

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS
GERAIS – *CAMPUS* FORMIGA
MESTRADO PROFISSIONAL EM ADMINISTRAÇÃO

Denis Ribeiro da Silva

**APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL PARA CONHECIMENTO E
REGISTRO DIGITAL DE PRODUTOS COM ÊNFASE NA CLASSIFICAÇÃO
TRIBUTÁRIA**

Formiga

2023

DENIS RIBEIRO DA SILVA

**APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL PARA CONHECIMENTO E
REGISTRO DIGITAL DE PRODUTOS COM ÊNFASE NA CLASSIFICAÇÃO
TRIBUTÁRIA**

Trabalho de conclusão de curso apresentado ao Programa de Pós-graduação em Administração do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais - IFMG - Campus Formiga, como requisito para conclusão do curso Mestrado Profissional em Administração.

Orientadores: Prof. Dr. Alexandre Pimenta
Prof. Dr. Daniel Fonseca Costa

Formiga

2023

S586a Silva, Denis Ribeiro da.

Aplicação da inteligência artificial para conhecimento e registro digital de produtos com ênfase na classificação tributária / Denis Ribeiro da Silva. - Formiga, 2023

126 p. : il. color.

Dissertação (Mestrado Profissional em Administração) – Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – Campus Formiga, 2023.

Orientador: Dr. Alexandre Pimenta.

1. Inteligência Artificial. 2. Tomadas de Decisão. 3. Contabilidade. 4. Tributação. 5. Processamento de Dados. I. Silva, Denis Ribeiro da. II. Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – Campus Formiga. Título.

CDD:657:336.22

Catálogo: Livia Renata Santos- CRB/6-2561

Denis Ribeiro da Silva

**APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL PARA CONHECIMENTO E
REGISTRO DIGITAL DE PRODUTOS COM ÊNFASE NA CLASSIFICAÇÃO
TRIBUTÁRIA**

Trabalho de conclusão de curso apresentado ao Programa de Pós-graduação em Administração do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais - IFMG - Campus Formiga, como requisito para conclusão do curso Mestrado Profissional em Administração.

Aprovado em ____/____/____ pela banca examinadora:

Prof. Dr. Alexandre Pimenta - IFMG (Orientador)

Prof. Dr. Daniel Fonseca - IFMG (Coorientador)

Wagner Antônio Marques - UFES (Avaliador)

Washington Santos da Silva (Avaliador)

Dedico esta dissertação à minha amada filha, Sophia, e à minha esposa, Fernanda.

AGRADECIMENTOS

Agradeço aos meus pais, Osmar e Neusa; ao meu irmão, Igor; e à minha esposa, Fernanda, por sempre me apoiarem em todos os momentos e situações.

Agradeço aos orientadores, Alexandre e Daniel, por acreditarem neste trabalho e me apoiarem durante todos os momentos e etapas.

Agradeço à direção da empresa Touch Comp por apoiar e disponibilizar os materiais necessários ao andamento e conclusão deste trabalho.

“Qualquer tecnologia suficientemente nova é indistinguível da magia”.

Arthur C. Clarke

RESUMO

A Inteligência Artificial está cada vez mais presente em nossas vidas, desde os equipamentos que fazem parte de nosso uso diário, como nossos *smartphones*, computadores e assistentes virtuais, até os serviços que utilizamos, como em compras na internet, interações nas redes sociais digitais, portfólios e análise de investimentos, previsão do tempo, avaliação de riscos, dentre diversos outros. A Inteligência Artificial vem sendo aplicada nas atividades que são trabalhosas, onerosas, que fornecem risco à saúde humana, como atividades em ambientes hostis, ou, ainda, atividades que o ser humano não consegue realizar, seja por tempo, capacidade ou custo, como análises preditivas de doenças, análises comportamentais, auditorias e outras. Uma das atividades que são onerosas e trabalhosas, proporcionadas pelas milhares de normas e leis tributárias brasileiras, são aquelas relacionadas à Tributação. Segundo o IBPT (2020), desde 1988, são editadas 829 normas fiscais por dia útil. Neste trabalho, pesquisamos o uso da Inteligência Artificial nessa área, assim como os trabalhos e as aplicações existentes, e propusemos novos produtos técnicos que poderão facilitar as atividades relacionadas à área tributária para os contribuintes. Encontramos lacunas de pesquisa e a falta de ferramentas voltadas ao contribuinte, sendo que as existentes buscam auxiliar o Estado em coibir a evasão de tributos. Diante desse cenário, foram criados um produto bibliográfico, três produtos técnicos relacionados aos objetivos específicos e, por fim, um produto técnico referente ao objetivo principal deste trabalho. Além disso, foram elaborados um artigo de revisão de literatura, que foi publicado no *USP International Conference in Accounting 2022*; um Algoritmo de Leitura e Processamento de Arquivos XML de Notas Fiscais; um Banco de Dados de Produtos, Pessoas e Dados Fiscais; um Algoritmo Processamento e Exportação de Dados Cadastrais e Fiscais; e, pelo objetivo principal, um Algoritmo de Inteligência Artificial baseado em Árvores de Decisão para a Classificação Tributária ICMS. Esses produtos foram implementados e unificados em um único programa. Nas medições realizadas, os produtos técnicos apresentaram bons resultados, como o tempo de leitura, processamento e persistência dos dados no banco de dados, com uma taxa de 165 documentos por segundo. A precisão do Algoritmo de Inteligência Artificial alcançou 97,2%, obtidos após 2,42 minutos de processamento e treinamento sobre uma base de 2.839.599 movimentos fiscais.

Palavras-chave: Inteligência Artificial. Tomadas de Decisão. Contabilidade. Tributação. Processamento de Dados.

ABSTRACT

Artificial Intelligence is increasingly present in our lives, from devices that are part of our daily use, such as our smartphones, computers, and virtual assistants, to services we use, such as online shopping, digital social media interactions, portfolios and investment analysis, weather forecasting, risk evaluation, and many others. AI is being applied to activities that are laborious, costly, and pose risks to human health, such as activities in hostile environments or those that humans cannot perform due to time, capacity, or cost constraints, such as predictive disease analysis, behavioral analysis, audits, and others. One of the activities that is laborious and costly, due to the thousands of Brazilian tax laws and regulations, is taxation-related activities. According to IBPT (2020), since 1988, 829 tax regulations are edited per business day. In this work, we researched the use of AI in this area, the existing works and applications, and proposed new technical products that could facilitate tax-related activities for taxpayers. We found research gaps and the lack of tools aimed at taxpayers, where existing tools seek to assist the government in curbing tax evasion. Given this scenario, a bibliographic product, three technical products related to specific objectives, and finally, a technical product related to the main objective of this work were created. A literature review article was created, which was published in the USP International Conference in Accounting 2022; XML Invoice Reading and Processing Algorithm; Database of Products, People, and Tax Data; Cadastral and Fiscal Data Processing and Export Algorithm; and, for the main objective, an ICMS Tax Classification Algorithm based on Artificial Intelligence. These products were implemented and unified into a single program. In the measurements performed, the technical products presented good results, such as the reading, processing, and data persistence time in the database, with a rate of 165 documents per second, the precision of the Artificial Intelligence Algorithm, which reached 97.2% accuracy, obtained after 2.42 minutes of processing and training on a database of 2,839,599 tax movements.

Keywords: Artificial Intelligence. Decision making. Accounting. Taxation. Data processing.

LISTA DE ABREVIATURAS E SIGLAS

IDC - *International Data Corporation*

PIS - Programa Integração Social

COFINS - Contribuição para Financiamento da Seguridade Social

IPI - Impostos sobre Produtos Industrializados

ICMS - Imposto sobre a Circulação de Mercadorias e Serviços de Transporte Interestadual e Intermunicipal e Comunicação

ICMSST - Imposto sobre a Circulação de Mercadorias e Serviços de Transporte Interestadual e Intermunicipal e Comunicação Substituição Tributária

SPED - Sistema Público de Escrituração Digital

NFE - Notas Fiscais Eletrônicas

XML - *Extended Markup Language*

CST - Código Situação Tributária

IA – Inteligência Artificial

ERP - *Enterprise Resource Planning*

RFID - *Radio Frequency Identification*

CFOP - Código Fiscal de Operações e Prestações

CNAE - Código Nacional Atividade Econômica

NCM - Nomenclatura Comum Mercosul

CRT - Código Regime Tributário

CEST - Código de Especificação Substituição Tributária

SVM - *Support Vector Machine*

CNPJ - Cadastro Nacional Pessoa Jurídica

DTO - *Data Transfer Object*

IBGE - Instituto Brasileiro de Geografia e Estatística

PCD - Pessoa com Deficiência

UF - Unidade Federativa

SUMÁRIO

1 INTRODUÇÃO GERAL	11
2 REFERENCIAL TEÓRICO.....	16
2.1 Inteligência Artificial.....	16
2.2 Inteligência Artificial nas organizações.....	18
2.3 Inteligência Artificial e contabilidade.....	20
2.4 Inteligência Artificial e tributos.....	22
2.5 Tributação, ICMS e Nota Fiscal Eletrônica.....	24
2.6 Sistemas de Aprendizado de Máquina – Classificadores.....	26
3 METODOLOGIA	28
3.1 Materiais.....	29
4 CONSIDERAÇÕES FINAIS.....	31
REFERÊNCIAS.....	34
PARTE 2: PRODUTOS BIBLIOGRÁFICOS E TÉCNICOS.....	39
PRODUTO 1 (bibliográfico): A Influência da Inteligência Artificial na Contabilidade e na Tributação das Organizações: uma revisão de literatura.....	40
PRODUTO 2 (Técnico/tecnológico): Algoritmo de Leitura de Arquivos XML de Notas Fiscais Eletrônicas.....	63
PRODUTO 3 (Técnico/tecnológico): Banco de dados.....	71
PRODUTO 4 (Técnico/Tecnológico): Algoritmo de Exportação de dados.....	81
PRODUTO 5 (Técnico/tecnológico): Algoritmo de Inteligência Artificial.....	95
APÊNDICE A – BIBLIOTECA TRIBUO.....	114
APÊNDICE B – INTERFACE DA APLICAÇÃO.....	116
APÊNDICE C – REPOSITÓRIO DO PROJETO.....	125

1 INTRODUÇÃO

O advento da Internet, a evolução da capacidade computacional dos dispositivos e o crescimento da web provocaram aumentos de ordem exponencial na geração de dados. Segundo a *International Data Corporation* (IDC), a quantidade de dados gerados dobraria a cada dois anos. Entretanto, 90% de todos os dados gerados pela Internet foram produzidos em apenas dois anos. Dados gerados em volume, em variedade, com baixa confiabilidade e em alto valor, bem como a velocidade crescente com que esses dados se acumulam, a ponto de desafiar as capacidades computacionais atuais, são conhecidos como *Big Data* (HARIRI; FREDERICKS; BOWERS, 2019).

Esse grande volume de dados é produzido pelas mais variadas conexões e dispositivos que usamos atualmente, tais como *smartphones*, sensores, computadores e sistemas cada vez mais inteligentes e conectados. Essa variedade de fonte de dados provoca uma baixa veracidade, por se tratar de dados sem padronização. Por outro lado, se tratados de maneira adequada, podem gerar informações de grande valor para as organizações nos processos de tomada de decisão (DUBEY *et al.*, 2020).

Para tratar esses dados e gerar informações valiosas, técnicas computacionais tradicionais de desenvolvimento de software não são eficientes, face à grande variabilidade e volume de dados e ao número de regras e informações existentes. Nesse aspecto, a Inteligência Artificial e suas subáreas ganharam notoriedade e importância, com novos métodos de tratamento de dados que conseguem identificar padrões, tendências e extrair informações úteis e, assim, aprender a partir dos dados disponíveis. Com isso, a Inteligência Artificial consegue realizar novas classificações e inferências em novas situações até então desconhecidas (DUBEY *et al.*, 2020).

Além disso, a Inteligência Artificial proporcionou um salto na performance dos afazeres e na diminuição de custos relativos aos computadores e hardwares. Isso ocorreu pela melhoria das tecnologias dos equipamentos, inclusive pelo advento da criação de hardwares específicos para aplicações baseadas em Inteligência Artificial (MA; SUN, 2020; NGUYEN *et al.*, 2019), o que tem provocado a popularização do uso dessa tecnologia.

A conjunção desses fatores relacionados ao uso da Inteligência Artificial, tal como a disponibilidade de dados e de recursos computacionais, está transformando diversas áreas de pesquisa pela possibilidade de se analisar grandes bases de dados, levando a avanços tecnológicos (NGUYEN *et al.*, 2019). Com isso, a Inteligência Artificial tem sido utilizada para processar, interpretar, correlacionar, realizar previsões e avaliar o comportamento das pessoas

e do mercado, a fim de criar produtos, serviços e aumentar a eficiência operacional, evitando erros, objetivando resultados e acelerando o processo de tomada de decisões (DUAN; EDWARDS; DWIVEDI, 2019).

Nessa tendência, as organizações têm usado a Inteligência Artificial de diversas formas para cumprir com seus objetivos e com suas obrigações. Assim, as empresas comerciais estão utilizando a Inteligência Artificial para otimizar as vendas e melhorar a divulgação dos produtos (FERFERNIG; TERPAN; GULA, 2007; GREWAL; ROGGEVEEN; NORDFÄLT, 2017; MA; SUN, 2020); as empresas do setor financeiro estão utilizando *chatbots* (HILDEBRAND; BERGNER, 2020) e ferramentas para a captação e análise do perfil de clientes, para a criação de portfólios de investimentos e para a análise de riscos e de *scores* de seus clientes (CUNHA; SILVEIRA, 2020; GIUDICI, 2018); as indústrias empregam a Inteligência Artificial em sistemas de monitoramento e em sensores para diminuir perdas e aumentar a eficiência nos processos produtivos (X. ZHAN; MING, 2021; ZHENG *et al.*, 2018); por sua vez, os governos têm aplicado a Inteligência Artificial no combate à evasão fiscal (FAÚNDEZ-UGALDE; MELLADO-SILVA; ALDUNATE-LIZANA, 2020); e as empresas de contabilidade, na automação de suas tarefas (KOKINA; DAVENPORT, 2017; C. S. LEE; TAJUDEEN, 2020; Y. ZHANG *et al.*, 2020).

Por um lado, temos aplicações promissoras da Inteligência Artificial; por outro, existem poucos estudos que a aplicam na área tributária (FAÚNDEZ-UGALDE; MELLADO-SILVA; ALDUNATE-LIZANA, 2020; OLIVEIRA; SANTOS, 2020; HÖGLUND, 2017; CHEN *et al.*, 2011; MABE-MADISA, 2018), sendo que os existentes são mais voltados para coibir a evasão fiscal do que para auxiliar as empresas quanto à classificação, reconhecimento e apuração correta dos tributos.

No Brasil, temos uma situação similar. Os estudos existentes buscam coibir a evasão fiscal e não propriamente auxiliar o contribuinte no reconhecimento e apuração correta dos tributos (OLIVEIRA; SANTOS, 2020; DIAS FILHO; SANTOS, 2021). Neste aspecto, pode-se, inclusive, destacar o projeto Harpia, criado para auxiliar na identificação de contribuintes suspeitos de praticarem a evasão de tributos (FAÚNDEZ-UGALDE; MELLADO-SILVA; ALDUNATE-LIZANA, 2020).

Além disso, o Brasil tem uma situação peculiar em comparação com outros países, por existir um vasto e complexo emaranhado de tributos e de legislações tributárias (FELICIO; MARTINEZ, 2018). Um estudo realizado pelo Instituto Brasileiro de Planejamento e Tributação (IBTP) identificou que, desde 1988, foram publicadas 829 normas tributárias por dia útil (IBTP, 2022). Dentre os tributos brasileiros que influenciam as organizações (COSTA

et al, 2016), podem-se citar o Imposto de Renda, a Contribuição Social sobre o Lucro, o Programa Integração Social (PIS), a Contribuição para Financiamento da Seguridade Social (Cofins) e o Imposto sobre Produtos Industrializados (IPI), de responsabilidade legislativa federal; o Imposto sobre a Circulação de Mercadorias e Serviços de Transporte Interestadual e Intermunicipal e Comunicação (ICMS), de responsabilidade estadual, e o Imposto sobre Serviços de Qualquer Natureza (ISSQN), o qual é de competência dos municípios (FELICIO; MARTINEZ, 2018).

Por outro lado, tivemos grandes avanços tecnológicos no Brasil com a instituição do Sistema Público de Escrituração Digital (SPED) (RECEITA FEDERAL DO BRASIL, 2022c) e, especialmente, com o subprojeto da Nota Fiscal Eletrônica (NFE) (RECEITA FEDERAL DO BRASIL, 2022b), que acabou por padronizar e homogeneizar essas operações realizadas pelas empresas, contribuindo, dessa forma, com a criação de grandes bases de dados de tais operações.

Diante do exposto, observa-se que existe a necessidade de uma aplicação que utilize a Inteligência Artificial, capaz de gerenciar, classificar e apurar os tributos que influenciam as empresas. Essa necessidade se destaca em função da complexidade do sistema tributário brasileiro; do número de impostos, taxas e contribuições; por haver poucos trabalhos e pesquisas nessa área; e pela dificuldade de utilização de técnicas convencionais de programação para parametrizar todas as regras e leis que versam sobre as operações tributárias realizadas pelas empresas.

Dos principais tributos que influenciam as empresas brasileiras, pode-se destacar o ICMS como um dos mais complexos, uma vez que há 26 Estados e o Distrito Federal que legislam em seu favor. Esse fato faz com que uma empresa que negocie produtos em todo o Brasil tenha que se atentar sobre a legislação que versa sobre o ICMS em cada Estado, para, assim, cumprir corretamente as obrigações acessórias e a obrigação principal. Além disso, o ICMS tem diversas situações tributárias relacionadas com o produto e com a operação, as quais precisam ser observadas nas entradas e nas saídas dos produtos. Diante desta complexidade, o processo de definição quanto ao ICMS é realizado manualmente pelos profissionais da área fiscal, com base em seu conhecimento e interpretação da legislação.

Devido à complexidade e ao trabalho em se classificar e definir a tributação de um produto, surge a necessidade de se desenvolver um sistema que seja capaz de aprender, com os dados existentes, a classificar a tributação do ICMS em relação aos produtos e às operações futuras.

Diante desse problema, o objetivo deste trabalho é desenvolver um sistema, baseado

em Inteligência Artificial, que seja responsável pela determinação da classificação tributária do ICMS. O Código de Situação Tributária do ICMS (CST) é composto por 3 dígitos. O primeiro define a procedência do produto, se nacional ou importado, e os dois últimos dígitos determinam propriamente a tributação. Por exemplo, o código 000 caracteriza o produto como nacional e tributado integralmente (RECEITA FEDERAL DO BRASIL, 2022b).

Portanto, para o desenvolvimento do sistema de classificação do ICMS, baseado em Inteligência Artificial, foi necessário:

- Realizar uma revisão de literatura sobre o tema;
- Desenvolver uma interface de comunicação e de leitura dos arquivos XML das Notas Fiscais Eletrônicas (NFE);
- Desenvolver uma base de dados para cadastro e armazenamento de informações sobre os produtos, incluindo as informações tributárias;
- Desenvolver uma interface de exportação de dados, de cadastros e de movimentações fiscais;
- Desenvolver um algoritmo, baseado em Inteligência Artificial, capaz de realizar o reconhecimento e registro digital de produtos, com ênfase na Classificação Tributária do ICMS.

Frente ao propósito deste trabalho, a execução dos objetivos específicos permitiu alcançar o objetivo geral, contribuindo com a área de tributação e com as empresas, ao criar uma aplicação que facilitará a classificação e os registros tributários. Para tanto, foram produzidos os seguintes produtos bibliográficos e técnicos:

- Um artigo de revisão de literatura que identifica as principais tendências de utilização de Inteligência Artificial em situações tributárias, conforme Produto 1 (Bibliográfico): Artigo de Revisão de Literatura;
- Um algoritmo de leitura e processamento dos arquivos XML das Notas Fiscais Eletrônicas (NFE), conforme Produto 2 (Técnico/tecnológico): Algoritmo de Leitura de Arquivos XML de Notas Fiscais Eletrônicas, com foco na performance de leitura de grande volume de documentos NFE;
- Uma base de dados (banco de dados) para cadastro e armazenamento de informações sobre os produtos, inclusive de informações tributárias, conforme Produto 3 (Técnico/tecnológico): Banco de Dados;
- Um algoritmo para exportação de dados de produtos, clientes, fornecedores, naturezas de operação e tributações em formato CSV, conforme Produto 4

(Técnico/tecnológico): Algoritmo Exportação Dados;

Algoritmo de Inteligência Artificial baseado em Árvores de Decisão, capaz de realizar a classificação tributária do ICMS em relação aos produtos e operação, conforme Produto 5 (Técnico/tecnológico): Algoritmo Inteligência Artificial.

O sistema será composto por quatro produtos técnicos distintos: Algoritmo de Leitura e Processamento dos arquivos XML; o Banco de Dados; Algoritmo de Exportação de dados; e o Algoritmo de Inteligência Artificial. Esses produtos podem ser utilizados de forma independente, em outros aplicativos. Neste trabalho, todos os produtos foram adicionados em um único programa, visando atingir o objetivo principal, que é realizar a classificação tributária do CST ICMS com a utilização de Inteligência Artificial. Esse programa conta com menus e interface gráfica para a imputação de parâmetros, dados e utilização de todos os produtos técnicos citados anteriormente. Essa interface gráfica e os detalhes de uso estão especificados no Apêndice B.

Este trabalho está dividido em duas partes. A primeira abrange a introdução, referencial teórico, metodologia utilizada em todos os produtos e as considerações finais sobre o trabalho. A segunda parte detalha cada produto técnico/bibliográfico, considerando a metodologia específica, resultados e considerações sobre cada produto. Ao final, temos três apêndices que trazem informações adicionais sobre o presente estudo.

2 REVISÃO DE LITERATURA

2.1 Inteligência Artificial

A Inteligência Artificial (IA) auxilia na simulação e na criação de soluções para problemas específicos de maneira mais eficiente e precisa. Assim, a IA ajuda os indivíduos a solucionarem problemas em diversas áreas e setores, bem como em complexos fluxos de trabalho (MUTHUKRISHNAN *et al.*, 2020). A Inteligência Artificial pode ser definida como o comportamento das máquinas em aprender, a partir de experiências, e se adaptar, conforme o ambiente, para simular a inteligência humana em tomadas de decisão e em processos racionais (MUTHUKRISHNAN *et al.*, 2020; FAN *et al.*, 2020; KAPLAN; HAENLEIN, 2019; BORGES *et al.*, 2021). Sua história começa com Alan Turing, cujo teste proposto por ele parte da premissa de que quando uma pessoa interroga remotamente um computador e não consegue dizer se é um computador ou outra pessoa, o sistema é descrito como inteligente (COOPER; LEEUWEN, 2013).

Ao longo da história e da construção do conceito da Inteligência Artificial, pode-se compreender a sua ligação com a inteligência e com o cérebro humano. John McCarthy definiu, em 1955, a Inteligência Artificial como: “todos os aspectos da aprendizagem — ou qualquer outra característica da inteligência — podem, em princípio, serem descritos tão precisamente que uma máquina será capaz de simulá-los” (MORGENSTERN; MCILRAITH, 2011). Muitos dos algoritmos e tecnologias relacionadas à Inteligência Artificial surgiram diretamente de estudos ligados ao funcionamento do cérebro humano, em que podemos destacar as redes neurais e seus subtipos, inspiradas nas conexões e nas sinapses entre os neurônios (FAN *et al.*, 2020). Por meio de redes neurais, mais precisamente redes neurais profundas (Deep Learning), é possível aprender novas tarefas e preservar o conhecimento adquirido da experiência de antigas tarefas, simulando uma das principais capacidades de nossa inteligência, o aprendizado contínuo (FAN *et al.*, 2020).

A Inteligência Artificial pode ser classificada, quanto à aprendizagem, em três tipos: supervisionada, não supervisionada e por reforço (NGUYEN *et al.*, 2019; MA; SUN, 2020). Os autores dizem que, de acordo com as características dos dados a serem analisados, um ou mais tipos de algoritmos poderão ser indicados e utilizados.

Para Nguyen *et al.* (2019) e Bruyn *et al.* (2020), a aprendizagem supervisionada é utilizada nos casos em que os dados são conhecidos e rotulados, e a aprendizagem não supervisionada é indicada para situações em que os dados não são conhecidos ou rotulados, com a finalidade de buscar padrões e comportamentos desconhecidos. Por outro lado, a

aprendizagem por reforço define como a Inteligência Artificial se relaciona com o ambiente e aprende com ele, de forma a selecionar as ações apropriadas conforme a situação e para maximizar as respostas. Basicamente, os algoritmos aprendem com a tentativa e o erro até conseguirem as respostas apropriadas, e, assim, a Inteligência Artificial aprende e evolui, tornando suas respostas cada vez mais precisas (NGUYEN *et al.*, 2019; BRUYN *et al.*, 2020).

Além disso, a Inteligência Artificial pode ser categorizada em três grupos: Analítica (*Analytical AI*), Inspiração Humana (*Human-Inspired AI*) e Humanizada (*Humanized AI*). A Inteligência Artificial Analítica tem apenas características cognitivas e, de acordo com as experiências passadas, consegue informar sobre decisões futuras. Para esse tipo de Inteligência Artificial, as soluções são mais especializadas e focadas em questões e problemas mais específicos. É nessa classificação que se encontra a maioria dos sistemas de Inteligência Artificial atualmente. A Inteligência Artificial Inspiração Humana é definida como sistemas cognitivos que conseguem entender emoções humanas, tais como expressões de surpresa, tristeza, felicidade, entre outras. Por fim, a Inteligência Artificial Humanizada é entendida como sistemas completos, cognitivos, interpretativos e sociais que conseguem ter consciência e aprendizado próprio e que ainda não existem (KAPLAN; HAENLEIN, 2019).

Frente a isso, as aplicações que utilizam a Inteligência Artificial são cada vez mais frequentes em todos os ramos da economia, sendo empregada nas esferas pública, privada, corporativa e social. Essa utilização tem se dado em maior escala devido, principalmente, à disponibilidade e quantidade de dados (*Big Data*), os quais começam a fazer parte do cotidiano, influenciando como os consumidores realizam suas compras e como empresas tomam suas decisões e interagem com os seus acionistas (HAENLEIN; KAPLAN, 2019). Na atualidade, robôs têm sido amplamente utilizados para conversar e interagir com as pessoas, principalmente em empresas fundadas com premissas tecnológicas. Nesse contexto, empresas do setor financeiro e as *fintechs* usam robôs, inclusive, para captar novos clientes, definir perfil dos investidores e, assim, escolher as melhores carteiras de investimentos. Nesse ponto, quanto mais interativos forem os robôs, maior será a confiança dos clientes nos produtos dessas empresas (HILDEBRAND; BERGNER, 2020).

Com esses avanços, pelo volume de dados gerados e disponíveis, aumento da capacidade computacional, somados à principal característica multiuso da Inteligência Artificial, de se adaptar e resolver diversos e variados problemas, as soluções baseadas em algoritmos de Inteligência Artificial se destacam (BORGES *et al.*, 2021).

Além disso, a Inteligência Artificial tem sido utilizada para solucionar problemas e melhorar processos em diversas áreas, em evasão de tributos (FAÚNDEZ-UGALDE;

MELLADO-SILVA; ALDUNATE-LIZANA, 2020; OLIVEIRA; SANTOS, 2020; HÖGLUND, 2017; CHEN *et al.*, 2011; MABE-MADISA, 2018), em processos contábeis e de auditoria (MESERVY; DENNA; HANSEN, 1992; KOKINA; DAVENPORT, 2017; BEHRENS; STEINBART, 1992; LEE; TAJUDEEN, 2020), em otimização de diversos processos, como classificar e categorizar melhor os produtos (PAWŁOWSKI, 2021), aumentar as vendas e melhorar a experiência dos clientes em suas compras (GREWAL; ROGGEVEEN; NORDFÄLT, 2017) e otimizar processos produtivos (ZHENG *et al.*, 2018; LEE; KAO; YANG, 2014), sendo esses alguns dos exemplos em que a Inteligência Artificial está otimizando processos nas organizações.

2.2 Inteligência Artificial nas organizações

Nas organizações, a Inteligência Artificial é utilizada, principalmente, no auxílio de tomadas de decisão, ao se tornar uma ferramenta estratégica para diferenciar as organizações em um mercado cada vez mais competitivo (BORGES *et al.*, 2021). As soluções mais empregadas nas organizações são aplicadas de forma especializada e para resolver problemas mais específicos (KAPLAN; HAENLEIN, 2019).

Em várias situações, os modelos de auxílio a tomadas de decisão propostos por soluções de Inteligência Artificial são, inclusive, melhores e mais eficientes do que os realizados por humanos (BORGES *et al.*, 2021), principalmente em avaliar grande volume de dados, para encontrar padrões e sugerir soluções ou percepções de comportamento (DUAN; EDWARDS; DWIVEDI, 2019).

Identificar comportamentos e perfis de consumidores é um dos principais objetivos de algoritmos de Inteligência Artificial utilizados em soluções de *e-commerce*, seja em uma melhor definição e categorização de produtos (PAWŁOWSKI, 2021) ou para avaliar as melhores estratégias de se chegar ao público-alvo, principalmente, com características e novidades que devem chamar a atenção dos clientes (GREWAL; ROGGEVEEN; NORDFÄLT, 2017). O objetivo principal é trazer uma experiência positiva a eles e, assim, agregar maior valor aos produtos (MA; SUN, 2020).

No mundo virtual, quanto mais assertivos forem a publicidade e os produtos, maiores serão o interesse e a chance de uma nova venda ser concretizada (GREWAL; ROGGEVEEN; NORDFÄLT, 2017). Nesse aspecto, as empresas passaram a investir, principalmente, em algoritmos de Inteligência Artificial que aprendem as preferências do cliente com base em suas pesquisas e comportamentos de consumo, agrupando os clientes e

criando modelos de aprendizado conforme as características destes (BORGES *et al.*, 2021). Com um planejamento mais aprimorado em vendas, as empresas conseguem se preparar melhor para atender ao fluxo de pedidos, principalmente no caso de pequenas empresas, que, geralmente, têm recursos limitados (NASCIMENTO *et al.*, 2020).

Com o fluxo de vendas melhor definido, é possível também otimizar o setor de produção, planejar melhor as compras de insumos e matéria-prima, além de recursos humanos e maquinário, de forma a atender em tempo e não atrasar as entregas. É nesse setor que a Inteligência Artificial pode iniciar uma próxima revolução e a criação de uma nova era, que será a união dos mundos físico e virtual, por meio de sistemas *cyber-physical*, um marco da revolução industrial 4.0. O uso de sistemas inteligentes em diversas subáreas da manufatura irá acelerar as tomadas de decisão, diminuir falhas e trazer economia às indústrias. O aumento do uso de sensores nas máquinas e a comunicação *on-line* com sistemas de monitoramento agilizam intervenções, quando necessárias, evitam paradas repentinas e aumentam a qualidade dos produtos (ZHENG *et al.*, 2018; LEE; KAO; YANG, 2014).

A integração desses mundos e a descoberta de novas tecnologias também proporcionam uma maior percepção aos consumidores quanto à inovação, qualidade e variedade de produtos, além de diminuir o lead time de fabricação e aumentar a velocidade de entrega aos clientes, o que minimiza os prejuízos com perdas de vendas e de produtos (LEE; KAO; YANG, 2014; NASCIMENTO *et al.*, 2020). Nascimento *et al.* (2020) demonstram como o uso de técnicas de Inteligência Artificial pode auxiliar na previsão de vendas, com assertividade de 81% em pequenas empresas. Se há uma previsão de vendas assertiva, a produção, quando aplicável, e o estoque podem ser melhor previstos, o que resulta em menos perdas e mais lucratividade às empresas.

No setor de finanças, observa-se uma maior preocupação tanto em relação às análises de riscos de insolvência e quanto à segurança financeira das empresas, (MELNYCHENKO, 2020; LIN; HSU, 2017), provavelmente devido ao maior esforço de investidores em avaliar a saúde financeira das empresas.

Em finanças, a maioria dos estudos e trabalhos com o uso da Inteligência Artificial se concentra no próprio ramo financeiro, ou seja, em empresas cujo negócio são finanças, principalmente as *fintechs*, que estão revolucionando o mercado financeiro, com *chatbots* cada vez mais interativos (HILDEBRAND; BERGNER, 2020). Essas ferramentas se provam eficazes para a captação e análise do perfil de clientes, criação de portfólios de investimentos, bem como análise de riscos e *scores* (GIUDICI, 2018; CUNHA; SILVEIRA, 2020). Cunha e Silveira (2020), inclusive, avaliaram a implantação de um software que tem recursos de

Inteligência Artificial para avaliação dos contratos de uma instituição financeira brasileira. Houve redução na avaliação de contratos de cinco para zero dias e uma redução de 19% de custos (CUNHA; SILVEIRA, 2020).

2.3 Inteligência Artificial e contabilidade

A contabilidade estuda o patrimônio e suas variações, tendo como objetivo gerar informações financeiras, econômicas e patrimoniais para a tomada de decisão. Nesse sentido, os sistemas de informações contábeis registram a entrada de dados de informações internas e externas, processam os dados e realizam a saída por meio de informações úteis para os diversos usuários da contabilidade (HENDRIKSEN; VAN BREDA; 1999).

No setor contábil, a tecnologia introduziu diversas mudanças, as quais geraram um grande impacto no trabalho dos profissionais do setor e no processamento das informações. Os sistemas integrados, tais como o *Enterprise Resource Planning* (ERP), trouxeram para a área da contabilidade inúmeras vantagens: disponibilidade de dados de forma mais rápida e eficiente, redução de tempo na apresentação de resultados, aumento na eficiência e na qualidade quanto à tomada de decisão, auditorias de relatórios e, ainda, na própria redução do número de contadores nas empresas (KANELLOU; SPATHIS, 2013).

Por outro lado, os robôs diminuem trabalhos repetitivos e humanamente demorados, liberando os profissionais de contabilidade para tarefas que requerem mais mensuração, ciência, análise de dados e tomadas de decisão (ZHANG *et al.*, 2020). Com a introdução da Inteligência Artificial, as empresas podem aumentar ainda mais sua performance e se manterem cada vez mais competitivas e sustentáveis (LEE; TAJUDEEN, 2020).

A utilização da Inteligência Artificial já é pautada há vários anos no setor contábil. Meservy, Denna e Hansen (1992) já abordavam o uso em *softwares* desenvolvidos na *Brigham Young University*, em Utah, nos Estados Unidos. Esses *softwares* atuavam no auxílio de tomadas de decisão, tributos e, principalmente, auditoria (MESERVY *et al.*, 1992).

Lee e Tajudeen (2020) mostraram o uso de tecnologias baseadas em Inteligência Artificial em empresas na Malásia. As empresas analisadas já utilizam soluções para automatizar processos, principalmente contas a pagar, com automações de leitura, armazenamento e cadastro de *invoices*, auditoria de processos e liberações, bem como auditoria quanto às ações executadas pelos usuários. Como principais resultados, concluíram: aumento da eficiência, da produtividade e da qualidade dos serviços e produtos aos seus clientes, maior flexibilidade aos colaboradores, melhoria de processos e diminuição de tempo nas tarefas

executadas pelos colaboradores.

As principais empresas do setor contábil, conhecidas como *The Big Four*, também já investem em tecnologias de robotização de tarefas com o uso de IA, as quais são capazes de efetuar reconhecimento automático de dados, escriturar documentos e gerar relatórios financeiros. Esses sistemas permitem que outros profissionais, sem conhecimento contábil, realizem tarefas que, antes, somente contadores podiam executar (ZHANG *et al.*, 2020). Essas empresas têm estratégias diferentes, entre criar suas próprias ferramentas ou, ainda, realizar parcerias com outras empresas. A exemplo, a KPMG firmou parcerias com a IBM para utilizar o Watson AI em ferramentas de auditoria. Por outro lado, a *Pricewaterhouse Coopers* (PCW) e a Deloitte optaram por desenvolver sistemas próprios, voltados para análise de dados, como é o caso do Halo, desenvolvido pela PCW, e dos *softwares* Argus e Optix, elaborados pela Deloitte (KOKINA; DAVENPORT, 2017).

Muitas das soluções de Inteligência Artificial no setor contábil estão, principalmente, voltadas a processos de auditoria, tais como auditorias de estoque, processos de compras e autorizações, contratos, dentre outros (KOKINA; DAVENPORT, 2017; BEHRENS; STEINBART, 1992; LEE; TAJUDEEN, 2020), devido, principalmente, à dificuldade em analisar grandes volumes de documentos. Assim, as ferramentas de Inteligência Artificial conseguem escanear, extrair e elucidar somente as partes que realmente possam suscitar possíveis fraudes, poupando o tempo de contadores e auditores (KOKINA; DAVENPORT, 2017).

Diante disso, os novos produtos, com o uso da Inteligência Artificial, trazem uma diminuição do tempo dispendido pelos contadores em diversas tarefas. Por outro lado, a Inteligência Artificial tem colocado a profissão de contador na 21ª posição em um ranking de 366 profissões que podem ser eliminadas no futuro (ZHANG *et al.*, 2020). Com isso, os profissionais que atuam em contabilidade devem se adaptar a esse novo cenário, adicionando novas habilidades ao currículo, como conhecimentos computacionais mais avançados, inclusive em linguagens de programação, bem como conhecimentos em ciência e análise de dados (ZHANG *et al.*, 2020; KOKINA; DAVENPORT, 2017).

Por um lado, vemos essa necessidade de atualização das habilidades dos profissionais do setor contábil; por outro, a automação, a robotização e a IA trazem inúmeras vantagens. O uso dessas automações e de novas tecnologias traz mais segurança e confiabilidade das informações, por exemplo, tecnologias de *Radio Frequency Identification* (RFID), que automatizam e reduzem drasticamente os erros em operações que envolvem estoque e expedição (ZHANG *et al.*, 2020), o que permite aos profissionais contábeis se

dedicarem a tarefas mais complexas e de tomadas de decisão (LEE; TAJUDEEN, 2020).

2.4 Inteligência Artificial e tributos

Os tributos são necessários aos Estados para que possam concretizar seus deveres junto à sociedade, sendo baseados em legislações fiscais, conforme cada Estado e suas particularidades, definindo sua matriz tributária e os valores cobrados, de acordo com os atributos tributários desejáveis: eficiência, equidade, simplicidade, capacidade contributiva e segurança jurídica (AFONSO; SOARES; CASTRO, 2013).

No Brasil, muitos desses atributos não são totalmente atendidos, por possuímos uma alta carga tributária, além de um emaranhado de leis e de normas (FELICIO; MARTINEZ, 2018) e uma matriz tributária focada no consumo e tributos indiretos que acabam por penalizar quem pode contribuir menos (GASSEN; D'ARAÚJO; PAULINO, 2013). Além disso, ainda trazem um aumento de custos às organizações no cumprimento da legislação tributária e um aumento dos valores dos produtos e serviços (COSTA *et al.*, 2016).

Atualmente, os poucos estudos na área de tributos relacionados à Inteligência Artificial (FAÚNDEZ-UGALDE; MELLADO-SILVA; ALDUNATE-LIZANA, 2020; OLIVEIRA; SANTOS, 2020; HÖGLUND, 2017; CHEN *et al.*, 2011; MABE-MADISA, 2018) estão mais voltados para coibir evasão fiscal dos contribuintes, em função de suas obrigações para com os entes governamentais (países, Estados e Municípios), do que para auxiliar as empresas quanto à classificação, reconhecimento e apuração correta dos tributos.

Desse modo, alguns países já aplicam a Inteligência Artificial para fiscalizar os contribuintes, com o objetivo de identificar e coibir a evasão fiscal e prevenir crimes tributários (FAÚNDEZ-UGALDE; MELLADO-SILVA; ALDUNATE-LIZANA, 2020). Inclusive, o Brasil, com o projeto Harpia, e outros países sul-americanos já investem há algum tempo em ferramentas para auxiliar na identificação de contribuintes que praticam evasão fiscal (FAÚNDEZ-UGALDE; MELLADO-SILVA; ALDUNATE-LIZANA, 2020). Embora existam estudos, ainda são poucas as pesquisas sobre previsão do risco fiscal, dada a relevância do tema (OLIVEIRA; SANTOS, 2020; HÖGLUND, 2017).

Mesmo com a redução da evasão fiscal pelo uso de tecnologias, os prejuízos aos cofres públicos ainda são grandes. Segundo o Instituto Brasileiro de Planejamento e Tributação (IBPT), a sonegação de tributos no Brasil, em 2019, foi de R\$417 bilhões, cuja soma dos atos inflacionários foi de R\$273,1 bilhões (IBPT, 2020). Em função disso, além da redução quanto ao ônus provocado pela evasão, o uso de Inteligência Artificial auxilia na redução dos custos

de fiscalização, uma vez que essa tarefa exige pessoas extremamente qualificadas, tempo e muita dedicação (CHEN *et al.*, 2011; MABE-MADISA, 2018).

Diante desse cenário, alguns estudos abordaram o uso de tecnologias relacionadas à Inteligência Artificial no campo tributário. Oliveira e Santos (2020) aplicaram redes neurais para identificar contribuintes com indícios de práticas de evasão fiscal e, como resultado, obtiveram números que demonstraram como as redes neurais podem auxiliar o fisco na seleção mais acurada de contribuintes propensos a sonegar. A partir do treinamento da rede neural, foram identificados e classificados corretamente 71% dos contribuintes como passíveis de autuação ou não. Por outro lado, em relação a contribuintes que foram autuados, foram alcançados 94% de acertos, demonstrando, assim, a boa performance da rede neural (OLIVEIRA; SANTOS, 2020).

Em Taiwan, Chen *et al.* (2011) e Lin *et al.* (2012) também utilizaram redes neurais para detectar relatórios tributários gerados de forma errônea. Chen *et al.* (2011) observaram empresas do ramo de construção, conseguindo uma performance de acerto de 80%. Lin *et al.* (2012) utilizaram dados de empresas e indivíduos, a fim de detectar a evasão fiscal com maior rapidez e efetividade.

Na Finlândia, Høglund (2017) utilizou algoritmos genéticos para prever padrões de pagamento de impostos. Em seu estudo, solvência, liquidez e período de pagamento foram as principais variáveis que impactaram as empresas na inadimplência fiscal (HÖGLUND, 2017). Por outro lado, Mabe-Madisa (2018) utilizou uma combinação de técnicas, árvores de decisão e o algoritmo *Naive Bayes* para descobrir padrões e classificar os contribuintes quanto à sua atribuição fiscal (MABE-MADISA, 2018).

Em diversos setores das empresas, observa-se um aumento no uso da Inteligência Artificial; entretanto, nas áreas tributária e fiscal, ainda são poucos os estudos, sendo áreas carentes tanto de estudos quanto de criação de ferramentas baseadas em Inteligência Artificial para otimizar o trabalho. A Inteligência Artificial pode auxiliar em diversos pontos ao falarmos de tributos, como na classificação, no cálculo, na conferência das informações e na auditoria e prevenção de evasão fiscal. No Brasil, esse cenário é agravado com inúmeros tributos, quantidade de alíquotas e pelo emaranhado de leis e normas tributárias (COSTA *et al.*, 2016).

2.5 Tributação, ICMS e Nota Fiscal Eletrônica

Os tributos são a principal forma de os Estados financiarem os serviços públicos que prestam à sua população e, assim, concretizarem seus deveres constitucionais, em que

cada Estado, de acordo com a sua constituição e cultura, tem deveres, legislações e tributos distintos (FELICIO; MARTINEZ, 2018).

No Brasil, possuímos uma alta e complexa carga tributária, focada, principalmente, no consumo, que acaba por tornar nossa matriz tributária uma das mais regressivas e injustas do mundo, ao não diferenciar o poder econômico do consumidor, em que os mais pobres acabam pagando o mesmo valor de impostos que os mais ricos (GASSEN; D'ARAÚJO; PAULINO, 2013). Para as empresas, os tributos influenciam diretamente em seu resultado e em seu patrimônio, além de requererem controles e apurações específicas (COSTA, 2014; COSTA *et al.*, 2016).

O ICMS é um tributo focado no consumo, como o próprio nome menciona: Imposto Sobre Circulação de Mercadorias e Serviços, de competência estadual, ou seja, cada Estado define suas regras e atribuições (GASSEN; D'ARAÚJO; PAULINO, 2013). Logo, uma empresa que realiza operações em todos os Estados terá de conhecer as particularidades de cada um dos 26 Estados e do Distrito Federal. Dentro dos Estados, ainda pode existir regimes especiais de tributação, quando o Estado define algum tipo de exceção especial a um ramo de atividade ou até a uma empresa, o que deixa ainda mais complexas as matérias sobre o ICMS (GASSEN; D'ARAÚJO; PAULINO., 2013).

O ICMS e outros tributos, tais como o PIS e a Cofins, incidem nas operações entre as empresas e, por isso, devem ser declarados na Nota Fiscal Eletrônica. A manifestação desses tributos na Nota Fiscal Eletrônica possibilita aos entes fiscais fiscalizar de forma eletrônica os contribuintes e, assim, coibir a evasão de tributos, automatizar os processos e simplificar a fiscalização com a padronização da declaração das informações, por definição de *layouts* e de manuais de preenchimento (RECEITA FEDERAL DO BRASIL, 2022b).

A Nota Fiscal Eletrônica, conhecida como NFE, foi instituída em 20/12/2005, em sua versão 1.0. Foi colocada em uso de forma gradativa aos contribuintes brasileiros a partir de 2006 e implantada, inicialmente, em grandes empresas no Brasil, de forma seletiva e, posteriormente, ao longo dos anos, imposta obrigatoriamente às demais empresas (RECEITA FEDERAL DO BRASIL, 2022b). A versão atual da NFE é a 4.0. De forma cronológica, a versão 1.0 foi implantada em 2006; a 2.0, em 2010; a 3.1, em 2015; e a versão 4.0, em 2018 (RECEITA FEDERAL DO BRASIL, 2022b). A NFE está presente em praticamente todas as operações de saída de produtos e serviços das empresas - itens da nota fiscal - independentemente do objetivo: venda, empréstimo, descarte e diversas outras operações, denominadas nas notas fiscais como Natureza de Operação. Portanto, as Notas

Fiscais Eletrônicas substituem as antigas notas em papel, dos modelos 1, 1-A, 4 e modelo 2 (OLIVEIRA *et al.*, 2021).

De natureza digital, as Notas Fiscais Eletrônicas são arquivos assinados digitalmente pelo emissor da nota fiscal, os quais são transmitidos a servidores dos órgãos fiscalizadores que autorizam ou denegam a operação (OLIVEIRA *et al.*, 2021). Logo, para padronizar os arquivos XML (*Extended Markup Language*) (W3C) das notas fiscais, a Receita Federal do Brasil disponibiliza manuais que especificam os campos e valores a serem informados (PORTAL DA NOTA FISCAL ELETRÔNICA, 2021).

Os arquivos das notas fiscais têm diversas informações: sobre o emissor, destinatário, produtos comercializados e, principalmente, em relação aos tributos vinculados à operação comercial. Os tributos passíveis de serem informados na nota fiscal são: Imposto Sobre Serviços (ISS), de competência municipal; Imposto sobre Circulação de Mercadorias e Serviços (ICMS), de competência estadual; Imposto sobre Produtos Industrializados (IPI), Programa Integração Social (PIS) e Contribuição para o Financiamento da Seguridade Social (Cofins), de competência federal (AFONSO; SOARES; CASTRO, 2013), 2013).

Para que os órgãos fiscalizadores possam fiscalizar a operação comercial, foram criadas diversas tabelas de classificação, que relacionam desde o ramo de atividade do emissor da nota fiscal até o tipo de produto comercializado. Dentre essas classificações, tabelas e siglas, temos, por exemplo: o Código Fiscal de Operações e Prestações (CFOP), Nomenclatura Comum Mercosul (NCM) e Código Nacional Atividades Econômicas (CNAE).

O Código Fiscal de Operações e Prestações (CFOP) determina o tipo de operação realizada na NFE, sendo, de certa forma, uma representação codificada e detalhada da Natureza de Operação da Nota Fiscal. É composto por quatro dígitos: o primeiro representa a origem *versus* destino da operação (estadual, interestadual, exportação e importação), e os três últimos identificam propriamente a operação realizada, se é uma operação de venda, empréstimo, devolução, dentre outras (CONFAZ, 2022b).

A NCM é uma tabela de códigos definidos pelos países membros do Mercosul, cujo objetivo é classificar os produtos conforme sua origem e características, e, assim, padronizar as movimentações desses produtos dentro do bloco comercial (RECEITA FEDERAL DO BRASIL, 2022a).

O CNAE, por sua vez, classifica a empresa de acordo com sua área de atuação, bem como com quais tipos de produtos trabalham e quais serviços prestam a seus clientes (IBGE, 2022). Ligado à empresa, também temos o Código Regime Tributário (CRT), que,

conforme o manual da Nota Fiscal Eletrônica, é definido em três: Simples Nacional, Simples Nacional Excesso de Receita e Regime Normal (RECEITA FEDERAL DO BRASIL, 2022b). Enquanto, normalmente, o Simples Nacional é voltado para micro e pequenas empresas, o Regime Normal é para médias e grandes empresas. Esse tipo de regime influencia diretamente o cálculo de impostos e, principalmente, o ICMS, que tem, inclusive, tabelas de CST específicas para Simples Nacional e Regime Normal (RECEITA FEDERAL DO BRASIL, 2022b).

Existem tabelas para classificar o produto, contribuinte e operação, assim como também há outras que classificam e indicam o cálculo do imposto ou, ainda, além da NCM, categorizam o produto perante o tributo em questão conforme o tipo de movimentação, da Natureza da Operação da NFE. Nesse sentido, temos, principalmente, o Código de Situação Tributária (CST), que define se determinado produto é tributado ou não (PORTAL DA NOTA FISCAL ELETRÔNICA, 2021).

No caso do ICMS, além do CST especificado na NFE, tem-se a Tabela de Modalidade de Base Cálculo do Imposto, que determina como a base de cálculo do imposto foi definida, e a Tabela de Motivos de Desoneração de ICMS, que deve ser utilizada em casos em que o Estado concede desoneração tributária. Para o ICMS ST (ICMS cobrado por substituição tributária), uma forma de cobrança antecipada do ICMS, existem tabelas de Modalidade de Base de Cálculo e Desoneração (PORTAL DA NOTA FISCAL ELETRÔNICA, 2021) e, ainda, o Código de Especificação Substituição Tributária (CEST), que determina a classificação do produto perante o cálculo do ICMS ST (CONFAZ, 2022a).

2.6 Sistemas de Aprendizado de Máquina - Classificadores

Os Sistemas de Classificação têm como principal atributo classificar eventos e dados, a partir de diversas categorias predefinidas; as variáveis de saída, com base nos dados e características desses eventos, e as variáveis de entrada (TAN *et al.*, 2019). Podem ser utilizados em vários campos, como financeiro, marketing e até em campos de engenharia e medicina (ROKACH; MAIMON, 2015). Estão presentes desde a classificação de mensagens de *e-mail*, como SPAM, analisando o assunto e conteúdo da mensagem, até na classificação de galáxias, conforme seu formato (TAN *et al.*, 2019).

Tan *et al.* (2019) definem o processo de classificação como: “a classificação é a tarefa de aprender uma função alvo que mapeia cada conjunto de atributos x para uma das classes predefinidas de rótulos y ”. Esse processo é ilustrado na Figura 1.

Figura 1 – Modelo de Classificação



Fonte: Adaptado de TAN *et al.* (2019).

Existem diversos tipos de algoritmos de Aprendizagem de Máquina utilizados em problemas de classificação, por exemplo: redes neurais, algoritmos genéticos, árvores de decisão e máquina de vetores de suporte (SVM) (TAN ET AL., 2019; ROKACH; MAIMON, 2015). Esses algoritmos são amplamente empregados por pesquisadores na solução de problemas específicos.

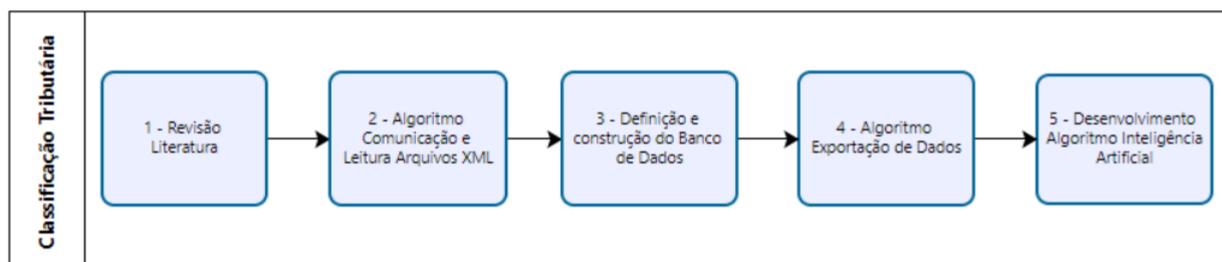
A existência de diversos tipos de algoritmos de Aprendizagem de Máquina, utilizados em sistemas de classificação, decorre da grande pluralidade de problemas, sendo que, de acordo com as especificidades e características do problema, um algoritmo pode proporcionar melhores resultados do que outro (ROKACH; MAIMON, 2015).

Podemos observar essa diversidade de técnicas e algoritmos nos trabalhos que abordaram problemas de classificação relacionados a questões tributárias. Algoritmos baseados em redes neurais foram utilizados por Oliveira e Santos (2020) para identificar contribuintes propícios a sonegarem impostos. Höglund (2017) utilizou algoritmos genéticos para identificar e classificar contribuintes inadimplentes na Finlândia. Por outro lado, Abedin *et al.* (2021), Lin *et al.* (2012), Mabe-Madisa (2018) e Lahann *et al.* (2019) optaram por empregar algoritmos baseados em árvores de decisão, com o objetivo de classificar os contribuintes passíveis de praticar evasão de tributos.

3 METODOLOGIA

Busca-se, neste tópico, delinear os percursos metodológicos utilizados para se chegar ao objetivo deste trabalho. Assim, a execução desta pesquisa foi dividida em cinco etapas, como demonstrado na Figura 2.

Figura 2 – Etapas de Desenvolvimento do Projeto



Fonte: Elaborado pelo autor, 2022.

A primeira etapa consistiu na revisão da literatura e em avaliar como os pesquisadores estão abordando o tema deste estudo, quais os trabalhos e as ferramentas utilizados e, principalmente, quais os resultados obtidos.

Na segunda etapa, foram desenvolvidas a interface de comunicação e a leitura dos arquivos XMLs das Notas Fiscais Eletrônicas, tanto as emitidas pelas empresas participantes quanto as notas fiscais recebidas pelas compras efetuadas. Somente foram utilizadas as notas fiscais que representaram operações de comercialização, relativas a vendas de produtos pelas empresas participantes do projeto, sendo que as compras executadas por elas, resumidamente, foram as vendas efetuadas por seus fornecedores.

Na terceira etapa, foi construído o banco de dados, que tem natureza relacional. Nele, foram criadas tabelas distintas para armazenar cada tipo de informação, a fim de reduzir redundâncias de informações e espaço de armazenamento, otimizar a performance e diminuir o tempo de resposta das consultas aos dados.

Na quarta etapa, foram desenvolvidos o algoritmo e as interfaces de exportação de dados cadastrais e fiscais, relacionados aos documentos fiscais processados e armazenados no banco de dados.

Na quinta etapa, desenvolveu-se o Algoritmo da Inteligência Artificial responsável pela classificação fiscal dos produtos relativos ao ICMS. Nessa etapa, também foram efetuados o treinamento da Inteligência Artificial e os testes sobre seu funcionamento e as classificações realizadas. O algoritmo utilizado pelo robô foi baseado em *Machine Learning*, com o emprego de técnicas de aprendizado supervisionado, baseado em árvores de decisão, uma vez que a variável resposta, que é a classificação tributária, e os parâmetros já são conhecidos (NGUYEN *et al.*, 2019).

Na Parte 2 deste trabalho, foi adicionada uma seção sobre a metodologia específica utilizada para o desenvolvimento de cada produto técnico, trazendo um detalhamento acerca dos métodos empregados em cada produto.

3.1 Materiais

Para o desenvolvimento dos produtos técnicos/tecnológicos, foi utilizada a Linguagem de Programação Java, que é livre e de código aberto. Como ambiente de desenvolvimento, utilizou-se a *Integrated Development Enviroment* (IDE) Netbeans. Ademais, o banco de dados relacional empregado foi o *Firebird*.

Como *framework*, para acelerar o desenvolvimento e manter o foco na lógica de negócios, foi utilizado o *Spring Boot*.

As principais bibliotecas utilizadas no desenvolvimento foram a *Stax* API, para leitura dos arquivos XML das notas fiscais, e a biblioteca Tribuo, da Oracle, para o desenvolvimento do algoritmo de Inteligência Artificial. Essa biblioteca tem modelos (*templates*) de algoritmos de Aprendizado de Máquina prontos, que facilitam e agilizam o desenvolvimento do algoritmo de Aprendizado de Máquina, o qual será responsável por aprender e, depois, prever as classificações fiscais. A biblioteca Tribuo pode ser encontrada no site <https://tribuo.org/>.

A Tribuo foi escolhida por proporcionar mais detalhes de uso, diminuição de possibilidade de erros e por ser nativa ao Java, o que possibilita a interoperabilidade entre o Java e outras linguagens e sistemas operacionais. Mais detalhes e características da biblioteca são descritos no Apêndice E.

A Tabela 1 expõe os equipamentos, *softwares* e versões utilizados no desenvolvimento desta pesquisa.

Tabela 1 – Materiais Utilizados

Item	Especificação	Observações
Hardware	Notebook Acer Nitro 5	- Memória RAM: 24GB - HD: 1 TB SSD - Placa de Vídeo: Geforce GTX1650 - Processador: AMD Ryzen 7 5800H
Sistema Operacional	Windows 11 Pro 21H2	
Linguagem de Programação	JAVA 8	JDK 1.8 – 302
IDE de Desenvolvimento	NetBeans 16	

Framework	Spring Boot 2.6.7
Banco de Dados	Firebird SuperServer 3.0.8.33535
Biblioteca Leitura Arquivos XML	Stax API 1.0-2
Biblioteca Inteligência Artificial	Oracle Tribuo 4.2.0

Fonte: Elaborado pelo autor, 2022.

Foram utilizadas ferramentas e tecnologias gratuitas e de código aberto, como o Java, a IDE *Netbeans*, o banco de dados *Firebird*, o *Framework Spring Boot* e as bibliotecas *Stax* e *Oracle Tribuo*. De forma a manter o foco nos objetivos deste trabalho, utilizamos o *framework Spring Boot* para o *core* da aplicação e as bibliotecas: *Stax* para leitura e processamento de arquivos XML, e a *Oracle Tribuo*, que possui diversos algoritmos de Inteligência Artificial já preparados para serem utilizados.

4 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo principal desenvolver um sistema baseado em Inteligência Artificial com a capacidade de definir a classificação tributária CST ICMS.

Para alcançar esse objetivo, foram desenvolvidos cinco produtos, os quais estão relacionados com os objetivos específicos do trabalho, sendo um bibliográfico e quatro técnicos.

O produto bibliográfico, denominado Produto 1, trata-se de uma revisão de literatura acerca do tema abordado neste trabalho, discorrendo sobre como a Inteligência Artificial está sendo aplicada na área de tributos e como auxiliar os profissionais dessa área na execução de suas atividades. De forma geral, o artigo apresenta as mudanças que a utilização da Inteligência Artificial vem provocando na economia e nas organizações, na automação de tarefas, diminuição de custos e até na criação de novos produtos e serviços. Com este estudo, chegou-se à conclusão da existência de um *gap* de pesquisa, da aplicação de soluções de Inteligência Artificial na área tributária, principalmente para os contribuintes. As soluções e pesquisas existentes buscam coibir e identificar a evasão de tributos, ao invés de auxiliar e prevenir erros por parte dos contribuintes.

O Produto 2 consiste em um algoritmo performático capaz de ler e processar os arquivos fiscais nas notas eletrônicas no menor tempo possível, aproveitando todo o

potencial das configurações de computadores atuais, como a capacidade *multithread*. Esse objetivo foi concluído utilizando-se uma combinação de *threads*, conceitos e tecnologias, como o processamento dos arquivos XMLs por *streaming* e o cache dinâmico de registros para diminuir as consultas ao banco de dados. Esse algoritmo alcançou uma taxa de 165,45 documentos processados por segundo. O produto pode ser aplicado em outros *softwares* que necessitam realizar a leitura e o processamento de muitos documentos, de forma simultânea e concomitante, em que o tempo de processamento e a performance são primordiais.

O terceiro produto, denominado Produto 3, refere-se a uma base de dados que foi construída e populada a partir dos dados obtidos no processamento dos arquivos XMLs. Essa base de dados, além de fomentar o objetivo principal do trabalho, por si só, torna-se uma base de referência e consulta tanto dos dados dos produtos quanto das classificações fiscais. Esse produto pode ser utilizado como um banco de dados para consulta de informações sobre pessoas (clientes e fornecedores), produtos e características tributárias das operações comerciais, aproveitando as pesquisas, cadastros e as próprias operações comerciais realizadas pelas empresas. Essa base de dados também poderá ser utilizada para análises das operações comerciais efetuadas, bem como para estudos e pesquisas acadêmicas e científicas.

O quarto produto, designado Produto 4, consiste em um algoritmo de exportação de dados, o qual utiliza a própria base de dados criada no Produto Técnico 3, que realiza a leitura e análise das informações do banco de dados para produzir arquivos padronizados sobre as operações comerciais realizadas pelas empresas. São gerados arquivos com as informações de produtos, clientes, fornecedores, naturezas de operação e dados tributários referentes ao IPI, PIS, Cofins e ICMS. Esses arquivos podem ser utilizados e aplicados em processos de análise das movimentações tributárias e na troca de *softwares* de gestão nas empresas. Especialmente, as trocas de *softwares* de gestão são operações trabalhosas e morosas. Afinal, todos os cadastros devem ser refeitos no novo *software*. Portanto, esse produto aumenta a assertividade e diminui o trabalho e o tempo, usando as próprias notas fiscais emitidas e recebidas pelas empresas para gerar esses arquivos e dados de transição e para automatizar o processo.

Por fim, o Produto 5 diz respeito à aplicação do algoritmo de Inteligência Artificial na previsão do CST ICMS. Esse algoritmo alcançou 97,2% de precisão, evidenciando ser possível aplicar a Inteligência Artificial em uma área considerada complexa no Brasil, com poucas soluções que auxiliam os profissionais que atuam na área de tributos. Esse produto pode ser aplicado em *softwares* de gestão ou diretamente pelos profissionais

da área tributária para consultar o CST ICMS, conforme a operação comercial a ser realizada pela empresa ao imputar alguns parâmetros. Assim, mesmo que a operação não exista na base de dados, o algoritmo é capaz de informar qual CST utilizar, acertando em 97,2% dos casos. Se, atualmente, profissionais da área realizam a leitura de várias páginas acerca das normas e legislação tributária sobre o ICMS para interpretar e identificar o CST a ser utilizado, ao saber qual o CST, eles já poderão consultar diretamente a parte específica da legislação que aborda o CST sugerido e, assim, confirmarem sua utilização na operação comercial em questão.

Portanto, todos objetivos deste trabalho foram concluídos - os objetivos específicos e o principal. Os produtos 2, 3, 4 e 5 estão disponíveis em um repositório na Internet, conforme detalhado no Apêndice C. Esses produtos foram encapsulados em uma única aplicação, com uma interface gráfica para que os usuários sem conhecimento computacional consigam utilizá-los de forma simples e direta. Com a abordagem adotada, além da disponibilização de uma aplicação com todos os produtos, cada objetivo pôde se tornar um produto técnico distinto. Estes produtos podem ser utilizados isoladamente ou em conjunto para auxiliar os profissionais que atuam na área de tributos e computação em diversas tarefas, como na análise, importação, consulta e exportação de dados, além da principal atividade, associada ao uso da Inteligência Artificial com a capacidade preditiva de qual classificação tributária do CST ICMS a ser utilizada, conforme as características da operação comercial.

Por fim, retoma-se o propósito do trabalho, informando que ele contribui com a área tributária e com as empresas de modo geral, as quais poderão simplificar e automatizar a classificação tributária do CST ICMS, evitando, assim, erros materiais causados pela complexidade da legislação tributária brasileira. O trabalho contribui com a área computacional, com a análise e aplicação de diferentes algoritmos de Inteligência Artificial baseados em árvores de decisão e os resultados alcançados.

As limitações encontradas no desenvolvimento deste trabalho estão relacionadas, principalmente, à base de dados utilizada. Essa base foi limitada a um conjunto de empresas de alguns ramos específicos, e não de todos os ramos existentes, ficando circunscrita a operações específicas, especialmente a de vendas, e ao estado de Minas Gerais. O Produto 5 se limitou apenas à análise da previsão do CST ICMS, não avaliando as demais características do ICMS, como alíquotas e situações tributárias do próprio ICMS e de outros tributos.

Apesar das limitações, o trabalho é consistente com a necessidade da gestão

tributária no Brasil e embrionário para o desenvolvimento de novas aplicações tributárias baseadas em Inteligência Artificial. Para trabalhos futuros, sugere-se popular a base de dados com mais empresas, produtos e movimentações, de Minas Gerais e demais estados. Sugere-se, também, otimizar o algoritmo para definir as alíquotas e as situações tributárias do ICMS, bem como do PIS, da Cofins e do IPI.

REFERÊNCIAS

- BEHRENS, Michael L.; STEINBART, Paul John. Integrating expert systems and artificial intelligence in accounting: a description of the academic program at Memphis State University. **Expert Systems with Applications**, 4 (2), p. 219–223, 1992. DOI: 10.1016/0957-4174(92)90113-7.
- BORGES, Aline F.S.; LAURINDO, Fernando J.B.; SPÍNOLA, Mauro M.; GONÇALVES, Rodrigo F.; MATTOS, Claudia A. The strategic use of artificial intelligence in the digital era: systematic literature review and future research directions. **International Journal of Information Management**, 57, p.102225, 2021. DOI: 10.1016/j.ijinfomgt.2020.102225.
- BRUYN, Arnaud de; VISWANATHAN, Vijay; BEH, Yean Shan; BROCK, Jürgen Kai-Uwe; WANGENHEIM, Florian von. Artificial Intelligence and Marketing: pitfalls and opportunities. **Journal of Interactive Marketing**, 51, p. 91–105, 2020. DOI: 10.1016/j.intmar.2020.04.007.
- CHEN, Jieh-Haur; SU, Mu-Chun; CHEN, Chang-Yi; HSU, Fu-Hau; WU, Chin-Chao. Application of neural networks for detecting erroneous tax reports from construction companies. **Automation in Construction**, 20 (7), p. 935–939, 2011. DOI: 10.1016/j.autcon.2011.03.011.
- CONFAZ. **Código Especificador Substituição Tributária (CEST)**. (2022a). Disponível em: https://www.confaz.fazenda.gov.br/legislacao/convenios/2015/CV146_15. Acesso em: 3 mar. 2022.
- CONFAZ. **Código Fiscal Operações e Prestações (CFOP)**. 2022b. Disponível em: https://www.confaz.fazenda.gov.br/legislacao/ajustes/sinief/cfop_cvsn_70_vigente. Acesso em: 3 mar. 2022.
- COOPER, S. Barry; VAN LEEUWEN, Jan. Computing Machinery and Intelligence. Alan Turing: His Work and Impact: **Elsevier**, p. 551–621, 2013.
- COSTA, Daniel Fonseca. Utilização do orçamento no planejamento tributário de uma pequena empresa. **Revista Brasileira de Contabilidade**, 2014, (206), p. 26–39. Disponível em: <http://rbc.cfc.org.br/index.php/rbc/article/view/1147>.
- COSTA, Daniel Fonseca; CHAIN, Caio Peixoto; CARVALHO, Francisval de Melo; MOREIRA, Bruno César de Melo. O custo financeiro dos tributos sobre consumo nas cadeias de suprimento brasileiras: uma proposta metodológica. **R. Contemp. Contab.**, 13 (29), p. 91, 2016. DOI: 10.5007/2175-8069.2016v13n29p91.

CUNHA, Carlos; SILVEIRA, Heber. Inteligência artificial na formalização de contratos - análise do impacto em uma instituição financeira brasileira de médio porte. **G&T**, 20 (2), p. 256–279, 2020. DOI: 10.20397/2177-6652/2020.v20i2.1872.

DIAS FILHO, José Maria; SANTOS, Luis Paulo Guimarães dos. Administração Tributária Estadual: um modelo para combater focos de sonegação do ICMS. **RN**, 25 (4), p. 35, 2021. DOI: 10.7867/1980-4431.2020v25n4p35-47.

DUAN, Yanqing; EDWARDS, John S.; DWIVEDI, Yogesh K. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. **International Journal of Information Management**, 48, p. 63–71, 2019. DOI: 10.1016/j.ijinfomgt.2019.01.021.

DUBEY, R., GUNASEKARAN, A., CHILDE, S. J., BRYDE, D. J., GIANNAKIS, M., FOROPON, C., ROUBAUD, D., & HAZEN, B. T. Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: a study of manufacturing organisations. **International Journal of Production Economics**, 226, 107599, 2020. DOI: <https://doi.org/10.1016/j.ijpe.2019.107599>.

FAN, Jingtao; FANG, Lu; WU, Jiamin; GUO, Yuchen; DAI, Qionghai. From Brain Science to Artificial Intelligence. **Engineering**, 6 (3), p. 248–252, 2020. DOI: 10.1016/j.eng.2019.11.012.

FAÚNDEZ-UGALDE, Antonio; MELLADO-SILVA, Rafael; ALDUNATE-LIZANA, Eduardo. Use of artificial intelligence by tax administrations: an analysis regarding taxpayers' rights in Latin American countries. **Computer Law & Security Review**, 38, p. 105441, 2020. DOI: 10.1016/j.clsr.2020.105441.

FELICIO, Raphael Maleque; MARTINEZ, Antônio Lopo. Sistema tributário brasileiro: análise da percepção dos operadores do direito tributário à luz dos conceitos de eficiência e justiça fiscal. **Ambiente**, 11 (1), p. 156–181, 2018. DOI: 10.21680/2176-9036.2019v11n1ID13975.

FERFERNIG, Alexander., TERPAN, Erich.; GULA, Bartosz. Knowledge-based recommender technologies for marketing and sales. **International Journal of Pattern Recognition and Artificial Intelligence**, 21 (02), 333–354, 2007. DOI: <https://doi.org/10.1142/S0218001407005417>

GASSEN, Valcir; D'ARAÚJO, Pedro Júlio Sales; PAULINO, Sandra. Tributação sobre Consumo: o esforço em onerar mais quem ganha menos. **Seq.**, 34 (66), 2013. DOI: 10.5007/2177-7055.2013v34n66p213.

GIUDICI, Paolo. Fintech Risk Management: a research challenge for Artificial Intelligence in finance. **Frontiers in artificial intelligence**, 1, p. 1, 2018. DOI: 10.3389/frai.2018.00001.

GREWAL, Dhruv; ROGGEVEEN, Anne L.; Nordfält, Jens. The Future of Retailing. **Journal of Retailing**, 93 (1), p. 1–6, 2017. DOI: 10.1016/j.jretai.2016.12.008.

HAENLEIN, Michael; KAPLAN, Andreas. A Brief History of Artificial Intelligence: on the

past, present, and future of Artificial Intelligence. **California Management Review**, 61 (4), p. 5–14, 2019. DOI: 10.1177/0008125619864925.

HARIRI, Reihaneh H.; FREDERICKS, Erik M.; BOWERS, Kate M. Uncertainty in big data analytics: survey, opportunities, and challenges. **Journal of Big Data**, 6 (1), 2019. DOI: <https://doi.org/10.1186/s40537-019-0206-3>.

HENDRIKSEN, Eldon S.; VAN BREDA, Antonio Zoratto: **Teoria da contabilidade**. São Paulo: Atlas, 1999.

HILDEBRAND, Christian; BERGNER, Anouk. Conversational robo advisors as surrogates of trust: onboarding experience, firm perception, and consumer financial decision making. **J. of the Acad. Mark. Sci.** 2020. DOI: 10.1007/s11747-020-00753-z.

HÖGLUND, Henrik. Tax payment default prediction using genetic algorithm-based variable selection. **Expert Systems with Applications**, 88, p. 368–375, 2017. DOI: 10.1016/j.eswa.2017.07.027.

INSTITUTO BRASILEIRO DE PLANEJAMENTO E TRIBUTAÇÃO. Estudo sobre a Sonegação Fiscal das Empresas Brasileiras. Public. IBPT, 2022. Disponível em: <https://ibpt.com.br/estudo-autos-de-infracao-e-sonegacao-fiscal/>, 2020, Última atualização em 2020, Última verificação em 04/03/2022.

_____. **Quantidade de normas Editadas no Brasil**: 34 anos da Constituição Federal de 1988. Public. Disponível em: <https://ibpt.com.br/quantidade-de-normas-editadas-no-brasil-34-anos-da-constituicao-federal-de-1988/>.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Código Nacional de Atividades Econômicas (CNAE)**, 2022. Disponível em <https://concla.ibge.gov.br/classificacoes/por-tema/atividades-economicas>.

KANELLOU, Alexandra; SPATHIS, Charalambos. Accounting benefits and satisfaction in an ERP environment. **International Journal of Accounting Information Systems**, 14 (3), p. 209–234, 2013. DOI: 10.1016/j.accinf.2012.12.002.

KAPLAN, Andreas; HAENLEIN, Michael. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. **Business Horizons**, 62 (1), p. 15–25., 2019. DOI: 10.1016/j.bushor.2018.08.004.

KOKINA, Julia; DAVENPORT, Thomas H. The Emergence of Artificial Intelligence: How Automation is Changing Auditing. **Journal of Emerging Technologies in Accounting**, 14 (1), p. 115–122, 2017. DOI: 10.2308/jeta-51730.

LEE, Cheah Saw; TAJUDEEN, Farzana Parveen. Usage and Impact of Artificial Intelligence on Accounting: 213 Evidence from Malaysian Organisations. **AJBA**, 13 (1), p. 213–240, 2020. DOI: 10.22452/ajba.vol13no1.8.

LEE, Jay; KAO, Hung-An; YANG, Shanhu. Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment. **Procedia CIRP**, 16, p. 3–8, 2014. DOI: 10.1016/j.procir.2014.02.001.

LIN, Sin-Jin; HSU, Ming-Fu. Incorporated risk metrics and hybrid AI techniques for risk management. **Neural Comput & Applic**, 28 (11), p. 3477–3489, 2017. DOI: 10.1007/s00521-016-2253-4.

MA, Liye; SUN, Baohong. Machine learning and AI in marketing – connecting computing power to human insights. **International Journal of Research in Marketing** 37 (3), p. 481–504, 2020. DOI: 10.1016/j.ijresmar.2020.04.005.

MABE-MADISA, G. V: A Decision Tree and Naïve Bayes algorithm for income tax prediction. **African Journal of Science, Technology, Innovation and Development** 10 (4), p. 401–409, 2020. DOI: 10.1080/20421338.2018.1466440.

MELNYCHENKO, Oleksandr. Is Artificial Intelligence Ready to Assess an Enterprise's Financial Security? **JRFM**, 13 (9), p. 191. DOI: 10.3390/jrfm13090191.

MESERVY, Rayman D.; DENNA, Eric L.; HANSEN, James V. Application of artificial intelligence to accounting, tax, and audit services: Research at Brigham Young University. **Expert Systems with Applications**, 4 (2), p. 213–2181992. DOI: 10.1016/0957-4174(92)90112-6.

MORGENSTERN, Leora; MCILRAITH, Sheila A. John McCarthy's legacy. **Artificial Intelligence**, 175 (1), p. 1–24, 2011, 2011. DOI: 10.1016/j.artint.2010.11.003.

MUTHUKRISHNAN, Nikesh; MALEKI, Farhad; OVENS, Katie; REINHOLD, Caroline; FORGHANI, Behzad; FORGHANI, Reza. Brief History of Artificial Intelligence. **Neuroimaging clinics of North America**, 30 (4), p. 393–399, 2011. DOI: 10.1016/j.nic.2020.07.004.

NASCIMENTO, Alexandre Moreira; MELO, Vinicius Veloso; QUEIROZ, Anna Carolina Muller; BRASHEAR-Alejandro, Thomas; Meirelles, Fernando de Souza: Inteligência Artificial Aplicada a Pequenas Empresas: o uso da engenharia automática de recursos e do aprendizado de máquina para um planejamento mais preciso. **Revista de Contabilidade e Organizações**, 2020.

NGUYEN, Giang; DLUGOLINSKY, Stefan; BOBÁK, Martin; TRAN, Viet; GARCÍA, Álvaro López; HEREDIA, Ignacio; MALÍK, Peter; HLUCHÝ, Ladislav. Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. **Artif Intell Rev.**, 52 (1), p. 77–124, 2019. DOI: 10.1007/s10462-018-09679-z.

OLIVEIRA, Francisco Nobre de; SANTOS, Luis Paulo Guimarães dos. Estratégias para Combater a Sonegação Fiscal: um modelo para o ICMS baseado em redes neurais artificiais. **RGFC**, 10 (1), p. 42–64, 2020. DOI: 10.18028/rgfc.v10i1.7474.

OLIVEIRA, Lana Cristina; TEODORO, Pítias; VASCONCELLOS, Yury; BONFIM, Mariana Pereira. Nota Fiscal Eletrônica (NF-E): benefícios propagados pelos responsáveis pelo projeto versus benefícios percebidos por profissionais de contabilidade. **RACEF**, 12 (2), 2021. DOI: 10.13059/racef.v12i2.740.

PAWŁOWSKI, Mieczysław. Machine Learning Based Product Classification for

eCommerce. **Journal of Computer Information Systems**, p. 1–10, 2021. DOI: 10.1080/08874417.2021.1910880.

PORTAL DA NOTA FISCAL ELETRÔNICA. **Manual de Orientação ao Contribuinte - MOC - versão 7.0 - NF-e e NFC-e**. 2020. Disponível em: <https://www.nfe.fazenda.gov.br/portal/exibirArquivo.aspx?conteudo=LrBx7WT9PuA=>. Acesso em: 3 mar. 2022.

RECEITA FEDERAL DO BRASIL. NCM - **Nomenclatura Comum Mercosul**. 2022a. Disponível em: <https://www.gov.br/receitafederal/pt-br/assuntos/aduana-e-comercio-exterior/classificacao-fiscal-de-mercadorias/ncm>. Acesso em: 3 mar. 2022.

RECEITA FEDERAL DO BRASIL: **Portal da Nota Fiscal Eletrônica (NFe)**. 2022b. Disponível em: <https://www.nfe.fazenda.gov.br/>. Acesso em: 3 mar. 2022.

RECEITA FEDERAL DO BRASIL. **Sistema Escrituração Pública Digital (SPED)**, 2022c. Disponível em: <http://sped.rfb.gov.br/>. Acesso em: 3 mar. 2022.

AFONSO, Jose Roberto Rodrigues; SOARES, Julia Morais; CASTRO, Kleber Pacheco de. **Avaliação da estrutura e do desempenho do sistema tributário Brasileiro**: livro branco da tributação Brasileira. Banco Interamericano de Desenvolvimento, 2013. Disponível em: <https://publications.iadb.org/pt/avaliacao-da-estrutura-e-do-desempenho-do-sistema-tributario-brasileiro-livro-branco-da-tributacao>.

ROKACH, Lior; MAIMON, Oded. Data mining with decision trees. Theory and applications. 2nd ed. Hackensack, **New Jersey**: World Scientific (Series in machine perception and artificial intelligence, vol. 81, 2015.

TAN, Pang-Ning; STEINBACH, Michael; KARPATNE, Anuj; KUMAR, Vipin. **Introduction to data mining. Second edition**. NY: Pearsonm 2019.

ZHANG, Yingying; XIONG, Feng; XIE, Yi; FAN, Xuan; GU, Haifeng. The Impact of Artificial Intelligence and Blockchain on the Accounting Profession. **IEEE**, 2020. Access 8, p. 110461–110477. DOI: 10.1109/ACCESS.2020.3000505.

ZHENG, Pai; WANG, Honghui; SANG, Zhiqian; ZHONG, Ray Y.; LIU, Yongkui; LIU, Chao; MUBAROK, Khamdi; YU, Shiqiang; XU, Xun. Smart manufacturing systems for Industry 4.0: conceptual framework, scenarios and future perspectives. **Front. Mech. Eng.**, 13 (2), p. 137–150, 2018. DOI: 10.1007/s11465-018-0499-5.

PRODUTOS BIBLIOGRÁFICOS E TÉCNICOS

PRODUTO 1 (bibliográfico): A Influência da Inteligência Artificial na Contabilidade e na Tributação das Organizações: uma revisão de literatura

A Influência da Inteligência Artificial na Contabilidade e na Tributação das Organizações: uma revisão de literatura¹

DENIS RIBEIRO DA SILVA

Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais

DANIEL FONSECA DA COSTA

Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais

ALEXANDRE PIMENTA

Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais

Resumo

A Inteligência Artificial (IA) está cada vez presente no cotidiano, no trabalho, nas organizações e em sistemas inteligentes e integrados. Neste estudo, pretende-se abordar a questão de como as tecnologias baseadas em Inteligência Artificial estão sendo abordadas nas pesquisas científicas e identificar como elas estão sendo aplicadas no ambiente organizacional e sua influência nas áreas tributária e contábil. Para isso, o trabalho tem como objetivo analisar como as organizações estão utilizando técnicas de Inteligência Artificial em seus negócios, com especial enfoque nos setores Contábil e Fiscal. Diante disso, será realizada uma revisão de literatura narrativa, de forma a estudar e apresentar os trabalhos de maior relevância, que demonstram como as organizações estão aplicando a Inteligência Artificial em seus negócios e na sua gestão. A abordagem da revisão se dará em três pontos específicos: a) aplicação da inteligência artificial nas organizações; b) aplicação da inteligência artificial nas atividades de contabilidade; e c) aplicação da contabilidade nas atividades fiscais e de tributação. Identificamos o uso de soluções especialistas pelas empresas, atribuídas especialmente em resoluções de problemas ligados aos objetivos fim das empresas, de seus negócios e em processos de auditoria. Encontramos lacunas de pesquisa quanto ao uso da Inteligência Artificial nos setores contábil e de tributos. Quanto ao contábil, existem diversos trabalhos e pesquisas vinculados a empresas onde o negócio da empresa é justamente soluções contábeis, onde tais soluções estão conectadas a automação de seus negócios e não propriamente em analisar os resultados contábeis e assim encontrar padrões e sugerir melhorias na gestão do negócio de seus clientes. Relativo ao setor de tributos, não encontramos trabalhos e pesquisas com uso de Inteligência Artificial para auxiliarem o trabalho das empresas. Os trabalhos encontrados estão relacionados ao combate de evasão de tributos, voltados principalmente a organizações governamentais.

Palavras-chave: Inteligência Artificial; Tomadas de Decisão; Contabilidade; Tributação; Organizações.

¹ Versão inicial apresentada no *USP International Conference in Accounting 2022*

1. Introdução

A Inteligência Artificial (IA) está cada vez mais presente nas vidas das pessoas e nas organizações, mesmo que indiretamente. Se antes os indivíduos eram restritos ao uso de computadores nada portáteis, hoje estão rodeados de equipamentos, sensores e sistemas cada vez mais inteligentes e conectados (Hariri et al., 2019). O advento da internet e as conexões levam a números exponenciais de geração de dados. Com isso, a International Data Corporation (IDC) estimou que a quantidade de dados gerados dobraria a cada dois anos. Entretanto, 90% de todos os dados gerados pela Internet, foi gerado em apenas dois anos (Hariri et al., 2019).

Este volume de dados e as informações geradas pelas conexões na internet são conhecidos por Big Data e tem como características: grande volume, baixa veracidade, alta velocidade, alta variedade e alto valor para as empresas (Hariri et al., 2019). É pela análise de grandes volumes de dados (Big Data Analytics), por meio de técnicas de Mineração de dados (Data Mining), que a Inteligência Artificial e suas subáreas tomam notoriedade, afinal o uso de ferramentas e algoritmos tradicionais são ineficientes ou simplesmente não funcionam para o gerenciamento e extração de informações dessa massa de dados (Dubey et al., 2020). Por outro lado, tem-se, também, um salto na performance e na diminuição de custos relativos aos computadores e hardware. Isso ocorre pelo aumento das tecnologias dos equipamentos, inclusive pelo advento da criação de hardwares específicos para aplicações baseadas em Inteligência Artificial (Ma & Sun, 2020; Nguyen et al., 2019), o que tem provocado uma popularização do uso dessa tecnologia.

A conjunção de fatores relacionados ao uso da Inteligência Artificial, tal como a disponibilidade de dados e de recursos computacionais, está transformando diversas áreas de pesquisa pela possibilidade de se analisar grandes bases de dados, levando a grandes avanços tecnológicos (Nguyen et al., 2019). Com isso, a Inteligência Artificial tem sido utilizada para processar, interpretar, correlacionar, realizar previsões e avaliar o comportamento das pessoas e do mercado a fim de criar produtos, serviços e aumentar a eficiência operacional, evitando erros, objetivando resultados e acelerando o processo de tomada de decisões (Duan et al., 2019).

Assim, as organizações têm usado a Inteligência Artificial de diversas formas para cumprir com seus objetivos fins e com suas obrigações. Nesse sentido, as empresas comerciais têm utilizado a Inteligência Artificial para otimizar as vendas e melhorar a divulgação dos produtos (Ferfernig et al., 2007; Grewal et al., 2017; Ma & Sun, 2020), as

empresas do setor financeiro estão utilizando chatbots (Hildebrand & Bergner, 2020) e ferramentas para a captação e análise do perfil de clientes, para a criação de portfólios de investimentos e para a análise de riscos e de scores de seus clientes (Cunha & Silveira, 2020; Giudici, 2018), as indústrias utilizam a Inteligência Artificial em sistemas de monitoramento e em sensores para diminuir perdas e aumentar a eficiência nos processos produtivos (X. Zhang & Ming, 2021; Zheng et al., 2018), por sua vez, os governos têm aplicado a Inteligência Artificial no combate à evasão fiscal (Faúndez-Ugalde et al., 2020) e as empresas de contabilidade na automação de suas tarefas (Y. Zhang et al., 2020).

Embora haja um volume considerável de trabalhos que abordam sobre o uso da Inteligência Artificial, há de se ressaltar que as pesquisas têm percorrido caminhos diversos, em função de suas várias aplicações (Borges et al., 2021). Logo, é importante entender como a Inteligência Artificial poderá auxiliar as empresas em suas tarefas cotidianas, especialmente as áreas de contabilidade e de tributação. Com isso, torna-se importante a necessidade de se realizar uma revisão de literatura na busca de consolidar os principais achados e as principais contribuições da Inteligência Artificial para a gestão dos serviços de contabilidade e de apuração dos tributos nas organizações.

A contabilidade e a tributação são atividades importantes nas Organizações. A contabilidade tem como objetivo gerar informações econômico-financeiras para a tomada de decisões (Hendriksen et al., 1999) e a atividade de tributação, além de influenciar no resultado e no patrimônio, requer controles e apurações específicas (Costa, 2014; Costa et al., 2016).

Com isso, o objetivo desta pesquisa é analisar e discutir como a Inteligência Artificial tem sido utilizada na contabilidade e na tributação das organizações. De forma específica, o estudo busca discutir como a inteligência artificial tem sido aplicada e tem influenciado à gestão contábil e fiscal nas organizações, bem como evidenciar uma agenda de pesquisa sobre o tema. Esta pesquisa é importante e se justifica por buscar compreender a relação da Inteligência Artificial com as organizações, entender o seu uso pelas organizações, conhecer as aplicações nas áreas de contabilidade e de tributação e, principalmente, levantar as lacunas de pesquisa que possam ser suprimidas para melhorar os processos contábeis e fiscais nas empresas.

Ressalta-se, que trabalhos de revisão de literatura que abordem esse tema, são escassos. No entanto, é possível observar os realizados por Behrens e Steinbart (1992), Zhang et al. (2020), Kokina e Davenport (2017), Lee e Tajudeen (2020) no setor contábil que discutem a aplicação da Inteligência Artificial em automatizar processos contábeis;

Faúndez-Ugalde et al. (2020), Oliveira e Santos (2020), Höglund (2017), Chen et al. (2011) e Mabe-Madisa (2018) no setor de tributos, que discutem o uso da Inteligência Artificial para coibir evasão de tributos. Diante disso, o presente estudo difere dos demais pelo fato de discutir, por meio de uma revisão de literatura, a aplicação da Inteligência Artificial nos setores de contabilidade e tributação das organizações.

Para atingir o objetivo proposto, utilizou-se uma revisão sistemática narrativa em relação aos trabalhos disponíveis sobre o tema. A revisão foi realizada na base de dados da Coleção Principal (Clarivate) da Web of Science (WOS), com termos compatíveis com o objetivo desta pesquisa. A opção pela pesquisa sistemática narrativa se deve pela especificidade do assunto e pelo pequeno número de estudos encontrados (Donthu et al., 2021) sobre a ligação entre a Inteligência Artificial e a gestão das organizações, especialmente em relação à contabilidade e à tributação.

A revisão sistemática narrativa busca sintetizar os trabalhos que abordam Inteligência Artificial e sua influência na contabilidade e na tributação das Organizações. Sendo assim, a abordagem da revisão se dará em três pontos específicos: a) aplicação da inteligência artificial nas organizações; b) aplicação da inteligência artificial nas atividades de contabilidade; e c) aplicação da contabilidade nas atividades fiscais e de tributação.

O trabalho contribui trazendo conceitos e informações sobre a Inteligência Artificial, sua ascensão nos últimos anos, sua utilização pelas organizações em suas atividades e negócios e especialmente como as organizações estão utilizando a Inteligência Artificial nas atividades de contabilidade, tributária e no apoio a tomadas de decisão. Além disso, o estudo traz uma agenda de pesquisa relacionando a inteligência artificial com a contabilidade e tributação das organizações.

Para isso, o trabalho foi dividido, além desta introdução, em mais seis seções para contextualizar o uso da Inteligência Artificial pelas organizações. A segunda seção traz uma introdução conceitual sobre a Inteligência Artificial. A terceira seção demonstra como algumas organizações estão utilizando a Inteligência Artificial em seus negócios, apresentando a amplitude e a generalização do uso da Inteligência Artificial pelas organizações. A quarta seção avalia o uso da Inteligência Artificial nas atividades de contabilidade. A quinta seção apresenta o uso da Inteligência Artificial nas atividades de gestão fiscal e tributária. A sexta seção faz uma discussão sobre os achados e apresenta uma proposta de agenda de pesquisas e, por fim, a sétima seção apresenta as considerações finais e as limitações do trabalho.

2. Inteligência Artificial

A Inteligência Artificial auxilia na simulação e na criação de soluções para problemas em geral, de maneira mais eficiente e precisa. Assim, a IA ajuda os indivíduos a solucionarem problemas em diversas áreas e setores e em complexos fluxos de trabalho (Muthukrishnan et al., 2020). A Inteligência Artificial pode ser definida como o comportamento das máquinas em aprender, a partir de experiências, e se adaptar, conforme o ambiente, para simular a inteligência humana em tomadas de decisão e em processos racionais (Borges et al., 2021; Fan et al., 2020; Kaplan & Haenlein, 2019; Muthukrishnan et al., 2020). Sua história começa com Alan Turing, cujo teste proposto por ele parte da premissa de que se uma pessoa ao interrogar remotamente um computador não conseguir dizer se é um computador ou outra pessoa, o sistema é descrito como inteligente (Cooper & van Leeuwen, 2013).

Ao longo da história e da construção do conceito da Inteligência Artificial, pode-se compreender a sua ligação com a Inteligência e com o Cérebro Humano. John McCarthy definiu, em 1955, a Inteligência Artificial como “Todos os aspectos da aprendizagem — ou qualquer outra característica da inteligência — podem, em princípio, serem descritos tão precisamente que uma máquina será capaz de simulá-los” (Morgenstern & Mcilraith, 2011). Muito dos algoritmos e tecnologias relacionadas a Inteligência Artificial surgiram diretamente de estudos ligados ao funcionamento do cérebro humano, onde podemos destacar as Redes Neurais e seus subtipos, inspiradas nas conexões e as sinapses entre os neurônios (Fan et al., 2020). Por meio de redes neurais, mais precisamente Redes Neurais Profundas (Deep Learning), é possível aprender novas tarefas e preservar conhecimento adquirido da experiência de antigas tarefas, simulando uma das principais capacidades de nossa inteligência, o aprendizado contínuo (Fan et al., 2020).

A Inteligência Artificial pode ser classificada, quanto a aprendizagem em três tipos: supervisionada, não supervisionada e por reforço (Ma & Sun, 2020; Nguyen et al., 2019). Os autores dizem que, de acordo com as características dos dados a serem analisados, um ou mais tipos de algoritmos poderão ser indicados e utilizados.

Para Nguyen et al. (2019) e Bruyn et al. (2020) a aprendizagem supervisionada é utilizada onde os dados são conhecidos e rotulados e a aprendizagem não supervisionada é indicada para situações em que os dados não são conhecidos ou rotulados, com a finalidade de buscar padrões e comportamentos desconhecidos. Por outro lado, a aprendizagem por reforço define como a Inteligência Artificial se relaciona com o ambiente e aprende com ele,

de forma a selecionar as ações apropriadas conforme a situação e maximizar as respostas. Basicamente, os algoritmos aprendem com tentativa e erro até conseguir as respostas apropriadas e, assim, a Inteligência Artificial aprende e evolui, tornando suas respostas cada vez mais precisas (Bruyn et al., 2020; Nguyen et al., 2019).

Além disso, a Inteligência Artificial pode ser categorizada em três grupos: Analítica (Analytical AI), Inspiração Humana (Human-Inspired AI) e Humanizada (Humanized AI). A Inteligência Artificial Analítica possui apenas características cognitivas e, de acordo com as experiências passadas, conseguem informar sobre decisões futuras. Para esse tipo de Inteligência Artificial, as soluções são mais especializadas e focadas em questões e problemas mais específicos. É nesta classificação que se encontram a maioria dos sistemas de Inteligência Artificial atualmente. Em seguida, a Inteligência Artificial Inspiração Humana é definida como sistemas cognitivos que conseguem entender emoções humanas, tais como expressões de surpresa, tristeza, felicidade, entre outras. Por fim, a Inteligência Artificial Humanizada são sistemas completos, cognitivos, interpretativos, sociais, que conseguem ter consciência e aprendizado próprio e que ainda não existem (Kaplan; Haenlein, 2019).

Frente a isso, as aplicações que utilizam Inteligência Artificial são cada vez mais frequentes em todos os ramos da economia, sendo aplicada nas esferas públicas, privadas, corporativas e sociais. Essa utilização tem se dado em maior escala devido, principalmente, a disponibilidade e quantidade de dados (Big data), os quais começam a fazer parte do cotidiano, influenciando como os consumidores realizam suas compras e como empresas tomam suas decisões e interagem com os seus acionistas, (Haenlein & Kaplan, 2019). Na atualidade, robôs têm sido amplamente utilizados para conversar e interagir com as pessoas, principalmente em empresas fundadas com premissas tecnológicas, como, por exemplo, as startups. Nesse contexto, as startups financeiras e as fintechs usam robôs inclusive para captar novos clientes, definir perfil dos investidores e assim escolher as melhores carteiras de investimentos. Neste ponto, quanto mais interativo forem os robôs, maior será a confiança dos clientes nos produtos destas empresas (Hildebrand & Bergner, 2020).

Com estes avanços, pelo volume de dados gerados e disponíveis, aumento da capacidade computacional, somados a principal característica multiuso da Inteligência Artificial, em se adaptar e a resolver diversos e variados problemas, que soluções baseadas em algoritmos de Inteligência Artificial se destacam (Borges et al., 2021).

Além disso, a Inteligência Artificial tem sido utilizada para solucionar problemas e melhorar processos em diversas áreas, em evasão de tributos (Chen et al., 2011; Faúndez-

Ugalde et al., 2020; Höglund, 2017; Mabe-Madisa, 2018; Oliveira & Santos, 2020), em processos contábeis e de auditoria (Behrens & Steinbart, 1992; Kokina & Davenport, 2017; C. S. Lee & Tajudeen, 2020; Meservy et al., 1992) em otimização de diversos processos como classificar e categorizar melhor os produtos (Pawłowski, 2021), aumentar as vendas e melhorar a experiência dos clientes em suas compras (Grewal et al., 2017) e otimizar processos produtivos (J. Lee et al., 2014; Zheng et al., 2018) são alguns dos exemplos onde a Inteligência Artificial está otimizando processos nas organizações.

3. Aplicação da Inteligência Artificial nas Organizações

Nas organizações, a Inteligência Artificial é utilizada principalmente no auxílio de tomadas de decisão, ao se tornar em uma ferramenta estratégica em diferenciar as organizações em um mercado cada vez mais competitivo (Borges et al., 2021). As soluções mais utilizadas nas organizações são aplicadas de forma especializada e para resolver problemas mais específicos (Kaplan & Haenlein, 2019).

Em várias situações, os modelos de solução, de auxílio a tomadas de decisão propostos por soluções de Inteligência Artificial são inclusive melhores e mais eficientes do que o realizado por humanos (Borges et al., 2021), principalmente em avaliar grande volume de dados, para encontrar padrões, e sugerir soluções ou percepções de comportamento (Duan et al., 2019).

Identificar comportamentos, perfis de consumidores é um dos principais objetivos de algoritmos de Inteligência Artificial utilizados em soluções de e-commerce, seja em uma melhor definição e categorização de produtos (Pawłowski, 2021), avaliar as melhores estratégias de se chegar ao público alvo, principalmente, com características e novidades que devem chamar a atenção dos clientes, (Grewal et al., 2017). O objetivo principal é trazer uma experiência positiva aos clientes e assim agregar um valor maior aos produtos (Ma & Sun, 2020).

No mundo virtual, quanto mais assertivo for a publicidade e os produtos, maior será o interesse e a chance de uma nova venda ser concretizada (Grewal et al., 2017). Neste aspecto, as empresas passaram a investir principalmente em algoritmos de Inteligência Artificial que aprendem as preferências do cliente com base nas suas pesquisas e comportamento de consumo, agrupando os clientes e criando modelos de aprendizado conforme as características destes (Borges et al., 2021). Com um planejamento melhor em vendas, as empresas conseguem se preparar melhor para atender o fluxo de pedidos, principalmente no caso de pequenas empresas que geralmente possuem recursos limitados,

(Nascimento et al., 2020).

Com o fluxo de vendas melhor definido, é possível também otimizar o setor de produção, planejar melhor as compras de insumos e matéria prima, planejar recursos humanos e maquinário de forma a atender em tempo e não atrasar as entregas. É neste setor que a Inteligência Artificial pode iniciar uma próxima revolução e a criação de uma nova era, que será a união dos mundos físicos e virtuais através de sistemas cyber-physical, um marco da revolução industrial 4.0. O uso de sistemas inteligentes em diversas subáreas da manufatura irá acelerar as tomadas de decisões, diminuir falhas e trazer economia as indústrias. O aumento de uso de sensores nas máquinas e a comunicação online com sistemas de monitoramento agilizam intervenções quando necessárias, evitam paradas repentinas e aumentam a qualidade dos produtos (J. Lee et al., 2014; Zheng et al., 2018).

A integração destes mundos e a descoberta de novas tecnologias proporcionam uma maior percepção aos consumidores quanto a inovação, qualidade e variedade de produtos e diminui o lead time de fabricação e aumenta a velocidade de entrega aos clientes, diminuindo prejuízos com perdas de vendas e de produtos (J. Lee et al., 2014; Nascimento et al., 2020). Nascimento et al. (2020) demonstraram como uso de técnicas de Inteligência Artificial podem auxiliar na previsão de vendas, com assertividade de 81% em pequenas empresas. Se há uma previsão de vendas assertiva, a produção quando aplicável e o estoque podem ser melhores previstos, resultando em menos perdas e mais lucratividade as empresas.

No setor de finanças, observa-se uma preocupação maior, relativa às análises de riscos de insolvência e uma preocupação com a segurança financeira das empresas, (Lin & Hsu, 2017; Melnychenko, 2020) provavelmente devido ao maior esforço de investidores em avaliar a saúde financeira das empresas.

Em finanças, a maioria dos estudos e trabalhos com o uso da Inteligência Artificial está concentrado no próprio ramo financeiro. Ou seja, empresas onde o negócio é finanças, principalmente as fintechs, que estão revolucionando o mercado financeiro, com chatbots cada vez mais interativos (Hildebrand & Bergner, 2020). Estas ferramentas se provam eficazes para a captação e análise do perfil de clientes, criação de portfólios de investimentos, análise de riscos e scores (Cunha & Silveira, 2020; Giudici, 2018). Cunha e Silveira (2020) inclusive avaliaram a implantação de um software que possui recursos de Inteligência Artificial, para avaliação dos contratos de uma instituição financeira brasileira. Houve redução na avaliação de contratos de cinco para zero dias e uma redução de 19% de custos (Cunha & Silveira, 2020).

4. Aplicação da Inteligência Artificial nas atividades de Contabilidade

A contabilidade estuda o patrimônio e suas variações, tem como objetivo gerar informações financeiras, econômicas e patrimoniais para a tomada de decisões. Nesse sentido, o sistema de informação contábil registra a entrada de dados e informações internas e externas, processa os dados e realiza a saída, por meio de informações úteis para os diversos usuários da contabilidade (Hendriksen et al., 1999).

Portanto, a tecnologia introduziu diversas mudanças no setor contábil, gerando impacto no trabalho dos profissionais do setor e no processamento das informações. Sistemas integrados, tais como o ERP (Enterprise Resource Planning), trouxeram para a área de contabilidade inúmeras vantagens, como a disponibilidade de dados de forma mais rápida e eficiente, redução de tempo na apresentação de resultados, aumento na eficiência e na qualidade quanto a tomada de decisões, auditorias e de relatórios e ainda na própria redução do número de contadores nas empresas (Kanellou & Spathis, 2013).

Por outro lado, os robôs diminuem trabalhos repetitivos e humanamente demorados, liberando os profissionais de contabilidade para tarefas que requerem mais mensuração, ciência, análise de dados e tomadas de decisão (Y. Zhang et al., 2020). Com a introdução da Inteligência Artificial, as empresas podem aumentar ainda mais sua performance e se manterem cada vez mais competitivas e sustentáveis (C. S. Lee & Tajudeen, 2020).

A utilização da Inteligência Artificial já é pautada há vários anos no setor Contábil. Meservy et al (1992) já abordava o uso em softwares desenvolvidos na Universidade Brigham Young University em Utah nos Estados Unidos. Estes softwares atuavam no auxílio de tomadas de decisão, tributos e principalmente auditoria (Meservy et al., 1992).

Lee e Tajudeen (2020) abordaram o uso de tecnologias baseadas em Inteligência Artificial em empresas na Malásia. As empresas analisadas já utilizam soluções para automatizar processos, principalmente contas a pagar, com automações de leitura, armazenamento e cadastro de invoices, auditoria de processos e liberações, e auditoria quanto as ações executadas pelos usuários. Como principais resultados, concluíram: aumento da eficiência, aumento da produtividade, aumento da qualidade dos serviços e produtos aos seus clientes, maior flexibilidade aos colaboradores, melhoria de processos e diminuição de tempo nas tarefas executadas pelos colaboradores.

As principais empresas do setor contábil, conhecidas como The Big Four, também já investem em tecnologias de robotização de tarefas com o uso de IA, que são capazes de realizar reconhecimento automático de dados, escriturar documentos e gerar relatórios

financeiros. Estes sistemas permitem que outros profissionais, sem conhecimento contábil, realizem tarefas que antes, somente contadores podiam realizar (Y. Zhang et al., 2020). Estas empresas possuem estratégias diferentes, entre criar suas próprias ferramentas ou ainda realizar parcerias com outras empresas. A exemplo, a KPMG realizou parcerias com a IBM, para utilizar o Watson AI em ferramentas de auditoria. Por outro lado, a Pricewaterhouse Coopers (PCW) e a Deloitte optaram em desenvolver sistemas próprios, voltados para análise de dados, como é o caso do Halo, desenvolvido pela PCW e dos softwares Argus e Optix, elaborados pela Deloitte (Kokina & Davenport, 2017).

Muitas das soluções de Inteligência Artificial no setor contábil estão principalmente voltadas a processos de auditoria, tais como auditorias de Estoque, Processo de Compras e autorizações, Contratos, dentre outros (Behrens & Steinbart, 1992; Kokina & Davenport, 2017; C. S. Lee & Tajudeen, 2020), devido, principalmente, a dificuldade em analisar grandes volumes de documentos. Assim, as ferramentas de Inteligência Artificial conseguem escanear, extrair e elucidar somente as partes que realmente possam suscitar possíveis fraudes, poupando o tempo de contadores e auditores (Kokina & Davenport, 2017).

Diante disso, os novos produtos, com o uso da Inteligência Artificial, trazem uma diminuição do tempo dispendido pelos contadores em diversas tarefas. Por outro lado, a Inteligência Artificial tem colocado a profissão de Contador na vigésima primeira posição, em um ranking de 366 profissões, que podem ser eliminadas no futuro (Y. Zhang et al., 2020). Com isso, os profissionais que atuam em contabilidade devem se adaptar a este novo cenário, adicionando novas habilidades ao currículo, como conhecimentos computacionais mais avançados, inclusive em linguagens de programação, bem como conhecimentos em ciência e análise de dados (Kokina & Davenport, 2017; Y. Zhang et al., 2020).

Se por um lado vemos esta necessidade de atualização das habilidades dos profissionais do setor contábil, por outro a automação, a robotização e a IA trazem inúmeras vantagens. O uso destas automações e de novas tecnologias trazem mais segurança e confiabilidade das informações, como por exemplo tecnologias de RFID (Radio Frequency Identification) que automatizam e reduzem drasticamente os erros, em operações que envolvem estoque e expedição (Y. Zhang et al., 2020), permitindo aos profissionais contábeis em se dedicar a tarefas mais complexas e de tomadas de decisão (C. S. Lee & Tajudeen, 2020).

5. Aplicação da Inteligência Artificial nas Atividades Fiscais e de Tributação

Os Tributos são necessários aos Estados para que possam concretizar seus deveres

junto a sociedade, sendo baseado em legislações fiscais conforme cada Estado e suas particularidades, definindo sua matriz tributária e os valores cobrados conforme os atributos tributários desejáveis: eficiência, equidade, simplicidade, capacidade contributiva e a segurança jurídica (Rodrigues et al., 2013).

No Brasil, muitos desses atributos não são totalmente atendidos, por possuímos uma alta carga tributária, emaranhado de leis e de normas (Felicio & Martinez, 2018) e uma matriz tributária focada no consumo e de tributo indiretos, que acabam por penalizar quem pode contribuir menos (Gassen et al., 2013) e ainda trazer um aumento de custos as organizações em cumprir a legislação tributária e um aumento dos valores dos produtos e serviços (Costa et al., 2016).

Atualmente, os poucos estudos na área de tributos relacionados a Inteligência Artificial (Chen et al., 2011; Faúndez-Ugalde et al., 2020; Höglund, 2017; Mabe-Madisa, 2018; Oliveira & Santos, 2020) estão mais voltados para coibir evasão fiscal dos contribuintes, em função de suas obrigações para com os entes governamentais (países, estados e municípios), do que para auxiliar as empresas quanto à classificação, reconhecimento e apuração correta dos tributos.

Logo, alguns países já aplicam a Inteligência Artificial para fiscalizar os contribuintes com o objetivo de identificar e coibir evasão fiscal e prevenir crimes tributários (Faúndez-Ugalde et al., 2020). Inclusive, o Brasil, com o projeto Harpia, e outros países sul-americanos já investem há algum tempo em ferramentas para auxiliar na identificação de contribuintes que praticam evasão fiscal (Faúndez-Ugalde et al., 2020). Embora existam estudos, ainda são poucas as pesquisas sobre previsão do risco fiscal, dada a relevância do tema (Höglund, 2017; Oliveira Santos, 2020).

Mesmo com a redução da evasão fiscal pelo uso de tecnologias, os prejuízos aos cofres públicos ainda são grandes. Segundo o IBPT (2020), a sonegação de tributos no Brasil em 2019 foi de 417 bilhões de reais, onde a soma dos atos inflacionários foi de R\$273,1 bilhões (IBPT, 2020). Em função disso, além da redução quanto ao ônus provocado pela evasão, o uso de Inteligência Artificial auxilia na redução dos custos de fiscalização, uma vez que esta tarefa exige pessoas extremamente qualificadas, tempo e muita dedicação (Chen et al., 2011; Mabe-Madisa, 2018).

Diante deste cenário, alguns estudos abordaram o uso de tecnologias relacionadas a Inteligência Artificial no campo tributário. Oliveira e Santos (2020) aplicaram redes neurais para identificar contribuintes com indícios de práticas de evasão fiscal e como resultado obtiveram números que demonstraram como as redes neurais podem auxiliar o fisco na

seleção mais acurada de contribuintes propensos a sonegar. A partir do treinamento da rede neural, a mesma identificou e classificou corretamente 71% dos contribuintes como passíveis de autuação ou não. Por outro lado, em relação a contribuintes que foram autuados, obteve 94% de acerto, demonstrando assim a boa performance da rede neural (Oliveira & Santos, 2020).

Em Taiwan, Chen et al (2011) e Lin et al (2012) também utilizaram Redes Neurais para detectar Relatórios tributários gerados de forma errônea. Chen et al (2011) em seu trabalho observou empresas do ramo de construção, conseguindo uma performance de acerto de 80%. Lin et al (2012) em sua pesquisa, utilizou dados de empresas e indivíduos, a fim de detectar evasão fiscal com maior rapidez e efetividade.

Na Finlândia, Hoglund (2017) utilizou Algoritmos Genéticos para prever padrões de pagamento de impostos. Em seu estudo, Solvência, Liquidez e período de pagamento foram as principais variáveis que impactaram as empresas na inadimplência fiscal (Höglund, 2017). Por outro lado, Mabe-Madisa (2018) utilizou uma combinação de técnicas, Árvores de Decisão e o algoritmo Naive Bayes, para descobrir padrões e classificar os contribuintes quanto a sua atribuição fiscal (Mabe-Madisa, 2018).

Se em diversos setores das empresas observa-se um aumento no uso da Inteligência Artificial, na área tributária e fiscal, ainda são poucos os estudos, sendo uma área carente tanto de estudos quanto a criação de ferramentas baseadas em Inteligência Artificial para otimizar o trabalho. A Inteligência Artificial pode auxiliar em diversos pontos ao falarmos de tributos, como na classificação, no cálculo, na conferência das informações e na auditoria e prevenção de evasão fiscal. No Brasil, este cenário é agravado com inúmeros tributos, quantidade de alíquotas e pelo emaranhado de leis e normas tributárias (Costa et al., 2016).

6. Discussão e Proposta de Agenda de Pesquisa

No decorrer dos tempos, a Inteligência Artificial está despontando como uma das tecnologias mais promissoras em criar inovações, realizar predições e reinventar ou moldar novos negócios capazes de revolucionar o mundo (Duan et al., 2019), sendo um dos catalisadores da Indústria 4.0 (C. S. Lee & Tajudeen, 2020).

As soluções mais utilizadas nas organizações possuem características da Inteligência Artificial categorizada como Analítica, o primeiro nível e onde a maioria das soluções se encontram, sendo aplicada de forma especializada e para resolver problemas mais específicos (Kaplan & Haenlein, 2019) das organizações. Mesmo assim, já espantam com suas características, a ponto de soluções de Inteligência Artificial serem melhores do que

humanos, e até em ganhar do melhor jogador do mundo de Go, como o fez o sistema AlphaGo do Google em 2016 (Duan et al., 2019).

Se a Inteligência Artificial já é discutida desde 1955 (Morgenstern & Mcilraith, 2011), é somente a partir dos últimos anos que diversos produtos tomaram destaque em diversas áreas e nas organizações, influenciando as tomadas de decisão e o setor contábil (Kokina & Davenport, 2017; C. S. Lee & Tajudeen, 2020). Esta ascensão está ocorrendo principalmente a dois pontos cruciais: Disponibilidade de dados e de Potência Computacional (Ma & Sun, 2020; Nguyen et al., 2019).

Algumas organizações já realizam investimentos em novas soluções para potencializar as organizações e os seus negócios, mudando a cultura e as formas de se tomar decisões (Di Vaio et al., 2020), inclusive com a criação de novos negócios. Google, Amazon, Microsoft, Salesforce, e IBM já comercializam infraestruturas para aprendizado de máquina na nuvem (Borges et al., 2021). Qualquer empresa pode comprar e usar estes produtos e criar suas próprias soluções de Inteligência Artificial.

No setor contábil, a Inteligência Artificial já é apontada como o principal catalisador para aumentar a performance e manter as empresas competitivas, reduzindo o tempo dispendido em tarefas repetitivas e até humanamente inviáveis ou ao menos dispendiosas, como na escrituração contábil e auditoria (C. S. Lee & Tajudeen, 2020). Outra tecnologia que pode somar a Inteligência Artificial para modificar profundamente a contabilidade é o Blockchain. O uso de Blockchain deve trazer muito mais segurança e confiabilidade nas operações eletrônicas, com características de inviolabilidade, auditoria e rastreabilidade de operações, facilitando principalmente processos de auditoria. O próprio advento do Blockchain inclusive pode levar a mudanças profundas na contabilidade, como até a alteração do tradicional do modo de duas entradas: débito e crédito para três entradas, onde está terceira etapa seria um novo conjunto de plano de contas para explicar as receitas (Cai, 2021).

Por outro lado, no setor de tributos, vemos o uso mais acentuado por organizações governamentais e estudos associados a prevenir a evasão fiscal. Os Governos já utilizam tecnologias e softwares baseados em Inteligência Artificial para fiscalizarem evasão de tributos, como o projeto Harpia, utilizado pelo governo brasileiro (Faúndez-Ugalde et al., 2020).

Identificar perfis de contribuintes com maiores chances de praticarem evasão de divisas é uma das principais estratégias adotadas. Utilizando o histórico e as características de contribuintes que já sonegaram, com a Inteligência Artificial e aprendizagem profunda,

possibilitam um maior acerto em determinar e encontrar possíveis contribuintes corruptos (Oliveira; Santos, 2020). Esta provavelmente é a melhor estratégia para o Brasil, que possui características peculiares, alta carga tributária e principalmente um emaranhado de diversas leis (Costa et al., 2016).

Frente ao exposto, fica evidente que a Inteligência Artificial tem exercido um papel importante na sociedade, principalmente no que diz respeito à redução de tarefas, agilidade de processos e diminuição de erros de rotinas (Y. Zhang et al., 2020). Nas organizações, é possível observar, pela literatura revisada, que a aplicação das tecnologias de Inteligência Artificial aparece com mais ênfase nas atividades fins, tais como vendas e produção (Ferfernig et al., 2007; Grewal et al., 2017; Ma & Sun, 2020; X. Zhang & Ming, 2021; Zheng et al., 2018).

Diante disso, observa-se que os administrativos, tais como o de contabilidade e o fiscal, ainda são menos beneficiados pela IA. No caso da contabilidade, há diversos estudos, porém a maioria nas áreas de Auditoria e Análises Financeiras (C. S. Lee & Tajudeen, 2020). No fiscal, voltados principalmente a esfera governamental, para coibir a evasão de tributos (Chen et al., 2011; Faúndez-Ugalde et al., 2020; Höglund, 2017; Mabe-Madisa, 2018; Oliveira & Santos, 2021). Assim, infere-se que a aplicação de tecnologias de Inteligência Artificial nas atividades de contabilidade e de tributação podem auxiliar no desempenho e na eficácia das organizações. Portanto, se faz necessário apresentar uma agenda de pesquisa para o desenvolvimento de tecnologias baseadas em Inteligência Artificial aplicadas a esses setores.

A contabilidade tem como objeto de estudo o patrimônio e como objetivo gerar informações econômico-financeiras para a tomada de decisões. Assim, o sistema de informação contábil, consiste em entradas, processamento e saída de informações úteis para que usuários internos e externos tomem decisões (Hendriksen et al., 1999). Na atualidade, a Inteligência Artificial tem sido aplicada no reconhecimento dos fatos e auditoria, ou seja, na execução dos lançamentos contábeis (entrada) no sistema de contabilidade (Kokina & Davenport, 2017; C. S. Lee & Tajudeen, 2020; Y. Zhang et al., 2020). Entretanto, observa-se uma carência de estudos que utilizem ou apliquem a Inteligência Artificial nas outras partes no sistema de informação contábil, ou seja no processamento e na saída de informações.

Logo, a utilização da Inteligência Artificial no reconhecimento dos fatos pode ser ampliada, por meio de buscas automáticas dos eventos que influenciam o patrimônio. Além disso, a aplicação de Inteligência Artificial no processamento pode ser útil para melhorar e

agilizar o sistema de informação contábil e garantir a correta interpretação e aplicação das normas de contabilidade (IFRS). Ademais, a Inteligência Artificial pode melhorar o processo de saída das informações, por meio da elaboração das demonstrações contábeis de forma correta e ágil, além de proporcionar análises econômico-financeiras sofisticadas e úteis. Com isso, pesquisas baseadas na aplicação de Inteligência Artificial podem contribuir para que a informação contábil seja divulgada de forma tempestiva atendendo corretamente as normas de contabilidade (IFRS).

Por outro lado, observa-se que a aplicação da Inteligência Artificial nas questões tributárias são mais presentes na ótica governamental (Chen et al., 2011; Faúndez-Ugalde et al., 2020; Höglund, 2017; Mabe-Madisa, 2018; Oliveira & Santos, 2020) e não nas atividades fiscais presentes nas organizações. As questões fiscais são complexas, pois os governos agem como sujeitos ativos e os contribuintes como passivos perante a obrigação tributária, o que gera um conflito de interesse por natureza (Zugman, 2016). Por um lado, os entes públicos buscam um aumento na arrecadação e, pelo outro, os contribuintes procuram formas de racionalizar o ônus tributário (Zugman, 2016).

De maneira geral, os tributos são divididos em tributos sobre o patrimônio, sobre a renda e sobre o consumo. Dessas classes tributárias, destaca-se os tributos sobre a renda e sobre o consumo por influenciarem as atividades e os resultados empresariais. Os tributos sobre a renda incidem sobre o lucro das organizações e os tributos sobre o consumo influenciam o preço de venda dos produtos e dos serviços (Gassen et al., 2013). Além disso, há de se destacar que imensidão de Leis que regulamentam os tributos existentes e as diferenças tributárias entre os diversos países, são complicadores que trazem complexidade para o entendimento, apuração e reconhecimento dos tributos.

Os tributos sobre a renda possuem impacto no resultado e no valor das empresas e, conseqüentemente, no retorno para os investidores (Hanlon & Heitzman, 2010), o que leva, em alguns, casos ao gerenciamento de resultado (Machado & Nakao, 2012). A Inteligência Artificial pode ajudar a melhorar o processo de apuração e reconhecimento dos tributos sobre a renda, principalmente pelo fato de que esses tributos afetam o resultado das empresas. Assim, a Inteligência Artificial pode contribuir com a diferenciação da legislação tributária com as normas de contabilidade, promover aplicações mais coerentes da legislação, melhorar a elisão fiscal e evitar a evasão fiscal.

Os tributos indiretos são complexos e influenciam o preço de venda dos produtos (Gassen et al., 2013). Os tributos indiretos podem incidir sobre o valor agregado ou sobre o preço de venda (Gassen et al., 2013). Vários países utilizam a modalidade de imposto sobre

o valor agregado (IVA), o que traz aplicações e interpretações diversas. Além disso, existe diversidade de Leis que tratam sobre o assunto e uma complexidade na apuração dos tributos a serem pagos, além do impacto no preço de venda e influência nas margens de lucro. Ademais, o reconhecimento dos tributos indiretos na contabilidade não é tarefa fácil em função de sua diversidade e de sua aplicação. Com isso, a Inteligência Artificial pode auxiliar na aplicação e interpretação da legislação, na apuração dos tributos a recolher e no reconhecimento e contabilização desses tributos.

Por fim, há estudos (Costa, 2014; Costa et al., 2018; Redivo et al., 2019) que tratam do planejamento tributário. O planejamento tributário é a forma lícita de racionalizar tributos, no entanto quando é realizado de forma agressiva (Hanlon & Heitzman, 2010) pode gerar evasão fiscal. Diante disso, o planejamento tributário é um tema importante para melhorar o resultado das organizações. No entanto, a Inteligência Artificial pode auxiliar na execução de planejamentos legais que propiciem benefícios às empresas, sem incorrer em riscos fiscais.

7. Considerações Finais

A Inteligência Artificial, bem como outras tecnologias associadas como os robôs, está provocando profundas mudanças na economia e nas organizações e está levando as empresas a novos horizontes, novos e revolucionários produtos e serviços, trazendo muito mais eficiência em suas operações, rapidez e assertividade nas tomadas de decisões, a diminuição de custos em tarefas corriqueiras e aumento de investimentos na pesquisa e desenvolvimento de novos produtos e serviços.

Estas mudanças não se restringem somente a esfera privada. Diversos países também já utilizam tais tecnologias para otimizar e reduzir os prejuízos quanto a sua eficiência operacional, principalmente quando falamos da fiscalização relativa à evasão fiscal, onde os maiores esforços se concentram.

É possível observar uma utilização especialista da Inteligência Artificial, principalmente nos alvos de negócios das empresas, ou seja, nos seus produtos e serviços, do que propriamente na sua gestão conforme elucidado inicialmente. Uma exceção acontece em empresas especialistas em gestão, como por exemplo as The Big Four. Estas estão criando tecnologias e robôs para aplicar na gestão e automação de tarefas em seus clientes. A exceção se dá, pois o negócio destas empresas é justamente este.

São inúmeras as áreas, dentro das organizações, onde a Inteligência Artificial ainda pode ser aplicada quando consideramos os setores Contábil e Fiscal. Porém, provavelmente,

isso somente deverá ocorrer quando as empresas já tiverem aplicado efetivamente estas tecnologias nas suas atividades alvo, ou através de outras empresas de consultoria e software de gestão que possuam nativamente tais tecnologias e atendam tais setores nestas empresas, como por exemplo as The Big Four (Y. Zhang et al., 2020).

No setor Fiscal especialmente, identificamos lacunas de pesquisa e o pouco uso de técnicas de Inteligência Artificial pelas Organizações. Praticamente inexistem pesquisas e trabalhos que relacionam a Inteligência Artificial com este setor, na gestão, automação e análise de dados e resultados, excetuando-se questões de auditoria. Esta descoberta traz luz a um gap de pesquisa com muitas possibilidades e ainda inexplorado pelo mercado e pela academia, mesmo com a importância deste setor ao funcionamento das Organizações.

Mesmo que pouco explorado, os profissionais destas áreas, contábil e fiscal, devem se preparar para as futuras soluções e ferramentas para atender estes setores, o que irá trazer a estes profissionais novos desafios e mudanças na forma de trabalho. Estes profissionais deverão se adaptar e reaprender suas funções, que serão voltadas mais a sua capacidade de análise dos dados. Se antes tínhamos os analfabetos sobre leitura e escrita, eis que teremos os profissionais que terão de se digitalizar, multidisciplinar, aprender conceitos e até linguagens de programação, para não se tornarem novos analfabetos, porém digitais.

Podemos concluir, que as soluções de Inteligência Artificial usadas pelas organizações são especialistas, ligadas aos objetivos fins das organizações e seus negócios. Existem lacunas de pesquisa e utilização da Inteligência Artificial nos setores Contábil e Fiscal. Na contabilidade, a maioria das aplicações estão ligadas a processos de auditoria e fiscalização, visto que tarefas ligadas a este campo são por si complexas, morosas e difíceis aos profissionais, onde a Inteligência Artificial consegue inclusive desempenhar análises mais profundas e precisas do que os humanos. É carente no setor contábil, aplicações que auxiliem nas outras partes do sistema contábil, quanto ao seu processamento e na análise qualitativa dos dados e do resultado contábil. Na área Fiscal, os estudos e ferramentas se concentram na área governamental, para coibir a evasão de divisas. É latente a demanda por estudos e ferramentas que auxiliem as organizações a realizar a tributação correta de suas operações, a evitar erros e agilizar as parametrizações e cálculos, principalmente no Brasil, com várias regras e uma legislação complexa.

Portanto, este trabalho contribui trazendo conceitos e informações sobre a Inteligência Artificial e sobre sua ascensão nos últimos anos, sua utilização pelas organizações em suas atividades e negócios, no apoio a tomadas de decisão e, especialmente, nas atividades de contabilidade e tributação. Além disso, o estudo traz uma agenda de

pesquisa relacionando a inteligência artificial com a contabilidade e tributação das empresas, contribuindo, assim, com pesquisadores e com empresas que tenham interesse sobre o tema.

8. Referências

Behrens, M. L., & Steinbart, P. J. (1992). Integrating expert systems and artificial intelligence in accounting: A description of the academic program at Memphis State University. *Expert Systems with Applications*, 4(2), 219–223. [https://doi.org/10.1016/0957-4174\(92\)90113-7](https://doi.org/10.1016/0957-4174(92)90113-7)

Borges, A. F., Laurindo, F. J., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, 57, 102225. <https://doi.org/10.1016/j.ijinfomgt.2020.102225>

Bruyn, A. de, Viswanathan, V., Beh, Y. S., Brock, J. K.-U., & Wangenheim, F. von (2020). Artificial Intelligence and Marketing: Pitfalls and Opportunities. *Journal of Interactive Marketing*, 51, 91–105. <https://doi.org/10.1016/j.intmar.2020.04.007>

Cai, C. W. (2021). Triple-entry accounting with blockchain: How far have we come? *Accounting & Finance*, 61(1), 71–93. <https://doi.org/10.1111/acfi.12556>

Chen, J.-H., Su, M.-C., Chen, C.-Y., Hsu, F.-H., & Wu, C.-C. (2011). Application of neural networks for detecting erroneous tax reports from construction companies. *Automation in Construction*, 20(7), 935–939. <https://doi.org/10.1016/j.autcon.2011.03.011>

Cooper, S. B.; Leeuwen, J. V. (2013). Computing Machinery and Intelligence. In *Alan Turing: His Work and Impact* (pp. 551–621). Elsevier. <https://doi.org/10.1016/B978-0-12-386980-7.50023-X>

Costa, D. F. (06/2014). Utilização do orçamento no planejamento tributário de uma pequena empresa. *Revista Brasileira De Contabilidade*, 2014(206), pp. 26–39. <http://rbc.cfc.org.br/index.php/rbc/article/view/1147>

Costa, D. F., Chain, C. P., Carvalho, F. D. M., & Moreira, B. C. d. M. (2016). O custo financeiro dos tributos sobre consumo nas cadeias de suprimento brasileiras: uma proposta metodológica. *Revista Contemporânea De Contabilidade*, 13(29), 91. <https://doi.org/10.5007/2175-8069.2016v13n29p91>

Costa, D. F., Silva, A. C. M., Moreira, B. C. d. M., Costa, M. F., & Andrade, L. P. (2018). Proposta de um modelo de previsão do resultado para o planejamento tributário de pequenas empresas. *Enfoque: Reflexão Contábil*, 37(3), 93.

<https://doi.org/10.4025/enfoque.v37i3.33607>

Cunha, C.;Silveira, H. (2020). Inteligência artificial na formalização de contratos - análise do impacto em uma instituição financeira brasileira de médio porte. *Revista Gestão & Tecnologia*, 20(2), 256–279. <https://doi.org/10.20397/2177-6652/2020.v20i2.1872>

Di Vaio, A., Palladino, R., Hassan, R., & Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. *Journal of Business Research*, 121, 283–314. <https://doi.org/10.1016/j.jbusres.2020.08.019>

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63–71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>

Dubey, R., Gunasekaran, A., Childe, S. J., Bryde, D. J., Giannakis, M., Foropon, C., Roubaud, D., & Hazen, B. T. (2020). Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations. *International Journal of Production Economics*, 226, 107599. <https://doi.org/10.1016/j.ijpe.2019.107599>

Fan, J., Fang, L., Wu, J., Guo, Y., & Dai, Q. (2020). From Brain Science to Artificial Intelligence. *Engineering*, 6(3), 248–252. <https://doi.org/10.1016/j.eng.2019.11.012>

Faúndez-Ugalde, A., Mellado-Silva, R., & Aldunate-Lizana, E. (2020). Use of artificial intelligence by tax administrations: An analysis regarding taxpayers' rights in Latin American countries. *Computer Law & Security Review*, 38, 105441. <https://doi.org/10.1016/j.clsr.2020.105441>

Felicio, R. M., & Martinez, A. L. (2018). Sistema tributário brasileiro: análise da percepção dos operadores do direito tributário à luz dos conceitos de eficiência e justiça fiscal. *REVISTA AMBIENTE CONTÁBIL - Universidade Federal Do Rio Grande Do Norte - ISSN 2176-9036*, 11(1), 156–181. <https://doi.org/10.21680/2176-9036.2019v11n1ID13975>

Ferfernig, A., Terpan, E., & Gula, B. (2007). KNOWLEDGE-BASED RECOMMENDER TECHNOLOGIES FOR MARKETING AND SALES. *International Journal of Pattern Recognition and Artificial Intelligence*, 21(02), 333–354.

<https://doi.org/10.1142/S0218001407005417>

Gassen, V., D'Araújo, P. J. S., & Paulino, S. (2013). Tributação sobre Consumo: o esforço em onerar mais quem ganha menos. *Seqüência: Estudos Jurídicos E Políticos*, 34(66). <https://doi.org/10.5007/2177-7055.2013v34n66p213>

Giudici, P. (2018). Fintech Risk Management: A Research Challenge for Artificial Intelligence in Finance. *Frontiers in Artificial Intelligence*, 1, 1. <https://doi.org/10.3389/frai.2018.00001>

Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The Future of Retailing. *Journal of Retailing*, 93(1), 1–6. <https://doi.org/10.1016/j.jretai.2016.12.008>

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4), 5–14. <https://doi.org/10.1177/0008125619864925>

Hanlon, M., & Heitzman, S. (2010). A review of tax research. *Journal of Accounting and Economics*, 50(2-3), 127–178. <https://doi.org/10.1016/j.jacceco.2010.09.002>

Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0206-3>

Hendriksen, E. S., van Breda, M. F., A. Z. (1999). *Teoria da contabilidade*. Atlas.

Hildebrand, C., & Bergner, A. (2020). Conversational robo advisors as surrogates of trust: onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science*. Advance online publication. <https://doi.org/10.1007/s11747-020-00753-z>

Höglund, H. (2017). Tax payment default prediction using genetic algorithm-based variable selection. *Expert Systems with Applications*, 88, 368–375. <https://doi.org/10.1016/j.eswa.2017.07.027>

IBPT. (2020). Estudo sobre a Sonegação Fiscal das Empresas Brasileiras. Instituto Brasileiro de Planejamento e Tributação. <https://ibpt.com.br/estudo-autos-de-infracao-e-sonegacao-fiscal/>

Kanellou, A., & Spathis, C. (2013). Accounting benefits and satisfaction in an ERP environment. *International Journal of Accounting Information Systems*, 14(3), 209–234. <https://doi.org/10.1016/j.accinf.2012.12.002>

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25. <https://doi.org/10.1016/j.bushor.2018.08.004>

Kokina, J., & Davenport, T. H. (2017). The Emergence of Artificial Intelligence: How Automation is Changing Auditing. *Journal of Emerging Technologies in Accounting*, 14(1), 115–122. <https://doi.org/10.2308/jeta-51730>

Lee, C. S., & Tajudeen, F. P. (2020). Usage and Impact of Artificial Intelligence on Accounting: 213 Evidence from Malaysian Organisations. *Asian Journal of Business and Accounting*, 13(1), 213–240. <https://doi.org/10.22452/ajba.vol13no1.8>

Lee, J., Kao, H.-A., & Yang, S. (2014). Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment. *Procedia CIRP*, 16, 3–8. <https://doi.org/10.1016/j.procir.2014.02.001>

Lin, S.-J., & Hsu, M.-F. (2017). Incorporated risk metrics and hybrid AI techniques for risk management. *Neural Computing and Applications*, 28(11), 3477–3489. <https://doi.org/10.1007/s00521-016-2253-4>

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing – Connecting computing power to human insights. *International Journal of Research in Marketing*, 37(3), 481–504. <https://doi.org/10.1016/j.ijresmar.2020.04.005>

Mabe-Madisa, G. V. (2018). A Decision Tree and Naïve Bayes algorithm for income tax prediction. *African Journal of Science, Technology, Innovation and Development*, 10(4), 401–409. <https://doi.org/10.1080/20421338.2018.1466440>

Machado, M. C., & Nakao, N. (2012). DIFERENÇAS ENTRE O LUCRO TRIBUTÁVEL E O LUCRO CONTÁBIL DAS EMPRESAS BRASILEIRAS DE CAPITAL ABERTO. *Revista Universo Contábil*, 100–112. <https://doi.org/10.4270/ruc.2012324>

Melnychenko, O. (2020). Is Artificial Intelligence Ready to Assess an Enterprise's Financial Security? *Journal of Risk and Financial Management*, 13(9), 191. <https://doi.org/10.3390/jrfm13090191>

Meservy, R. D., Denna, E. L., & Hansen, J. V. (1992). Application of artificial intelligence to accounting, tax, and audit services: Research at Brigham Young University. *Expert Systems with Applications*, 4(2), 213–218. [https://doi.org/10.1016/0957-4174\(92\)90112-6](https://doi.org/10.1016/0957-4174(92)90112-6)

Morgenstern, L., & McIlraith, S. A. (2011). John McCarthy's legacy. *Artificial Intelligence*, 175(1), 1–24. <https://doi.org/10.1016/j.artint.2010.11.003>

Muthukrishnan, N., Maleki, F., Ovens, K., Reinhold, C., Forghani, B., & Forghani, R. (2020). Brief History of Artificial Intelligence. *Neuroimaging Clinics of North America*, 30(4), 393–399. <https://doi.org/10.1016/j.nic.2020.07.004>

Nascimento, A. M., Melo, V. V., Queiroz, A. C. M., Brashear-Alejandro, T., & Meirelles, F. d. S. Inteligência Artificial Aplicada a Pequenas Empresas: O Uso da Engenharia Automática de Recursos e do Aprendizado de Máquina para um Planejamento mais Preciso. *Revista De Contabilidade E Organizações*, 2020.

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López García, Á., Heredia, I., Malík, P., & Hluchý, L. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, 52(1), 77–124. <https://doi.org/10.1007/s10462-018-09679-z>

Oliveira, F. N. de, & Santos, L. P. G. d. (2021). 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. <https://doi.org/10.18028/rgfc.v10i1.7474>

Pawłowski, M. (2021). Machine Learning Based Product Classification for eCommerce. *Journal of Computer Information Systems*, 1–10. <https://doi.org/10.1080/08874417.2021.1910880>

Redivo, J., Almeida, D., & Beuren, I. (2019). Reflexos dos Controles de Gestão no Planejamento Tributário: Um estudo em uma pequena Empresa Industrial. *Pensar Contábil - CRCRJ*, 2020/22(77), 55–67. <http://www.spell.org.br/documentos/ver/58062/reflexos-dos-controles-de-gestao-no-planejamento-tributario--um-estudo-em-uma-pequena-empresa-industrial--/i/pt-br>

Rodrigues, J., Soares, A., & Castro, K. (Eds.). (2013). Avaliação da estrutura e do desempenho do sistema tributário Brasileiro. Banco Interamericano de Desenvolvimento. <https://publications.iadb.org/pt/avaliacao-da-estrutura-e-do-desempenho-do-sistema-tributario-brasileiro-livro-branco-da-tributacao>

Zhang, X., & Ming, X. (2021). An implementation for Smart Manufacturing Information System (SMIS) from an industrial practice survey. *Computers & Industrial Engineering*, 151, 106938. <https://doi.org/10.1016/j.cie.2020.106938>

Zhang, Y., Xiong, F., Xie, Y., Fan, X., & Gu, H. (2020). The Impact of Artificial Intelligence and Blockchain on the Accounting Profession. *IEEE Access*, 8, 110461–110477. <https://doi.org/10.1109/ACCESS.2020.3000505>

Zheng, P., wang, H., Sang, Z., Zhong, R. Y., Liu, Y., Liu, C., Mubarak, K., Yu, S., & Xu, X. (2018). Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives. *Frontiers of Mechanical Engineering*, 13(2), 137–150. <https://doi.org/10.1007/s11465-018-0499-5>

Zugman, D. L. (2016). Reflexões sobre as possíveis razões para não ocorrer uma reforma tributária no Brasil. *Revista Direito GV*, 12(3), 610–631. <https://doi.org/10.1590/2317-6172201625>

PRODUTO 2 (Técnico/tecnológico): Algoritmo de Leitura de Arquivos XML de Notas Fiscais Eletrônicas

INTRODUÇÃO

Nos últimos anos, as empresas realizaram a transição das notas fiscais emitidas em papel para as Notas Fiscais Eletrônicas. São emitidas milhares de notas fiscais diariamente. Segundo o site da NFE, disponibilizado pela Receita Federal do Brasil, foram emitidos 35,765 bilhões de notas fiscais desde 2006, quando o projeto foi implantado (RECEITA FEDERAL DO BRASIL, 2022b). Essa informação foi obtida em 11 de fevereiro de 2023, uma vez que o site é atualizado diariamente.

Notas Fiscais Eletrônicas são arquivos digitais estruturados, assinados e autenticados digitalmente por certificados digitais, que concedem autenticidade sobre o autor - emissor - da nota fiscal (RECEITA FEDERAL DO BRASIL, 2022b). Esses arquivos têm *layout* específico, no formato XML (*Extend Markup Language*) (W3C), como demonstrado na Figura 1.

Figura 1 – Estrutura de *tags* de um arquivo XML de uma NFE

```

<ide>
  <cUF>31</cUF>
  <cNF>10307106</cNF>
  <natOp>VENDA</natOp>
  <mod>65</mod>
  <serie>1</serie>
  <nNF>155929</nNF>
  <dhEmi>2021-10-14T10:56:09-03:00</dhEmi>
  <tpNF>1</tpNF>
  <idDest>1</idDest>
  <cMunFG>3133808</cMunFG>
  <tpImp>4</tpImp>
  <tpEmis>1</tpEmis>
  <cDV>0</cDV>
  <tpAmb>1</tpAmb>
  <finNFe>1</finNFe>
  <indFinal>1</indFinal>
  <indPres>1</indPres>
  <indIntermed>0</indIntermed>
  <procEmi>0</procEmi>
  <verProc>39.32</verProc>
</ide>

```

Fonte: Elaborado pelo autor, 2022.

Os arquivos XML são formados por uma estrutura de chaves - também denominada de *tags* - e valor. Cada abertura de *tag*/elemento é delimitada por < >, e cada elemento de encerramento de *tag*, por </>. Os dados estão entre as duas *tags*, de abertura e fechamento. A título de exemplo, temos <cUF>31</cUF>, que é o código da unidade federativa do emissor da nota fiscal. A *tag* <cUF> é a de abertura; a </cUF>, a de fechamento; e 31, o

valor da *tag*.

Portanto, essas notas fiscais são uma grande fonte de dados e informações que podem ser utilizadas pelas empresas, desde a automação de tarefas cotidianas e para extraírem informações estatísticas até no cadastro de novos produtos, clientes e fornecedores com a importação desses arquivos. Estes podem ser empregados em ferramentas avançadas, com o uso de Inteligência Artificial, para identificar e coibir a evasão de tributos (FAÚNDEZ-UGALDE; MELLADO-SILVA; ALDUNATE-LIZANA, 2020; OLIVEIRA; SANTOS, 2020; HÖGLUND, 2017; CHEN *et al.*, 2011; MABE-MADISA, 2018) ou auxiliar os contribuintes na determinação de qual classificação e características tributárias utilizar nas operações comerciais, que é o objetivo principal deste trabalho.

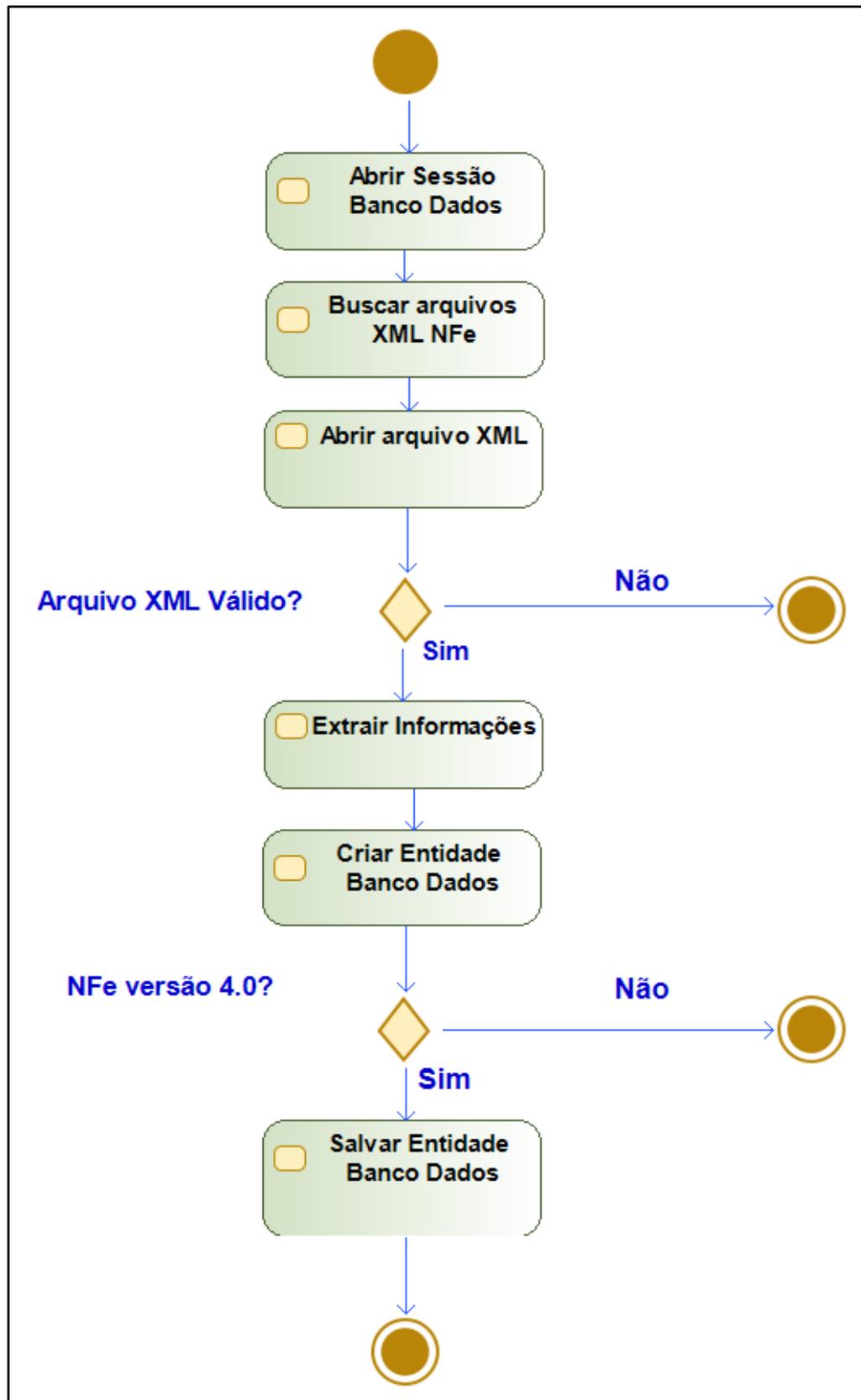
Dessa forma, os algoritmos que trabalham com o processamento desses arquivos devem ter como característica, principalmente, a performance, para processar o maior número de arquivos no menor tempo.

Esse algoritmo irá interagir e ser responsável por obter e extrair os dados dos arquivos XML para popular a base de dados definida no Produto Técnico 3, denominado Produto 3. O algoritmo também poderá ser empregado em outros *softwares*, em especial, sua arquitetura e lógica de programação, para processamento de arquivos de notas fiscais eletrônicas e até outros arquivos, uma vez que tem a capacidade de ser expansível.

METODOLOGIA

O algoritmo proposto neste produto técnico foi desenvolvido com a capacidade de processar diversos arquivos simultaneamente, aproveitando todo o potencial do computador, desde a utilização de memória até a capacidade de processamento dos computadores, uma vez que deverá processar milhões de documentos. Para atingir este objetivo, foram utilizadas ferramentas e técnicas especiais, como a utilização de leitura do arquivo XML por *streaming* e o uso de cache em memória de dados, para evitar buscas excessivas e repetitivas no banco de dados. O fluxo de desenvolvimento empregado neste algoritmo é apresentado na Figura 2.

Figura 2 – Fluxo Algoritmo Leitura Arquivos XML



Fonte: Elaborado pelo autor, 2022.

A arquitetura desse algoritmo foi projetada para ser abastecida por diversos tipos de

fontes de dados, visto que os arquivos das Notas Fiscais Eletrônicas podem ser armazenados em pastas ou diretamente em bancos de dados. Outro ponto associado à arquitetura é a sua extensibilidade.

Inicialmente, o algoritmo tem a capacidade de ler e processar os arquivos XML das notas fiscais. Porém, existem outras fontes de dados que podem ser utilizadas e adicionadas no futuro, por exemplo, arquivos fiscais: Sped Fiscal, Sped Contribuições, Sintegra, dentre outras (RECEITA FEDERAL DO BRASIL, 2022c). Essa abordagem permite que novas fontes de dados sejam adicionadas sem grandes refatorações de código e do projeto. Esses arquivos são obrigações legais da empresa e incluem todas as movimentações passíveis de fiscalização pelos órgãos competentes, a fim de coibir a evasão de tributos; portanto, têm diversas informações sobre produtos, notas fiscais, estoque, dentre outras. A adição dessas funcionalidades, no futuro, aumentará a capacidade desse algoritmo e programa em fornecer mais informações às empresas.

Outra consideração na arquitetura foi em relação ao processamento de eventos associados à nota fiscal (RECEITA FEDERAL DO BRASIL, 2022b). Inicialmente, o evento considerado foi de cancelamento. Algumas empresas salvam de forma separada o XML da nota fiscal e o XML referente ao evento de cancelamento. Dessa forma, o sistema é capaz de processar ambos os arquivos e interpretar que determinada nota fiscal foi cancelada, e, assim, desconsiderá-la nas futuras análises, uma vez que documentos cancelados não têm valor e podem ter sido cancelados por erros de classificação fiscal, o que representaria ruídos e erros nos demais produtos técnicos, os quais se baseiam nesses dados.

A Tabela 1 demonstra a área de atuação econômica e a quantidade de empresas utilizadas, cujas notas fiscais foram processadas pelo algoritmo.

Tabela 1 – Segmentos de Empresas Utilizadas

Ramo/área de atuação	Notas Emitidas	Notas Recebidas	Quantidade de empresas
Comércio – Loja Produtos Agropecuários	434.579	32.457	13
Comércio – Supermercados e Departamentos	586.705	32.467	19
Indústrias – Transformação, Mineração, Siderurgia, Rações	335.080	199.921	35
Distribuidoras	41.424	10.283	4

Serviços	1.638	4.601	6
Comércio Gás e Combustíveis	125.182	8.817	4
Total	1.524.608	288.546	81

Fonte: Elaborado pelo autor, 2022.

O algoritmo processou 1.813.154 arquivos de 81 empresas diferentes. Foram processados tanto os arquivos emitidos quanto os arquivos recebidos pelas empresas, em que estas foram os destinatários da operação. Consideramos separadamente matriz e filial, uma vez que, em um mesmo grupo empresarial, pode haver ramos de atividades diferentes.

No algoritmo, eliminamos automaticamente documentos fiscais que não foram emitidos na versão 4.0. Esse descarte foi necessário, uma vez que a versão 4.0 é a mais recente e tem mais informações sobre as operações do que as versões anteriores. A eliminação, neste momento, irá otimizar a performance do algoritmo, de forma a não salvar no banco de dados notas fiscais que não serão utilizadas pelos outros produtos técnicos deste estudo. Todavia, o algoritmo irá processar o arquivo para identificar a sua versão.

A biblioteca Stax API, disponível em <https://docs.oracle.com/javase/tutorial/jaxp/stax/index.html>, foi selecionada para processar os arquivos XML devido à sua performance, pois realiza a leitura dos arquivos de forma linear, por *streaming*, gerando eventos a cada bloco de informação, a cada *tag* do arquivo do XML. Enquanto outras bibliotecas baseadas em *Document Object Model* (DOM) carregam todo o arquivo XML em memória, a Stax executa leituras por *streaming*, ou seja, à medida que processa o arquivo XML, os dados são carregados em memória, disponibilizados ao algoritmo e, posteriormente, descartados, liberando a memória utilizada (WANG; CHANG, 2019).

Para acelerar a leitura dos documentos, foi realizado *cache* das informações salvas no banco de dados. Ou seja: Registros de pessoas - Emissores e Destinatários das Notas Fiscais, Produtos e Dados Fiscais foram salvos apenas uma vez no banco de dados. Portanto, quando uma nova nota fiscal for processada, caso já exista o produto, por exemplo, foi utilizado o produto já salvo e em *cache* na memória (JING; FAN, 2011). A biblioteca utilizada para esse processo foi a *Ehcache*, disponível em <https://www.ehcache.org/>.

O algoritmo também foi desenvolvido baseado em paralelismo (*multithread*), com a capacidade de realizar a leitura e o processamento dos documentos simultaneamente. Enquanto um processo (*thread*) efetuava a busca dos documentos das notas fiscais e

alimentava uma fila, outras *threads* realizavam o consumo dos documentos dessa fila, processavam e os armazenavam no banco de dados. Dessa forma, foi aproveitado todo o potencial *multithread* do computador.

De forma a assegurar as responsabilidades do código e das classes, foram utilizados objetos/classes intermediários. Portanto, os dados coletados não são armazenados diretamente nas classes de entidades que representam as tabelas do banco de dados. Esses dados são salvos em classes temporárias de transferência de dados, chamadas de *Data Transfer Object* (DTO) (GRIFFIN, 2021). A escolha dessa arquitetura visa preservar a lógica de leitura e processamento dos arquivos XML da lógica das classes de entidades do banco de dados, além de delinear as fronteiras entre o Produto Técnico 2: Algoritmo de Leitura de Arquivos XML e o Produto Técnico 3: Banco de dados. Essa escolha facilita a manutenção do código e evita erros de programação em isolar a leitura e o processamento dos dados dos arquivos XML, das entidades e do banco de dados.

RESULTADOS

A implementação desse algoritmo resultou em uma estrutura preparada para ler e processar diversos tipos de arquivos, nos quais foram implementados a leitura e o processamento dos arquivos XML das Notas Fiscais e dos Eventos Associados. Instalaram-se duas fontes de dados: Arquivos XML locais, em pastas do computador, e Arquivos XML armazenados em banco de dados relacionais, especialmente *Firebird*.

O algoritmo é capaz de processar ambos os arquivos fiscais, a Nota Fiscal e Eventos de Nota Fiscal, em qualquer ordem, e interpretar que determinada nota fiscal foi cancelada, para que seja desconsiderada em análises futuras em outros produtos técnicos.

Outra característica importante desse algoritmo foi a sua performance, como mencionado anteriormente. Portanto, nos testes realizados, foram processados 2.690.206 documentos, obtendo-se êxito em 2.689.497 deles. Essa diferença se deve à existência de documentos inválidos estruturalmente, com problemas no arquivo. Foram salvos 1.752.084, dos quais 937.413 foram desconsiderados. Na seção Metodologia, descrevem-se os critérios de eliminação.

Foram processados 2.689.497 documentos, salvando-se 1.752.084 no banco de dados em 16.255 segundos, 270.91 minutos, 4,51 horas. Ou seja, a taxa de processamento de arquivos obtida pelo algoritmo foi de 165,45 documentos processados por segundo. Os testes foram efetuados nos equipamentos descritos na seção Metodologia. Adicionalmente, essas

medições incluem o processamento e a persistência desses dados no banco de dados, uma vez que realizar esses processos simultaneamente, na forma *streaming*, seria mais rápido do que propriamente processar todos os documentos e, posteriormente, persisti-los no banco de dados. Além disso, otimizou-se o uso de memória do computador.

CONSIDERAÇÕES FINAIS

O algoritmo criado neste produto técnico pode ser empregado isoladamente em outros *softwares* escritos em Java. Neste trabalho, o algoritmo foi utilizado em conjunto com o Produto Técnico 3 para a população de um banco de dados com as informações das Notas Fiscais Eletrônicas.

Em outros *softwares*, pode ser utilizado para a leitura e o processamento de um grande volume de arquivos, em que é necessário processar o maior número de arquivos no menor tempo possível, para que se possa aproveitar todo o potencial do computador: memória e capacidade de processamento paralelo (*multithread*). Apesar de escrito em Java, a lógica de programação empregada pode ser replicada em outras linguagens, adotando-se as mesmas estratégias, como a utilização de DTO, *cache* de dados e gerenciamento de filas de arquivos para processamento paralelo.

A limitação desse algoritmo está em processar apenas arquivos XML de Notas Fiscais Eletrônicas. Em trabalhos futuros, ele pode ser estendido ao processamento de outros tipos de arquivos, por exemplo, arquivos fiscais que são gerados mensalmente pelas empresas, como os arquivos SPED, que têm diversas outras informações sobre as operações comerciais efetuadas pelas empresas.

Esse algoritmo está disponível em um repositório na internet, conforme detalhado no Apêndice C.

REFERÊNCIAS

CHEN, Jieh-Haur; SU, Mu-Chun; CHEN, Chang-Yi; HSU, Fu-Hau; WU, Chin-Chao. Application of neural networks for detecting erroneous tax reports from construction companies. **Automation in Construction**, 20 (7), p. 935–939, 2011. DOI: 10.1016/j.autcon.2011.03.011.

FAÚNDEZ-UGALDE, Antonio; MELLADO-SILVA, Rafael; ALDUNATE-LIZANA, Eduardo. Use of artificial intelligence by tax administrations: an analysis regarding taxpayers' rights in Latin American countries. **Computer Law & Security Review**, 38, p. 105441, 2020. DOI: 10.1016/j.clsr.2020.105441.

GRIFFIN, Jesse (Ed.). **Domain-Driven Laravel**. Berkeley, CA: Apress, 2021. ISBN 978-1-

4842-6022-7.

HÖGLUND, Henrik. Tax payment default prediction using genetic algorithm-based variable selection. **Expert Systems with Applications**, 88, p. 368–375, 2017. DOI: 10.1016/j.eswa.2017.07.027.

JING; FanJing, WANG; Fan, Rui. The research of Hibernate cache technique and application of EhCache component. **2011 IEEE 3rd International Conference on Communication Software and Networks**. 2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN). Xi'an, China. 26/05/2011 - 28/05/2011: IEEE, 2011. ISBN 978-1-61284-485-5, p. 160–162.

MABE-MADISA, G. V: A Decision Tree and Naïve Bayes algorithm for income tax prediction. **African Journal of Science, Technology, Innovation and Development** 10 (4), p. 401–409, 2020. DOI: 10.1080/20421338.2018.1466440.

OLIVEIRA, Francisco Nobre de; SANTOS, Luis Paulo Guimarães dos. Estratégias para Combater a Sonegação Fiscal: um modelo para o ICMS baseado em redes neurais artificiais. **RGFC**, 10 (1), p. 42–64, 2020. DOI: 10.18028/rgfc.v10i1.7474.

RECEITA FEDERAL DO BRASIL: **Portal da Nota Fiscal Eletrônica (NFe)**. 2022b. Disponível em: <https://www.nfe.fazenda.gov.br/>. Acesso em: 3 mar. 2022.

RECEITA FEDERAL DO BRASIL. **Sistema Escrituração Público Digital (SPED)**, 2022c. Disponível em: <http://sped.rfb.gov.br/>. Acesso em: 3 mar. 2022.

WANG, Yulin; CHANG, Chin-Chen (Ed.). **Proceedings of the 3rd International Conference on Cryptography, Security and Privacy**. New York, NY, USA: ACM, 2019. ISBN 9781450366182.

PRODUTO 3 (Técnico/tecnológico): Banco de dados

INTRODUÇÃO

Neste produto técnico, criamos um banco de dados para armazenar as informações extraídas e analisadas pelo Produto 2 para serem utilizadas posteriormente nos Produtos 4 e 5. Este produto também poderá ser utilizado isoladamente como uma base de consulta de cadastros e características de Produtos, Pessoas (Clientes e Fornecedores), Naturezas de Operação e Dados Tributários.

Bancos de dados são estruturas importantes para armazenar, com segurança e rapidez, grandes volumes de dados de forma segura, consistente e estável (NISHTHA *et al.*, 2012). Nesse contexto, a criação do banco de dados deste trabalho foi necessária para evitar redundâncias das informações e otimizar as análises efetuadas, uma vez que realizar as consultas em banco de dados é mais rápido do que processar as informações diretamente nos arquivos XML, devido, principalmente, à redundância de informações.

Nos arquivos das notas fiscais, um mesmo produto e cliente podem aparecer inúmeras vezes, conforme cada operação comercial realizada. No banco de dados, os cadastros serão salvos uma única vez. Ou seja, independentemente do número de operações comerciais executadas com um produto, no banco de dados, esse produto será salvo apenas uma vez.

Optamos por armazenar, também, as demais informações tributárias dos impostos que estavam nas notas fiscais, como IPI, PIS e Cofins (RECEITA FEDERAL DO BRASIL, 2022b). Esses dados foram arquivados para utilizações futuras, por exemplo, a própria expansão do *software* em realizar também a classificação tributária desses impostos.

Apesar de os Produtos 2 e 3 serem produtos técnicos distintos, os dois foram interligados para objetivar a performance, pois, à medida que o algoritmo desenvolvido no Produto Técnico 2 finaliza o processamento de um arquivo, ele o envia em sequência para ser salvo no banco de dados pelo Produto Técnico 3.

METODOLOGIA

As tabelas do banco de dados foram separadas em três grupos para melhor explicação: Grupo de Tabelas de Classificação; Grupo de Tabelas de Movimentações e Grupo de Tabelas Estatísticas. Esses grupos são demonstrados nas Tabelas 1, 2 e 3, sendo que, em cada uma, expõem-se as informações relativas ao nome da tabela do banco de dados,

uma breve descrição e observações inerentes a ela.

As tabelas de banco de dados, destacadas na Tabela 1, armazenam os dados de classificação das operações, responsáveis por determiná-las e categorizá-las.

Tabela 1 – Tabelas de Banco de Dados Classificação

Tabela	Descrição	Observação
CEST	Código do CEST conforme tabela definida pela Receita Federal	Salva os dados referentes ao Código Especificador da Substituição Tributária
CFOP	Código do CFOP conforme tabela definida pela Receita Federal	Salva os Códigos de Operações Prestações
CIDADE	Código IBGE definido para cada cidade	Salva as informações da Cidade
CLASSE_ENQ_IPI	Código conforme tabela definida pela Receita Federal	Salva classes de enquadramento referentes ao imposto IPI
CNAE	Código do CNAE conforme tabela definida pelo Governo Brasileiro	Salva os códigos de Classificação Nacional de Atividades Econômicas Código conforme tabela definida pelo Governo Brasileiro
INCIDENCIA_ICMS	Código conforme tabela definida pela Receita Federal	Salva as classificações fiscais (CST) do ICMS
INCIDENCIA_IPI	Código conforme tabela definida pela Receita Federal	Salva as classificações fiscais (CST) do IPI
INCIDENCIA_PIS_COFINS	Código conforme tabela definida pela Receita Federal	Salva as classificações fiscais (CST) do PIS/Cofins
MODALIDADE_BC_ICMS	Código conforme tabela definida pela Receita Federal	Salva os tipos, modalidades de definição da base de cálculo do ICMS
MODALIDADE_BC_ICMS_ST	Código conforme tabela definida pela Receita Federal	Salva os tipos, modalidades de definição da base de cálculo do ICMSST
MODELO_DOCUMENTO	Modelo de Documento Fiscal	Determina o modelo de documento: 55 NFE e 65 NFCE.
MOTIVO_DESONERACAO_ICMS	Código conforme tabela	Salva os motivos de

	definida pela Receita Federal	desoneração para benefícios fiscais relativos ao ICMS.
MOTIVO_DESONERACAO_ICMS_ST	Código conforme tabela definida pela Receita Federal	Salva os motivos de desoneração para benefícios fiscais relativos ao ICMSST.
NCM	Código conforme tabela definida pelos países integrantes do Mercosul.	Salva códigos de definição do produto, Nomenclatura Comum Mercosul
PAIS	Código IBGE definido para cada país	Salva as informações referentes aos países
PESSOA	Chave Composta: CNPJ + Inscrição Estadual	Salva as informações dos emissores e destinatários da nota fiscal
PESSOA_ENDERECO	Vinculado à Pessoa	Salva as informações de endereço dos emissores e destinatários da nota fiscal
PROCEDENCIA_PRODUTO	Código conforme tabela definida pela Receita Federal	Salva as procedências dos produtos relativos à sua origem
PRODUTO	Foi utilizado o código de barras do produto. Em caso de não existir o código de barras, foi utilizada uma combinação do CNPJ do emissor do documento + código do produto utilizado pelo emissor	Salva as informações dos produtos utilizados nas notas fiscais
TIPO_EVENTO_NFE	Salva os tipos de eventos associados à NFE	Cancelamento; carta de correção
TIPO_TRIBUTACAO_EMPRESA	Código do Regime, conforme definido no manual NFE pela Receita Federal	Salva o tipo de tributação
TIPO_TRIBUTACAO_EMPRESA	Salva o Tipo de Tributação das empresas, conforme definido pela Receita	1 – Simples Nacional 2 – Simples Nacional, excesso Receita 3 – Tributação Normal
UNIDADE_FEDERATIVA	Código IBGE definido para cada unidade federativa	Salva as informações referentes às unidades federativas. No caso de operações com unidades federativas, foi utilizada

UNIDADE_MEDIDA	Sigla da unidade de medida utilizada nas notas fiscais	apenas uma unidade: EX. Salva as unidades de medida dos produtos.
----------------	--	--

Fonte: Elaborado pelo autor, 2022.

Esse é o maior grupo de tabelas, face às várias classificações e cadastros existentes, desde a classificação do emissor da nota fiscal até o produto e a movimentação em si. Essas tabelas serão populadas dinamicamente a partir da leitura e processamento dos arquivos. Tomemos como exemplo a tabela NCM: ao realizar a leitura de um código NCM, caso ele não exista, será criado; por outro lado, caso exista, será utilizado o registro vigente. Dessa forma, não haverá registros e códigos duplicados.

Na Tabela 2, em menor quantidade de tabelas no banco de dados, estão aquelas responsáveis por armazenar as movimentações.

Tabela 2 – Tabela de Banco de Dados Movimentações

Tabela	Descrição	Observação
ASSOCIACAO_COFINS	Dados Fiscais Cofins	Salva os dados tributários relativos ao Cofins
ASSOCIACAO_ICMS	Dados Fiscais ICMS	Salva os dados tributários relativos ao ICMS
ASSOCIACAO_ICMS_ST	Dados Fiscais ICMS ST	Salva os dados tributários relativos ao ICMS ST
ASSOCIACAO_IPI	Dados Fiscais IPI	Salva os dados tributários relativos ao IPI
ASSOCIACAO_PIS	Dados Fiscais PIS	Salva os dados tributários relativos ao PIS
ASSOCIACAO_TRIBUTARIA	Dados Fiscais Correlações Tributárias	Salva as Associações Tributárias, as correlações entre os tributos PIS, Cofins, ICMS, ICMS ST, IPI
ASSOCIACAO_TRIB_EMISSOR	Associações Tributárias - Emissores	Salva as vinculações de Associações Tributárias e Emissores/Pessoas
ASSOCIACAO_TRIB_NAT_OP	Associações Tributárias – Natureza Operação	Salva as vinculações de Associações Tributárias e

		Naturezas de Operação
ASSOCIACAO_TRIB_NCM	Associações Tributárias - NCM	Salva as vinculações de Associações Tributárias e NCM
ASSOCIACAO_TRIB_PROD	Associações Tributárias – Produtos	Salva as vinculações de Associações Tributárias e Produtos
ASSOCIACAO_TRIB_UF	Associações Tributárias – Unidades Federativas	Salva as vinculações de Associações Tributárias e Unidades Federativas
DOCUMENTO_NFE	Documentos Fiscais	Salva os dados da nota fiscal
DOCUMENTO_NFE_ITEM	Vinculado à nota fiscal	Salva os itens da nota fiscal
DOCUMENTO_NFE_ITEM_COFINS	Vinculado ao item da nota fiscal	Salva as informações de Cofins de cada item da nota fiscal
DOCUMENTO_NFE_ITEM_PIS	Vinculado ao item da nota fiscal	Salva as informações de PIS de cada item da nota fiscal
DOCUMENTO_NFE_ITEM_IPI	Vinculado ao item da nota fiscal	Salva as informações de IPI de cada item da nota fiscal
DOCUMENTO_NFE_ITEM_ICMS	Vinculado ao item da nota fiscal	Salva as informações de ICMS de cada item da nota fiscal
DOCUMENTO_NFE_ITEM_ICMS_ST	Vinculado ao item da nota fiscal	Salva as informações de ICMSST de cada item da nota fiscal.

Fonte: Elaborado pelo autor, 2022.

Para armazenar as movimentações, utilizaram-se 18 tabelas no banco de dados. Esse número de tabelas foi necessário para separar pela importância e pelo grupo de informações de cada imposto e funcionalidade. Assim, fica mais fácil interpretar e organizar os dados. Nesse ponto, temos duas tabelas principais: DOCUMENTO_NFE e DOCUMENTO_NFE_ITEM, que armazenam as informações relativas às notas fiscais e seus itens. As demais tabelas, vinculadas a DOCUMENTO_NFE_ITEM, salvam as informações tributárias de cada imposto relativo a cada item da nota fiscal.

As tabelas referentes à Associação Tributária, como ASSOCIACAO_COFINS, ASSOCIACAO_IPI, ASSOCIACAO_TRIBUTARIA e outras, salvam as características tributárias. Elas são populadas após serem processados os dados pelo Produto Técnico 2.

Mais detalhes sobre o processamento estão no Produto Técnico 2.

Por fim, o terceiro grupo, demonstrado na Tabela 3, armazena as informações estatísticas relativas aos dados utilizados.

Tabela 3 – Tabelas de Banco de Dados Estatísticas

Tabela	Descrição	Observação
SOURCE_DATA	Caminho do banco de dados de arquivos das notas fiscais no computador	Salva informações dos bancos de dados utilizados
SOURCE_DATA_STATS	Vinculado à tabela de bancos utilizados	Salva informações de cada tipo de dado coletado: 1-Notas Emitidas 2-Notas Recebidas

Fonte: Elaborado pelo autor, 2022.

Nesse grupo de tabelas, ficam armazenados os dados relativos à origem, ou seja, de quais pastas de arquivos, de quais bancos de dados as notas fiscais foram extraídas, dentre outras informações, como o número total de documentos e de documentos válidos. Também foram armazenados os tempos de processamento dos arquivos de cada fonte de dados utilizada, para análises estatísticas.

Para simplificar a inserção e manutenção dos dados no banco, utilizamos o framework Hibernate, disponível em <https://hibernate.org/>. Dessa forma, foi possível utilizar a linguagem Java, que é orientada a objetos com o banco de dados que tem estrutura relacional. A utilização do Hibernate também possibilita o uso de outros bancos de dados, e não somente o utilizado neste trabalho (Firebird) (BAUER *et al.*, 2016). Portanto, se forem utilizados outros tipos de banco de dados no futuro, não será necessário reescrever o algoritmo.

RESULTADOS

No diagrama das tabelas do banco de dados, conforme Figura 1, estão demonstradas as tabelas de relevância. Algumas tabelas foram suprimidas para melhorar a visualização. Em ponto central, temos as tabelas DOCUMENTO_NFE e DOCUMENTO_NFE_ITEM, que são as principais do banco de dados, onde foram armazenadas as notas fiscais e seus

DOCUMENTO_NFE_ITEM_ICMS	8.850.445
DOCUMENTO_NFE_ITEM_ICMSST	8.850.445
DOCUMENTO_NFE_ITEM_IPI	8.850.445
DOCUMENTO_NFE_ITEM_PIS	8.850.445
EVENTO_NFE	97.114
INCIDENCIA_ICMS	22
INCIDENCIA_IPI	14
INCIDENCIA_PIS_COFINS	33
MODALIDADE_BC_ICMS	4
MODALIDADE_BC_ICMS_ST	7
MODELO_DOCUMENTO	6
MOTIVO_DESONERACAO_ICMS	12
MOTIVO_DESONERACAO_ICMSST	12
NCM	12.914
NCM_CAPITULO	98
NCM_ITEM	10.311
NCM_POSICAO	1.514
NCM_SECAO	21
NCM_SUBPOSICAO	6.842
PAIS	242
PESSOA	56.284
PESSOA_ENDERECO	56.284
PROCEDENCIA_PRODUTO	9
PRODUTO	250.797
SOURCE_DATA	23
SOURCE_DATA_STATS	92
TIPO_EVENTO_NFE	7

TIPO_TRIBUTACAO_EMPRESA	3
UNIDADE_FEDERATIVA	28
UNIDADE_MEDIDA	895

Fonte: Elaborado pelo autor, 2022.

Portanto, foram armazenados, no banco de dados, 55.357.394 registros em 40 tabelas. O banco de dados ficou com 7,17 GB, e as tabelas com maior número de registros foram as relacionadas com os itens da nota fiscal, com 53.102.670 registros.

CONSIDERAÇÕES FINAIS

O banco de dados deste produto técnico pode ser utilizado, principalmente, para estudos e pesquisas pelas próprias empresas ou pela academia, à medida que crescer e tiver um maior volume de dados. Neste trabalho, o banco de dados foi utilizado em conjunto com os Produtos 4 e 5.

No Produto 4, foi empregado para análise e extração de dados, para a geração de informações cadastrais e tributárias das empresas.

No Produto 5, os dados foram utilizados pela Inteligência Artificial, a fim de aprender com as movimentações realizadas pelas empresas e, assim, sugerir o CST de ICMS para operações desconhecidas, que ainda não foram efetuadas pelas empresas.

As empresas também podem utilizar esse banco de dados para automatizarem processos cadastrais e parametrizações tributárias. Afinal, se uma empresa comercializou um produto A, e outra deseja comercializar esse produto, ela pode utilizar os dados de cadastros e parametrizações já informados pela primeira empresa e, assim, evitar digitações manuais.

A limitação deste produto está no número e no ramo de atividades das empresas em que os dados foram avaliados e coletados. Portanto, como trabalho futuro, sugere-se a adição de mais empresas e ramos de atividade, para que a base de dados se torne cada vez maior. Sugere-se, também, a criação de integrações via *webservices* entre esse banco de dados e os *softwares* de gestão das empresas, para que a adição de novas informações ocorra de forma *on-line* e simultânea, sem a necessidade de importação assíncrona de arquivos.

O banco de dados e o algoritmo responsável por criá-lo e manuseá-lo estão disponíveis em um repositório na internet, conforme detalhado no Apêndice C.

REFERÊNCIAS

BAUER, Christian; KING, Gavin; GREGORY, Gary. **Java persistence with Hibernate**. 2. ed. Greenwich, Conn.: Manning, 2016. 580 p. ISBN 9781617290459.

NISHTHA, Jatana; PURI, Sahil; AHUJA, Mehak; KATHURIA, Ishita; GOSAIN, Dishant. A Survey and Comparison of Relational and Non-Relational Database. **International Journal of Engineering Research & Technology (IJERT)**, 2012. Disponível em: <<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=14897e1d7a48840ac77809966376a29fe9a09d2b>>.

RECEITA FEDERAL DO BRASIL: **Portal da Nota Fiscal Eletrônica (NFe)**. 2022b. Disponível em: <https://www.nfe.fazenda.gov.br/>. Acesso em: 3 mar. 2022.

PRODUTO 4 (Técnico/tecnológico): Algoritmo de Exportação de dados

INTRODUÇÃO

Com o Produto Técnico 3 finalizado, foi desenvolvido o algoritmo capaz de exportar os dados vinculados a um CNPJ.

Esses dados poderão facilitar a migração de informações pelas empresas e a análise das movimentações tributárias realizadas. Podemos ilustrar, como exemplo de uso, a necessidade de migração de dados entre *softwares*, quando a empresa decide trocar de *software* de gerenciamento fiscal.

O processo de migração e substituição de *softwares* acaba proporcionando muito trabalho às empresas, uma vez que a base de dados anterior pode estar poluída, com cadastros repetidos e antigos, que não são mais utilizados. Adicionalmente, pode ser difícil de se obter esses dados com o *software* anterior, devido à incompatibilidade entre os *softwares*, uma vez que, em muitas situações, os usuários devem cadastrar manualmente essas informações (BUZATTO; SILVEIRA, 2016).

Com este produto, as empresas poderão utilizar os próprios documentos fiscais emitidos e os recebidos de seus fornecedores, que têm dados confiáveis, uma vez que tais documentos foram autorizados pela Sefaz, evitando erros, cadastros duplicados, bem como outros vícios e ruídos nos dados.

METODOLOGIA

O algoritmo será capaz de exportar os seguintes dados: Produtos, Clientes, Fornecedores, Naturezas de Operação, Dados Fiscais IPI, Dados Fiscais PIS, Dados Fiscais Cofins, Dados Fiscais ICMS e Dados Fiscais Associações Tributárias, no formato CSV. Esses dados estão disponíveis no arquivo XML da NFE (RECEITA FEDERAL DO BRASIL, 2022b).

Como parâmetro, receberá o CNPJ alvo, em que realizará as buscas no banco de dados para retornar os dados associados, conforme o tipo de dado a ser exportado e, assim, gerar o arquivo de dados no formato CSV.

Para a análise desse algoritmo, serão considerados todos os documentos associados ao CNPJ, no qual o CNPJ informado seja o mesmo do emissor ou do destinatário das notas fiscais.

O algoritmo irá gerar uma chave única para cada dado armazenado, de acordo com a correlação de campos de cada tipo de registro avaliado, conforme detalhado na Tabela 1: Produtos, Fornecedores, Clientes etc.

Tabela 1 – Registros e Campos das Chaves

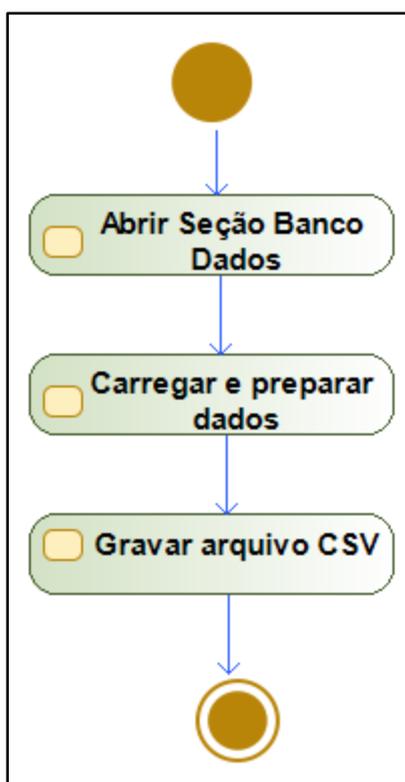
Item	Campos-Chave
Produto	<p>- Caso Produto tenha Código Barras: código_barras + código_ncm + código_cest</p> <p>- Caso Produto não tenha Código Barras: cnpj_emissor + código_produto + código_ncm + código_cest</p>
Cliente (Pessoa)	cnpj + inscrição_estadual
Fornecedores (Pessoa)	cnpj + inscrição_estadual
Naturezas Operação	cnpj + inscrição_estadual + código_hash (descrição_nat_operacao)
Dados Fiscais IPI	código_incidencia_ipi
Dados Fiscais PIS/Cofins	código_incidencia_pis_cofins
Dados Fiscais ICMS	código_incidencia_icms + código_modalidade_bc + código_mot_desoneracao
Dados Fiscais ICMS ST	código_incidencia_icms + código_modalidade_bc + código_mot_desoneracao
Dados Fiscais Associações Tributárias	código_cfop + chave_ipi + chave_pis_cofins + chave_icms + chave_icms_st

Fonte: Elaborado pelo autor, 2022.

As chaves representarão cada registro de forma única, sendo que, relativo à Associação Tributária, o algoritmo irá correlacionar os impostos nas notas fiscais, as combinações de IPI, PIS/Cofins, ICMS/ICMS ST e CFOP. Especialmente em relação ao produto, haverá uma condição para montagem da chave. Caso haja código de barras, que é uma identificação única do produto, ele será utilizado. Por outro lado, por não ser obrigatório, caso não tenha sido informado, serão utilizados o CNPJ do emissor e o código do produto por ele definido. Assim, será possível exportar os produtos referentes a cada CNPJ, mesmo que haja diversos outros produtos, de diferentes CNPJ, no mesmo banco de dados.

Para cada item, será gerado um arquivo CSV, conforme Figura 1. Entretanto, para o item Associação Tributária, serão gerados arquivos CSV adicionais, com as condições em que a associação tributária foi utilizada e em quais produtos, NCMs, estados e naturezas de operação utilizou-se cada Associação Tributária.

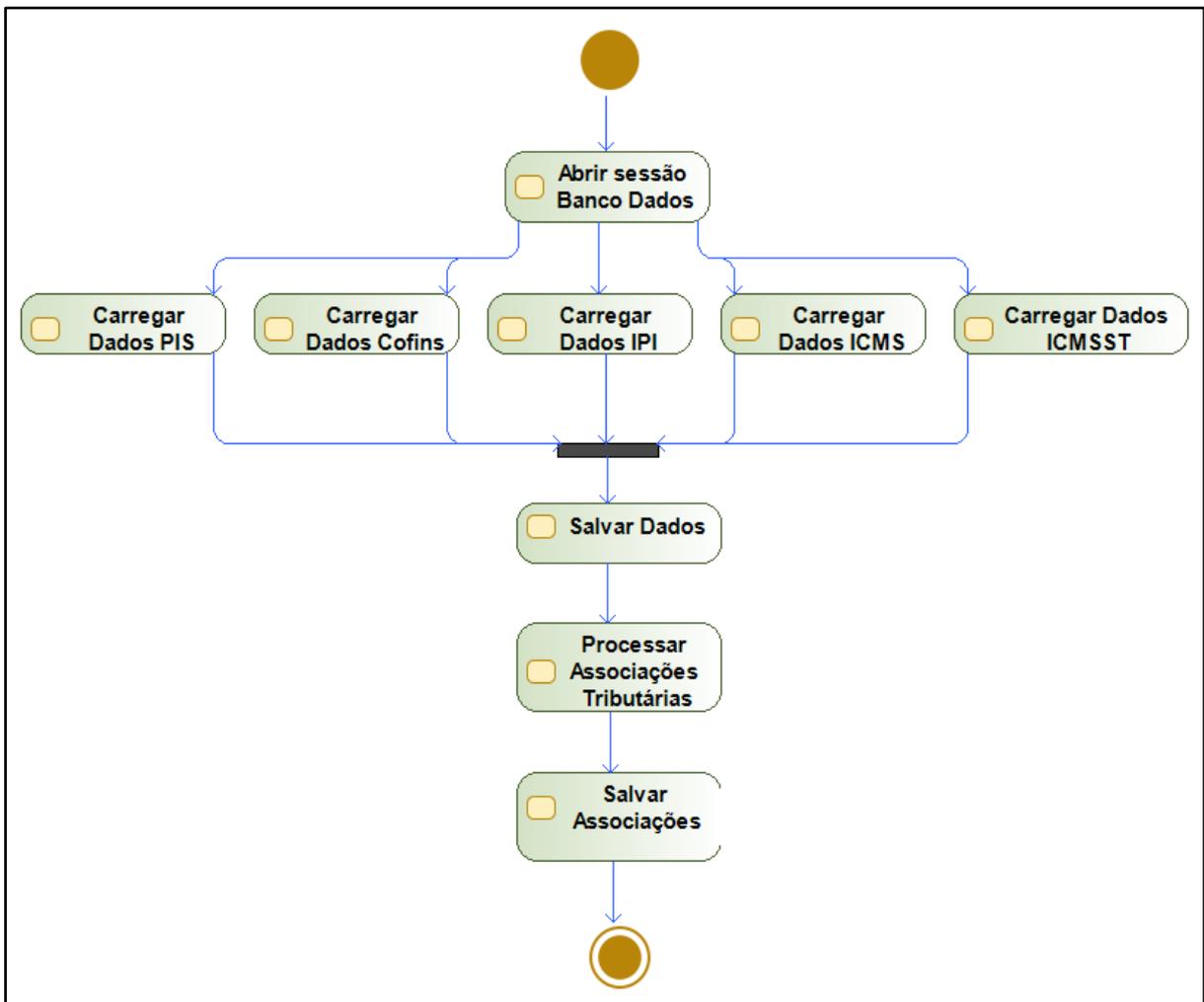
Figura 1 – Fluxo algoritmo exportação dados em arquivo CSV



Fonte: Elaborado pelo autor, 2023.

A exportação de Clientes, Fornecedores, Produtos e Naturezas de Operação será realizada de forma direta, conforme os dados disponíveis no banco de dados. Por outro lado, com relação aos dados tributários, relativos ao IPI, PIS, Cofins, ICMS/ICMS ST e Associações Tributárias, será necessário realizar um pré-processamento antes da exportação dos dados, pelo menu e botões disponibilizados na interface, conforme detalhado no Apêndice B. A Figura 2 ilustra o diagrama de atividades realizado pelo algoritmo nesse processamento.

Figura 2 – Fluxo algoritmo processamento de Dados e Associações Tributárias



Fonte: Elaborado pelo autor, 2023.

O processamento da associação tributária será necessário para que o algoritmo crie, de forma única, cada tipo de tributação. Realizar esse processo durante a leitura e processamento dos arquivos XML demandaria mais recursos do computador, sendo mais eficiente realizá-lo após todos os dados importados e salvos.

Sobre a exportação de produtos, o algoritmo poderá exportar os dados demonstrados na Tabela 2.

Tabela 2 – Dados do Produto

Campo	Descrição
COD_PRODUTO	Código do Produto
PRODUTO	Descrição do Produto

CODIGO_BARRAS	Código de barras do produto
UM	Sigla Unidade Medida
UN_TRIBUTADA	Sigla Unidade Medida Tributada
NCM	Código do NCM
CEST	Código do CEST, Código Especificação Tributária
ALIQ_COFINS_VLR_QTD	Alíquota Cofins Valor ou Quantidade. Determina se o produto é calculado por alíquota percentual ou por unidade comercializada
ALIQ_COFINS	Alíquota Cofins
ALIQ_COFINF_QTD	Alíquota Cofins Quantidade
ALIQ_PIS_VLR_QTD	Alíquota PIS Valor ou Quantidade. Determina se o produto é calculado por alíquota percentual ou por unidade comercializada
ALIQ_PIS	Alíquota PIS
ALIQ_PIS_QTD	Alíquota PIS Quantidade
ALIQ_ICMS	Alíquota ICMS
PERC_RED_BC_ICMS	Percentual Redução Base Cálculo ICMS
ALIQ_ICMSST	Alíquota ICMS ST
PERC_RED_BC_ICMSST	Percentual Redução Base de Cálculo ICMSST
PMVAST_ICMSST	Percentual da Margem de Valor Adicionado do ICMS ST
ALIQ_IPI_QTD	Alíquota IPI por unidade comercializada do produto
ALIQ_IPI	Alíquota IPI do produto
COD_CLASSE_ENQ_IPI	Código da Classe de enquadramento do produto junto ao IPI
CHAVE	Chave do produto

Fonte: Elaborado pelo autor, 2022.

São exportados 22 campos relacionados aos produtos, que são os campos necessários para realizar o faturamento e a comercialização dos produtos por nota fiscal.

No que se refere aos Clientes e Fornecedores, a Tabela 3 apresenta os dados que são

exportados.

Tabela 3 – Dados de Fornecedores e Clientes

Campo	Descrição
NOME	Nome
NOME_FANTASIA	Nome Fantasia
CNPJ	CNPJ
INSCRICAO_ESTADUAL	Inscrição Estadual
INSCRICAO_ESTADUAL_ST	Inscrição Estadual Substituição Tributária
INSCRICAO_MUNICIPAL	Inscrição Municipal
LOGRADOURO	Logradouro
NUMERO	Número
BAIRRO	Bairro
CEP	CEP
COD_IBGE_CIDADE	Código IBGE da Cidade
COD_IBGE_UF	Código IBGE da UF
COD_IBGE_PAIS	Código IBGE do País
TELEFONE	Telefone

Fonte: Elaborado pelo autor, 2022.

Em relação aos Clientes e Fornecedores, são exportados 14 campos. O algoritmo, por meio dos campos de Emissor e Destinatário, classificará os fornecedores e clientes de acordo com o papel deles nas notas fiscais.

Se o CNPJ informado estiver no campo Emissor, serão exportados os dados do campo destinatário como Clientes; por outro lado, se o CNPJ informado estiver no campo Destinatário, serão exportados os dados do campo Emissor como Fornecedores.

No que diz respeito aos dados referentes à Naturezas de Operação, a Tabela 4 apresenta os dados que são exportados.

Tabela 4 – Dados de Natureza Operação

Campo	Descrição
IDENTIFICADOR	Identificador/Código da Natureza
CNPJ_EMPRESA	CNPJ da Empresa
DESCRICA0	Descrição da Natureza de Operação
COD_MOD_DOC_FISCAL	Código do Modelo de Documento Fiscal: NFE (55) e NFCe (65)
ENT_SAI	Movimentação de Entrada (2) ou Saída (1)
CHAVE	Chave Natureza Operação

Fonte: Elaborado pelo autor, 2022.

Para as Naturezas de Operação, são exportados seis campos para a identificação das naturezas de operação e tipos de documentos associados.

Os dados de Tributação relacionados ao IPI são demonstrados na Tabela 5.

Tabela 5 – Dados de Tributação IPI

Campo	Descrição
INCIDENCIA_IPI	Código da Incidência IPI
DESCRICA0	Descrição da Incidência IPI
ENT_SAI	Movimentação de Entrada (2) ou Saída (1)
CNPJ_EMPRESA	CNPJ da Empresa
CHAVE	Chave Tributação IPI

Fonte: Elaborado pelo autor, 2022.

Para o IPI, são exportados cinco campos de informações. Como é um tributo federal, existem menos variações de campos, e as principais estão vinculadas ao cadastro do Produto, como NCM e Classe de Enquadramento, além das alíquotas. Os dados relativos ao PIS e Cofins são demonstrados nas Tabelas 5 e 6.

Tabela 5 – Dados de Tributação PIS

Campo	Descrição
INCIDENCIA_PIS	Código da Incidência PIS

DESCRICA0	Descrição da Incidência PIS
ENT_SAI	Movimentação de Entrada (2) ou Saída (1)
CNPJ_EMPRESA	CNPJ da Empresa
CHAVE	Chave Tributação PIS

Fonte: Elaborado pelo autor, 2022.

Tabela 6 – Dados de Tributação Cofins

Campo	Descrição
INCIDENCIA_COFINS	Código da Incidência Cofins
DESCRICA0	Descrição da Incidência Cofins
ENT_SAI	Movimentação de Entrada (2) ou Saída (1)
CNPJ_EMPRESA	CNPJ da Empresa
CHAVE	Chave Tributação Cofins

Fonte: Elaborado pelo autor, 2022.

PIS e Cofins também possuem menos campos, da mesma forma que o IPI, por também serem tributos federais. Por outro lado, sobre o ICMS, são exportados mais campos, como exposto na Tabela 7.

Tabela 7 – Dados de Tributação ICMS e ICMS ST

Campo	Descrição
CST_ICMS	Código da Incidência ICMS
INCIDENCIA_ICMS	Descrição da Incidência ICMS
COD_MOD_BC_ICMS	Código Modalidade BC ICMS
MODALIDADE_BC_ICMS	Descrição Modalidade BC ICMS
COD_MOT_DESONERACAO_ICMS	Código Motivo Desoneração ICMS
MOTIVO_DESONERACAO_ICMS	Descrição Motivo Desoneração ICMS
COD_MOD_BC_ICMSST	Código Modalidade BC ICMS ST
MODALIDADE_BC_ICMSST	Descrição Modalidade BC ICMS ST

COD_MOT_DESONERACAO_ICMSST	Código Motivo Desoneração ICMS ST
MOTIVO_DESONERACAO_ICMSST	Descrição Motivo Desoneração ICMS ST
ENT_SAI	Movimentação de Entrada (2) ou Saída (1)
CNPJ_EMPRESA	CNPJ da Empresa
CHAVE_ICMS	Chave Tributação ICMS
CHAVE_ICMS_ST	Chave Tributação ICMS ST

Fonte: Elaborado pelo autor, 2022.

Para o ICMS e ICMSST, são exportados 12 campos: a Incidência de ICMS, que determina se haverá ou não o cálculo de ICMS e/ou ICMS ST, e as modalidades de base de cálculo e motivos de desoneração. Se a Incidência de ICMS define a tributação, a modalidade de base de cálculo determina sobre qual valor será calculado o ICMS/ICMS ST. Por outro lado, o motivo de desoneração define e justifica o não recolhimento ou cálculo do ICMS. Esses casos ocorrem quando o governo estadual desonera o tributo em movimentações específicas, por exemplo, vendas de automóveis para taxistas ou PCD (RECEITA FEDERAL DO BRASIL, 2022b).

Por último, o algoritmo exporta os dados relativos às combinações de impostos, os quais são especificados e detalhados nas notas fiscais. Denominamos essas combinações, que serão efetuadas pelas chaves de cada imposto, de Associações Tributárias. Adicionalmente, são criados quatro arquivos, além do próprio arquivo da Associação Tributária: Associações Tributárias Natureza Operação; Associações Tributárias NCM; Associações Tributárias UF e Associações Tributárias Produtos. Os dados exportados no arquivo de associações tributárias e os demais estão detalhados nas Tabelas 8, 9, 10, 11 e 12. O objetivo desses arquivos é especificar em quais situações cada Associação Tributária será aplicada, em quais produtos, NCM, estados e tipos de movimentações (Naturezas de Operações).

Tabela 8 – Dados de Associações Tributárias

Campo	Descrição
CFOP	Código CFOP da Operação
CHAVE	Código da Associação Tributária

CHAVE_ICMS	Chave do Imposto ICMS
CHAVE_ICMS_ST	Chave do Imposto ICMS ST
CHAVE_COFINS	Chave do Imposto COFINS
CHAVE_PIS	Chave do Imposto PIS
CHAVE_IPI	Chave do Imposto IPI
CNPJ_EMPRESA	CNPJ da Empresa
ENT_SAI	Movimentação de Entrada (2) ou Saída (1)

Fonte: Elaborado pelo autor, 2022.

Tabela 9 – Dados de Associações Tributárias – Naturezas Operação

Campo	Descrição
CHAVE_ASS_TRIBUTARIA	Chave da Associação Tributária
CHAVE_NAT_OP	Chave da Natureza de Operação
DESCRICA0	Descrição da Natureza Operação
CNPJ_EMPRESA	CNPJ da Empresa

Fonte: Elaborado pelo autor, 2022.

Tabela 10 – Dados de Associações Tributárias – NCM

Campo	Descrição
CHAVE_ASS_TRIBUTARIA	Chave da Associação Tributária
CODIGO_NCM	Código do NCM
NCM	Descrição do NCM
CNPJ_EMPRESA	CNPJ da Empresa

Fonte: Elaborado pelo autor, 2022.

Tabela 11 – Dados de Associações Tributárias – UF

Campo	Descrição
CHAVE_ASS_TRIBUTARIA	Chave da Associação Tributária
UF	Sigla da Unidade Federativa
CNPJ_EMPRESA	CNPJ da Empresa

Fonte: Elaborado pelo autor, 2022.

Tabela 12 – Dados de Associações Tributárias – Produtos

Campo	Descrição
CHAVE_ASS_TRIBUTARIA	Chave da Associação Tributária
CHAVE_PRODUTO	Chave do Produto
CODIGO_PRODUTO	Código do Produto
PRODUTO	Descrição do Produto
EAN	Código Barras do Produto
CNPJ_EMPRESA	CNPJ da Empresa

Fonte: Elaborado pelo autor, 2022.

É importante observar que, no arquivo principal, temos a chave da Associação Tributária, que é replicada em todos os demais arquivos, além das chaves de cada tributo. Nos arquivos adicionais, no caso de Produto e Natureza de Operação, temos as respectivas chaves de cada Produto e Natureza de Operação vinculadas à Associação Tributária. Para as Unidades Federativas, a chave utilizada foi a própria sigla da UF, que é única. Por outro lado, para o NCM, utilizou-se o próprio código, uma vez que não existe mais de um NCM com o mesmo código, sendo este um campo único.

RESULTADOS

Foram criados menus na interface gráfica do aplicativo para possibilitar a exportação dos dados com base no CNPJ informado. Detalhes da interface gráfica podem ser verificados no Apêndice B.

Portanto, é gerado um arquivo para cada tipo de dado solicitado: Produtos, Fornecedores, Clientes, Naturezas de Operação, Dados Fiscais IPI, Dados Fiscais PIS/Cofins, Dados Fiscais ICMS e Dados Associações Tributárias utilizados nas notas fiscais, cujo CNPJ do emissor é igual ao CNPJ informado.

Para as relações tributárias, o algoritmo efetua uma leitura nas informações do banco de dados e carrega as combinações e parâmetros únicos de cada imposto: IPI, PIS/Cofins, ICMS/ICMS ST, sendo essas as parametrizações fiscais utilizadas pela empresa. Sobre as Associações Tributárias, o algoritmo correlaciona as combinações dos impostos nas notas

fiscais, ou seja, correlaciona os tributos IPI x PIS x Cofins x ICMS/ICMS ST, uma vez que, na nota fiscal, todos esses impostos são calculados e informados.

Para mensurar a capacidade do algoritmo e performance, foi criada uma base de dados separada, na qual foram importados apenas as notas fiscais e os arquivos XMLs de um CNPJ específico. Esses arquivos estavam armazenados em um banco de dados de uma das empresas participantes do projeto.

Na Tabela 13, são demonstradas as estatísticas do algoritmo, em relação ao número de registros exportados e o tempo gasto na exportação das informações desse banco de dados utilizado como exemplo.

Tabela 13 – Dados de Fornecedores e Clientes

Campo	Descrição
Número de Documentos Processados	315.318
Número de Documentos Salvos/Considerados	218.207
Tempo Gasto na Importação	1.172 segundos, 19,53 minutos

Fonte: Elaborado pelo autor, 2022.

Pode-se observar que 97.111 documentos foram desconsiderados, conforme detalhado na seção Metodologia. Considerando-se a importação de apenas um CNPJ, o algoritmo alcançou 186,18 documentos processados por segundo.

Levando em conta a exportação dos dados, a Tabela 14 detalha os resultados obtidos pelo algoritmo.

Tabela 14 – Estatísticas exportação Dados

Descrição	Nr. Registros	Tempo
Produtos	2.693	24 segundos
Clientes	1912	<1 segundo
Fornecedores	241	<1 segundo
Natureza Operação	23	<1 segundo
Dados/relações tributárias IPI	2	5 segundos
Dados/relações tributárias PIS/Cofins	7	5 segundos

Dados/relações tributárias ICMS	10	10 segundos
Dados/relações Associações Tributárias	44	<1 segundo

Fonte: Elaborado pelo autor, 2022.

Podemos observar que o algoritmo obteve excelentes resultados temporais tanto na importação dos documentos das notas fiscais quanto na exportação de dados.

Adicionalmente, foi desenvolvida uma interface para busca e atualização dos dados cadastrais das pessoas, emissores e destinatários, dos documentos fiscais junto à Receita Federal (RECEITA FEDERAL DO BRASIL, 2022b). Nessa interface, o sistema busca, utilizando um certificado digital e em determinado endereço da Sefaz, os dados de cada pessoa presente no banco de dados. Dessa forma, os dados das pessoas ficam atualizados, conforme o cadastro atual desses contribuintes na Receita Federal. Essa interface é apresentada no Apêndice B.

CONSIDERAÇÕES FINAIS

Este produto poderá auxiliar em análises tributárias realizadas pelas empresas, informando quais parametrizações tributárias utilizam e em quais cenários. Por meio da análise dessas informações, torna-se possível verificar parametrizações incorretas, com possíveis perdas tributárias, como o pagamento de impostos em operações que não seriam devidas.

Outro enfoque importante desse algoritmo é a possibilidade de as empresas utilizarem esses dados na migração e cadastro em outros softwares, por exemplo, quando a empresa decide trocar o software responsável pela emissão das notas fiscais. Portanto, esse algoritmo diminui o tempo gasto nessas tarefas, analisando e processando os arquivos XML das próprias notas fiscais emitidas ou recebidas pela empresa. Como se trata de informações oficiais, que foram, inclusive, validadas pela Receita Federal e Estadual, esses dados têm maior coerção e garantem maior assertividade ao processo do que propriamente em realizar o cadastro manualmente, quando poderá ocorrer erros de digitação. Obter esses dados diretamente do outro *software* pode ser dispendioso, uma vez que ele pode não oferecer uma interface para essa função. Adicionalmente, o *software* pode exportar todos os dados, inclusive aqueles que não são mais utilizados e estão inativos, gerando, portanto, informações que não terão

utilidade e deverão poluir o novo *software* se não forem devidamente tratadas.

Como trabalho futuro, podem ser adicionadas mais informações, se implementadas no Produto 2 outras fontes de dados, como os arquivos SPED, em que existem informações que não estão presentes nas notas fiscais, adicionando, desse modo, mais informações aos arquivos.

O algoritmo deste produto técnico está disponível em um repositório na internet, conforme detalhado no Apêndice C.

REFERÊNCIAS

BUZATTO, Jardel; Silveira; SIDNEI, Renato. Implantação de um Sistema ERP: um estudo de caso na empresa Ramos e Copini Autopeças. **Anais do EATI**, p. 108–115, Nov/2016. Disponível em: <<http://2016.eati.info/assets/anais/Longos/108.pdf>>.

RECEITA FEDERAL DO BRASIL: **Portal da Nota Fiscal Eletrônica (NFe)**. 2022b. Disponível em: <https://www.nfe.fazenda.gov.br/>. Acesso em: 3 mar. 2022.

PRODUTO 5 (Técnico/tecnológico): Algoritmo de Inteligência Artificial

INTRODUÇÃO

Neste produto técnico, desenvolvemos um algoritmo de Inteligência Artificial para prever qual o CST ICMS a ser utilizado conforme a operação comercial realizada pelas empresas.

Foram utilizados os dados armazenados no banco de dados criado no Produto Técnico 4. Por se tratar de dados conhecidos e rotulados, empregou-se a técnica de aprendizado supervisionada (NGUYEN *et al.*, 2019; BRUYN *et al.*, 2020).

Este produto foi concebido em duas fases distintas: a primeira foi realizada com menos empresas e, portanto, menos documentos, e na segunda, foram adicionadas mais empresas e documentos fiscais. Esses resultados são apresentados nas seções Resultados Primeira Etapa e Resultados Segunda Etapa. É importante ressaltar que essa adição proporcionou um maior treinamento à Inteligência Artificial e, portanto, melhores resultados, de 89% para 97,2% de acuracidade, o que demonstra sua capacidade de adquirir conhecimento e proporcionar melhores resultados com um número maior de dados de treinamento.

METODOLOGIA

No desenvolvimento deste produto técnico, utilizamos técnicas de Inteligência Artificial com o uso de algoritmos de Árvores de Decisão. Avaliamos e comparamos os resultados de três tipos de algoritmos de Árvore de Decisão: *CART*, *XGBoost* e *RandomForest*. A seção Testes de Algoritmos de Classificação apresenta os resultados de cada algoritmo, em que o *XGBoost* obteve o melhor resultado. Por outro lado, a seção Medições de Performance exibe as fórmulas utilizadas para avaliar o resultado dos algoritmos, principalmente a variável Precisão. Pretendemos, com este algoritmo, encontrar como variável-resposta o CST ICMS, cujos valores que o CST pode assumir estão demonstrados na Tabela 1.

Tabela 1 – Tabela de Valores Variável-Resposta, CST

Código	Descrição
00	Tributada integralmente
10	Tributada e com cobrança do ICMS por

	substituição tributária
20	Com redução de base de cálculo
30	Isenta ou não tributada e com cobrança do ICMS por substituição tributária
40	Isenta
41	Não tributada
50	Suspensão
51	Diferimento
60	ICMS cobrado anteriormente por substituição tributária
70	Com redução de base de cálculo e cobrança do ICMS por substituição tributária
90	Outras

Fonte: Elaborado pelo autor, 2022.

Portanto, a variável-resposta pode assumir 11 valores: 00, 10, 20, 30, 40, 41, 50, 51, 60, 70 e 90. É essa variável, CST (Código Situação Tributária), que determina se a operação será ou não tributada, quais informações o contribuinte deve prestar sobre a operação e como realizar o preenchimento da nota fiscal (RECEITA FEDERAL DO BRASIL, 2022b).

As variáveis de entrada, demonstradas na Tabela 2, são responsáveis por determinar a variável-resposta. Essas variáveis foram escolhidas com base na definição e categorização da operação, relativa ao emissor, destinatário e ao produto, referente ao ICMS.

Tabela 2 – Variáveis de Entrada Utilizadas

Código	Tipo	Origem	Descrição
CNAE Emissor Nota Fiscal	Catagórica	/NFe/infNFe/ide/	Os Estados têm legislações que podem variar conforme o CNAE da empresa, em prol de dar incentivos fiscais a determinados grupos de empresa, por exemplo, o agronegócio. Existem mais de 1000 códigos, compostos de 7 dígitos, que classificam o ramo de operação das empresas.
Tipo de Destinatário: inscrito ou	Catagórica	/NFe/infNFe/dest/indIEDest	Existem, também, diferenças para destinatários inscritos ou não no Estado. Os tipos são: 1 – Contribuinte ICMS; 2 –

não no Estado			Contribuinte Isento de Inscrição; 9 – Não Contribuinte.
Consumidor Final	Catagórica	/NFe/infNFe/ide/indFinal	Indicativo na nota fiscal, informado pelo emissor, indicando se o destinatário irá utilizar o produto (consumidor final) ou se ainda o repassará a terceiros. Os tipos são: 0 – Normal; 1 – Consumidor Final.
CFOP	Catagórica	/NFe/infNFe/det/prod/CFOP	Código que identifica a operação em si. Apesar de todas as operações serem venda, temos variações da venda, se comercializado diretamente pelo fabricante/produtor ou por uma revenda. Existem mais de 600 códigos, compostos de 4 dígitos, que representam cada operação.
NCM	Catagórica	/NFe/infNFe/det/prod/NCM	Código que identifica o tipo de produto. Existem mais de 12000 códigos, compostos de 8 dígitos, para classificar os produtos.
NCM Item	Catagórica	Tabela Códigos NCM Item	Código que identifica o tipo de produto, de forma sintética, responsável por agrupar NCM. Existem mais de 10000 códigos, compostos de 7 dígitos, para classificar os NCM.
Procedência do Produto	Catagórica	/NFe/infNFe/det/imposto/ICMS/ICMSxx/orig	Procedência do produto, relativa à sua origem, se produto de origem nacional ou importado. As procedências são: 0 - Nacional, exceto as indicadas nos códigos 3, 4, 5 e 8; 1 - Estrangeira - Importação direta, exceto a indicada no código 6; 2 - Estrangeira - Adquirida no mercado interno, exceto a indicada no código 7; 3 - Nacional, mercadoria ou bem com Conteúdo de Importação superior a 40% (quarenta por cento) e inferior ou igual a 70% (setenta por cento); 4 - Nacional, cuja produção tenha sido feita em conformidade com os processos produtivos básicos de que tratam o Decreto-Lei nº 288/67, e as Leis nº 8.248/91, 8.387/91, 10.176/01 <i>Add to Citavi project</i> by DOI e 11.484/07; 5 - Nacional, mercadoria ou bem com Conteúdo de Importação inferior ou igual a 40% (quarenta por cento); 6 - Estrangeira - Importação direta, sem similar nacional, constante em lista de Resolução CAMEX e gás natural; 7 - Estrangeira - Adquirida no mercado interno, sem similar nacional, constante

Pessoa Física ou Jurídica	Catagórica	Determinado conforme: /NFe/infNFe/dest/CNPJ ou /NFe/infNFe/ide/CPF	em lista de Resolução CAMEX e gás natural; 8 - Nacional, mercadoria ou bem com Conteúdo de Importação superior a 70% (setenta por cento). Determina se o destinatário da operação é pessoa física ou pessoa jurídica (empresa): 0 – Pessoa física; 1 – Pessoa Jurídica.
Operação Dentro ou Fora do Estado	Catagórica	/NFe/infNFe/ide/idDest	Identifica se a operação é estadual ou interestadual: 0 – Operação dentro do Estado; 1 – Operação fora do Estado; 3 – Operação com exterior.

Fonte: Elaborado pelo autor, 2022.

A utilização destas variáveis na definição da variável-resposta é realizada com base no entendimento da legislação de ICMS de cada Estado por profissionais da área, uma vez que não existem ou não foram encontradas na academia regras claras de quais variáveis definem ou não a classificação fiscal de forma objetiva. É importante ressaltar que, na Tabela 2, estão todas as características utilizadas e que houve variações das características empregadas nas etapas 1 e 2, conforme detalham as seções Resultados Primeira e Segunda etapas. As variáveis são todas do tipo categóricas, utilizadas para categorizar a operação, desde o emissor da nota fiscal até as características do produto e informações do destinatário.

Todas as variáveis utilizadas são do tipo categóricas, sendo códigos que classificam a operação. Na coluna origem, descrevemos o caminho até o local onde a variável é definida no arquivo XML da nota fiscal. Em “descrição”, além de uma breve explanação sobre a variável, também listamos os valores que ela pode assumir.

A etapa de seleção de características (variáveis de entrada) com a utilização de métodos estatísticos de análise não foi necessária. A sua supressão ocorreu conforme foi possível analisar no Manual da Nota Fiscal Eletrônica (RECEITA FEDERAL DO BRASIL, 2022b), assim como o entendimento de especialistas sobre a legislação do ICMS e quais características presentes na Nota Fiscal Eletrônica caracterizam a operação relativa à classificação tributária do CST ICMS. Conforme a seção Resultados, podemos ver que as características selecionadas foram efetivas, com base na própria estatística gerada pelo algoritmo da relevância das características, bem como a precisão alcançada.

Foram considerados apenas documentos com os aspectos demonstrados na Tabela 3. Essa restrição foi devido à avaliação das movimentações comuns, efetuadas pelas empresas

participantes, que, portanto, têm maior número de amostras e perfazem a maioria das notas fiscais. Por outro lado, como a tributação varia conforme a operação (venda, devolução, empréstimos etc.) e conforme o Estado, não necessariamente todas as categorias (variáveis respostas) poderão estar presentes nos dados coletados em número suficiente para treinar o algoritmo de Inteligência Artificial e, assim, inferir os demais tipos de operações.

Tabela 3 – Regras de Seleção

Regras de Seleção	Origem	Valores
1: Somente notas fiscais que representam movimentações de vendas, conforme o código CFOP de cada item na nota fiscal.	/NFe/infNFE/det/prod/CFOP	5101, 5102, 5103, 5104, 5105, 5106, 5113, 5114, 5115, 5116, 5117, 5118, 5119, 5120, 5122, 5123, 5252, 5401, 5402, 5403, 5405, 5551, 5652, 5653, 5655, 5656, 6101, 6102, 6105, 6106, 6107, 6108, 6109, 6110, 6116, 6118, 6119, 6122, 6123, 6252, 6401, 6403, 6404, 6551, 6656, 7101, 7102, 7106.
2: Somente notas fiscais com a Finalidade Normal, determinado na nota fiscal.	/NFe/infNFE/ide/finNFe	1
3: Somente notas fiscais do tipo/operação de Saída, determinado na nota fiscal.	/NFe/infNFE/ide/tpEmis	1
4: Somente notas fiscais cujo emissor seja do Regime Normal, determinado pelo Tipo de Regime Tributário do Emissor da Nota Fiscal.	/NFe/infNFE/emit/CRT	4
5: Somente notas fiscais cujo emissor esteja situado em Minas Gerais, determinado pelo código da UF do emissor da nota fiscal.	/NFe/infNFE/ide/cUF	31

Fonte: Elaborado pelo autor, 2022.

Por meio do Código Fiscal Operações e Prestações (CFOP), foi possível separar as operações que efetivamente são de vendas.

Na nota fiscal, temos também o campo que determina sua finalidade (RECEITA FEDERAL DO BRASIL, 2022b), sendo seus possíveis valores:

- 1 - NF-e normal;
- 2 - NF-e complementar;
- 3 - NF-e de ajuste;
- 4 - Devolução de mercadoria.

Da mesma forma, devido ao volume de dados disponíveis, serão utilizadas apenas as notas fiscais de finalidade 1, ou seja, as notas fiscais cotidianas denominadas como “NFE Normal” (RECEITA FEDERAL DO BRASIL, 2022b). As demais finalidades são casos específicos associados a situações especiais, não sendo vendas em si, apesar de poderem ser associadas diretamente a operações de venda, como no caso de devoluções.

Outra restrição ocorreu na utilização de notas fiscais de “Saída” de produtos das empresas, visto que podemos ter, também, notas de “Entrada” em operações específicas. As operações de venda são, em grande maioria, operações de “Saída”. Essa informação é determinada no arquivo XML, por meio do campo tpNF (RECEITA FEDERAL DO BRASIL, 2022b), onde:

- 0 – Notas de Entrada;
- 1 – Notas de Saída.

Temos, no Brasil, dois grupos de tributação distintos quanto às empresas e contribuintes, que diferem o tipo de tributação do contribuinte de forma exclusiva. Neste estudo, foram utilizados apenas documentos fiscais emitidos por contribuintes do Regime Normal. O campo que informa o regime, o tipo de tributação na NFE é o CRT (RECEITA FEDERAL DO BRASIL, 2022b):

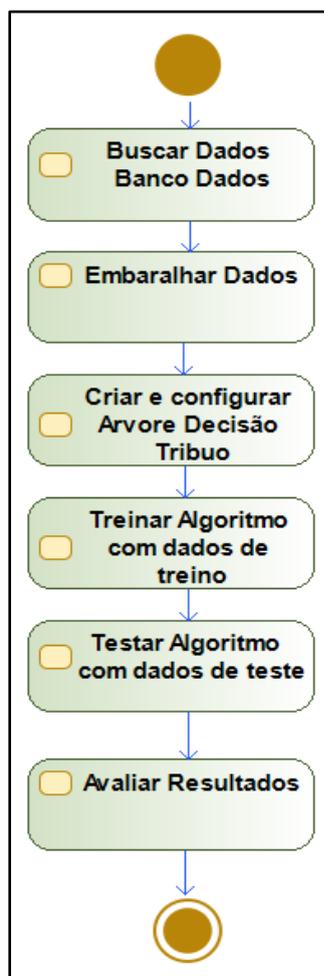
- 1 – Simples Nacional;
- 2 – Simples Nacional, excesso sublimite de receita bruta;
- 3 – Regime Normal.

A última restrição ocorreu na utilização de notas fiscais de empresas sediadas em Minas Gerais, onde se encontra a maioria dos documentos XML disponíveis para serem utilizados, uma vez que as bases de dados empregadas neste estudo são de empresas de Minas Gerais. Por outro lado, o ICMS é um imposto de competência estadual, em que cada Estado define sua legislação. Portanto, as regras e legislação do ICMS de Minas Gerais podem ser diferentes das dos demais estados.

Na implementação do algoritmo, foi escolhida a biblioteca Tribuo, da Oracle. A estrutura do algoritmo é demonstrada na Figura 1 e apresentada no Apêndice A. A base de aprendizado de máquina selecionada foi do tipo árvores de decisão, especificamente o

algoritmo *XGBoost*, cuja escolha foi em virtude do tipo de dados utilizados e pelo fato de as variáveis serem categóricas, tanto as de entrada quanto as de saída desejada, e também pelo problema proposto ser do tipo de aprendizagem supervisionada. A escolha por árvores de decisão se baseia em trabalhos em que tais algoritmos foram utilizados, como no de Abedin *et al.*(2021), Lin *et al.* (2012), Mabe-Madisa (2018) e Lahann *et al.* (2019). Lahann *et al* (2019) utilizaram algoritmos de árvore de decisão em seus trabalhos, o que demonstra que eles são eficientes em problemas de classificação. Adicionalmente, foram realizados testes reais com outros tipos de algoritmos e com a mesma base de dados, para confirmar a melhor eficiência do *XGBoost*. Os resultados desses testes estão detalhados na seção Testes de Algoritmos de Classificação.

Figura 1 – Fluxograma Algoritmo Classificação Tributária



Fonte: Elaborado pelo autor, 2022.

Árvores de decisão são estruturas organizacionais que têm uma estrutura de nós interconectados, os quais, de acordo com a posição, são denominados: nó raiz, ou principal;

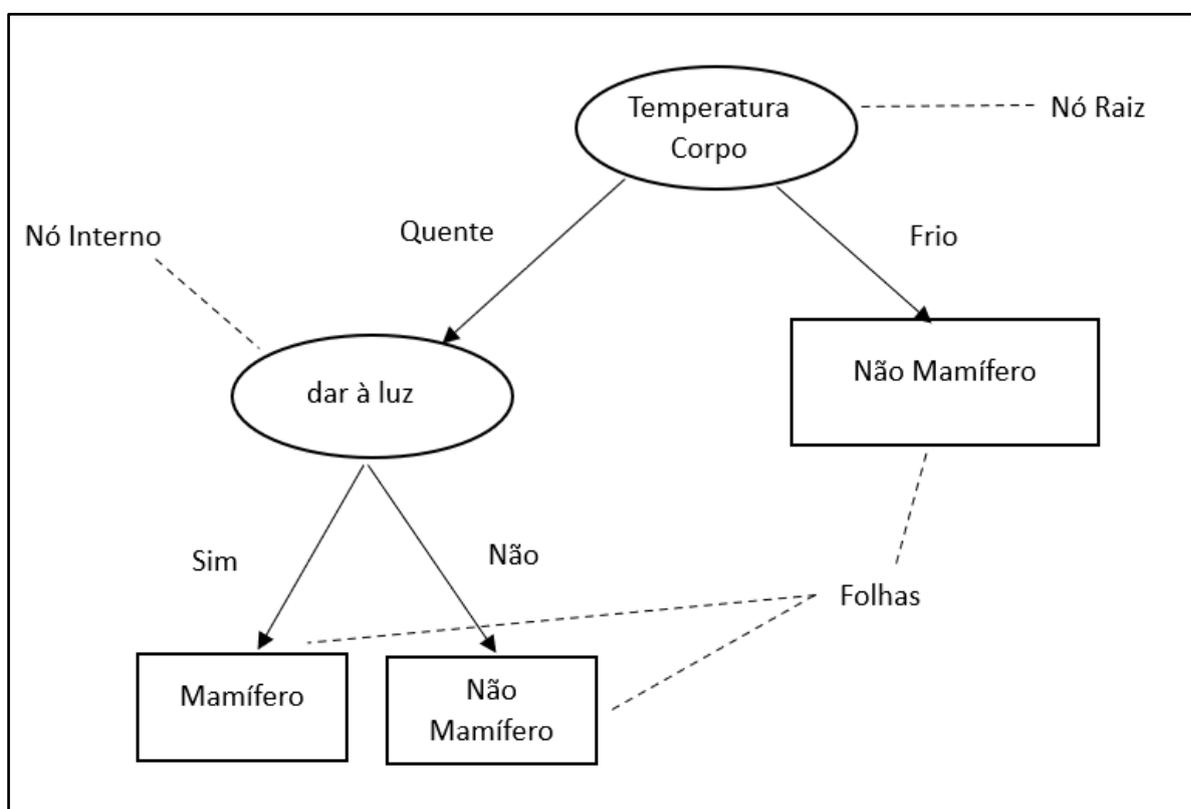
nós intermediários, ou galhos; e nós terminais, ou folhas (ROKACH; MAIMON, 2015).

Cada nó intermediário possui atributos de testes, responsáveis por separar os registros conforme suas características. A Figura 2 ilustra o exemplo de uma árvore de decisão que consiste em avaliar se um animal é ou não um mamífero, onde podemos visualizar as perguntas em cada nó:

1. Temperatura corporal: quente ou frio;
2. Em relação à reprodução: dar à luz ou não.

As folhas são os nós terminais, que são nós de categorização, que, neste exemplo, consistem em apenas dois valores: mamíferos (Mammals) e não mamíferos (Non-mammals).

Figura 2 – Estrutura de um arquivo XML – ICMS



Fonte: Adaptado de TAN *et al.* (2019).

No caso de alguns algoritmos de aprendizado de máquina baseados em estruturas de árvores de decisão, como o *RandomForest*, e o algoritmo utilizado neste estudo, *XGBoost*, são implementados vários conjuntos de árvores, onde são utilizados *scores* ou pontuações que definem cada folha ao invés de valores categóricos. A soma desses *scores* define posteriormente as classificações, de acordo com os maiores valores (WANG *et al.*, 2022).

XGBoost difere-se de outros métodos baseados em árvore de decisão, como do *RandomForest*, no treinamento de seu algoritmo, com a utilização de *Gradient Boosting*. No

Extreme Gradient Boosting, cada preditor e conjunto de árvores são avaliados e, com base em seus erros, são modelados os predecessores, isto é, as próximas árvores. As futuras árvores são selecionadas conforme as maiores pontuações, os maiores *scores*, que apresentaram o melhor desempenho, proporcionando uma melhor performance, acuracidade e velocidade de aprendizagem do algoritmo (WANG *et al.*, 2022).

Medições de Performance

Os resultados do algoritmo de Inteligência Artificial serão medidos, principalmente, pelas fórmulas de Precisão e Acurácia, ou Acuracidade. Também serão apresentadas as medidas de Sensibilidade e F-Score.

A Acuracidade avalia o percentual de acertos realizados pelo algoritmo, por meio da divisão do total de acertos (TA) pelo total de itens (TI) (ROKACH; MAIMON, 2015). É uma métrica simples e muito importante de avaliar o resultado do algoritmo.

$$acuracidade = \frac{TA}{TI} \quad (1)$$

A medida de precisão avalia a quantidade de verdadeiros positivos (VP) sobre a soma de todos os valores positivos, a soma dos verdadeiros positivos e os falsos positivos (FP) (ROKACH; MAIMON, 2015). Portanto, avalia quantos exemplos classificados como positivos são, realmente, positivos.

$$precisão = \frac{VP}{(VP + FP)} \quad (2)$$

A Sensibilidade (Recall) avalia a capacidade do algoritmo em detectar com sucesso os resultados classificados como positivos (ROKACH; MAIMON, 2015).

$$sensibilidade = \frac{VP}{(VP + FN)} \quad (3)$$

O F1, F-Measure ou F-Score é uma medida harmônica calculada com base na Precisão e pela Sensibilidade ou Recall do algoritmo. O F1 tem como objetivo medir a confiabilidade da Acuracidade (ROKACH; MAIMON, 2015).

$$f - measure = \frac{2 * (precisão * sensibilidade)}{(precisão + sensibilidade)} \quad (4)$$

Será utilizado o método de Validação Cruzada (*Cross Validation*), por meio do K-Fold, para avaliar o desempenho do algoritmo. A Validação Cruzada consiste em particionar os dados em conjuntos (partes), em que um conjunto é utilizado para treino, e outro, para teste e avaliação do desempenho do modelo. Nesse método, todos os dados serão utilizados para testar o algoritmo e avaliar o seu desempenho (ROKACH; MAIMON, 2015).

Testes de Algoritmos de Classificação

Executaram-se testes nos algoritmos do tipo árvores de decisão disponibilizados na biblioteca Tribuo. O intuito foi verificar qual algoritmo proporcionaria os melhores resultados, para determinar qual o melhor a ser utilizado.

A Tabela 4 detalha esses algoritmos de forma resumida, os quais foram testados, inicialmente, com os parâmetros padrões, sugeridos pela Tribuo.

Tabela 4 – Algoritmos testados

Código	Descrição	Tempo Treinamento	Precisão
CART / CARTClassificationTrainer	Algoritmo baseado em árvores de decisão J. Friedman, T. Hastie, & R. Tibshirani. "The Elements of Statistical Learning". Springer 2001	0,31 minutos	96,90%
XGBoost	Árvores de decisão Friedman JH. "Greedy Function Approximation: a Gradient Boosting Machine". Annals of statistics, 2001.	2.42 minutos	97,20%
RandomForest	Árvores de decisão J. Friedman, T. Hastie, & R. Tibshirani. "The Elements of Statistical Learning". Springer 2001.	50.07 minutos	96,90%

Fonte: Elaborado pelo autor, 2022.

Portanto, *XGBoost* obteve os melhores índices de precisão, além de ser um dos algoritmos mais rápidos na realização do treinamento, pois gastou 2,42 minutos para efetuar o seu treinamento e testes sobre um conjunto de 2.839.599 amostras. Por outro lado, o algoritmo CART foi o mais rápido de todos, levando menos de 1 minuto para executar o treinamento.

RESULTADOS

O algoritmo baseado em árvores de decisão, *XGBoost*, foi o escolhido para ser utilizado neste estudo, de acordo com os testes realizados, pois obteve a melhor eficiência e um bom tempo de processamento.

Inicialmente, os dados foram separados em dois grupos: treinamento e testes. Para o treinamento, foram utilizados 70% dos dados, e, para testes, 30%, para avaliar a capacidade de aprendizado da Inteligência Artificial.

Em um primeiro momento, em caráter inicial e preliminar, os testes foram executados em uma base de dados menor, cuja eficiência global foi 89%. Posteriormente, o algoritmo foi aprimorado, tunando parâmetros e, principalmente, adicionando mais dados de treinamento para o algoritmo de Inteligência Artificial, quando a eficiência global atingiu 97,2%. As seções de Resultado Primeira e Segunda Etapa detalham esses resultados.

Foram criadas interfaces gráficas no *software*, descritas no Apêndice B, para facilitar a utilização e visualização das etapas de treinamento e testes da Inteligência Artificial, sendo que, na interface de treinamento, é possível alterar os parâmetros de população da base de dados e de treinamento da Inteligência Artificial.

Primeira Etapa

Nesta primeira etapa, trabalhamos com um menor número de dados e empresas, em comparação com a segunda etapa. O objetivo foi simular os parâmetros e a implementação do próprio algoritmo, a fim de verificar a sua aplicabilidade no problema, em definir a variável-resposta. Os parâmetros utilizados no algoritmo, nesta etapa, são apresentados na Tabela 5.

Tabela 5 – Parâmetros Utilizados

Código	Descrição	Valores
Número Árvores	Número de árvores a serem utilizadas no processo de treinamento.	25
colsample_bytree	Parâmetro utilizado para amostrar, de forma aleatórias, as variáveis usadas para treinar cada árvore. O objetivo é evitar <i>overfitting</i> e descorrelacionar as árvores.	1
tree_method	Tipo de algoritmo de construção da árvore utilizada. O valor Auto especifica para ser utilizado heurística para escolher o método mais rápido.	Auto
Seed	Valor inicial utilizado para embaralhar os dados.	12345
num_class	Número de Atributos	8
max_depth	Profundidade máxima da árvore. É utilizado para evitar <i>overfitting</i> do modelo.	6
Booster	Algoritmo <i>booster</i> utilizado. O <i>Gbtree</i> é um modelo baseado em árvore.	Gbtree
Objective	Especifica o tipo de aprendizagem utilizado. O multi:softprob é especial para problemas de classificação multiclasse.	multi:softprob
Lambda	Parâmetro correspondente a taxas de regularização que determina os pesos nas folhas da árvore. O objetivo é evitar <i>overfitting</i> , e, quanto maior o valor, mais forte é a regularização.	1.0
Eta	Parâmetro utilizado para ajustar a taxa de aprendizado. Valores menores deixam o modelo mais robusto; em contrapartida, aumentam o tempo de treinamento.	0.3
Alpha	Parâmetro utilizado para regularização dos pesos e evitar <i>overfitting</i> .	0.0
Subsample	Proporção de subamostra das instâncias de treinamento.	1.0
Gamma	Parâmetro utilizado para definir a redução mínima da perda necessária para realizar uma divisão adicional em um nó da árvore.	0.0
min_child_weight	Parâmetro para regularização e limitação da profundidade da árvore.	1.0

Verbosity	Opção para impressão de mensagens de construção do algoritmo.	0
-----------	---	---

Fonte: Elaborado pelo autor, 2022.

Foi definida a utilização de 25 árvores, com uma profundidade máxima de seis níveis como parâmetros que melhoram os resultados.

Após a execução do algoritmo, obtiveram-se bons resultados, com uma precisão global de 89%. Além disso, foram identificadas as características que mais contribuíram para determinar corretamente a variável-resposta, conforme demonstrado na Tabela 6.

Tabela 6 – Variáveis de Relevância

Código	Descrição
NCM	2671
CNAE Emissor	885
CFOP	278
Consumidor Final	203
Inscrito Estado	135
Procedência	40

Fonte: Elaborado pelo autor, 2022.

Podemos observar que o NCM que classifica os produtos foi o mais relevante em determinar o CST do ICMS na operação, seguido pelo tipo de empresa definido pelo código CNAE. Estes códigos seguem, respectivamente, o CFOP que define a operação, ou seja, se é consumidor final ou não, e se o destinatário tem Inscrição Estadual. Por último, com menor relevância, temos a Procedência do Produto, que define sua origem, por exemplo, importado ou nacional. Na Tabela 7, os resultados de cada classificação e demais estatísticas são apresentados.

Tabela 7 – Resultados do Algoritmo de Inteligência Artificial

Class CST's ICMS	N Número de amostras	Tp Verdadeiros Positivos	Fn Falsos negativos	Fp Falsos Positivos	Recall Sensibilidade	Prec Precisão	F1 F-Score
00	53.208	47.704	5.504	2.332	0,897	0,953	0,924
20	28.914	23.884	5.030	5.593	0,826	0,810	0,818

40	109.383	93.018	16.365	6.560	0,85	0,934	0,890
41	1	0	1	0	0,000	0,000	0,000
50	1	0	1	0	0,000	0,000	0,000
51	1	0	1	0	0,000	0,000	0,000
60	1.476	850	626	44	0,576	0,951	0,717
90	3	2	1	0	0,667	1	0,800
TOTAL	260.188	231.548	28.640	28.640			
Accuracy Acuracidade					0,890		
Micro Average Micromédia					0,890	0,890	0,890
Marco Average Média Macro					0,600	0,684	0,631
Balanced Error Rate Taxa Equilíbrio Erro					0,400		

Fonte: Elaborado pelo autor, 2022.

Após introduzidos os conceitos e fórmulas de análise do resultado do algoritmo e apresentados os resultados, podemos observar que o algoritmo teve melhores resultados com os CSTs: 00, com 95,3% de eficiência; seguidos pelos CSTs 60, com 95,1%; 40, com 93,4%; 20, com 82,6%; e, por último, o CST 20, com 81% de eficiência. Os demais CSTs não tiveram medições realizadas, sendo que, pela quantidade pequena de amostras, demonstram possíveis erros de classificação efetuados pelas empresas.

A existência de CSTs com poucas amostras ou que não existem na massa de dados utilizadas, como o CST 10, 30 e 70, pode ser explicada por: operações equivocadas, como citado anteriormente; operações muito específicas; operações e regimes especiais; operações não utilizadas pelas empresas participantes deste estudo; ou, ainda, situações não previstas na legislação de ICMS de Minas Gerais em operações de Venda.

A Tabela 8 exibe a tabela da Matrix de Confusão, ou Matrix de Erro. A Matrix de Confusão demonstra as predições corretas e incorretas, sendo uma forma de avaliar a performance do algoritmo (TAN *et al.*, 2019). Se observarmos a linha que representa o CST 00 e coluna 00, temos o valor de 47.704 amostras que foram classificadas corretamente. Por outro lado, se observarmos a linha 00 e a coluna 50, vemos 2.704 amostras que deveriam ser classificadas como 00, mas foram classificadas erroneamente como CST 50.

Tabela 8 – Matriz de Confusão

	00	20	40	41	50	51	60	90
00	47.704	0	38	0	2.704	646	0	2.116
20	506	0	1	0	2892	1631	0	23.884
40	1.494	0	4	0	93.018	11.418	0	3.449
41	1	0	0	0	0	0	0	0
50	1	0	0	0	0	0	0	0
51	133	0	1	0	920	66.090	0	28
60	197	0	850	0	13	416	0	0
90	0	2	0	0	1	0	0	0

Fonte: Elaborado pelo autor, 2022.

Segunda Etapa

Na segunda etapa de testes, procuramos, principalmente, aumentar o número de amostras utilizadas para que fosse possível verificar os ganhos no resultado do algoritmo com uma maior disponibilidade de dados. Desse modo, algumas alterações foram realizadas nos parâmetros, contribuindo, assim, para resultados melhores. Os parâmetros utilizados nessa etapa são apresentados na Tabela 9.

Tabela 9 – Parâmetros Utilizados

Código	Descrição	
Número Árvores	Número de árvores a serem utilizadas no processo de treinamento.	50
colsample_bytree	Parâmetro utilizado para amostrar, de forma aleatórias, as variáveis usadas para treinar cada árvore. O objetivo é evitar <i>overfitting</i> e descorrelacionar as árvores.	1
tree_method	Tipo de algoritmo de construção da árvore utilizada. O valor Auto especifica para ser utilizado heurística para escolher o método mais rápido.	exact
Seed	Valor inicial utilizado para embaralhar os dados.	12345
num_class	Número de Atributos.	7

max_depth	Profundidade máxima da árvore. É utilizado para evitar <i>overfitting</i> do modelo.	2147483647
Booster	Algoritmo <i>booster</i> utilizado. O Gbtree é um modelo baseado em árvore.	Gbtree
Objective	Especifica o tipo de aprendizagem utilizado. O multi:softprob é especial para problemas de classificação multiclasse.	multi:softprob
Lambda	Parâmetro correspondente a taxas de regularização que determina os pesos nas folhas da árvore. O objetivo é evitar <i>overfitting</i> , e, quanto maior o valor, mais forte é a regularização.	1.0
Eta	Parâmetro utilizado para ajustar a taxa de aprendizado. Valores menores deixam o modelo mais robusto; em contrapartida, aumentam o tempo de treinamento.	0.3
Alpha	Parâmetro utilizado para regularização dos pesos e evitar <i>overfitting</i> .	0.0
Subsample	Proporção de subamostra das instâncias de treinamento.	0.5
Gamma	Parâmetro utilizado para definir a redução mínima da perda necessária para realizar uma divisão adicional em um nó da árvore.	0.5
min_child_weight	Parâmetro para regularização e limitação da profundidade da árvore.	5.0
Verbosity	Opção para impressão de mensagens de construção do algoritmo.	1

Fonte: Elaborado pelo autor, 2022.

Além do número de amostras utilizadas, alguns parâmetros também foram alterados: o número de árvores, de 25 para 50; o tipo de árvore, de auto para *exact*; profundidade (max_depth), de 6 para o máximo permitido; subsample, de 1 para 0.5; gamma, de 0 para 0.5; e min_child_weight, de 1 para 5. A propriedade *verbosity* também foi ativada, para detalhar apenas avisos/log de relevância.

A Tabela 10 detalha o ranking das variáveis de entrada utilizadas.

Tabela 10 – Variáveis de Relevância

Código	Descrição
NCM	16.519

CNAE Emissor	4.092
CFOP	1.637
Pessoa Física ou Jurídica (Destinatário)	1.039
Consumidor Final (Destinatário)	886
Inscrito Estado (Destinatário)	847
Operação Dentro ou Fora do Estado	543
NCM Item	183
Procedência do Produto	96

Fonte: Elaborado pelo autor, 2022.

Nessa etapa, foram adicionadas e testadas outras variáveis de entrada. Porém, destacamos aqui somente as que tiveram impacto na definição da variável-resposta, sendo que NCM, CNAE do Emissor e CFOP continuaram como as primeiras em defini-la. Pessoa Física ou Jurídica, Operação Dentro ou Fora do Estado e NCM Item correspondem às novas variáveis de entrada adicionadas, que também impactaram na definição da variável-resposta, além das que já existiam: Consumidor Final, Inscrito no Estado e Procedência do Produto.

A Tabela 11 detalha os resultados de cada variável-resposta. Podemos observar que a classe 00 é majoritária, correspondendo a 71% do número de amostras. Nessa etapa, foram desconsideradas amostras de classes/CST com poucas movimentações, e os CST que tiveram menos de 1% de amostragem foram descartados.

Tabela 11 – Resultados do Algoritmo de Inteligência Artificial

Class CST's ICMS	N Número de amostras	Tp Verdadeiros Positivos	Fn Falsos negativos	Fp Falsos Positivos	Recall Sensibilidade	Prec Precisão	F1 F-Score
00	2.023.314	2.004.861	18.453	12.817	0,991	0,994	0,992
10	147.720	147.677	43	157	1	0,999	0,999
20	73.283	63.583	9.700	19.409	0,868	0,766	0,814
40	318.757	272.376	46.381	13.990	0,854	0,951	0,9
51	176.727	173.371	3.356	33.036	0,981	0,84	0,905
60	79.578	77.940	1.638	162	0,979	0,998	0,989
70	20.220	20.213	7	7	1	1	1

TOTAL	2.839.599	2.760.021	79.578	79.578			
Accuracy Acuracidade					0,972		
Micro Average Micromédia					0,972	0,972	0,972
Marco Average Média Macro					0,953	0,935	0,943
Balanced Error Rate Taxa Equilíbrio Erro					0,047		

Fonte: Elaborado pelo autor, 2022.

As variáveis-respostas que tiveram maiores precisões foram as CST 70, 10, 60, 00 e 40. Por outro lado, as CST 20 e 51 atingiram os piores resultados. Porém, o algoritmo alcançou 97,2% de precisão de acerto, o que demonstra a sua efetividade em prever a variável-resposta.

A matriz de confusão é detalhada na Tabela 12 e demonstra os acertos e onde ocorreram os erros nas previsões de cada variável-resposta.

Tabela 12 – Matriz de Confusão

	00	10	20	40	51	60	70
00	2.004.861	1	6.867	8.939	2.526	120	0
10	4	147.677	0	28	0	4	7
20	3.841	0	63.583	2.742	3.108	8	0
40	7.576	2	12.341	272.376	26.437	25	0
51	933	0	200	2.222	173.371	1	0
60	462	152	0	59	965	77.940	0
70	0	2	1	0	0	4	20.213

Fonte: Elaborado pelo autor, 2022.

A variável-resposta CST 20, que revelou a pior precisão, teve uma distribuição de erros para os CST 00, 40 e 51, com, respectivamente, 5,24%, 3,74% e 4,24%, pelo total de amostras 73.283. Ou seja, não houve tendência para uma única classe em especial.

Foi realizado o processo de Validação Cruzada (*Cross Validation*), conforme detalhado na Tabela 13. O método empregado foi K-Fold, que consiste em dividir os dados em K conjuntos, de forma que cada subconjunto será utilizado para teste em algum momento

da avaliação do modelo.

Tabela 13 – Resultado do *Cross Validation*

Class CST's ICMS	Tp Verdadeiros Positivos	Tn Verdadeiros Negativos	Fn Falsos negativos	Fp Falsos Positivos	Recall Sensibilidade	Prec Precisão	Prec Precisão Média	F1 F-Score	AUCROC
00	286178,30	115088,60	2644,10	1745,90	0,99085	0,99394	0,99974	0,99239	0,99938
10	21172,40	384458,50	6,90	19,10	0,99967	0,99910	0,99998	0,99939	1,00000
20	8618,80	392931,50	1862,30	2244,30	0,82233	0,79342	0,90116	0,80759	0,99693
40	39576,70	357495,50	6063,80	2520,90	0,86714	0,94012	0,97200	0,90216	0,99627
51	24811,10	375653,00	466,40	4726,40	0,98155	0,83999	0,95170	0,90527	0,99775
60	11141,40	394266,30	231,30	17,90	0,97966	0,99840	0,99640	0,98894	0,99991
70	2882,80	402772,60	0,60	0,90	0,99979	0,99969	0,99996	0,99974	1,00000
Micro Average Micromédia	394381,50	2422666,00	11275,40	11275,40	0,97220	0,97220		0,94221	
Macro Average Média Macro	56340,21	346095,14	1610,77	1610,77	0,94871	0,94871		0,97220	
Balanced Error Rate Taxa Equilíbrio Erro						0,05129			

Fonte: Elaborado pelo autor, 2022.

Portanto, os dados foram divididos em 10 subconjuntos, ou *folds*. Os resultados se mantiveram conforme os testes iniciais, com baixas taxas de erro, demonstrando a efetividade do aprendizado da inteligência artificial.

CONSIDERAÇÕES FINAIS

Este foi o principal produto produzido por meio deste trabalho. O algoritmo desenvolvido neste produto técnico se revela promissor em auxiliar os profissionais da área tributária na classificação das operações comerciais em relação ao CST ICMS, conforme detalhado na seção Resultados.

Um dos principais desafios durante o desenvolvimento deste produto foi a escolha do algoritmo de Inteligência Artificial que melhor se adequasse em resolver o problema proposto. Durante a etapa de fundamentação teórica, detectamos que algoritmos do tipo árvore de decisão foram os mais utilizados em problemas similares. Desse modo, foram

executados testes sobre os algoritmos baseados em árvore de decisão implementados pela biblioteca escolhida para este estudo, a *Tribuo*. No *benchmark* realizado, as bibliotecas obtiveram resultados semelhantes, mas o *XGBoost* se destacou com o melhor resultado.

Além do próprio resultado do algoritmo, foi possível verificar outro ponto interessante nos resultados, sendo o principal a verificação de quais características realmente definem o CST ICMS, ou seja, quais são mais relevantes em classificar a operação.

Uma das limitações deste produto é em relação aos dados utilizados no treinamento da Inteligência Artificial, em que foram preteridas as operações de maior ocorrência: operações de vendas e empresas sediadas em Minas Gerais.

Constitui-se, também, uma limitação o número de empresas utilizadas para treinamento da Inteligência Artificial. Como a tributação varia por ramo de atividade, o CNAE da empresa seria importante para a adição de empresas com os mais variados CNAE, o que não seria possível nesse momento, considerando-se as dezenas de classificações existentes.

O algoritmo também se limitou em prever somente qual o CST ICMS a ser utilizado. Nas operações comerciais, existem outros campos que devem ser informados, como a Alíquota, Modalidade Base de Cálculo e Motivo Desoneração ICMS.

Como trabalhos futuros, sugerimos a adição de mais empresas, de variados ramos de atividades, para aumentar a acuracidade do algoritmo, bem como a adição de outras operações, e não somente as de vendas. Recomendamos, também, a adição preditiva das demais características do ICMS, como alíquota, modalidade de base de cálculo e motivo de desoneração, bem como a extensão na previsão das características de outros tributos, como o IPI, PIS e Cofins. Por fim, sugerimos adicionar uma interface de interação e *feedback* dos usuários, para que a Inteligência Artificial possa se aprimorar continuamente com a experiência e resposta dos próprios usuários.

Portanto, este produto se mostra promissor no auxílio aos profissionais da área de tributos na definição do CST ICMS, reduzindo o tempo dispendido nessa operação e trazendo maior assertividade ao processo.

O algoritmo deste produto técnico está disponível em um repositório na internet, conforme exposto no Apêndice C.

REFERÊNCIAS

BRUYN, Arnaud de; VISWANATHAN, Vijay; BEH, Yean Shan; BROCK, Jürgen Kai-Uwe; WANGENHEIM, Florian von. Artificial Intelligence and Marketing: pitfalls and

opportunities. **Journal of Interactive Marketing**, 51, p. 91–105, 2020. DOI: 10.1016/j.intmar.2020.04.007.

NGUYEN, Giang; DLUGOLINSKY, Stefan; BOBÁK, Martin; TRAN, Viet; GARCÍA, Álvaro López; HEREDIA, Ignacio; MALÍK, Peter; HLUCHÝ, Ladislav. Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. **Artif Intell Rev** 52 (1), p. 77–124, 2019. DOI: 10.1007/s10462-018-09679-z.

RECEITA FEDERAL DO BRASIL: **Portal da Nota Fiscal Eletrônica (NFe)**. 2022b. Disponível em: <https://www.nfe.fazenda.gov.br/>. Acesso em: 3 mar. 2022.

ROKACH, Lior; MAIMON, Oded. Data mining with decision trees. Theory and applications. 2nd ed. Hackensack, **New Jersey**: World Scientific (Series in machine perception and artificial intelligence, vol. 81, 2015).

TAN, Pang-Ning; STEINBACH, Michael; KARPATNE, Anuj; KUMAR, Vipin. **Introduction to data mining. Second edition**. NY: Pearsonm 2019.

WANG, Kui; LI, Meixuan; CHENG, Jingyi; ZHOU, Xiaomeng; LI, Gang. Research on personal credit risk evaluation based on XGBoost. **Procedia Computer Science**, v. 199, p. 1128–1135, 2022. DOI:10.1016/j.procs.2022.01.143.

APÊNDICE A – Biblioteca Tribuo

A biblioteca Tribuo, da Oracle, disponível no site <https://tribuo.org>, é de código aberto, gratuita e escrita em Java. Tem diversas ferramentas para desenvolvimento de Algoritmos e Sistemas de Aprendizado de Máquina, como ferramentas para classificação, multiclassificação, regressão, *clustering* e detecção de anomalias, como demonstrado em seu site, na Figura A1.

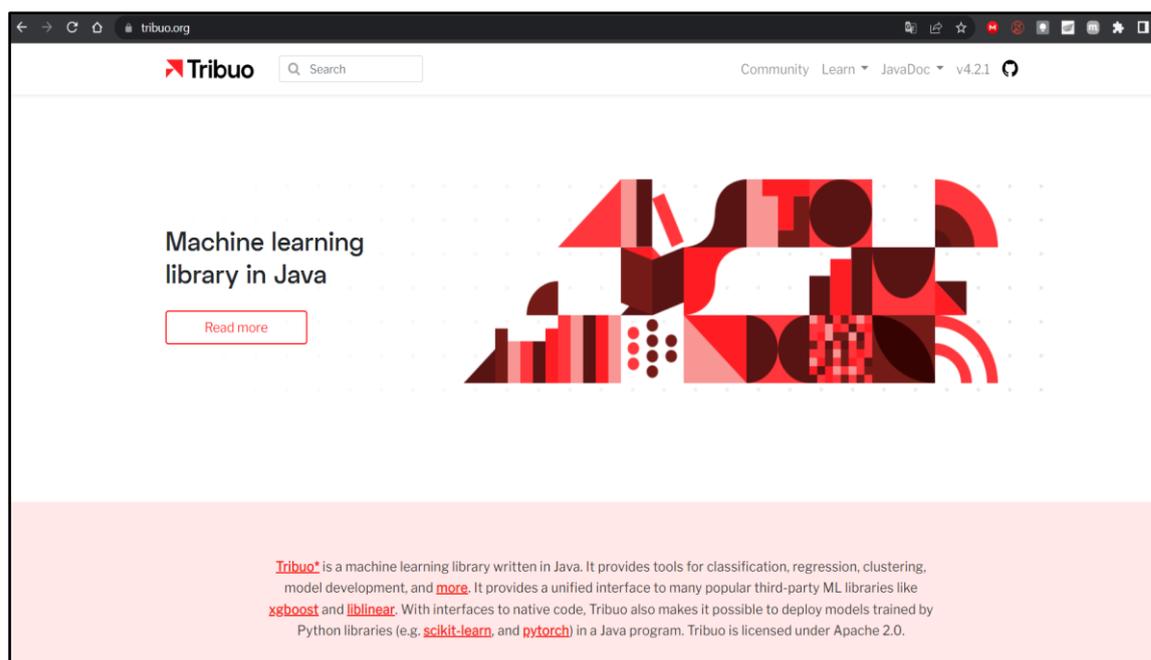
Suas principais características são a simplicidade de configuração e de uso, além das citadas em seu site: *Provenance*, *Typesafe* e *Interoperable*. Tem ampla documentação, com tutoriais e exemplos, principalmente em seu repositório no *Github*: <https://github.com/oracle/tribuo>.

Provenance, ou Proveniência, diz respeito à sua capacidade de rastreabilidade em verificar exatamente quais modelos, conjuntos de dados e avaliadores foram utilizados, bem como quais parâmetros e transformações foram realizados. Essa característica é importante, principalmente, para reconstruir modelos já criados.

Typesafe é a capacidade da biblioteca em evitar erros em produção/execução, característica inerente ao próprio Java, em que seu core foi desenvolvido, de modo a ser fortemente tipado. Dessa forma, cada modelo sabe as saídas que produz, quais entradas espera e quais componentes estão envolvidos no algoritmo.

Interoperable é a característica interoperabilidade na Tribuo e outras bibliotecas de Aprendizado de Máquina, inclusive de outras linguagens, em que a Tribuo pode tanto exportar quanto importar modelos de informações, por exemplo, usando arquivos ONNX.

Figura A1 – Algoritmo de Inteligência Artificial

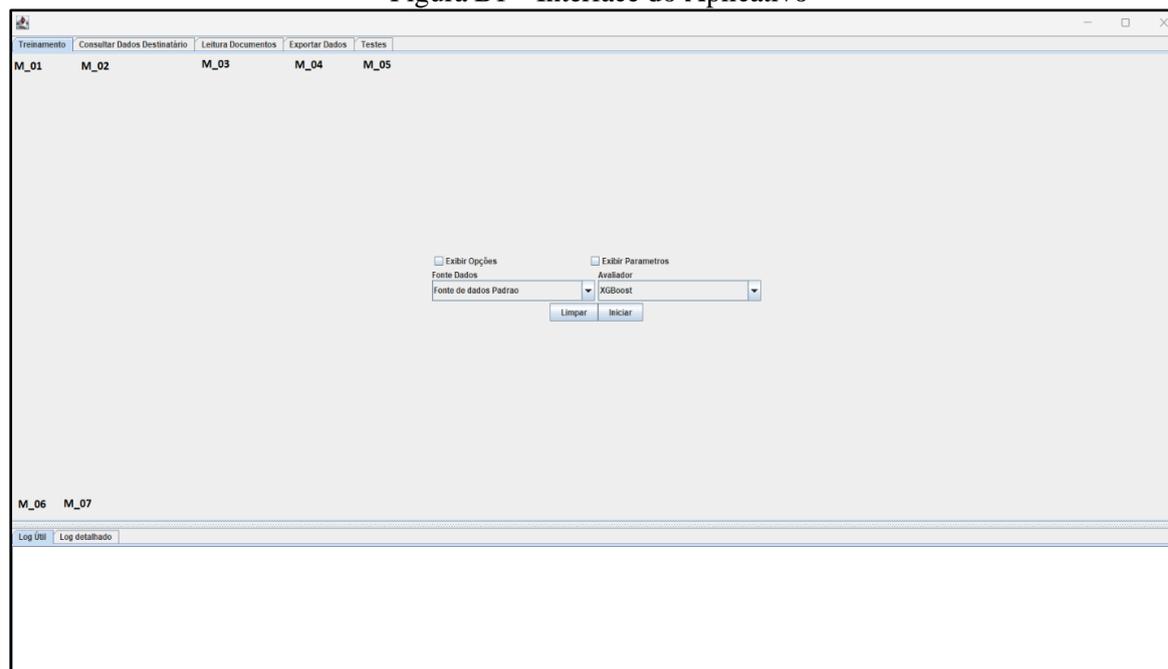


Fonte: Elaborado pelo autor, 2022.

APÊNDICE B – Interface da Aplicação

Os produtos técnicos descritos na parte 2 deste trabalho foram disponibilizados dentro de um único aplicativo. Para facilitar as interações junto ao usuário, foi desenvolvida uma interface gráfica com os menus necessários para operar cada produto. A Figura B1 exhibe a interface gráfica do aplicativo.

Figura B1 – Interface do Aplicativo



Fonte: Elaborado pelo autor, 2022.

As funcionalidades dos menus são detalhadas na Tabela B1. Cada menu de interação está denominado coma sigla ‘M’, seguida de um número sequencial.

Tabela B1 – Interface do Aplicativo e Funções

Código	Descrição
M_01	Menu com opções para busca dos dados e para treinamento da Inteligência Artificial
M_02	Menu para consulta e atualização de dados das pessoas envolvidas no documento fiscal, Emissores e Destinatários na Receita Estadual
M_03	Menu para leitura, processamento e armazenamento de dados dos documentos XML (notas fiscais)
M_04	Menu para exportação de dados
M_05	Menu para testes da aplicação

M_06	Área de texto (Log) em que são exibidas as informações importantes do aplicativo
M_07	Área de texto (Log) em que são exibidas as informações detalhadas do aplicativo

Fonte: Elaborado pelo autor, 2022.

A Figura B2 detalha os campos de interação e configurações que o usuário pode realizar sobre a etapa de treinamento da Inteligência Artificial.

Figura B2 – Interface do Menu de Treinamento

The screenshot shows a web application interface for training. It features a top navigation bar with tabs: 'Treinamento', 'Consultar Dados Destinatário', 'Leitura Documentos', and 'Exportar Dados'. The main content area is divided into two sections: 'Exibir Opções' (left) and 'Exibir Parametros' (right). The 'Exibir Opções' section contains several checkboxes: F_01 (Salvar Dados de Teste), F_02 (Salvar Dados de Teste, somente com erro), F_03 (Salvar Modelo), F_04 (Realizar Cross Validation), and F_05 (Gerar Arquivo Dados). Below these are input fields for 'Perc. dados Teste' (F_06) and 'Remover Dados, baixo volume %' (F_07). The 'Exibir Parametros' section contains dropdown menus for 'Finalidade Nfe' (F_08), 'Tipo Nfe' (F_10), 'Unidade Federativa' (F_12), 'Tipo Tributacao' (F_09), and 'Classificação CFOP' (F_11). At the bottom, there are dropdowns for 'Fonte Dados' (F_13) and 'Avaliador' (F_14), along with 'Limpar' (F_15) and 'Iniciar' (F_16) buttons.

Fonte: Elaborado pelo autor, 2022.

As funcionalidades de cada campo são detalhadas na Tabela B2. Cada campo é denotado com a sigla F e um número sequencial.

Tabela B2 – Interface do Menu de Treinamento

Código	Descrição
F_01	Opção que determina se o aplicativo deve salvar os dados de testes, realizados pela IA
F_02	Opção que determina para salvar somente os dados de Testes em que a IA errar, ou seja, quando o valor previsto pela IA for diferente do valor da variável-resposta, conforme as demais variáveis
F_03	Determina se deve ser salvo o arquivo resultado/modelo, de treinamento da IA, para ser usado em previsões. O arquivo é salvo dentro do diretório da aplicação. Esse arquivo contém o conhecimento adquirido pela IA, que poderá ser utilizado para novas previsões.
F_04	Realizar validação cruzada dos dados (<i>Cross Validation</i>)
F_05	Opção para gerar arquivo de dados, fonte de dados a ser utilizada no

	treinamento da IA. Caso desmarcada, a aplicação irá usar o último arquivo gerado
F_06	Percentual de dados a ser utilizado para testar a IA
F_07	Percentual de dados limite em que, caso a ocorrência dos registros de determinado valor da variável-resposta for menor que esse percentual, os registros referentes a essa variável-resposta serão excluídos da fonte de dados. Usando essa opção, é possível eliminar antecipadamente ruídos dos dados, ou seja, variáveis-respostas com poucas ocorrências
F_08	Finalidade dos documentos da NFE: Venda, Complementar, Ajuste e Devolução
F_09	Tipo de Tributação do emissor da NFE: simples, simples com excesso de receita e normal
F_10	Tipo de saída informado nos documentos da NFE: Saída e Entrada Classificação dos CFOP utilizados nos itens da NFE:
F_11	CFOP Entrada: Compras; CFOP Entrada: Empréstimos, Comodato; CFOP Entrada: Devoluções Vendas; CFOP Entrada: Retorno Mercadorias; CFOP Entrada: Transferências; CFOP Saída: Vendas; CFOP Saída: Remessas; CFOP Saída: Devoluções; CFOP Saída: Outras; CFOP Saída: Transferências Outras Filiais
F_12	Unidade Federativa do Emissor da NFE: Acre; Alagoas; Amazonas; Amapá; Bahia; Ceará; Distrito Federal; Espírito Santo; Goiás; Maranhão; Minas Gerais; Mato Grosso do Sul; Mato Grosso; Pará; Paraíba; Pernambuco; Piauí; Paraná; Rio de Janeiro; Rio Grande do Norte; Rondônia; Roraima; Rio Grande do Sul; Santa Catarina; Sergipe; São Paulo; Tocantins, EX (Exportação/Exterior).
F_13	Fonte de dados para treinamento da IA. Atualmente, existe apenas uma fonte de dados, que é a própria base de dados gerada pela Aplicação ao importar os XMLs
F_14	Tipo de Algoritmo para treinamento da IA: <i>XGBoost</i> ; <i>Random Forest</i> ; <i>AdaBoost</i> ; CART e Todos os Classificadores. Em todos os classificadores, são adicionados: <i>LinearSGDTrainer</i> ; <i>CRFTrainer</i> ; <i>MultinomialNaiveBayesTrainer</i> e <i>LibLinearClassificationTrainer</i> . Esses três últimos não têm menus separados, uma vez que não tiveram resultados relevantes nas simulações realizadas
F_15	Limpar os parâmetros
F_16	Iniciar o treinamento da IA

Fonte: Elaborado pelo autor, 2022.

Os campos F_01 até F_06 são parâmetros utilizados para treinamento e avaliação da

capacidade de aprendizado da IA. Os campos F_08 até F_12 são empregados para exportação dos dados das notas fiscais que estejam salvas no banco de dados do aplicativo, que já foram previamente importadas.

O campo F_13 determina a origem dos dados, havendo, nesse momento, apenas uma opção: dos dados importados, processados e armazenados na própria base de dados do aplicativo. Essa opção foi criada para que, no futuro, outras fontes de dados possam ser adicionadas, por exemplo, a leitura direta de um arquivo CSV com os dados necessários ao treinamento da Inteligência Artificial.

O campo F_14 lista as opções de algoritmos disponíveis para treinamento da Inteligência Artificial. Atualmente, as opções são: *XGboost*; *Random Forest*; *CART*, que foram testadas neste trabalho e obtiveram melhores resultados. Outros algoritmos foram adicionados, mas, por não serem baseados em árvores de decisão, não foram utilizados. Por último, existe uma opção genérica que permite o mapeamento e a utilização de outros algoritmos da Tribuo, cujas propriedades devem ser mapeadas em arquivos XML, no arquivo `/configs/full-config-file.xml`.

A Figura B3 detalha a interface para consulta de atualização dos dados das pessoas armazenados no banco de dados junto à Receita Federal. A Tabela B3 expõe as descrições de cada campo. Essa interface visa à busca e atualização dos cadastros dos emissores e destinatários, as pessoas envolvidas nos documentos fiscais, conforme o cadastro destas na Receita Federal.

Figura B3 – Interface do Menu de Consulta de Dados

Fonte: Elaborado pelo autor, 2022.

Tabela B3 – Funções do Menu de Consulta de Dados

Código	Descrição
--------	-----------

F_01	Arquivo do Certificado a ser utilizado para consulta dos dados
F_02	Keystore, com a cadeia de certificados confiáveis, utilizados para a consulta dos dados
F_03	Senha do certificado
F_04	Senha do arquivo <i>Keystore</i>
F_05	UF base de consulta, conforme o certificado do contribuinte informado
F_06	Botão para iniciar a consulta dos dados junto à Sefaz

Fonte: Elaborado pelo autor, 2022.

A Figura B4 detalha a interface para leitura, processamento e importação de documentos XML, enquanto a Tabela B4 exhibe as funções de cada campo.

Figura B4 – Interface Leitura de Documentos

Fonte: Elaborado pelo autor, 2022.

Tabela B4 – Funções da Interface de Leitura de Documentos

Código	Descrição
F_01	Fonte de dados: Banco de Dados ou Arquivos diretamente do computador
F_02	Número de <i>threads</i> de execução paralela, para processamento dos arquivos. Esse número de <i>threads</i> deve ser menor do que o número de cores de processamento do processador do computador
F_03	Caso Fonte Dados Banco de Dados, qual o tipo de banco de dados. Atualmente, a opção suportada é <i>Firebird</i>

F_04	Diretório base para processamento dos bancos de dados ou dos arquivos XML
F_05	Caso Fonte Dados Banco de Dados, informar o usuário do Banco de Dados
F_06	Caso Fonte Dados Banco de Dados, informar a senha do Banco de Dados
F_07	Caso Fonte Dados Banco de Dados, informar a porta do Banco de dados
F_08	Caso Fonte Dados Banco de Dados, informar o nome da tabela onde estão salvos os XMLs
F_09	Caso Fonte Dados Banco de Dados, informar a coluna onde está o XML da nota fiscal
F_11	Informar a extensão do arquivo do banco de dados. No caso do <i>Firebird</i> , os arquivos têm extensão .FDB
F_12	Lista de fonte de dados carregada
F_13	Clique para listar todas as pastas ou banco de dados no diretório-base informado F_04, conforme a seleção de fonte de dados, F_01
F_14	Clique para remover uma fonte de dados
F_15	Clique para iniciar o processamento das fontes de dados (Pastas ou Bancos de Dados) carregados

Fonte: Elaborado pelo autor, 2022.

Nessa interface, podemos visualizar dois grupos de funções específicas, conforme Fonte de Dados selecionada: Banco de Dados ou Arquivos locais da máquina.

No caso de banco de dados, se o emissor dos documentos armazena os XMLs em um banco, essa opção é a indicada, pois os arquivos serão lidos diretamente do banco de dados, sem a necessidade de exportá-los para pastas do computador. Atualmente, o programa consegue carregar esses arquivos de bancos *Firebird*, podendo ser adicionados outros tipos de bancos no futuro.

Se os arquivos são salvos localmente, deve-se utilizar a Fonte de Dados Arquivos. Nesse caso, deve-se configurar o aplicativo para a pasta onde estão os arquivos. Caso existam subpastas, o aplicativo será capaz de automaticamente identificar as subpastas e os arquivos dentro delas.

Na Figura B5, temos a interface para exportação de dados e, na Tabela B5, as funções

dos campos.

Figura B5 – Interface do Menu de Exportação de Dados

Fonte: Elaborado pelo autor, 2022.

Tabela B5 – Funções do Menu de Exportação de dados

Código	Descrição
F_01	Analisa e define as relações tributárias de IPI
F_02	Analisa e define as relações tributárias de PIS
F_03	Analisa e define as relações tributárias de COFINS
F_04	Analisa e define as relações tributárias de ICMS
F_05	Analisa e define as relações tributárias de ICMS ST
F_06	Analisa e define as relações tributárias Associações Tributárias
F_07	Informe o CNPJ alvo em que deseja exportar os dados
F_08	Exportar os Produtos em CSV
F_09	Exportar dados/relações tributárias IPI em CSV
F_10	Exportar os Clientes em CSV
F_11	Exportar dados/relações tributárias PIS/Cofins em CSV
F_12	Exportar os Fornecedores em CSV
F_13	Exportar dados/relações tributárias ICMS em CSV

F_14

Exportar as Naturezas Operação em CSV

F_15

Exportar dados/relações Associações Tributárias em CSV

Fonte: Elaborado pelo autor, 2022.

Nessa interface, o usuário poderá gerar arquivos de Produtos, Clientes, Fornecedores, Naturezas de Operação e demais dados fiscais relacionados aos impostos IPI, ICMS, PIS e Cofins.

Na interface apresentada na Figura B6, o usuário poderá realizar consultas e testes junto ao aplicativo, informando os parâmetros necessários, sendo que, ao clicar em testar, o sistema irá carregar o modelo de aprendizado da Inteligência Artificial e, assim, aplicá-lo nos parâmetros. Como resultado, exibirá qual o CST a ser utilizado.

Tabela B6 – Funções do Menu de Testes IA

Fonte: Elaborado pelo autor, 2022.

Tabela B6 – Funções do Menu de Testes IA

Código	Descrição
F_01	Clique para adicionar um registro
F_02	Clique para remover um registro
F_03	Clique para avaliar os dados
F_04	Informe o código NCM
F_05	Informe o código CNAE
F_06	Informe o tipo de operação
F_07	Informe código CFOP
F_08	Informe a procedência do produto
F_09	Informe o tipo de inscrição estadual
F_10	Informe o tipo de consumidor

F_11

Informe o tipo de destinatário

F_12

Campo onde é exibido o CST sugerido

Fonte: Elaborado pelo autor, 2022.

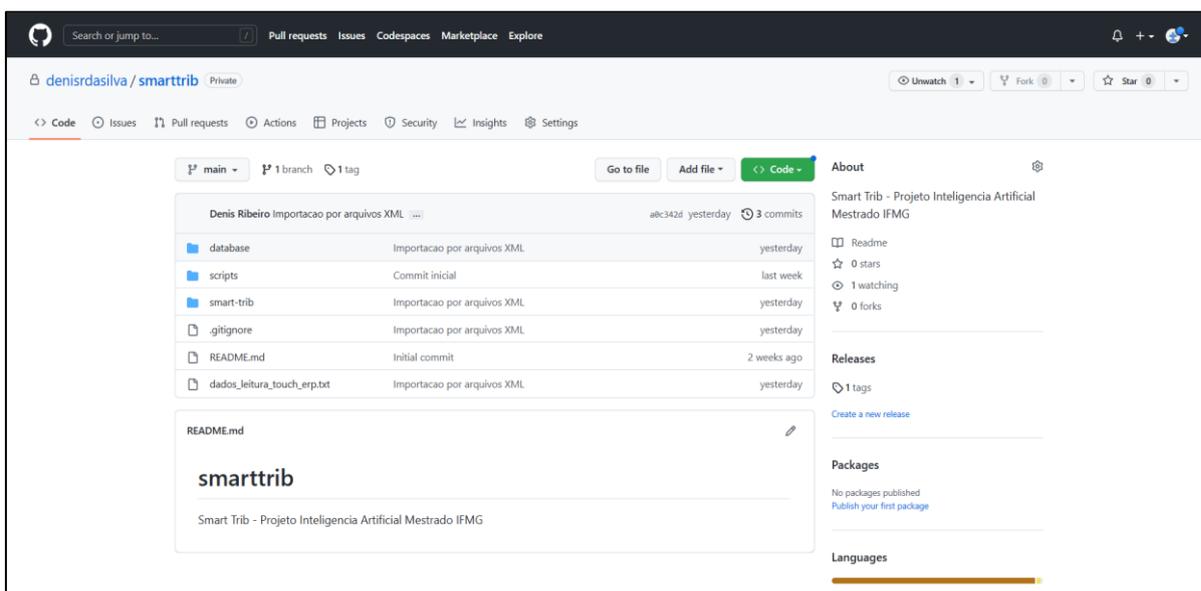
Podem ser adicionados diversos registros, e, após, deve-se clicar no botão de Avaliação/Teste. Desse modo, todos os registros serão avaliados, e os respectivos CSTs, retornados.

APÊNDICE C – Repositório do Projeto

Este projeto foi hospedado em um repositório privado da plataforma *GitHub*, conforme Figura C1, no endereço <https://github.com/denisrdasilva/smarttrib>.

Caso tenha interesse em visualizar o projeto, solicite liberação pelo e-mail deniscomputacao@gmail.com.

Figura C1 – Repositório do Projeto



Fonte: Elaborado pelo autor, 2023.