

Concrétiser l'utilisation du datamining pour lutter contre la discrimination sur le marché du travail

Résumé

Suite à l'ordonnance du 16 novembre 2017, l'inspection régionale de l'emploi à Bruxelles (IRE/GWI) s'est vue confier la mission de réaliser des tests de situation dans le cadre d'infractions aux ordonnances anti-discrimination du 4 septembre 2008. Ceci faisait suite au constat de la difficulté d'apporter des preuves de pratiques discriminatoires sur le marché du travail. Ces tests sont étroitement cadrés par l'ordonnance : ils ne peuvent intervenir que dans le cas où des indices sérieux et un signalement laissent entendre l'existence de discrimination.

Cette recommandation développe la possibilité d'utiliser les techniques de datamining en tant qu'indice sérieux de pratiques discriminatoires. Le datamining aboutit à la création de témoins d'alertes correspondant à des indices de discriminations potentielles permettant des contrôles ciblés (cf. point 1 et 2).

Nous proposons plusieurs manières de faire du datamining. D'une part du datamining *stricto sensu*, nécessitant la création d'algorithmes dont l'objectif est de chercher de manière systématique dans les bases de données des comportements pouvant correspondre à des pratiques potentiellement discriminatoires de la part des entreprises (cf. point 3 et 4). D'autre part, des possibilités intermédiaires de techniques de datamining existent également, telle que par exemple des analyses permettant de comparer la composition du personnel d'une entreprise avec la composition moyenne des salariés du secteur et l'identification des entreprises qui diffèrent fortement de cette moyenne. Une meilleure exploitation des bases de données des contrôles déjà réalisés ainsi que des signalements rapportés peut aussi constituer une possibilité de mettre en œuvre un datamining « simplifié » (cf. section 5).

Unia plaide pour la mise en œuvre de moyens financiers et humains pour développer les techniques de datamining. Ceci permettra d'évoluer d'un mode de contrôle réactif qui a montré ses limites, telles qu'illustrées dans le récent rapport d'Unia relatif aux tests de situation en région bruxelloise, vers une stratégie de ciblage des contrôles sur la base de témoins d'alerte, correspondant à des indices objectifs de discriminations potentielles. Le travail concernant le datamining, réalisé au sein des services d'inspections tant fédéral que régional bruxellois est complémentaire. Unia recommande la mise en place d'un partenariat entre le CLS/TSW et l'IRE/GIW pour réaliser un projet de datamining conjoint, permettant de mettre en commun les bases de données fédérales et régionales, les avancées techniques en termes d'analyse datamining et l'expertise des contrôles. D'autres acteurs, détenant des données ou de l'expertise devraient être alliés à ce projet conjoint, par exemple, l'ONSS, Unia, l'Institut pour l'égalité entre les femmes et les hommes, Actiris, etc. Ce projet innovant permettra aux actions des différents acteurs de se renforcer mutuellement et d'offrir une vraie réponse pour la lutte contre la discrimination comme priorité politique en Région de Bruxelles-Capitale.

Contenu

1	Qu'est-ce que le datamining ?.....	1
2	Le datamining et les services d'inspection sociale: des fraudes sociales et fiscales à la discrimination	1
3	Mettre les bases de données en lien avec l'expertise « discrimination »	1
4	Des scénarios de discrimination concrets.....	2
5	Du « datamining » simplifié	4
6	Recommandations	5

1 Qu'est-ce que le datamining ?

Le datamining (aussi parfois appelé « exploration de données » en français) désigne tout une série d'outils statistiques et informatiques qui permettent l'analyse d'une grande quantité de données. L'objectif du datamining est de faire apparaître des liens dans les données et des dynamiques récurrentes qui ne sont pas visibles directement quand on considère la masse de données dans son ensemble. Il s'agit donc de révéler de l'information, contenue dans la masse de données. Cette information peut ensuite aider à prendre des décisions.

2 Le datamining et les services d'inspection sociale : des fraudes sociales et fiscales à la discrimination

Le datamining est déjà utilisé de manière satisfaisante au niveau fédéral en Belgique pour combattre les fraudes sociales et fiscales. En mettant en lien des données tirées des milliers de contrôles déjà effectués par l'Inspection sociale et des données « sociales » concernant les entreprises et leurs salariés, le datamining a permis de repérer les entreprises ou secteurs à risques pour la fraude sociale, permettant ainsi de cibler les contrôles et de découvrir plus d'infractions. En matière de fraude fiscale, l'inspection spéciale des impôts a été une des pionnières en Europe dans l'utilisation du datamining pour analyser des réseaux frauduleux, rapportant, dès 2014, près d'un milliard aux contributions publiques¹. A titre d'exemple, le datamining a permis au fisc de traiter des millions de données issues des Offshore Leaks pour détecter et documenter l'étendue des patrimoines soustraits à l'impôt.

Le même type de raisonnement peut être appliqué à la recherche d'infractions concernant la discrimination. Le datamining met en avant des indices indiquant des pratiques potentiellement discriminatoires au sein de secteurs spécifiques, des entreprises privées, des organismes publics ou chez des intermédiaires de l'emploi. Ces indices constitueront donc des témoins d'alertes pour le service d'inspection (Contrôle des lois sociales, inspection régionale de l'emploi (IRE/GWI), etc.), permettant des investigations et des contrôles plus ciblés. En d'autres termes, il s'agit d'optimiser la stratégie de ciblage, d'indiquer où il faut chercher les situations de discriminations, pour augmenter la pertinence des contrôles. L'exploitation des données permet de ne viser que les entreprises ou les instances publiques dans lesquelles les données agrégées révèlent des indices de discrimination potentielle. Le datamining ne suffira pas, seul, à conclure à l'existence d'une discrimination, mais il pourrait contribuer à la création d'un faisceau d'indices établissant l'existence d'une discrimination.

3 Mettre les bases de données en lien avec l'expertise « discrimination »

Les services d'inspection disposent de riches bases de données administratives, notamment nourries par les employeurs, dans le cadre de leurs devoirs administratifs. Ces bases de données contiennent des informations sur les entreprises (taille, secteur, etc.) et sur les personnes qui y travaillent, leur statut et durée d'emploi ainsi que certaines de leur caractéristiques personnelles (le sexe, l'âge, la nationalité, le domicile, etc.). Il est également techniquement possible de les coupler à d'autres bases de données administratives, issues de la Banque Carrefour de la Sécurité Sociale par exemple. Ceci permettrait notamment d'ajouter des informations sur des caractéristiques personnelles, notamment l'origine nationale de la personne (permettant d'aller au-delà de la nationalité). L'ONSS a

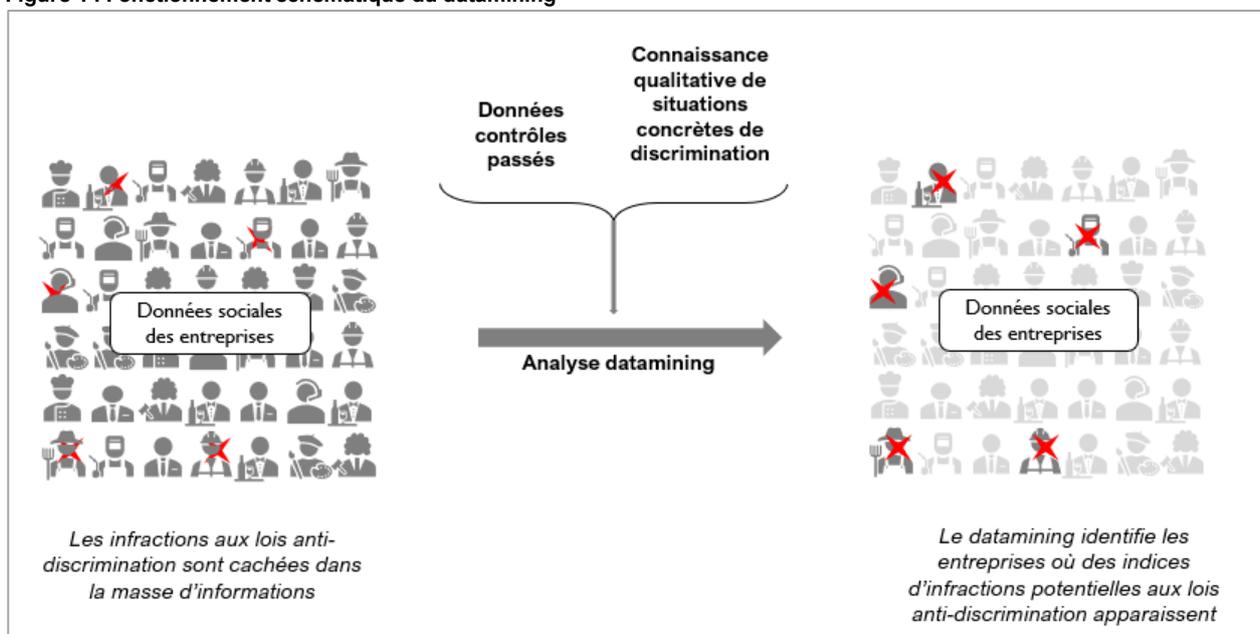
¹ Voir ici : <https://www.agoria.be/fr/Big-Data-comment-le-fisc-a-recupere-plus-d-1-milliard-d-euros-dans-la-lutte-contre-la-fraude>.

déjà accès à ce type de données et le service d'inspection CLS/TSW a également réalisé une demande d'accès à ce type de données à l'Autorité de protection des données.

Appliquer le datamining à la lutte contre la discrimination reviendrait à analyser ces bases de données pour y repérer des scénarios-types de discrimination (« patterns ») à l'aide d'algorithmes adéquats. Ces scénarios-types de discrimination doivent être élaborés en prenant en compte d'une part les informations récoltées lors des nombreux contrôles « discrimination » déjà réalisés par le Service d'inspection et d'autre part la connaissance qualitative des situations concrètes de discrimination dans le monde du travail par des experts de la discrimination, notamment dans les différents services qui récoltent et traitent des signalements et plaintes de discrimination dans le domaine de l'emploi (notamment Unia, IEFH, Service anti-discrimination d'Actiris Inclusive).

L'analyse des données selon ces scénarios-types permettra de fournir une liste d'entreprises ou d'organismes publics pour lesquels un indice de pratique discriminatoire aura été trouvé. Le datamining constitue ainsi un système de témoins d'alerte pour l'organisation de contrôles anti-discrimination ciblés, assurés par l'Inspection Régionale de l'Emploi (IRE).

Figure 1 : Fonctionnement schématique du datamining



4 Des scénarios de discrimination concrets

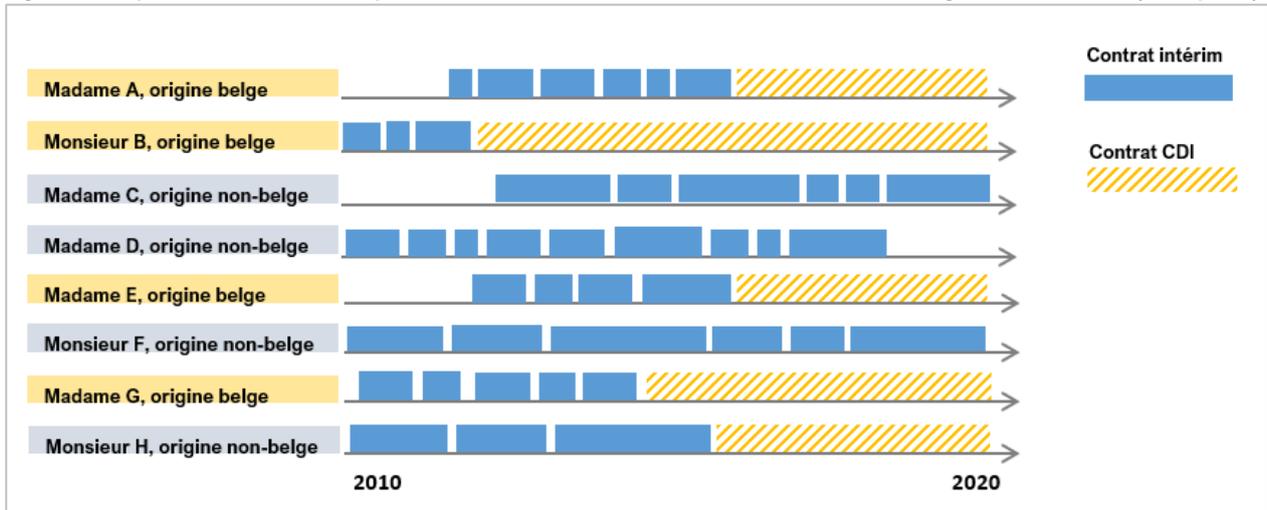
Il y a de nombreux scénarios-types qui peuvent être définis pour caractériser des profils de comportements à risque. Voici ci-dessous des exemples de scénarios de discrimination qu'il s'agit d'aller rechercher dans les données :

- 1) une entreprise qui n'embauche pas de travailleurs âgés ;
- 2) une organisation qui embauche des personnes d'origine belge et d'origine étrangère via une agence intérimaire, mais qui ne propose des emplois salariés fixes qu'aux personnes d'origine belge ;
- 3) une organisation subsidiée (> 50%) par la RBC qui licencie systématiquement des personnes après une période d'incapacité de travail ;
- 4) une entreprise qui n'engage pas de femmes de moins de 30 ans ;
- 5) une entreprise qui fait des différences de salaires entre les personnes d'origine étrangère ou d'origine belge, ou selon le sexe des salariés ;
- 6) ...

La figure 2 ci-dessous permet de visualiser l'analyse de données pour l'exemple n°2. Il s'agit de créer une base de données contenant :

- ✓ les informations sur l'entrée et la sortie de salariés dans l'entreprise (Dimona-in et -out) et le type de contrat
- ✓ les informations sur l'origine nationale des personnes (information disponible dans la BCSS)
- ✓ pour toutes les entreprises bruxelloises d'un secteur entre 2010 à 2020.

Figure 2. Représentation schématique d'un scénario de discrimination sur base de l'origine des salariés (exemple 2).



Dans cette base de données, il s'agit de rechercher les scénarios « intérim + CDI » et « uniquement intérim ». Dans notre exemple, il s'agit de 88% des salariés de l'entreprise d'origine belge qui bénéficient d'un « scénario intérim + CDI », alors qu'il s'agit de 12% pour les salariés d'origine étrangère (= Monsieur H). Cette différence de proportion déclenche un témoin d'alerte, indiquant une pratique potentiellement discriminatoire dans cette entreprise. Bien entendu, ce témoin doit être interprété, il ne s'agit pas d'une indication effective de discrimination. Dans cet exemple, nous avons simplifié les profils. Il faut prendre en compte plus d'information avant de comparer les groupes de salariés (niveau de diplômes, ancienneté, etc.).

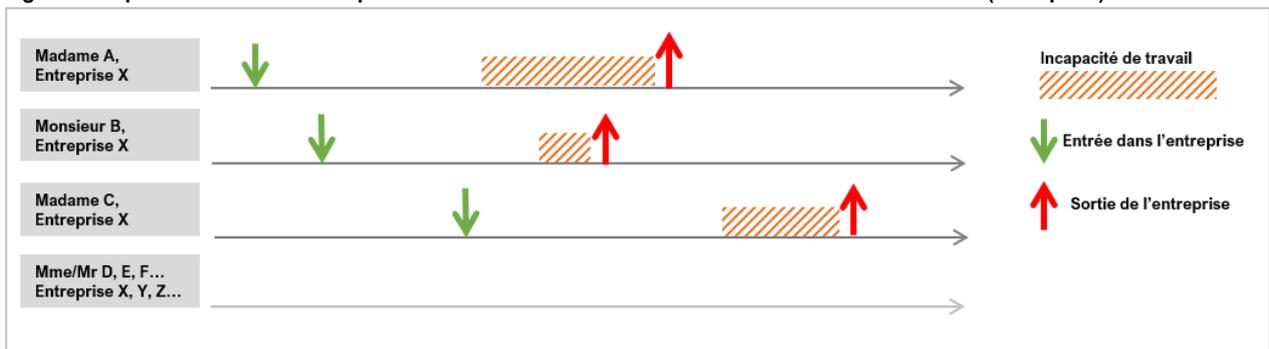
Le schéma ci-dessous permet de rendre visuellement l'action du datamining en ce qui concerne l'exemple n°3. Dans cet exemple, on exploite une base de données contenant les informations suivantes :

- ✓ sur l'entrée et la sortie de l'entreprise (Dimona-in et out)
- ✓ les périodes d'incapacité de travail,
- ✓ pour toutes les entreprises bruxelloises d'un secteur de 2010 à 2020

Dans cette masse de données, il s'agit de repérer les comportements individuels suivants : des travailleurs qui quittent une entreprise rapidement après une période d'invalidité. On appellera ce comportement le scénario-type « période d'invalidité + départ ». Ainsi, la puissance d'analyse du datamining permet de considérer en une fois le parcours de tous les travailleurs des entreprises d'un secteur, et de voir si ce scénario-type « période d'invalidité + départ » est présent de manière répétée pour plusieurs travailleurs au sein d'une entreprise.

Ceci se fera par la création d'un algorithme permettant de repérer dans la base de données ce scénario-type « période d'invalidité + départ ». L'algorithme indiquera les entreprises dans lesquelles ce scénario aura été repéré, de manière répétée ou systématique et cela constituera « un témoin d'alerte » indiquant un profil de comportement à risque pour l'organisme. Ces signaux devront être interprétés par l'inspection régionale et cela permettra de décider d'un contrôle ciblé pour l'entreprise visée par le signal.

Figure 3. Représentation schématique d'un scénario de discrimination sur base de l'état de santé (exemple 3).



L'étape de l'interprétation des témoins d'alerte par l'inspection régionale est importante, dans la mesure où il s'agit d'une indication de profil de comportement à risque et non un constat de discrimination effective. Ainsi, dans ce cas concret, il faudra pouvoir comparer le comportement de l'entreprise visée par le témoin d'alerte au comportement moyen d'entreprises comparables (secteur, taille, ...).

5 Du « datamining » simplifié

Cette dernière démarche peut également constituer une manière « simplifiée » d'utiliser certains aspects du datamining. En effet, les bases de données permettent d'avoir des informations sur le profil socio-démographique de la population de salariés de chaque entreprise, par secteur. On peut donc définir un « profil moyen » des salariés d'un secteur et simplement comparer la population des salariés des entreprises pour déterminer quelles sont les entreprises qui diffèrent fortement du « profil » moyen des entreprises de ce secteur.

Voici un exemple concret : si l'on s'intéresse à la discrimination sur base de l'origine nationale. On peut pour un secteur déterminé calculer la proportion moyenne de salariés d'origine belge et d'origine non-belge (éventuellement en faisant une distinction UE/non-UE). Ensuite, on compare ces proportions moyennes avec les proportions dans chaque entreprise de ce secteur. Cette comparaison a pour but de repérer les entreprises dont la proportion de salariés d'origine belge est très différente de la moyenne (très élevée), ce qui est une première indication d'une situation potentiellement problématique. Ces comparaisons peuvent également se faire pour la proportion d'hommes et de femmes, ou la proportion de différents groupes d'âge (salariés de 50 ans et plus, par exemple).

La pertinence des résultats de cette analyse sera déterminée par la pertinence du groupe d'entreprises sélectionnées. En d'autres termes, il faut pouvoir comparer des entreprises comparables, en termes de taille par exemple, ou de type d'activité, ou de temporalité d'emploi (en cas de travail saisonnier par exemple). L'analyse de la comparaison est également importante : à partir de quelle ampleur la différence entre la proportion moyenne et la proportion dans l'entreprise peut-elle être considérée comme indicative d'un potentiel problème ? Ceci devra faire l'objet de discussions et de tests afin de parvenir à un outil pertinent d'aide à la décision.

L'exploitation des bases de données des contrôles déjà réalisés et des listes d'entreprises ayant déjà fait l'objet de plaintes pour de discrimination est également une piste à explorer pour nourrir les analyses de datamining.

6 Recommandations

Dans la lignée de sa recommandation « Recherche et surveillance en matière de discrimination dans les relations de travail par l'inspection fédérale du travail » de mars 2017², Unia plaide pour la mise en œuvre de moyens financiers et humains pour développer les techniques de datamining. Ceci permettra d'évoluer d'un mode de contrôles réactifs qui a montré ses limites, telles qu'illustrées dans le récent rapport d'Unia relatif aux tests de situation en région bruxelloise, vers une stratégie de ciblage des contrôles sur base de témoins d'alerte, correspondant à des indices objectifs de discriminations potentielles.

Le travail concernant le datamining réalisé au sein des services d'inspections tant fédéral que régional bruxellois est complémentaire. Unia recommande la mise en place d'un partenariat entre le CLS/TSW et l'IRE/GIW pour réaliser un projet de datamining conjoint, permettant de mettre en commun les bases de données, les avancées techniques en termes d'analyse datamining et l'expertise des contrôle. D'autres acteurs, détenant des données ou de l'expertise doivent être alliés à ce projet conjoint : l'ONSS, Unia, l'Institut pour l'égalité entre les femmes et les hommes et Actiris et view.brussels (cette liste n'est pas exhaustive). Un tel partenariat permettra aux acteurs de se renforcer mutuellement et d'offrir une vraie réponse pour la lutte contre la discrimination comme priorité politique en Région de Bruxelles-Capitale.

Il est utile de noter que l'ONSS et le service d'inspection sociale fédérale -à l'initiative du ministre de l'emploi du précédent gouvernement- ont déjà entamé la réflexion ainsi que des travaux techniques permettant d'appliquer le datamining à la recherche d'indices de discrimination. Un partenariat serait donc certainement bénéfique aux travaux de la Région Bruxelles-Capitale à ce sujet.

Concrètement, cette note de recommandation contient plusieurs propositions de pistes de travail pour réaliser des analyses de datamining. En ce qui concerne les techniques de datamining les plus poussées il sera nécessaire de développer 1) des scénarios-types de discrimination, 2) modélisables par des algorithmes que 3) l'on peut faire opérer dans des bases de données existantes (ou à créer). La technicité de ce programme est importante, ce qui requiert la mobilisation de savoir-faire spécialisés et sans doute une phase d'expérimentation. C'est pourquoi Unia recommande ce partenariat entre les spécialistes utilisant déjà le datamining et les acteurs de la lutte contre les discriminations pour concevoir et tester des algorithmes pertinents.

Par ailleurs, le SIRS (Service d'information et de recherche sociale), en tant qu'organe de coordination qui soutient les services d'inspection sociale dans leur lutte contre la fraude sociale, pourrait être également un acteur partenaire dans le cadre du datamining contre la discrimination. A ce titre, c'est un interlocuteur de choix pour recevoir des informations et des bonnes pratiques. Enfin des possibilités d'information et d'offre de service pour parvenir à faire des analyses de datamining existent également au sein de SMALS, en tant qu'ASBL chargée de soutenir et seconder la gestion ICT des organismes publics.

Contact Unia

Bruno Martens (bruno.martens@unia.be)

Louise Callier (louise.callier@unia.be)

²https://www.unia.be/files/Documenten/Aanbevelingen-advies/FED_FR_-_165_-_Recherche_et_contr%C3%B4le_par_les_services_dinspection_du_travail.pdf