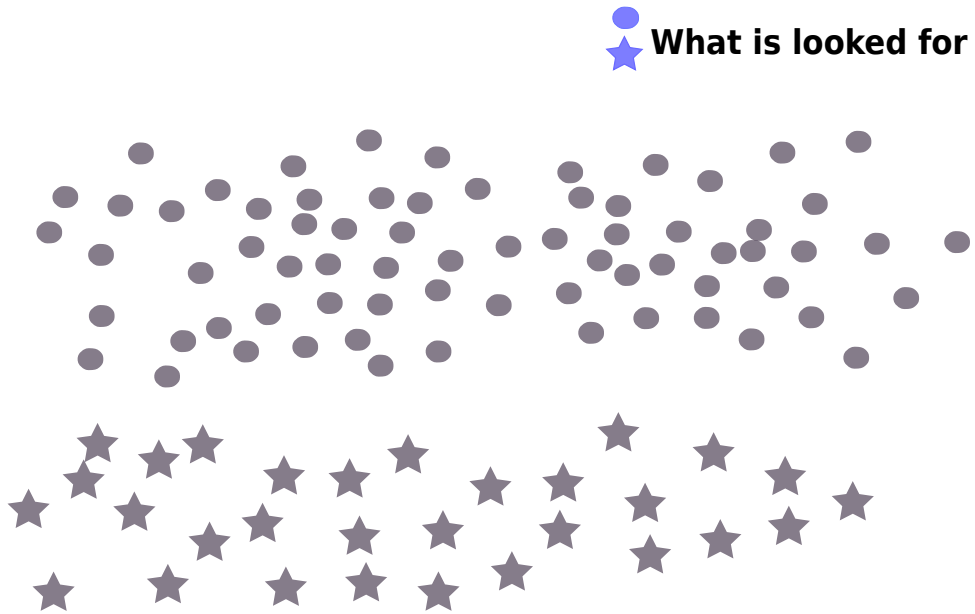
Algorithmes d'apprentissage et modèles prédictifs

Fabien Tarissan

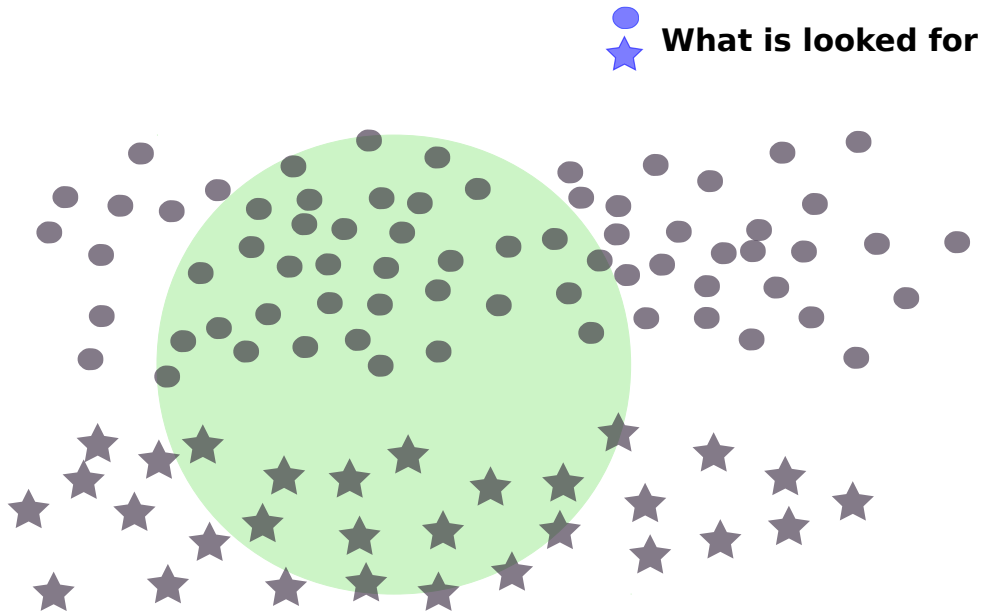
CNRS – ENS Paris Saclay

ENUM

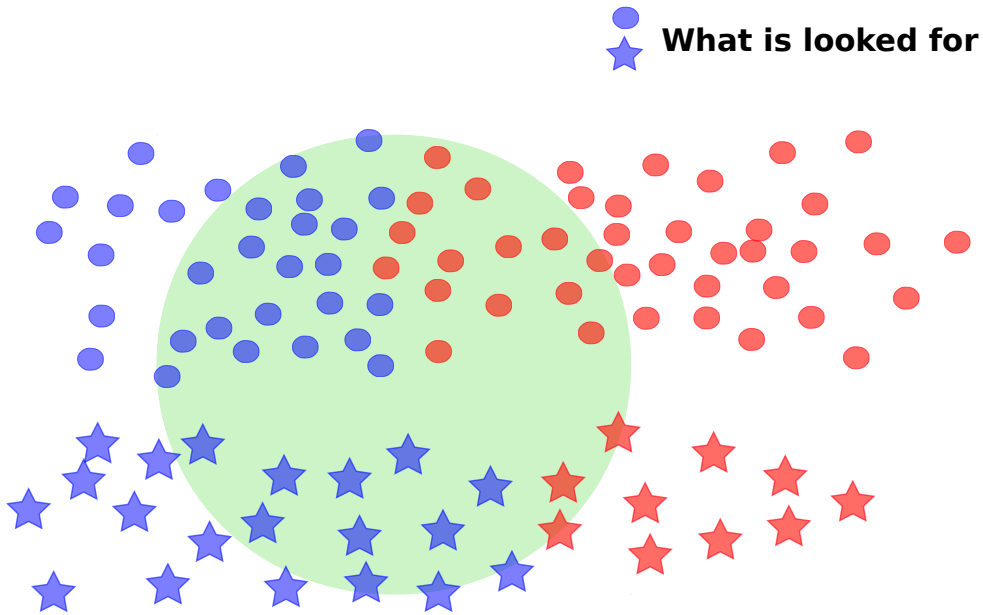
Avant de commencer



Avant de commencer

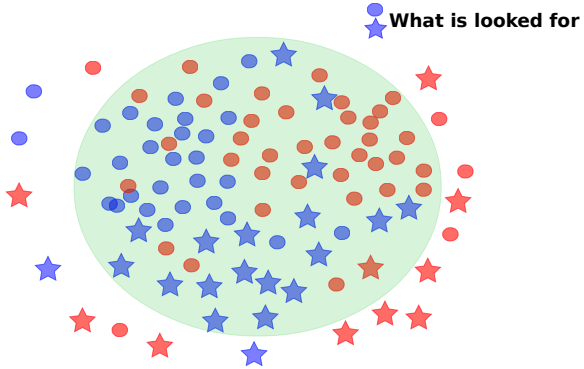


Avant de commencer

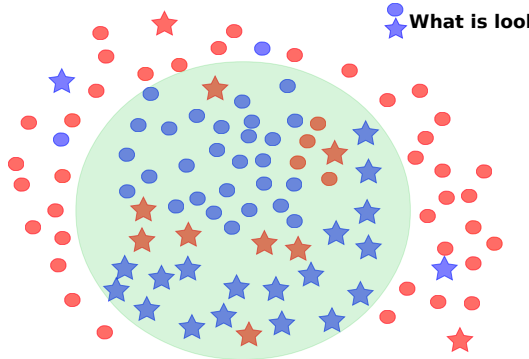


Avant de commencer

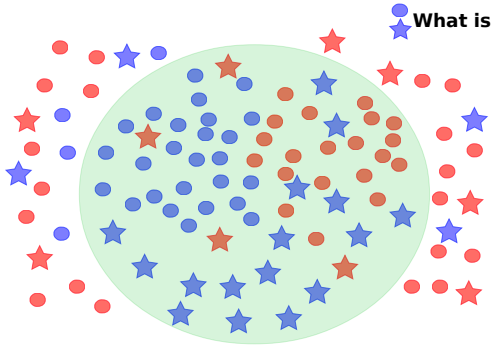
A



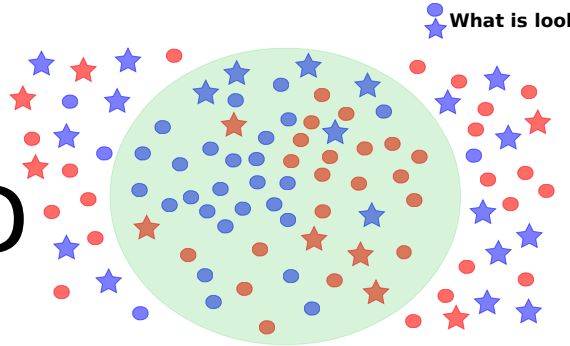
B



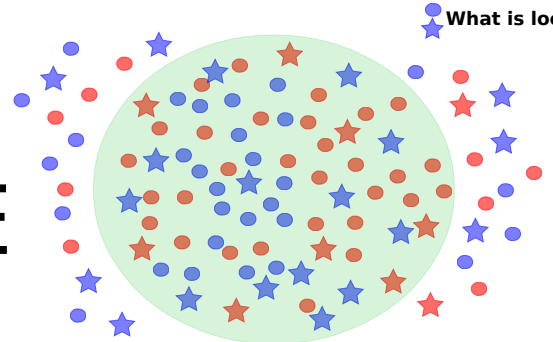
C



D



E



Algorithmes d'apprentissage

Plan

Qu'est-ce qu'un algorithme ?

Qu'est-ce qu'un algorithme ?

Qu'est-ce qu'un algorithme ?

Qu'est-ce qu'un algorithme ?

Un algorithme est séquence d'opérations :

- simples
- non ambiguës
- qui résout un problème donné

Exemples

- une recette de cuisine (comment faire un gâteau au chocolat ?)
- un itinéraire (comment se rendre à l'ENS depuis son domicile ?)
- algorithme de Dijkstra (comment trouver un plus court chemin entre 2 points ?)

Qu'est-ce qu'un algorithme ?

Qu'est-ce qu'un algorithme ?

Un algorithme est séquence d'opérations :

- simples
- non ambiguës
- qui résout un problème donné

Exemples

- une recette de cuisine (comment faire un gâteau au chocolat ?)
- un itinéraire (comment se rendre à l'ENS depuis son domicile ?)
- algorithme de Dijkstra (comment trouver un plus court chemin entre 2 points ?)

En informatique, on s'intéresse en général à des algorithmes travaillant sur des symboles (caractères, chiffres, ...)

Qu'est-ce qu'un algorithme ?

Qu'est-ce qu'un algorithme ?

Un algorithme est séquence d'opérations :

- simples
- non ambiguës
- qui résout un problème donné

Exemples

- une recette de cuisine (comment faire un gâteau au chocolat ?)
- un itinéraire (comment se rendre à l'ENS depuis son domicile ?)
- algorithme de Dijkstra (comment trouver un plus court chemin entre 2 points ?)

En informatique, on s'intéresse en général à des algorithmes travaillant sur des symboles (caractères, chiffres, ...), qui résolvent des problèmes génériques

Qu'est-ce qu'un algorithme ?

Qu'est-ce qu'un algorithme ?

Un algorithme est séquence d'opérations :

- simples
- non ambiguës
- qui résout un problème donné

Exemples

- une recette de cuisine (comment faire un gâteau au chocolat ?)
- un itinéraire (comment se rendre à l'ENS depuis son domicile ?)
- algorithme de Dijkstra (comment trouver un plus court chemin entre 2 points ?)

En informatique, on s'intéresse en général à des algorithmes travaillant sur des symboles (caractères, chiffres, ...), qui résolvent des problèmes génériques et qui sont exécutables par des machines.

Comment étudier un algorithme ?

Résoudre un problème (pour un.e informaticien.ne)
c'est proposer une méthode générique
qui fournit la bonne réponse à chaque instance du problème

Analyser un algorithme, c'est donc :

- comprendre quel problème il cherche à résoudre
- savoir comment il procède
- identifier les avantages / les dangers / les biais dans sa mise en application

À chaque type de problème ...

... correspond une famille d'algorithmes (un paradigme)

Comment étudier un algorithme ?

Résoudre un problème (pour un.e informaticien.ne)
c'est proposer une méthode générique
qui fournit la bonne réponse à chaque instance du problème

Analyser un algorithme, c'est donc :

- comprendre quel problème il cherche à résoudre
- savoir comment il procède
- identifier les avantages / les dangers / les biais dans sa mise en application

À chaque type de problème ...

- Faire la somme de deux nombres ?

... correspond une famille d'algorithmes (un paradigme)

- Procédure explicite ? $\Longrightarrow$ Description séquentielle

Comment étudier un algorithme ?

Résoudre un problème (pour un.e informaticien.ne)
c'est proposer une méthode générique
qui fournit la bonne réponse à chaque instance du problème

Analyser un algorithme, c'est donc :

- comprendre quel problème il cherche à résoudre
- savoir comment il procède
- identifier les avantages / les dangers / les biais dans sa mise en application

À chaque type de problème ...

- Faire la somme de deux nombres ?
- ...

... correspond une famille d'algorithmes (un paradigme)

- Procédure explicite ? $\Longrightarrow$ Description séquentielle
- ... $\Longrightarrow$ Algorithmes "Diviser pour Régner", "Glouton", "Probabiliste", "Programmation Dynamique", ...

Comment étudier un algorithme ?

Résoudre un problème (pour un.e informaticien.ne)
c'est proposer une méthode générique
qui fournit la bonne réponse à chaque instance du problème

Analyser un algorithme, c'est donc :

- comprendre quel problème il cherche à résoudre
- savoir comment il procède
- identifier les avantages / les dangers / les biais dans sa mise en application

À chaque type de problème ...

- Faire la somme de deux nombres ?
- ...
- Étiqueter une image ? Proposer un montant compensatoire? Estimer un risque?

... correspond une famille d'algorithmes (un paradigme)

- Procédure explicite ? $\Longrightarrow$ Description séquentielle
- ... $\Longrightarrow$ Algorithmes "Diviser pour Régner", "Glouton", "Probabiliste", "Programmation Dynamique", ...

Comment étudier un algorithme ?

Résoudre un problème (pour un.e informaticien.ne)
c'est proposer une méthode générique
qui fournit la bonne réponse à chaque instance du problème

Analyser un algorithme, c'est donc :

- comprendre quel problème il cherche à résoudre
- savoir comment il procède
- identifier les avantages / les dangers / les biais dans sa mise en application

À chaque type de problème ...

- Faire la somme de deux nombres ?
- ...
- Étiqueter une image ? Proposer un montant compensatoire? Estimer un risque?

... correspond une famille d'algorithmes (un paradigme)

- Procédure explicite ? $\Longrightarrow$ Description séquentielle
- ... $\Longrightarrow$ Algorithmes "Diviser pour Régner", "Glouton", "Probabiliste", "Programmation Dynamique", ...
- Exemples de solutions ? $\Longrightarrow$ Algorithmes d'apprentissage (supervisé)

Éthique et algorithmes

Observation : les algorithmes ont pénétré énormément de systèmes qui ont un impact sur nos sociétés.

1. Donne lieu à un débat (passionnel) entre :
   - les promoteurs : efficacité, suppression des biais humains, etc ...
   - les opposants : impacte l'économie, l'emploi, introduit de nouveaux biais (discriminations), impact imprévisible sur le long terme, ...
2. Lié à différentes problématiques en terme d'éthique
   - transparence/confiance : système APB/ParcourSup, vote électronique...
   - responsabilité : véhicule autonomes, armes létales autonomes, ...
   - vie privée: traces de navigations, plateforme en lignes, ... (cf cours 3)
   - diversité: algorithmes de classement, recommandations, ... (cf cours 2)
   - équité/discrimination : prise de décision (prêt bancaire, décision juridique)
3. Chacune dépend de plusieurs aspects: but, contexte, utilisateurs, données, ...
   - offres d'emploi attractifs (via google ad) : proposés aux hommes
   - images renvoyées pour "C.E.O." (google) : seulement 11% de femmes
   - complétion de "transgenders are" : ...

Algorithmes d'apprentissage

Une grande diversité de méthodes

Supervisé

Non supervisé

Classification:

- K plus proches voisins
- Naïve bayésienne
- Réseaux de neurones
- Machine à vecteurs de support (SVM)
- Forêts aléatoires
- ...

Régression:

- Régression linéaire
- Régression polynomiale
- Arbres de décision
- Réseaux de neurones
- ...

Clustering:

- K-means
- Analyse en Composantes Principales (ACP)
- Apprentissage profond
- ...

Apprentissage supervisé

Un algorithme d'apprentissage supervisé (simplifié) dispose :

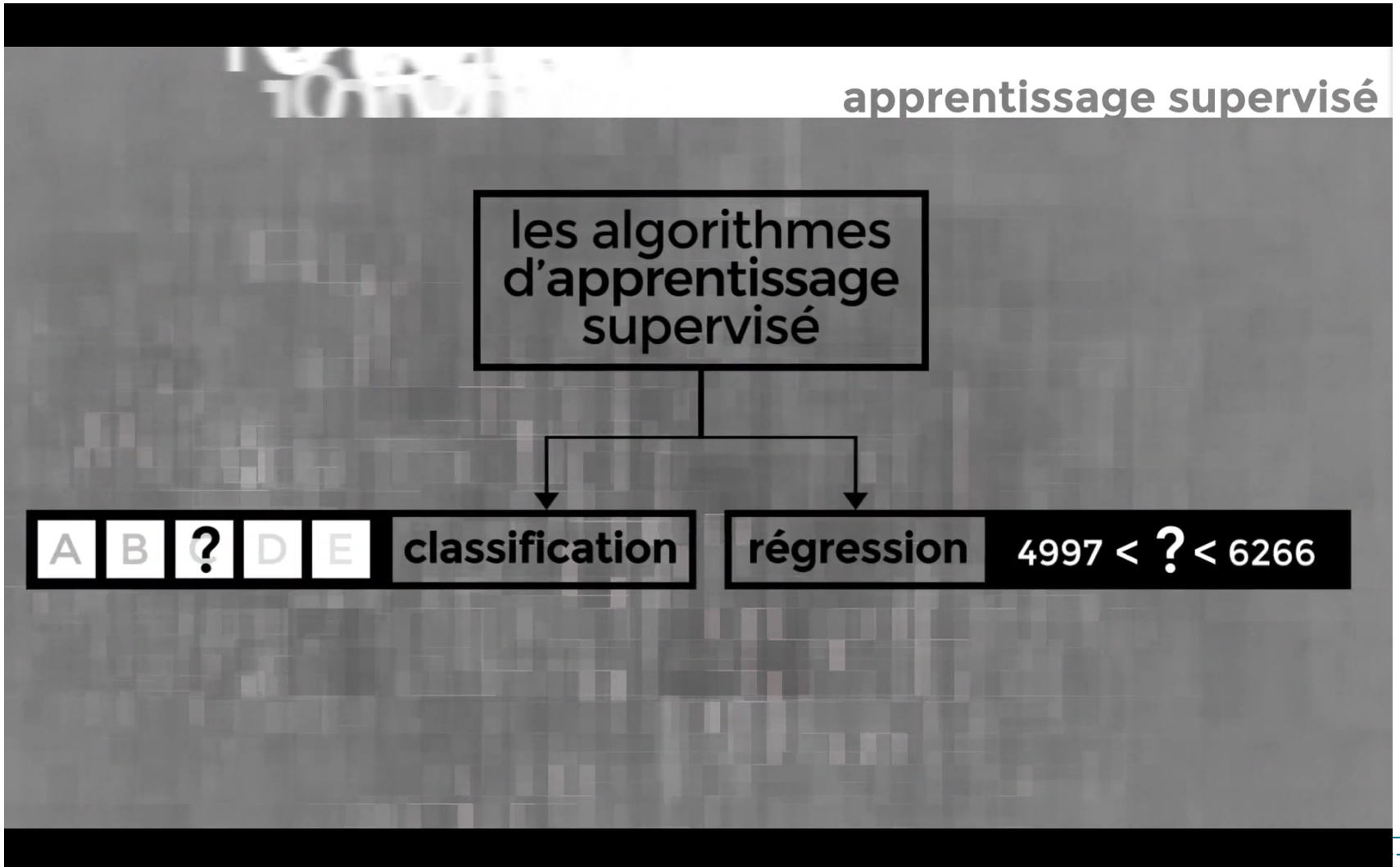
- d'une question (un critère)
- de données (observées) auxquelles poser cette question
- des réponses correctes pour ces données

Le but : avoir une procédure de décision qui associe le plus souvent possible la bonne réponse vis à vis de ces données.

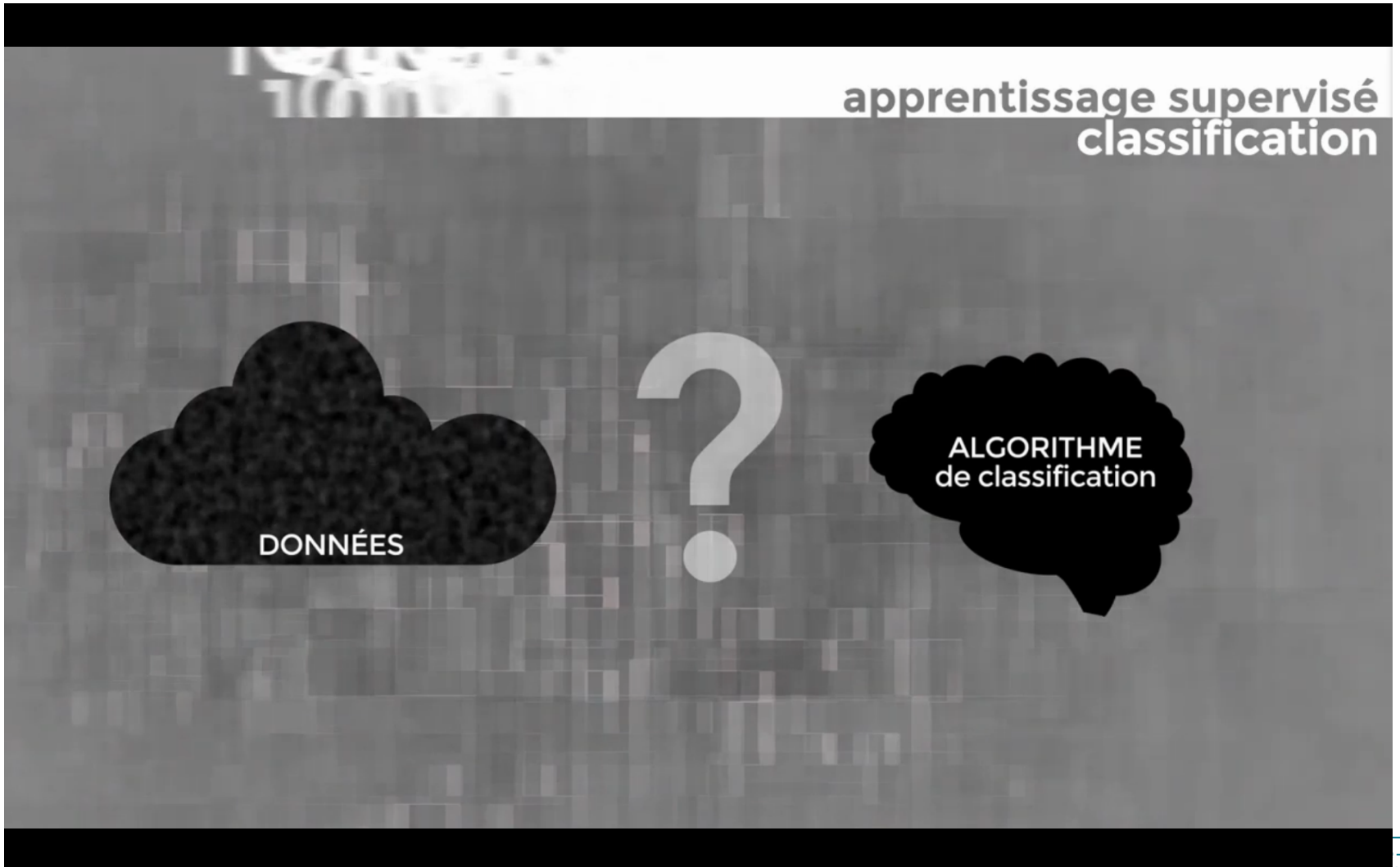
Exemple de problème à résoudre :

- étiqueter le contenu d'une image
- estimer le risque de récidive d'un prévenu
- estimer un montant compensatoire

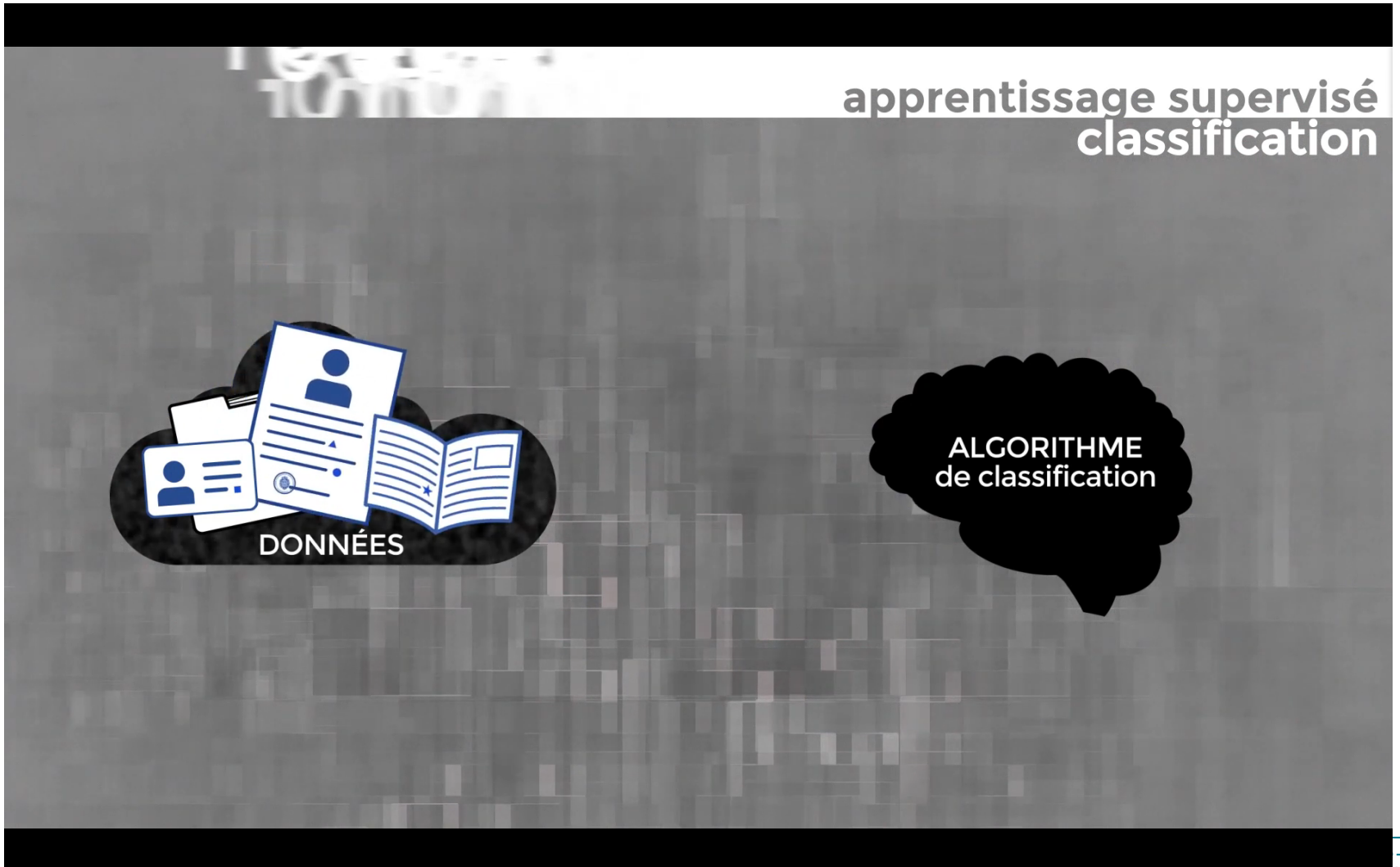
Apprentissage supervisé



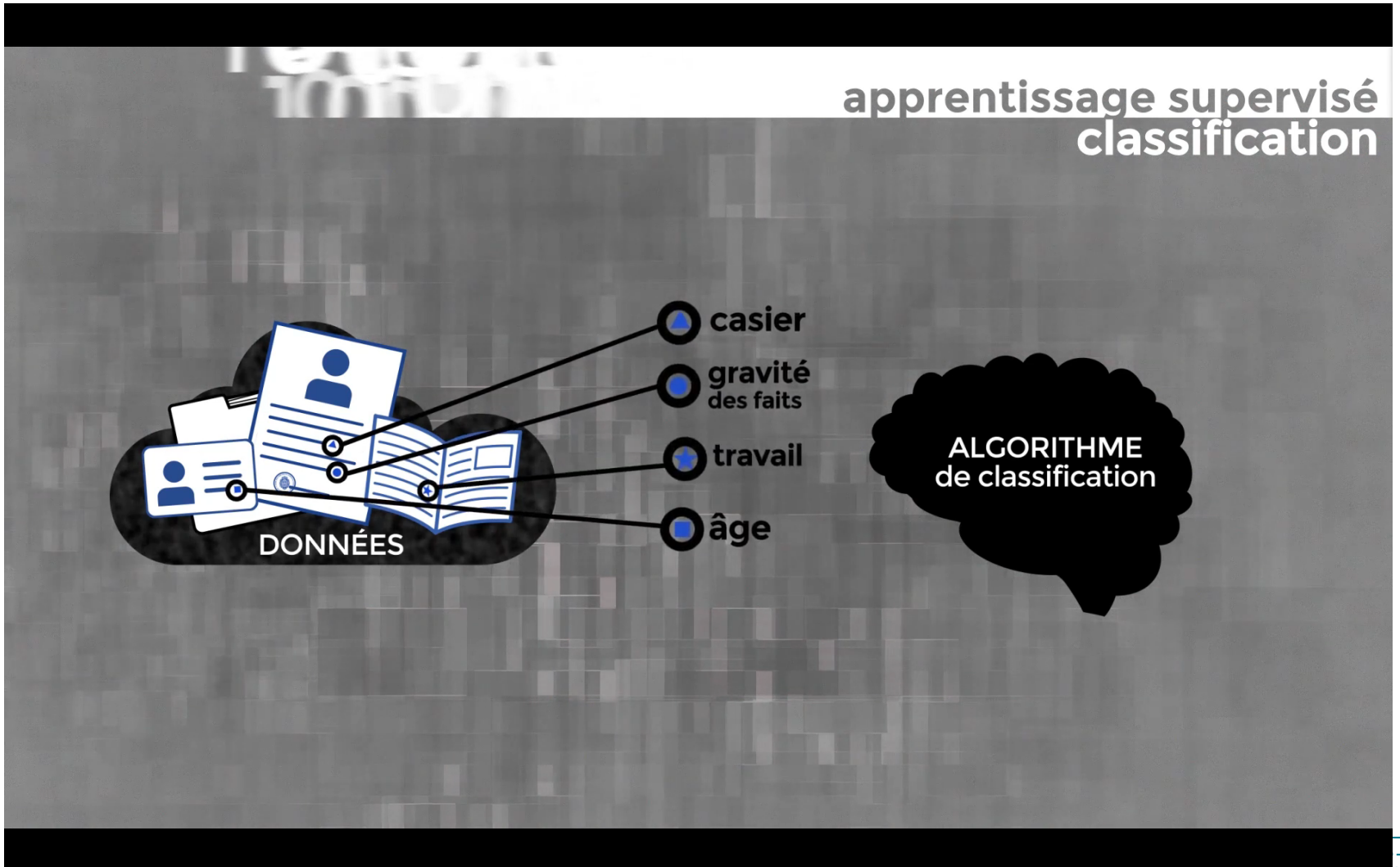
Classification



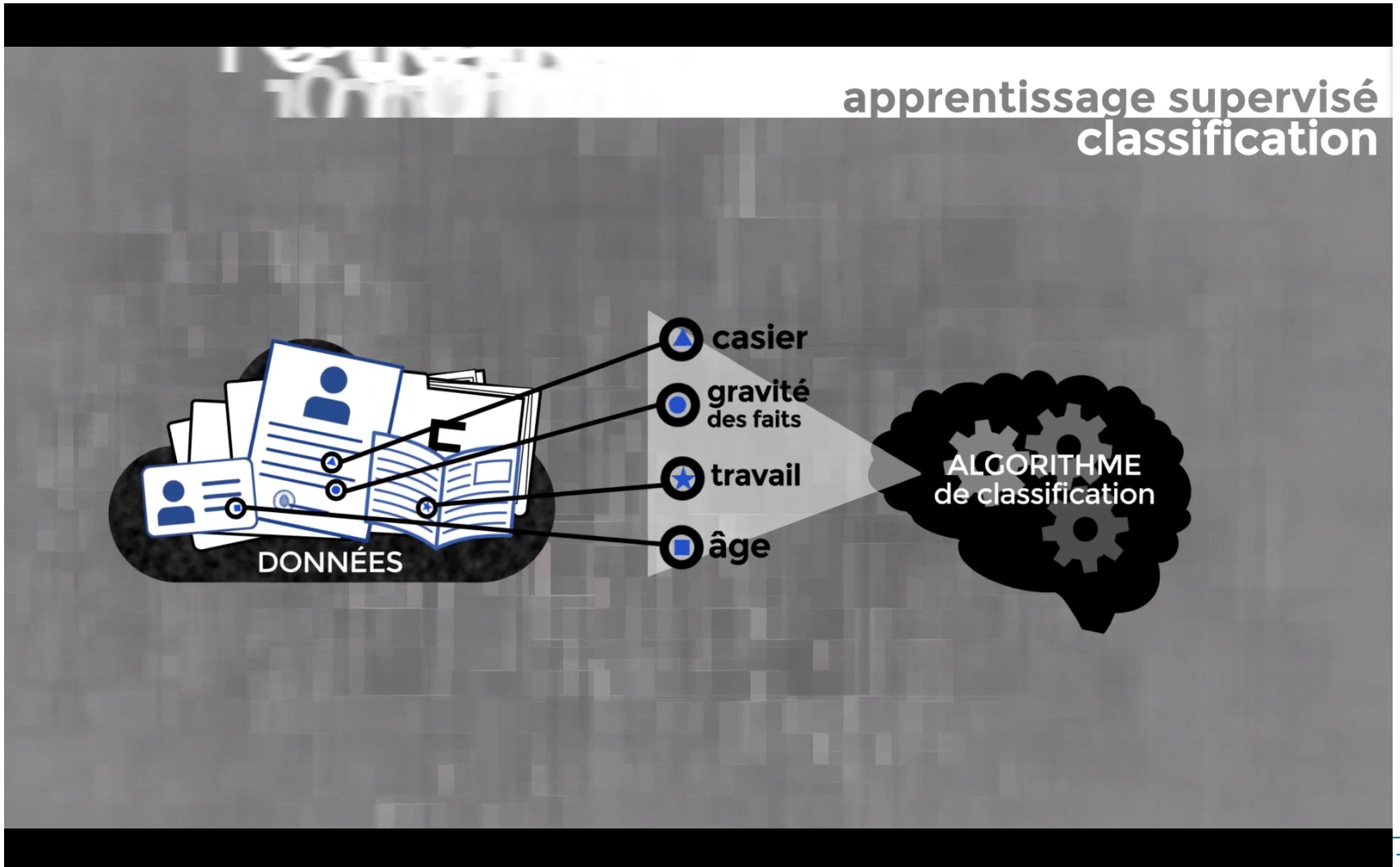
Classification



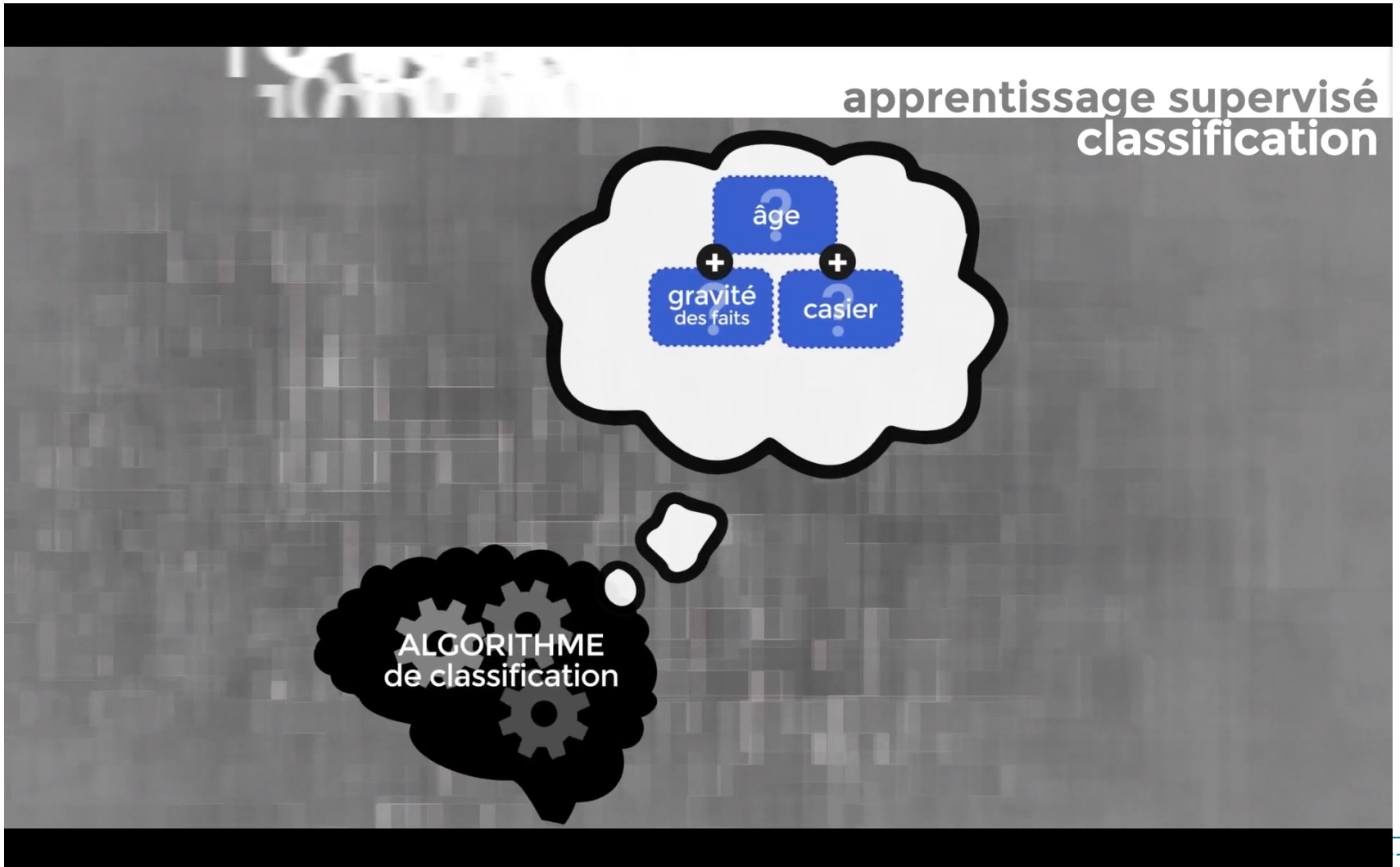
Classification



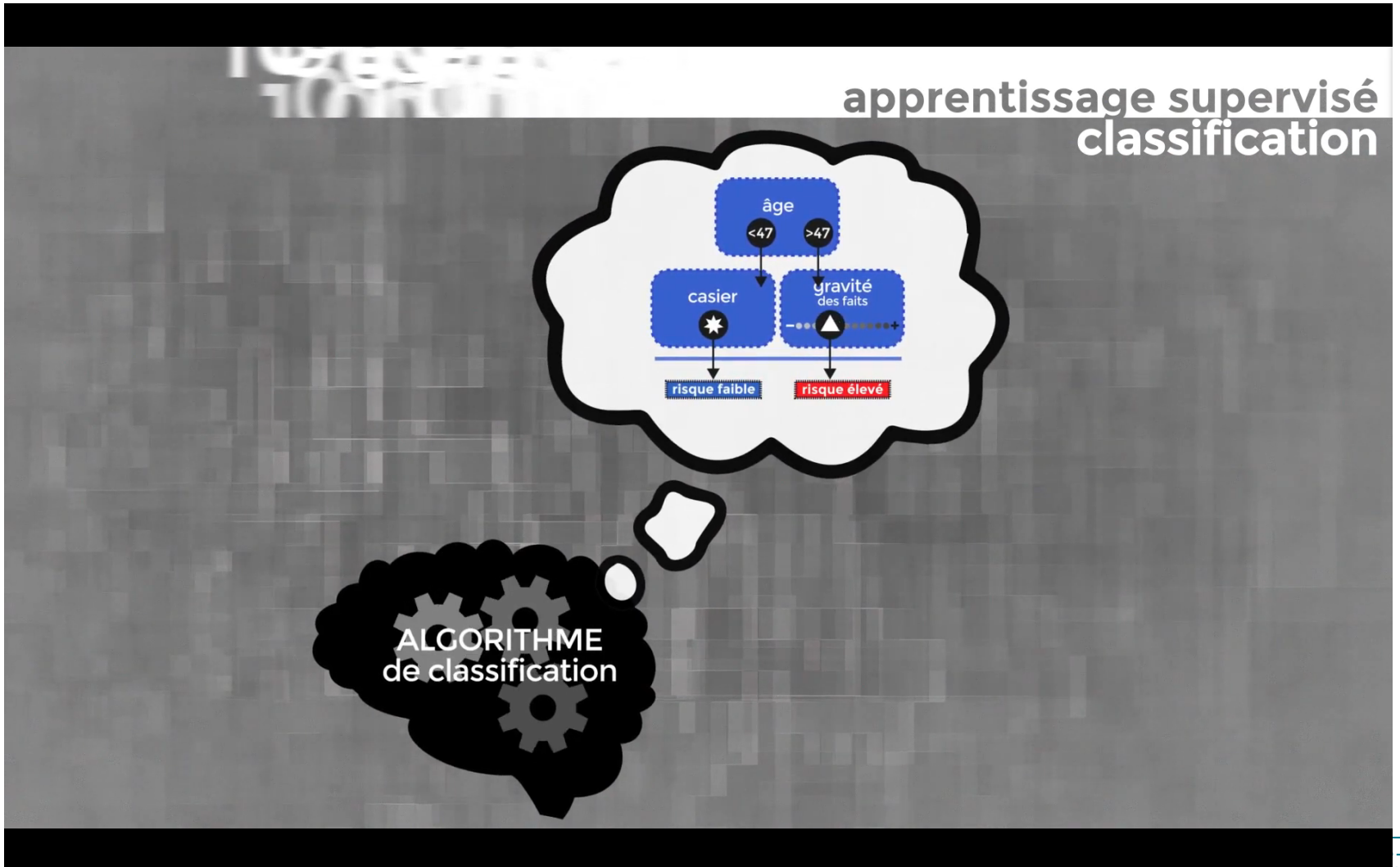
Classification



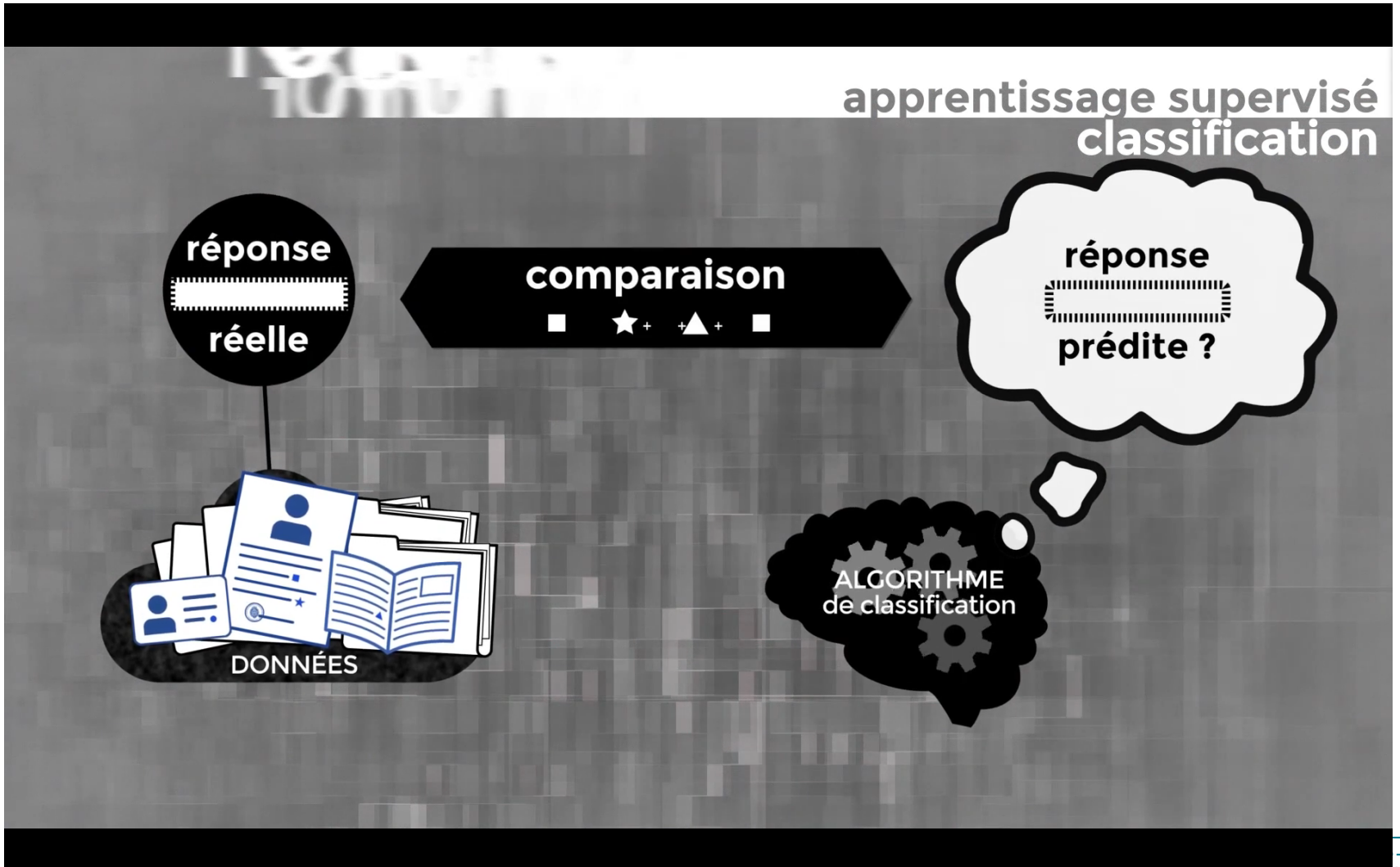
Classification



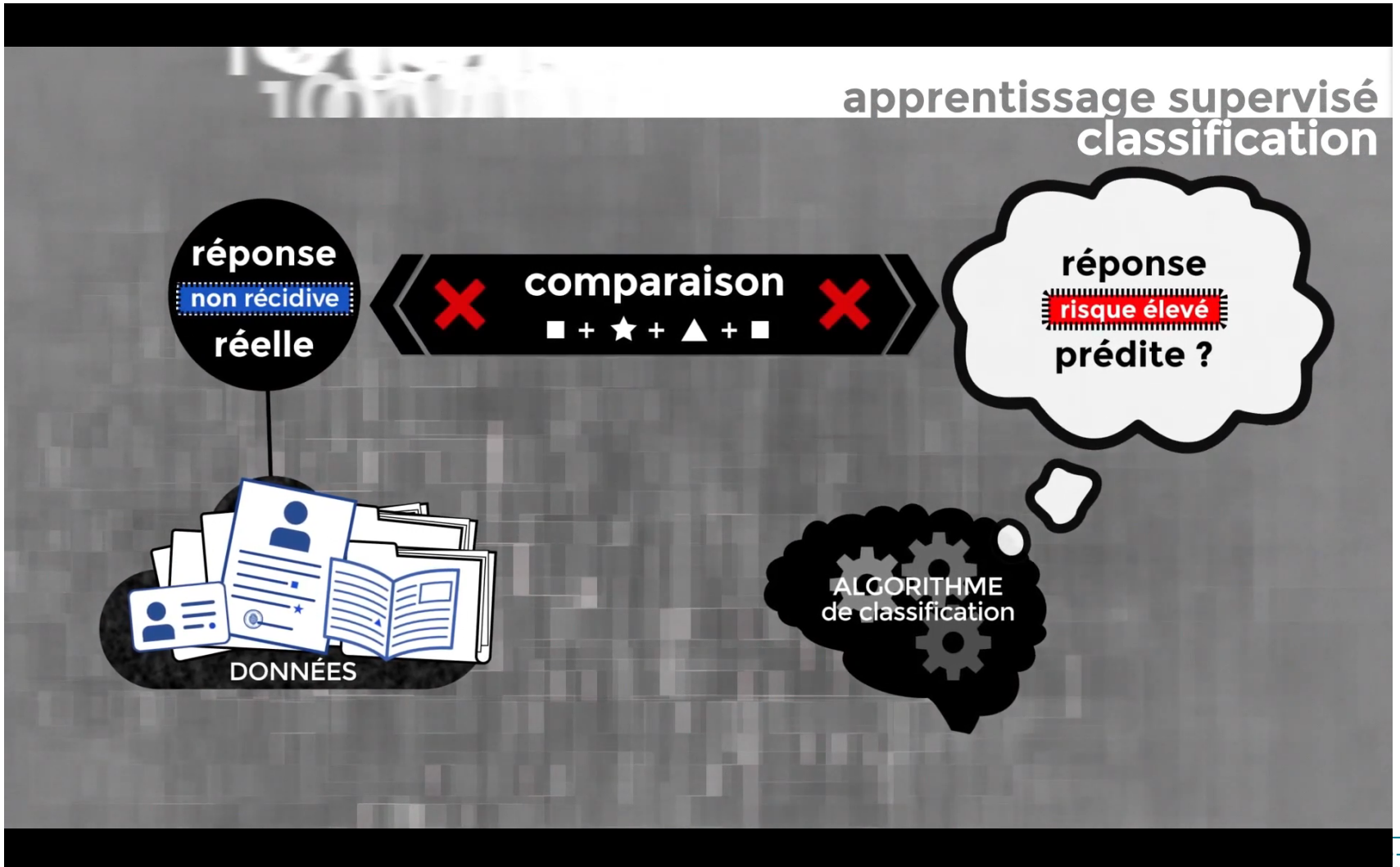
Classification



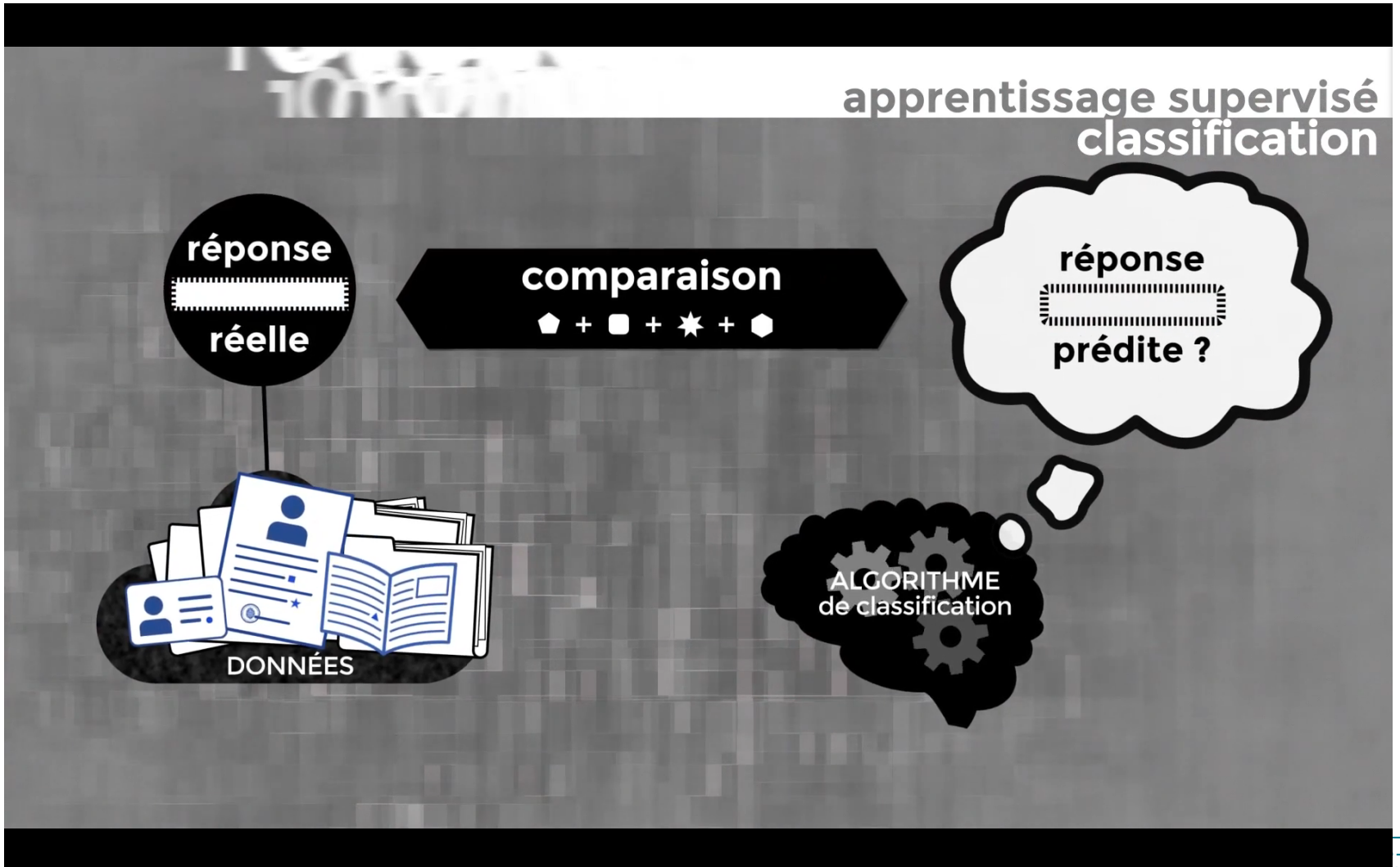
Classification



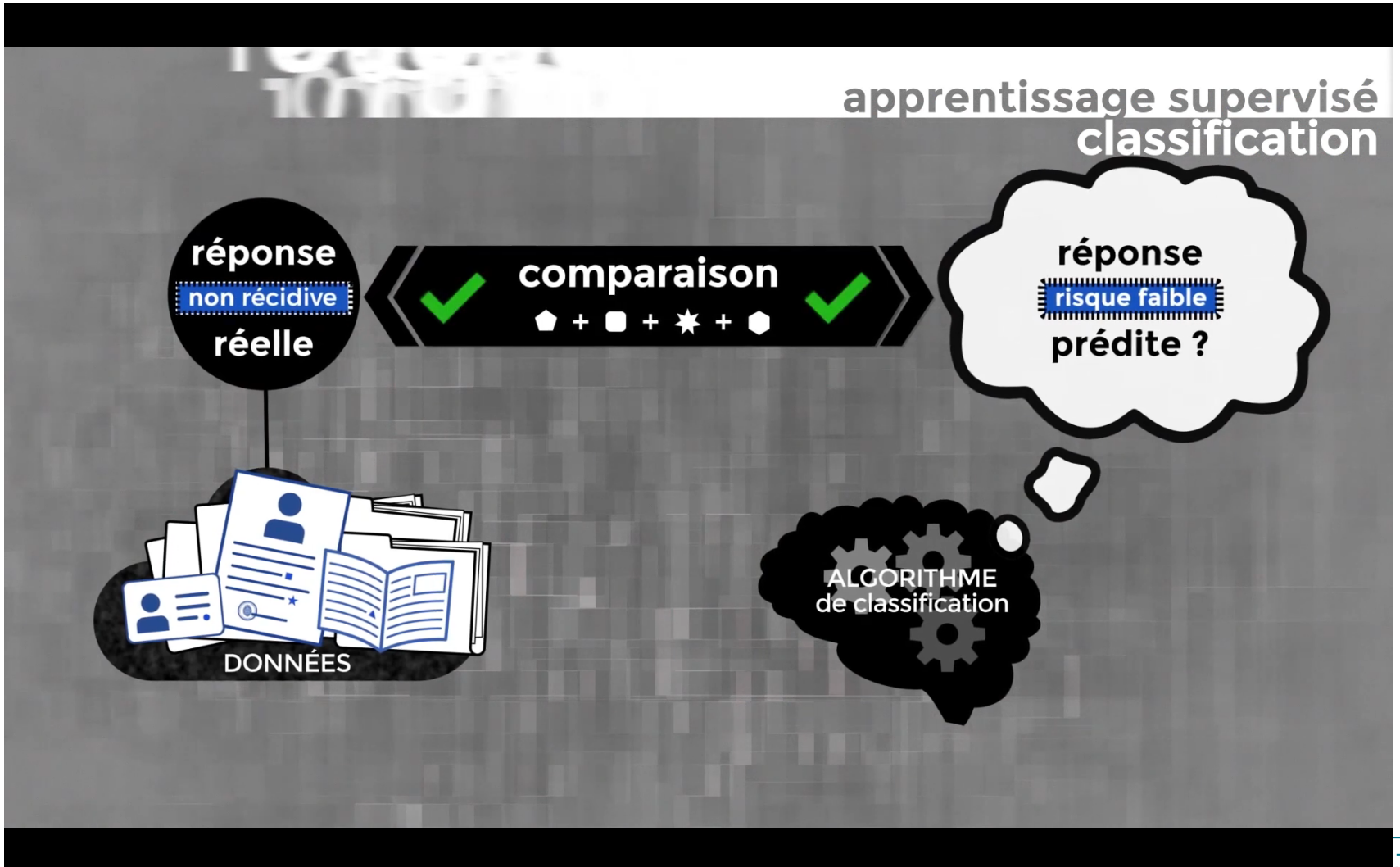
Classification



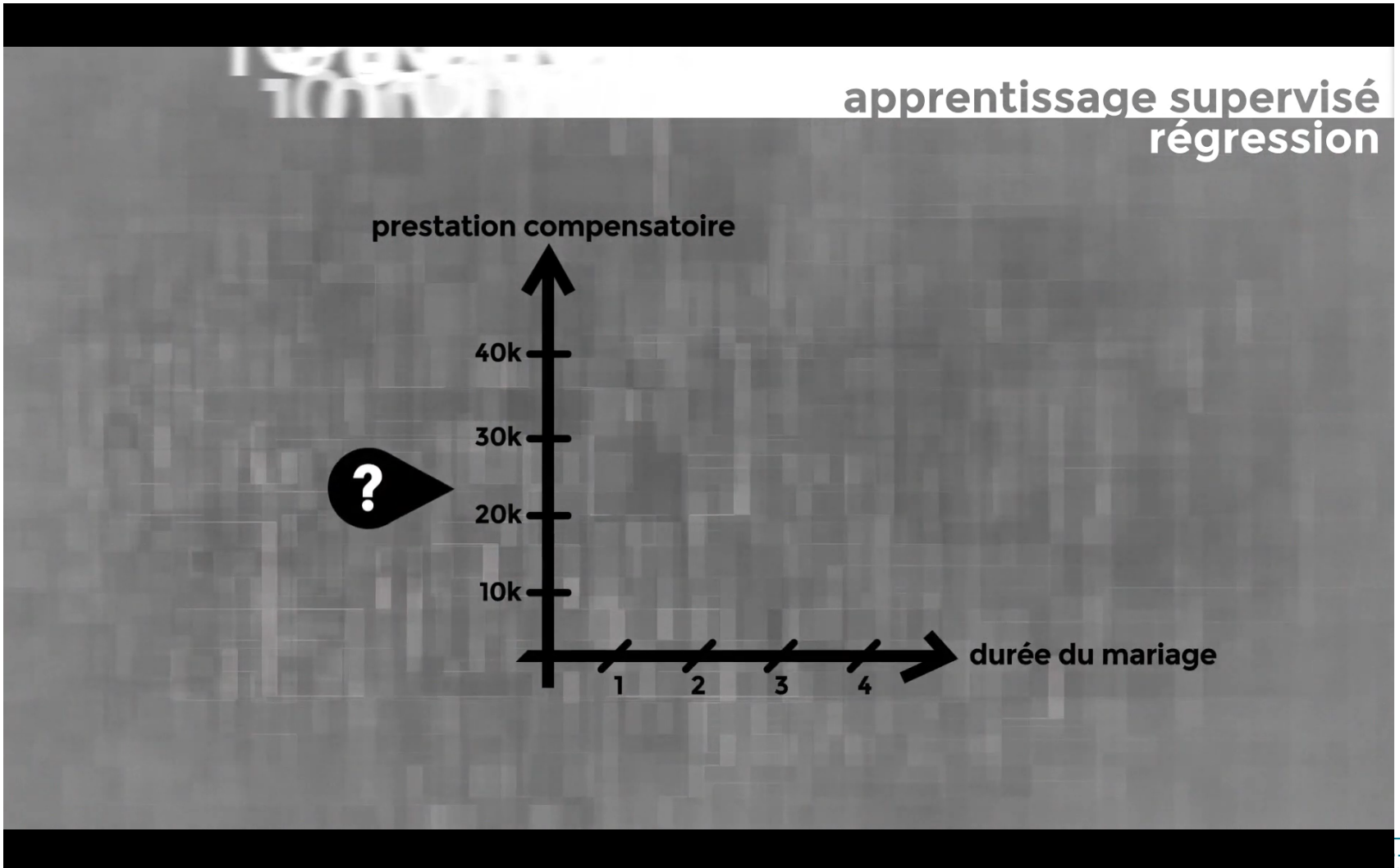
Classification



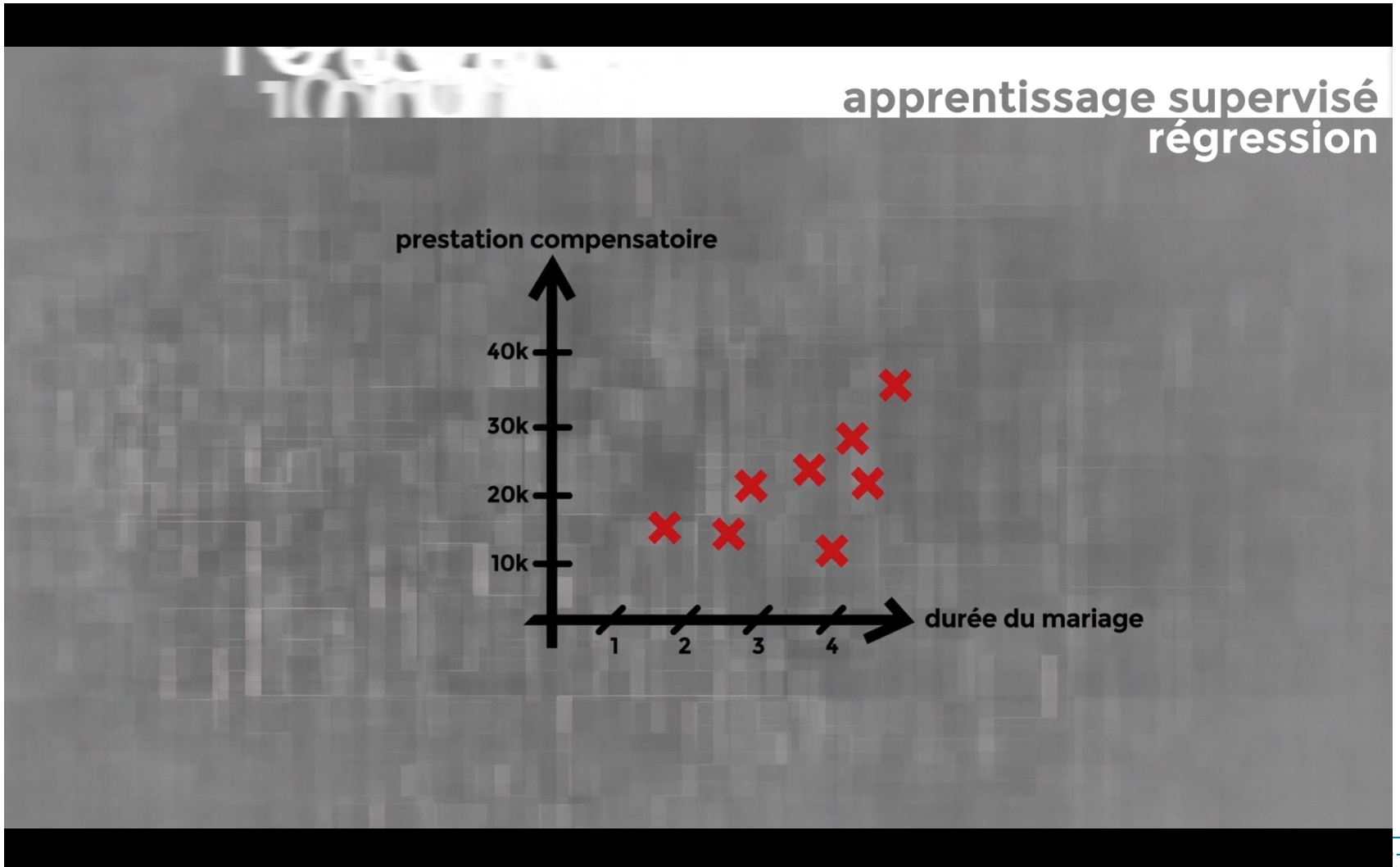
Classification



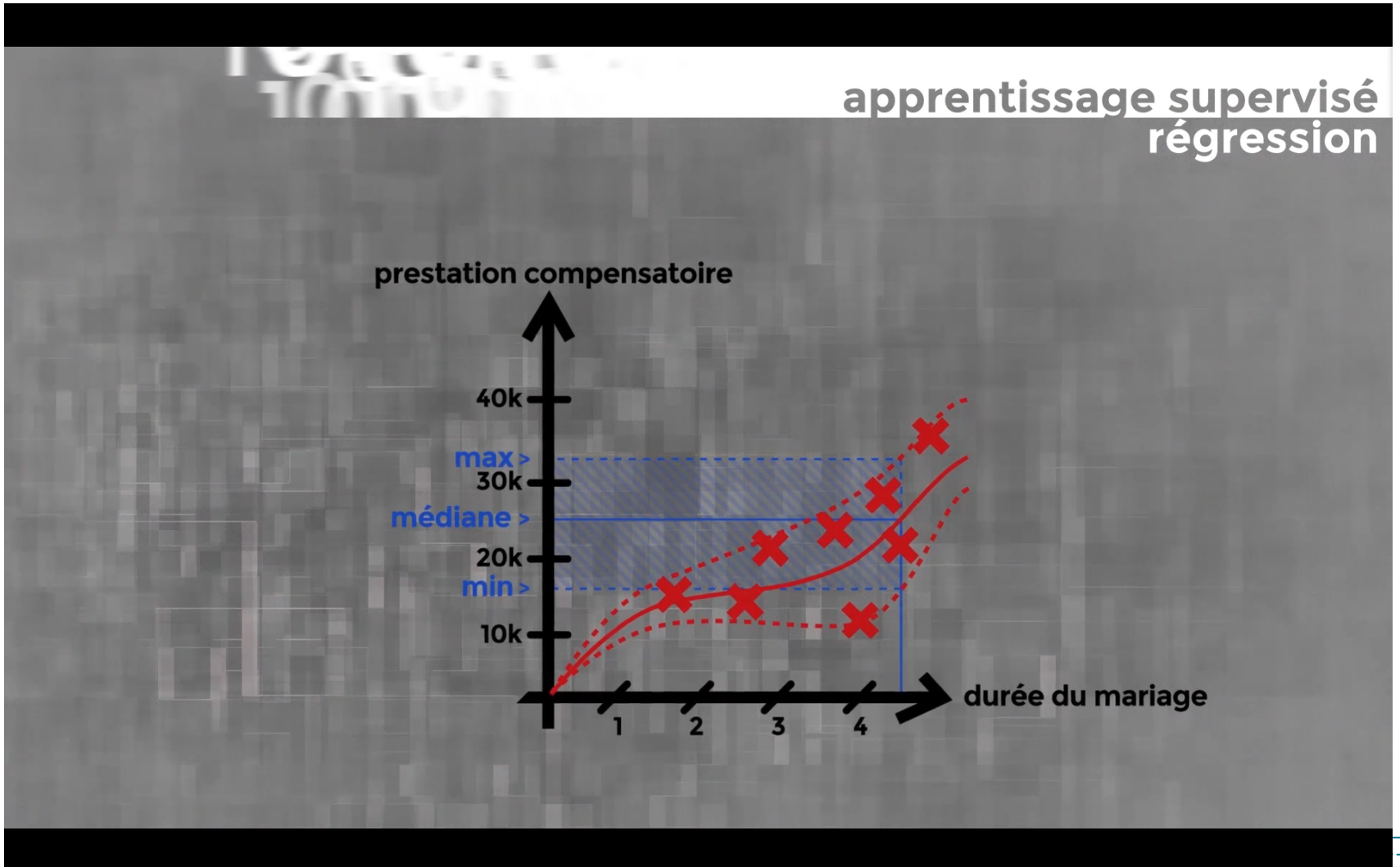
Régressions



Régressions

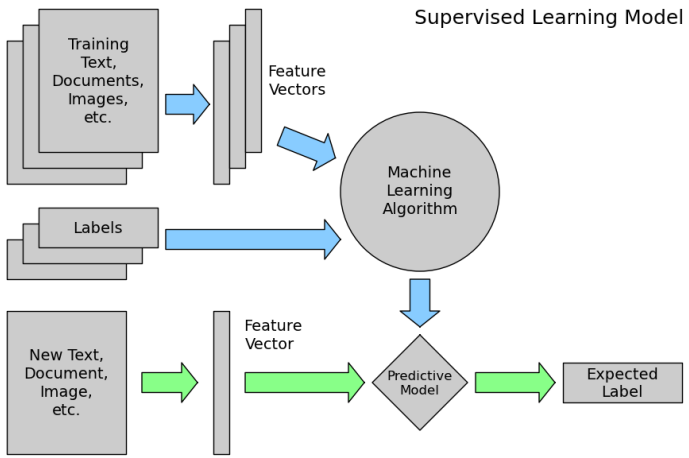


Régressions



Apprentissage supervisé

- on extrait une sélection d'information (critères) à partir des données
- on donne des valeurs à ces critères (poids et seuils)
- on cherche les valeurs qui optimisent le succès de la décision



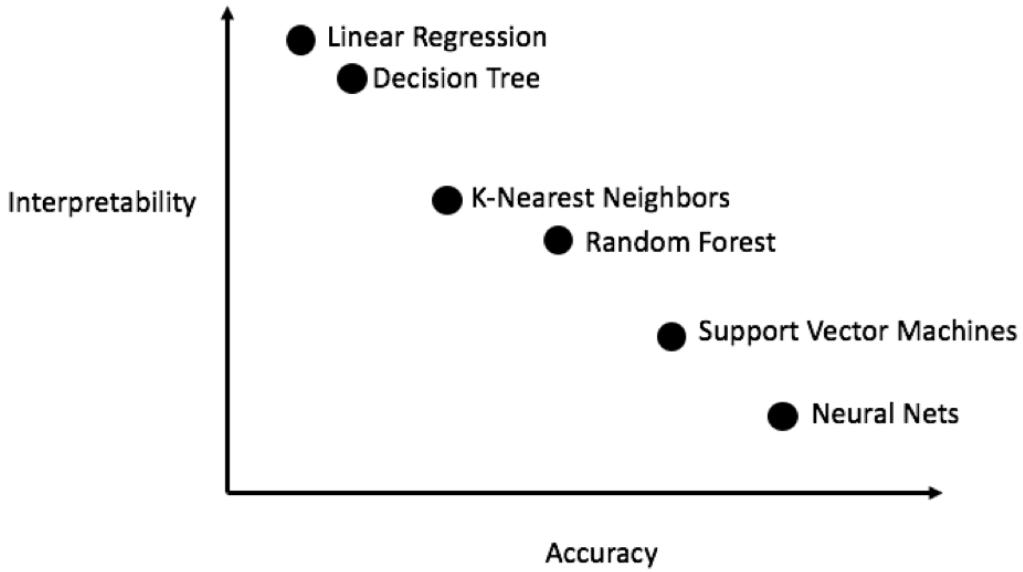
$\Longrightarrow$ La sortie de l'algorithme c'est la valeur de ces poids/seuils.

Documents : affaires, jugements

Vecteurs : profil d'un prévenu (âge, domicile, travail, casier, ...), d'un contrat (durée, âge, revenu, ...)

Étiquettes : discrète (récidive) ou continue (montant compensatoire)

Précision vs. interprétabilité



Classification:

- K plus proches voisins
- Réseaux de neurones
- SVM
- Forêts aléatoires
- ...

Régression:

- Régression linéaire
- Arbres de décision
- Réseaux de neurones
- ...

L'apprentissage dans le contexte juridique

Différents modèles (économiques)

États-Unis

Évaluation du risque lié à une liberté provisoire, liberté conditionnelle, ...
$\longrightarrow$ logiciels utilisés dans les juridictions

- Compas (Northpointe / Equivant) : *"Risk assessment tool for criminal justice practitioner"*

France

Chance de succès d'une affaire, montants (compensatoires, dommages et intérêts, ...)
$\longrightarrow$ logiciels proposés aux cabinets d'avocats

- Predictice (2016) : *"Optimisez votre stratégie juridique."*
- Case Law Analytics (2017) : *"Quantifier le risque juridique par l'Intelligence Artificielle."*

Royaume-Uni

Chance de succès d'une affaire, prédiction de décisions
$\longrightarrow$ chatbot / mise en relation clients $\longleftrightarrow$ avocats (plus maintenant)

- CaseCrunch (2017) : *"Find truth in law."* et *"Solve law"*

Point de vue académique

Human Decisions and Machine Predictions

Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig. The Quarterly Journal of Economics, 2018.

- Problème : détention ou liberté provisoire ?
- Résultats (simulations) : réduction des crimes (24.7 %) ou de la surpopulation carcérale (41.9 %)

exemple : 1% des plus "dangereux" sont relâchés dans 48.5% des cas. Parmi ceux-ci, 56.3% sont absents au procès et 62.7% commettent de nouvelles infractions.

A general approach for predicting the behavior of the US Supreme Court

Daniel M. Katz, Michael J. Bommarito, Josh Blackman. Plos One, 2017.

- Problème : prédiction des décisions (USSC) et des votes (juges)
- Résultats : prédiction correcte dans 70.2 % des décisions, 71.9 % des votes.

Predicting judicial decisions of the ECHR: a Natural Language Processing perspective

Nikolas Aletras, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro, et al. PeerJ in Computer Science, 2016.

- Problème : y a-t-il violation d'un article de la Convention ?
- Résultats : prédiction correcte dans 79 % des affaires

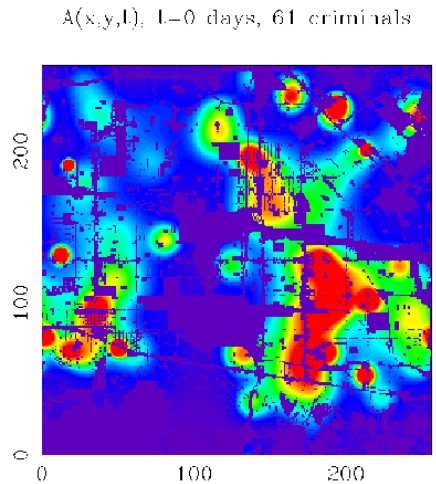
Évaluer un modèle prédictif

L'exemple de PredPol

Cas de PredPol (Predictive Police)

- Problème : où et quand déployer les forces de police ?
- Utilisé depuis 2013 à LA.
- Étude en 2016 sur l'utilisation du logiciel :

To predict and serve?, Kristian Lum and William Isaac, Significance, 13(5), 2016.



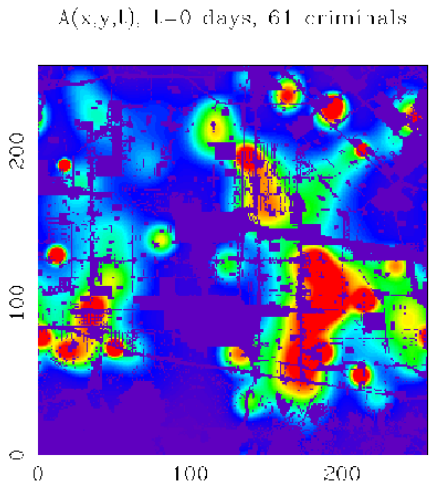
L'exemple de PredPol

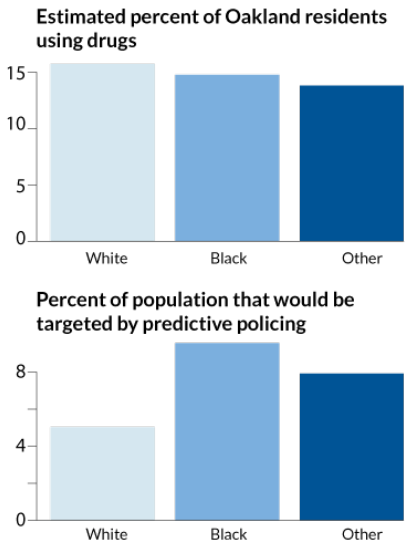
Cas de PredPol (Predictive Police)

- Problème : où et quand déployer les forces de police ?
- Utilisé depuis 2013 à LA.
- Étude en 2016 sur l'utilisation du logiciel : renforce les discriminations

⟶ Prophétie auto-réalisatrice (lien algorithme / données)

To predict and serve?, Kristian Lum and William Isaac, Significance, 13(5), 2016.





Un focus sur l'outil Compas

Compas (Equivant) — *Risk assessment tool for criminal justice practitioner*

Vecteur caractéristique de 137 variables issues d'un questionnaire
(rempli par le prévenu + casier judiciaire)

https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html

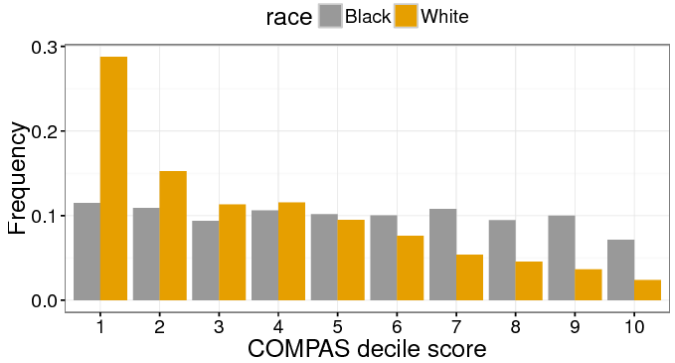
Note to Screener: The following Criminal History Summary questions require you to add up number of specific types of offenses in the person's criminal history. Count an offense type if the charges or counts within an arrest event. Exclude the current case for the following questions.

11. How many times has this person been arrested for a felony property offense that included an element
☐0 ☑1 ☐2 ☐3 ☐4 ☐5+
12. How many prior murder/voluntary manslaughter offense arrests as an adult?
☑0 ☐1 ☐2 ☐3+
13. How many prior felony assault offense arrests (not murder, sex, or domestic violence) as an adult?
☑0 ☐1 ☐2 ☐3+
14. How many prior misdemeanor assault offense arrests (not sex or domestic violence) as an adult?
☑0 ☐1 ☐2 ☐3+
15. How many prior family violence offense arrests as an adult?
☑0 ☐1 ☐2 ☐3+
16. How many prior sex offense arrests (with force) as an adult?
☑0 ☐1 ☐2 ☐3+
17. How many prior weapons offense arrests as an adult?
☑0 ☐1 ☐2 ☐3+
18. How many prior drug trafficking/sales offense arrests as an adult?
☑0 ☐1 ☐2 ☐3+
19. How many prior drug possession/use offense arrests as an adult?
☑0 ☐1 ☐2 ☐3+
20. How many times has this person been sentenced to jail for 30 days or more?
☐0 ☑1 ☐2 ☐3 ☐4 ☐5+
21. How many times has this person been sentenced (new commitment) to state or federal prison?
☐0 ☑1 ☐2 ☐3 ☐4 ☐5+
22. How many times has this person been sentenced to probation as an adult?
☑0 ☐1 ☐2 ☐3 ☐4 ☐5+

The next statements are about your feelings and beliefs about various things. Again, the or wrong' answers. Just indicate how much you agree or disagree with each statement.

127. "A hungry person has a right to steal."
☑ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
128. "When people get into trouble with the law it's because they have no chance to get a decent job."
☐ Strongly Disagree ☑ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
129. "When people do minor offenses or use drugs they don't hurt anyone except themselves."
☑ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
130. "If someone insults my friends, family or group they are asking for trouble."
☐ Strongly Disagree ☐ Disagree ☑ Not Sure ☐ Agree ☐ Strongly Agree
131. "When things are stolen from rich people they won't miss the stuff because insurance will cover the lo
☑ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
132. "I have felt very angry at someone or at something."
☐ Strongly Disagree ☐ Disagree ☐ Not Sure ☑ Agree ☐ Strongly Agree
133. "Some people must be treated roughly or beaten up just to send them a clear message."
☑ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
134. "I won't hesitate to hit or threaten people if they have done something to hurt my friends or family."
☐ Strongly Disagree ☐ Disagree ☑ Not Sure ☐ Agree ☐ Strongly Agree
135. "The law doesn't help average people."
☑ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree

Le débat



Évaluation de Compas (effectuée par Propublica)

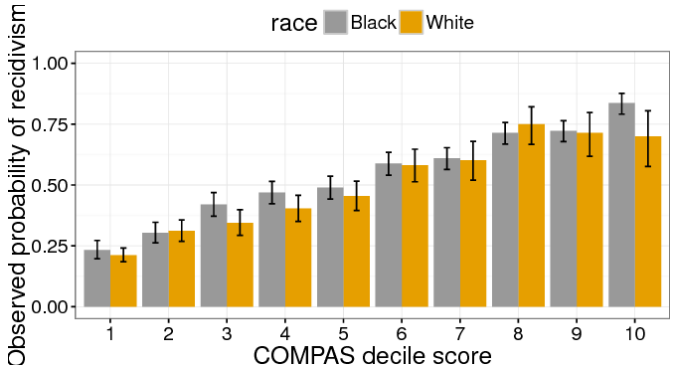
Machine bias. J. Angwin, J. Larson, S. Mattu, L. Kirchner. *ProPublica.* 2016, May.

`https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing`

> *There's software used across the country to predict future criminals. And it's biased against blacks.*

FPR parity	Afro-américains qui n'ont pas récidivé sont plus souvent catégorisés à haut risque que les blancs (45 % vs. 23 %)
FNR parity	Blancs qui ont récidivé sont plus souvent catégorisés à faible risque que les afro-américains (48 % vs. 28 %)

Le débat



Réponse de Northpointe

COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity. W. Dieterich, C. Mendoza, T. Brennan. *Northpointe Inc. Research Department*. 2016, July.

> *We strongly reject the conclusion that the COMPAS risk scales are racially biased against blacks. ProPublica [...] did not take into account the different base rates of recidivism for blacks and whites. ProPublica [...] wrongly defined classification terms and measures of discrimination.*

faux positifs	Afro-américains catégorisés à haut risque ont la même probabilité de ne pas récidiver que les blancs (37 % vs. 41 %)
faux négatifs	Afro-américains catégorisés à faible risque ont la même probabilité de récidiver que les blancs (35 % vs. 29 %)

Évaluer un modèle prédictif

Le problème

- Considérons une tache de classification binaire

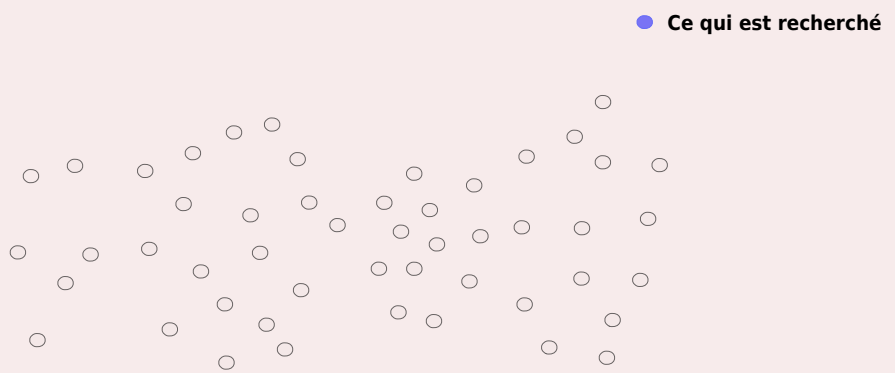
  $\longrightarrow$ récidive / non récidive

- Considérons un modèle prédictif (issu par exemple d'un algorithme d'apprentissage supervisé).

  $\longrightarrow$ risque récidive élevé / faible

Comment évaluer la qualité du modèle ?

Matrice de confusion



Évaluer un modèle prédictif

Le problème

- Considérons une tache de classification binaire

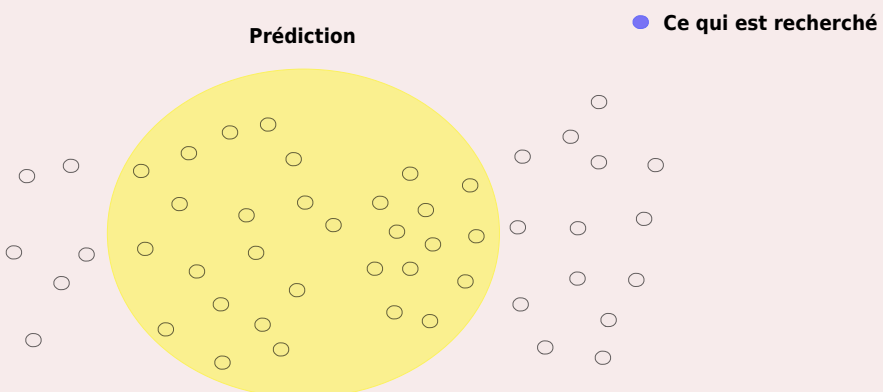
  $\longrightarrow$ récidive / non récidive

- Considérons un modèle prédictif (issu par exemple d'un algorithme d'apprentissage supervisé).

  $\longrightarrow$ risque récidive élevé / faible

Comment évaluer la qualité du modèle ?

Matrice de confusion



Évaluer un modèle prédictif

Le problème

- Considérons une tache de classification binaire

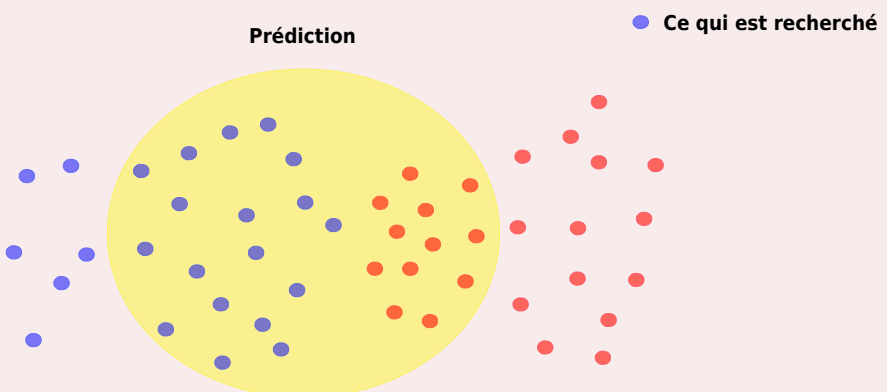
  $\longrightarrow$ récidive / non récidive

- Considérons un modèle prédictif (issu par exemple d'un algorithme d'apprentissage supervisé).

  $\longrightarrow$ risque récidive élevé / faible

Comment évaluer la qualité du modèle ?

Matrice de confusion



Évaluer un modèle prédictif

Le problème

- Considérons une tache de classification binaire

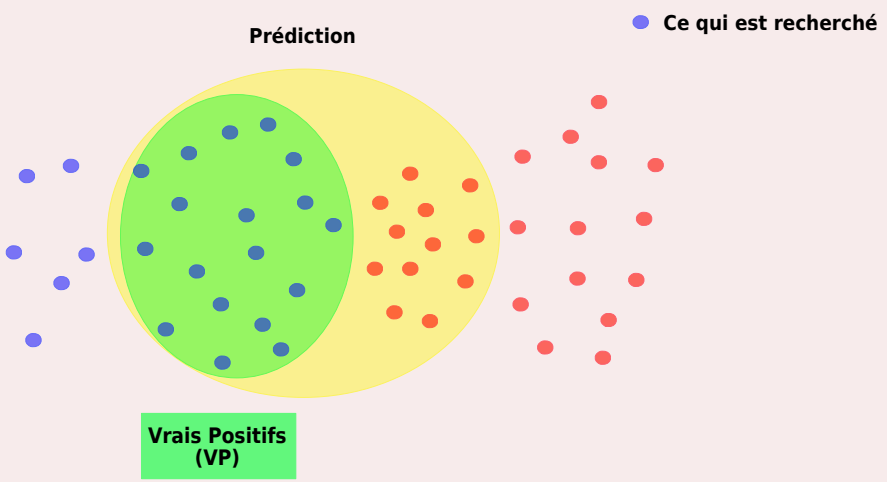
  $\longrightarrow$ récidive / non récidive

- Considérons un modèle prédictif (issu par exemple d'un algorithme d'apprentissage supervisé).

  $\longrightarrow$ risque récidive élevé / faible

Comment évaluer la qualité du modèle ?

Matrice de confusion



Évaluer un modèle prédictif

Le problème

- Considérons une tache de classification binaire

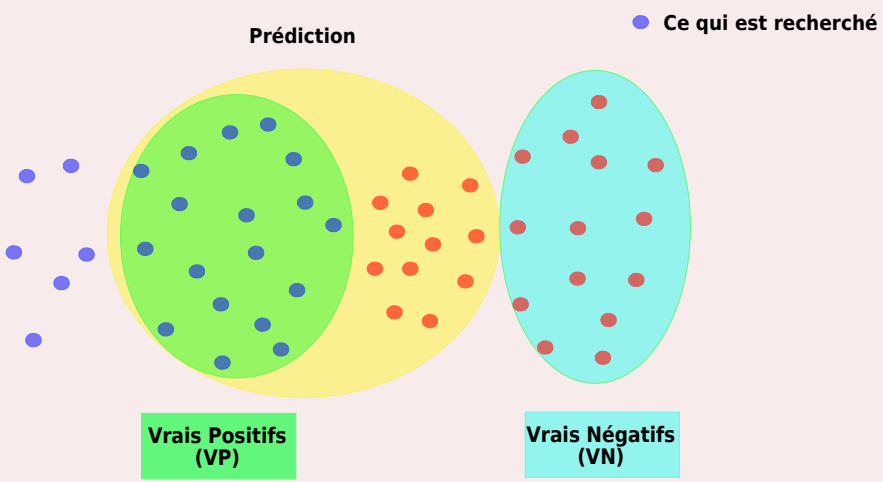
  $\longrightarrow$ récidive / non récidive

- Considérons un modèle prédictif (issu par exemple d'un algorithme d'apprentissage supervisé).

  $\longrightarrow$ risque récidive élevé / faible

Comment évaluer la qualité du modèle ?

Matrice de confusion



Évaluer un modèle prédictif

Le problème

- Considérons une tache de classification binaire

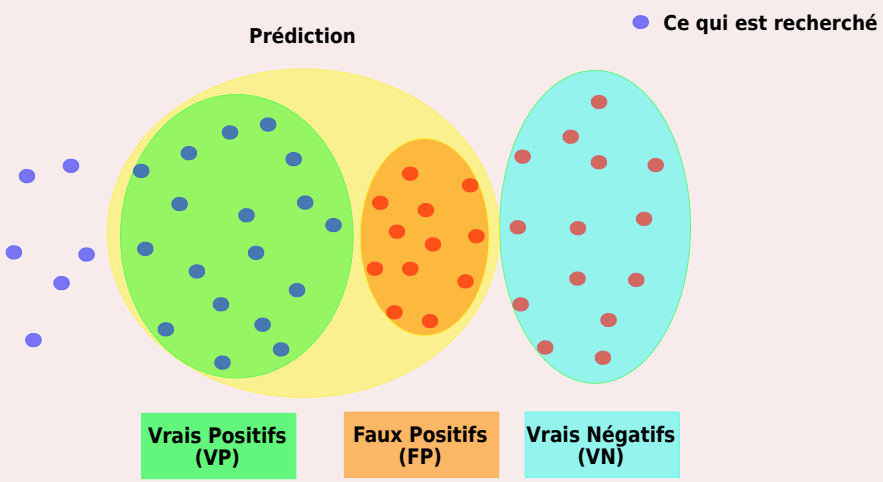
  $\longrightarrow$ récidive / non récidive

- Considérons un modèle prédictif (issu par exemple d'un algorithme d'apprentissage supervisé).

  $\longrightarrow$ risque récidive élevé / faible

Comment évaluer la qualité du modèle ?

Matrice de confusion



Évaluer un modèle prédictif

Le problème

- Considérons une tache de classification binaire

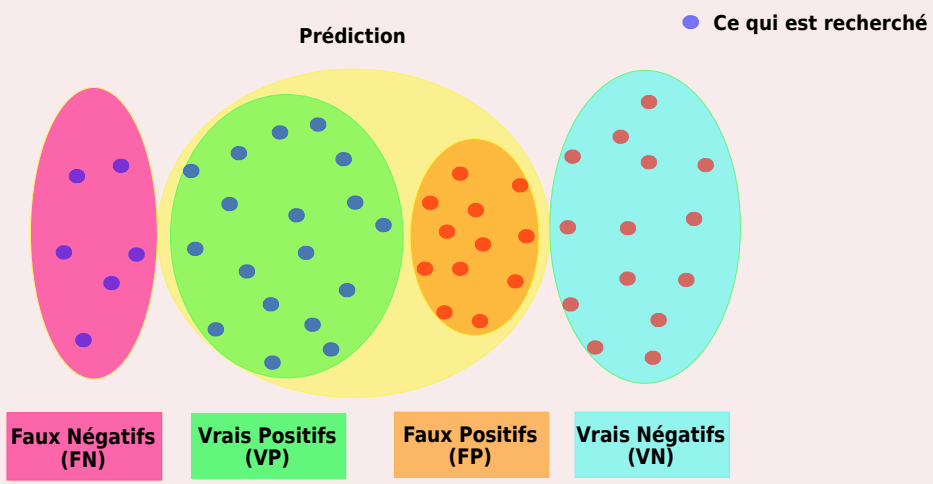
  $\longrightarrow$ récidive / non récidive

- Considérons un modèle prédictif (issu par exemple d'un algorithme d'apprentissage supervisé).

  $\longrightarrow$ risque récidive élevé / faible

Comment évaluer la qualité du modèle ?

Matrice de confusion



Évaluer un modèle prédictif

Le problème

- Considérons une tache de classification binaire

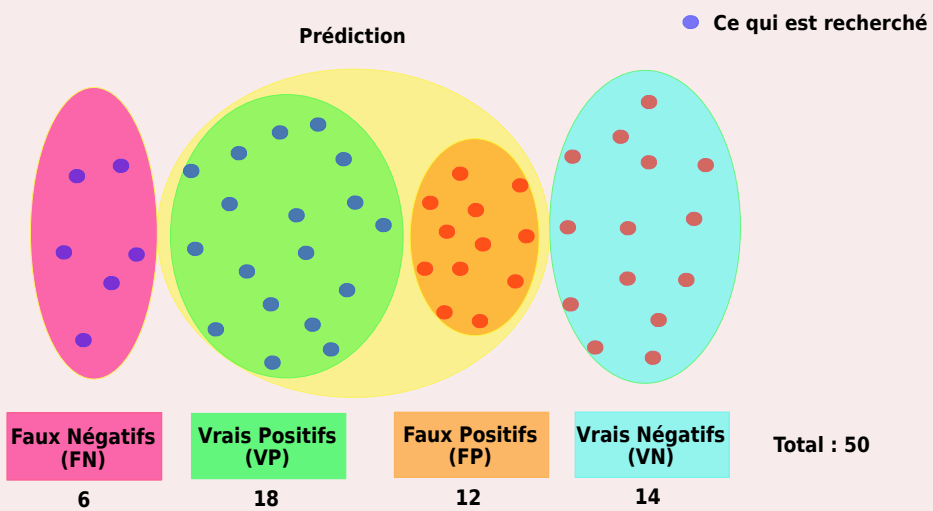
  $\longrightarrow$ récidive / non récidive

- Considérons un modèle prédictif (issu par exemple d'un algorithme d'apprentissage supervisé).

  $\longrightarrow$ risque récidive élevé / faible

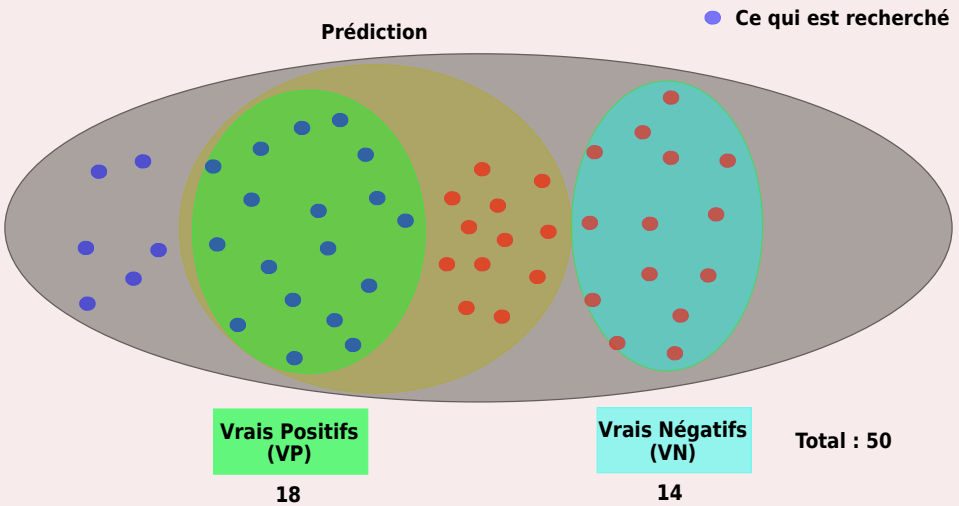
Comment évaluer la qualité du modèle ?

Matrice de confusion



Évaluer un modèle prédictif

Matrice de confusion

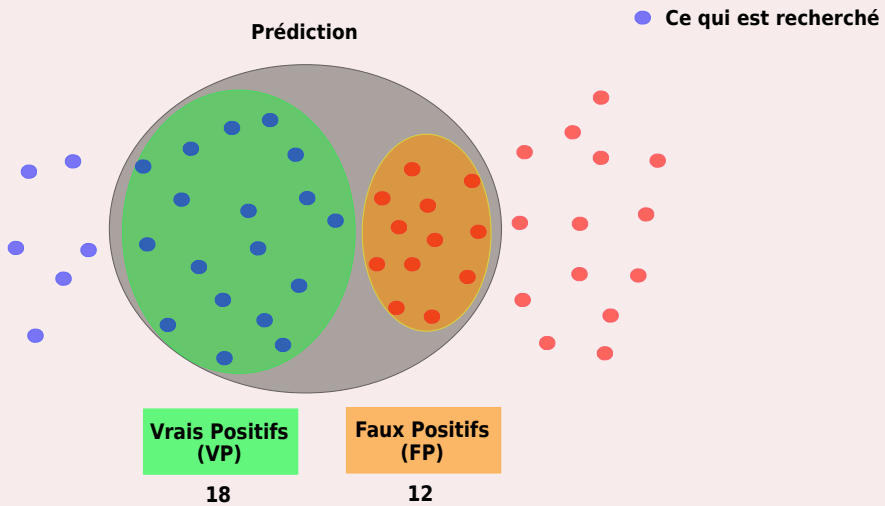


Quantifier la qualité (succès) du modèle

Accuracy $\mathrm{ACC} = \frac{VP+VN}{P+N} = \frac{\textit{bonne predictions}}{\textit{total}} = \frac{32}{50} = 0.64$ exactitude / justesse

Évaluer un modèle prédictif

Matrice de confusion



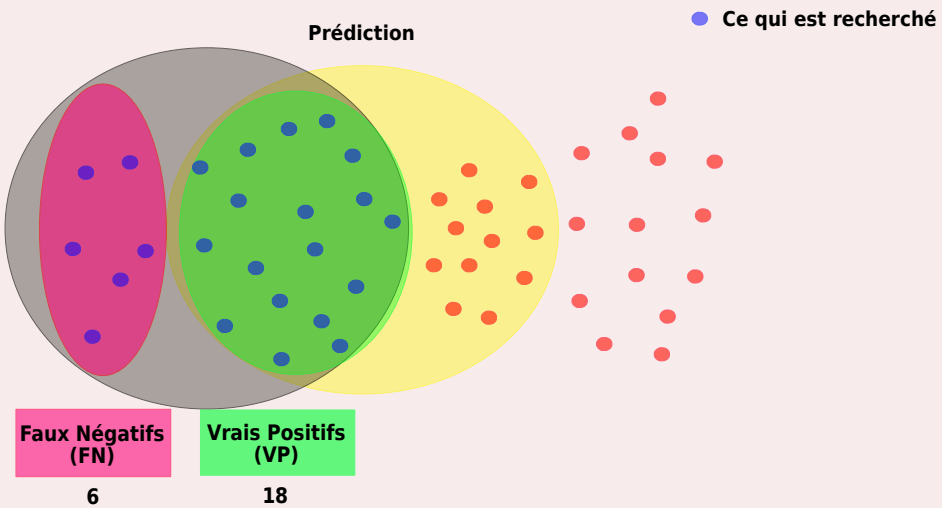
Quantifier la qualité (succès) du modèle

Accuracy $\mathrm{ACC} = \frac{VP+VN}{P+N} = \frac{\textit{bonne predictions}}{\textit{total}} = \frac{32}{50} = 0.64$ exactitude / justesse

Precision $\mathrm{PR} = \frac{VP}{VP+FP} = \frac{\textit{positifs predits}}{\textit{nb predictions}} = \frac{18}{30} = 0.6$ précision

Évaluer un modèle prédictif

Matrice de confusion



Quantifier la qualité (succès) du modèle

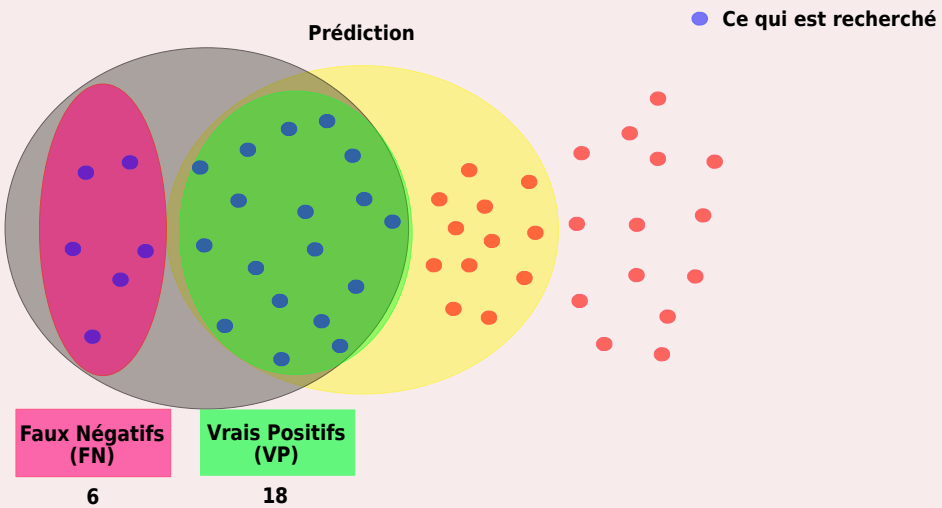
Accuracy $\mathrm{ACC} = \frac{VP+VN}{P+N} = \frac{\textit{bonne predictions}}{\textit{total}} = \frac{32}{50} = 0.64$ exactitude / justesse

Precision $\mathrm{PR} = \frac{VP}{VP+FP} = \frac{\textit{positifs predits}}{\textit{nb predictions}} = \frac{18}{30} = 0.6$ précision

Recall $\mathrm{RC} = \frac{VP}{VP+FN} = \frac{\textit{positifs predits}}{\textit{nb positifs}} = \frac{18}{24} = 0.75$ rappel / sensibilité

Évaluer un modèle prédictif

Matrice de confusion



Quantifier la qualité (succès) du modèle

Accuracy $\mathrm{ACC} = \frac{VP+VN}{P+N} = \frac{\textit{bonne predictions}}{\textit{total}} = \frac{32}{50} = 0.64$ — exactitude / justesse

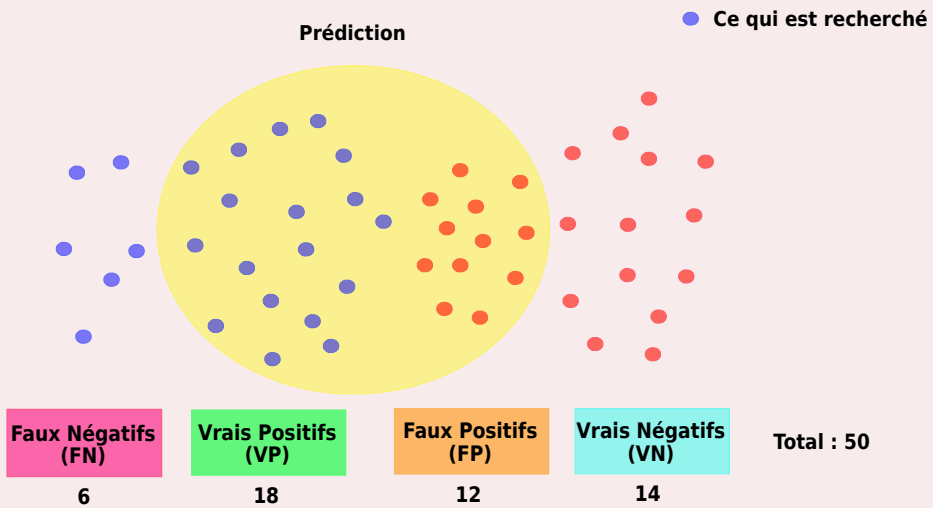
Precision $\mathrm{PR} = \frac{VP}{VP+FP} = \frac{\textit{positifs predits}}{\textit{nb predictions}} = \frac{18}{30} = 0.6$ — précision

Recall $\mathrm{RC} = \frac{VP}{VP+FN} = \frac{\textit{positifs predits}}{\textit{nb positifs}} = \frac{18}{24} = 0.75$ — rappel / sensibilité

F-score $\mathrm{F} = \frac{2\times \mathrm{PR}\times \mathrm{RC}}{\mathrm{PR}+\mathrm{RC}} = 0.67$

Évaluer un modèle prédictif

Matrice de confusion



Quantifier la qualité (succès) du modèle

Accuracy $\mathrm{ACC} = \frac{VP+VN}{P+N} = \frac{\textit{bonne predictions}}{\textit{total}} = \frac{32}{50} = 0.64$ — exactitude / justesse

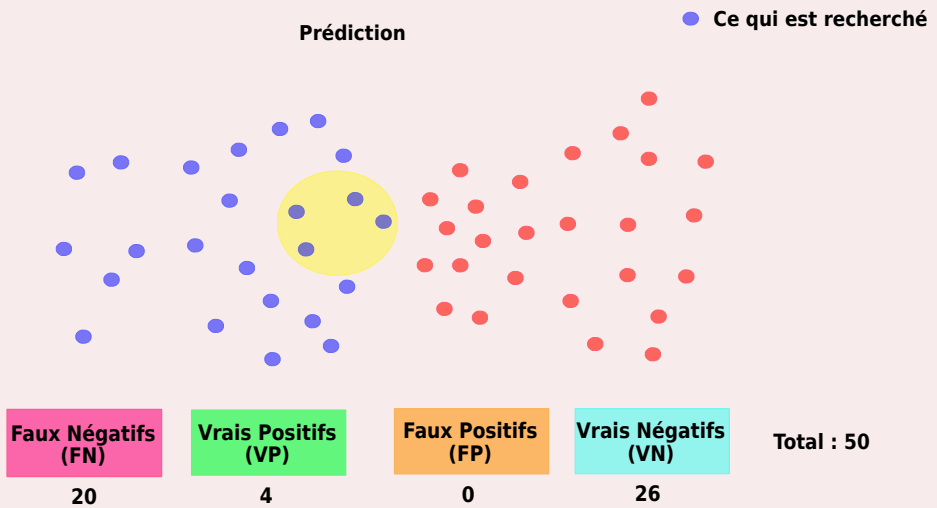
Precision $\mathrm{PR} = \frac{VP}{VP+FP} = \frac{\textit{positifs predits}}{\textit{nb predictions}} = \frac{18}{30} = 0.6$ — précision

Recall $\mathrm{RC} = \frac{VP}{VP+FN} = \frac{\textit{positifs predits}}{\textit{nb positifs}} = \frac{18}{24} = 0.75$ — rappel / sensibilité

F-score $\mathrm{F} = \frac{2\times \mathrm{PR}\times \mathrm{RC}}{\mathrm{PR}+\mathrm{RC}} = 0.67$

Évaluer un modèle prédictif

Matrice de confusion



Quantifier la qualité (succès) du modèle

Accuracy $\mathrm{ACC} = \frac{VP+VN}{P+N} = \frac{\textit{bonne predictions}}{\textit{total}} = \frac{30}{50} = 0.6$ — exactitude / justesse

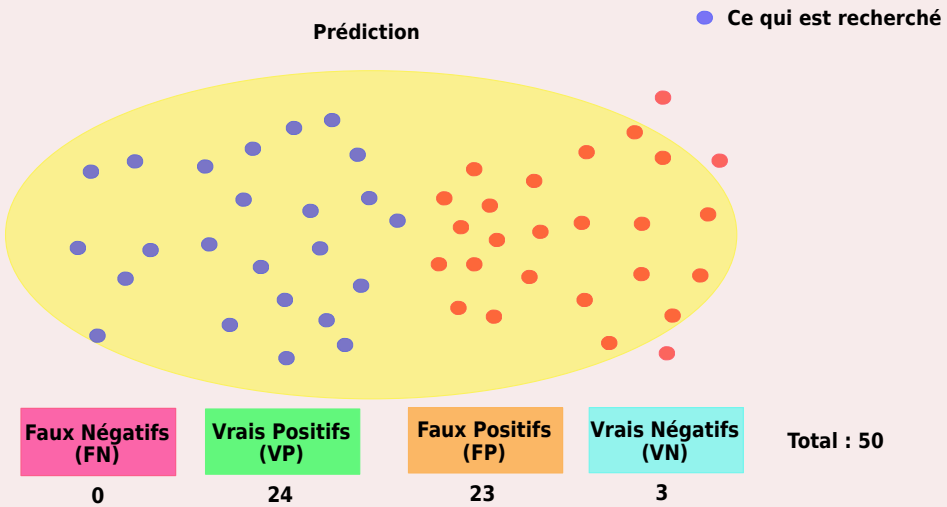
Precision $\mathrm{PR} = \frac{VP}{VP+FP} = \frac{\textit{positifs predits}}{\textit{nb predictions}} = \frac{4}{4} = 1.0$ — précision

Recall $\mathrm{RC} = \frac{VP}{VP+FN} = \frac{\textit{positifs predits}}{\textit{nb positifs}} = \frac{4}{24} = 0.17$ — rappel / sensibilité

F-score $\mathrm{F} = \frac{2 \times \mathrm{PR} \times \mathrm{RC}}{\mathrm{PR}+\mathrm{RC}} = 0.27$

Évaluer un modèle prédictif

Matrice de confusion



Quantifier la qualité (succès) du modèle

Accuracy $\mathrm{ACC} = \frac{VP+VN}{P+N} = \frac{\textit{bonne predictions}}{\textit{total}} = \frac{27}{50} = 0.54$ — exactitude / justesse

Precision $\mathrm{PR} = \frac{VP}{VP+FP} = \frac{\textit{positifs predits}}{\textit{nb predictions}} = \frac{24}{47} = 0.51$ — précision

Recall $\mathrm{RC} = \frac{VP}{VP+FN} = \frac{\textit{positifs predits}}{\textit{nb positifs}} = \frac{24}{24} = 1.0$ — rappel / sensibilité

F-score $\mathrm{F} = \frac{2 \times \mathrm{PR} \times \mathrm{RC}}{\mathrm{PR}+\mathrm{RC}} = 0.68$

Différentes notions d'équité

Le problème

- Considérons une tache de classification binaire (observation O)

  $\longrightarrow$ récidive / non récidive

- Considérons un modèle prédictif M (issu par exemple d'un algorithme d'apprentissage supervisé).

  $\longrightarrow$ risque récidive élevé / faible

Différentes notions d'équité

Le problème

- Considérons une tache de classification binaire (observation O)

  $\longrightarrow$ récidive / non récidive
- Considérons un modèle prédictif M (issu par exemple d'un algorithme d'apprentissage supervisé).

  $\longrightarrow$ risque récidive élevé / faible
- Considérons deux groupes G distincts : ○ et ☆

  $\longrightarrow$ blancs / afro-américains

  $\longrightarrow$ hommes / femmes

Comment évaluer si les groupes sont traités de manière équitable par le modèle ?

Différentes notions d'équité

Le problème

- Considérons une tache de classification binaire (observation O)
  $\longrightarrow$ récidive / non récidive
- Considérons un modèle prédictif M (issu par exemple d'un algorithme d'apprentissage supervisé).
  $\longrightarrow$ risque récidive élevé / faible
- Considérons deux groupes G distincts : ○ et ☆
  $\longrightarrow$ blancs / afro-américains
  $\longrightarrow$ hommes / femmes

Comment évaluer si les groupes sont traités de manière équitable par le modèle ?

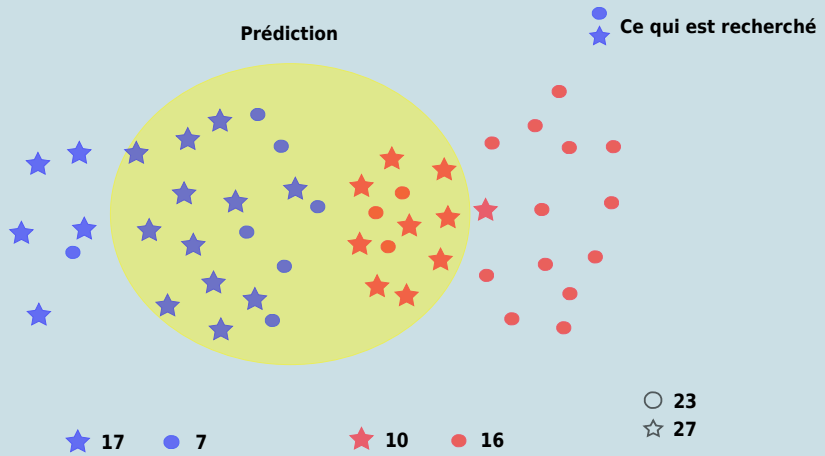
Plusieurs propositions (*group fairness*) :

- Parité démographique (*Demographic parity*)
- Égalité en opportunité (*Equal opportunity*)
- Parité en taux de faux-positifs (*False positive rate parity*)
- Égalité des chances (*Equalized odds*)
- Parité en taux de prédiction (*Predictive rate parity*)
- ...

Quelle relation entre ces notions ?

Parité démographique

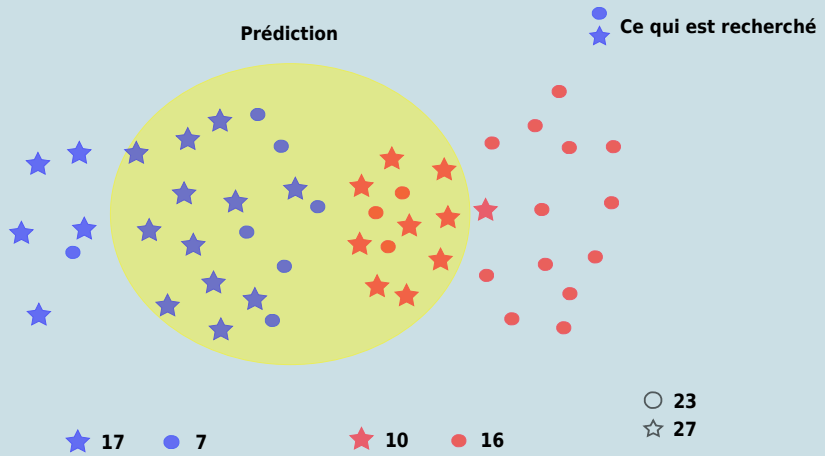
Matrice de confusion



Demographic parity (ou *statistical parity*)

Parité démographique

Matrice de confusion



Demographic parity (ou *statistical parity*)

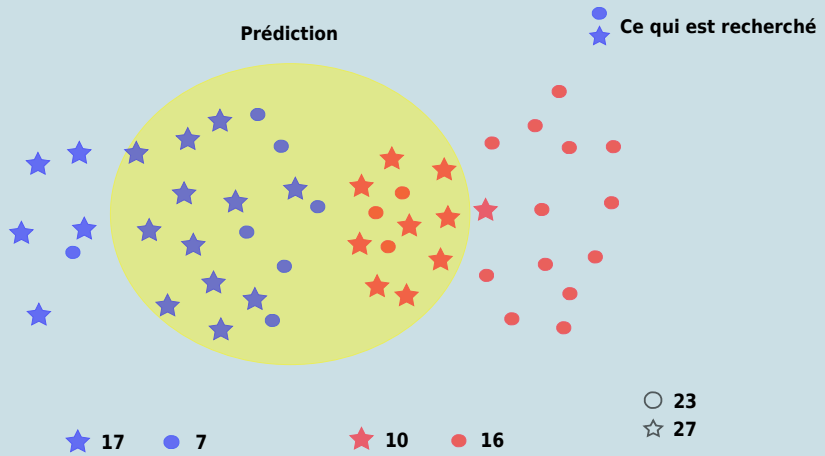
Un modèle est réputé préserver la parité démographique si :

$$P(M = \bullet \,|G = \bigcirc) = P(M = \bigstar \,|G = \bigstar)$$
$$P(M = \bullet \,|G = \bigcirc) = P(M = \bigstar \,|G = \bigstar)$$

La probabilité qu'un individu reçoive une prédiction (positive ou négative) doit être la même quelque soit le groupe auquel il appartient.

Parité démographique

Matrice de confusion



Demographic parity (ou *statistical parity*)

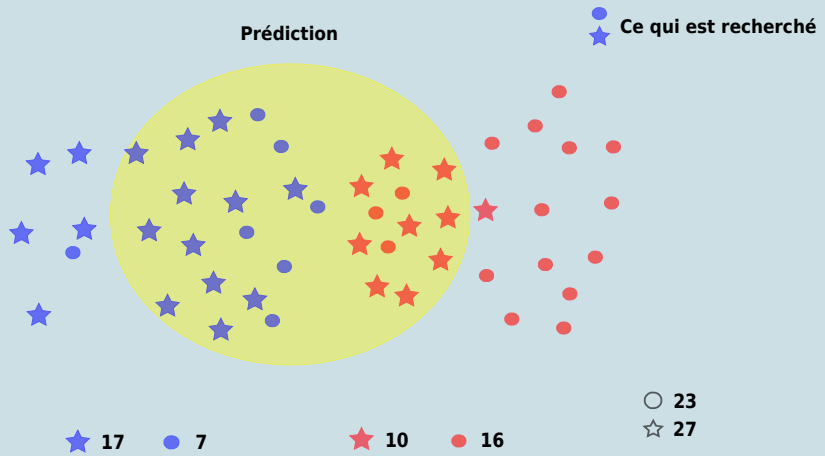
Un modèle est réputé préserver la parité démographique si :

$$0.39 \leftarrow \quad P(M = \bullet \,|G = \bigcirc) = P(M = \bigstar \,|G = \bigstar) \quad \rightarrow 0.78$$
$$0.61 \leftarrow \quad P(M = \bullet \,|G = \bigcirc) = P(M = \bigstar \,|G = \bigstar) \quad \rightarrow 0.22$$

La probabilité qu'un individu reçoive une prédiction (positive ou négative) doit être la même quelque soit le groupe auquel il appartient.

Égalité en opportunité

Matrice de confusion



Equal opportunity

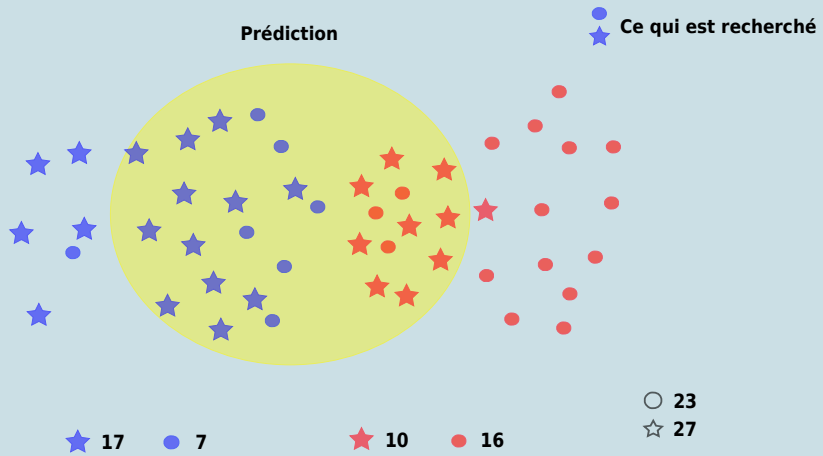
Un modèle est réputé préserver l'égalité en opportunité si :

$$P(M = \bullet \,|G = \circ, O = \bullet) = P(M = \bigstar \,|G = \star, O = \bigstar)$$

La probabilité qu'un individu positif reçoive une prédiction positive doit être la même quelque soit le groupe auquel il appartient.

Égalité en opportunité

Matrice de confusion



Equal opportunity

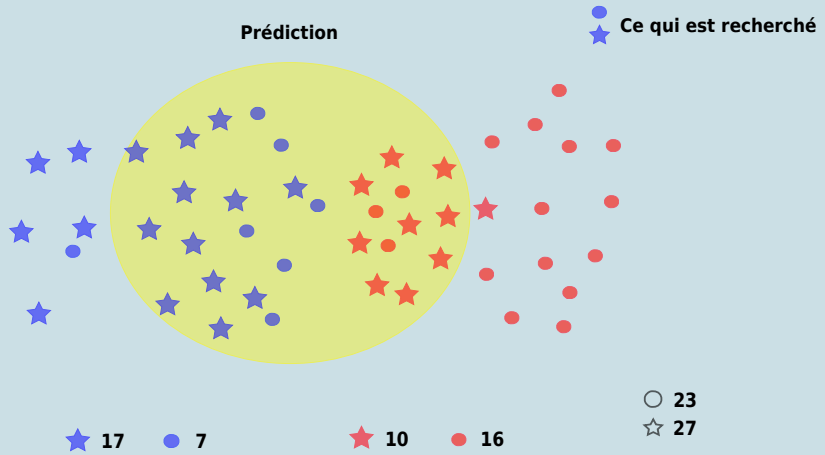
Un modèle est réputé préserver l'égalité en opportunité si :

$$0.86 \leftarrow \quad P(M = \bullet \,|G = \bigcirc, O = \bullet) = P(M = \bigstar \,|G = \bigstar, O = \bigstar) \quad \rightarrow 0.71$$

La probabilité qu'un individu positif reçoive une prédiction positive doit être la même quelque soit le groupe auquel il appartient.

Parité vis-à-vis des faux-positifs

Matrice de confusion



False positive rate parity

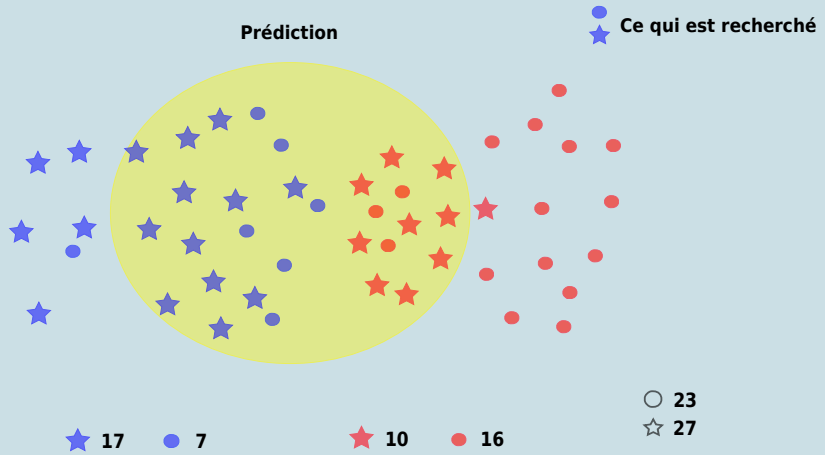
Un modèle est réputé préserver la parité en taux de faux-positifs si :

$$P(M = \bullet \,|G = \circ, O = \bullet) = P(M = \bigstar \,|G = \star, O = \bigstar)$$

La probabilité qu'un individu négatif reçoive une prédiction positive doit être la même quelque soit le groupe auquel il appartient.

Parité vis-à-vis des faux-positifs

Matrice de confusion



False positive rate parity

Un modèle est réputé préserver la parité en taux de faux-positifs si :

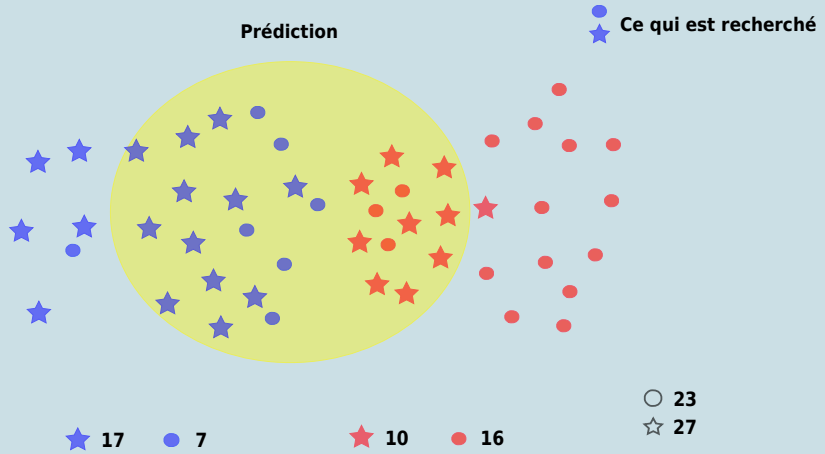
$$0.19 \leftarrow \quad P(M = \bullet \,|G = \circ, O = \bullet) = P(M = \bigstar \,|G = \star, O = \bigstar) \quad \rightarrow 0.90$$

La probabilité qu'un individu négatif reçoive une prédiction positive doit être la même quelque soit le groupe auquel il appartient.

C'est exactement ce qui est mis en défaut par ProPublica !!

Parité vis-à-vis des prédictions

Matrice de confusion



Prédictive rate parity

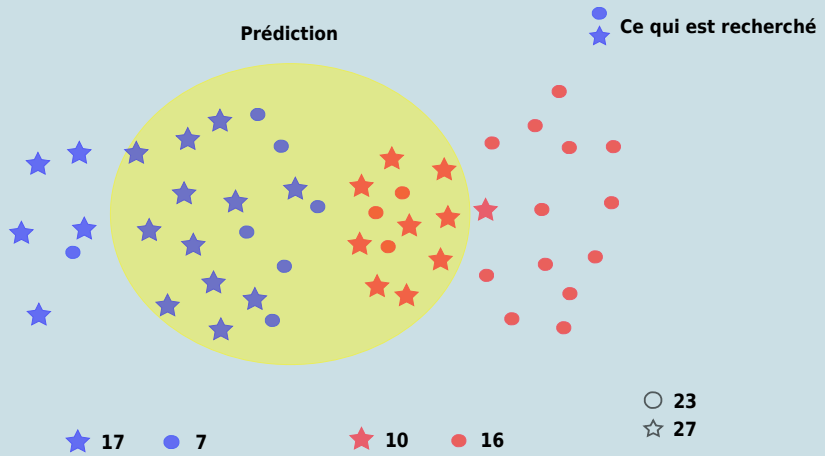
Un modèle est réputé préserver la parité en taux de prédiction si :

$$P(O = \bullet \,|G = \circ, M = \bullet) = P(O = \bigstar \,|G = \star, M = \bigstar)$$

La probabilité qu'un individu prédit comme positif soit effectivement positif (resp. négatif) doit être la même quelque soit le groupe auquel il appartient.

Parité vis-à-vis des prédictions

Matrice de confusion



Prédictive rate parity

Un modèle est réputé préserver la parité en taux de prédiction si :

$$0.67 \leftarrow \quad P(O = \bullet \,|G = \bigcirc, M = \bullet) = P(O = \bigstar \,|G = \bigstar, M = \bigstar) \quad \rightarrow 0.57$$

La probabilité qu'un individu prédit comme positif soit effectivement positif (resp. négatif) doit être la même quelque soit le groupe auquel il appartient.

C'est exactement ce qui est mis en avant par NORTHPOINT/EQUIVANT !!

Que faire de ces notions ?

Toutes ces notions d'équité sont souhaitables mais ...

Que faire de ces notions ?

Toutes ces notions d'équité sont souhaitables mais ...

Théorème d'impossibilité

À moins d'avoir un modèle prédictif parfait (oracle) ou que les groupes aient exactement la même proportion d'individus positifs (*equal base rate*), un modèle ne peut pas satisfaire à la fois :

- la parité démographique
- l'équalité des chances (liée à la propriété mise en avant par ProPublica)
- la parité en taux de prédiction (mis en avant par Equivant)

Que faire de ces notions ?

Toutes ces notions d'équité sont souhaitables mais ...

Théorème d'impossibilité

À moins d'avoir un modèle prédictif parfait (oracle) ou que les groupes aient exactement la même proportion d'individus positifs (*equal base rate*), un modèle ne peut pas satisfaire à la fois :

- la parité démographique
- l'équalité des chances (liée à la propriété mise en avant par ProPublica)
- la parité en taux de prédiction (mis en avant par Equivant)

Pire : ces trois propriétés sont incompatibles deux à deux !

Discussion

Impact sur la justice

- Comment les juges utilisent ces outils ? (affaire *Paul Zilly*)
- Quelles recours ? (affaire *State vs. Loomis*)

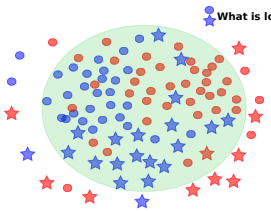
Réflexions en vrac

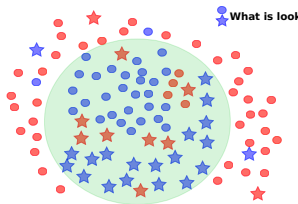
- Théorème d'impossibilité : pas seulement pour l'apprentissage ...
- sources de biais (23) : *selection bias*, *information bias*, *cofounding factors*, ...
- Intégrer de l'équité dans les modèles : approches *preprocessing*, *postprocessing* et *fairness-aware algorithms*.
- Biais cognitifs : choix des critères, concepts d'*automation bias*, *moral buffer*, ...
- Outils présentés comme utiles au niveau individuel mais basés sur calcul global
- Corrélation n'est pas causalité.

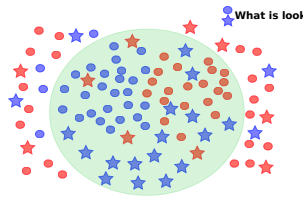
Constat pas très rassurant .... mais et si nous pouvions :

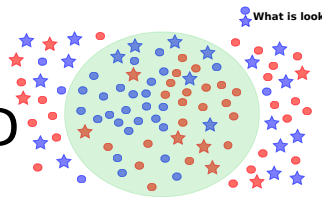
- controler la distance à une équité optimale pour les différentes notions ? (en relachant les contraintes)
  Cela va de pair avec une perte en efficacité ...
- mettre les praticiens (juges, avocats, ...) en position d'experts ?
  Cela va de pair avec un coût (energie et investissement dans la formation) ...

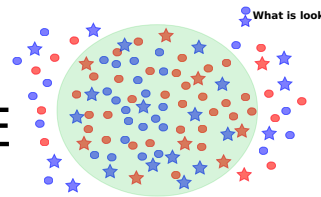
Quel type de juge êtes-vous ?

A 

B 

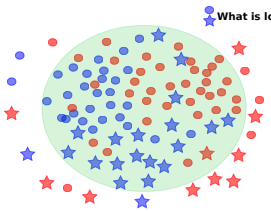
C 

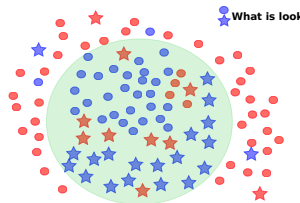
D 

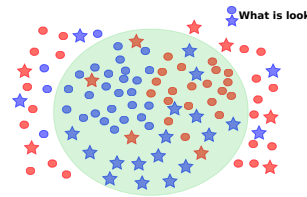
E 

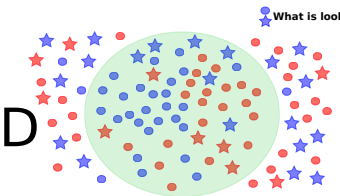
Modèle	Acc	Pr	Recall	F-score	Dem Parity	FP Parity	PR Parity
A							
B							
C							
D							
E							

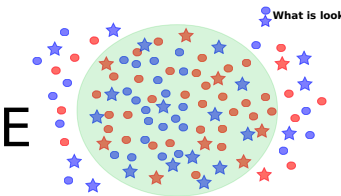
Quel type de juge êtes-vous ?

A 

B 

C 

D 

E 

Modèle	Acc	Pr	Recall	F-score	Dem Parity	FP Parity	PR Parity
A	0.60	0.56	0.92	0.69	0.90 0.63	0.87 0.10	0.43 0.95
B	0.83	0.78	0.92	0.84	0.47 0.87	0.12 0.80	0.85 0.69
C	0.69	0.64	0.88	0.74	0.65 0.79	0.49 0.56	0.57 0.78
D	0.57	0.56	0.64	0.60	0.66 0.37	0.50 0.50	0.57 0.55
E	0.45	0.46	0.66	0.55	0.71 0.70	0.75 0.80	0.40 0.62

Questions?

http://tarissan.complexnetworks.fr/