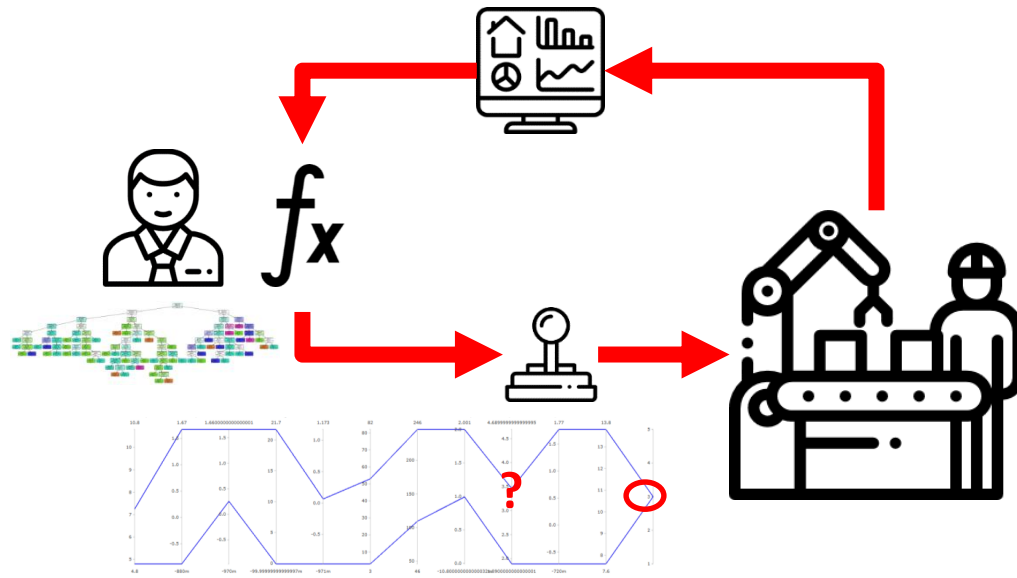


Comment améliorer votre production via les techniques de Machine Learning?

IA pour l'amélioration des performances des systèmes industriels



Jean-Yves DANTAN
Lazhar HOMRI
Wahb ZOUHRI
Alain ETIENNE

Population à former	Technicien supérieur, Ingénieur, Manager, Enseignant
Objectifs stratégiques	Augmenter l'efficacité des stratégies et méthodes de process control et de maîtrise de la qualité via le déploiement de techniques d'IA
Objectifs de la formation	Etre capable de: <ul style="list-style-type: none">- Comprendre et maîtriser les finalités des techniques d'IA- Maîtriser les concepts clés des techniques de Machine Learning- Déployer certaines techniques de Machine Learning
Durée	3 jours (6 séances de 3h.)
Besoins: <ul style="list-style-type: none">- Support informatique pour la formation sur l'utilisation des comptes scientifiques- Modification des codes Python et R pour une utilisation sur les comptes scientifiques- Installation des librairies nécessaires	

A QUOI CA SERT ?

Découvrir des relations entre variables.

Clustering
Regrouper un ensemble d'objets en fonction de leurs caractéristiques.

Classification
Prédire la catégorie (classe) d'une nouvelle observation.

Boîte à outils de Data science

Trouver une fonction qui modélise les données.
Prédire

A quoi ça sert ?

- ✓ Finalités des techniques de machine learning

COMMENT DEPLOYER ?

Exemple:

41 Paramètres Process (PP)
12 Caractéristiques Produit (KC)
809 points
42877 données

COMMENT DEPLOYER ?

Identifier les causes d'une non-conformité Produit

Comparison

Paramètres discriminants de la spécification ?

Paramètres influents KC7

Pourquoi et comment déployer des techniques d'IA ? Comment exploiter les résultats ? avec des applications

- ✓ Techniques d'association: ACP,...
- ✓ Techniques de clustering: K-MEANS, KNN, ...
- ✓ Techniques de classification: Decision tree, Random Forest, SVM, ...
- ✓ Techniques de régression: régression polynomiale, RFR, ANN, ...

Comment ça marche ?

Analyse de données - Analyse en Composantes Principales

Matrice des données (individus selon les lignes et grandeurs selon les colonnes)

Matrice des données réduite

Matrice des corrélations

Valeurs propres (ordonnées dans l'ordre décroissant)

Vecteurs propres correspondants

Matrice de passage des colonnes vers les vecteurs propres

Matrice des données dans la nouvelle base des vecteurs propres (possibilité de normaliser e_i)

le nouvel espace

$$M = \begin{matrix} \downarrow & X_j \\ \downarrow & \downarrow \\ \downarrow & \downarrow \\ \downarrow & \downarrow \end{matrix}$$
$$m = \begin{matrix} \downarrow & X_j - \bar{X}_j \\ \downarrow & \downarrow \\ \downarrow & \downarrow \\ \downarrow & \downarrow \end{matrix}$$
$$r = \begin{matrix} \downarrow & \sum_{i=1}^m X_j X_{i,j} \\ \downarrow & \downarrow \\ \downarrow & \downarrow \\ \downarrow & \downarrow \end{matrix}$$
$$r_{ii} = \frac{\sum_{i=1}^m X_j X_{i,j}}{n-1}$$
$$detr - \lambda I = 0$$
$$w_i = \frac{\lambda_i}{p}$$
$$r e_i^T = \lambda_i e_i^T$$
$$m^T = m P$$

Comment ça marche ?

avec des mises en œuvre

Comment exploiter les résultats

Prédire les non-conformités Produit

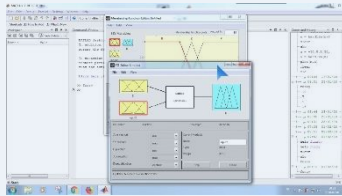
Traitements: Random Forest, Analyse

Résultats

½ Journée JYD	<ul style="list-style-type: none">✓ Estimation, Estimateur, biais, ... (notions clés de statistiques)✓ Scenarii de déploiement des techniques de Machine Learning pour l'amélioration des performances des systèmes de production (illustrés sur une étude de cas)<ul style="list-style-type: none">✓ Identifier les paramètres clés de mon système de production et de mes produits✓ Identifier les stratégies de réglage du système de production✓ Prédire le taux de non conformité Produit✓ Identifier les causes d'une non conformité Produit ou de défaillance Process
½ Journée WZ	<ul style="list-style-type: none">✓ Préparer les données pour mieux les exploiter<ul style="list-style-type: none">✓ Nettoyage des valeurs manquantes✓ Codage des valeurs non numériques✓ Transformation et mise à l'échelle des données✓ Réduction de la dimensionnalité; réduire le nombre de paramètres en fonction de leur pertinence✓ Applications sur une étude de cas mise en oeuvre avec Python
½ Journée LH	<ul style="list-style-type: none">✓ Réduire le volume et/ou les dimensions de données à traiter, Extraire des règles de réglage<ul style="list-style-type: none">✓ Réduction de la dimensionnalité, et étude de corrélation entre les paramètres : Analyse de composantes principales, ...✓ Extraction des règles qui régissent un jeu de données : Arbres de décision, Random Forest✓ Régression logistique✓ Applications sur une étude de cas mise en oeuvre avec Python
½ Journée WZ	<ul style="list-style-type: none">✓ Prédire la production non conforme/défaillante ; Identifier les causes de non conformité ou de défaillance.<ul style="list-style-type: none">✓ Identification des groupes de données similaires (ex: gammes de production) : K MEANS✓ Prédiction de la non conformité : KNN & SVM✓ Applications sur une étude de cas mise en oeuvre avec Python
1 Journée AE	<ul style="list-style-type: none">✓ Prédire les performances des systèmes de production<ul style="list-style-type: none">✓ Base de la conception d'un réseau de neurones en classification et régression✓ Amélioration des prédictions (par configuration des hyperparamètres du réseau)✓ Applications sur plusieurs études de cas mise en oeuvre en python et Tensorflow



Intelligence artificielle



Statistiques



Machine learning

Deep learning



Neurosciences

Introduction

1. Estimation, Estimateur, biais, ... (notions clés de statistiques)
2. Scenarii de déploiement des techniques de Machine Learning pour l'amélioration des performances des systèmes de production (illustrés sur une étude de cas)
 - a. Identifier les paramètres clés de mon système de production et de mes produits
 - b. Identifier les stratégies de réglage du système de production
 - c. Prédire le taux de non-conformité Produit
 - d. Identifier les causes d'une non-conformité Produit ou de défaillance Process

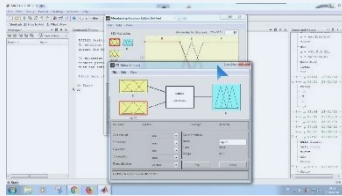


« Le **Data Science** est l'extraction de connaissance d'ensembles de données. »

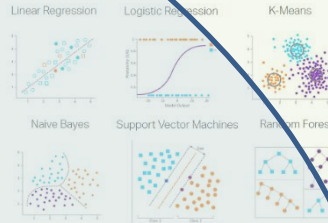
« Le **Data Science** s'appuie sur des outils mathématiques, de statistiques, d'informatique et de visualisation des données. »



Intelligence artificielle



Statistiques



Machine learning

Deep learning

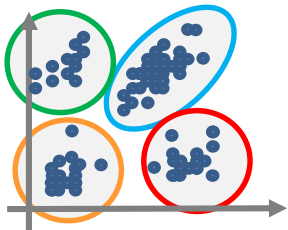


Neurosciences



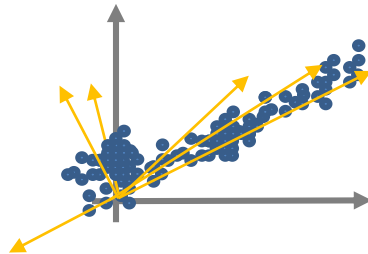
Clustering

Regrouper un ensemble d'objets en fonction de leurs caractéristiques.



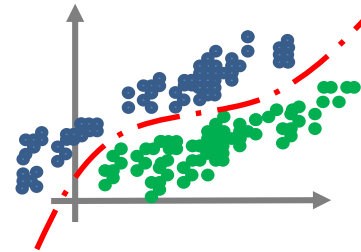
Association

Découvrir des relations entre variables.



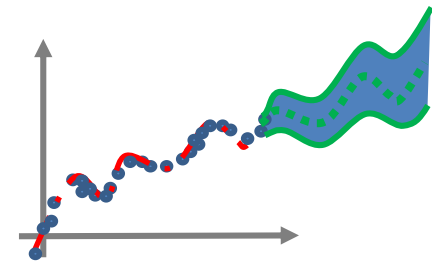
Classification

Prédire la catégorie (classe) d'une nouvelle observation.



Regression

Trouver une fonction qui modélise les données.
Prédire





Clustering

Regrouper un ensemble d'objets en fonction de leurs caractéristiques.

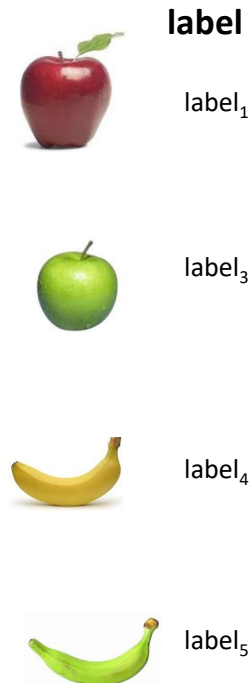


Unsupervised learning
Apprentissage non supervisé

Regrouper les données

Classification

Prédire la catégorie (classe) d'une nouvelle observation.



Modèle de
prédiction
du label

Supervised learning
Apprentissage supervisé



Association

Découvrir des relations entre variables.

Analyser les données en tenant compte de leur caractère multidimensionnel

ACP : Analyse en Composantes Principales, pour les tableaux de variables quantitatives.

AFTD : Analyse Factorielle d'un Tableau de Distances, pour les tableaux de distances.

AFC : Analyse Factorielle des Correspondances, pour les tableaux de contingence.

ACM : Analyse des Correspondances Multiples, pour les tableaux de variables qualitatives.

STATIS : Structuration des Tableaux A Trois Indices de la Statistique

AFM : Analyse Factorielle Multiple

DACP : Double Analyse en Composante Principale

La liste n'est pas exhaustive.

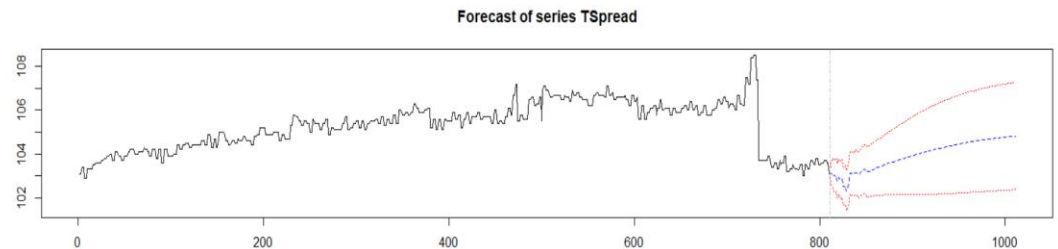
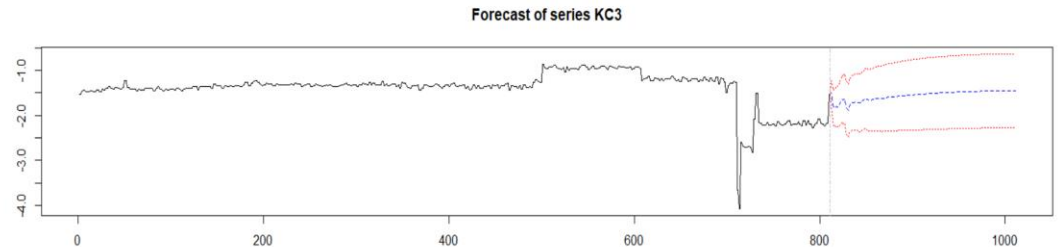
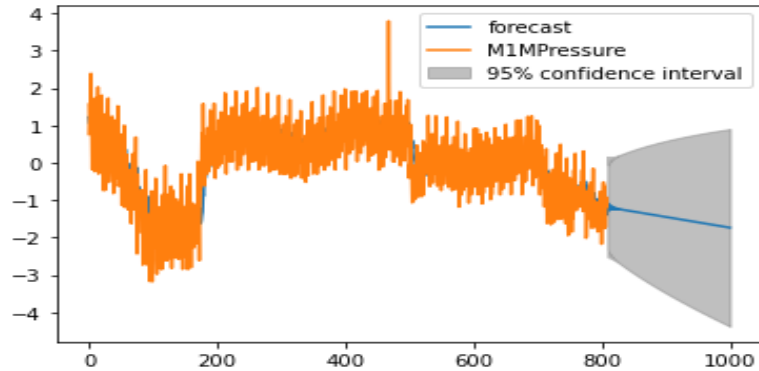


Prédire une variable Y (variable expliquée) à l'aide
d'un ensemble de variables X_1, X_2, \dots, X_p (variables
explicatives)

Regression

Trouver une fonction
qui modélise les
données.

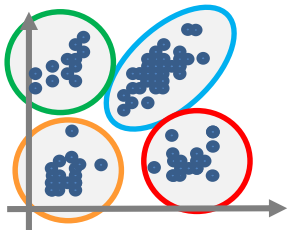
Prédire





Clustering

Regrouper un ensemble d'objets en fonction de leurs caractéristiques.



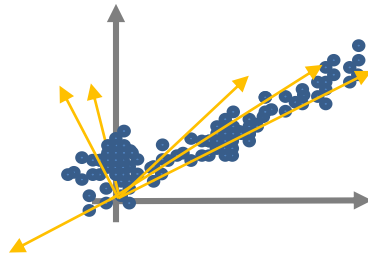
Maintenance

- Prédire les pannes
- Optimiser la disponibilité des équipements



Association

Découvrir des relations entre variables.

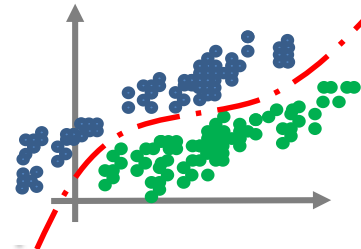


Qualité

- Prédire les défauts
- Trouver les paramétrages optimaux

Classification

Prédire la catégorie (classe) d'une nouvelle observation.

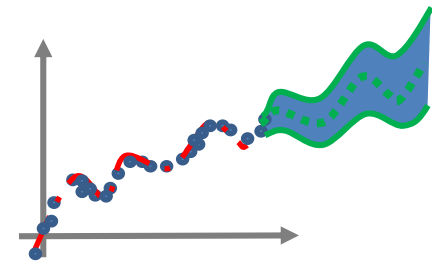


Supply Chain

- Contrôler l'inventaire
- Prédire le comportement du client

Regression

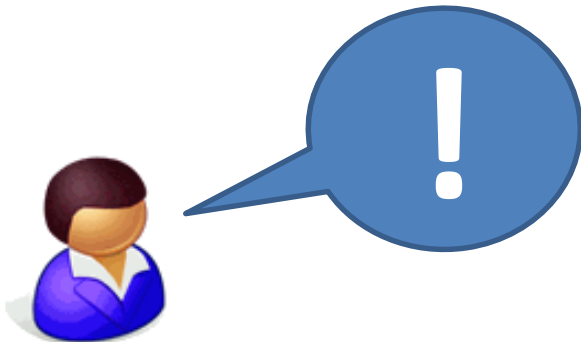
Trouver une fonction qui modélise les données.
Prédire



Ordonnancement

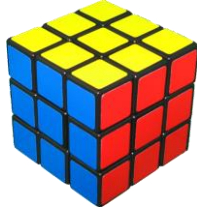
- Minimiser les temps morts entre machines

Data science **Attention !**



Attention à la manipulation des données

- Savez vous calculer la surface d'un carré ?



$$Surface = Coté^2$$

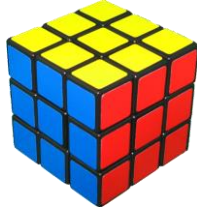
Mais d'un point de vue statistique

$$Surface\ moyenne = \left(\frac{\sum Coté_i}{n} \right)^2$$

$$Moyenne\ des\ Surfaces = \frac{\sum Coté_i^2}{n}$$

Attention à la manipulation des données

- Savez vous calculer la surface d'un carré ?



$$Surface\ moyenne = \left(\frac{\sum Coté_i}{n} \right)^2$$

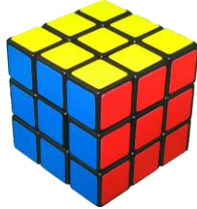
Espérance (Coté) = 25

i	Coté		
1	21,292		
2	24,423		
3	26,176		
4	27,959	Coté moyen	25,431
5	21,924	Surface moyenne	646,740
6	27,057		
7	26,399		
8	27,112		
9	22,845		
10	24,991		
11	22,214		
12	28,562		
13	30,391		
14	26,590		
15	23,533		



Attention à la manipulation des données

- Savez vous calculer la surface d'un carré ?

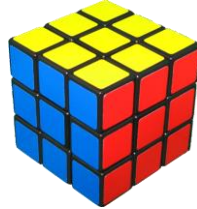


$$\text{Moyenne des Surfaces} = \frac{\sum \text{Coté}_i^2}{n}$$

i	Coté	Surface moyenne		
1	21,292	453,337		
2	24,423	596,476		
3	26,176	685,171		
4	27,959	781,701		
5	21,924	480,657	Moyenne des surfaces	653,432
6	27,057	732,064		
7	26,399	696,911		
8	27,112	735,037		
9	22,845	521,879		
10	24,991	624,574		
11	22,214	493,441		
12	28,562	815,806		
13	30,391	923,612		
14	26,590	707,015		
15	23,533	553,801		

Attention à la manipulation des données

- Savez vous calculer la surface d'un carré ?



$$\text{Surface moyenne} = \left(\frac{\sum \text{Coté}_i}{n} \right)^2$$

$$\text{Moyenne des Surfaces} = \frac{\sum \text{Coté}_i^2}{n}$$

i	Coté			i	Coté	Surface moyenne		
1	25,916			1	25,916	671,660		
2	26,781			2	26,781	717,241		
3	22,502			3	22,502	506,321		
4	25,874	Coté moyen	25,130	4	25,874	669,438		
5	27,959	Surface moyenne	631,540	5	27,959	781,690	Moyenne des surfaces	641,028
6	21,024			6	21,024	442,027		
7	26,186			7	26,186	685,700		
8	19,416			8	19,416	376,992		
9	25,130			9	25,130	631,528		
10	27,231			10	27,231	741,521		
11	27,221			11	27,221	740,986		
12	28,046			12	28,046	786,552		
13	24,558			13	24,558	603,084		
14	29,690			14	29,690	881,500		
15	24,636			15	24,636	606,909		
16	28,192			16	28,192	794,768		
17	24,247			17	24,247	587,895		
18	22,510			18	22,510	506,691		
19	24,681			19	24,681	609,146		
20	23,427			20	23,427	548,840		
21	24,918			21	24,918	620,908		
22	19,448			22	19,448	378,219		
23	28,238			23	28,238	797,384		
24	22,161			24	22,161	491,090		
25	19,812			25	19,812	392,524		
26	25,304			26	25,304	640,299		
27	27,006			27	27,006	729,319		
28	24,200			28	24,200	585,639		
29	25,858			29	25,858	668,611		

Echantillon de 100 observations

Soit X une variable aléatoire dont la densité de probabilité $f(x, \theta)$ dépend d'un paramètre θ appartenant à \mathbb{R}^k .

Au vu d'un échantillon de X il s'agit de déterminer au mieux la « vraie » valeur θ_0 de θ . On peut utiliser deux méthodes :

- Estimation ponctuelle : on calcule une valeur vraisemblable $\hat{\theta}$ de θ_0 .
- Estimation par intervalle : on cherche un intervalle dans lequel θ_0 se trouve avec une probabilité élevée.

Probabilité

X variable aléatoire de paramètre θ

N variables aléatoires: (X_1, X_2, \dots, X_n)

Estimateur: T

$$T = \varphi(X_1, X_2, \dots, X_n)$$

$$E(T) = E(\varphi(X_1, X_2, \dots, X_n))$$

Statistiques

N mesures: (x_1, x_2, \dots, x_n)

Estimation: $\hat{\theta}$ $[a, b]$

$$\hat{\theta} = \varphi(x_1, x_2, \dots, x_n)$$

Soit X une variable aléatoire de densité $f(x, \theta)$, θ appartenant à un intervalle I de \mathbb{R} .

Définition . Un n -échantillon de X est un n -uplet (X_1, X_2, \dots, X_n) tel que les X_k ont la même loi que X et sont indépendantes.

Définition . Une réalisation de l'échantillon est un n -uplet (x_1, x_2, \dots, x_n) où les x_k sont des nombres réels, résultats d'expériences.

Définition . Une statistique de l'échantillon est une variable aléatoire $\varphi(X_1, X_2, \dots, X_n)$ où φ est une application de \mathbb{R}^n dans \mathbb{R} .

Définition . Un estimateur T de θ est une statistique à valeurs dans I .

Exemple : $T_n = \frac{1}{n} \sum_{k=1}^n X_k$ est un estimateur de l'espérance mathématique de X .

Définition . Le biais de l'estimateur est $(\mathbf{E}(T) - \theta_0)$. S'il est nul T est un estimateur sans biais.

Définition . L'estimateur est asymptotiquement sans biais si

$$\lim_{n \rightarrow \infty} \mathbf{E}(T_n) = \theta_0.$$

Définition . *L'estimateur est convergent si la suite (T_n) converge en probabilité vers θ_0 :*

$$\forall \varepsilon > 0, \mathbb{P}(|T_n - \theta_0| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$$

Théorème . *Un estimateur convergent est asymptotiquement sans biais. Pour qu'un estimateur asymptotiquement sans biais soit convergent il suffit que $\mathbf{Var}(T_n) \xrightarrow{n \rightarrow \infty} 0$.*

Définition . *Soient T_1 et T_2 deux estimateurs de θ . On dira que T_1 est plus efficace que T_2 si $\mathbf{E}((T_1 - \theta_0)^2) \leq \mathbf{E}((T_2 - \theta_0)^2)$.*

Si T_1 et T_2 sont des estimateurs sans biais cette condition s'écrit : $\mathbf{Var}(T_1) \leq \mathbf{Var}(T_2)$.

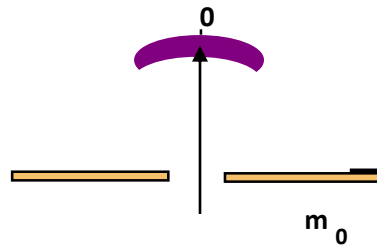
Définition . *Une estimation est la valeur de l'estimateur correspondant à une réalisation de l'échantillon.*

- Savez vous peser des masses ?

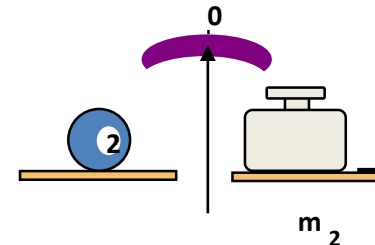


Mais d'un point de vue statistique

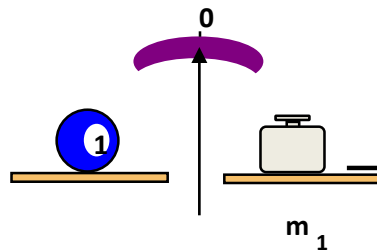
Tarage de
la balance



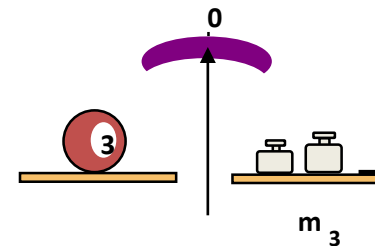
Pesée de
l'objet n° 2



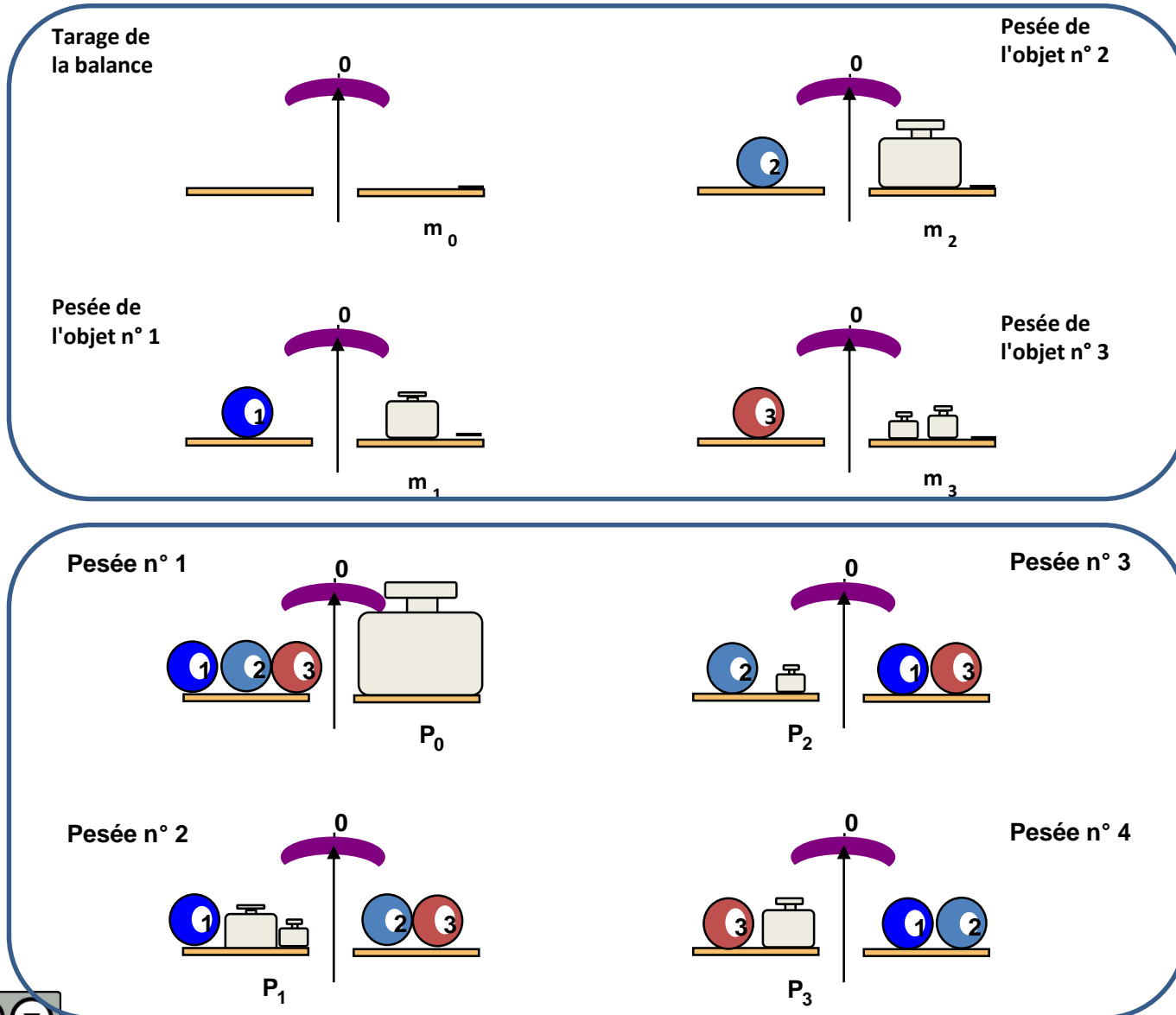
Pesée de
l'objet n° 1



Pesée de
l'objet n° 3



- Savez vous peser des masses ?





Attention à la manipulation des données

- Savez vous peser des masses ?

1

$$m_1 \quad 2\sigma^2$$

2

$$m_2 \quad 2\sigma^2$$

3

$$m_3 \quad 2\sigma^2$$

$$m_1 = \frac{P_0 + P_1 - P_2 - P_3}{4} \quad \frac{\sigma^2}{4}$$

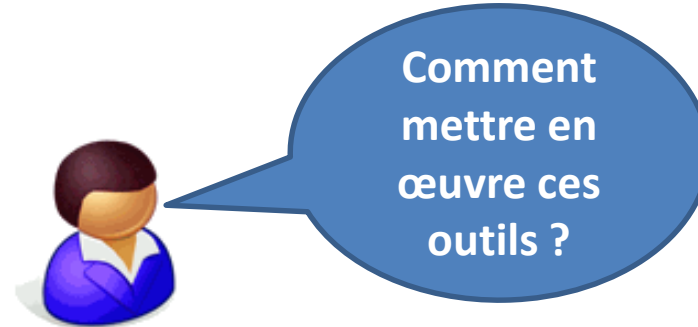
$$m_2 = \frac{P_0 + P_2 - P_1 - P_3}{4} \quad \frac{\sigma^2}{4}$$

$$m_3 = \frac{P_0 + P_3 - P_1 - P_2}{4} \quad \frac{\sigma^2}{4}$$

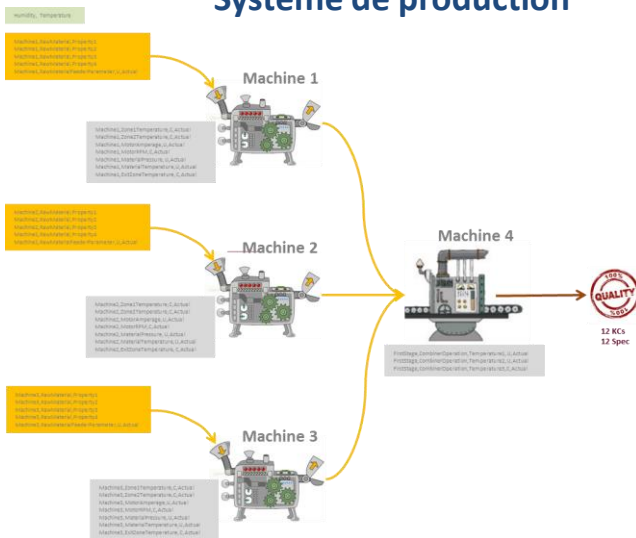
Data science

COMMENT DEPLOYER ? Les finalités

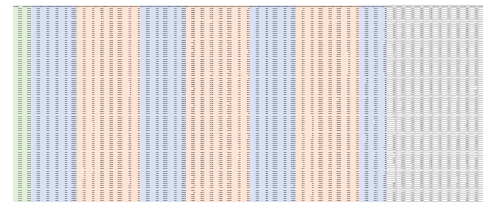




Système de production



Données

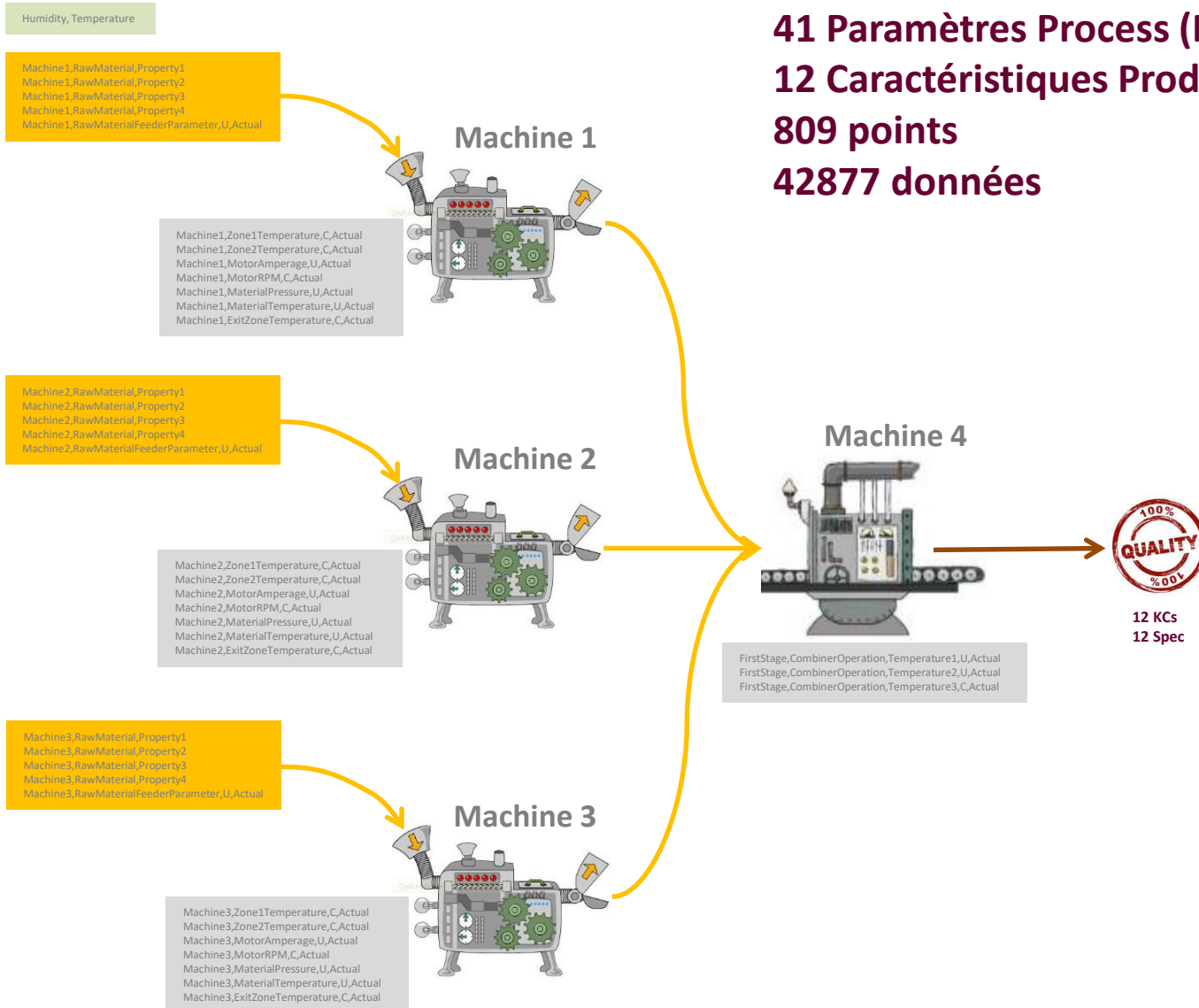


A large table representing data, with many columns and rows of text. The text is too small to read but appears to be a list of data points or a log.

Boite à outils de fouille de données



Exemple:



41 Paramètres Process (PP)
12 Caractéristiques Produit (KC)
809 points
42877 données



Identifier les paramètres clés de mon système de production

Données

Tous les paramètres
process et produits

Traitement(s)

Analyse



ACP, kACP

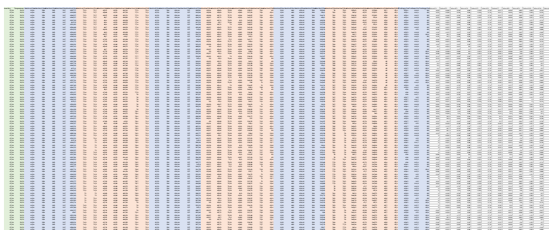


Analyser les causalités
entre les variabilités des
paramètres

Identifier les paramètres
process et produits
« libres »



Identifier les paramètres clés de mon système de production



Traitement
ACP

Analyse

ERREUR

Machine1,RawMaterial,Property1
Machine1,RawMaterial,Property2
Machine1,RawMaterial,Property3
Machine1,RawMaterial,Property4

Constants

Machine1,RawMaterialFeederParameter,U,Actual

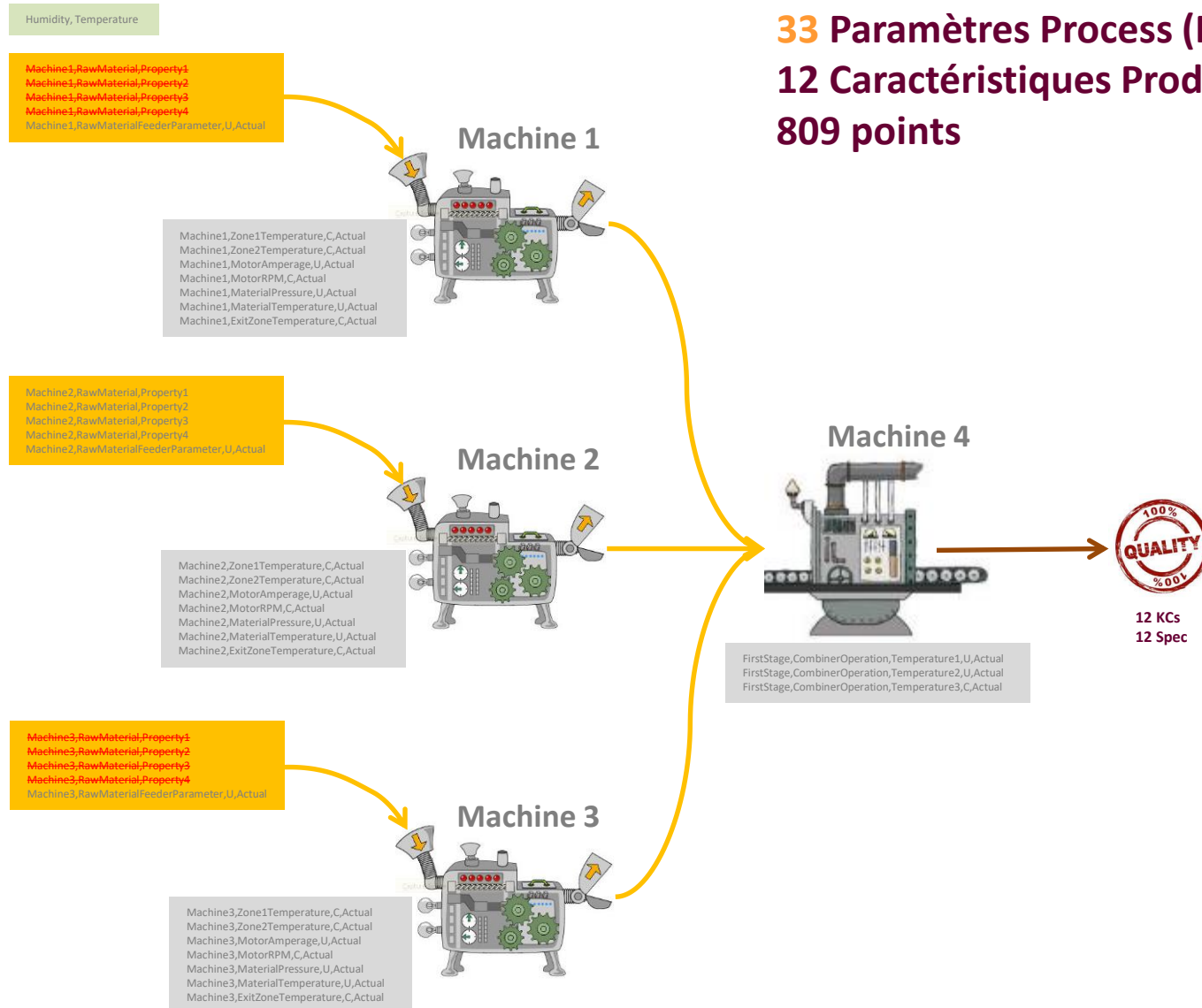
Machine3,RawMaterial,Property1
Machine3,RawMaterial,Property2
Machine3,RawMaterial,Property3
Machine3,RawMaterial,Property4

Constants

Machine3,RawMaterialFeederParameter,U,Actual

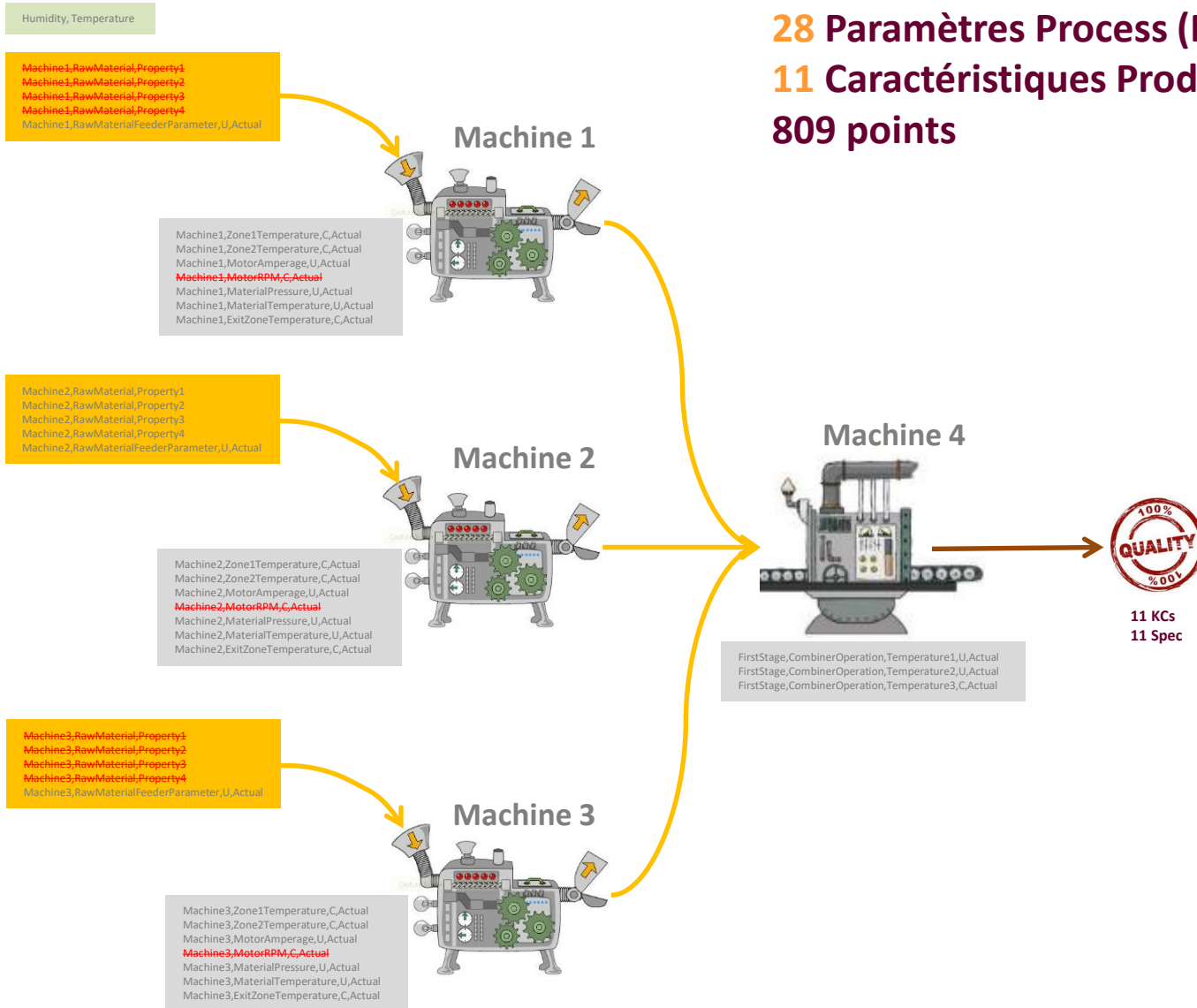
Aucune analyse ne peut être
faite sur ces paramètres

Exemple:



33 Paramètres Process (PP)
12 Caractéristiques Produit (KC)
809 points

Exemple:



28 Paramètres Process (PP)
11 Caractéristiques Produit (KC)
809 points



Identifier les paramètres clés de mes produits

Données
Tous les paramètres
produits

Traitement

Analyse



ACP, kACP



Analyser les causalités
entre les variabilités des
paramètres

Identifier les KCs des
produits à surveiller

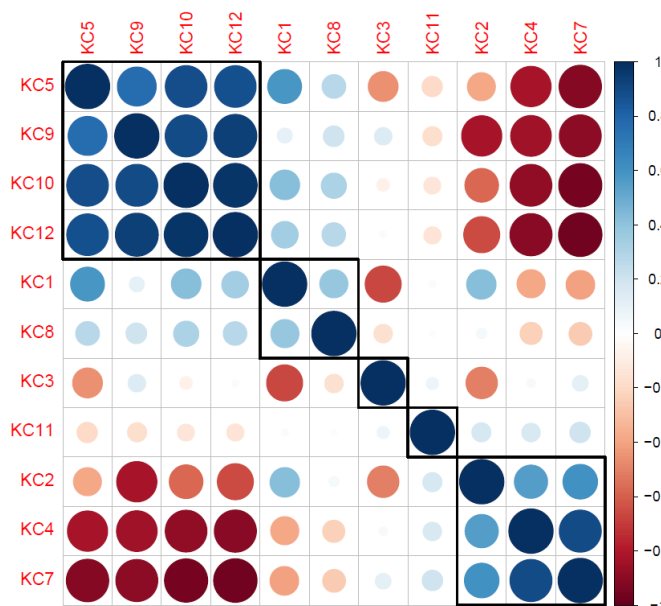
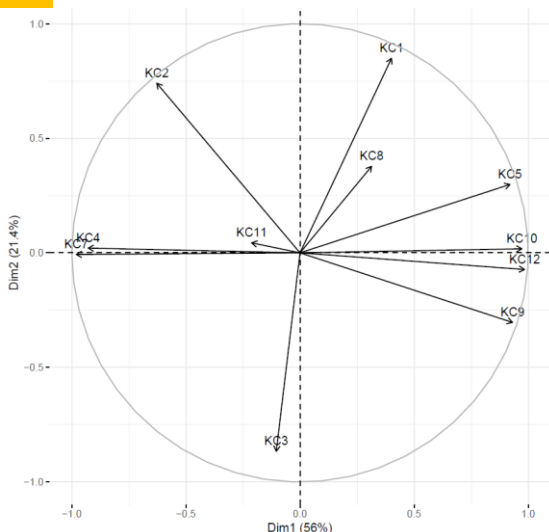


Identifier les paramètres clés de mes produits

Données
Tous les paramètres
produits

Traitement
ACP

Analyse



KC4, KC7, KC10, KC12 très fortement liées, avec aussi KC5 et KC9.

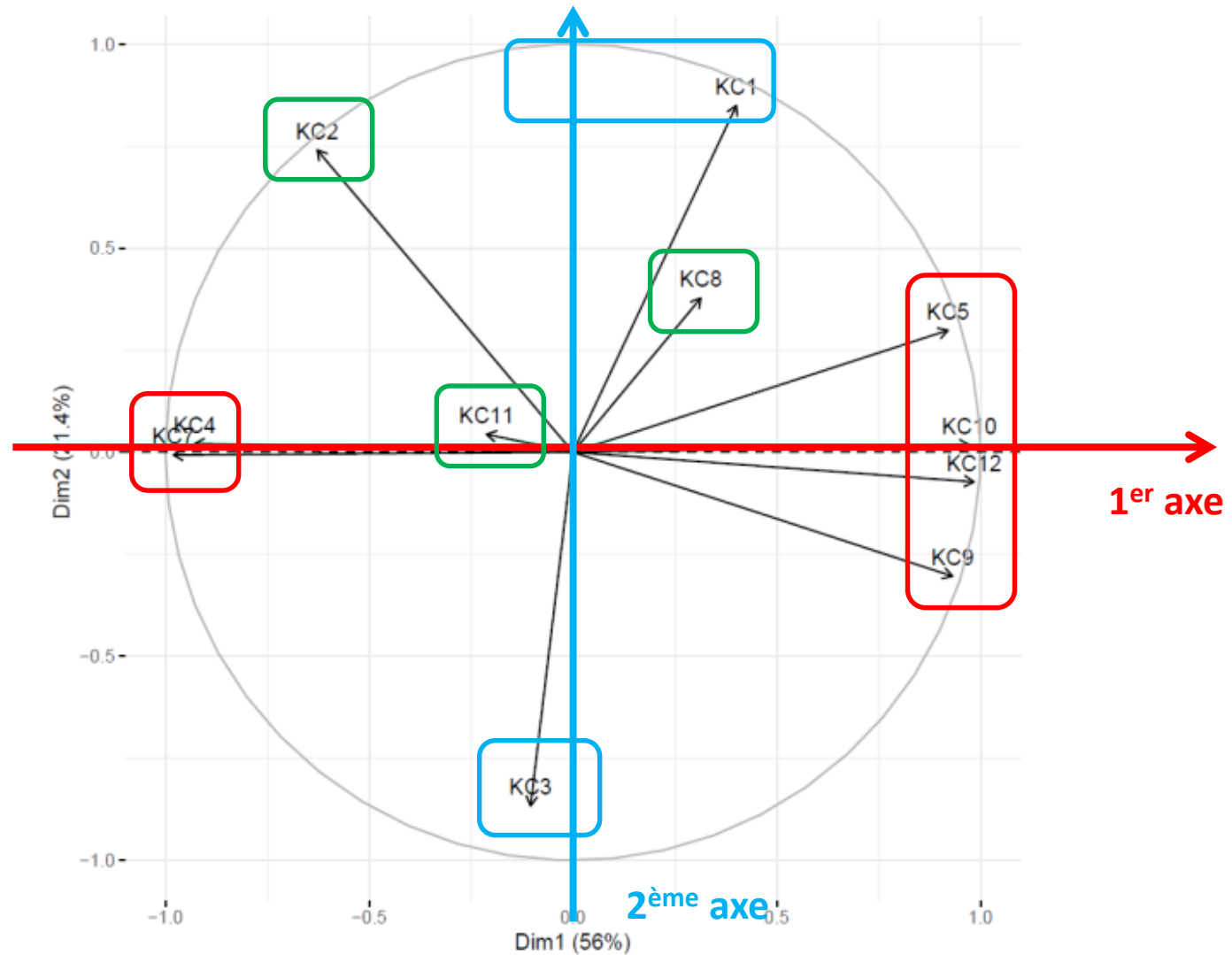
> Possibilité de surveiller qu'une de ces KCs

KC1, KC2, KC3, KC8 et KC11 indépendantes

> Surveiller chacune d'elles



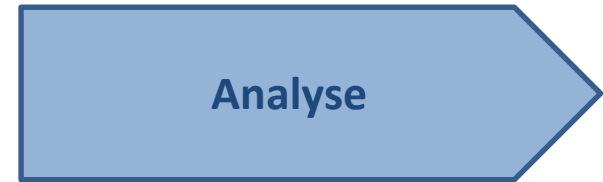
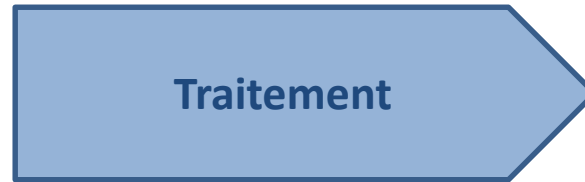
Identifier les paramètres clés de mes produits





Identifier les stratégies de réglage du système de production

Données
Tous les paramètres
process



ACP, kACP



Analyser les causalités
entre les variabilités des
paramètres

Identifier les PPs pilotant

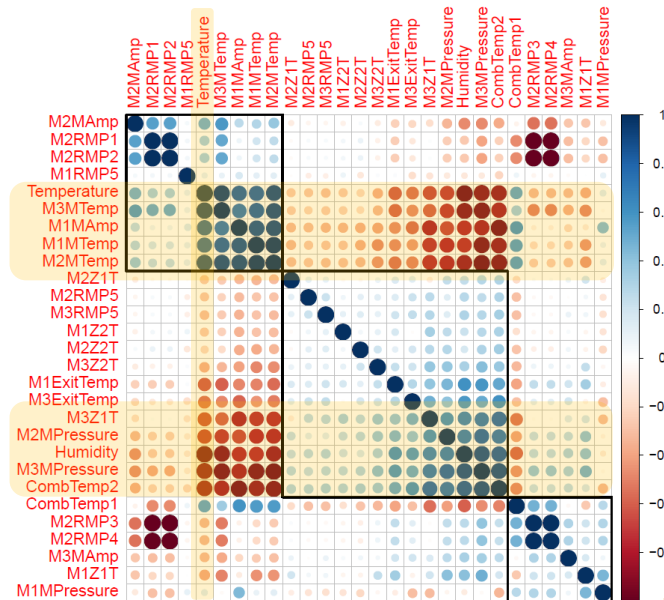
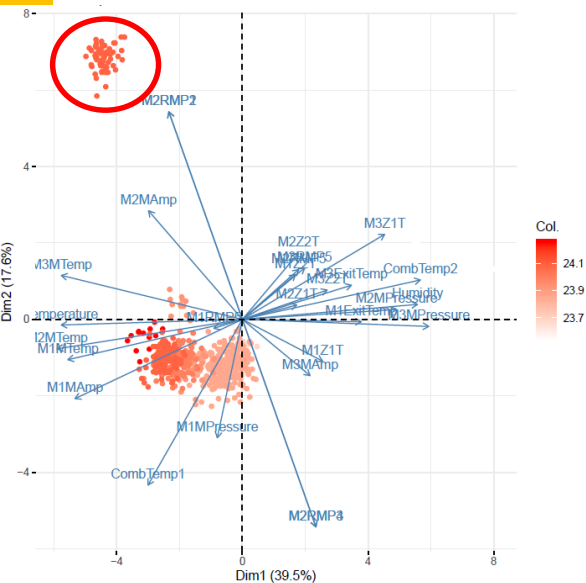


Identifier les stratégies de réglage du système de production

Données
Tous les paramètres
process

Traitement
ACP

Analyse



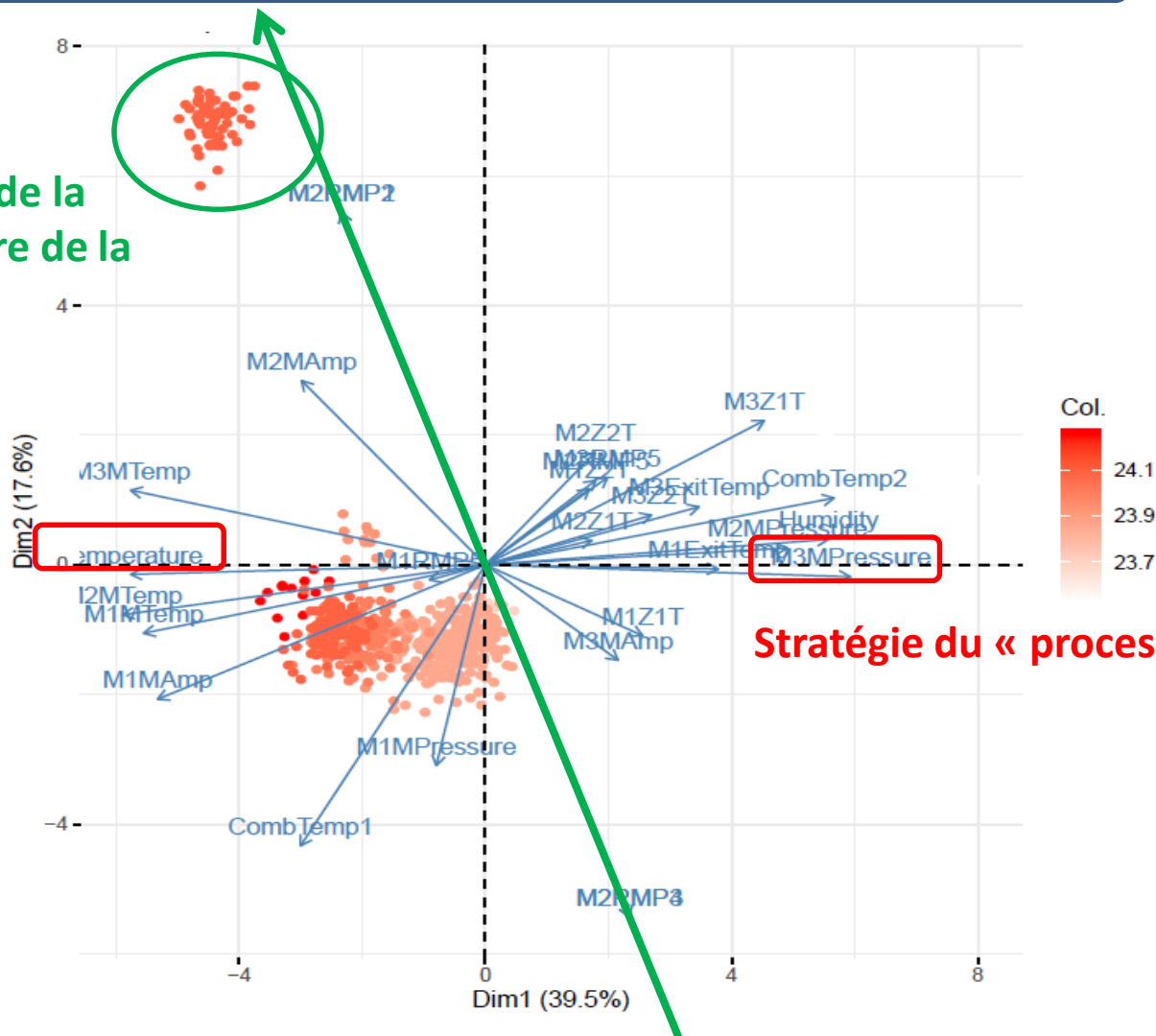
La température externe influence la stratégie de réglage
> M3MTemp, M1MAmp, ...
sont influencées par la température externe

Deux groupes de stratégies d'appro de la machine 2 apparaissent
> Analyse des réglages spécifiques



Identifier les stratégies de réglage du système de production

Caractéristique de la matière première de la machine 2



Stratégie du « process control »



Identifier les causes d'une non-conformité Produit

Données
Tous les paramètres
process et la KC

Traitement

Analyse

ACP, kACP

C4.5, SVM,
Random Forest,
Logistic Regression
SHAP



Analyser les causalités entre
les variabilités des paramètres

Identifier les PPs « pilotant » la
KC

Extraire des règles

Extraire des plages de réglages

Extraire les paramètres
discriminants des règles

Identifier les PPs « pilotant » la
non conformité

COMMENT DEPLOYER ?

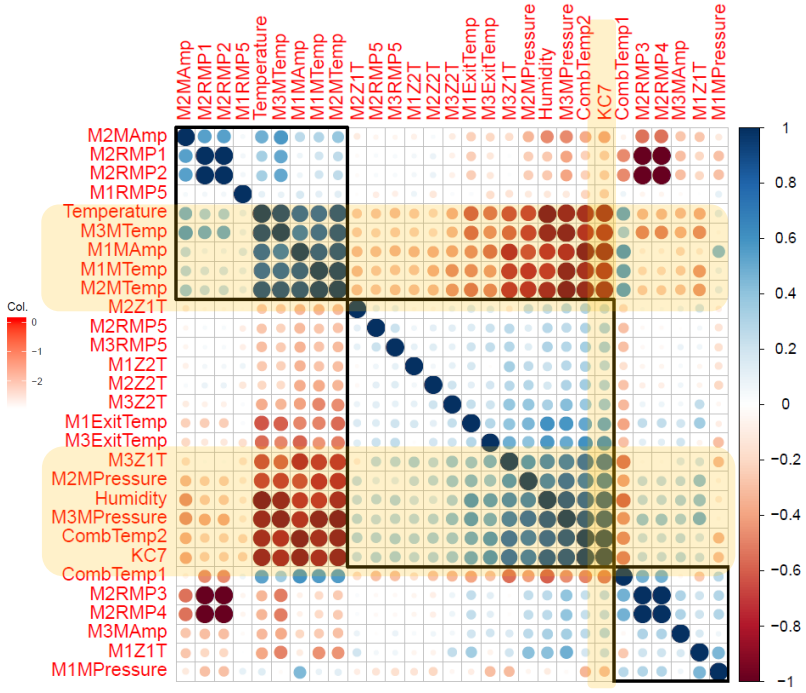
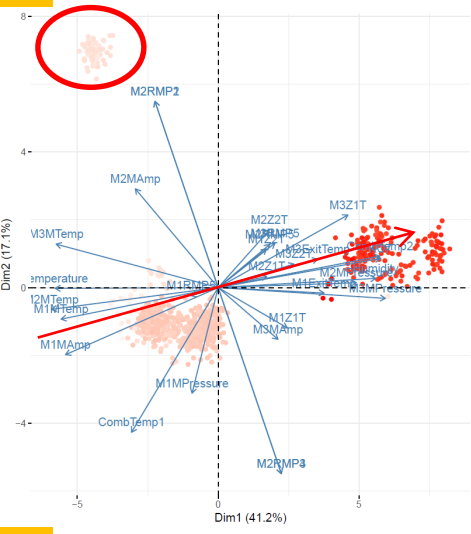


Identifier les causes d'une non-conformité Produit

Données
 Tous les paramètres process et la KC
 KC7

Traitement
 ACP

Analyse

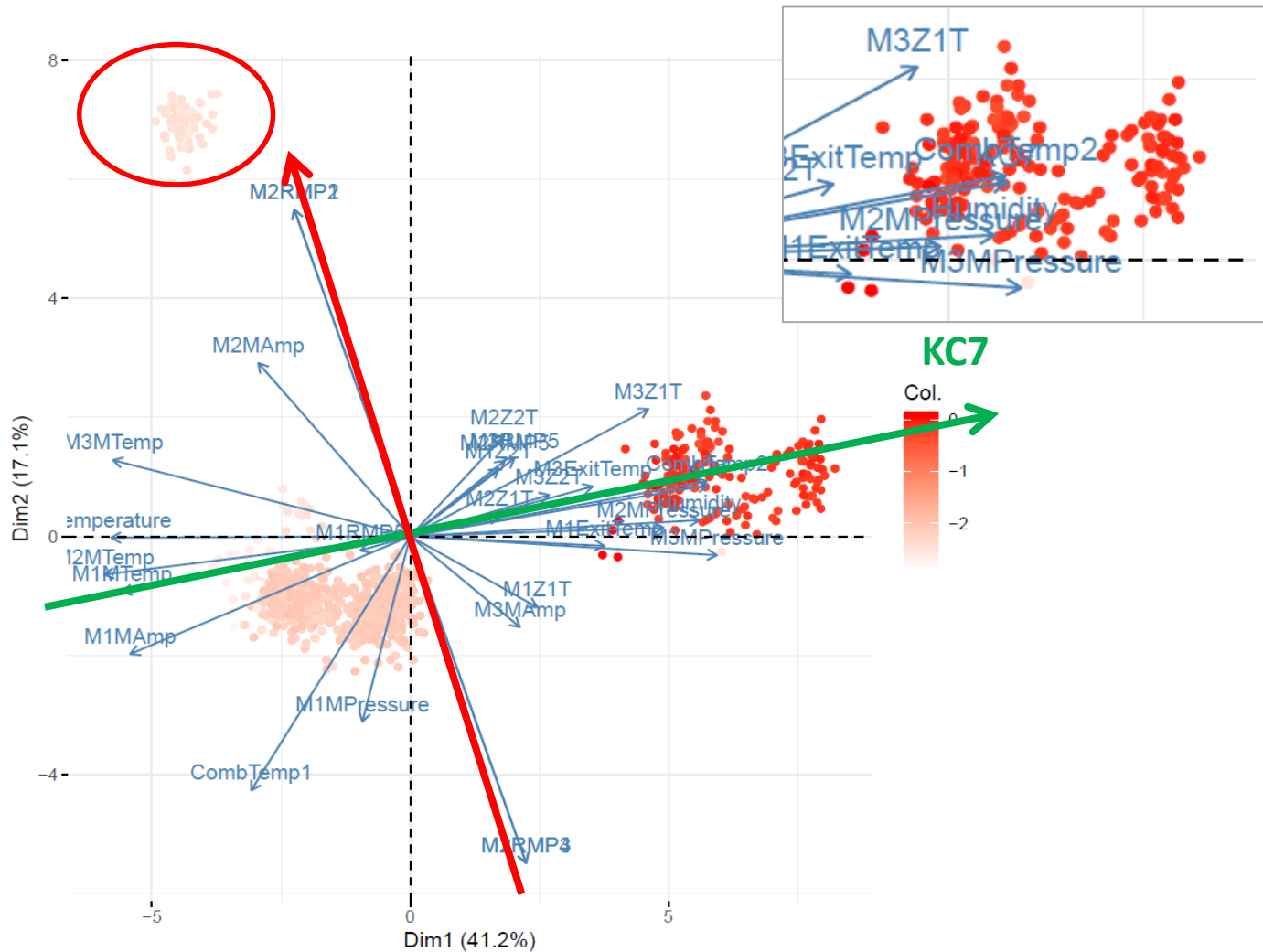


L'analyse des corrélations et l'ACP montrent que la KC7 est fortement corrélée donc influencée par certains paramètres process:
 M3MTemp, M1MAmp, ...
 > Je dois définir une stratégie de réglage de ces paramètres afin de garantir la conformité





Identifier les causes d'une non-conformité Produit



COMMENT DEPLOYER ?

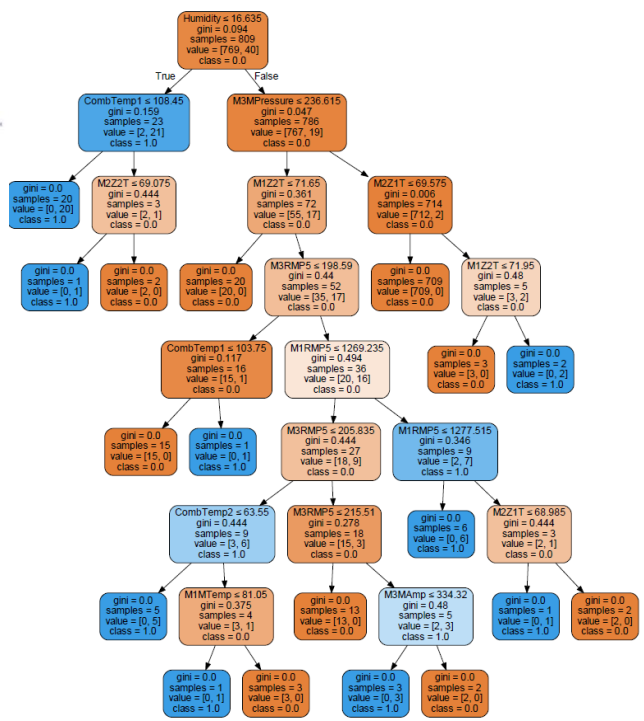
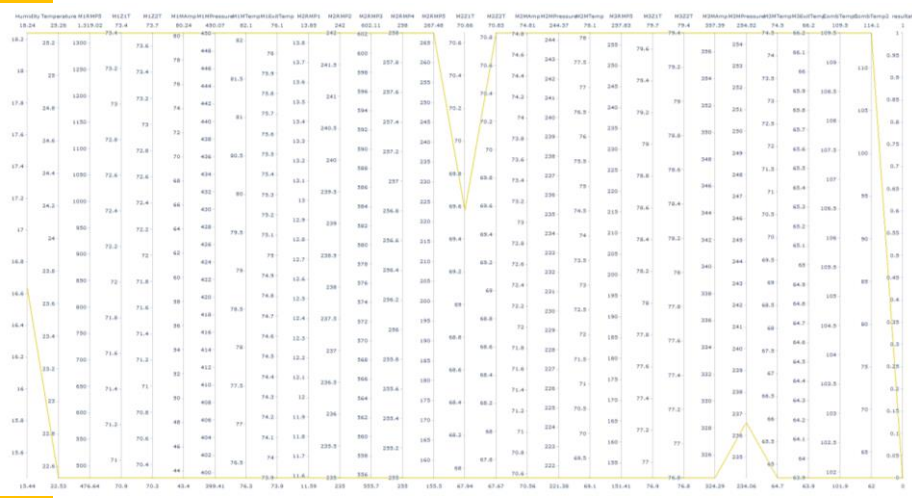


Identifier les causes d'une non-conformité Produit

Données
 Tous les paramètres
 process et la conformité
 KC7

Traitement
 C4.5

Analyse



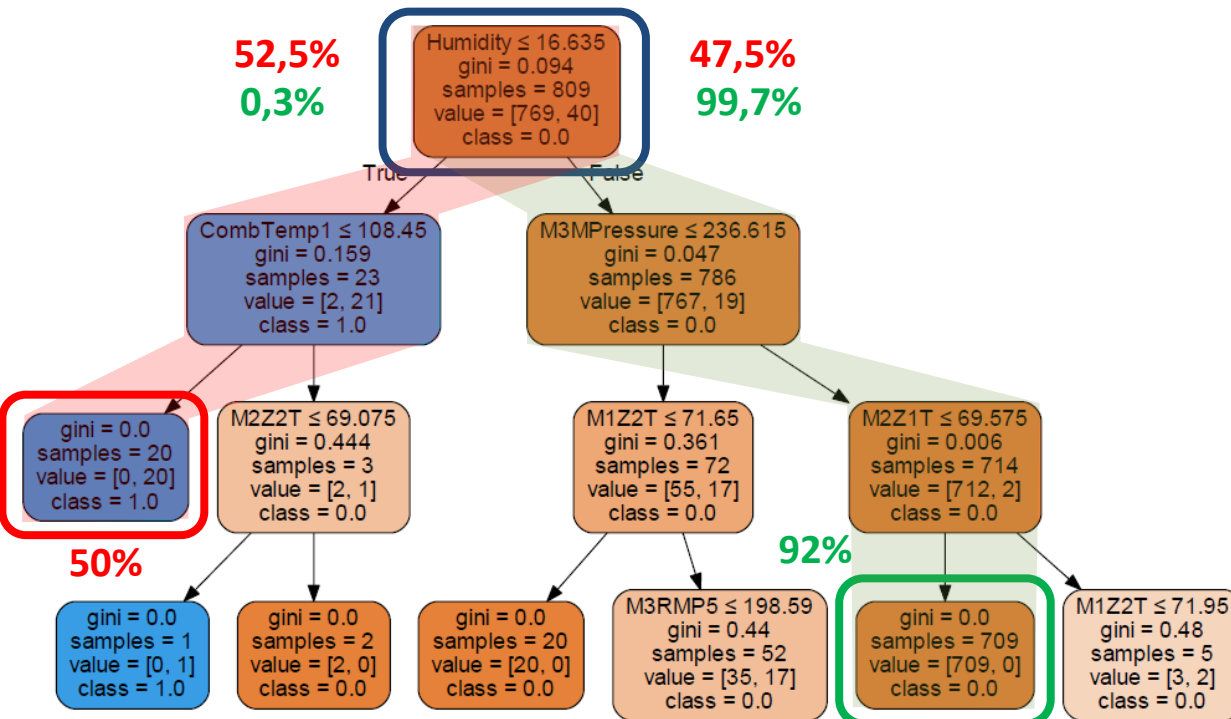


Identifier les causes d'une non-conformité Produit

Données
Tous les paramètres
process et la conformité
KC7

Traitement
C4.5

Analyse



La classification réalisée par C4.5 fait apparaître un certain nombre de règles et une règle majeure (92,2% des conformités):

Conformité si

**Humidity > 16,635 &
M3MPressure > 236,615 &
M2Z1T < 69,575**

Attention, l'analyse de la stratégie de réglage existante montre que ces paramètres sont liés à d'autres.



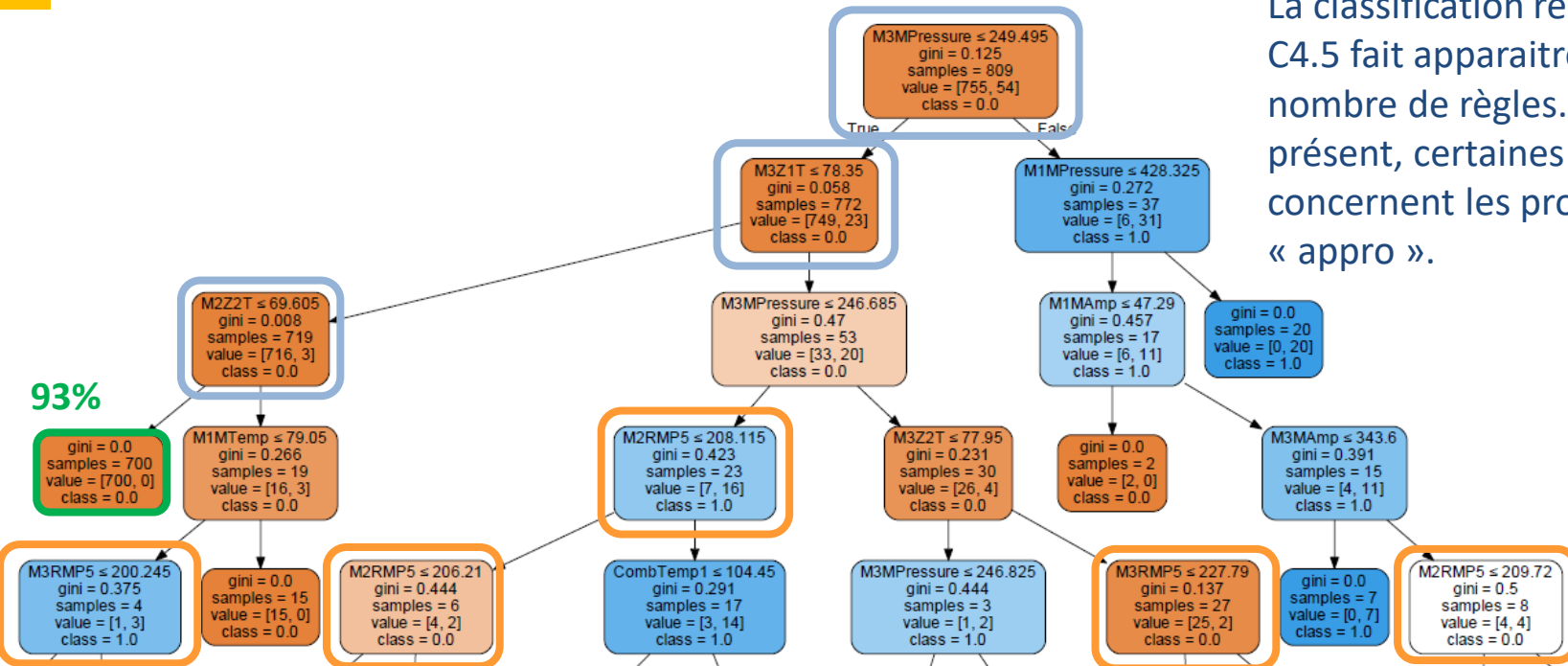
Identifier les causes d'une non-conformité Produit

Données
Tous les paramètres
process et la conformité
KC10

Traitement
C4.5

Analyse

La classification réalisée par C4.5 fait apparaître un certain nombre de règles. Dans le cas présent, certaines règles concernent les propriétés des « appro ».



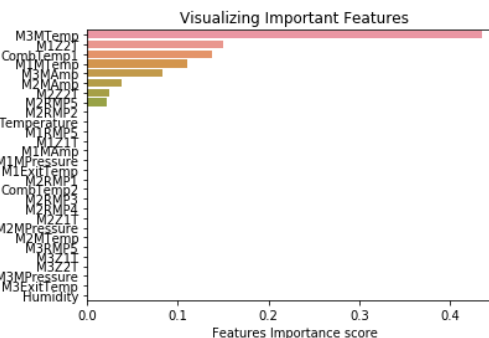
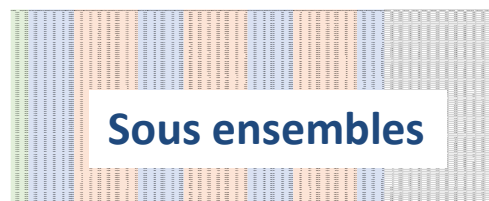


Identifier les causes d'une non-conformité Produit

Données
Tous les paramètres
process et la conformité
KC7

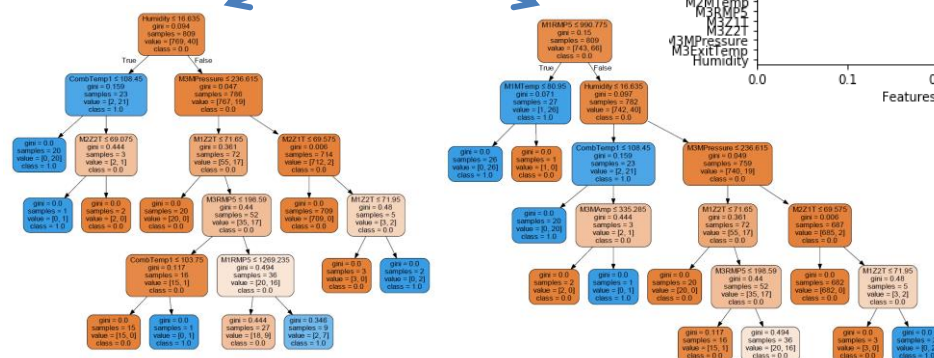
Traitement
Random Forest

Analyse



Les classifications réalisées par Random Forest permettent une prédiction plus robuste que C4.5.

L'extraction des règles n'est plus possible facilement, mais l'identification des paramètres discriminants est possible afin d'identifier les paramètres process à surveiller.



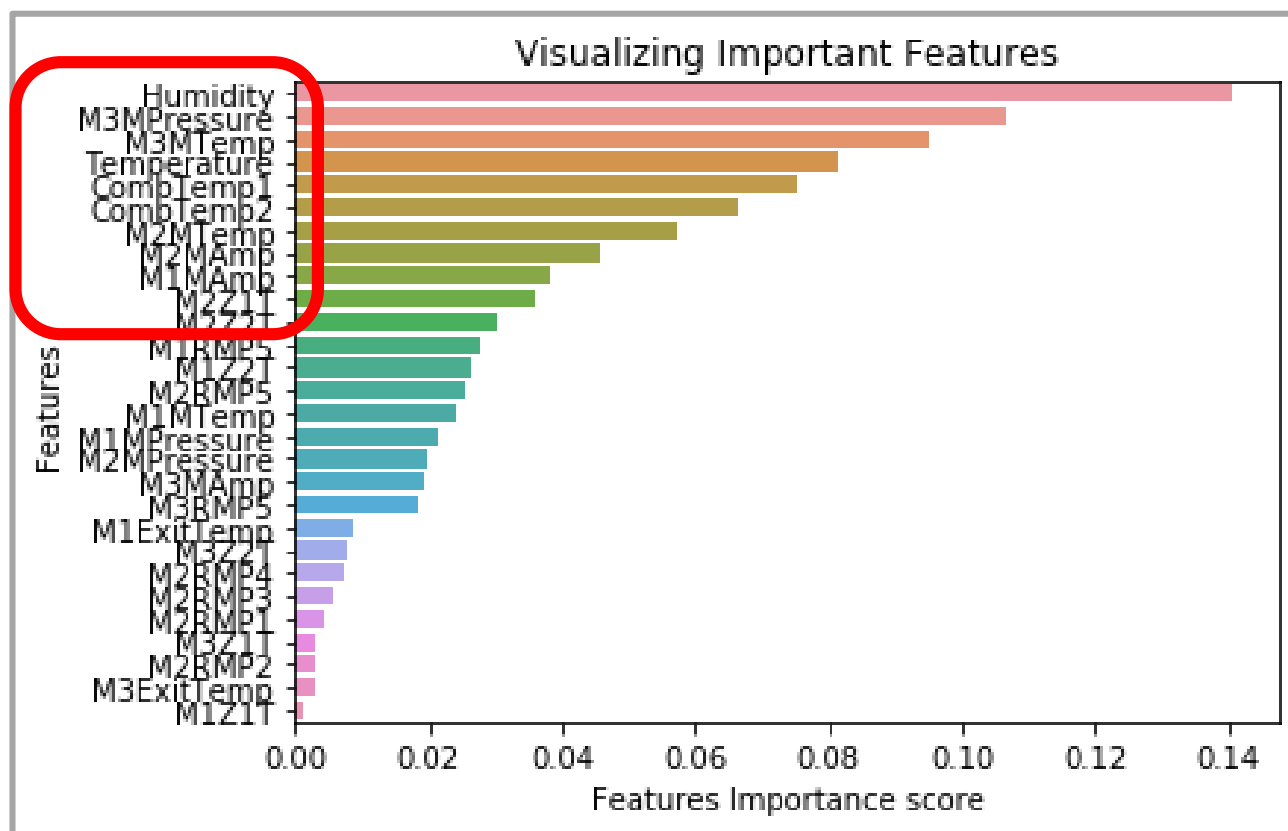
Construction de l'arbre très sensible à la présence de points erronés





Identifier les causes d'une non-conformité Produit

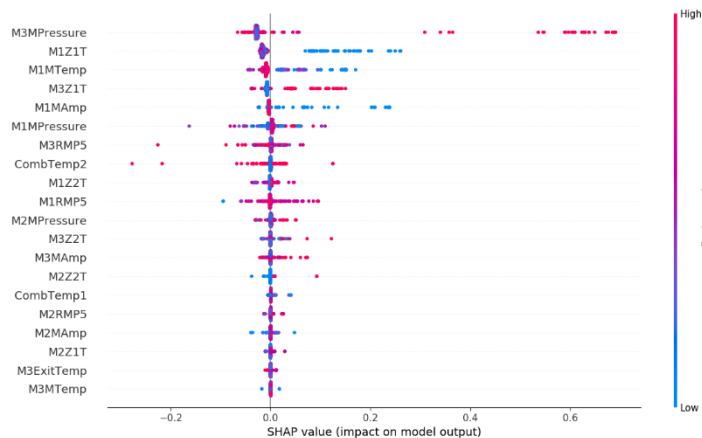
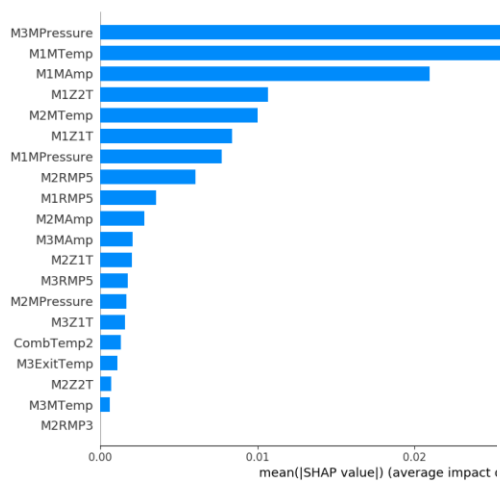
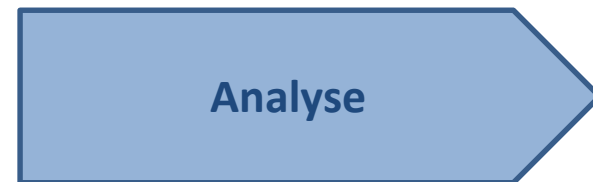
**Paramètres
discriminants**





Identifier les causes d'une non-conformité Produit

Données
Tous les paramètres
process et la conformité
KC10



Les paramètres process à surveiller car ils sont les principales causes de non-conformité sont:

M3MPressure
M1MTemp

...

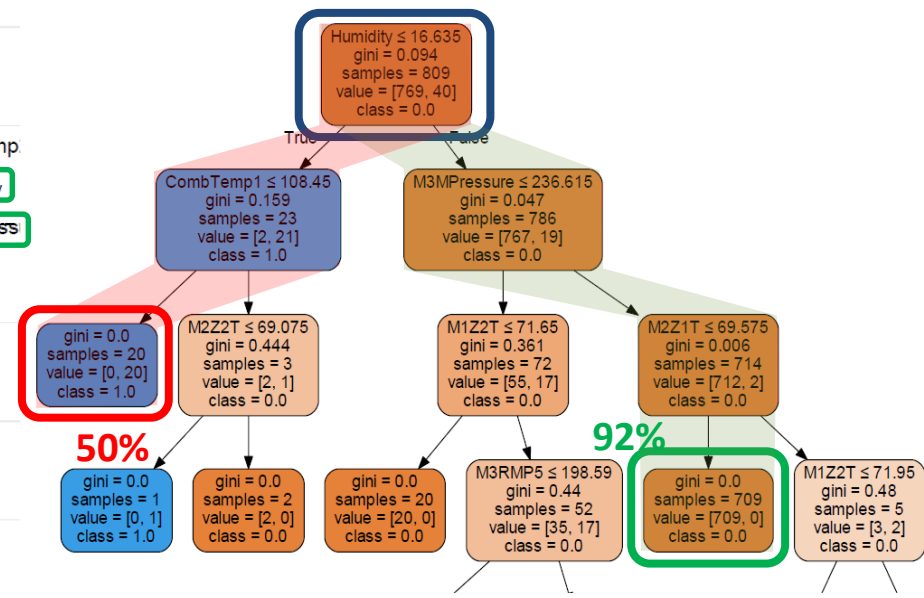
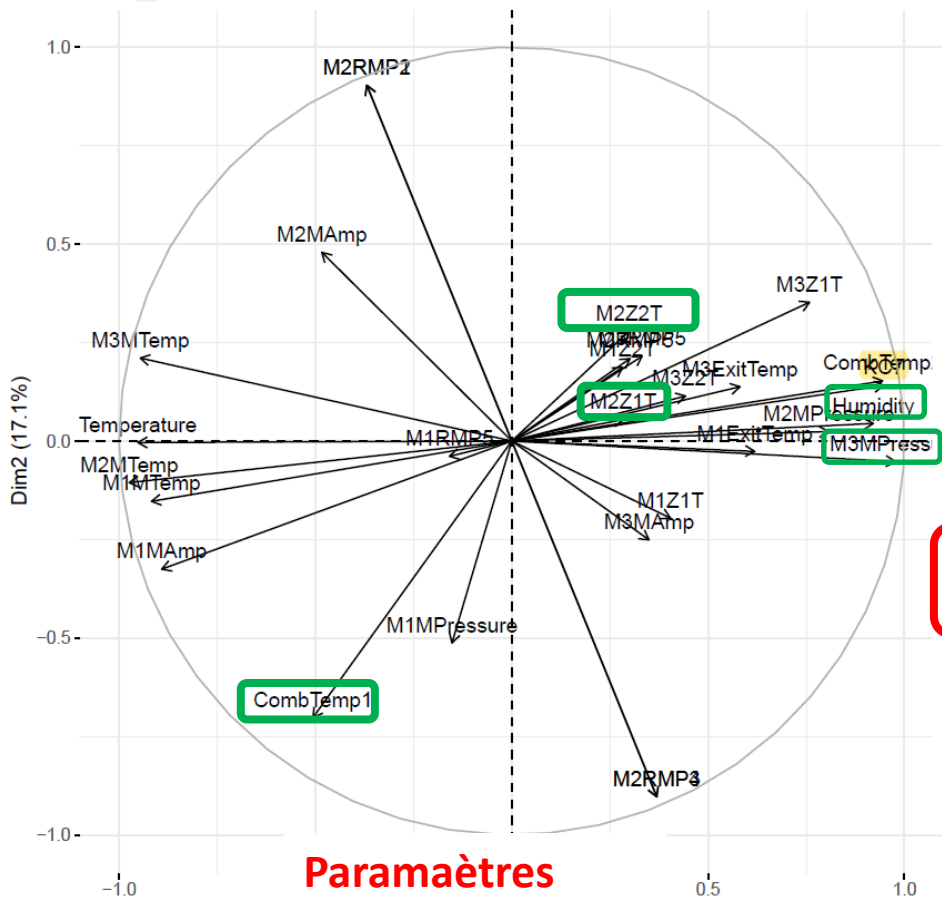
COMMENT DEPLOYER ?

Comparaison



Identifier les causes d'une non-conformité Produit

Paramètres discriminants de la spécification 7





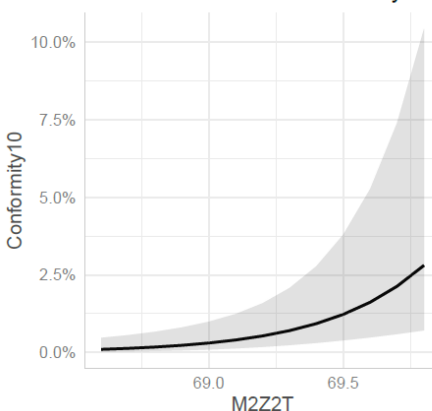
Prédire le taux de non-conformité Produit

Données
Tous les paramètres
process et la conformité
KC10

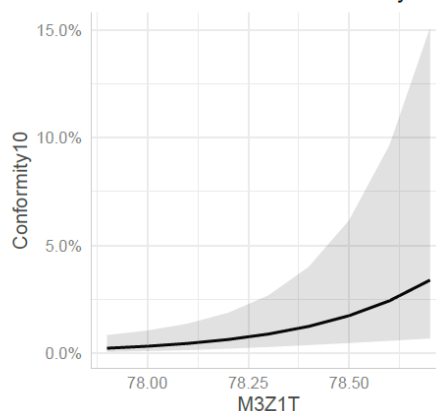
Traitement
Logistic Regression

Analyse

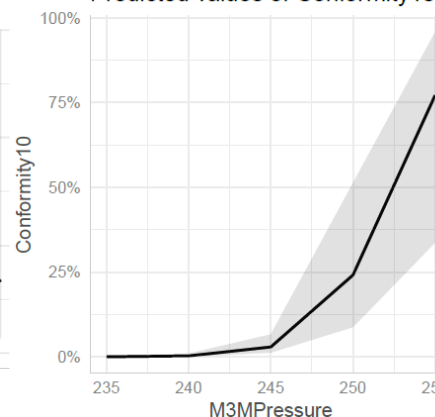
Predicted values of Conformity10



Predicted values of Conformity10



Predicted values of Conformity10



Le modèle de prédiction déterminé a un taux d'erreur de 4,6% et montre que les paramètres influents sont:

- M3MPressure
- M2Z2T
- M3Z2T

Paramètre	Coeff modèle	Risque 1 ^{ère}
(Constante)	-580.02862	1.08e-06 ***
M2Z2T	2.81189	0.000846 ***
M3Z2T	3.41257	0.004764 **
M3MPressure	0.47227	5.83e-08 ***

	0	1
0 prédit	738	20
1 prédit	17	34

Attention, l'analyse de la stratégie de réglage existante montre que ces paramètres sont liés à d'autres, et que la régression logistique ne s'applique que sur des paramètres indépendants.

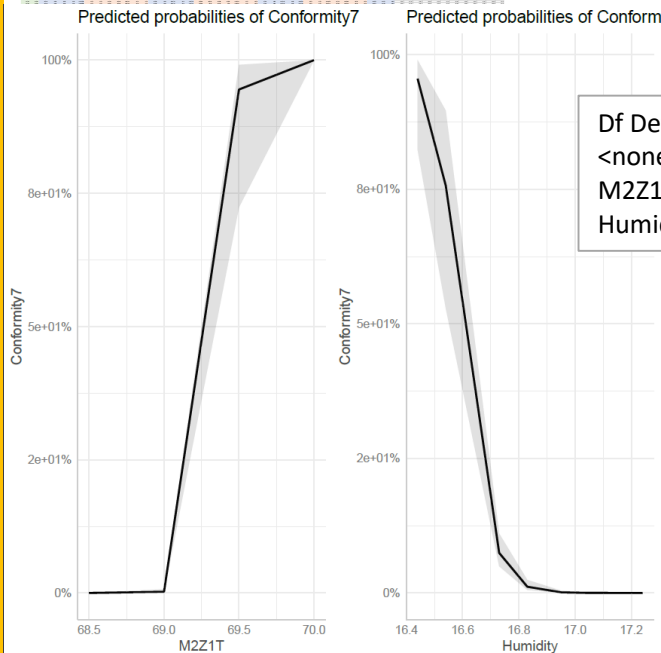


Prédire le taux de non-conformité Produit

Données
Tous les paramètres
process et la conformité
KC7

Traitement
Logistic Regression

Analyse



	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		171.43	177.43		
M2Z1T	1	202.72	206.72	31.288	2.225e-08 ***
Humidity	1	317.32	321.32	145.891	< 2.2e-16 ***

	0	1
0 prédit	767	19
1 prédit	2	21

Le modèle de prédiction déterminé a un taux d'erreur de 4,6% et montre que les paramètres influents sont:
M2Z1T
Humidity

Attention, l'analyse de la stratégie de réglage existante montre que ces paramètres sont liés à d'autres, et que la régression logistique ne s'applique que sur des paramètres indépendants.



Test d'hypothèse = Procédure de détermination de la plausibilité d'une hypothèse.

Hypothèse nulle (ou principale) : H₀,

- ✓ celle que l'on cherche à réfuter,
- ✓ celle qui est supposée « vraie », tant qu'on n'ait pas démontré le contraire
- ✓ hypothèse avec laquelle on dispose d'une loi de probabilité
- ✓ Hypothèse optimiste

Hypothèse alternative (ou contraire) : H₁,

- ✓ hypothèse contraire de l'hypothèse nulle
- ✓ celle que l'on cherche à démontrer
- ✓ mais souvent, on ne dispose pas d'une loi de probabilité
- ✓ Hypothèse pessimiste

Test scolaire:

- ✓ H₀: L'étudiant a toutes les compétences
- ✓ H₁: L'étudiant n'a pas les compétences suffisantes

Carte de contrôle:

- ✓ H₀: le système de production est bien réglé
- ✓ H₁: le système de production est dérégulé

Structure d'un test statistique:

1. Définition de H_0

Énoncé de l'hypothèse nulle H_0 et de l'hypothèse alternative H_1 .

2. Définition de la variable de décision

Calcul d'une variable de décision correspondant à une mesure de la distance entre les deux échantillons dans le cas de l'homogénéité, ou entre l'échantillon et la loi statistique dans le cas de la conformité. Plus cette distance sera grande et moins l'hypothèse nulle H_0 sera probable. En règle générale, cette variable de décision se base sur une statistique qui se calcule à partir des observations.

3. Définition de la limite d'acceptation et de refus

Calcul de la probabilité, en supposant que H_0 soit vraie, d'obtenir une valeur de la variable de décision au moins aussi grande que la valeur de la statistique que l'on a obtenue avec notre échantillon. Cette probabilité est appelée la p-value.

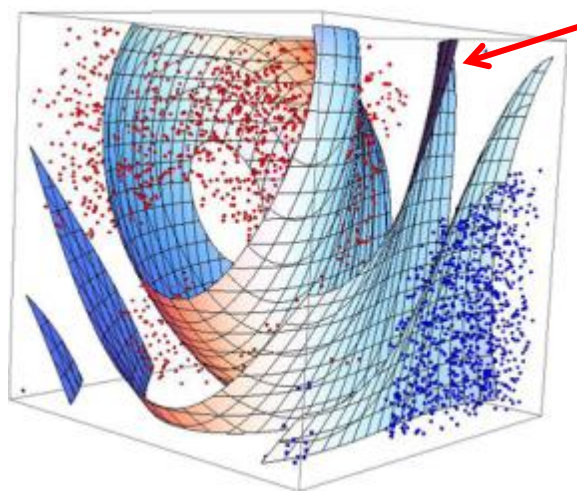
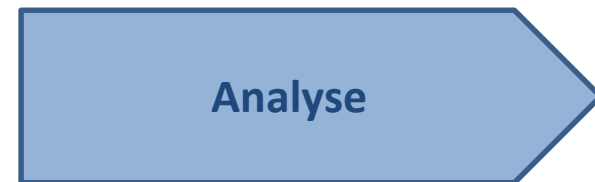
Conclusion du test, en fonction d'un risque seuil α_{seuil} , en dessous duquel on est prêt à rejeter H_0 . Souvent, un risque de 5% est considéré comme acceptable (c'est-à-dire que dans 5% des cas quand H_0 est vraie, l'expérimentateur se trompera et la rejettera).

	Hypothèse H_0 vraie	Hypothèse H_0 fausse
Hypothèse H_0 acceptée		Risque de type II (β)
Hypothèse H_0 refusée	Risque de type I (α)	



Prédire le taux de non-conformité Produit

Données
Tous les paramètres
process et la conformité
KC10



Séparateur non linéaire
Maximisation de la marge
entre les surfaces
séparatrices

Aucune représentation possible dans notre cas

La performance de prédiction de cet outil est supérieure aux outils précédents par contre la visualisation des résultats n'est pas possible.



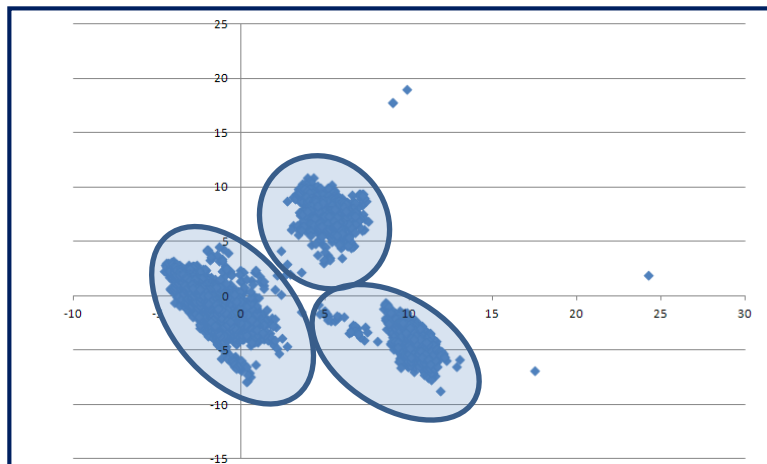
Détecter des anomalies

Données
Tous les paramètres
process et la conformité
KC10

Traitement

Analyse

Clustering



Le **clustering** (la classification non supervisé) est la recherche d'une partition, ou répartition des individus en classes, ou catégories; Ceci est fait en optimisant un critère visant à regrouper les individus dans des classes



Détecter des anomalies

Données
Tous les paramètres
process et la conformité
KC10

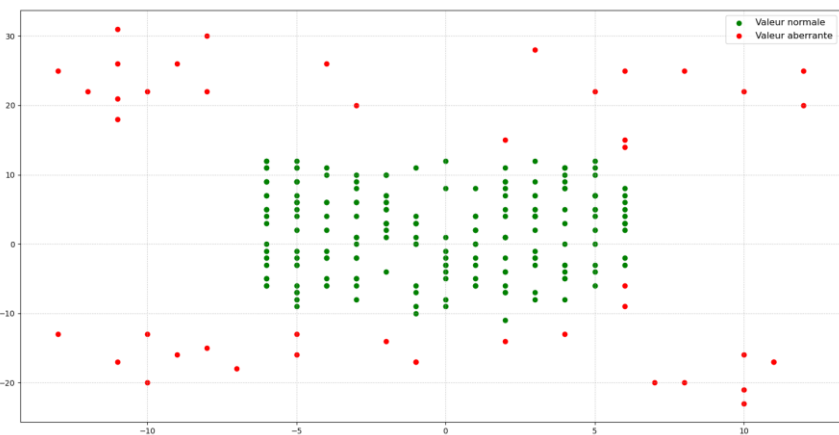
Traitement

Analyse

Novelty detection

Algorithme **Local OutlierFactor** (LOF) → basée sur K-NN (k plus proches voisins)

NoveltyDetection:
Algorithme **One-class SVM** → basée sur SVM (Support Vector Machines).



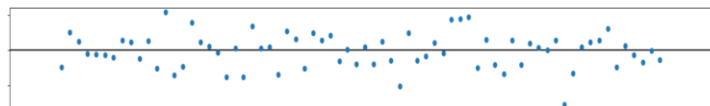
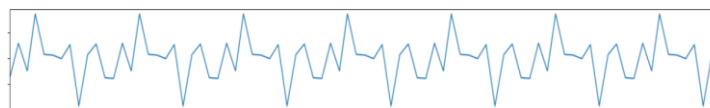
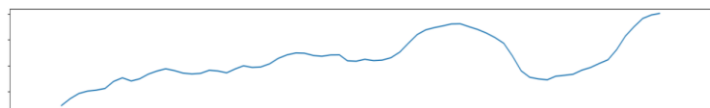


Analyser des dérives

Données
Tous les paramètres
process et la conformité
KC10

Traitement

Analyse



Décomposition d'une série temporelle

Tendance Lissage de la série
Évolution à long terme

Saisonnalité
Périodicité de la série

Résidu



Prédire les dérives

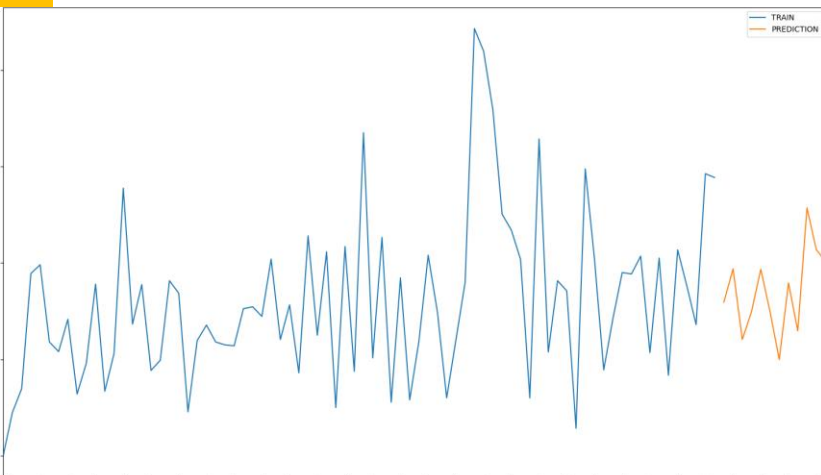
Données
Tous les paramètres
process et la conformité
KC10

Traitement

Analyse

Méthode de Holt-Winters(lissage exponentiel triple)

Pour prédire l'évolution d'une série temporelle, il existe plusieurs méthodes, la méthode de Holt-Winters permet de prendre en compte de grandes variabilités dans les données avec une saisonnalité potentielle.



COMMENT DEPLOYER ?



Identifier les paramètres clés de mon système de production

Identifier les paramètres clés de mes produits

Identifier les stratégies de réglage du système de production

Identifier les causes d'une non-conformité Produit

Prédire le taux de non-conformité Produit



ACP, kACP, ...

C4.5, SVM,
Random Forest,
...

SHAP, ...

SVM, LogitR, ...

KNN, Novelty
detection,

Traitement

Analyser les causalités entre les variabilités des paramètres

Extraire des règles

Extraire les paramètres discriminants des règles

Optimiser un modèle de prédiction des classes

Analyser et détecter des dérives

Analyse

Identifier les influences des paramètres sur les variabilités process

Extraire des plages de réglages

Identifier les PPs « pilotant » la non conformité

Prédire les non conformités

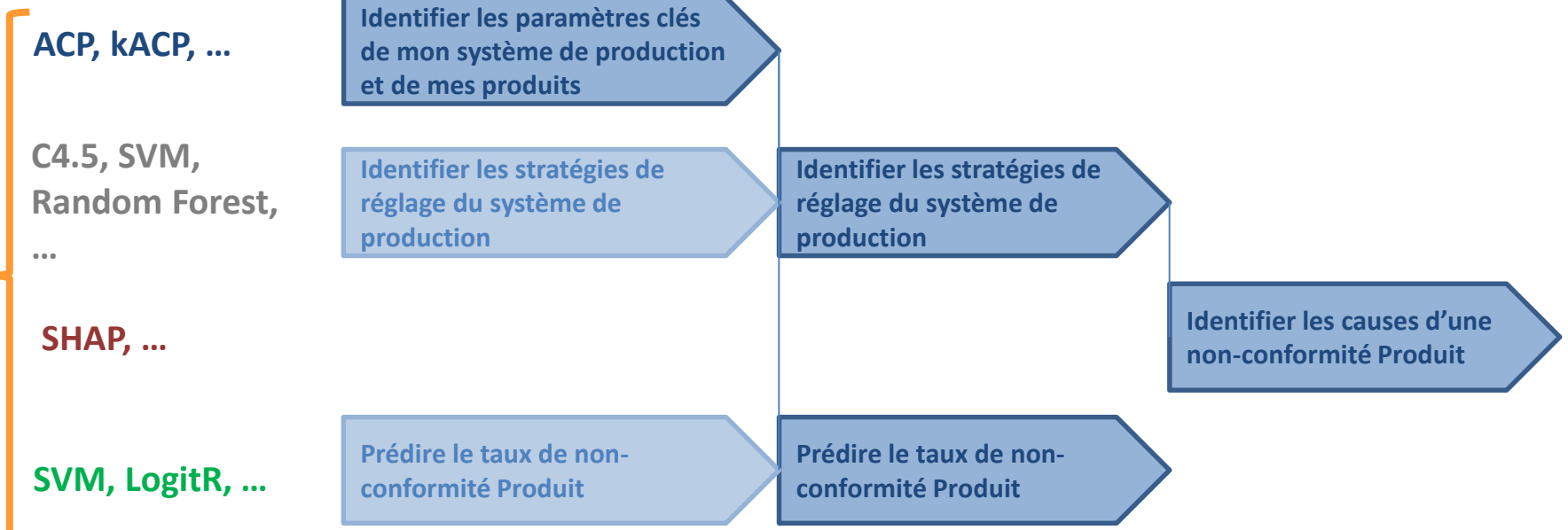
Prédire les dérives

COMMENT DEPLOYER ?



- Identifier les paramètres clés de mon système de production
- Identifier les paramètres clés de mes produits
- Identifier les stratégies de réglage du système de production
- Identifier les causes d'une non-conformité Produit
- Prédire le taux de non-conformité Produit

Démarche



COMMENT DEPLOYER ?



Identifier les paramètres clés de mon système de production

Identifier les paramètres clés de mes produits

Identifier les stratégies de réglage du système de production

Identifier les causes d'une non-conformité Produit

Prédire le taux de non-conformité Produit

Démarche

Identifier des configurations d'exploitation du système de production

Identifier les stratégies de réglage du système de production

Identifier les causes d'une non-conformité Produit

KNN, ...

C4.5, SVM,
Random Forest,
...

SHAP, ...



COMMENT DEPLOYER ?



Identifier les paramètres clés de mon système de production

Identifier les paramètres clés de mes produits

Identifier les stratégies de réglage du système de production

Identifier les causes d'une non-conformité Produit

Prédire le taux de non-conformité Produit

Démarche

ACP, kACP, ...

Identifier les paramètres clés de mon système de production et de mes produits

KNN, ...

Identifier des configurations d'exploitation du système de production

Random Forest Regressor, ANN

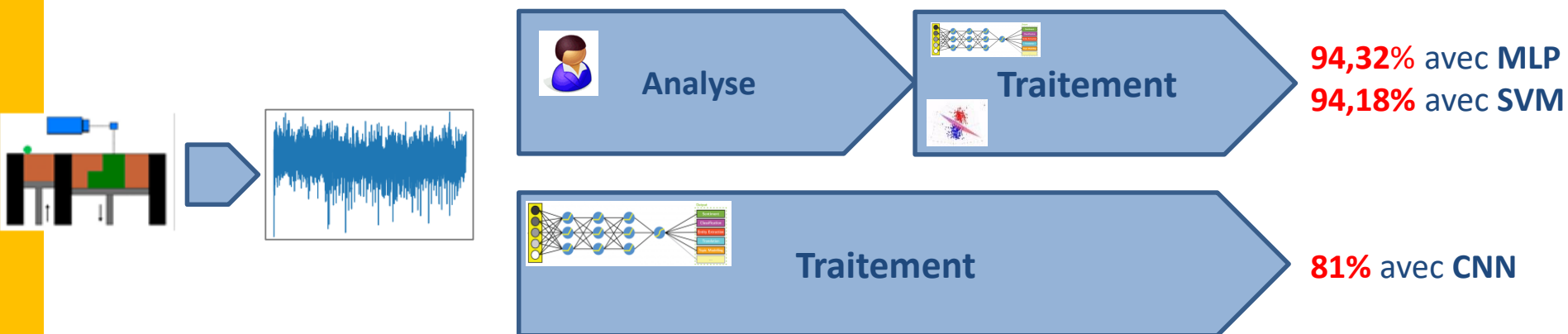
Prédire les performances du système de production

...



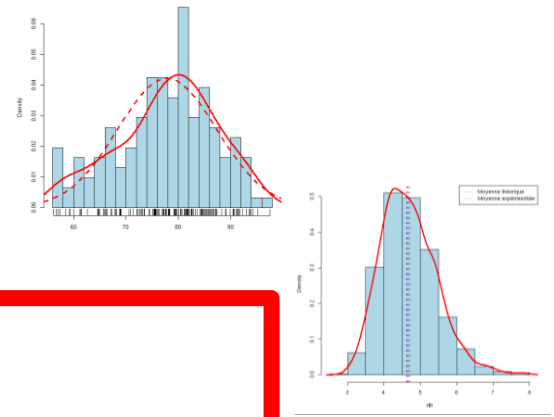


Exemple de prédiction de densité:

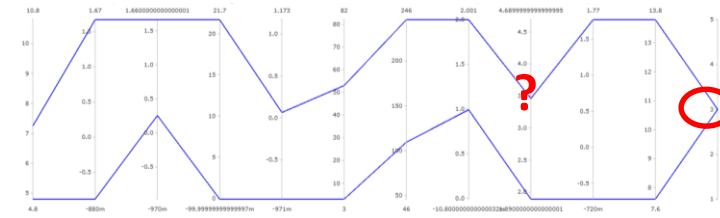
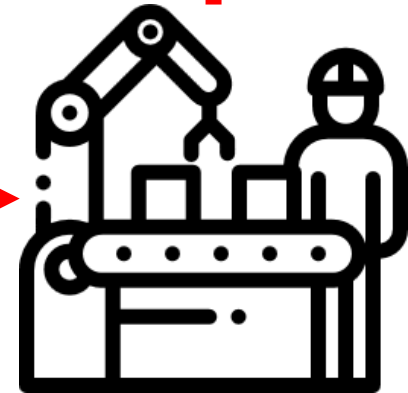




Attention, nos données
sont entachées
d'incertitude de
mesure !

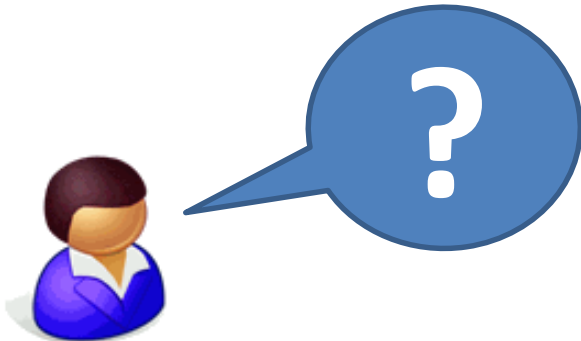


f_x

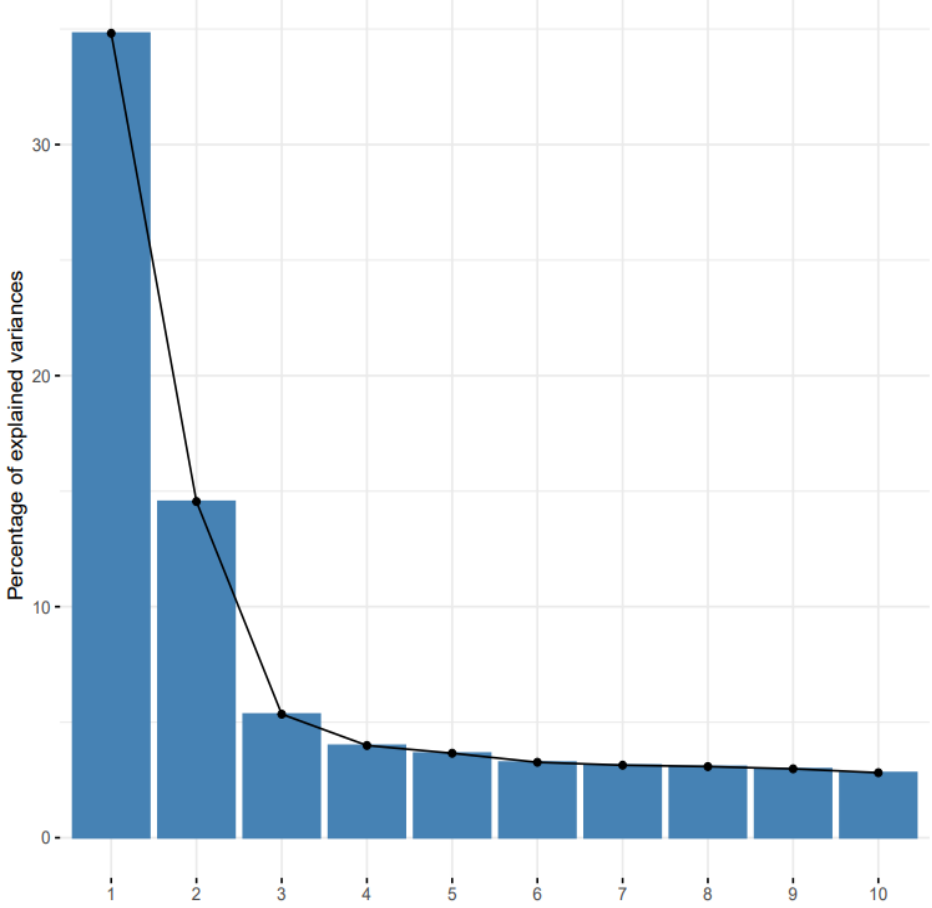


Exemples

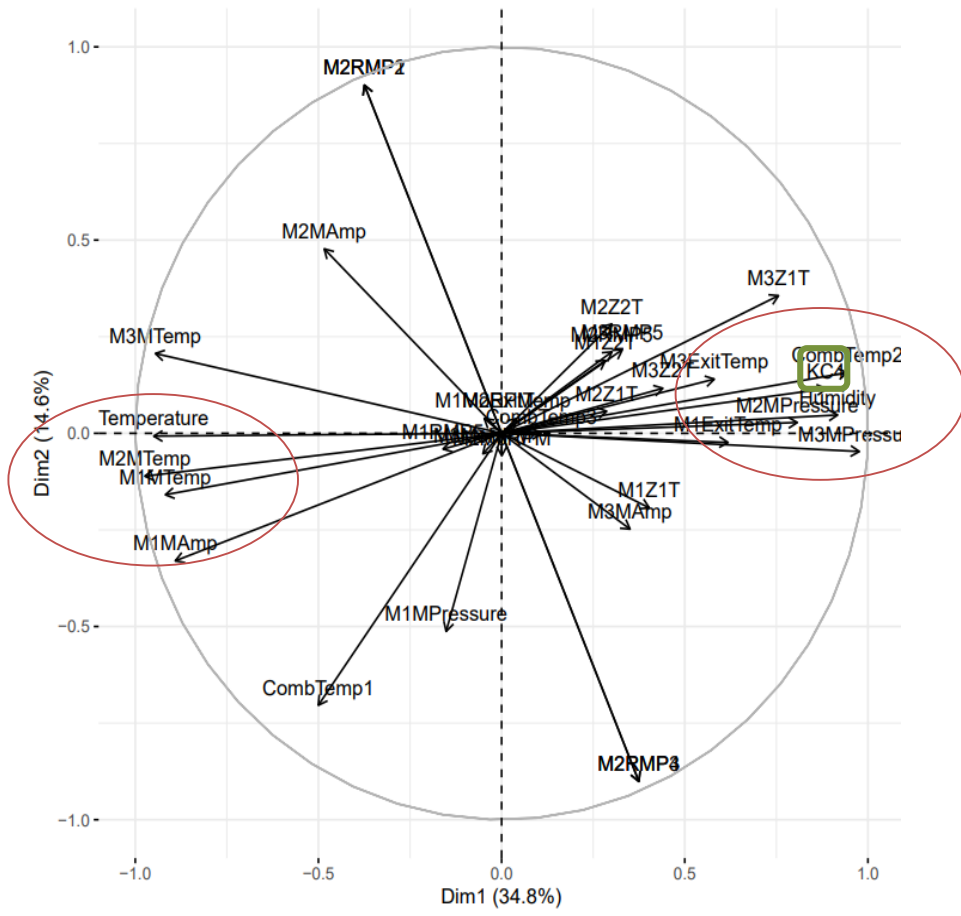
KC4 du système de production



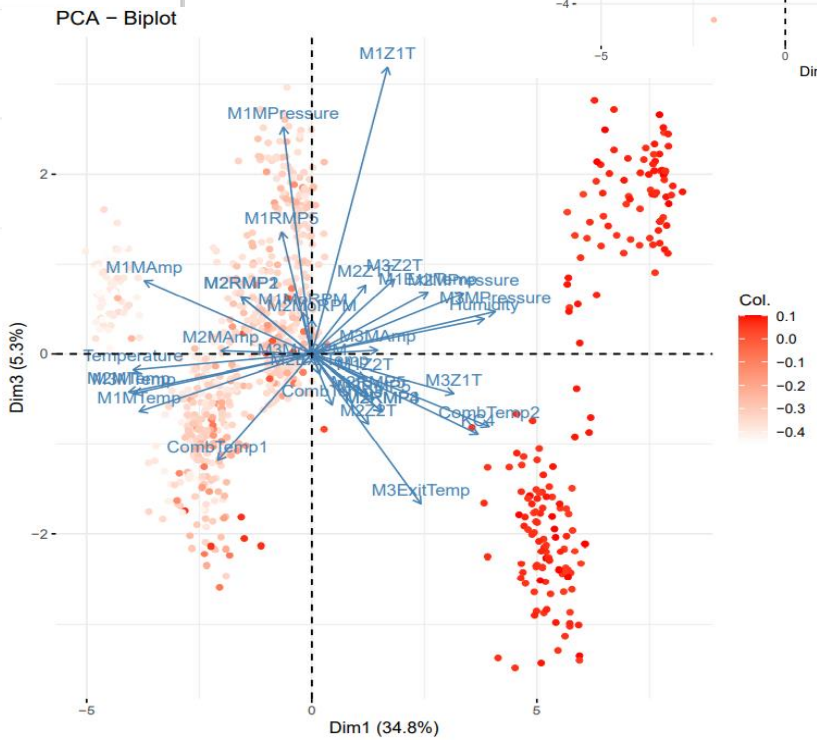
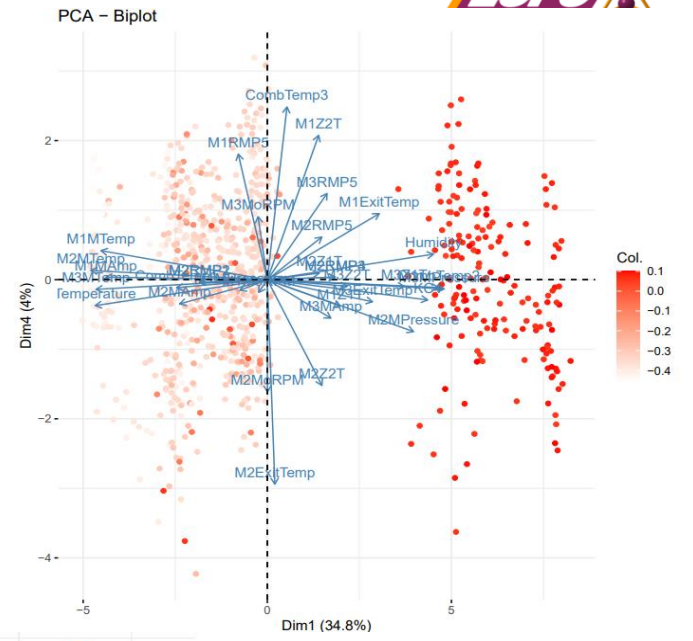
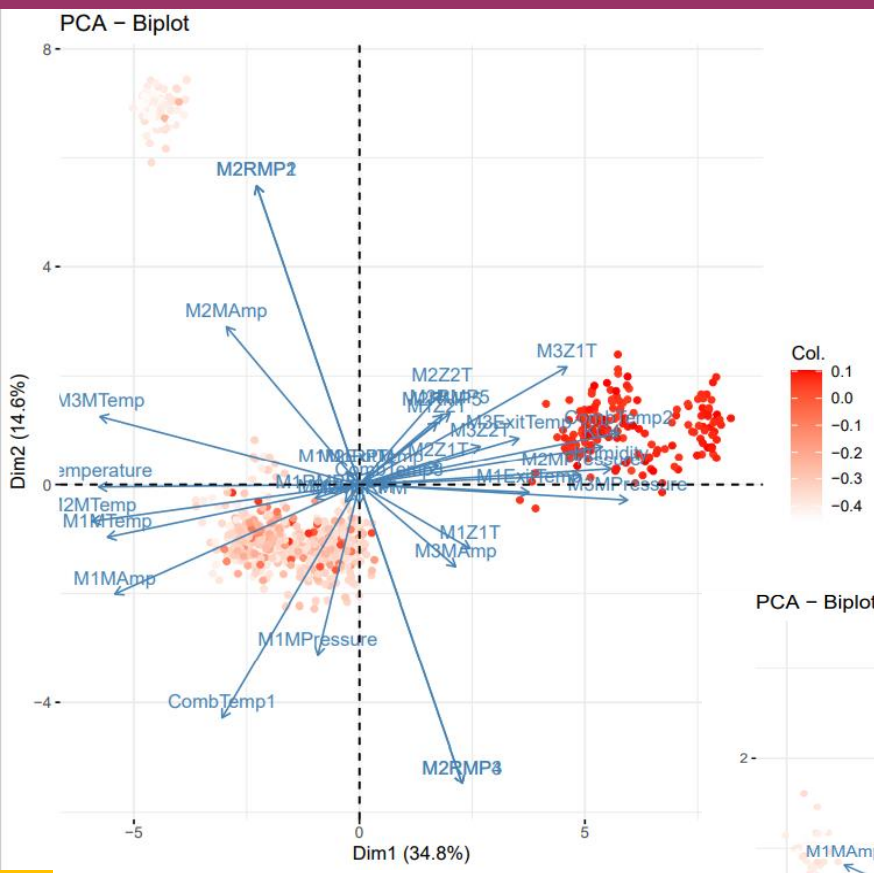
Scree plot KC4



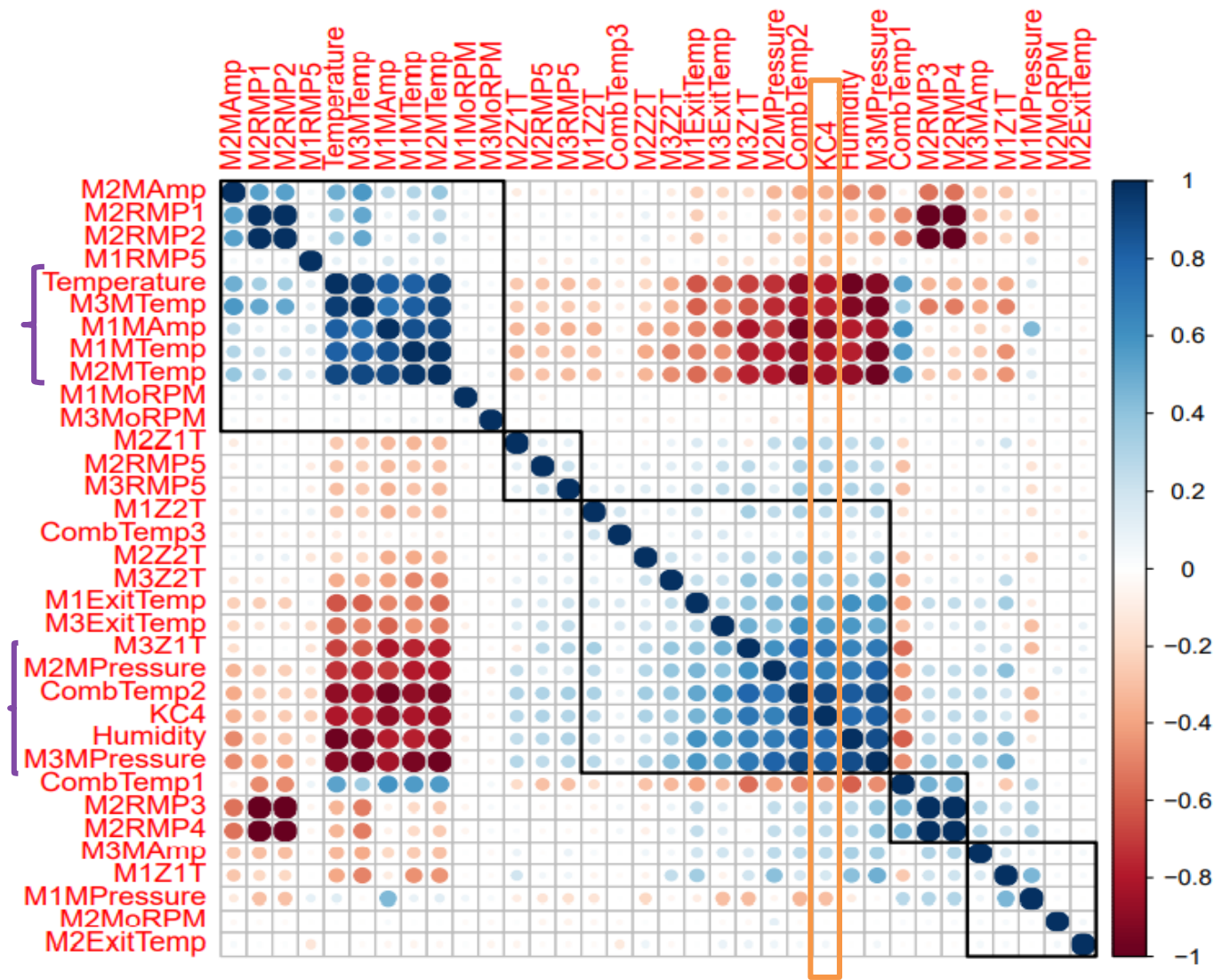
Variables - PCA



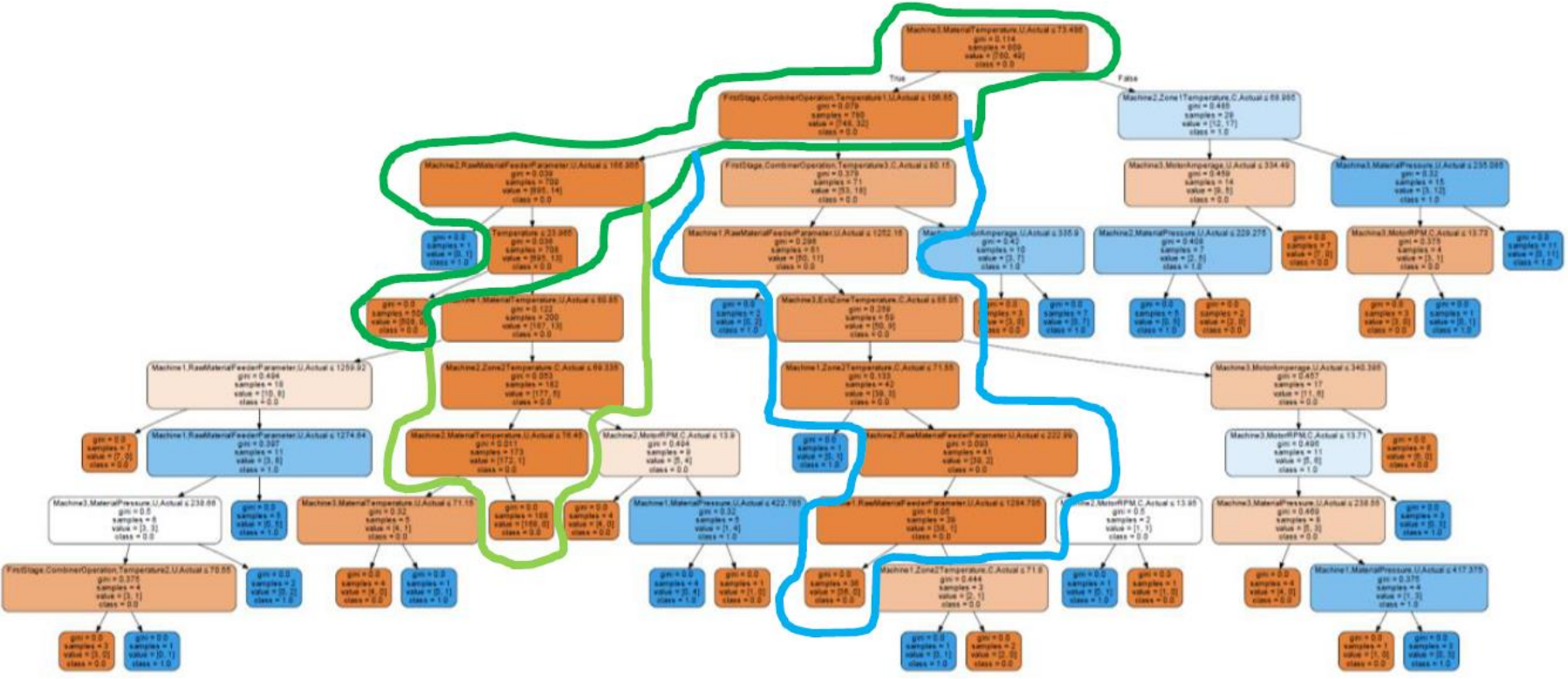
Exemples



Exemples

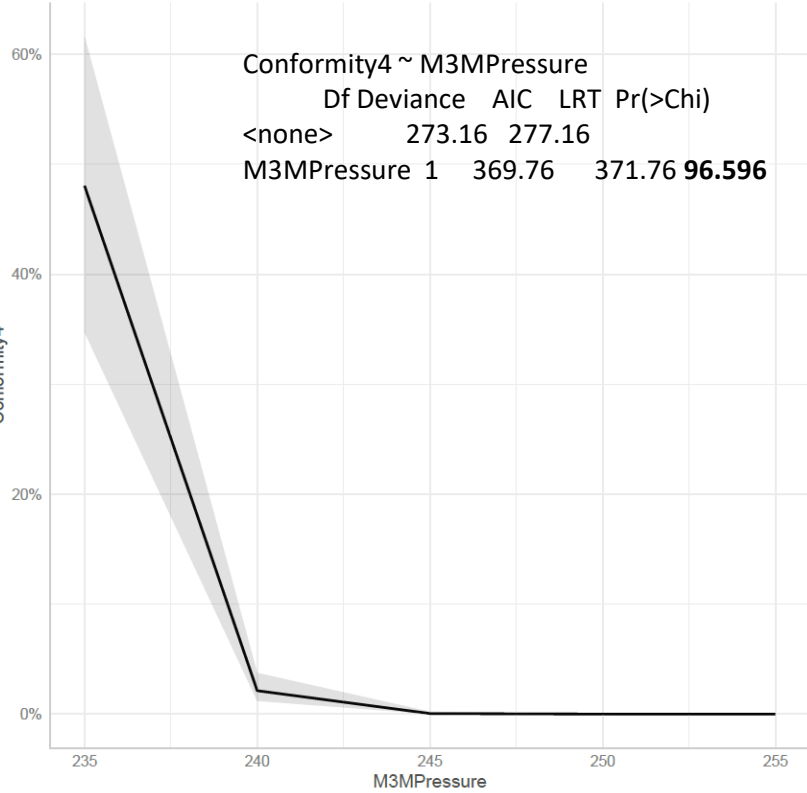


KC4 Decision Tree



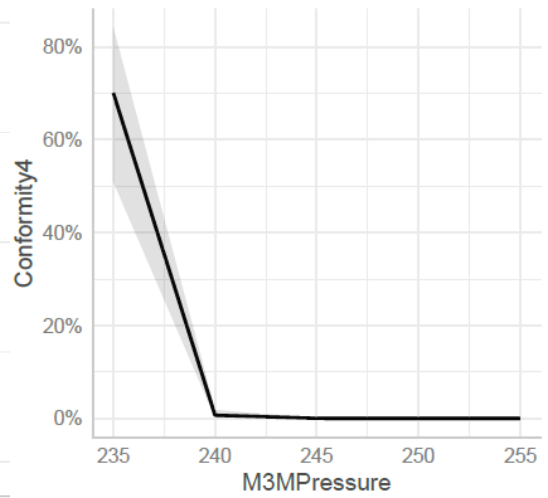
Exemples

Predicted probabilities of Conformity4

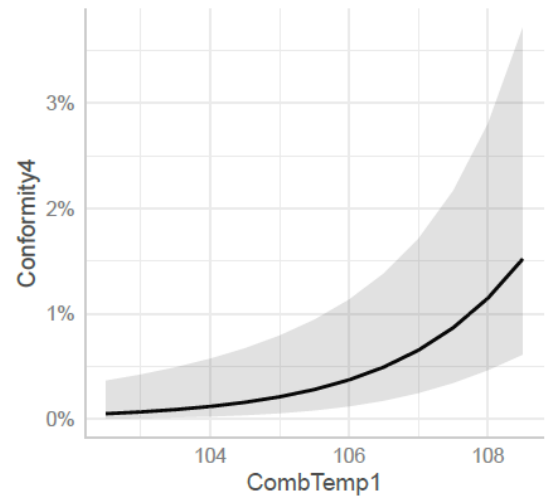


Conformity4 ~ M3MPressure
 Df Deviance AIC LRT Pr(>Chi)
 <none> 273.16 277.16
 M3MPressure 1 369.76 371.76 **96.596 < 2.2e-16 *****

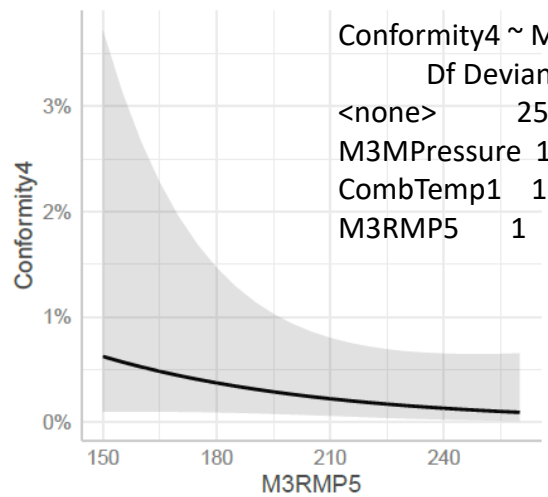
Predicted probabilities of Conformity4



Predicted probabilities of Conformity4



Predicted probabilities of Conformity4



Conformity4 ~ M3MPressure + CombTemp1 + M3RMP5
 Df Deviance AIC LRT Pr(>Chi)
 <none> 254.35 262.35
 M3MPressure 1 362.37 368.37 108.027 < 2.2e-16 ***
 CombTemp1 1 268.03 274.03 13.688 0.0002158 ***
 M3RMP5 1 256.15 262.15 1.801 **0.1796225**