CPA COMPTABLES PROFESSIONNELS AGRÉÉS CANADA

Se doter de données de qualité pour atteindre ses objectifs de transformation numérique

Septembre 2021

CPA Canada
Voir demain
RÉIMAGINER LA PROFESSION.

21-2136

À PROPOS DE CPA CANADA

Comptables professionnels agréés du Canada (CPA Canada) travaille en collaboration avec les ordres de CPA des provinces, des territoires et des Bermudes, et représente la profession comptable canadienne sur les scènes nationale et internationale. La profession canadienne peut ainsi faire la promotion de pratiques exemplaires, favorables aux entreprises et à la société en général, et préparer ses membres aux défis posés par un contexte en évolution constante, marqué par des changements sans précédent. Forte de plus de 220 000 membres, CPA Canada est l'une des plus grandes organisations comptables nationales au monde. cpacanada.ca

La version électronique de ce document est disponible sur le site **cpacanada.ca**.

Table des matières

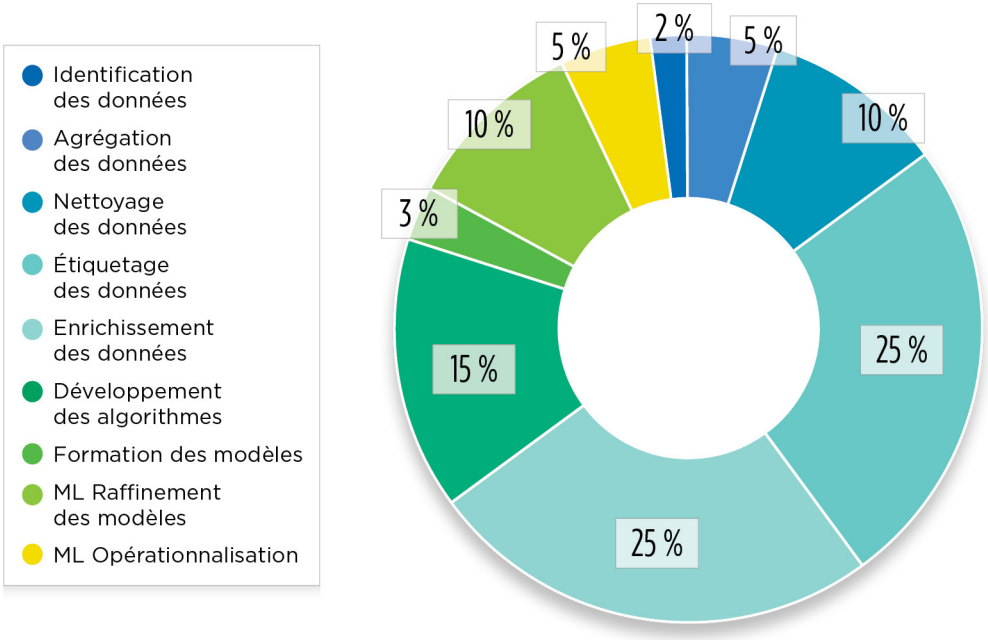
Dans le domaine informatique, on sait depuis longtemps que la qualité des données détermine la qualité des résultats. Ce principe s'applique tout particulièrement à l'apprentissage automatique, qui sous-tend des millions de systèmes basés sur l'intelligence artificielle (IA) autour du globe. C'est pourquoi la qualité des données est devenue la priorité des dirigeants d'entreprise. Dans une étude menée par Gartner, on estime que l'IA générera cette année 2 900 milliards de dollars américains en valeur pour les entreprises et permettra de dégager des gains de productivité correspondant à 6,2 milliards d'heures de travail[1].

Au cours des 10 dernières années, les organisations ont réalisé des investissements substantiels dans le développement et la mise à l'essai de nouveaux algorithmes et outils d'apprentissage automatique. Parallèlement, des experts en science des données ont déployé des dizaines de milliers d'algorithmes. Résultat : des applications basées sur l'IA sont maintenant offertes au moyen de plateformes à code source libre, d'abonnements à des plateformes infonuagiques et de contrats de licence. De plus, il est possible d'adapter et de former ces applications en vue d'usages variés. Les grandes organisations qui ont déjà entrepris le virage numérique se tournent maintenant vers la création (ou l'acquisition) d'ensembles de données de qualité.

1 Dataiku (2021). *Getting the Most out of AI in 2021: Insights from 10+ Industry Trailblazers*. https://pages.dataiku.com/getting-the-most-out-of-ai-2021

Comme les organisations cherchent à créer des algorithmes spécialisés visant à régler des problèmes de plus en plus précis, il en coûte de plus en plus cher de produire des données toujours plus détaillées et plus nuancées. Dans une étude récente menée par Cognilytica, on estimait que plus de 80 % du temps consacré aux projets d'IA et d'apprentissage automatique était employé à la préparation des données. Ce pourcentage inclut le temps passé à identifier les données, à les agréger, à les nettoyer, à les étiqueter et à les enrichir.

Temps consacré aux diverses tâches dans les projets d'apprentissage automatique



« La qualité des données est possiblement le principal facteur de réussite, estime Jeff McMillan, chef des données et de l'analyse des données chez Morgan Stanley. Effectivement, si les données sont inexactes, rien ne va plus. L'absence de données de qualité est probablement la principale raison pour laquelle les organisations ne parviennent pas à mener à bien leurs projets en lien avec les données. » Selon une étude menée par Appen, société spécialisée dans l'étiquetage de données, auprès d'entreprises qui ont adopté l'IA de manière précoce, 93 % des répondants sont d'avis qu'il est essentiel de disposer de données de formation de qualité pour qu'un projet fondé sur l'IA soit

une réussite[2]. Les organisations doivent composer avec de nombreux défis, notamment des données non étiquetées ou mal étiquetées, des ensembles de données incohérents ou désorganisés, des sources de données trop nombreuses et un manque d'outils leur permettant de traiter adéquatement les problèmes de qualité et d'éliminer les goulots d'étranglement[3].

Le présent document d'information vise à aider les CPA à s'assurer que les ensembles de données utilisés dans leurs organisations satisfont aux critères de qualité minimaux essentiels à la réussite des initiatives de transformation numérique. Il présente aussi différentes approches pour la mise en place et la gestion des systèmes et des contrôles de qualité des données.

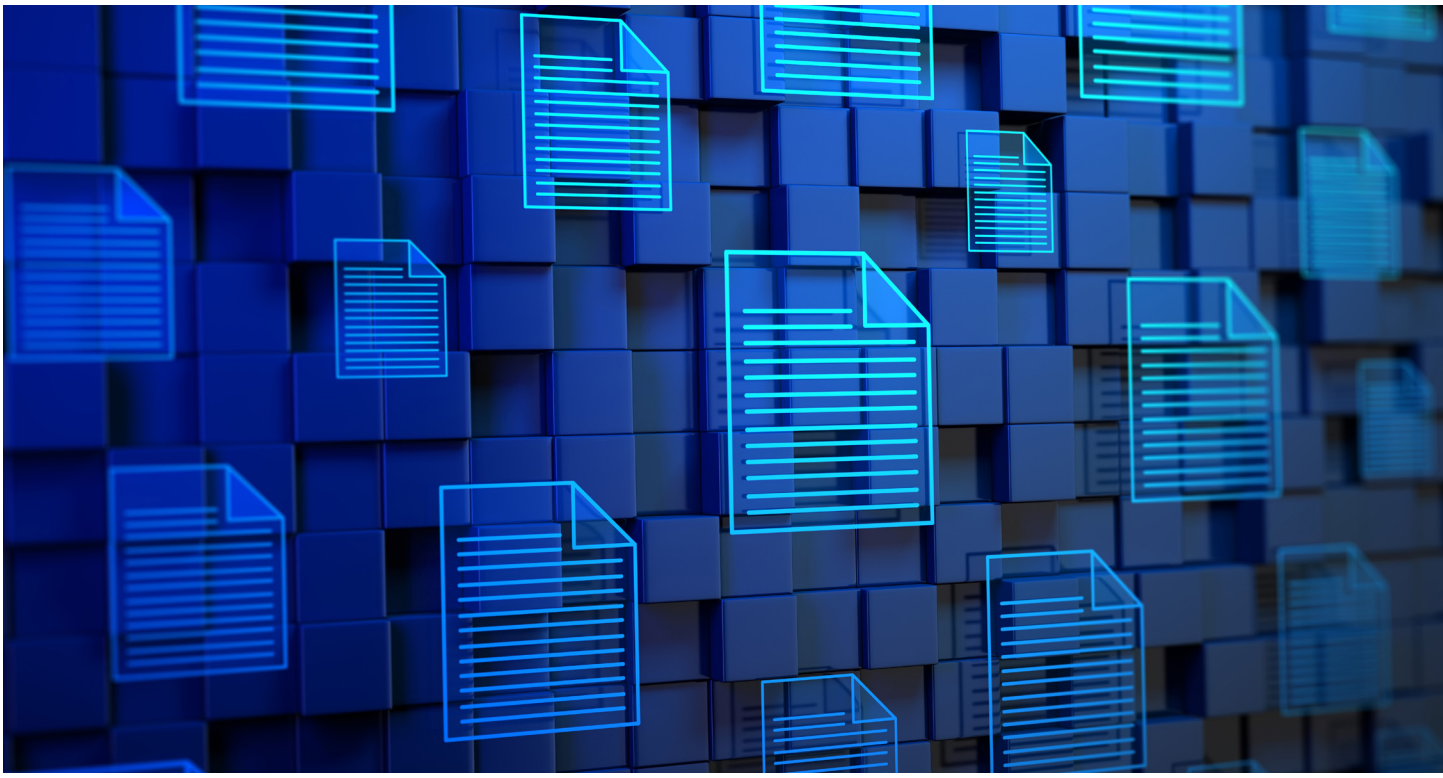
2 Appen (2020). *The State of AI and Machine Learning*. https://resources.appen.com/wp-content/uploads/2020/06/Whitepaper-State-of-Ai-2020-Final.pdf

3 Dataiku (2021). « 2021 Trends: Where Enterprise AI Is Headed Next », p. 19. https://www.dataiku.com/stories/2021-trends-where-enterprise-ai-is-headed-next/

Les risques associés à l'utilisation de données de mauvaise qualité

L'utilisation d'ensembles de données de qualité médiocre pour l'exécution d'algorithmes et d'outils d'apprentissage automatique présente tout un éventail de risques, au premier chef le risque financier. Selon l'International Data Corporation, les investissements dans l'IA à l'échelle mondiale devraient doubler au cours des quatre prochaines années. On prévoit qu'ils passeront de 50 milliards de dollars américains en 2020 à plus de 110 milliards en 2024[4]. Et pourtant, une étude de 2019 menée auprès d'entreprises indiquait qu'environ 87 % des projets liés aux sciences des données ne parviennent jamais au stade de la production. L'une des principales raisons pour lesquelles ces

4 International Data Corporation (2020). « Worldwide Spending on Artificial Intelligence Is Expected to Double in Four Years, Reaching $110 Billion in 2024, According to New IDC Spending Guide ». https://www.idc.com/getdoc.jsp?containerId=prUS46794720



projets échouent est la mauvaise qualité des données. « Quand on parle de données, explique Deborah Leff, chef de la technologie, de la science des données et de l'IA à IBM, le problème, c'est qu'elles prennent toujours une multitude de formes : structurées ou non structurées, vidéo, texte, images, etc. Elles sont conservées à divers endroits et soumises à différentes exigences de sécurité et de protection des renseignements personnels. Par conséquent, les projets sont freinés dès leur création parce qu'il faut rassembler et nettoyer toutes ces données[5]. »

Les outils d'apprentissage automatique, s'ils sont formés au moyen de données de mauvaise qualité ou incomplètes, peuvent aussi mener à un accroissement des risques réglementaires. Par exemple, des outils mal formés peuvent générer des résultats biaisés ou erronés qui sont susceptibles de causer de la discrimination en amenant une organisation à refuser de servir des groupes sous-représentés. Lorsque les données de formation sont biaisées ou erronées, les organisations s'exposent à des risques juridiques accrus. Dans un article du *Columbia Law Review*, le professeur Frank Pasquale s'est penché sur le risque de contamination de l'apprentissage automatique par les données inexactes ou inappropriées. « Les entreprises qui s'appuient sur des données erronées, prévient-il, peuvent être tenues de dédommager les personnes qui ont subi un préjudice en lien avec l'utilisation de ces données. » En effet, dans

5 Venture Beat (2019). « Why do 87% of data science projects never make it into production? ». https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/ [Traduction libre]

certaines positions doctrinales et réglementaires publiées récemment, on semble reconnaître l'émergence d'un devoir de diligence pouvant engager la responsabilité des personnes morales dans le cas où une organisation ne veille pas à l'application de mesures de sécurité adéquates à l'égard des données. Ce devoir de diligence comprend l'obligation d'accorder une attention particulière aux données inexactes, inappropriées ou obtenues illégalement et de « veiller à ce que les données de formation utilisées à des fins d'apprentissage automatique reflètent adéquatement le domaine qu'elles cherchent à représenter[6] ».

6 Frank Pasquale (2020). « Data Informed Duties in AI Development ». *Columbia Law Review*, vol. 119, pp. 1917-1940. https://columbialawreview.org/content/data-informed-duties-in-ai-development/ [Traduction libre]

Premiers pas pour prendre le virage numérique

Investir dans la création de données de qualité est une étape importante dans la transformation numérique d'une organisation, mais ce n'est pas la seule. Voici d'autres étapes clés pour vous permettre de préparer le terrain et d'assurer une transformation réussie :

- adopter une politique de gestion des données afin d'établir des règles de gouvernance concernant la réutilisation des données;
- faire approuver la stratégie numérique et le budget;
- mettre sur pied une équipe hybride composée de spécialistes, d'experts des données et de professionnels des TI;
- cerner les problèmes à résoudre ou les occasions à saisir par l'organisation, créer et consigner des cas d'utilisation, et sélectionner les bonnes solutions d'IA ou les bons outils d'apprentissage automatique à former.

Une fois ces premières étapes franchies, l'organisation devrait tourner son attention vers la création de données de qualité. En effet, il est crucial que les données soient épurées, exactes, complètes et étiquetées convenablement, ce dont l'organisation peut s'assurer lors de la collecte et de la préparation des données.

Collecte et préparation des données

La première chose à faire pour générer des données de qualité est d'identifier les sources de données pertinentes dans l'organisation, puis de trouver une façon de permettre à l'équipe hybride de voir les données, d'y accéder et de les utiliser. À cette étape, il faut mettre en correspondance tous les ensembles de données à partir des sources vers les sites de stockage. Les ensembles de données peuvent inclure du texte, des chiffres, des images fixes, des fichiers audio (y compris des fichiers contenant des paroles), du contenu vidéo, des objets, des séries chronologiques, des flux de données provenant de capteurs, des données sur les clics en ligne ou encore des informations sur les unités de gestion des stocks. L'examen des ensembles de données doit englober les processus de transformation (mesures à prendre pour regrouper des ensembles de données aux formats différents en un seul ensemble principal), les activités d'harmonisation (harmonisation des ensembles de données en les renommant selon une convention de nommage, d'horodatage et de séquencement) et des résultats finaux traités (un ensemble de données principal harmonisé et adapté à l'usage prévu). Les variables différeront selon que les données sont recueillies automatiquement (comme pour les clics en ligne), proviennent de capteurs (comme d'appareils connectés à l'Internet des objets) ou sont saisies manuellement dans une base de données (comme dans le cas d'un bon de commande).

Nettoyage, étiquetage et annotation des données

Les données non épurées sont d'une valeur limitée pour l'alimentation d'algorithmes et la formation d'outils d'apprentissage automatique. Il convient de s'assurer que les données sont conformes aux politiques internes avant de procéder au nettoyage, à l'étiquetage et à l'annotation.

Lors du **nettoyage des données**, on vérifie notamment que les champs vides ont été remplis, que les doublons ont été éliminés, que les définitions et les acronymes utilisés dans les différents ensembles de données ont été harmonisés et que les conventions d'horodatage et de séquencement ont été uniformisées. Il pourrait aussi être nécessaire de s'assurer que les renseignements personnels qui pourraient permettre d'identifier les personnes sont supprimés afin de garantir le respect les obligations de protection des renseignements personnels.

Les tâches **d'étiquetage et d'annotation des données** prennent énormément de temps. Grâce à elles, les ingénieurs de données qui programment les outils d'apprentissage automatique disposeront d'informations exactes sur les données qu'ils utilisent. Lors de la formation des outils d'apprentissage automatique, l'objectif est de montrer au modèle d'intelligence artificielle le résultat que l'on souhaite qu'il apprenne à prédire. En fournissant suffisamment d'exemples au modèle d'intelligence artificielle et en lui expliquant ce qu'il doit chercher et comment il doit le faire, on parvient après un certain temps à lui faire reconnaître par lui-même certaines des caractéristiques des objets non étiquetés et à l'amener à prendre une décision ou des mesures en fonction de

ces caractéristiques[7]. La nature des tâches d'étiquetage dépend généralement du type de données utilisées. Les principaux cas d'utilisation sont le balisage et l'étiquetage de fichiers textes et d'enregistrements vocaux, la reconnaissance, la classification et l'annotation d'objets ainsi que la classification, l'étiquetage et l'annotation d'images.

L'équipe hybride d'une organisation peut s'occuper du nettoyage, de l'étiquetage et de l'annotation de petits ensembles de données, mais elle aura besoin d'aide pour s'attaquer à des ensembles de données plus importants ou hétéroclites. « Bien souvent, explique John Singleton, co-fondateur de Watchful, fournisseur de solutions d'annotation des données, l'annotation des données est effectuée par une petite équipe d'experts en science des données qui est déjà débordée et qui ne peut donc pas se concentrer sur son véritable travail : développer et mettre en œuvre des modèles utiles[8]. »

Depuis 2018, on assiste à l'arrivée sur le marché d'une vague d'outils d'annotation des données, qui offrent des solutions complètes englobant le flux complet d'opérations d'étiquetage des données. Les organisations ont maintenant l'embarras du choix. Elles peuvent télécharger un logiciel clé en main et embaucher du personnel pour qu'il réalise manuellement les tâches d'annotation des données (même s'il est difficile de garantir la qualité du travail effectué par du personnel temporaire, à moins qu'il soit adéquatement formé et supervisé). Il existe aussi une grande variété d'outils d'étiquetage de données en libre-service sur le marché, dont Prodigy et Label Studio. Git-Hub répertorie d'autres excellents outils à code source libre ou payants (comme l'extrêmement populaire Dataturk).

La conclusion d'accords de type logiciel-service (SaaS) pour obtenir l'accès à une plateforme d'étiquetage des données en libre-service est une option de plus en plus populaire. Ces plateformes intègrent souvent des outils d'automatisation et facilitent la collaboration à grande échelle sur de gros ensembles de données. L'apprentissage automatique peut se révéler fort utile pour les tâches suivantes :

- l'étiquetage du contenu;
- la réalisation de contrôles de qualité sur le travail effectué par des humains;

7 Cloud Factory (2020). *Data Annotation Tools for Machine Learning: Choosing the Best Data Annotation Tool For Your Project*. https://go.cloudfactory.com/hubfs/02-Contents/2-eBooks/Data%20Annotation%20Tools%20for%20Machine%20Learning%20(Evolving%20Guide).pdf?utm_campaign=RTQD%20%7C%20Data%20Annotation&utm_medium=email&_hsenc=p2ANqtz-8i3RosIvhxfqIn1ZCEspQQNZd_wOxhmw0aYBhTkfcyLfrQV_zuzjje3rETW5tKOHMINA1N0IUJyujsuNRUu2A6Djz5FQ&_hsmi=88137733&utm_content=88137733&utm_source=hs_automation&hsCtaTracking=7bf603ac-2909-467b-b505-0f34b50289a0%7C408970b5-5de5-4328-91b7-7535da77716f

8 Synced (2019). « Data Annotation: The Billion Dollar Business Behind AI Breakthroughs ». https://medium.com/syncedreview/data-annotation-the-billion-dollar-business-behind-ai-breakthroughs-d929b0a50d23 [Traduction libre]

- l'identification des différents types de données;
- le repérage des valeurs aberrantes dans la structure d'une colonne de données et l'aide aux utilisateurs dans le nettoyage des données.

Des fournisseurs comme Alegion offrent des plateformes en libre-service pour l'annotation et l'étiquetage de vidéos que les spécialistes au sein des organisations peuvent utiliser pour étiqueter et annoter des données en interne[9].

Les organisations qui sont pleinement investies dans la transformation numérique pourraient quant à elles décider de créer des systèmes complets d'exploitation des données (DataOps) afin d'aider leurs équipes de spécialistes à générer rapidement et à répétition des données prêtes à l'usage à partir des diverses sources de données de l'organisation. Le terme « DataOps » a été inventé en 2019 par O'Reilly, une société qui mène des recherches dans le domaine de l'informatique, pour décrire les systèmes et les équipes spécialisés mis en place par les organisations dans le but d'exploiter leurs données en tant qu'actif stratégique[10]. Les organisations qui se sont dotées de telles équipes peuvent investir dans l'embauche d'experts capables d'effectuer de l'annotation de données hautement spécialisée en interne pour alimenter des algorithmes complexes.

9 Appen (2020). « Key Considerations for Developing a Data Annotation Solution for Your AI Models ». https://appen.com/blog/build-or-buy-data-annotation-tool/

10 Andy Palmer, Michael Stonebreaker, Nik Bates-Haus, Liam Cleary et Mark Marinelli (2019). *Getting DataOps Right*, O'Reilly, 58 pages. https://get.oreilly.com/ind_getting-dataops-right.html

Certaines entreprises offrent maintenant des logiciels qui permettent d'accélérer l'annotation des données pour les modèles d'apprentissage profond[11]. Cloud Factory a d'ailleurs récemment publié un guide complet des outils d'annotation offerts sur le marché.

Beaucoup d'organisations choisissent plutôt de faire appel à des tiers qui offrent à la fois des services de préparation des données et des conseils en la matière. De grands cabinets de services-conseils comme Deloitte, EY, KPMG et PwC proposent maintenant toute une gamme de services de soutien aux organisations qui se lancent dans la transformation numérique, dont des conseils et un encadrement en matière de préparation des données[12].

Les plateformes de traitement des données prennent aussi de plus en plus de place. En 2021, Gartner a publié un rapport exhaustif dans lequel on comparait 20 plateformes de science des données et d'apprentissage automatique qui permettent aux organisations de recueillir des données, de créer des modèles et d'opérationnaliser l'apprentissage automatique. Toutes les plateformes évaluées dans le rapport offraient des logiciels, des outils et des conseils pour gérer la préparation et l'exploration de données. Certaines plateformes visent des créneaux précis du marché, mais beaucoup des grands fournisseurs ont mis au point des approches qui peuvent être adaptées à une multitude de cas d'utilisation. Comme le montre le graphique ci-après, les chefs de file du domaine sont notamment SAS, IBM, Dataiku, MathWorks, Databricks et TIBCO Software[13].

11 Par exemple, voir l'offre de Supervisely, dont les services sont utilisés par plus de 25 000 organisations en activité dans un nombre croissant de secteurs : https://supervise.ly/

12 Canadian Accountant (2018). « Consulting in Canada: Where the Big Four firms are making money ». http://canadian-accountant.com/content/business/consulting-in-canada-where-the-big-four-firms-are-making-money

13 Gartner (2020). « Magic Quadrant for Unified Communications as a Service, Worldwide ». https://www.gartner.com/doc/reprints?id=1-24M0T9F3&ct=201117&st=sb

En date de janvier 2021

Les organisations peuvent aussi songer à confier la préparation des données à un tiers. Les fournisseurs ont créé des outils, qui se fondent souvent sur l'apprentissage automatique, pour améliorer la qualité de l'étiquetage et de l'annotation et accélérer ces fonctions. Ces tâches deviennent de plus en plus spécialisées et propres à chaque domaine. Par exemple, les nouvelles techniques d'imagerie médicale s'appuient sur des modèles d'apprentissage automatique capables d'identifier un large éventail de pathologies comme des caillots, des fractures, des tumeurs et des obstructions. Les données utilisées doivent donc permettre de diagnostiquer les bonnes pathologies grâce à un étiquetage cohérent, ce qui nécessite une bonne connaissance du domaine de la part des nombreux techniciens en étiquetage et en annotation des données.

Les fournisseurs de services d'étiquetage s'adaptent rapidement afin d'offrir des services spécialisés en fonction du domaine. D'après un rapport récent de Cognilytica, il existe actuellement plus de 35 entreprises qui fournissent de la main-d'œuvre pour l'étiquetage et l'annotation des données. Certains de ces fournisseurs utilisent des méthodes générales d'externalisation ouverte, tandis que d'autres font appel à leur propre bassin de main-d'œuvre formée pour répondre aux besoins en étiquetage de données dans des domaines précis. Parmi les étoiles montantes de ce secteur, mentionnons Scale, qui compte sur des outils d'annotation basés sur l'IA et 30 000 employés contractuels pour l'étiquetage de texte, d'audio, d'images et de vidéo. On s'attend à ce que la demande de services d'étiquetage de données externes connaisse une forte croissance et que la valeur du marché passe de 1,7 milliard de dollars américains en 2019 à plus de 4,1 milliards en 2024[14].

Enfin, une autre possibilité consiste à acheter les données qui serviront à former les algorithmes et les outils d'apprentissage automatique. Des organisations comme Thompson Reuters, Orbital Insights, Appen et Collibra Marketplace offrent déjà la location ou l'achat d'ensembles de données soigneusement choisies et traitées. On trouve aussi de plus en plus d'ensembles de données en libre accès rendus disponibles par des administrations publiques, des OSBL et des universités[15]. Gartner estime que d'ici 2022, 35 % des grandes organisations vont vendre ou acheter des données sur des marchés de données en ligne. Elles étaient 25 % à le faire en 2020.

Les marchés et les bourses de données regroupent les données offertes par des tiers. L'accès centralisé qu'ils procurent permet la réalisation d'économies d'échelle qui font baisser le coût des données fournies par des tiers. Toutefois, dans un article publié en 2020 sur le site de Gartner, on soulignait que « pour monnayer les actifs de données grâce aux marchés de données, les chefs de file du secteur doivent établir un cadre juste et transparent fondé sur des principes de gouvernance des données et en lequel ceux qui participent à l'écosystème peuvent avoir confiance[16] ».

14 Cognilytica (2019). *Data Engineering, Preparation, and Labeling for AI 2019*. https://www.cognilytica.com/2019/03/06/report-data-engineering-preparation-and-labeling-for-ai-2019/

15 https://appen.com/open-source-datasets/; https://www.thomsonreuters.com/en/artificial-intelligence/machine-learning.html; https://orbitalinsight.com/

16 Gartner (2020). « Gartner Top 10 Trends in Data and Analytics for 2020 ». https://www.gartner.com/smarterwithgartner/gartner-top-10-trends-in-data-and-analytics-for-2020/ [Traduction libre]

Exactitude, contrôle de la qualité et classement

La mise en place de contrôles est essentielle assurer un bon nettoyage des données avant qu'elles soient soumises à l'analyse. Les organisations devraient passer en revue les ensembles de données combinées ou agrégées pour vérifier leur exhaustivité, leur exactitude et leur fiabilité.

C'est ce qu'on appelle le classement des données. Ce processus vise à déterminer si les données évaluées conviennent à l'utilisation prévue. Toutes les données ne sont pas égales et, selon le type de décision que l'on cherche à prendre, il convient d'utiliser des données dont le niveau de fiabilité est plus ou moins élevé. En effet, l'utilisation de données dont on ne comprend pas les limites sur le plan de la fiabilité peut donner lieu à un piètre résultat, en particulier lorsque les données sont présumées plus fiables qu'elles ne le sont réellement. Lorsqu'il est question d'IA et d'apprentissage automatique, les CPA devraient donc rechercher : a) des données de grande qualité pour éviter les biais induits par des ensembles de données incomplets ou erronés et b) des algorithmes et des outils d'apprentissage automatique bien conçus pour éviter les biais découlant d'hypothèses erronées. Toutefois, la collecte de données très fiables nécessite temps et argent, raison pour laquelle le fait d'exiger un tel degré de qualité pour toutes les décisions devant être prises pourrait conduire à des occasions manquées.

Voici un exemple pour illustrer l'idée d'adaptation à l'objectif : les données qui sous-tendent l'information financière (rapports externes) sont très fiables et généralement certifiées comme telles. À l'inverse, les données utilisées à des fins internes peuvent être plus ou moins fiables. Les professionnels

comptables doivent comprendre les différences et être en mesure d'informer les décideurs de la fiabilité des données qui sous-tendent une décision. Pour ce faire, ils doivent notamment être en mesure d'évaluer si les données répondent aux exigences de fiabilité correspondant au type de décision à prendre. Par exemple, les exigences de fiabilité des données pour une analyse des options sont moins rigoureuses que celles qui s'appliquent lorsqu'il s'agit de décider d'aller de l'avant ou non avec un projet. Le calibrage de la fiabilité des données selon les décisions à prendre peut améliorer la rapidité de la prise de décisions, la communication des risques et la transparence. Il est essentiel que des processus officiels de vérification soient en place afin de s'assurer que les décisions sont validées à mesure que les données sont mises à jour, sous peine de devoir revenir sur des décisions judicieuses déjà prises du fait d'événements postérieurs[17].

Les CPA sont bien placés pour gérer les systèmes et les fonctions de contrôle de la qualité des données d'organisations engagées dans une transformation numérique. Comme on l'explique dans l'article Comprendre les chaînes de valeur des données de CPA Canada, de nouvelles catégories de professionnels font leur apparition pour remplir une série de nouvelles tâches et de fonctions. Les experts en science des données jouent un rôle crucial dans l'élaboration de nouveaux algorithmes. C'est là qu'entrent en scène les ingénieurs de données, ingénieurs en apprentissage automatique et autres ingénieurs en logiciels, qui assurent la gestion des activités de collecte de données et des systèmes et interfaces de gestion des données, et veillent à la programmation qui sous-tend l'apprentissage automatique. La gestion des solutions d'accès et de partage de données, l'organisation des tableaux de bord de données et leur mise à jour, le suivi des demandes de données, et l'observation des contrats de partage de données et de la réglementation applicable entraîneront une forte demande de contrôleurs de données. Dans l'optique où cette nouvelle catégorie professionnelle n'offrait strictement qu'un service public, elle pourrait combiner des compétences telles que la bibliothéconomie, le droit contractuel, le droit en matière de protection des renseignements personnels, la cybersécurité, la vérification de la conformité et la production de rapports de conformité.

17 CPA Canada et IFAC (2021). *Le rôle du professionnel comptable dans la gestion des données*, document de travail, 44 pages.

Même si l'évaluation et la vérification de la qualité des données sont devenues des priorités, tout comme la communication d'informations à ce sujet, il n'existe encore aucune catégorie professionnelle qui possède la bonne combinaison d'aptitudes et de compétences pour s'occuper des systèmes et des contrôles de qualité des données. Il s'agit toutefois d'un rôle que les CPA sont bien placés pour assumer. À l'avenir, les CPA bénéficieraient de l'élaboration et de la tenue à jour de normes sur la qualité des données qui viendraient encadrer adéquatement ces nouveaux systèmes et contrôles. CPA Canada entend appuyer l'élaboration d'indications appropriées qui permettront de favoriser la collecte et l'utilisation de données de qualité.

Pour de plus amples informations

Série d'articles Maîtrise des données de CPA Canada

- La politique de gestion des données et ses éléments
- Élaborer une stratégie numérique pour votre organisation
- Comprendre les chaînes de valeur des données

Outils d'annotation et d'étiquetage des données

- Prodigy
- Label Studio
- Dataturk
- Git-Hub
- Cloud Factory
- Scale

COMPTABLES PROFESSIONNELS AGRÉÉS CANADA

277, RUE WELLINGTON OUEST
TORONTO (ONTARIO) CANADA M5V 3H2
TÉL. : 416 977.3222 TÉLÉC. : 416 977.8585
CPACANADA.CA