

## APLICAÇÃO DA LÓGICA NEBULOSA NA DETECÇÃO DE FRAUDE DE ICMS

**Rafael R. Aguiló\*** e **Giovane Quadrelli**

*Centro de Engenharia e Computação, Universidade Católica de Petrópolis, Petrópolis, RJ, Brasil*

**Palavras-chave:** Fraude, ICMS, Lógica Nebulosa, Rio de Janeiro, SEFAZ-RJ.

**Resumo.** Nos estados brasileiros, a arrecadação de ICMS (Imposto sobre a circulação de mercadorias e serviços de transporte interestadual e intermunicipal e de comunicação) é a principal fonte de recursos que permite ao governo ter saúde financeira para prover todos os benefícios à sua população. A percepção de comportamentos suspeitos, que indiquem uma conduta de cometimento de fraude, em relação ao recolhimento do imposto devido, é imprescindível ao trabalho de se manter uma arrecadação equilibrada e consistente com a economia estadual. A modernização dos métodos atualmente utilizados com esse fim é essencial para que o estado melhore a eficácia da sua função. Este estudo tem o objetivo de mostrar que a utilização da lógica nebulosa pode melhorar o trabalho de recuperação dessa importante receita, selecionando mais assertivamente os possíveis fraudadores. Um modelo nebuloso foi construído a fim de ser aplicado na identificação de possíveis desvios de conduta, caracterizados por um esquema de fraude, onde empresas fornecedoras são utilizadas para emitir notas fiscais falsas, com o propósito de gerar crédito tributário às empresas compradoras. O resultado seria uma maior e mais rápida recuperação da receita do estado, melhorando sua situação financeira e, consequentemente, permitindo mais investimentos no bem-estar de sua população. Apesar de aplicado ao Rio de Janeiro, o resultado obtido neste estudo tem o potencial de contribuir em favor de qualquer estado brasileiro, pois trata-se de um problema nacional.

---

E-mail addresses: [rafael.41940043@ucp.br](mailto:rafael.41940043@ucp.br)\*, [giovane.quadrelli@ucp.br](mailto:giovane.quadrelli@ucp.br).

## **APPLYING FUZZY LOGIC IN ICMS TAX FRAUD DETECTION**

**Keywords:** Fraud, Fuzzy Logic, ICMS, Rio de Janeiro, SEFAZ-RJ.

### **Abstract.**

In Brazilian states, the collection of ICMS (Tax on the circulation of goods and services of interstate and intercity transport and communication) is the main source of resources that allows government to have financial health to provide all the benefits to its population. The perception of suspicious behavior, which indicates a conduct of committing fraud, in relation to the collection of the tax due, is essential to the work of maintaining a balanced and consistent collection with the state economy. The modernization of the methods currently used for this purpose is essential for the state to improve the effectiveness of its function. This study aims to show that the use of fuzzy logic improves the recovery work of this important revenue, selectively selecting potential fraudsters more assertively. A Fuzzy Logic model was built to be applied in the identification of possible misconduct, characterized by a fraud scheme, where supplier companies are used to issue false invoices, with the purpose of generating tax credit for other companies. The results are a greater and faster recovery of the state's revenue, improving its financial situation and, consequently, allowing more investments in the well-being of its population. Although applied to Rio de Janeiro, the result obtained in this study has the potential to contribute in favor of any Brazilian state, as it is a national problem.

## 1 INTRODUÇÃO

Junto com o aumento da facilidade na abertura de novas empresas nos últimos anos, infelizmente, percebe-se também o aumento da criação de empresas fantasmas para fraudar o ICMS. Conhecidas como noteiras, essas empresas fantasmas emitem notas fiscais falsas com o propósito de gerar crédito tributário a supostas empresas compradoras. Identificar e impedir o funcionamento desse tipo de empresa, o mais rápido possível, é primordial para evitar a perda de receitas no sistema de arrecadação do Estado do Rio de Janeiro. Assim, é necessário que o trabalho de fiscalização seja aperfeiçoado para se ter maior eficiência e eficácia na procura de fraudadores que prejudicam a economia do estado. O objetivo deste trabalho é construir um modelo de lógica nebulosa com a finalidade de obter um considerável ganho sobre o modelo atual de identificação de fraudadores, buscando maior rapidez e impessoalidade para o processo.

As empresas noteira fraudam o ICMS por meio de uma prática que se aproveita do direito ao crédito do valor do imposto devido pela empresa fornecedora, responsável pelo recolhimento inicial desse tributo. Supondo um exemplo de funcionamento normal de arrecadação, uma empresa *A* vende uma mercadoria a uma empresa *B* que, por sua vez, vende a mesma mercadoria, beneficiada ou não, a uma empresa *C*. A empresa *A* recolhe ICMS por conta da venda à empresa *B* e esta também recolhe o ICMS, agora sobre o valor final, ou seja, parte do imposto estaria sendo cobrado novamente. Acontece que de acordo com a história do ICMS [9], seguindo o princípio da não-cumulatividade, créditos são gerados à empresa *B* de modo que, após a apuração de suas entradas e saídas de mercadorias, esta recolha ao estado somente o imposto sobre a diferença entre o que entrou e saiu. A Figura 1 ilustra o processo normal de recolhimento de ICMS.

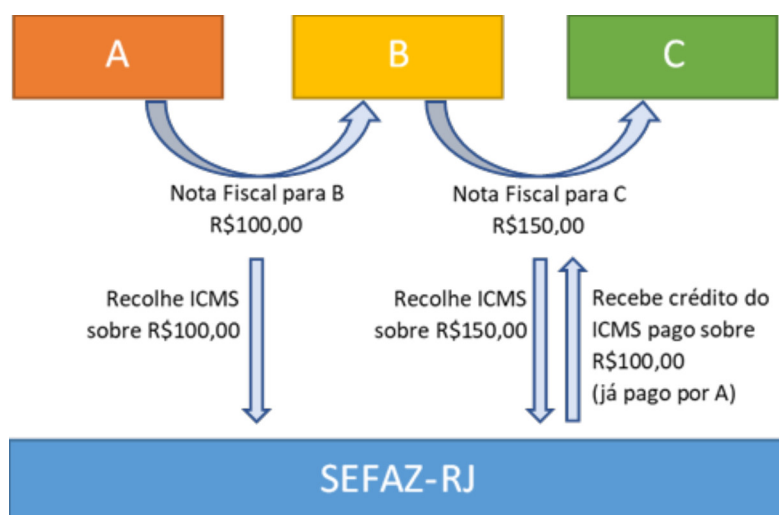


Figura 1: Processo de recolhimento de ICMS

A fraude tem início quando as empresas fantasmas (*A*) são criadas com o único intuito de emitir notas fiscais “frias”, ou seja, sem que aconteçam vendas reais, contra outras empresas (*B*) para que esta se beneficie dos créditos anteriormente citados.

O problema acontece, efetivamente, quando a empresa emissora (*A*) não paga os impostos devidos e encerra ou abandona suas atividades, nunca recolhendo o ICMS devido, como destacado na Figura 2.

Identificar tais contribuintes é um processo trabalhoso que depende da função humana para realizar diversos cruzamentos de dados em busca de padrões de comportamento que se adéquam

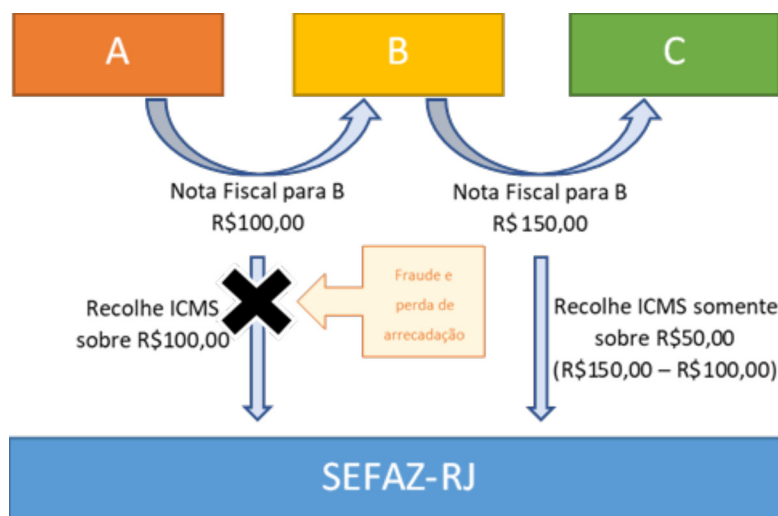


Figura 2: Processo de fraude

ao perfil das empresas noteiras para, então, se chegar à decisão de elencar as empresas que devem passar por um processo de fiscalização para a confirmação de sua existência e da sua idoneidade. A seleção rígida, obtida pela aplicação da lógica clássica, não se mostra eficaz quando, por exemplo, seleciona-se um conjunto de empresas por faixa de arrecadação. Supondo, num caso hipotético, que sejam selecionadas empresas que arrecadem mais que R\$10.000,00 (dez mil reais). Isto não quer dizer que as empresas que arrecadaram R\$9.999,00 (nove mil, novecentos e noventa e nove reais), ou seja, apenas 1 real a menos, não sejam relevantes. Se utilizarmos a teoria clássica dos conjuntos, elas serão ignoradas enquanto que, com a aplicação da teoria nebulosa, elas serão consideradas.

Atualmente, o trabalho de análise e investigação de fraude de ICMS (Imposto sobre Circulação de Mercadorias e Serviços) [6], feito pela Secretaria de Estado de Fazenda do Rio de Janeiro [7], é realizado por meio de ferramentas tradicionais de bancos de dados que utilizam a linguagem SQL [2], baseada na álgebra relacional criada pelo matemático britânico Edgar Codd [3]. Nesse método, realizam-se as operações básicas da teoria clássica dos conjuntos, como junções e interseções, para fazer todos os filtros e cruzamentos de informações necessários a fim de se chegar à identificação de possíveis fraudadores. A desvantagem em se utilizar tais ferramentas está na sua imprecisão em medir o “meio-termo”. Isso não é possível utilizando teoria clássica dos conjuntos porque ela se baseia na lógica Booleana [1], ou seja, são considerados apenas os valores verdadeiro ou falso para tomadas de decisão.

A lógica nebulosa [10] é a ferramenta que possibilita a resolução de problemas que estão entre a precisão da lógica clássica e a imprecisão do mundo real. Ao permitir a utilização de variáveis linguísticas, ao invés de numéricas, ela se aproxima da razão, permitindo que computadores simulem o pensamento humano. Por exemplo, a precisão de números consegue ser traduzida em imprecisão de termos, como “provável” e “improvável”, e vice-versa, de modo que, ao lidar com sistemas nebulosos, o ser humano fica mais próximo da realidade sem abdicar da precisão matemática.

Em [8], utilizou-se a lógica nebulosa como primeira etapa do processo de detecção de empresas fraudulentas. Redes neurais artificiais foram utilizadas em uma segunda etapa. O objetivo dos autores foi o de elencar todas as empresas provavelmente fraudulentas para uma futura ação fiscalizatória. Um outro trabalho, desenvolvido na Secretaria de Fazenda do Estado da Bahia [4], utilizou apenas redes neurais artificiais e atua na detecção de sonegação e não da fraude.

A solução proposta procurou atuar na prevenção da fraude e contribui para o entendimento do comportamento tributário do contribuinte.

## 2 SOBRE LÓGICA NEBULOSA

A lógica nebulosa foi fundamentada pela teoria dos conjuntos nebulosos por Lofti Zadeh [10], nos Estados Unidos. É uma proposta que oferece uma maneira, próxima à percepção humana, para traduzir informações imprecisas e vagas em entidades capazes de serem processadas por um computador. Por exemplo, “arrecada pouco” ou “arrecada o esperado”, que definem os valores percebidos na arrecadação de ICMS de uma empresa. As funções de pertinência [11] são escolhidas, apropriadamente, de acordo com o intervalo de uma variável linguística a ser representada e então, é aplicado um método de *defuzzificação*, como os citados por M. Mizumoto em [5]. Para isso, é fundamental que um especialista no assunto defina os intervalos numéricos para os seus valores linguísticos.

Um sistema de lógica nebulosa é um processo composto por um conjunto de etapas onde, primeiro, ocorrem transformações de variáveis numéricas em variáveis linguísticas e, depois, utilizam-se regras construídas através de expressões linguísticas do tipo *Se  $U_1$  é  $A_1$  e  $U_2$  é  $A_2$  então  $R$  é  $B$*  [11]. Por fim, ocorre a transformação das variáveis linguísticas de volta em variáveis numéricas. A Figura 3 mostra o esquema de um sistema nebuloso.

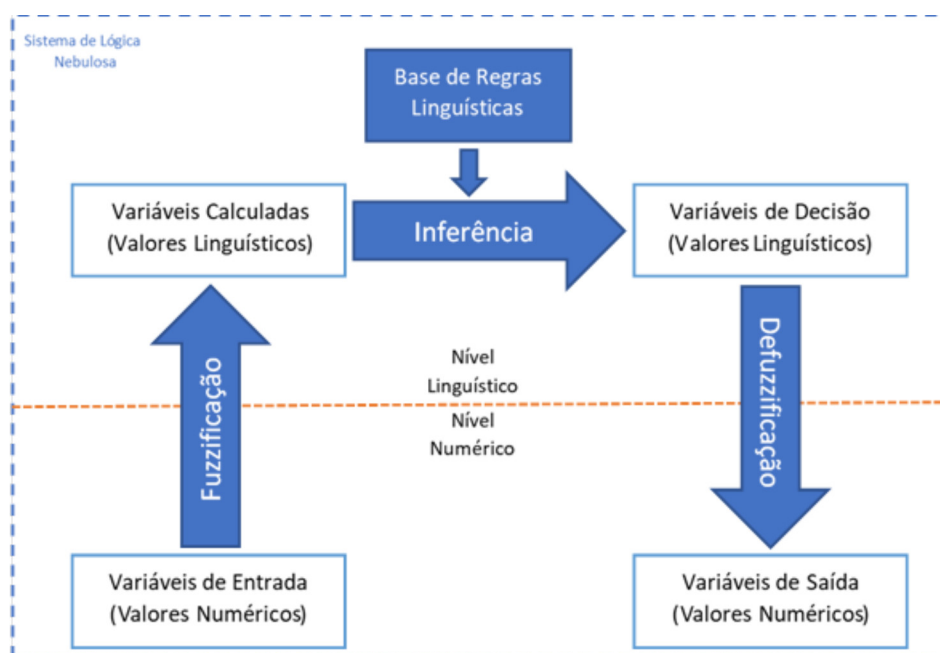


Figura 3: Sistema Nebuloso

A entrada são os dados a serem analisados, representados por valores numéricos. O processo de *fuzzificação* é onde ocorre a conversão das variáveis numéricas em variáveis linguísticas, com seu respectivo grau de pertinência para cada conjunto definidos pelas funções de pertinência. O sistema de inferência realiza a avaliação do conjunto de regras utilizando a função de implicação:

$$\mu_{A \rightarrow B}(u, v) = (\mu_A(u) \wedge \mu_B(v)) \vee (1 - \mu_A(u)) \quad (1)$$

proposta por [10] e gera um conjunto nebuloso de saída contendo todos os resultados das regras.

Por fim, o processo de *defuzzificação* utilizará o conjunto nebuloso resultante para obter um valor numérico para cada variável de saída, aplicando-se um dos métodos de *defuzzificação* [5], como o Método do Centro de Gravidade definido por

$$w^0 = \frac{\sum_i \mu_C(w_i) \cdot w_i}{\sum_i \mu_C(w_i)} \quad (2)$$

### 3 METODOLOGIA

Para a implementação de um sistema nebuloso, os seguintes passos foram seguidos:

- escolha das variáveis de entrada e saída;
- definição dos intervalos do universo discurso das variáveis de entrada e de saída;
- elaboração das regras de inferência;
- escolha das funções de pertinência;
- discretização do universo de entrada e saída;
- escolha do tipo de implicação; e
- implementação do motor de inferência nebuloso.

Começando pelas variáveis de entrada, a variável “Idade da Empresa” representa a quantidade de meses que uma empresa tem desde a data de sua inscrição, e possui três conjuntos nebulosos, como definidos na Tabela 1.

Tabela 1: Conjuntos nebulosos de “Idade da Empresa”

Muito Jovem	até 7 meses
Jovem	entre 4 e 12 meses
Não Jovem	a partir de 10 meses

Para que seja considerado um sistema nebuloso, os intervalos, definidos pelas funções de pertinência das variáveis de entrada, devem ter áreas de sombra, ou seja, as extremidades dos intervalos devem se sobrepôr de modo que esta área de transição demonstre a pertinência de um valor em mais de um conjunto, simultaneamente. As funções de pertinência utilizadas para cada conjunto da variável “Idade da Empresa” foram, respectivamente, função decrescente, função trapezoidal e a função crescente, como pode ser visto na Figura 4.

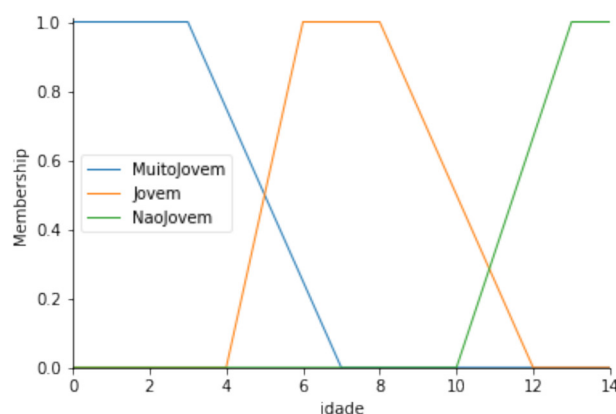


Figura 4: Funções de pertinência de “Idade da Empresa”

A variável “Compras x Vendas” representa a relação entre valores de notas fiscais de entradas e as notas fiscais de saída no mesmo período, e tem quatro conjuntos nebulosos, como descritos na Tabela 2

Tabela 2: Conjuntos Nebulosos de “Compras x Vendas”

Compra Pouquíssimo	entre 0 e 10
Compra Pouco	entre 7,5 e 45
Compra o Esperado	entre 40 e 75
Compra Muito	a partir de 62,5

As funções de pertinência utilizadas para cada conjunto da variável “Compras x Vendas” foram, respectivamente, função decrescente, função triangular, função triangular novamente e a função crescente, como pode ser visto na Figura 5.

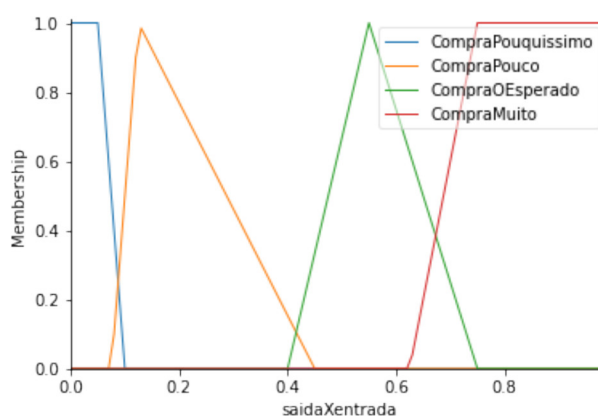


Figura 5: Funções de Pertinência de “Compras x Vendas”

A última variável de entrada utilizada, “Arrecadação x Vendas”, representa a arrecadação esperada em relação às notas fiscais de saída, ou seja, às vendas realizadas no mesmo período, e tem quatro conjuntos nebulosos, como mostra a Tabela 3. As funções de pertinência utilizadas para cada conjunto dessa variável foram, respectivamente, função decrescente, função triangular, função triangular novamente e a função crescente, como pode ser visto na Figura 6.

Tabela 3: Conjuntos Nebulosos de “Arrecadação x Vendas”

Arrecada Pouquíssimo	entre 0 e 1
Arrecada Pouco	entre 0,8 e 7
Arrecada o Esperado	entre 6 e 14
Arrecada Muito	a partir de 12

Um conjunto de 48 regras foram determinadas para o sistema de inferência e, na Tabela 4, são mostradas algumas regras para servir de exemplo. O antecedente é a condição formada pelas variáveis de entrada “Idade da Empresa”, “Compras x Vendas” e “Arrecadação x Vendas”, e o consequente é o resultado indicado por um dos valores da variável de saída “Grau Fraudadora”, que traduz o grau de possibilidade de uma empresa ser considerada como fraudadora, ou seja, “Baixo”, “Médio” e “Alto”, conforme Tabela 5. Três funções de pertinência foram utilizadas para definir as três faixas de saída da variável “Grau Fraudadora”. Cada faixa foi definida por uma função triangular, como podem ser vistas na Figura 7.

Definidas as variáveis de entrada e saída e seus universos discursivos, construiremos as regras de inferência. O sistema proposto possui 48 regras e, por brevidade e exemplificação, somente serão listadas nove delas, como mostra a Tabela 6.



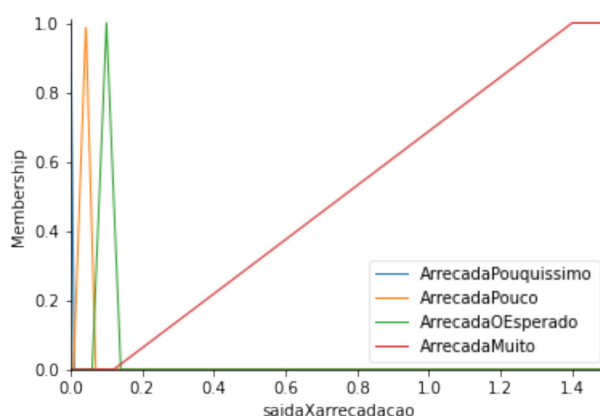


Figura 6: Funções de Pertinência de “Arrecadação x Vendas”

Tabela 4: Regras do sistema nebuloso de inferência

Se	Então
Idade é “Muito Jovem”, “Compra Pouquíssimo” e “Arrecada Pouquíssimo”	o grau é “Alto”
Idade é “Jovem”, “Compra Pouco” e “Arrecada Pouco”	o grau é “Médio”
Idade é “Não Jovem” e “Compra o Esperado” e “Arrecada o Esperado”	o grau é “Baixo”

Tabela 5: Conjuntos Nebulosos de “Grau Fraudadora”

Baixo	Entre 0 e 60
Médio	Entre 50 e 80
Alto	Entre 70 e 100

Tabela 6: Regras do Sistema Nebuloso de Inferência

Regras	Variáveis Linguísticas de Entrada			Variável Linguística de Saída
	Idade da Empresa	Compras x Vendas	Arrecadação x Vendas	Grau Fraudadora
R1	Muito Jovem	Compra Pouquíssimo	Arrecada Pouquíssimo	Alto
R2	Muito Jovem	Compra Pouquíssimo	Arrecada O Esperado	Médio
R3	Jovem	Compra Pouco	Arrecada Pouquíssimo	Alto
R4	Jovem	Compra Pouco	Arrecada Pouco	Médio
R5	Não Jovem	Compra O Esperado	Arrecada Pouquíssimo	Alto
R6	Não Jovem	Compra Muito	Arrecada Pouco	Médio
R7	Muito Jovem	Compra O Esperado	Arrecada O Esperado	Baixo
R8	Jovem	Compra Muito	Arrecada Pouco	Médio
R9	Não Jovem	Compra Pouquíssimo	Arrecada O Esperado	Baixo

Para cada entrada de dados, todas as regras são avaliadas pelo sistema de inferência que formará um conjunto nebuloso resultado para que, então, seja submetido ao processo de *defuzzificação* a fim de se obter a variável numérica de saída. A Tabela 6 mostra alguns exemplos de valores numéricos de entrada que produzirão suas respectivas saídas numéricas e seu valor linguístico final, resultado da *defuzzificação*.

Com a finalidade de saber se o sistema nebuloso de inferência seria assertivo na detecção de fraudadores, foram submetidos a ele dois conjuntos de empresas. O primeiro conjunto possui um grupo com cerca de mil e cem empresas que já foram fiscalizadas e sofreram cancelamento ou impedimento por conta da comprovação do ato de fraude. O segundo conjunto é formado por um grupo de, aproximadamente, cento e vinte mil empresas, das quais não se tem nenhuma certeza que praticaram fraude, pois não foram submetidas a nenhum processo de fiscalização



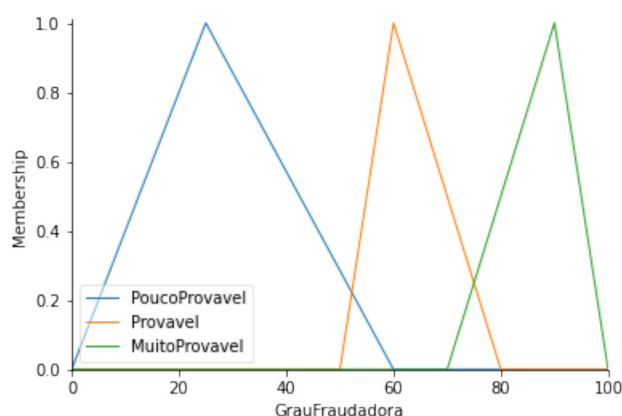


Figura 7: Funções de Pertinência de “Grau Fraudadora”

para comprovação da ilegalidade. Os resultados podem ser utilizados para a seleção do conjunto de empresas a serem fiscalizadas.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Empresas Fraudadoras

Após aplicação do modelo nebuloso ao primeiro conjunto de empresas, já fiscalizadas e comprovadas como fraudadoras, obtive-se o resultado cuja breve amostra é listada na Tabela 7. Em destaque, estão quatro resultados definidos com um grau baixo de possibilidade de serem fraudadoras. Assim, de um total de 1.176 empresas classificadas como fraudadoras, as quatro destacadas foram as únicas exceções que contrariaram o método atual, ou seja, tivemos apenas 0,34% de divergência entre o método manual e o método nebuloso.

Tabela 7: Aplicação do sistema nebuloso em empresas já conhecidas como fraudadoras

Entrada	Variáveis Numéricas de Entrada			Saída	
	Idade da Empresa	Compras x Vendas	Arrecadação x Vendas	Numérica	Linguística (Grau Fraudadora)
E1	26,97	0,007	0,000	0,833	Alto
E2	70,50	0,234	0,021	0,793	Alto
E3	215,53	0,000	0,000	0,833	Alto
E4	144,70	0,003	0,066	0,258	Médio
E5	41,67	0,000	0,086	0,818	Médio
E6	60,47	0,000	1322,42	0,904	Baixo
E7	79,43	0,0	27,18	0,904	Baixo
E8	66,90	0,000	3.457,00	0,904	Baixo
E9	50,43	0,0	1.900,83	0,904	Baixo

Após uma nova análise, mais minuciosa, feita pelos especialistas que as haviam selecionado manualmente essas empresas possivelmente fraudadoras e que tiveram sua fraude confirmada, a falha do método nebuloso ocorreu por razões que não estavam sendo consideradas pelas variáveis utilizadas. Assim, observa-se a necessidade de se aprimorar o sistema nebuloso adicionando-se mais variáveis importantes à tomada de decisão como informações sobre consumo de energia elétrica, confirmação da existência em endereço cadastrado etc.

### 4.2 Empresas Possivelmente Fraudadoras

Aplicou-se o modelo nebuloso ao segundo conjunto de empresas, das quais não se tem ainda nenhuma informação sobre a prática de fraude. De um total de, aproximadamente, 120.000

empresas, mais de 80.000 empresas foram classificadas com grau 'Alto', como possíveis fraudadoras, enquanto que, pouco mais de 7.000 foram classificadas com grau 'Médio', também com possibilidade de serem fraudadoras, e cerca de 33.000 com grau 'Baixo' podem ser descartadas de fiscalização. Este resultado pode ser utilizado para compor a lista de empresas a serem fiscalizadas presencialmente numa ação futura e também ser confrontado com o método manual.

## 5 CONCLUSÃO

Os resultados foram baseados na comparação entre o método atualmente utilizado pela SEFAZ-RJ e o método proposto que utiliza um modelo nebuloso para constatação de fraude a um imposto estadual brasileiro. O método atual analisa vários dados sobre uma empresa, obtidos utilizando-se a teoria dos conjuntos e resolve, por decisão humana e intuitiva, indicar uma empresa para uma ação de fiscalização a fim de verificar a ocorrência da fraude suspeita. O novo método proposto exige que um especialista no assunto elabore as regras de inferência, uma única vez, e passe a utilizá-lo de forma que a análise das informações sejam processadas computacionalmente, ou seja, agilizando o processo e evitando variações de interpretação na produção dos resultados.

A aplicação da teoria nebulosa na detecção de fraude de ICMS mostrou que é possível contar com uma ferramenta confiável para a tomada de decisão em situações onde o conhecimento e razão humanos são essenciais. Tornou-se possível tomar decisões de uma forma mais natural, impessoal, simplificada e rápida. O alto grau de acerto superou as expectativas e pode ser justificativa para a utilização da mesma técnica em outras situações de detecção de fraude envolvendo novas variáveis. A consistência nos resultados indica que o novo método pode melhorar a qualidade e confiança de todo o trabalho realizado na SEFAZ-RJ no combate à fraude.

Desse modo, é possível perceber que a teoria nebulosa pode ser utilizada com grande eficácia na detecção de fraudadores de ICMS e que esta técnica deve ser explorada em problemas similares, não somente no estado do Rio de Janeiro, mas em todos os estados da federação, a fim de contribuir para uma maior e mais rápida recuperação das suas receitas, melhorando suas situações financeiras e, conseqüentemente, viabilizando mais investimentos no bem-estar de suas populações.

## REFERÊNCIAS

- [1] G. Boole. *The mathematical analysis of logic*. Philosophical Library, 1847.
- [2] D. D. Chamberlin e R. F. Boyce. Sequel: A structured english query language. In *Proceedings of the 1974 ACM SIGFIDET (now SIGMOD) workshop on Data description, access and control*. 1974. <https://doi.org/10.1145/800296.811515>.
- [3] E. F. Codd. A relational model of data for large shared data banks. *Association for Computing Machinery*, 13:377–387, 1970. <https://dl.acm.org/doi/10.1145/362384.362685>.
- [4] F. N. de Oliveira. Estratégias para combater a sonegação fiscal: Um modelo para o icms baseado em redes neurais artificiais. *Revista de Gestão, Finanças e Contabilidade*, 10(1):42–64, 2020. <https://doi.org/10.18028/rgfc.v10i1.7474>.
- [5] M. Mizumoto. *Handbook of Fuzzy Computation*. IOP Publishing Ltd, 1998.
- [6] Secretaria de Estado de Fazenda do Rio de Janeiro. *Legislação Básica do ICMS*. SEFAZ-RJ, 2019. Disponível em <[https://www.fazenda.rj.gov.br/sefaz/faces/menu\\_structure/legislacao/legislacao-tributaria-basica-navigation/folder0/basicalCMS](https://www.fazenda.rj.gov.br/sefaz/faces/menu_structure/legislacao/legislacao-tributaria-basica-navigation/folder0/basicalCMS)>. Acessado em julho de 2022.

- [7] Secretaria de Estado de Fazenda do Rio de Janeiro. *Web site*. SEFAZ-RJ, Rio de Janeiro, Brasil, 2019. Disponível em <<https://www.fazenda.rj.gov.br/>>. Acessado em julho de 2022.
- [8] C. Thang, P. Q. Toan, E. W. Cooper e K. Kamei. Application of soft computing to tax fraud detection in small businesses. In *First International Conference on Communications and Electronics*. 2006. <https://doi.org/10.1109/CCE.2006.350887>.
- [9] C. Yamao. A história do imposto sobre circulação de mercadorias – do IVM ao ICMS. *Revista Jurídica UNICURITIBA*, 3:36, 2014. Disponível em: <<http://revista.unicuritiba.edu.br/index.php/RevJur/article/view/990/681>>. Acessado em julho de 2022.
- [10] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
- [11] L. A. Zadeh e R. A. Aliev. *Fuzzy Logic Theory and Applications: Part I and Part II*. World Scientific, 2018.