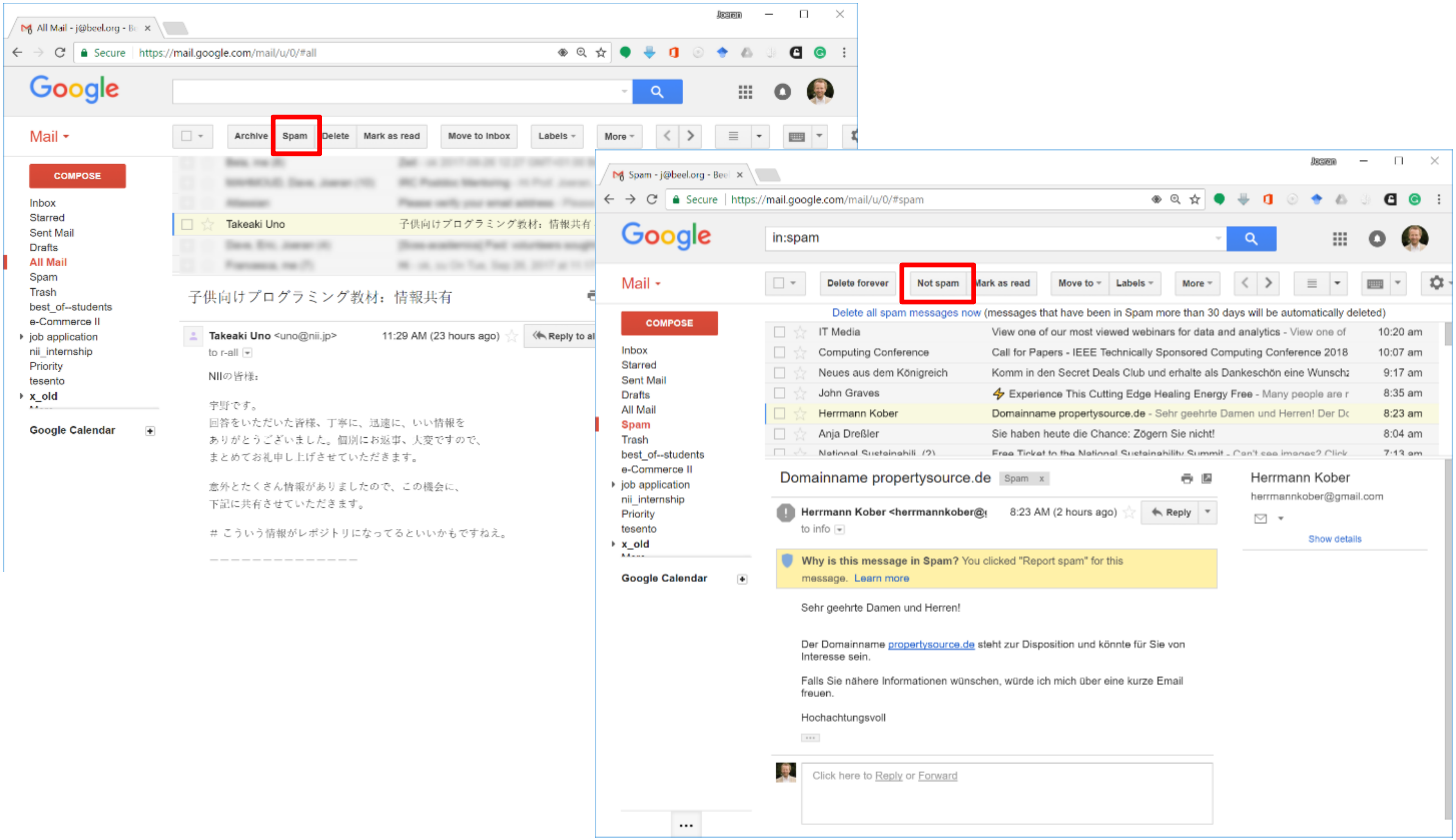
Updating Machine-Learning

- **You update the data, not the algorithm**
- **For instance, once there is an updated email dataset (with emails that contain e.g. the term „p0rn“ and are flagged as spam), the algorithm learns that emails containing the term „p0rn“ are also spam.**
- **Can be automated**



A. Géron, *Hands on Machine Learning with scikit-learn and Tensorflow*. O'Reilly Media, 2017.

Spam Detection in Gmail

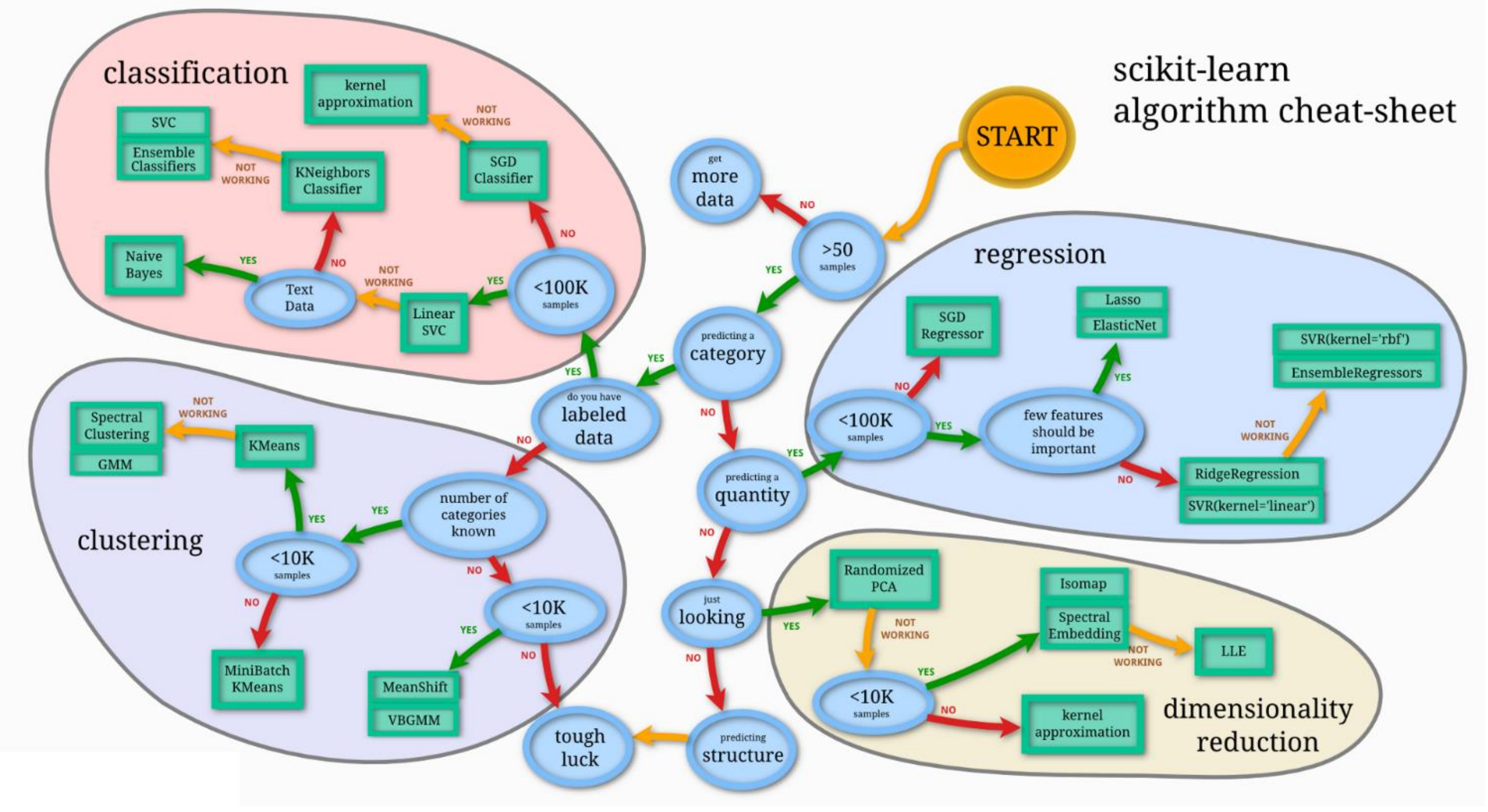


Machine Learning as General Problem Solver

- **One algorithm can solve different problems. For instance, Support Vector Machines can solve:**
  - Spam Detection
  - Hand Written Character Recognition
  - Image Classification
  - …
- **However, there is not „the one“ machine-learning algorithm that can do everything**

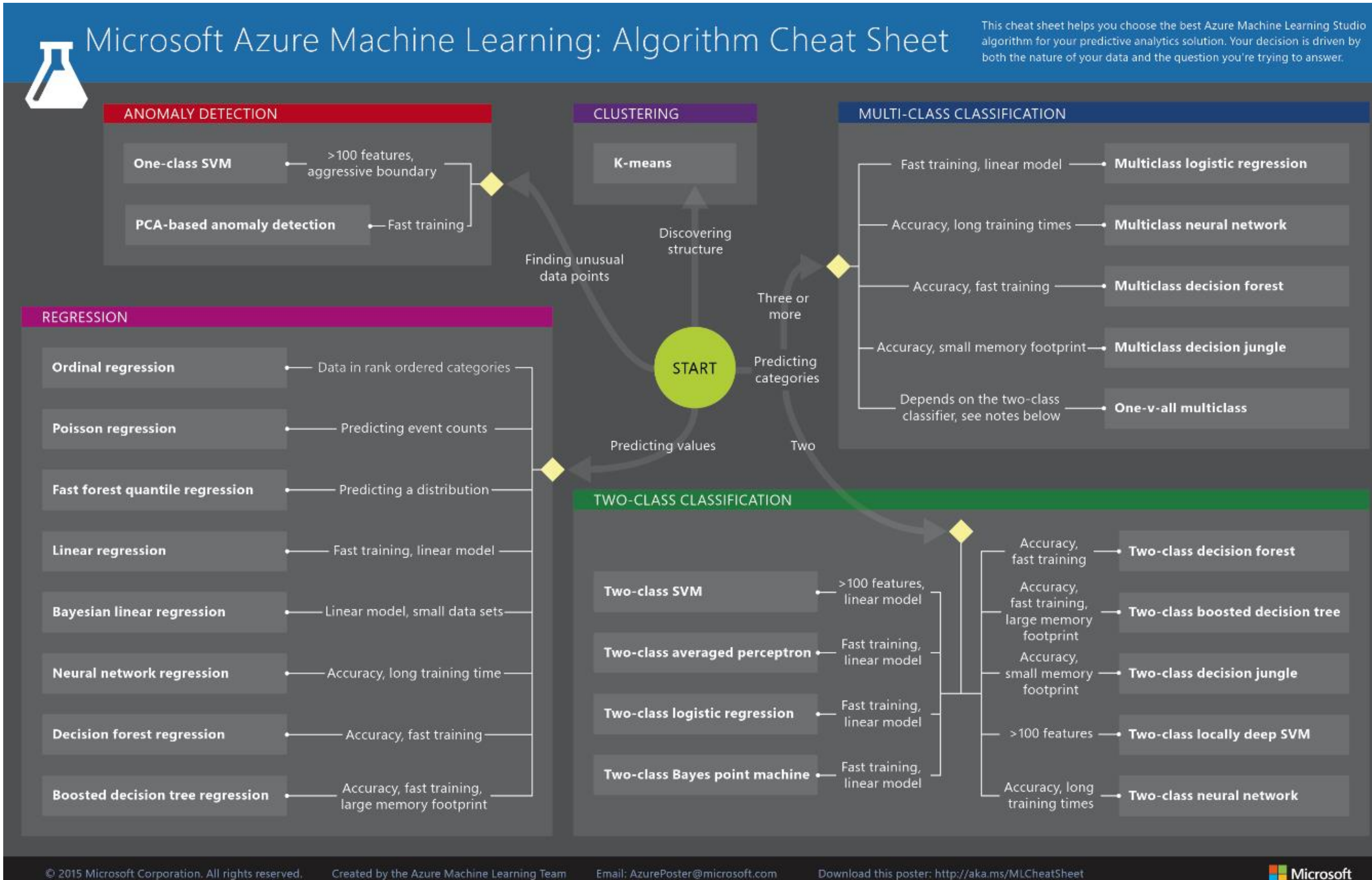
The Machine Learning Landscape(s)

Scikit-learn



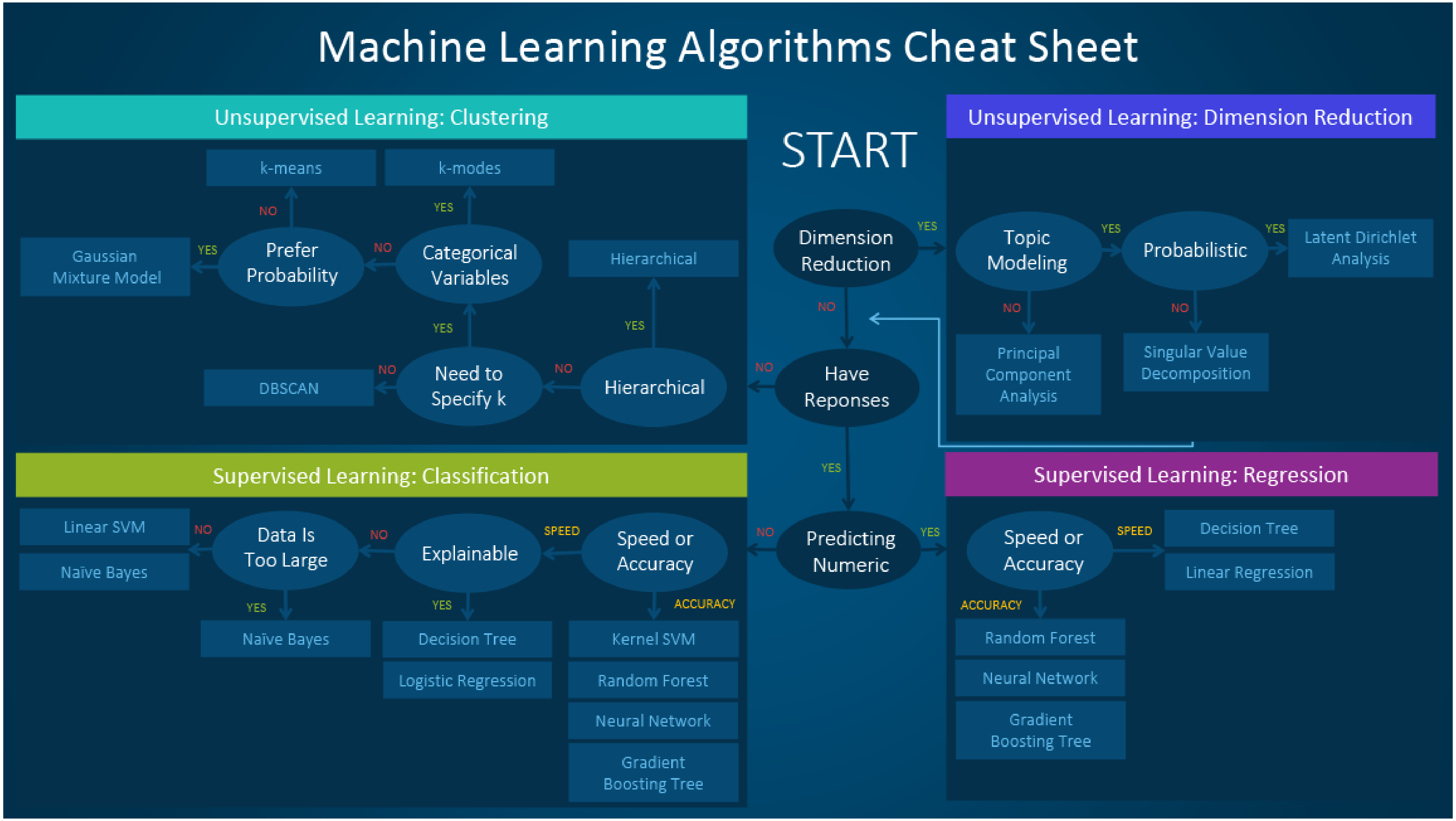
https://cdn-images-1.medium.com/max/1920/1*kJiLzEawYtmD7t-VQ3AXmw.png

Microsoft Azure



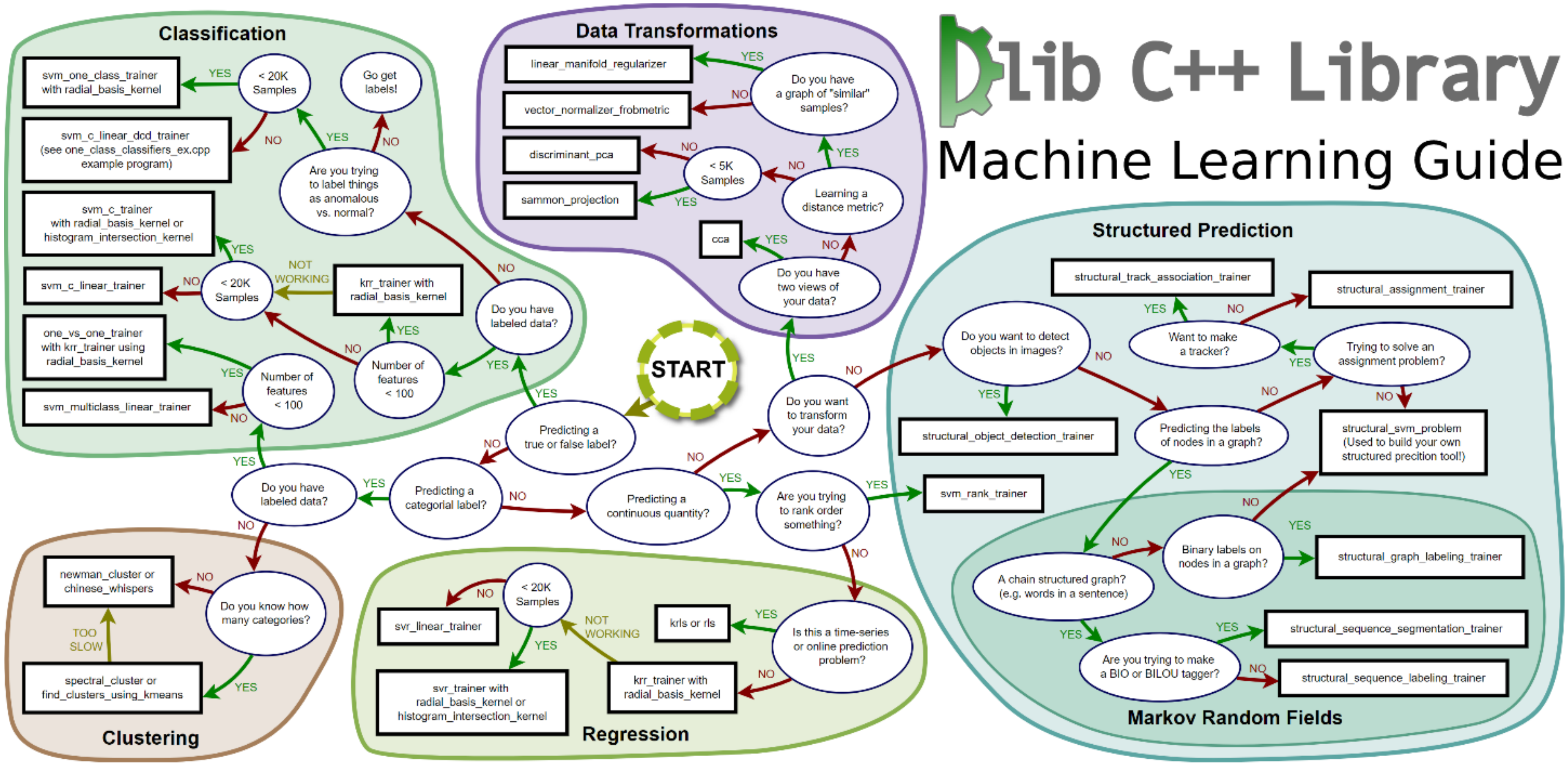
https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-algorithm-cheat-sheet

SAS



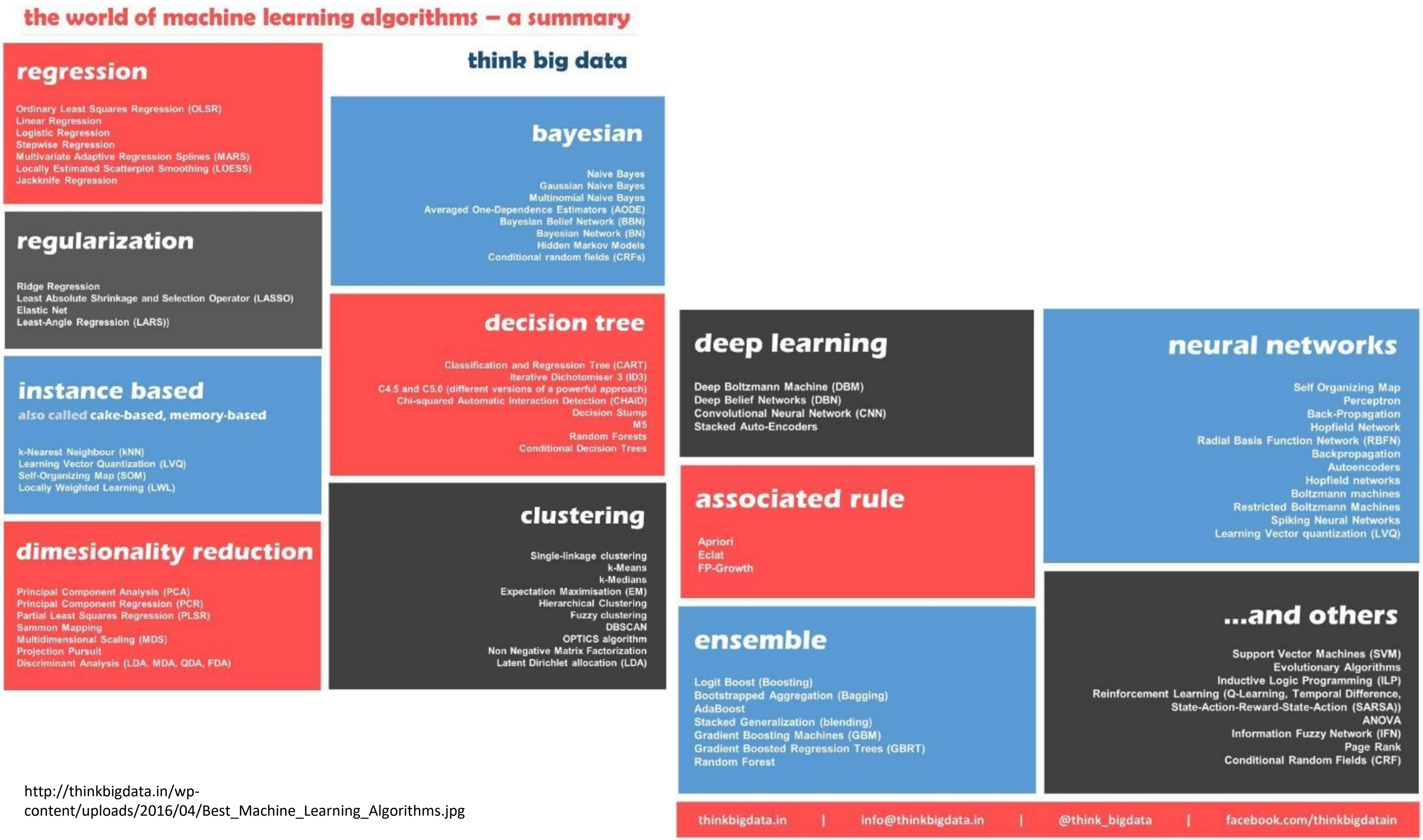
http://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/

DLib



http://dlib.net/ml_guide.svg

Facebook



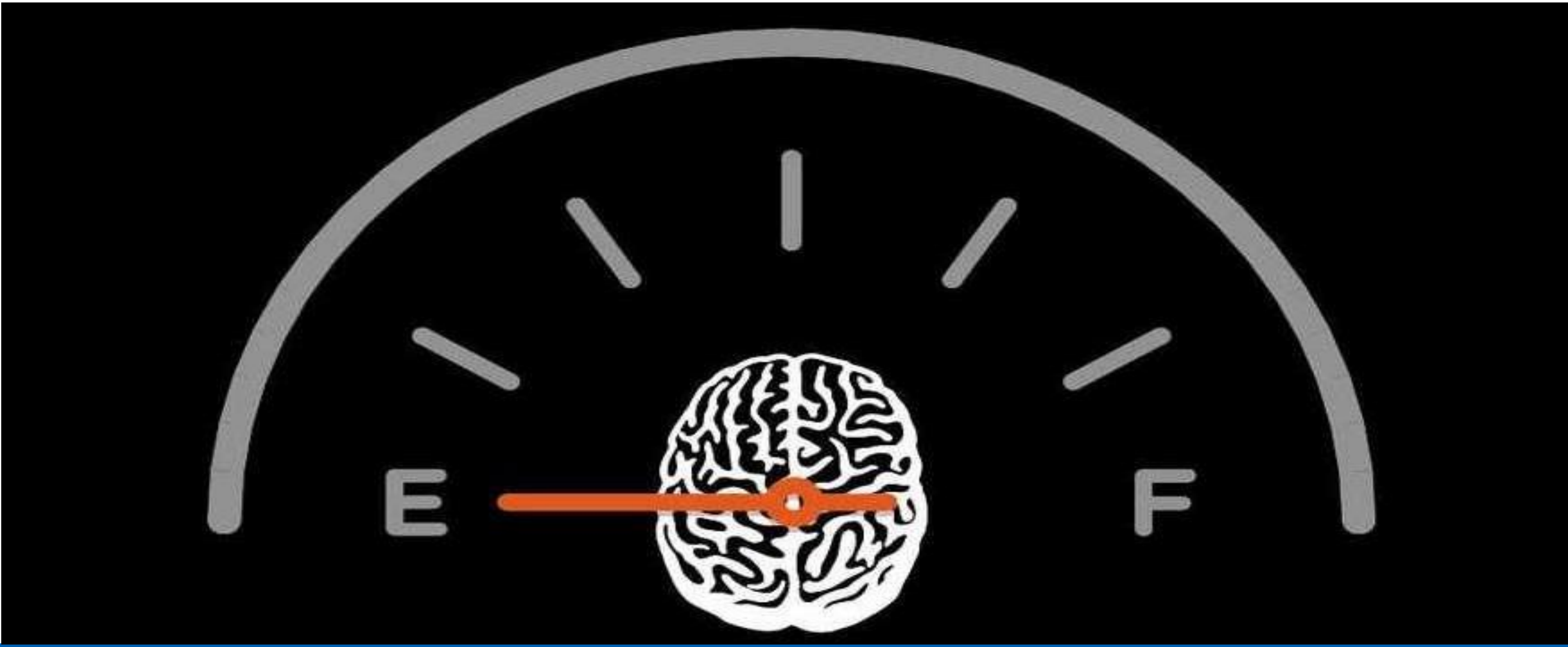
http://thinkbigdata.in/wp-content/uploads/2016/04/Best_Machine_Learning_Algorithms.jpg

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Strengths and Weaknesses of Machine Learning

Machine Learning is good for

1. Problems that would require many rules to solve
2. Problems that are too complex to be solved by rules (e.g. handwriting or speech recognition)
3. Problems that have changing environments/data (e.g. spammers who try to cheat the system)
4. Problems where (lots of) data is already available (e.g. a corpus of labelled emails)

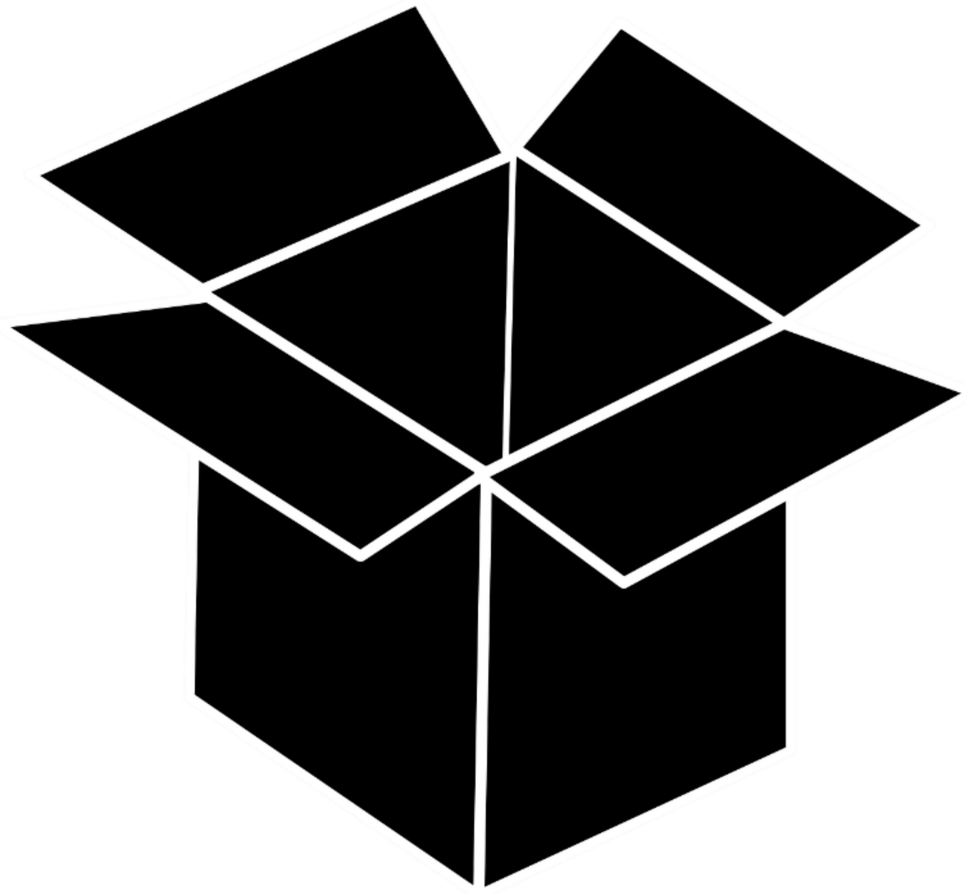


Machine Learning is not good for

- **Random number generation**
- **En/Decryption**
- **Simple operations such as copying data**
- **Executing programs (following algorithms)**

Blackbox

- **It is often not immediately clear why a machine learning system creates the outputs it creates**
- **Understanding the reasoning behind the output, requires additional analyses; sometimes it's very difficult or impossible to fully understand the reasoning**
- **Legal implications**



https://s3-us-west-2.amazonaws.com/courses-images-archive-read-only/wp-content/uploads/sites/903/2016/01/23225801/black-box-310220_1280.png

Legal Implications

WIRED — Tesla Bears Some Blame for Self-Driving Crash Death, Feds Say

SHARE

SHARE 359

TWEET

COMMENT

EMAIL

AARIAN MARSHALL TRANSPORTATION 08.13.17 07:00 AM

TESLA BEARS SOME BLAME FOR SELF-DRIVING CRASH DEATH, FEDS SAY

TESLA

IT'S BEEN NEARLY a year and a half since Joshua Brown became the first person to die in a car driving itself. In May 2016, Brown was on a Florida highway in his Tesla Model S using Autopilot, the semi-autonomous driver assist feature that handles steering and speed during highway driving.

Tesla has always warned drivers that Autopilot isn't perfect. According to car's driving manual and the disclaimer drivers accept before they can engage it, the system should only

European Commission - Press release

Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service

Brussels, 27 June 2017

The European Commission has fined Google €2.42 billion for breaching EU antitrust rules. Google has abused its market dominance as a search engine by giving an illegal advantage to another Google product, its comparison shopping service.

The company must now end the conduct within 90 days or face penalty payments of up to 5% of the average daily worldwide turnover of Alphabet, Google's parent company.

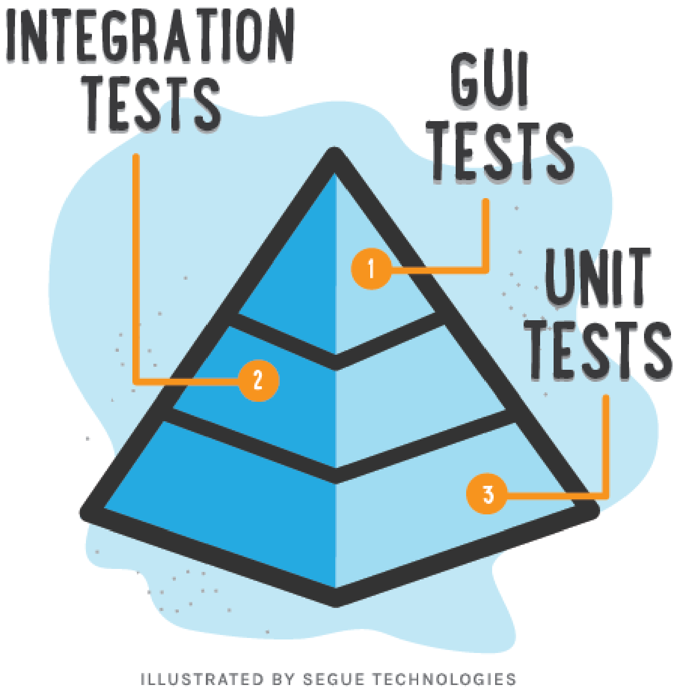
Commissioner Margrethe **Vestager,** in charge of competition policy, said: "*Google has come up with many innovative products and services that have made a difference to our lives. That's a good thing. But Google's strategy for its comparison shopping service wasn't just about attracting customers by making its product better than those of its rivals. Instead, Google abused its market dominance as a search engine by promoting its own comparison shopping service in its search results, and demoting those of competitors.*

http://europa.eu/rapid/press-release_IP-17-1784_en.htm

https://www.wired.com/story/tesla-ntsb-autopilot-crash-death/

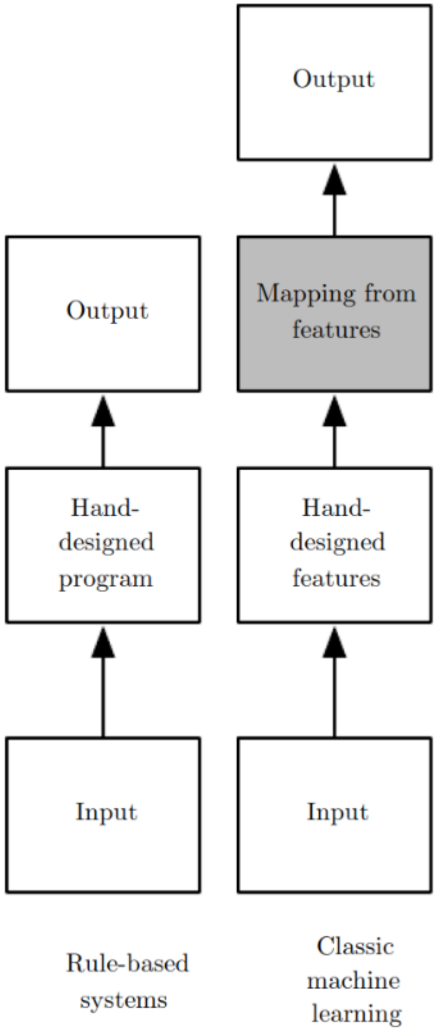
No/Difficult Testing

- **With traditional approaches, unit testing is easy**
- **For instance, title detection – detected title for a specific document will not change when e.g. new documents are added to the corpus.**



https://ekiy5aot90-flywheel.netdna-ssl.com/wp-content/uploads/2014/10/segue-blog-benefits-unit-testing.png

Thank you



Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning* (MIT press, 2016).