



la sécurité sociale

caf.fr



Contrôles et datamining à la Caf

Des prestations au juste droit

pour tous



Qu'est-ce que le datamining ?

Le datamining est une démarche scientifique d'étude statistique. Elle vise à déterminer des probabilités de risque pour un individu, un dossier, un traitement. Elle s'appuie sur de grandes quantités d'informations quantitatives, présentes dans le système d'informations. Ces études sont réalisées par des experts des statistiques et de l'informatique.

En résumé : c'est une démarche d'étude statistique sur de grands volumes de données, menée par des experts.



Le datamining, dans les Caf, en quoi ça consiste ?

Dans les Caf, le datamining a commencé à être utilisé en 2011, après avoir été autorisé par la Cnil, et contrairement à une idée reçue, il n'est pas construit pour lutter contre la fraude.

Il existe, au sein des Caf, trois modèles datamining qui permettent d'identifier :

1. des potentielles situations d'erreurs déclaratives ;
2. des potentielles situations de non-recours ;
3. des potentielles erreurs internes de traitement des dossiers.



Dans le premier cas, le modèle datamining, construit par des humains, ne permet que d'identifier des dossiers à analyser par d'autres humains.

Dans certains cas, ces contrôles permettent d'établir des indus, mais aussi des rappels, des erreurs, et parfois des cas de fraudes.



Le datamining, comment ça fonctionne ?

La mise en place et le fonctionnement du datamining peuvent être découpés en 3 étapes :

- création d'un modèle, par une équipe d'experts, sur la base de cas authentiques et récents d'erreurs de déclaration et de cas sans erreur ;
- application du modèle pour déterminer les dossiers nécessitant une analyse humaine ;
- analyse du dossier par un agent dument habilité pour vérifier le paiement à bons droits.



Qu'est-ce que le score de risque ?

Chaque dossier allocataire se voit attribuer par le modèle datamining un score de risque à partir duquel la Caf va :

- classer les dossiers en fonction du niveau de score, donc de l'importance du risque qu'il présente ;
- sélectionner les dossiers qui présentent les scores les plus élevés afin d'engager une action, pour sécuriser le risque, en engageant une action de contrôle ou en contactant l'allocataire pour vérifier qu'il bénéficie bien de tous ses droits.

Les modèles datamining sont des outils au service du versement de tous les droits, et rien que les droits.

Le score de risque est le fruit d'une sélection, sur seuls critères statistiques, des caractéristiques les plus fréquemment rencontrées dans ces cas d'erreur. Ces caractéristiques sont scientifiquement pondérées pour maximiser les chances de trouver une erreur lorsque le contrôle est dispensé sur les dossiers ayant les scores les plus élevés.



Comment fonctionnent les contrôles à la Caf ?

Grâce au datamining, différents dossiers avec un score de risque considéré comme important sont identifiés et des contrôles peuvent alors être diligentés par les Caf. Ceux-ci peuvent être réalisés sur pièces, ou sur place. Les contrôles sur place sont réalisés par l'un des 700 contrôleurs assermentés qui travaillent au sein des Caf pour vérifier que l'allocataire concerné perçoit bien les prestations auxquelles il a droit, et qu'ils les touchent au juste droit.

Si ce n'est pas le cas, soit l'allocataire doit rembourser les sommes trop perçues, soit il bénéficiera d'un rappel de ce qui aurait dû lui être versé.



Quelques résultats

En 2021 :

- près de 4 millions de contrôles sur pièces (dont 166 000 issus du datamining)
- 128 000 contrôles sur place (dont près de 70 % issus du datamining).
- moins de 1% de tous ces contrôles aboutissent à une suspicion de fraude.
- près de 32 millions de contrôles automatisés grâce à des échanges de données avec d'autres organismes (ex : Pôle Emploi).
- 50 % des allocataires sont contrôlés chaque année et un quart des montants régularisés sont des rappels.
- Le datamining représente 1 % des contrôles diligentés par les Caf.





Le vrai/faux du datamining

Les contrôleurs de la Caf ont accès à toutes les infos qu'ils souhaitent à l'insu des allocataires.

✘ Faux

Les contrôleurs ont principalement accès aux données que la Caf possède et qui ont donc été fournies par les allocataires. Mais la branche Famille peut aussi bénéficier d'informations émanant de tiers, de partenaires ou d'auteurs de dénonciation. Aucune demande n'est faite à d'autres administrations, sans en avoir averti, auparavant l'allocataire, aucune procédure vis-à-vis d'un tiers n'est engagée à l'insu de celui-ci.

Plus de 1 000 données concernant les allocataires sont utilisées dans le modèle de datamining des caf.

✘ Faux

Seules les données possédées par les Caf sont utilisées, il s'agit essentiellement d'informations portant sur la situation familiale des allocataires (couple ou personne isolée, situation des enfants à charge), sur la situation professionnelle (en activité, demandeur d'emploi...), sur la situation financière (revenus imposables et non imposables, avantages en nature...), ainsi que sur la résidence en France. Trois cents informations environ sont nécessaires pour instruire le droit aux différentes prestations, mais seulement une quarantaine sont utilisées pour calculer le score de risque.



Les contrôles se font uniquement par un logiciel.

 Faux

Il n'y a pas de logiciel, mais un modèle statistique construit par des experts. Le datamining permet d'octroyer à chaque allocataire un score de risque afin de mettre en avant ceux qui présentent le plus grand risque de régularisation et qui ont donc plus de possibilité d'être contrôlés.

Le contrôle, qu'il soit réalisé sur pièces ou sur place, est toujours réalisé par un humain. Le contrôle sur place est réalisé par l'un des 700 contrôleurs assermentés des Caf.

Les scores de risques les plus élevés concernent toujours les personnes les plus pauvres.

 Faux

Il est erroné de dire que seules les personnes aux revenus les plus faibles sont visées par le datamining. Les scores de risque sont calculés pour tous les allocataires, sans distinction. Par ailleurs, le score de risque n'intègre pas comme seule donnée la situation financière.

Un allocataire peut être contrôlé plusieurs fois.

 Vrai

Un allocataire peut subir plusieurs contrôles, mais avec des délais entre chaque. Ainsi, la Caf ne peut réaliser un autre contrôle sur place que 18 mois après le premier, si celui-ci n'avait pas eu de conséquence financière. Dans le cas inverse, le délai est ramené à 12 mois.

Les allocations sont suspendues pendant le contrôle.

 Faux

Les allocations ne sont pas suspendues, elles continuent à être versées, sauf dans les cas où l'allocataire ne répond pas aux sollicitations ou refuse de fournir les éventuels éléments complémentaires nécessaires à l'analyse du dossier.

Les contrôleurs sont payés aux résultats.

❌ Faux

Les contrôleurs n'ont pas de prime sur leurs résultats. S'il existe bien un montant prévu, il s'agit seulement d'un indicateur présenté dans la Convention d'objectif et de gestion de la branche Famille et en aucun cas un objectif à atteindre.

Les contrôles sont réalisés uniquement pour récupérer de l'argent.

❌ Faux

Les contrôles sont réalisés pour permettre à chacun de toucher son juste droit. Dans le cas des indus, la Caf récupère effectivement de l'argent versé en trop mais 27 % des montants des régularisations réalisées sont des rappels de prestations qui auraient dues être versés aux allocataires.

Le datamining stigmatise les bénéficiaires des minima sociaux et les considère comme des potentiels fraudeurs.

❌ Faux

Ayant pour objectif le paiement à bons droits, le datamining recherche les situations qui peuvent générer des indus et non des fraudes. Moins de 1 % des contrôles aboutissent à une suspicion de fraude. 27 % des montants des régularisations réalisées sont des rappels de prestations qui auraient dues être versés aux allocataires.





Caisse nationale des Allocations familiales
32 avenue de la Sibelle - 75014 Paris



@cnaf_actus



facebook.com/cafalllocationsfamiliales



youtube.com/c/AllocationsfamilialesFrance



linkedin.com/company/caf-allocations-familiales