

Algorithmes d'apprentissage

Le cas de la justice "prédictive"

Fabien Tarissan

CNRS – ENS Paris Saclay

PHILOSOPHIE DES SCIENCES
ET INTELLIGENCE ARTIFICIELLE

Plan

- 1 Algorithmes d'apprentissage
 - Qu'est-ce qu'un algorithme
 - Les principes de l'apprentissage supervisé

- 2 L'apprentissage dans le contexte juridique
 - Différents acteurs pour différents modèles
 - Articles scientifiques
 - Études empiriques
 - Discussions

Algorithmes d'apprentissage

Qu'est-ce qu'un algorithme ?

Algorithme

Qu'est-ce qu'un algorithme ?

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'ISC-PIF depuis son domicile ?)
- algorithme de Dijkstra (comment trouver un plus court chemin entre 2 points ?)

Qu'est-ce qu'un algorithme ?

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'ISC-PIF 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, ...) et qui résolvent des problèmes **génériques**.

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

Une grande diversité de méthodes

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
- ...

Non supervisé

Clustering:

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

Apprentissage supervisé

Un algorithme d'apprentissage supervisé (simplifié) dispose :

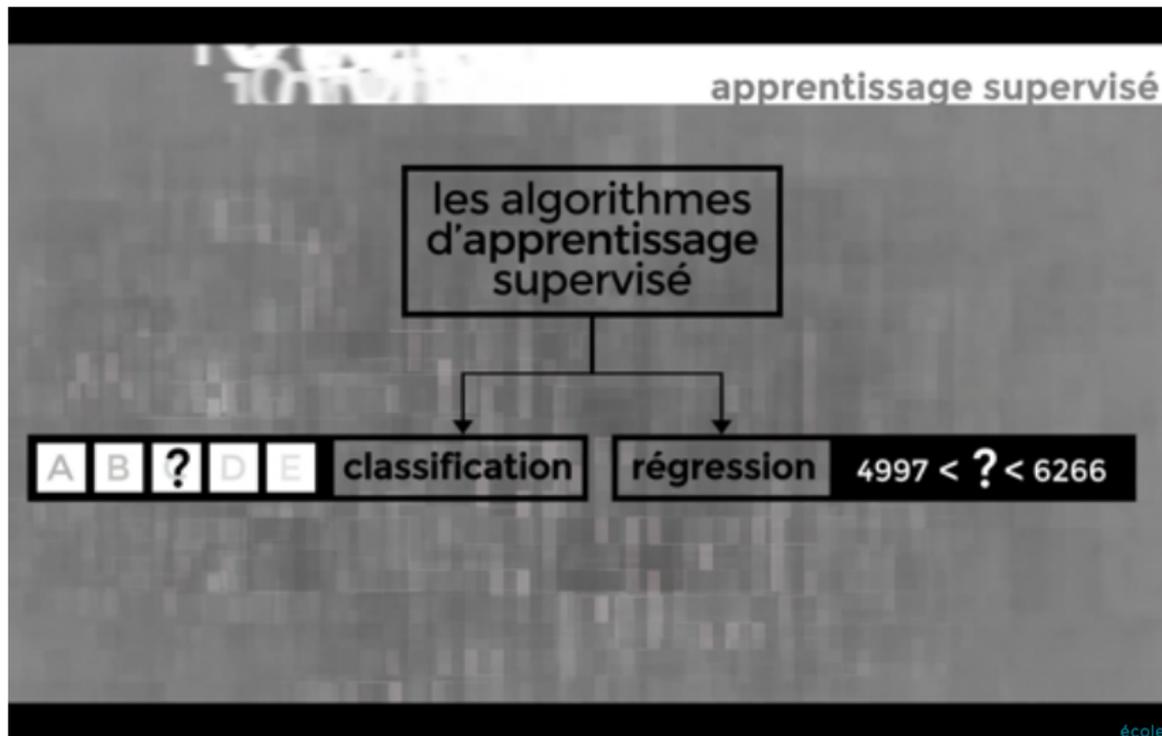
- d'une question (un critère)
- de données (observées) auxquelles poser cette question
- des réponses correctes pour ces données

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

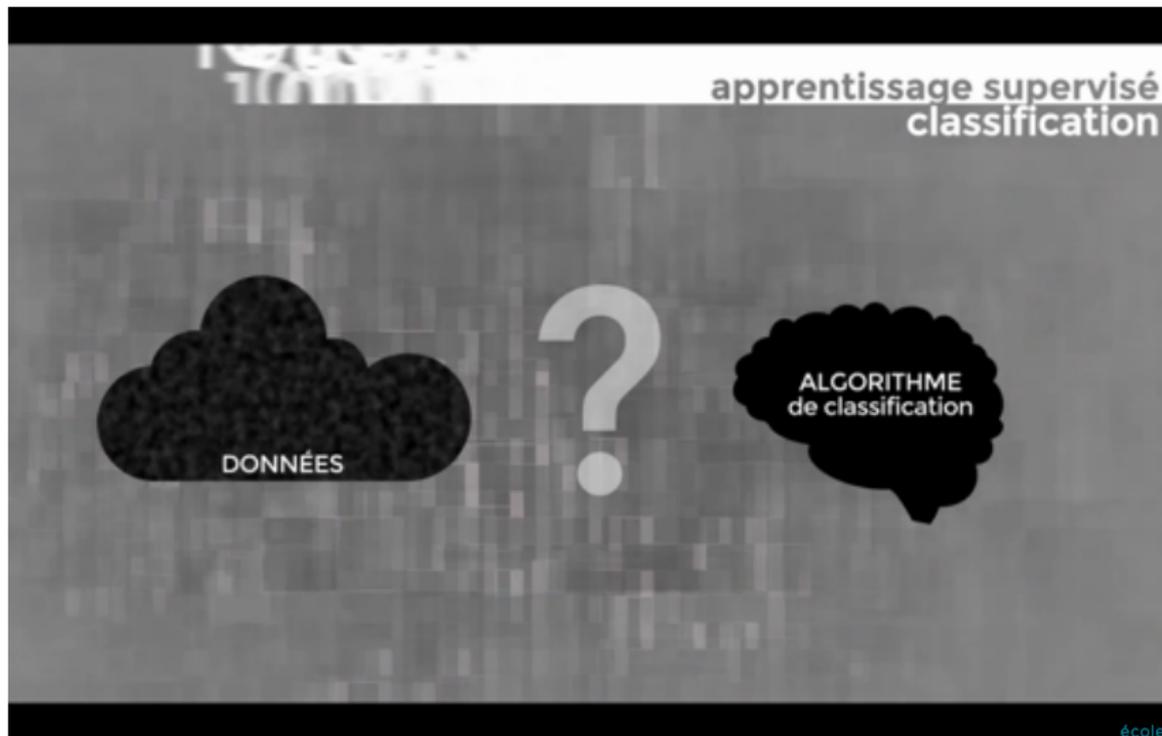
Exemple de problème à résoudre :

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

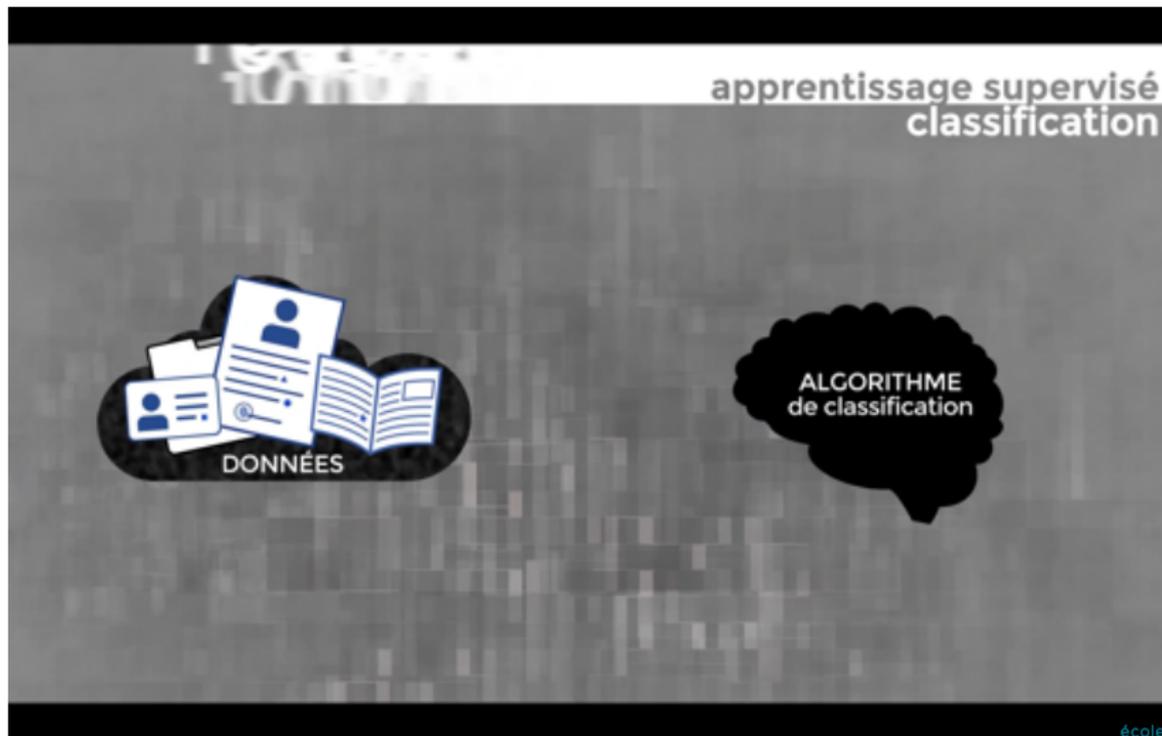
Apprentissage supervisé



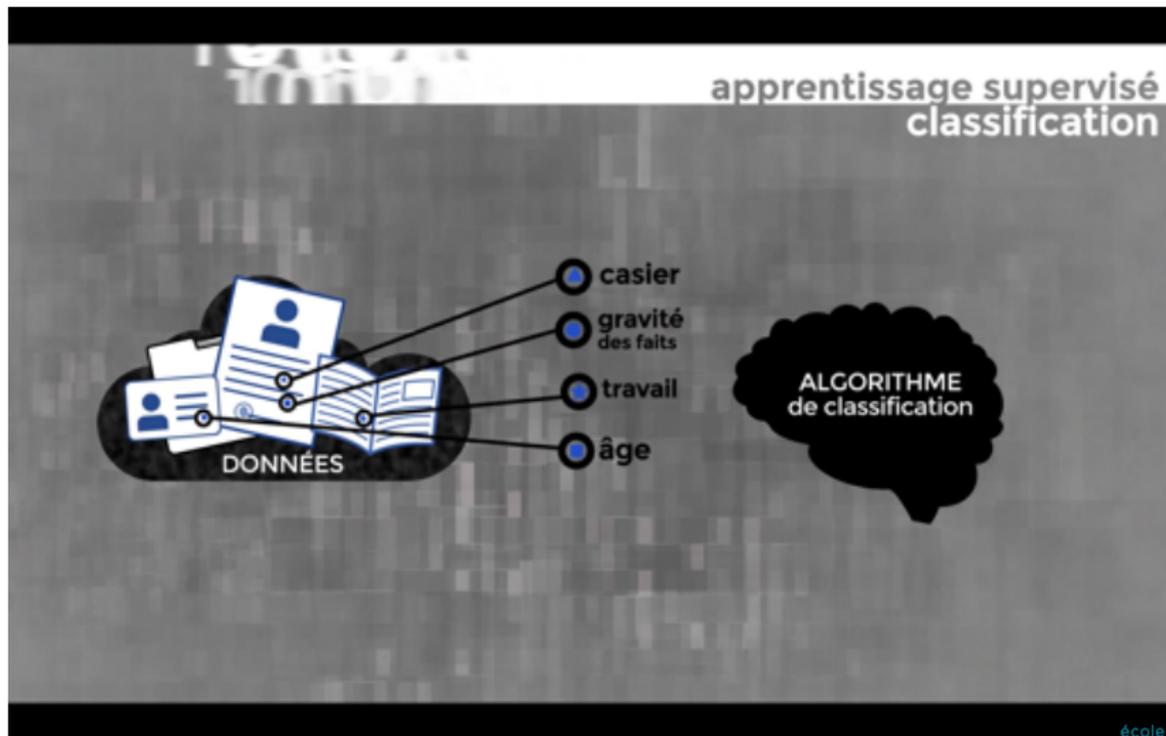
Classification



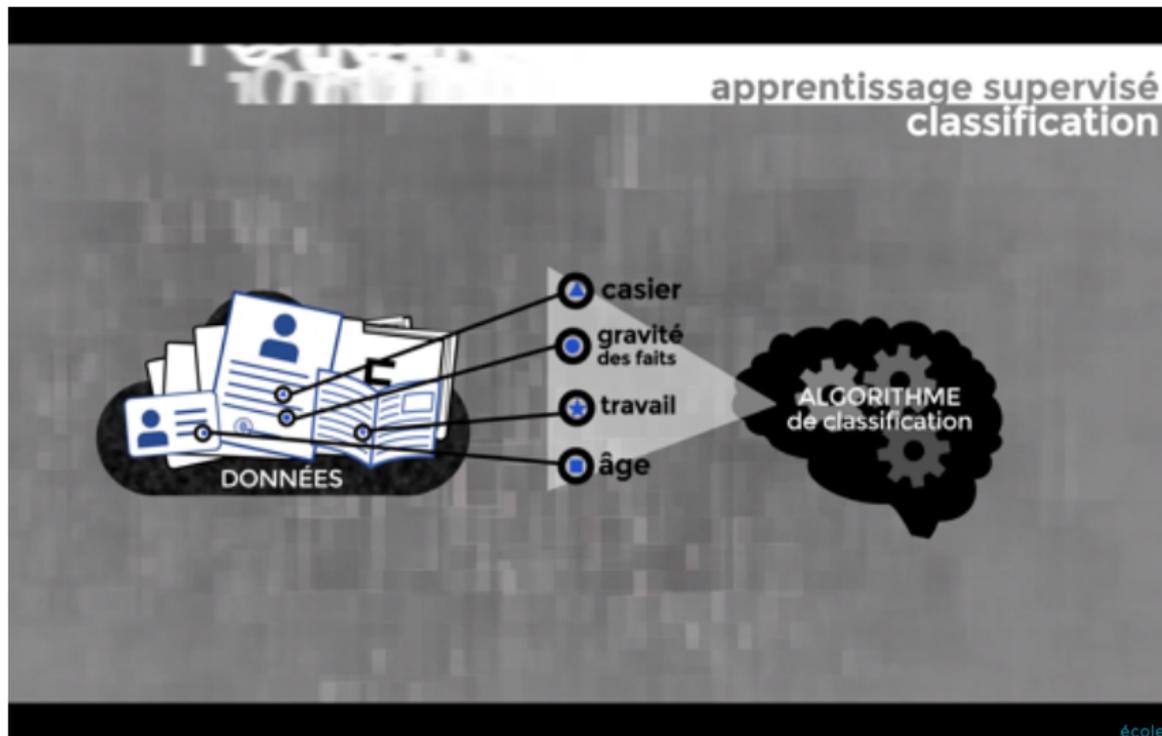
Classification



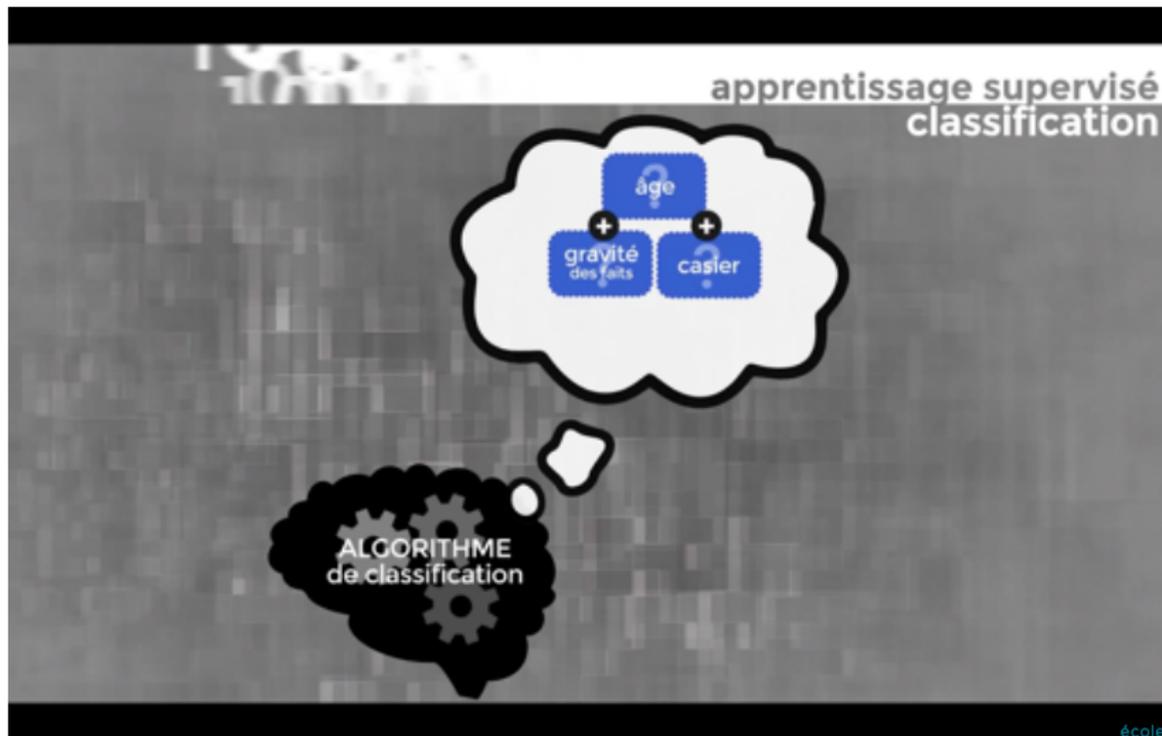
Classification



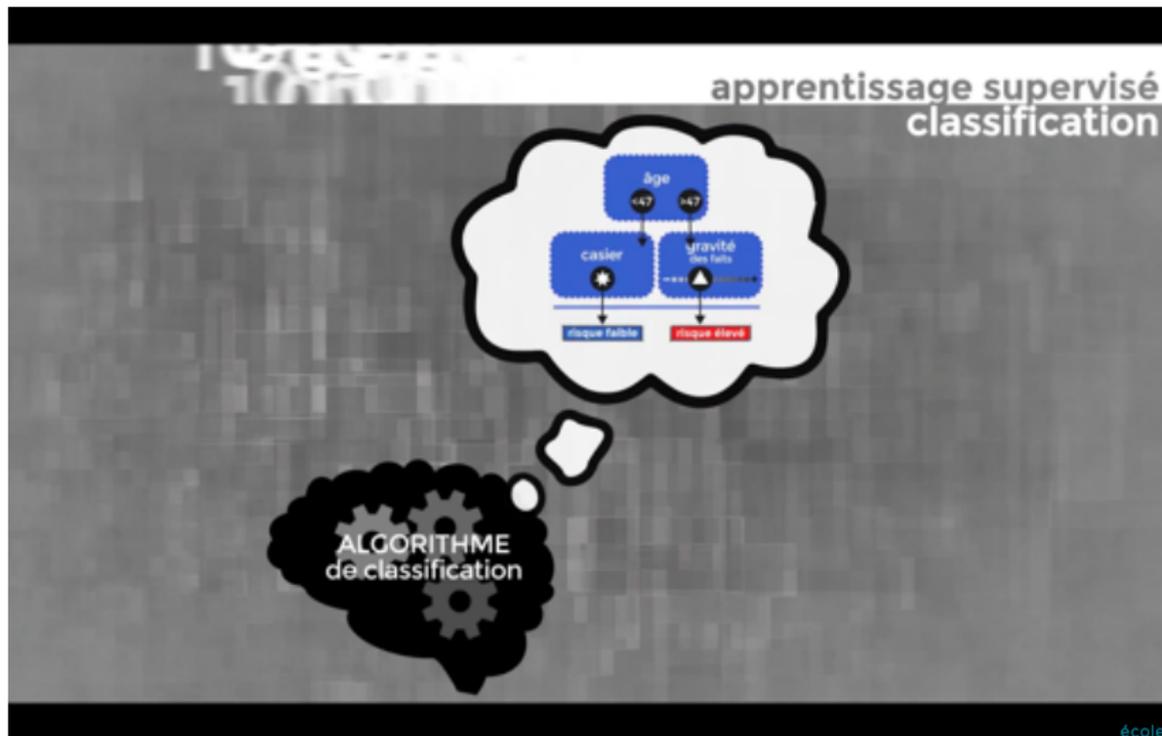
Classification



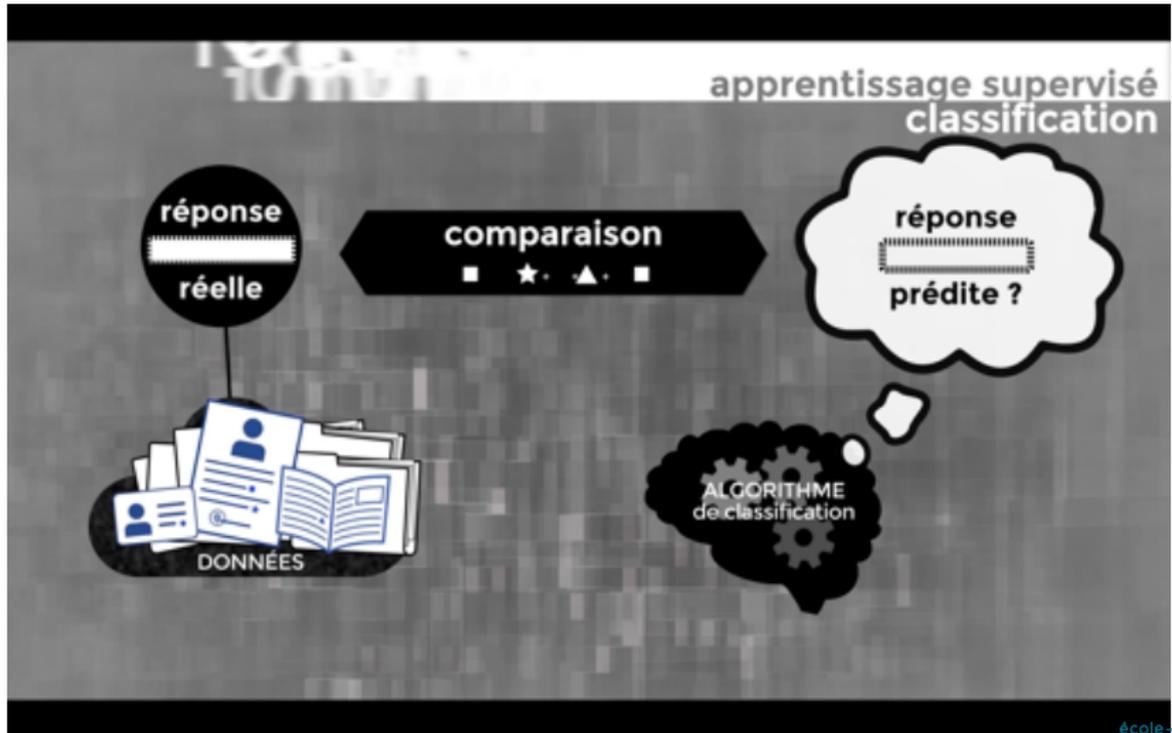
Classification



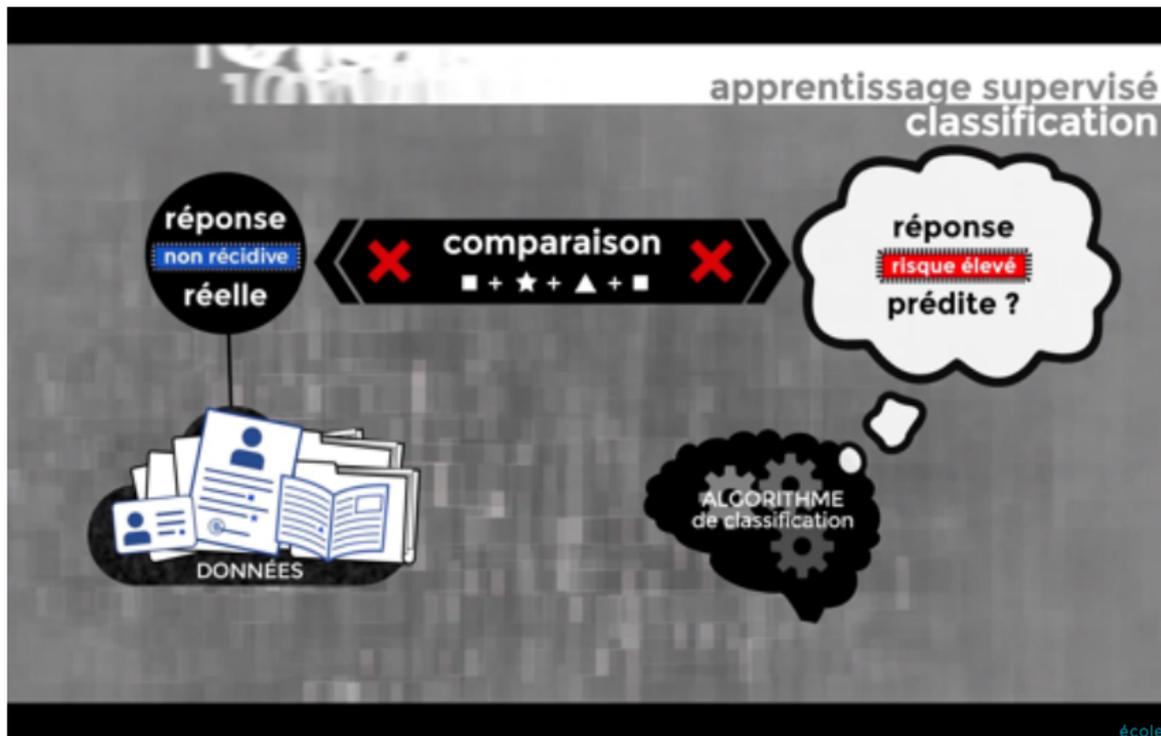
Classification



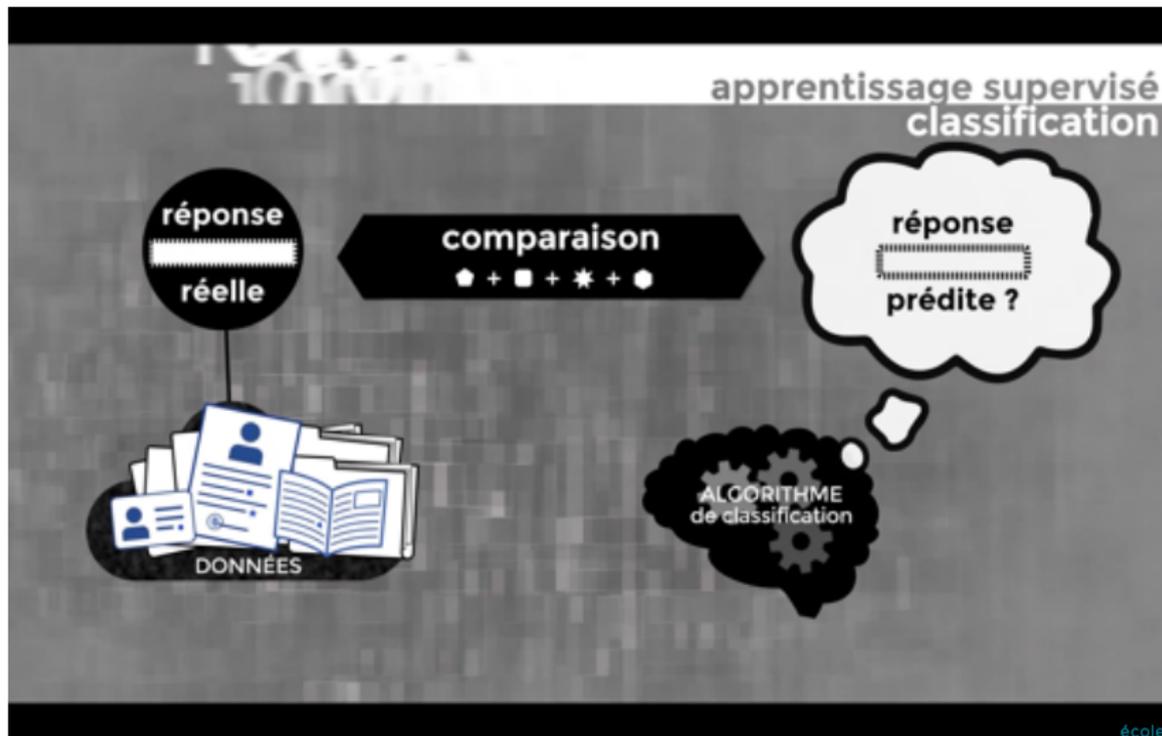
Classification



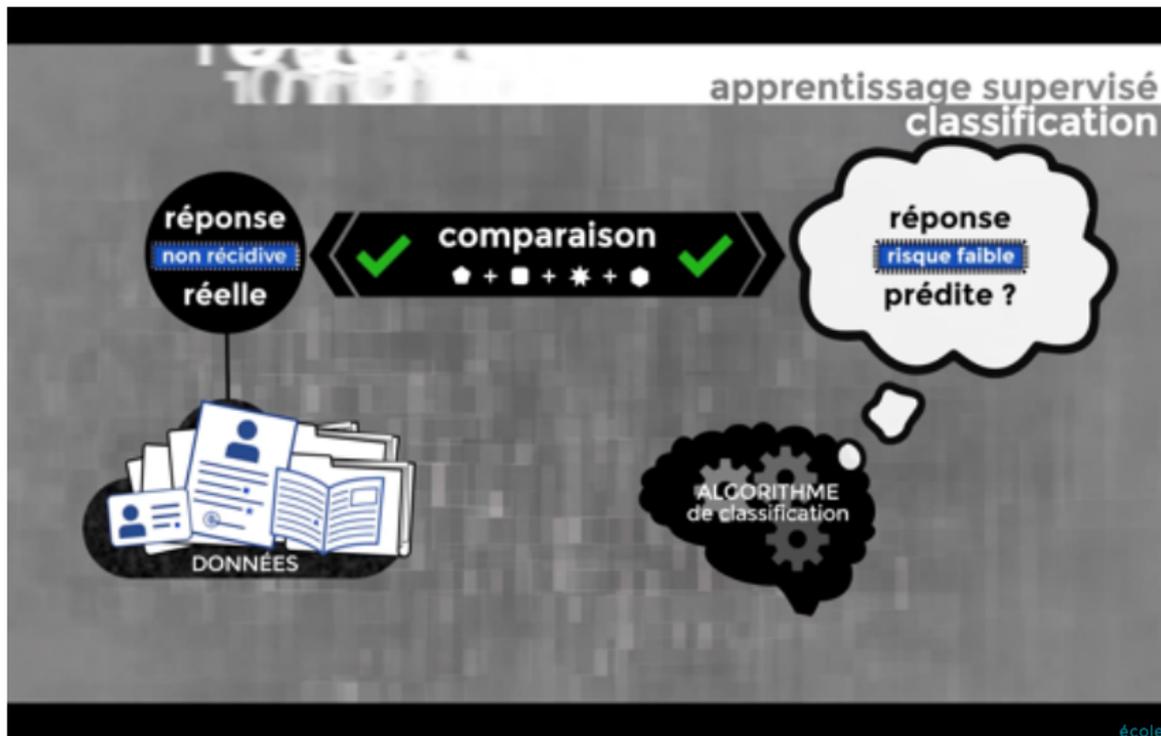
Classification



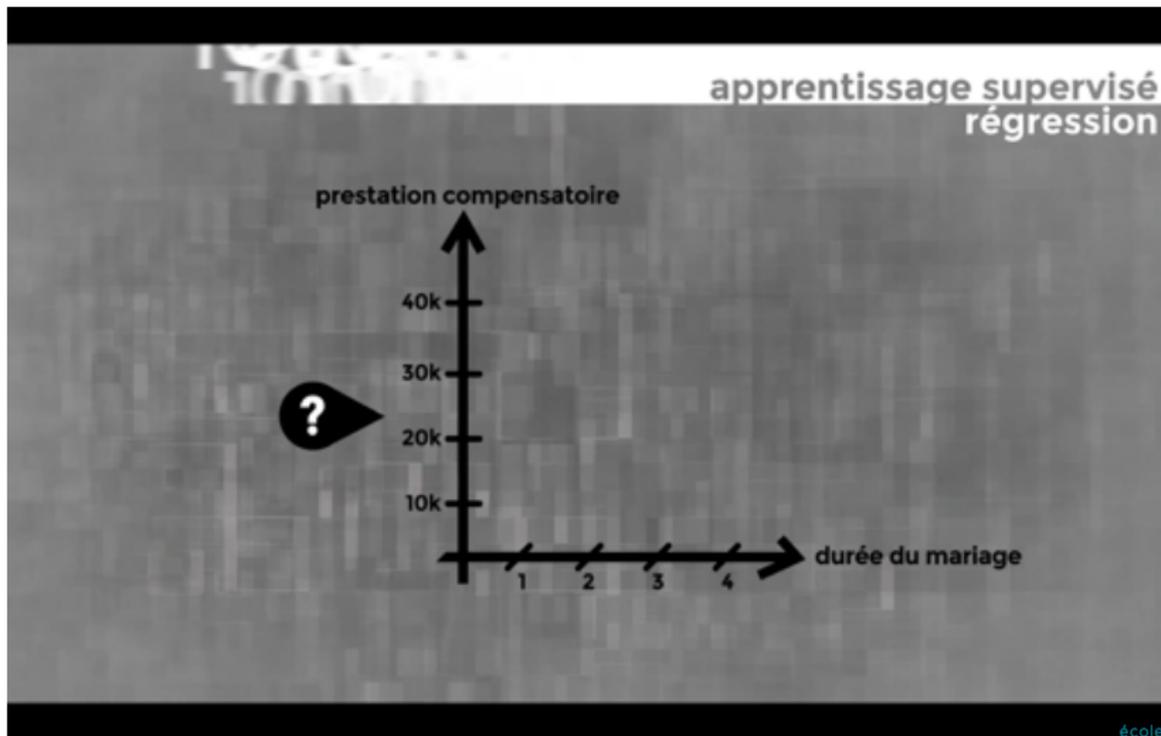
Classification



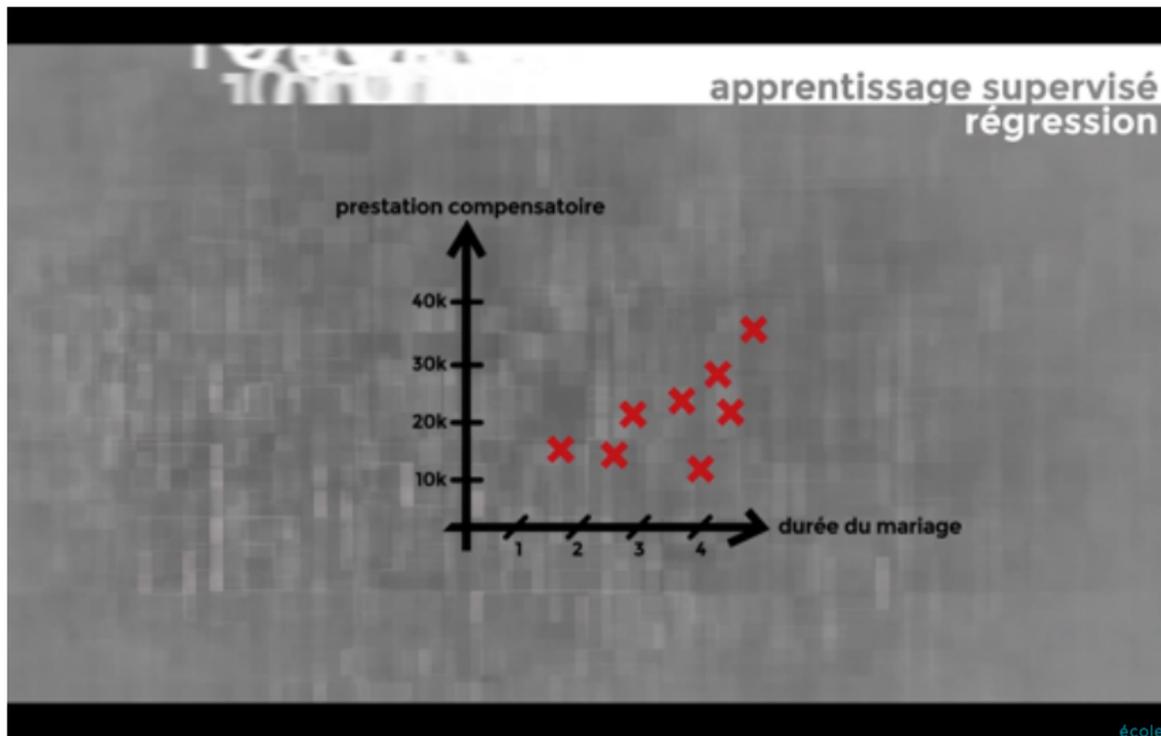
Classification



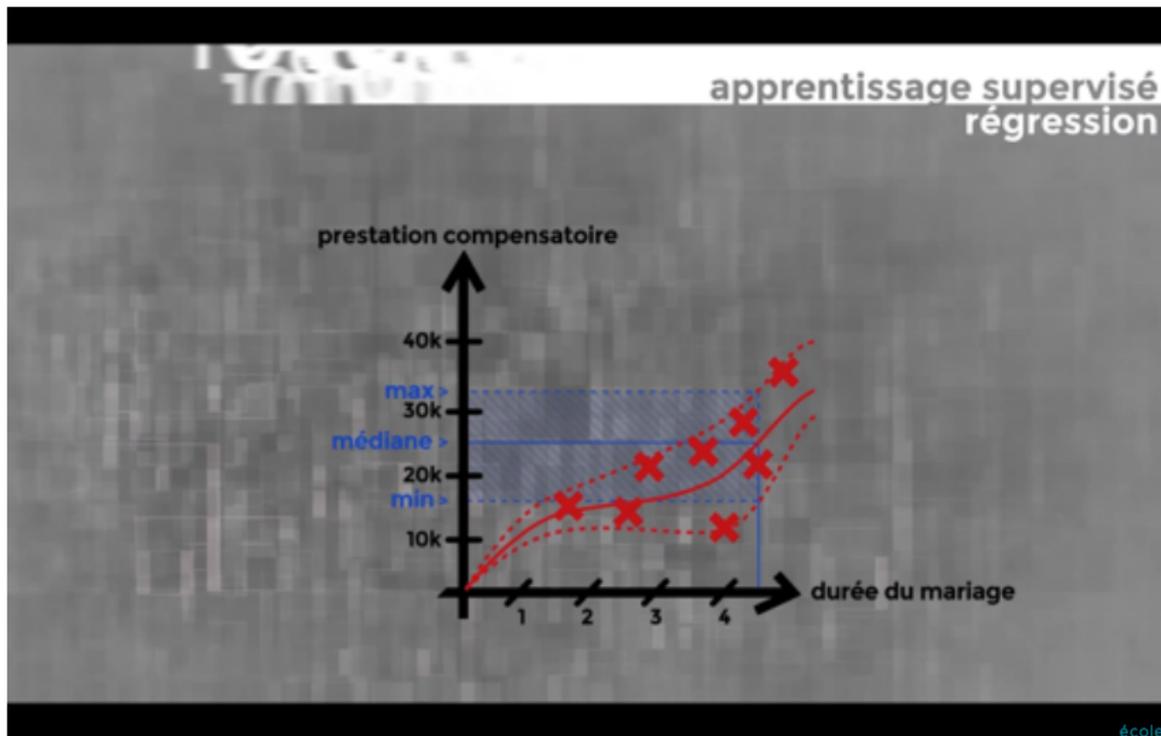
Régressions



Régressions

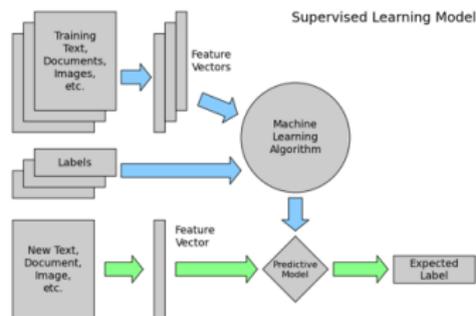


Régressions



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



⇒ 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)

*L'apprentissage dans le
contexte juridique*

Différents modèles (économiques)

États-Unis

Évaluation du risque lié à une liberté provisoire, liberté conditionnelle, ...
→ logiciels utilisés dans les **juridictions**

- COMPAS (2010 ?) : *"Risk assessment tool for criminal justice practitioner"*

France

Chance de succès d'une affaire, montants (compensatoires, dommages et intérêts, ...)
→ 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**
→ **chatbot** / mise en relation clients ↔ 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.

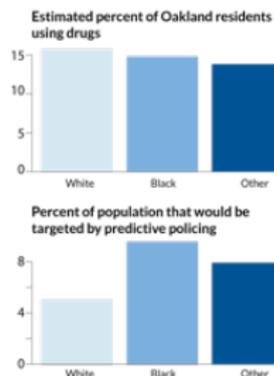
- Problème : y a-t-il violation d'un article de la Convention ?
- Résultats : prédiction correcte dans 79 % des affaires (**plutôt 64 % en fait**)

Quels biais ? Quelle précision ?

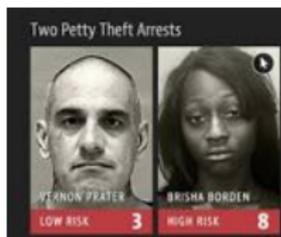
Cas de PREDPOL (Predictive Police)

- 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)



Quels biais ? Quelle précision ?



Évaluation de COMPAS (effectuée par PROPUBLICA)

Plusieurs études sur ses biais et son efficacité

- faux négatifs** Afro-américains qui n'ont pas commis de crimes sont plus souvent catégorisés à **haut risque** que les blancs (45 % vs. 23 %)
- faux positifs** Blancs qui ont commis un crime sont plus souvent catégorisés à **faible risque** que les afro-américains (48 % vs. 28 %)
- Prédictions** COMPAS vs. humains non juristes : équivalent (67 % vs. 65 %)

Évaluation de CASECRUNCH (effectuée par ... CASECRUNCH)

Prédictions CASECRUNCH vs. avocats : meilleur (86.6 % vs. 62.3 %)

Discussion

Quelques pistes à creuser

- Comment les juges utilisent ces outils ? (affaire *Paul Zilly*)
- Quelles recours ? (affaire *State vs. Loomis*)
- Impressions contrastées de ceux qui pratiquent ces outils (avocats vs. magistrats)

Réflexions en vrac

- Décision de critères pertinents : Qui décide ? Avec quelle autorité ? Y a-t-il des biais ?
- Biais dans les données d'apprentissage.
- Concept d'*automation bias* (et *moral buffer*, ...)
- Expérience ENM ...
- Question des nouvelles données.
- Corrélation n'est pas causalité.
- Cas des tribunaux vs. cabinets d'avocats.
- Outils présentés comme utiles au niveau **individuel** mais basé sur un calcul **global**

Questions?

<http://tarissan.complexnetworks.fr/>