

Veillez trouver ci-dessous une demande d'accès aux documents administratifs à la CNAF qui vous est adressée en votre qualité de DPO de cette institution.

Comme le livre III du code des relations entre le public et l'administration le prévoit si nous avons mal identifié la personne qui est susceptible de répondre à notre requête, nous vous prions de bien vouloir la transmettre au service/personne à même d'y répondre.

En application de la loi n° 78-575 du 17 juillet 1978 relative aux documents administratifs, l'association "La Quadrature du Net" souhaite recevoir communication des documents administratifs relatifs à l'algorithme de scoring (profilage) utilisé par la CAF dans le cadre de sa politique de contrôle.

Pour préciser la demande, nous nous intéressons ici à l'algorithme de calcul de "scores de risque" utilisé depuis une dizaine d'années par la CNAF dans le cadre de sa politique de contrôle.

Nous tenons à préciser que nous avons pris connaissance des documents (communiqués et délibération CNIL) mis en ligne par la CNAF à ce sujet sur sa page informatique et libertés.

Le communiqué est ici: <https://www.caf.fr/sites/default/files/m> (<https://www.caf.fr/sites/default/files/m>)...

L'avis de la CNIL est lui disponible ici: <https://www.caf.fr/sites/default/files/m> (<https://www.caf.fr/sites/default/files/m>)...

L'association La Quadrature du Net considère cependant que ces deux documents ne donnent pas les informations nécessaires à une compréhension de ce traitement et de son impact pour les allocataires de la CNAF. En particulier, très peu, voire aucune, information n'est fournie sur les variables utilisées et leur pondérations statistiques.

En conséquence, l'association demande la communication des documents suivants:

a) Le détail du ou des algorithmes utilisés pour son calcul, c'est à dire le ou les tableaux de régressions statistique comprenant:

- La liste exhaustive des données (variables d'entrée) utilisées pour son calcul. Par exemple: décile de revenu, âge, fréquence d'interaction avec la CAF... Nous tenons à avoir le détail de chaque variable: nom, découpage, seuils etc de manière à être à même de comprendre clairement le modèle.
- Pour chacune de ces variables, leur impact/pondération (estimation du coefficient, écart-type et éventuellement odds ratio dans le cadre d'une régression logistique)

- Les éventuels pré traitements sur ces variables
- Les éventuels post-traitements

Le plus simple concernant ce premier point serait la transmission des tables statistiques (aussi appelées tables de résultats) des modèles utilisés et présentant la liste des variables et pour chacune d'entre elles l'estimation des coefficients, leur écart-type et leurs éventuels odds-ratio.

Un exemple d'une telle table peut être trouvé ici: <https://www.researchgate.net/figure/Summ> (<https://www.researchgate.net/figure/Summ>)...

b) Nous demandons aussi la liste exhaustive des variables utilisées lors de la phase d'entraînement du modèle, même si ces dernières n'ont pas été retenues pour le modèle final.

Ceci inclut donc l'ensemble des variables utilisées lors de la phase d'entraînement du modèle et non retenues pour le modèle final, soit parce que leur impact statistique est jugé négligeable (non significatif) soit pour une autre raison.

c) Si, comme il est probable, le modèle a évolué depuis l'année 2010, date à laquelle la délibération de la CNIL a été rendue et/ou est mis à jour de manière régulière, nous demandons aussi l'accès rétrospectif aux tables statistiques des différents modèles ayant été utilisés depuis sa création.

Pour chacun de ces modèles "historiques", nous demandons les mêmes informations qu'en a) et b) de cette demande d'accès.

d) Nous demandons par ailleurs les informations suivantes:

- La durée de conservation de ce "score de risque" ;
- La fréquence de mise à jour de ce "score de risque" ;
- La liste des éventuelles entreprises privées et/ou administrations publiques ayant contribué au développement de l'algorithme.

e) Enfin, de manière à appréhender le mieux possible les enjeux internes autour de la mise en place des techniques de "datamining" appliquées à la politique de contrôle de la CAF, nous souhaiterions nous voir communiquer les documents et échanges des groupes de travail/équipes travaillant, ou ayant travaillé, à leur développement. Ceci inclut notamment:

- L'ensemble des documents internes/de travail, et non rendus publiques, à ce sujet;
- Les comptes rendus de réunions des équipes en charge du développement de

l'algorithme;

- Les échanges avec la CNIL
- Les échanges à ce sujet avec d'autres administrations publiques et ou entreprises privées.

Nous souhaitons recevoir ces documents dans un format numérique, ouvert et réutilisable. Pour ce faire, veuillez nous indiquer leur adresse de téléchargement ou les envoyer en pièce jointe.

A défaut de réponse dans un délai d'un mois, ou dans l'hypothèse d'une réponse incomplète, nous saisirons la Commission d'accès aux documents administratifs (CADA).

Nous tenons à ajouter que nous nous tenons à votre disposition pour toute question relative à notre demande à l'adresse: [email address].