



## Welcome to Applied Machine Learning

**Dr Paul Yoo**

Dept CSIS

03/10/19

Birkbeck, University of London

1

© Copyright 2019

## Overview



We will cover:

- Module Overview
- Industry 4.0
- ML Experts
- Predictive Modelling
- The Analytic Workflow
- UCI ML Repository
- Python
- Loading ML Data
  - Pima Indians Data
  - Python, NumPy and Pandas
  - Some statistics

Birkbeck, University of London

7

© Copyright 2019

**ILO**

By the end of this module, you will be able to:

- identify and use Python tools and libraries for machine learning based analytics tasks
- evaluate and identify appropriate machine learning methods and techniques to analyse data
- critically analyse and interpret machine learning results
- use machine learning tools to solve practical problems in real-life scenarios
- demonstrate deep understanding of a range of complex real-life topics in applied machine learning.

**Timetable**

Week	Date	Lecture (G12, Torrington, UCL)	Lab (MAL 414-417)
1	03/10/19	Introduction, Workflow and Loading	Loading data and descriptive statistics
2	10/10/19	Data preparation	Preparing data
3	17/10/19	Feature selection and re-sampling	Selecting features and re-sampling
4	24/10/19	DT and RF	Comparing ML algorithms
5	31/10/19	LR and NN	Automating the process
6	07/11/19	TensorFlow and Keras	MLP with Keras
7	14/11/19	Project Briefing	Project (30%)
8	21/11/19		
9	28/11/19	Image processing	Deep learning - CNN
10	05/12/19	RNN and sequential data	Deep learning - RNN
11	12/12/19	Real-life case	Deep learning - LSTM

Autumn term: 30/09/2019 to 13/12/2019

## Assessment



- Final exam worth 70% of your total mark
- A report (inc. individual section) of a group project worth 30% of your total mark
  - Publication Date: 11/11/19
  - Deadline: 15/12/19
  - Late cut-off deadline: 29/12/19
  - Mark return: 05/01/20
- More details will be provided at the project briefing (W7)

## Discussion



### What is Industry 4.0?

In small groups discuss what you think Industry 4.0 is and their impacts are.

You have **5 minutes** and then we will discuss your answers.

## Wikipedia says...



### Industry 4.0

From Wikipedia, the free encyclopedia

*For other uses, see [4.0](#).*

**Industry 4.0** is a name given to the current trend of [automation](#) and [data exchange](#) in [manufacturing](#) technologies. It includes [cyber-physical systems](#), the [Internet of things](#), [cloud computing](#)<sup>[1][2][3]</sup><sup>[4]</sup> and [cognitive computing](#). Industry 4.0 is commonly referred to as the [fourth industrial revolution](#).<sup>[5]</sup>

Industry 4.0 fosters what has been called a "smart factory". Within modular structured smart factories, cyber-physical systems monitor physical processes, create a virtual copy of the physical world and make decentralized decisions. Over the Internet of Things, cyber-physical systems communicate and cooperate with each other and with humans in real-time both internally and across organizational services offered and used by participants of the value chain.<sup>[1]</sup>

Birkbeck, University of London

12

© Copyright 2019

## Ford's conveyer belt (Industry 2.0)



Birkbeck, University of London

13

© Copyright 2019

## Tesla's Fremont factory (Industry 4.0)



Birkbeck, University of London

14

© Copyright 2019

## Traditional shoes factory (Nike)

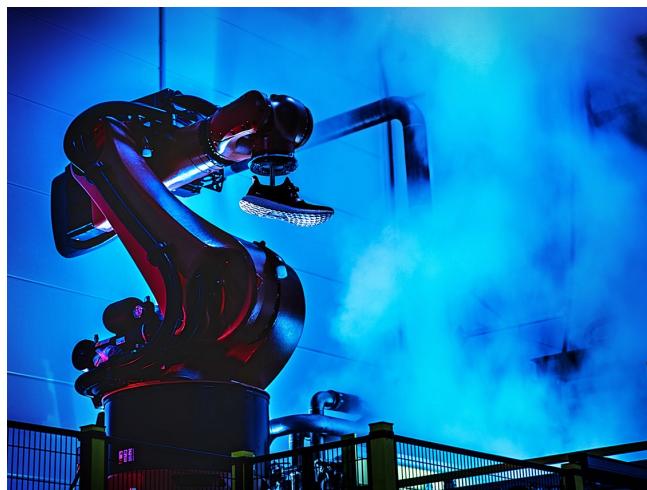


Birkbeck, University of London

15

© Copyright 2019

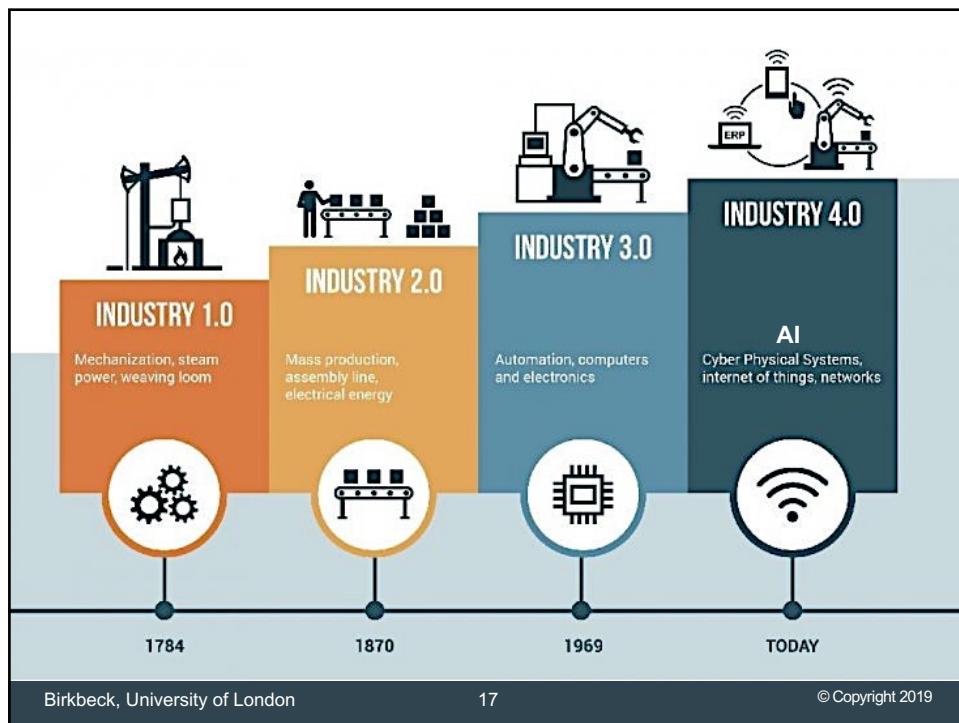
## Adidas' speed factory (Industry 4.0)



Birkbeck, University of London

16

© Copyright 2019



Birkbeck, University of London

17

© Copyright 2019

 **NATO REVIEW** magazine

NATO Review Newsletter   
Stay informed! Join our free newsletter

Home Themes Videos Previously in NATO Review Subscribe free About NATO Review Contact us

[NATO Review / Autonomous military drones: no longer science fiction](#)

## Autonomous military drones: no longer science fiction

 28/07/2017

The 'PERDIX' is a micro-drone swarm system developed for the US DoD/Naval Air Systems Command together with MIT Lincoln Laboratory.

Birkbeck, University of London 18 © Copyright 2019

**NEWS**

Home | UK | World | Business | Politics | Tech | Science | Health | Family & Education | Entertainment & Arts | Stories | Video & Audio | In Pictures | Newsbeat | More ▾

World | Africa | Asia | Australia | Europe | Latin America | Middle East | US & Canada

**NOW PLAYING**  0:44 Islamic State claims drone attack  
① 25 Feb 17 | Middle East

**UP NEXT**  1:23 Mosul: roadside bombs and gunfights  
① 24 Feb 17 | UK

 1:26 Iraqi forces 'within sight of Mosul'  
① 20 Feb 17 | Middle East

 0:06 Iraqi forces go into battle to music  
① 19 Feb 17 | Middle East

 So-called Islamic State 'carried out drone attacks'  
So-called Islamic State releases footage apparently showing a drone attack they carried out on Mosul, Iraq.  
① 25 Feb 2017 | Middle East

f t o e Share

Birkbeck, University of London 19 © Copyright 2019



## The New AI Toilets Will Scan Your Poop To Diagnose Your Ailments

Navin Bondade 4 Comments Artificial Intelligence, Data Science, Deep Learning, Machine Learning

**41.2k Shares**

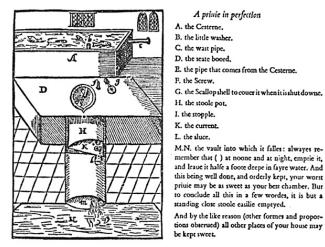
Well, it's sound weird but in the upcoming future, your toilet will be your mini doctor. A company called Micron is developing a smart artificial intelligence-powered toilet that will reportedly be able to diagnose your state of health and risk of disease by analyzing your bodily waste.

The major goal of this technology is to analyze bowel movements and recognize the signs for health issues or ailments early on. This information will also use to understand early symptoms of certain diseases.

**Science Focus**  
THE HOME OF BBC SCIENCE FOCUS MAGAZINE

FUTURE TECHNOLOGY NATURE SPACE THE HUMAN BODY EVERYDAY SCIENCE PLANET EARTH Q

Home > Everyday science > Who really invented the flushing toilet?

**Who really invented the flushing toilet?**



**Sir John Harington** in 1596. Known as a water closet, it was installed in Richmond Palace. Yet despite this royal support, the device was long rejected by the public, who saw it as an expensive indulgence.

**How Computers are Learning to be Creative by Blaise Agüera y Arcas**

URL: [https://youtu.be/uSUOdu\\_5MPc](https://youtu.be/uSUOdu_5MPc)

5:45 – 17:34

Birkbeck, University of London

22

© Copyright 2019

**IEEE SPECTRUM** Engineering Topics ▾ Special Reports ▾ Blogs ▾ Multimedia ▾ The Magazine ▾ Professional Resources ▾

Search ▾ Tech Talk | Computing | Hardware

27 Sep 2019 | 17:35 GMT

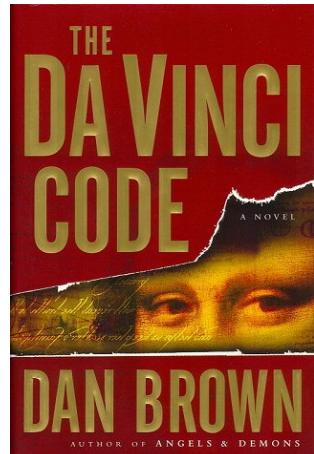
## What Google's Quantum Supremacy Claim Means for Quantum Computing

Leaked details about Google's quantum supremacy experiment stirred up a media frenzy about the next quantum computing milestone

By Jeremy Hsu

The Sycamore quantum computing device sampled the random quantum circuit one million times in just three minutes and 20 seconds. When the team simulated the same quantum circuit on classical computers, it found that even the [Summit supercomputer](#) that is currently ranked as the most powerful in the world would require approximately 10,000 years to perform the same task.

Photo: Google/NASA  
Google used a 54-qubit processor called Sycamore to carry out its quantum supremacy experiment. Because one qubit did not function properly, the experiment involved 53 qubits.



## God to be replaced by AI: New religion to be created by computers, Dan Brown claims

RELIGION as we know it will be a thing of the past when artificial intelligence creates a new form of collective consciousness, US author Dan Brown believes.

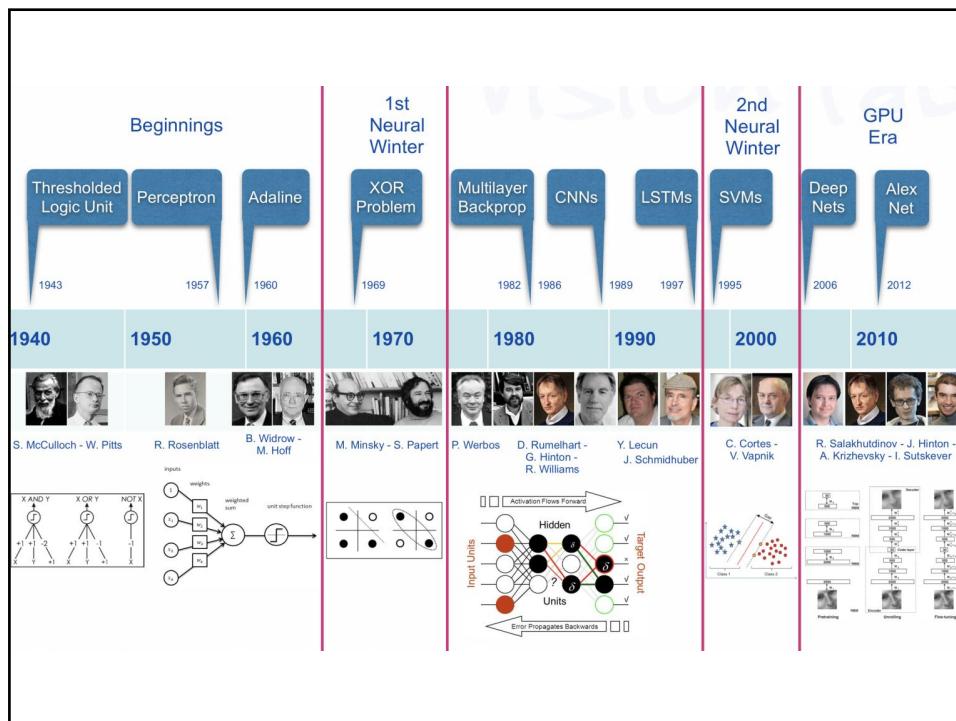
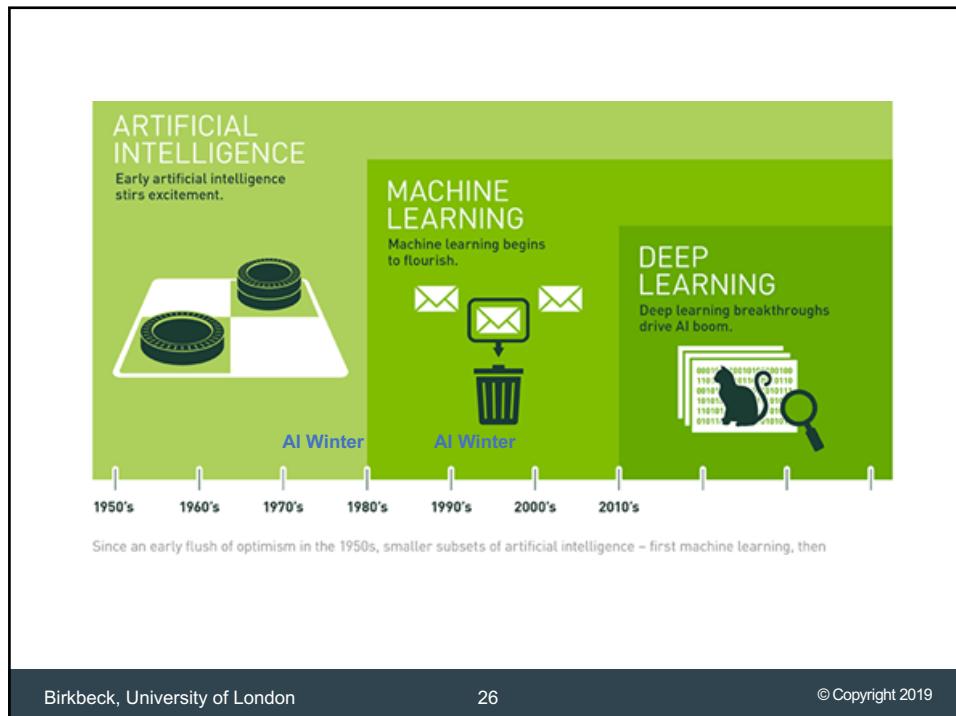
Brown made the provocative remark during a talk to promote his latest novel 'Origin', the fifth in a series following Harvard professor Robert Langdon, the protagonist of 'The Da Vinci Code', which questioned the history of Christianity.

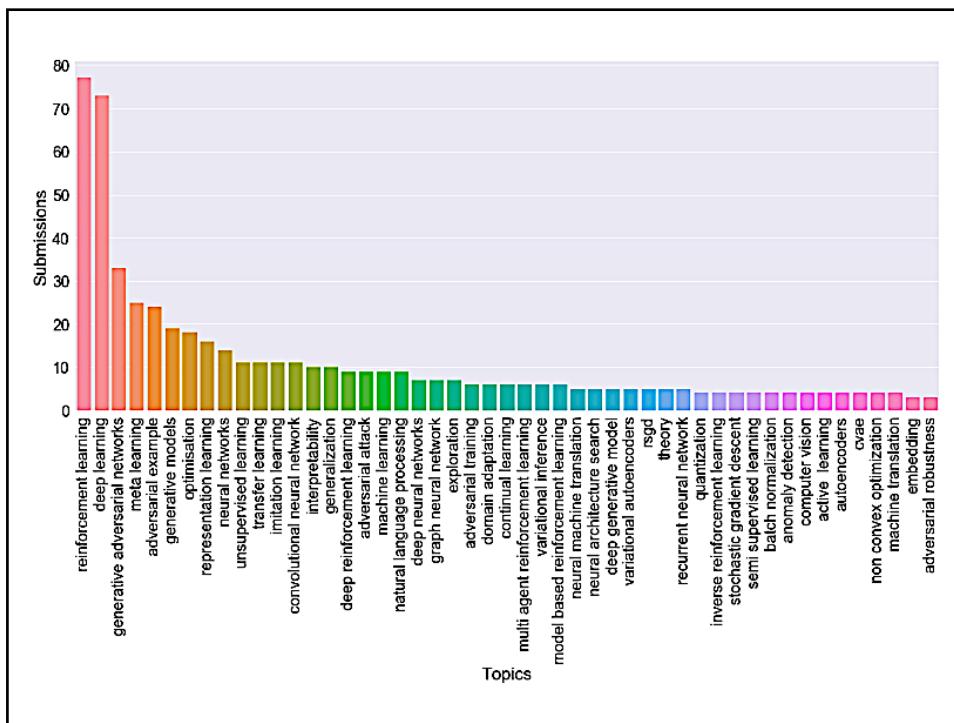
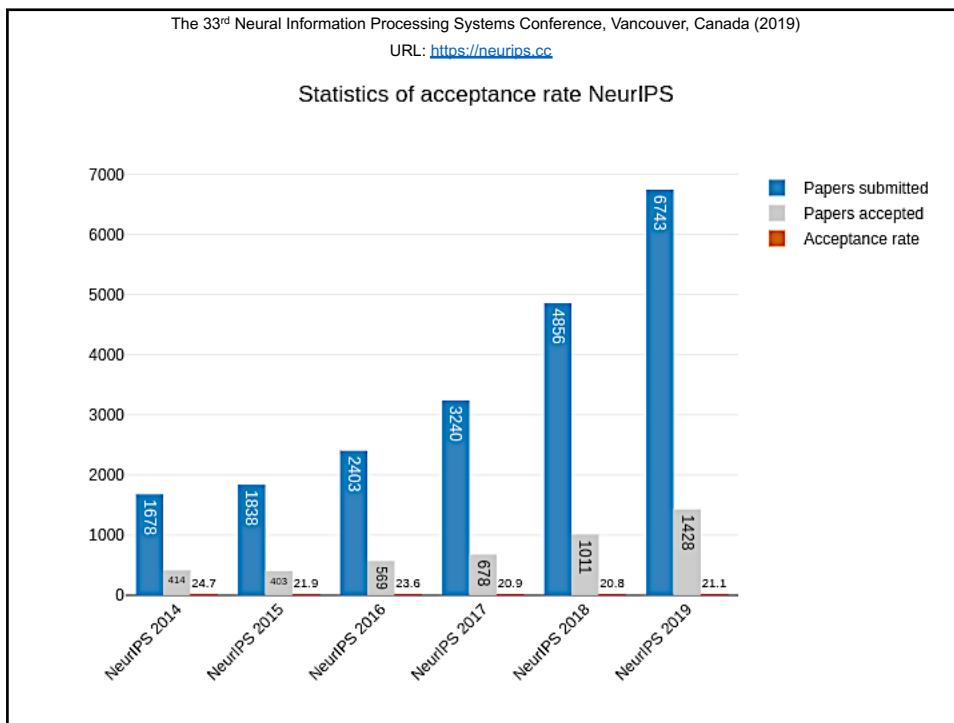
'Origin' was inspired by the question "Will God survive science?", said Brown, adding this had never happened in the history of humanity.

## Machine Learning Experts You Need to Know



- Geoffrey Hinton – backpropagation (1980s), Boltzmann machines and CapsNet (URL: <https://youtu.be/uAu3jQWaN6E>)
- Michael I Jordan – RNN (1980s)
- Yann LeCun – CNN with backpropagation
- Yoshua Bengio – RNN (1996)
- Jürgen Schmidhuber – LSTM (1992)
- Andrew Ng – Coursera, deeplearning.ai, Google Brain project, Landing AI (SaaS)
- Vladimir Vapnik – SVM (1963)
- Ian Goodfellow – GANs (2014)
- Blaise Agüera y Arcas – Google TPU 3 teraops ( $10^{12}$  per sec) with 1 watt





# Google DeepMind



**AlphaZero taught itself to play three different games**



AlphaGo Vs Lee Sedol

Game	AlphaZero vs. Stockfish	AlphaZero vs. Elmo	AlphaZero vs. AGO
Chess	W: 29.0% D: 70.6% L: 0.4%	W: 84.2% D: 2.2% L: 13.6%	W: 68.9% D: 31.1% L: 0.0%
Shogi	W: 98.2% D: 0.6% L: 1.2%	W: 98.2% D: 0.6% L: 1.2%	W: 53.7% D: 46.3% L: 0.0%
Go	W: 2.6% D: 97.2% L: 0.2%	W: 1.8% D: 98.2% L: 0.0%	W: 1.8% D: 98.2% L: 0.0%

AZ wins (green) AZ draws (grey) AZ loses (red) AZ white (circle) AZ black (cross)

AlphaZero

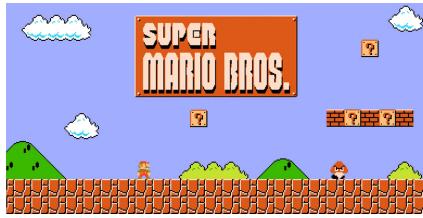
[Source] <https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go/>

Birkbeck, University of London      30      © Copyright 2019

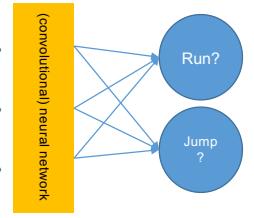
# Google DeepMind continues



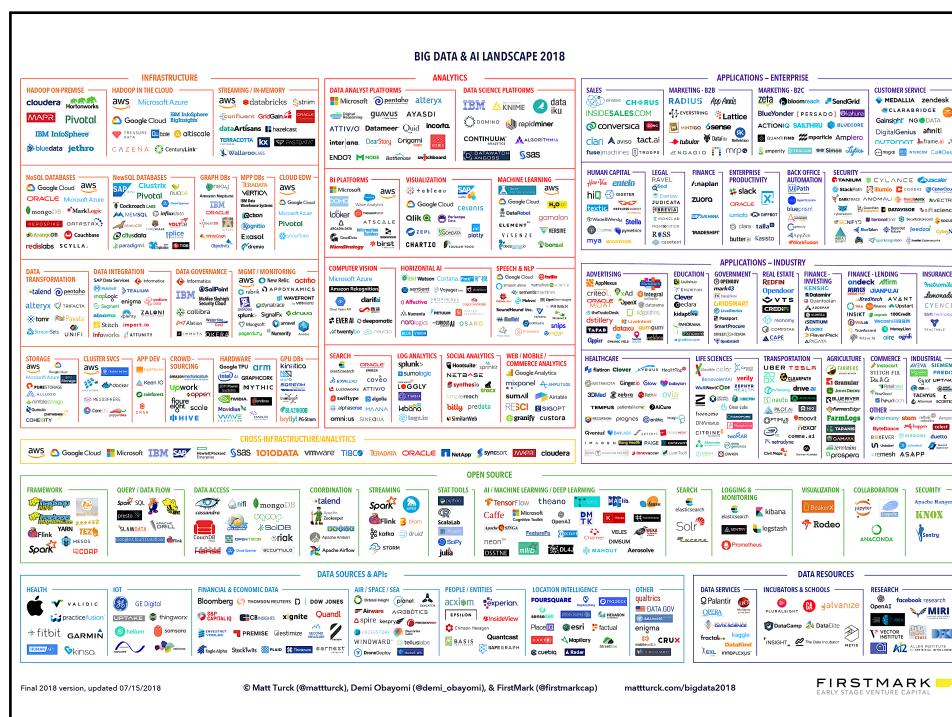
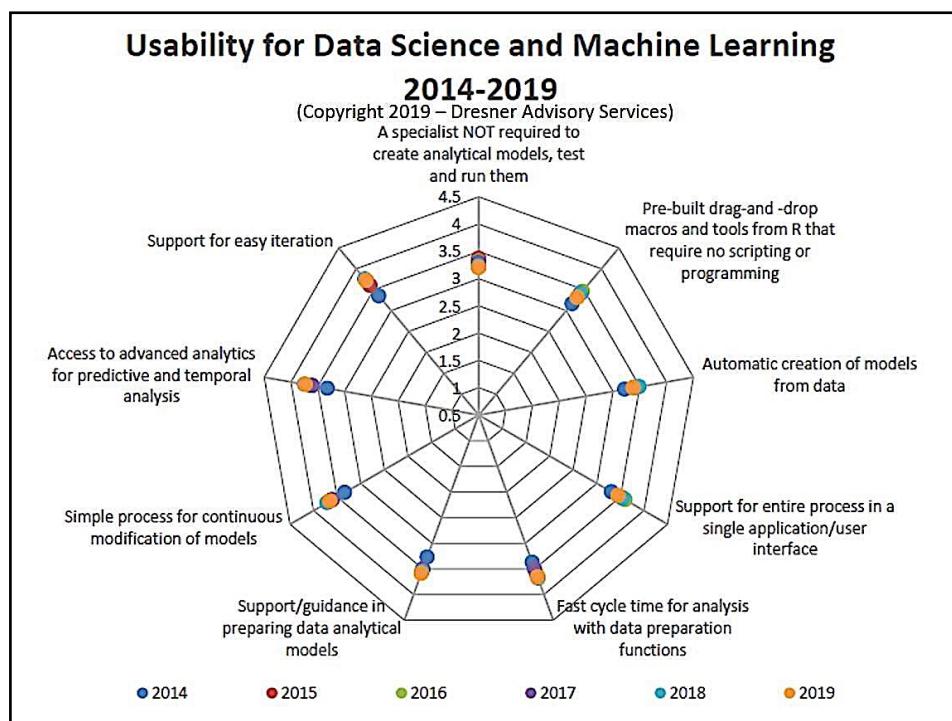
**"Reinforcement learning is concerned with the problem of finding suitable actions to take in a given situation in order to maximize a reward."**

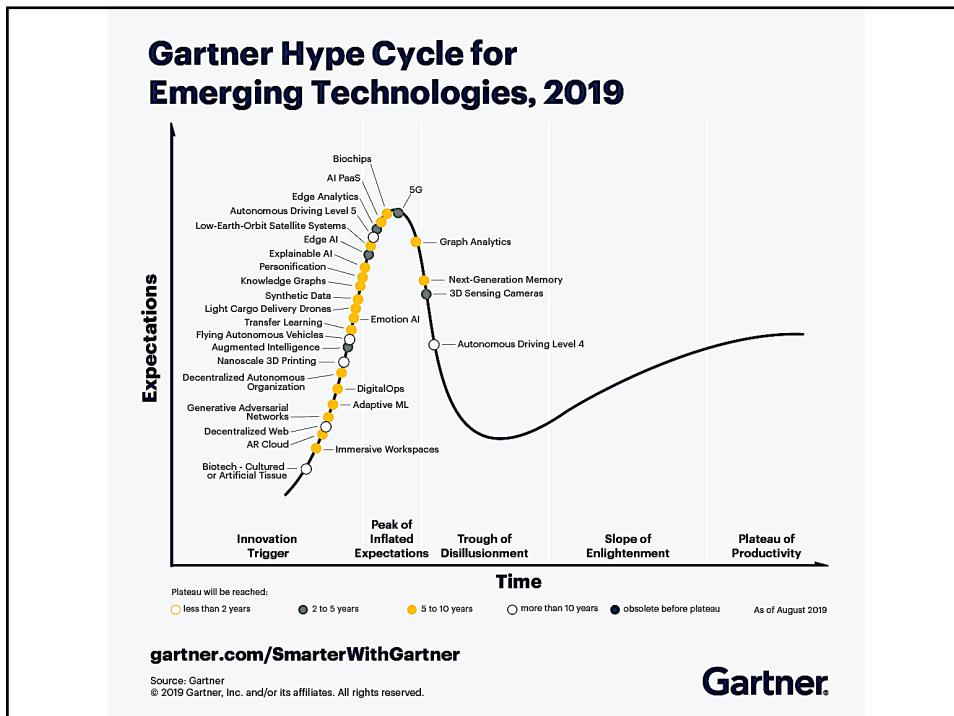


Input image

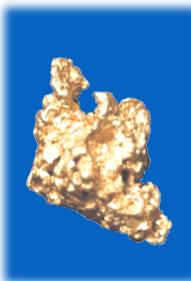

  
 Possible Actions

Birkbeck, University of London      31      © Copyright 2019





## Predictive Modelling




The Essence of Data Mining

***“Most of the big payoff [in data mining] has been in predictive modeling.”***

– Herb Edelstein

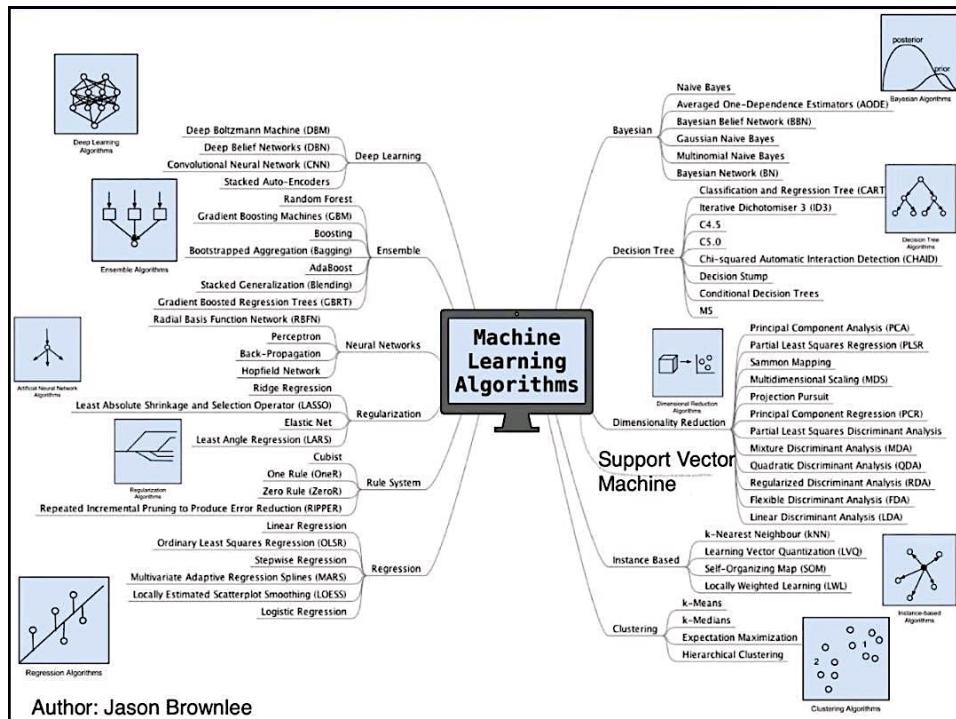
This module focuses on a specific sub-field of machine learning called predictive modeling.

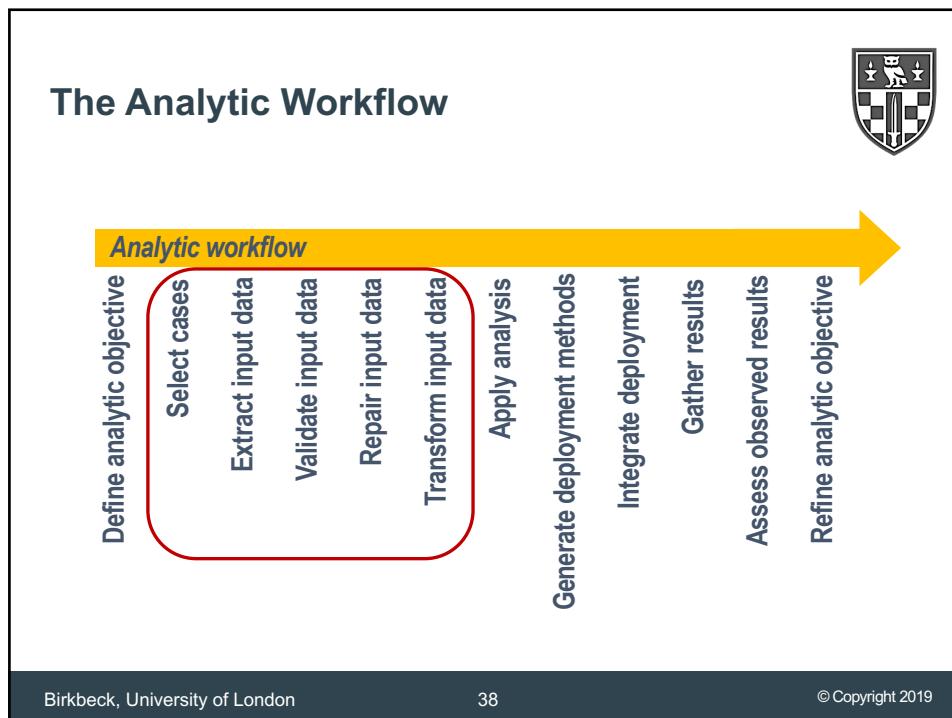
Birkbeck, University of London      35      © Copyright 2019

## Predictive Modelling ML Steps



1. **Define Problem:** Investigate and characterise the problem in order to better understand the goals of the project.
2. **Analyse Data:** Use descriptive statistics and visualisation to better understand the data you have available.
3. **Prepare Data:** Use data transforms in order to better expose the structure of the prediction problem to modeling algorithms.
4. **Evaluate Algorithms:** Design a test harness to evaluate a number of standard algorithms on the data and select the top few to investigate further.
5. **Improve Results:** Use algorithm tuning and ensemble methods to get the most out of well-performing algorithms on your data.
6. **Present Results:** Finalise the model, make predictions and present results.





**UCI Machine Learning repository**



<http://archive.ics.uci.edu/ml/index.php>

- Small – fit into memory and model them in reasonable time
- Well behaved – don't need to do a lot of feature engineering
- Benchmarks – many people have used them

Birkbeck, University of London      40      © Copyright 2019

**Python**

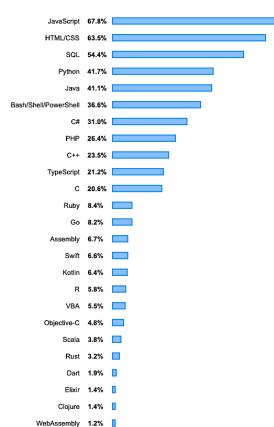


**It is consistently appearing in the top 10 programming languages in surveys on StackOverflow.**

Overview

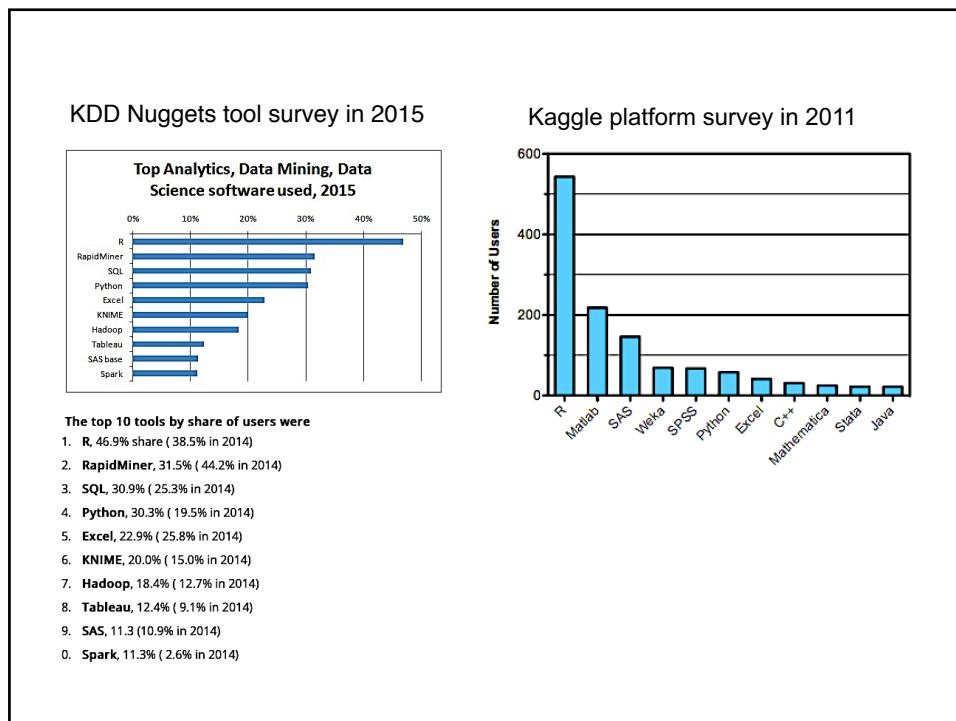
This year, nearly 90,000 developers told us how they learn and level up, which tools they're using, and what they want.

URL: <https://insights.stackoverflow.com/survey/2019>



87,354 responses; select all that apply

Birkbeck, University of London      41      © Copyright 2019



## SciPy



SciPy is a free and open-source Python library used for scientific computing and technical computing.

- It is an add-on to Python that you will need for machine learning.
- It contains modules for optimisation, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers and other tasks common in science and engineering.
- It is comprised of the following core modules relevant to machine learning:
  - NumPy: A foundation for SciPy that allows you to efficiently work with data in arrays.
  - Matplotlib: Allows you to create 2D charts and plots from data.
  - Pandas: Tools and data structures to organise and analyse your data. (to load explore and better understand your data)

Birkbeck, University of London      43      © Copyright 2019

## SciPy ecosystem

```

graph TD
    SciPyEcosystem[SciPy ecosystem] --- SciPyLibrary[SciPy Library  
Fundamental library for scientific computing]
    SciPyEcosystem --- IPython[IP[y]: IPython  
Enhanced interactive console]
    SciPyEcosystem --- SymPy[SymPy  
Symbolic mathematics]
    SciPyEcosystem --- NumPy[NumPy  
Based N-dimensional array package]
    SciPyEcosystem --- Pandas[Pandas  
Data structures & analysis]
    SciPyEcosystem --- Matplotlib[Matplotlib  
Comprehensive 2D Plotting]
  
```

The diagram illustrates the SciPy ecosystem, centered around the SciPy ecosystem logo. It branches out to show various components:

- SciPy Library**: Fundamental library for scientific computing.
- IP[y]: IPython**: Enhanced interactive console.
- SymPy**: Symbolic mathematics.
- NumPy**: Based N-dimensional array package.
- Pandas**: Data structures & analysis.
- Matplotlib**: Comprehensive 2D Plotting.

Birkbeck, University of London      44      © Copyright 2019

## scikit-learn

The scikit-learn library is how you can develop and practice ML in Python.

- scikit = SciPy +toolkit
- It is built upon and requires the SciPy.
- ML algorithms for classification, regression, clustering and etc.
- Tools for evaluating models, tuning parameters and pre-processing data.

Birkbeck, University of London      45      © Copyright 2019

## Python Installation



### Python 3.7.2

- Python Beginners Guide  
<https://wiki.python.org/moin/BeginnersGuide/Download>
- python --version
- pip - Python package management tool
- *pip install jupyter scipy numpy matplotlib pandas sklearn tensorflow theano keras seaborn subprocess.run graphviz pydot*
- *Anaconda 2019.03 for Windows Installer (Python 3.7 version)*

## Some Python codes



```
# define an array
import numpy
mylist = [1, 2, 3]
myarray = numpy.array(mylist)
print(myarray)                                [[1 2 3]
                                             (3, )
print(myarray.shape)

# access values
import numpy
mylist = [[1, 2, 3], [3, 4, 5]]
myarray = numpy.array(mylist)
print(myarray)                                [[[1 2 3]
                                             [3 4 5]]
                                             (2, 3)
print(myarray.shape)                          First row: [1 2 3]
                                             Last row: [3 4 5]
                                             Specific row and col: 3
                                             Whole col: [3 5]
print("First row: %s" % myarray[0])
print("Last row: %s" % myarray[-1])
print("Specific row and col: %s" % myarray[0, 2])
print("Whole col: %s" % myarray[:, 2])
```

```
# arithmetic
import numpy
myarray1 = numpy.array([2, 2, 2])
myarray2 = numpy.array([3, 3, 3])
print("Addition: %s" % (myarray1 + myarray2))
print("Multiplication: %s" % (myarray1 * myarray2))

# basic line plot
import matplotlib.pyplot as plt
import numpy
myarray = numpy.array([1, 2, 3])
plt.plot(myarray)
plt.xlabel('some x axis')
plt.ylabel('some y axis')
plt.show()

# basic scatter plot
import matplotlib.pyplot as plt
import numpy
x = numpy.array([1, 2, 3])
y = numpy.array([2, 4, 6])
plt.scatter(x,y)
plt.xlabel('some x axis')
plt.ylabel('some y axis')
plt.show()
```

Addition: [5 5 5]  
Multiplication: [6 6 6]

Birkbeck, University of London

48

© Copyright 2019

```
# series
import numpy
import pandas
myarray = numpy.array([1, 2, 3])
rownames = ['a', 'b', 'c']
myseries = pandas.Series(myarray, index=rownames)
print(myseries)

print(myseries[0])                                1
print(myseries['a'])                             1

# dataframe
import numpy
import pandas
myarray = numpy.array([[1, 2, 3], [4, 5, 6]])
rownames = ['a', 'b']
colnames = ['one', 'two', 'three']
mydataframe = pandas.DataFrame(myarray, index=rownames, columns=colnames)
print(mydataframe)

print("method 1:")
print("one column:\n%s" % mydataframe['one'])
print("method 2:")
print("one column:\n%s" % mydataframe.one)
```

	one	two	three
a	1	2	3
b	4	5	6

method 1:  
one column:  
a 1  
b 4  
method 2:  
one column:  
a 1  
b 4

Birkbeck, University of London

49

© Copyright 2019

## Summary



This lecture we've covered the basics of AML including:

- Module Overview
- Industry 4.0
- ML Experts
- Predictive Modelling
- The Analytic Workflow
- UCI ML Repository
- Python, NumPy and Pandas

Next week

- Data Preparation

Labs

- MAL 414–417

## Questions?

[paul@dcs.bbk.ac.uk](mailto:paul@dcs.bbk.ac.uk)