Lecture 1: Machine Learning for Earth Science

MIR LAB LECTURE SERIES

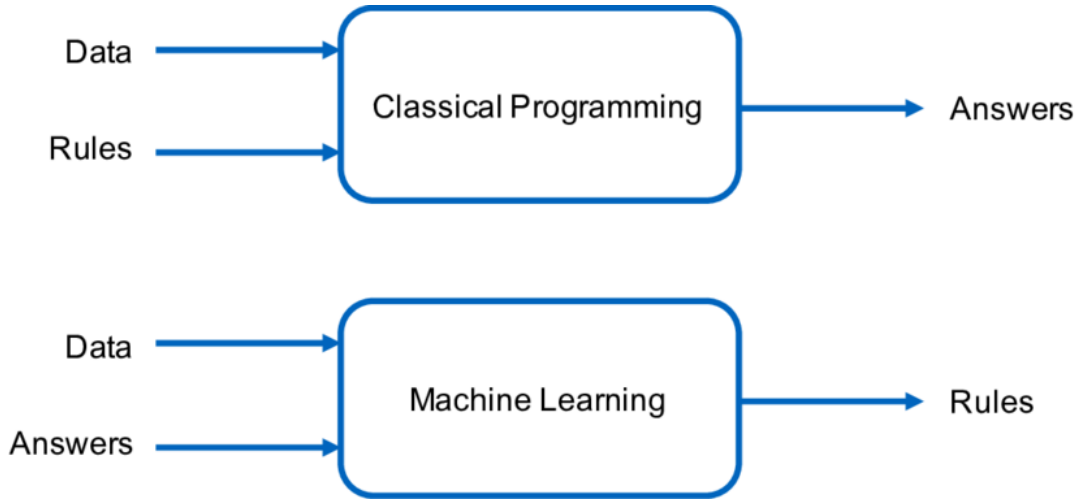
Highlights

1. Introduction to Machine Learning
2. ML in real-life, academia, and earth science applications
3. ML Taxonomy
4. End-to-end ML Pipeline
5. Getting started

Machine Learning (ML)

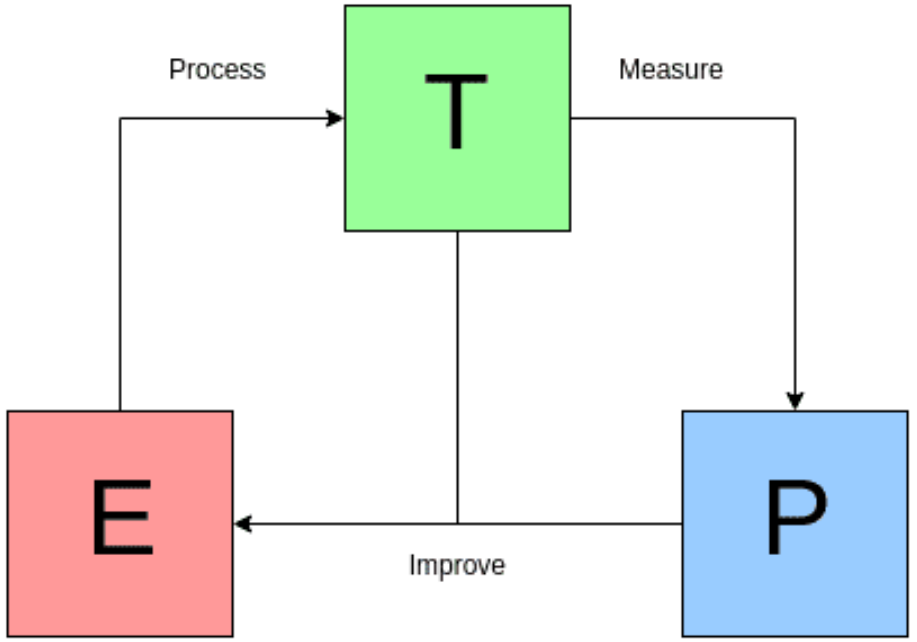
How it differs from Traditional/Classical Programming?

- The term **Machine Learning** was coined by Arthur Samuel 1959, an American pioneer in the field of computer gaming and artificial intelligence, and stated that "**it gives computers the ability to learn without being explicitly programmed.**"
- Simply, it can be said that ML enables our program to achieve a certain level of **cognition** or **awareness**.

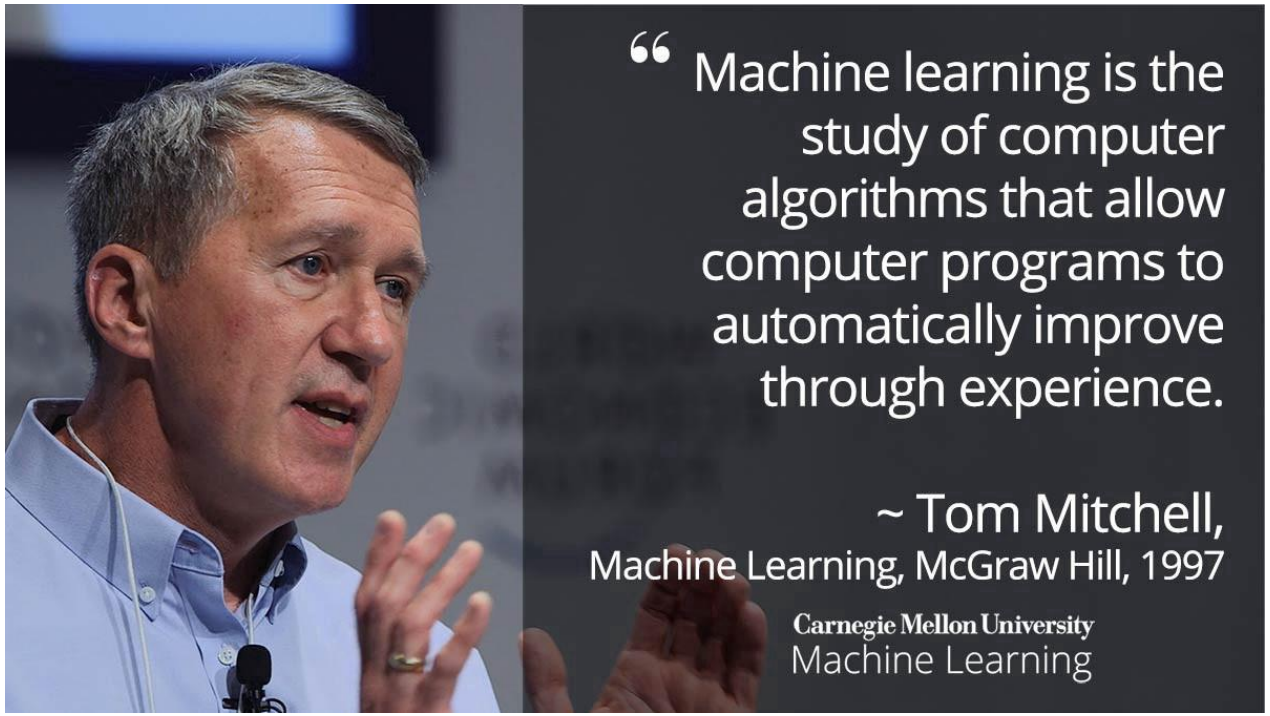


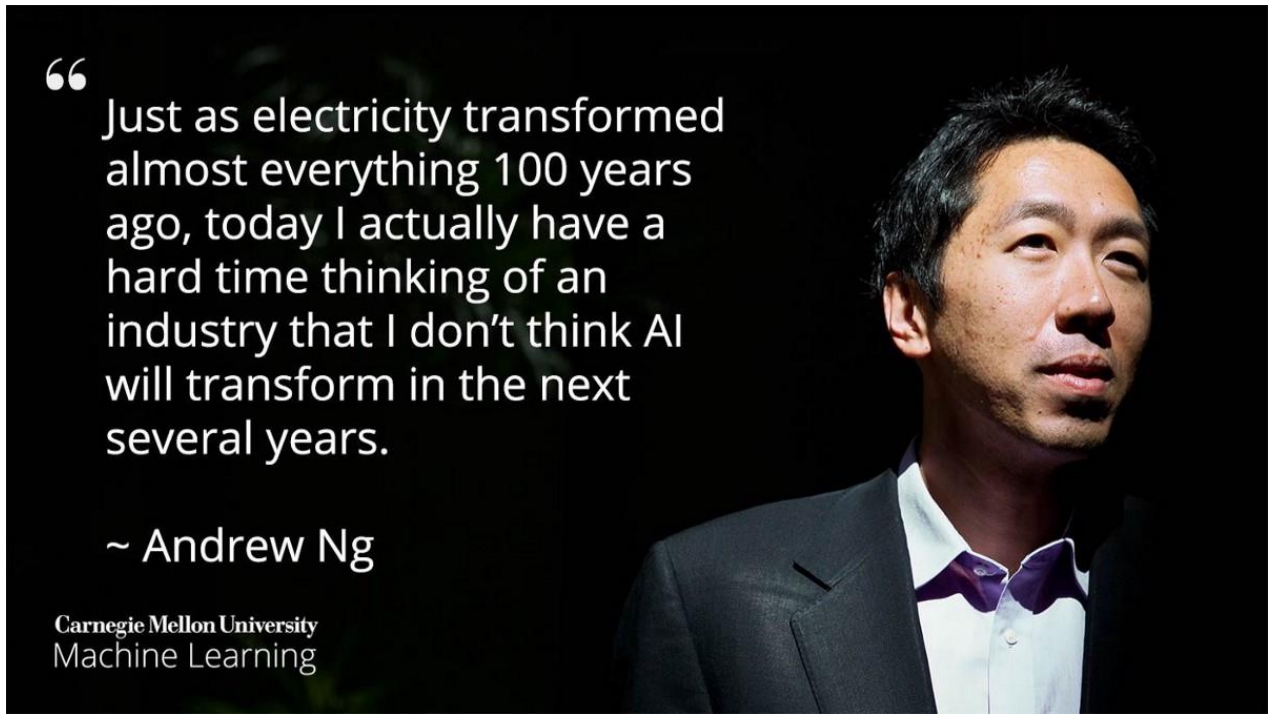
Machine Learning
A better definition?

- In 1997, Tom Mitchell gave a ***well-posed mathematical and relational definition*** – "A computer program is said to learn from **experience E** concerning some **task T** and some **performance measure P**, if its performance on T, as measured by P, **improves** with experience E."



A thought to ponder…





However,

1. What is AI (Artificial Intelligence)?
2. How is it similar or different from ML?
3. Can these terms be used interchangeably?

Observations: 

1. Lack of a precise, universally accepted definition of AI, but instead of confusion has **stimulated the field to grow**.
2. Simply, The ability of a machine to ***learn from experience*** and perform *tasks* ***normally attributed to human intelligence***, for example, problem-solving, reasoning, and understanding natural language.
3. Though used interchangeably, ML falls under the umbrella of AI.
4. However, 'Defining AI' exhibits significant changes across domains and with time. (Read: Link, and Link)

What ML isn’t?

It Isn’t a magic wand that will simply solve any problem.

What we can say ML is?

- Though not apparent for a beginner, there's a rationale behind every successful and failed ML pipeline.
- It isn't 'Garbage-in, Garbage-out'.
- Every ML model necessitates that the data it uses is descriptive of the task it is training for.
- Most moving parts are intuitive as performance varies with many factors. This makes it more art-like as this becomes a muscle memory for seasoned ML practitioners.
- **Analogy**: Painting, as an art, involves abundant intricacies, but seasoned painters attain a high level with experience. That doesn't mean you can do '***Math-free ML***', instead you should know at least enough.

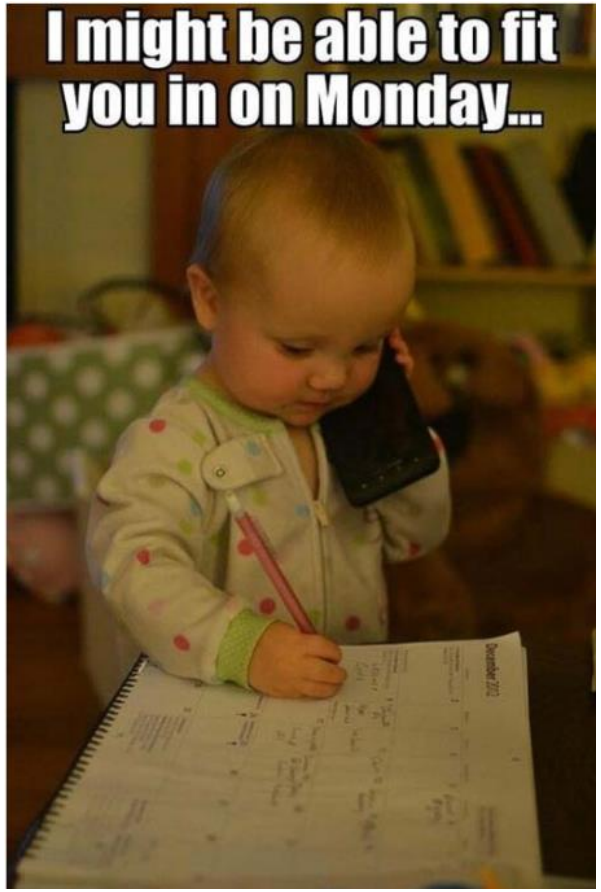
ML Applications

In real life

1. Google Map
2. Recommender systems in advertisements, online shopping, Netflix and more.
3. E-mail Spam filters
4. Smart replies and word completion
5. Smart assistants like Apple Siri, Google Now, Windows Cortana
6. Just Walk Out by Amazon

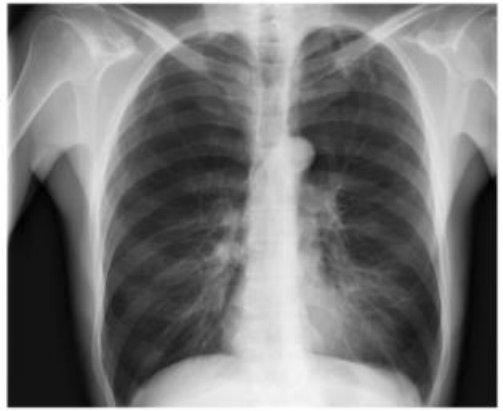
ML Applications: Google Duplex

In real life



https://youtu.be/D5VN56jQMWM

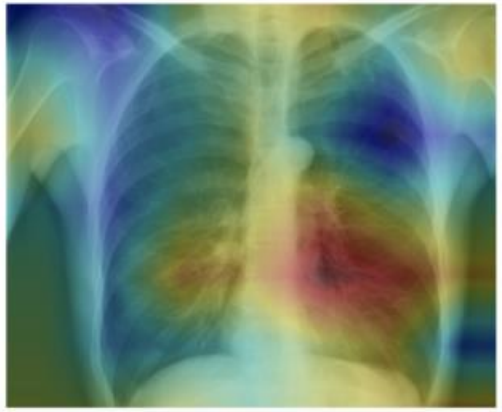
ML Applications: **Healthcare**
In real life



Input
Chest X-Ray Image

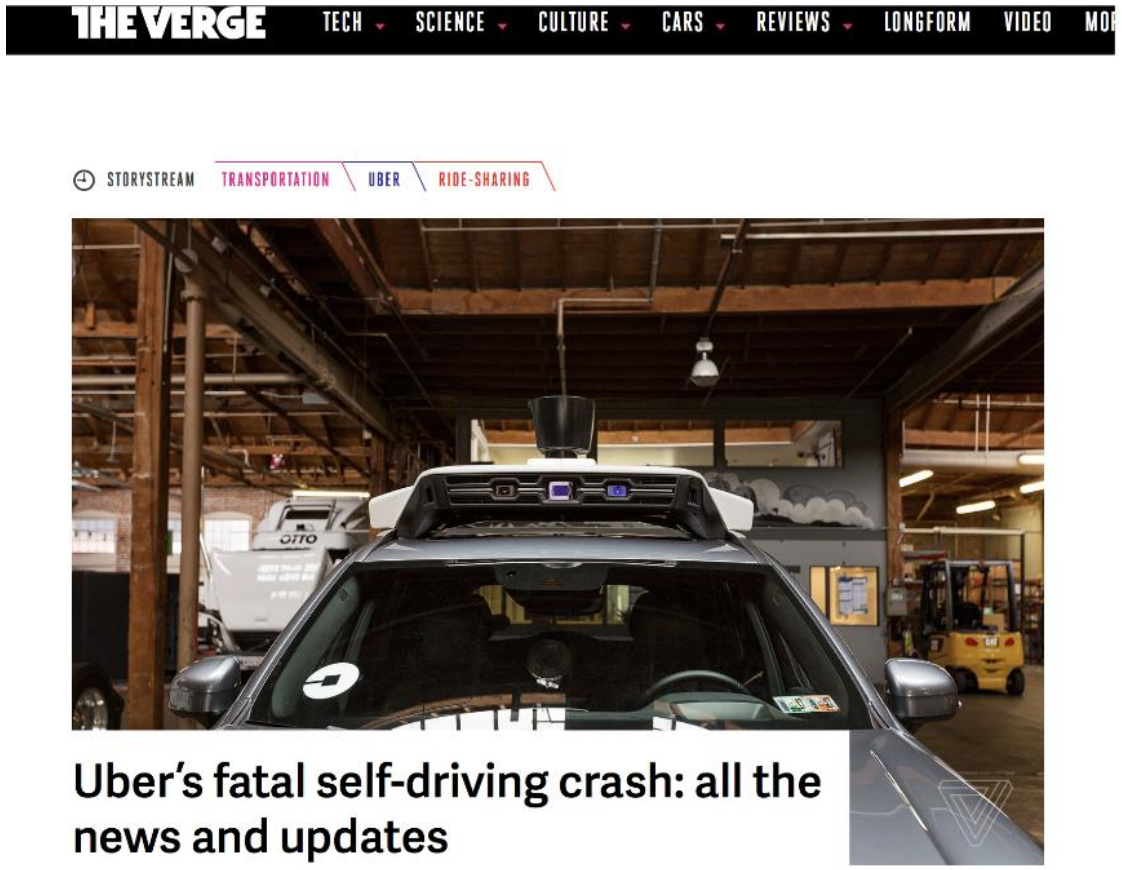
CheXNet
121-layer CNN

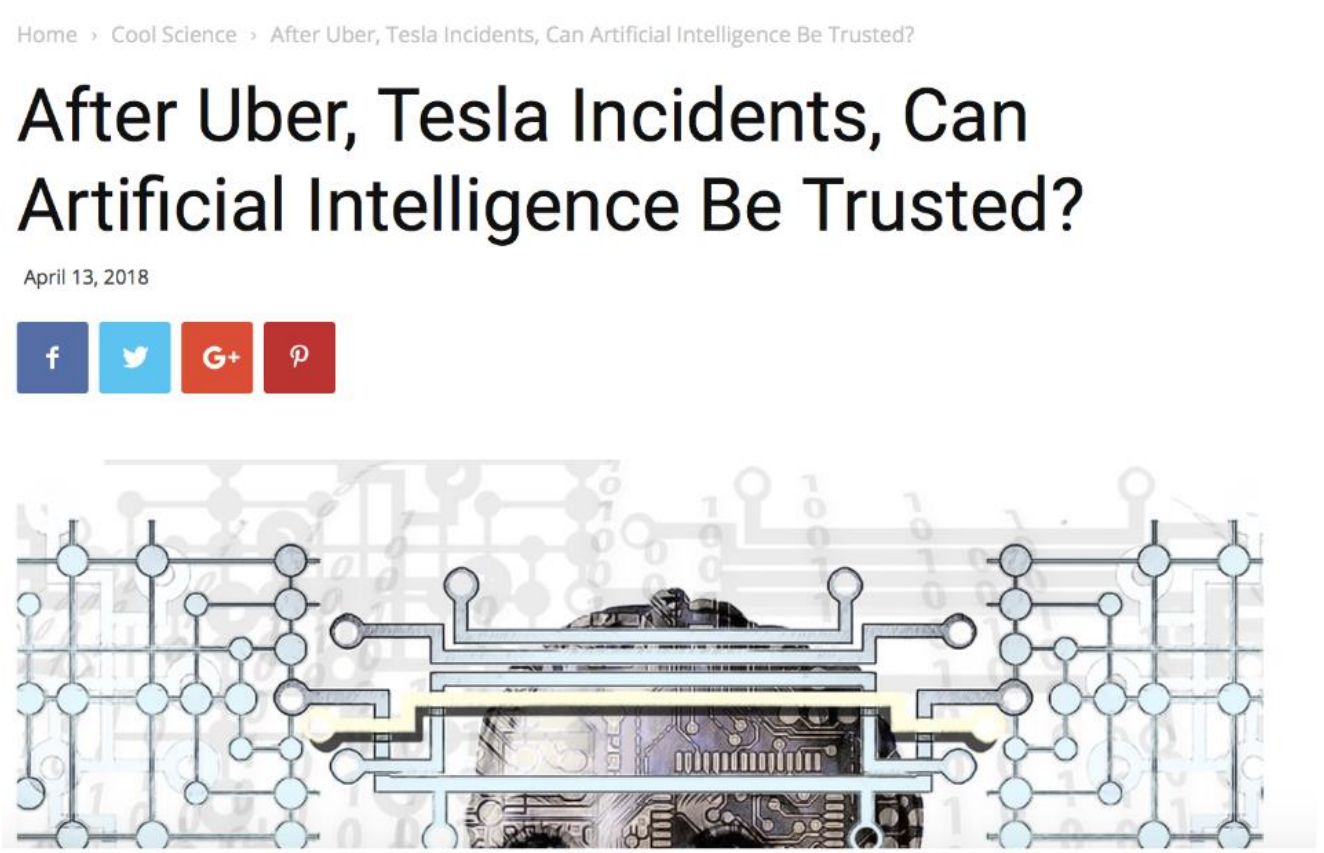
Output
Pneumonia Positive (85%)



ML Applications: **Caution Stories**

In real life





ML Applications

In academia

1. Replacing the traditional toolkits with ML counterparts
2. Addressing systems where enough information isn't known or comprehensible
3. Solving ill-posed problems
4. Achieving generality, robustness, fairness, continuality, resilience, interpretability, and/or comprehensiveness for complex chaotic systems
5. Discovering unknown physics

and more.

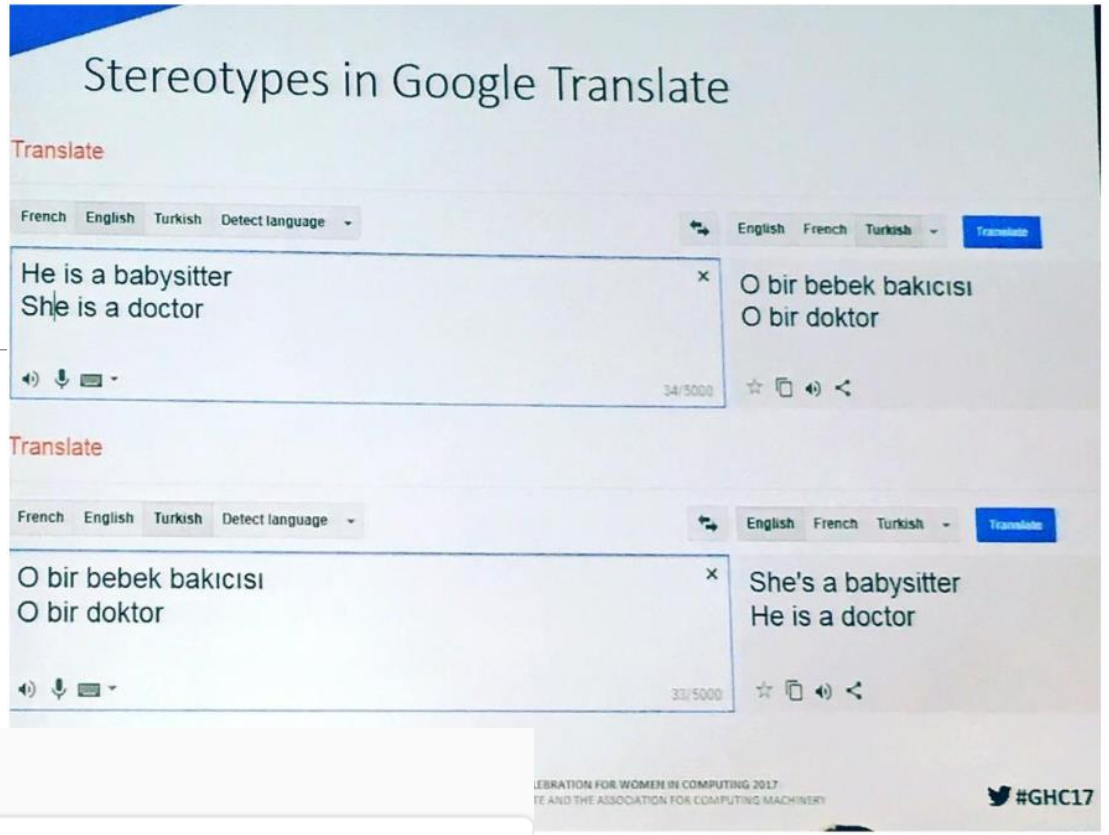
ML Applications

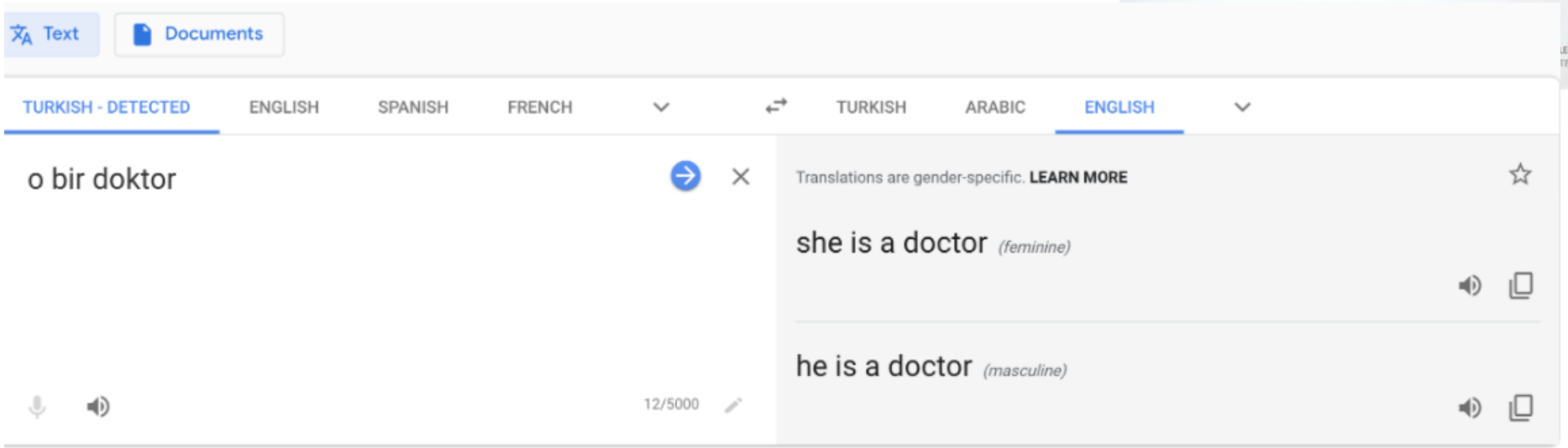
In Earth Sciences

1. Image Segmentation
2. Downscaling
3. Forecasting
4. Sensor Fusion

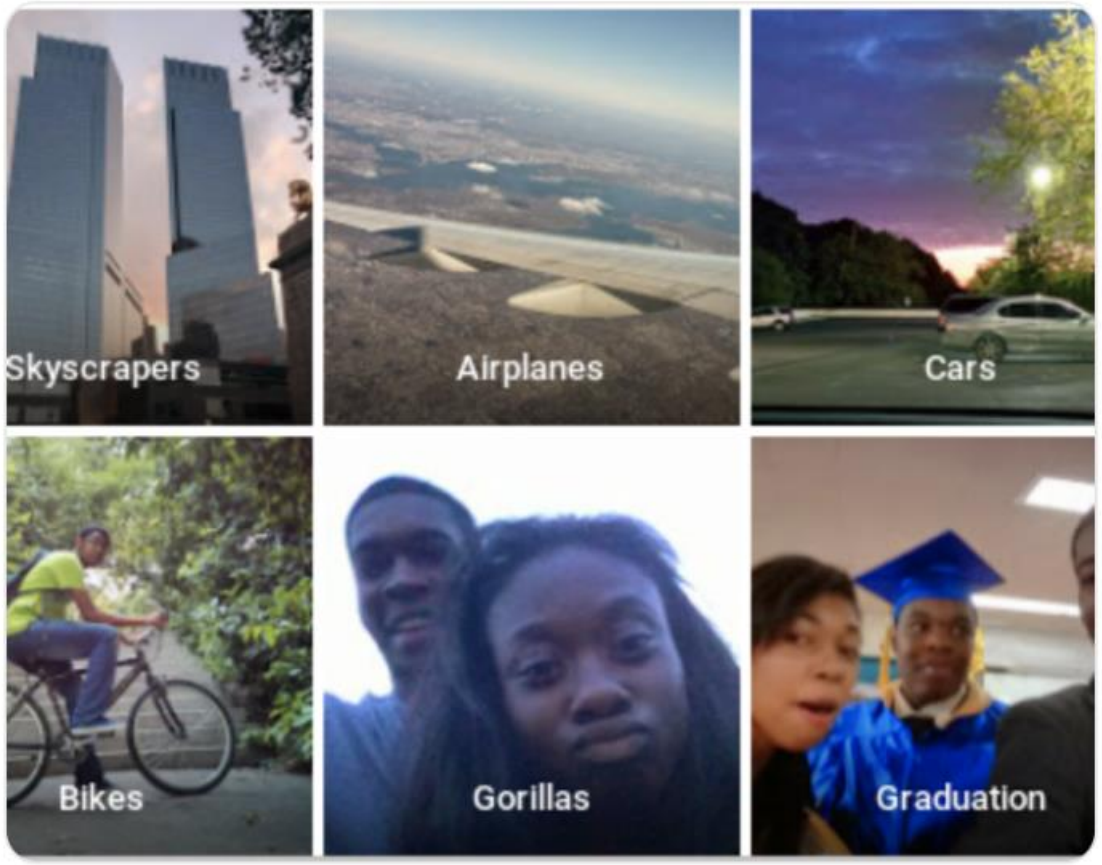
and more.

Bias in ML


Stereotypes in Google Translate
Translate
French English Turkish Detect language
English French Turkish
He is a babysitter
She is a doctor
O bir bebek bakıcısı
O bir doktor
34/5000
Translate
French English Turkish Detect language
English French Turkish
O bir bebek bakıcısı
O bir doktor
She's a babysitter
He is a doctor
33/5000
#GHC17


Text
Documents
TURKISH - DETECTED
ENGLISH
SPANISH
FRENCH
TURKISH
ARABIC
ENGLISH
o bir doktor
12/5000
Translations are gender-specific. LEARN MORE
she is a doctor (feminine)
he is a doctor (masculine)

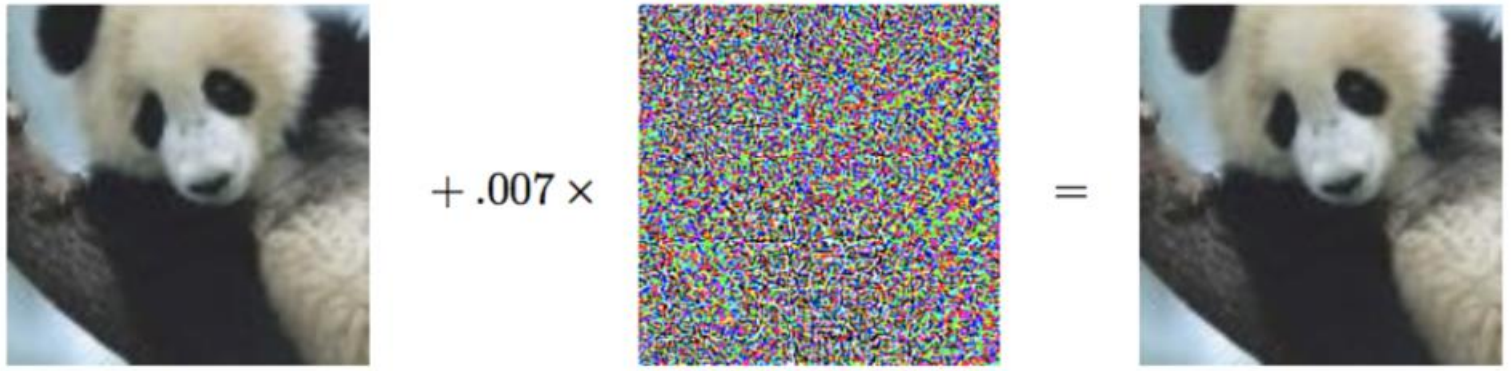
Bias in ML



Bias in ML

Where is the bride?

Adversaries



Trivia Time

Coupon Generators:

Trivia Time

Coupon Generators:

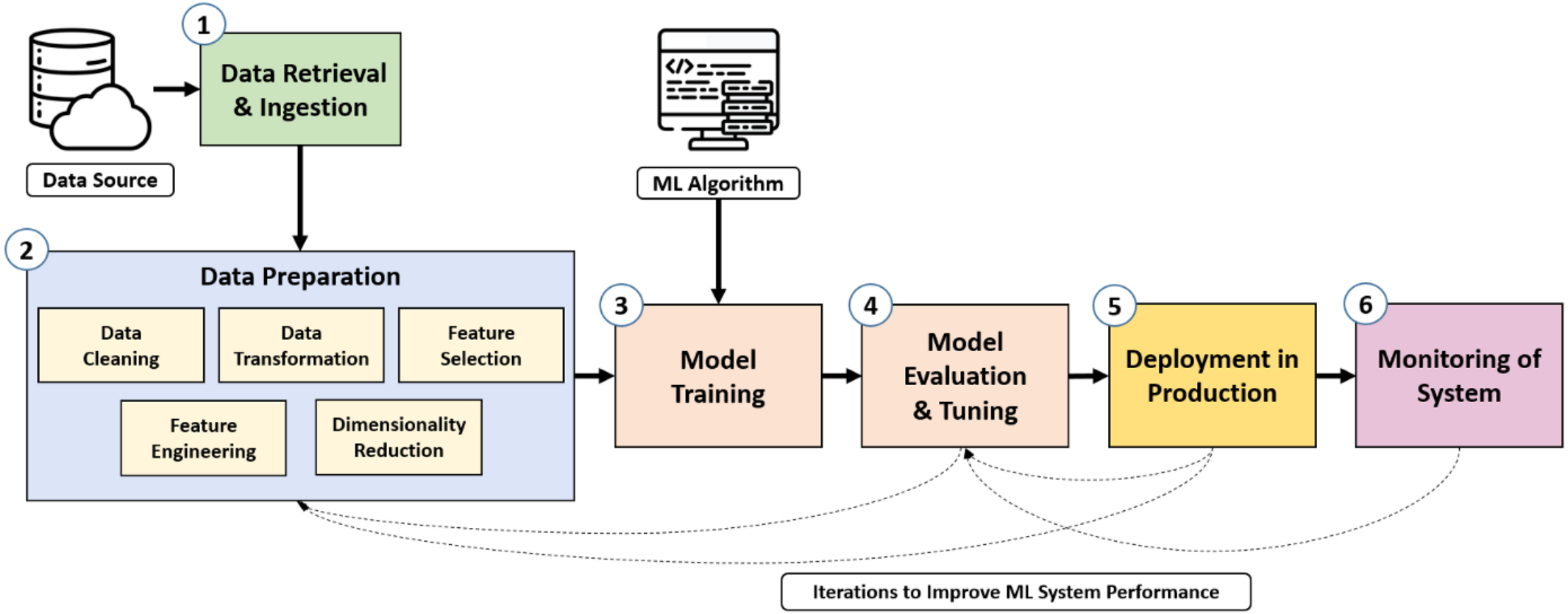
- They spit out coupons and offers based on your previous purchases.
- You see these recommendation engines at checkouts in drug stores and grocery stores.
- An algorithm predicts when you're likely to run out of a previous purchase (such as dog food) and spits out a "right on time" coupon to entice you to buy it again.
- It **does not know why you buy what you bought**, and it doesn't learn your preferences or behaviour changes over time.

Trivia Time

- **Online ad matching**
- **Chatbots**
- **Car diagnostics**
- **Internet of Things (IoT)**

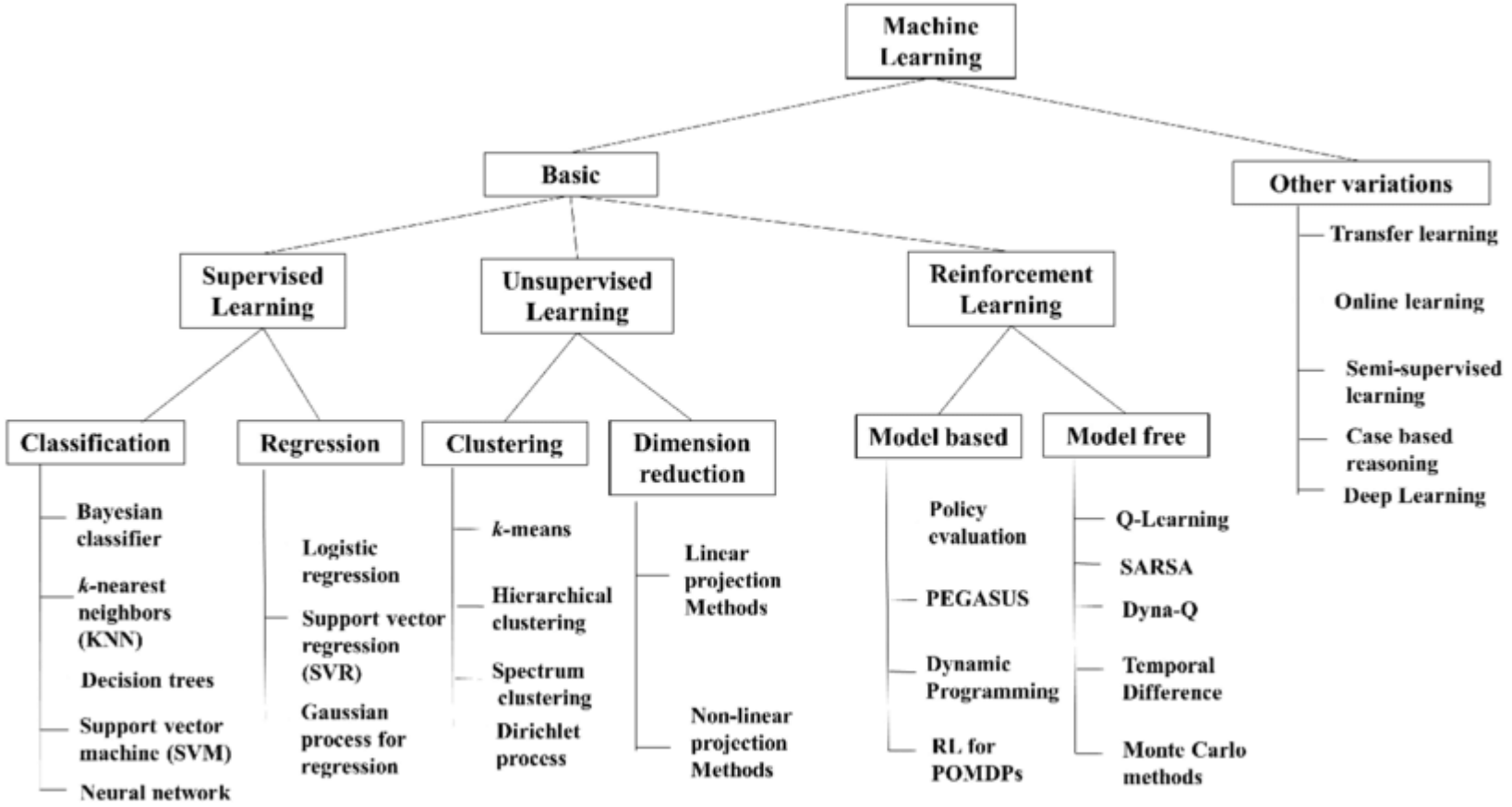
Conclusion: Understand the difference between **ML model** and an **algorithm**

A generic end-to-end ML Pipeline



ML Taxonomy

[ML Mindmap]



Expectations for getting started with this course

- Using Local installation (Assuming Windows)
  1. Git
  2. Anaconda
  3. Visual Studio Code

- Using Google Colab

Note:

- Conda environment yml file will be uploaded to the GitHub repository along with this presentation.
- For idle training and heavy computation, we'll employ the lab servers.

Thank You