Machine learning applied to the manufacturing industry: some case studies

Gagné, Marc-Olivier; Thériault, Benoit

NRC·CNRC

NRC.CANADA.CA

Machine learning applied to the manufacturing industry: some case studies

Marc-Olivier Gagné

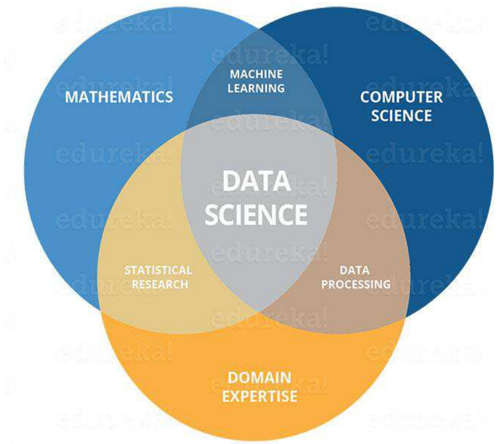
Benoit Thériault

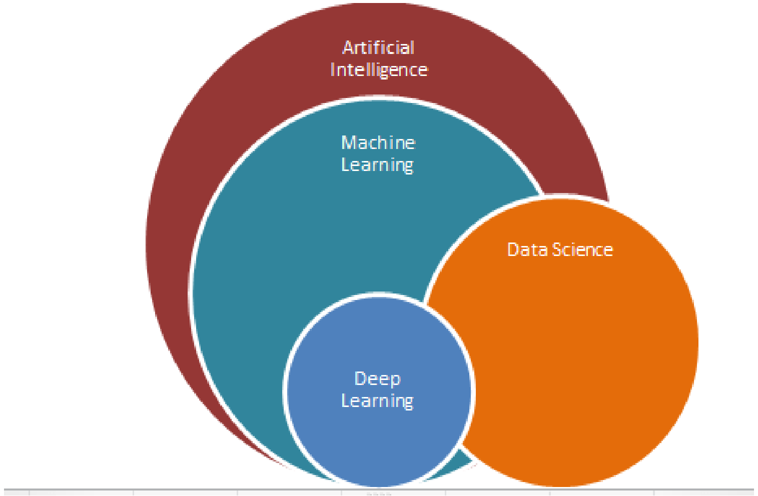
October 16th, 2019

What is Machine Learning?

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

https://www.expertsystem.com/machine-learning-definition/

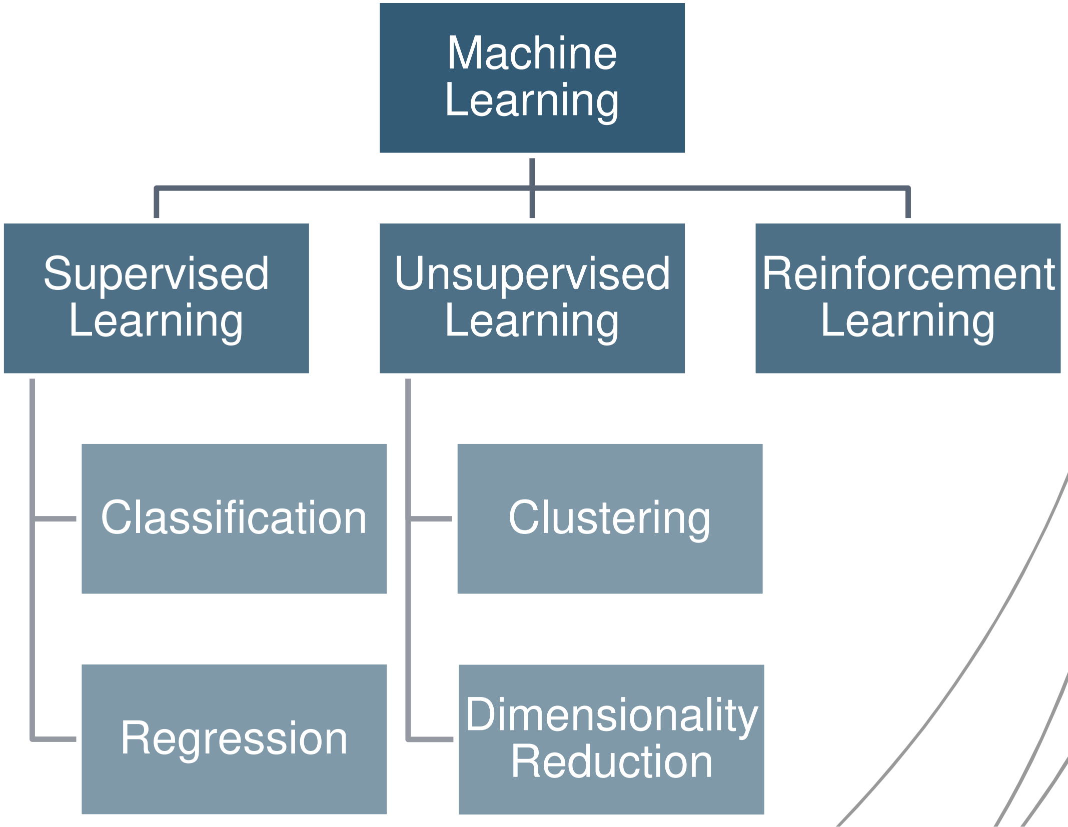




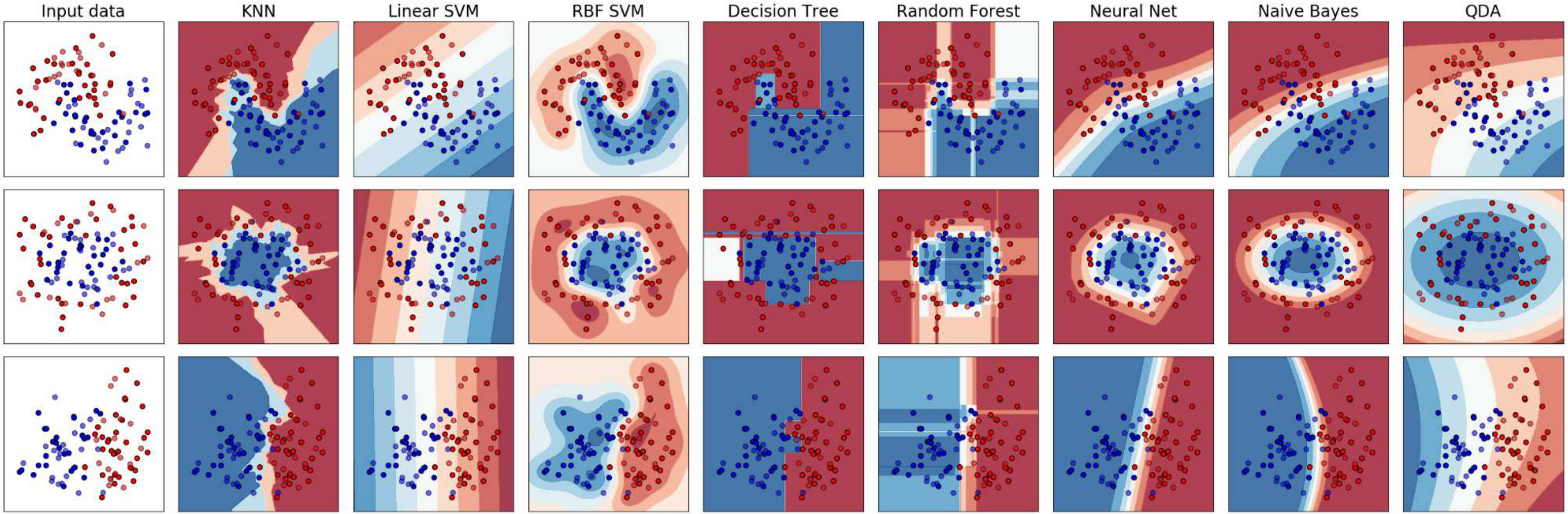
Different types of machine learning algorithms

Supervised learning: The model aim is to map inputs to outputs based on known inputs-outputs values

Unsupervised learning: The model seeks to find hidden patterns or data grouping from a set on inputs without an explicit objective or labeled response



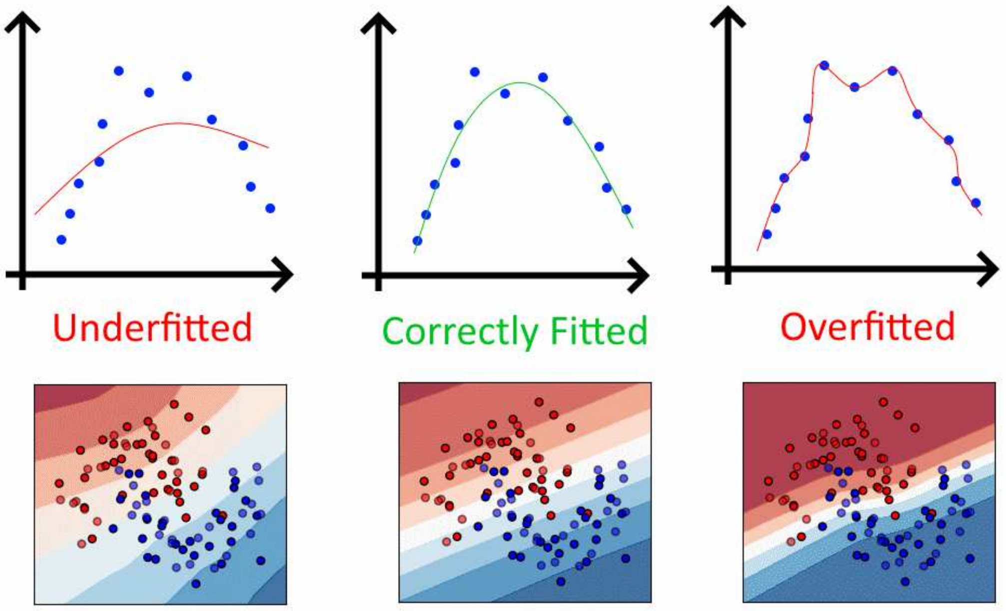
Different models will have different effects for the same classification problem



Modified from: scikit-learn.org, by Gaël Varoquaux, Andreas Müller

Models are checked and improved by different means

- K-Fold cross-validation, resampling
- Check for under or overfitting
- Compare performance from different models
- Optimize for best meta-parameters



Machine Learning: Great but not perfect

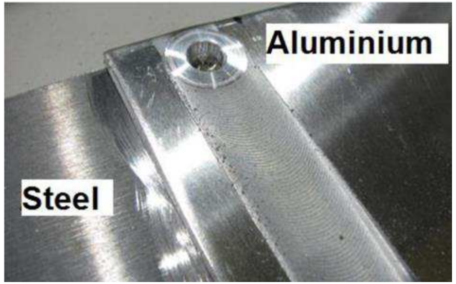
Advantages:

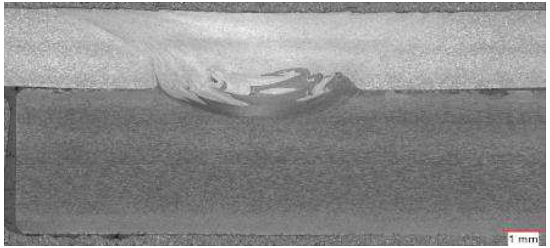
- Take into account large number of parameters
- Orders of magnitudes faster than FEA

Limitations:

- Incapable of prediction outside the bounds of training sets
- Only as good as the training set
- Hard to interpret more “Black Box” models

Friction Stir Welding (FSW) mixes the material at the interface of two metals to bond them together

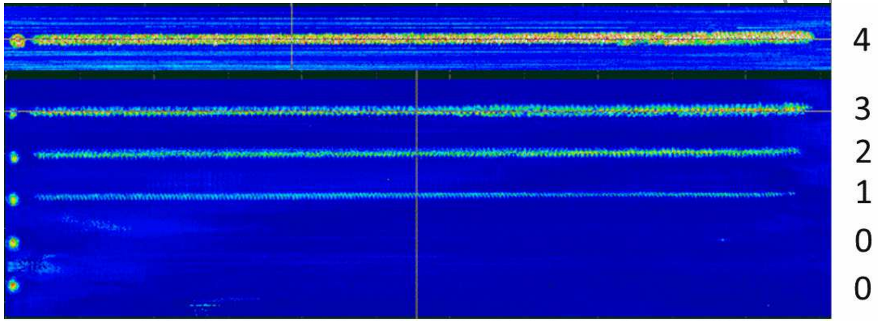


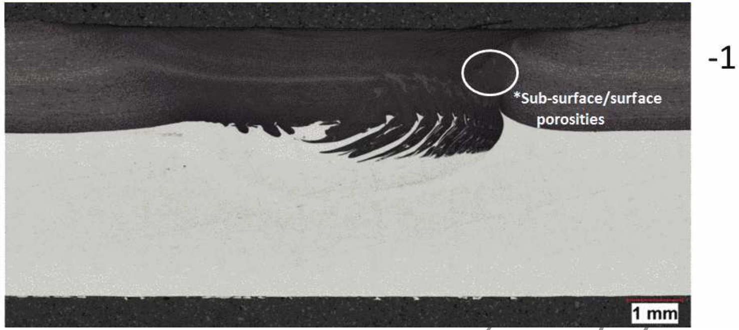


Defect index were created to categorize different types of defects and their severity

- 5 years worth of welds
- Inspected for defects sizes down to 100 μm

Defect index nomenclature	
Defect type	**Index value**
Insipient melting (too hot; reach solidus)	-1
No volumetric defect	0
Intermittent volumetric defect	1
Small volumetric defect	2
Medium-size volumetric defect	3
Large volumetric defect	4





The dataset contains all known weld parameters

Geometrical parameters	Material properties	Machine Setup		Live measurements	Attempted prediction
Joint configuration (Lap joint, butt joint, …)	Yield Stress	Rotational Speed	Pin diameter	Longitudinal & lateral forces	Defect Index From -1 To 4
Pin Diameter	Ultimate Stress	Travel Speed	Tool “agressivity score”	Torque	
Pin Feature	Solidus Temperature	Forge force	Pin & shoulder feature	Specific energy	
Top & Bottom sheet thicknesses	Insipient melting Temperature	Pin tool material		Temperature	

Number of samples ≈ 500

Building a classification model using RStudio

Example with K-Nearest Neighbors

```
# Splitting the dataset into the Training set and Test set
# install.packages('caTools')
library(caTools)
set.seed(123)
split = sample.split(all_knn$defect_index, SplitRatio = 0.75)
training_set = subset(all_knn, split == TRUE)
test_set = subset(all_knn, split == FALSE)

# Encoding the target feature as factor
training_set$defect_index = factor(training_set$defect_index, levels = c(0, 1))
test_set$defect_index = factor(test_set$defect_index, levels = c(0, 1))

# Feature Scaling
training_set[-50] = scale(training_set[-50])
test_set[-50] = scale(test_set[-50])

# Fitting K-NN to the Training set and Predicting the Test set results
cl = training_set$defect_index
library(class)
y_pred = knn(train = training_set[, -50],
             test = test_set[, -50],
             cl = cl,
             k = 5,
             prob = FALSE)
```

- Several models are evaluated similarly
- Model performances are compared to find the best model
- Predictors importance in evaluated to identify the most relevant ones

Evaluating model performance

For this particular problem, a custom metric was developed to take into account how far the prediction is from the real value.

Scoring Grid

Real Values \ Predicted Values	-1	0	4	2	3	4
-1	0	1	2	3	4	5
0	1	0	1	2	3	4
1	2	1	0	1	2	3
2	3	2	1	0	1	2
3	4	3	2	1	0	1
4	5	4	3	2	1	0

MLP

Average error: 0,689

Real Values \ Predicted Values	-1	0	1	2	3	4
-1	0.4	0.1	0.2	0	0	0
0	0.9	3.4	1.5	0.3	0.1	0
1	0.5	0.6	3	1.5	0.5	0.4
2	0.2	0.5	0.3	1.3	0.7	0.3
3	0.2	0.6	0.1	0.3	0.2	0.1
4	0.1	0.3	0.5	0.9	0	0

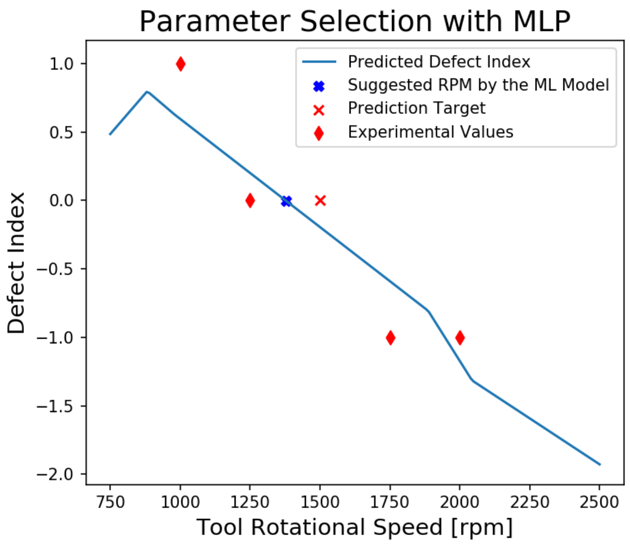
KNN

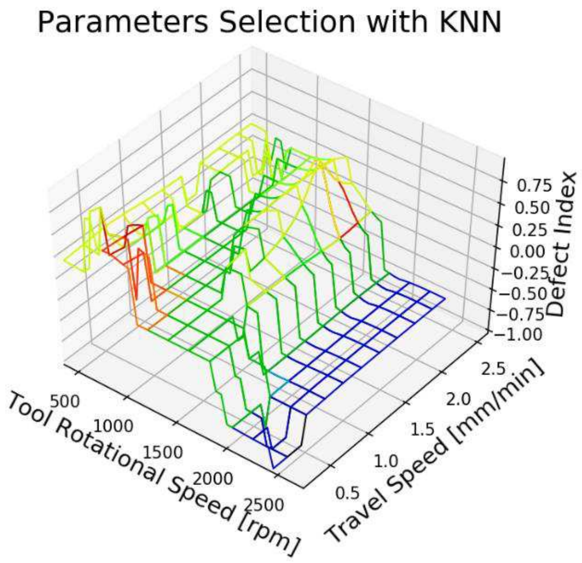
Average error: 0,553

Real Values \ Predicted Values	-1	0	1	2	3	4
-1	1.6	1	0.1	0	0	0
0	0.5	7.9	2.1	0.5	0.1	0
1	0	1.5	1.1	0.2	0.1	0.1
2	0	0.7	0.3	0.1	0.1	0.1
3	0	0.1	0.5	0.3	0	0.1
4	0	0.2	0.5	0.1	0.1	0

Machine Learning could be used to find ultivariate process windows

Objective: For a particular weld, find the parameters for lowest possible defect probability





Opportunities for future developments

- Improve predictive power by instrumenting the machines
- Meta parameters optimization (including predictors scaling)
- “Map” the parameters space with FEA and additional experimental tests (data twinning)
- Improve safety by implementing warnings if:
    - The prediction comes from FEA instead of Experimental values
    - The parameter space is sparse near the required prediction

On-road Corrosion data

On-road exposure provides a lot of data on the weather conditions and corrosion development of the various samples exposed

Data type – Info wirelessly transmitted	Number / 10 mins / vehicle
CFRP / AA6022, SS201LN / AA5083-SPF, Usibor 1500 (stamped, with Al-Si coating) / AA6022 galvanic assemblies, using bare material	9
Painted (Zinc phosphate + sealant, e-coat, primer, top coat, clear coat) & scribed AA6061, AA6022 (as rolled & sanded) and AA5083-SPF, with galvanic coupling with SS201LN: 4 data / 10 mins	4
Ambient temperature & %RH	2
Surface temperature	2
ToW & Conductivity	2
Vehicle position	1
Vehicle speed	1
U-bend specimens with crack detection	6
TOTAL	27 *

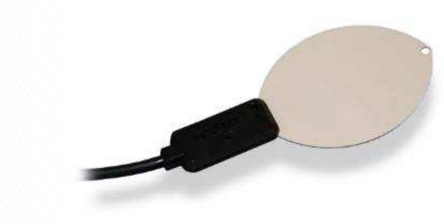
* 27 data / 10 mins / vehicle
≡ 3,888 data / day / vehicle
≡ 1,419,120 data / year / vehicle
≡ 5,676,480 data / year / 4 vehicles
≡ 14,191,200 data / 2 ½ years / 4 vehicles

ALTec "Development of a Design Support Tool and Accelerated Structural Integrity Qualification Procedure for Corrosion Resistance Assessment of Multi-Materials Assemblies Incorporating High Strength Aluminum Alloys", report #3, July 18th 2018, D. Gallant

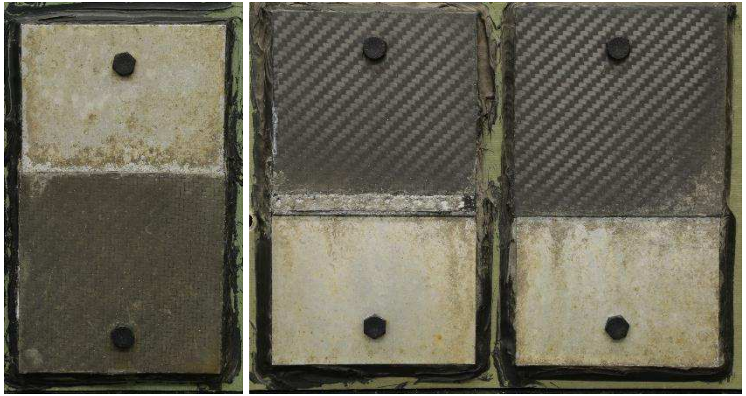
Development of a new lab corrosion chamber

- The first objective of the new chamber is to reproduce the environment and corrosion behavior encountered on the road.
- We need to reproduce the environmental conditions: temperature, relative humidity and wetness.
- Then, we need to make sure that the same environmental conditions produces the same corrosion behavior.

T/RH probe



Time of wetness sensor

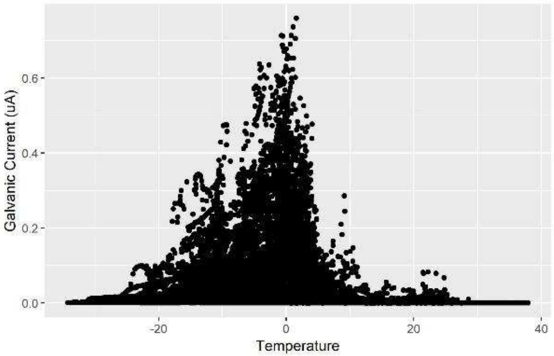


Galvanic corrosion samples

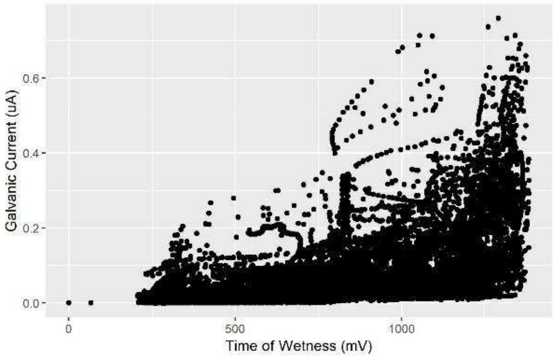
Development of a new lab corrosion chamber

To do this, let's take a one year dataset from one of the vehicles consisting of 53116 points. How do we choose which values to send to the chamber to test and calibrate it?

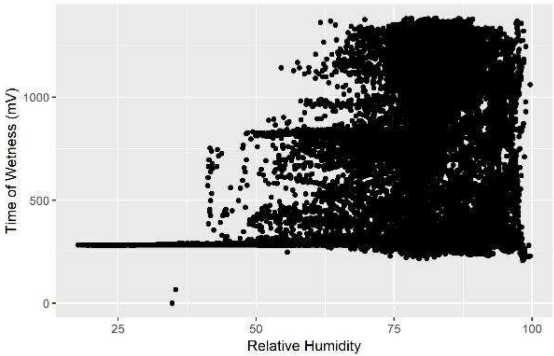
We need to find a way to reduce the 53116 points to a few values that best represent the whole dataset?



T-134 1 year data, Galvanic current vs temperature



T-134 1 year data, Galvanic current vs time of wetness



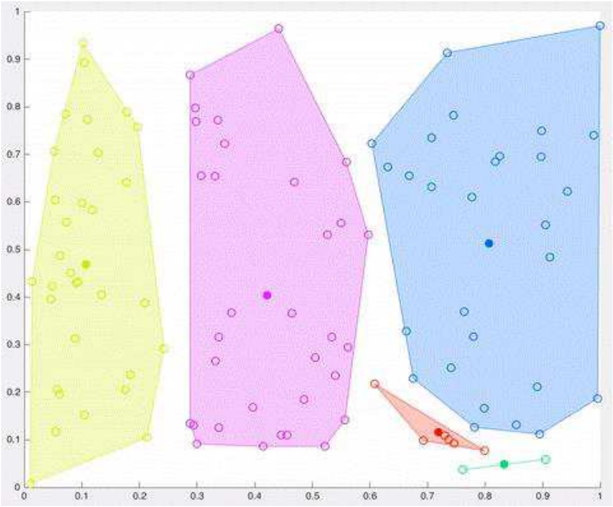
T-134 1 year data, Time of wetness vs relative humidity

K-means clustering

K-means clustering is an unsupervised technique that regroup similar data points into clusters. The similarity is evaluated using the Euclidean distances between the points.

How does K-means clustering works:

1. Initiate K random centroids;
2. Every data point is associated to the nearest centroid, creating the clusters;
3. The position of the centroid is then relocated to the center of the cluster using the mean value of all the cluster’s points;
4. Iterate between 2. and 3. until the position of the centroid do not change anymore.

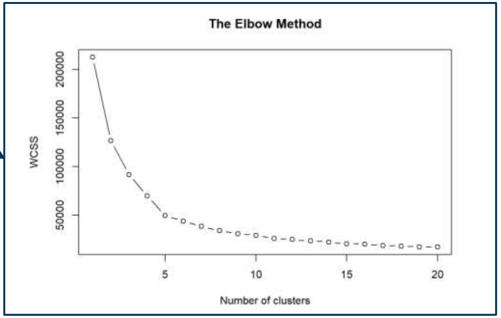


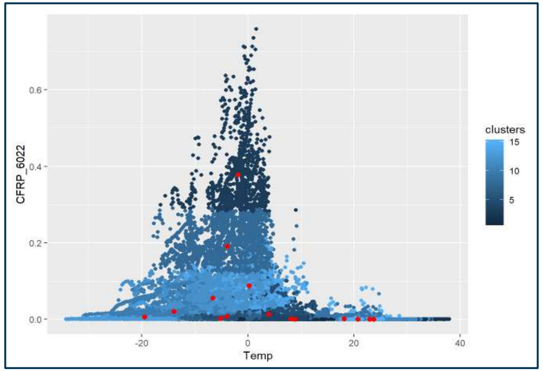
http://ai.stanford.edu

K-means clustering results

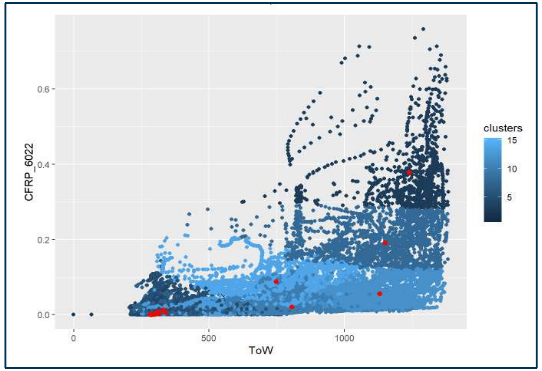
The optimal number of clusters is found by looking at the metric called Within Clusters Sum of Square

Once the optimal number of clusters is known, the clustering is applied and the clusters centroids values are used to calibrate the chamber

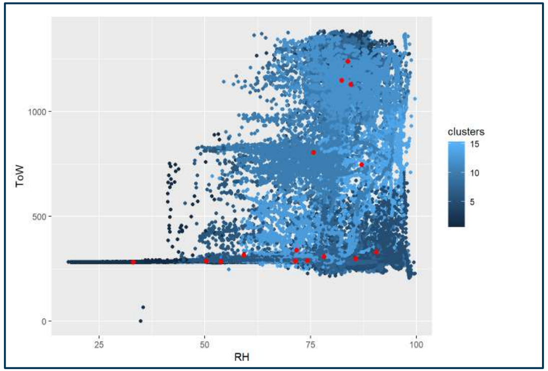




T-134 1 year data, Galvanic current vs temperature, with clustering



T-134 1 year data, Galvanic current vs time of wetness, with clustering



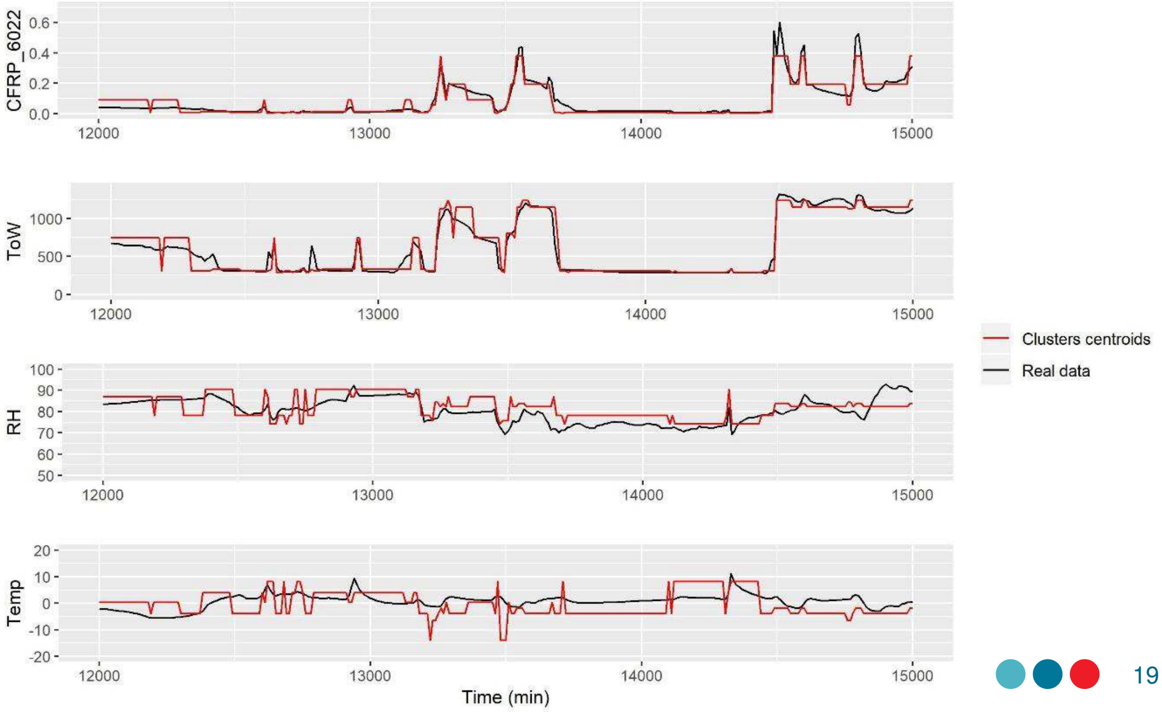
T-134 1 year data, Time of wetness vs relative humidity, with clustering

K-means clustering results

In reality, the data from the vehicles are time series. To validate how well the clusters represent the reality, the best way is to overlap the cluster centers values for each data point over the observed values

Clusters centroids values

	Temp	RH	ToW	CFRP_6022
1	8.9171643	50.26270	288.1716	3.413533e-04
2	-3.9008992	82.28551	1149.2753	1.914509e-01
3	-13.9415911	75.62794	804.8076	2.133191e-02
4	-5.1725809	59.17310	315.7853	2.012694e-03
5	20.7178756	71.41062	287.7615	3.284107e-04
6	-3.9910389	78.10910	309.2486	7.931014e-03
7	8.0601338	74.17808	289.8110	1.855113e-03
8	3.9385269	90.52851	329.9877	1.288955e-02
9	-19.5227145	71.65825	339.2690	6.631807e-03
10	-1.9210623	83.75269	1239.0953	3.785124e-01
11	18.1199816	85.61228	298.0711	1.126341e-03
12	22.9539303	33.05150	282.1629	1.602352e-05
13	0.2712112	86.97196	746.5987	8.843956e-02
14	23.6833980	53.81430	284.2089	7.543513e-05
15	-6.6173114	84.46848	1129.4778	5.587546e-02



Opportunities

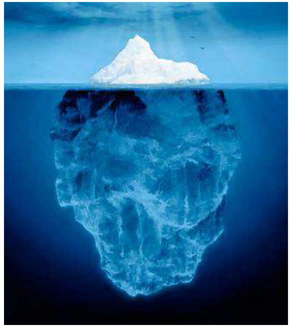
We have reduced the 53000 data points to 15, now we need to:

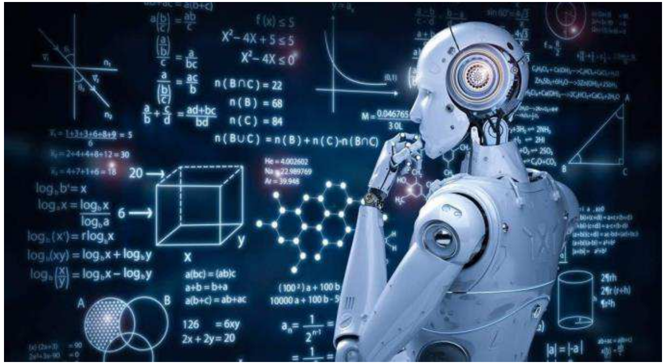
- Optimize the clustering
  - Test different scaling methods
  - Test other clustering methods
  - Using data from other vehicles/samples
- Using the clustering to test the new chamber (Can we reproduce real life behavior with only 15 values?)
- Duplicate the on-road conditions in real time in a simplified manner:

Real life data acquisition → Clustering → Lab

Conclusion

- Machine learning is a very useful tool to get the most out of your data as long as it is applied correctly.
- It can be used with almost any data given some preprocessing to prepare them.
- If you have a lot of data that you don't know what to do with and how to get the most out of it. Chances are good that machine learning can help you!
- In manufacturing, field engineers and technicians are the best placed to accomplish this task since a deep understanding of the process is needed to implement machine learning in a meaningful way





THANK YOU

Marc-Olivier Gagné • Marc-Olivier.Gagne@nrc-cnrc.gc.ca

Benoit Thériault • Benoit.Theriault@nrc-cnrc.gc.ca