

# Analyse de la causalité pour l'aide à la décision dans un contexte de supervision pour l'industrie 4.0

Kenza AMZIL

**Directeur de thèse :**

M. Lionel ROUCOULES

**Encadrantes :**

Mme. Esma Yahia

Mme. Nathalie Klement

04/12/2020

# LA SUPERVISION DES SYSTÈMES DE PRODUCTION POUR L'AIDE À LA DÉCISION



La **supervision** consiste à surveiller le fonctionnement des systèmes de production et leur environnement, alerter des éventuelles **dérives**, et pouvoir les **éviter**.



La **supervision** doit permettre une prise de décision :

- Dérives de processus ou système de production ;
- Optimisation du rendement de production ;
- Amélioration de la qualité ;
- Optimisation des coûts.

Contexte général et  
problématique

Etat de l'art et  
proposition

Cas d'étude

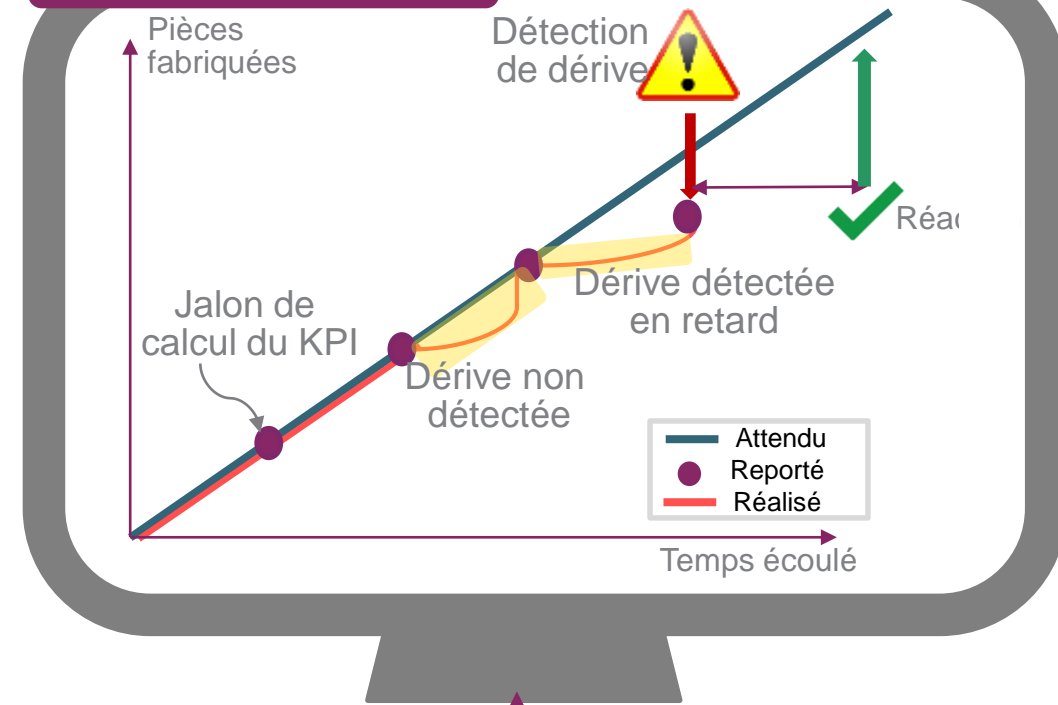
Conclusion et  
perspectives

# LA SUPERVISION DES SYSTÈMES DE PRODUCTION POUR L'AIDE À LA DÉCISION

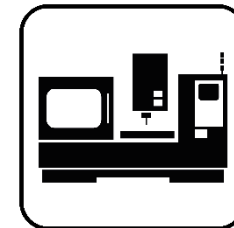
- Il faut **prévoir les dérives** et les détecter en temps réel.
- **Identifier les liens de causalité** entre les KPI et les données, de façon **exhaustive et objective**.
- **Prioriser causes identifiées** pour agir efficacement.

Reporting

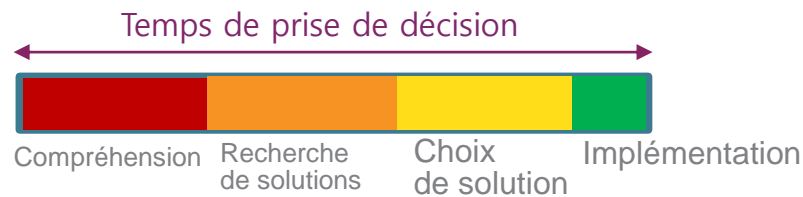
KPI : quantité fabriquée



Récupération des données



Sources de données



Contexte général et problématique

Etat de l'art et proposition

Cas d'étude

Conclusion et perspectives

# LA SUPERVISION DES SYSTÈMES DE PRODUCTION POUR L'AIDE À LA DÉCISION

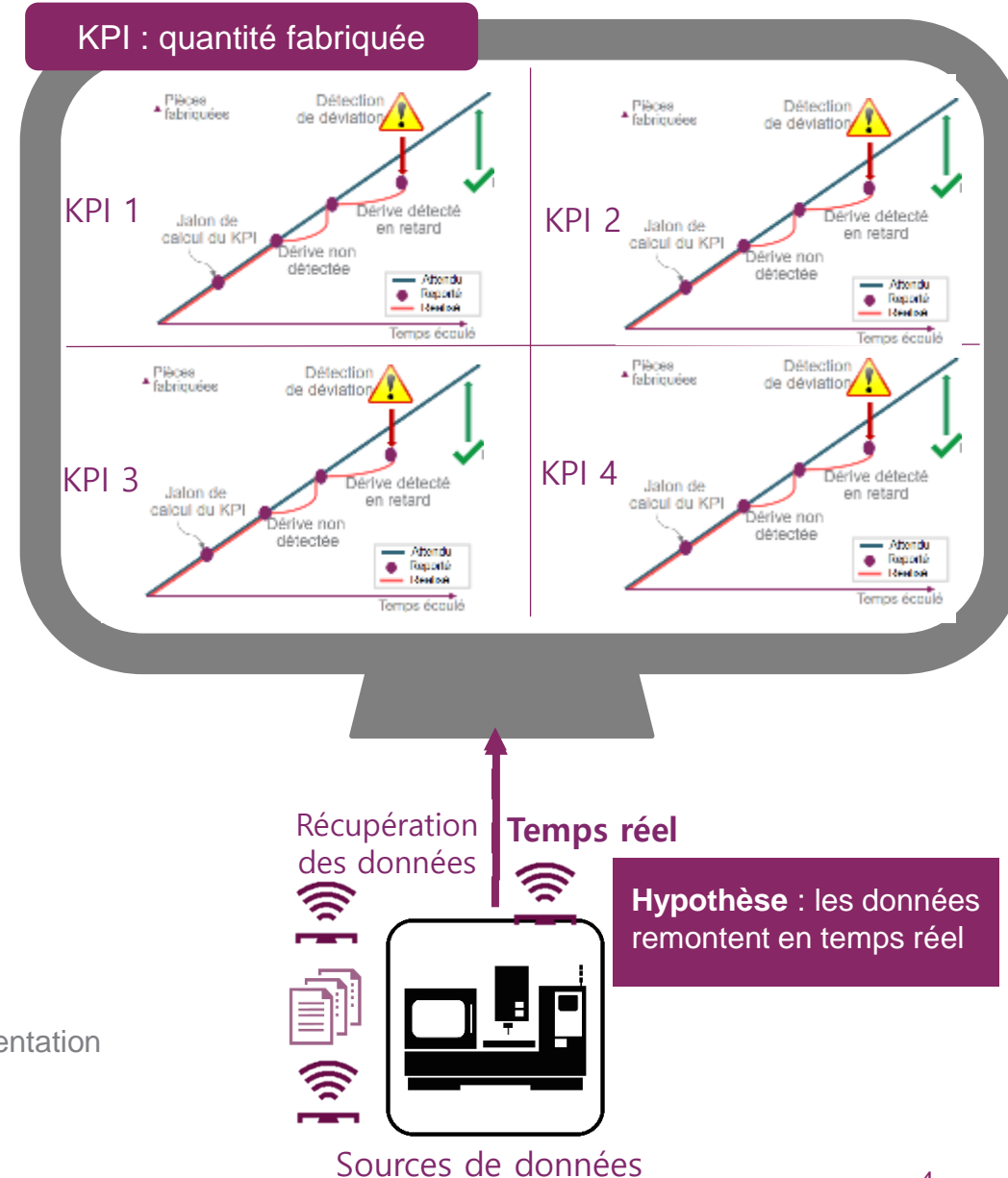
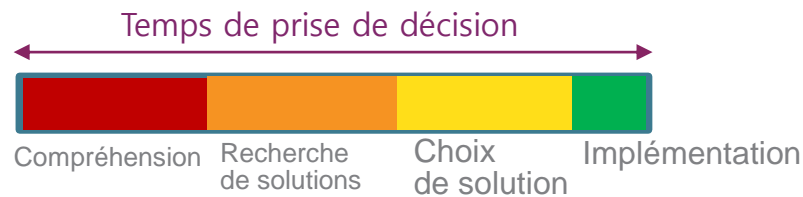
- Il faut **prévoir les dérives** et les détecter en temps réel.
- **Identifier les liens de causalité** entre les KPI et les données, de façon **exhaustive et objective**.
- **Prioriser causes identifiées** pour agir efficacement.
- Il faut que l'approche soit **générique**

Contexte général et problématique

Etat de l'art et proposition

Cas d'étude

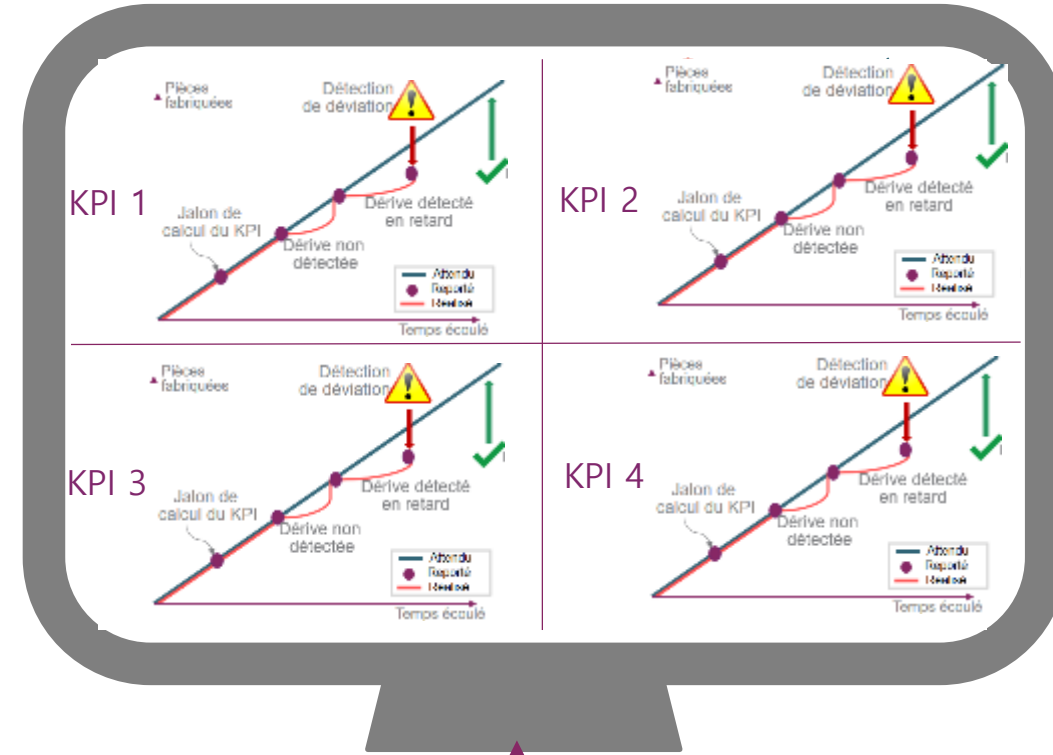
Conclusion et perspectives



# LA SUPERVISION DES SYSTÈMES DE PRODUCTION POUR L'AIDE À LA DÉCISION

- Il faut **prévoir les dérives** et les détecter en temps réel.
- **Identifier les liens de causalité** entre les KPI et les données, de façon **exhaustive et objective**.
- **Prioriser causes identifiées** pour agir efficacement.
- Il faut que l'approche soit **générique**

**Postulat :** L'environnement de production peut être équipé de capteurs et de sources de données dans un contexte industrie 4.0



Récupération des données **Temps réel**



**Hypothèse :** les données remontent en temps réel

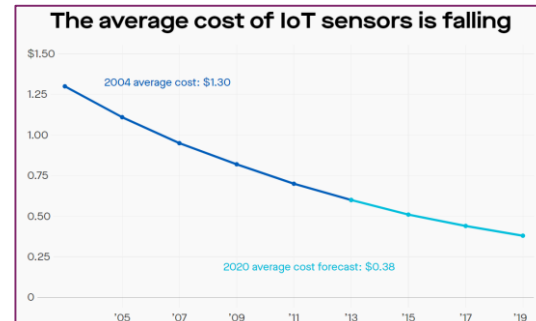
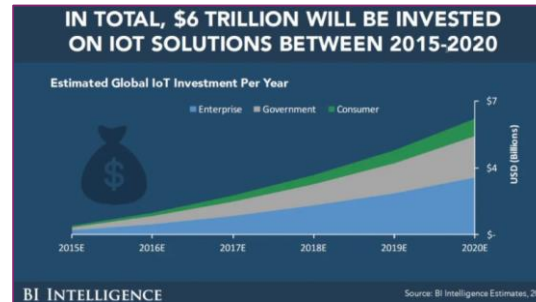
Sources de données

Contexte général et problématique

Etat de l'art et proposition

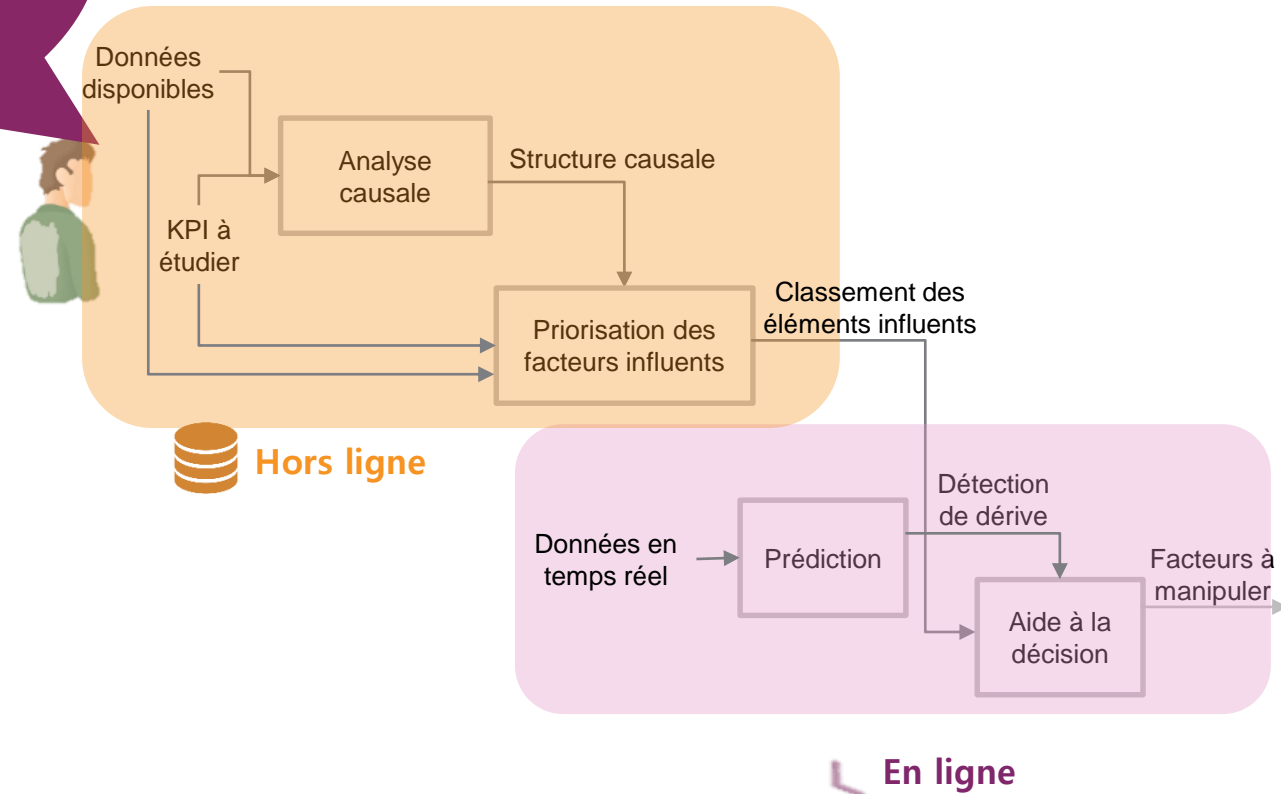
Cas d'étude

Conclusion et perspectives



# LA SUPERVISION DES SYSTÈMES DE PRODUCTION POUR L'AIDE À LA DÉCISION

- Il faut **prévoir les dérives** et les détecter en temps réel.
- **Identifier les liens de causalité** entre les KPI et les données, de façon **exhaustive et objective**.
- **Prioriser causes identifiées** pour agir efficacement.
- Il faut que l'approche soit **générique**



Contexte général et problématique

Etat de l'art et proposition

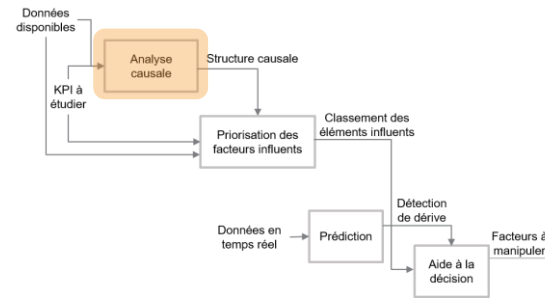
Cas d'étude

Conclusion et perspectives

# ANALYSE CAUSALE : CARACTÉRISTIQUES DE LA CAUSALITÉ



La notion de causalité, pourtant abordée par plusieurs domaines de recherche, n'a pas encore de définition figée et vérifiée.



## Caractéristiques de la causalité :

- Aspect probabiliste ;
- Ordre temporel ;
- Dépendance de l'espace de recherche
- Représentation graphique;
- Taille de l'effet ;
- Stabilité ;
- Plausibilité ;
- Vérificabilité;

- Acyclicité ;
- Consistance ;
- Régularité ;
- Relation dose-effet ;
- Spécificité

**Caractéristiques non retenues dans notre contexte**

Ces caractéristiques, en cas de leur vérification, représentent une série d'arguments; qui, ajoutés les uns aux autres, nous permettrons d'approcher la présomption de relation causale.

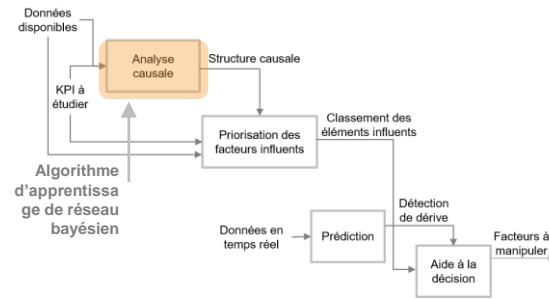
Contexte général et problématique

Etat de l'art et proposition

Cas d'étude

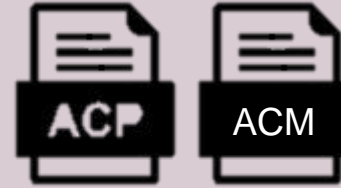
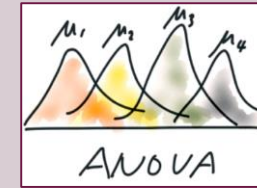
Conclusion et perspectives

# ANALYSE CAUSALE : LES DIFFÉRENTES MÉTHODES

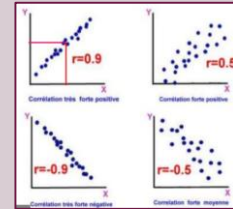


## Méthodes descriptives :

## Méthodes exploratoires et/ou analytiques pour identifier les relations entre variables :

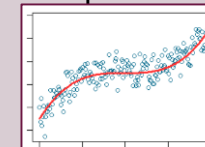


Exp	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	Y <sub>exp</sub>
1	-1	-1	-1	18,1
2	+1	-1	-1	16,0
3	-1	+1	-1	17,1
4	+1	+1	-1	17,0
5	-1	-1	+1	17,8
6	+1	-1	+1	17,2
7	-1	+1	+1	18,1
8	+1	+1	+1	17,0

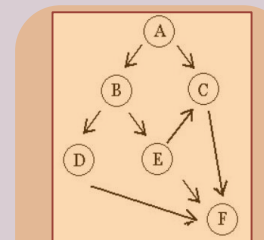


### Tests de corrélation

### Plans d'expériences



### Régression



### Réseaux bayésiens

## Caractéristiques de la causalité :

- ✓ - Aspect probabiliste ;
- ✓ - Ordre temporel ;
- ✓ - Dépendance de l'espace de recherche
- ✓ - Représentation graphique;
- Taille de l'effet ;
- Stabilité ;
- Plausibilité ;
- ✓ - Vérificabilité;



- Non exhaustives ;
- Subjectives ;
- Temps de résolution important ;
- Mobilisation des ressources humaines ;
- Possibilité de surestimation du problème ;
- Difficulté de choix de la meilleure solution.



Algorithme d'apprentissage de réseau bayésien



- Construction du graphe à partir des connaissances à priori => subjectivité
- Temps de construction important ;
- Mobilisation des ressources humaines ;
- Difficulté de choix de la meilleure solution.

Contexte général et problématique

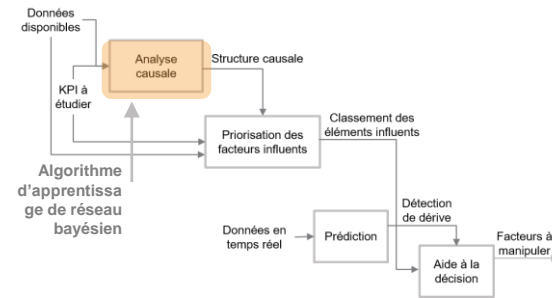
Etat de l'art et proposition

Cas d'étude

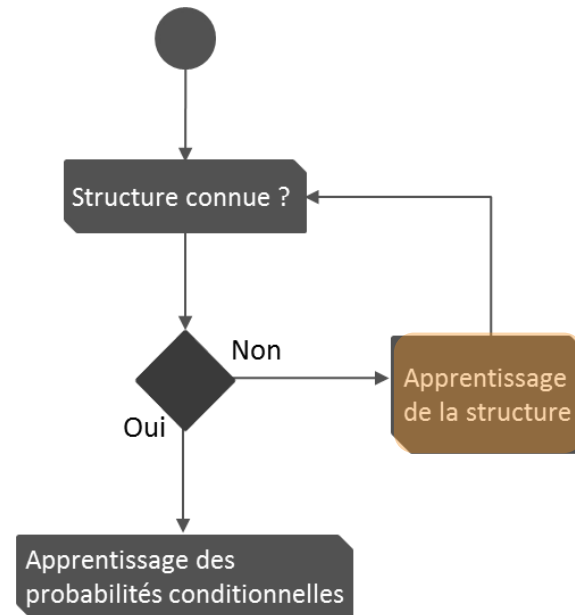
Conclusion et perspectives



# ANALYSE CAUSALE : APPRENTISSAGE DE RÉSEAU BAYÉSIEN



Apprentissage de réseau bayésien



Prétraitement des données

Apprentissage de la structure

Recherche d'indépendances conditionnelles (Pearl et Verma)

Méthodes basées sur un score

- Explosion du nombre de tests d'indépendance à effectuer.
- A quel ordre s'arrêter ?
- Sensibilité au seuil du test d'indépendance
- L'orientation des arcs est faite à posteriori

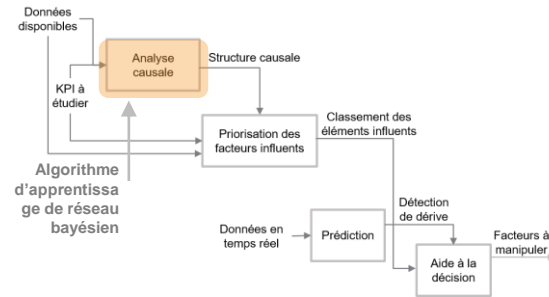
Contexte général et problématique

Etat de l'art et proposition

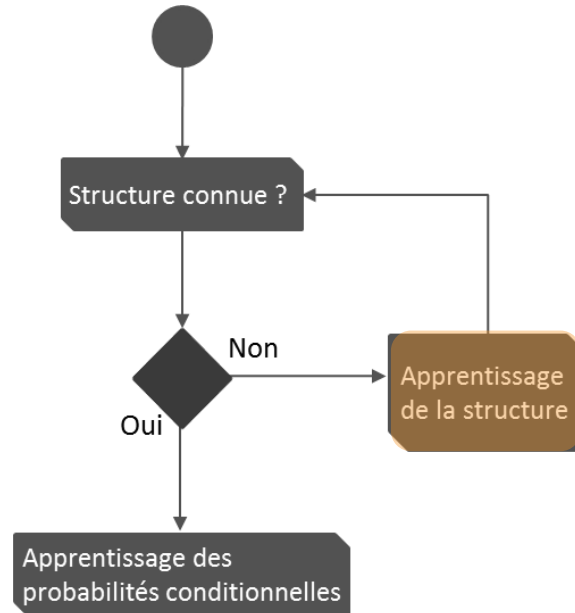
Cas d'étude

Conclusion et perspectives

# ANALYSE CAUSALE : APPRENTISSAGE DE RÉSEAU BAYÉSIEN



Apprentissage de réseau bayésien



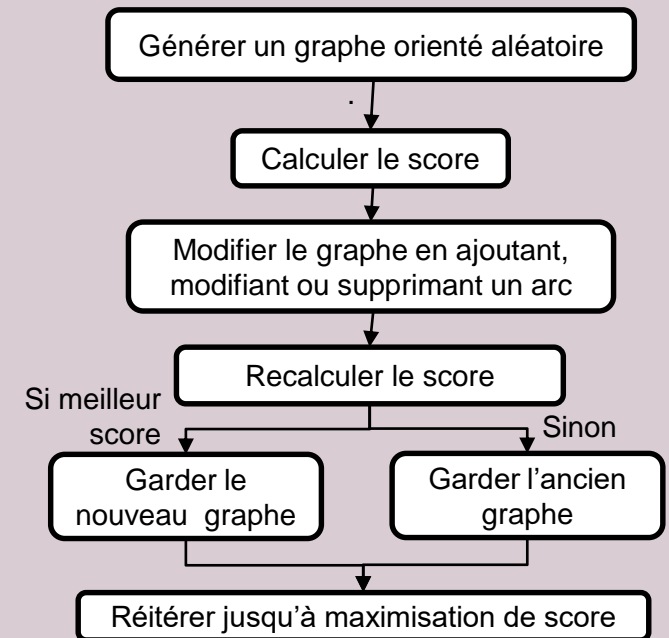
Prétraitement des données

Apprentissage de la structure

Recherche d'indépendances conditionnelles (Pearl et Verma)

Méthodes basées sur un score

Fonction de score basée sur la vraisemblance que les données observées soient générées à partir du graphe évalué



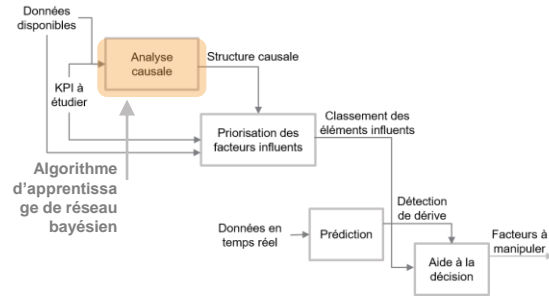
Contexte général et problématique

Etat de l'art et proposition

Cas d'étude

Conclusion et perspectives

# ANALYSE CAUSALE : APPRENTISSAGE DE RÉSEAU BAYÉSIEN



Prétraitement des données

Apprentissage de la structure

Méthodes basées sur un score

## Apprentissage de la structure

Algorithme Hill climbing

Générer un graphe orienter aléatoire

Calculer le score

Modifier le graphe en ajoutant, modifiant ou supprimant un arc

Recalculer le score

Si meilleur score

Garder le nouveau graphe

Sinon

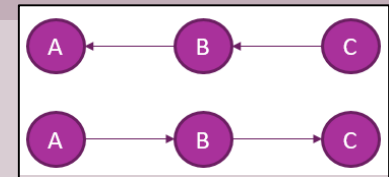
Garder l'ancien graphe

Réitérer jusqu'à maximisation de score

Algorithme Hill climbing

Algorithme Hill climbing

Sélection des meilleurs graphes



Éliminer les graphes contenant des relations impossibles

Garder le graphe contenant des relations causales déjà connues

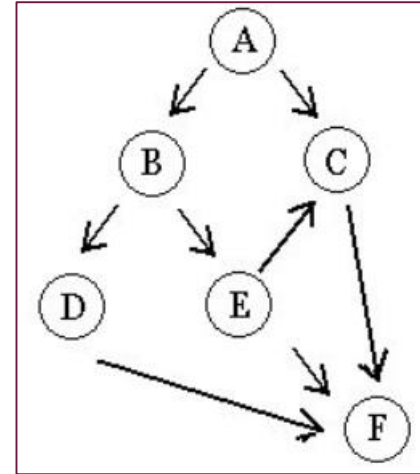
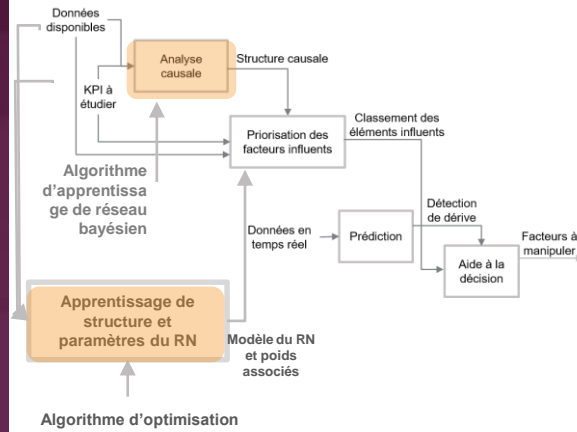
Contexte général et problématique

Etat de l'art et proposition

Cas d'étude

Conclusion et perspectives

# PRIORISATION DES ÉLÉMENTS INFLUENTS



C	E	F=5	F=10
1	1	0,95	0,05
1	0	0,64	0,36
0	1	0,35	0,65
0	0	0,20	0,80

**Les probabilités ne sont pas faciles à interpréter quand il s'agit de classer les variables indépendamment des valeurs qu'elles peuvent prendre.**

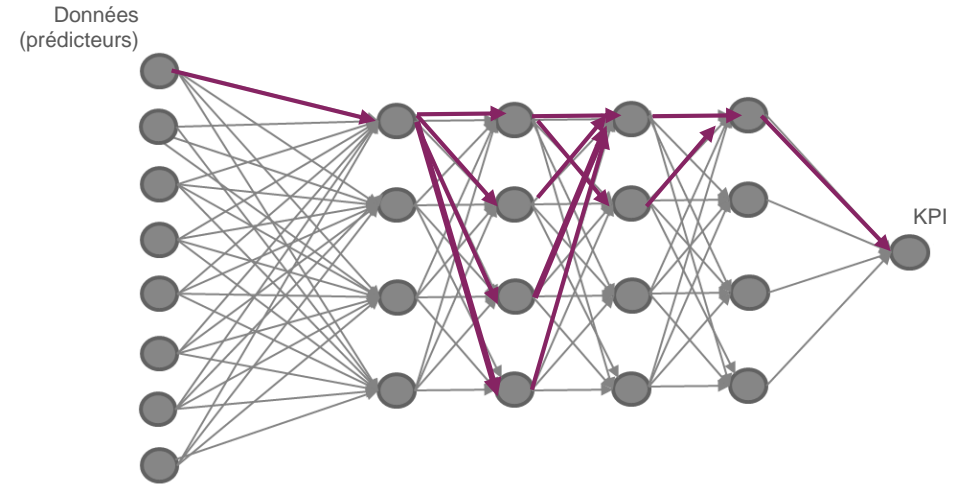
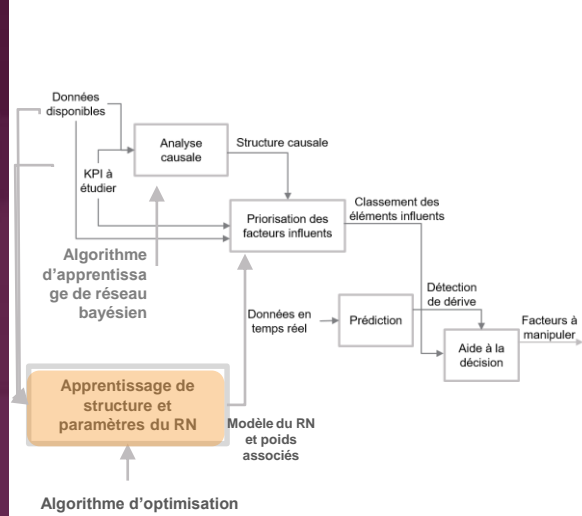
Contexte général et problématique

Etat de l'art et proposition

Cas d'étude

Conclusion et perspectives

# PRIORISATION DES ÉLÉMENTS INFLUENTS

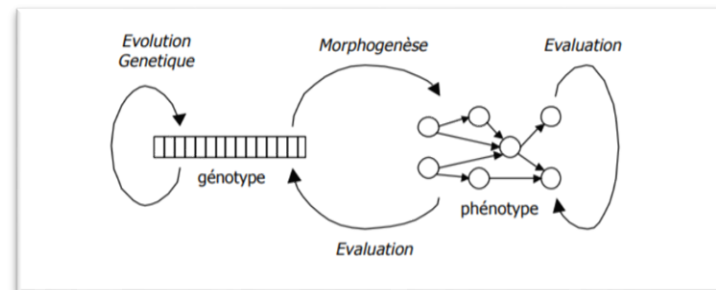
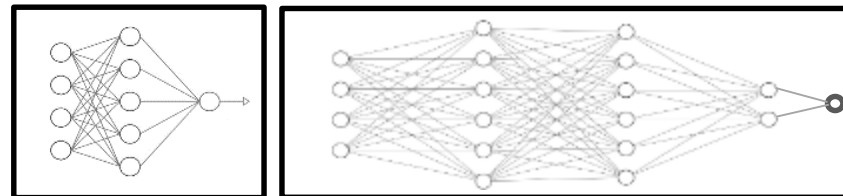


Contexte général et problématique

Etat de l'art et proposition

Cas d'étude

Conclusion et perspectives



Algorithme génétique

Généricité

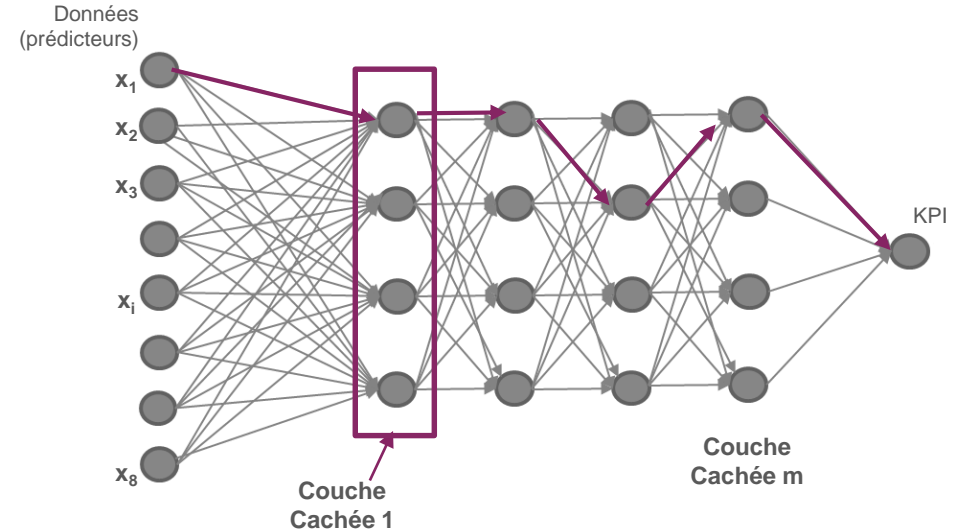
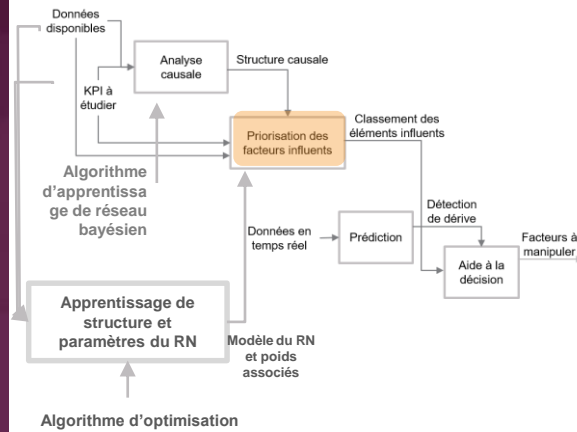
Classer les éléments influents en se basant sur les poids finaux d'un réseau de neurones avec un bon pouvoir de prédiction

Trouver la structure et les paramètres RN

- Approche expérimentale composée d'une **série d'affinages successifs** des paramètres pour avoir une prédiction optimale
- ⇒ **Pour chaque KPI, il faut trouver la structure permettant la meilleure prédiction**

Neuroévolution : Algorithme d'optimisation permettant d'automatiser le processus de construction des réseaux de neurones.

# PRIORISATION DES ÉLÉMENTS INFLUENTS : CLASSEMENT



- Les poids issus d'un réseau de neurones ayant un bon pouvoir de prédiction **représentent la force des connexions entre les unités**
- Ils mettent en avant **les degrés d'importance** des valeurs des entrées

## Règle de mise à jour des poids (avec descente de gradient)

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \lambda e_i^{(l)} x_j^{(l-1)}$$

où  $\lambda$  représente le taux d'apprentissage (de faible magnitude et compris entre 0,0 et 1,0).

- D'un modèle à l'autre, les poids changent, en fonction des données sélectionnées en entrées
- **L'influence dépend donc des poids finaux et de la variabilité des données d'entrée**

Classement de l'élément  $i$

Moyenne des valeurs de  $i$

Nb de chemins possibles de  $i$  à la sortie

Nb de couches cachées

Poids liant la couche  $k$  à  $k+1$  sur le chemin  $j$

$$Cl_i = \frac{\sum_{j=0}^n \sum_{k=0}^m \mu_i \cdot |w_{kj \rightarrow kj+1}|}{\sigma_i \cdot m..}$$

Ecart-type des valeurs de  $i$

$$CL = [Cl_1, Cl_2, \dots, Cl_i, \dots, Cl_8]$$

Classement des éléments influents par ordre décroissant

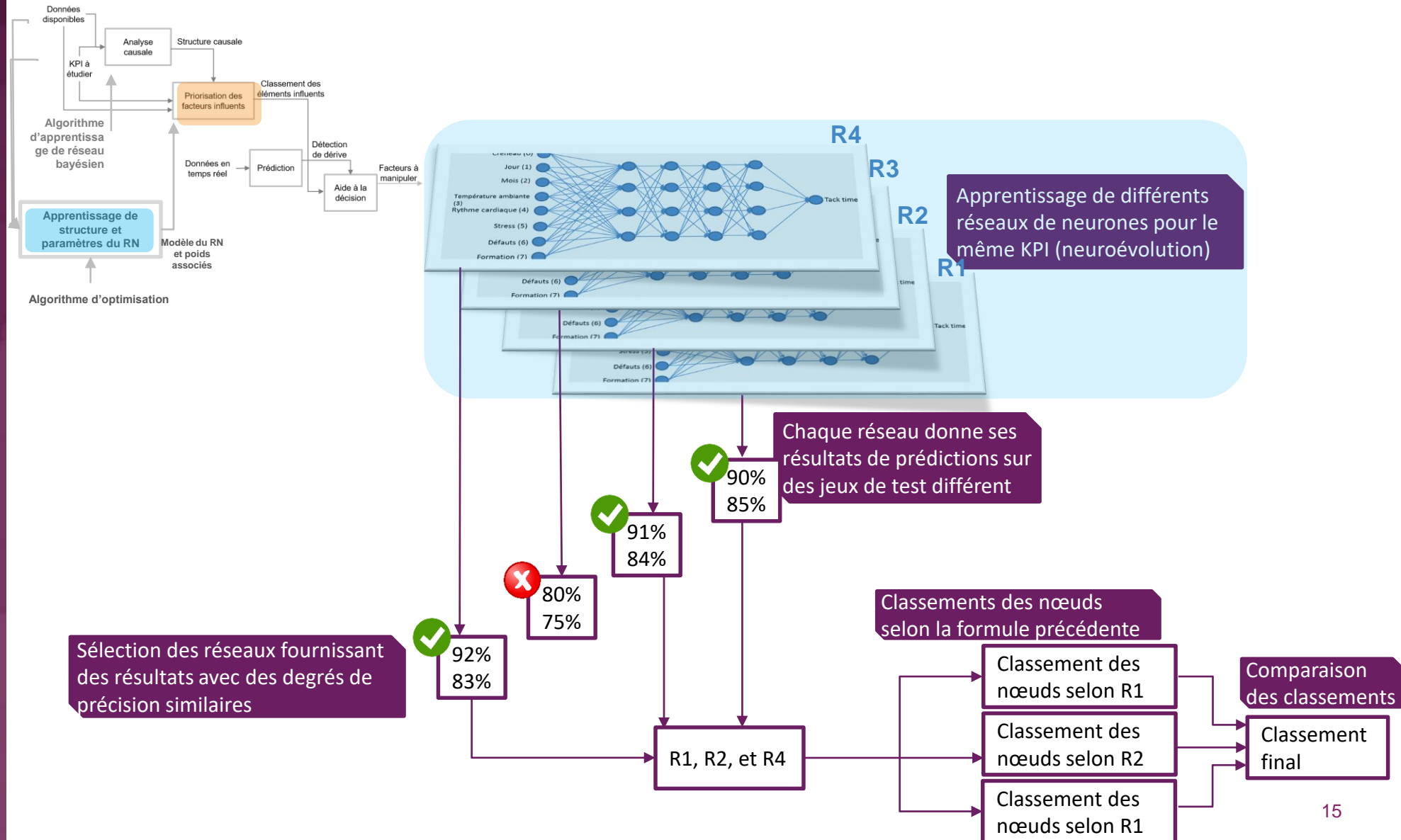
Contexte général et problématique

Etat de l'art et proposition

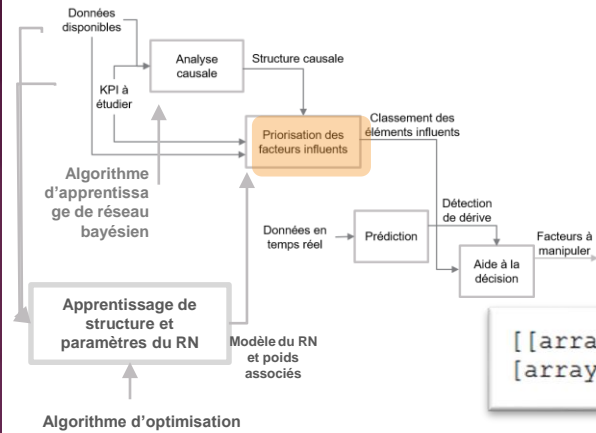
Cas d'étude

Conclusion et perspectives

# PRIORISATION DES ÉLÉMENTS INFLUENTS : VALIDATION DU CLASSEMENT



# PRIORISATION DES ÉLÉMENTS INFLUENTS : VALIDATION DU CLASSEMENT



```
[[array([2021.85]), 4], [array([1967.68]), 0], [array([1934.16]), 3], [array([1861.07]), 1],
[array([1852.13]), 2], [array([1839.75]), 5], [array([1779.06]), 6], [array([1724.95]), 7]]
```

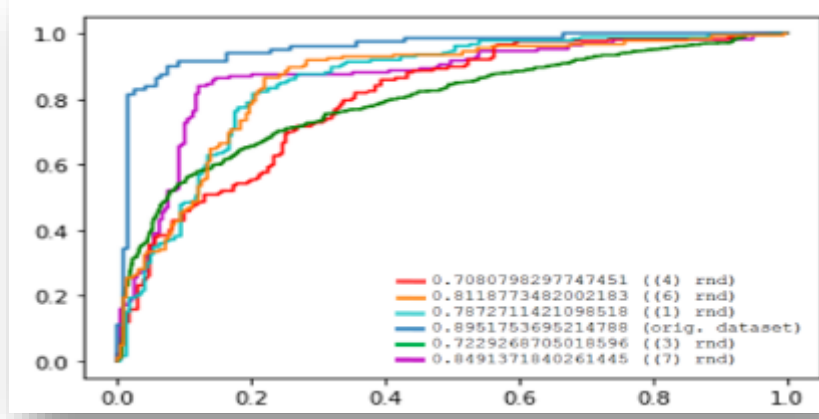
Contexte général et problématique

Etat de l'art et proposition

Cas d'étude

Conclusion et perspectives

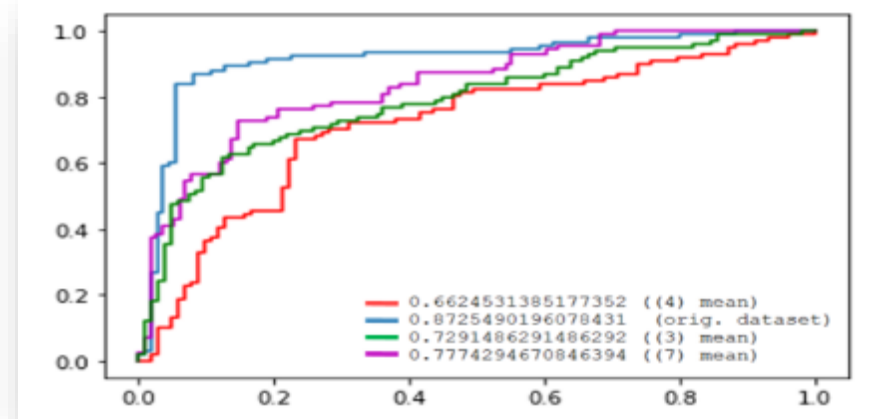
Sensibilité (Taux de vrais positifs)



1-spécificité (Taux de faux positifs)

Courbes ROC de prediction après remplacement de chaque variable par des valeurs aléatoires de son intervalle de variation.

Sensibilité



1-spécificité

Courbes ROC de prediction après remplacement de chaque variable par sa moyenne.



# PRÉDICTION ET AIDE À LA DÉCISION

- Il faut **prévoir les dérives** et les détecter en temps réel.
- **Identifier les liens de causalité** entre les KPI et les données, de façon **exhaustive et objective**.
- **Prioriser causes identifiées** pour agir efficacement.
- Il faut que l'approche soit **générique**

## Caractéristiques de la causalité :

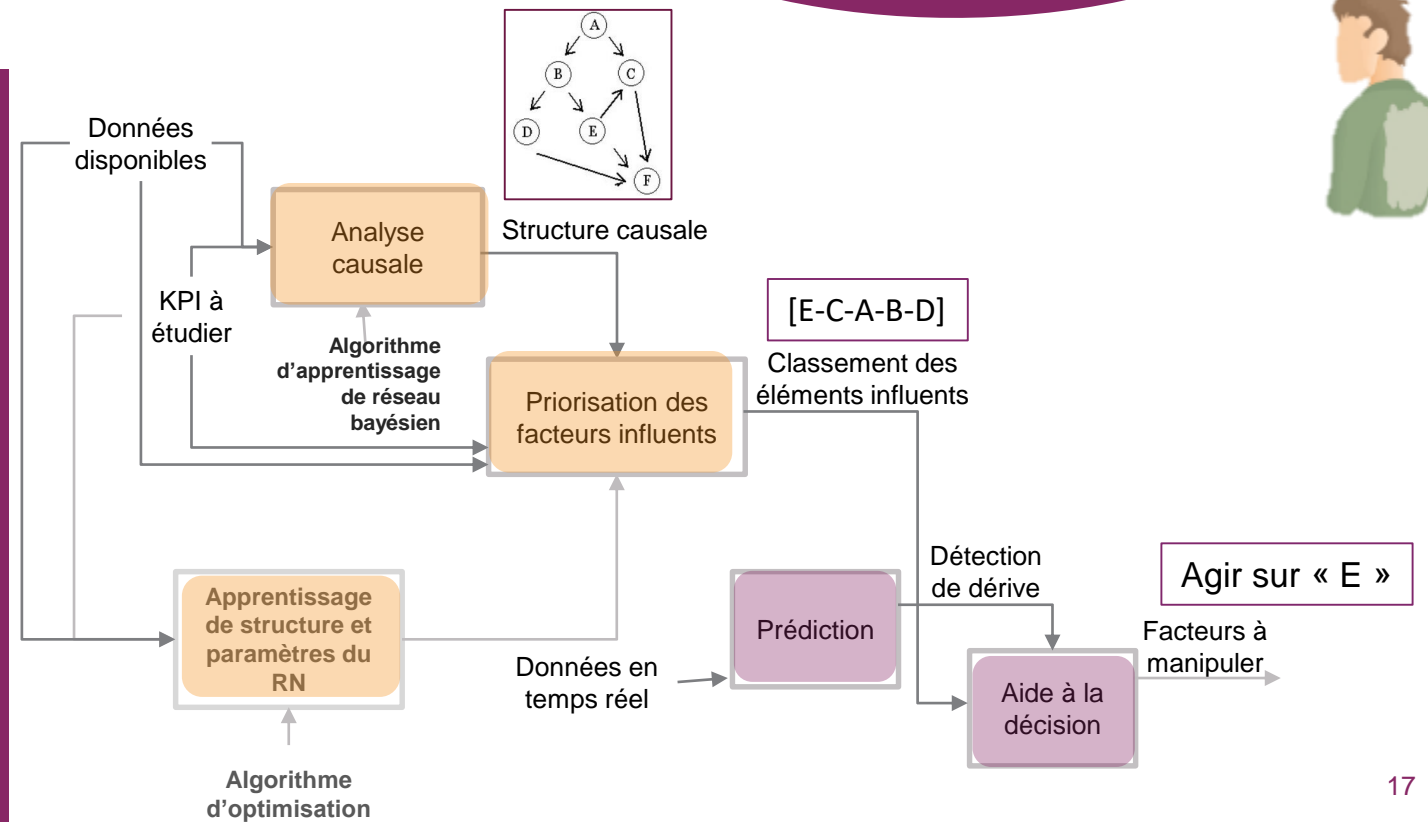
- ✓ - Aspect probabiliste ;
- ✓ - Ordre temporel ;
- ✓ - Dépendance de l'espace de recherche
- ✓ - Représentation graphique;
- ✓ - Taille de l'effet ;
- ✓ - Stabilité ;
- ✓ - Plausibilité ;
- ✓ - Vérificabilité;

Contexte général et problématique

Etat de l'art et proposition

Cas d'étude

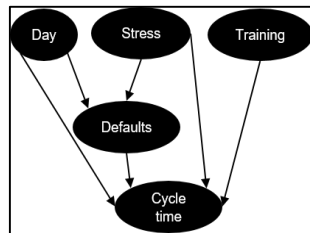
Conclusion et perspectives



# VÉRIFICATION PAR CAS D'ÉTUDE

- Données de synthèse avec connaissance préalable des liens de causalités

- Classement et validation du classement

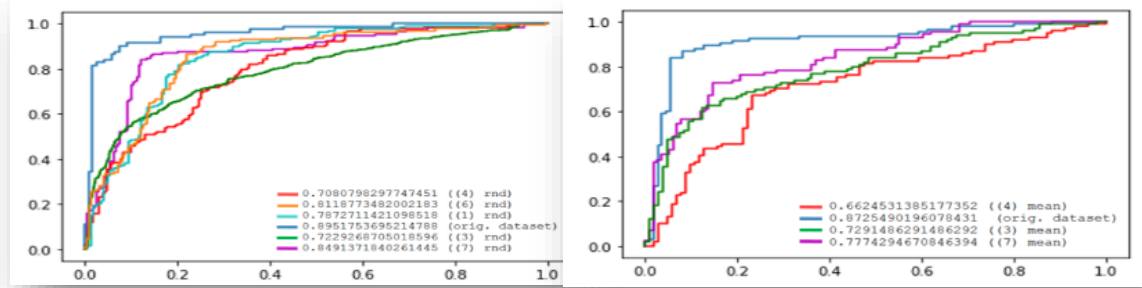


```

[[array([2021.85]), 4], [array([1967.68]), 0], [array([1934.16]), 3], [array([1861.07]), 1],
[array([1852.13]), 2], [array([1839.75]), 5], [array([1779.06]), 6], [array([1724.95]), 7]]

[[array([25390.9]), 4], [array([22044.01]), 0], [array([21106.]), 3], [array([16625.73]), 1],
[rray([14845.48]), 2], [array([10886.93]), 5], [array([5121.37]), 6], [array([4620.99]), 7]]
  
```

Sensibilité



1-spécificité

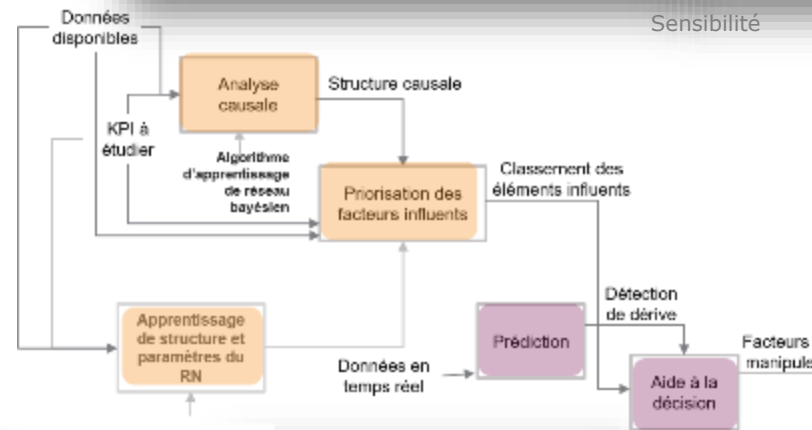
- Apprentissage de structure

Contexte général et problématique

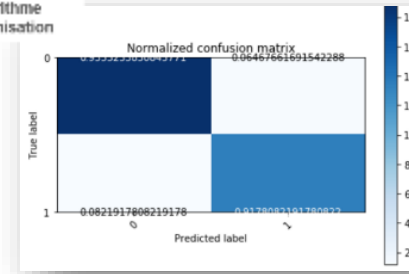
Etat de l'art et proposition

Cas d'étude

Conclusion et perspectives



*GEN00000000000000000000	0.46568627450988393*
*GEN00000000000000000001	0.5441176470588235*
*GEN00000000000000000002	0.5784313725490197*
*GEN00000000000000000003	0.803921568627451*
*GEN00000000000000000004	0.5441176470588235*
*GEN00000000000000000005	0.6274509803921569*
*GEN00000000000000000006	0.5245098039215687*
GEN27-5620.03630900383	0.9264705882352942
GEN22-5253.9048981666565	0.9215686274509803
GEN13-4681.029283761978	0.9166666666666666
GEN21-5184.305323123932	0.9166666666666666
GEN2-2336.9909529685974	0.9068627450980392
GEN11-4577.829998970032	0.9068627450980392
GEN20-5163.891398906708	0.9068627450980392



- Prédiction

Normalized confusion matrix	
[[0.94 0.06]	
[0.08 0.92]]	

# CONCLUSION ET PERSPECTIVES

Contexte général et problématique

Etat de l'art et proposition

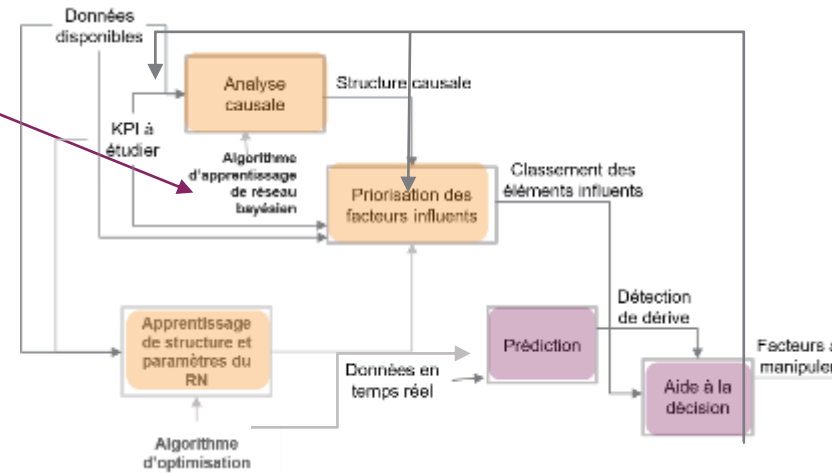
Cas d'étude

Conclusion et perspectives

Automatisation de sélection du réseau bayésien à retenir parmi plusieurs réseaux équivalents

Mise en lien des différentes briques

Application sur cas d'étude réel (En cours sur l'usine école à Aix, et la plateforme « Usine du futur » au campus de Lille)



MERCI  
😊