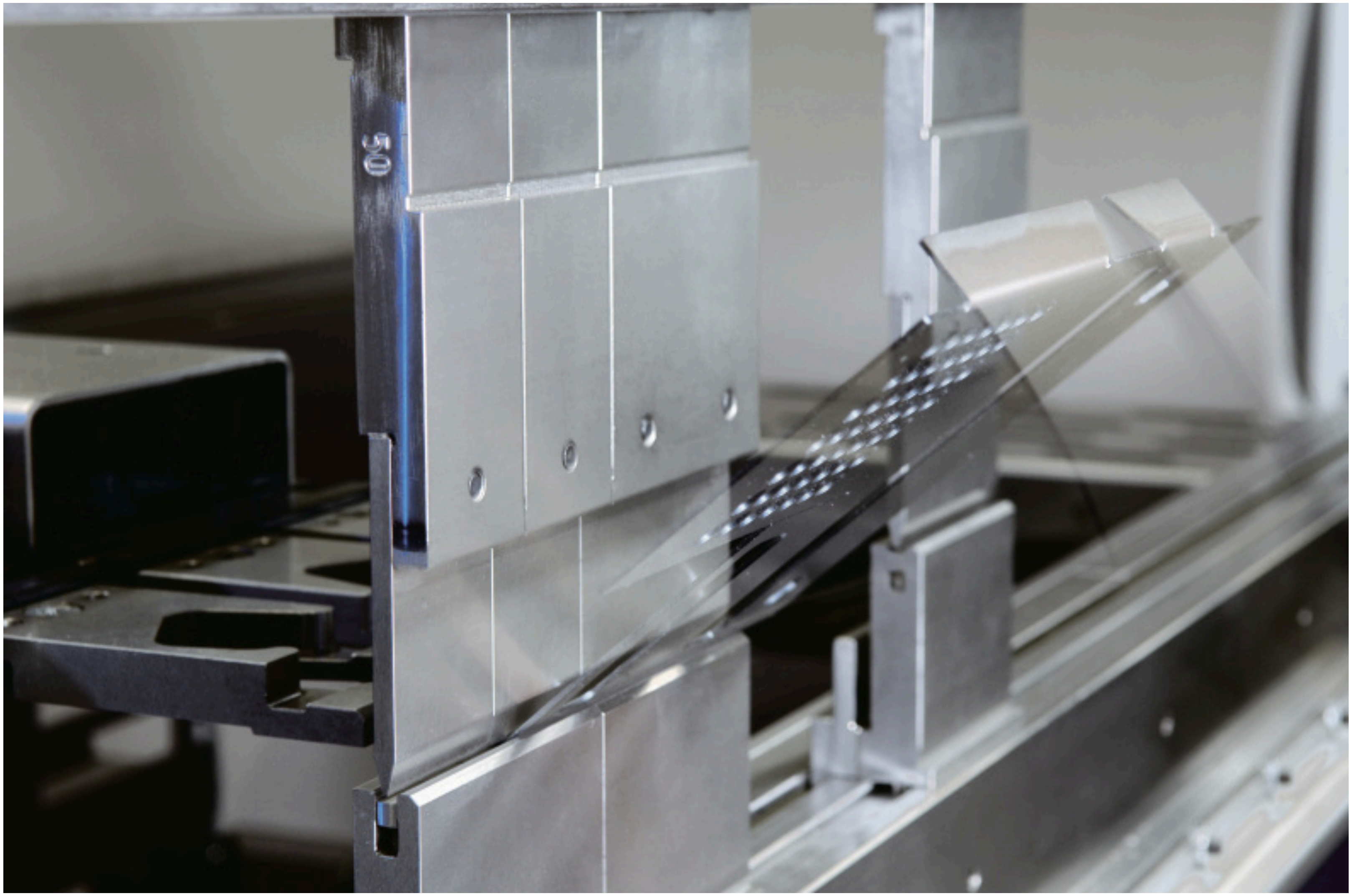
# Sheet metal
# Cost calculator



# Agenda

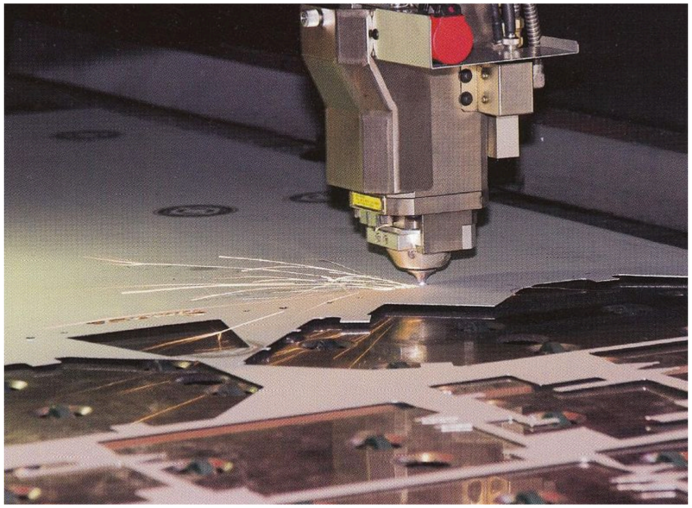
1. Project definition
2. Formed part process
3. Product design Process
4. Dataset creation
5. Dataset analysis
6. Prediction system
7. Conclusions

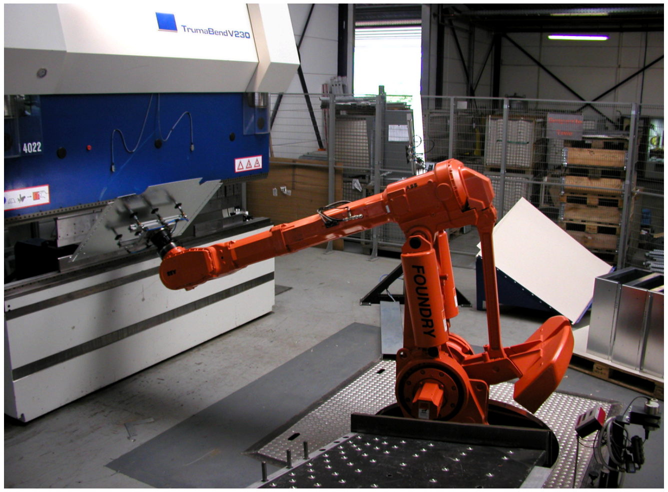
# Project definition

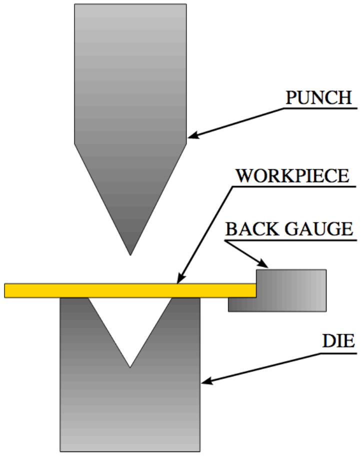
- Study limited to sheet metal parts
- Extract information from CAD files (.dxf)
- Define a cost calculator system for sheet metal parts
- Create a user friendly program

# Sheet metal forming

- The CAD system generates a flat contour of the sheet metal part in DXF
- The laser cutting machine will cut patterns in raw material
- If needed, bends are added with the press brake

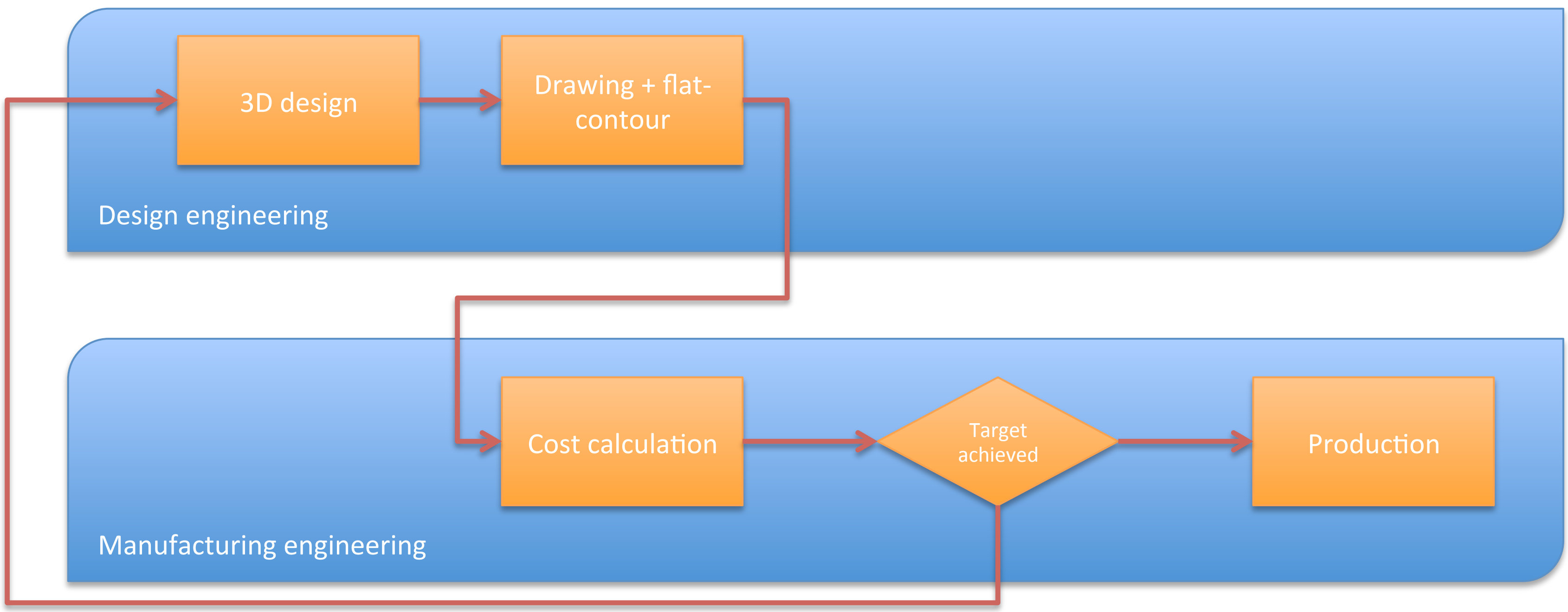




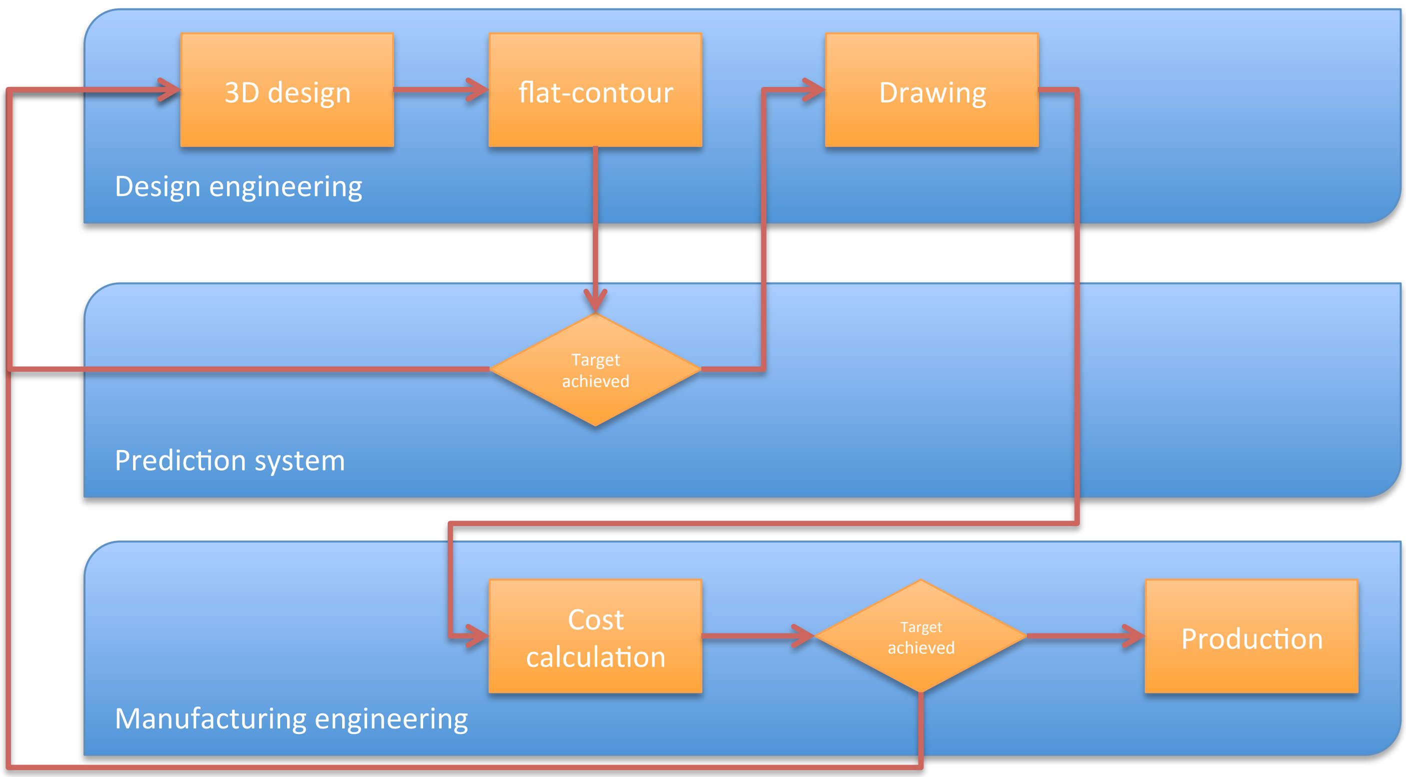
# Standard design process

- DE must create drawings of an uncertain design for ME in order to get a cost
- ME may work on designs too expensive
- ME may have other priorities



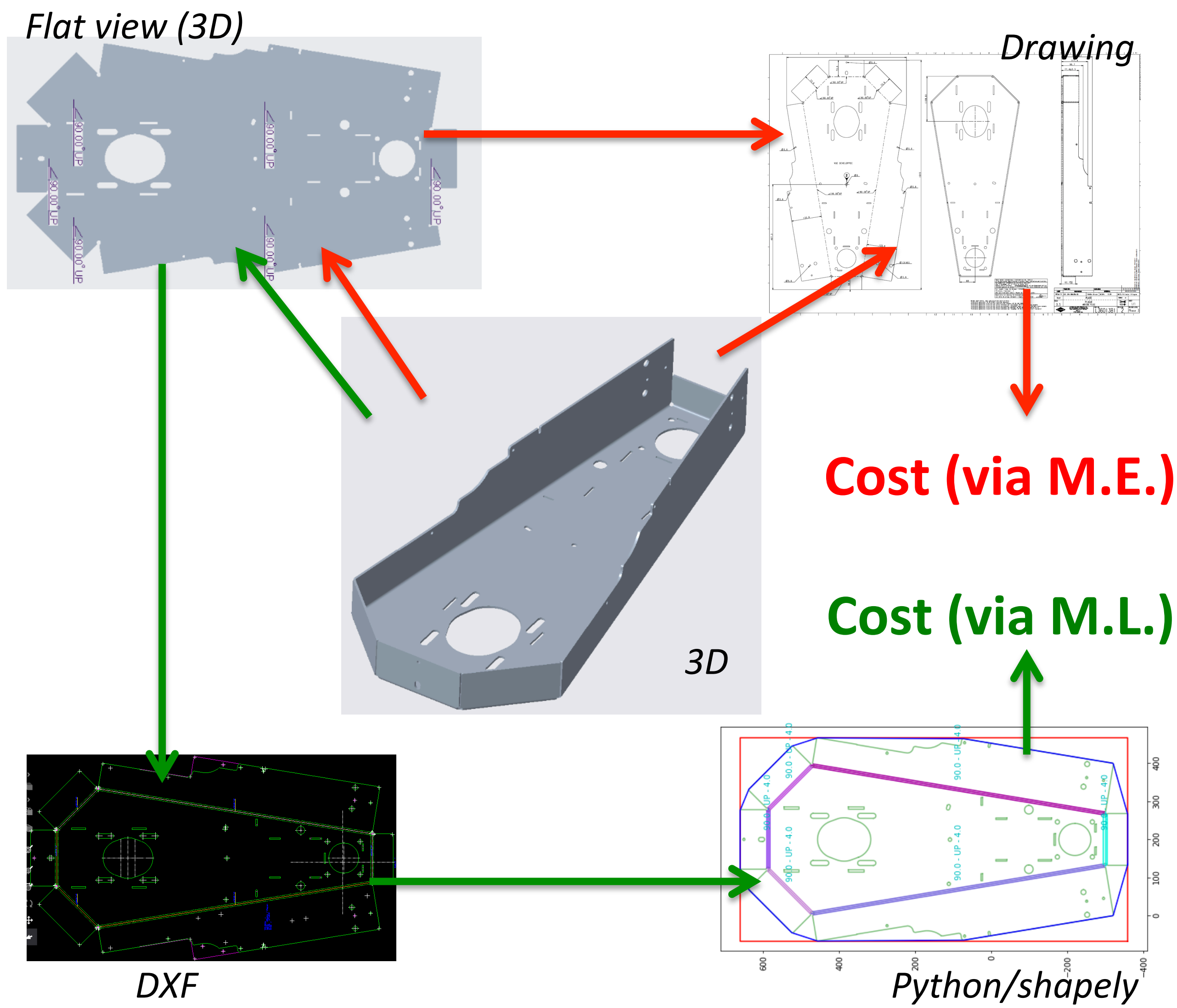
# Improved design process

- DE will only work on drawing after cost prediction
- ME will only work on design likely to achieve target price



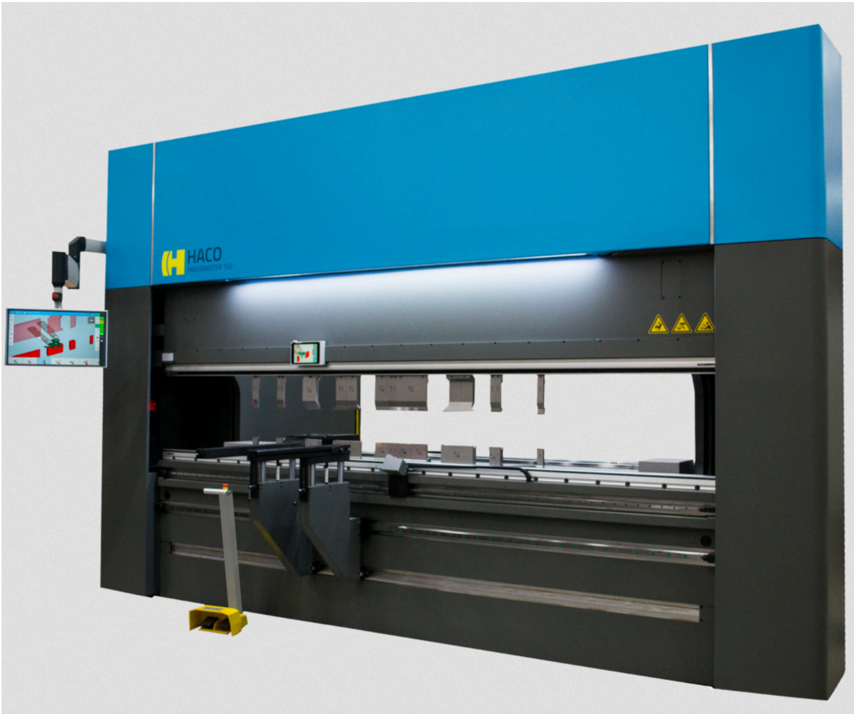
# Different part representations

- 3D design
  - Created by design engineer with CAD system
- Flat view (3D)
  - Generated by CAD system
- Drawing
  - Created by (DE) for (ME) and production
- DXF (2D)
  - Standard file, adapted for our needs
- Python/Shapely
  - Used for our study



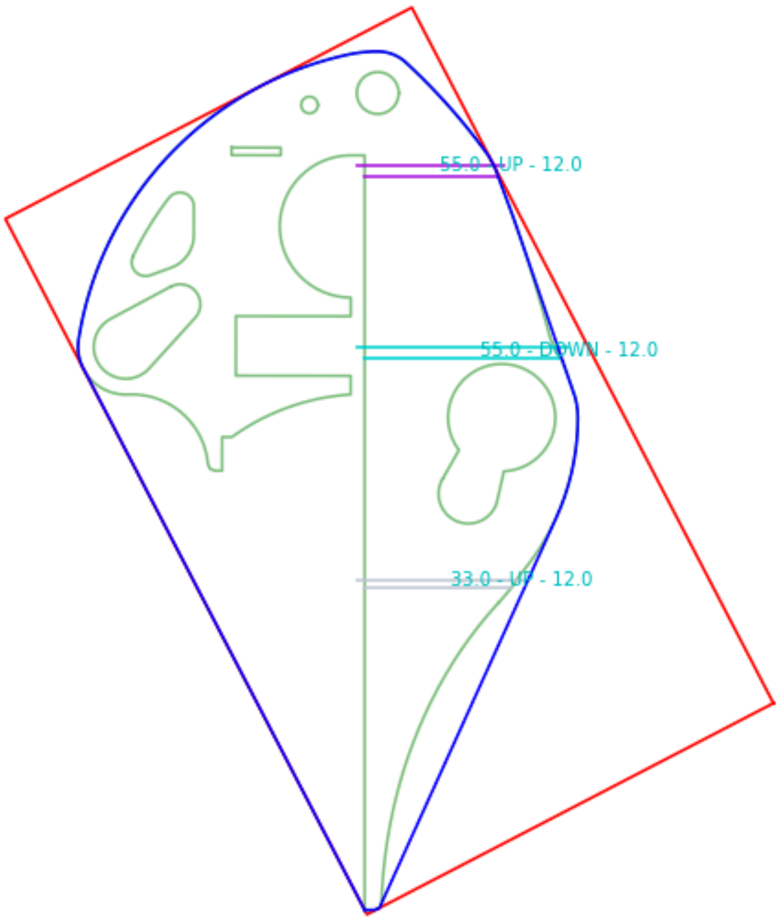
# Data available

- Approx. 11000 DXF files
- Information extracted from DXF:
  - Pattern counts
  - Bends
  - Thickness
  - Material
  - Areas



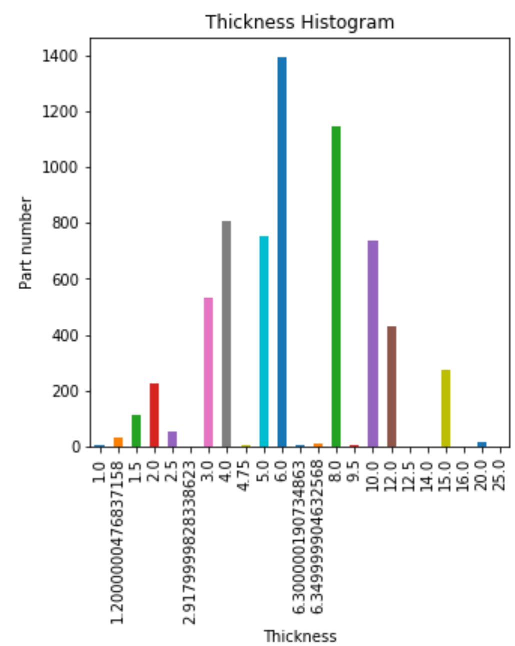
# DXF parser

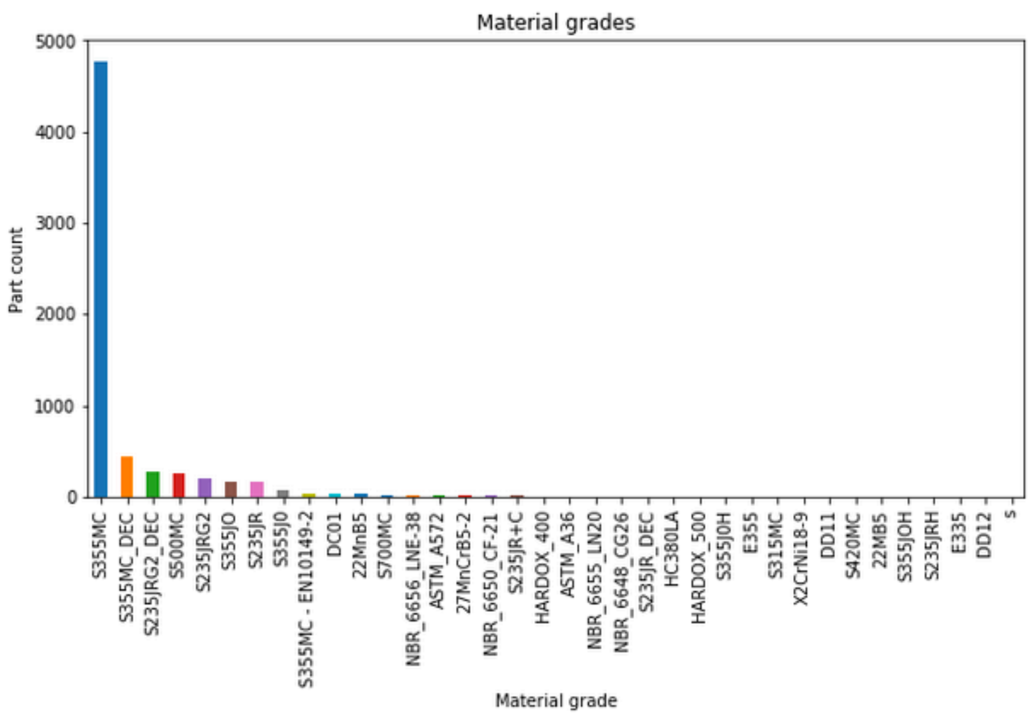
- Improved DXF contour file used, more details [here]()
- General parsing procedure:
  - Read general information (thickness, material)
  - List all the edges entities
  - Re-construct the patterns
  - Calculate areas and count patterns
  - Detect bends
  - Analyze bends (placement, length…)



# Dataset analysis

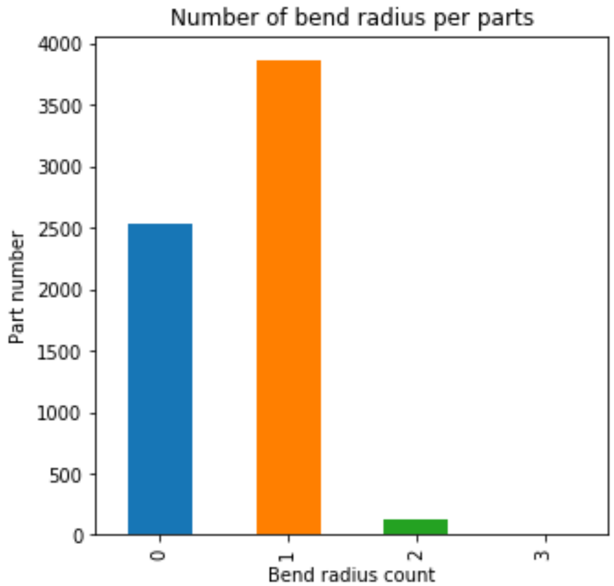
- Some parts with imperial units thickness
- Most popular material is S355

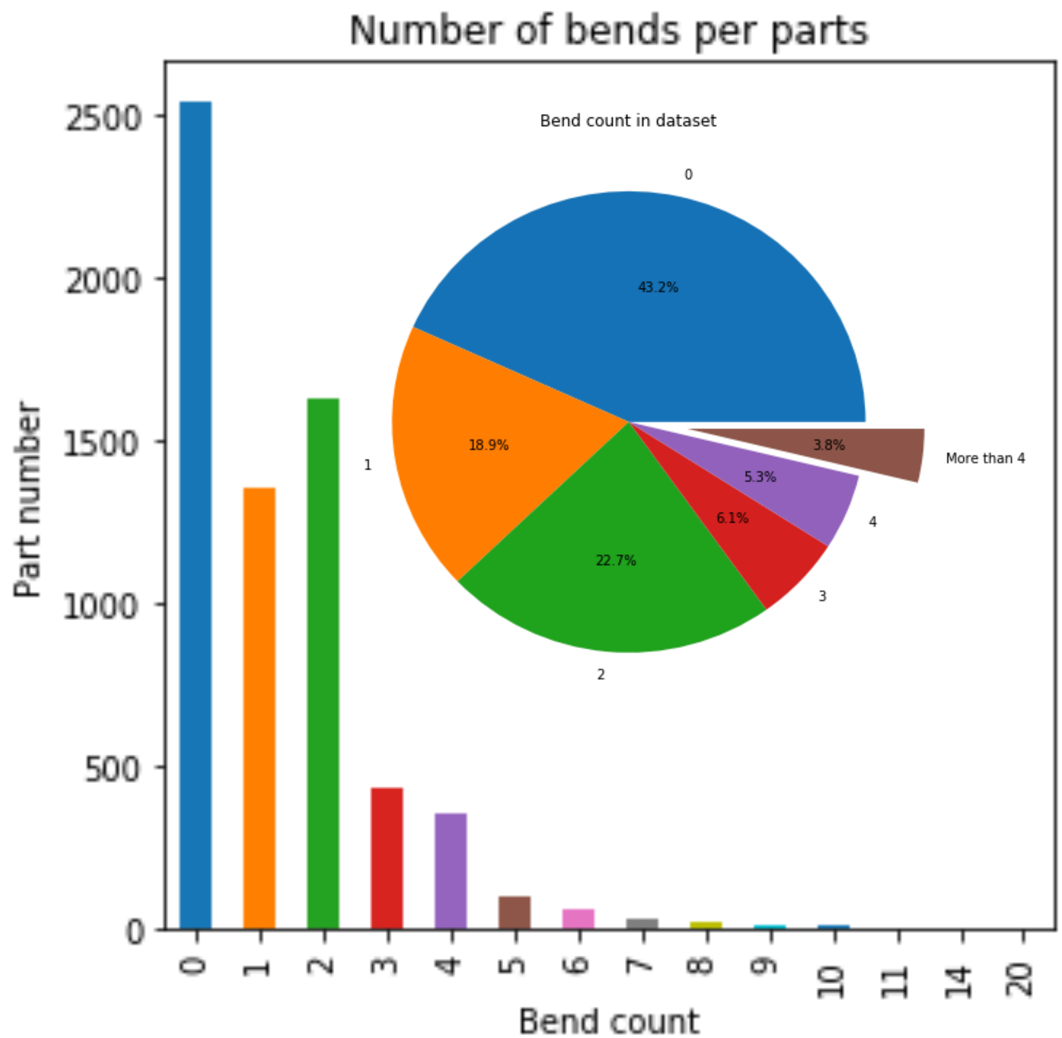


# Dataset analysis - bends

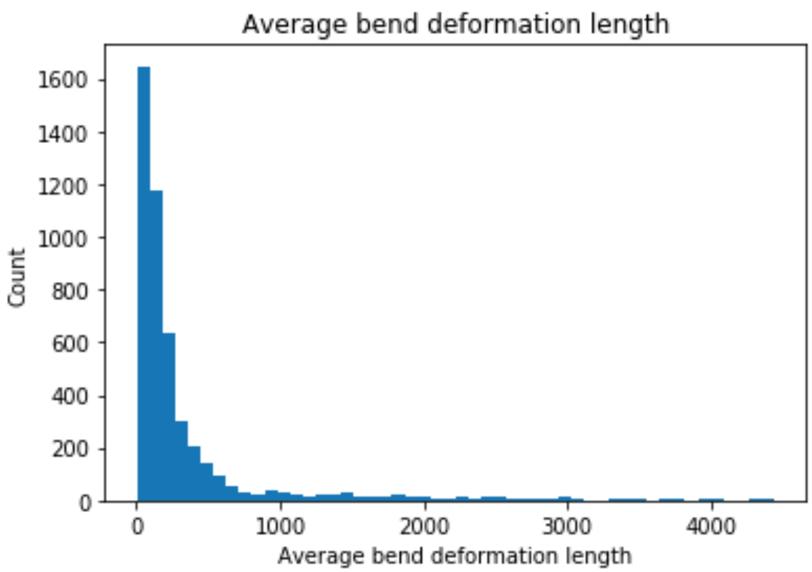
- Some rare parts have more than 1 bend radius
- 40% of the parts have at least 2 bends
- 4% of the parts have more than 4 bends



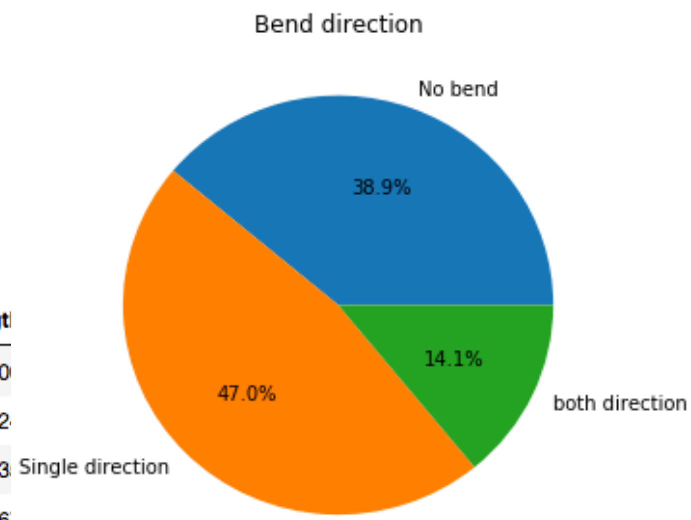



# Data analysis - bends

- Bend length is up to 4 meters
- 50% of the bends have a length inferior to 140mm
- 14% of parts must be flipped during forming operation




| | avg_deformation_lengt |
|---|---|
| count | 4745.00000 |
| mean | 340.24272 |
| std | 593.36003 |
| min | 2.66666 |
| 25% | 70.000000 |
| 50% | 139.387005 |
| 75% | 297.649699 |
| max | 4441.000000 |

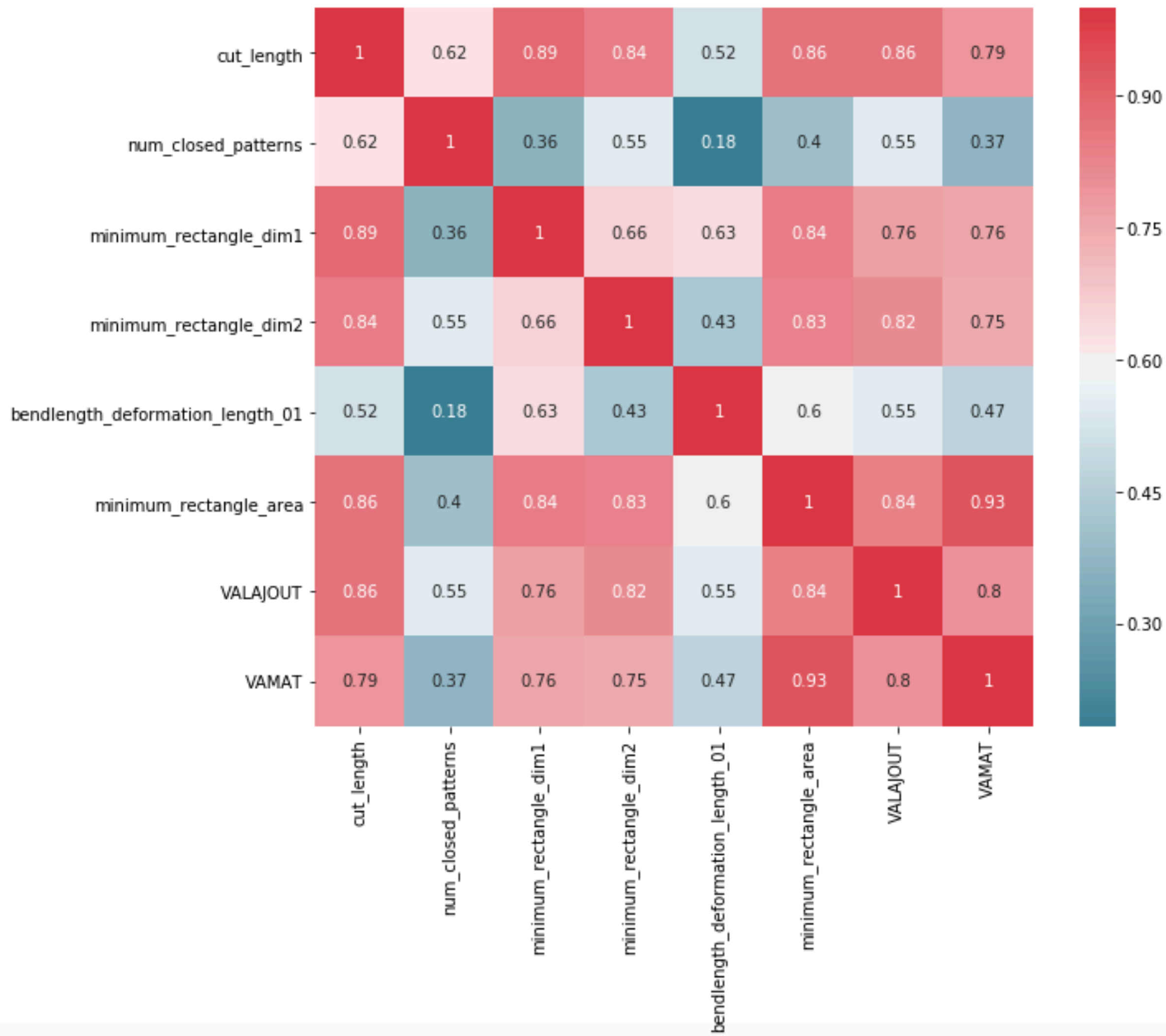


# Dataset analysis – thickness 6mm

OPENCLASSROOMS

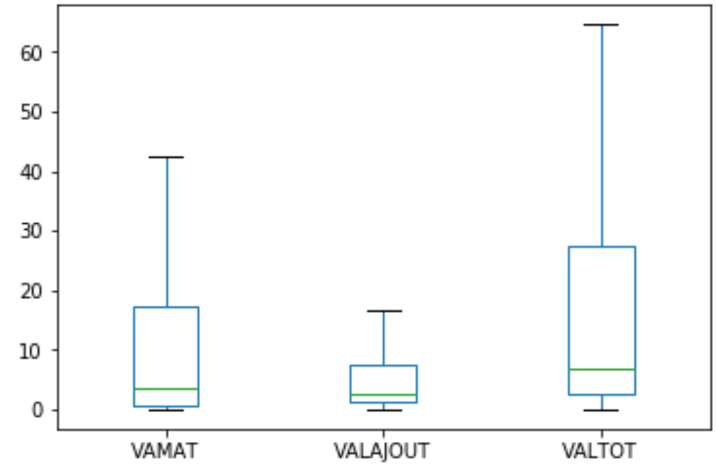
CentraleSupélec

- VALAJOUT and VAMAT are correlated
- Cut length and minimum rectangle dimensions have high impact on cost



# Dataset analysis - Cost

- Cost composition:
  - VAMAT: material cost
  - VALAJOUT: Added value (HR, amortization…)
  - VALTOT: VAMAT + VALAJOUT
- Outliers dropped

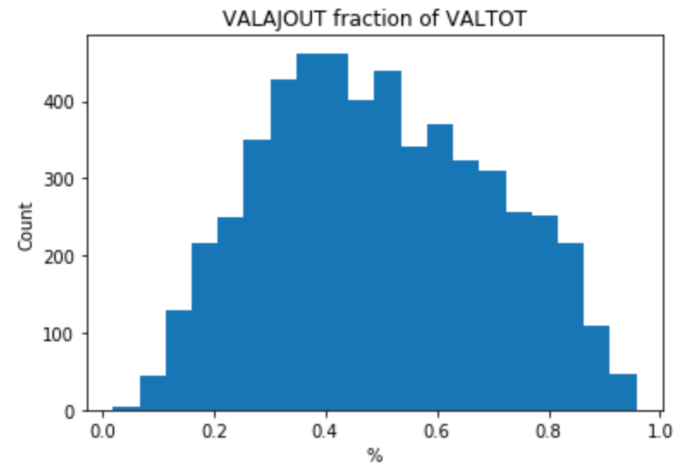



| | VAMAT | VALAJOUT | VALTOT |
|---|---|---|---|
| count | 31067.000000 | 31067.000000 | 31067.000000 |
| mean | 122.925028 | 71.520884 | 194.445912 |
| std | 532.884074 | 303.707643 | 819.256172 |
| min | 0.000000 | 0.000000 | 0.000000 |
| 25% | 0.700000 | 1.170000 | 2.520000 |
| 50% | 3.500000 | 2.670000 | 6.820000 |
| 75% | 17.390000 | 7.370000 | 27.380000 |
| max | 15597.310000 | 5468.790000 | 20659.900000 |

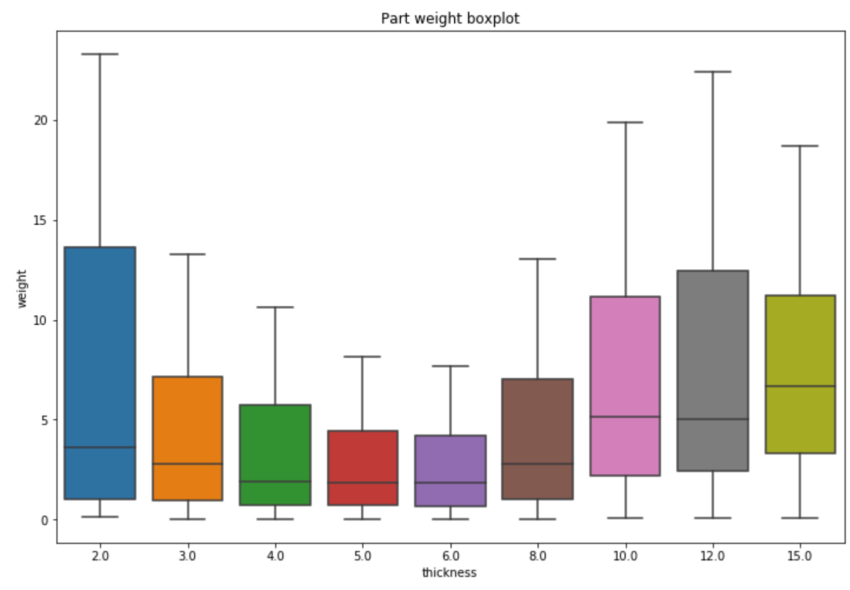
```
# and nothing should be more expensive than ... euros...
priceset = priceset[priceset['VAMAT']<45]
priceset = priceset[priceset['VALAJOUT']<20]
priceset = priceset[priceset['VALTOT']<65]
```

# Dataset analysis - Cost

- VALAJOUT is usually about 50% of VALTOT
- The parts weight is usually in a range that makes manual handling possible

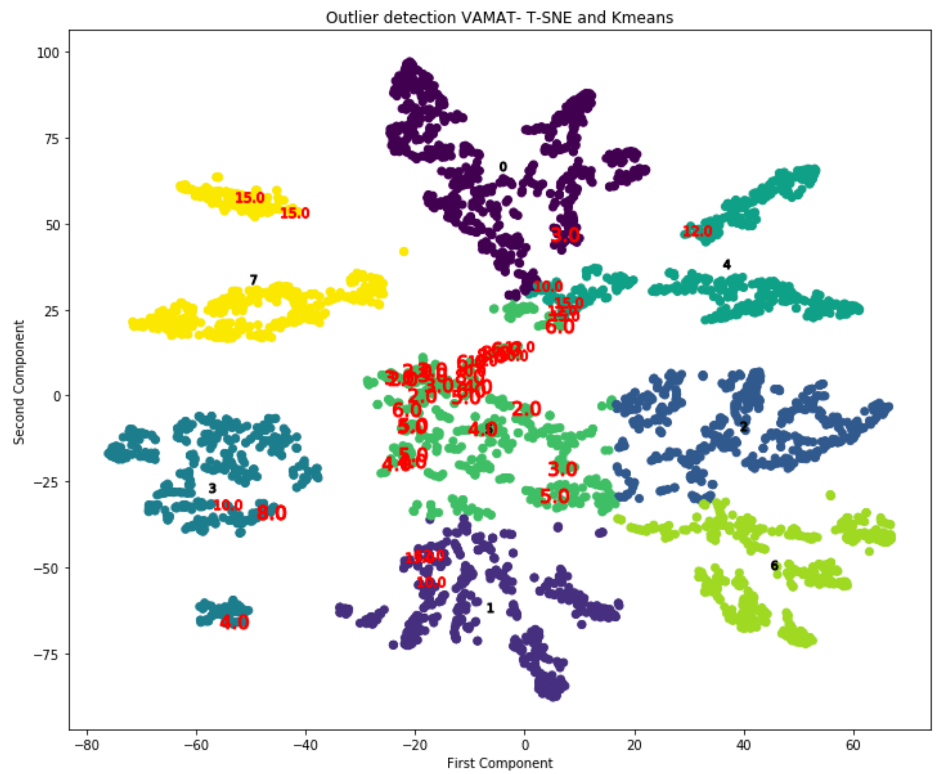



| | weight | | |
|---|---|---|---|
| | mean | median | count |
| **thickness** | | | |
| **2.0** | 7.361038 | 3.610330 | 115 |
| **3.0** | 5.206124 | 2.788439 | 438 |
| **4.0** | 5.844878 | 1.914116 | 694 |
| **5.0** | 4.681015 | 1.816365 | 631 |
| **6.0** | 4.499785 | 1.838196 | 1197 |
| **8.0** | 5.768910 | 2.770102 | 995 |
| **10.0** | 9.901327 | 5.173511 | 687 |
| **12.0** | 9.504931 | 5.036549 | 384 |
| **15.0** | 11.022819 | 6.704310 | 217 |



# Outlier detection

- Possible outlier have been detected in the dataset (poor prediction)
- K-Means clustering with T-SNE decomposition may help to detect unusual parts
- No access to the part routing to pursue investigation
- Data cleaned with quantiles (Top 5% and Bottom 2%) for estimator training



# Data cleaning

- Error may occur during parsing:
  - Improper DXF (-500 files)
  - No thickness (-1000 files)
  - Negative values
- Parts with no cost or possible outliers
  - Routing errors in dataset
- Only 1 one material grade (S355)
- At least 50 occurrences of each thickness

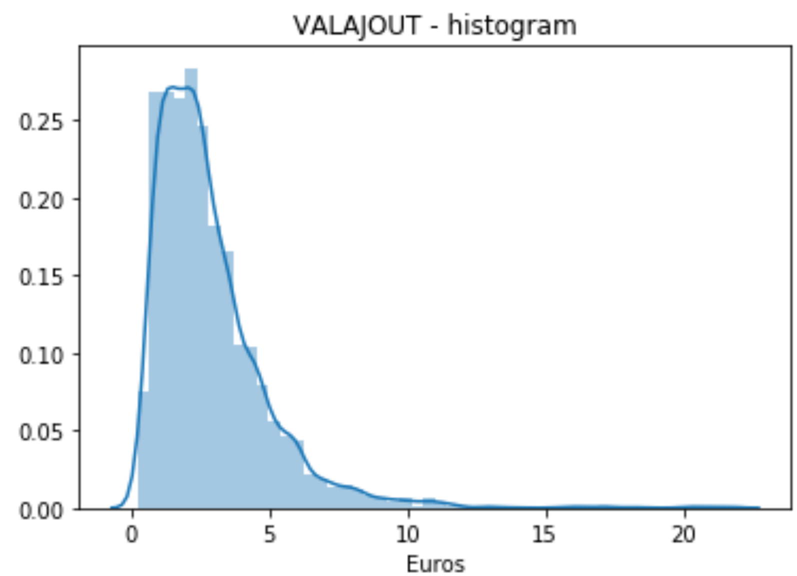
- Final dataset approx. 5000 parts

# Metrics and cross validation

- MSE (mean squared error)
- MAPE (mean absolute percentage error)
- BOXCOX treatment used to un-skew the dataset
- Test set size = 30% dataset
- Calculation done with 3 folds
- Possible outliers suppressed
- Estimator optimized with gridsearch and hyperopt package

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right|$$



# Baseline definition

- The cost can be crudely defined with linear algebra:

*VALTOT= “surface” X “K1” +*

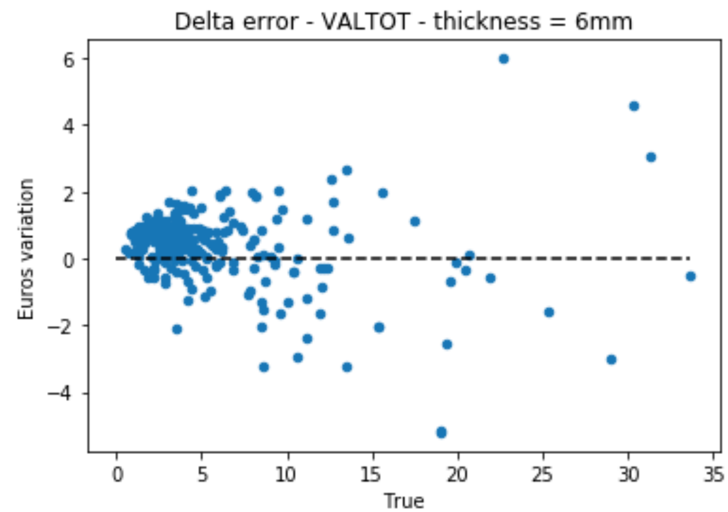
*“Number of bends” X “K2*

| thickness | K1 | K2 |
|---|---|---|
| 2.0 | 0.000027 | 0.381410 |
| 3.0 | 0.000033 | 0.835805 |
| 4.0 | 0.000039 | 0.656602 |
| 5.0 | 0.000051 | 0.775819 |
| 6.0 | 0.000058 | 0.816917 |
| 8.0 | 0.000081 | 0.795512 |
| 10.0 | 0.000099 | 0.896394 |
| 12.0 | 0.000108 | 0.432202 |
| 15.0 | 0.000123 | 1.215107 |

- MSE : 3.59 (euros$^2$)
- MAPE : 21.27 (%)

# Baseline definition

- Method simple requiring no special M.L. tools
- Solution can be tedious, requires to measure the smallest possible rectangle
- Bends cost inconsistent




| | MSE_s2 | | MAPE_error_s2 | | COART |
|---|---|---|---|---|---|
| | mean | max | mean | max | count |
| bend_group | | | | | |
| 0 | 1.077050 | 35.748383 | 38.470766 | 86.914728 | 117 |
| 1-2 | 0.918265 | 10.511479 | 17.256390 | 49.832241 | 159 |
| 3-5 | 2.646685 | 21.113920 | 13.512030 | 59.611360 | 26 |
| 6+ | 20.857343 | 27.010366 | 21.593645 | 27.367824 | 3 |

# Light GBM presentation

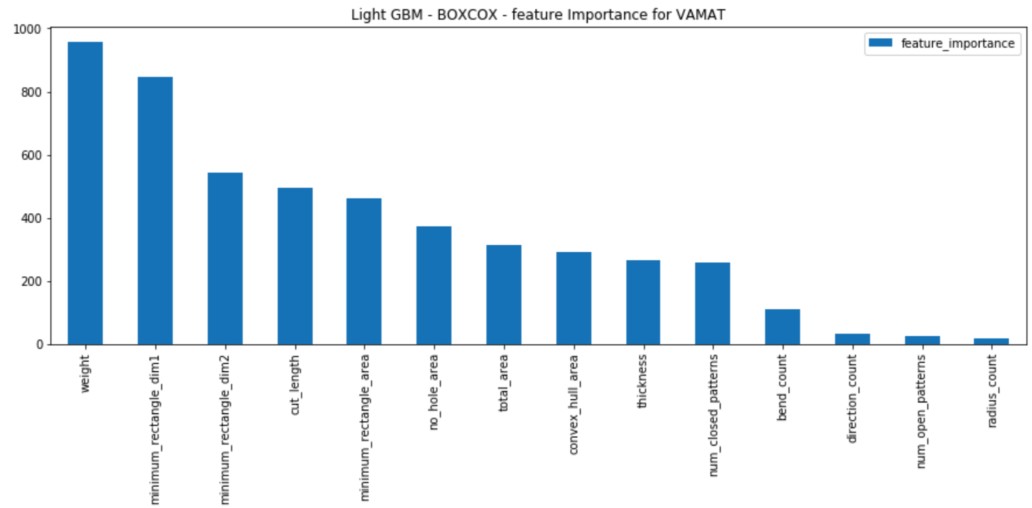
- Light GBM is recent algorithm using ensemble methods and gradient boost
- Results similar to the really popular XGBoost algorithm
- Embedded with GOSS and EFB features
- Trees growing leaf-wise
- Much… much faster to train than XGBoost (allows more optimization runs)

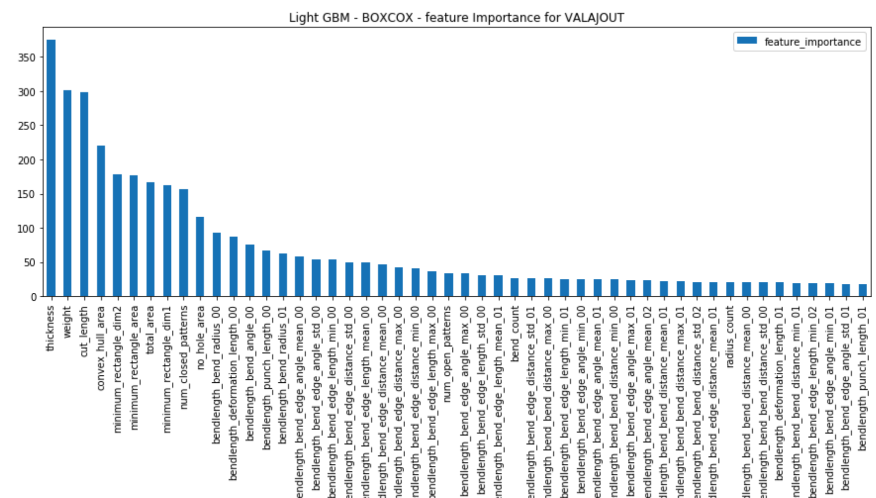
# Engineered features

- VAMAT and VALAJOUT will use 2 different dataset
- VAMAT size : 14 features:
  - (cut length, general dimension, thickness…)
- VALAJOUT size : 144 features:
  - VAMAT features
  - 5 first bend information ordered by length (angle, radius, deformed length…)

# Light GBM

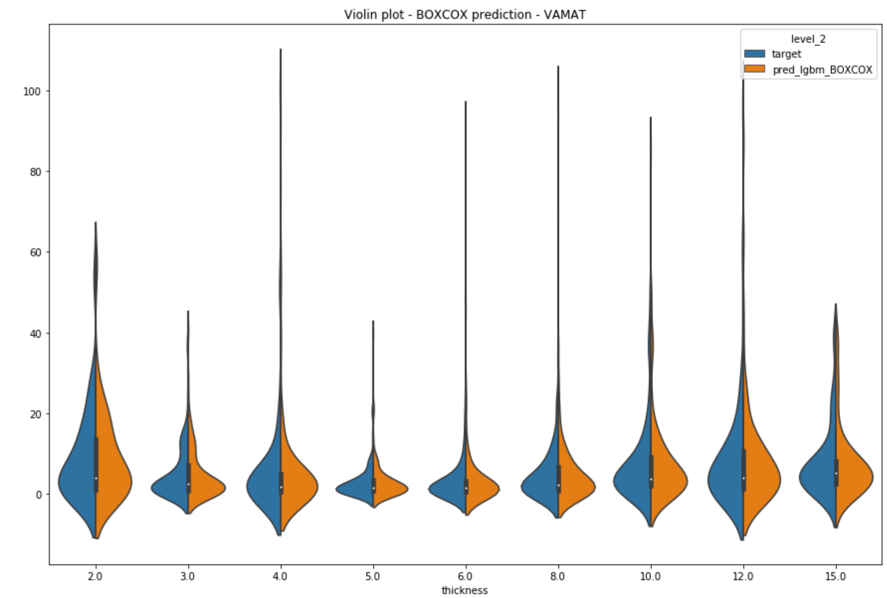
- 135 fits done in 2m44s (i5-2.7GHz)
- The most important features are common to both predicator
- Main Parameter used:
  - Num_leaves
  - Min_data_leaves
  - Alpha and lambda regularization
  - Min_child_weight
  - Bagging parameters

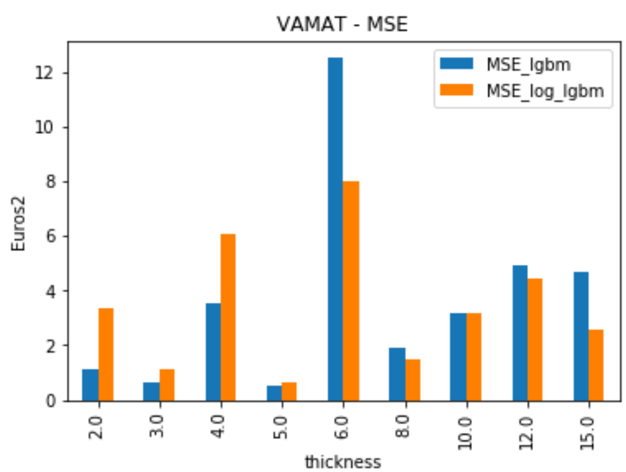


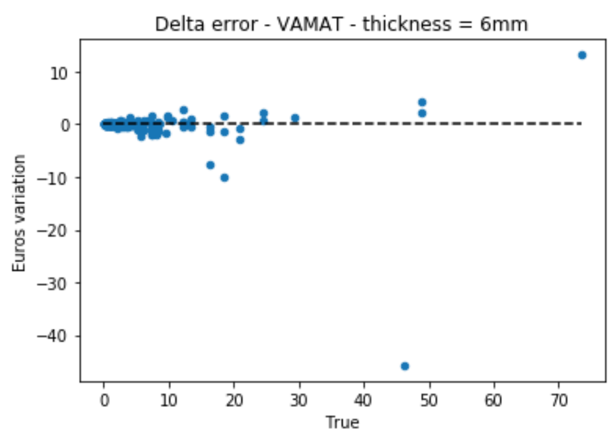



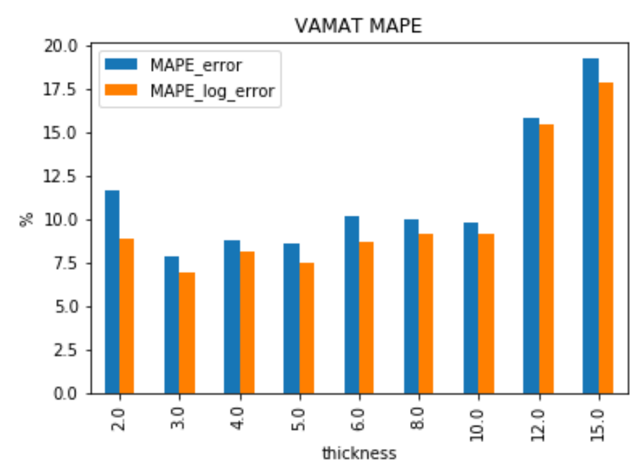
# Result analysis - VAMAT

- MSE : 3.83 (euros$^2$)
- MAPE : 9.22 (%)
- Error not evenly spread across different thicknesses



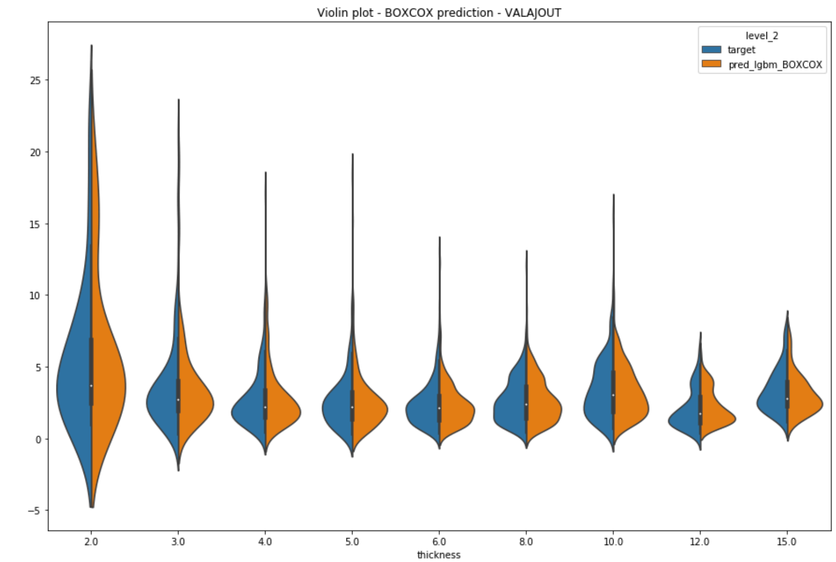



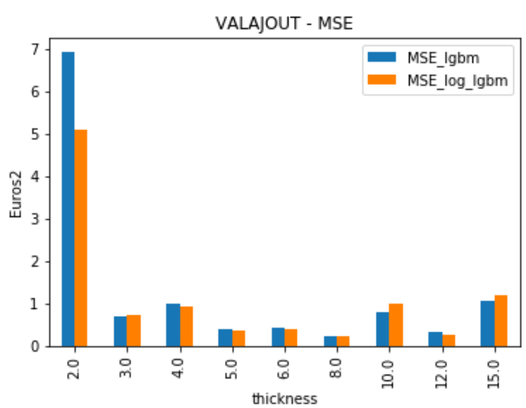



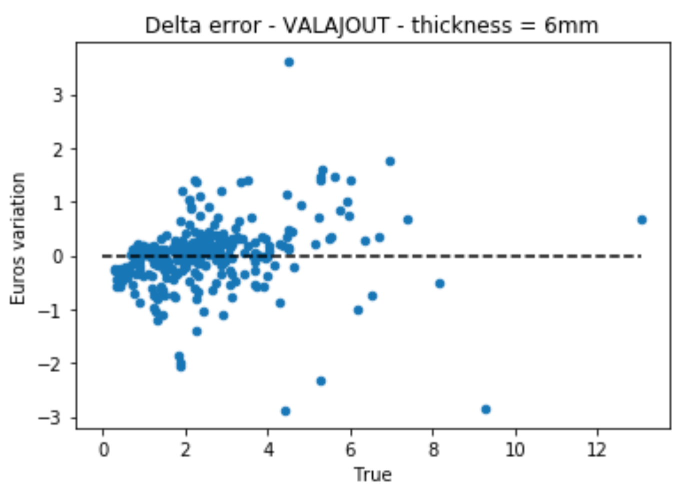



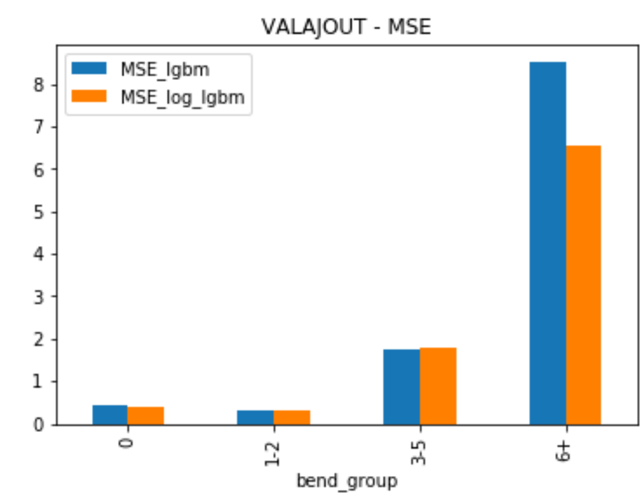
# Result analysis - VALAJOUT

- MSE : 0.62 (euros$^2$)
- MAPE : 19.68 (%)
- Algorithm showing poor result with parts having more than 6 bends (probably due to a dataset too small for those parts)



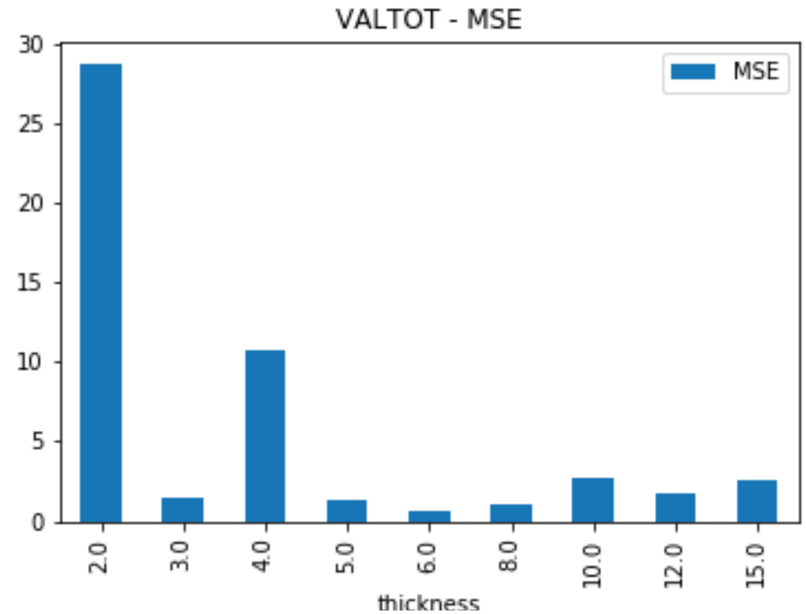



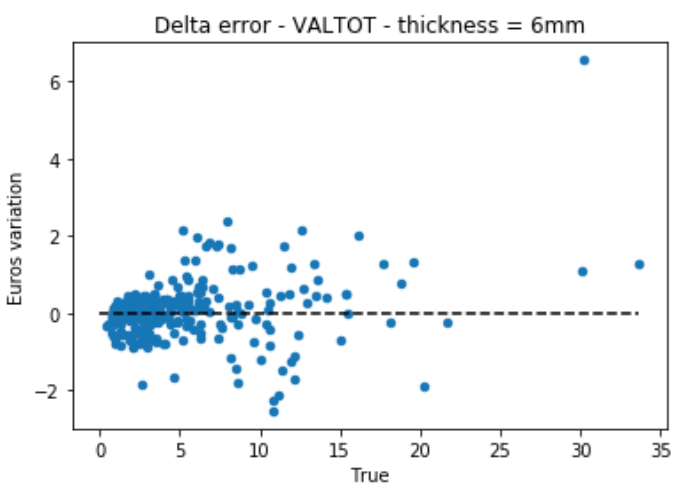



# Result analysis - VALTOT

- VALTOT = VAMAT+VALAJOUT
- Test set size : 1295
- MSE : 3.26 (euros$^2$)
- MAPE : 11.71 (%)
- Highest errors for thickness=2mm



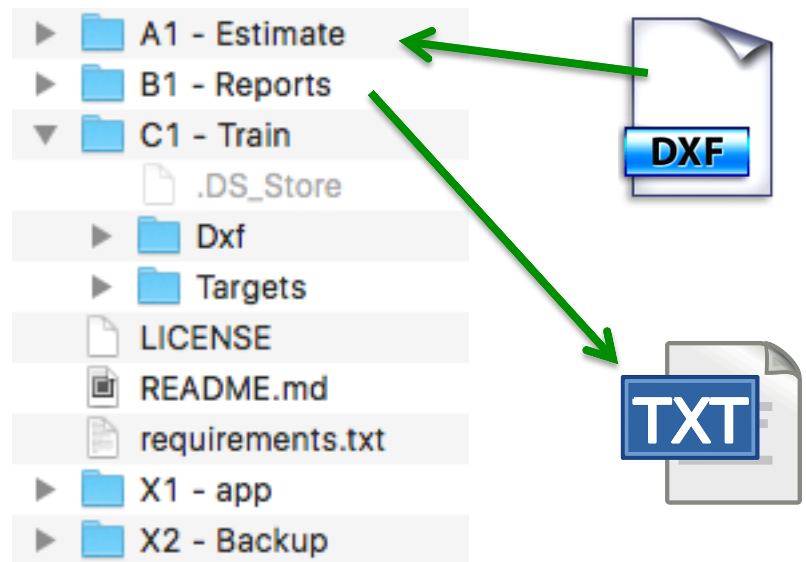




| | MSE | | MAPE | | COART |
|---|---|---|---|---|---|
| | mean | max | mean | max | count |
| bend_group | | | | | |
| 0 | 0.283994 | 6.402269 | 13.336109 | 80.740488 | 112 |
| 1-2 | 0.466230 | 5.769868 | 10.446275 | 65.530253 | 149 |
| 3-5 | 2.185605 | 42.917305 | 11.384887 | 70.584991 | 37 |
| 6+ | 0.217894 | 0.217894 | 3.455154 | 3.455154 | 1 |

*MSE and MAPE for thickness=6mm*

# API presentation

- Git-Hub link here
- (DE) will drop DXFs in folder A1
- Program generates result report in folder B1
- (ME) uses folder C1 to train model
- Trained model saved in folder X2



# API presentation

- Prediction report will contain various information:
  - Costs and accuracy observed on test set
  - Bend count
  - Patterns count
  - Cut length
- Training report will contain:
  - Performance report
  - Worst predictions on testing set

```
###################################
##         l1600620_flat1        ##
###################################

Date : 2018-05-05 00:25

VAMAT          (euros) : 8.99355306126939
MSE       (+/- euros2) : 0.8925799543406079
MAPE           (+/- %) : 7.512572323977111

VALAJOUT       (euros) : 4.588705394023897
MSE       (+/- euros2) : 0.18168545616218
MAPE           (+/- %) : 14.87350159625641

VALTOT         (euros) : 13.582258455293287

###################################
##          GENERAL INFO         ##
###################################

Thickness                (mm) : 3.0
Rectangle dim1           (mm) : 1140.0
Rectangle dim2           (mm) : 346.314
Rectangle surface       (mm2) : 394797.96
Closed pattern count     (mm) : 1
Open pattern count            : 0
Bend count                    : 2
```

*Prediction report sample*

# Conclusions and future directions

- MSE only slightly improved, but MAPE improved by 50%
- The algorithm can be used by (DE) to speed-up design process and (ME) to confirm cost
- The cost of a part is usually calculated within a second
- Outliers must be detected before training to improve results

***Baseline:***
*MSE : 3.59 (euros$^2$)*
*MAPE : 21.27(%)*

***Light GBM:***
*MSE : 3.26 (euros$^2$)*
*MAPE : 11.71 (%)*

# Conclusions and future directions

- This study is only the first step of a more global cost prediction system covering other processes such as welding, painting or machining
- Research papers have been found describing technics to read CAD files (STL), those technics will request a much bigger dataset to be used for cost prediction
- While final MAPE score may seem high, the final product is a sum of multiple parts that will dilute the error
- The dataset must increase in size to improve performance for parts with many bends or rare thickness

# Questions ?