

**AVALIAÇÃO EXPERIMENTAL DE UM CLASSIFICADOR PARA APOIAR A
DETECÇÃO DE FRAUDES EM NOTAS FISCAIS PÚBLICAS DE MEDICAMENTOS**

**EXPERIMENTAL EVALUATION OF A CLASSIFIER TO SUPPORT THE
DETECTION OF FRAUD IN PUBLIC INVOICES FOR MEDICINES**

**EVALUACIÓN EXPERIMENTAL DE UN CLASIFICADOR PARA RESPALDAR LA
DETECCIÓN DE FRAUDES EN LAS FACTURAS PÚBLICAS DE
MEDICAMENTOS**

Cited as:

Fontes, Raphael S., Colaço Júnior, Methanias, Oliveira, Frederico R. de & Benicasa, Alcides X. (2024). Avaliação experimental de um classificador para apoiar a detecção de fraudes em notas fiscais públicas de medicamentos. Revista Gestão & Tecnologia. Journal of Management and Technology. v 24, nº 1. p .112-148

Raphael Silva Fontes

Mestrado em Ciência da Computação pela Universidade Federal de Sergipe
<https://orcid.org/0000-0003-3160-3384>

Methanias Colaço Júnior

Professor da Universidade Federal de Sergipe.
<https://orcid.org/0000-0002-4811-1477>

Frederico Resende de Oliveira

Universidade Federal de Sergipe
<https://orcid.org/0000-0002-9503-4160>

Ricardo Alexsandro de Medeiros Valentim

Universidade Federal do Rio Grande do Norte
<https://orcid.org/0000-0002-9216-8593>

Editor Científico: José Edson Lara
Organização Comitê Científico
Double Blind Review pelo SEER/OJS
Recebido em 25/09/2022
Aprovado em 05/03/2024



This work is licensed under a Creative Commons Attribution – Non-Commercial 3.0 Brazil

RESUMO

Objetivo do estudo: Desenvolver e avaliar a eficácia de um classificador hierárquico customizado, no contexto de Notas Fiscais Eletrônicas Públicas de medicamentos, analisadas pelo Ministério Público (MP), Grupo de Atuação Especial de Combate ao Crime Organizado (GAECO) e Secretaria da Fazenda Estaduais.

Metodologia/abordagem: Após o desenvolvimento e a parametrização inicial do classificador proposto, foi executado um experimento controlado com notas fiscais públicas custodiadas pelo MP, respeitando o sigilo fiscal dos envolvidos.

Originalidade/relevância: O problema de descrições de Notas Fiscais incongruentes é nacional e as soluções gerais publicadas na literatura não apresentam uma altíssima precisão. Neste sentido, o trabalho apresenta a implementação de um classificador de texto capaz de identificar descrições de medicamentos, suas subclasses hierárquicas e associações presentes em Notas Fiscais Eletrônicas Públicas, cujo resultado servirá para construção de análises e investigações que apoiarão decisões estratégicas.

Principais resultados: Considerando a significância estatística, o classificador foi capaz de identificar as descrições de medicamentos e suas subclasses hierárquicas, com os seguintes resultados médios: acurácia de 99.81%, precisão de 100%, revocação ou sensibilidade de 99.64% e medida-F1 de 99.82%.

Contribuições teóricas/metodológicas: Foi esmiuçado o problema de reconhecimento incorreto de produtos em Notas Fiscais, facilitando discussões e construções de novas propostas. Além disso, foi apresentada uma solução prática baseada em Mineração de Textos, evidenciando o algoritmo que obteve a melhor eficácia para automatizar a extração das subclasses hierárquicas das descrições dos medicamentos e identificá-las, viabilizando investigações e detecção de fraudes.

Palavras-chave: Medicamentos, Corrupção, Mineração de Dados, Nota Fiscal Eletrônica Pública, Investigação.

ABSTRACT

Objective: Develop and evaluate the effectiveness of a customized hierarchical classifier, in the context of drugs descriptions in the Public Electronic Invoices, analyzed by the Prosecution Office (MP), Special Action Group to Combat Organized Crime (GAECO) and the State Finance Department.

Methodology/approach: After the development and initial parameterization of the classifier, a controlled experiment was carried out with public invoices held by the MP, respecting the fiscal confidentiality of those involved.

Originality/relevance: The problem of incongruent invoice descriptions is national and the general solutions published in the literature are not very precise. In this sense, this work presents the implementation of a text classifier able to identify drug descriptions, their hierarchical subclasses and associations present in Public Electronic Invoices, the result of which will serve to build analyzes and investigations that will support strategic decisions.

Main results: Considering the statistical significance, the classifier was able to identify the drug descriptions and their hierarchical subclasses, with the following average results: accuracy of 99.81%, precision of 100%, recall or sensitivity of 99.64% and F1-measure of 99.82%.

Theoretical/methodological contributions: The problem of incorrect recognition of products in invoices was investigated, facilitating discussions and construction of new proposals. In addition, a practical solution based on Text Mining was presented, showing the algorithm that obtained the best efficiency to automate the extraction of hierarchical subclasses of drug descriptions and to identify them, enabling investigations and fraud detection.

Keywords: Medicines, Corruption, Data Mining, Public Electronic Invoice, Investigation.

RESUMEN

Objetivo de la investigación: Desarrollar y evaluar la efectividad de un clasificador jerárquico personalizado, en el contexto de las Facturas Electrónicas Públicas de medicamentos, analizadas por el Ministerio Público (MP), Grupo de Acción Especial para Combatir el Crimen Organizado (GAECO) y la Secretaría de Hacienda del Estado.

Metodología/enfoque: Luego del desarrollo y parametrización inicial del clasificador, se realizó un experimento controlado con facturas públicas en poder del MP, respetando la confidencialidad fiscal de los involucrados.

Originalidad/relevancia: El problema de las descripciones de facturas incongruentes es nacional y las soluciones generales publicadas en la literatura no son muy precisas. En este sentido, el trabajo presenta la implementación de un clasificador de texto capaz de identificar descripciones de medicamentos, sus subclases jerárquicas y asociaciones presentes en las Facturas Electrónicas Públicas, cuyo resultado servirá para construir análisis e investigaciones que sustentarán decisiones estratégicas.

Principales resultados: Considerando la significancia estadística, el clasificador pudo identificar las descripciones de los medicamentos y sus subclases jerárquicas, con los siguientes resultados promedio: exactitud del 99,81%, precisión del 100%, recall o sensibilidad del 99,64% y medida-F1 del 99,82%

Aportes teórico/metodológicos: Se investigó el problema del reconocimiento incorrecto de los productos en las facturas, facilitando las discusiones y la construcción de nuevas propuestas. Además, se presentó una solución práctica basada en Text Mining, mostrando el algoritmo que obtuvo la mayor eficiencia para automatizar la extracción de subclases jerárquicas de descripciones de medicamentos e identificarlas, permitiendo investigaciones y detección de fraudes.

Palabras clave: Medicamentos, Corrupción, Minería de datos, Factura Electrónica Pública, Investigación.

1 INTRODUÇÃO

A corrupção é definida como o uso indevido do poder para a obtenção de vantagens ilegais, podendo ser encontrada em quase todos os tipos de organizações ou grupos, incluindo órgãos públicos, empresas privadas e entidades do terceiro setor (Kratcoski & Edelbaher, 2018). A Convenção contra a Corrupção da Organização das Nações Unidas (ONU) descreve-a como uma praga insidiosa, que tem uma ampla gama de efeitos corrosivos nas sociedades: mina a democracia e o estado de direito, leva a violações dos direitos humanos, distorce os mercados, prejudica a qualidade de vida e permite que a organização do crime, o terrorismo e outras ameaças à segurança humana floresçam (United Nations, 2003).

Por se tratar de um fenômeno cujas dimensões, componentes e envolvidos têm natureza oculta, a quantificação da corrupção é uma tarefa desafiadora. Sua avaliação é muitas vezes feita de maneira indireta, por meio da medida de sua percepção pela população. Nesta categoria, o Índice de Percepção da Corrupção produzido pela ONG Transparência Internacional é o mais difundido. Utilizando treze fontes diferentes de dados, tais como pesquisas feitas com *experts* no tema e com executivos de vários setores da economia, além de indicadores socioeconômicos quantitativos e qualitativos, o índice permite comparar o nível de percepção da corrupção em diferentes países (Transparency International, 2019).

Dadas as dificuldades de mensuração da corrupção, seus efeitos na economia também são estimados utilizando-se modelos teóricos. No Brasil, um estudo a esse respeito foi realizado pela Federação das Indústrias do Estado de São Paulo (FIESP) (2010). Utilizando um modelo neoclássico de crescimento econômico, baseado em Mankiw, Romer e Weil (1992), seus autores estimaram uma faixa compreendida entre 1,38 e 2,30% para as perdas do PIB devidas à corrupção no Brasil, o que, a valores do PIB brasileiro de 2019, representa recursos entre R\$ 100 e 160 bilhões, nos cálculos do Instituto Brasileiro de Geografia e Estatística (IBGE, 2020). Dentro do contexto das oportunidades de corrupção, os gastos governamentais representam uma parcela significativa da economia brasileira. O orçamento federal, por exemplo, previu o montante de quatro trilhões de reais para o exercício de 2020 (CGU, 2020). Tais recursos são, de forma recorrente, alvo de ações corruptas no Brasil, notadamente aqueles ligados às áreas de saúde e educação públicas. A Constituição Federal estabelece que os estados deverão gastar no

mínimo 12% de sua receita com a área de saúde; que os Municípios terão que aplicar pelo menos 15% dos seus recursos nesta seara e que a União deverá utilizar 15% da sua Receita Corrente Líquida para financiar o setor (Constituição da República Federativa do Brasil, 1988) (Sulpino, Piola & Benevides, 2019).

No fornecimento de medicamentos, há particularidades que tornam seu fornecimento especialmente suscetível às mencionadas fraudes, quando da entrega dos itens contratados. Ocorre que é necessário que estejam especificadas as subclasses, tais como: os princípios ativos, as formas de apresentações, as dosagens e a quantidade de unidades por embalagem de cada lote licitado, para que o produto seja devidamente identificado. O descumprimento do atendimento de uma ou mais dessas subclasses pode facilmente resultar em prejuízo aos cofres públicos e estratégias devem ser implementadas pelos gestores públicos com a finalidade de mitigar tais riscos potenciais (Receita Federal, 2016).

Os registros das subclasses também são feitos nas Notas Fiscais Eletrônicas (NF-es), as quais são instrumentos que podem ser utilizados para ratificar o fato de um serviço ou bem adquirido condizer com a especificação passada pela Administração Pública, servindo como prova jurídica das operações comerciais realizadas pelo contribuinte (Constituição da República Federativa do Brasil, 1988). A partir da NF-e, a Administração Pública arrecada recursos financeiros, na forma de tributos, que são necessários para custear seu funcionamento e para lastrear investimentos públicos, notadamente, no âmbito estadual, pelo Imposto sobre Circulação de Mercadorias e Prestação de Serviços (ICMS) (Receita Federal, 2016). No processamento do ICMS, a Administração Fazendária utiliza o campo Nomenclatura Comum do Mercosul (NCM), presente na NF-e, como meio para definir qual a alíquota que incidirá sobre o item. Acontece que nem sempre o NCM inserido pelo contribuinte condiz com a descrição da mercadoria que ele está vendendo (Carvalho, Paiva, Rocha & Mendes, 2014).

Os registros das subclasses também são feitos nas Notas Fiscais Eletrônicas (NF-es), as quais são instrumentos que podem ser utilizados para ratificar o fato de um serviço ou bem adquirido condizer com a especificação passada pela Administração Pública, servindo como prova jurídica das operações comerciais realizadas pelo contribuinte (Constituição da República Federativa do Brasil, 1988). A partir da NF-e, a Administração Pública arrecada

recursos financeiros, na forma de tributos, que são necessários para custear seu funcionamento e para lastrear investimentos públicos, notadamente, no âmbito estadual, pelo Imposto sobre Circulação de Mercadorias e Prestação de Serviços (ICMS) (Receita Federal, 2016). No processamento do ICMS, a Administração Fazendária utiliza o campo Nomenclatura Comum do Mercosul (NCM), presente na NF-e, como meio para definir qual a alíquota que incidirá sobre o item. Acontece que nem sempre o NCM inserido pelo contribuinte condiz com a descrição da mercadoria que ele está vendendo (Carvalho, Paiva, Rocha & Mendes, 2014).

Além desse fator, se for considerado o uso das descrições como uma alternativa de identificação, estas, tanto para produtos em geral quanto para medicamentos, são apresentadas em texto escrito de forma livre pelo fornecedor, o que pode causar a omissão de informações ou dificultar a identificação e extração de subclasses hierárquicas de medicamentos, por exemplo (Carvalho, Paiva, Rocha & Mendes, 2014). Essas subclasses representam a composição dos medicamentos e podem ser: o nome comercial ou genérico do medicamento, a dosagem, a quantidade do produto e a forma farmacêutica. A ausência de uma destas subclasses pode acarretar a não-identificação do medicamento, uma vez que ele pode ser descrito pelo seu nome comercial ou princípio ativo, possuir mais de uma dosagem, diferentes quantidades e formas farmacêuticas.

Essa ausência prejudica a recuperação de informações, pois as descrições dos medicamentos nas NF-es são utilizadas para consultas, tanto por parte dos emissores quanto dos receptores, com a finalidade de facilitar a identificação das características das mercadorias. Essas nomenclaturas também são utilizadas pelos órgãos de controle para identificação dos itens nas compras efetuadas por entidades públicas, os quais enfrentam o mesmo problema das falhas nas descrições (Carvalho, Paiva, Rocha & Mendes, 2013).

Diante dos contextos supracitados, associações incorretas dos NCMs das mercadorias e as falhas nas descrições das subclasses dos medicamentos, o objetivo deste trabalho é apresentar e avaliar, por meio do método de experimentação controlada, um classificador hierárquico de descrições incongruentes de medicamentos, denominado DescMiner, sob o contexto das notas fiscais eletrônicas públicas do Estado de Sergipe, analisadas pelo Ministério Público. A aplicação desse classificador poderá servir efetivamente aos trabalhos investigativos realizados

pelos órgãos de controle, pois permitirá a extração de todas as subclasses dos medicamentos, presentes nas descrições desestruturadas das notas fiscais, bem como permitirá a localização de medicamentos cadastrados com NCMs errados.

Tal utilidade é ilustrada pela identificação de fraudes no fornecimento de medicamentos, conexas com as quantidades de unidades fornecidas e seus respectivos preços, auxiliando na descoberta de despesas com notas fiscais inidôneas, na identificação de pagamentos indevidos ou realizados sem a correspondente execução do objeto, ou ainda na falta de prestação de contas. Além destes cenários, o classificador também pode auxiliar na identificação de necessidade da compra de um determinado medicamento para uma cidade ou região de saúde, a qual, em apertada síntese, consiste num agrupamento de municípios limítrofes que tem a finalidade de integrar a organização, o planejamento e a execução de ações e serviços de saúde, conforme Resolução Normativa n. 259, da Agência Nacional de Saúde (ANS) (2011). Por fim, com os dados classificados e agrupados corretamente, é possível averiguar desvios nas compras dos medicamentos, tais como: compras incompatíveis com a população, medicamentos fora da validade ou no limite desta e preços anômalos, muito acima dos praticados pelos diversos entes públicos ou acima do preço máximo de venda ao governo (PMVG) - valor máximo de preço para compras públicas estabelecido pela Agência Nacional de Vigilância Sanitária (ANVISA).

Após a execução do experimento, o qual contrapôs medidas de similaridade e de dissimilaridade da área de Mineração de Textos (Vijayarani & Janani, 2016), foi evidenciado que o classificador DescMiner, ao utilizar a medida mais efetiva experimentalmente, detectou as subclasses hierárquicas dos medicamentos e, conseqüentemente, as descrições presentes nas NF-es públicas, com uma acurácia média de 99.81%, precisão de 100%, revocação ou sensibilidade de 99.64% e medida-F1 de 99.82%. Por se tratar de um contexto no qual não pode haver erros no momento da identificação das subclasses hierárquicas, o classificador se mostrou eficaz no que se refere à extração destas subclasses e à classificação precisa dos medicamentos. Além disso, este trabalho confirma resultados encontrados na literatura, os quais evidenciam o cosseno como medida de similaridade eficaz para classificação de textos (Feldman & Sanger, 2006).

2. Base Conceitual

2.1 Controle da Atividade Pública

Dos tributos de responsabilidade da esfera estadual, o ICMS representa a maior fonte de receitas próprias, responsável por 19,7% da arrecadação pública do país, sendo um imposto não-cumulativo, pelo qual se compensa o que for devido em cada operação relativa à circulação de mercadorias ou prestação de serviços, com o montante cobrado nas operações anteriores, segundo o IBTP - Instituto Brasileiro de Planejamento e Tributação (IBTP, 2020). No estado de Sergipe, o ICMS representou 90,13% do que foi arrecadado no primeiro semestre de 2019 pela Sefaz - Secretaria de Fazenda de Sergipe (Sefaz, 2019).

Neste contexto, a Nota Fiscal Eletrônica é uma das principais provas de validade jurídica das operações comerciais realizadas pelo contribuinte. Sendo um programa de âmbito nacional, a Nota Fiscal Eletrônica (NF-e) foi desenvolvida pela Administração Tributária, a qual instituiu um modelo unificado de documento fiscal em meio eletrônico, substituindo o documento fiscal em papel. A NF-e tem como principal objetivo a modernização da Administração Tributária por meio da redução de custos e entraves burocráticos, facilitando o cumprimento das obrigações tributárias pelo contribuinte, além de fortalecer o controle e fiscalização pelos órgãos de controle (Emenda Constitucional n. 42, 2003).

A representação gráfica (física) da NF-e é denominada de Documento Auxiliar da Nota Fiscal Eletrônica (DANFE). O DANFE possui os mesmos campos definidos nos modelos anteriores à NF-e, este serve apenas como um instrumento auxiliar na consulta à NF-e, pois contém impressa a chave de acesso, a qual permite a validação do documento no site da Sefaz. O DANFE não é uma nota fiscal e nem a substitui.

A partir do DANFE é possível identificar os bens que foram faturados. Estes itens, produtos ou serviços possuem um código identificador que é utilizado pelo emissor da NF-e para sua identificação em seu sistema de faturamento, uma descrição textual do item para leitura e fácil identificação ao que se refere o item e o código NCM - Nomenclatura Comum do Mercosul -, além das informações tributárias, quantidades e valores do bem (Receita Federal, 2016). Um NCM nada mais é que um número, cujo seu objetivo é facilitar a identificação da mercadoria para a fiscalização e tributação desta pelas autoridades competentes. Cada NCM

possui um valor de alíquota atrelado, correspondente ao imposto que incidirá sobre o bem, além disso, cada item na nota deve ser vinculado obrigatoriamente ao código correspondente na tabela NCM.

Todavia, alguns contribuintes podem associar códigos referentes à tabela de NCMs que não correspondam, de fato, aos itens dos produtos descritos nas NF-e, de modo que pode haver a incidência de alíquotas erradas sobre os produtos fornecidos. Como apresentado por Carvalho, Paia, Rocha e Mendes (2014), este fato impacta a arrecadação do Estado, uma vez que o imposto devido não é corretamente recolhido. Além disso, há a impossibilidade de autuação do contribuinte envolvido, por falta de meios que evidenciem a existência de irregularidades na emissão da NF-e. Finalmente, códigos incompatíveis ou incongruentes com as verdadeiras descrições dos produtos podem inviabilizar as investigações de órgãos de controle, que ficam impedidos de averiguar com precisão variáveis como preço médio, valores e quantidades exorbitantes de produtos adquiridos, muitas vezes contabilizados em categorias totalmente diferentes das suas. Este fato ocorre com as notas públicas de medicamentos, os quais possuem uma vasta taxionomia, elucidada a seguir.

2.2 Medicamentos e a Problemática das Notas Fiscais Públicas com Descrições Incongruentes, como Base para Investigações e para o Cálculo do Preço Médio

De 25% a 70% do gasto em saúde nos países em desenvolvimento correspondem a medicamentos, em comparação a menos de 15% nos países desenvolvidos. Nesta mesma linha, de 50% a 70% das consultas médicas geram prescrição de medicamentos (Aquino, 2008). Esses medicamentos possuem um composto principal, chamado de princípio ativo, bem como as especialidades farmacêuticas, também conhecidas como fórmulas ou formas farmacêuticas, que podem ser comercializadas sob as mais diferentes maneiras (Silva, 2002). O conjunto desses princípios ativos e formas farmacêuticas são organizados hierarquicamente e compõem, juntamente com as dosagens e a quantidade, as descrições dos medicamentos. Na NF-e, a descrição deve ser completa, conforme apresentado na Figura 1.

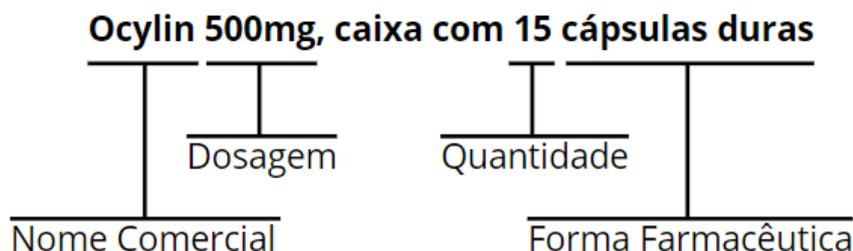


Figura 1 – Subclasses Hierárquicas da Descrição de um Medicamento.

Contudo, as descrições de medicamentos nas NF-es são expressas em textos desestruturados que possuem informações omitidas sobre as subclasses hierárquicas, as quais representam a composição dos medicamentos e contêm seus nomes comerciais ou genéricos, suas dosagens, suas quantidades e suas formas farmacêuticas. Por exemplo, é comum o exemplo da Figura 2 vir descrito assim:

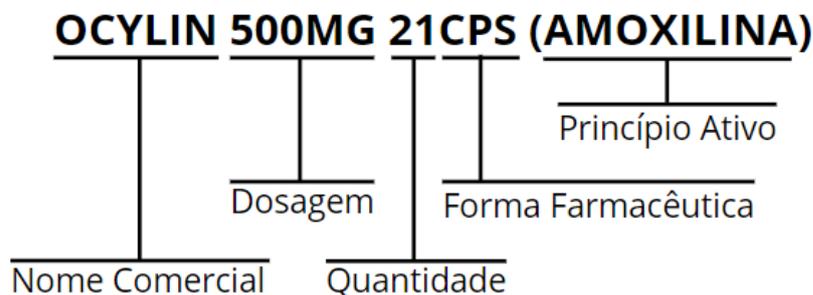


Figura 2 – Subclasses Hierárquicas na Descrição de um Medicamento, em uma NF-e Pública.

Como visto anteriormente, também é comum o medicamento estar com o NCM (código) errado ou um outro produto qualquer estar com o código de um medicamento. Desta forma, a extração das subclasses dos medicamentos, juntamente com as NF-es, pode auxiliar a comprovação de despesas com notas fiscais inidôneas, a identificação de pagamentos indevidos ou realizados sem a correspondente execução do objeto, ou ainda a detecção da falta de prestação de contas. Além destes cenários, a extração correta dessas subclasses também pode auxiliar a identificação de necessidade da compra de um determinado medicamento para uma cidade ou região de saúde, a qual, em apertada síntese, consiste num agrupamento de municípios limítrofes cujo a finalidade é de integrar a organização, o planejamento e a execução

de ações e serviços de saúde na região ANS (2011). Por fim, com os dados classificados e agrupados corretamente, os órgãos de controle poderão averiguar desvios nas compras dos medicamentos, tais como compras incompatíveis com a população, medicamentos fora da validade ou no limite desta, bem como com preços anômalos, muito acima do praticado pelos diversos entes públicos ou acima do Preço Máximo de Venda ao Governo (PMVG) (CMED, 2004).

2.3 Mineração de Textos e Medida de Importância de uma Palavra (TF-IDF)

Este trabalho utilizou técnicas de mineração de textos para fundamentar e construir o classificador aqui proposto. Dentre essas técnicas, destacam-se o Modelo Vetorial, cálculo de pontuação (importância ou *score*) de uma palavra e formas de medir similaridade entre documentos ou entre descrições, abordados nesta e na próxima seção. A mineração de textos consiste em um processo de conhecimento intensivo, no qual um usuário interage com uma coleção de documentos utilizando um conjunto de ferramentas de análises, com a finalidade de extrair informações úteis desses documentos, por meio da identificação e exploração dos padrões de interesses definidos (Feldman & Sange, 2006).

Dentro desse contexto, um documento pode ser qualquer dado textual discreto, tais como relatórios, memorandos, e-mails, artigos de pesquisas, notícias etc. Diferente de um banco de dados, os documentos podem não ter uma estrutura de dados bem definida, pois cada documento apresenta sua própria estrutura de sintaxe e semântica. A estrutura do documento é constituída por letras, palavras, pontuações, caracteres especiais, números etc. Para normalizar essa diferença na estruturação dos documentos, é efetuado um pré-processamento no texto, empregando-se técnicas que transformam o documento desestruturado em uma estrutura simétrica ao conjunto de documentos.

No primeiro momento do pré-processamento, o texto é quebrado em palavras, que poderão ser chamadas de *tokens*. Os *tokens* são quebrados, de modo geral, pelos espaços, e cada palavra que esteja entre espaços torna-se um *token*. Este processo de *tokenização* é utilizado para identificar palavras-chave que façam sentido e representem o documento (Vijayarani & Janani, 2016). Neste processo, pontuação e espaços podem ser removidos ou não, dependendo

do contexto que está sendo analisado. Os *tokens* extraídos do documento farão parte de um dicionário, cujo objetivo é agrupar aqueles que representam o documento.

Depois da extração dos *tokens*, é possível transformar o dicionário em uma planilha (modelo vetorial) na qual cada linha corresponde a um documento e cada coluna a um token. A célula na planilha apresenta o valor correspondente do *token* na coluna. Os valores podem ser obtidos por alguns processos de extração, tais como *Term Frequency* (TF) e *Term Frequency-Inverse Document Frequency* (TF-IDF).

A Frequência do Termo, do inglês *Term-frequency* (TF), corresponde ao número de vezes que o termo aparece no documento. Os termos que são frequentemente mencionados em determinados documentos podem servir como discriminantes.

Para um cálculo mais contextual, a medida estatística TF-IDF, apresentada por Salton, Fox e Wu (1983), leva em conta a importância de uma palavra no *corpus* (dicionário completo de palavras), estando estruturada ou não. Para calcular do TF-IDF, é atribuído um valor para cada termo a partir da frequência do termo no próprio documento (TF já visto) e em todo *corpus* (IDF), indicando sua importância.

A importância do termo é calculada pela equação de Frequência Inversa do Documento, *Inverse Document-Frequency* (IDF), (SALTON; BUCKLEY, 1988). O IDF é definido pela equação a seguir, a qual consiste no logaritmo da razão entre o número total de documentos, N_D , e a frequência de documentos que possuem o termo t , df_t . Quanto maior o IDF do termo, maior é a sua representatividade.

$$IDF_t = \log\left(\frac{N_D}{df_t}\right)$$

O peso final do termo é atribuído pela equação do TF-IDF, na qual o peso é associado à proporção da frequência do termo no documento (TF) e à proporção inversa do número de documentos no qual o termo aparece pelo menos uma vez (IDF). O TF-IDF é representado pela equação a seguir:

$$tf - idf = TF \times IDF$$

Caso a palavra apareça em todos os documentos, seguindo a fórmula, a IDF será o Log de 1, ou seja, será zero, pois o Log de 1 em qualquer base é zero. Como o IDF será zero, quando

multiplicado por qualquer valor do TF, o resultado também será zero. Em outras palavras, quanto mais a palavra aparecer no *corpus*, menor será o seu peso, podendo ser zero, caso esteja presente em todos os documentos (pouco poder de discriminação).

2.4 Medidas de Distância, Similaridade e Dissimilaridade

Ao comparar documentos, a medição se dará por sua similaridade, por exemplo: o quanto dois documentos são similares? Um ponto de avaliação compartilhado entre os documentos são as mesmas palavras. É possível ranquear os documentos que possuem a maior quantidade de palavras compartilhadas, dessa forma, quanto mais palavras, mais similares serão os documentos. Neste trabalho, as medidas de similaridade foram usadas em uma das etapas de classificação, para comparar as descrições incongruentes com as descrições corretas de medicamentos.

2.4.1 Distância de Levenshtein

A distância de Levenshtein, ou o algoritmo de Levenshtein, foi escrito em 1966 pelo matemático Vladimir I, baseado na distância de *Hamming* (Levenshtein, 1965). O princípio do algoritmo é definir a distância entre duas palavras com base no número de operações necessárias para torná-las iguais. Cada operação para modificar a palavra tem um determinado custo, o qual é acumulado de acordo com as edições realizadas, sendo inserção, exclusão ou substituição dos caracteres da palavra. Logo, quanto menos operações existirem, menor é a distância e mais similares são as palavras. Por exemplo, para que a palavra “amoxilina” se torne “**amoxicilina**”, ocorrerá a inserção de 02 (dois) caracteres, sendo: o *i* na posição 4 e o *c* na nova posição 4 (após a inserção do *i*), perfazendo uma distância de 2 entre as duas palavras. Essa distância foi usada para comparar nomes de medicamentos incongruentes das notas com nomes de medicamentos corretos, quando houve a necessidade de desempate, no uso das distâncias abordadas nas próximas seções.

2.4.2 Similaridade do Cosseno

A Similaridade do Cosseno é uma abordagem utilizada para comparar dois documentos e é a medida de similaridade mais comum para trabalhar com textos (Weiss, Indurkha & Zhang Damerau, 2004; Feldman & Sange, 2006). A medida utiliza o ângulo formado pelos vetores correspondentes aos documentos, no espaço dimensional reduzido, para calcular o cosseno deste ângulo e medir a similaridade entre duas cadeias de palavras-chave (*tokens*) (Li & Han, 2013). O peso de um *token* em um documento, $w(j)$, é calculado pela equação de TF-IDF, conforme apresentado na seção anterior.

Abstraindo como o cálculo é feito, até porque este é feito geralmente pelas máquinas, é importante entender que o valor do cosseno, que varia de 0 a 1, indica a similaridade entre os documentos. Quanto mais próximo a 1, mais similares são os documentos, uma vez que o ângulo *theta* formado entre os dois vetores iguais é igual a 0 e $\cos(0)$ é 1. Consequentemente, quanto mais próximo de 0, menos similares são (Al-Anzi & Abuzeina, 2017).

2.4.3 Dissimilaridade com Distâncias: Manhattan e Euclidiana

Uma distância possui algumas características diferentes das métricas de similaridade, tais como o coeficiente de correlação e a similaridade do cosseno. Para as medidas de similaridade, quanto mais próximas do valor 1, mais similares são os documentos. Já para as medidas de dissimilaridade, quanto mais próximas de 0, menor é a distância entre os documentos, consecutivamente, são mais similares.

A dissimilaridade é apresentada pela equação a seguir, na qual o parâmetro k define a forma de chavear as distâncias, Manhattan ($k = 1$) e Euclidiana ($k = 2$), variando de 1 a infinito (Skiena, 2017).

$$d_k(p, q) = \sqrt[k]{\sum_{i=1}^d |p_i - q_i|^k} = \left(\sum_{i=1}^d |p_i - q_i|^k \right)^{\frac{1}{k}}$$

2.5 Métricas de Qualidade

Para avaliação do classificador, foram utilizadas as métricas: acurácia, revocação, precisão e medida-F1 (Zhu et al., 2010). Essas medidas são mensuradas a partir das seguintes frequências:

- a. True Positive (TP): Total de instâncias de produtos das notas, presentes na base anotada (base de treinamento do algoritmo, contendo descrições corretas de medicamentos), que foram classificadas corretamente;
- b. True Negative (TN): Total de instâncias de produtos, que **não** são medicamentos, e foram classificadas corretamente;
- c. False Positive (FP): Total de instâncias de outros produtos que foram classificados **incorretamente** como pertencentes à classe de medicamentos;
- d. False Negative (FN): Total de instâncias de produtos da classe de medicamentos que **não** foram classificadas corretamente.

Por se tratar de um classificador hierárquico, as frequências acima relacionadas foram avaliadas para os quatro níveis da hierarquia e as seguintes métricas foram calculadas, também para cada nível.

2.5.1 Acurácia

A acurácia (*accuracy*) representa o percentual de instâncias (notas fiscais) que foram classificadas de forma correta, sendo definida por:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2.5.2 Revocação

A revocação (*recall*), também conhecida como a taxa de verdadeiros positivos, sensibilidade ou cobertura real da amostragem positiva, é o percentual de instâncias positivas que foram classificadas corretamente:

$$recall = \frac{TP}{TP + FN}$$

2.5.3 Precisão

A precisão (*precision*) é a razão entre as instâncias classificadas como "verdadeiro positivo" e todas as instâncias classificadas como positivas:

$$precision = \frac{TP}{TP + FP}$$

2.5.4 Medida-F1

A medida-F1 (*F1-Score*) é a métrica que combina dois indicadores de desempenho, sendo a expressão da média harmônica da precisão e da revocação:

$$F = \frac{2 \times precision \times recall}{precision + recall}$$

3 METODOLOGIA

Este artigo é um estudo experimental, o qual segue os passos apresentados em (Colaço Júnior & Oliveira, 2018), para avaliação dos resultados das classificações de medicamentos, obtidas a partir da implementação do classificador hierárquico de descrições incongruentes, o DescMiner, o qual utiliza as medidas de similaridade e de dissimilaridade descritas na subseção 2.4.2, avaliadas conforme métricas de qualidade apresentadas na subseção 2.5 e implementadas a partir das bibliotecas mencionadas na subseção 6.2.8.

Inicialmente, foi construída uma base anotada, com as descrições corretas dos medicamentos, identificação dos seus princípios ativos e com o número de referência de cadastro no Ministério da Saúde, que serviu para identificar o medicamento na base cadastral do Ministério, incluindo informações sobre, por exemplo, o preço máximo de venda, tanto para o consumidor quanto para o governo. Após a construção dessa base, foram definidas as etapas de classificação, originadas a partir das subclasses das descrições. Essas etapas serão introduzidas na subseção 5.2 e cada uma delas possui um tratamento exclusivo para seleção dos *tokens* (palavras) que são úteis ao processo de classificação. Na execução do experimento, as descrições das NF-es públicas foram comparadas com as descrições dos medicamentos presentes na base anotada e foram extraídas as informações importantes para auditoria.

No processo de classificação, foi definido o limiar de aceitação para a distância entre a descrição da nota e uma descrição correta, responsável por ser a “linha de corte” entre o que será classificado automaticamente e o que será encaminhado para análise do analista de dados. Além dessa definição, foi feita a seleção dos parâmetros já implementados no DescMiner, tais como: a seleção da medida de similaridade ou de dissimilaridade a ser usada; a seleção do modelo vetorial (TF, TF-IDF); a opção facultativa para selecionar quais os dicionários de palavras serão utilizados ou ignorados; a opção facultativa para informar quais os *tokens* são extras e indesejados, para que possam ser excluídos do processo de classificação; e a aplicação facultativa de um filtro de seleção de parte da base de dados a ser classificada.

Após o desenvolvimento e a parametrização inicial do classificador, foi executado um experimento controlado, *in vitro*, uma vez que os objetos foram retirados da base original para serem manipulados dentro de um ambiente controlado. Este método possibilitou analisar e avaliar os resultados utilizando análise estatística, bem como permitirá uma replicação mais fiel dos procedimentos apresentados neste trabalho.

A replicação de experimentos é uma característica importante para qualquer área da ciência. Assim, na área de software, devem haver métodos que permitam a replicação e a avaliação dos trabalhos, a fim de evitar que novos métodos, técnicas, linguagens e ferramentas sejam apenas sugeridos, publicados ou apresentados para venda, sem experimentação e validação (Travassos, Gurov & Amaral, 2020).

A definição e o planejamento do experimento estão descritos detalhadamente, com suas **metodologias autocontidas**, na Seção 6. Em resumo, o experimento está dividido em quatro etapas principais: o planejamento; a operação de limpeza dos dados, coleta e geração da base experimental; a comparação de métodos; e a análise dos resultados.

3 TRABALHOS RELACIONADOS

Apesar do grande potencial de pesquisa existente nas NF-es públicas brasileiras, estas não são uma área com grande exploração acadêmica. Essa afirmação dar-se-á pelo fato de terem sido encontrados poucos trabalhos utilizando as NF-es como objeto de classificações, utilizando técnicas de aprendizado de máquina ou de mineração de textos. Para que haja um embasamento

deste trabalho, estão referidas, a seguir, pesquisas que classificaram NF-es e outras que fizeram uma forma similar de classificação textual.

A tese de mestrado de Madeira (2015) fez uso de um documento fiscal do município do Rio de Janeiro, denominado de Nota Carioca, para classificar e extrair informações úteis ao planejamento de ações fiscais sobre os tributos municipais. O trabalho atingiu, ao final, índice de acerto de 50%. Esse trabalho, na sua primeira fase, considerando uma situação de complexidade menor do que a deste artigo, assemelha-se a um dos objetivos aqui propostos: a separação de notas em grupos similares.

A monografia de Oliveira (2017) utiliza o algoritmo de Levenshtein para computar a distância entre as descrições de medicamentos, presentes em dados de licitações governamentais dos municípios do Estado do Mato Grosso, e as descrições da base anotada, originada do cadastro oficial de materiais do governo federal (CATMAT). Este trabalho apresentou resultados iniciais exploratórios, sem ainda tratar as perdas de informações, considerando a hierarquia dos dados de medicamentos e descrições incompletas.

O trabalho de Batista (2017) teve como objetivo a classificação automática dos códigos de NCM, por meio das descrições dos produtos presentes nas NF-es emitidas no estado do Rio Grande do Sul. As bases foram distinguidas pelas suas complexidades, avaliadas como simples, média ou complexa, as quais obtiveram, respectivamente, 98%, 90% e 83% de acurácia. O trabalho possui grande similaridade com este artigo, apesar de um objetivo final distinto.

O trabalho de Ribeiro et al. (2018) aplicou a técnica de classificação denominada de Campos Condicionais Aleatórios, para classificação dos *tokens* das descrições das NF-es em rótulos, ou entidades. Segundo os autores, houve 99% de acurácia para algumas categorias, porém ainda não há uma seção destinada aos testes e análises estatísticas dos resultados das classificações.

É importante salientar que não foram encontrados trabalhos que apresentem uma análise comparativa entre algoritmos para a classificação de Notas Fiscais, com a utilização de uma abordagem experimental, assim como não houve propostas de identificação de microinformações relacionadas a subclasses de medicamentos, como é abordado neste trabalho.

Estes fatos impediram uma análise comparativa direta.

4 DESCMINER

Nesta seção, são apresentados a arquitetura de funcionamento, o pré-processamento das NF-es públicas e da base anotada (base de treinamento do algoritmo, contendo descrições corretas de medicamentos), a classificação hierárquica e a estrutura do classificador hierárquico de descrições incongruentes de medicamentos, denominado DescMiner.

4.1 Arquitetura de Funcionamento

A arquitetura de um *software* é a estrutura do sistema, sendo composta por elementos de *software*, propriedades externamente visíveis desses elementos e pelos relacionamentos entre eles (Bass, Clements & Kazman, 2003). A arquitetura de *software* define o que é o sistema, em termos de componentes, os relacionamentos entre estes, bem como os padrões que guiam as composições e suas restrições.

Partindo dos princípios introdutórios da arquitetura de *software*, o DescMiner foi desenvolvido com base no modelo apresentado na Figura 3, exibindo as principais funcionalidades e suas relações.

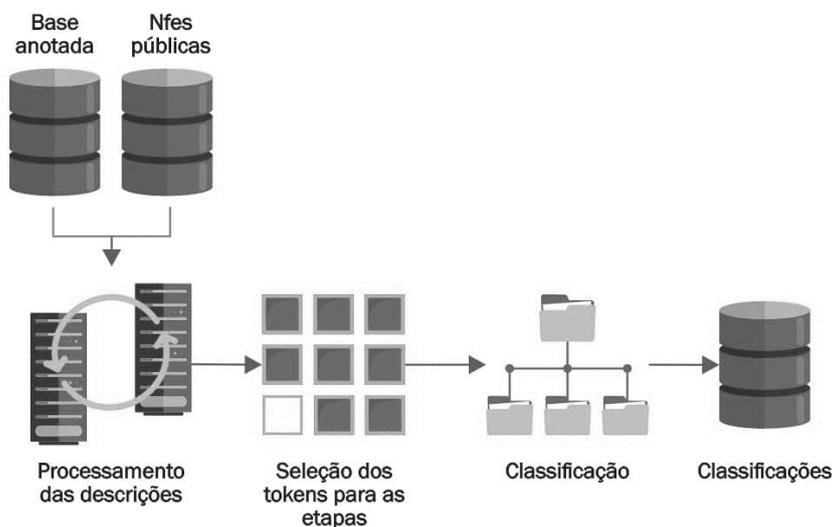


Figura 3 – Fluxo básico de funcionamento do DescMiner.

4.2 Pré-processamento das Notas Fiscais e da Base Anotada

O DescMiner é capaz de minerar textos e realizar o pré-processamento, cujo finalidade é a de extração/seleção dos *tokens* úteis para as classificações hierárquicas. Essa extração é realizada condicionalmente e de maneira ordenada, de modo que os *tokens* selecionados são mantidos numa tabela própria que contém: o número de identificação, o próprio *token*, a sua posição, o número da etapa e a identificação de qual descrição ele faz parte.

O processo é iniciado pela tarefa de normalização da descrição, com o intuito de padronizar as descrições. A normalização, neste trabalho, é responsável por manter o texto com caracteres minúsculos e por remover, caso exista na descrição, os seguintes *tokens* ou caracteres: especiais (" \ [! @ # \$ % ^ & * () [] { } ; : . < > ? \ | ~ = _ -); data; acentuação, mantendo a letra original; espaços em branco em excesso; pontuação; e *stopwords* (preposição, artigo, etc).

Após a tarefa de normalização, é iniciado o processo de seleção dos *tokens* que serão utilizados na classificação hierárquica. Para exemplificar, a Figura 4 representa o processo que a descrição passa até a extração dos *tokens* a serem utilizados na classificação hierárquica.

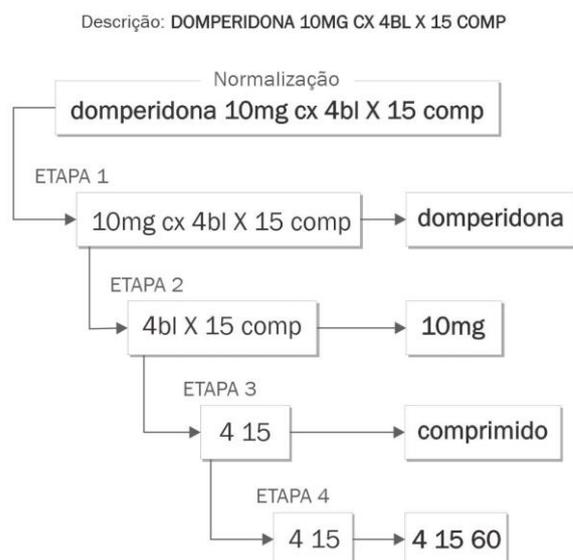


Figura 4 – Exemplo do pré-processamento de uma descrição presente na NFe pública

Na primeira etapa hierárquica, ou também conhecida como Etapa 1, os *tokens* representam os possíveis princípios ativos, selecionados a partir da comparação com a base anotada de medicamentos. O classificador busca sempre os *tokens* elegíveis para essa etapa, além disso, nas descrições dos medicamentos, podem existir possíveis princípios ativos que são compostos e, para cada um deles, é realizado o mesmo processo de aceitação, o qual é efetuado de acordo com a quantidade de princípios ativos que existem na descrição.

Após a execução da primeira etapa, inicia-se a extração dos *tokens* para a dosagem, ou Etapa 2. Os *tokens* dessa etapa são selecionados a partir das unidades conhecidas de dosagem, tais como: miligrama (mg), grama (g), mililitro (ml), unidade internacional (ui) e etc. O DescMiner busca, nos *tokens* restantes, oriundos da etapa anterior, aquele que possui a unidade de medida, fazendo uso das diversas possibilidades de escrita, as quais são extraídas da própria base de dados.

Na Etapa 3, é realizada a verificação e extração da forma farmacêutica presente na descrição do medicamento. O DescMiner possui dicionários já definidos para cada forma farmacêutica comercializada, conforme publicado no Vocabulário Controlado de Formas Farmacêuticas, via Administração Embalagem de Medicamentos (ANVISA, 2011). Cada

vocabulário possui a nomenclatura correta da forma farmacêutica, que serve como chave de sua identificação, bem como as suas variações e diferentes formas de escrita, as quais são utilizadas como possíveis valores de sua associação. Assim que o DescMiner localiza a variação da forma farmacêutica, o *token* é substituído pela chave de identificação, servindo como método de normalização dos *tokens* utilizados. Por exemplo: o *token comp* é normalizado para *comprimido*, uma vez que o *token* é uma variação da forma farmacêutica para comprimido (chave).

Por fim, na Etapa 4, ou etapa da seleção da quantidade, é procurado o *token* que representa a quantidade do medicamento. São observadas as maneiras que as quantidades são dispostas, averiguando a possibilidade de representação de uma unidade ou de uma caixa, o que poderá exigir que o classificador infira as quantidades finais, baseadas nas quantidades de caixas e em quanto cada caixa comporta.

Por fim, a atividade de pré-processamento das descrições acontece de forma automática, sempre antes de uma nova descrição ser armazenada no banco de dados do DescMiner, ocorrendo em momentos diferentes para a base anotada e para a base de NF-es públicas. Na base anotada, ou base de treinamento do algoritmo com as descrições corretas de medicamentos, o pré-processamento ocorre na inserção ou atualização das amostras que compõem os dicionários de treinamento, já na base de NF-es públicas, o processo ocorre quando as descrições são classificadas.

4.3 Classificação Hierárquica

Ao iniciar o processo de classificação hierárquica, a base anotada é bloqueada para que não haja nenhum tipo de modificação em seus registros, enquanto o procedimento de classificação está sendo realizado. Esse procedimento se rege pela preservação dos parâmetros e *tokens* selecionados para a classificação. Após o bloqueio, o DescMiner coleta as descrições da base de NF-es, normaliza e extrai os possíveis *tokens*, mantendo o mesmo método realizado nas descrições da base anotada. Após o primeiro processo de classificação, as descrições alvo da classificação possuem seus textos e os *tokens* armazenados no banco de dados do DescMiner, para que possam ser reaproveitados durante novas classificações.

Como visto anteriormente, na Etapa 1, é realizado o procedimento de identificação da descrição como referente ou não a um medicamento, pois o valor da similaridade ou dissimilaridade deve estar dentro do limiar de aceitação determinado para a classificação. Caso a descrição não alcance o resultado esperado, essa descrição é classificada como não sendo um medicamento. A definição do limiar utilizado para as medidas de similaridade será detalhada na seção experimental.

Após a descrição alcançar o limiar de aceitação informado, ela é classificada para a classe de medicamento. Neste momento, sua descrição é associada ao dicionário do princípio ativo identificado. Esse procedimento de relação facilita a redução de dimensionalidade das possíveis amostras que a descrição pode ser similar, pois nas etapas seguintes são utilizados apenas as amostras do dicionário classificado na Etapa 1. Todas as etapas concluídas resultarão, hierarquicamente, no medicamento da NFe pública mais similar ao medicamento presente no dicionário de medicamentos da base anotada, com a vantagem da informação classificada em quatro níveis, os quais podem ser usados para relatórios gerenciais e investigativos, agrupados por cada etapa da hierarquia.

As descrições de medicamentos da base de notas públicas podem se repetir e, caso isso ocorra, o DescMiner realiza a classificação apenas uma vez, reaproveitando uma classificação anterior e mantendo a eficiência do processo. Por fim, existe a possibilidade do classificador desconhecer um medicamento oriundo da base de notas públicas, fazendo com que a nota seja classificada como “outros”. Para evitar esse processo, é possível adicionar a nova referência na base anotada, caso seja um novo tipo de medicamento, ou uma variação da apresentação do medicamento, para uma descrição já existente.

5 Avaliação Experimental

5.1 Definição do Objetivo

O objetivo deste trabalho é fazer uma análise experimental dos principais métodos utilizados para verificar correspondência de textos encontrados na literatura, para detecção de medicamentos em descrições desestruturadas de NF-es públicas avaliando e validando o que melhor se adequa ao contexto.

Utilizando o modelo GQM (Goal Question Metric) (Basili & Weiss, 1984), foi possível formalizar o objetivo deste estudo da seguinte maneira: **analisar** um classificador hierárquico customizado, utilizando a medida de similaridade cosseno, **contra a** utilização das distâncias Euclidiana e de Manhattan pelo mesmo classificador, **com a finalidade de** avaliar, **com relação** à eficácia, **sob o ponto de vista** de Cientistas de Dados, Auditores Fiscais Fazendários e Investigadores, **no contexto de** notas fiscais eletrônicas públicas do estado de Sergipe, analisadas pelo Ministério Público, GAECO e Secretaria da Fazenda Estaduais.

5.2 Planejamento

5.2.1 Seleção do Contexto

O experimento foi realizado “*in vitro*”, análise em ambiente controlado, no qual foi utilizado o classificador hierárquico para classificar os medicamentos em Notas Fiscais Eletrônicas Públicas. Além disso, o experimento considerou todos os registros de notas fiscais públicas de 2019 e 2020, disponibilizados pelo Ministério Público de Sergipe, fruto de um convênio entre a Universidade Federal de Sergipe, a Secretaria da Fazenda e o Ministério Público de Sergipe.

5.2.2 Questões de Pesquisa

Para guiar o experimento, foi elaborada a seguinte questão principal de pesquisa, cuja resposta visa cumprir o objetivo do trabalho: um classificador, baseado em mineração de textos, pode automatizar o agrupamento correto de medicamentos, independente de seus NCMs?

Além da questão principal, uma questão secundária será respondida neste experimento: Entre os algoritmos selecionados, qual o melhor em termos de eficácia?

5.2.3 Variáveis Independentes

Para esta experimentação, foi elaborado um dicionário a partir de uma base de medicamentos com as descrições corretas. Além disso, também são variáveis independentes a base de Notas Fiscais Eletrônicas Públicas, com as descrições alvo a serem classificadas, o classificador DescMiner, as medidas de similaridade e dissimilaridade, bem como seus limiares.

5.2.4 Variáveis Dependentes

As classificações do DescMiner (descrições dos medicamentos, oriundas das Notas Fiscais Eletrônicas Públicas, classificadas nas subclasses corretas e sob o mesmo NCM), das quais podem ser derivadas as métricas: Acurácia, Precisão, Revocação e Medida-F1.

5.2.5 Formulação de Hipóteses

Como já exposto acima, para avaliar a questão secundária de pesquisa, foram utilizadas quatro métricas: Acurácia, Precisão, Revocação e Medida-F1. Sendo assim, com os objetivos e métricas definidas, serão consideradas as hipóteses teóricas a seguir (**para cada métrica**). A avaliação rejeitou ou não a hipótese nula (H_0):

H_0 : As medidas de similaridade Cosseno, Distância Euclidiana e de Manhattan, combinadas com a distância de Levenshtein, apresentam os mesmos resultados para a métrica de eficácia avaliada.

H_1 : As medidas de similaridade Cosseno, Distância Euclidiana e de Manhattan, combinadas com a distância de Levenshtein, apresentam resultados diferentes para a métrica de eficácia avaliada.

5.2.6 Seleção dos Objetos

Os objetos deste experimento são as descrições oriundas da base de NF-es públicas a serem classificadas. Vale ressaltar que, entre as notas, existem registros de produtos, tal como, por exemplo, algodão, que não é medicamento e auxilia na avaliação de Falsos Positivos e Verdadeiros Negativos. A base fornecida possui registros de 2019 a 2020, perfazendo 587.428 notas fiscais de produtos farmacêuticos.

Para o cálculo da amostra (Seward & Doane, 2014), foi considerado o cálculo de uma amostra para uma população infinita, perfazendo 800 descrições. Vale ressaltar que a amostra, em verdade, foi arredondada e é um pouco maior do que uma amostra para uma população infinita, a qual, considerando uma confiabilidade de 95% e uma margem de erro de 3,5%, seria de 784 descrições. Neste caso, considerando a mesma confiabilidade, a margem de erro é de

aproximadamente 3,46%. Calculada a amostra, do ponto de vista da seleção, esta foi feita de forma aleatória, por meio de uma função de randomização programada no Python, com a qual foram sendo sorteados números de linhas da base de dados de descrições, sendo 400 notas do ano de 2019 e 400 notas do ano de 2020.

5.2.7 Projeto do Experimento

Uma das métricas utilizadas neste trabalho foi a Acurácia, a qual exige o balanceamento dos dados das classes. Uma vez que os dados coletados já estavam balanceados, não foi necessário planejar a adoção de um método de balanceamento (Machado, 2007).

Além disso, neste experimento, foi utilizada uma adaptação da abordagem de *K-Fold Cross-Validation* (Hastie, Tibshirani & Friedman, 2011). O conjunto de descrições das NF-es públicas foi dividido em 30 partes, sendo obtidos 30 valores de cada métrica, para cada método avaliado. Posteriormente, foram calculadas as médias das métricas, para validação estatística.

Quanto à definição do limiar para a similaridade utilizada no trabalho, foi observado que as descrições dos medicamentos presentes na base de NF-es públicas apresentam, em sua grande maioria, apenas os nomes dos princípios ativos ou comerciais, contudo, na base anotada, as descrições possuem também os nomes dos laboratórios responsáveis pela fabricação. Dessa forma, a partir de testes iniciais e por observação, foi possível estabelecer os limiares iniciais para a similaridade do Cosseno, a distância de Manhattan e a Euclidiana: 0.7, 1.0 e 1.0, respectivamente.

A partir desse momento, foram realizadas diversas e árduas avaliações (automáticas e manuais) com faixas de valores abaixo e acima dos iniciais, com o intuito de calibrar o limiar do experimento, porém os resultados produzidos foram muito ruins. Ou seja, os resultados foram mais promissores para correspondências entre descrições com Cosseno ≥ 0.7 e distâncias ≤ 1 . Estes valores foram usados para efetuar as classificações a partir dos 3 métodos avaliados.

5.2.8 Instrumentação

O processo de instrumentação consistiu na configuração do ambiente para a realização do experimento controlado. Os materiais/recursos utilizados foram: biblioteca *Scikit-learn* de aprendizado de máquina, de código aberto; linguagem de programação Python; o *framework Django*, para otimizar o processo de desenvolvimento; o pacote *SQLAlchemy*, responsável por realizar a interface geral com a base de dados; e o SPSS, para avaliação dos resultados. Por fim, para construção da base anotada, foi utilizado o *BeautifulSoup*, aplicando a técnica de *Web Scraping*, responsável pela obtenção de informações relevantes de medicamentos, publicadas na *Web*.

6 OPERAÇÃO DO EXPERIMENTO

6.1 Preparação

Os dados anotados utilizados neste experimento foram obtidos a partir da coleta em bulas médicas de diferentes fontes de dados, com as descrições corretas dos medicamentos. Para obtenção dessas descrições, foi utilizada uma técnica chamada *Web Scraping*. Desta forma, foi desenvolvido um programa na linguagem Python, o qual coletou 33.074 (trinta e três mil e setenta e quatro) amostras de medicamentos, para 2.171 (dois mil, cento e setenta e um) diferentes princípios ativos.

Antes de utilizar os métodos de similaridade, foi realizada a etapa de pré-processamento de texto em ambas as bases, a qual manteve apenas informações necessárias para efetuar o cálculo de similaridade, removendo qualquer informação desnecessária. Após a normalização das descrições, foram extraídos os possíveis *tokens* para cada etapa. A seguir, é possível observar dois exemplos de medicamentos, com suas descrições completas e corretas, e duas descrições presentes nas NF-es públicas, antes e depois do pré-processamento e classificação por etapas.

Exemplos de medicamentos presentes na base anotada:

- Exemplo 01:
 - *Antes do processamento*: Citrato de Fentanila Hipolabor 50mcg/mL, caixa com 25 ampolas com 5mL

- *Depois do processamento*: citrato fentanila hipolar 50mcgml 25 ampola 5ml
 - *Tokens selecionados*: Etapa 1 (medicamento): citrato fentanila; Etapa 2 (dosagem): 50mcgml 5ml; Etapa 3 (forma farmacêutica): ampola; Etapa 4 (quantidade): 25.
- Exemplo 02:
 - *Antes do processamento*: Amoxicilina Aché 250mg/5mL, caixa com 1 frasco com pó para suspensão de uso oral + 1 seringa dosadora (capacidade do frasco de 150mL).
 - *Depois do processamento*: amoxicilina ache 150ml 250mg 5ml frasco seringa po 1.
 - *Tokens selecionados*: Etapa 1 (medicamento): amoxicilina ache; Etapa 2 (dosagem): 150ml 250mg 5ml; Etapa 3 (forma farmacêutica): frasco po; Etapa 4 (quantidade): 1.

Exemplos de medicamentos presentes na base de NF-es públicas:

- Exemplo 01:
 - *Antes do processamento*: SERTRALINA 50MG C/490 CPR. (GEN)
 - *Depois do processamento*: sertralina 50mg 490 comprimido
 - *Tokens*: Etapa 1 (medicamento): sertralina; Etapa 2 (dosagem): 50mg; Etapa 3 (forma farmacêutica): comprimido; Etapa 4 (quantidade) 490.
- Exemplo 02:
 - *Antes do processamento*: (+)HIDROCLOROTIAZIDA 25MG C/400 CPRS
 - *Depois do processamento*: hidroclorotiazida 25mg comprimidos 400
 - *Tokens selecionados*: Etapa 1 (medicamento): hidroclorotiazida; Etapa 2 (dosagem): 25mg; Etapa 3 (forma farmacêutica): comprimido; Etapa 4 (quantidade): 400.

6.2 Execução

A execução do experimento consistiu na realização da classificação das descrições dos medicamentos presentes nas NF-es públicas, de acordo com as descrições corretas presentes na

base anotada, conforme planejado na Subseção 6.2, para cada método avaliado, utilizando o dicionário discutido na Subseção 3.

Para cada nível hierárquico extraído e classificado da NFe pública, verificava-se, num processo árduo, se a descrição condizia com a etapa correspondente do medicamento, sendo atribuído um rótulo para cada etapa, segundo a classificação apresentada na Seção 2.5. Por exemplo: primeira etapa, medicamento, classificada corretamente (TP1, ou *True Positive* para etapa 1); segunda etapa, dosagem, classificada corretamente (TP2); terceira etapa, forma farmacêutica, classificada corretamente (TP3); e quarta etapa, quantidade, classificada corretamente (TP4). O mesmo vale para os outros possíveis resultados, tais como Falsos Positivos e Negativos (FP e FN) e Verdadeiros Negativos (TN).

O processo de atribuição dos rótulos foi executado por meio da comparação a “olho nu”, conforme exemplo apresentado na Tabela 1. Considerando 4 etapas e 800 registros analisados manualmente, esse processo foi realizado 3.200 vezes.

Descrição na NFe Pública	Etapa de Classificação	Princípio Ativo	Medicamento Classificado Até a Etapa Correspondente	Rótulo
Tramadol 50mg C/500 Caps	Medicamento	Cloridrato de Tramadol	Cloridrato de Tramadol Cápsula Eurofarma 50mg, caixa com 10 cápsulas gelatinosas	TP1
Tramadol 50mg C/500 Caps	Dosagem	Cloridrato de Tramadol	Cloridrato de Tramadol Cápsula Eurofarma 50mg, caixa com 10 cápsulas gelatinosas	TP2
Tramadol 50mg C/500 Caps	Forma Farmacêutica	Cloridrato de Tramadol	Cloridrato de Tramadol Cápsula Eurofarma 50mg, caixa com 10 cápsulas gelatinosas	TP3
Tramadol 50mg C/500 Caps	Quantidade	Cloridrato de Tramadol	Cloridrato de Tramadol Cápsula Hipolabor 50mg, caixa com 500 cápsulas gelatinosas duras (embalagem hospitalar)	TP4

Tabela 1 – Verificação de uma classificação por etapa

6.3 Validação dos Dados

Para análise, interpretação e validação, foram utilizados 03 (três) tipos de testes estatísticos: *Shapiro-Wilk*, *Friedman* e *Wilcoxon* (vide próxima seção).

O teste de *Shapiro-Wilk* (Shapiro & Wilk, 1965) foi utilizado para testar a normalidade dos dados, já os testes de *Friedman* (Friedman, 1937) e *Wilcoxon* (Wilcoxon, 1945) foram utilizados para a comparação das medianas da Medida-F1 e da Acurácia, uma vez que, para ambas as métricas, no caso dos resultados da medida de similaridade do cosseno, o teste de normalidade indicou uma distribuição de dados não-normal. Todos os testes estatísticos foram feitos utilizando a ferramenta SPSS (IBM, 2021).

7 RESULTADOS

7.1 Análise e Interpretação dos Dados

Para responder às questões de pesquisas elencadas na Seção 6.2, ocorreu a execução do experimento e foram obtidos os resultados das classificações alcançadas por cada uma das três medidas utilizadas, as quais tiveram o intuito de verificar as correspondências das descrições dos medicamentos avaliados. Na Figura 5 e na Tabela 2, são apresentadas as médias das métricas.

Comparativo das Métricas entre as Medidas

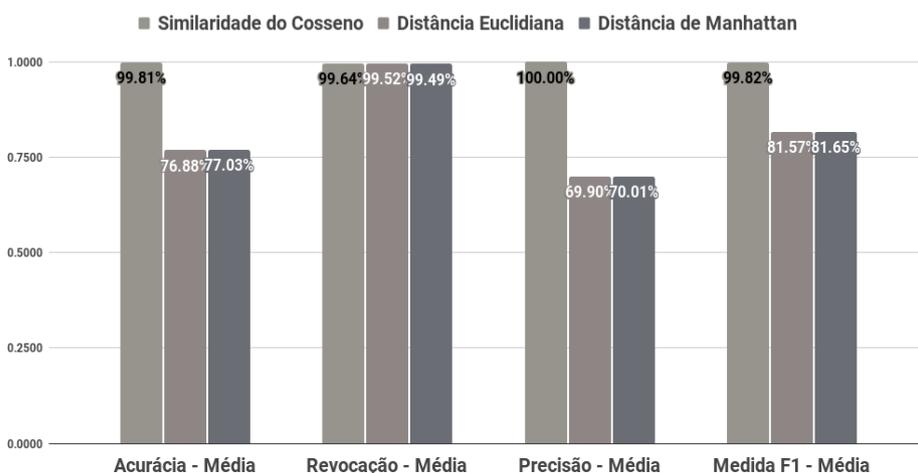


Figura 5 – Comparativo das médias das métricas entre as medidas.

Medidas	Acurácia	Revocação	Precisão	Medida-F1
Similaridade do Cosseno	0.9981	0.9964	1.0000	0.9982
Distância Euclidiana	0.7688	0.9952	0.6990	0.8157
Distância de Manhattan	0.7703	0.9949	0.7001	0.8165

Tabela 2 – Comparativo das médias das métricas entre as medidas

Estes resultados foram utilizados para responder à questão de pesquisa definida na subseção 6.2. Como é perceptível, os métodos obtiveram médias de Acurácia distintas e as distâncias Euclidiana e de Manhattan obtiveram médias abaixo da Similaridade do Cosseno, todavia, não é possível fazer essas afirmações sem evidências estatísticas suficientemente conclusivas.

Neste sentido, definiu-se um nível de significância (α) de 0,05 para todo o experimento. Ao aplicar o teste de *Shapiro-Wilk*, para análise da normalidade da distribuição dos dados, foram obtidos os *p-values* apresentados na Tabela 3, na qual se observa que as distribuições das métricas para a Similaridade do Cosseno alcançaram valores abaixo do nível de significância adotado, ou seja, a hipótese de normalidade foi refutada.

Medidas	Acurácia	Medida-F1
Similaridade do Cosseno	0.000	0.000
Distância Euclidiana	0.420	0.865
Distância de Manhattan	0.675	0.972

Tabela 3 – Resultado do teste de shapiro-wilk, para análise da normalidade dos dados

Uma vez que o teste de normalidade indicou distribuições não-normais, para o caso do Cosseno, foi necessário aplicar o teste não paramétrico de *Friedman*, com o objetivo de verificar se há pelo menos uma diferença entre as medianas das métricas. O teste foi aplicado para a Acurácia e para a Medida-F1, ressaltando que esta última métrica harmoniza as métricas de precisão e de revocação.

Com a aplicação do teste de *Friedman*, verificou-se um *p-value* fortemente menor que o nível de significância adotado, resultando em aproximadamente 0 (zero), tanto para a métrica de Acurácia quanto para a Medida-F1. Desta forma, foi possível confirmar a evidência das diferenças entre as medianas, ou seja, as hipóteses H_0 , de que os métodos possuem a mesma Medida-F1 e a mesma Acurácia, foram rejeitadas, dentro do contexto do experimento realizado.

Assim sendo, foi evidenciado que ao menos um método se diferencia dos demais, porém, não é possível afirmar qual é o mais discrepante. Para isso, foi utilizado o teste de *Wilcoxon*. A Tabela 4 apresenta os resultados deste teste, evidenciando que, após uma análise “*post-hoc*”, ou posterior, aplicando a correção de *Bonferroni* ($\alpha = \alpha / 6$), para a Acurácia e para a Medida-F1, encontramos a seguinte ordem relacionada aos métodos: Similaridade do Cosseno > Distância de Manhattan; Distância Euclidiana. Ou seja, não houve significância estatística para rejeitar que as Distâncias de Manhattan e Euclidiana sejam equivalentes, todavia, a Similaridade do Cosseno foi superior às duas concorrentes. Em outras palavras, cada método foi comparado aos outros, em avaliações dois a dois. A significância estatística é verificada nas linhas em que o *p-value* é menor que 0,05. Cada linha da Tabela 4, por exemplo, testa a hipótese nula de que as distribuições da Medida-F1 do Método 1 e da Medida-F1 do Método 2 são iguais.

Comparações Dois a Dois		
Método 1-Método 2	<i>p-value</i> Medida-F1	<i>p-value</i> Acurácia
Similaridade do Cosseno - Distância Euclidiana	0.000	0.000
Similaridade do Cosseno - Distância de Manhattan	0.000	0.000
Distância Euclidiana - Distância de Manhattan	0.169	0.307

Tabela 4 – Teste de *Wilcoxon*, dois a dois

7.2 Ameaças à Validade

Embora os resultados do experimento tenham se mostrado satisfatórios, os mesmos apresentam ameaças à sua validade que devem ser comentadas.

Ameaças à validade externa:

Apesar da similaridade existente entre as Notas Fiscais Públicas, os resultados apresentados são baseados em um banco de dados de Notas Fiscais Públicas, com CNPJs e Razões Sociais mascaradas, cedido pelo Ministério Público de Sergipe, no qual estão presentes apenas notas do referido estado. Com dados de outros estados, podem ocorrer variações nas métricas.

Ameaças à validade de construção:

- (1) Natureza das descrições utilizadas na base anotada: é necessário que haja uma atualização com medicamentos recém-surgidos no mercado, de modo que o classificador seja capaz de identificar os novos produtos. Para mitigar essa ameaça, é permitido o registro de novos medicamentos, a partir de alertas e buscas periódicas em bases públicas de remédios.
- (2) Implementações dos algoritmos de similaridade e de dissimilaridade: as implementações dos métodos abordados neste trabalho devem atender aos requisitos teóricos apresentados, assim, alterações podem comprometer seus resultados. Por garantia, utilizou-se a biblioteca Scikit-Learn (Pedregosa Varoquaux et al., 2011), a qual possui citações em estudos relacionados.

8 CONCLUSÃO

Nas operações de controle a respeito sobre gastos com saúde, para atestar se o medicamento licitado condiz com o que foi adquirido, faz-se necessária a utilização das Notas Fiscais Eletrônicas (NF-es), nas quais devem constar, para que o produto seja identificado, as especificações das subclasses, tais como: os princípios ativos, as formas de apresentações, as dosagens e a quantidade de unidades por embalagem de cada lote licitado. O descumprimento da especificação de uma ou mais dessas subclasses pode facilmente resultar em prejuízo aos cofres públicos e estratégias devem ser implementadas pelos gestores, com a finalidade de mitigar tais riscos potenciais. Além disso, nem sempre o NCM da NFe condiz com a descrição de uma mercadoria e tal associação falha impede a contabilização precisa dos produtos em

investigações e pode acarretar uma cobrança indevida do imposto, bem como a sua não-arrecadação.

Neste contexto, visando a necessidade de automatizar a classificação dos medicamentos, independente do NCM, e o reconhecimento das subclasses hierárquicas presentes nas suas descrições, este trabalho apresentou os resultados de um classificador que apoiará a detecção de fraudes. Denominado de DescMiner, ele utiliza uma base anotada com as descrições corretas dos medicamentos e combina medidas de similaridade. Para avaliar sua viabilidade, foi realizado em um processo experimental, pois ferramentas realmente eficazes se mostram cada vez mais necessárias, à medida que a corrupção também evolui com os avanços tecnológicos.

O experimento verificou significância estatística para a superioridade da Similaridade do Cosseno no reconhecimento dos medicamentos, com os seguintes resultados médios: acurácia de 99.81%, revocação de 99.64%, precisão de 100% e Medida-F1 de 99.82%. Além destes resultados, foi possível evidenciar que é factível automatizar a extração das subclasses hierárquicas das descrições dos medicamentos, permitindo a identificação precisa e detalhada dos produtos adquiridos, a qual viabiliza que os investigadores realizem análises de desvios.

Para trabalhos futuros, está em fase de testes uma interface para apresentar indícios de compras irregulares, permitindo *insights* visuais e maiores assertividade e agilidade no processo de tomada de decisão dos auditores fiscais fazendários e dos investigadores dos Ministérios Públicos, Grupos de Atuação Especial de Combate ao Crime Organizado (GAECOs) e Delegacias Especiais.

REFERÊNCIAS

- Al-Anzi, F. S. & Abuzeina, D. (2017). Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing. *Journal of King Saud University Computer and Information Sciences*, 29(2), pp. 189-195. doi: 10.1016/j.jksuci.2016.04.001.
- Agência Nacional de Vigilância Sanitária, Anvisa (2011). *Vocabulário Controlado de Formas Farmacêuticas, Vias de Administração e Embalagens de Medicamentos* [Manual]. Agência Nacional de Vigilância Sanitária, Brasília, DF, Brasil.
- Aquino, D.S. (2008). Por que o uso racional de medicamentos deve ser uma prioridade. *Ciência & Saúde Coletiva*, Rio de Janeiro, RJ, Brasil, 13.

- Basili, V. R. & Weiss, D. M. (1984). A methodology for collecting valid software engineering data. *IEEE Transactions of Software Engineering*, 10(6). doi: 0098-5589/84/1100-0728\$01.00.
- Bass, L., Clements, P. & Kazman, R. (2003). *Software architecture in practice* (3a ed.). New Jersey: Addison-Wesley Longman Publishing Co.
- Batista, R. de A., Bagatini, D. D. S. & Frozza, R. (2017). Classificação Automática de Códigos NCM utilizando o Algoritmo de Naïve Bayes. *Revista Brasileira de Sistemas de Informação*, 11(2), pp. 4-29. Recuperado de <http://www.seer.unirio.br/index.php/isy/article/view/6401>.
- Constituição da República Federativa do Brasil, de 05 de outubro de 1988. Diário Oficial da União. Brasília, DF.
- Emenda Constitucional n. 42, de 19 de dezembro de 2003. Diário Oficial da União. Brasília, DF.
- Câmara de Regulação do Mercado de Medicamentos (2004). Resolução N 2, de 5 de Março de 2004.
[http://portal.anvisa.gov.br/documents/374947/2932039/Resolu%C3%A7%C3%A3o+n%C2%BA+2+de+5+de+mar%C3%A7o+de+2004+\(PDF\).pdf/b6d68347-a134-4465-a2f1-e5ed0cab747](http://portal.anvisa.gov.br/documents/374947/2932039/Resolu%C3%A7%C3%A3o+n%C2%BA+2+de+5+de+mar%C3%A7o+de+2004+(PDF).pdf/b6d68347-a134-4465-a2f1-e5ed0cab747).
- Carvalho, R., Paiva, E., Rocha, H. & Mendes, G. (2014). Using Clustering and Text Mining to Create a Reference Price Database. Learning & Nonlinear Models. *Journal of the Brazilian Society on Computational Intelligence*, 12(1), pp. 38-52. <http://abricom.org.br/wp-content/uploads/sites/4/2016/07/vol12-no1-art3.pdf>.
- Carvalho, R., Paiva, E., Rocha, H. & Mendes, G. (2013). Methodology for Creating the Brazilian Government Reference Price Database. *X Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*.
https://www.researchgate.net/publication/270820493_Methodology_for_Creating_the_Brazilian_Government_Reference_Price_Database.
- Colaço Júnior, M. & Oliveira, R. N. (2018). Experimental Analysis of Stemming on Jurisprudential Documents Retrieval. Information. *ICEIS 2017 19th International Conference on Enterprise Information Systems*. Porto, Portugal. doi: 9. 28. 10.3390/info9020028.
- Controladoria-Geral da União. Portal da Transparência – Orçamento Anual. <http://www.portaltransparencia.gov.br/orcamento?ano=2020>.
- Feldman, R. & Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press.
- Federação de Indústrias do Estado de São Paulo (2010). *Corrupção: custos econômicos e propostas de combate* [Relatório]. São Paulo, SP, Brasil.
- Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, 32(200), pp. 675-701. doi: 10.2307/2279372.
- Hastie, T., Tibshirani, R. & Friedman, J. (2011). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2a ed.). New York: Springer.

- Instituto Brasileiro de Geografia e Estatística. Produto Interno Bruto – PIB. <https://www.ibge.gov.br/explica/pib.php#:~:text=O%20%20PIB%20do%20Brasil%20em,das%20Unidades%20da%20Federa%C3%A7%C3%A3o%20brasileiras>.
- IBM. SPSS software – Bridge the gap between data science and data understanding. Disponível <https://www.ibm.com/analytics/spss-statistics-software>.
- Instituto Brasileiro de Planejamento e Tributação. Boletim do Impostômetro mostra que ICMS tem a maior fatia de impostos recolhidos no país. <https://ibpt.com.br/noticia/2804/Boletim-do-Impostometro-mostra-que-ICMS-tem-a-maior-fatia-de-impostos-recolhidos-no-pais>.
- Kratcoski, P. C. & Edelbacher, M. (2018). *Fraud and Corruption: Major Types, Prevention, and Control* (1a ed.) (p. 3010. New York: Springer.
- Levene, H. (1960). Robust tests for equality of variances. *International Journal of Machine Learning and Cybernetics*, pp. 278-292.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10, pp. 707-710.
- Li, B. & Han, L. (2013). *Distance weighted cosine similarity measure for text classification. International Conference on Intelligent Data Engineering and Automated Learning* (pp. 611-618). Heidelberg: Springer.
- Machado, E. (2007). *Um estudo de limpeza em base de dados desbalanceada e com sobreposição de classes* (Dissertação de mestrado). Universidade de Brasília, Brasília, DF, Brasil.
- Madeira, R. (2015). *Aplicação de técnicas de mineração de texto na detecção de discrepâncias em documentos fiscais* (Dissertação de mestrado). Fundação Getúlio Vargas, Rio de Janeiro, RJ, Brasil.
- Mankiw, N. G., Romer, D. & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, 107(2), pp. 407-437. doi 10.2307/2118477.
- Nota Técnica n. 002 da Receita Federal do Brasil de Novembro de 2016. <https://www.nfe.fazenda.gov.br/portal/listaConteudo.aspx?tipoConteudo=tW+YMyk/50s=>.
- Oliveira, R. S. (2017). *Utilizando o algoritmo de levenshtein e mongodb em dados de licitações governamentais* (Monografia). Universidade Federal de Mato Grosso, Cuiabá, MT, Brasil.
- Pedregosa, F et al. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, pp. 2825-2930. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>.
- Portaria n. 3.992 do Ministério da Saúde, de 28 de dezembro de 2017. Diário Oficial da União, Brasília, DF, Brasil.
- Resolução Normativa n. 259 da Agência Nacional de Saúde, de 17 de junho de 2011. Diário Oficial da União. Brasília, DF.
- Ribeiro, L. V. et al. (2018) Reconhecimento de entidades nomeadas em itens de produto da nota fiscal eletrônica. *Revista do Tribunal de Contas do Estado de Minas Gerais*, 36.
- Salton, G. & Buckley, C. (1988) Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing & Management*, 24, pp. 513-523. doi 10.1016/0306-4573(88)90021-0.
- Salton, G., Fox, E. A. & Wu, H. (1983). Extended Boolean information retrieval. *Commun. Association for Computing Machinery* 26(11), pp. 1022–1036. doi: 10.1145/182.358466.
-  **Journal of Management & Technology, Vol. 24, n. 1, p. 112-148, 2024**

- Seward, L. E. & Doane, D. P. (2014). *Estatística Aplicada à Administração e Economia* (4a ed.). AMGH Editora: Rio Grande do Sul.
- Secretaria de Estado da Fazenda de Sergipe. Portal da Transparência. <http://www.transparenciasergipe.se.gov.br/>.
- Shapiro, S. S. & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *International Journal of Machine Learning and Cybernetics*, 52(3/4), pp. 591-611. doi: 10.2307/2333709.
- Silva, L. P. A. (2019). *Orçamento temático de acesso a medicamentos: Análise de 10 anos de recursos federais destinados à assistência farmacêutica* [Manual]. Instituto de Estudos Socioeconômicos, Brasília, DF, Brasil.
- Skiena, S. S. (2017). *Distance and network methods. The data Science Design Manual* (pp. 303-339). New York: Springer.
- Sulpino, F. V. (2018). *Evolução do gasto com medicamentos do Sistema Único de Saúde no período de 2010 a 2016* [Relatório]. Instituto de Pesquisas Econômicas Aplicadas, Brasília, DF, Brasil.
- Sulpino, F. V.; PIOLA, S.; BENEVIDES, R. (2019). *Vinculação orçamentária do gasto em saúde no Brasil: resultados e argumentos a seu favor* [Relatório]. Instituto de Pesquisas Econômicas Aplicadas, Brasília, DF, Brasil.
- Transparência Brasil (2019). Métodos de detecção de fraude e corrupção em contratações públicas. [Manual]. <https://www.transparencia.org.br/downloads/publicacoes/Metodos%20Detec%C3%A7%C3%A3o%20de%20Fraude.pdf>.
- Transparency International (2019). *Corruption Perceptions index 2019: Full source description* [Relatório]. Transparency International, Berlim, Alemanha.
- Travassos, G. H., Gurov, D. & Amaral, E. (2020). *Introdução à Engenharia de Software Experimental* [Relatório]. Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.
- United Nations (2003). United Nations Convention Against Corruption. *United Nations Office on Drugs and Crime*, Vienna, Áustria.
- Vijayarani, S & Janani, M. (2016). Text Mining: open Source Tokenization Tools – An Analysis. *Advanced Computational Intelligence: An International Journal*, 3(1), pp. 11. doi: 10.5121/acii.2016.3104.
- Weiss, S., Indurkha, N., Zhang, T. & Damerou, F. (2004). Text Mining: Predictive Methods for Analyzing Unstructured Information. doi: 10.1007/978-0-387-34555-0.
- Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6), pp. 80-83. doi:10.2307/3001968.
- World Health Organization (1972). *International Drug Monitoring, the Role of National Centers* [Relatório]. World Health Organization, Genebra, Suíça.
- Zhu, W. et al. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. NESUG proceedings: health care and life sciences. *Shanghai Arch Psychiatry*, 19, pp. 67. <https://www.semanticscholar.org/paper/1-Sensitivity-%2C-Specificity-%2C-Accuracy-%2C-Associated-Zhu-Zeng/d1e5c3097daf99db2c8dce3ac0edc3c5ade41460>.