

L'usage du Big Data (BD) appliqué à l'optimisation des fonctions de l'administration fiscale : les enseignements tirés de la littérature

Harouna KINDA

CERDI - Ecole d'Economie, University Clermont Auvergne, France

septembre, 2019

Abstract

Cette étude analyse la littérature sur les effets de l'application du BD sur l'optimisation des fonctions de l'administration fiscale, dans l'optique de réduction des écarts fiscaux. L'étude s'intéresse particulièrement à l'efficacité de l'administration fiscale et la lutte contre la fraude et l'évasion fiscale des contribuables. Une large littérature théorique et des évidences empiriques indiquent que l'usage du BD garantit un système de collecte de recettes fiscales efficace, facile et moins coûteux. Cette technologie permet aussi de combattre avec efficacité le comportement frauduleux des contribuables, que cela concerne la fiscalité interne ou internationale. Le travail de la caisse nationale de sécurité sociale (CNSS) au Maroc qui a consisté en un recoupement des données de déclarations de salaires de la CNSS avec celles de l'administration des impôts ainsi que le ministère de l'équipement et du transport a relevé respectivement 91% et 93% de taux de non-conformité. Faciliter l'échange d'informations entre les différentes administrations nationales et/ou internationales et la formation du personnel sur l'usage des nouvelles technologies s'avèrent donc très important pour le renforcement de l'efficacité de ces dernières dans la lutte contre la l'évasion fiscale.

1. Introduction

La conférence internationale d'Addis-Abeba relative au financement du développement et des Objectifs du Développement Durable a souligné le rôle prioritaire des recettes fiscales intérieures. Les pays en développement se sont engagés dans des réformes de leurs politiques et administrations fiscales avec l'appui des institutions de Bretton Woods, de la Commission Européenne ou des agences d'aide nationales dans le but de réduire les écarts fiscaux. La réforme de transition fiscale au Maroc, pour compenser la chute des recettes tarifaires causée par les accords de libre-échange et commerciaux de 1996, a favorisé une progression des recettes publiques (de 23.6% en 2006 contre 25.3 % du PIB en 2000). Cependant, cette réforme a été caractérisée d'atypique en raison de l'étroitesse de l'assiette fiscale et les mesures dérogatoires successives (Brun et al., 2007).

Des administrations fiscales, y compris celle du Maroc, ont vu leur situation s'améliorer grâce à la révolution numérique. Plusieurs travaux indiquent que l'exploitation des données volumineuses (BD) constitue un facteur majeur non seulement en raison de l'amélioration des techniques de collecte fiscale, mais aussi l'efficacité de la lutte contre la fraude et l'évasion fiscale¹. L'usage du BD dans le do-

maine de la fiscalité a facilité l'effort requis pour atteindre des niveaux élevés d'exactitude et a augmenté l'utilisation de l'information fiscale dans le monde entier (Vamsi, 2017). La question principale est de savoir : en quoi l'usage du BD permet d'optimiser les fonctions de l'administration fiscale ? Spécifiquement, comment optimiser la collecte des recettes publiques dans l'usage du BD ? Comment combattre la fraude et l'évasion fiscale dans l'usage du BD ?

L'objectif de cette étude est d'analyser les effets de l'usage du BD sur l'optimisation des fonctions de l'administration fiscale. De façon spécifique, il s'agit d'abord de montrer comment l'usage du BD permet d'optimiser l'administration fiscale en matière de collecte de recettes fiscales, puis de montrer comment cette nouvelle technologie permet-elle de combattre la fraude et l'évasion fiscale.

Dans la suite de notre analyse, la section 2 présente la définition de concepts, la section 3 analyse les éléments de réponse à la question principale, la section 4 fait un état des lieux concernant l'usage du BD dans l'administration Marocaine, la section 5 conclut avec quelques recommandations pour réussir les politiques que visent les BD.

2. Définitions de concepts

2.1. Big Data (BD)

Le terme «Big Data» a été officiellement introduit dans le monde informatique par Roger Magoulas en 2005 pour décrire généralement des ensembles de données si volumineuses ou complexes que la technologie et les systèmes

1. (Rosemary Williams DBA, 2008; Hoover, 2009; Ameur and Tkiouat, 2012; Wu et al., 2012, 2013; González and Velásquez, 2013; Hall et al., 2014; Sharma, 2015; Tian et al., 2016, 2017; Vamsi, 2017; Malaszczyk and Purcell, 2017; Kumar, 2018; Atanasijević et al., 2019; Pérez López et al., 2019)

d'information traditionnels ne sont pas adaptés pour leur traitement et analyse². Il fait souvent référence à l'utilisation de l'analyse prédictive, du comportement des utilisateurs ou des méthodes avancées d'analyse de données qui cherchent à extraire une valeur à partir de données, et rarement à une taille particulière de données. Sans doute, les quantités de données actuellement disponibles sont très importantes. Mais, cette caractéristique n'est pas la plus pertinente de ce nouvel environnement de données. Il couvre les techniques et technologies innovantes permettant de capturer, stocker, distribuer, gérer et analyser des ensembles de données de très grande taille avec diverses structures (Manyika, 2011).

Le BD est caractérisé par un nombre défini de V (volume, vitesse, véracité, variété). Laney (2001) dans le rapport Gartner propose les trois propriétés du BD les plus courantes (figure 1). Initialement on avait, le volume qui correspond à l'importance de la quantité d'informations; vitesse qui est liée à la vitesse à laquelle l'information circule au sein d'une organisation (exemple : le traitement des informations financières); la véracité qui caractérise les données provenant de sources différentes augmentant le risque d'incertitude dans les données; et la variété qui caractérise les types de structures de données, telles que des données structurées, semi-structurées et non structurées (Zhang et al., 2015).

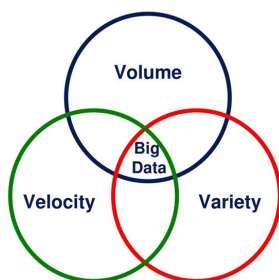


FIGURE 1: Les 3 V du BD

Théoriquement, le BD est le moyen d'étudier de très grandes quantités de données afin d'établir des modèles originaux offrant une vision plus fine de la réalité. Cela permet aux analystes, aux chercheurs et aux utilisateurs professionnels de prendre des décisions plus pertinentes et rapides à partir des données qui auparavant étaient inaccessibles ou inutilisables. Comme exemple, les achats en ligne, les transactions électroniques et par carte, les achats hors ligne facturés ainsi que les réactions aux avis d'imposition constituent des données du cheminement fiscal d'un particulier utiles pour l'analyse comportementale d'un contribuable (Kumar, 2018).

L'utilisation des données constitue un facteur clé dans la prise de décisions d'orientation stratégique et opérationnelle des organisations. Selon IBM (IBM BD analytics),

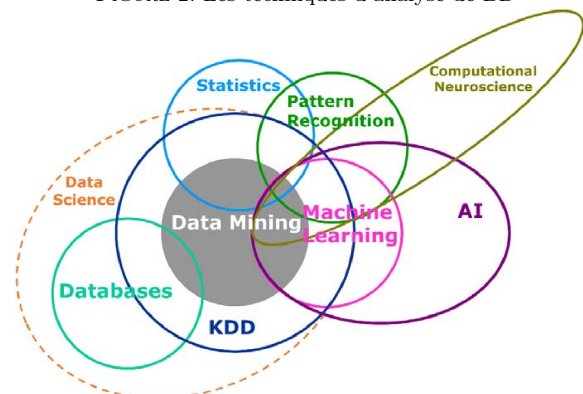
2. (Ularu et al., 2012; Wu et al., 2013; Cao et al., 2015; Vasarhelyi et al., 2015; Warren Jr et al., 2015)

80% des données générées actuellement sont non structurées et se présentent sous différents formats, tels que texte, vidéo, son, diagrammes, images et combinaisons de deux ou plusieurs formats. On note également actuellement l'importance des données collectées par des agences gouvernementales à des fins d'enregistrement, de transaction et de tenue de registres (Elias, 2014).

2.2. Méthodes d'analyse

Plusieurs définitions du BD semblent se baser spécifiquement sur un type de données, à savoir les données de médias sociaux ou des entreprises. Pourtant il existe divers types de données qui n'étaient pas disponibles auparavant pour la recherche en sciences sociales, pouvant également être qualifiées de BD. Ce qui nécessite des approches analytiques différentes selon la nature spécifique des données pour entreprendre des analyses appropriées (Connelly et al., 2016). Les techniques d'analyse du BD sont nombreuses et complémentaires (Data mining, Machine Learning, neural network, Business intelligence, etc.) (fig 2).

FIGURE 2: Les techniques d'analyse de BD



Source : Hall et al. (2014)

2.3. Évasion fiscale, fraude, optimisation

la fraude ou l'évasion fiscale décrit tout comportement visant à réduire ou à éliminer la dette fiscale d'un contribuable. L'emploi des termes « évasion », « optimisation » ou « fraude » fiscale porte souvent à confusion. Ils sont intimement liés mais revêtent néanmoins des aspects différents. La définition de l'évasion fiscale est plus complexe, puisqu'elle revêt à la fois de l'optimisation et de la fraude. L'évasion entre dans la catégorie de l'optimisation si elle utilise des moyens légaux. À l'inverse, elle s'apparente à la fraude lorsqu'elle s'appuie sur de moyens illégaux ou dissimule la portée véritable de ses acteurs. Selon l'OCDE la fraude fiscale est définie comme « toute action du contribuable qui implique une violation de la loi, lorsqu'on peut prouver que l'intéressé a agi dans le dessin délibéré d'échapper à l'impôt ». Elle se réfère à l'utilisation de moyens illégaux pour réduire les montants dus des impôts ou d'en échapper totalement. L'optimisation fiscale quant à elle

consiste à l'utilisation de la législation pour échapper à l'impôt par différents moyens légaux (régimes dérogatoires, niches fiscales...). Mais sa légitimité peut être contestée. Juridiquement, l'optimisation fiscale peut être considérée comme illégale dans la mesure où elle constitue un abus de droit, ce qui reste difficile à démontrer en pratique pour l'administration fiscale (lafinancepourtous.com, 2019). La figure 3 représente la composition de l'évasion fiscale.

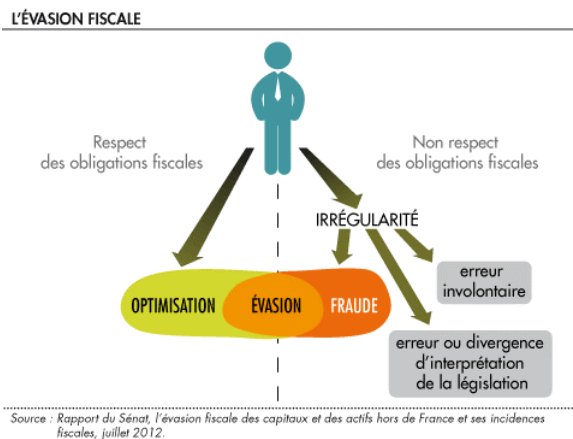


FIGURE 3: Représentation schématique de l'évasion fiscale

3. Comment la fiscalité moderne peut-elle tirer profit de l'analyse du BD ?

Ces dernières années, l'administration fiscale s'est considérablement améliorée dans plusieurs pays en développement. Cela s'explique en partie par la révolution numérique à travers la conjugaison d'une collecte améliorée et d'une puissance accrue de traitement des données permettant aux autorités d'affiner les méthodes existantes de recouvrement des taxes ([Gupta et al., 2017](#)). L'une des innovations majeures est l'introduction de la déclaration électronique de toutes les obligations fiscales. Le BD et l'intelligence artificielle (AI) quoiqu'ils soient complémentaires, sont les piliers de la transformation numérique. Avec le BD, les autorités fiscales peuvent analyser les données d'interaction, cartographier le parcours fiscal et comprendre le comportement d'un particulier ou d'une société en matière de conformité fiscale. Cette approche plutôt proactive que réactive, permet aux autorités fiscales d'analyser le potentiel fiscal, interagir avec les contribuables, prévoir des scénarios pour une utilisation efficace des ressources et répondre à leurs attentes et à celles des contribuables ([Kumar, 2018](#)).

3.1. L'efficacité de la collecte fiscale

L'efficacité de toute observation et de collecte fiscale dépend fortement de la garantie d'un ensemble de données fiables et d'un aperçu du parcours des contribuables. Le développement des services numériques facilite l'échange d'informations entre les autorités fiscales nationales. Par exemple le FATCA (Foreign Account Tax Compliance Act)

introduit par l'Internal Revenue Service (IRS) en 2010 dépend de la déclaration des titulaires de comptes américains par les institutions bancaires étrangères.

Tout d'abord, le BD aide les autorités fiscales et les services financiers à stocker et à traiter efficacement des données en quantités énormes. Cela consiste à automatiser le mouvement de données et de réduire les processus d'analyses manuelles et fastidieuses, ce qui améliore à la fois la qualité et la vitesse du processus fiscal. En outre, l'usage du BD garantit plus de précision dans les prévisions des rendements fiscaux, dans la mesure où les autorités fiscales peuvent créer une image complète d'un individu ou d'une entité pour tous les comptes ou toutes les zones géographiques ([Vamsi, 2017](#)). Aussi, l'un des moyens les plus importants dans l'usage du BD est le renforcement de la fonction analytique. Cela permet de détecter des risques et d'optimiser l'exécution des tâches de contrôle fiscal sur la base des catégories de contribuables les plus risquées. L'étude de [Atanasijević et al. \(2019\)](#) sur le cas de la Serbie souligne l'identification d'activités secondaires, le renforcement du rôle consultatif des inspecteurs et l'amélioration de la qualité des données à partir du développement des registres et de l'échange d'informations. Enfin, l'usage de nouvelles techniques de stockage et de traitement des données permet d'éviter les rapprochements coûteux et les écarts de rapports, ce qui réduit les audits coûteux en vue de faire une prévision stratégique et une meilleure planification fiscale.

3.2. La lutte contre la fraude et l'évasion fiscale

La fraude fiscale constitue un problème majeur pour l'environnement des affaires économiques dans le monde et en particulier dans les économies en développement. Elle affecte la concurrence, réduit la mobilisation des ressources propres et rend difficile l'application des politiques publiques ([Abrantes and Ferraz, 2016](#)). Les entreprises qui déclarent et paient légalement leurs impôts font face à une détérioration de leurs activités commerciales sur le marché au profit de leurs concurrentes fraudeuses à cause du différentiel des prix des biens et services. Toutefois, l'usage du BD permet de détecter des cas d'évasion et de fraude fiscale ([Malaszczyk and Purcell, 2017](#)).

Tout d'abord, l'usage des données historiques sur la réaction des contribuables à la réception des avis d'imposition permet d'anticiper sur leur prochain mouvement, ce qui permet de supprimer les possibilités de fraudes avant même qu'elles ne se produisent. Cela peut se traduire par l'analyse de la tendance, l'identification des problèmes d'audit potentiels et des cas de hauts risques à partir d'une enquête approfondie ([Kumar, 2018](#)). Ensuite, le recoupement d'un ensemble d'informations provenant de sources différentes permet d'avoir une vision intégrée des déclarants et des déclarations de revenus et d'utiliser les ressources de manière optimale, ce qui permet de faire face aux abus fiscaux ([Sharma, 2015](#)). Dans la même logique, les statistiques miroirs permettent d'identifier les secteurs à risque

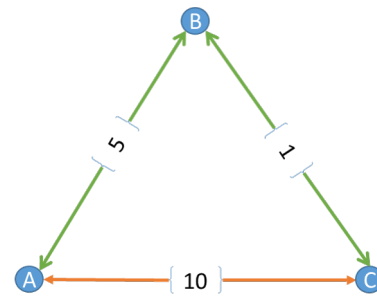
dans la mesure où elles peuvent servir à confronter les statistiques des exportations au départ des pays fournisseurs à destination du pays client et vice-versa des importations introduites par le pays client venant des pays fournisseurs.

De même l'exploration de données (BD mining) contribue significativement à détecter des cas de fraudes ou des anomalies des administrations fiscales. Selon [Wu et al. \(2012\)](#), elle améliore la détection de l'évasion fiscale à travers un filtrage des déclarations de TVA non conformes susceptibles de faire l'objet d'un audit supplémentaire. En se basant sur un ensemble d'algorithmes, [Mwanza and Phiri \(2016\)](#) ont analysé les paiements d'impôts domestiques en Zambie afin de détecter les sous-paiements et les sur-paiements, indiqués comme des valeurs aberrantes.

Par ailleurs, le BD à partir de la théorie des réseaux est de plus en plus utilisé dans la lutte contre les groupes suspects d'évasion fiscale ([Tian et al., 2016, 2017](#); [Santiago, 2018](#)). [González and Velásquez \(2013\)](#) utilisent des algorithmes de mise en grappes pour identifier des groupes de comportements similaires dans l'univers des contribuables. Les auteurs utilisent également des arbres de décision, des réseaux de neurones artificiels et des réseaux bayésiens pour identifier des schémas de fraude. Selon l'étude de [Pérez López et al. \(2019\)](#) sur l'Espagne, l'utilisation des réseaux de neurones a permis de segmenter les contribuables et de calculer la probabilité concernant la propension pour un contribuable à tenter de se soustraire à l'impôt.

De nos jours, la fraude et l'évasion fiscale se diversifient de plus en plus dans un environnement international en raison de la complexité des régimes fiscaux aussi provoquée par les conventions fiscales bilatérales. D'une part, les pays signent des conventions fiscales dans le but d'améliorer les entrées nettes d'IDE, et d'autre part, les taux de retenue à la source seraient considérablement plus élevés en l'absence de conventions fiscales ([Arel-Bundock, 2017](#)). Les multinationales utilisent la structure du réseau de traités pour réduire leur fardeau fiscal. L'utilisation du BD avec la technique d'analyse de réseau permet de combattre les formes de fraude fiscale qui découlent des opérations intragroupes des entreprises multinationales (fig 4). En supposant X une société résidente du pays C qui a une filiale implantée dans le pays A. On considère qu'il n'y a pas convention de double imposition (CDI) entre ces deux pays, et donc la retenue à la source bilatérale est de 10%. En supposant que la CDI est en vigueur entre A et B avec un taux réduit de 5%, de même entre B et C avec un taux réduit de 1%. Le groupe de la société X peut rapatrier son dividende en suivant le chemin AC ou ABC. Il serait rentable d'implanter une société mère écran située en B pour bénéficier d'un taux réduit de 5+1 au lieu de 10. L'identification des stratégies de réduction de l'impôt permet à l'administration fiscale du pays A d'évaluer des options de renégociation des accords. Le pays peut supprimer ou augmenter le taux réduit entre A et B ou négocier une convention avec C pour rendre le trajet plus concurrentiel.

FIGURE 4: Taxation des dividendes à la retenue à la source



Source : auteurs, à partir de [Santiago \(2018\)](#)

4. L'usage du BD dans l'administration fiscale marocaine

4.1. Etats des lieux sur l'usage du BD

L'étude de [Ameur and Tkiouat \(2012\)](#) utilisant l'exploration de données (datamining) indique que les autorités fiscales marocaines n'ont pas été épargnées par la fraude fiscale, y compris la sous-déclaration et la réduction des paiements. Dans le but de lutter efficacement contre les montages fiscaux abusifs, le pays a misé sur la collaboration internationale par des signatures de conventions fiscales bilatérales. En effet, le Maroc a signé près de 40 traités fiscaux bilatéraux à travers le monde selon *Price Waterhouse Coopers (PWC)* Maroc en 2019.

Parmi les pays qui utilisent les BD dans le but d'améliorer les fonctions de l'administration fiscale, le Maroc en est à ses débuts. Le développement du réseau internet et l'interconnexion de nos jours a stimulé une avancée des technologies numériques. En 2013, une enquête réalisée pour le compte d'IBM sur l'adoption de l'analyse de BD au Maroc indique que 44% des organisations ne prévoient pas de mettre en œuvre le BD, 28% sont à l'étape de la planification, et 22% des personnes ont déjà des projets BD donnant des résultats ([IBM, 2013](#)). En 2015, une étude a été lancée par le Haut-Commissariat au Plan (HCP) dans le but d'identifier les principaux défis liés aux projets BD au Maroc à l'horizon 2020.

4.2. Prise en compte des bilans doubles des entreprises

Pour lutter contre la minoration des bases de cotisations sociales, la CNSS a procédé en un recoupement des données de déclarations de salaires au niveau de la CNSS avec celles l'administration des impôts ainsi que le ministère de l'équipement et du transport. Avec les données de l'administration fiscale le rapprochement entre la déclaration du nombre de salariés et les registres de la CNSS a relevé un taux de non-conformité de 91%. Cela a permis de régulariser et de redresser 44 salariés en moyenne. La masse salariale d'écart est d'environ 1,3 million de dirham. De même, le rapprochement avec les données du ministère de l'équipement et du transport relève 93% de taux de non-conformité, ce qui a permis de régulariser et de redresser 29 salariés en moyenne avec une masse salariale moyenne de 900 000 dirham dérogée.

5. Conclusion et recommandations

L'objectif de cette étude était d'analyser la littérature sur les effets de l'application du BD sur l'optimisation des fonctions de l'administration fiscale. A travers des analyses théoriques et des évidences empiriques, cette analyse indique que l'usage du BD permet non seulement d'optimiser la collecte de recettes fiscales, mais constitue aussi un moyen incontournable de nos jours dans la lutte contre les comportements frauduleux des contribuables. La preuve en est que l'analyse du BD permet de réduire l'espace d'arbitrage fiscal ou l'amélioration des procédures réglementées relatives à la déclaration fiscale afin d'éliminer certainement des risques de fraude fiscale. Nous recommandons ainsi fortement aux pays en développement d'adopter les technologies du BD. Faciliter l'échange d'informations entre les différentes administrations nationales et/ou internationales. Les pays doivent également intégrer l'analyse du BD dans la formation du personnel de l'administration fiscale pour faciliter son usage. L'un des enjeux importants de l'usage du BD consiste à préserver la confidentialité sur les contribuables.

Références

- Abrantes, P. C. and Ferraz, F. (2016). Big data applied to tax evasion detection : A systematic review. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 435–440.
- Ameur, F. and Tkiouat, M. (2012). Taxpayers fraudulent behavior modeling the use of datamining in fiscal fraud detecting moroccan case. *Applied mathematics*, 3(10) :1207.
- Arel-Bundock, V. (2017). The unintended consequences of bilateralism : Treaty shopping and international tax policy. *International Organization*, 71(2) :349–371.
- Atanasijević, J., Jakovetić, D., Krejić, N., Krklec-Jerinkić, N., and Marković, D. (2019). Using big data analytics to improve efficiency of tax collection in the tax administration of the republic of serbia. *Ekonomika preduzeća*, 67(1-2) :115–130.
- Brun, J.-F., Chambas, G., and Laurent, M. (2007). Economie politique de la réforme de transition fiscale : le cas du maroc. *Afrique contemporaine*, (3) :309–324.
- Cao, M., Chychyla, R., and Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting Horizons*, 29(2) :423–429.
- Connelly, R., Playford, C. J., Gayle, V., and Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59 :1–12.
- Elias, P. (2014). Administrative data. *Facing the Future : European Research Infrastructures for the Humanities and Social Sciences*. Berlin : Scivero Verlag, page 47.
- González, P. C. and Velásquez, J. D. (2013). Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Syst. Appl.*, 40 :1427–1436.
- Gupta, S., Keen, M., Shah, A., Verdier, G., and Walutowy, M. F. (2017). *Digital revolutions in public finance*. International Monetary Fund Washington, DC.
- Hall, P., Dean, J., Kabul, I. K., and Silva, J. (2014). An overview of machine learning with sas® enterprise miner™. *SAS Institute Inc*.
- Hoover, J. (2009). States use bi, data warehousing to recoup unpaid taxes. *Intelligent Enterprise*, 12.
- IBM (2013). Big data trends in morocco. *IDG Connect*, 12.
- Kumar, R. (August 8, 2018). How big data can improve tax collection and reduce tax evasion ? Technical report.
- lafinancepourtous.com (2019). Évasion fiscale : définition, causes et impact sur l'activité économique : Mise À jour le 05 avril 2019.
- Laney, D. (2001). 3d data management : Controlling data volume, velocity and variety. *META group research note*, 6(70) :1.
- Malaszczyk, K. and Purcell, B. M. (2017). Big data analytics in tax fraud detection. *Northeastern Association of Business, Economics and Technology*, page 233.
- Manyika, J. (2011). Big data : The next frontier for innovation, competition, and productivity. <http://www.mckinsey>.
- Mwanza, M. and Phiri, J. (2016). Fraud detection on bulk tax data using business intelligence data mining tool : A case of zambia revenue authority. *International Journal of Advanced Research in Computer and Communication Engineering*, 5(3) :2278–1021.
- Pérez López, C., Delgado Rodríguez, M. J., and de Lucas Santos, S. (2019). Tax fraud detection through neural networks : An application using a sample of personal income taxpayers. *Future Internet*, 11(4) :86.
- Rosemary Williams DBA, C. (2008). Using data mining to ensure payment integrity. *The Journal of Government Financial Management*, 57(2) :22.
- Santiago, D. d. S. M. (2018). Taxation, big data and network analytics : An introductory analysis to the global network of double taxation treaties. *Inter-American Center of Tax Administrations (CIAT). Working Papers*, page 28p.
- Sharma, T. (MAY 15, 2015). How can big data be used to reduce tax evasion ? Technical report.
- Tian, F., Lan, T., Chao, K.-M., Godwin, N., Zheng, Q., Shah, N., and Zhang, F. R. (2016). Mining suspicious tax evasion groups in big data. *IEEE Transactions on Knowledge and Data Engineering*, 28 :2651–2664.
- Tian, F., Lan, T., Zheng, Q., Chao, K.-M., Godwin, N., Shah, N., and Zhang, F. R. (2017). Mining suspicious tax evasion groups in big data. *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 25–26.
- Ularu, E. G., Puican, F. C., Apostu, A., Velicanu, M., et al. (2012). Perspectives on big data and big data analytics. *Database Systems Journal*, 3(4) :3–14.
- Vamsi, C. (2017). How modern taxation can benefit from big data analytics.
- Vasarhelyi, M. A., Kogan, A., and Tuttle, B. M. (2015). Big data in accounting : An overview. *Accounting Horizons*, 29(2) :381–396.
- Warren Jr, J. D., Moffitt, K. C., and Byrnes, P. (2015). How big data will change accounting. *Accounting Horizons*, 29(2) :397–407.
- Wu, R.-S., Ou, C.-S., Lin, H.-y., Chang, S.-I., and Yen, D. C. (2012). Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications*, 39(10) :8769–8777.
- Wu, X., Zhu, X., Wu, G.-Q., and Ding, W. (2013). Data mining with big data. *IEEE transactions on knowledge and data engineering*, 26(1) :97–107.
- Zhang, J., Yang, X., and Appelbaum, D. (2015). Toward effective big data analysis in continuous auditing. *Accounting Horizons*, 29(2) :469–476.