

Prova de Conceito de um Classificador de OPMEs em Notas Fiscais

Wesckley Gomes^{1,2,3}, Methanias Colaço Júnior^{1,2,3}, Raphael Fontes^{1,2}, Rodrigo Silva¹, Bruno Nunes⁴, Caldeira Silva^{1,2}, Jailton Paiva^{1,2}, Ricardo Valetim^{1,2}

¹Laboratório de Inovação Tecnológica em Saúde (UFRN)

²Núcleo Avançado de Inovação Tecnológica (IFRN)

³PROCC (UFS)

⁴Auditoria Geral do Sistema Único de Saúde – Ministério da Saúde - Brasília/DF, Brasil

{wesckley.gomes, raphaelf.ti}@gmail.com, mjrse@hotmail.com,
{rodrigo.silva, caldeira.silva, jailton.paiva, ricardo.valetim}@lais.huol
.ufrn.br, bruno.nunes@saude.gov.br

Abstract. Context: *Despite the advancement of technology, many services and information systems, especially in the public sector, still use unstructured descriptions in natural language of products, services or events, making their classification and analysis difficult. For efficient audits, it is necessary automatically classify and total invoices issued for the purchase of products.* **Objective:** *Preliminarily evaluate a tool, based on artificial intelligence, to classify OPME invoices.* **Method:** *A proof-of-concept study was carried out in the OPMinEr project.* **Results:** *The results show that it is possible to identify and classify OPMEs in invoices.* **Conclusion:** *The use of artificial intelligence techniques helped to mitigate the problem of classifying and analyzing invoices, thus helping in auditing, investigation and anti-corruption processes.*

Resumo. Contexto: *Apesar do avanço da tecnologia, muitos serviços e sistemas de informação, especialmente no setor público, ainda utilizam descrições não estruturadas em linguagem natural de produtos, serviços ou eventos, dificultando suas classificações e análises. Para auditorias eficientes, é necessário classificar e totalizar automaticamente faturas emitidas para compra de produtos.* **Objetivo:** *Avaliar preliminarmente uma ferramenta, baseada em Inteligência Artificial (IA), o OPMinEr, para selecionar e classificar notas fiscais de OPMEs (Órteses, Próteses e Materiais Especiais).* **Método:** *Prova de conceito do OPMinEr.* **Resultados:** *Com uma acurácia de 99%, os resultados mostraram que é possível identificar e classificar OPMEs em notas fiscais.* **Conclusão:** *O uso de IA ajudou a mitigar o problema de classificação de notas, beneficiando processos de auditoria, investigação e combate à corrupção na área de saúde.*

1. Introdução

O avanço da internet e das tecnologias digitais impulsionaram o comércio eletrônico e o surgimento da nota fiscal eletrônica (NFe). Isso trouxe grandes oportunidades de análise de dados relativos à economia dos países, como, por exemplo, no âmbito da

Administração Pública, possibilitando inúmeras ações de controle, tanto do ponto de vista do montante (materialidade) das compras públicas quanto da legalidade (divergências de preços, compras em desacordo com limites legais, controle de licitações, entre outras) (Ribeiro, et al., 2018). Apesar deste avanço, muitos processos ainda utilizam descrições de produtos e serviços em linguagem natural não estruturada, sem a preocupação com a exatidão dos termos.

A Administração Fazendária utiliza o campo Nomenclatura Comum do Mercosul (NCM) (Batista, Bagatini, & Frozza, 2017), presente na NFe, como meio para definir qual alíquota incidirá sobre o item. Todavia, a codificação errada e fragilidades nas descrições de produtos, serviços, diárias e centros de custo é um problema geral e recorrente na Administração Pública (Santos, et al., 2015) (Ribeiro, et al., 2018). Com as Notas Fiscais, similarmente, nem sempre o NCM inserido pelo contribuinte condiz com a descrição da mercadoria vendida, o que impede a identificação automática e precisa de produtos, prejudicando investigações e auditorias.

Em se tratando da área de Saúde, o Mercado de Órteses, Próteses e Materiais Especiais (OPME) é marcado pela quantidade de produtos existentes e pela diversidade das tecnologias utilizadas. É um setor com o domínio de empresas multinacionais, porém também conta com pequenas e médias empresas locais, o que causa uma heterogeneidade desses dispositivos, concentrando o conhecimento em especialistas e produzindo assimetria de informação (Carmo de Souza Cruz, 2022). Isto, alinhado com a dificuldade de padronização da nomenclatura na comercialização de OPMEs, dificulta a comparação dos produtos, seja para uniformização de preços ou padronização de uso, o que dá margem para atos de corrupção.

Em janeiro de 2015, a imprensa nacional noticiou indícios da ocorrência de esquema fraudulento envolvendo a compra e utilização de órteses, próteses e materiais especiais (OPME), o que ficou conhecido como “máfia das próteses”. O suposto esquema envolveria uma série de atores – fabricantes, distribuidores, hospitais, médicos e advogados – e diversos tipos de irregularidades – venda de dispositivos com sobrepreço, recebimento de comissões irregulares, fraudes, desvios, entre outros (TCU 2016).

Com base neste contexto, o presente artigo tem como objetivo avaliar preliminarmente uma ferramenta, baseada em inteligência artificial, para identificar e classificar notas fiscais de OPME, o OPMinEr, apresentando um estudo de prova de conceito (PoC). A ferramenta utiliza o algoritmo *Naïve Bayes Multinomial*, com Frequência Inversa, e, apenas a partir da descrição das notas, classifica-as em: Tipo de OPME, Classe à qual o produto pertence, e Procedimento do Sistema Único de Saúde (SUS) no qual o produto pode ter sido usado. A classificação permite a identificação automática de OPMEs pelos auditores do Departamento Nacional de Auditoria do SUS, AudSUS, habilitando a execução de trilhas de auditoria de OPMEs, para as quais o classificador é uma condição *sine qua non*.

O presente trabalho está subdividido da seguinte forma: na seção 2, será descrita a motivação deste artigo; na seção 3, serão apresentados alguns trabalhos relacionados ao tema; a seção 4 apresenta os materiais e métodos utilizados; a seção 5 apresenta a prova de conceito, e, finalmente, na seção 6, as considerações finais e trabalhos futuros são apresentados

2. Motivação

As descrições das OPMEs em notas fiscais emitidas são apresentadas em texto escrito em linguagem natural não estruturada, de forma livre pelo fornecedor, o que pode causar a omissão de informações ou dificultar a identificação dos produtos, uma vez que os códigos utilizados não são confiáveis. Isto, alinhado com o grande volume de notas fiscais emitidas, dificulta o processo de auditoria e investigação de fraudes por parte dos órgãos competentes. Esse problema não é exclusivo do Brasil, em 2017, por exemplo, o Tribunal de Contas Europeu comunicou que as formas de evasão de pagamento amplamente aplicadas são a subavaliação, classificação errada ao mudar para uma classificação de um produto com alíquota mais baixa, e a descrição errada das mercadorias (Spichakova & Haav, 2020).

3. Trabalhos Relacionados

O trabalho de Batista, Bagatini, & Frozza (2017) teve como objetivo a classificação automática de códigos NCM, por meio das descrições presentes nas NFs emitidas no estado do Rio Grande do Sul. As bases de dados foram distinguidas pelas suas complexidades, avaliadas como simples, média ou complexa, as quais obtiveram, respectivamente, 98%, 90% e 83% de acurácia. O trabalho de Correa & Leal (2018) teve como objetivo identificar sobrepreço na aquisição de medicamentos adquiridos pelo governo federal do Brasil, disponibilizados pelos portais de transparência, por meio de técnicas de mineração de texto e agrupamento para classificar os produtos. Apesar do problema das OPMEs ser considerado mais complexo, a classificação de forma automática dos medicamentos assemelha-se ao trabalho aqui proposto, além disso, a identificação de sobrepreço é uma das trilhas de auditoria viabilizadas pelo *OPMinEr*.

4. Materiais e métodos

Inicialmente, foi implementada a ferramenta *OPMinEr*. O algoritmo utilizado nesta primeira versão foi o *Naïve Bayes Multinomial*, com Frequência Inversa, o qual foi selecionado baseado nos resultados de um mapeamento sistemático sobre classificação de descrições associadas a códigos errados (Gomes & Colaço, 2022), e nos resultados apresentados em (Santos, et al., 2015), os quais evidenciaram, entre os algoritmos mais eficientes para mineração de textos, a boa eficácia desta abordagem.

De fato, a ferramenta realiza três classificações, em primeiro lugar é feita a predição do Tipo de OPME: Órtese, Prótese ou Material Especial. Neste caso, quando uma Nota não é uma OPME, esta é classificada como Outros. A próxima classificação indica à qual classe pertence o produto. Existem 83 classes diferentes: Stent; Cateter; Prótese Ortopédica; Cânula, etc. Por fim, é realizada a última classificação, a qual prediz o Procedimento do Sistema Único de Saúde (SUS) no qual o produto pode ter sido usado. Existem 319 procedimentos diferentes: Balão Destacável; Cateter Atrial / Peritoneal; Fio de Kirschner; Haste Intramedular Tibio-Tarsica, etc.

Ato contínuo, foi realizada uma análise e discussão da ferramenta, utilizando o método de prova de conceito. Uma prova de conceito (PoC), do inglês *proof of concept*, é a evidência de que um produto ou um serviço potencial pode atender determinado objetivo, documentada a partir de testes, servindo para verificar e validar um projeto antes que este seja executado e disponibilizado para uso geral (Cruz et al., 2022). Na

próxima seção, de forma autocontida, a prova de conceito e sua metodologia são detalhadas.

5. Prova de Conceito

5.1 Objetivo

Avaliar preliminarmente a ferramenta *OPMinEr*, na tarefa de identificação e classificação de notas fiscais de OPMEs.

5.2 Planejamento

5.2.1 Seleção de Participantes

Os dados utilizados para a realização da prova de conceito foram obtidos de notas fiscais disponibilizadas pelo governo do Rio Grande do Norte (RN), custodiadas pelo AudSUS, Laboratório de Inovação Tecnológica em Saúde (LAIS) e Ministério Público Federal. Ao todo, foram 465.726 notas fiscais, que correspondem ao período de janeiro de 2020 a maio de 2022. Pelo fato de existirem bem mais notas de produtos em geral, os quais não são OPME, foi selecionada uma amostra de 4.549 registros, mantendo uma proporção aproximada da quantidade de notas de OPMEs.

5.2.2 Dicionário (Modelo de Conhecimento)

O modelo de conhecimento utilizado para aprendizado de como as OPMEs são descritas foi composto com registros oriundos da Agência Nacional de Vigilância Sanitária (Anvisa), cadastro oficial de materiais do governo federal (Catmat), descrições avulsas (inseridas manualmente) escritas por auditores e descrições das próprias notas fiscais do Rio Grande do Norte (RN), rotuladas manualmente. A Similaridade do Cosseno foi utilizada para localizar OPMEs de forma semiautomática nos dados da Anvisa e do Catmat, em comparação com os procedimentos realizados pelo SUS. Descrições com Cosseno acima de 0,6, de um máximo de 1, foram conferidas manualmente e inseridas no dicionário.

Além disso, um dicionário de conflação foi criado, o qual uniformiza palavras. Por exemplo, quando a palavra “parafus.” é encontrada numa nota, esta é automaticamente transformada em “parafuso”.

5.2.3 Instrumentação

Os materiais e recursos utilizados foram:

- Ferramenta *OPMinEr*; PostgreSQL 12.12; Python 3.10.8; Django 4.1.7

5.3 Operação

5.3.1 Preparação

Em síntese, foi preparado o ambiente para a realização da prova de conceito, ou seja, o carregamento de todos os dados para o banco de dados.

5.3.2 Execução

O processo foi iniciado pelas tarefas de normalização e conflação das descrições, com intuito de padronização e mitigação de erros de classificação. 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, artigos, etc). Ao contínuo, foi iniciado o processo de transformação dos dados em um modelo vetorial.

Após a transformação dos dados em um modelo vetorial, foi realizado o treinamento e avaliação do modelo, utilizando uma técnica chamada *k-fold cross validation*. O processo de validação cruzada envolve a divisão do conjunto de dados disponível em várias partes (*folds*), geralmente em k partes iguais. Para cada iteração do processo de validação cruzada, uma das k partes é usada como conjunto de teste, enquanto as outras k-1 partes restantes são usadas como conjunto de treinamento. Neste trabalho, os dados do dicionário foram divididos em 10 conjuntos (*10-fold*), ou seja, o modelo foi treinado e testado dez vezes, perfazendo, para cada métrica avaliada, a média das 10 interações. Vale ressaltar que como a ferramenta realiza três tipos de classificação diferentes, foi necessário construir e treinar três modelos distintos. Ao término da execução, foram obtidas as seguintes métricas: acurácia, sensibilidade, medida-F1, precisão e tempo médio de execução. Os resultados desses dados coletados serão apresentados na próxima seção.

5.4 Resultados

Após a etapa de execução, foram obtidos o tempo médio e métricas para cada tipo de classificação. A Tabela 1 apresenta o tempo médio de execução de todo o processo de classificação, para cada modelo. O tempo médio de pré-processamento é o tempo levado para pré-processar as descrições e gerar o modelo vetorial, sendo o mesmo para todos os modelos, uma vez os modelos utilizam os mesmos dados de entrada. É possível perceber que, apesar do ótimo desempenho apresentado, quanto maior o número de classes alvo, maiores são os tempos de treinamento e classificação.

	Pré Processamento	Treinamento	Classificação
Tipo	1,71 s	5,09 s	0,007 s
Classe	1,71 s	7,62 s	0,014 s
Procedimento	1,71 s	12,00 s	0,044 s

Tabela 1. Tempo médio de execução

A Tabela 2 apresenta as métricas após a execução da classificação, para os três modelos. A ferramenta teve um ótimo desempenho em relação às métricas obtidas. A abordagem obteve uma acurácia de 99% em relação ao modelo para classificar o Tipo de OPME. Isso já assegura a recuperação de quase todas as notas de OPME, permitindo a localização completa de produtos considerados mais importantes. Numa linha mais sofisticada de classificação, o modelo de associação a Procedimentos do SUS obteve uma boa acurácia, 92%, e uma precisão de 93%, mesmo considerando uma classificação para 319 Procedimentos distintos.

	Acurácia	Sensibilidade	Medida-F1	Precisão
Tipo	0,99	0,99	0,99	0,99
Classe	0,97	0,97	0,97	0,98
Procedimento	0,92	0,92	0,92	0,93

Tabela 2. Métricas do Processamento

Por fim, vale ressaltar que existe a hipótese de melhorias dos resultados, considerando que os dicionários serão incrementados com notas de todo o Brasil.

6. Considerações finais e trabalhos futuros

O problema envolvendo a codificação errada e descrições de OPMEs associadas a códigos errados em notas fiscais, bem como o uso destes artefatos para tributação, fiscalização e investigação envolve muitas áreas e tecnologias. Neste trabalho, foi apresentada uma prova de conceito de uma ferramenta preliminar, a qual identifica e classifica notas fiscais de OPMEs. A partir da descrição da NF, a ferramenta foi capaz de realizar três tipos de classificação: Tipo, Classe e Procedimento. Um vídeo ilustrando o funcionamento da ferramenta está disponível em <http://bit.ly/vopminer>. Além disso, é possível acessar a ferramenta por meio do link <http://bit.ly/opminer>, utilizando *admin* como usuário e senha.

Essas classificações otimizarão a realização de auditorias pelo AudSUS, as quais poderão encontrar indícios tais como preços anormalmente altos, quantidades de OPMEs compradas por habitante, etc. Finalmente, apesar da ferramenta contar apenas com dados do Rio Grande do Norte, os resultados foram promissores, principalmente para identificação de OPMEs e seus tipos específicos.

Como trabalhos futuros, o dicionário de dados será incrementado com informações de notas fiscais de outros estados, bem como outros tipos de algoritmos serão utilizados e confrontados, para criação de um metamodelo híbrido e ainda mais preciso.

Referências

- Batista, R. d., Bagatini, D. D., & Frozza, R. (2017). Classificação Automática de Códigos NCM utilizando o Algoritmo de Naïve Bayes. *Revista Brasileira de Sist. de Inf.*, 13, 4-29.
- Carmo de Souza Cruz, R. a. (2022). Análise do impacto do Banco de Preços em Saúde (BPS) para redução das assimetrias de informação dos preços de compras de Órteses, Prótese e Materiais Especiais (OPME). *JMPHC Journal of Manag. & Primary Health Care*, 14.
- Correa, M. A., & Leal, A. (2018). Identification of Overpricing in the Purchase of Medication by the Federal Government of Brazil, Using Text Mining and Clustering Based on Ontology. *ICCBDC'18: Proceedings of the 2018 2nd International Conference on Cloud and Big Data Computing*, 66-70. doi:10.1145/3264560.3264569
- Cruz, R., Colaço Júnior, M., & Gois, V. (2022). Quão experimentais e estratégicas são as aplicações de Business Intelligence (BI) e Data Mining? ; HOW EXPERIMENTAL AND STRATEGIC ARE BUSINESS INTELLIGENCE (BI) AND DATA MINING APPLICATIONS? *Revista Ibero-Americana de Estratégia*, 21, e17689.
- Gomes, W., & Colaço Júnior, M. (2022). Applications of Artificial Intelligence for Auditing and Classification of Incongruent Descriptions in Public Procurement. *Proceedings of the Brazilian Symposium on Information Systems* (pp. 1-8).
- Ribeiro, L., Brandão, W., Marques, Í., Andrade, P., Júnior, R., Oliveira, F., & Kelles, R. (2018). Reconhecimento de entidades nomeadas em itens de produto da nota fiscal eletrônica., 36, pp. 116-126.
- Santos, B., Colaço Júnior, M., Meneses Santos, R., & Nascimento, A. (2015). Análise Comparativa de Algoritmos de Mineração de Texto Aplicados a Históricos de Contas Públicas. *Proceedings of the Brazilian Symposium on Information Systems*.

Spichakova, M., & Haav, H.-m. (2020). Using Machine Learning for Automated Assessment of Misclassification of Goods for Fraud Detection., (pp. 144-158). doi:10.1007/978-3-030-57672-1_12

TCU. 2016). *Auditoria em órtese, Prótese e Materiais Especiais (OPME)*. Fonte: Portal TCU: <https://portal.tcu.gov.br/biblioteca-digital/auditoria-em-ortese-protese-e-materiais-especiais-opme.htm>