Aprendizado de Máquina e Reconhecimento de Padrões 2021.2

The Machine Learning Landscape

Prof. Samuel Martins (Samuka)
samuel.martins@ifsp.edu.br

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
São Paulo
Campus Campinas

What is Machine Learning?

What is Machine Learning?

Arthur Samuel

"The field of study that gives computers the ability to learn without being explicitly programmed."

What is Machine Learning?

Arthur Samuel

"The field of study that gives computers the ability to learn without being explicitly programmed."

*"A computer program is said to **learn** from **experience E** with respect to some **class of tasks T** and **performance measure P**, if its performance at tasks in **T**, as measured by **P**, improves with **experience E**."*

Tom Mitchell

What is Machine Learning?

Arthur Samuel

"The field of study that gives computers the ability to learn without being explicitly programmed."

*"A computer program is said to **learn** from **experience E** with respect to some **class of tasks T** and **performance measure P**, if its performance at tasks in **T**, as measured by **P**, improves with **experience E**."*

Tom Mitchell

Ex: Spam detector.

T = Predict whether an email is spam or not.
E = Collection of emails labeled as spam or not.
P = Spam classifier accuracy.

What is Machine Learning?

Arthur Samuel

"The field of study that gives computers the ability to learn without being explicitly programmed."

*"A computer program is said to **learn** from **experience E** with respect to some **class of tasks T** and **performance measure P**, if its performance at tasks in **T**, as measured by **P**, improves with **experience E**."*

Tom Mitchell

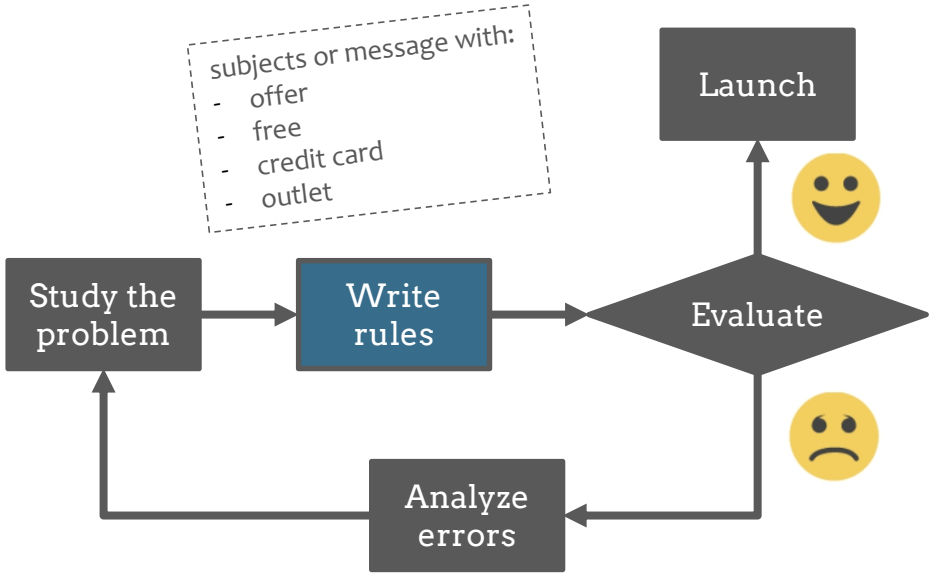
Ex: Spam detector.

T = Predict whether an email is spam or not.
E = Collection of emails labeled as spam or not.
P = Spam classifier accuracy.


Springboard
NETFLIX
IBM
facebook
UBER
Google
amazon

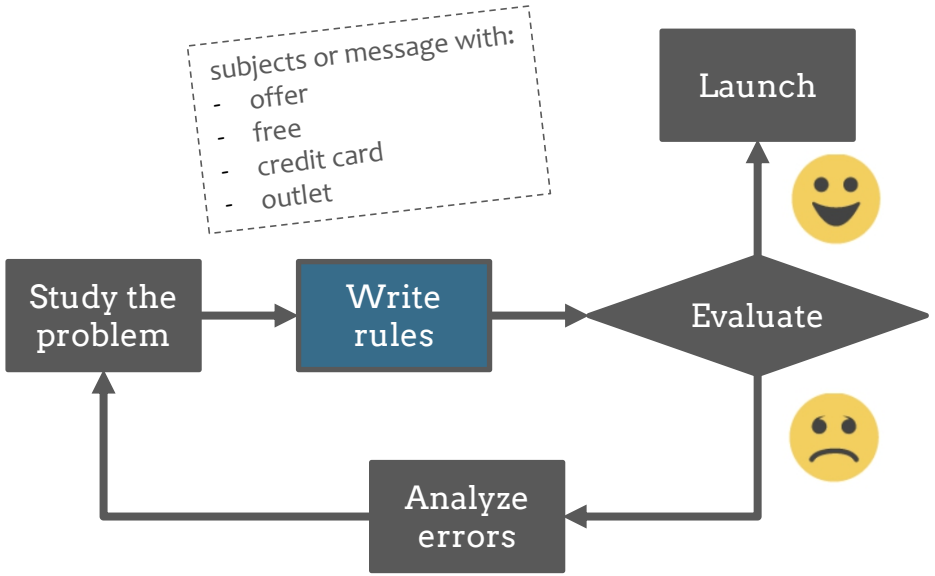
Why Use Machine Learning?

Why Use Machine Learning?

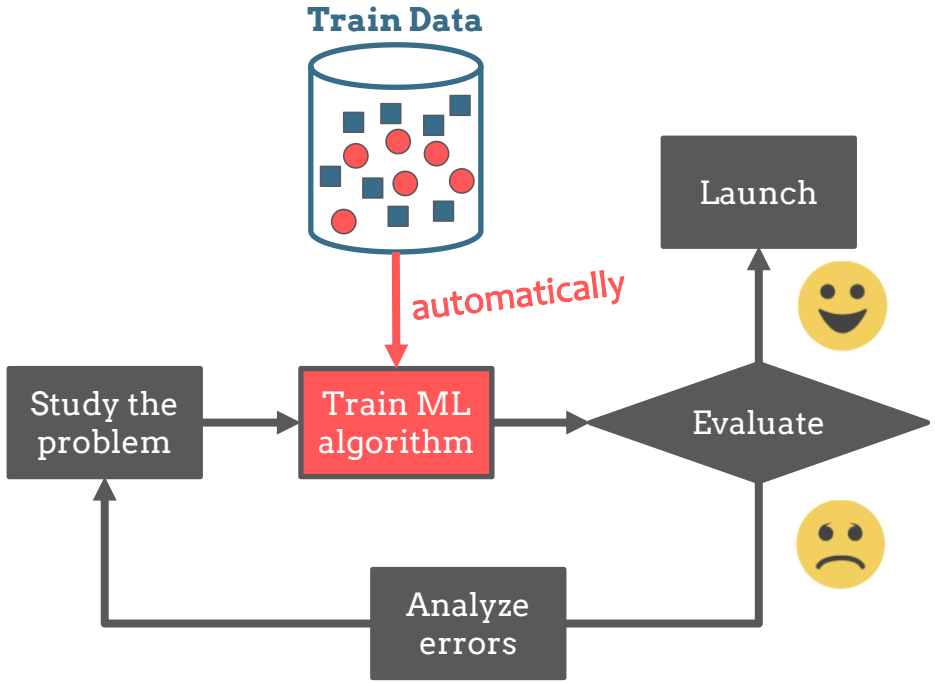


The traditional approach

Why Use Machine Learning?

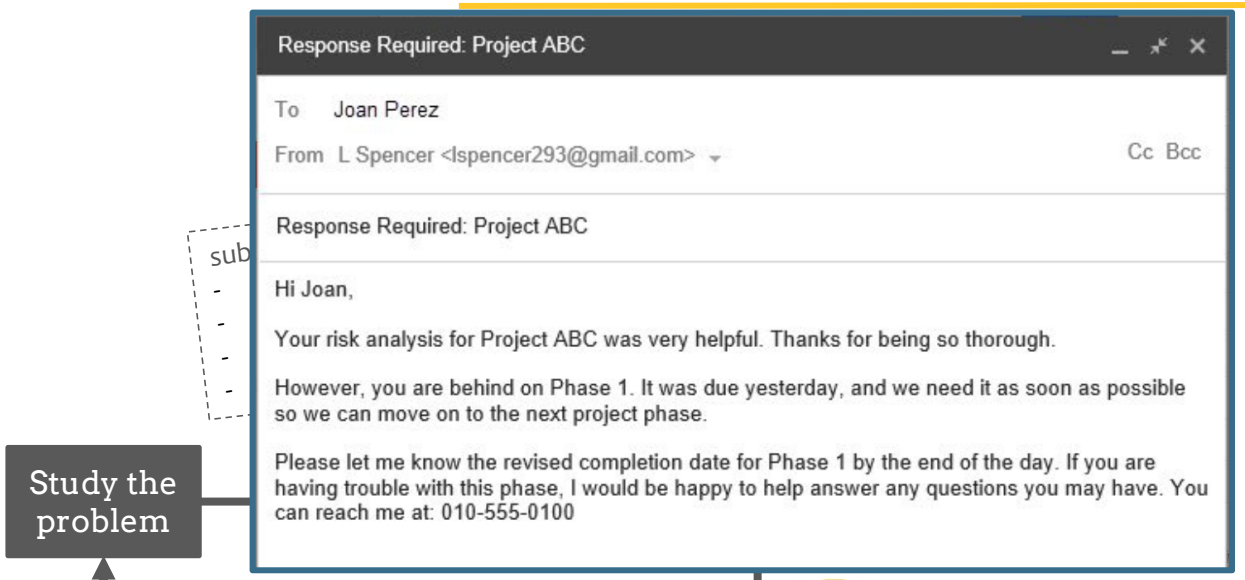


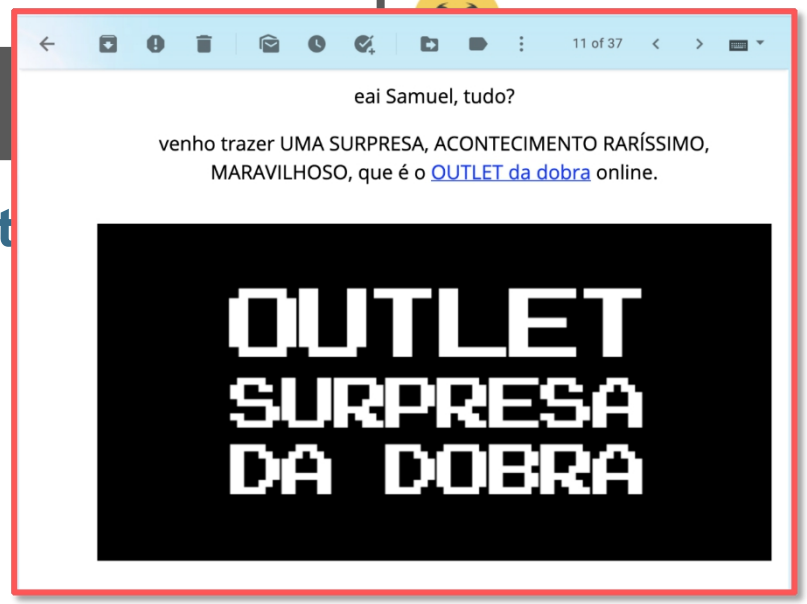
The traditional approach



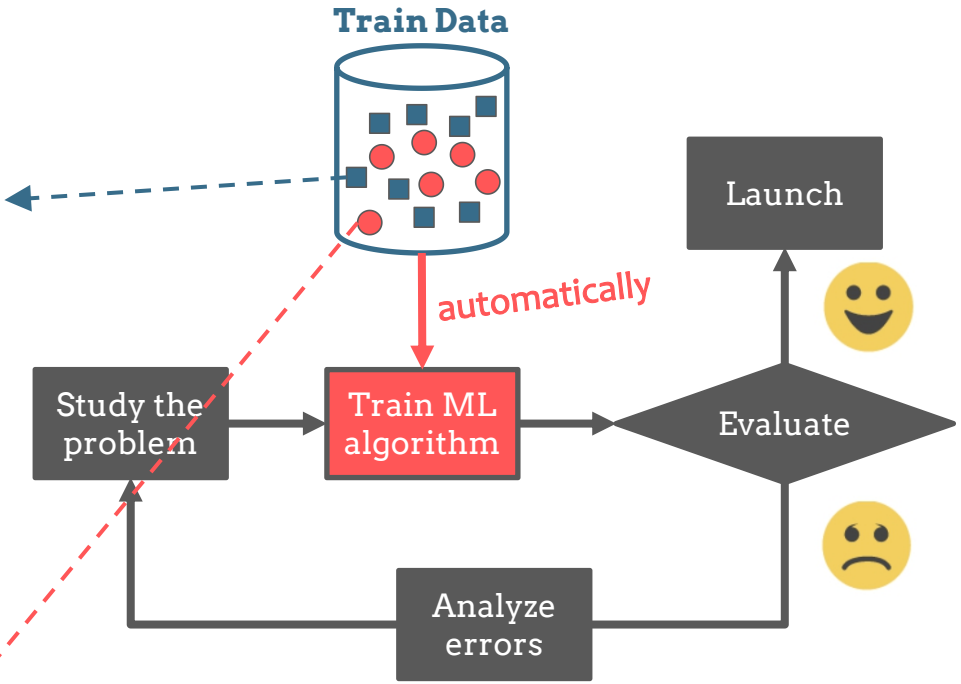
The machine learning approach

Why Use Machine Learning?



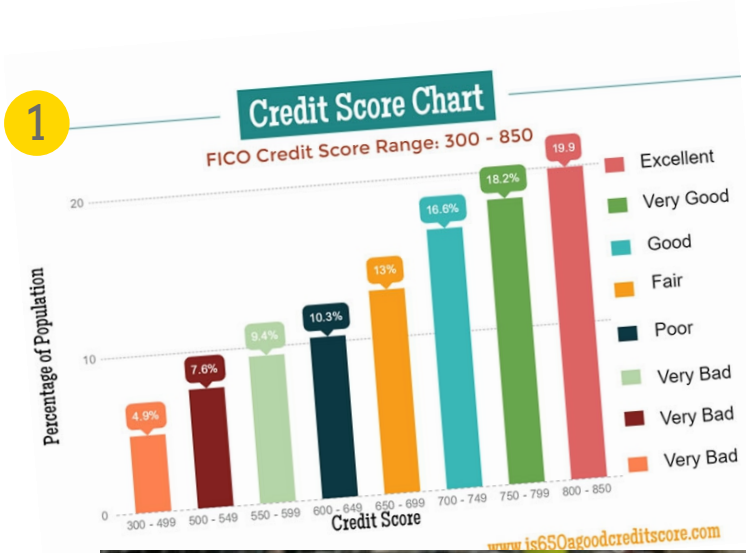


The tradit

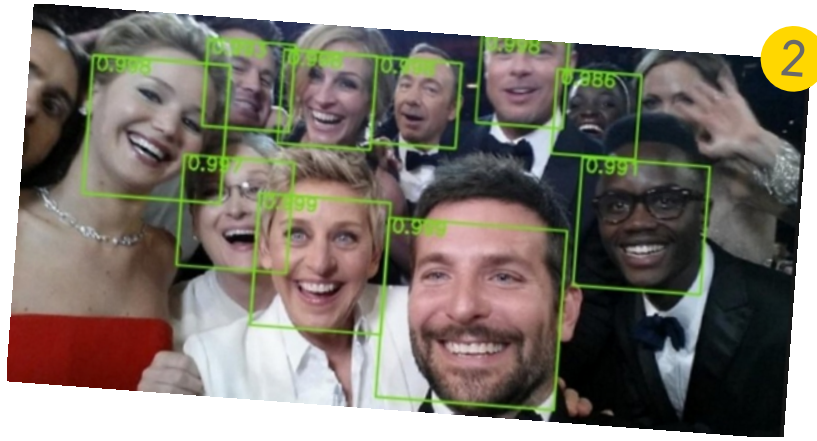


The machine learning approach

1


Credit Score Chart
FICO Credit Score Range: 300 - 850
Percentage of Population
Credit Score
4.9%
7.6%
9.4%
10.3%
13%
16.6%
18.2%
19.9
300 - 499
500 - 549
550 - 599
600 - 649
650 - 699
700 - 749
750 - 799
800 - 850
Excellent
Very Good
Good
Fair
Poor
Very Bad
Very Bad
Very Bad
www.isGSOagoodcreditscore.com

2



3


Speak now

4

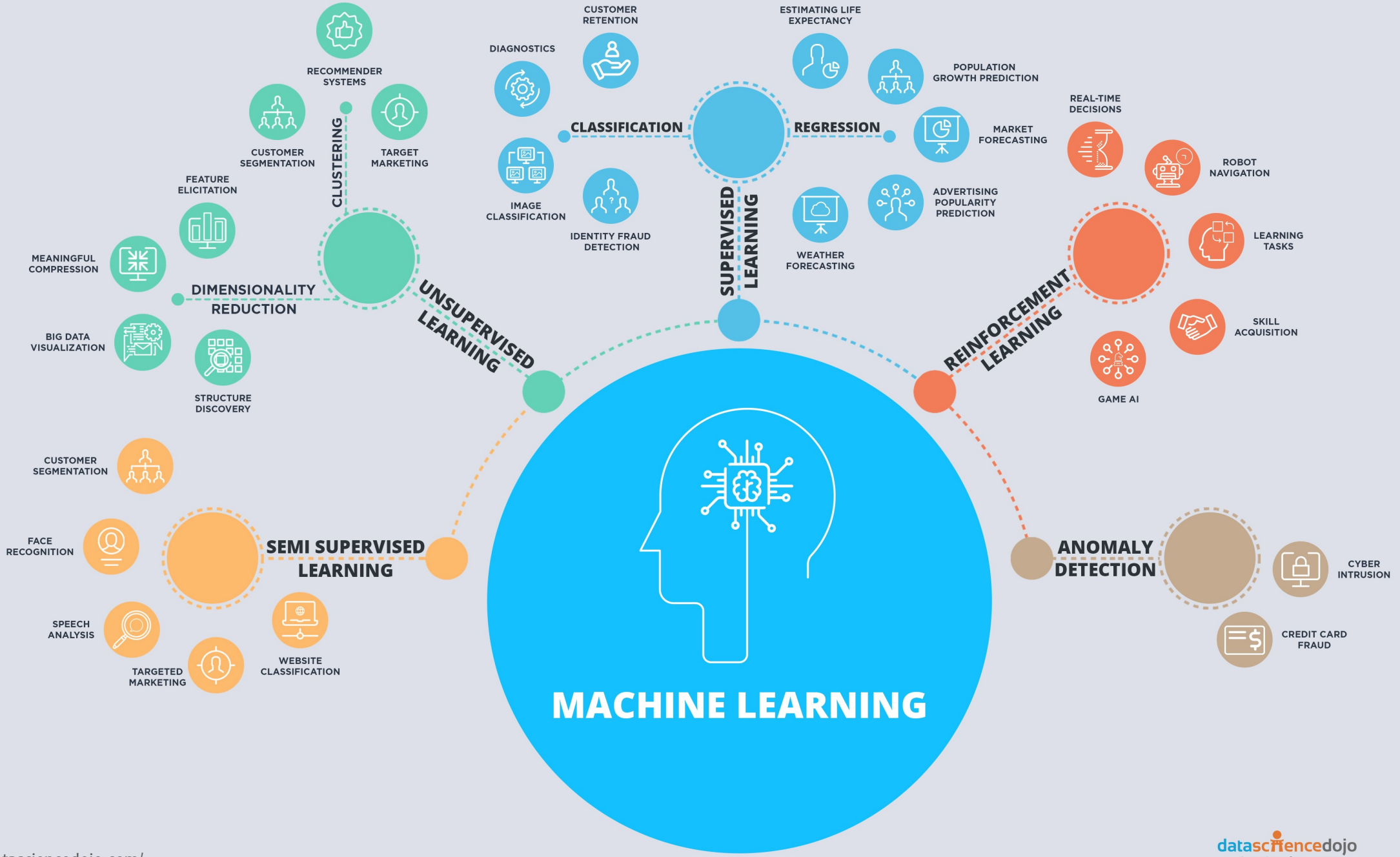
5


KINECT
SPORTS

https://medium.com/@narengowda/netflix-system-design-dbec30fede8d

https://internetofbusiness.com/nvidia-ai-synthetic-mri-diagnostics/

Types of Machine Learning Systems


MACHINE LEARNING
SUPERVISED LEARNING
CLASSIFICATION
DIAGNOSTICS
CUSTOMER RETENTION
IMAGE CLASSIFICATION
IDENTITY FRAUD DETECTION
REGRESSION
ESTIMATING LIFE EXPECTANCY
POPULATION GROWTH PREDICTION
MARKET FORECASTING
ADVERTISING POPULARITY PREDICTION
WEATHER FORECASTING
UNSUPERVISED LEARNING
CLUSTERING
RECOMMENDER SYSTEMS
CUSTOMER SEGMENTATION
TARGET MARKETING
DIMENSIONALITY REDUCTION
FEATURE ELICITATION
MEANINGFUL COMPRESSION
BIG DATA VISUALIZATION
STRUCTURE DISCOVERY
REINFORCEMENT LEARNING
REAL-TIME DECISIONS
ROBOT NAVIGATION
LEARNING TASKS
SKILL ACQUISITION
GAME AI
SEMI SUPERVISED LEARNING
CUSTOMER SEGMENTATION
FACE RECOGNITION
SPEECH ANALYSIS
TARGETED MARKETING
WEBSITE CLASSIFICATION
ANOMALY DETECTION
CYBER INTRUSION
CREDIT CARD FRAUD

Supervised Learning

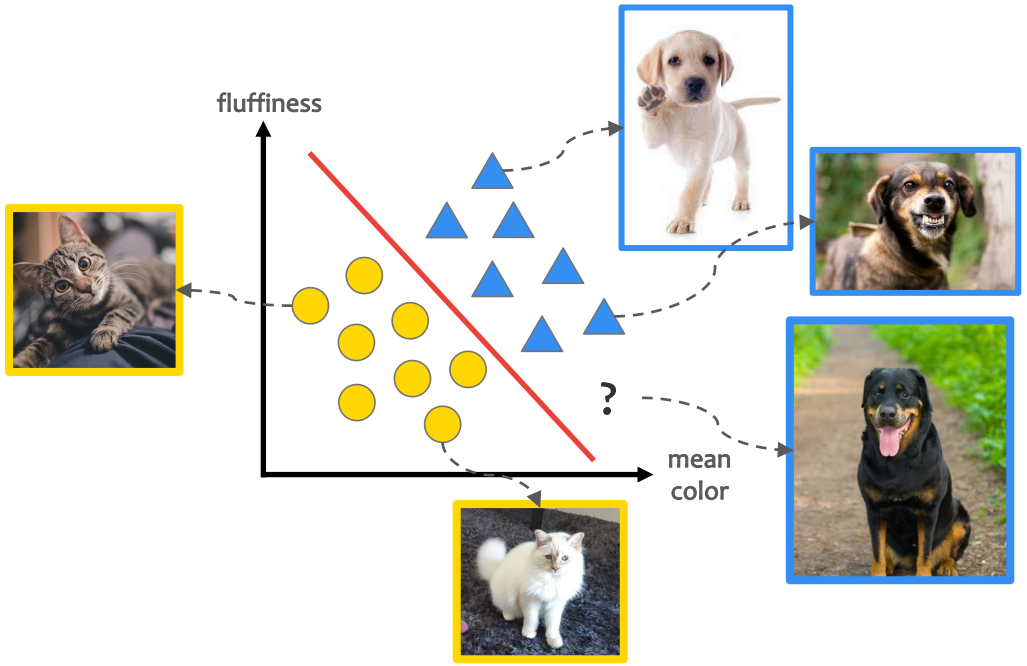
Supervised Learning

Classification

(usually) annotated by **humans:** time-consuming and costly

The **training set** includes annotations / labels / classes

Prediction of **categorical variables (labels)**



- k-Nearest Neighbors (KNN)
- Logistic Regression
- Support Vector Machines (SVM)
- Decision Trees and Random Forest
- Neural Networks
- ...

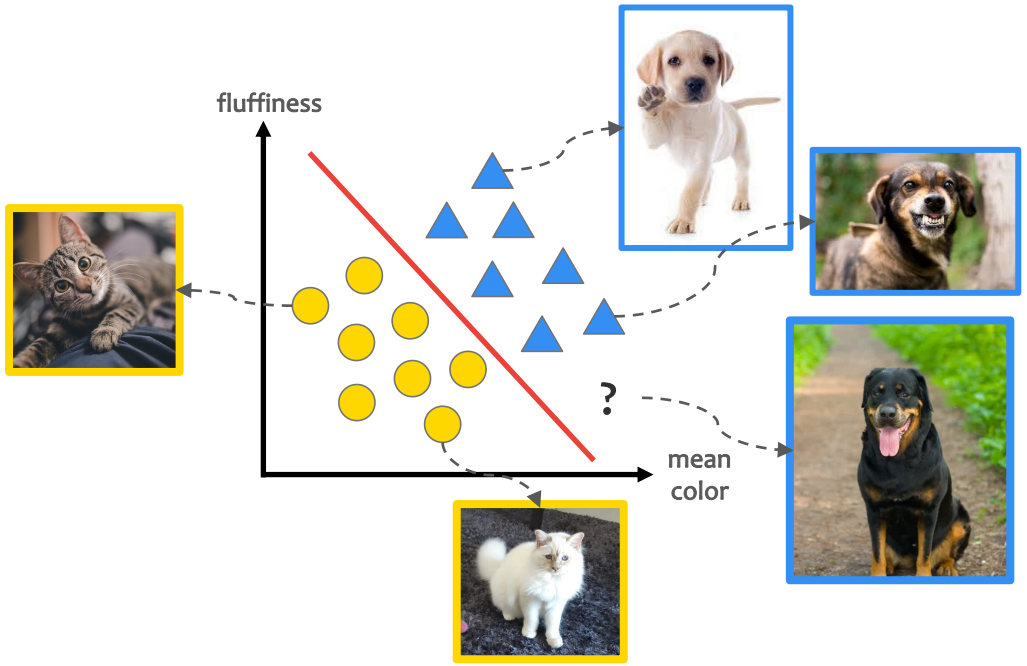
Supervised Learning

Classification

(usually) annotated by **humans:** time-consuming and costly

The **training set** includes annotations / labels / classes

Prediction of **categorical variables (labels)**

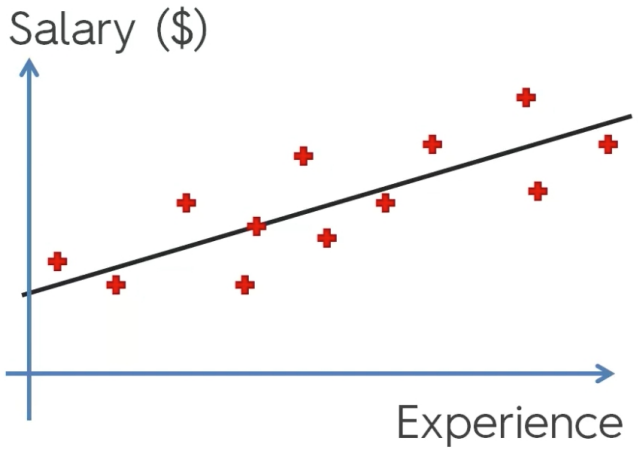


- k-Nearest Neighbors (KNN)
- Logistic Regression
- Support Vector Machines (SVM)
- Decision Trees and Random Forest
- Neural Networks
- ...

Regression

Prediction of **numeric variables.**

Experience	Salary
60	10,000.00
75	12,000.00
100	15,000.00
...	...



- Linear Regression
- Polynomial Regression
- KNN Regression
- SVM Regression
- Decision Tree Regression
- ...

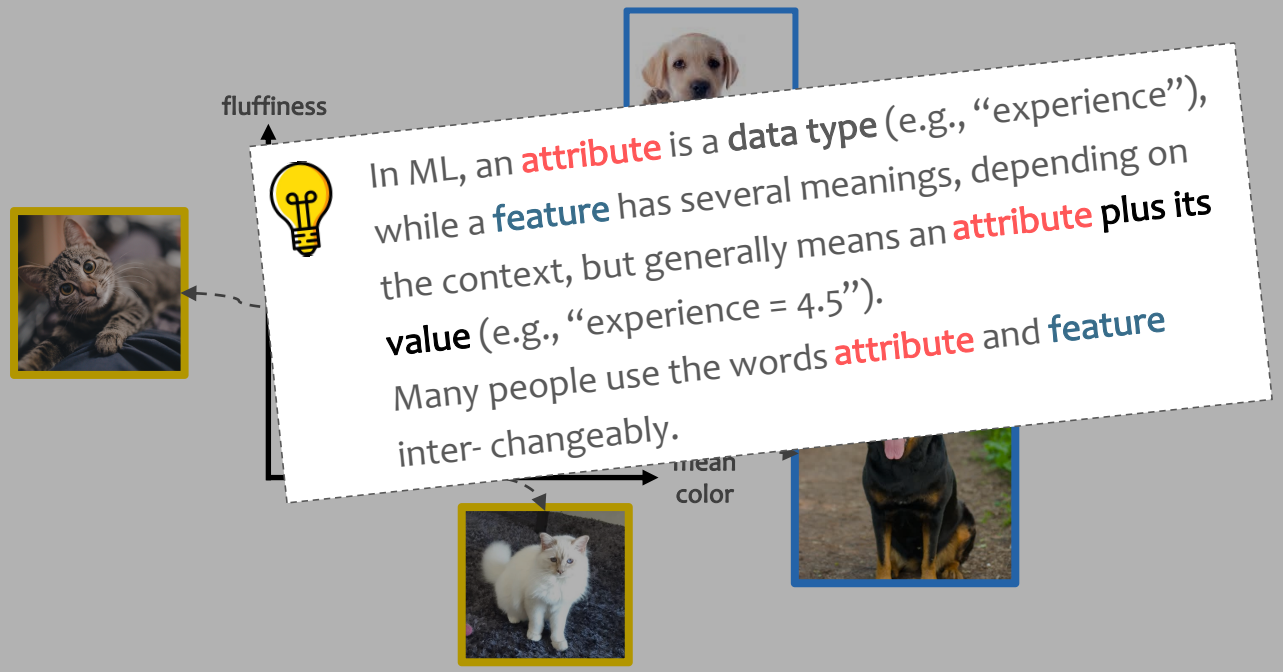
Supervised Learning

Classification

(usually) annotated by **humans:** time-consuming and costly

The **training set** includes annotations / labels / classes

Prediction of **categorical variables (labels)**



In ML, an **attribute** is a **data type** (e.g., "experience"), while a **feature** has several meanings, depending on the context, but generally means an **attribute** plus its **value** (e.g., "experience = 4.5").

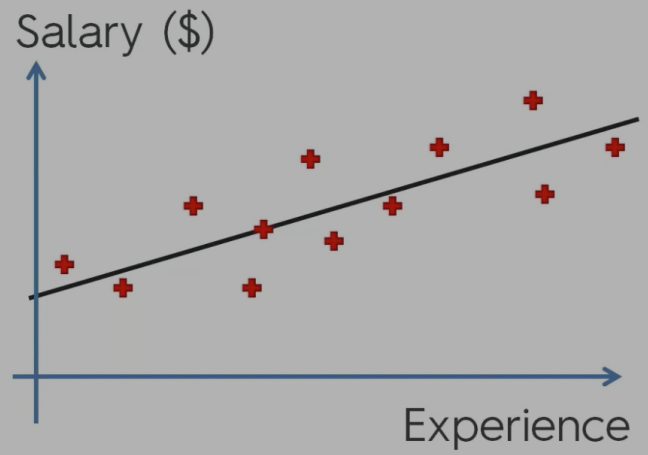
Many people use the words **attribute** and **feature** inter- changeably.

- k-Nearest Neighbors (KNN)
- Logistic Regression
- Support Vector Machines (SVM)
- Decision Trees and Random Forest
- Neural Networks
- ...

Regression

Prediction of **numeric variables.**

Experience	Salary
60	10,000.00
75	12,000.00
100	15,000.00
...	...



- Linear Regression
- Polynomial Regression
- KNN Regression
- SVM Regression
- Decision Tree Regression
- ...

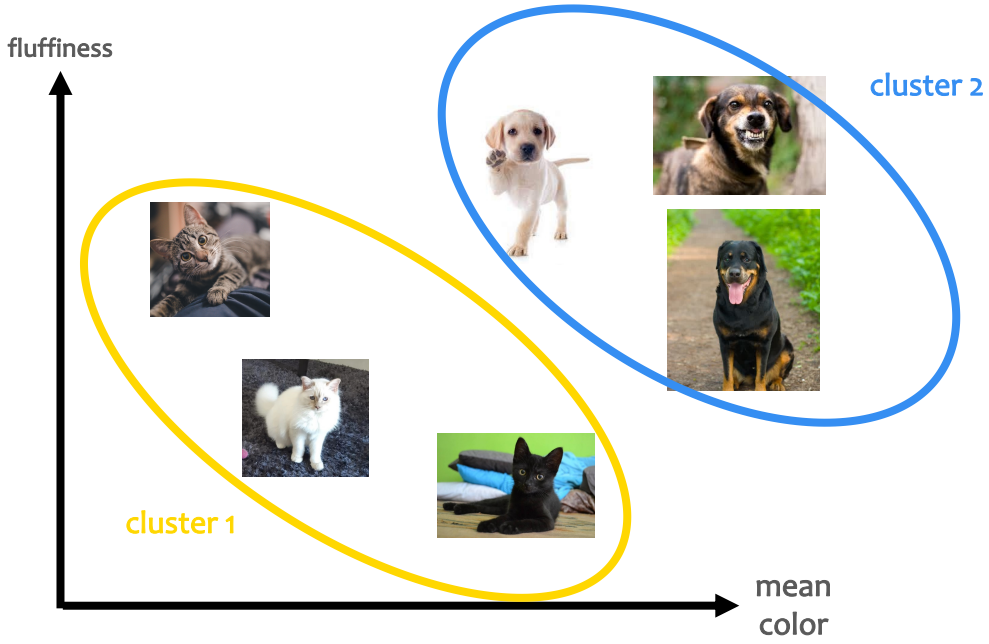
Unsupervised Learning

no supervision → no labels / annotations

Unsupervised Learning

no supervision → no labels / annotations

Clustering

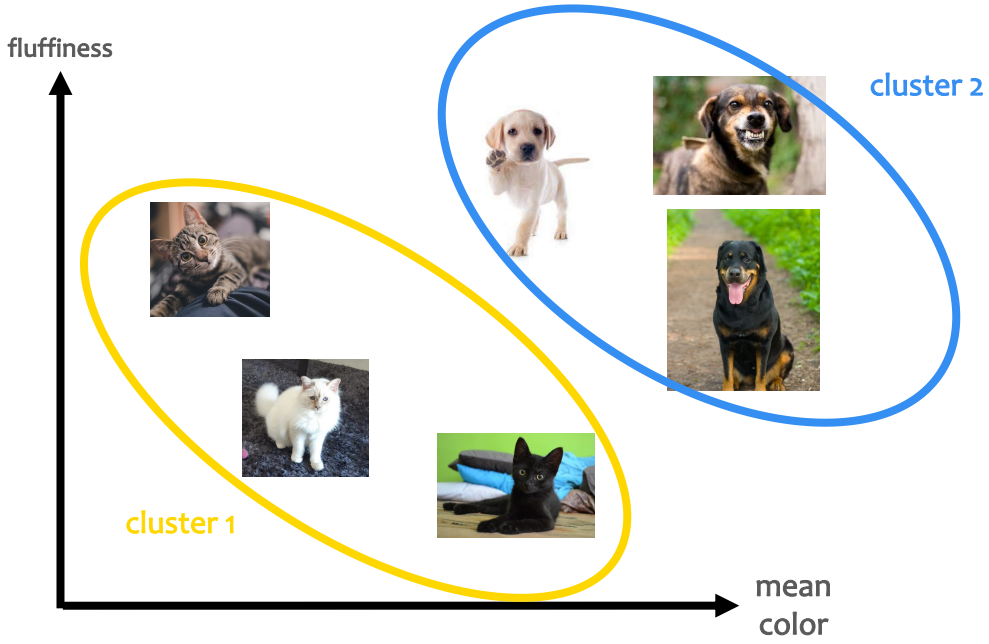


- k-Means
- DBSCAN
- Mean-shift
- OPF Clustering
- Hierarchical Clustering
- ...

Unsupervised Learning

no supervision → no labels / annotations

Clustering

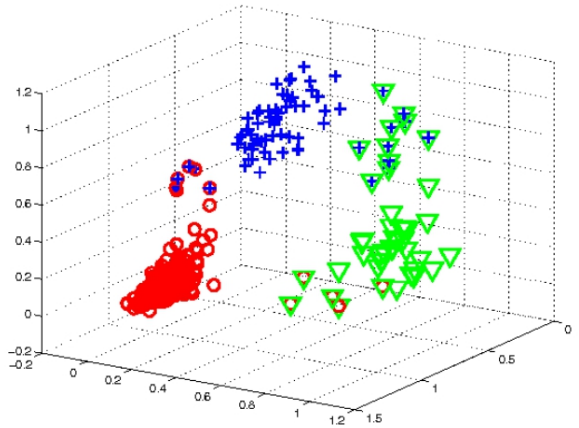


- k-Means
- DBSCAN
- Mean-shift
- OPF Clustering
- Hierarchical Clustering
- ...

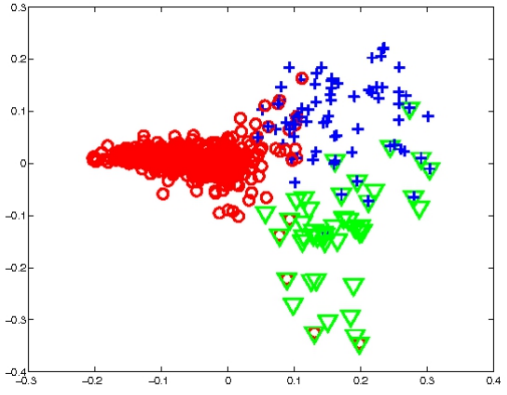
Dimensionality Reduction / Projection

Project a **high-dimensional data** on to a **lower-dimensional feature space** without losing much information.

Projection on to a **2D** or **3D** space is used to **visualize and understand** the data.



high-dimensional data



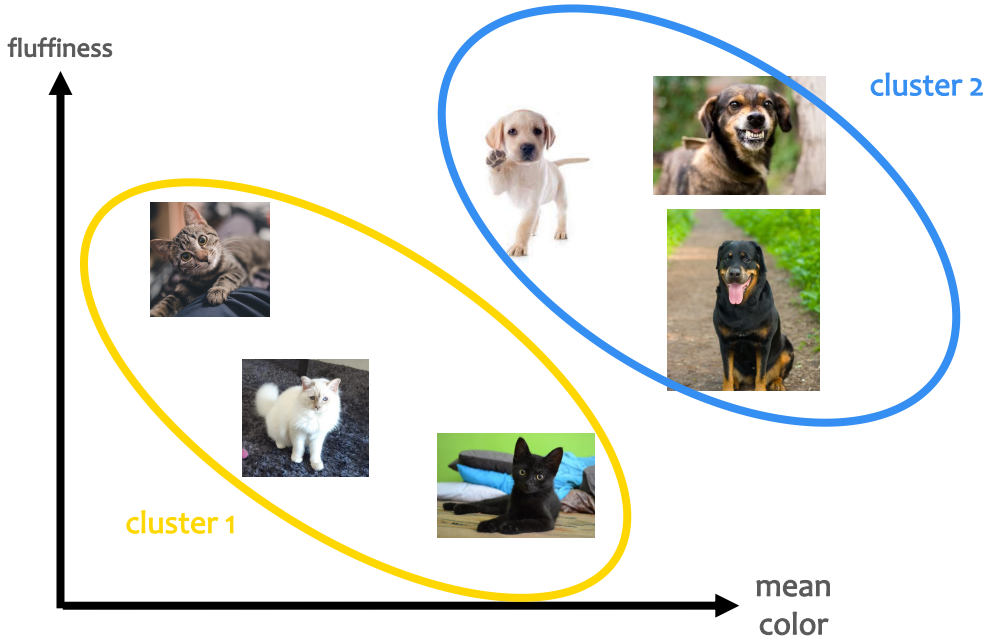
projection

- Principal Component Analysis (PCA)
- Locally Linear Embedding (LLE)
- t-SNE
- UMAP
- ...

Unsupervised Learning

no supervision → no labels / annotations

Clustering

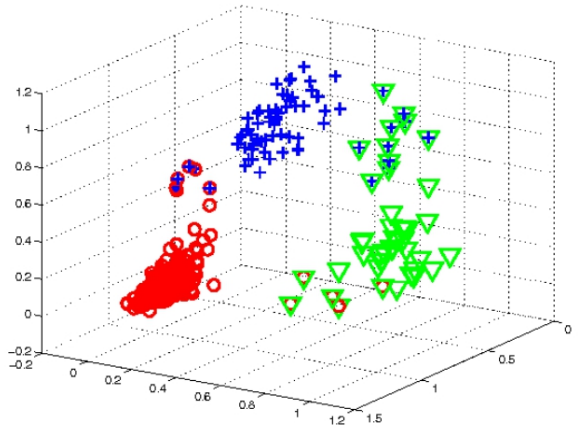


- k-Means
- DBSCAN
- Mean-shift
- OPF Clustering
- Hierarchical Clustering
- ...

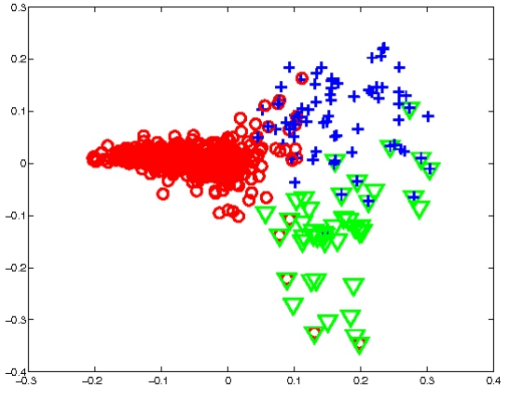
Dimensionality Reduction / Projection

Project a **high-dimensional data** on to a **lower-dimensional feature space** without losing much information.

Projection on to a **2D** or **3D** space is used to **visualize and understand** the data.



high-dimensional data



projection

There are some algorithms that use **label information** during **projection**.
E.g.: Linear Discriminant Analysis (LDA)

- ...nent Analysis (PCA)
- ...bedding (LLE)
- UMAP
- ...

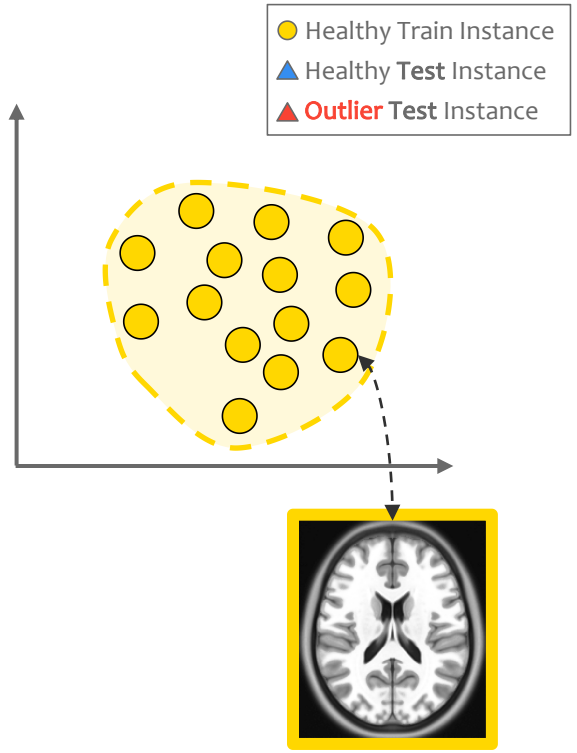
Unsupervised Learning

Anomaly/Novelty/Outlier Detection

The algorithm is fed with **training data** containing **only** a **single class** (normal, control, healthy).

New unseen instances (test) that **DOES NOT** look a **normal one** is an **outlier**, being considered as an **anomaly**.

The **training set** must be very **clean** and **representative**.



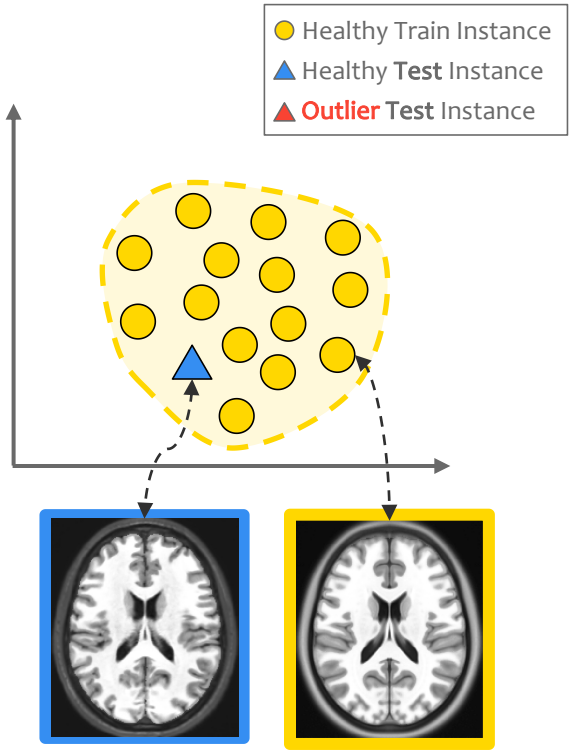
Unsupervised Learning

Anomaly/Novelty/Outlier Detection

The algorithm is fed with **training data** containing **only** a **single class** (normal, control, healthy).

New unseen instances (test) that **DOES NOT** look a **normal one** is an **outlier**, being considered as an **anomaly**.

The **training set** must be very **clean** and **representative**.



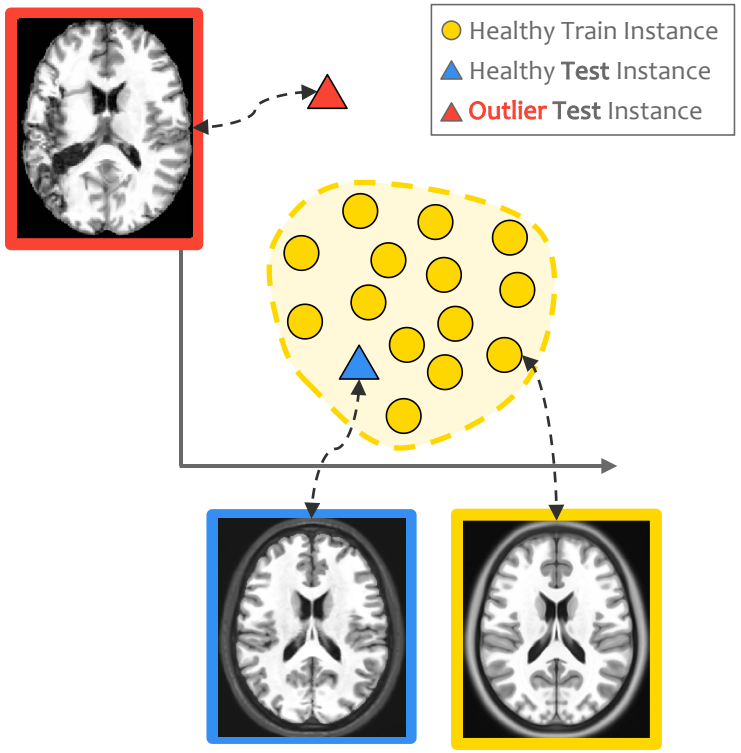
Unsupervised Learning

Anomaly/Novelty/Outlier Detection

The algorithm is fed with **training data** containing **only** a **single class** (normal, control, healthy).

New unseen instances (test) that **DOES NOT** look a **normal one** is an **outlier**, being considered as an **anomaly**.

The **training set** must be very **clean** and **representative**.



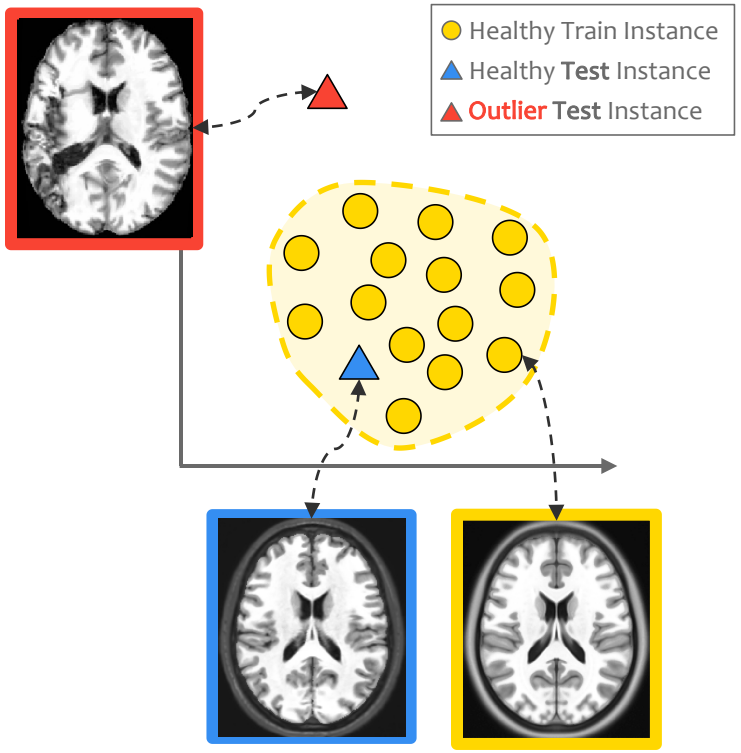
Unsupervised Learning

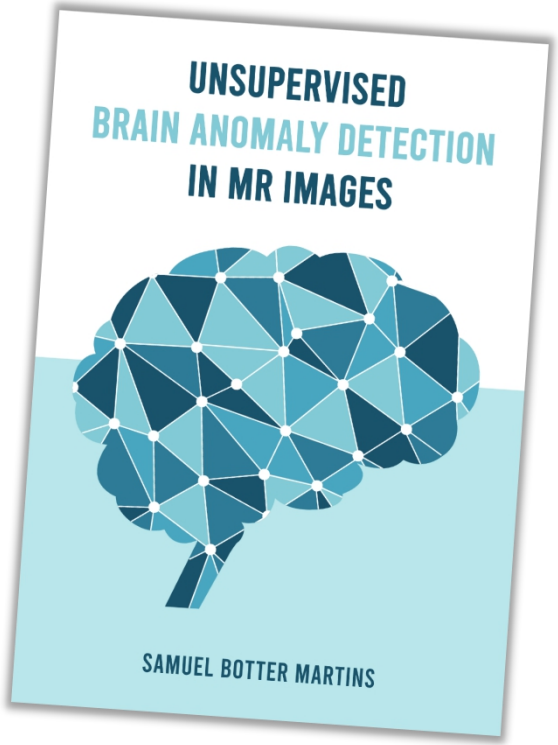
Anomaly/Novelty/Outlier Detection

The algorithm is fed with **training data** containing **only** a **single class** (normal, control, healthy).

New unseen instances (test) that **DOES NOT** look a **normal one** is an **outlier**, being considered as an **anomaly**.

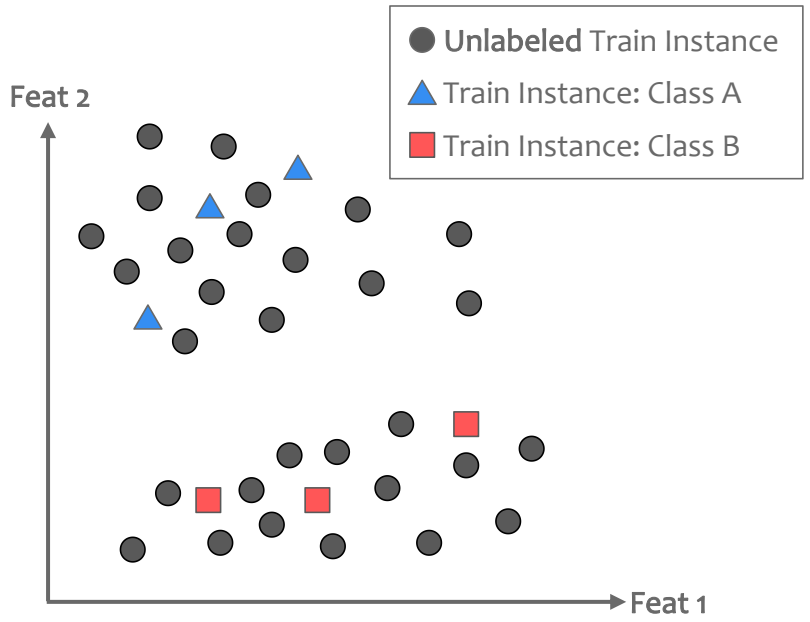
The **training set** must be very **clean** and **representative**.



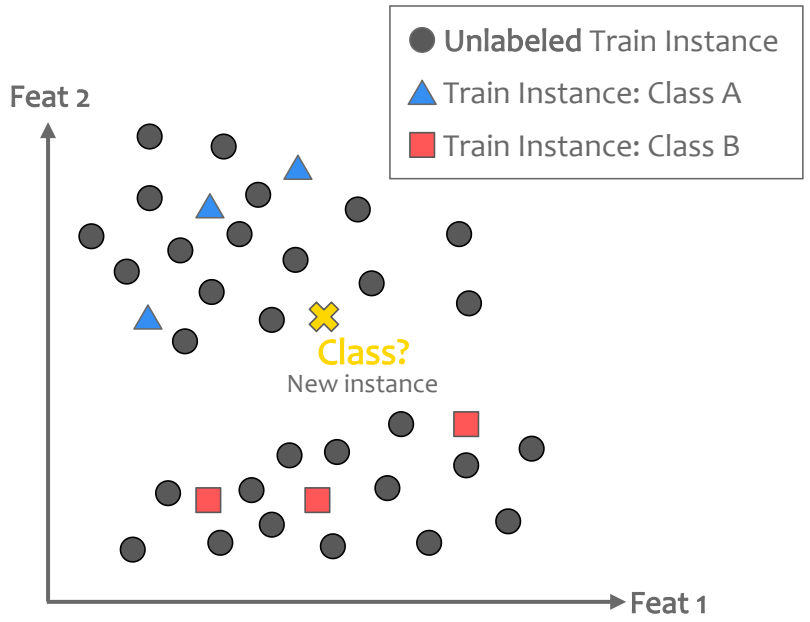


https://research.rug.nl/en/publications/unsupervised-brain-anomaly-detection-in-mr-images

Semisupervised Learning

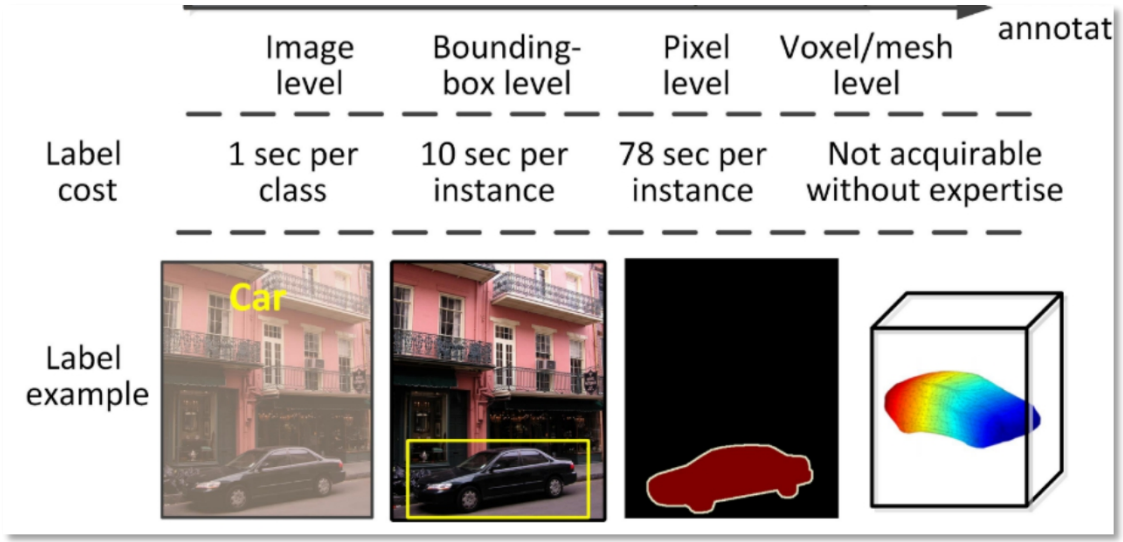


Semisupervised Learning



Weakly Supervised Learning

Supervision with **noisy labels**.

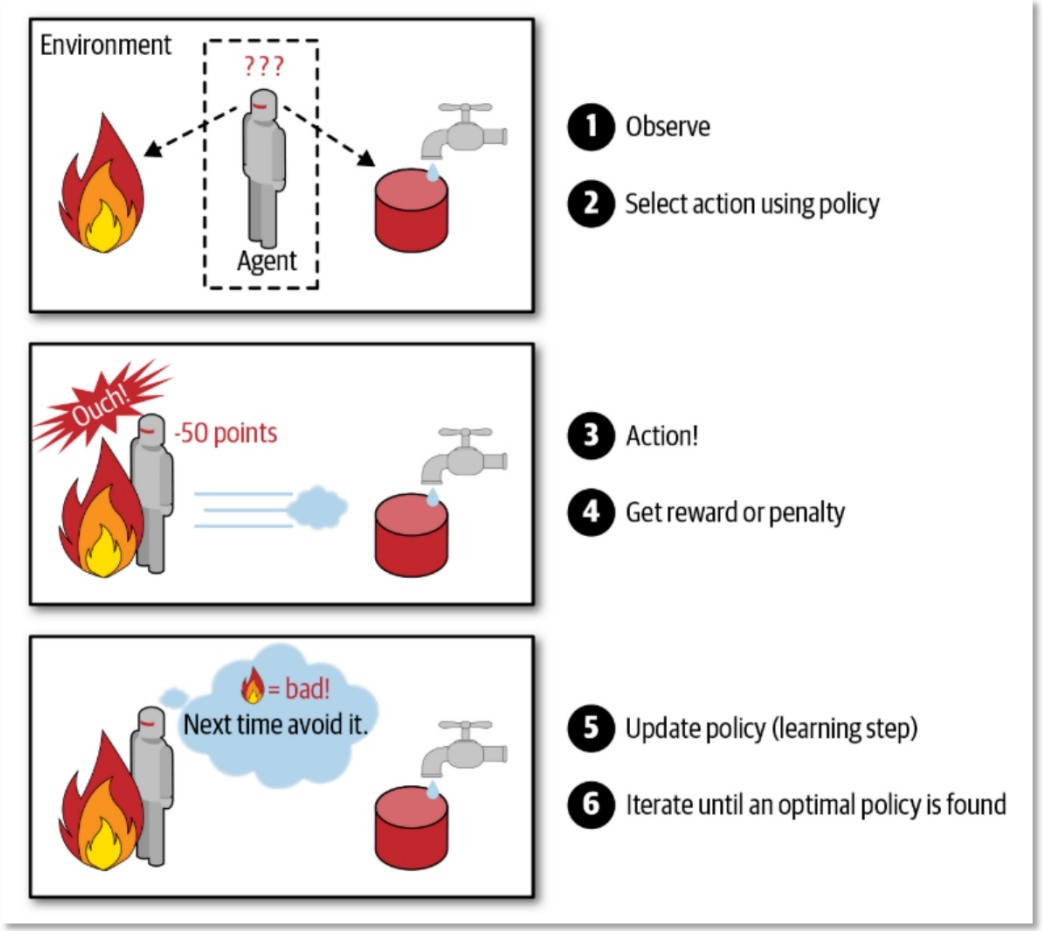


Zhang, Dingwen, et al. **"Weakly Supervised Object Localization and Detection: A Survey."** IEEE transactions on pattern analysis and machine intelligence (2021).

Figure 1. Weakly supervised object localization problem setting. Given a collection of training images with image-level annotations, our goal is to train object detectors for localizing objects in unseen images.

Li, Dong, et al. **"Weakly supervised object localization with progressive domain adaptation."** Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

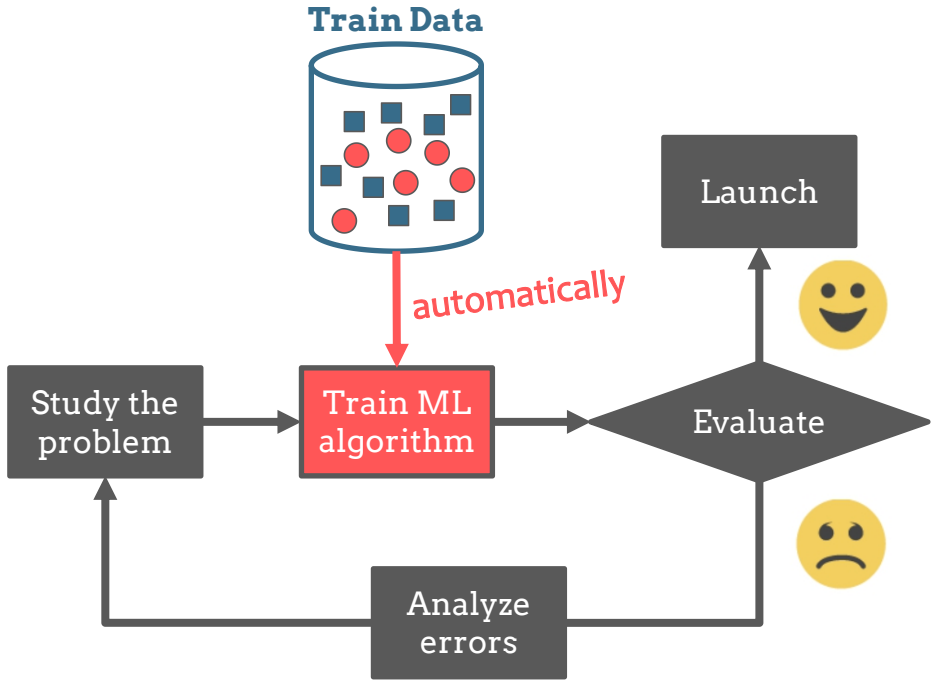
Reinforcement Learning



Géron, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.** O'Reilly Media, 2019.

Batch Learning

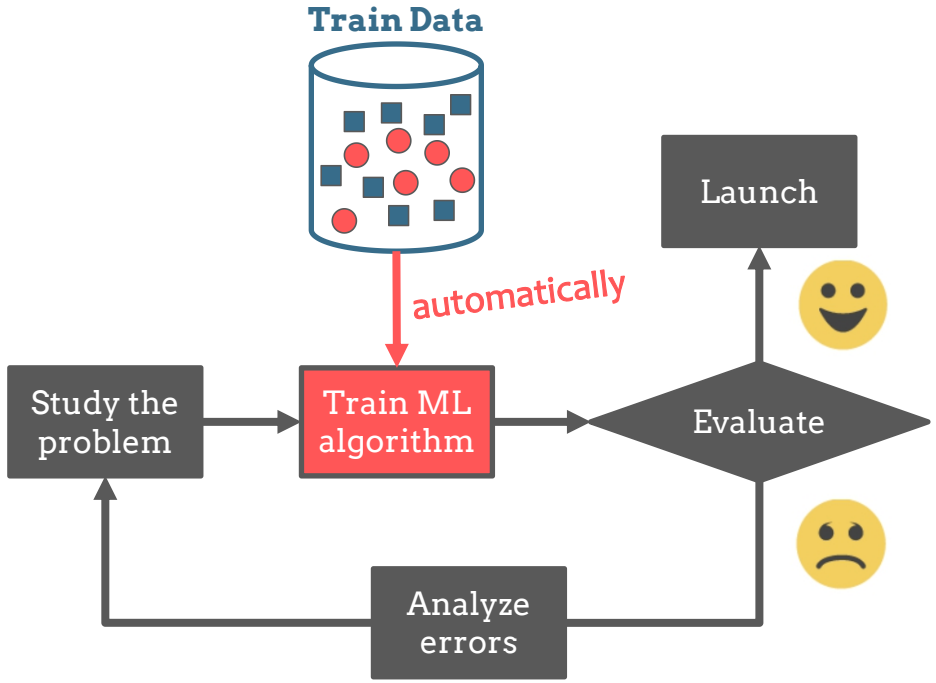
- The system use **all avaiable data** for *offline* training once and launched into production **without learning anymore** → no incremental updates.
- Time-consuming and computational expensive.
- Knowing new data requires to **train a new version** of the system **from scracth.**



Online Learning

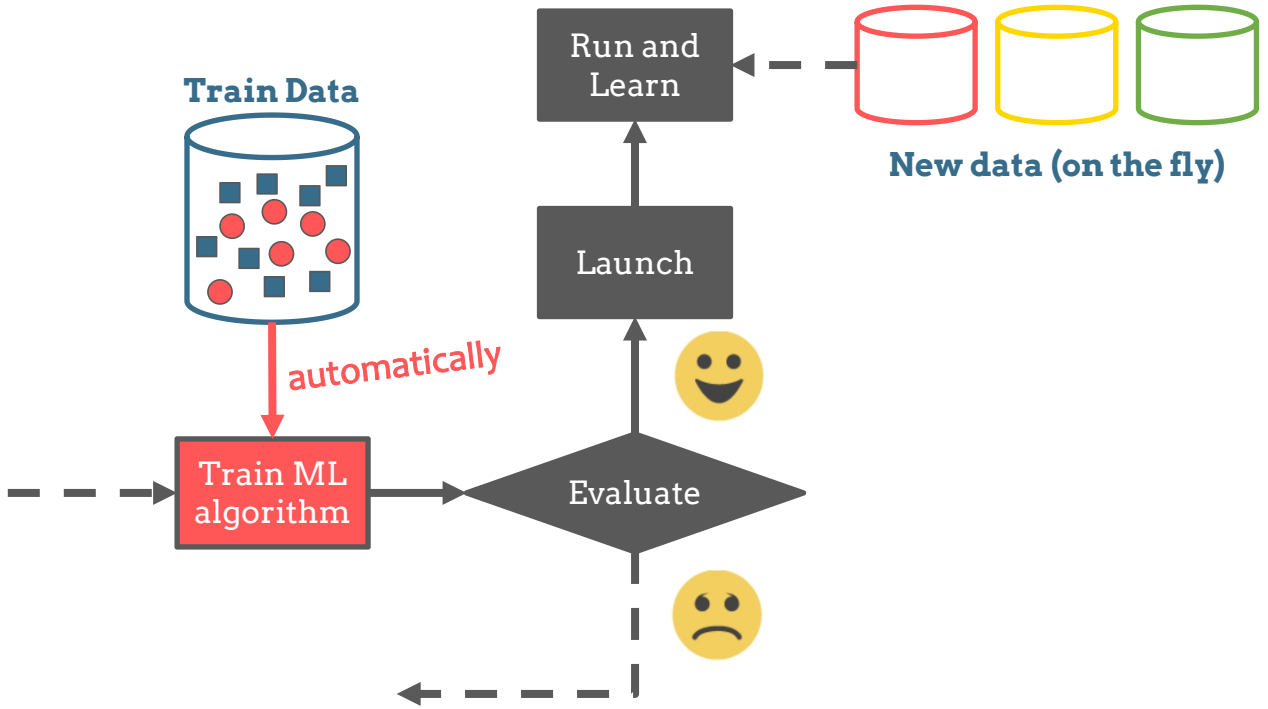
Batch Learning

- The system use **all avaiable data** for *offline* training once and launched into production **without learning anymore** → no incremental updates.
- Time-consuming and computational expensive.
- Knowing new data requires to **train a new version** of the system **from scracth**.



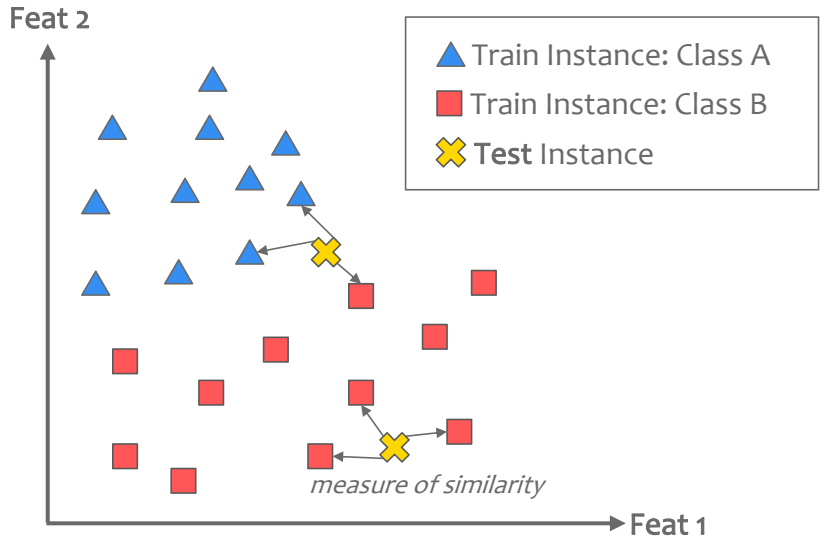
Online Learning

- Train the system **incrementally** by feeding it data instances sequentially, either individually or in small groups called mini-batches;
- Each learning step is **fast and cheap**, so the system can learn about new data **on the fly**, as it arrives.



Instance-based learning

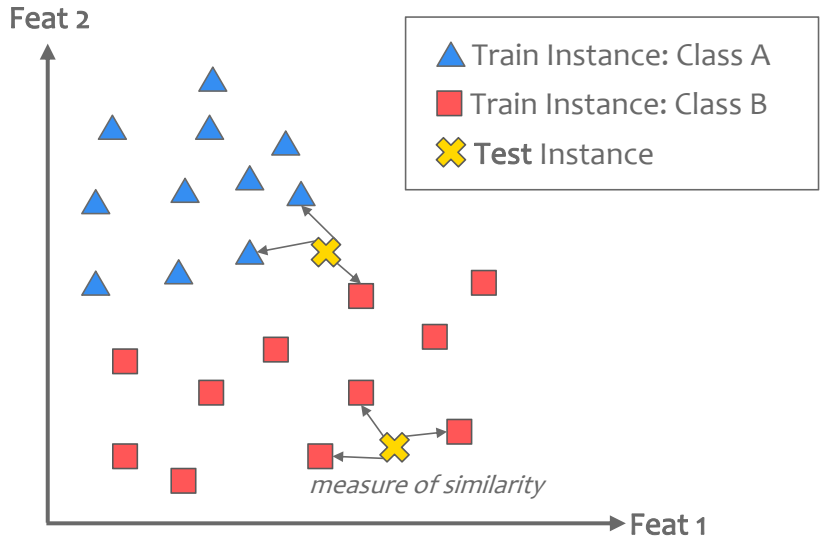
- The system learns the examples **by heart**, then generalizes to **new cases** by using **a similarity measure** to compare them **to the learned examples** (or a subset of them) → **"there is no training"**.



Model-based learning

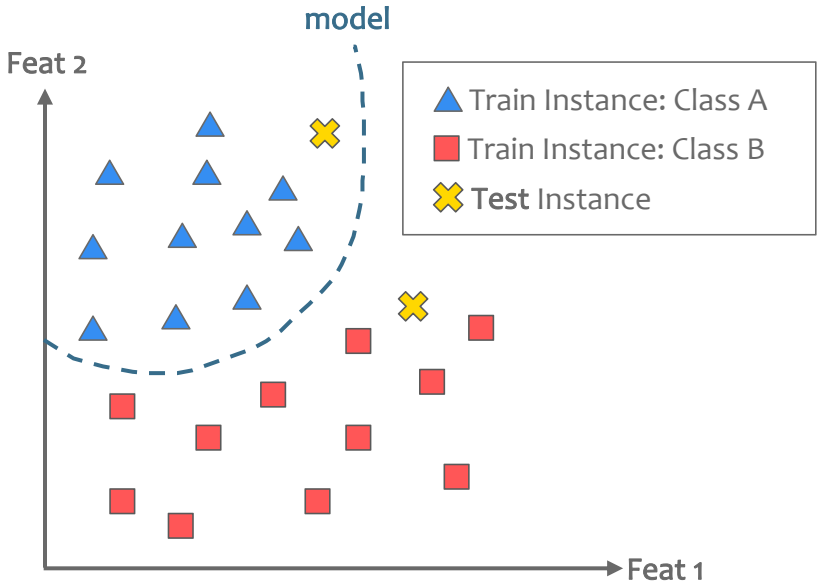
Instance-based learning

- The system learns the examples by heart, then generalizes to **new cases** by using **a similarity measure** to compare them **to the learned examples** (or a subset of them) → "there is no training".



Model-based learning

- The system **trains a** model (function, neural network, ...) from a training data to make *predictions.*

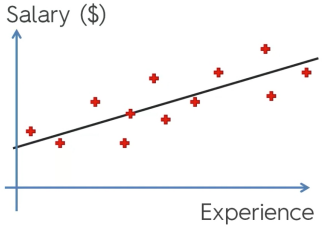


Main Challenges of Machine Learning

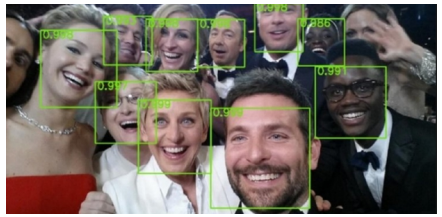
Insufficient Quantity of Train Data

Insufficient Quantity of Train Data

ML algorithms requires a **considerable quantity of train data** to work properly.



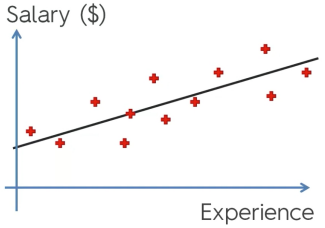
simple problems
(**thousands** of examples)



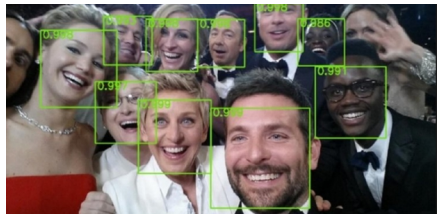
complex problems
(**millions** of examples)

Insufficient Quantity of Train Data

ML algorithms requires a **considerable quantity of train data** to work properly.



simple problems
(**thousands** of examples)

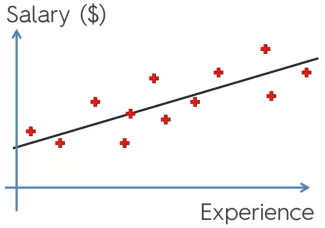


complex problems
(**millions** of examples)

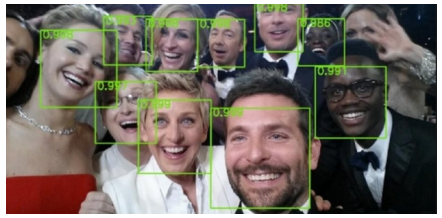
Some authors consider that **"data matters more than algorithms for complex problems".**

Insufficient Quantity of Train Data

ML algorithms requires a **considerable quantity of train data** to work properly.



simple problems
(**thousands** of examples)



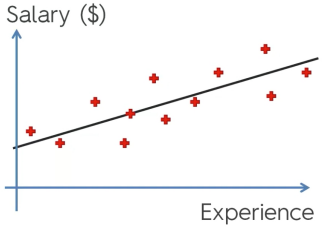
complex problems
(**millions** of examples)

Some authors consider that **"data matters more than algorithms for complex problems".**

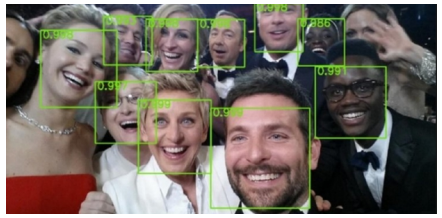
Poor-Quality Data

Insufficient Quantity of Train Data

ML algorithms requires a **considerable quantity of train data** to work properly.



simple problems
(**thousands** of examples)



complex problems
(**millions** of examples)

Some authors consider that **"data matters more than algorithms for complex problems".**

Poor-Quality Data

- Noise
- Missing values
- Outliers

flawed data

+

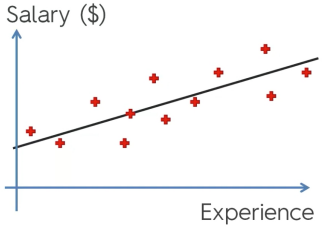
fancy
ML algorithms

flawed
results/conclusions

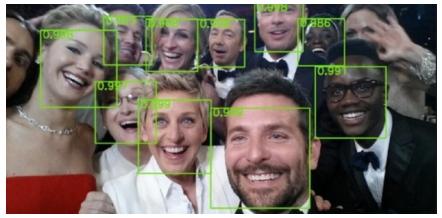
Garbage In → **Garbage Out**

Insufficient Quantity of Train Data

ML algorithms requires a **considerable quantity of train data** to work properly.



simple problems
(**thousands** of examples)



complex problems
(**millions** of examples)

> Some authors consider that **"data matters more than algorithms for complex problems".**

Nonrepresentative Train Data

Poor-Quality Data

- Noise
- Missing values
- Outliers

flawed data

+

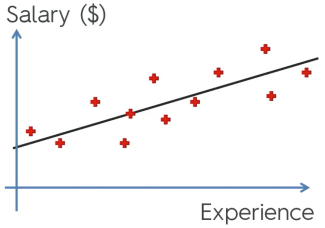
fancy
ML algorithms

flawed
results/conclusions

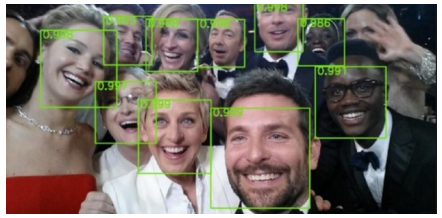
Garbage In → **Garbage Out**

Insufficient Quantity of Train Data

ML algorithms requires a **considerable quantity of train data** to work properly.



simple problems
(**thousands** of examples)



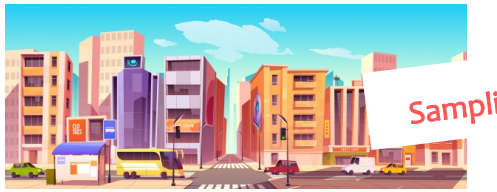
complex problems
(**millions** of examples)

Some authors consider that **"data matters more than algorithms for complex problems".**

Nonrepresentative Train Data

It is crucial to use a **training set** that is **representative** of the cases you want to **generalize to.**

E.g.: Predict the risk of infant mortality of a **city population**.



fancy neighborhood

Sampling bias

poor neighborhood

Poor-Quality Data

- Noise
- Missing values
- Outliers

flawed data

+

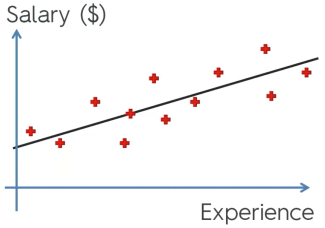
fancy ML algorithms

flawed results/conclusions

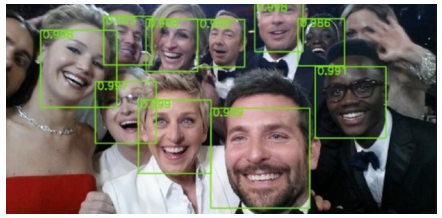
Garbage In → **Garbage Out**

Insufficient Quantity of Train Data

ML algorithms requires a **considerable quantity of train data** to work properly.



simple problems
(**thousands** of examples)



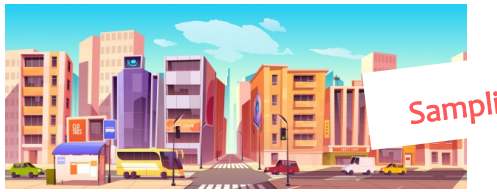
complex problems
(**millions** of examples)

Some authors consider that **"data matters more than algorithms for complex problems".**

Nonrepresentative Train Data

It is crucial to use a **training set** that is **representative** of the cases you want to **generalize to.**

E.g.: Predict the risk of infant mortality of a **city population**.



fancy neighborhood

Sampling bias

poor neighborhood

Poor-Quality Data

- Noise
- Missing values
- Outliers

flawed data

+

fancy ML algorithms

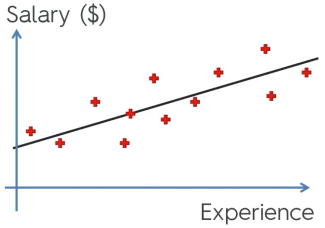
flawed results/conclusions

Garbage In → **Garbage Out**

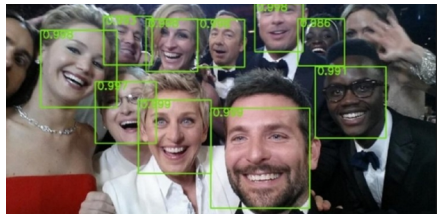
Irrelevant Features

Insufficient Quantity of Train Data

ML algorithms requires a **considerable quantity of train data** to work properly.



simple problems
(**thousands** of examples)



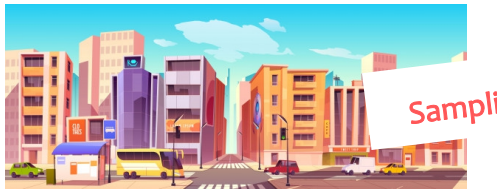
complex problems
(**millions** of examples)

Some authors consider that **"data matters more than algorithms for complex problems".**

Nonrepresentative Train Data

It is crucial to use a **training set** that is **representative** of the cases you want to **generalize to**.

E.g.: Predict the risk of infant mortality of a **city population**.



fancy neighborhood

Sampling bias

poor neighborhood

Poor-Quality Data

- Noise
- Missing values
- Outliers

flawed data

+

fancy ML algorithms

flawed results/conclusions

Garbage In → **Garbage Out**

Irrelevant Features

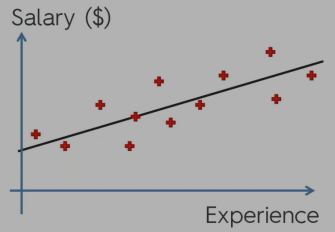
E.g.: Predict employees' salaries.

Experience	Education	Soccer Team	Favorite Color	Salary
60	Undergraduation	Barcelona	Blue	10,000.00
75	MSc	Real Madrid	Red	12,000.00
100	PhD	Botafogo	Black	15,000.00
...	...	...	..	...

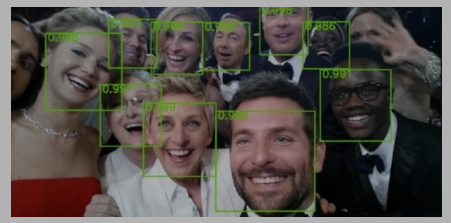
Feature engineering = feature selection + feature extraction

Insufficient Quantity of Train Data

ML algorithms requires a **considerable quantity of train data** to work properly.



simple problems
(**thousands** of examples)



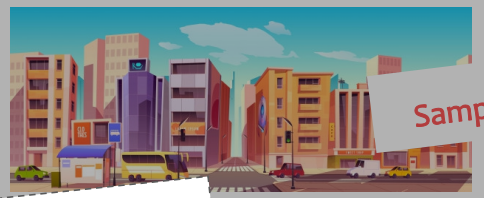
complex problems
(**millions** of examples)

Some authors consider that **"data matters more than algorithms for complex problems".**

Nonrepresentative Train Data

It is crucial to use a **training set** that is **representative** of the cases you want to **generalize to**.

E.g.: Predict the risk of infant mortality of a **city population**.



ighborhood

Sampling bias

poor neighborhood

Other challenges of ML, such as **overfitting** and **underfitting**, will be discussed later in this course.

Poor-Quality Data

- Noise
- Missing values
- Outliers

flawed data

+

fancy ML algorithms

flawed results/conclusions

Garbage In → **Garbage Out**

Irrelevant Features

E.g.: Predict employees' salaries.

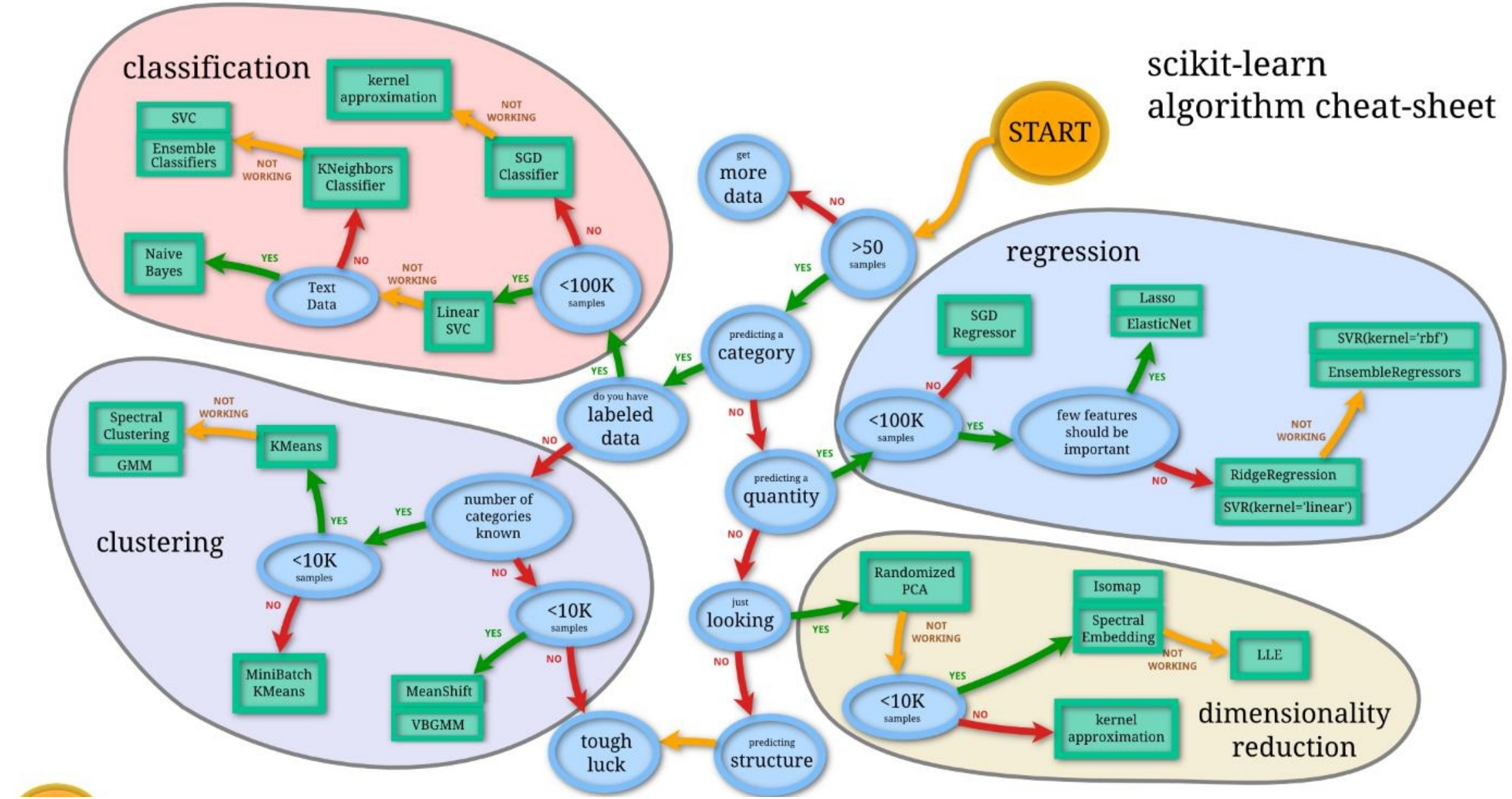
Experience	Education	Soccer Team	Favorite Color	Salary
60	Undergraduation	Barcelona	Blue	10,000.00
75	MSc	Real Madrid	Red	12,000.00
100	PhD	Botafogo	Black	15,000.00
...	...	...	..	...

Feature engineering = **feature selection** + feature extraction

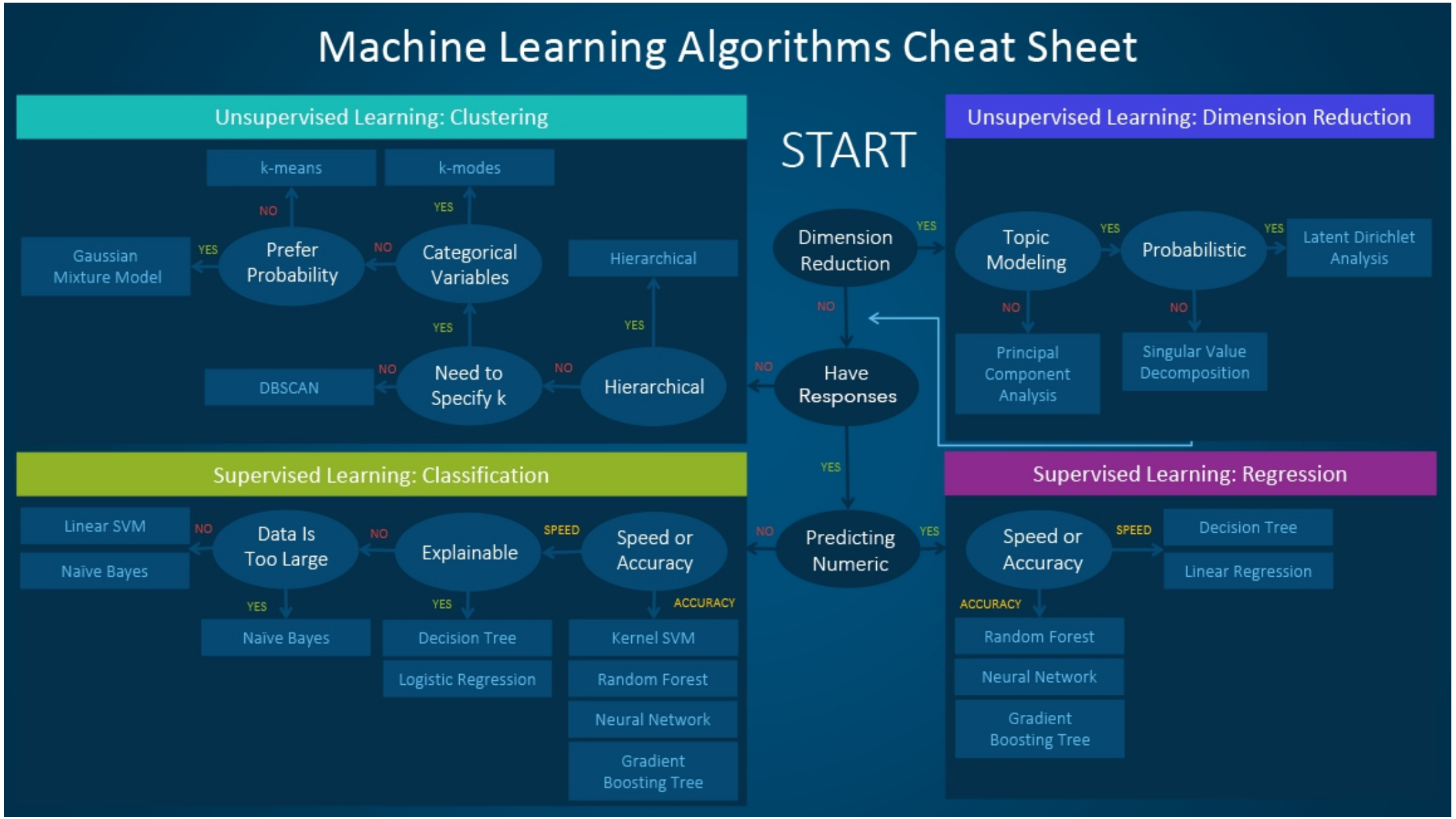
No Free Lunch Theorem

- **There is no** model that is ***a priori*** guaranteed to work better for some problems/datasets;
- The only way to know for sure which model is best is **to evaluate them all**.
- Since **this is not possible,** in practice you make some **reasonable assumptions about the data** and **evaluate only** a **few reasonable models.**

Wolpert, David H. **"The lack of a priori distinctions between learning algorithms."** Neural computation 8.7 (1996): 1341-1390.



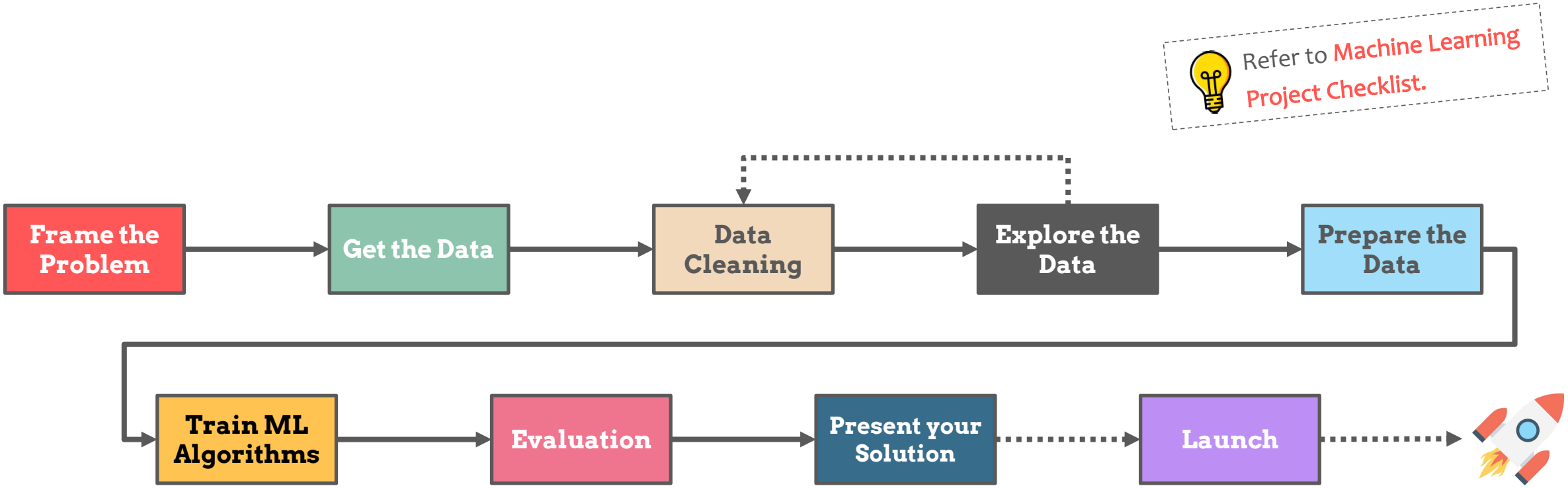
https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html



https://blogs.sas.com/content/subconsciousmusings/2020/12/09/machine-learning-algorithm-use/

The Machine Learning Pipeline

An adapted version from: Géron, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.** O'Reilly Media, 2019.


Refer to Machine Learning Project Checklist.
Frame the Problem
Get the Data
Data Cleaning
Explore the Data
Prepare the Data
Train ML Algorithms
Evaluation
Present your Solution
Launch

Aprendizado de Máquina e Reconhecimento de Padrões 2021.2

The Machine Learning Landscape

Prof. Samuel Martins (Samuka)
samuel.martins@ifsp.edu.br

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
São Paulo
Campus Campinas