

O Uso de Inteligência Artificial no Combate à Evasão Fiscal: Uma Revisão Sistemática da Literatura

Glauco de Vasconcelos Soares¹, Rodrigo C. L. V. Cunha¹, Fernando Erico de Medeiros Filho¹

¹Centro de Estudos e Sistemas Avançados do Recife (CESAR SCHOOL)
AV. Cais do Apolo, 77 – Recife, PE – Brasil

{gvs2, rclvc, femf}@cesar.school

Abstract. *Tax evasion is a problem faced by governments around the world. The use of AI has been a viable alternative to combat this problem. In this context, this work aims to identify how AI helps to combat tax evasion. For this, a systematic literature review was developed, finding 738 articles, of which 18 were selected for this review. The review investigates the types of algorithms used, how to validate the models and the challenges encountered during the research. The results demonstrate an overview of the latest research in this area, gaps in current research and classification of techniques used recently.*

Resumo. *Evasão fiscal é um problema enfrentado pelos governos de todo mundo. O uso de IA tem sido uma alternativa viável para combater esse problema. Neste contexto, esse trabalho tem como objetivo identificar como a IA auxilia no combate à evasão fiscal. Para isso, foi desenvolvido uma revisão sistemática da literatura, encontrando 738 artigos, dos quais 18 foram selecionados para esta revisão. A revisão investiga os tipos de algoritmos utilizados, a forma de validação dos modelos e os desafios encontrados durante as pesquisas. Os resultados demonstram um panorama geral das últimas pesquisas desta área, as lacunas nas pesquisas atuais e classificação das técnicas utilizadas recentemente.*

1. Introdução

A evasão fiscal é um fenômeno global que afeta a sociedade como um todo. Este fenômeno pode ser descrito como um ato intencional de manipulação sobre uma declaração fiscal com o objetivo de obter ilegalmente um benefício financeiro, reduzindo assim sua carga tributária [de Roux et al. 2018]. O não cumprimento das obrigações fiscais tem impacto negativo na qualidade de vida dos cidadãos, pois sem receita tributária não é possível manter os serviços essenciais, tais como: serviços de saúde, saneamento, mobilidade, segurança, educação, entre outros [Mathews et al. 2018, Matos et al. 2019, Xiangyu et al. 2018].

Estima-se que uma empresa típica tenha perdas equivalentes a 5% do seu faturamento anual por motivos de fraude [Baesens et al. 2015]. Esse tipo de problema não é enfrentado apenas pelas empresas, fraude fiscal é um dos principais problemas enfrentados pelos governos, tendo um forte impacto econômico, político e social [Matos et al. 2019].

Segundo o Sindicato Nacional dos Procuradores da Fazenda Nacional (SINDPROFAZ), em 2018 o Brasil deixou de arrecadar 345 bilhões de reais em razão de sonegação fiscal¹. Outros dados apresentados por [Didimo et al. 2020] demonstram que os Estados Unidos estimam perdas de 458 bilhões de dólares anuais, enquanto que os países do grupo Europeu têm perdas estimadas de 147 bilhões de euros por ano. Já a China, segundo [Wu et al. 2019], perde 9.99% do seu produto interno bruto por evasão fiscal.

Para lidar com esse problema, autoridades tributárias do mundo inteiro estão implementando sistemas de faturamento eletrônico [Adamov 2019]. Tais sistemas auxiliam na gestão tributária, e tem como objetivo aumentar a fiscalização e arrecadação, ao mesmo tempo que facilita o recolhimento do imposto perante os contribuintes² [Dias and Becker 2017]. Embora esses sistemas tenham agilizado o processo de coleta de informação, o número de fraudes ainda é alto, representando perdas de 25% da arrecadação devido a fraudes recorrentes [Matos et al. 2015]. Contudo, as informações contidas nessas declarações trazem novas oportunidades para detecção automática das atividades fraudulentas, já que a partir delas é possível encontrar detalhes das transações realizadas pelos contribuintes.

O interesse pela prevenção e detecção automática de fraudes tem tido um crescimento significativo nos últimos anos [Wu et al. 2019]. Conforme mencionado por Baensens et al. (2015), diferentes tipos de técnicas de análise preditiva de fraudes vêm sendo desenvolvidas tanto na indústria quanto na academia, e tem como base uma variedade de disciplinas como, estatística, aprendizado de máquina, inteligência artificial, reconhecimento de padrões e mineração de dados.

Diante da problemática apresentada, o objetivo deste trabalho é compreender como é usada atualmente a IA para combater a evasão fiscal nos governos.

O restante deste artigo está estruturado da seguinte forma. A Seção 2 apresenta o protocolo aplicado que conduziu a revisão sistemática. Nas Seções 3 e 4 são descritos os resultados da pesquisa e algumas discussões. Por fim, a Seção 5 apresenta algumas conclusões a partir dos resultados encontrados e direções para trabalhos futuros.

2. Protocolo Aplicado

O objetivo desta revisão é identificar o estado da arte das soluções de IA no contexto da evasão fiscal. Para esse fim, foi adaptado o protocolo baseado em [Kitchenham and Charters 2007] e o modelo de revisão utilizado por [Dybå and Dingsøyr 2008]. Este processo de revisão é composto pelas seis etapas seguintes: (1) definição do protocolo, (2) identificação dos critérios de inclusão e exclusão, (3) pesquisa dos estudos relevantes, (4) avaliação crítica, (5) extração de dados, e (6) síntese.

Para identificar os estudos disponíveis relacionados ao tópico de interesse desta revisão, foi elaborada a seguinte pergunta:

¹ <https://congressoemfoco.uol.com.br/economia/brasil-deixou-de-arrecadar-mais-de-r-345-bilhoes-por-sonegacao-em-2018/> (acessado em 18/04/2020)

² Contribuinte é o sujeito passivo de uma obrigação tributária. Em outros termos, é aquele que se sujeita, por previsão legal, ao pagamento de tributos ao fisco.

- *Como o uso de IA pode auxiliar no combate à evasão fiscal?*

A partir dessa pergunta principal, outras questões de pesquisa (QP) foram desenvolvidas para ajudar na compreensão do problema:

- QP1: *Quais tipos de técnicas e algoritmos estão sendo aplicados no combate à evasão fiscal?*
- QP2: *Como são avaliados e quais os resultados obtidos com o uso de IA na administração tributária?*
- QP3: *Quais os principais desafios encontrados no processo de detecção automática de sonegadores de impostos?*

2.1. Critérios de inclusão e exclusão

Para esta revisão foram considerados os estudos relacionados a aprendizado de máquina e mineração de dados, cujo objetivo seja construir modelos ou ferramentas que ajudem os governos a combater a evasão fiscal.

Com o objetivo de avaliar as pesquisas mais recentes, esta revisão limitou-se a analisar apenas os trabalhos publicados a partir de 2017, além deste filtro, também foram adicionados os seguintes critérios de exclusão:

- Estudos não publicados na língua inglesa;
- Estudos cujo a solução do problema não esteja no contexto de IA;
- Artigos resumidos ou ainda em progresso, pôsteres e estudos secundários;
- Que não demonstrem evidências empíricas;

2.2. Estratégia de busca

Para encontrar os estudos utilizados nesta revisão, considerou-se as seguintes bases de dados eletrônicas de publicação:

- ACM Digital Library;
- IEEE Xplore;
- ScienceDirect - Elsevier;

O termo de busca utilizado está relacionado ao objetivo central deste trabalho, e com base nas questões de pesquisa, foi elaborada a seguinte *string* de busca:

- *("tax fraud" OR "tax evasion" OR "fiscal fraud" OR "tax audit" OR "taxpayer" OR "tax compliance") AND ("mining" OR "machine learning" OR discover* OR detect* OR identif* OR classif* OR predict*)*

Para o processo de extração de informação nas bases de dados, foi aplicada a *string* de busca separadamente em cada base eletrônica. As buscas foram realizadas entre março e abril de 2020, os resultados foram agrupados por base, conforme demonstrado na Tabela 1.

Tabela 1. Total de estudos selecionados em cada base de dados

Base de dados	Quantidade de artigos
ACM Digital Library	91
IEEE Xplorer	18
ScienceDirect - Elsevier	629

2.3. Processo de seleção dos artigos

Nesta etapa da revisão, foi utilizado o aplicativo Rayyan, desenvolvido pelo Qatar Computing Research Institute (QCRI), como uma ferramenta auxiliar para arquivamento, organização e seleção dos artigos [Ouzzani et al. 2016].

O processo de seleção dos estudos foi realizado em quatro etapas, aplicando-se os critérios de inclusão e exclusão, conforme descrito na Tabela 2. Na primeira etapa, foi aplicada a *string* de busca nas bases de dados, encontrando-se 738 artigos, que foram catalogados na ferramenta Rayyan por título, autor, ano e base de dados.

Na segunda etapa, realizou-se a leitura dos títulos, excluindo-se todos os trabalhos cujos temas abordados não eram compatíveis com os tópicos de interesse deste estudo, restando, ao final desta etapa, 38 artigos.

Na terceira etapa, após a leitura dos resumos, restaram 22 artigos. E após a leitura dos artigos na íntegra (quarta etapa), sobrou um total de 18 estudos primários.

Tabela 2. Resultado das etapas do processo de seleção dos artigos

Etapas do processo de seleção	Quantidade de artigos
Estudos encontrados	738
Análise do título	38
Análise do resumo	22
Leitura completa	18

2.4. Avaliação da Qualidade

Para determinar a credibilidade, rigor e relevância dos artigos, foram elaboradas 7 questões baseadas em [Kitchenham and Charters 2007], que foram utilizadas para guiar a avaliação da qualidade das evidências coletadas:

- Q1: *O estudo avalia o uso de IA como ferramenta de apoio a evasão fiscal?*
- Q2: *O estudo apresenta um modelo para detecção de evasão fiscal?*
- Q3: *Os objetivos do estudo são informados de forma clara?*
- Q4: *O contexto do estudo é descrito de maneira adequada?*
- Q5: *O projeto de pesquisa foi adequado para alcançar os objetivos da pesquisa?*
- Q6: *Os resultados da pesquisa foram validados adequadamente?*
- Q7: *O estudo contribui para a pesquisa ou traz melhorias significativas no ambiente onde foi aplicado?*

O processo de avaliação de qualidade será descrito na Seção de resultados, no qual serão avaliados os 18 estudos selecionados.

3. Resultados

Esta Seção apresenta uma visão geral dos estudos selecionados para o processo de extração de dados e síntese, apresentados na Seção 2. Em seguida, serão apresentados os resultados da aplicação dos critérios de qualidade, e uma breve análise das soluções.

3.1. Análise quantitativa

O desenvolvimento desta pesquisa resultou em um processo de análise de 738 estudos encontrados nas já citadas bases digitais, dos quais 18 foram selecionados por responder as perguntas que guiaram esta revisão. Estes trabalhos foram publicados por 64 autores vinculados a diversas instituições localizadas em diferentes países, situadas em 4 regiões do globo (Europa, Oriente Médio, Ásia e América do Sul).



Figura 1. Nuvem de palavra das palavras-chaves

Foram identificadas 84 palavras-chaves nos trabalhos selecionados, das quais, as mais relevantes com suas respectivas frequências são: *tax evasion* (6), *data mining* (4), *tax fraud detection* (4), *tax evasion detection* (2), *social network analysis* (2). A figura 1 apresenta a nuvem de palavras geradas com todas as palavras-chaves.

3.2. Análise qualitativa

Na fase de análise qualitativa, conforme descrito na Seção de avaliação de qualidade, cada um dos estudos foi examinado e investigado com base nas sete questões de qualidade. Os critérios de qualidade definidos fornecem uma boa medida para avaliar o quanto cada estudo é relevante para esta revisão.

A tabela 3 apresenta o resultado da avaliação de qualidade. Cada linha representa um estudo e as colunas de Q1 a Q7 representam as 7 questões descritas na Seção anterior. Para cada questão, a resposta '1' representa que o trabalho respondeu positivamente, e '0' que o estudo não respondeu.

Tabela 3. Análise qualitativa dos estudos primários

Estudo	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Total
[de Roux et al.2018]	1	1	1	1	1	1	1	7
[Ying et al. 2019]	1	1	1	0	0	0	0	3
[Wei et al. 2019]	1	1	1	1	1	1	1	7
[Jupri and Sarno 2018]	1	1	1	0	1	0	0	4
[Xiangyu et al. 2018]	1	1	1	0	0	1	0	4
[Zhu et al. 2018]	1	1	1	1	1	1	1	7
[Mathews et al. 2018]	1	1	0	0	1	0	0	3
[Mehta et al. 2018]	1	1	0	0	1	1	0	4
[Didimo et al. 2020]	1	1	1	1	1	1	1	7
[Wu et al. 2019]	1	1	1	1	1	1	1	7
[Mehta et al. 2019]	1	1	1	1	0	1	0	5
[Matos et al. 2019]	1	0	1	1	1	1	1	6
[Rahimikia et al. 2017]	1	1	1	1	1	1	1	7
[Kleanthous and Chatzis 2019]	1	1	1	1	1	1	1	7
[Lismont et al. 2018]	1	1	1	1	1	1	1	7
[Goumagias et al. 2018]	1	1	1	0	1	0	1	5
[Ruan et al. 2019]	1	1	1	1	1	1	1	7
[Vanhoeyveld et al. 2020]	1	1	1	1	1	1	1	7
Total	18	17	16	11	15	14	11	

Observamos que todos os estudos analisados nesta fase tiveram respostas positivas para a questão Q1. Apenas um dos estudos respondeu negativamente à questão Q2, embora este permaneça relevante para o estudo, pois nele é apresentado um método de seleção de atributos específico para modelos de detecção de fraude. Em quatro estudos, os autores não validaram adequadamente seu modelo de predição (Q6) e sete trabalhos não apresentaram contribuições significativas (Q7).

Do total de 18 artigos, 10 estudos obtiveram pontuação máxima, respondendo positivamente a todas as questões de qualidade, enquanto que 02 estudos responderam positivamente a apenas 3 questões.

3.3. Análise das soluções

Após a leitura e a avaliação dos estudos primários, as soluções propostas pelos autores foram categorizadas em 3 grupos com base no tipo de técnica utilizada no modelo de detecção de evasão fiscal. Alguns trabalhos utilizaram mais de uma técnica, nestes, procurou-se identificar qual abordagem mais proeminente foi utilizada na solução, e esta prevaleceu para o enquadramento do estudo. A Tabela 4 sumariza as categorias identificadas juntamente com o número de trabalhos enquadrados em cada grupo.

Tabela 4. Técnicas utilizadas para detecção de evasão fiscal

Técnica	Estudo	Total
Supervisionado	[Ying et al. 2019], [Jupri and Sarno 2018], [Xiangyu et al. 2018], [Zhu et al. 2018], [Wu et al. 2019], [Rahimikia et al. 2017], [Goumagias et al. 2018], [Kleanthous and Chatzis 2019]	8
Não supervisionado	[de Roux et al. 2018], [Wei et al. 2019], [Mehta et al. 2019], [Vanhoeyveld et al. 2020], [Mathews et al. 2018], [Mehta et al. 2018]	6
Redes complexas	[Didimo et al. 2020], [Lismont et al. 2018], [Ruan et al. 2019], [Matos et al. 2019]	4

Os trabalhos enquadrados na categoria de aprendizado de máquina “supervisionado” consistem em soluções baseadas em dados históricos, ou seja, dados de auditoria que foram previamente realizadas, informando que um determinado contribuinte já foi fiscalizado por suspeita de evasão.

No entanto, a capacidade dos auditores de identificar os sonegadores é baixa devido às limitações de recursos dos órgãos de fiscalização e ao alto número de contribuintes e do volume de dados, refletindo diretamente na quantidade de dados rotulados [de Roux et al. 2018, Wei et al. 2019, Wu et al. 2019, Vanhoeyveld et al. 2020, Zhu et al. 2018].

Para lidar com esse problema, o trabalho de [Wu et al. 2019] propõe um modelo para detecção de evasão fiscal para ambientes que contenham poucos dados rotulados utilizando a técnica *Positive and Unlabeled Learning (PU)*. O intuito é calcular a probabilidade de evasão baseada nos dados históricos, desta forma é possível aumentar a quantidade de dados rotulados para que o modelo de previsão possa ser melhor treinado. [Kleanthous and Chatzis 2019] utiliza *Deep Neural Network* para seleção de casos de auditoria em bases com poucos dados rotulados, a solução é baseada no Imposto sobre Valor Agregado (IVA).

Em [Zhu et al. 2018] é apresentado um novo método de identificação dos sonegadores inter-regional, essa abordagem utiliza técnicas de Aprendizado por Transferência, ou seja, o modelo transfere o conhecimento de uma região para outra, evitando eu necessite de um novo treinamento para a nova região com pouco ou nenhum dado rotulado.

Na categoria de aprendizado de máquina “não supervisionado”, os sonegadores são identificados através da análise do comportamento individual perante um grupo, desta forma, os chamados *outliers* são marcados como suspeitos. Neste contexto, [de Roux et al. 2018] utilizou *Spectral Clustering* para detecção de declarações subfaturadas para o setor de construção civil. No trabalho de [Vanhoeyveld et al. 2020] foi proposto uma solução baseada em *K-Nearest Neighbor (KNN)* para detecção de fraude em declarações de IVA para 10 setores econômicos na Bélgica.

Na última categoria foram classificados os trabalhos que utilizaram “redes complexas” para detecção de evasão fiscal, esta técnica permite um mapeamento de redes sociais dos contribuintes, onde é possível identificar características e comportamento fraudulento entre as empresas e seus membros, construindo novas

variáveis que foram utilizadas no modelo de predição, por tanto, todas elas são abordagens híbridas [Didimo et al. 2010, Lismont et al. 2018, Ruan et al. 2019].

4. Discussão

Esta Seção sumariza os resultados da revisão sistemática realizada, considerando as três questões de pesquisa definidas na Seção 2.

4.1. IA como ferramenta de combate à evasão fiscal

Durante esta revisão foi possível identificar vários cenários onde foram aplicadas soluções de IA no contexto de evasão fiscal. Nos trabalhos [Ying et al. 2019, Zhu et al. 2018] foram desenvolvidos modelos inteligentes para seleção automática de casos de auditoria. A seleção de casos é parte fundamental do processo de fiscalização realizado pelas autoridades fiscais.

Em [de Roux et al. 2018, Mehta et al. 2019], são apresentados modelos de IA para reconhecimento de fraude em declarações fiscais subfaturadas, ou seja, os dados das declarações são analisados com o objetivo de encontrar indícios de sonegação fiscal.

Além disso, o uso de IA também tem sido frequentemente utilizado para combater a inadimplência, de modo a serem asseguradas a regularidade e a previsibilidade da arrecadação [Jupri and Sarno 2018, Rahimikia et al. 2017]. Paralelo a isso, as previsões de receita através de regressões lineares também foram apontadas no estudo [Mathews et al. 2018].

4.2. Avaliação dos resultados

Existem várias métricas possíveis para apresentar e analisar o resultado de uma classificação. Geralmente é calculado a acurácia do modelo, ou seja, a proporção de exemplos que são classificados corretamente. Nos estudos avaliados as principais métricas utilizadas foram as seguintes: Acurácia, *F-Measure*, Curva ROC e Matriz de confusão.

Em [Jupri and Sarno 2018] são comparados 5 algoritmos de classificação no contexto de conformidade fiscal, são eles: *Decision Trees C4.5*, *Support Vector Machine* (SVM), KNN, *Naive Bayes* e *Multilayer Preceptron* (MPL). O algoritmo que obteve melhor resultado foi o *Decision Trees C4.5*, com 98.93% de acurácia, enquanto que o MPL teve 89.26% de acurácia, ficando na última posição do *ranking*.

A solução proposta em [Zhu et al. 2018] utiliza o algoritmo *LightGBM* para seleção de casos de auditoria, este algoritmo superou os classificadores SVM e MLP, com 86.80% para a métrica *F-Measure*.

No estudo [Lismont et al. 2018], foi construído um modelo de predição de elisão fiscal por meio de análise de redes sociais. Durante os experimentos, os autores comparam os algoritmos *Logistic Regression* (LR) e *Random Forest* (RF), sendo que o primeiro obteve um melhor resultado de acurácia, com 89.53%, contra 89.13% do segundo classificador. No entanto, analisando a métrica da área sob a curva ROC, o RF apresentou a melhor pontuação, com 84.12%, enquanto que o LR apresentou 83.94%.

4.3. Desafios identificados

Um dos desafios reportados por [de Roux et al. 2018, Wei et al. 2019, Vanhoeyveld et al. 2020] está relacionado à escassez de dados históricos de evasão fiscal. Este problema tende a diminuir na medida que sistemas de detecção estão sendo implantados e validados pelas agências fiscalizadoras.

Outro desafio apresentado por [Kleanthous and Chatzis 2019, Wu et al. 2019] diz respeito à falta de uma base pública de dados de fiscalização, isso dificulta o desenvolvimento de modelos de detecção, já que a maioria dos desenvolvedores não tem acesso a esse tipo de dado.

[Didimo et al. 2019, Rahimikia et al. 2017, de Roux et al. 2018, Wu et al. 2019] afirmam que o processo de auditoria é muito demorado, isso atrapalha a evolução dos modelos de predição, dado que a validação *in loco* leva em média de 4 a 6 meses.

Por fim, outro desafio reportado por [Didimo et al. 2019, Matos et al. 2019] está relacionado à seleção de variáveis para treinamento do modelo. As variáveis fiscais geralmente têm baixa ou nenhuma correlação, o que dificulta a seleção de variáveis, já que os algoritmos de *machine learning* não trabalham bem com um número grande de variáveis.

5. Conclusão

A partir da revisão sistemática apresentada neste trabalho foi possível traçar um panorama geral de como a IA tem contribuído no combate à evasão fiscal. Foram encontrados 738 artigos na primeira etapa de busca, dos quais 18 foram classificados como estudos primários a partir de critérios de seleção e qualidade pré-definidos.

Durante a pesquisa, foi possível classificar os tipos de técnicas utilizadas nas soluções apresentadas nos trabalhos: aprendizado supervisionado, aprendizado não supervisionado e análise de redes sociais. Também foram identificados diversos algoritmos de IA e métodos de avaliação dos modelos que podem e devem ser utilizados em conjunto, desde que os resultados obtidos superem o uso individual das técnicas.

A principal limitação desta pesquisa é dada pelo fato de que tanto a seleção dos estudos quanto a análise qualitativa foram realizadas apenas por um pesquisador. Isso prejudica a imparcialidade do protocolo, possibilitando posturas tendenciosas durante as análises, uma vez que nem todos os trabalhos são apresentados de forma clara. Para minimizar esta questão, o protocolo e as questões de pesquisa foram elaborados antecipadamente.

Para futuras pesquisas a serem realizadas, pode-se estender o número de bases científicas de busca, e também podem ser alterados os critérios de inclusão e exclusão. Os desafios destacados na Seção anterior também estão abertos para resolução dos problemas.

Referências

Adamov, A. Z. (2019) “Machine Learning and Advanced Analytics in Tax Fraud Detection”, 2019 IEEE 13th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan, pp. 1-5.

- Baesens, B., Vlasselaer V. V. and Verbeke W. (2015) "Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection". New Jersey: John Wiley & Sons.
- de Roux, D.; Pérez, B.; Moreno, A.; Villamil, P. and Figueroa, C. (2018). "Tax Fraud Detection for Under-Reporting Declarations Using an Unsupervised Machine Learning Approach". In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18), London, UK, pp. 2015-222.
- Dias, M. and Becker, K. (2017). "Identificação de Candidatos à Fiscalização por Evasão do Tributo ISS". In: 5th Symposium on Knowledge Discovery, Mining and Learning, Uberlândia, MG, 2017.
- Didimo, W., Grilli, L., Liotta, G., Menconi, L., Montecchiani, F. and Pagliuca D. (2020) "Combining Network Visualization and Data Mining for Tax Risk Assessment", in IEEE Access, vol. 8, pp. 16073-16086.
- Dybå, T. and Dingsøyr, T. (2008). "Empirical studies of agile software development: A systematic review". Information and Software Technology. vol. 50, pp. 833-859.
- Goumagias, N., Hristu-Varsakelis, D., and Assael, Y. (2018). "Using Deep Q-learning to understand the tax evasion behavior of risk-averse firms". Expert Systems with Applications, vol. 101, pp. 258-270.
- Jupri, M. and Sarno R. (2018) "Taxpayer compliance classification using C4.5, SVM, KNN, Naive Bayes and MLP". 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, pp. 297-303.
- Kitchenham, B. and Charters, S. (2007) "Guidelines for performing systematic literature reviews in software engineering".
- Kleanthous, C., and Chatzis, S.P. (2019). "Gated Mixture Variational Autoencoders for Value Added Tax audit case selection". Knowledge-Based Systems, vol. 188, 105048.
- Lismont, J., Cardinaels, E., Bruynseels, L., Groote, S.D., Baesens, B., Lemahieu, W., and Vanthienen, J. (2018). "Predicting tax avoidance by means of social network analytics". Decision Support Systems, vol. 108, pp. 13-24.
- Mathews, J., Mehta, P., Kuchibhotla, S., Bisht, D., S. Chintapalli, B. and Rao S. V. K. V.(2018) "Regression Analysis towards Estimating Tax Evasion in Goods and Services Tax", 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Santiago, pp. 758-761.
- Matos, T., Macedo, J. A., Lettich, F., Monteiro, J. M., Renso, C., Perego and R., Nardini, F. M. (2019). "Leveraging Feature Selection To Detect Potential Tax Fraudsters". Expert Systems with Applications. vol. 145, 113128.
- Matos, T.; Macedo, J. A. and Monteiro, J. M. (2015) "An Empirical Method for Discovering Tax Fraudsters". In: IDEAS '15: Proceedings of the 19th International Database Engineering & Applications Symposium, Yokohama, Japão, p. 41-48.

- Mehta, P., Mathews, J., Rao, S. V. K.V., Kumar, K. S., Suryamukhi, K., Babu C. S. (2019) "Identifying Malicious Dealers in Goods and Services Tax", 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA), Suzhou, China, pp. 312-316.
- Mehta, P., Mathews, J., Suryamukhi, K., Kumar, K. S. and Babu, C. S. (2018) "Predictive Modeling for Identifying Return Defaulters in Goods and Services Tax", 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, pp. 631-637.
- Okoli, C., Schabram, K. (2010) "A Guide to Conducting a Systematic Literature Review of Information Systems Research," . Sprouts: Working Papers on Information Systems, 10(26).
- Ouzzani M., Hammady H., Fedorowicz Z. and Elmagarmid A. (2016) "Rayyan-a Web and Mobile App for Systematic Reviews". Syst Rev. 2016; 5(1): 210.
- Ying, Q., Xiaoxin, H. and Weige, J. (2019). "Research on Tax Inspection Case Selection Model Based on Bayesian Network". In Proceedings of the 2019 2nd International Conference on Information Management and Management Sciences (IMMS 2019). Association for Computing Machinery, New York, NY, USA, pp. 198–202.
- Rahimikia, E., Mohammadi, S., Rahmani, T. and Ghazanfari, M. (2017). "Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran". International Journal of Accounting Information Systems. Vol. 25, pp. 1-17.
- Ruan, J., Yan, Z., Dong, B., Zheng, Q., Qian, B. (2018). "Identifying Suspicious Groups of Affiliated-Transaction-based Tax Evasion in Big Data". Information Sciences, vol. 477, pp. 508-532.
- Vanhoeyveld, J., Martens, D., Peeters, B. (2020). "Value-added tax fraud detection with scalable anomaly detection techniques". Applied Soft Computing, vol. 86, 105895.
- Wei, R., Dong, B., Zheng, Q., Zhu, X., Ruan, J. and He, H. (2019). "Unsupervised Conditional Adversarial Networks for Tax Evasion Detection". 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, pp. 1675-1680.
- Wu, Y., Zheng, Q., Gao, Y., Dong, B., Wei, R., Zhang, F. and He, H. (2019) "TEDM-PU: A Tax Evasion Detection Method Based on Positive and Unlabeled Learning", 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, pp. 1681-1686.
- Xiangyu, X., Youlin, Y. and Qicheng, X. (2018) "Intelligent Identification of Corporate Tax Evasion Based on LM Neural Network", 2018 37th Chinese Control Conference (CCC), Wuhan, pp. 4507-4511.
- Ying, Q., Xiaoxin, H. and Weige, J. (2019). "Research on Tax Inspection Case Selection Model Based on Bayesian Network". In: Proceedings of the 2019 2nd International Conference on Information Management and Management Sciences (IMMS 2019), Chengdu, China, pp. 198-202.

Zhu, X., Yan, Z., Ruan, J., Zheng, Q., Dong, B. (2018) “IRTED-TL: An Inter-Region Tax Evasion Detection Method Based on Transfer Learning,” 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, pp. 1224-1235.