

Aplicação de rede neural artificial para auxiliar a fiscalização tributária na identificação de empresas fantasmas

Ricardo Costa Pinto^{1*}, Patrícia Belfiore Fávero²

¹ Bacharel em Ciência da Computação pela Universidade Federal da Bahia. Quadra 201, lote 03 - Águas Claras;
71937-540 Brasília, Distrito Federal, Brasil

² Pós-Doutorado em Pesquisa Operacional e Logística pela Universidade de Columbia, Nova Iorque. Rua
Alexandre Herculano, 120 - Vila Monteiro; 13418-445 Piracicaba, São Paulo, Brasil

*autor correspondente: ricardocpinto@hotmail.com

Aplicação de rede neural artificial para auxiliar a fiscalização tributária na identificação de empresas fantasmas

Resumo

A administração tributária, parte integrante da Administração Pública, é responsável por acompanhar, controlar e fiscalizar o cumprimento das normas que tratam das receitas tributárias. Assim sendo, o objetivo deste trabalho é a construção de uma rede neural artificial capaz de auxiliar essa entidade na identificação de empresas fantasmas. O modelo foi construído a partir de variáveis que demonstram características relevantes de cada contribuinte, tais como: localização, tipo de atividade econômica, porte, área do estabelecimento, notas fiscais de compras, fornecedores, sócios, entre outras. A partir da obtenção dos resultados, foi possível concluir que as variáveis preditoras foram capazes de explicar a variável dependente, com isso, restou demonstrado que é possível identificar empresas fantasmas utilizando redes neurais artificiais.

Palavras-chave: detecção de irregularidades, inteligência artificial, modelo preditivo, ICMS, mineração de dados

Abstract

The tax administration, an integral part of the Public Administration, is responsible for monitoring, controlling and inspecting compliance with the rules that deal with tax revenues. Therefore, the objective of this work is the construction of an artificial neural network capable of assisting this entity in the identification of the constitution of shell companies. The model was built from variables that demonstrate relevant characteristics of each taxpayer, such as: location, type of economic activity, size, area of the establishment, purchase invoices, suppliers, partners, among others. From the obtaining of the results, it was possible to conclude that the predictor variables were able to explain the dependent variable, with this, it remained demonstrated that it is possible to identify shell companies using artificial neural networks.

Keywords: irregularity detection, artificial intelligence, predictive model, ICMS, data mining

Introdução

O principal desafio da Administração Tributária diz respeito ao combate da sonegação fiscal. O problema é tão antigo quanto a própria existência do tributo, pois em geral os indivíduos não são adeptos ao pagamento de impostos. A evasão fiscal tem impacto direto nos investimentos realizados pelo Estado, haja visto que os recursos em infraestrutura, educação, saúde, entre outros, são fruto da arrecadação dos tributos, ou seja, é preciso entender que a sociedade é destinatária dos recursos arrecadados pelo governo, e ela é quem mais perde com a sonegação.

Este tipo de crime provoca prejuízos incalculáveis quando comparamos com outros previstos no Código Penal, isso ocorre em virtude que neste tipo de infração penal o prejuízo é em desfavor da Fazenda Pública, tendo como consequência a diminuição do Estado em prover saúde, educação e saneamento básico, e em virtude desse efeito perverso, ocasiona

doenças, analfabetismo, pobreza, desemprego, e oneração daqueles que contribuem de acordo com a lei (Lemos, 1999).

Ademais, com o objetivo de promover a integração dos Fiscos e tornar mais veloz a identificação de ilícitos tributários, foi instituído em 2007 o Sistema Público de Escrituração Digital – SPED, constituindo um grande avanço tecnológico na relação Fisco e Contribuinte. Nesse novo cenário tecnológico, as informações passaram a ser realizadas de forma eletrônica e on-line, de tal modo que atualmente tanto a Receita Federal, quanto as Receitas Estaduais, possuem uma grande quantidade de dados à disposição, que vem crescendo de forma exponencial ao passar do tempo (Brasil, 2007).

No mesmo sentido da evolução dos processos fiscalizatórios, rapidez no acesso às informações e ações mais efetivas, o modus operandi dos ilícitos tributários também passaram por transformações e inovações. A forma tradicional de sonegação, venda sem documento fiscal, foi dificultada a partir de 2007 em razão da obrigatoriedade de emissão de nota fiscal eletrônica em toda cadeia produtiva (indústria, atacado, varejo e serviço).

Nessa senda, a Administração Tributária tem observado nos últimos anos um aumento significativo de empresas que estão sendo constituídas somente no papel (empresas fantasmas), com o fim específico de sonegar impostos, tendo em vista que esse tipo de empresa não existe fisicamente (Mendroni, 2015).

Oportuno se tornar dizer que segundo relatório do Tribunal de Contas do Estado de Goiás, a recuperação de créditos provenientes de dívida ativa não chega a 1%, além disso, o estoque desse tipo de dívida só vem aumentando ao longo dos anos (TCE, 2020). Um fator que influencia diretamente o acréscimo do passivo tributário e a baixa recuperação de crédito, é a constituição de empresas fantasmas com interpostas pessoas no quadro societário.

Assim sendo, é necessário o aprimoramento das técnicas de fiscalização tributária, utilizando ferramentas e tecnologias capazes de identificar padrões de empresas que são constituídas para praticar o crime de sonegação. Dessa forma, a fiscalização estará trabalhando de forma preventiva ao invés de remediadora, haja vista que efetuar auditoria fiscal/contábil em empresas fantasmas não possui resultado satisfatório, é ineficiente e só aumenta o estoque da dívida ativa estadual.

Neste passo de inovação tecnológica, as redes neurais artificiais são sistemas flexíveis, de aprendizado adaptativo que assimilam os padrões observáveis nos dados e estabelecem relações não lineares que são capazes de prover previsões confiáveis (Samarasinghe, 2006). Elas possuem uma arquitetura inspirada nas operações do cérebro humano e são capazes de adquirir conhecimento baseado no processo de treinamento de tentativa e erro. Essa técnica de inteligência artificial possui vasta aplicação na vida moderna,

e ultimamente vem sendo explorada em pesquisas para auxiliar as Administrações Tributárias na detecção de fraudes e combate à sonegação fiscal.

O presente estudo, por sua vez, possui como objetivo aplicar predição baseada em redes neurais artificiais capaz de identificar padrões típicos de contribuintes desaparecidos. O modelo classifica as empresas, a partir do conjunto de 25 (vinte e cinco) variáveis explicativas, como suspensas (desaparecidas) ou ativas. Dessa forma é possível determinar com mais assertividade as pessoas jurídicas a serem fiscalizadas in loco, para averiguação/confirmação da fraude praticada.

Material e Métodos

A pesquisa em questão é de natureza aplicada, de cunho exploratório, haja vista o caráter tecnológico, que objetiva auxiliar a fiscalização tributária na detecção de fraudes no processo de constituição de empresas utilizando redes neurais artificiais.

Quanto à abordagem, o viés é predominantemente quantitativo, com objetivo explicativo, pois estaremos diante de um estudo que utiliza técnicas de mineração de dados, tendo como essência a interpretação numérica dos eventos apresentados e investigados.

Neste estudo foram selecionados dados de 27.709 empresas do cadastro de contribuintes do Estado de Goiás, pessoas jurídicas de diversos ramos de atividades econômicas localizadas na região do entorno sul do Distrito Federal, que congrega os municípios de Águas Lindas de Goiás, Cidade Ocidental, Cristalina, Luziânia, Novo Gama, Santo Antônio do Descoberto e Valparaíso de Goiás. Esse território vivencia nos últimos anos uma explosão demográfica, e com isso alguns setores da economia estão em constante crescimento, dentre eles o comércio varejista, indústria, serviços e demais atividades relacionadas ao agronegócio.

Ademais, em virtude dos contribuintes classificados como MEI – Microempreendedor Individual e Produtor Urbano/Rural possuem particularidades específicas em relação às questões cadastrais, os mesmos foram excluídos dessa amostra coletada, pois poderiam gerar distorções nos resultados esperados.

Além disso, em virtude do vultoso volume de dados existentes relacionados às movimentações econômicas/fiscais da amostra de empresas acima, optou-se por restringir o período a ser coletado referente a esse tipo informação.

Convém ressaltar que todo o trabalho foi realizado em consonância com a Lei Geral de Proteção de Dados Pessoais (LGPD) e à Instrução Normativa nº 1.455/2020-GSE da Secretaria de Economia do Estado de Goiás. Portanto os dados foram trabalhados de forma agregada e genérica, assegurando e preservando o sigilo fiscal das informações.

O procedimento metodológico deste trabalho está apresentado na figura 1 a seguir:



Figura 1. Etapas do processo metodológico

Fonte: Dados originais da pesquisa

Análise do Problema

É bem verdade que estamos diante de um problema bastante complexo, identificar se uma empresa de fato existe ou não, baseado nas informações cadastrais e econômicas relacionadas a ela.

Nessa esteira, em razão da inexistência de estudo/pesquisa/artigo específico acerca da identificação de empresas fantasmas utilizando redes neurais artificiais, foi necessário efetuar análises detalhadas dos casos reais já identificados, e somado à vasta experiência profissional do pesquisador na identificação deste tipo de ilícito, foi possível selecionar preliminarmente as variáveis mais aderentes ao objetivo do estudo.

Definição de variáveis explicativas

Com relação às variáveis explicativas, após uma análise pormenorizada do problema, foram selecionadas 25 (vinte e cinco) que demonstram características relevantes de cada contribuinte, conforme descritas a seguir:

- **PORTE**: variável qualitativa textual que classifica o contribuinte de acordo com a faixa de faturamento (Micro/Pequeno Porte, Médio Porte, Grande Porte ou não informado).
- **TIPO_ENQUADRAMENTO**: variável qualitativa textual que classifica o contribuinte quanto ao regime de tributação (Normal ou Simples Nacional).
- **FX_AREA_ESTAB**: variável qualitativa textual que define a faixa de área do estabelecimento em metros. Exemplo: até 20 m, de 21 a 50, de 51 a 100, etc.
- **GRUPO_ECONOMICO**: variável qualitativa que classifica o contribuinte de acordo com o grupo da atividade econômica desenvolvida. Exemplo: Industrial, Comerciante Atacadista, Comerciante Varejista, Transportador, entre outros.
- **COD_ATIV_ECONOMICA**: variável qualitativa numérica que identifica o subgrupo da atividade econômica desenvolvida pelo contribuinte, como exemplo, o código 57 refere-se ao varejo de alimentos. A relação completa dos códigos e descrição está apresentada no Anexo I.

- IND_BLOQUEIO_COMPRA: variável qualitativa que identifica se o contribuinte já teve bloqueio/desbloqueio para as notas fiscais eletrônicas de compra.
- IND_BLOQUEIO_VENDA: variável qualitativa que identifica se o contribuinte já teve bloqueio/desbloqueio para as notas fiscais eletrônicas de venda.
- CRED_DTE: variável qualitativa binária (S/N) que identifica se o contribuinte está credenciado no DTE – Domicilio Tributário Eletrônico.
- MUNICIPIO: variável qualitativa textual que identifica o município do contribuinte.
- QTDE_SOCIO_UF_GO: variável quantitativa que indica a quantidade de sócios do contribuinte que residem no estado de Goiás.
- QTDE_SOCIO_UF_OUTRO: variável quantitativa que indica a quantidade de sócios do contribuinte que residem fora do estado de Goiás.
- IND_LOCAL: variável qualitativa que identifica se a empresa é de Goiás ou de outro estado. Exemplo: 1 - Goiás, 2 – outro estado.
- IND_MATRIZ_FILIAL: variável qualitativa que identifica se a empresa é uma matriz ou filial. Exemplo: 1 - Matriz, 2 – Filial.
- VL_MED_ARRECADACAO: variável quantitativa que informa o valor médio recolhido pelo contribuinte – é soma dos valores recolhidos dividido pela quantidade de recolhimentos efetivados.
- VL_MED_CARTAO_DEBITO: variável quantitativa que informa o valor médio de vendas no cartão de débito efetuadas pelo contribuinte – é soma dos valores das vendas dividido pela quantidade de venda.
- VL_MED_CARTAO_CREDITO: variável quantitativa que informa o valor médio de vendas no cartão de crédito efetuadas pelo contribuinte – é soma dos valores das vendas dividido pela quantidade de venda.
- QTDE_UF_FORNECEDOR: variável quantitativa que determina a quantidade de Estados que estão localizados os fornecedores do contribuinte.
- QTDE_FORNECEDOR: variável quantitativa que determina a quantidade de fornecedores do contribuinte.
- VL_MIN_NFE_COMPRA: variável quantitativa que informa o menor valor da nota fiscal de compra, dentre todas realizadas.
- VL_MAX_NFE_COMPRA: variável quantitativa que informa o maior valor da nota fiscal de compra, dentre todas realizadas.
- VL_MED_NFE_COMPRA: variável quantitativa que informa o valor médio das notas fiscais de compra - é soma dos valores das notas fiscais de compra dividido pela quantidade notas de compras.

- VL_MED_ICMS_NFE_COMPRA: variável quantitativa que informa o valor médio do ICMS das notas fiscais de compra - é soma dos valores do ICMS das notas fiscais de compra dividido pela quantidade notas de compras.
- PERC_NFE_COMPRA_ENTRE_0_5000: variável quantitativa que informa o percentual de notas fiscais de compra com valores menores ou iguais à R\$ 5.000,00 - é soma da quantidade das notas fiscais de compra com valores ≤ 5.000 , dividido pela quantidade total de notas de compras.
- PERC_NFE_COMPRA_ENTRE_5000_10000: variável quantitativa que informa o percentual de notas fiscais de compra com valores maiores à R\$ 5.000,00 e menores ou iguais à R\$ 10.000,00 - é soma da quantidade das notas fiscais de compra com valores > 5000 e ≤ 10.000 , dividido pela quantidade total de notas de compras.
- PERC_NFE_COMPRA_MAIOR_10000: variável quantitativa que informa o percentual de notas fiscais de compra com valores maiores à R\$ 10.000,00 - é soma da quantidade das notas fiscais de compra com valores > 10.000 , dividido pela quantidade total de notas de compras.

Oportuno se torna a dizer que em relação às notas fiscais de compras, baseado em padrões já identificados nas fiscalizações tributárias realizadas pelo pesquisador, esse tipo de operação comercial efetivada por uma empresa fantasma quase sempre se distingue de um contribuinte regular.

Dessa forma, o modelo não ficou restrito aos valores nominais constantes nas notas fiscais de compras, foram criadas novas variáveis estatísticas identificando: o percentual da quantidade de notas fiscais por faixa de valores; o valor médio das notas fiscais; o valor médio do ICMS; e os valores mínimos e máximos praticados.

Outra consideração, diz respeito à variável que identifica a existência de bloqueio fiscal como medida preventiva, resta constatado nas fiscalizações realizadas que empresas desaparecidas (fantasmas), na maior parte dos casos, sofreu alguma intervenção desse gênero.

Nessa esteira, o intuito foi alimentar a rede neural artificial com informações diferenciadas a fim de prover mais inteligência, e conseqüentemente buscar um resultado com mais assertividade na predição.

Coleta de Dados

Inicialmente foram criadas 06 (seis) consultas específicas, vide figura 2, para extração dos dados do SAP BusinessObjects - software de Business Intelligence - ferramenta que

consolida os dados de diversos sistemas corporativos, da Secretaria de Estado da Economia de Goiás, através do conceito de universos de dados. Os universos selecionados para a criação dessas consultas foram: Arrecadação das Receitas, Contribuinte Cadastro, Contribuinte – Informações dos Sócios, ECF – Faturamento Cartão de Crédito-Débito e Nota Fiscal Eletrônica – Por Nota Fiscal, salientamos que fontes de dados de outros órgãos ou instituições não foram utilizados.

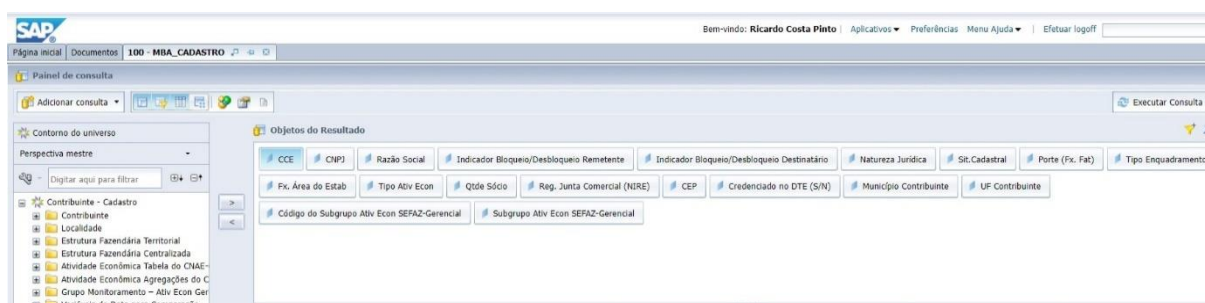


Título	Tipo
100 - MBA_ARRECADACAO	Web Intelligence
100 - MBA_CADASTRO	Web Intelligence
100 - MBA_CARTAO_CREDITO	Web Intelligence
100 - MBA_CARTAO_DEBITO	Web Intelligence
100 - MBA_COMPRAS	Web Intelligence
100 - MBA_SOCIO	Web Intelligence

Figura 2. Consultas criadas no SAP BusinessObjects

Fonte: Dados originais da pesquisa

A ferramenta de BI supracitada, possui interface visual para criação de consultas - documentos do tipo Web Intelligence. O processo de criação de tais documentos utiliza o conceito de Drag and Drop (arrastar e soltar), desse modo, foram selecionados os campos desejados dos universos de dados específicos para cada consulta gerada, conforme podemos observar na figura 3.



Objetos do Resultado
CCE
CHPJ
Razão Social
Indicador Bloqueio/Desbloqueio Remetente
Indicador Bloqueio/Desbloqueio Destinatário
Natureza Jurídica
Sit.Cadastral
Porte (Fv. Fat)
Tipo Enquadramento
Fv. Área do Estab
Tipo Ativ Econ
Qtde Sócio
Reg. Junta Comercial (NIRE)
CEP
Credenciado no DTE (S/N)
Município Contribuinte
UF Contribuinte
Código do Subgrupo Ativ Econ SEFAZ-Gerencial
Subgrupo Ativ Econ SEFAZ-Gerencial

Figura 3. Campos selecionados do universo Contribuinte – Cadastro

Fonte: Dados originais da pesquisa

Nesta etapa os dados foram previamente filtrados com o objetivo se adequar ao escopo da pesquisa, vide exemplo da figura 4. A seguir descrevemos os filtros que foram aplicados nos campos das consultas dessa fase inicial de coleta de dados:

- ✓ Município do Contribuinte igual à: Águas Lindas de Goiás, Cidade Ocidental, Cristalina, Luziânia, Novo Gama, Santo Antônio do Descoberto e Valparaíso de Goiás;

- ✓ Situação Cadastral do contribuinte igual a Ativo ou Suspenso – o objetivo é extrair a amostra de contribuintes que estejam ativos ou suspensos por desaparecimento (não existe de fato);
- ✓ Ano igual a 2021 e 2022 – filtrar o período dos dados referentes às notas fiscais de compra, faturamento de cartão de crédito-débito e recolhimentos de tributos;
- ✓ Tipo Enquadramento diferente de Simples Nacional/SIMEI - eliminar os contribuintes classificados com MEI - Micro Empreendedor Individual;
- ✓ Tipo Ativ Econ diferente de Produtor Rural e Produtor Urbano – eliminar os contribuintes classificados como produtores rurais e urbanos.
- ✓ CCE compreendido entre 100000000 e 109999999 – esse filtro foi utilizado para garantir a eliminação dos contribuintes fora do escopo da pesquisa, tendo em vista que as inscrições estaduais iniciadas por 11 são associadas à produtores rurais pessoa física, e iniciadas por 15 eram utilizadas para vendedores ambulantes.

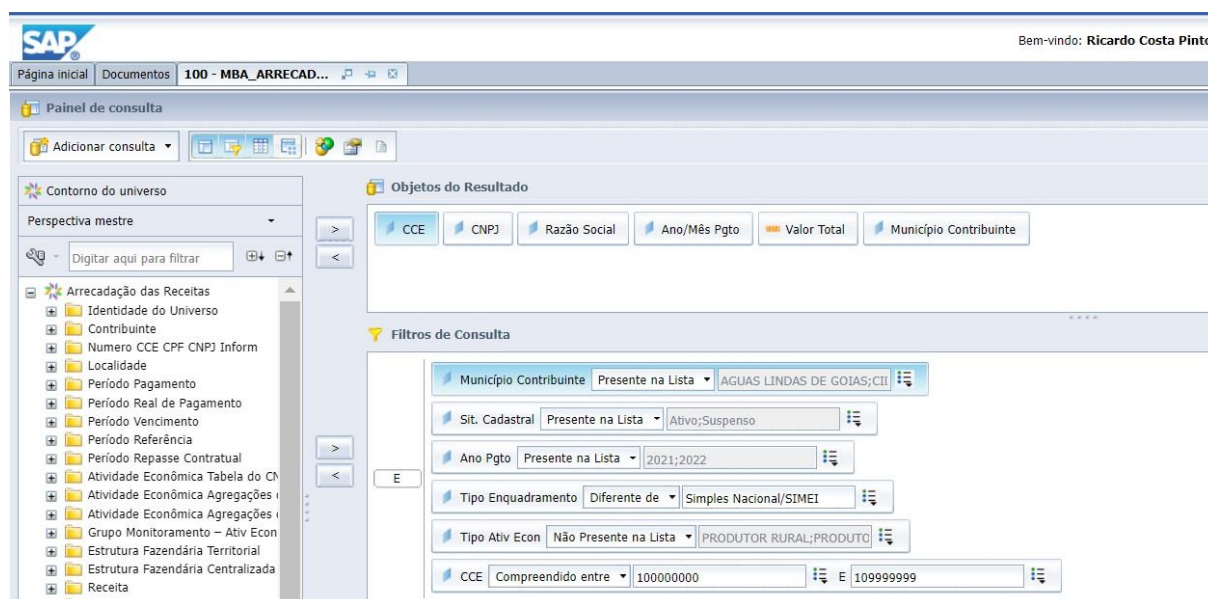


Figura 4. Exemplo de filtro aplicado no universo *Arrecadação das Receitas*
Fonte: Dados originais da pesquisa

Após a criação e execução das consultas no SAP BusinessObjects, os resultados dos processamentos foram exportados para arquivos em formato .csv, conforme listado a seguir:

- ARRECADACAO.csv, arquivo contendo os recolhimentos dos tributos efetuados pelos contribuintes de forma sintetizada.
- CONTRIBUINTE.csv, arquivo com os dados cadastrais dos contribuintes.
- INFO_SOCIO.csv, arquivo contendo as informações dos sócios dos contribuintes.

- FAT_CARTAO_CREDITO.csv, arquivo com vendas sintéticas dos contribuintes efetuadas com cartão de crédito.
- FAT_CARTAO_DEBITO.csv, arquivo com vendas sintéticas dos contribuintes efetuadas com cartão de débito.
- NFE_COMPRA.csv, arquivo com os dados sintéticos referentes às compras dos contribuintes.

Tratamento e Pré-processamento dos Dados

Esta etapa de preparação dos dados é de suma importância para buscar a máxima eficiência na execução do algoritmo de aprendizado de máquina, no caso em estudo, uma rede neural artificial.

A título exemplificativo, a presença de variáveis discriminantes categóricas, não métricas, exige transformação de forma que os dados simbólicos sejam transformados em classes numéricas, mais adequadas para utilização neste tipo de rede (Guimarães, 2006).

Desse modo foram utilizadas as técnicas de Data Wrangling para estruturação, padronização, organização, limpeza, normalização e validação dos dados. Uma vez os dados prontos para serem utilizados na rede neural artificial, procedeu-se a divisão de forma aleatória em dois subconjuntos, o primeiro para treinamento, e o segundo para teste

Em relação à plataforma tecnológica, foi utilizada a *Linguagem R* tanto nesse momento, quanto na próxima etapa de construção do modelo preditivo. Essa linguagem vem sendo amplamente utilizada por pesquisadores, engenheiros e estatísticos, por possuir características próprias, destacamos as seguintes: dinâmica; gratuita; independente; multi-paradigma orientada a objetos; dispõe de funções prontas para data wrangling e machine learning.

A linguagem R possui diversos pacotes para o processo de limpeza, transformação e organização dos dados, através de funções que foram desenvolvidas especificamente para este fim. Dessa forma, com poucas instruções de código é possível agrupar, modificar, inverter e transformar os dados ou ainda criar novos subconjuntos de dados. Nesse sentido, como exemplo temos os pacotes rpart, dplyr, tidyr, tidyverse, sendo que alguns deles foram utilizados nesse trabalho.

Ademais, foram criados scripts na linguagem R para a realização dos processos de preparação de dados (data wrangling). A seguir discriminamos textualmente e de forma sintetizada a execução dos algoritmos em questão, os detalhes das linhas de códigos poderão ser consultados no Apêndice I.

1. Carregamento individualizado por dataset dos arquivos .csv mencionados no tópico anterior, exportados do SAP BusinessObjects, contendo os dados de contribuinte, sócio, arrecadação, cartão de crédito, cartão de débito e compra.
2. Tratamento e transformação dos dados de cada dataset individualizado:
 - 2.1 Conversão das variáveis categóricas textuais para numéricas;
 - 2.2 Conversão das variáveis categóricas em fator;
 - 2.3 Criação de novas variáveis;
 - 2.4 Conversão de variáveis para numérico.
3. Junção das bases de dados individualizadas após tratamento.
4. Seleção das variáveis significativas para o modelo.
5. Eliminação das observações que contenham variáveis com valores NA (“Not Available”).
6. Normalização dos dados.
7. Divisão do conjunto de dados final pré-processado em dois dataset’s, treinamento e teste, nas proporções de 70% e 30%, respectivamente.

Construção do Modelo de Rede Neural

A fase de construção do modelo de rede neural converge para a concepção de extrair o conhecimento por meio dos dados do conjunto preparado na etapa anterior, focando na escolha do tipo, configuração e calibração da rede a ser utilizada, buscando otimizar os resultados esperados.

Neste estágio, a rede neural foi submetida ao processo de aprendizagem, ou seja, assim como o cérebro humano, a rede teve que “aprender a pensar” para que suas decisões fossem tomadas de forma coerente e precisa; nesse caso, as decisões foram os resultados das saídas com base nos valores dos pesos. Após esse processo, foi possível mensurar o desempenho da RNA em resposta às expectativas do modelo para que se possa realizar sua aplicação (Henriques, 2019).

Vale ressaltar que a seleção dos parâmetros de uma rede neural é baseada em heurísticas ou em simulações derivadas de experimentos limitados, haja vista que a escolha destes parâmetros não possui metodologia consolidada e varia de acordo com o problema a ser resolvido. (Cottrell, 1995).

A arquitetura ideal da rede neural deve ser tal que ela seja tão grande quanto necessária para conseguir obter as representações necessárias, mas ao mesmo tempo pequena o suficiente para se obter um treinamento mais rápido. Não existem regras claras

para se definir quantas unidades devem existir nas camadas escondidas, quantas camadas, ou como devem ser as conexões entre essas unidades (Siu, 1995).

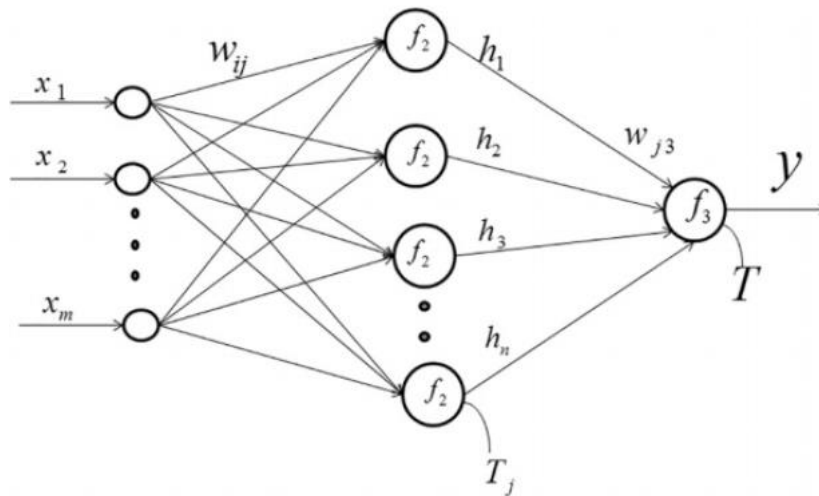


Figura 5. Arquitetura genérica de uma rede neural artificial.
Fonte: Dados originais da pesquisa

Hipoteticamente, uma rede neural artificial com uma camada de neurônios escondida e um número suficiente de neurônios nesta camada, seria capaz de aproximar qualquer tipo de função contínua. Na prática, redes neurais com uma ou duas camadas de neurônios escondidas são amplamente utilizadas e possuem performances bastante satisfatórias. Vale ressaltar que aumentar de forma crescente o número de camadas e o número de neurônios da rede neural, além de aumentar o custo computacional, também expõe a arquitetura da rede neural a problemas de *overfitting* (Frascaroli, 2006).

Esse problema ocorre quando um modelo de redes neurais possui muito poucos graus de liberdade, causando perda de poder de generalização fazendo com que a rede ganhe em memorização de características de observações individuais e perda do poder de predição. Logo, um modelo de redes neurais robusto deve seguir o critério da parcimônia, o que, além de diminuir o custo computacional, aumenta sua capacidade de generalização (Kaastra, 1996).

Nesse sentido, inicialmente foi definida uma rede neural com uma única camada oculta e um único neurônio, posteriormente foi aumentando o número de neurônios e/ou camadas, a depender dos resultados obtidos. Dessa forma, o objetivo foi avaliar as respostas em busca de uma calibração mais eficaz para o problema proposto, que é a identificação de empresas que não existem de fato.

De modo a obter um conhecimento dos resultados provenientes das diversas variações de topologias testadas (camadas x neurônios escondidos), os indicadores: acurácia (“accuracy”), sensibilidade (“sensitivity”), especificidade (“specificity”), kappa, MSE e RMSE,

foram tabulados e apresentados na próxima seção com o objetivo de embasar a seleção do melhor modelo de rede para resolução do problema proposto.

- Acurácia (“Accuracy”) – a proporção de observações que foram corretamente previstos, sejam eles verdadeiro positivo ou verdadeiro negativo.
- Sensibilidade (“Sensibility”) – a proporção de observações positivas que foram identificados corretamente.
- Especificidade (“Specificity”) - a proporção de observações negativas que foram identificados corretamente.
- Kappa - índice que retrata o grau de concordância dos dados, concebendo um aspecto de confiabilidade e precisão dos dados classificados.
- MSE – Erro médio quadrático, apresenta o desvio padrão do erro de predição do modelo, quando menor o erro melhor será o ajuste do modelo.
- RMSE – Raiz quadrada do erro médio.

Iniciando a etapa de criação do modelo, foram utilizados parâmetros fixos habitualmente mencionados na literatura técnica sobre o tema, conforme discriminados abaixo. A implementação do modelo de redes neurais foi realizada através da linguagem R, um dos estágios importantes é adoção das bibliotecas, nas quais estão contidos os pacotes necessários para a aplicação dos vários métodos de modelagem, nessa esteira foi utilizado o pacote neuralnet, necessário para todas as etapas de implementação e funcionamento da rede neural (Fritsch, 2019).

- Threshold = 0.1 - valor numérico especificando o limiar para os derivativos parciais da função de erro como critério de parada.
- Learningrate.factor = list(minus = 0.5, plus = 1.2) - lista contendo os fatores de multiplicação para a taxa de aprendizagem superior e mais baixa.
- algorithm = "rprop+" – selecionado algoritmo para calcular a rede neural com retro propagação resiliente com e sem retrocesso de peso.
- stepmax=1e7 - os passos máximos para o treinamento da rede neural, ao atingir esse máximo leva a uma parada do processo de treinamento da rede neural.
- act.fct = "logistic" – selecionada a função diferencial logística.

Além disso, como forma de otimização, evitando que a rede neural desperdice iterações adicionais de treinamento, as variáveis de fatores foram convertidas em variáveis dummy. Uma variável dummy transforma os N valores distintos e os converte em N-1 variáveis com resultados “zero” ou “um”, indicando a ausência ou presença de qualidades ou atributos. Para efetuar essa transformação foi utilizada a função *model.matrix* nos dataset's de treinamento e teste.

O código fonte em R referente ao processo de dummização pode ser visualizado na figura 6, enquanto que a figura 7, a título ilustrativo, retorna parcialmente a matriz de dados de treinamento com as variáveis dummizadas. Com a aplicação do processo de transformação dos fatores nos dataset's treinamento e teste, que possuem 26 variáveis, as novas matrizes de dados geradas passaram a apresentar 108 variáveis cada (ds_matrix_treinamento e ds_matrix_teste).

```

208 - #####
209 - ### Rede Neural pacote neuralnet
210 - #####
211 - # Elimina os NA
212 - ds_treinamento <- na.omit(ds_treinamento)
213 - ds_teste <- na.omit(ds_teste)
214 -
215 - formula_modelo <- - SIT_CADASTRAL + PORTE + TIPO_ENQUADRAMENTO + FX_AREA_ESTAB + GRUPO_ECONOMICO +
216 - COD_ATIV_ECONOMICA + IND_BLOQUEIO_COMPRA + IND_BLOQUEIO_VENDA + CRED_DTE +
217 - MUNICIPIO + QTDE_SOCIO_UF_GO + QTDE_SOCIO_UF_OUTRO + IND_LOCAL + IND_MATRIZ_FILIAL +
218 - VL_MED_ARRECADACAO +
219 - VL_MED_CARTAO_DEBITO + VL_MED_CARTAO_CREDITO +
220 - QTDE_UF_FORNECEDOR + QTDE_FORNECEDOR + VL_MIN_NFE_COMPRA + VL_MAX_NFE_COMPRA + VL_MED_NFE_COMPRA +
221 - VL_MED_ICMS_NFE_COMPRA + PERC_NFE_COMPRA_ENTRE_0_5000 + PERC_NFE_COMPRA_ENTRE_5001_10000 + PERC_NFE_COMPRA_MAIOR_10000
222 -
223 - # transformando o dataset em matrix dumizando variáveis de fatores
224 - ds_matrix_treinamento <- model.matrix(formula_modelo, data = ds_treinamento)
225 - ds_matrix_teste <- model.matrix(formula_modelo, data = ds_teste)
226 -
227 - # Cria a formula com o nome das variáveis para o modelo
228 - col_list <- paste(c(colnames(ds_matrix_treinamento)[-c(1,2)]), collapse="+")
229 - col_list <- paste(c("SIT_CADASTRAL1~", col_list), collapse="")
230 - f <- formula(col_list)
231 -

```

Figura 6. Código fonte do processo de transformação utilizando a função model.matrix
Fonte: Dados originais da pesquisa

	(Intercept)	SIT_CADASTRAL1	PORTE2	PORTE3	PORTE4	TIPO_ENQUADRAMENTO2	FX_AREA_ESTAB10	FX_AREA_ESTAB2	FX_AREA_ESTAB3	FX_AREA_ESTAB4	F
6316	1	1	0	0	0		1	0	0	1	0
3695	1	0	0	0	0		1	0	0	1	0
13385	1	0	0	0	1		1	0	0	0	0
6695	1	0	0	0	0		1	0	1	0	0
19521	1	1	0	0	0		0	0	0	1	0
8205	1	0	0	0	0		0	0	0	0	0
6242	1	1	0	0	0		1	0	0	1	0
625	1	0	0	0	0		0	0	0	0	1
10906	1	0	0	0	0		1	0	0	1	0
19673	1	1	0	0	1		1	0	0	0	0
9928	1	1	0	0	0		0	0	1	0	0
25363	1	1	0	0	1		0	0	0	0	0
7684	1	0	0	0	0		0	0	0	0	0
13854	1	1	0	0	1		0	0	0	0	0
6725	1	1	0	0	0		1	0	0	0	1

Figura 7. Exemplo de variáveis dummizadas
Fonte: Dados originais da pesquisa

Ao ensejo da conclusão desse tópico, a figura 8 ilustra o código fonte produzido em Linguagem R para implementação do modelo de rede neural artificial com duas camadas escondidas contendo três neurônios - c(2,1), enquanto que a figura 9 retorna à representação gráfica de uma rede com duas camadas escondidas com cinco neurônios - c(3,2).

```
231
232 # Rede Neural pacote neuralnet
233 set.seed(0)
234 print(paste('Inicio: ', Sys.time()))
235 mdl_redeneural <- neuralnet(f, data=ds_matrix_treinamento, hidden=c(2,1),
236                             threshold = 0.1,
237                             learningrate.factor = list(minus = 0.5, plus = 1.2),
238                             algorithm = "rprop+", stepmax=1e7, act.fct = "logistic" )
239 print(paste('Fim: ', Sys.time()))
240 plot(mdl_redeneural)
```

Figura 8. Código fonte de uma rede neural c(2,1)
Fonte: Dados originais da pesquisa

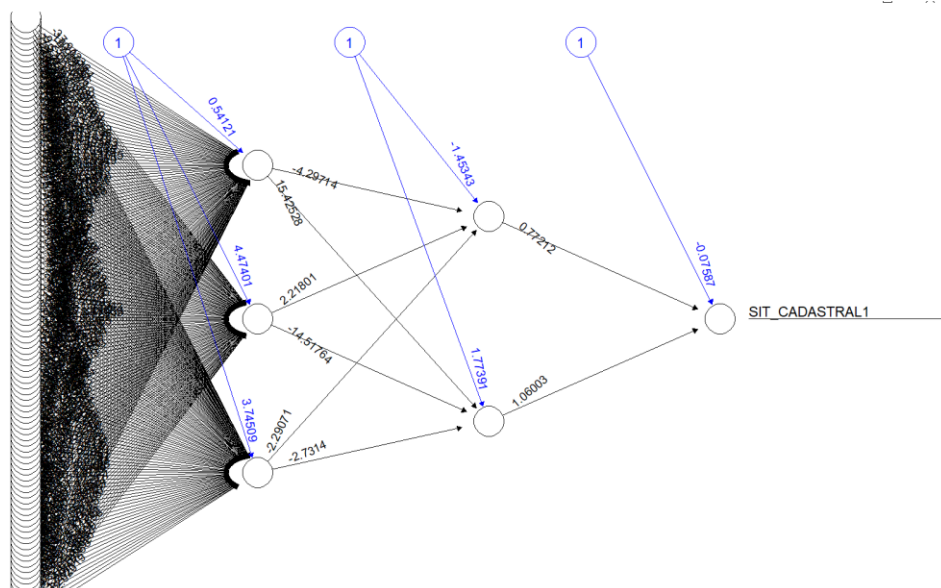


Figura 9. Representação gráfica de uma rede neural c(3,2)
Fonte: Dados originais da pesquisa

Resultados e Discussão

Os resultados atingidos pela aplicação do método discorrido anteriormente são apresentados nesse tópico. As análises foram realizadas a partir de métricas estatísticas para adoção do modelo que mais se adequa à capacidade de identificar padrões de empresas fantasmas, que são constituídas para praticar o crime de sonegação.

Para realizar a análise dos modelos criados quanto à qualidade e eficácia é essencial a utilização de métodos estatísticos. No que diz respeito à técnica de classificação, a avaliação pode ser considerada efetivada pelos conceitos taxa de erros, falso positivo, falso negativo e custo de erros de ajustes (Quinlan, 1993).

Neste sentido foram avaliadas 20 (vinte) topologias de redes neurais artificiais, que sofreram variações somente em relação à quantidade de: neurônios e camadas escondidas. Os indicadores estatísticos de cada topologia foram registrados ao término da fase de teste das redes neurais, e as métricas foram tabuladas conforme tabela 1.

Tabela 1. Resultado das métricas para as configurações avaliadas

Id	Configuração	Acurácia	Sensibilidade	Especificidade	Kappa	MSE	RMSE
1	hidden=1	0,838	0,927	0,773	0,679	0,118	0,343
2	hidden=2	0,835	0,910	0,772	0,778	0,113	0,336
3	hidden=3	0,850	0,853	0,848	0,695	0,108	0,329
4	hidden=5	0,857	0,862	0,854	0,710	0,104	0,322
5	hidden=c(1,1)	0,840	0,920	0,781	0,681	0,117	0,342
6	hidden=c(2,1)	0,844	0,923	0,786	0,689	0,115	0,339
7	hidden=c(2,2,1)	0,841	0,902	0,682	0,678	0,119	0,345
8	hidden=c(3,1)	0,859	0,907	0,824	0,717	0,103	0,322
9	hidden=c(4,1)	0,844	0,888	0,811	0,686	0,108	0,328
10	hidden=c(4,1,1)	0,841	0,875	0,815	0,680	0,120	0,346
11	hidden=c(4,1,1,1)	0,841	0,905	0,793	0,683	0,125	0,353
12	hidden=c(5,1)	0,862	0,885	0,846	0,721	0,101	0,318
13	hidden=c(5,2)	0,863	0,901	0,835	0,725	0,103	0,321
14	hidden=c(6,1)	0,857	0,864	0,853	0,710	0,105	0,325
15	hidden=c(6,2)	0,854	0,865	0,846	0,704	0,110	0,331
16	hidden=c(7,1)	0,859	0,877	0,845	0,714	0,107	0,327
17	hidden=c(8,1)	0,857	0,867	0,849	0,709	0,110	0,331
18	hidden=c(9,1)	0,861	0,871	0,853	0,717	0,104	0,322
19	hidden=c(10,1)	0,851	0,860	0,845	0,699	0,111	0,333
20	hidden=c(15,1)	0,849	0,857	0,843	0,694	0,121	0,347
	Mínimo	0,835	0,853	0,682	0,678	0,101	0,318
	Máximo	0,863	0,923	0,854	0,778	0,125	0,353
	Mediana	0,851	0,881	0,839	0,701	0,110	0,331
	Variância	6,4E-05	4,7E-04	1,7E-03	2,3E-04	4,9E-05	1,1E-04
	Desvio Padrão	8,9E-03	2,3E-02	4,2E-02	2,3E-02	6,8E-03	1,0E-02

Fonte: Resultados originais da pesquisa

Iniciando a análise pela estatística Kappa podemos classificar o desempenho do modelo como muito bom, desta forma, as amostras coletadas foram concordantes com as informações classificadas. Convém ressaltar que esse índice é utilizado para realizar a medida de concordância em escalas nominais fornecendo um cenário que apresenta o quanto as observações se afastam das observações esperadas, indicando assim o grau de genuinidade das interpretações feitas pelo modelo criado (Thompson, 2001). O resultado obtido pelo coeficiente Kappa, varia no intervalo de 0 a 1, sendo que quanto mais próximo a 1, melhor a qualidade dos dados classificados. Vários são os índices para agrupar esses dados quantitativos para qualitativos, entre eles, pode ser destacado o de Fonseca (2000), conforme tabela 2.

Tabela 2. Agrupamento qualitativo do coeficiente kappa

Índice Kappa	Desempenho
<0	Péssimo
$0 < k \leq 0,2$	Ruim
$0,2 < k \leq 0,4$	Razoável
$0,4 < k \leq 0,6$	Bom
$0,6 < k \leq 0,8$	Muito Bom
$0,8 < k \leq 1,0$	Excelente

Fonte: Fonseca (2000)

Em relação à acurácia, que é a porcentagem de observações classificadas corretamente (positivos ou negativos), o modelo se mostrou bastante eficiente em virtude da complexidade do problema, variando de 83,5% a 86,3%.

No que diz respeito ao erro a rede número 12 foi o que apresentou menor valor de MSE e RMSE, 0.101 e 0.298, respectivamente. Esse resultado, em conjunto com a acurácia de 0.862, poderia corroborar para o melhor modelo proposto. No entanto, a **sensibilidade** de 0.885 está relativamente próximo à mediana, e um pouco “distante” dos melhores valores registrados, métrica essa que é de suma importância para o problema apresentando.

Convém notar, que o aumento de neurônios e camadas não geraram ganhos de predição, tendo como consequência direta o aumento do custo computacional de processamento da rede neural na fase de aprendizado.

Ademais, o modelo de rede neural proposta tem como objetivo identificar empresas que não existem de fato, com isso explicaremos a seguir, baseado na matriz de confusão da tabela 3, por que a sensibilidade possui maior relevância do que a especificidade, levando em consideração que os demais indicadores não variaram substancialmente.

Tabela 3. Matriz de Confusão da rede 2

		Valores Reais	
		Positivo	Negativo
Valores Preditos	Positivo	3137	1020
	Negativo	310	3567

Fonte: Resultados originais da pesquisa

Vejamos a seguir as classificações dos dados realizada pela rede neural número 2, que possui duas camadas escondidas, dois neurônios na primeira camada e um neurônio na segunda camada:

- 3137 empresas foram classificadas como suspensas (desaparecidas), e estavam de fato suspensas;

- 310 empresas foram classificadas como ativas, porém estavam suspensas (falso negativo);
- 3567 empresas foram classificadas como ativas (não suspensas), e a princípio estão ativas.
- 1020 empresas foram classificadas como suspensas (desaparecidas), e a princípio estão ativas.

Pois bem, utilizamos o termo a *princípio* pois apesar dessas empresas estarem ativas no cadastro estadual, existe a possibilidade de não existirem de fato. Ou seja, nesse universo de 4587 contribuintes, 3567 verdadeiros negativos e 1020 falsos positivos, podem existir ocorrências de empresas fantasmas. Para atestarmos que realmente existem de fato, seria necessária uma vistoria *in loco* em cada uma delas, e nesse sentido deve-se priorizar as empresas a serem fiscalizadas.

Neste passo, um modelo com *maior sensibilidade*, proporção de casos positivos identificados corretamente, é o mais recomendado para identificar empresas fantasmas, levando em consideração que as demais métricas não devem variar substancialmente.

Logo, analisando os dados da tabela 1, concluímos que o melhor modelo de rede neural é o número 06, haja vista a sensibilidade de 0.923 e acurácia de 0.844. Salientamos que os desvios padrões e as variâncias da: acurácia, sensibilidade, especificidade, kappa, mse e rmse demonstraram que esses valores estão condensados próximos da média, com pouquíssima variação.

Portanto, a Administração Tributária deverá priorizar as fiscalizações *in loco* nos contribuintes ativos no cadastro estadual e classificados como desaparecidos pela rede neural artificial, dessa forma a probabilidade de assertividade para identificar as empresas fantasmas será muito maior do que pelos métodos tradicionais de seleção, que não utilizam aprendizado de máquina.

Apesar da literatura apresentar alguns trabalhos sobre a utilização de redes neurais artificiais aplicadas às fiscalizações tributárias, não localizamos uma pesquisa específica acerca de identificação de empresas fantasmas utilizando esse tipo de inteligência artificial.

Deste modo é perceptível a contribuição para o desenvolvimento de trabalhos futuros com a aplicação do mesmo método, bem como a necessidade de se trabalhar este tipo de modelo com um número maior de variáveis explicativas relevantes, tais como:

- dados pessoais dos sócios exclusivo da Receita Federal do Brasil (rendimentos, bens móveis e imóveis, entre outros).
- dados das notas fiscais de venda – não foram utilizados em virtude do custo computacional e restrições tecnológicas de processamento.

Convém ressaltar que as características dos contribuintes localizados nos municípios selecionados do entorno sul do Distrito Federal, não é o mesmo dos demais Estados brasileiros, caso o modelo seja aplicado em outras regiões ou entes federativos ou a tributos com particularidades diferentes do ICMS, este deverá ser adaptado levando em consideração o perfil econômico e diversidade regional.

Considerações Finais

A rede neural artificial implementada neste trabalho demonstrou que essa tecnologia auxilia a fiscalização tributária na identificação de empresas que não existem de fato, dessa forma coibindo a sonegação fiscal. A pesquisa procurou construir, treinar e testar a rede, buscando uma sistemática para determinação do provável crime tributário. Os resultados encontrados são de fundamental importância para a Administração Tributária, ao aprontar indicação de empresas com possíveis irregularidades, outrossim, a chance de planejamento de um trabalho de fiscalização mais assertiva e direcionada, demonstrando que é possível ampliar o que se faz para se chegar a uma sofisticação muito maior na rotina das autoridades fiscais. Ressaltamos que é notória a carência da aplicação de ferramentas desse tipo nas administrações tributárias, onde muitas vezes faltam investimentos na elaboração de tecnologias que garantam processos de automação e sofisticação dos procedimentos de fiscalização, otimizando custos e garantindo maior precisão nas ações realizadas. Este esforço permite uma análise que leva à constatação de que não basta elenco de leis e normas para definir comportamento por parte dos contribuintes. Faz-se necessário aprimorar a eficiência da ação fiscal através de conhecimentos e instrumentos modernos que já estão disponíveis e que possuem um grau destacado de maturidade. Reunindo técnicas de modelagem baseadas em dados, redes neurais artificiais, com outras baseadas em regras de conhecimento (Towell, 1993), e outras técnicas de inteligência artificial e computacional, é possível avançar sem limites rumo a ferramentas cada vez mais eficientes e poderosas para uso na fiscalização tributária. A não utilização dessas ferramentas modernas de abordagem provoca a perda de arrecadação de tributos que desaparecem em virtude da baixa eficiência da Administração Tributária no que concerne à estrutura cara e burocrática nos procedimentos e ações fiscais. Como se pode notar o ganho deste trabalho é mais vasto que demonstrado na narrativa aqui elaborada, desse modo a conclusão de melhor ênfase é a de que é factível ampliar o estudo científico, baseado em técnicas de Inteligência Artificial, voltado para a fiscalização tributária.

Agradecimento

Aos meus familiares e amigos que me apoiaram para a construção deste trabalho.

A minha orientadora professora Patrícia Belfiore Fávero, pela presteza e disposição apresentada.

À Secretaria de Economia do Estado de Goiás, em particular a Superintendência de Informações Fiscais e a Gerência de Inovação em Auditoria, por terem autorizados a coleta dos dados necessários para a realização desta pesquisa.

Referências

Amaral, G.L.; Amaral, L.M.F.; Olenike, J.E.; Yazbek, C.L.; Steinbruch, F. 2009. Estudo sobre sonegação fiscal das empresas brasileiras. IBPT – Instituto Brasileiro de Planejamento Tributário.

Brasil. 2007. Decreto Nº 6.022, de 22 de Janeiro de 2007. Institui o Sistema Público de Escrituração Digital – Sped. Disponível em: <<http://sped.rfb.gov.br>>. Acesso em: 17 mar. 2022.

Cottrell, M. 1995. Neural modeling for time series: A statistical stepwise method for weight elimination. IEEE Transactions on Neural Networks. Institute of Electrical and Electronics Engineers (IEEE): 1355–1364. Disponível em: <<https://doi.org/10.1109/72.471372>>. Acesso em: 02 jul. 2022.

Ferracini, L.A. 1996. Do crime de sonegação fiscal. 1ed. Editora de Direito, São Paulo, SP, Brasil.

Fonseca, L.M.G. 2000. Processamento digital de imagens. Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP, Brasil.

Frascaroli, B.F. 2006. Utilização de Redes Neurais Artificiais para Classificação de Ratings de Risco Soberano. Dissertação de Mestrado em Economia. Universidade Federal da Paraíba, João Pessoa, PB, Brasil.

Freitas, S.O. 2001. Utilização de um Modelo Baseado em Redes Neurais para Precificação de Opções. Dissertação de Mestrado em Administração. Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

Fritsch, S.; Guenther, F.; Wright, M.; Suling, M.; Mueller, S. 2019. *Package 'neuralnet'*. Disponível em: <<https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf>>. Acesso em: 15 abr 2022.

Gaspareto, O.B. 2008. Rede Neurais Artificiais Aplicadas ao Reconhecimento de Speed Cheating em Jogos Online de Computador. Dissertação de Mestrado em Ciência da Computação. Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil.

Guimarães L.R. 2006. Previsão de inadimplência e redes neurais artificiais. Dissertação de Pós-graduação em Modelagem Computacional. Universidade do Estado do Rio de Janeiro. Rio de Janeiro, RJ, Brasil.

- Gomes, A.H.T. 2006. Tributação e Sonegação Fiscal Um Estudo do Comportamento do Estado ante a Sonegação Fiscal. Dissertação de Mestrado em Direito Constitucional. Universidade de Fortaleza. Fortaleza, CE, Brasil.
- Junior, G.G.P; Silveira, M.R.M; Neto, R.F. 2004. Uma rede neural artificial de múltiplas camadas aplicada ao combate à sonegação fiscal de ICMS. In: XXXVI - Simpósio Brasileiro de Pesquisa Operacional, São João Del-Rei, MG, Brasil
- Henriques, J.A. 2019. Proposição de Modelo de Predição de Desempenho de Estações Convencionais de Tratamento de Água utilizando Redes Neurais Artificiais. Tese de Doutorado em Saneamento, Meio Ambiente e Recursos Hídricos. Universidade Federal de Minas Gerais, Belo Horizonte, BH, Brasil.
- Kaastra, L.; Boyd, M. 1996. Designing a neural network for forecasting financial and economic time series. Neurocomputing. 10(3): 215-236.
- Leite, L.P.C.V. 2011. O sistema tributário brasileiro: análise da sonegação fiscal nas empresas brasileiras. Monografia – Bacharelado em Ciências Econômicas. Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.
- Lemos, R. 1999. Legalidade, crime de sonegação fiscal e o Ministério público. Revista da Fesmpdft.
- Lima, S.P. 2005. Crime Organizado e Lavagem de Dinheiro: Uma aplicação das teorias dos jogos e de redes neurais para reconhecimento e descrição de padrões. Tese de Doutorado em Engenharia da Produção. Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.
- Mendroni, M.B. 2015. Crime organizado: aspectos gerais e mecanismos legais. 5ed. Editora Atlas, São Paulo, SP, Brasil.
- Menezes, D.H. 2014. Tributação e Sonegação Fiscal: Um Estudo da Sonegação do ICMS. Dissertação de Mestrado em Economia. Universidade Federal do Ceará, Fortaleza, CE, Brasil.
- Nascimento, L. P. 2016. Elaboração de projetos de pesquisa: monografia, dissertação, tese e estudo de caso, com base em metodologia científica. Editora Cengage Learning, São Paulo, SP, Brasil.
- Oliveira, F. N.; Santos, L. P. G. D. 2020. 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.
- Oliveira, V.D. 2019. Redes neurais artificiais aplicadas à identificação de riscos de inadimplência fiscal de ICMS e ISS no Distrito Federal. Dissertação de Mestrado em Computação Aplicada. Universidade de Brasília, Brasília, DF, Brasil.
- Piccirilli, T.L. 2013. Mineração de Dados Aplicada a Classificação dos Contribuintes do ISS. Dissertação de Mestrado em Engenharia de Produção e Sistemas. Universidade Católica de Goiás, Goiânia, GO, Brasil.
- Piva, G.O. 2021. Modelo de Previsão de vagas de Educação Infantil de um Município Brasileiro. Dissertação de Mestrado em Engenharia de Produção. Universidade de Caxias do Sul, Bento Gonçalves, RS, Brasil.

- Quinlan, J. R. 1993. C4.5: Programs for Machine Learning, San Mateo:, Morgan Kaufmann Publishers.
- Santos, A.L. 2021. Uso de Rede Neural para Desenvolvimento de Sistema Especialista para Diagnóstico de Doenças Foliaves em Eucalipto. Dissertação de Mestrado em Produção Sustentável. Universidade Federal de São Carlos, Sorocaba, SP, Brasil.
- Samarasinghe, S. 2006. Neural Networks for Applied Sciences and Engineering. 1st Edition, Auerbach Publications, New York, disponível em: <<https://doi.org/10.1201/9780849333750>>. Acesso em: 15 fev. 2022.
- Schier, A.C.R. 2016. A Institucionalização da Administração Tributária. 1ed. Editora Íthala, Curitiba, SC, Brasil.
- Simões, A.C.A. 2008. Mineração de Dados baseada em Árvores de Decisão para Análise do Perfil de Contribuintes. Dissertação de Mestrado em Ciência da Computação. Universidade Federal de Pernambuco, Recife, PE, Brasil.
- Sisnando, S.R.A.; Freitas, M.A.S. 2006. Previsão e Avaliação do Desempenho dos contribuintes do ICMS do Estado do Ceará Utilizando as Redes Neurais Artificiais. Revista Econômica do Nordeste 37(1).
- Siu K.Y; Roychowdhury V; Kailath T. 1995. Toward massively parallel design of multipliers. Journal of Parallel and Distributed Computing.
- Souza, R.M. 2016. Inteligência computacional aplicada ao controle externo: classificação de padrões utilizando redes neurais artificiais. Revista do TCU, Brasília, jan-abr.
- Venturini, F.C. 2020. Uso de modelos preditivos na gestão de riscos da Fiscalização Tributária. Dissertação de Mestrado em Computação Aplicada. Universidade de Brasília, Brasília, DF, Brasil.
- Thompson. J. R. 2001. Estimation equations for kappa statistics. Statistics in Medicine 20(19): 2895 - 2906.
- Towell, G.G.; Shavlik, J.W. 1993. The Extraction of Refined Rules from Knowledge-Based Neural Networks, Machine Learning 13: 71-101.
- Tribunal de Contas do Estado de Goiás [TCE]. 2020. Relatório Técnico Contas do Governo. Disponível em: <<https://portal.tce.go.gov.br/documents/20181/336411/Relatório%20da%20Unidade%20Técnica%20-%20Contas%20do%20Governador%202020/9496e817-817b-4596-bb75-bd51ab39429d>>. Acesso em: 05 ago. 2022.
- Wolfrann, J. 2018. Redes Neurais Artificiais para Estimar a Precipitação na Irrigação por Aspersão. Dissertação de Mestrado em Tecnologias Computacionais para o Agronegócio. Universidade Tecnológica Federal do Paraná, Medianeira, PR, Brasil.

Anexo I

Código	Subgrupo Atividade Econômica
0	CNAE NAO ENCONTRADO
1	ATACADISTA DE COMBUSTÍVEL - TRR
3	ATACADISTA DE LUBRIFICANTES
4	ATACADISTA/DISTRIBUIDOR DE COMBUSTÍVEL (EXCETO TRR)
5	ATACADO AGRONEGOCIO
6	ATACADO CONSTRUÇÃO CIVIL, MINERAÇÃO E MAQUINAS
7	ATACADO DE ALIMENTOS
8	ATACADO DE BEBIDAS
9	ATACADO DE CALÇADOS
10	ATACADO DE CARNES
11	ATACADO DE FUMO
12	ATACADO DE LÁCTEOS
13	ATACADO DE MEDICAMENTOS E PROD. HOSPITALARES
14	ATACADO DE MOVEIS, ELETROELETRÔNICOS, CINEFOTO, SOM E REFRIGERAÇÃO
15	ATACADO DE PRODUTOS QUÍMICOS, PROD. HIG. E LIMP. PAPELARIA, EMBALAGENS
16	ATACADO DE VEÍCULOS E PEÇAS
17	ATACADO DE VESTUÁRIO
18	DISTRIBUIDORA DE ENERGIA ELÉTRICA
19	EXTRAÇÃO DA CONSTRUÇÃO CIVIL, MINERAÇÃO E MAQUINAS
21	INDÚSTRIA DA CONSTRUÇÃO CIVIL, MINERAÇÃO E MAQUINAS
22	INDÚSTRIA DE ÁLCOOL E AÇÚCAR
23	INDÚSTRIA DE ALIMENTOS
24	INDÚSTRIA DE BEBIDAS
26	INDÚSTRIA DE CALÇADOS
27	INDÚSTRIA DE CARNES
29	INDÚSTRIA DE COMBUSTÍVEL DERIVADO DE PETRÓLEO E COQUE
31	INDÚSTRIA DE LÁCTEOS
33	INDÚSTRIA DE MEDICAMENTOS E PROD. HOSPITALARES
34	INDÚSTRIA DE MOVEIS, ELETROELETRÔNICOS, CINEFOTO, SOM E REFRIGERAÇÃO
35	INDUSTRIA DE VEÍCULOS E PEÇAS
36	INDÚSTRIA DE VESTUÁRIO
37	INDÚSTRIA DO AGRONEGÓCIO
38	INDÚSTRIA GERADORAS DE ENERGIA ELÉTRICA
39	INDÚSTRIA PRODUTOS QUÍMICOS, PROD. HIG. E LIMP. PAPELARIA, EMBALAGENS E OUTROS
40	INDÚSTRIA TRANSMISSORA DE ENERGIA ELÉTRICA
41	ATIVIDADE NÃO SUJEITA AO ICMS
42	PRESTAÇÃO DE SERVIÇO DE MOVEIS, ELETROELETRÔNICOS, CINEFOTO, SOM E REFRIGERAÇÃO
43	PRESTAÇÃO DE SERVIÇO DA CONSTRUÇÃO CIVIL, MINERAÇÃO E MAQUINAS
44	PRESTAÇÃO DE SERVIÇO DO AGRONEGÓCIO
45	PRESTAÇÃO DE SERVIÇOS DE VEÍCULOS E PEÇAS
46	PRESTAÇÃO DE SERVIÇOS OUTROS DE COMUNICAÇÃO
47	PRESTAÇÃO DE SERVIÇOS PRODUTOS QUÍMICOS, PROD. HIG. E LIMP, PAPELARIA

48	PRESTAÇÃO DE SERVIÇOS TELECOMUNICAÇÕES
49	PRODUÇÃO AGROPECUÁRIA
51	TRANSPORTES
52	VAREJISTA DE COMBUSTÍVEL
53	VAREJISTA DE GÁS LIQUEFEITO DE PETRÓLEO - GLP
54	VAREJISTA DE LUBRIFICANTES
55	VAREJO AGRONEGÓCIO
56	VAREJO CONSTRUÇÃO CIVIL, MINERAÇÃO E MAQUINAS
57	VAREJO DE ALIMENTOS
58	VAREJO DE BEBIDAS
59	VAREJO DE CALÇADOS
60	VAREJO DE CARNES
61	VAREJO DE FUMO
62	VAREJO DE LÁCTEOS
63	VAREJO DE MEDICAMENTOS E PROD. HOSPITALARES
64	VAREJO DE MOVEIS, ELETROELETRÔNICOS, CINEFOTO, SOM E REFRIGERAÇÃO
65	VAREJO DE PRODUTOS QUÍMICOS, PROD. HIG. E LIMP. PAPELARIA, EMBALAGENS E OUTROS
66	VAREJO DE VEÍCULOS E PEÇAS
67	VAREJO DE VESTUÁRIO
