



CT5165 Principles of Machine Learning

Dr Ihsan Ullah
School of Computer Science
University of Galway



About Me

Dr Ihsan Ullah

Lecturer Above the Bar
School of Computer Science

Ihsan.ullah@universityofgalway.ie

Research Area:

- Designing light weight neural networks
- Computer Vision
 - recognition of action, object, person or in health domain (e.g. Cardiovascular)

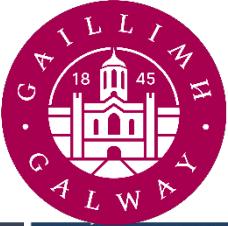




Module Learning Objectives

On successful completion of this module, you will be able to:

1. Define Machine Learning and explain what major categories of learning task entail
2. Demonstrate how to apply the machine learning and data mining process to practical problems
3. Explain and apply algorithms including decision tree learning, instance-based learning, probabilistic learning, linear regression, logistic regression, Q-learning, and others
4. Given a dataset and task to be addressed, select, apply and evaluate appropriate algorithms, and interpret the results
5. Discuss ethical issues and emerging trends in machine learning.



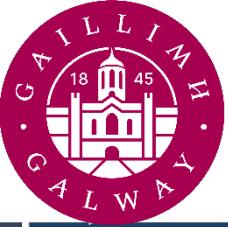
Overview of Topics in the Module

- Week 1: **Introduction to Machine Learning** (Definition, Major categories, etc.)
- Week 2 & 3: **Information-based Learning** (principles of supervised learning)
- Week 4 & 5: **Similarity-based Learning** (e.g. k-nearest neighbours)
- Week 6: **Reinforcement Learning** (introduction, model free)
- Week 7: **Evaluating Classifier Performance** (Practical Advice; Some ML Tools)
- Week 8 & 9: **Linear Regression in One and Multiple Variables**
- Week 10&11: **Linear Classifiers with Hard and Soft Thresholds**
- Week 12: **Probabilistic Machine Learning** (Naïve bayes and Bayesian)



Module Assessment

- Final exam (70%)
 - Related past exam papers are available from the ISS exam papers service
- Continuous assessment (30%)
 - 3 assignments
 - Details and due dates will be posted on Blackboard



Learning objectives for today's lecture

Having completed todays lecture successfully, you will be able to ...

1. Discuss definitions of Machine Learning
2. Describe what major categories of ML task entail: classification, regression, clustering, relationship discovery and reinforcement learning
3. Discuss the relationship with Data Mining
4. Explain the Data Mining process
5. Consider current and future applications of Machine Learning and Data Mining



Prerequisites

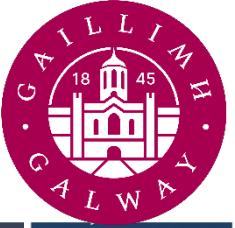
- This is for students who already have a degree or substantial experience in computer science, software development or a closely related subject area
- You need to understand:
 - How to program (any language)
 - Algorithm analysis
 - Basic statistics and probability
 - Knowledge of standard mathematical notation (i.e. how to read an equation)



Resources

- Course slides:
 - Necessary but insufficient!
- Recommended books:
 - List available on Blackboard
 - Will also provide references in individual sections
- Others:
 - Andrew Ng's Coursera Machine Learning Course
 - Sebastian Thrun's Udacity AI Course
 - Contributions welcome!

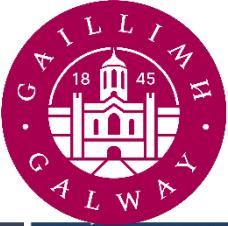
If you find useful links, email them either to me (ihsan.ullah@nuigalway.ie).



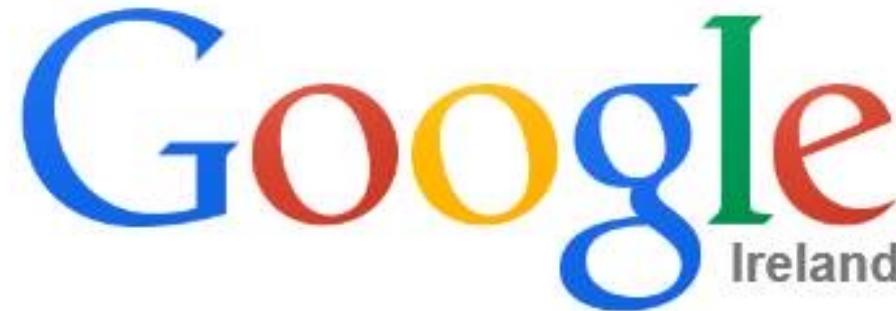
What is Machine Learning? [3]

Vevox.app

The image shows a screenshot of the Vevox app interface. On the left side, there is a sidebar with the Vevox logo, a QR code, and text encouraging users to join at vevox.app with ID 165-451-844. The main content area displays a presentation slide with the title "Machine Learning is" and a progress bar indicating 1/1 slide. The slide has a small circular icon with a play button and the number 0.



What is Machine Learning? [3]



machine learning is |

machine learning is **bullshit**

machine learning is **hard**

machine learning is **fun**

machine learning is **the future**

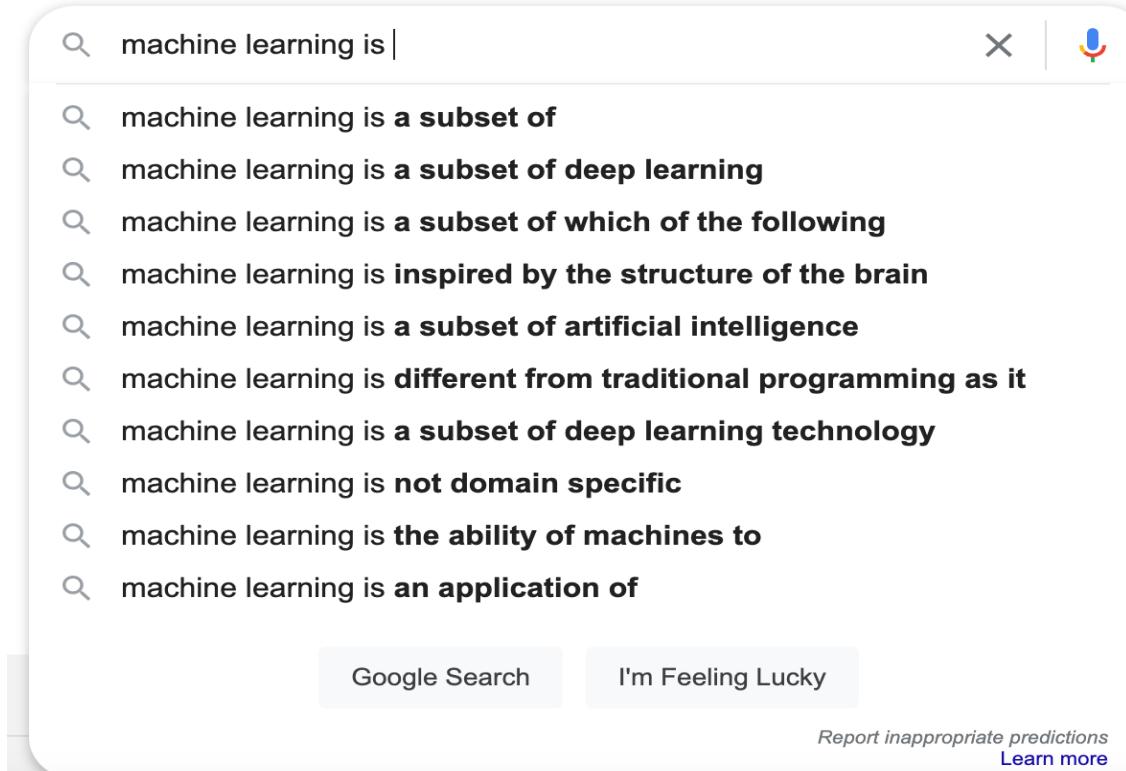
machine learning is **not as cool as it sounds**

machine learning is

machine learning is **math**



Google



A screenshot of a Google search interface. The search bar contains the query "machine learning is |". Below the search bar is a list of ten suggested search terms, each preceded by a magnifying glass icon. At the bottom of the interface are two buttons: "Google Search" and "I'm Feeling Lucky". A small link at the bottom right says "Report inappropriate predictions Learn more".

- machine learning is a subset of
- machine learning is a subset of deep learning
- machine learning is a subset of which of the following
- machine learning is inspired by the structure of the brain
- machine learning is a subset of artificial intelligence
- machine learning is different from traditional programming as it
- machine learning is a subset of deep learning technology
- machine learning is not domain specific
- machine learning is the ability of machines to
- machine learning is an application of

Google Search I'm Feeling Lucky

Report inappropriate predictions [Learn more](#)



What is Machine Learning? [1]

- Samuel, 1959:
 - "Field of study that gives computers the ability to learn without being explicitly programmed"
- Witten & Frank, 1999:
 - Learning is changing behaviour in a way that makes **performance** better in the future

Arthur Samuel, 1901-1990

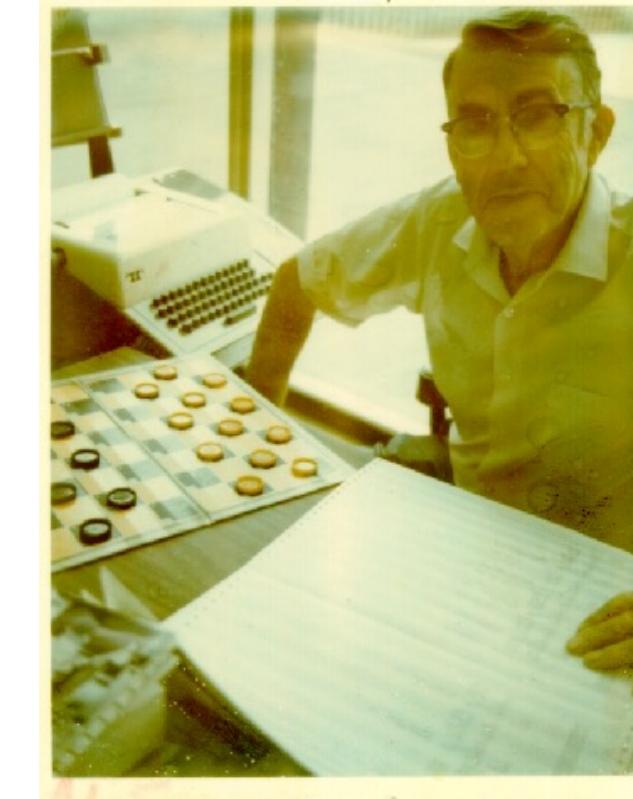


Image source: <http://www.computer.org/portal/web/awards/cp-samuel>

What is Machine Learning? [2]

- Mitchell, 1997:

- Improvement with experience at some task
- A well-defined ML problem:
 - Improve over task T
 - wrt **performance** measure P
 - based on experience E
- For draughts/checkers example, what are T , P , E ?



- Other possible definitions

- Philosophical and psychological considerations ...
- Relationship to Artificial Intelligence generally ...
- Artificial Intelligence \neq Machine Learning \neq Deep Learning
- **Artificial Intelligence \supsetneq Machine Learning \supsetneq Deep Learning**



Machine Learning Categories

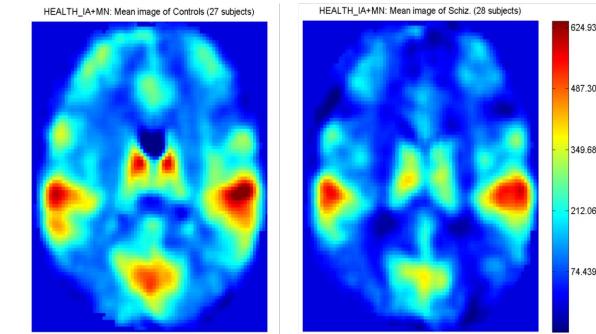
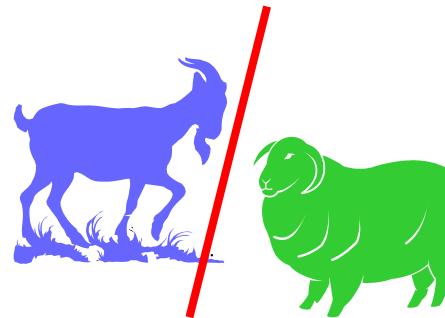
- Supervised learning
 - Vector with independent and dependent variable
- Unsupervised learning
 - Vector with independent variable
- Semi-supervised learning
 - Small subset with dependent variable
- Reinforcement learning
 - Feedback in form of reward

Others:
Self- Supervised Learning,
Transfer learning,
etc.



Major Types of Task [1]

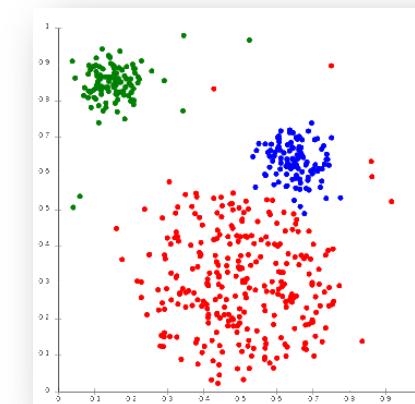
1. Classification



2. Regression



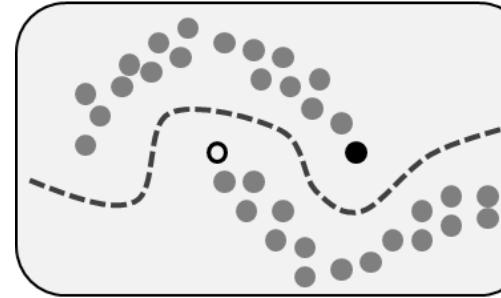
3. Clustering





Major Types of Task [2]

4. Co-Training



5. Relationship Discovery

beer \leftrightarrow diapers

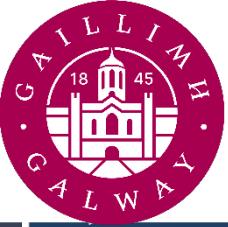
6. Reinforcement Learning





Techniques for these Tasks

- Classification
Decision trees, SVMs
 - Regression
Linear Regression, Neural nets; k-NN
(good for Classification too)
 - Clustering
k-Means, EM-clustering
 - Relationship Discovery
Association Rules; Bayesian nets
 - Learning From Part-Labelled Data
Co-Training; Transductive Learning
[Combines ideas from clustering & classification]
 - Reinforcement Learning
Q-Learning, SARSA
-
- The diagram illustrates the classification of machine learning tasks into four main categories, each represented by a red curly brace:
- Supervised**: Braces the first two items of the list.
 - Unsupervised**: Braces the third and fourth items of the list.
 - Semi-supervised**: Braces the fifth item of the list.
 - Reward-based**: Braces the sixth item of the list.



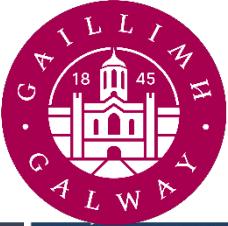
What do these have in common?

- In all cases, machine searches for a **hypothesis** that best describes the data presented to it
- Choices to be made:
 - How is hypothesis expressed?
mathematical equation, logic rules, diagrammatic form, table, parameters of a model (e.g. weights of an ANN), ...
 - How is search carried out?
systematic (breadth-first or depth-first), heuristic (most promising first), ...
 - How do we measure quality of hypothesis?
 - K-fold cross validation
 - What is appropriate format for data?
 - How much data is required?
 - Problems e.g. Overfitting, underfitting.



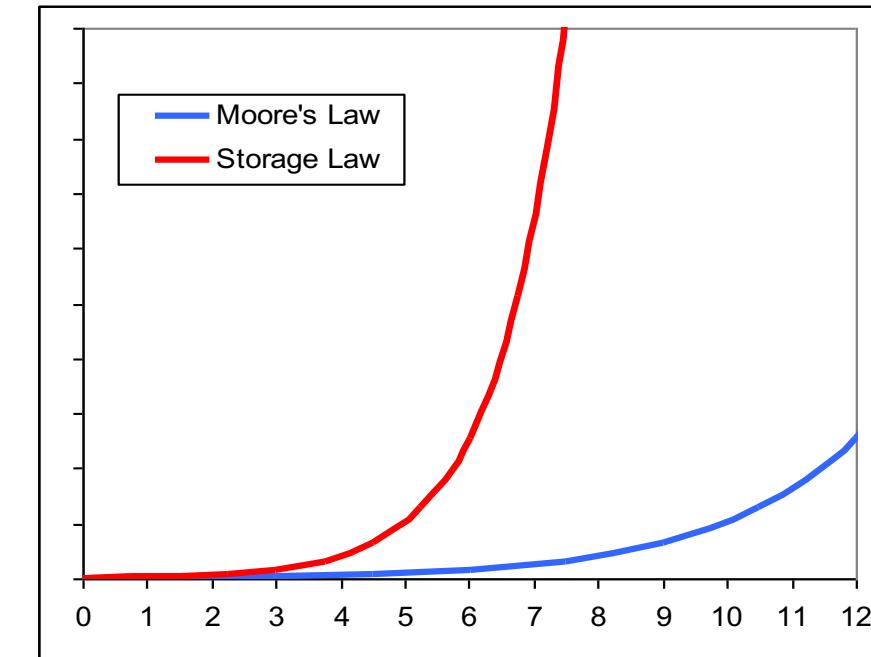
What else do we need to know about?

- To apply ML:
 - How to formulate a problem
 - How to prepare the data
 - How to select an appropriate algorithm
 - How to interpret the results
- To evaluate results and compare methods:
 - Separation between training, testing & validation
 - Performance measures:
simple metrics, statistical tests, graphical methods
 - To improve performance
 - Ensemble methods
 - Theoretical bounds on performance



Data Mining: What's the Link?

- Data Mining:
 - Extract **interesting** knowledge from **large unstructured** datasets
 - **non-obvious / comprehensible / meaningful / useful**
- Storage Law (Fayyad & Uthurusamy, Comms.ACM 2002)
 - Storage capacity **doubling** every year
 - Faster than Moore's law
 - Result: write-only “data tombs”
- Developments in ML essential to be able to process and exploit this lost data

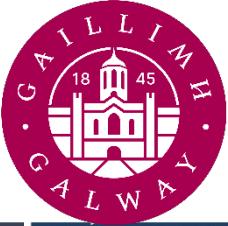


Big Data

Data sets of scale and complexity such that they can be difficult to process using current standard methods

- Standard DB tools & data management apps
- Moving target





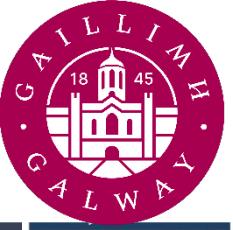
Data generated in 60 seconds

2019 *This Is What Happens In An Internet Minute*



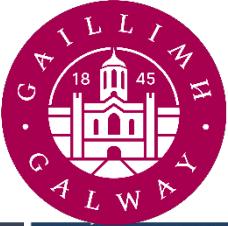
2021 *This Is What Happens In An Internet Minute*





Big Data

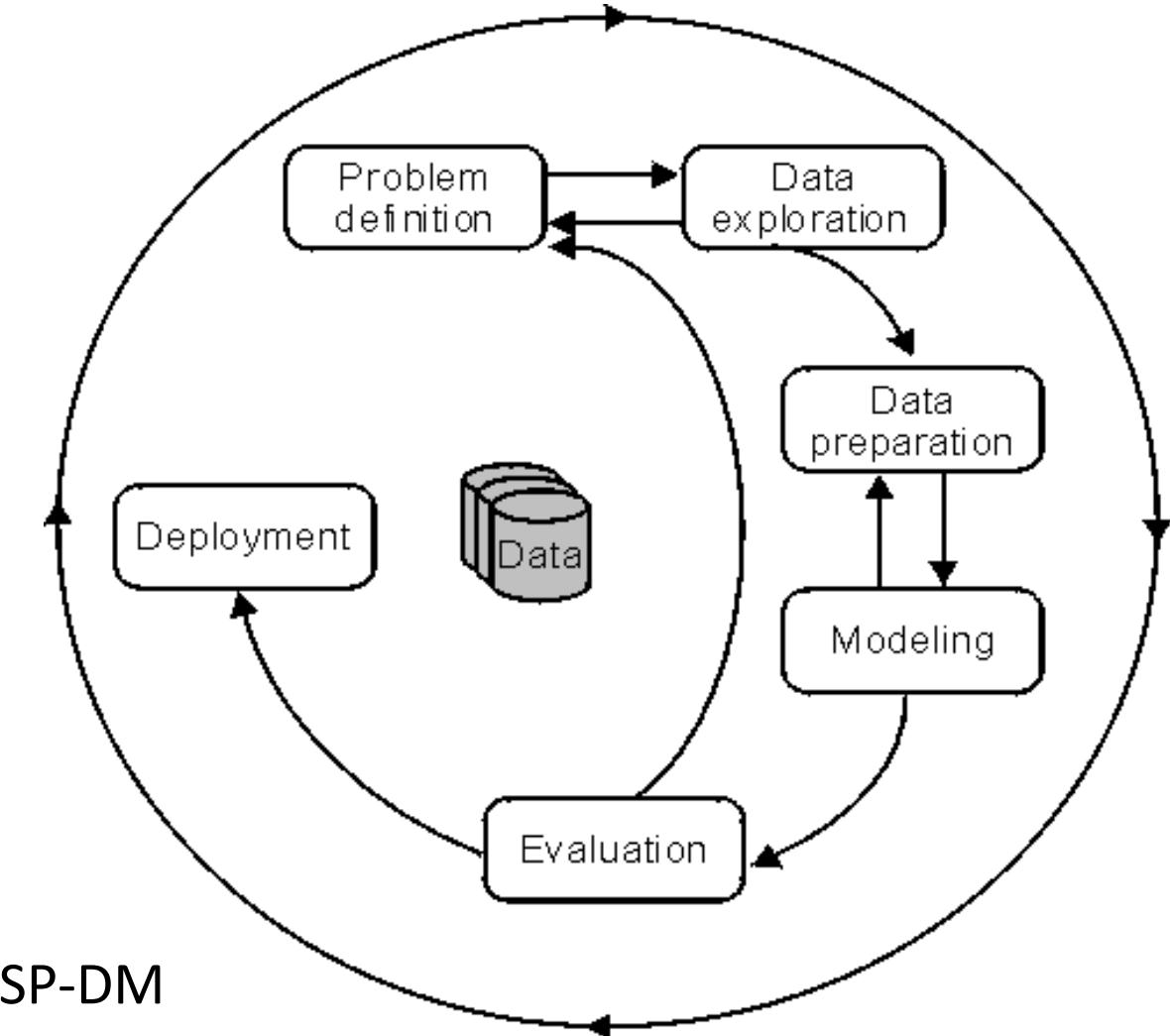
- Data scale dimensions (One or more of “3 Vs”):
 - **Volume**: terabytes and up
 - **Velocity**: from batch to streaming data
 - **Variety**: numeric, video, sensor, unstructured text ...
- Fashionable to add others that are not key ...
 - Veracity: quality & uncertainty associated with items
 - Variability: change / inconsistency over time
 - Value: for the organisation
- Key techniques:
 - Sampling; inductive learning; clustering; associations
 - Distributed programming methods



CRISP-DM Data Mining Process Model

Cross Industry Standard Process for Data Mining (CRISP-DM) process model

- Problem Definition
 - E.g. Customer retention problem
- Data Exploration
- Data Preparation
- Modelling
- Evaluation
- Deployment



This link gives a summary of the main steps in CRISP-DM

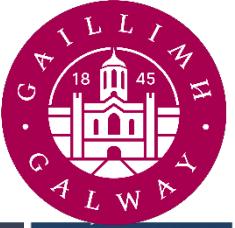
<https://www.sv-europe.com/crisp-dm-methodology/>



Current & Emerging Applications

Any ideas?

What companies use ML & DM?



The image shows a presentation slide from the vevox app. The slide has a teal background. On the left side, there is promotional text for the app: "Join at: vevox.app" and "ID: 165-451-844", along with a QR code. Below this text is an illustration of a hand holding a smartphone displaying the vevox logo, with two other people in the background also looking at phones. In the center of the slide, there is a white rectangular area containing the question "What company use Machine Learning & Data Minning?". In the top right corner of this white area, there is a small circular icon with a person icon and the text "0 / 0". At the bottom of the slide, there is a navigation bar with various icons, including a back arrow, a forward arrow, a search icon, a square (stop) button, and a speech bubble icon.

Press **esc** to exit full screen

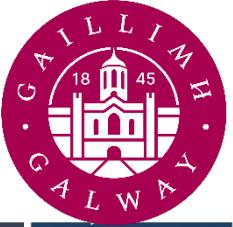
What company use Machine Learning & Data Minning?

0 / 0

Join at:
vevox.app

ID:
165-451-844

WHAT COMPANY USE
MACHINE LEARNING & ... 3/3



Users of ML & DM



amazon

NETFLIX

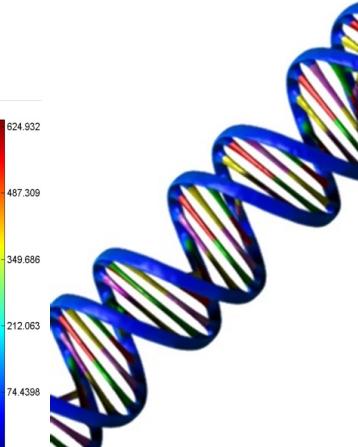
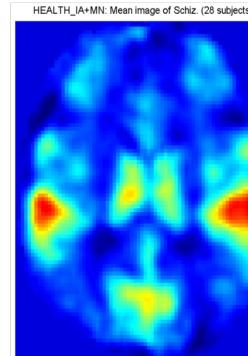
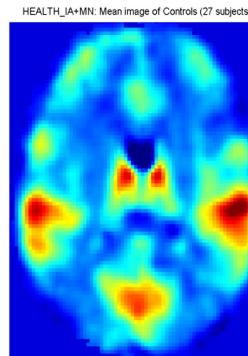
Microsoft

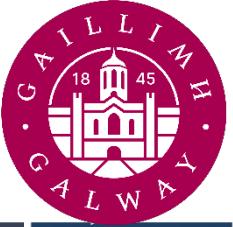
YAHOO!



facebook

Google™





High-Profile Examples ...

Forbes

New Posts
+4 posts this hour

Most Popular
Most Disliked Athletes



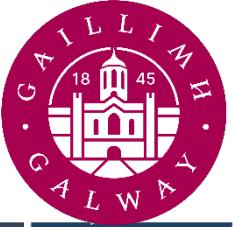
Kashmir Hill, Forbes Staff
Welcome to The Not-So Private Parts where technology & privacy collide
[+ Follow](#) (1,178)

TECH | 2/16/2012 @ 11:02AM | 1,930,513 views

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Forbes, 16 Feb 2012

<http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>



How Netflix is turning viewers into puppets

"House of Cards" gives viewers exactly what Big Data says we want. This won't end well

BY ANDREW LEONARD



House of Cards (BBC, 1990)

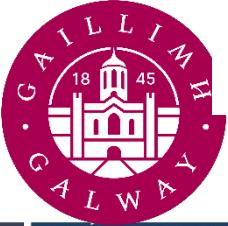
⇒ ★ ★ ★ ★

⇒ Kevin Spacey (Actor)

⇒ David Fincher (Dir.)

Salon, 1 Feb 2013

<https://www.salon.com/2013/02/01/how.netflix.is.turning.viewers.into.puppets/>



Deep Learning for Object Recognition: Hinton & colleagues, NIPS 2012

UNIVERSITY
OF GALWAY



mite

container ship

motor scooter

leopard

mite
black widow
cockroach
tick
starfish

container ship
lifeboat
amphibian
fireboat
drilling platform

motor scooter
go-kart
moped
bumper car
golfcart

leopard
jaguar
cheetah
snow leopard
Egyptian cat



grille



mushroom



cherry



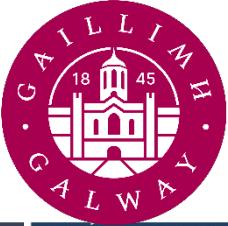
Madagascar cat

convertible
grille
pickup
beach wagon
fire engine

agaric
mushroom
jelly fungus
gill fungus
dead-man's-fingers

dalmatian
grape
elderberry
ffordshire bullterrier
currant

squirrel monkey
spider monkey
titi
indri
howler monkey



AI/ML for Autonomous Vehicles



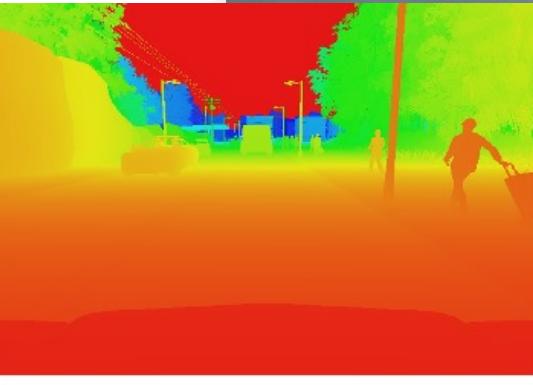
FiveAI: delivering autonomous vehicles to London in 2019

KEY COMPETENCIES WE'RE HIRING

Structure Form Motion (SFM) Depth and Pose Estimation
Stereo Reconstruction Optical Flow
Pixel-Wise Segmentation SLAM Multi-task Learning
Recurrent Neural Networks Unsupervised Learning POMDP
Interpretability / XAI Agent Intention

FiveAI/Careers info@five.ai @ FiveAI

Work on self-driving cars

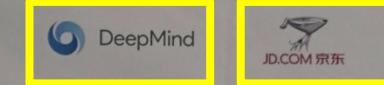


ICML would like to thank the generous sponsors who help make this conference possible each year.

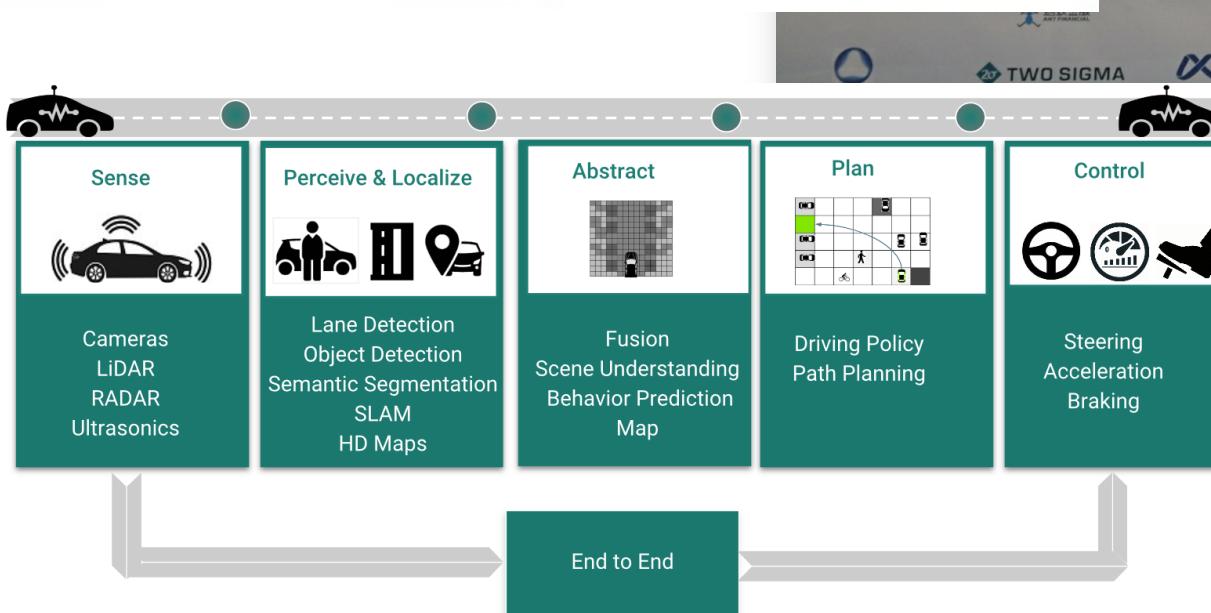
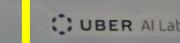
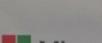
DIAMOND SPONSOR



PLATINUM SPONSORS



GOLD, SILVER & BRONZE SPONSORS





Learning from Experience Without a Teacher



Learns to play the game Go, just by playing games against itself

Starting from completely random play

<https://deepmind.com/blog/alphago-zero-learning-scratch/>





Generative adversarial networks



Obvious' "Portrait of Edmond Belamy" exceeded expectations at Thursday's sale (Courtesy of Obvious)

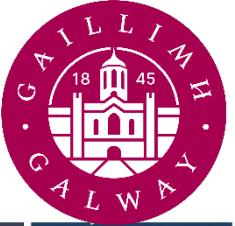
SMARTNEWS *Keeping you current*

Christie's Is First to Sell Art Made by Artificial Intelligence, But What Does That Mean?

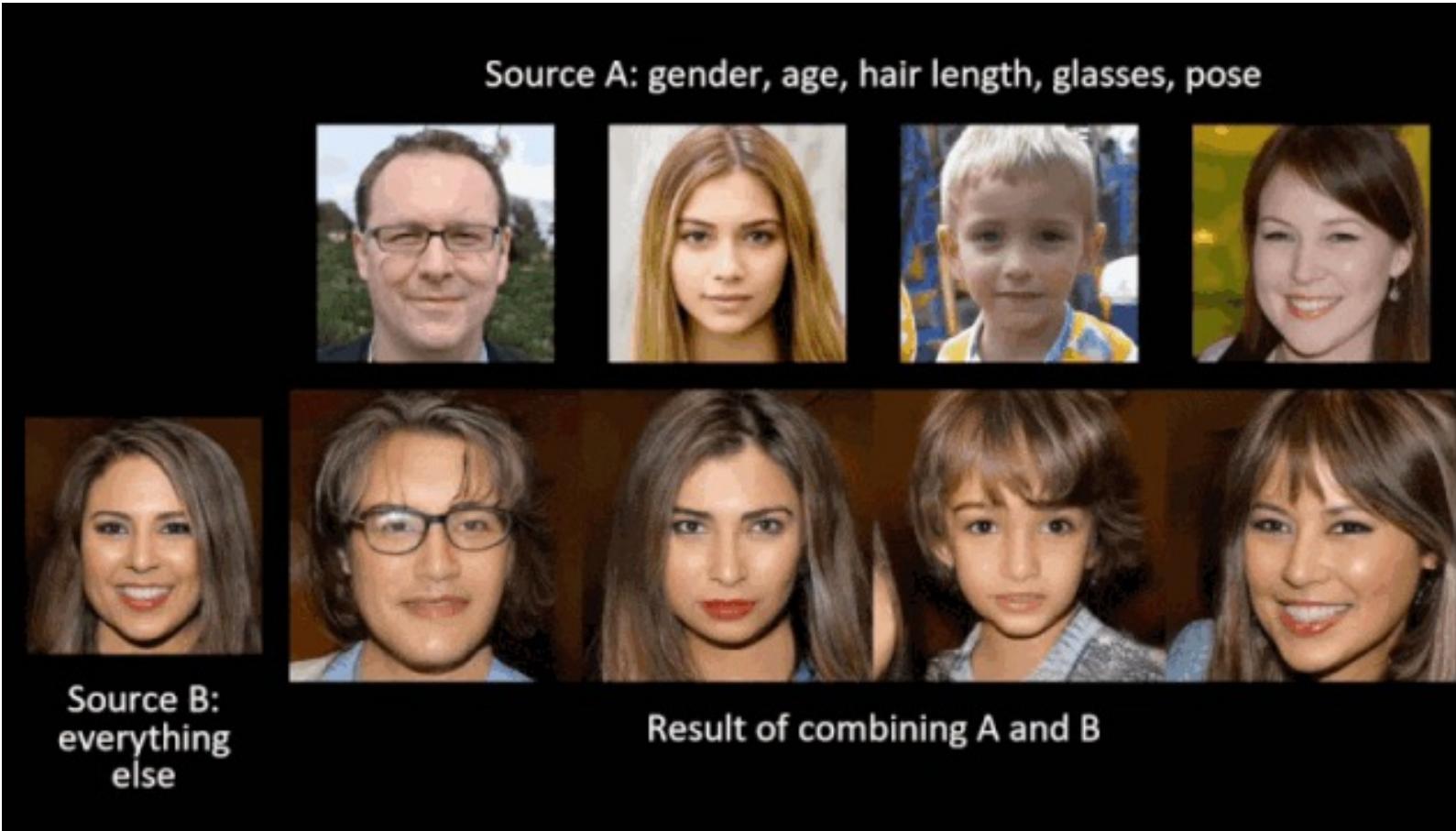
Paris-based art collective Obvious' 'Portrait of Edmond Belamy' sold for \$432,500, nearly 45 times its initial estimate

October 2018

<https://www.smithsonianmag.com/smart-news/christies-first-sell-art-made-artificial-intelligence-what-does-mean-180970642/>

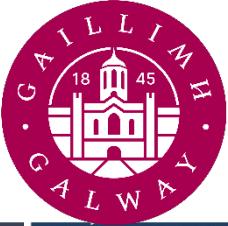


Generative adversarial networks



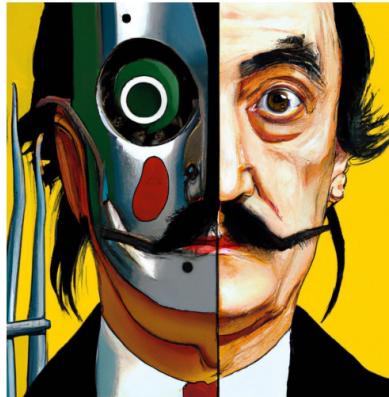
<https://medium.com/syncedreview/nvidia-open-sources-hyper-realistic-face-generator-stylegan-f346e1a73826>

<https://www.theverge.com/2019/3/19/18272602/ai-art-generation-gan-nvidia-doodle-landscapes>



Latest examples

- Image generation from natural language processing
 - Open AI model called Dall-E 1 and 2



vibrant portrait painting of Salvador Dalí with a robotic half face



a shiba inu wearing a beret and black turtleneck



a close up of a handpalm with leaves growing from it



an espresso machine that makes coffee from human souls, artstation

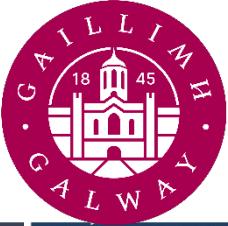


panda mad scientist mixing sparkling chemicals, artstation



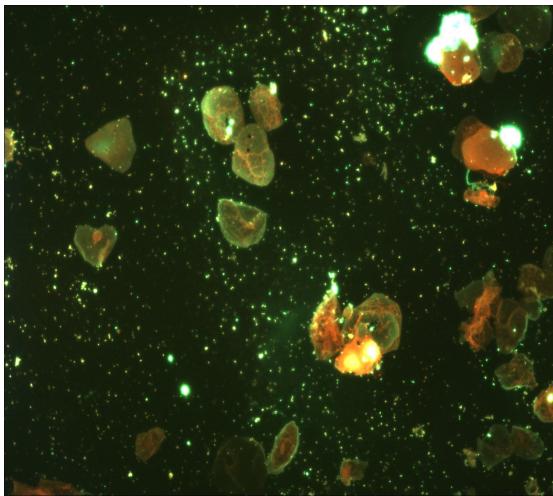
a corgi's head depicted as an explosion of a nebula

<https://analyticsindiamag.com/openai-to-change-the-digital-image-making-game-with-dall-e-2-its-text-to-image-generator/>

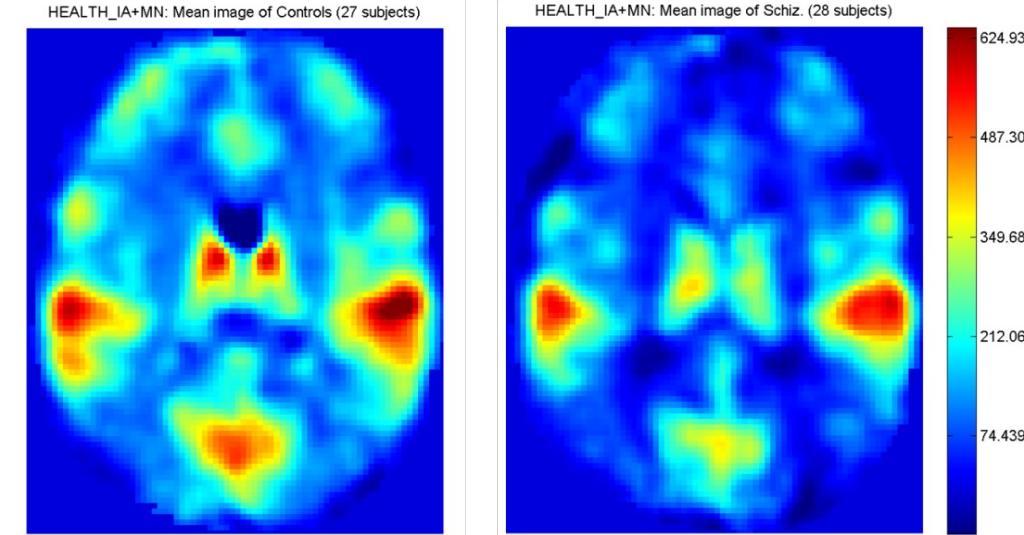


Some local examples at NUI Galway: Image & Sensor Data Mining

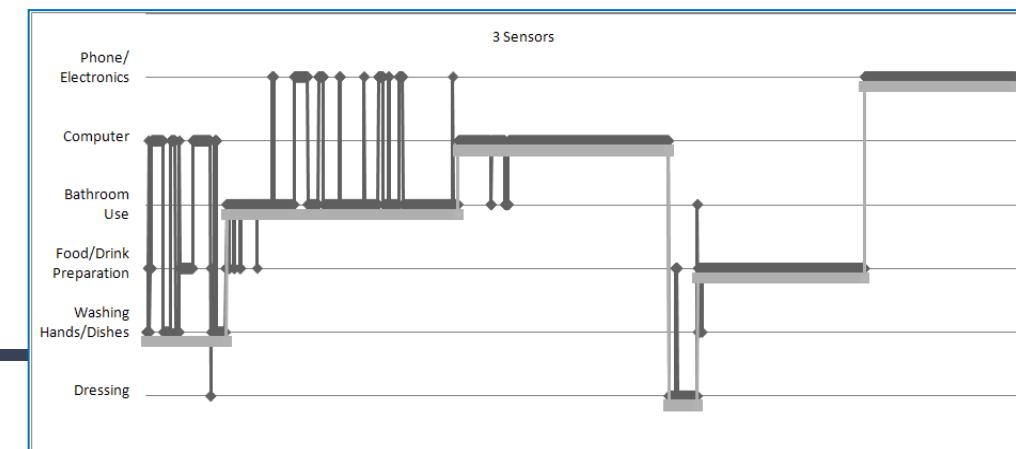
UC Irvine / NIH BIRN collaboration:
Using fMRI to distinguish subjects
with Schizophrenia from controls

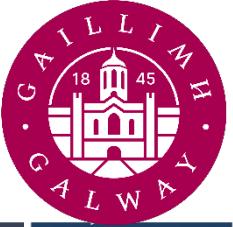


Analysing microscope images
of sputum to screen for TB:
Image processing, ML,
Sequential statistics



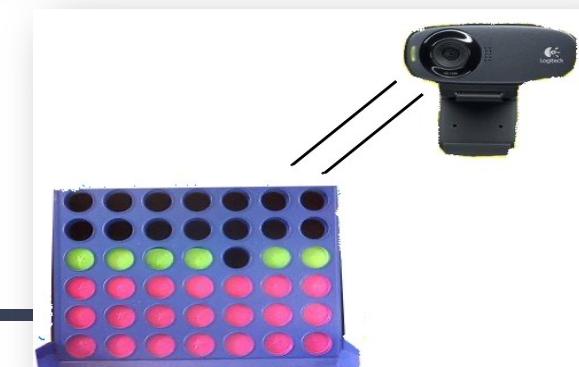
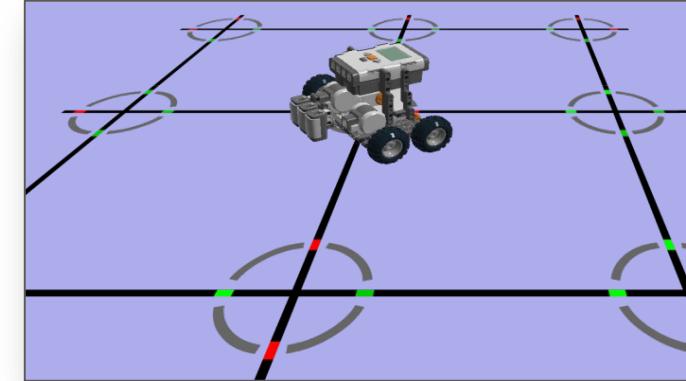
Identify Activities of Daily Living from
sensors: Ensemble DTW Classifier

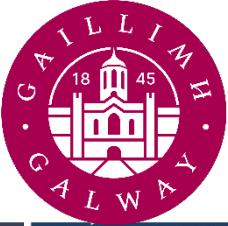




Some local examples at NUI Galway: Reinforcement Learning

- RL Agent that learns to play UT2004 through trial & error
 - Goal: human-like performance
- Robots that learn to navigate mazes & solve puzzles





OLSCOIL NA
GAILLIMHE

UNIVERSITY
OF GALWAY

ROCSAFE: Remotely Operated CBRNE Scene Assessment & Forensic Examination

RAVs with automatic navigation and routes optimised for finding zones of interest and scene overview



Video, images, relayed to Central Decision Management.
Command Centre with map-based GUI showing threat colour maps, etc.
Video & maps augmented with analysed sensor results.

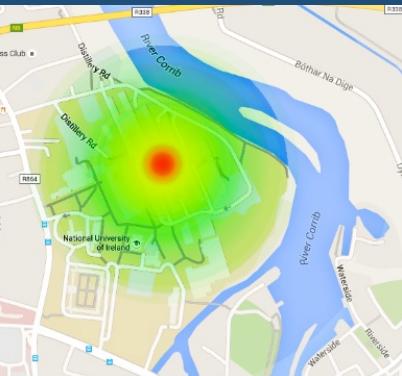


All data streamed to Central Decision Management

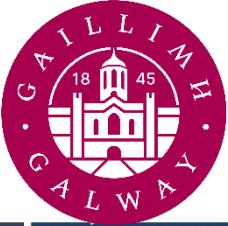
Forensic results, when available, transmitted to Central Decision Management



RGVs deployed directly to zones of interest. Equipped with tools for forensic evidence collection

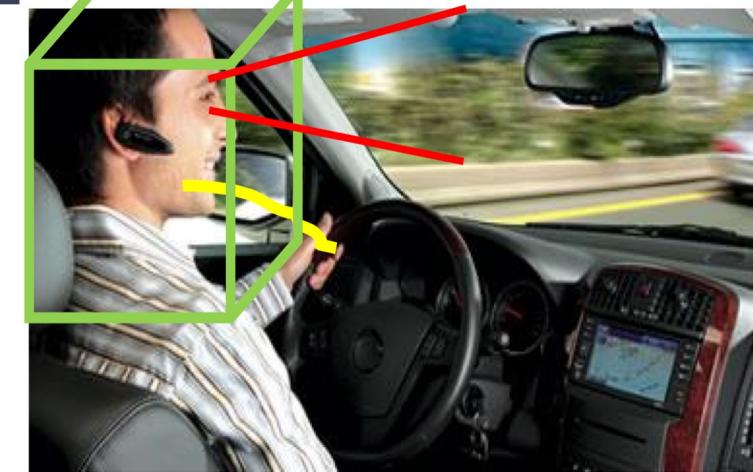


Forensic samples delivered to mobile lab [which is outside the scope of ROCSAFE]



Some of my projects at NUI Galway and outside: Regression, Classification, etc.

- **Driver Gaze Estimation:** FORD's project for driver assessment from gaze
- **FIM:** Light weight DNN for smart freezers inventory management
- **Cybercrime Analytics:** Open social data crime analytics to detect financial organization data breaches
- **ROCSAFE:** Remotely Operated CBRNe Scene Assessment for forensic examination
- **XplainIT:** Explainable deep models for structured data



· 3 Aug 2016

Replies to @Tesco

I phoned and the guy did it all in a minute. Someone **hacked** my account. The matter is resolved. Twitter didn't help!



· 3 Aug 2016

Replies to @Tesco

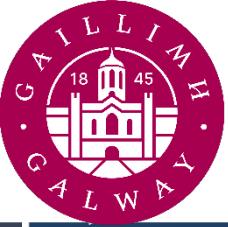
I haven't ordered anything. This needs resolving. I can't access my account because you want a password and now I've been **hacked**



3 Aug 2016

3rd time in three weeks @Tesco have locked my bloody account... Seriously - with passwords like y:+tB5[{yK4gM^h do you think it is **hacked**??





Learning Objectives: Review

If you have been paying close attention, you will now be able to ...

1. Discuss definitions of Machine Learning
2. Describe what major categories of ML task entail: classification, regression, clustering, relationship discovery and reinforcement learning
3. Discuss the relationship with Data Mining
4. Explain the Data Mining process
5. Consider current and future applications of Machine Learning and Data Mining