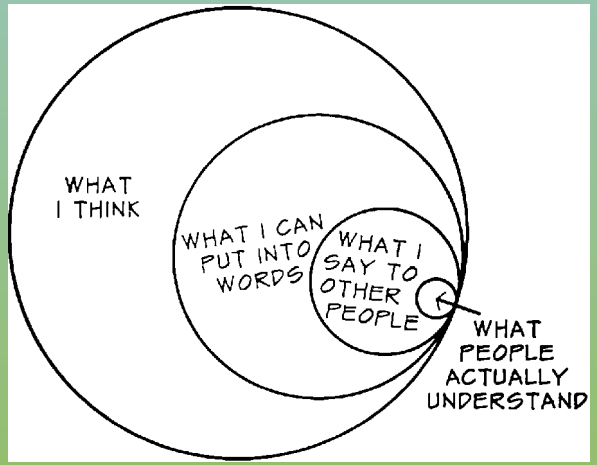
101 Machine Learning



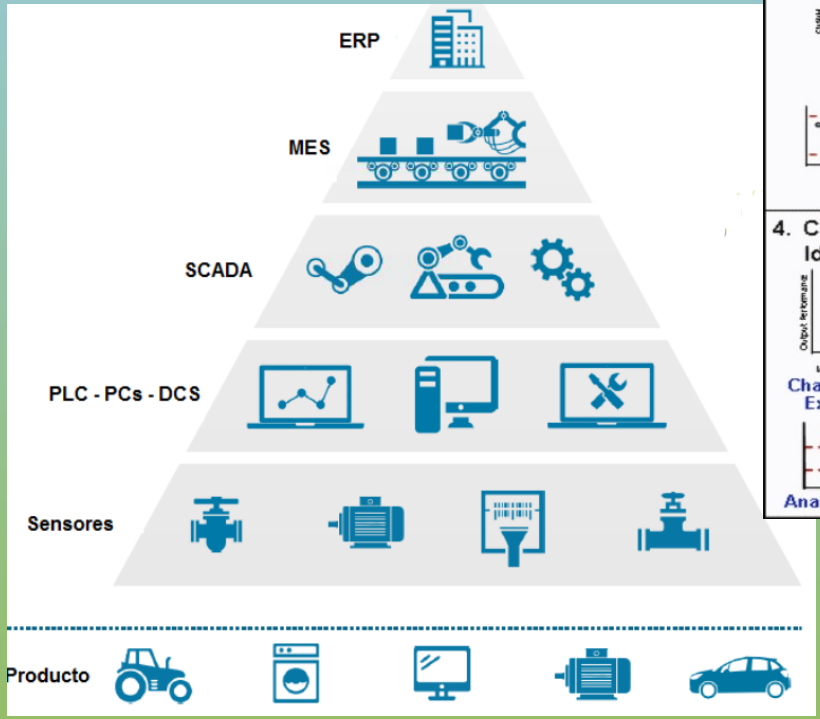
I. Evolution for computing environments: volume of data.

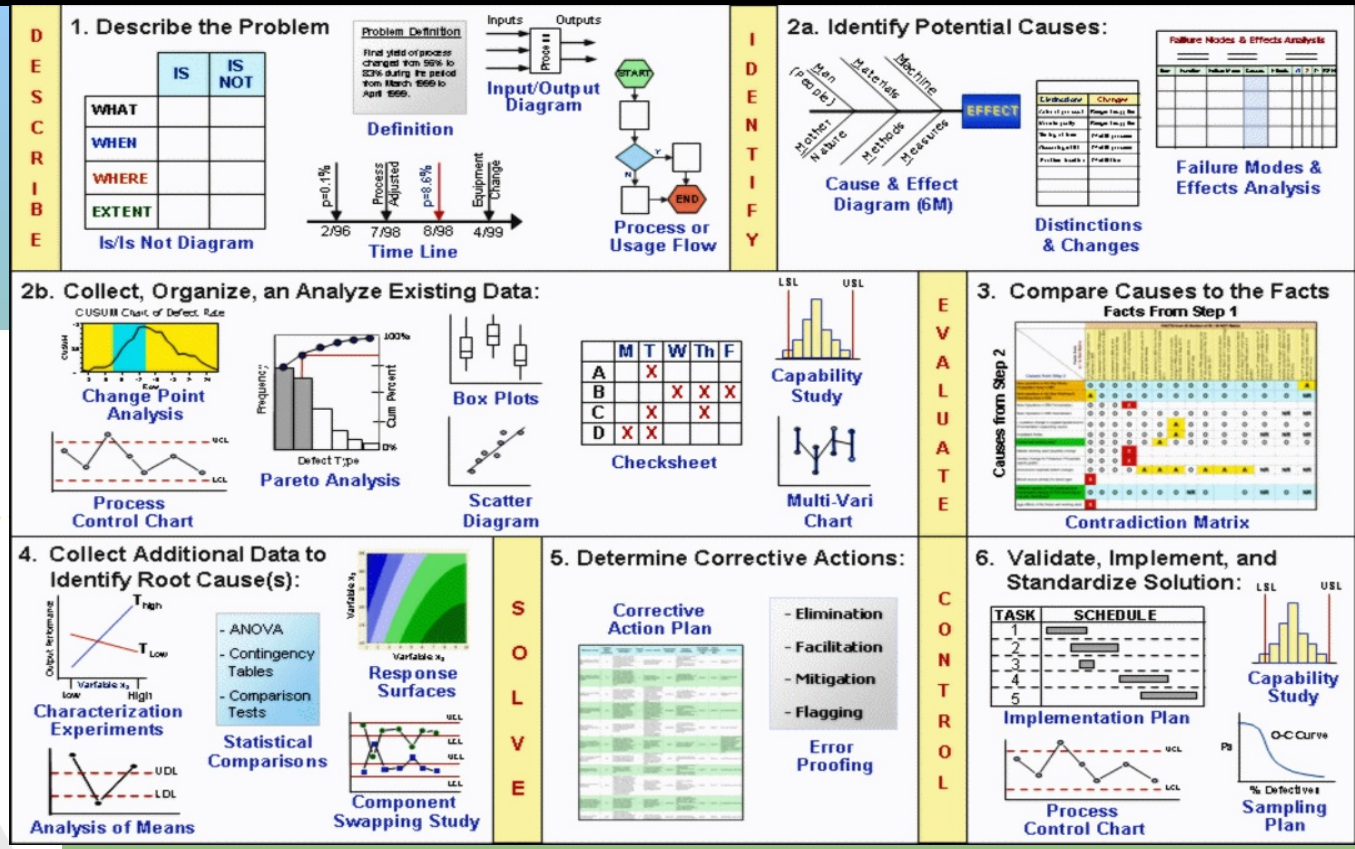
II. Choice with ML: types, phases and functionalities.

III. Checking how predictions were measured and compared.

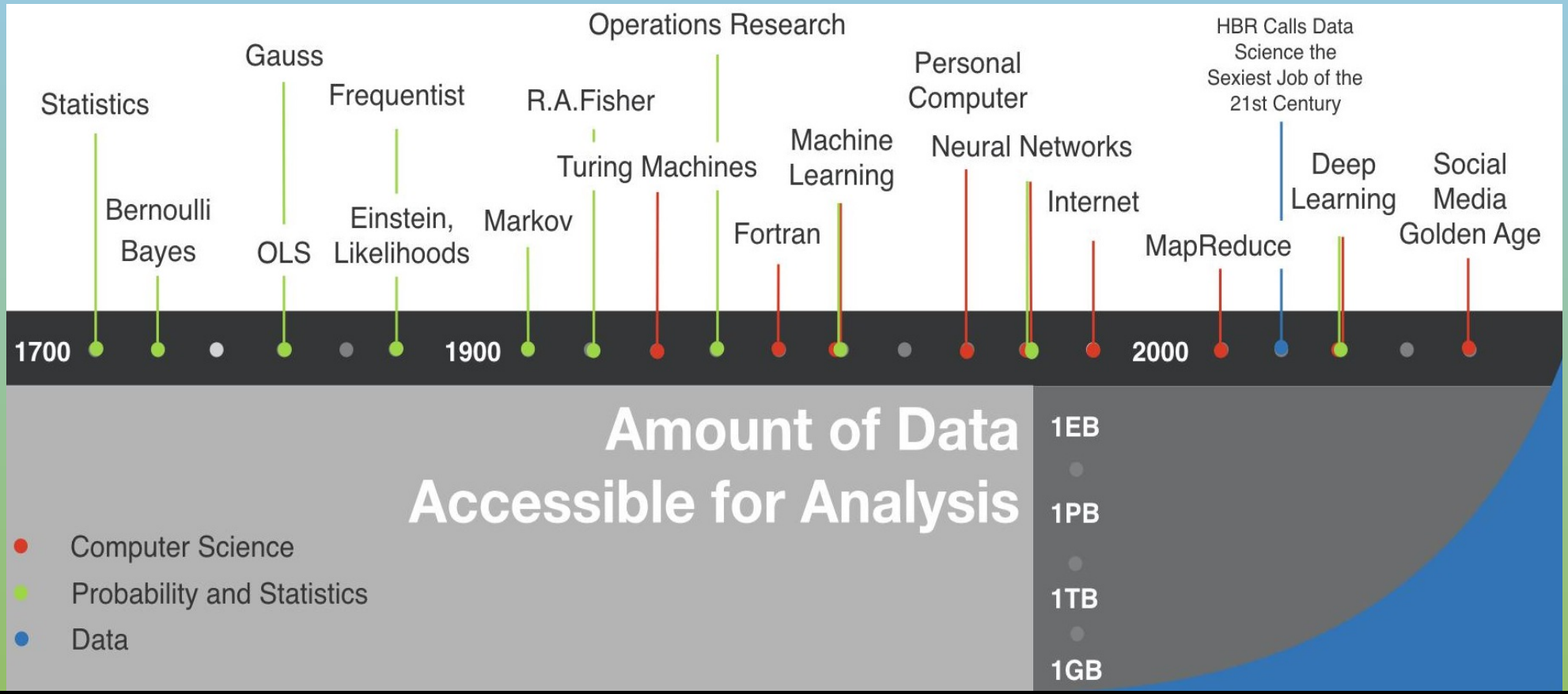
Previous environment

- Isolated processes/systems.
- Specialization: departments/ operations
- Distributed/isolated knowledge





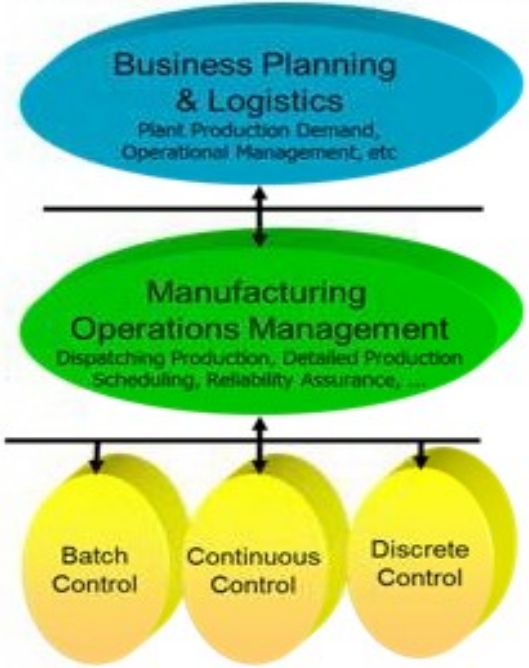
Resolution techniques of issues: not complicated operations, but it implies many problems of adaptation of referents and specializations towards an analysis.


Statistics
Bernoulli
Bayes
Gauss
OLS
Frequentist
Einstein,
Likelihoods
Markov
R.A.Fisher
Turing Machines
Operations Research
Fortran
Machine
Learning
Personal
Computer
Neural Networks
Internet
HBR Calls Data
Science the
Sexiest Job of the
21st Century
MapReduce
Deep
Learning
Social
Media
Golden Age
1700
1900
2000
Amount of Data
Accessible for Analysis
1EB
1PB
1TB
1GB
Computer Science
Probability and Statistics
Data

Smart Manufacturing Challenge

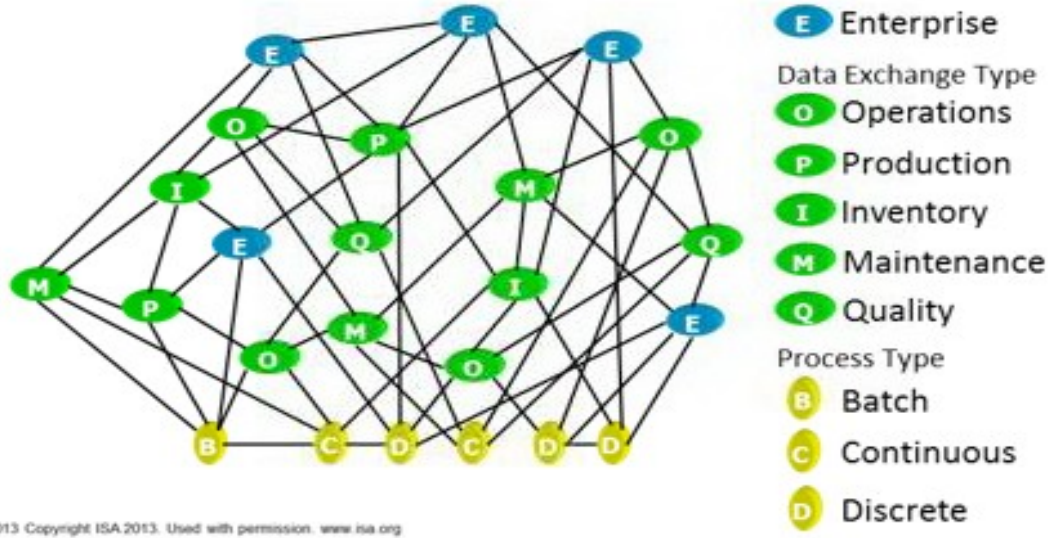
From here

Vertical Interoperability



To here

Smart Enterprises, Sites, Work Centers.. Devices



Design thinking involves observation to discover unmet needs within the context and constraints of a particular situation. It frames the opportunity and scope of innovation, generating creative ideas, testing and refining solutions. It creates a repeatable and scalable process for innovation.

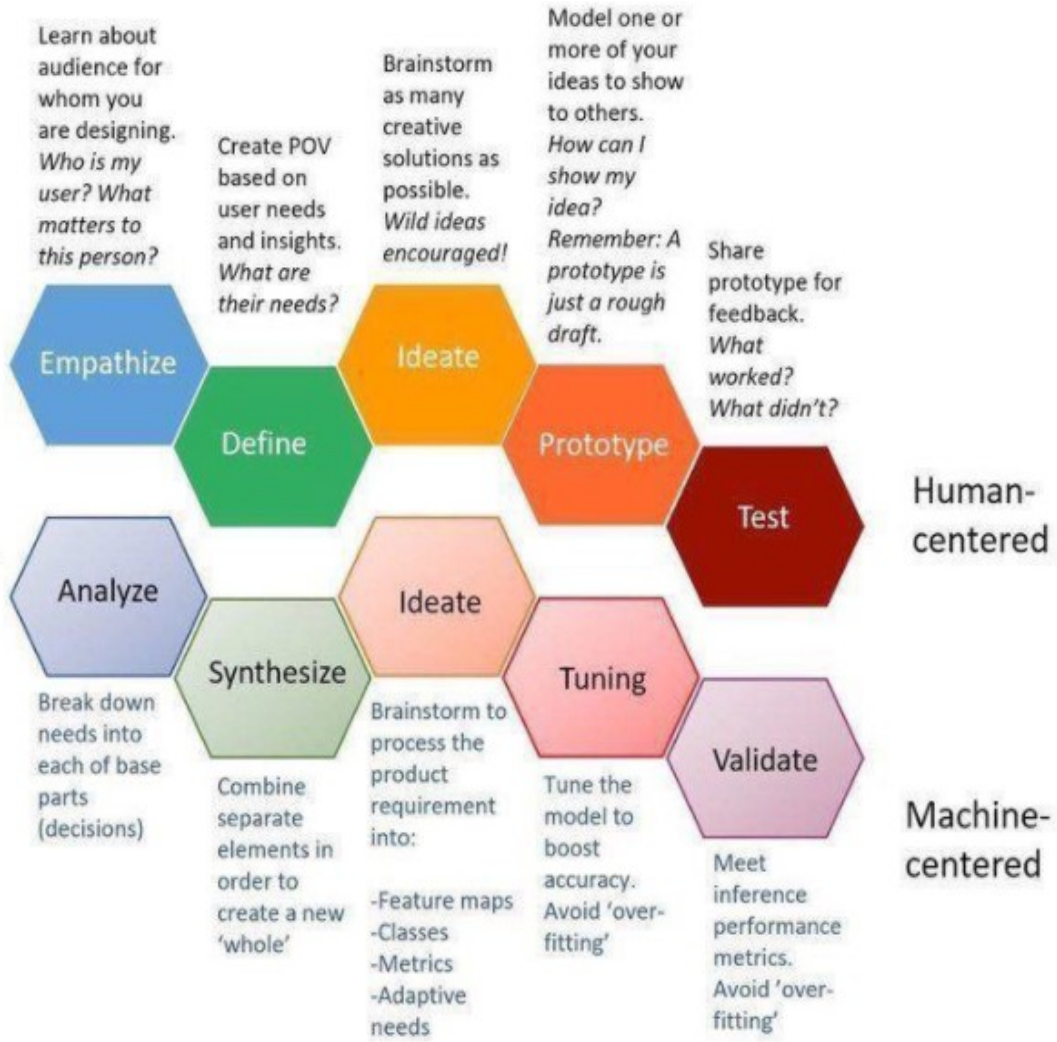
Design Thinking

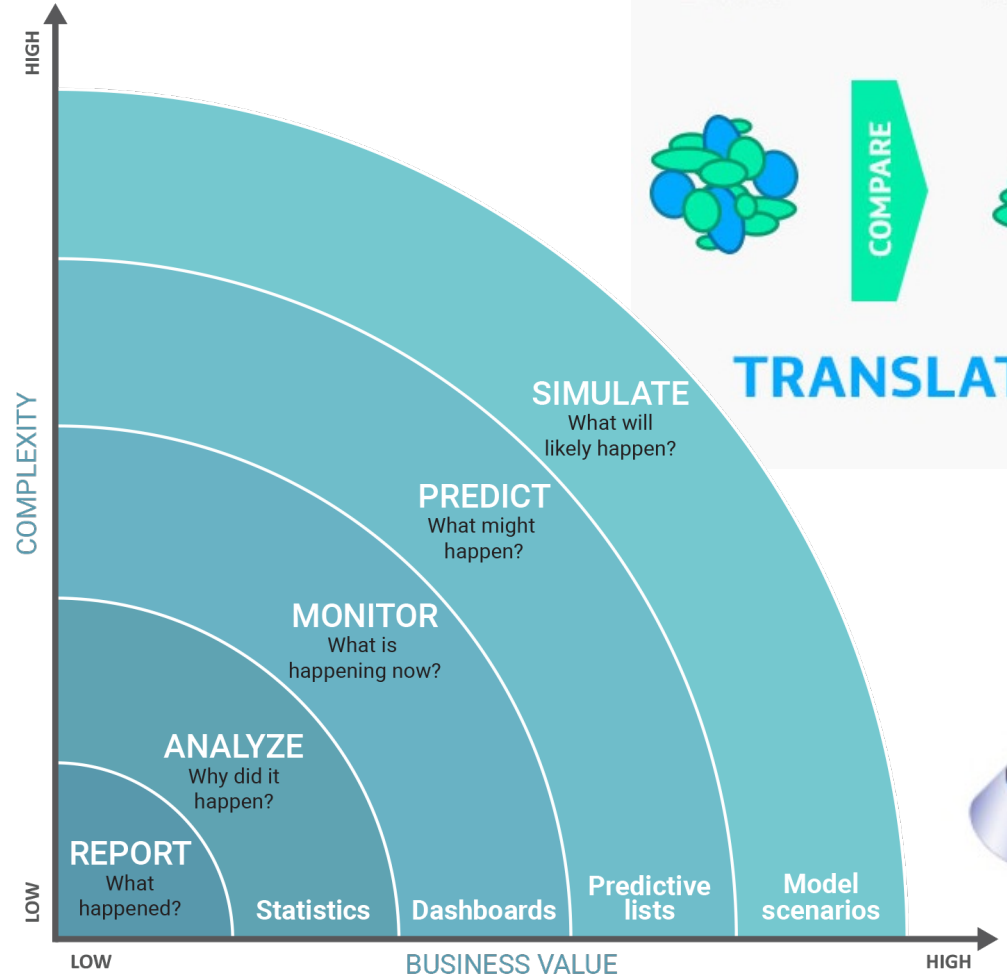
designs products this way…

refines, tunes and predicts this way…

Machine Learning

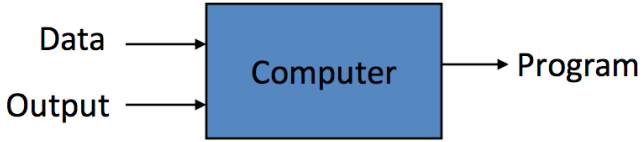
Machine learning is a method of data analysis that automates analytical model building. Using algorithms that iteratively learn from data, *machine learning* allows computers to find hidden insights without being explicitly programmed where to look.

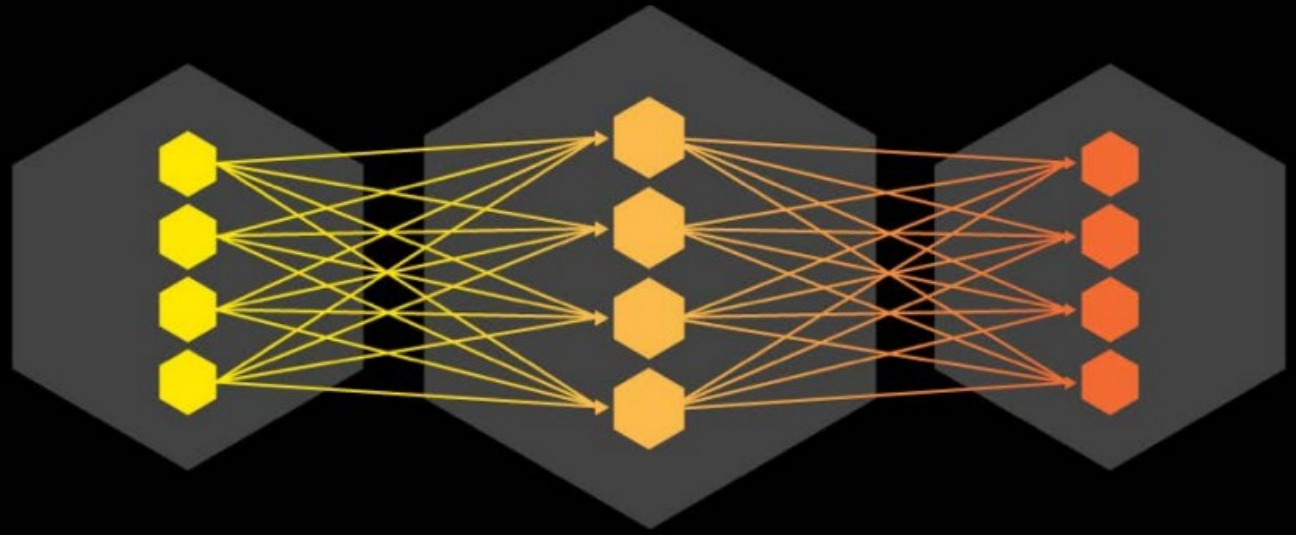




Traditional Programming

Machine Learning

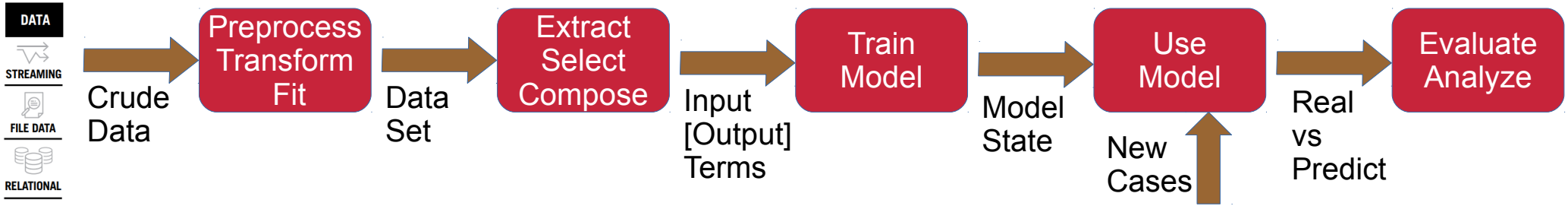




1. Entry data is chosen/prepared along with input conditions.
2. The machine learning algorithm is built and trained to accomplish a specific task.
3. The training data is augmented with the desired output information.

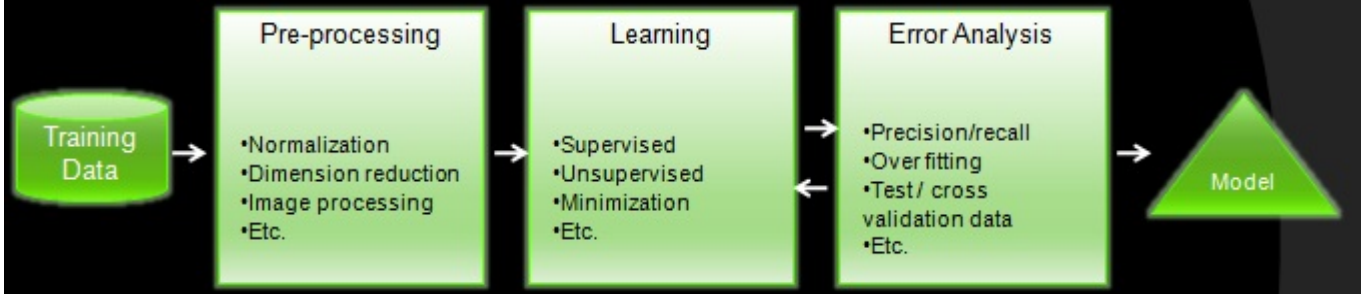
¿Is mission acomplished?

4. Evaluate model: unknown cases.
5. Analyze results.
6. Refine/change model.

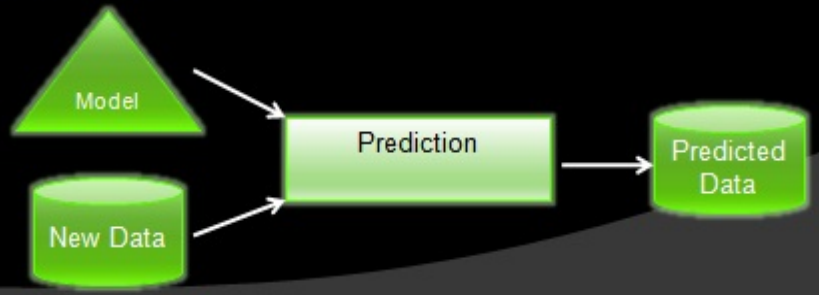


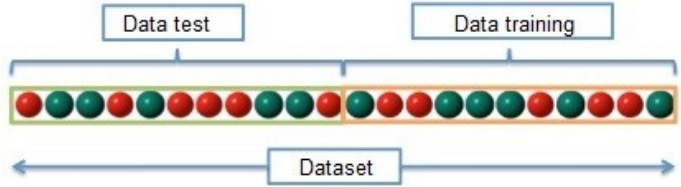
Machine Learning Process

- Phase 1) Learning

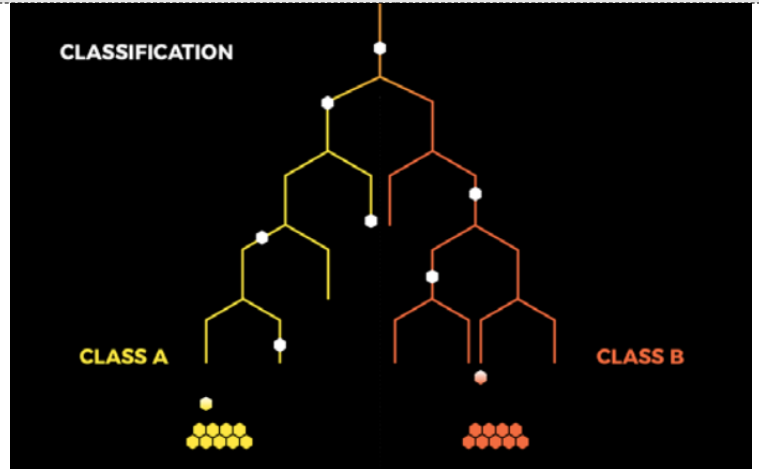


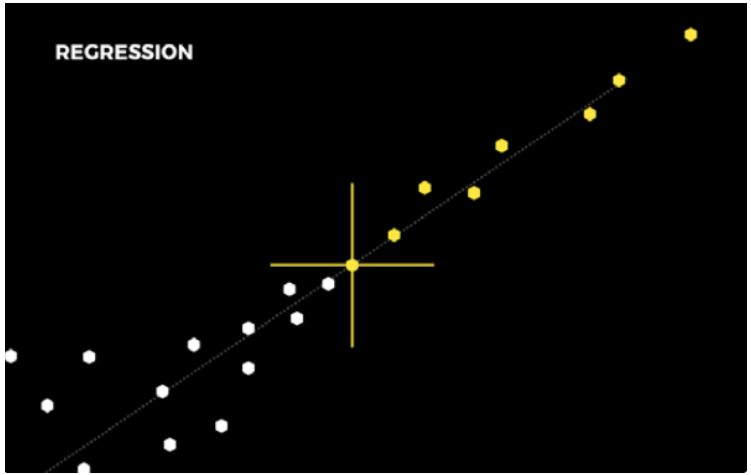
- Phase 2) Prediction



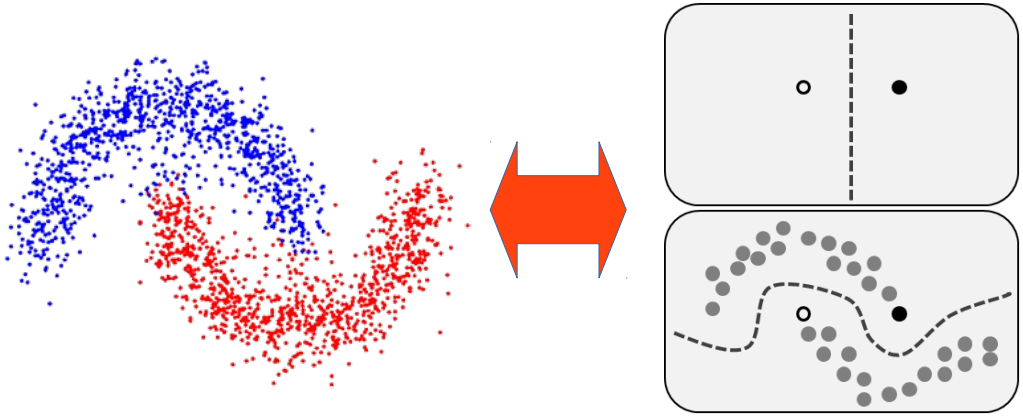


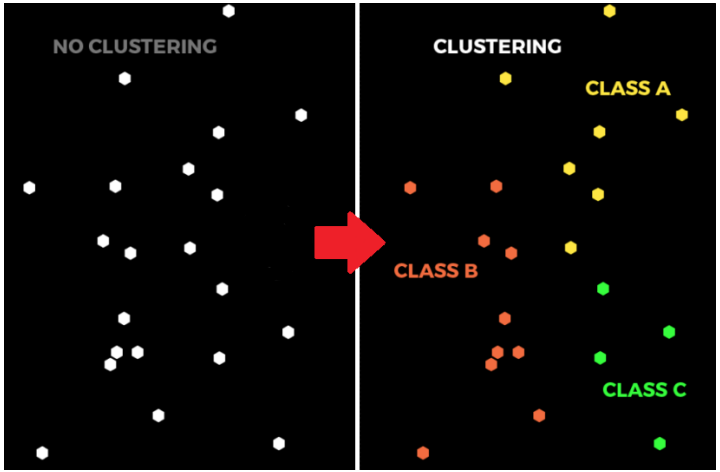
Supervised: Input+Outcome for training





Unsupervised: Input for training





Confusion Matrix

- *a* is the number of **correct** predictions that an instance is **negative**,
- *b* is the number of **incorrect** predictions that an instance is **positive**,
- *c* is the number of **incorrect** of predictions that an instance **negative**,
- *d* is the number of **correct** predictions that an instance is **positive**.

		Predicted	
		Negative	Positive
Actual	Negative	**a**	**b**
	Positive	**c**	**d**

Accuracy $AC = \frac{a+d}{a+b+c+d}$

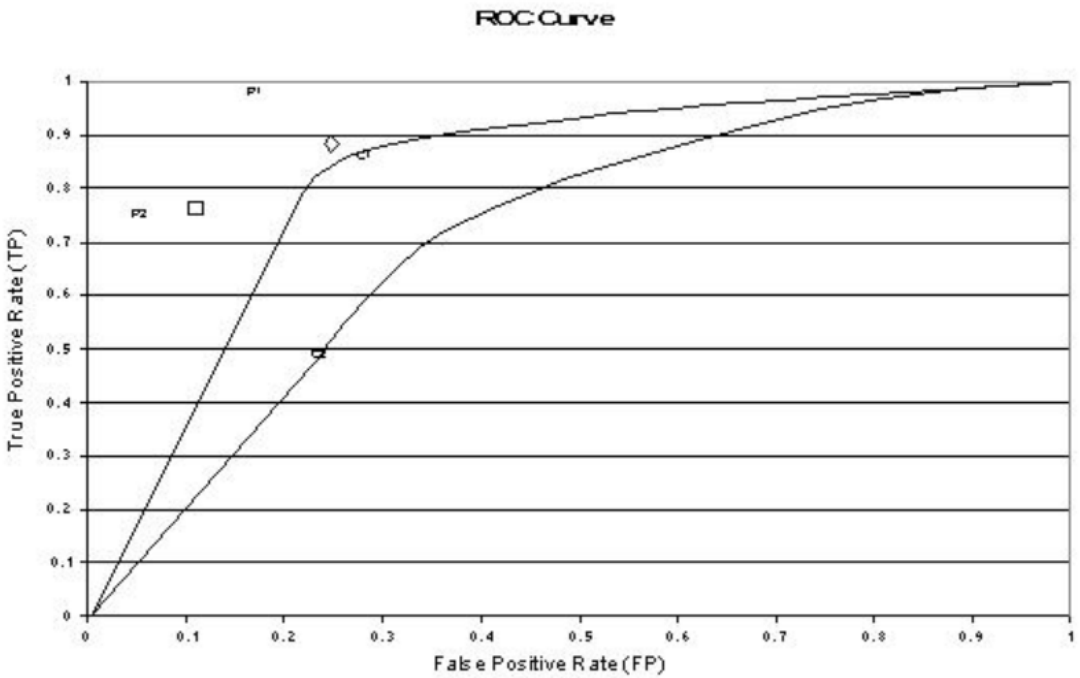
True positive rate (recall) $TP = \frac{d}{c+d}$

False positive rate $FP = \frac{b}{a+b}$

True negative rate $TN = \frac{a}{a+b}$

False negative rate $FN = \frac{c}{c+d}$

Precision $P = \frac{d}{b+d}$

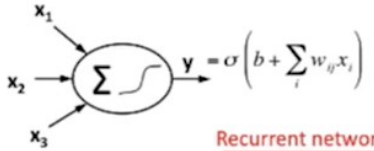


Geometric mean $g-mean_1 = \sqrt{TP * P}$

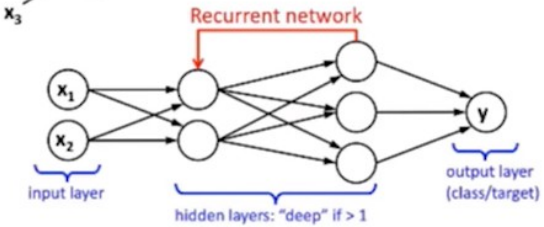
Geometric mean2 $g-mean_2 = \sqrt{TP * TN}$

F-Measure $F = \frac{(\beta^2+1) * P * TP}{\beta^2 * P + TP}$

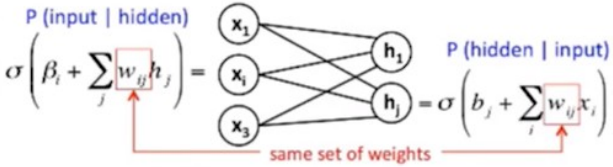
Types of Neural Networks



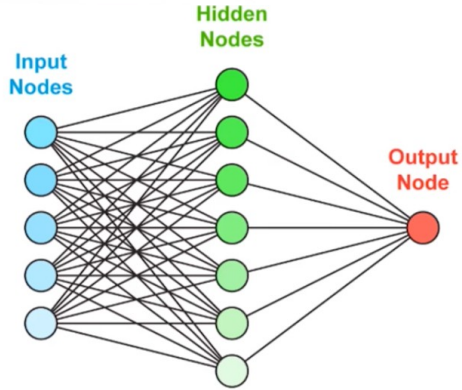
Single neuron: perceptron,
linear / logistic regression

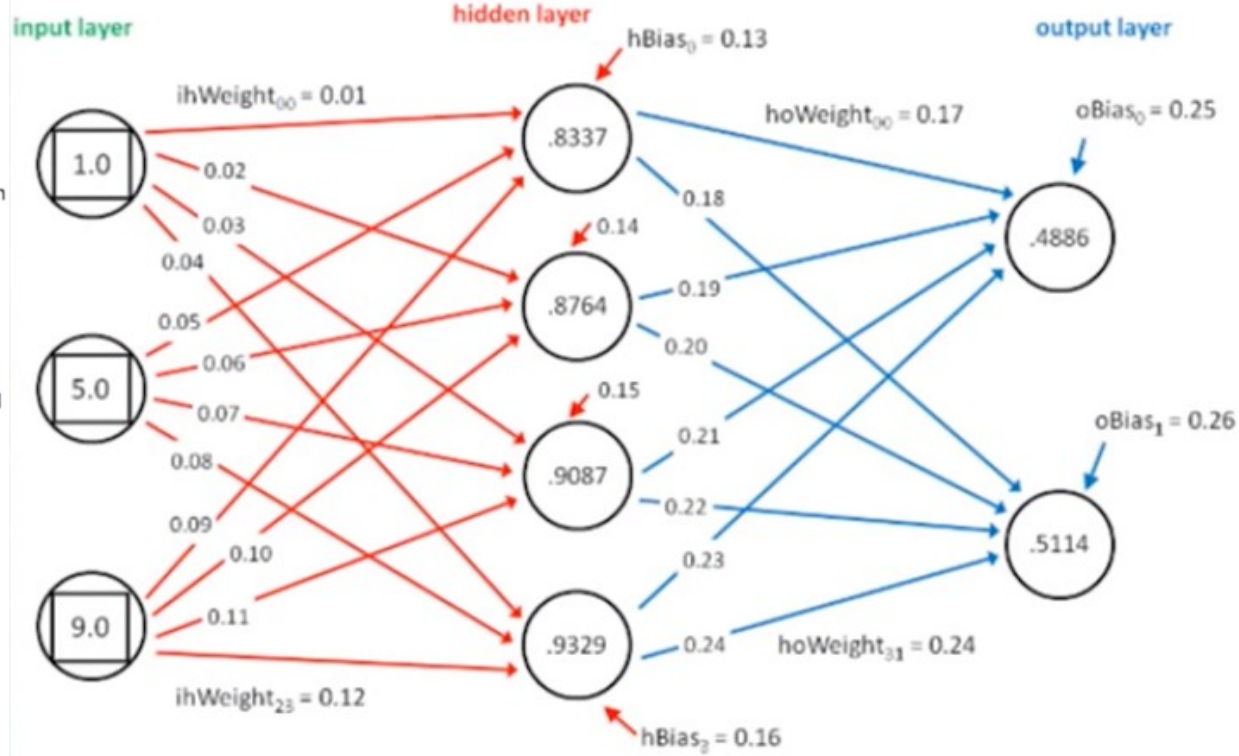


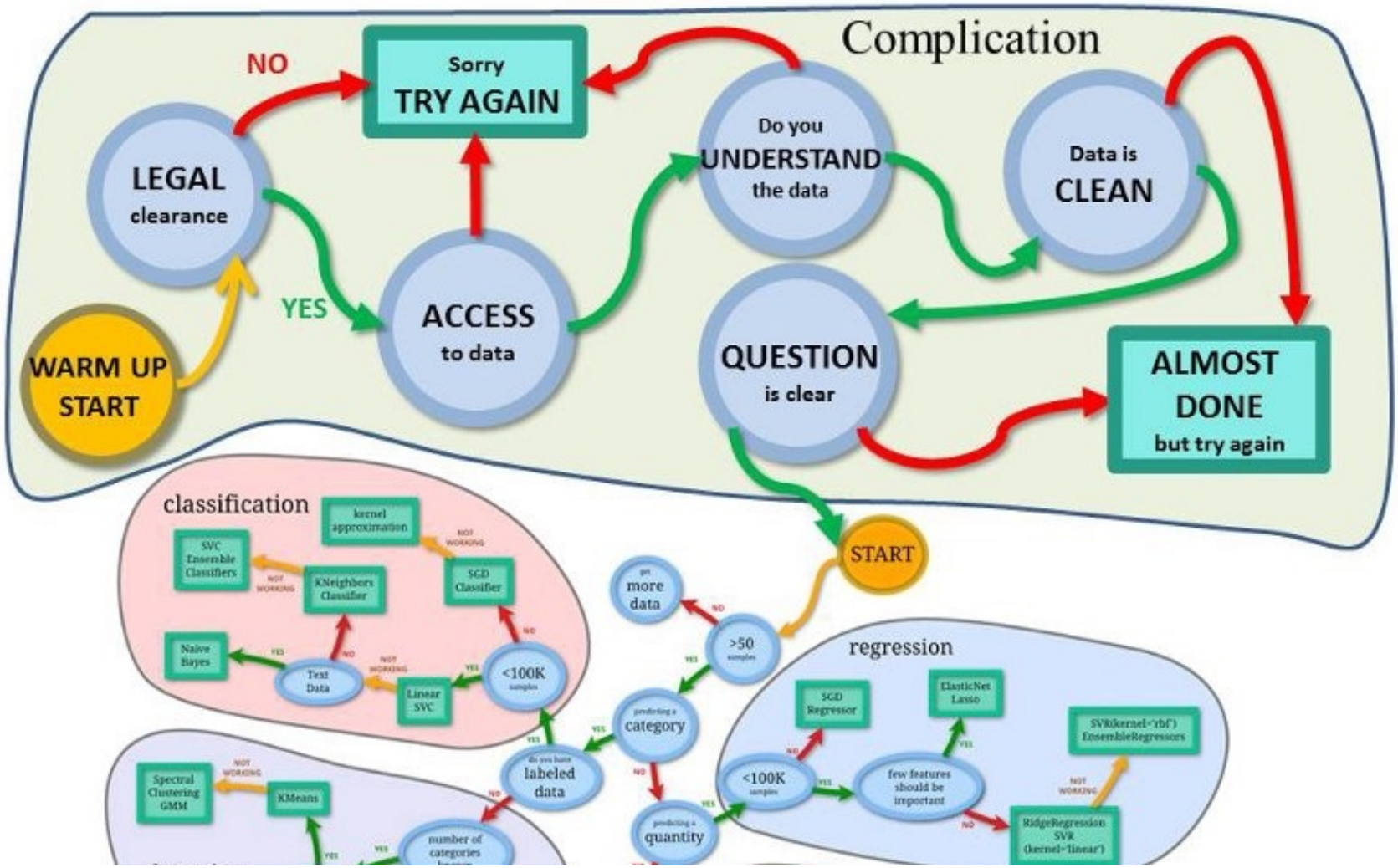
Feed-forward network
(no cycles) -- non-linear
classification & regression

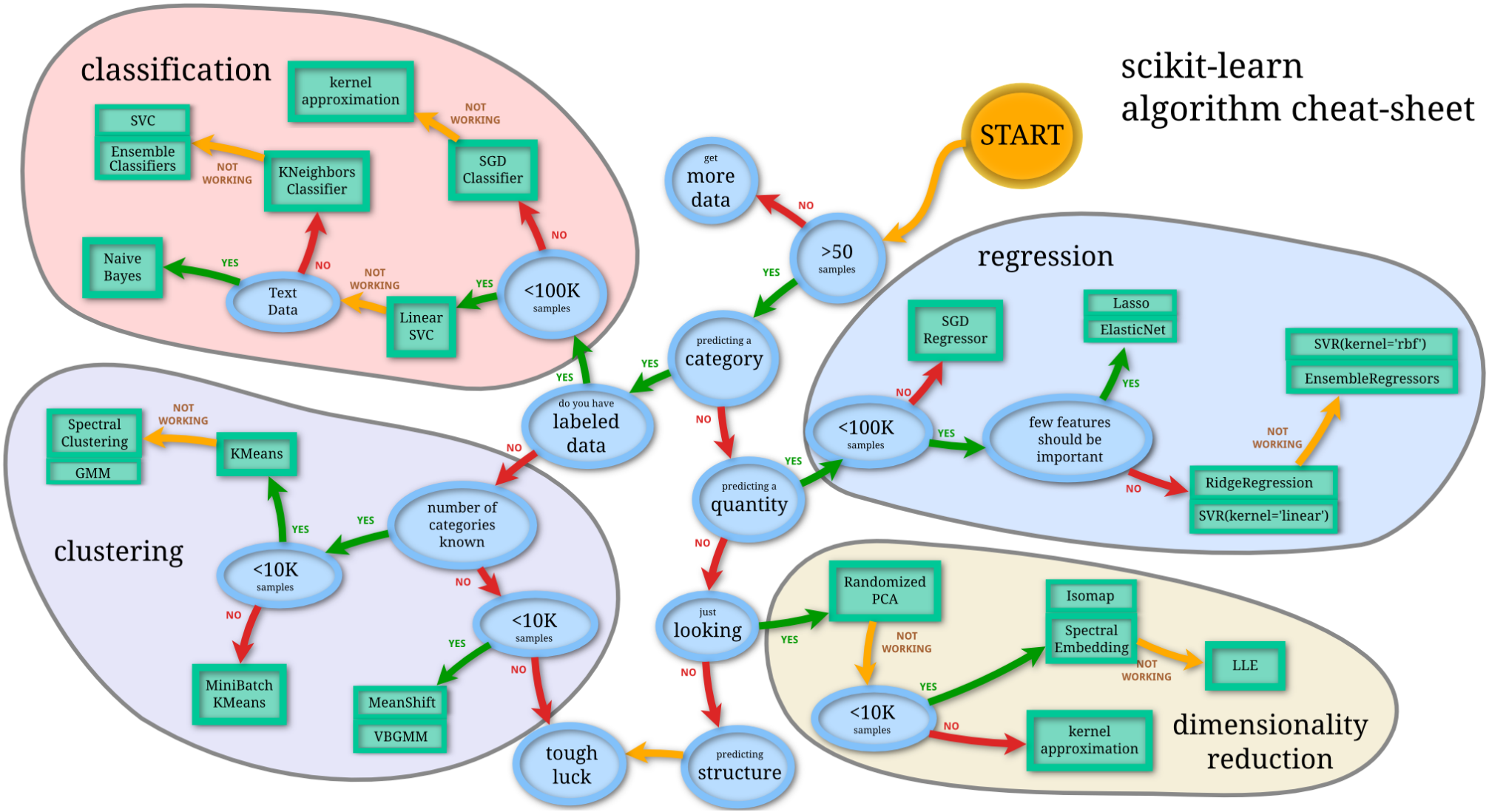


Symmetric (RBM)
unsupervised, trained
to maximize likelihood
of input data
a mixture model

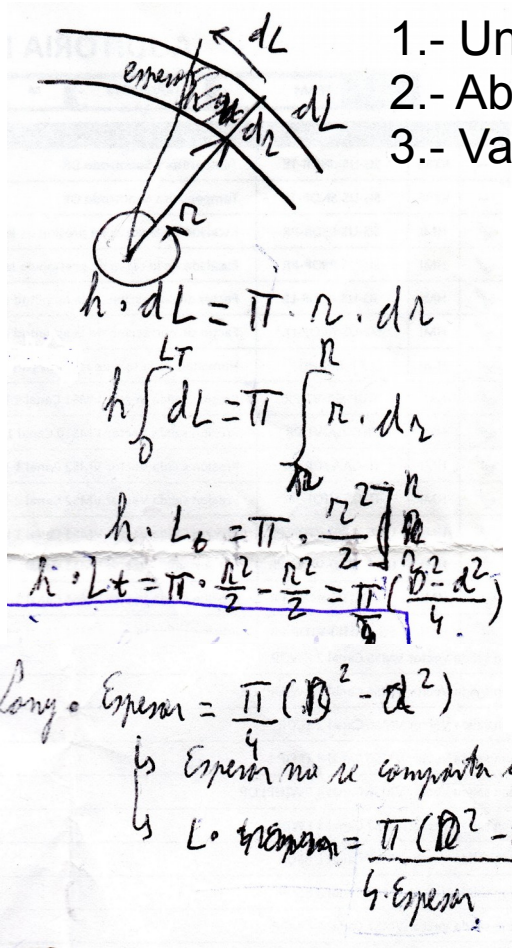




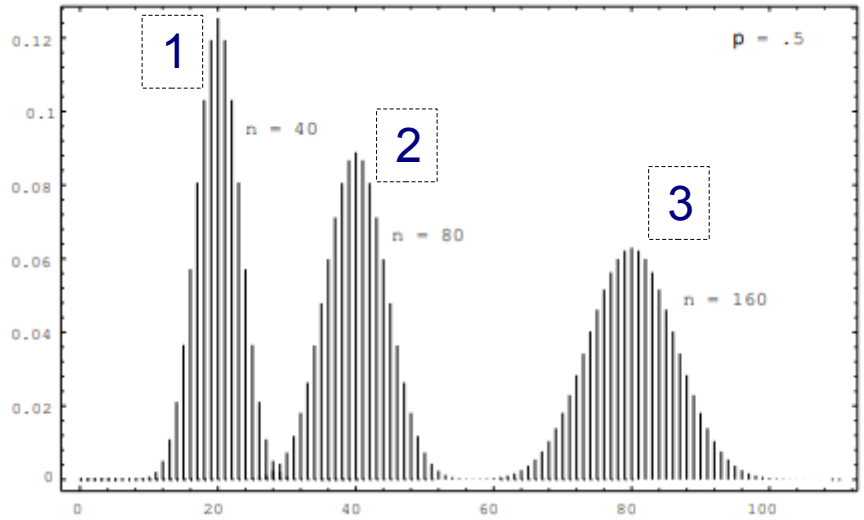

Complication
NO
Sorry
TRY AGAIN
LEGAL
clearance
YES
ACCESS
to data
Do you
UNDERSTAND
the data
Data is
CLEAN
WARM UP
START
QUESTION
is clear
ALMOST
DONE
but try again
START
classification
kernel approximation
SVC Ensemble Classifiers
KNeighbors Classifier
SGD Classifier
Naive Bayes
Text Data
Linear SVC
<100K
Spectral Clustering GMM
KMeans
number of categories
labeled data
get more data
>50
predicting a category
predicting a quantity
regression
SGD Regressor
ElasticNet Lasso
<100K
few features should be important
SVR(kernel='rbf') EnsembleRegressors
RidgeRegression SVR (kernel='linear')


scikit-learn
algorithm cheat-sheet
START
get more data
>50 samples
predicting a category
do you have labeled data
predicting a quantity
just looking
predicting structure
tough luck
classification
kernel approximation
SVC
Ensemble Classifiers
KNeighbors Classifier
SGD Classifier
Naive Bayes
Text Data
Linear SVC
<100K samples
regression
SGD Regressor
Lasso
ElasticNet
SVR(kernel='rbf')
EnsembleRegressors
<100K samples
few features should be important
RidgeRegression
SVR(kernel='linear')
clustering
Spectral Clustering
GMM
KMeans
number of categories known
<10K samples
MiniBatch KMeans
<10K samples
MeanShift
VBGMM
dimensionality reduction
Randomized PCA
Isomap
Spectral Embedding
LLE
<10K samples
kernel approximation
YES
NO
NOT WORKING

Hands on real 1: unwinding wires, hoses and flat material from winc, coils and rolls



1.- Unwinder speed or calculations tolerance: new options for running
2.- Abnormal crosslinks, thickness, density, rigidity of materials.
3.- Variable speed on some cases with angular speed constant state.



Exploratory Analisys provides good understanding of the dataset. Actionable Insigths vs. Tech Knowledge

Estimation for variability with dataframe transformation, panda's python lib. Generate fast run-time model for K constant speed intervals.

Hands on real 2: abnormal shape of the feet

SMART SPORT

MIND REDUCE

TEAM MEMBERS:

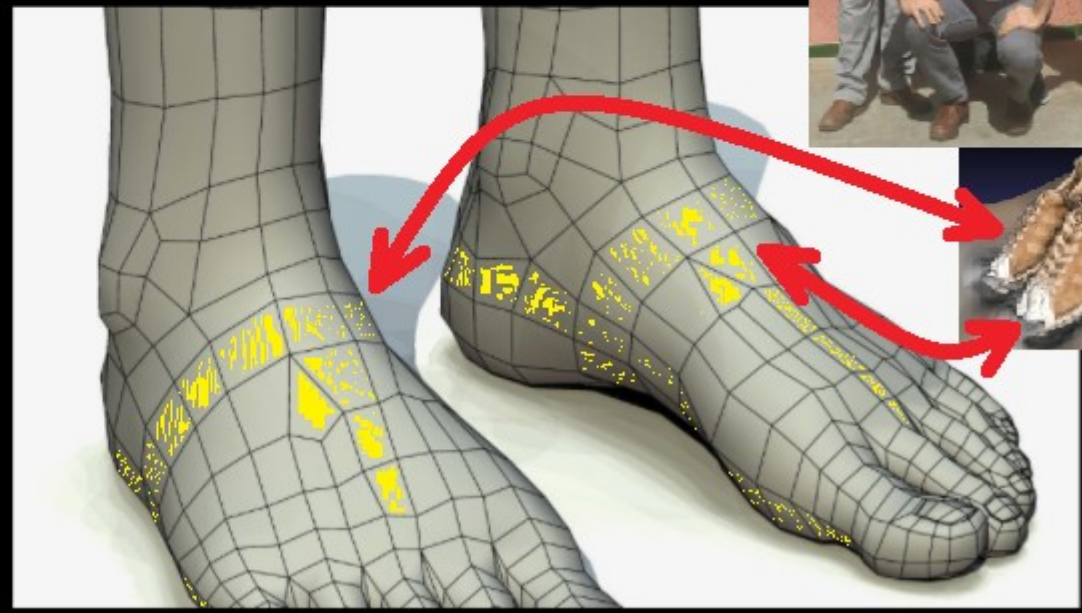
Santiago Frias , Tech Manager

JM Torrents, Research and Ideas

Alberto Pardo, Tech

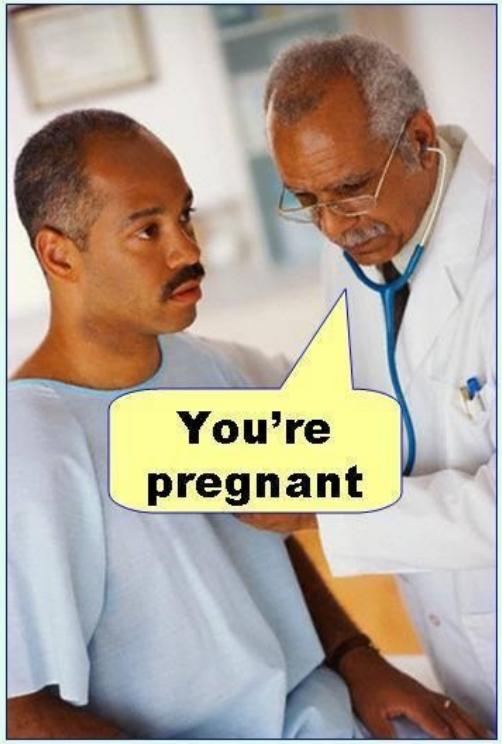
Jesus Girones, Big Data

Josep Romero, Marketing and Business

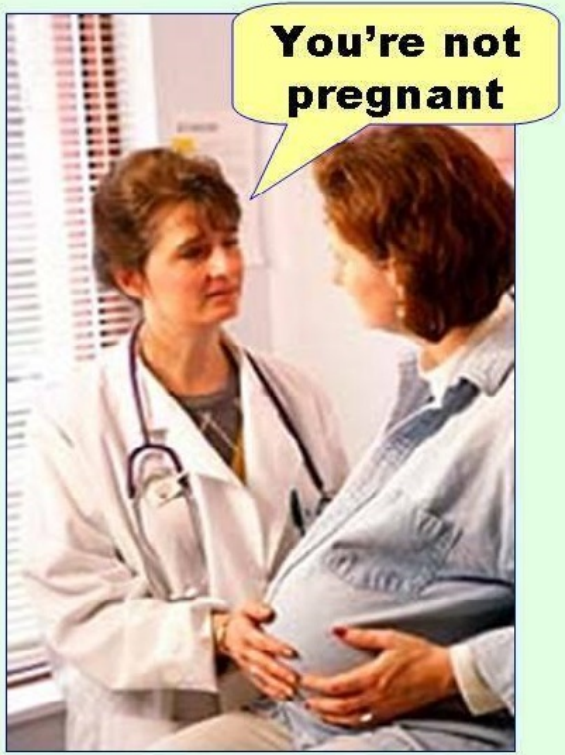


Remember!! Take care about this!!

Type I error
(false positive)



Type II error
(false negative)



Santiago Frias

sfrias1@gmail.com

https://www.linkedin.com/in/sfrias/

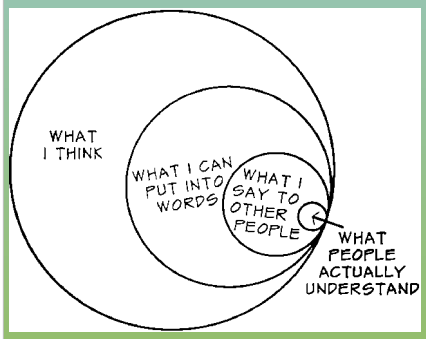
https://github.com/sfrias

https://www.kaggle.com/sfrias

https://stackoverflow.com/users/4538435/santi-fri

Link to this presentation

101 Machine Learning



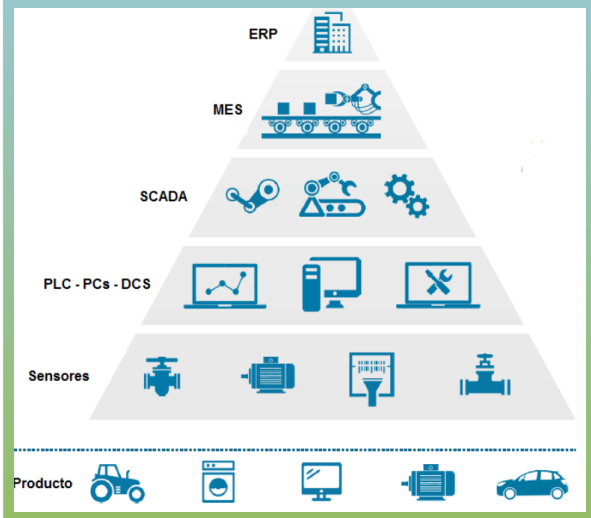
I. Evolution for computing environments: volume of data.

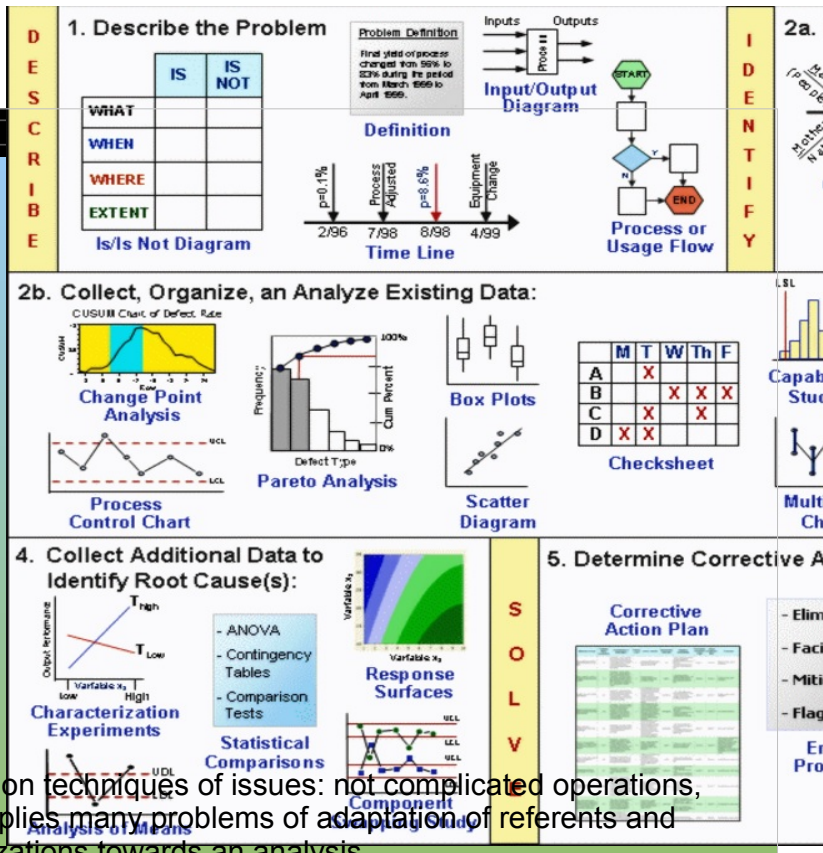
II. Choice with ML: types, phases and functionalities.

III. Checking how predictions were measured and compared.

Previous environment

- Isolated processes/systems.
- Specialization: departments/ operations
- Distributed/isolated knowledge





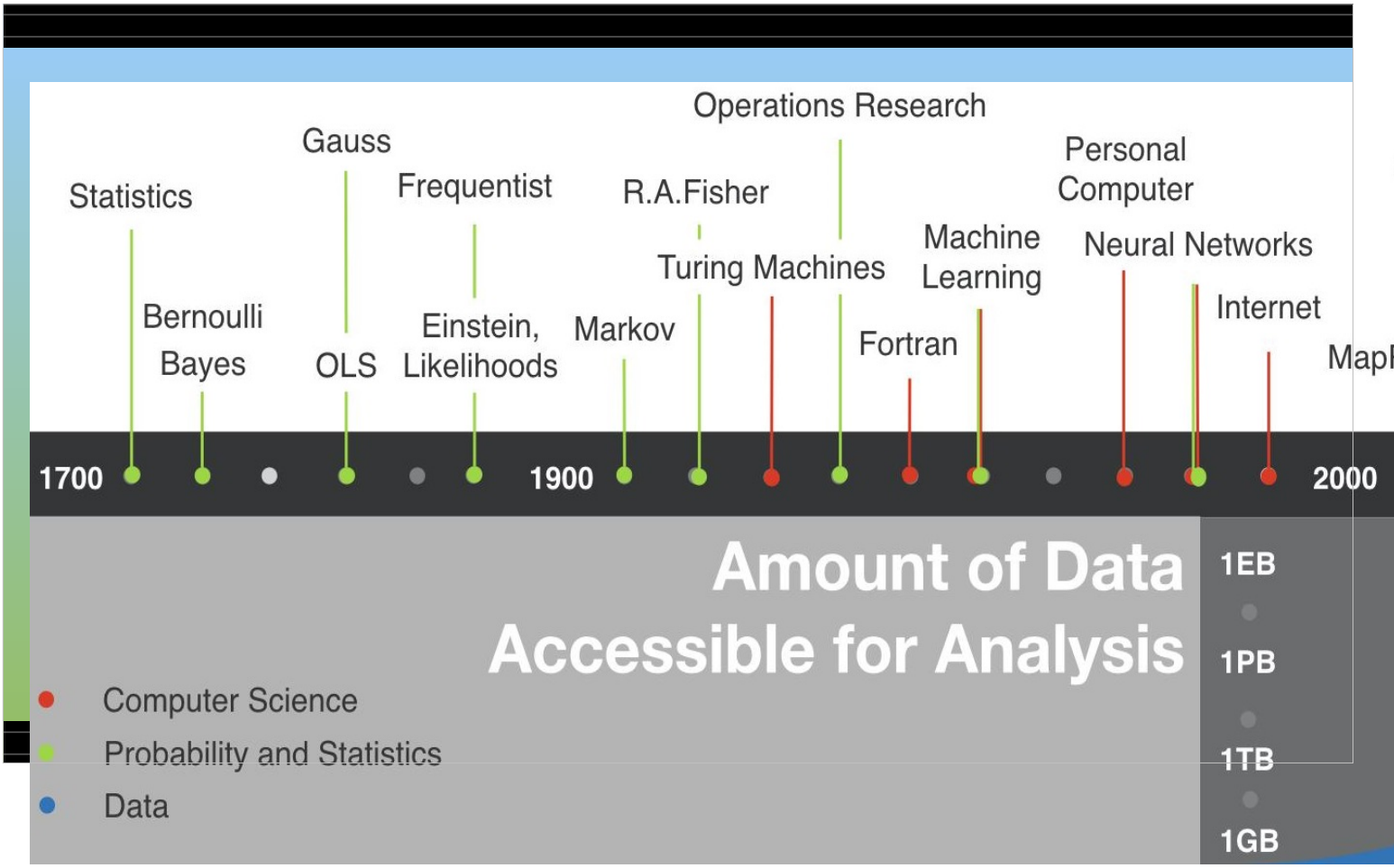
Resolution techniques of issues: not complicated operations, but it implies many problems of adaptation of referents and specializations towards an analysis.

2

The techniques of ML are candidates to be one of the tools that help to alleviate some situations, among them:

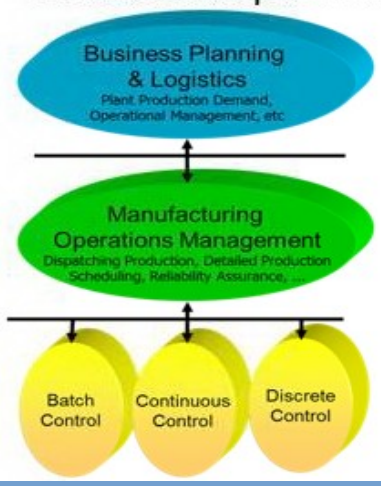
- The exponential increase in computing capacity and the availability of diverse and enormous volumes of data about the environment that surrounds us (big data) and its inter-organization with the processes that are related, with a strategy of data integration, related to the paradigm of organizations driven by data.

- The partial capacity of the people when solving some problems in the field of organizations and in the right decision making, in front of to the possibilities of establishing reliable prediction models, based on information extracted in part from these sources for training those same models.

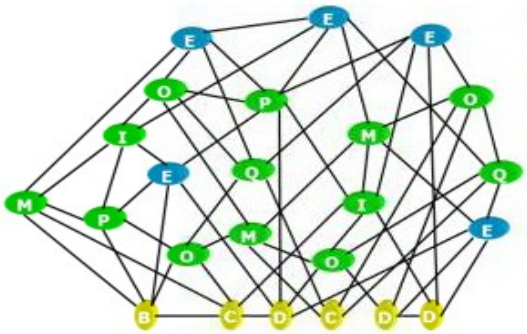

Operations Research
Gauss
Personal
Computer
Statistics
Frequentist
R.A.Fisher
Machine
Learning
Neural Networks
Turing Machines
Bernoulli
Bayes
OLS
Einstein,
Likelihoods
Markov
Fortran
Internet
1700
1900
2000
Amount of Data
Accessible for Analysis
1EB
1PB
1TB
1GB
Computer Science
Probability and Statistics
Data

Smart Manufacturing Challenge

From here
Vertical Interoperability



To here
Smart Enterprises, Sites, Work Centers.. Devices



E Enterprise

Data Exchange Type
O Operations
P Production
I Inventory
M Maintenance
Q Quality

Process Type
B Batch
C Continuous
D Discrete

Design thinking involves observation to discover unmet needs within the context and constraints of a particular situation. It frames the opportunity and scope of innovation, generating creative ideas, testing and refining solutions. It creates a repeatable and scalable process for innovation.

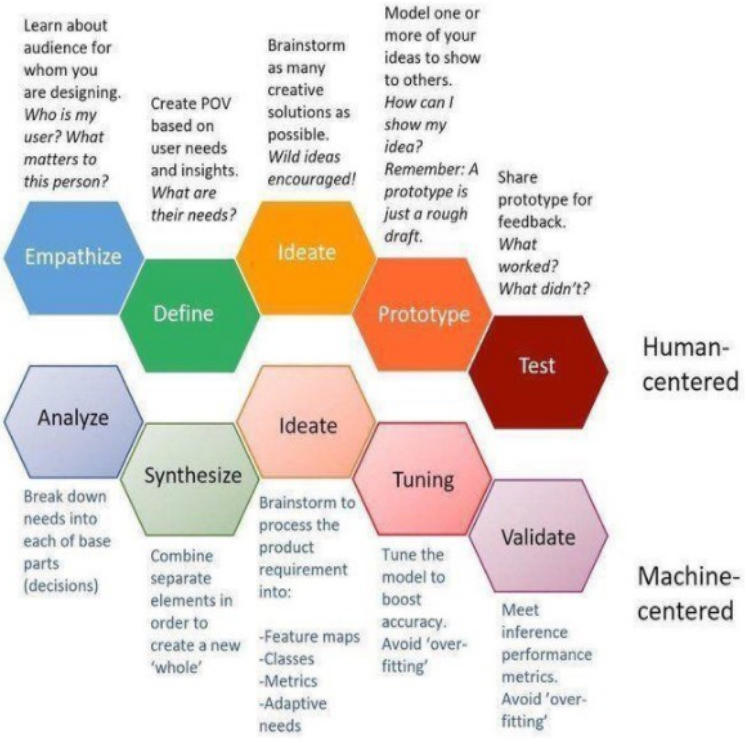
Design Thinking

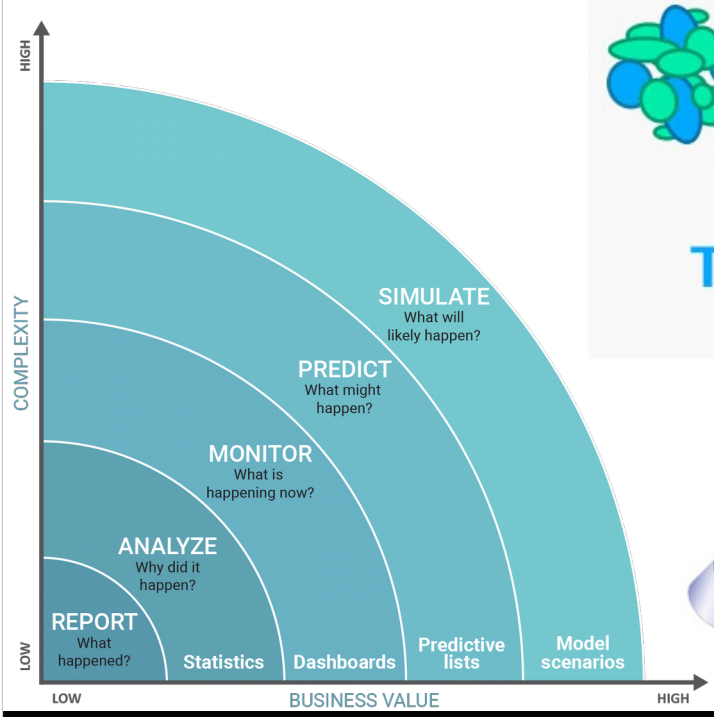
designs products this way…

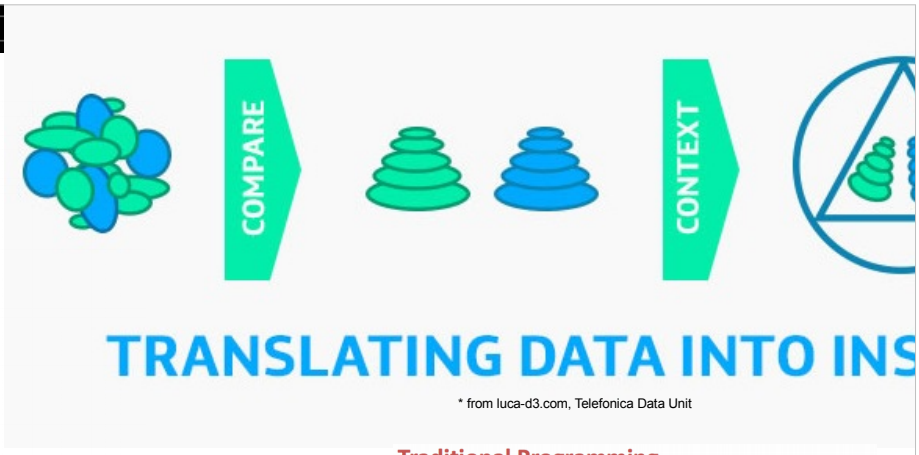
refines, tunes and predicts this way…

Machine Learning

Machine learning is a method of data analysis that automates analytical model building. Using algorithms that iteratively learn from data, *machine learning* allows computers to find hidden insights without being explicitly programmed where to look.



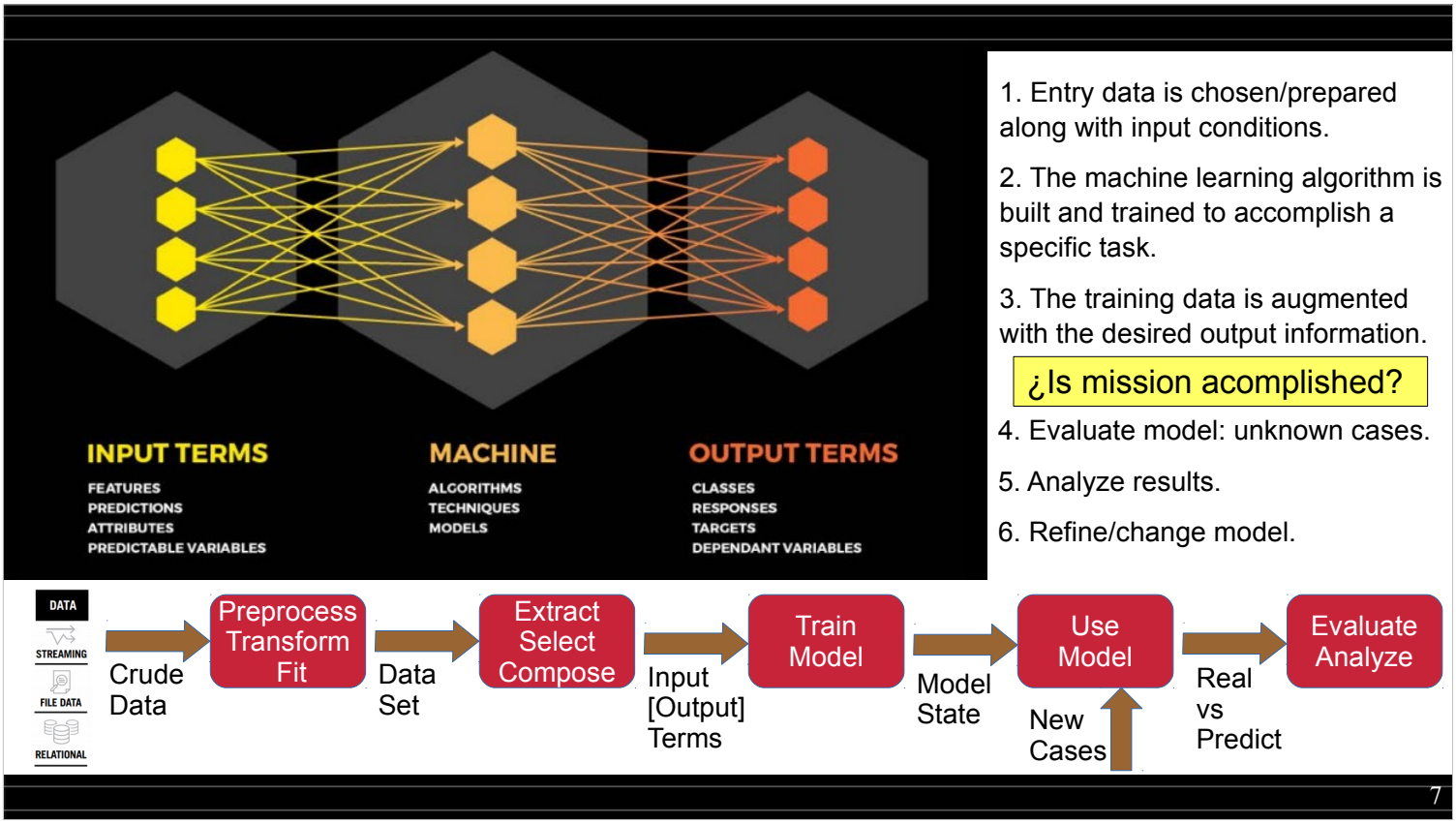




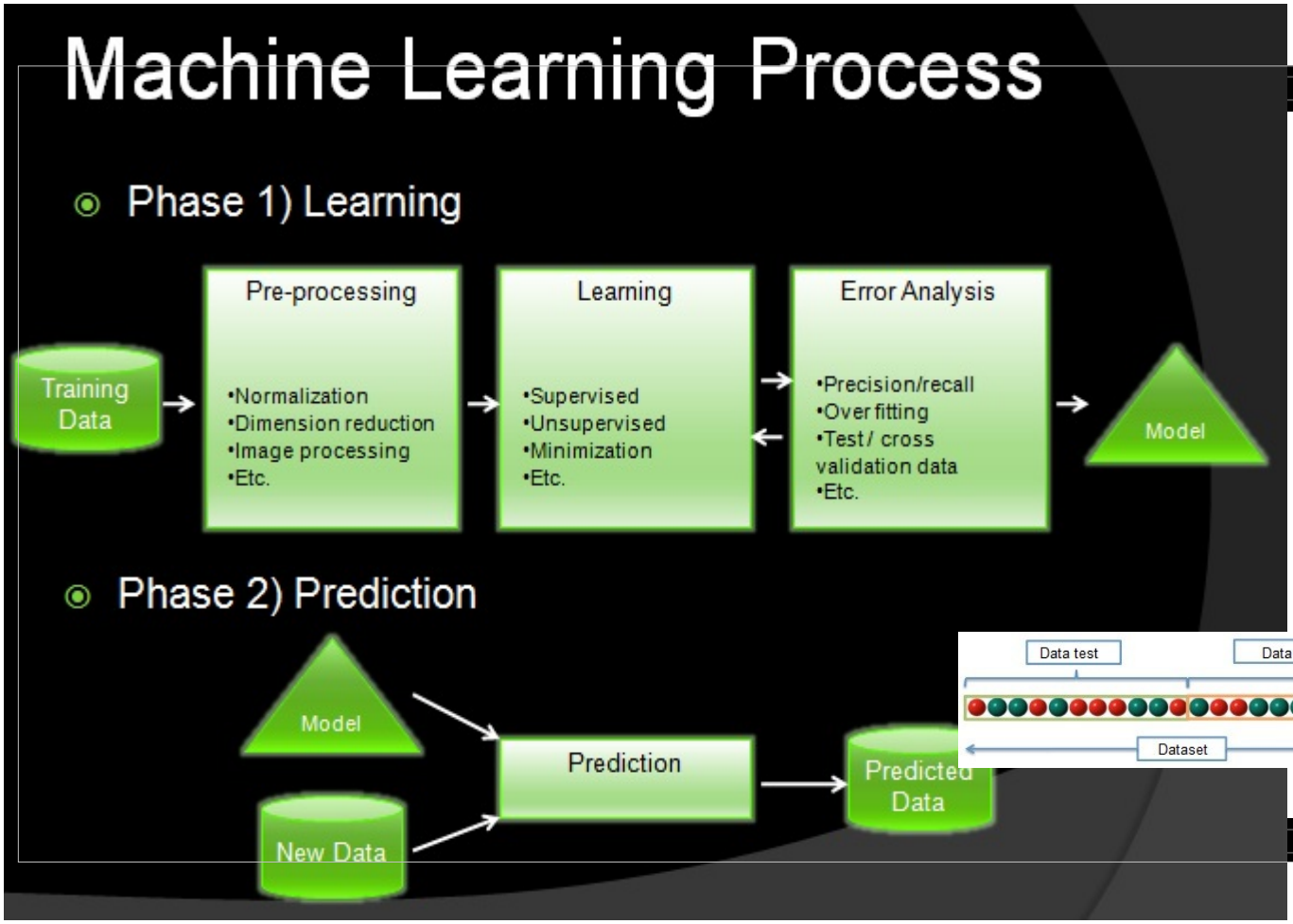
* from luca-d3.com, Telefonica Data Unit

Traditional Programming

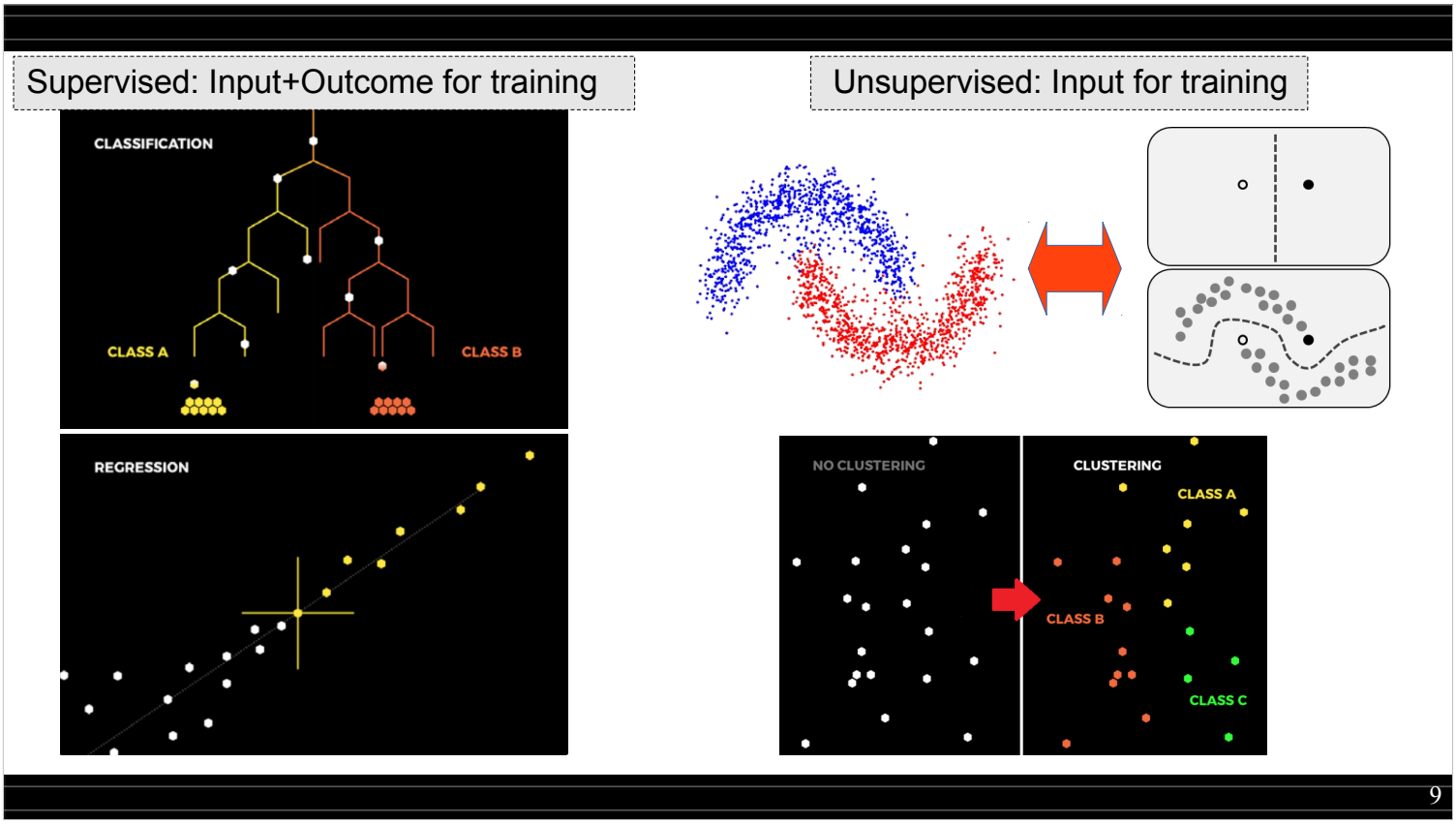
Machine Learning



1. Entry data is chosen and prepared along with input Conditions.
2. The machine learning algorithm is built and trained to accomplish a specific task.
3. The training data is augmented with the desired output information.



- Clean: Data can be generated by different sources,contained in different file formats, and expressedin different languages. It might be required to addor remove information from your data set, as someinstances might be missing information while others might contain undesired or irrelevant entries. Its preparation will impact its usability and the reliability of the outcome.
- Split: Depending on the size of your data set, only a portion of it may be required. This is usually referred to as sampling. From the chosen sample, your data should be split into two groups: one to train the algorithm and the other to evaluate it.
- Train: As commonly seen with neural networks,this stage aims to find the mathematical function that will accurately accomplish the chosen goal. Using a portion of your data set, the algorithm will attempt to process the data, measure its own performance and auto-adjust its parameters (also called backpropagation) until it can consistently produce the desired outcome with sufficient reliability.
- Evaluate: Once the algorithm performs well with the training data, its performance is measured again with data that it has not yet seen. Additional adjustments are made when needed. This process allows you to prevent overfitting, which happens when the learning algorithm performs well but only with your training data.
- Optimize: The model is optimized before integration within the destined application to ensure it is as lightweight and fast as possible.



Supervised learning: With this type of learning, the correct outcome for each data point is explicitly labelled when training the model. This means the learning algorithm is already given the answer when reading the data. Rather than finding the answer, it aims to find the relationship so that when unassigned data points are introduced, it can correctly classify or predict them.

Unsupervised learning: the learning algorithm is not given the answer during training. Its objective is to find meaningful relationships between the data points. Its value lies in discovering patterns and correlations.

Confusion Matrix

- a is the number of **correct** predictions that an instance is **negative**,
- b is the number of **incorrect** predictions that an instance is **positive**,
- c is the number of **incorrect** of predictions that an instance **negative**,
- d is the number of **correct** predictions that an instance is **positive**.

		Predicted	
		Negative	Positive
Actual	Negative	**a**	**b**
	Positive	**c**	**d**

Accuracy $AC = \frac{a+d}{a+b+c+d}$

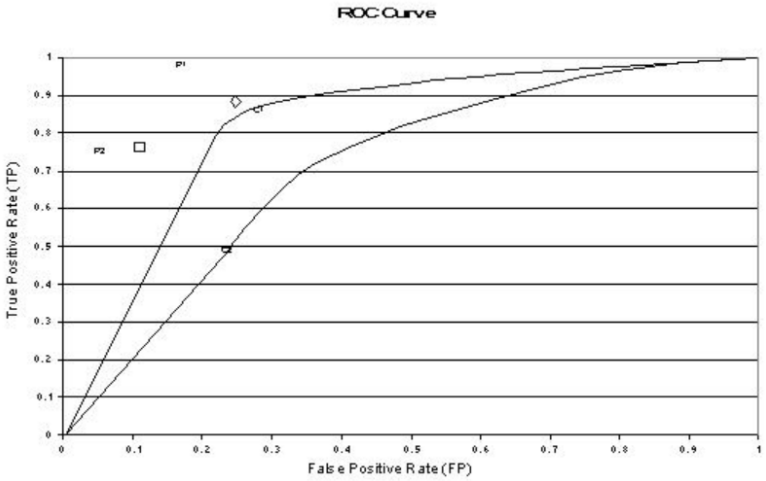
True positive rate (recall) $TP = \frac{d}{c+d}$

False positive rate $FP = \frac{b}{a+b}$

True negative rate $TN = \frac{a}{a+b}$

False negative rate $FN = \frac{c}{c+d}$

Precision $P = \frac{d}{b+d}$



Geometric mean $g - mean_1 = \sqrt{TP * P}$

Geometric mean2 $g - mean_2 = \sqrt{TP * TN}$

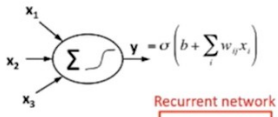
F-Measure $F = \frac{(\beta^2 + 1) * P * TP}{\beta^2 * P + TP}$

10

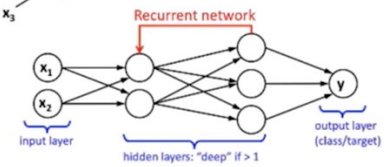
A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix.

- Accuracy is the proportion of the total number of predictions that were correct.
- recall or true positive rate is the proportion of positive cases that were correctly identified.
- false positive rate is the proportion of negatives cases that were incorrectly classified as positive.
- true negative rate is defined as the proportion of negatives cases that were classified correctly.
- false negative rate is the proportion of positives cases that were incorrectly classified as negative.
- precision is the proportion of the predicted positive cases that were correct
- Other measures are geometric means or F-Measure.

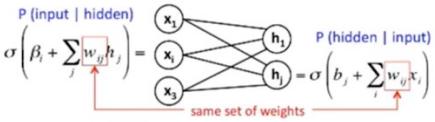
Types of Neural Networks



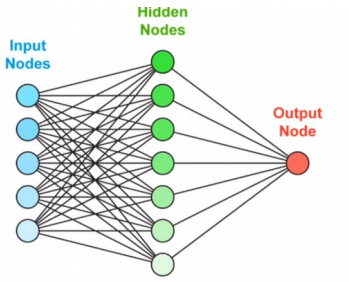
Single neuron: perceptron,
linear / logistic regression

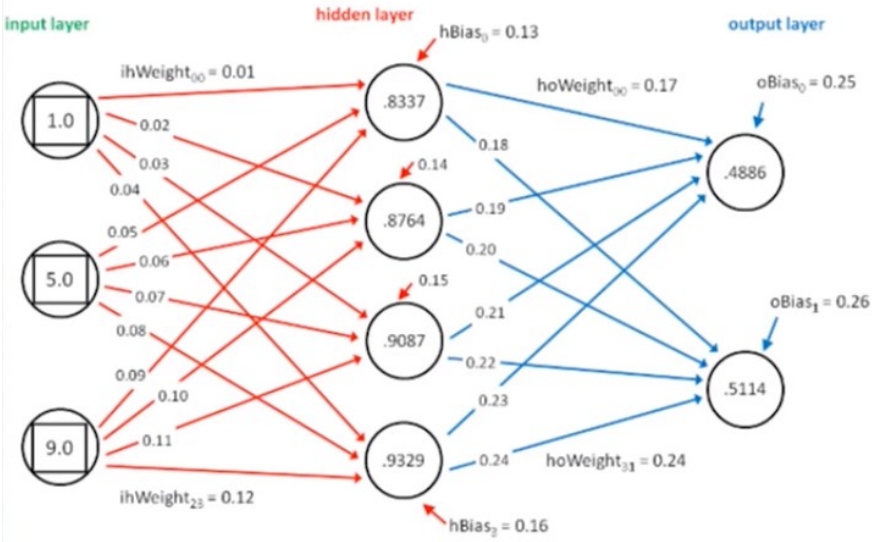


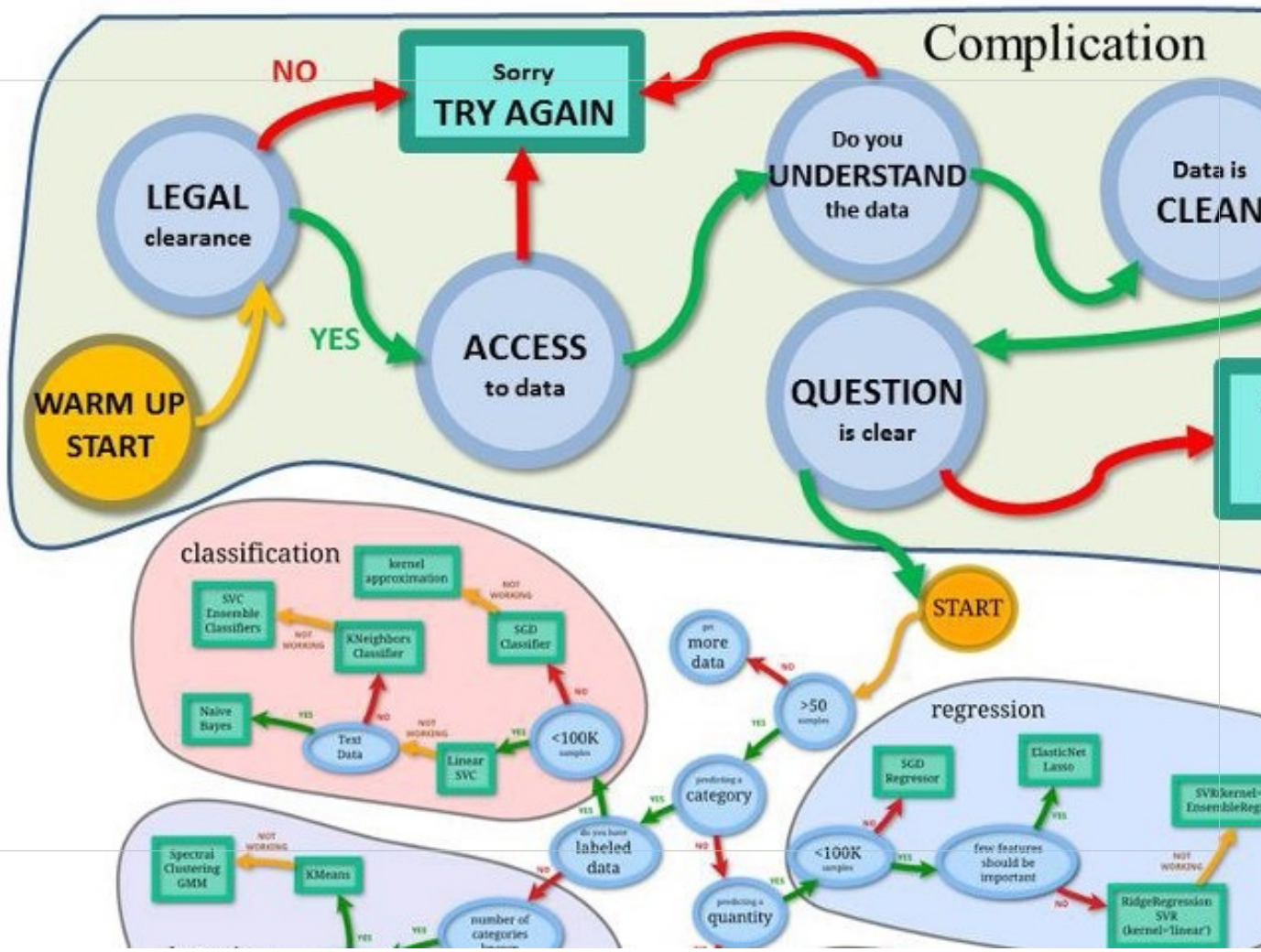
Feed-forward network
(no cycles) -- non-linear
classification & regression

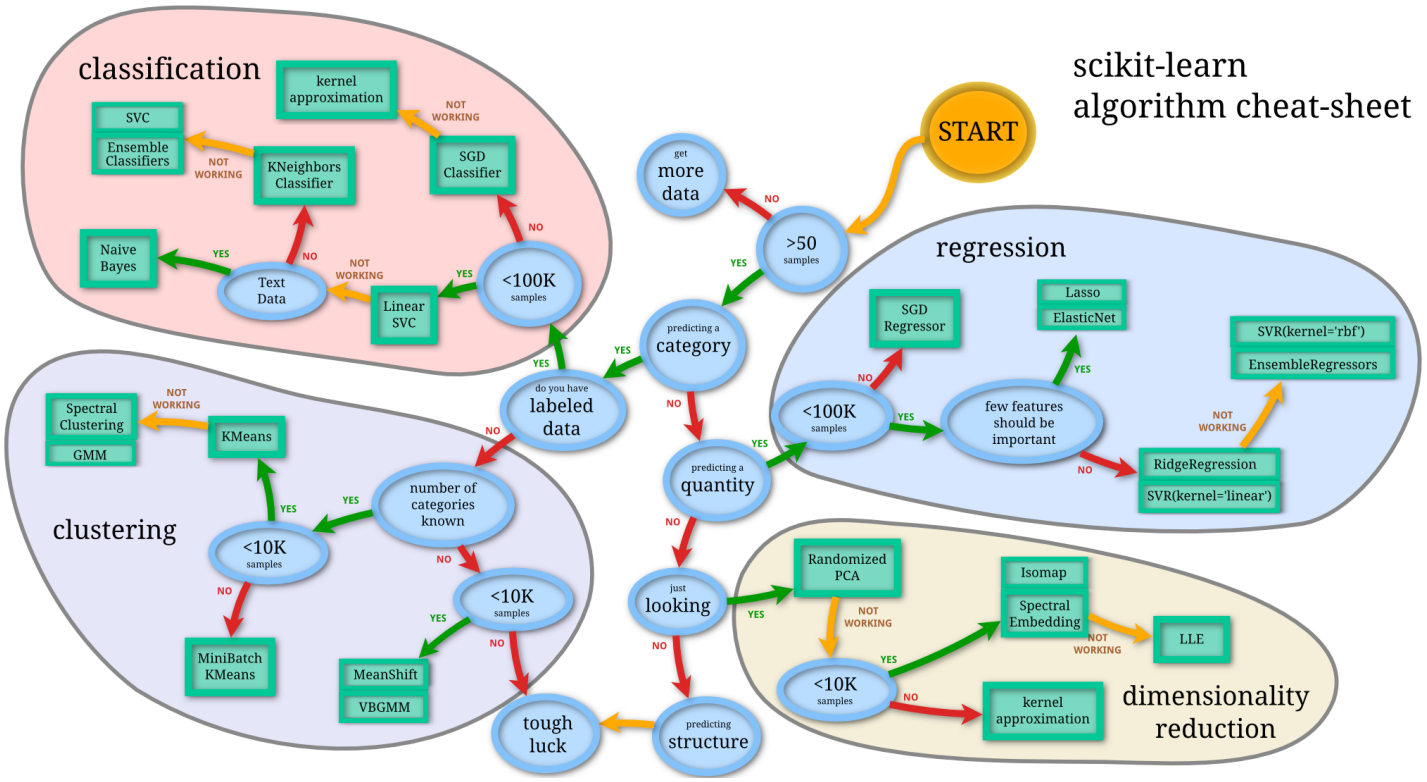


Symmetric (RBM)
unsupervised, trained
to maximize likelihood
of input data
a mixture model

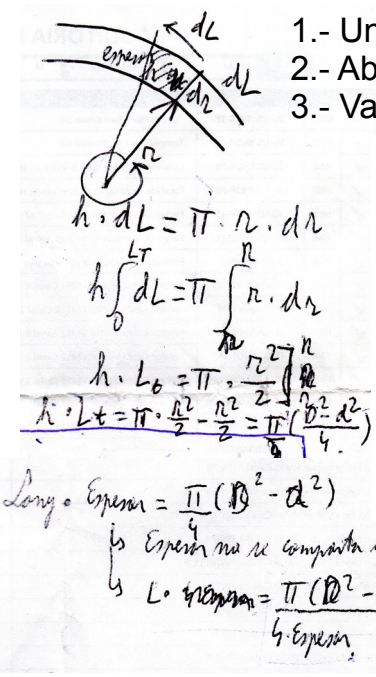




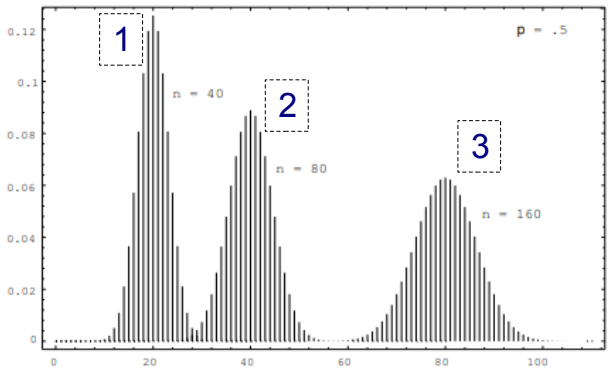

Complication
NO
Sorry
TRY AGAIN
LEGAL
clearance
YES
ACCESS
to data
WARM UP
START
Do you
UNDERSTAND
the data
Data is
CLEAN
QUESTION
is clear
START
classification
kernel approximation
SVC Ensemble Classifiers
KNeighbors Classifier
SGD Classifier
Naive Bayes
Text Data
Linear SVC
<100K
more data
>50
category
labeled data
quantity
regression
SGD Regressor
ElasticNet Lasso
<100K
few features should be important
RidgeRegression SVR (kernel='linear')
Spectral Clustering GMM
KMeans
number of categories


scikit-learn
algorithm cheat-sheet
START
>50 samples
get more data
predicting a category
do you have labeled data
predicting a quantity
just looking
predicting structure
tough luck
classification
kernel approximation
SVC
Ensemble Classifiers
KNeighbors Classifier
SGD Classifier
Naive Bayes
Text Data
Linear SVC
<100K samples
clustering
Spectral Clustering
GMM
KMeans
number of categories known
<10K samples
MiniBatch KMeans
MeanShift
VBGMM
regression
SGD Regressor
<100K samples
few features should be important
Lasso
ElasticNet
SVR(kernel='rbf')
EnsembleRegressors
RidgeRegression
SVR(kernel='linear')
dimensionality reduction
Randomized PCA
<10K samples
Isomap
Spectral Embedding
LLE
kernel approximation
YES
NO
NOT WORKING

Hands on real 1: unwinding wires, hoses and flat material from winc, coils and rolls



1.- Unwinder speed or calculations tolerance: new options for running
2.- Abnormal crosslinks, thickness, density, rigidity of materials.
3.- Variable speed on some cases with angular speed constant state.

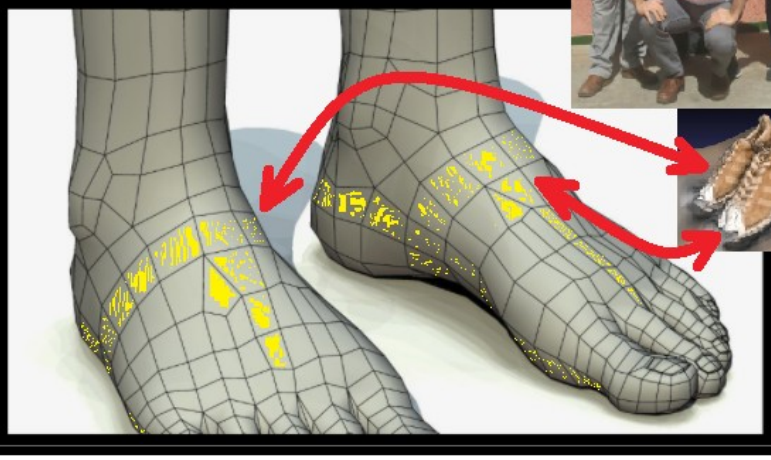


Exploratory Analisys provides good understanding of the dataset. Actionable Insigths vs. Tech Knowledge

Estimation for variability with dataframe transformation, panda's python lib. Generate fast run-time model for K constant speed intervals.

Hands on real 2: abnormal shape of the feet

SMART SPORT



MIND REDUCE

TEAM MEMBERS:

Santiago Frias , Tech Manager

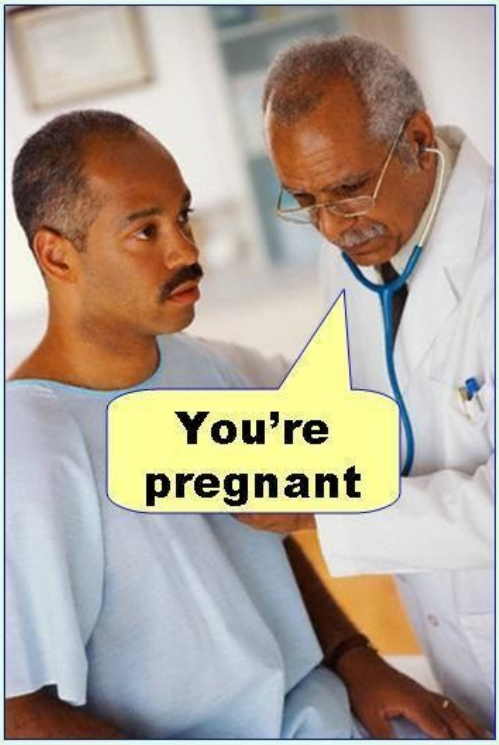
JM Torrents, Research and Ideas

Alberto Pardo, Tech

Jesus Girones, Big Data

Josep Romero, Marketing and Business

Type I error
(false positive)



Type II error
(false negative)

Santiago Frias

sfrias1@gmail.com

https://www.linkedin.com/in/sfrias/

https://github.com/sfrias

https://www.kaggle.com/sfrias

https://stackoverflow.com/users/4538435/santi-fri

Link to this presentation