

Ciblage des contrôles fiscaux et datamining par Frédérique Perrotin

Publié le 20/05/2020 - mis à jour le 22/05/2020 à 10H25

Bercy table sur une programmation efficace, pour optimiser l'activité des équipes en charge du contrôle fiscal. Cette stratégie devrait lui permettre de mieux appréhender les nouvelles formes de fraude et les dossiers aux enjeux financiers les plus lourds. Pour ce faire, la Direction générale des finances publiques (DGFIP) s'intéresse aux avancées de l'intelligence artificielle, notamment aux nouvelles modalités d'exploitation des données, les outils de *datamining*.

La DGFIP s'est engagée dans un processus de modernisation de ses travaux de ciblage des contrôles fiscaux en mettant en place la mission requêtes et valorisation (MRV) qui est chargée de développer l'utilisation des nouvelles méthodes d'analyse de données et en créant les pôles de programmation localisés dans les 8 directions spécialisées de contrôle fiscal.

Cette mission impulse les projets de *datamining*.

Composée de *data-scientists*, elle effectue des travaux permettant de caractériser une personne ou une entreprise frauduleuse et d'établir un profil de fraude qui sera appliqué à une population cible. Elle recueille des informations issues des applications professionnelles, personnelles et patrimoniales de la DGFIP pour détecter certaines anomalies, comme des incohérences entre différentes déclarations.

La mission est désormais constituée d'une équipe de 26 personnes. Les recrutements de *data-scientists* sont finalisés et la constitution de l'équipe est pratiquement terminée.

Quatre nouveaux recrutements sont en cours afin d'achever, en 2020, la phase de montée en charge prévue.

Les compétences recherchées n'étant pas toutes disponibles à la DGFIP, ce renforcement s'est effectué en grande partie à partir de recrutements extérieurs, des contractuels pour la plupart ainsi que deux attachés de l'INSEE. L'équipe

s'appuie par ailleurs sur des prestataires dont les travaux permettent une montée en compétences rapide.

Un nouvel outil 2

Des résultats prometteurs 2

Montée en puissance. 3

des listes relatives aux fraudes réalisées par les particuliers 3

Une expérimentation pour détecter les fraudes les plus graves 4

Des garanties suffisantes ?. 4

Les réserves de la CNIL 5

Elle a engagé également, début 2019, un partenariat avec un laboratoire de recherche du CNRS spécialisé dans l'utilisation des techniques d'analyse de données en matière de lutte contre la fraude.

Un nouvel outil

Bercy s'emploie à déployer de nouveaux outils informatiques à la fois pour mieux cibler la fraude fiscale et pour optimiser la programmation des contrôles fiscaux grâce aux avancées de l'intelligence artificielle.

Le développement du *deep learning*, une technique d'apprentissage automatique sans intervention humaine, basée sur des réseaux de neurones artificiels, a profondément transformé les possibilités d'actions de Bercy.

Les intelligences artificielles développées dans le cadre de cette technique de Neural Machine Translation (NTM) utilisent des corpus pour littéralement apprendre à égaler l'humain, voire à le surpasser. En se basant sur des contrôles anciens la machine apprend à repérer et à construire des modèles de fraude.

L'utilisation des algorithmes permet à l'administration fiscale de détecter les dossiers les plus complexes comme de traiter en masse des dossiers plus simples voire d'inciter les contribuables à régulariser leur situation.

La mission bénéficie à cet effet d'un outil informatique puissant, le projet de ciblage de la fraude et valorisation des requêtes (CVFR). Ce projet d'envergure développé dès 2014 a pour ambition d'améliorer l'efficacité des opérations de contrôle fiscal et s'inscrit dans le cadre de la modernisation des travaux d'analyse de données et de ciblage.

Financé par le fonds de transformation de l'action publique, **le CVFR explore les informations de 11 bases de données**, comme Sirius-Pro, le logiciel de

contrôle fiscal de la Direction générale des finances publiques (DGFIP), ou le Bulletin officiel des annonces civiles et commerciales (Bodacc).

En s'appuyant sur des méthodes exploratoires basées sur la statistique et des algorithmes et en permettant de modéliser des comportements, cet outil utilise les techniques de l'intelligence artificielle pour identifier de façon automatique des cas de fraudes fiscales possibles.

Des résultats prometteurs

Des investissements matériels et logiciels importants ont été réalisés en 2018 et 2019 et la nouvelle plate-forme informatique du projet est opérationnelle depuis le début de l'année 2019.

La mise en place de cette plate-forme a permis d'exploiter rapidement, dans la perspective de la programmation fiscale, les données internationales reçues dans le cadre des échanges automatiques d'information, dont les 5 millions de comptes bancaires détenus par des résidents français dans des établissements financiers étrangers ou les informations figurant dans les rapports pays par pays déclarés par les groupes multinationaux réalisant un chiffre d'affaires annuel de plus de 750 millions d'euros.

À titre d'exemple, au cours du mois de février 2019, la MRV a adressé, de façon automatisée et centralisée, 15 000 courriers invitant les détenteurs de comptes étrangers, n'ayant respecté les obligations déclaratives afférentes à la détention de comptes à l'étranger pour des montants significatifs, à contacter directement leurs services locaux. **Environ 4 000 personnes sont d'ores et déjà venues régulariser leur situation dans un service de la DGFIP.**

Par ailleurs, plus de 3 000 contrôles de détenteurs de comptes à l'étranger qui n'ont pas régularisé leur situation ont également été réalisés en 2019. Plus de 85 M€ de droits et pénalités ont été rappelés.

Montée en puissance

Plus de 6 M€ d'investissement ont été consenties sur 2018-2022 pour donner tous les moyens d'action nécessaires à la MRV.

Ces investissements ont été accompagnés d'une montée en puissance des travaux d'analyse qui s'est traduit par la progression significative des productions et de leur appropriation par les services.

Grâce à l'activité des équipes, le nombre de dossiers envoyés aux services est ainsi passé de quelques centaines en 2016 à plus de 100 000 en 2019.

Dans le domaine des fraudes des professionnels, la mission a développé une centaine de requêtes reposant sur des techniques d'apprentissage automatique ou d'analyse risque qui lui permettent de couvrir la plupart des risques fiscaux et de

déterminer, automatiquement, pour chaque entreprise relevant d'un régime déclaratif réel, une cotation traduisant le niveau de son risque fiscal.

Ces travaux ont permis aux productions de la mission d'être à l'origine de 22 % des opérations de contrôle fiscal concernant les entreprises programmées en 2019 (contre 14 % en 2018 et 11 % en 2017) avec pour perspective que cette part s'élève à 35 % en 2020 et à 50 % en 2022 pour les professionnels.

des listes relatives aux fraudes réalisées par les particuliers

Par ailleurs, conformément à l'avis favorable de la CNIL, la mission produit désormais, chaque trimestre, des listes relatives aux fraudes réalisées par les particuliers et est à l'origine de 11 % des contrôles relatifs aux particuliers terminés en 2019 et de 12 % des droits rappelés.

Au total, les droits et pénalités mis en recouvrement à l'issue des contrôles engagés à partir des dossiers qui ont été transmis par la MRV se sont élevés en 2019 à plus de 785 M€.

Les investissements matériels et techniques réalisés au cours des deux années précédentes permettent désormais, grâce à l'expérience et la puissance de calcul acquis, de mettre en œuvre des technologies d'analyse plus puissantes et de viser de nouveaux domaines d'investigations.

Ainsi, en 2020, la mission développera de nouvelles requêtes et modèles qui intégreront de nouvelles données (notamment les données transmises par les plates-formes collaboratives en application [de l'article 242 bis du CGI](#) issu de la loi relative à la lutte contre la fraude votée en 2018 ainsi que celles qui seront obtenues dans le cadre de l'analyse des réseaux sociaux).

Elle poursuivra également ses évolutions technologiques, notamment en développant sa capacité à traiter les données non structurées (*text-mining*).

Si les entreprises et leurs dirigeants constituaient les premières cibles du CVFR, le dispositif s'est étendu aux particuliers.

L'expérimentation relative à la collecte de données à caractère personnel publiées sur internet par les utilisateurs de plates-formes en ligne, va permettre aux *data-scientists* d'affiner leur méthode de profilage pour les personnes physiques.

Une expérimentation pour détecter les fraudes les plus graves

[L'article 154 de la loi de finances pour 2020](#) autorise en effet, les administrations fiscales et douanières, à titre expérimental pour une durée de trois ans, à collecter et exploiter au moyen de traitements automatisés les données publiées par les internautes sur les plates-formes de mise en relation en ligne (telles que les places de marché ou les réseaux sociaux).

Ce dispositif vise à renforcer les outils à disposition de l'administration pour détecter des fraudes fiscales ou douanières particulièrement graves, pour lesquels les moyens d'investigation traditionnels de l'administration sont insuffisants : fausse domiciliation fiscale à l'étranger, activité commerciale occulte, activités illicites telles que la contrebande de tabac ou le commerce de stupéfiants.

Ces traitements ne porteront que sur des données librement accessibles sur les sites internet concernés et rendues publiques par les utilisateurs eux-mêmes. L'administration agira en toute transparence, elle n'utilisera pas de pseudonyme et ne s'infiltrera pas dans des cercles restreints de personnes pour les recueillir.

Des garanties suffisantes ?

Cette expérimentation est strictement encadrée afin d'apporter des garanties quant au respect de la vie privée des utilisateurs des plates-formes.

L'examen approfondi de ce texte réalisé par l'Assemblée nationale et le Sénat a certes permis de renforcer les garanties qui entourent l'utilisation des données recueillies sur les plates-formes numériques à des fins de contrôle fiscal ou douanier.

Il a également permis d'informer largement les utilisateurs de plates-formes de l'exploitation qui pourrait être faite des données qu'ils rendent eux-mêmes accessibles à tous.

La mise en œuvre de ces traitements est ainsi entourée de nombreuses garanties prévues par la loi. En particulier, les délais de conservation des données collectées sont limités au strict nécessaire.

Et seuls des agents spécialement habilités pourront effectuer les traitements.

Aucun système de reconnaissance faciale ne pourra être utilisé. L'administration fiscale ne pourra pas sous-traiter la collecte ou le stockage des données.

Enfin les traitements ne pourront pas donner lieu à l'engagement automatique d'un contrôle.

Ainsi l'administration veillera, dans la mise en œuvre de ce dispositif qui s'inscrit dans le cadre fixé par la loi informatique et libertés de 1978, à ne pas réaliser de traitements informatiques intrusifs qui porteraient atteinte de manière disproportionnée à la liberté d'expression ou aux libertés publiques. L'administration se consacre actuellement aux mesures préparatoires à la mise en œuvre de l'expérimentation. Le texte sera soumis à l'avis de la CNIL au premier trimestre puis à l'avis du Conseil d'État, préalablement au lancement de l'expérimentation proprement dite. La conception des outils de collecte et d'analyse des données sera lancée dans l'année.

Un bilan intermédiaire sera dressé mi-2021, puis un bilan définitif sera établi en 2023. Ces deux synthèses seront transmis au Parlement et à la CNIL, afin d'évaluer la proportionnalité des traitements mis en œuvre et leur efficacité dans la lutte contre la fraude fiscale et douanière.

Les réserves de la CNIL

Ces garanties correspondent en partie à certaines des exigences de la CNIL. Consultée en amont sur le projet de loi, la CNIL a émis de nombreuses réserves sur cette expérimentation permettant la collecte de données sur les plates-formes en ligne([Délibération n° 2019-114, 12 sept. 2019, portant avis sur le projet d'article 9 du projet de loi de finances pour 2020](#)).

Elle a tenu à rappeler qu'une telle expérimentation, dont il revient au législateur d'apprécier l'opportunité, doit s'accompagner de garanties fortes afin de préserver les droits et libertés des personnes concernées.

Un dispositif de ce type est inédit.

Il témoigne d'un changement d'échelle dans l'utilisation de données personnelles par ces administrations.

Il traduit également un changement de technique, en permettant le développement d'algorithmes pour améliorer le ciblage des contrôles fiscaux à partir de l'exploitation de ces données.

La CNIL a constaté la légitimité des objectifs poursuivis et la présence de certaines garanties (absence de contrôles automatiques à partir des traitements mis en œuvre, nombre d'infractions limitées, etc.).

Elle a toutefois relevé que ce dispositif présente des enjeux très particuliers du point de vue des libertés, compte tenu de l'impact du dispositif sur la vie privée et ses possibles effets sur la liberté d'expression en ligne.

Dans ce contexte, elle a formulé plusieurs réserves de nature à préserver un strict équilibre entre l'objectif de lutte contre la fraude fiscale et le respect des droits et libertés des personnes concernées, notamment la nécessité de préciser les contours du dispositif envisagé tel que le caractère librement accessible des contenus visés, la nature des traitements envisagés ou celle des données pouvant être collectées et de renforcer les garanties existantes.

La Commission a notamment souligné la nécessité d'évaluer de manière approfondie le respect, par les administrations concernées, du principe de proportionnalité : seules les données réellement nécessaires à la détection de la fraude devant être exploitées.

Il conviendra de s'en assurer à tous les stades, lors de l'élaboration des textes réglementaires d'application, au cours de l'expérimentation et à l'issue de celle-

ci, a précisé la CNIL qui restera particulièrement attentive aux suites de ce texte ainsi qu'aux conditions de mise en œuvre effectives du dispositif.

Il est à noter à cet égard que le Conseil constitutionnel a globalement validé le dispositif. Il a considéré que le dispositif était assorti de garanties propres à assurer une conciliation équilibrée entre le droit au respect de la vie privée et l'objectif de valeur constitutionnelle de lutte contre la fraude et l'évasion fiscales.