



L'ANALYSE

Sachons concilier innovation et justice

Bruno Biais,

directeur de recherche, *Toulouse School of Economics* et CNRSPLUS SUR LE SITE WEB <http://www.agefi.fr/>

LES TECHNIQUES DE MACHINE LEARNING permettent de mieux prédire le risque de défaut. Supposons qu'on veuille établir un « rating ». Pour prendre un exemple simplifié, considérons trois classes de risque : dans la classe A figurent les ménages à risques de défaut faible, dans la classe B les risques moyens, dans la classe C les risques élevés. Les techniques de *machine learning* permettent d'utiliser de très grandes bases de données pour déterminer l'allocation des individus aux différentes classes en fonction d'un certain nombre de caractéristiques observables. Le prêteur dispose ainsi d'une analyse plus fine et plus précise du risque des emprunteurs potentiels.

Quelles seront les conséquences pour les emprunteurs ? Un article récent, du département de recherche de la Réserve fédérale américaine (Fed), intitulé « Predictably Unequal ? The Effects of Machine Learning on Credit Markets », apporte un éclairage intéressant sur la question.

Une évaluation plus précise du risque de défaut réduit les problèmes d'asymétrie d'information entre prêteur et emprunteur. Conformément aux analyses de George Akerlof (dans l'article fondateur « The Market for Lemons »), cela augmente le volume des prêts, grâce à une amélioration de l'accès au crédit pour les bons risques. Le revers de la médaille est un moindre accès au crédit pour les emprunteurs potentiels dont le risque de défaut est plus élevé.

CLASSES DE RISQUES

Les conséquences du *machine learning* ne sont donc pas les mêmes pour tous. Certains ménages qui, sans cette technologie, n'auraient pas accès au crédit y auront désormais droit. D'autres ménages, au contraire, se verront exclus. Cette évolution pose des problèmes éthiques. L'exclusion du crédit qu'implique pour certains l'utilisation de ces nouvelles techniques est-elle juste et socialement acceptable ?

Cette question prend un tour particulièrement alarmant en ce qui concerne l'information touchant l'appartenance ethnique ou raciale. Aujourd'hui, la loi interdit de prendre en compte de tels critères pour décider ou non d'accorder un crédit. C'est une bonne chose, au nom de l'égalité



«
L'EXCLUSION
DU CRÉDIT
QU'IMPLIQUE
POUR CERTAINS
L'UTILISATION
DE NOUVELLES
TECHNIQUES
EST-ELLE JUSTE
ET SOCIALEMENT
ACCEPTABLE ?
»

et de la justice. Mais les techniques du *big data*, en utilisant de manière approfondie des ensembles de données considérables, pourraient construire des algorithmes qui, effectivement, corrèleraient la décision d'attribution de crédit à la race, même si la base de données utilisée ne contenait aucune information explicite concernant l'appartenance ethnique ou raciale.

Dans l'article mentionné plus haut, les chercheurs de la Fed utilisent une grande base de données, contenant de l'information sur 10 millions de prêts hypothécaires accordés aux États-Unis entre 2009 et 2013. L'usage de ces données et des techniques de *machine learning* permet aux chercheurs de prédire la probabilité de défaut. Ils simulent ensuite des décisions de prêts, en supposant que ceux-ci sont accordés si la probabilité de défaut est inférieure à un seuil donné. Par rapport à une technique classique de *scoring* de crédit, l'analyse en termes de *machine learning* donne des évaluations plus précises. Cela amène à distinguer plus finement les différents emprunteurs potentiels. Pour reprendre l'exemple des trois classes de risques (A, B et C) mentionné plus haut, de nombreux emprunteurs classés B par des techniques traditionnelles se voient promus en A ou dégradés en C par les techniques de *machine learning*.

MAJORITAIRES ET MINORITAIRES

Alors que les algorithmes construits par une technique de *machine learning* n'utilisent pas de variable concernant l'appartenance ethnique ou raciale, les chercheurs de la banque centrale américaine constatent que les emprunteurs appartenant à l'ethnie majoritaire bénéficient de meilleurs *ratings* en comparaison de la méthode de *scoring* traditionnelle. A l'inverse, les emprunteurs des autres ethnies se voient attribuer de moins bons *ratings* par la technique de *machine learning*. En fait, si l'on détaille plus précisément les résultats, certains, parmi les emprunteurs issus des minorités ethniques, atteignent de meilleurs *ratings* par le *machine learning* que par le *scoring* traditionnel, mais un plus grand nombre sont dégradés.

Les pouvoirs politiques et les citoyens doivent être conscients des questions d'éthique et de société posées par le développement de nouvelles technologies de l'information. Il est vital aujourd'hui d'éviter une aggravation de l'exclusion et des inégalités. Sachons concilier l'innovation technologique et le souci de la justice. ■