MACHINE LEARNING & INDUSTRY

Review and trends

Dr. Esteban Guerrero
esteban@cs.umu.se

UMEÅ UNIVERSITY

AGENDA

1. Introduction
2. Machine Learning and Industry, European and Swedish context
3. ML approaches (hands-on)
4. Industry 4.0
5. ML in two industrial sectors (cases)
6. Concluding remarks
7. Activity: co-analysis of ML+IND scenarios

DR. ESTEBAN GUERRERO

- Current position: Researcher
- Industrial experience:
  - 4 years R&D engineer in a Colombian telecommunications company. Role: support back-end services.
- Education:
  - Ph.D. in Computing Science, Umeå University. Sweden.
  - Ph. Licentiate in Computing Science, Umeå University.
  - M.Sc. Master’s degree in Computer Science, Malmö University. Sweden.
  - M.Sc. Master’s studies in Telematics Engineering, University of Cauca. Colombia.
  - B.Eng. Bachelor degree in Electronic and Telecommunications Engineering, University of Cauca. Colombia.

STAR-C: Sustainable behaviour change for health supported by person-Tailored, Adaptive, Risk-aware digital Coaching in a social context

Partners: **Department of Computing Science, Department of Epidemiology and Global Health, Department of Culture and Media Studies** and **Department of Social Work** Umeå University

Project period

2018-12-01 – 2024-11-30

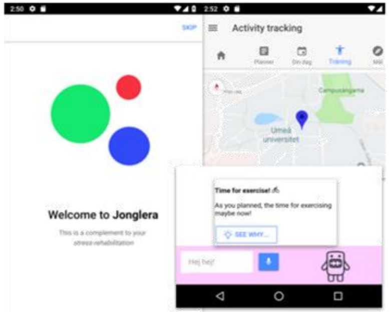
Funding agency

FORTE

Budget

14'940.000 SEK (Swedish krona)

Research subject: Public health and community medicine



Jonglera - an agent-b coaching system for s management

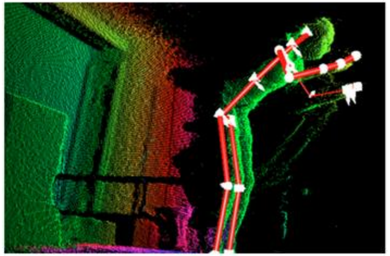
Partners: **Computing Science** dept. ar **Psychology**, Umeå Universit

Key words: multi-agent system; argu theory; coalitions; stress; psycho

2017-2019

Autonomous adap software agents support of human

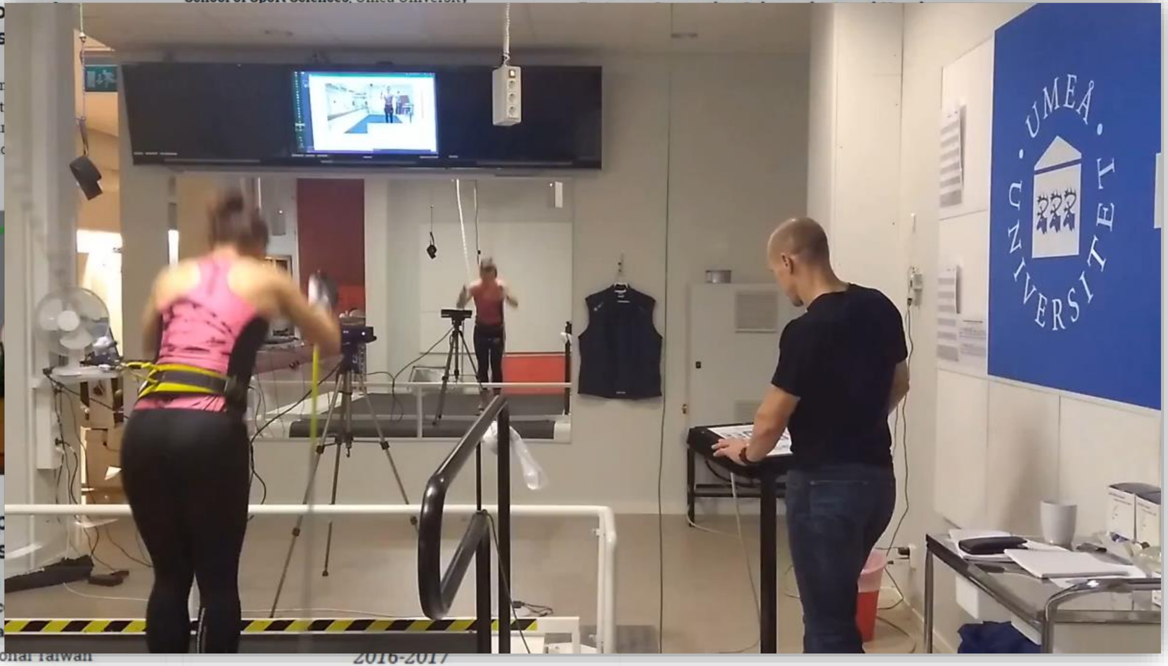
Partners: **Computing Scienc** University Sweden and **Huma and Technology Lab.**, National Taiwan University, Taiwan



Intelligent skiing coach

Partners: **Computing Science** dept. and **Umeå School of Sport Sciences**, Umeå University

Towards a *trusted* intelligent coach



2016-2017

https://people.cs.umu.se/esteban/

EUROPEAN CONTEXT

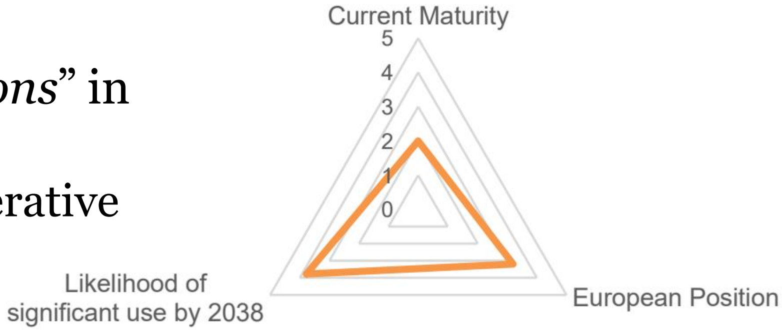
Main source:
https://ec.europa.eu/info/index_sv

Europeiska kommissionen

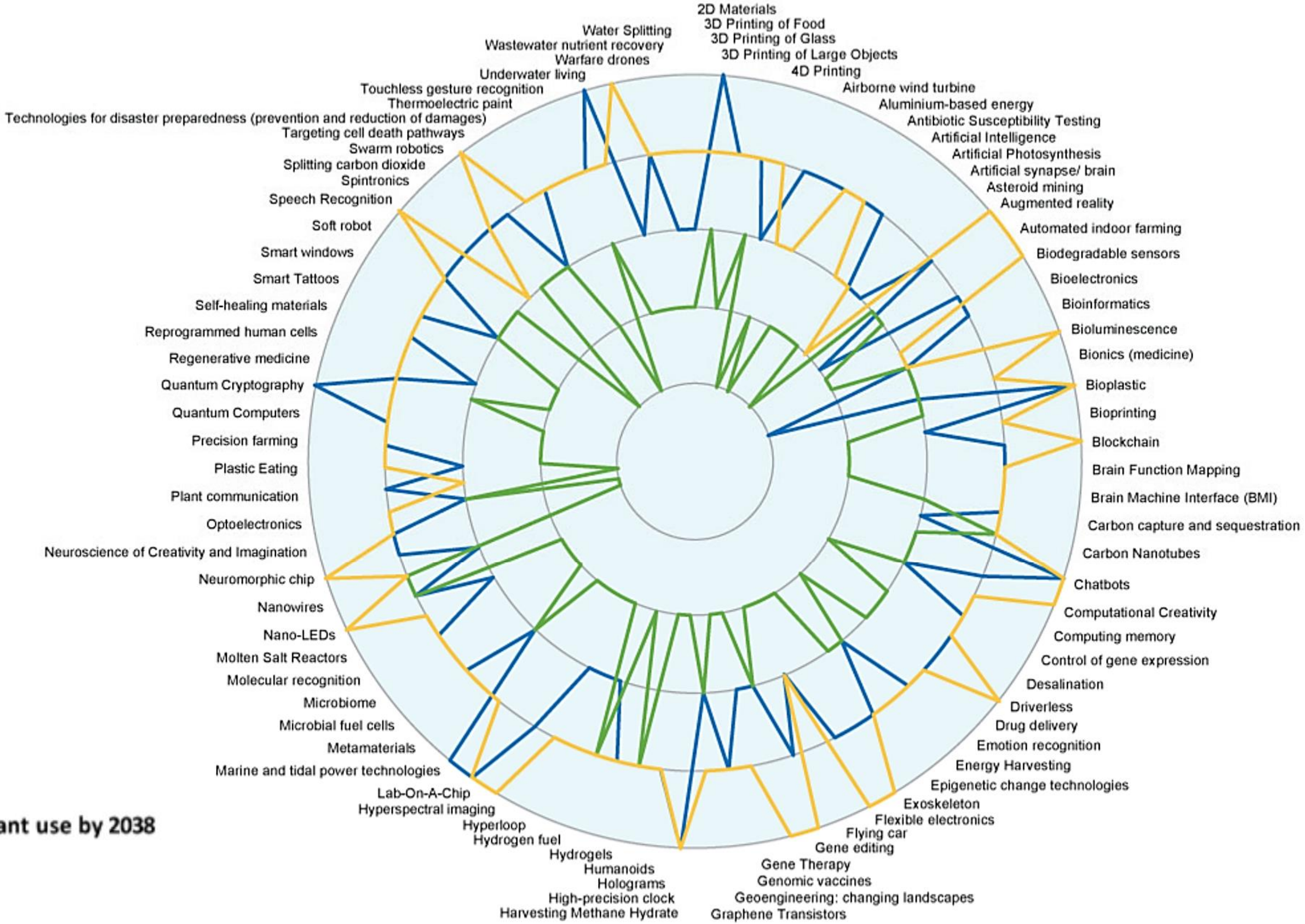
https://ec.europa.eu/info/sites/info/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_radical-innovation-breakthrough_052019.pdf

Key in the report:

- Summary of "*recent progress directions*" in artificial intelligence:
  - Duelling networks, also called generative adversarial networks (GANs)
  - Capsule Networks
- Long term perspectives:
  "In the next decade and beyond, one can expect significant research around *one-shot* and *zero-shot learning models* involving knowledge transfer."

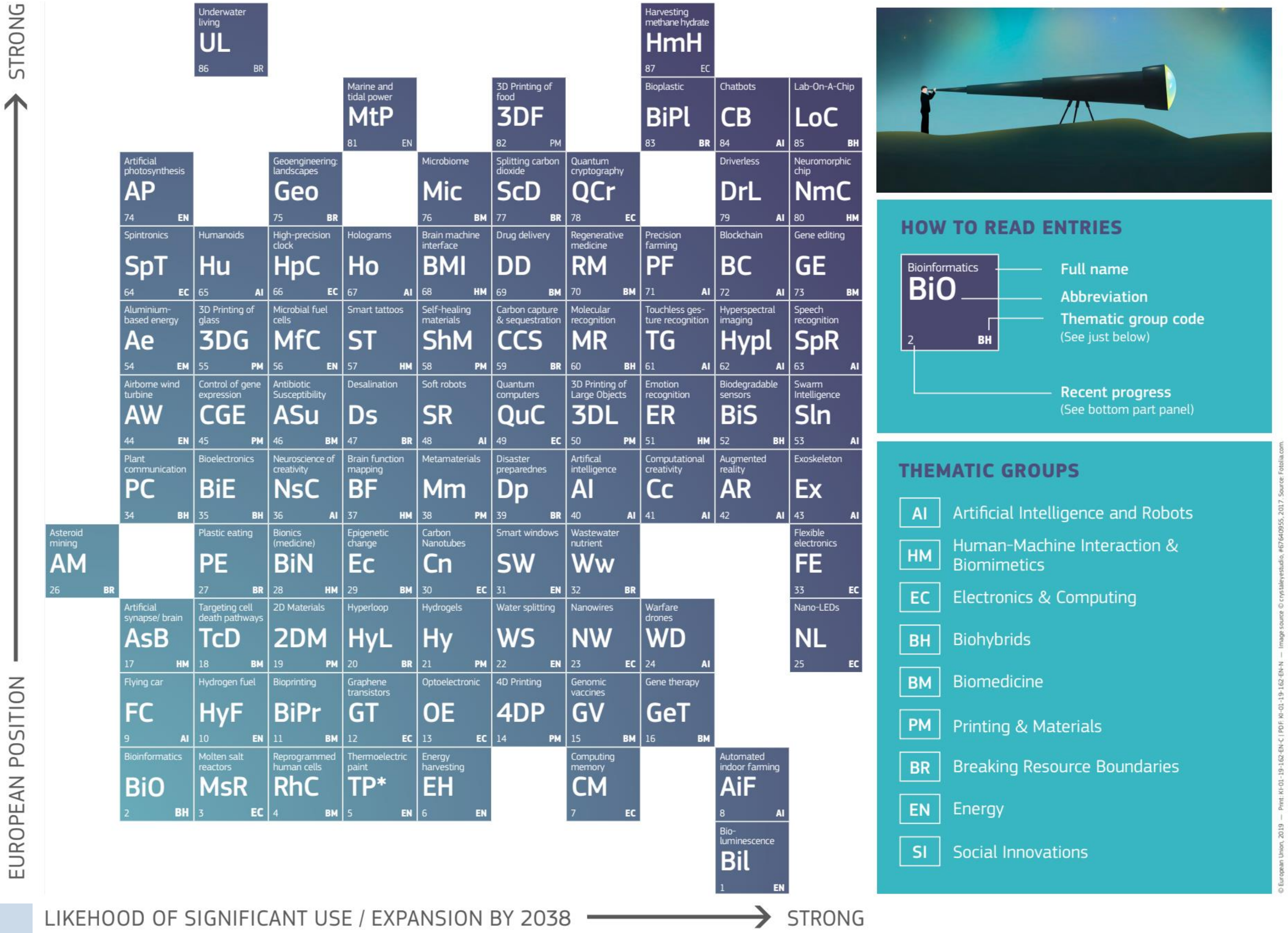


https://ec.europa.eu/info/sites/info/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_radical-innovation-breakthrough_052019.pdf



Current Maturity
European Position
Likelihood of significant use by 2038

2. Context



STRONG
EUROPEAN POSITION
LIKEHOOD OF SIGNIFICANT USE / EXPANSION BY 2038
STRONG
Underwater living UL 86 BR
Harvesting methane hydrate HmH 87 EC
Marine and tidal power MtP 81 EN
3D Printing of food 3DF 82 PM
Bioplastic BiPl 83 BR
Chatbots CB 84 AI
Lab-On-A-Chip LoC 85 BH
Artificial photosynthesis AP 74 EN
Geoengineering: landscapes Geo 75 BR
Microbiome Mic 76 BM
Splitting carbon dioxide ScD 77 BR
Quantum cryptography QCr 78 EC
Driverless DrL 79 AI
Neuromorphic chip NmC 80 HM
Spintronics SpT 64 EC
Humanoids Hu 65 AI
High-precision clock HpC 66 EC
Holograms Ho 67 AI
Brain machine interface BMI 68 HM
Drug delivery DD 69 BM
Regenerative medicine RM 70 BM
Precision farming PF 71 AI
Blockchain BC 72 AI
Gene editing GE 73 BM
Aluminium-based energy Ae 54 EM
3D Printing of glass 3DG 55 PM
Microbial fuel cells MfC 56 EN
Smart tattoos ST 57 HM
Self-healing materials ShM 58 PM
Carbon capture & sequestration CCS 59 BR
Molecular recognition MR 60 BH
Touchless gesture recognition TG 61 AI
Hyperspectral imaging HypI 62 AI
Speech recognition SpR 63 AI
Airborne wind turbine AW 44 EN
Control of gene expression CGE 45 PM
Antibiotic Susceptibility ASu 46 BM
Desalination Ds 47 BR
Soft robots SR 48 AI
Quantum computers QuC 49 EC
3D Printing of Large Objects 3DL 50 PM
Emotion recognition ER 51 HM
Biodegradable sensors BiS 52 BH
Swarm Intelligence SIn 53 AI
Plant communication PC 34 BH
Bioelectronics BiE 35 BH
Neuroscience of creativity NsC 36 AI
Brain function mapping BF 37 HM
Metamaterials Mm 38 PM
Disaster preparednes Dp 39 BR
Artifical intelligence AI 40 AI
Computational creativity Cc 41 AI
Augmented reality AR 42 AI
Exoskeleton Ex 43 AI
Asteroid mining AM 26 BR
Plastic eating PE 27 BR
Bionics (medicine) BiN 28 HM
Epigenetic change Ec 29 BM
Carbon Nanotubes Cn 30 EC
Smart windows SW 31 EN
Wastewater nutrient Ww 32 BR
Flexible electronics FE 33 EC
Artificial synapse/ brain AsB 17 HM
Targeting cell death pathways TcD 18 BM
2D Materials 2DM 19 PM
Hyperloop HyL 20 BR
Hydrogels Hy 21 PM
Water splitting WS 22 EN
Nanowires NW 23 EC
Warfare drones WD 24 AI
Nano-LEDs NL 25 EC
Flying car FC 9 AI
Hydrogen fuel HyF 10 EN
Bioprinting BiPr 11 BM
Graphene transistors GT 12 EC
Optoelectronic OE 13 EC
4D Printing 4DP 14 PM
Genomic vaccines GV 15 BM
Gene therapy GeT 16 BM
Bioinformatics BiO 2 BH
Molten salt reactors MsR 3 EC
Reprogrammed human cells RhC 4 BM
Thermoelectric paint TP* 5 EN
Energy harvesting EH 6 EN
Computing memory CM 7 EC
Automated indoor farming AiF 8 AI
Bio-luminescence BiI 1 EN
HOW TO READ ENTRIES
Bioinformatics BiO 2 BH
Full name
Abbreviation
Thematic group code
(See just below)
Recent progress
(See bottom part panel)
THEMATIC GROUPS
AI Artificial Intelligence and Robots
HM Human-Machine Interaction & Biomimetics
EC Electronics & Computing
BH Biohybrids
BM Biomedicine
PM Printing & Materials
BR Breaking Resource Boundaries
EN Energy
SI Social Innovations
© European Union, 2019 — Print: KI-01-19-162-EN-C | PDF KI-01-19-162-EN-N — Image source © crystaleyestudio, #67640955, 2017. Source: Fotolia.com

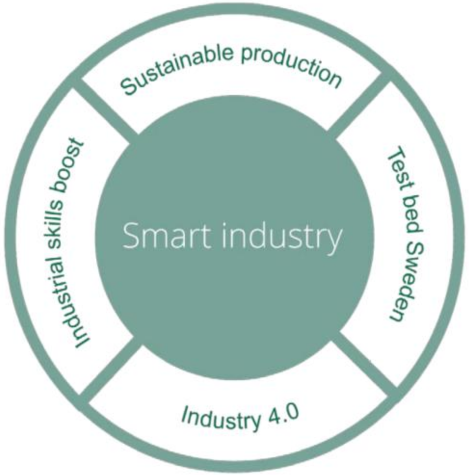
SWEDISH CONTEXT

Main sources: Vinnova, Regeringskansliet

SMART INDUSTRY - A STRATEGY FOR NEW INDUSTRIALISATION FOR SWEDEN

Four focus areas have been chosen:

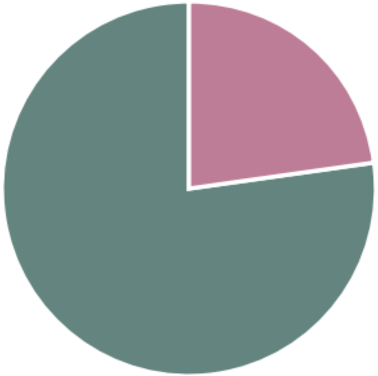
- Industry 4.0 – Exploit the potential of digitalisation
- Sustainable production – Improve the industrial sector's capacity for sustainable and resource-efficient production
- Industrial skills boost – Ensure the supply of skil to the industrial sector
- Test bed Sweden – Create attractive innovation environments



https://www.government.se/information-material/2016/04/smart-industry---a-strategy-for-new-industrialisation-for-sweden/

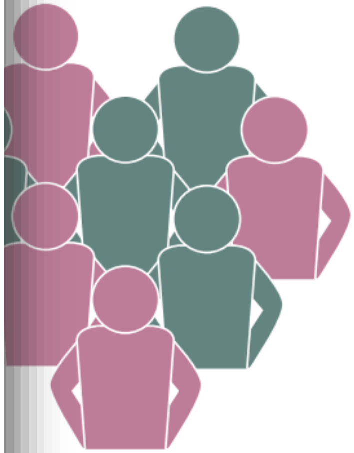
https://www.regeringen.se/

The industry sector
services sector acc



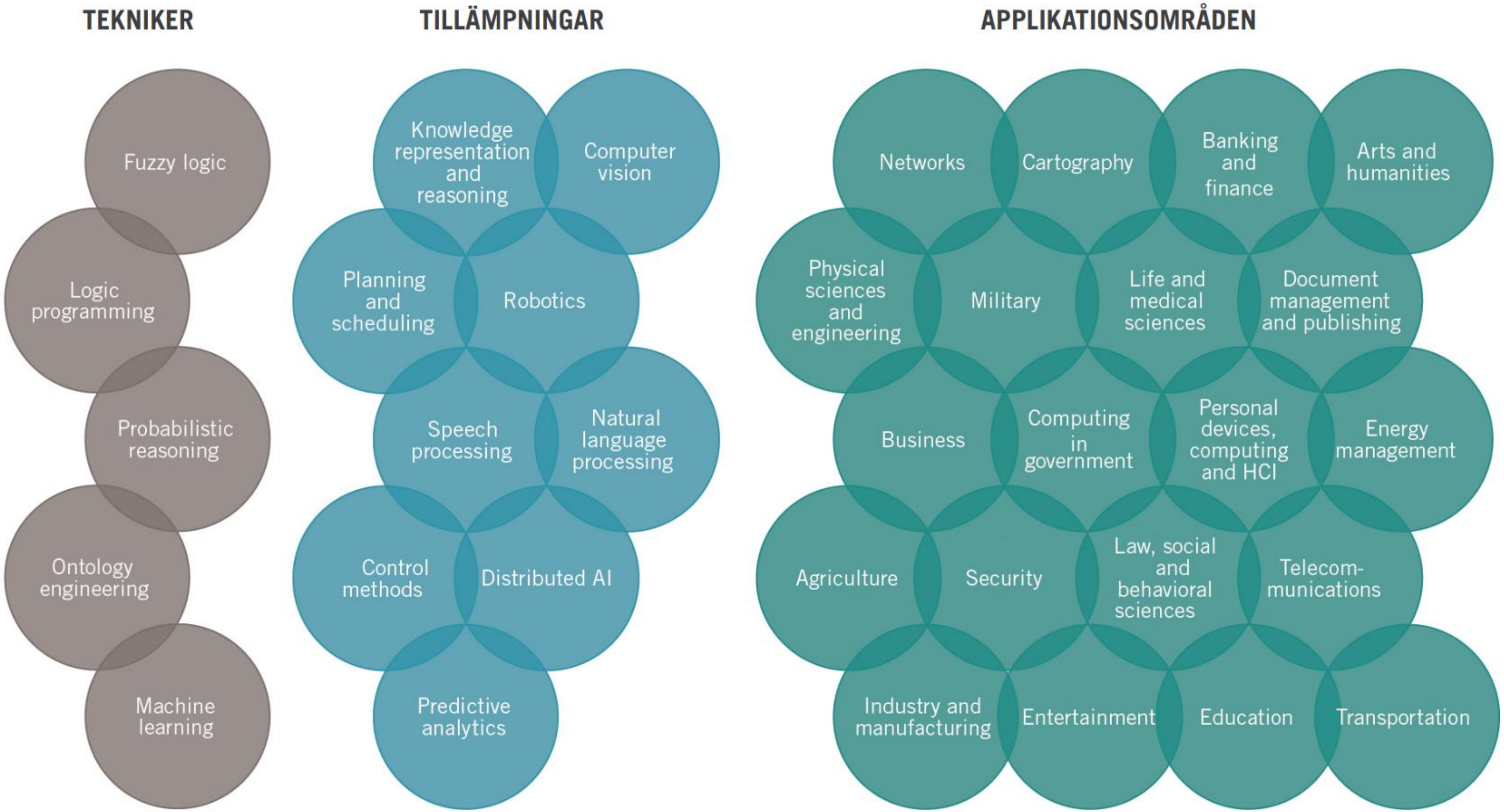
20 per cent of
Sweden's GDP

Sweden ranks highly in international indexes ...
...but is being overtaken by other countries in competitiveness rankings.



und 1 million jobs

https://www.government.se/4 … e4a7ad75f2f54/nist_a4_faktablad_160701_eng_web.pdf



Artificiell intelligens: tekniker, tillämpningar och applikationsområden

Källa: WIPO, Technology trends Artificial Intelligence (2019).

Svenska AI-miljöer med geografisk hemvist.

Västerås
Robotdalen
Synerleap
AOD-AI

Eskilstuna
MITC

Örebro
AASS

Skövde
SAIL

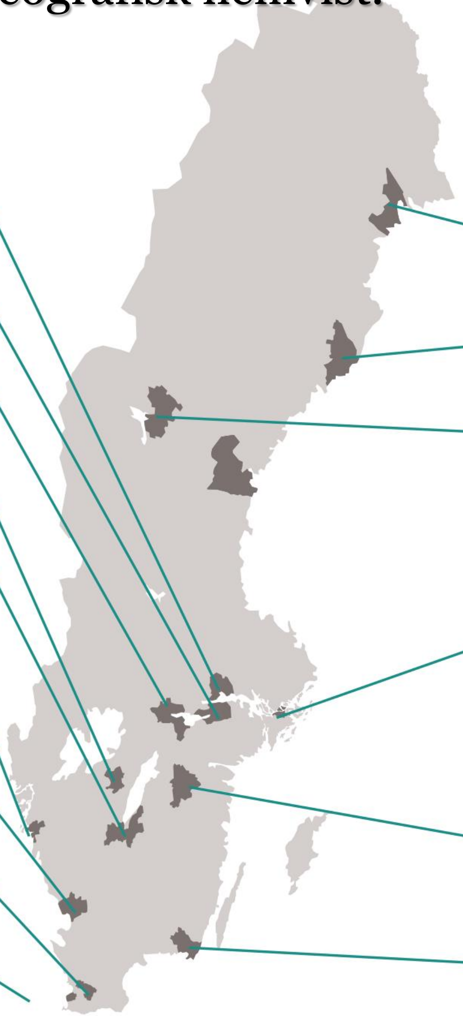
Jönköping
SPARK

Göteborg
AI Innovation of Sweden - Västra noden
CHAIR
SII-Lab
Software Center

Halmstad
CAISR
AIWay Halland

Lund
AI.ML

Malmö
IOATP



Nationella miljöer
AI Competence for Sweden
SAIS
WASP

Luleå
SICS ICE
Data Center Innovation Region

Umeå
UmeAI

Östersund & Sundsvall
NIIT
Transformative Technologies

Stockholm
RISE AI (Kista)
Stockholm AI
Sting Test Drive
Ignite Sweden
I-AID
Kista Science City
Urban ICT Arena
CDA

Linköping
AIDA
AgTech 2030
Visual Sweden (Norrköping)

Karlskrona
Big Data @BTH

Centers with focus on Machine Learning

39 AI-miljöer som arbetar för utveckling av artifciell intelligens

https://www.vinnova.se/globalassets/mikrosajter/ai/vr_19-05_190704.pdf

ML APPROACHES (BRIEF)
+
HANDS-ON

MACHINE LEARNING IS NOT ONLY PYTHON

python C++ Java Swift

Go julia

C# Haskell R JS

SUPERVISED LEARNING

- What is: “Supervised learning is a learning model built to make prediction, given an unforeseen input instance.”

- How it works: “With supervised learning you use labeled data, which is a data set that has been classified, to infer a learning algorithm. The data set is used as the basis for predicting the classification of other unlabeled data” [Talabis,et.al.2014].

- Two important approaches (among many):
  - Linear Regression (many sub-mechanisms)
  - Classification Techniques: Logistic regression, Linear discriminant analysis, etc. …(among many).

Talabis, M., McPherson, R., Miyamoto, I., & Martin, J. (2014). *Information Security Analytics: Finding Security Insights, Patterns, and Anomalies in Big Data*. Syngress.

Sugiyama, M. (2015). *Introduction to statistical machine learning*. Morgan Kaufmann.

SUPERVISED LEARNING

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} \mathbf{x}_1^\mathsf{T} \\ \mathbf{x}_2^\mathsf{T} \\ \vdots \\ \mathbf{x}_n^\mathsf{T} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

as Y.

- We assume the equation: $\boldsymbol{Y_e = \varepsilon + \beta X}$
- Y_e is the estimated or predicted value of Y based on our linear equation.

Goal: find statistically significant values of the **parameters *ε*** and ***β*** that minimize the difference between Y and Y_e.

If we are able to determine the optimum values of these two parameters, then we will have the **line of best fit** that we can use to predict the values of Y, given the value of X.

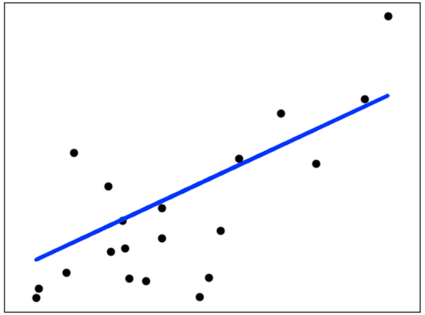
SUPERVISED LEARNING

- Example 0: Linear regression, a diabetes dataset with Python
  - Jupyter notebook using Python3, scikit-learn

https://colab.research.google.com/drive/1Ey0_E-fCtggCNxdd96v8gxbCnAKh7_ug

Patient	AGE x1	SEX x2	BMI x3	BP x4	··· Serum x5	x6	x7	x8	x9	Measurements ··· x10	Response y
1	59	2	32.1	101	157	93.2	38	4	4.9	87	151
2	48	1	21.6	87	183	103.2	70	3	3.9	69	75
3	72	2	30.5	93	156	93.6	41	4	4.7	85	141
4	24	1	25.3	84	198	131.4	40	5	4.9	89	206
5	50	1	23.0	101	192	125.4	52	4	4.3	80	135
6	23	1	22.6	89	139	64.8	61	2	4.2	68	97
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
441	36	1	30.0	95	201	125.2	42	5	5.1	85	220
442	36	1	19.6	71	250	133.2	97	3	4.6	92	57

Table 1. Diabetes study. 442 diabetes patients were measured on 10 baseline variables. A prediction model was desired for the response variable, a measure of disease progression one year after baseline.



https://web.stanford.edu/~hastie/Papers/LARS/LeastAngle_2002.pdf

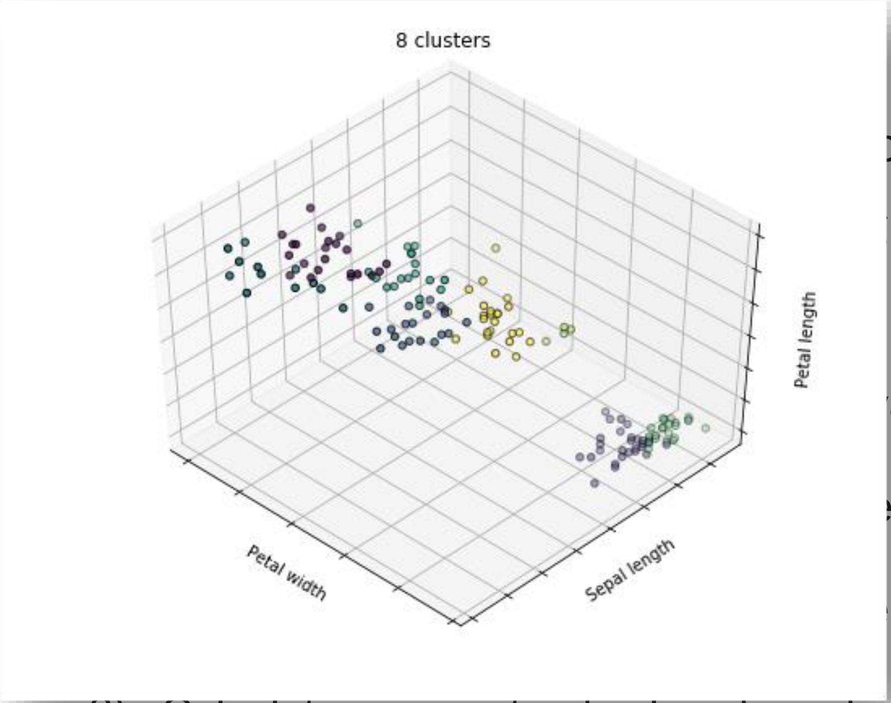
https://www4.stat.ncsu.edu/~boos/var.select/diabetes.tab.txt

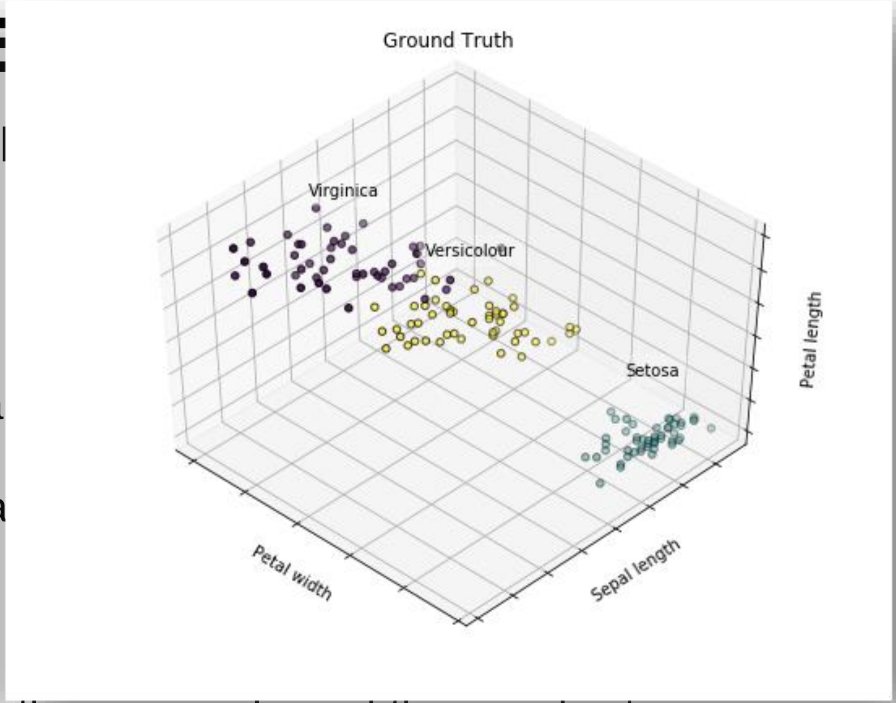
Out:
```
Coefficients:
 [938.23786125]
Mean squared error: 2548.07
Variance score: 0.47
```

UNSUPERVISED LEARNING

- What is: "Unsupervised learning finds structures in the data."
- How does it work?: "Labels for the data instances or other forms of guidance for training are not necessary. This makes unsupervised learning attractive in applications where data is cheap to obtain, but labels are either expensive or not available." [Wittek. 2014].
- Two important approaches (among many):
  - Clustering (many sub-mechanisms)
  - Principal Components Analysis (among many).

Wittek, P. (2014). *Quantum machine learning: what quantum computing means to data mining*. Academic Press.





3) Calculate new centroid values based on the mean values of the coordinates of all the data instances from the corresponding cluster.

UNSUPERVISED LEARNING

Example 1: K-means with Python + Scikit-learn

Iris flower data set clustering

	Sepal length	Sepal width	Petal length	Petal width	Class
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
⋮	⋮	⋮	⋮	⋮	⋮
150	5.9	3.0	5.1	1.8	virginica

Iris Versicolor   Iris Setosa   Iris Virginica

https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_iris.html#sphx-glr-auto-examples-cluster-plot-cluster-iris-py

UNSUPERVISED LEARNING

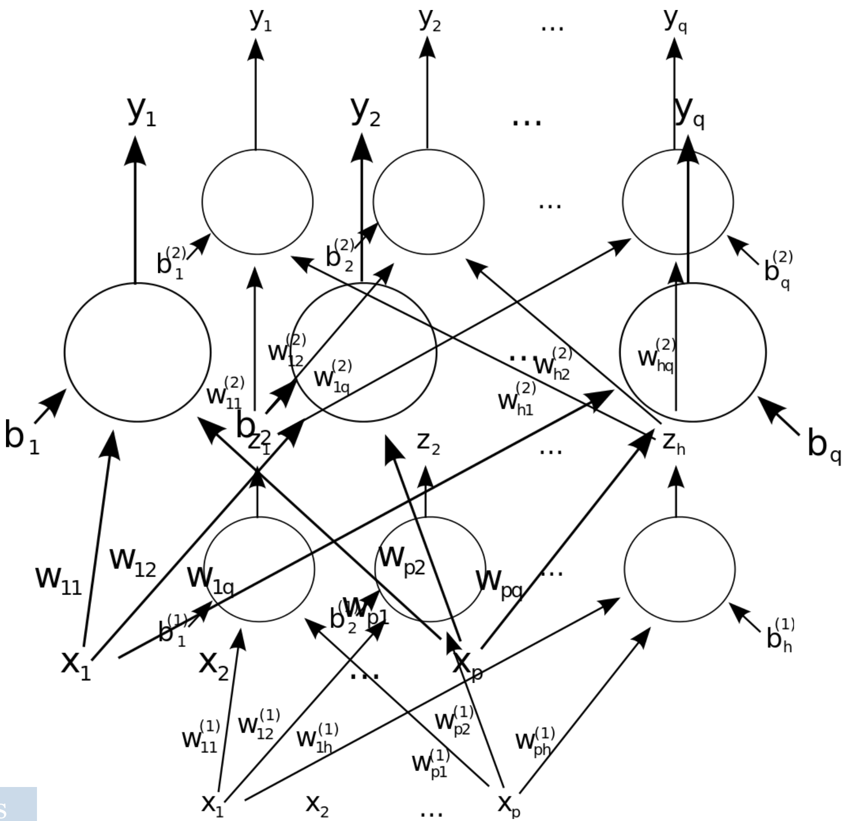
Example 1: K-means with Python + Scikit-learn

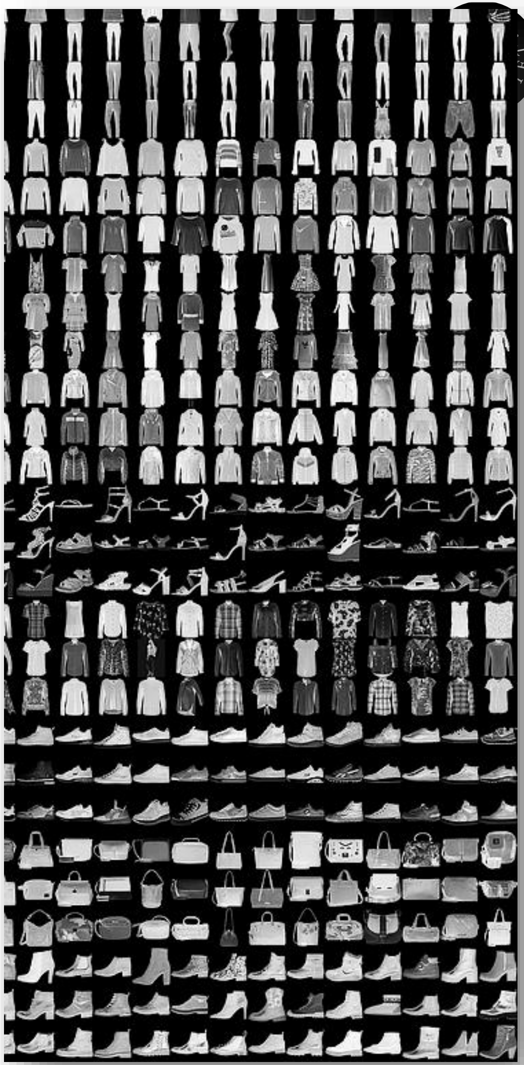
Jupyter notebook:

https://colab.research.google.com/drive/1dCb2q2ONQA3ge
g2VBZ03bvxSdmGV_JF0

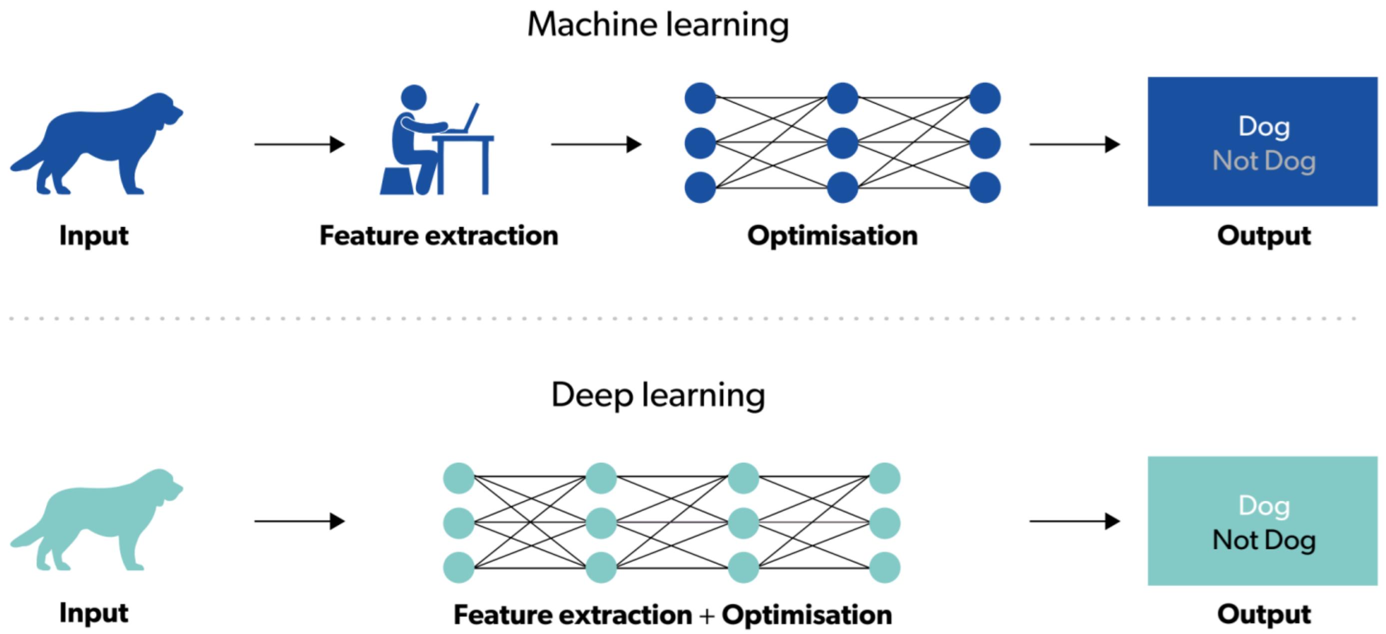
https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_iris.html#sphx-glr-auto-examples-cluster-plot-cluster-iris-py

DEEP LEARNING





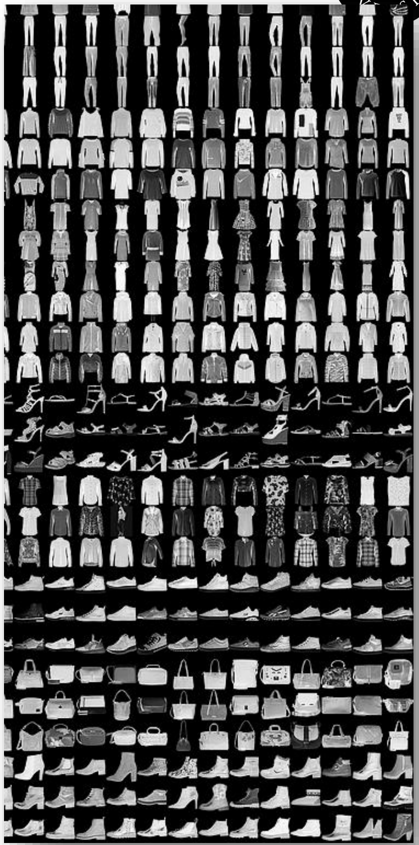
DEEP LEARNING



Source: MMC Ventures

Deep learning offloads the burden of feature extraction from a programmer to her program

DEEP LEARNING



Example 2: Fashion MNIST dataset + TensorFlow + Python

Goal: Classify images.

How: Train a network with 60000 examples. Evaluate with 10000.

Reference: https://www.tensorflow.org/tensorboard/tensorboard_in_notebooks

DEEP LEARNING

Example 2: Fashion MNIST dataset + TensorFlow + Python

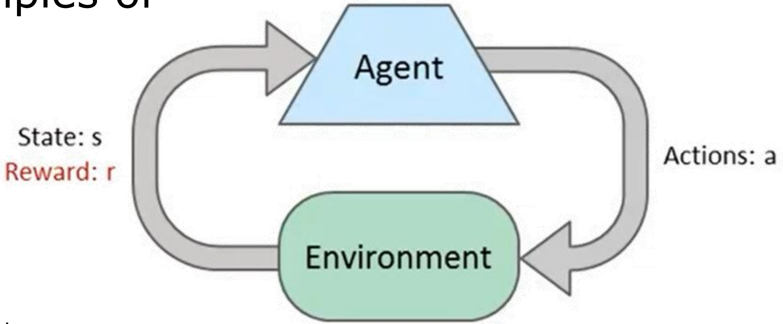
See Jupyter notebook:

https://colab.research.google.com/drive/13YG9iSbbDRDKPOrD4sE0G8w_Vd2Q-_sF

Reference: https://www.tensorflow.org/tensorboard/tensorboard_in_notebooks

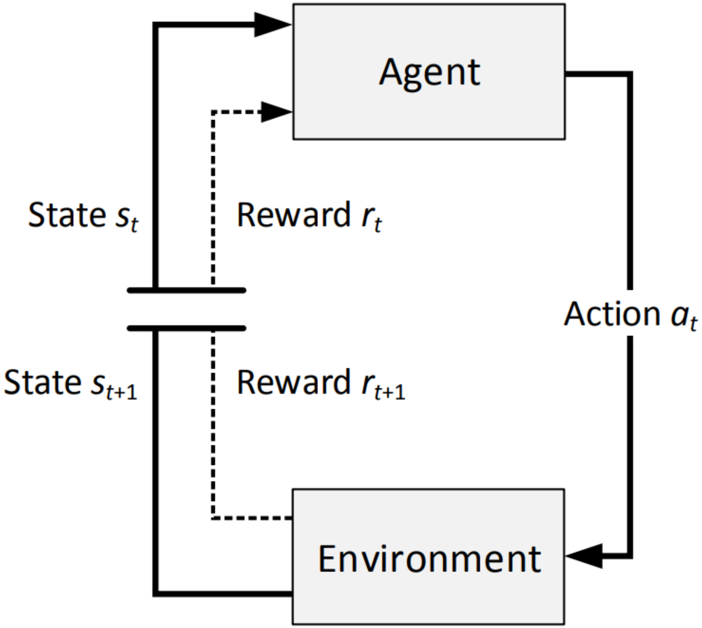
REINFORCEMENT LEARNING

- Basic idea:
  - Agent receives feedback in the form of rewards
  - Agent's utility is defined by the reward function
  - Must (learn to) act so as to maximize expected rewards
  - All learning is based on observed samples of outcomes



See CS188 Artificial Intelligence UC Berkeley. Prof. Pieter Abbeel https://www.youtube.com/watch?v=ifma8G7LegE

REINFORCEMENT LEARNING



At each step t, the agent:

- Executes action a_t
- Receives observation s_t
- Receives scalar reward r_t

The environment:

- Changes upon action a_t
- Emits observation s_{t+1}
- Emits scalar reward r_{t+1}

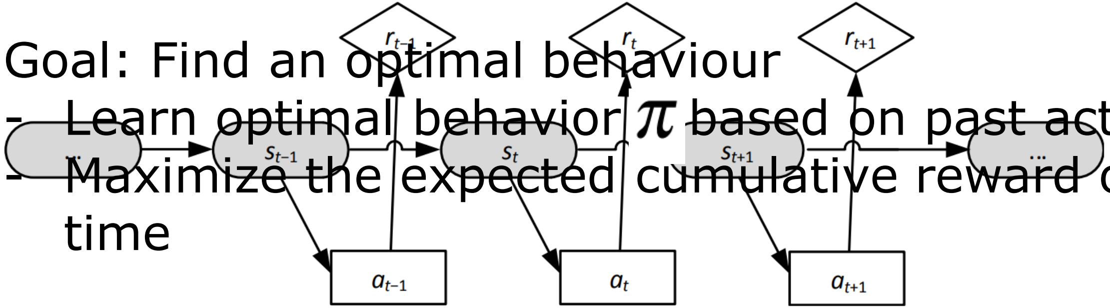
- Time step t is incremented after each iteration

REINFORCEMENT LEARNING

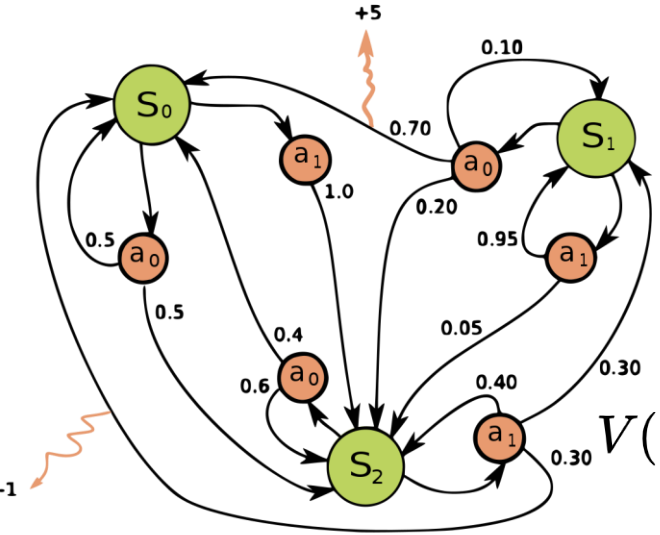
1	ENVIRONMENT	▸ You are in state 3 with 4 possible actions
2	AGENT	▸ I'll take action 2
3	ENVIRONMENT	▸ You received a reward of 5 units
		▸ You are in state 1 with 2 possible actions
⋮	⋮	⋮

Goal: Find an optimal behaviour

- Learn optimal behavior π based on past actions.
- Maximize the expected cumulative reward over time



REINFORCEMENT LEARNING



$$\Pr(s_{t+1} = s' \mid s_t = s, a_t = a)$$

..and we know that the action is determined by

$$\Pr(s_{t+1} = s' \mid s_t = s) \;\; \pi(s)$$

$$V(s) := \sum_{s'} P_{\pi(s)}(s, s') \left(R_{\pi(s)}(s, s') + \gamma V(s') \right)$$

Markov Decision Process

$$\pi(s) := \operatorname{argmax}_a \left\{ \sum_{s'} P(s' \mid s, a) \left(R(s' \mid s, a) + \gamma V(s') \right) \right\}$$

https://en.wikipedia.org/wiki/File:Markov_Decision_Process.svg

REINFORCEMENT LEARNING

- Example 3:OpenAI Gym
  https://gym.openai.com/

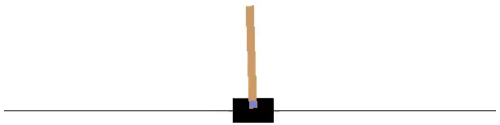
Gym

Gym is a toolkit for developing and comparing reinforcement learning algorithms. It supports teaching agents everything from walking to playing games like Pong or Pinball.

```
import gym
env = gym.make("CartPole-v1")
observation = env.reset()
for _ in range(1000):
  env.render()
  action = env.action_space.sample() # your agent here (this takes random actions)
  observation, reward, done, info = env.step(action)

  if done:
    observation = env.reset()
env.close()
```

Lubuntu
Python 3.7



INDUSTRY 4.0

INDUSTRI*E* 4.0

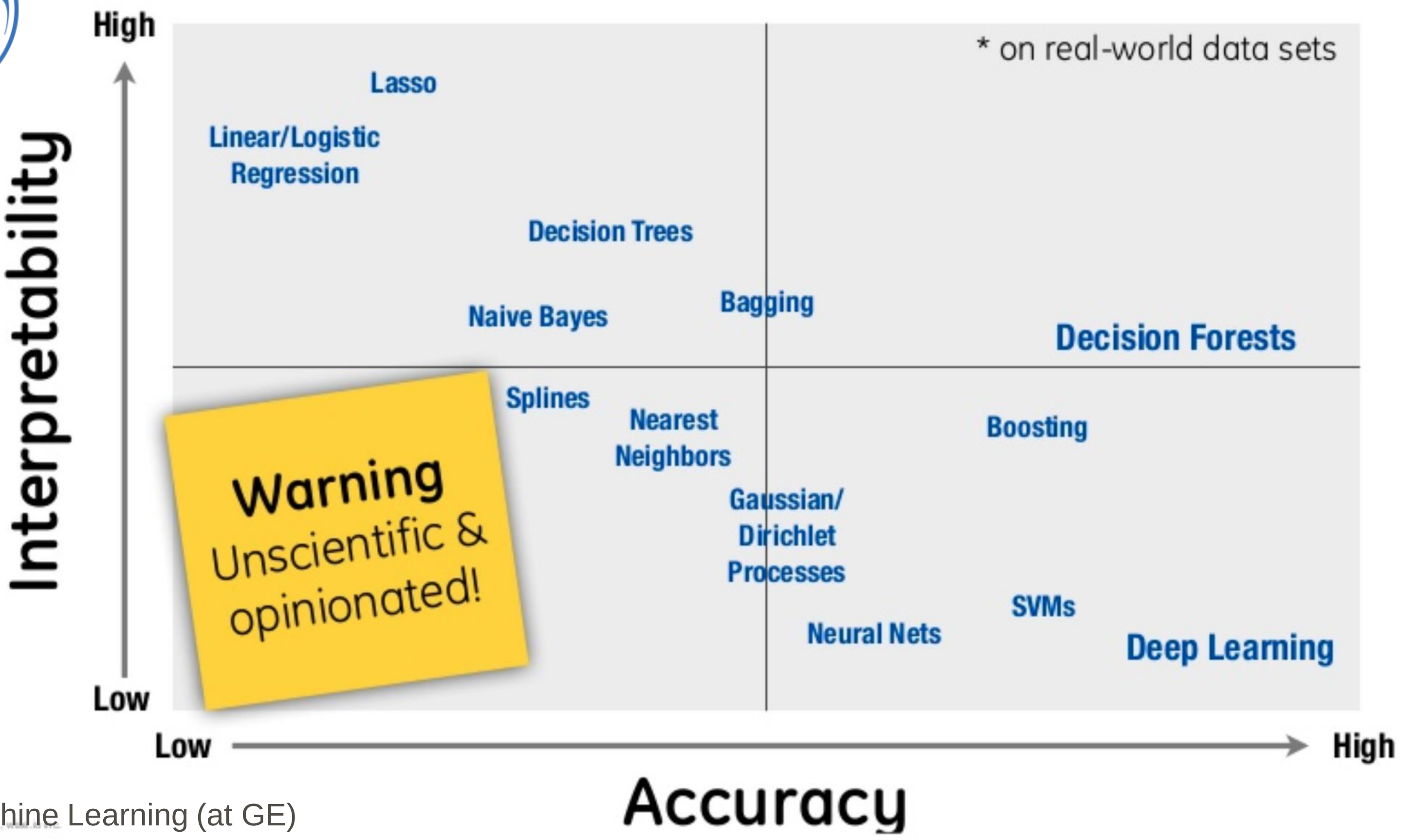
The Fourth Industrial Revolution aims to leverage differences between the physical, digital, and biological sphere. It integrates cyber-physical systems and the Internet of Things, big data_and cloud computing, robotics, artificial-intelligence based systems and additive manufacturing

Expected effects:

- On the business side: it drastically modifies customer expectations, product enhancement, collaborative innovation and organisational forms.
- On people: one of the greatest challenges is on privacy, on the notion of ownership, consumer patterns and how we devote time to develop skills.

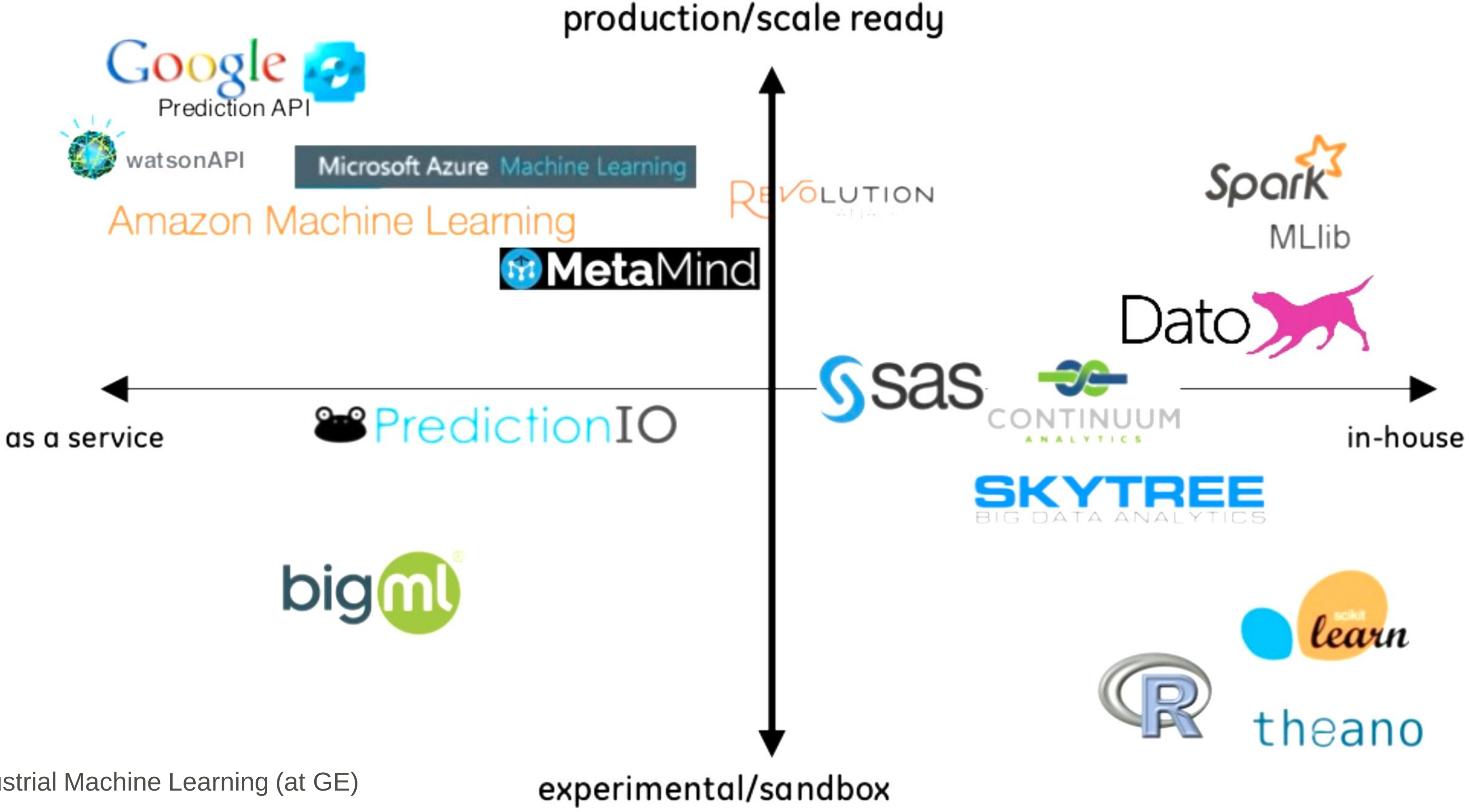
European Commission, strategy on digitising the European industry.
https://ec.europa.eu/digital-single-market/en/fourth-industrial-revolution

One ML Algorithmic Trade-Off



Industrial Machine Learning (at GE)

Joshua Bloom, Professor at UC Berkeley, CTO Wise/GE



Industrial Machine Learning (at GE)

Joshua Bloom, Professor at UC Berkeley, CTO Wise/GE

Sector	Core use cases:			
Asset Management	Investment strategy	Portfolio construction	Risk management	Client service
Healthcare	Diagnostics	Drug discovery	Monitoring	
Insurance	Risk assessment	Claims processing	Fraud detection	Customer service
Law & compliance	Case law	Discovery and due diligence	Litigation strategy	Compliance
Manufacturing	Predictive maintenance	Asset performance	Utility optimisation	
Retail	Customer segmentation	Content personalisation	Price optimisation	Churn prediction
Transport	Autonomous vehicles	Infrastructure optimisation	Fleet management	Control applications
Utilities	Supply management	Demand optimisation	Security	Customer experience

Source: MMC Ventures

https://www.mmcventures.com/wp-content/uploads/2019/02/The-State-of-AI-2019-Divergence.pdf

Challenges in industrial cyber-physical systems

Area	Key Challenges	Difficulty	Priority	Maturity in
CPS Capabilities	Real-time control of CPS systems	High	High	4–7 years
	Real-time CPS SoS	High	Medium	3–5 years
	Optimization in CPS and their application	High	Medium	4–7 years
	On-CPS advanced analytics	Medium	High	3–5 years
	Modularization and servification of CPS	low	High	3–5 years
	Energy efficient CPS	Medium	Medium	3–5 years
CPS Management	Lifecycle management of CPS	Medium	Medium	5–8 years
	Management of (very) large scale CPS and CPS-SoS	High	High	5–8 years
	Security and trust management for heterogeneous CPS	High	High	5–8 years
CPS Engineering	Safe programming and validation of CPS SoS	High	High	5–10+ years
	Resilient risk-mitigating CPS	High	High	5–10+ years
	Methods and tools for CPS lifecycle support	High	High	3–7 years
	New operating systems and programming languages for CPS and CPS SoS	Medium	Low	3–6 years
	Simulation of CPS and of CPS-SoS	Medium	High	3–6 years
CPS Infrastructures	Interoperable CPS services	Medium	High	2–5 years
	Migration solutions to emerging CPS infrastructures	Medium	High	3–6 years
	Integration of heterogeneous/mobile hardware and software technologies in CPS	Low	Medium	2–4 years
	Provision of ubiquitous CPS services	Medium	Medium	3–5 years
	Economic impact of CPS Infrastructure	High	High	3–6 years
CPS Ecosystems	Autonomic and self-* CPS	High	Medium	7–10+ years
	Emergent behavior of CPS	High	Medium	7–10+ years
	CPS with humans in the loop	High	High	2–5 years
	Collaborative CPS	Medium	Medium	5–8 years
CPS Information Systems	Artificial intelligence in CPS	High	High	7–10+ years
	Cross-domain large-scale information integration to CPS infrastructures	Medium	Low	6–9 years
	Transformation of CPS data and information analytics to actionable knowledge	High	High	4–8 years
	Knowledge-driven decision making/management	High	Medium	6–10+ years

Leitão, P., Colombo, A. W., & Karnouskos, S. (2016). Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges. *Computers in Industry*, *81*, 11-25.

REVIEW OF
ML IN TWO SPECIFIC SECTORS

Oil and Gas 4.0

Cybersecurity

SPECIFIC INDUSTRY AREAS

Case: Cybersecurity (1/3)

- Spam detection: "The detection of spam is based on the use of filters that analyse the content and decide whether or not they are spam or legitimate messages, blogs or websites" Two main strategies can be followed to detect spam: 1) textual analysis and 2) image-based analysis.

- Major technologies: Bayesian classifiers: Naive Bayes classifiers, Boolean Naive Bayes, etc.; Support vector machines (SVM), back-propagation neural networks, among others, Deep Belief Networks

Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cyber security. *Information*, *10*(4), 122.

Torres, J. M., Comesaña, C. I., & García-Nieto, P. J. (2019). Machine learning techniques applied to cybersecurity. *International Journal of Machine Learning and Cybernetics*, 1-14.

SPECIFIC INDUSTRY AREAS

Case: Cybersecurity (2/3)

- Malware detection: "The detection of spam is based on the use of filters that analyse the content and decide whether or not they are spam or legitimate messages, blogs or websites" Two main strategies can be followed to detect spam: 1) textual analysis and 2) image-based analysis.

- Main technologies: Bayesian classifiers: Naive Bayes classifiers, Boolean Naive Bayes, etc.; Support vector machines (SVM), convolutional neural networks (CNNs) and recurrent neural networks (RNNs)

Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cyber security. *Information*, *10*(4), 122.

Torres, J. M., Comesaña, C. I., & García-Nieto, P. J. (2019). Machine learning techniques applied to cybersecurity. *International Journal of Machine Learning and Cybernetics*, 1-14.

SPECIFIC INDUSTRY AREAS

Case: Cybersecurity (3/3)

- Phishing detection: "The detection of spam is based on the use of filters that analyse the content and decide whether or not they are spam or legitimate messages, blogs or websites" Two main strategies can be followed to detect spam: 1) textual analysis and 2) image-based analysis.

- Main technologies: Bayesian classifiers, SVMs, neural networks

Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cyber security. *Information*, *10*(4), 122.

Torres, J. M., Comesaña, C. I., & García-Nieto, P. J. (2019). Machine learning techniques applied to cybersecurity. *International Journal of Machine Learning and Cybernetics*, 1-14.

SPECIFIC INDUSTRY AREAS

Datasets for cybersecurity:

- Knowledge Discovery and Dissemination (KDD) 1999 dataset: https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
- NSL-KDD dataset: https://www.unb.ca/cic/datasets/nsl.html
- CTU-13 Dataset: https://www.stratosphereips.org/datasets-ctu13/
- Others: Contagio, Comodo, the Genome Project, Virus Share, VirusTotal, DREBIN, and Microsoft Malware Classification

Torres, J. M., Comesaña, C. I., & García-Nieto, P. J. (2019). Machine learning techniques applied to cybersecurity. *International Journal of Machine Learning and Cybernetics*, 1-14.

Application scenario		Benefit
Drilling		Reduce the safety risk an
Diagnosis and detection		With the help of unmanı
Weather monitoring system		Automatic weather senso … vance, reduce construction risks
Metering system		The intelligent sensor is … tically according to the data.

Lu, H., Guo, L., Azimi, M., & Huang, K. (2019). Oil and Gas 4.0 era: A systematic review and outlook. *Computers in Industry*, *111*, 68-90.

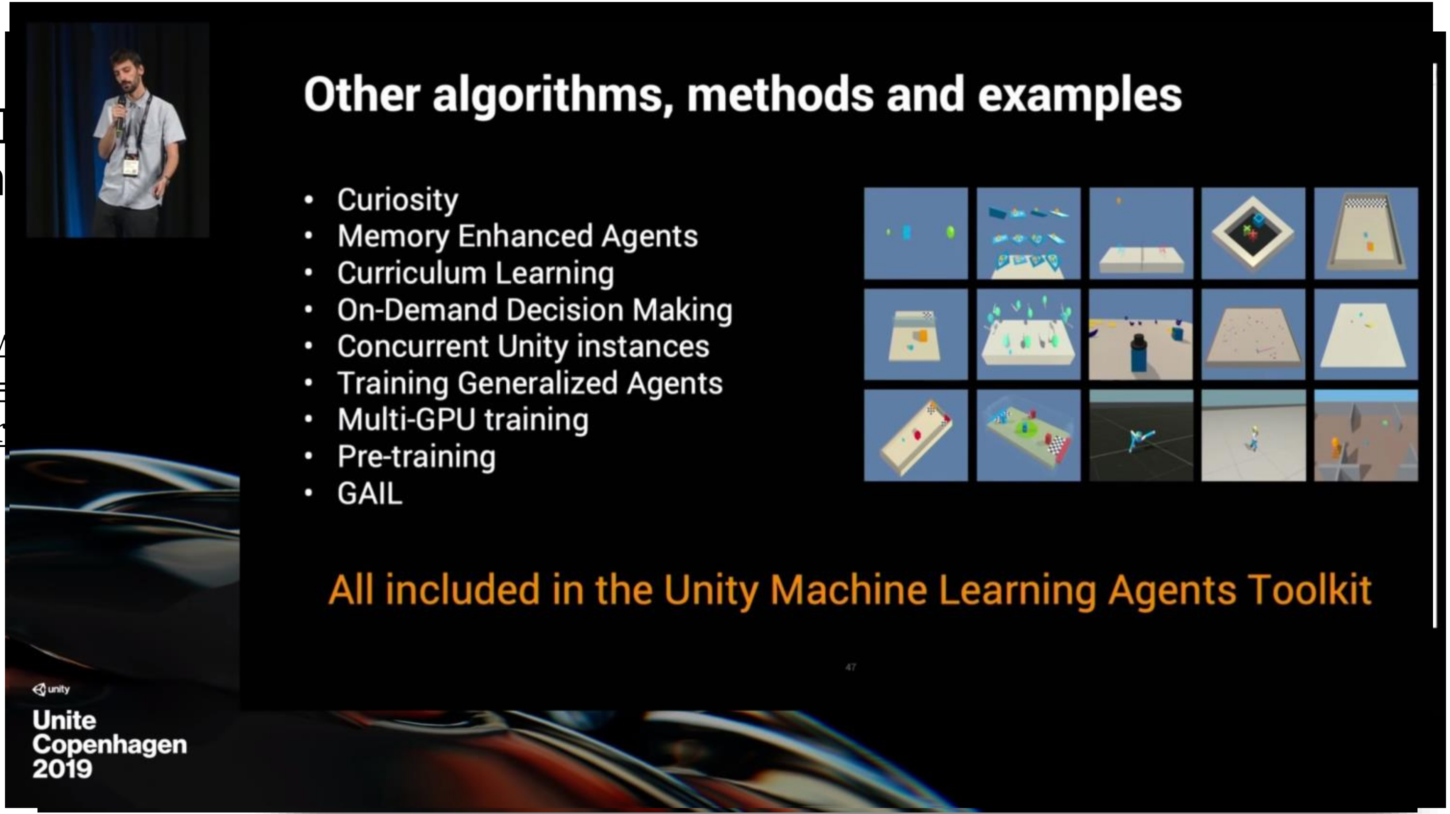
Application scenarios of big data in the oil and gas industry (Mohammadpoor and Torabi, 2018).

Domain	Application scenario
Exploration	Seismic data Micro-seismic data 1D, 2D, and 3D geological maps
Drilling	Drilling rig efficiency Drilling performance Invisible non-production time Reduce the risk of drilling operations Characterize the drill string dynamics
Reservoir engineering	Reservoir management application Closed-Loop Reservoir Management (CLRM) and Integrated Asset Modeling (IAM) Improve the CO_2 sequestration Optimization on heavy oil reservoirs Reservoir modeling for unconventional oil and gas resources Improve the modeling of hydraulically fractured reservoirs Optimize the application of EOR projects
Production engineering	Conduct automated decline analysis Production allocation technique Optimize the performance of electric submersible pumps (ESPs) Optimize the performance of rod pump wells Improve hydraulic fracturing projects Conduct field development
Refining	Petroleum asset management Management optimization of a comprehensive refinery in Spain Workflow to study the impact of well completion parameters on well productivity
Transportation	Improve shipping performance
Health and Safety Executive (HSE)	Develop an energy efficiency model during ship operations Improve the occupational safety of the oil and gas industry

GAMING INDUSTRY

- See AI
Copen

https://www
m/watch?v=
LQQ&featur



CONCLUDING

WHY DO BUSINESSES FAIL AT MACHINE LEARNING?
CHECKLIST

- ❑ Know what business you’re in.
- ❑ Do things in the right order. (don’t start with the algorithms; solve: what business problem I am solving?)
- ❑ Don’t reinvent the wheel
- ❑ Data is not the most important part (data not Data)
- ❑ To scale using ML in Industry, more humans are needed (more than data scientists / engineers; from: statistics, ethics, social work, etc.)
- ❑ Simplify where is possible
- ❑ Focus more in information ~~data~~ than the algorithms
- ❑ Design incentives that cannot be gamed (for reward functions in Reinforcement Learning: what you want the thing to learn?; and maybe for some Deep Learning approaches)

Cassie Kozyrkov
Head of Decision Intelligence, Google.

https://www.youtube.com/watch?v=dRJGyhS6gA0

Recommendations

Executives

- Familiarise yourself with the concepts of rules-based software, machine learning and deep learning.
- Explore why AI is important and its many applications
- Identify sources of AI expertise, and existing AI projects, within your organisation.

Entrepreneurs

- To identify opportunities for value creation, explore the many applications for AI
- Familiarise yourself with current developments in AI technology New approaches and novel techniques offer new possibilities.

https://www.ai-playbook.com/introduction

END

Questions?

UMEÅ UNIVERSITY

ACTIVITY

1. Individually or in group
2. Choose an industrial use case (it can be anonymized or hypothetical)
3. Simplify the case to I/ML/O
   1. Output: ? (key: think in business model)
   2. Input: data, human resources, infrastructure, etc.
   2. ML approach?
4. Take some minutes to reflect.
5. Group analysis

PROGRAMMING WITH PYTHON

Programming with Python

The best way to learn how to program is to do something useful, so this introduction to Python is built around a common scientific task: **data analysis**.

Arthritis Inflammation

We are studying **inflammation in patients** who have been given a new treatment for arthritis, and need to analyze the first dozen data sets of their daily inflammation. The data sets are stored in comma-separated values (CSV) format:

- each row holds information for a single patient,
- columns represent successive days.

The first three rows of our first file look like this:

Code

```
0,0,1,3,1,2,4,7,8,3,3,3,10,5,7,4,7,7,12,18,6,13,11,11,7,7,4,6,8,8,4,4,5,7,3,4,2,3,0,0
0,1,2,1,2,1,3,2,2,6,10,11,5,9,4,4,7,16,8,6,18,4,12,5,12,7,11,5,11,3,3,5,4,4,5,5,1,1,0,1
0,1,1,3,3,2,6,2,5,9,5,7,4,5,4,15,5,11,9,10,19,14,12,17,7,12,11,7,4,2,10,5,4,2,2,3,2,2,1,1
```

So, we want to:

1. Calculate the average inflammation per day across all patients.
2. Plot the result to discuss and share with colleagues.

To do all that, we'll have to learn a little bit about programming.

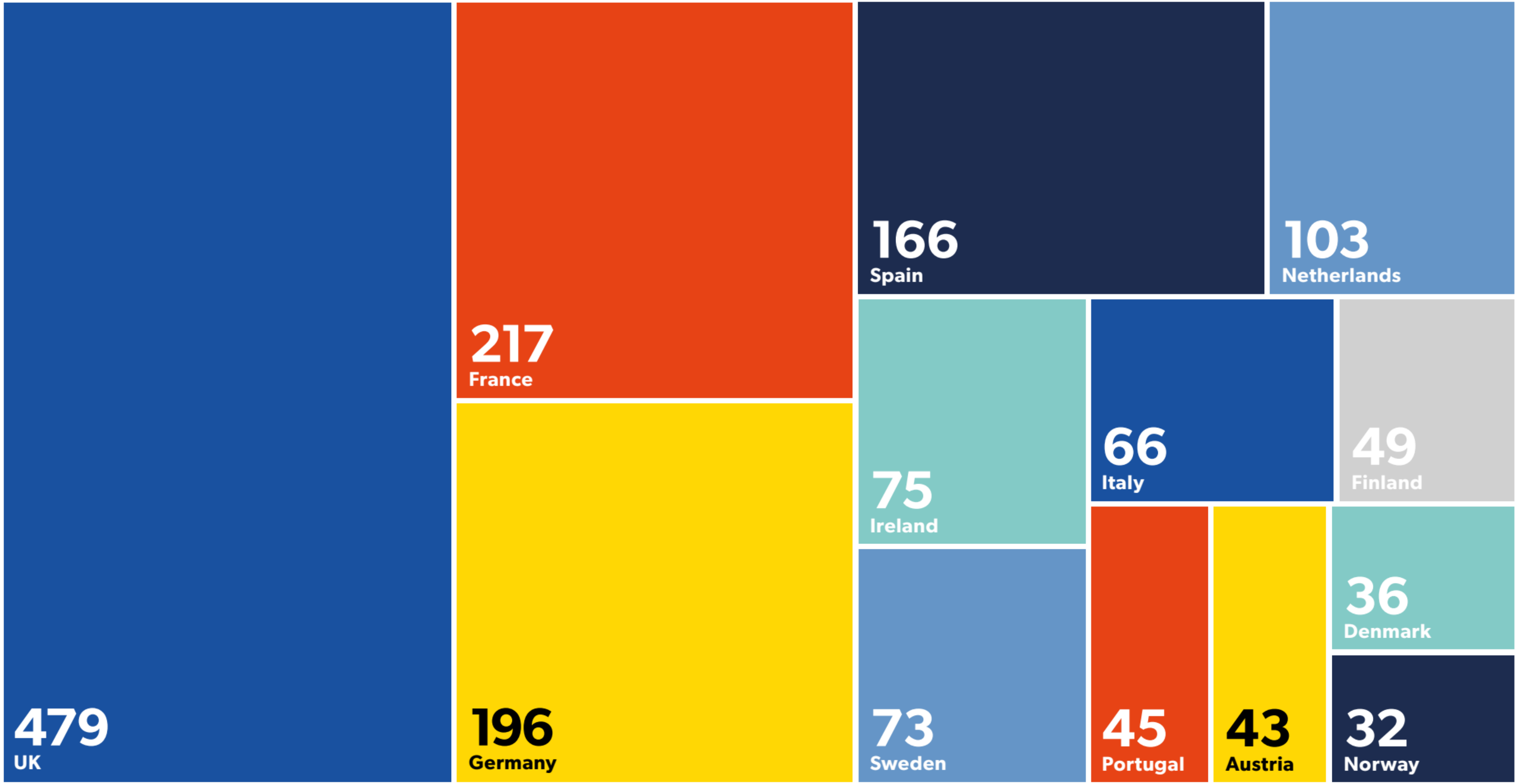
Prerequisites

You need to understand the concepts of **files** and **directories** and how to start a Python interpreter before tackling this lesson. This lesson sometimes references Jupyter Notebook although you can use any Python interpreter mentioned in the Setup.

The commands in this lesson pertain to **Python 3**.

https://www8.cs.umu.se/kursmaterial/ws-python-scientific-computing/

With twice as many AI startups as any other country, the UK is the powerhouse of European AI entrepreneurship



Source: MMC Ventures, Beauhurst, Crunchbase, Tracxn

https://www.mmcventures.com/wp-content/uploads/2019/02/The-State-of-AI-2019-Divergence.pdf

EVENT

Digitising European Indu
Stakeholder Forum 2019

- 13 November 2019 to mid November 2019
- "The focus of the 2019 DEI St Forum will be on the Europea for Digital Innovation Hubs, A Intelligence and industrial pa beyond 2020, including the re funding instruments and oppo

Program of the event

13/11 afternoon:

13:00 Welcome coffee

14:00 Keynote speeches

"We are all feeling the concrete effects of digitalisation" (Ursula von der Leyen) and EU industry needs to be digital. High level speakers from Spanish government, European Commission and EU Finnish presidency will discuss the overall political context and the future actions.

14:45 The future of European industry is digital (panel)

Digitalisation is no longer a question of whether it will take place or not, but a question of how best to accompany businesses in their digital transformation. Speakers will explore how to reinforce the offer of digital innovations across the EU, how to strengthen European industrial ecosystems of IoT, data, cloud, and AI in and across sectors, and how to reinforce the European network of Digital Innovation Hubs to support SMEs in their digital transformation.

16:15 Coffee break

16:45 The rise of Artificial Intelligence: opportunities for industry (panel)

In the industrial domain the potential impact of AI is extremely important, and can help industry to address the long-term challenges of competitiveness and sustainability. Stakeholders will showcase actual use cases in different industry sectors discussing legal considerations and challenges they had to overcome to make AI a cost effective solution to their business.

https://ec.europa.eu/digital-single-market/en/news/digitising-european-industry-stakeholder-forum-2019

Process Industrial IT and Automation

NEW REPORT (OCTOBER)

http://sip-piia.se/wp-content/uploads/2019/10/AI-rapport_2019-low.pdf