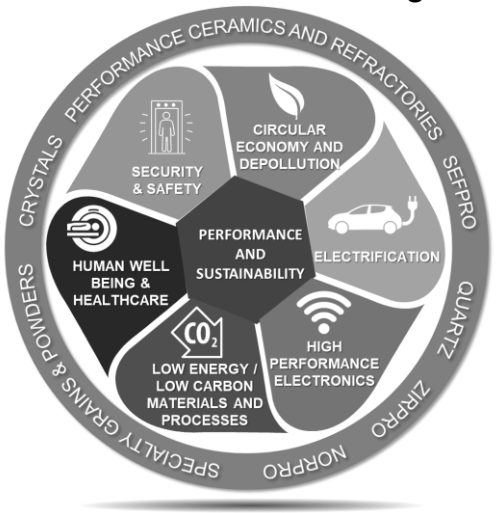
TRANSVERSAL R&D CENTERS BY SAINT-GOBAIN

OUTLINE

- Data project life cycle
- Saint-Gobain examples

CERAMICS *GRAND CHALLENGES*
Together we engineer a better, safer & greener world



CERAMIC MATERIALS ENGINEERING

THE ALLIANCE OF PERFORMANCE AND SUSTAINABILITY TO IMPROVE LIFE FOR ALL

SAINT-GOBAIN RESEARCH PROVENCE
Smart Manufacturing team

Y. EL HOUSNI, M. BOINET

SAINT-GOBAIN RESEARCH PROVENCE

SAINT-GOBAIN

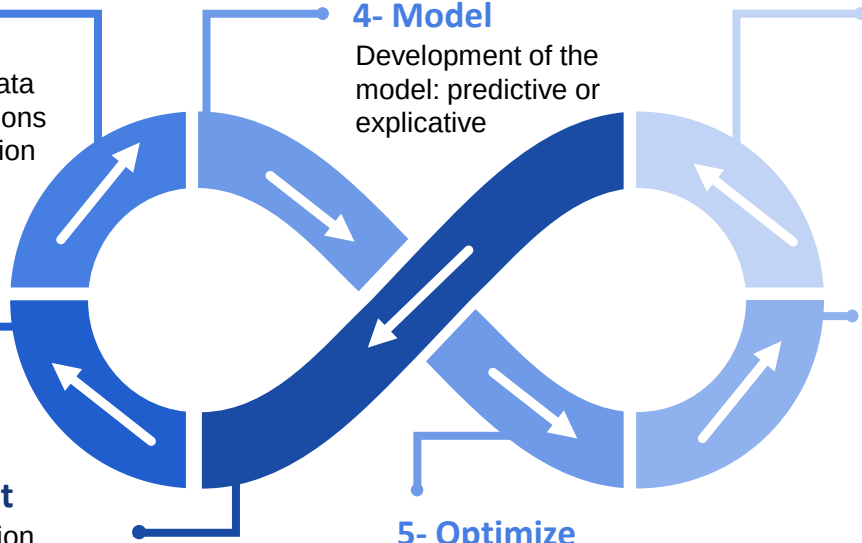
DATA PROJECT LIFECYCLE

Define / Measure / Analyse / Improve / Control

3- Explore
Dynamic visualization
Find the first patterns in the data
Create drag & drop visualizations with embedded data preparation

4- Model
Development of the model: predictive or explicative

7- Sustain & control
Model management and governance to follow model accuracy and continuously improve



2- Prepare
Collect & store data
Structure data in JMP
Clean data

6- Implement & act
Realize the value from analytics assets in action

1- Define the project
Yield / process optimization
Product quality
Energy optimization

5- Optimize
Embedding single or multiple advanced analytics models to optimize the performance & sustainability

Process understanding (variation in the process & causes of the variations)

STEP 1 - DEFINE THE PROJECT

What should be in place before starting a project?

Business need definition

- What is the **exact business challenge** / **opportunity** we want to solve?
  - Yield / process optimization
  - Product quality
  - Energy optimization
- What **type of analytics** is needed to solve the business challenge (i.e. **visualization vs analytics**)?
- Define the scope of what is and what is not covered
- **Ensures** the **impact** can be **quantified** so progress is measurable

Data availability & usability

- What **data do we need** & can we **access it**?
- Is the data of **good quality**, i.e. is it accurate, granular & of sussicient quantity?
- **Who is responsible to create an accessible data** environment for the project team?

Team setup

- What **capabilities** do we need in the project team to deliver the use case?
- **Who do we need** to make avaible for the use case?
- What additional capabilities do we need to train to deliver this use case?

STEP 2 – PREPARE

Collect & store data, Structure data in JMP, Clean data

TRANSVERSAL R&D CENTERS BY SAINT-GOBAIN

ACCESS

Access data & assess the data structure and content

INTEGRATE

Select data of interest, manipulate & structure it for analysis

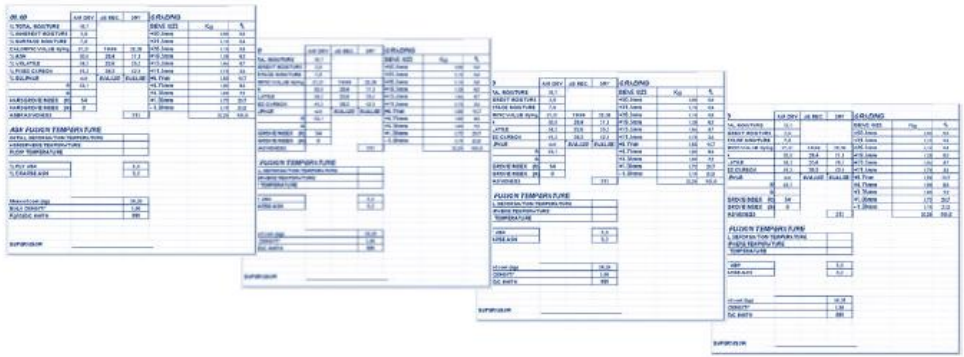
CLEANSE

Put data into a consistent, trusted format

GOVERN

Automate data preparation tasks, monitor jobs & share plans across users

Multiple data files



Structure & Clean

One table including all selected variables, and the target variable.

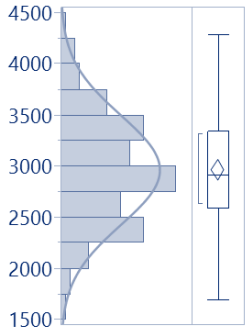
One row per observation.

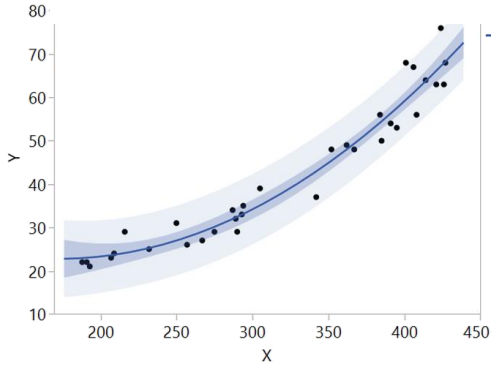
Year	Mont	Da	Date	Unit	TOTAL MOISTURE %	INHERENT MOISTURE %	SURFACE MOISTURE %	CALORIFIC VALUE Air Dry Mj/Kj
2010	3	26	26.03.2010	1	10,2	4	6,5	20,82
2010	3	26	26.03.2010	2	9,6	3,8	5,8	20,88
2010	3	26	26.03.2010	3	7,8	2,8	5,2	21,88
2010	3	26	26.03.2010	4	8,8	3,8	5	20,83
2010	3	26	26.03.2010	5	8,4	3,6	5,2	19,7
2010	3	26	26.03.2010	6	7,4	3,1	4,2	21,06
2010	3	27	27.03.2010	1	9,1	3,3	6	20,62
2010	3	27	27.03.2010	2	10,1	3,6	7	20,54
2010	3	27	27.03.2010	3	8,4	3,6	5,2	20,64
2010	3	27	27.03.2010	4	7,4	3	4,2	21,1
2010	3	27	27.03.2010	5	7,8	2,7	5,2	21,88
2010	3	27	27.03.2010	6	7,6	2,8	5	21,17
2010	3	28	28.03.2010	1	8,2	2,5	5,8	20,46
2010	3	28	28.03.2010	2	8	2,9	5,5	20,25
2010	3	28	28.03.2010	3	8.6	3	5.5	20.96

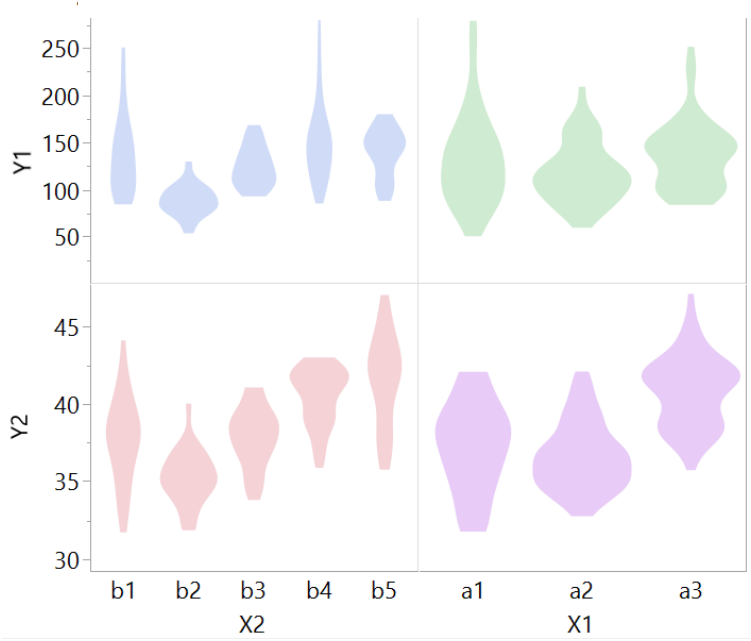
SAINT-GOBAIN

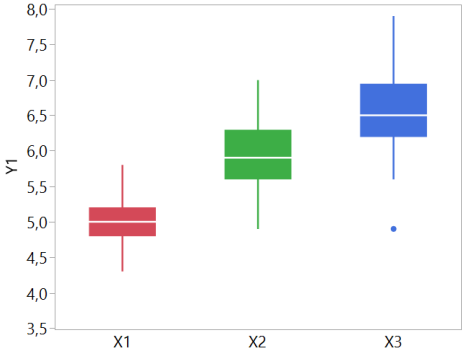
STEP 3 – EXPLORE

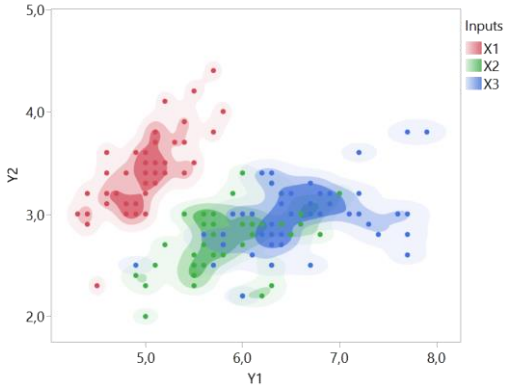
Dynamic visualization, Find the first patterns in the data, Create drag & drop visualizations











Data visualization capabilities & investigational tools

STEP 4 – MODEL

Development of the model: explicative or predictive

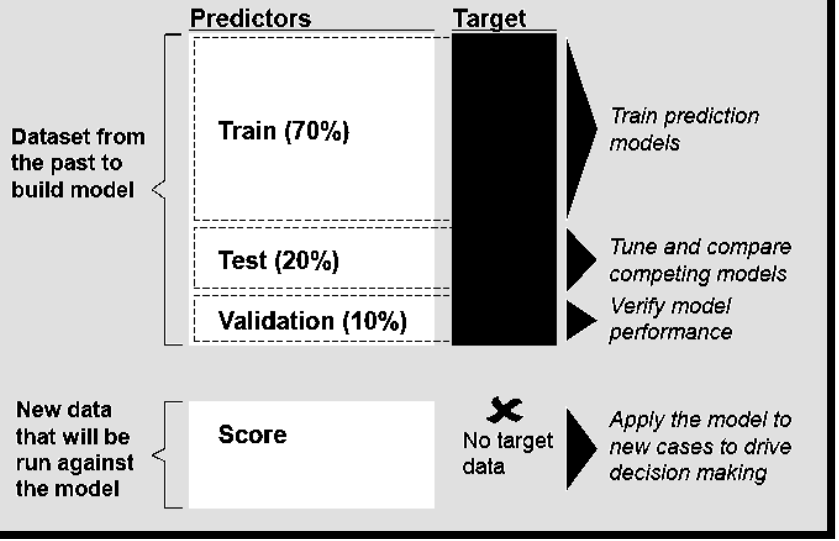
Explicative

- Finding links / dependences
- Understand and explain patterns in data

Predictive

- Finding a formula
- Explore, find and predict patterns in data

Our master table must be split into **3 sets**:
Train, Test & Validation



Supervised learning

Linear regression

Models the target by assuming a linear relatioship with the input

Random Forest

Models the target by building multiple decision trees and combining their input

Decision Tree

Models the target by using an iterative series of decisions on the input variables

Neural Network

Models the target by building a network of simple individual models that are connected among each other by optimized weights

Unsupervised learning

K-means clustering

Group data into clusters based on atributes of the input

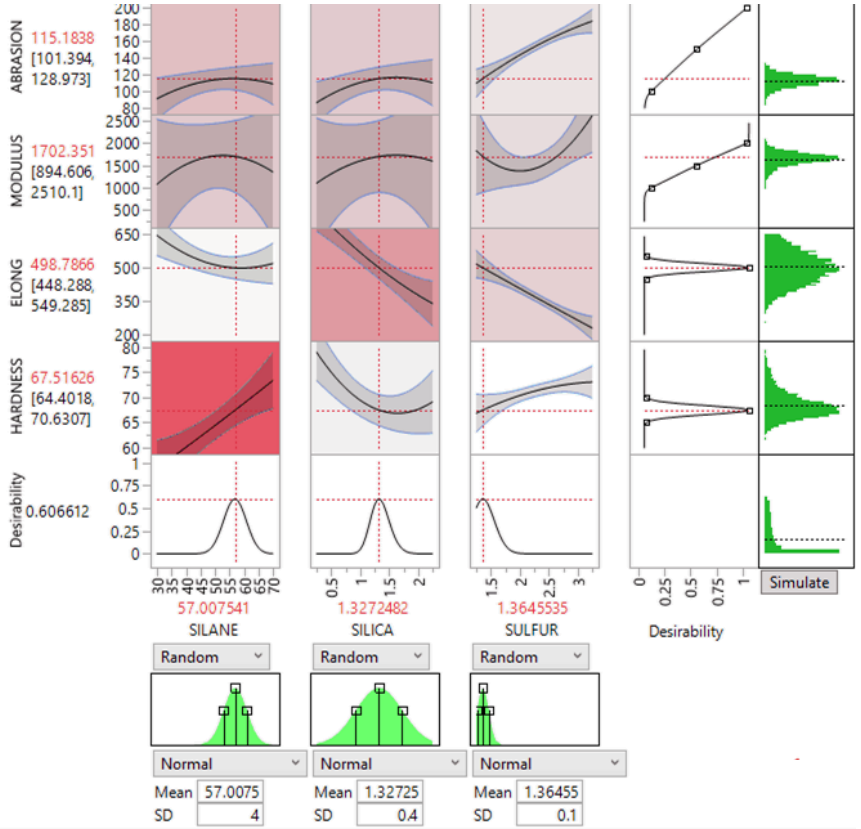
STEP 5 – OPTIMIZE

Embedding single or multiple advanced analytics models to optimize the performance & sustainability

FIND OPTIMAL SOLUTIONS

- Robust, intuitive algebraic optimization modeling language (including linear, mixed integer linear, nonlinear and quadratic and network optimization)
- Constraint programming
- Powerful optimization solvers and presolvers
- Network flow optimization
- Local search optimization (e.g. genetic algorithms) for (generally nonlinear) optimization problems

Prediction profiler



STEP 6 – IMPLEMENT & ACT

Realize the value from analytics assets in action

MANAGE MODEL INVENTORY

- Create a single model inventory for all model types
- Maintain documentation, versioning and model lineage
- Track changes, usage and deployment

GOVERN

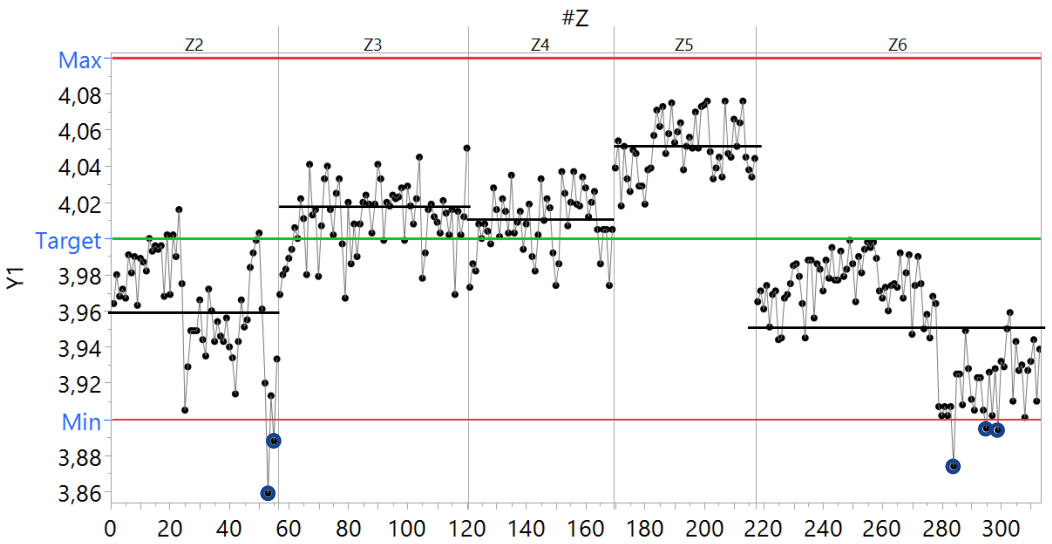
Provide oversight and governed

analytics

EMBED

Embed and execute models, leverage compute platforms

Control chart



STEP 7 – SUSTAIN & CONTROL

Model management and governance to follow model accuracy and continuously improve

MONITOR MODEL EFFECTIVENESS

- Create routine analysis of model performance
- Automate model re-training to maintain optimal performance
- Conduct champion/challenger test to determine

MONITOR

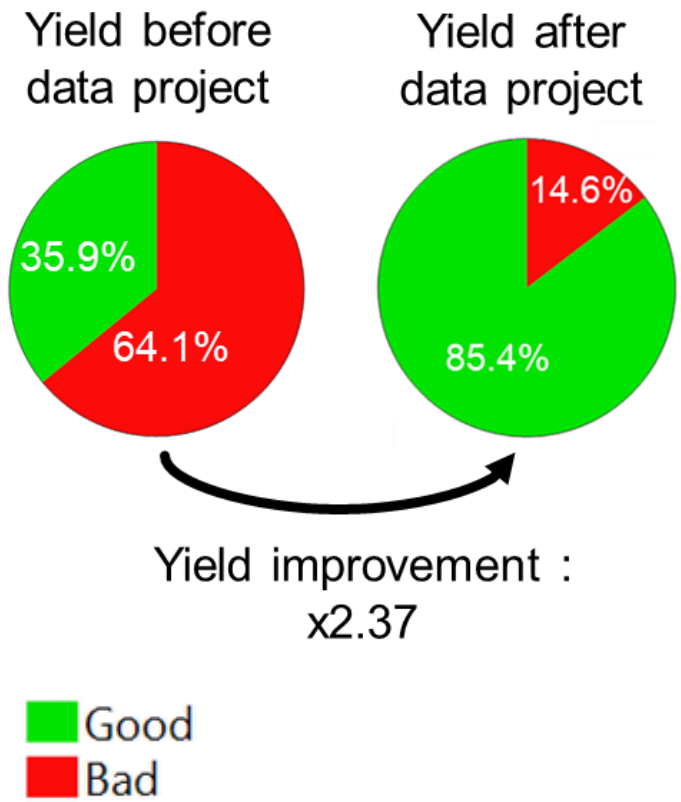
Gain visibility into analytic outcomes and approach

AUTOMATE

Move from insight to action

CONTINUOUS IMPROVEMENT

Prepare the step 1 for new projects



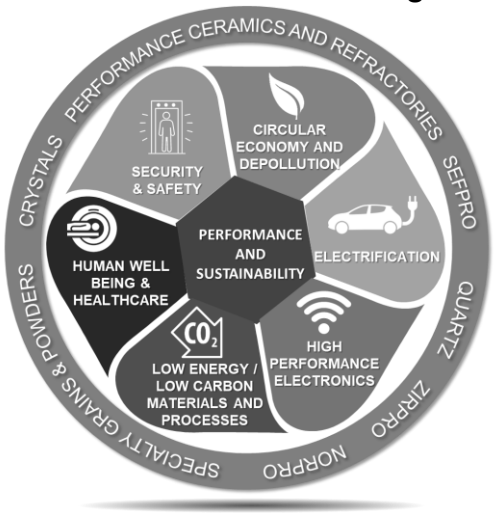
TRANSVERSAL R&D CENTERS BY SAINT-GOBAIN

OUTLINE

- Data project life cycle

Saint-Gobain examples

CERAMICS *GRAND CHALLENGES*
Together we engineer a better, safer & greener world



CERAMIC MATERIALS ENGINEERING

THE ALLIANCE OF PERFORMANCE AND SUSTAINABILITY TO IMPROVE LIFE FOR ALL

SAINT-GOBAIN RESEARCH PROVENCE
Smart Manufacturing team

Y. EL HOUSNI, M. BOINET

SAINT-GOBAIN RESEARCH PROVENCE

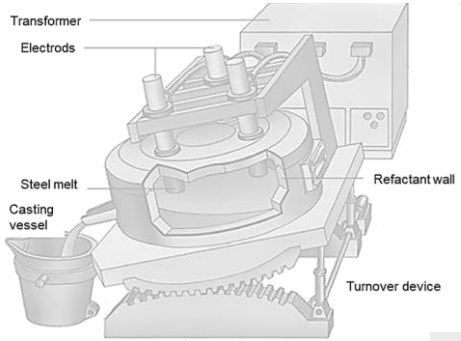
SAINT-GOBAIN

EXAMPLE #1

EXAMPLE #1

#1 DEFINE THE PROJECT

- Data analysis of Industrial electrofusion process
- More than **100** process variables (collected from different kind of sensors)
- Target: **explain the global yield of product JO7** in function of the **best relevant variables**



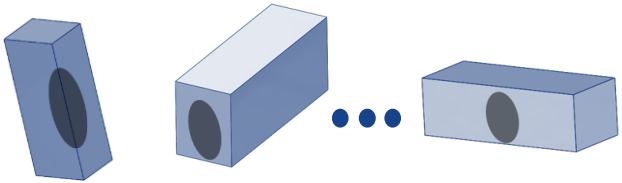
Process technology

Electric Arc Furnace

Sensors types

- What is the **customer's definition of quality ?**

Y% : percent of good pieces with no defects

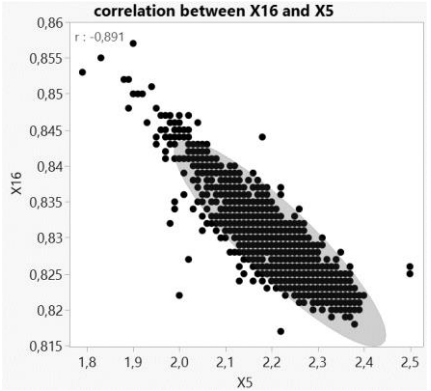


Defect #1 Defect #2 Defect #N

#2 PREPARE

PARETO	Outliers processing
Recoder of attributes	Missing values processing

#3 EXPLORE



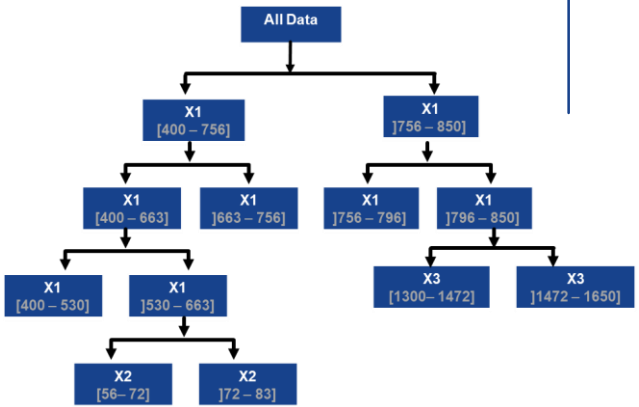
EXAMPLE #1

#4 MODEL

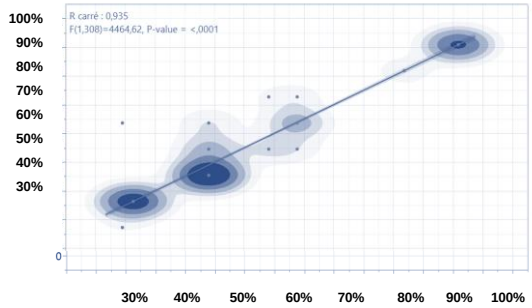
Target

(1) Explicative **(2) Predictive**

Decision trees



Y%



Predict Y%

Desnity mapping
Blue (very hight) – White (very low)

	Root square	Sample size
Train	**93,5%**	**310** products type J7O
Validation	**74%**	**185** products type J7O

#5 OPTIMIZE

Y%

3 process parameters drive the yield

X1 X2 X3

To maximize the Y% a specific setting for **X1** & **X3** has been defined

Yield improvement

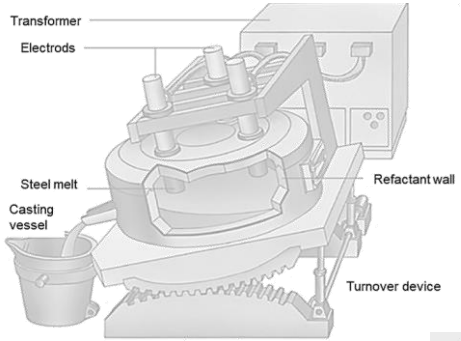
Physical understanding of each parameters

EXAMPLE #2

EXAMPLE #2

#1 DEFINE THE PROJECT

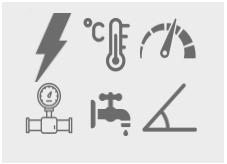
- Data analysis of Industrial electrofusion process
- More than **80** process variables (collected from different kind of sensors)
- Target: **explain the quality (good/bad) of product LO6** in function of the **best relevant variables**



Process technology

Electric Arc Furnace

Sensors types



- What is the **customer's definition of quality ?**

Y : categorial variable

Good piece with no defect D1

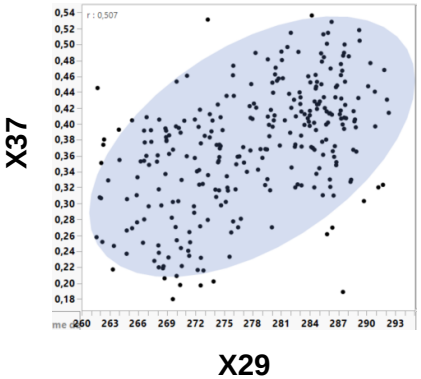
Bad piece with defect D1

#2 PREPARE

- Delete missing values
- Delete constant variables
- Detecting outliers & removing them

#3 EXPLORE

correlation: 51%



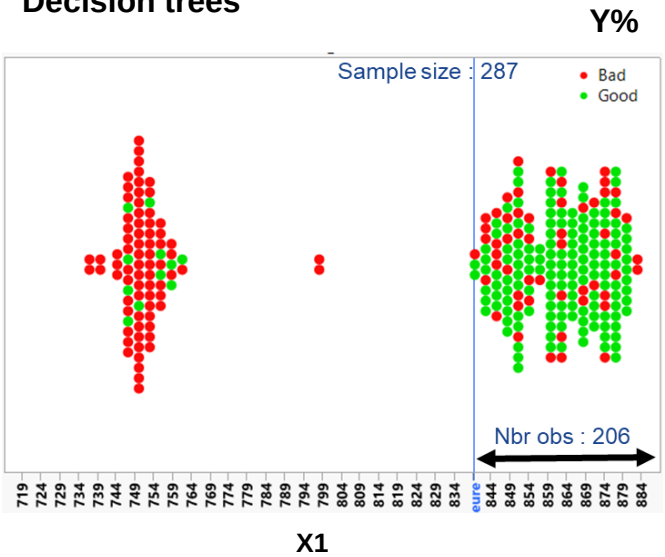
EXAMPLE #2

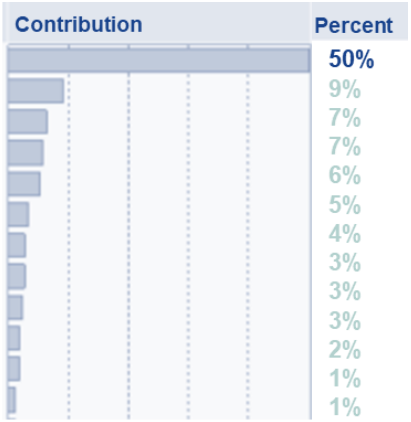
#4 MODEL

Target

(1) Explicative — **(2) Predictive**

Decision trees





R² (Root Square)	RMSE (Root Mean square error)	Total number of observations
84%	3%	287

(Cross validation method)

#5 OPTIMIZE

Y%

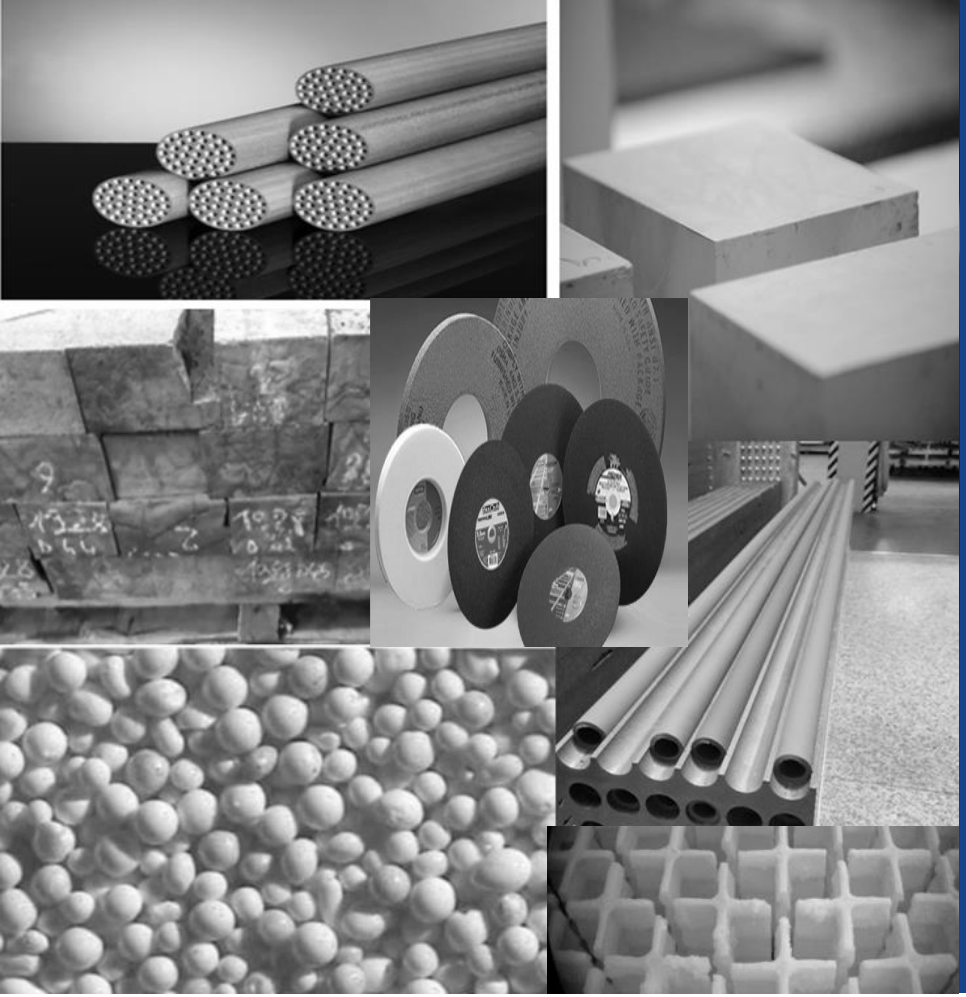
12 process parameters drive the yield

X1 | …. | X12

To maximize the Y% a specific setting for **X1** has been defined

Yield improvement

Physical understanding of each parameters




grazie谢谢ขอบคุณ
merci Σας ευχαριστ tak bedankt
Спасибо धन्यवाद ありがとう
tack thank you terima kasih
gracias obrigado
teşekkür ederim شكرا 고마워요
danke kiitos köszönjük


"Do you have any questions for me?"