

Classificação automática de produtos comercializados por órgãos públicos do Rio Grande do Norte através de comitê de classificadores

Automatic classification of products marketed by public agency of Rio Grande do Norte through a committee of classifiers

Clasificación automática de productos comercializados por órganos públicos de Rio Grande do Norte mediante un comité de clasificadores

Recebido: 16/06/2022 | Revisado: 29/06/2022 | Aceito: 30/06/2022 | Publicado: 10/07/2022

Elvis Rafael Ferreira Dias

ORCID: <https://orcid.org/0000-0003-3709-6221>
Universidade Federal do Rio Grande do Norte, Brasil
E-mail: elvisnaopresley@gmail.com

João Carlos Xavier Júnior

ORCID: <https://orcid.org/0000-0003-1517-2211>
Universidade Federal do Rio Grande do Norte, Brasil
E-mail: jxavier@imd.ufrn.br

Resumo

O uso de técnicas de mineração de texto tem aumentado consideravelmente nos últimos anos devido à grande quantidade de informações em texto sendo produzidas e armazenadas por sistemas eletrônicos e a necessidade de tornar esses dados informações para as organizações. Dentro desse contexto, o Tribunal de Contas do Estado do Rio Grande do Norte (TCE-RN) recebe diariamente uma grande quantidade de Notas Fiscais Eletrônicas (NF-e) contendo informações de compras de produtos que precisam ser analisadas para uso interno e externo ao órgão. Ainda, esses documentos possibilitam preenchimento livre, ou muitas vezes errôneo, de alguns dados por parte dos vendedores que emitem as Notas Fiscais. Dessa forma, as NF-e geradas não seguem um padrão, dificultando a realização de análises de forma prática e rápida por meio de ferramentas comuns para obtenção e filtragem de dados. Sendo assim, surge a necessidade de um processamento automatizado de forma a padronizar os dados, disponibilizá-los de forma rápida e possibilitar sua utilização como informações para fins de auditorias. Portanto, este trabalho apresenta uma solução baseada em mineração de texto e técnicas de aprendizado de máquina para o problema de identificação de produtos comercializados no estado do Rio Grande do Norte a partir do campo descrição das NF-e com o propósito de possibilitar a classificação de produtos únicos.

Palavras-chave: Mineração de texto; Notas Fiscais Eletrônicas; Processamento de dados; Aprendizado de máquina.

Abstract

The use of text mining techniques has increased considerably in recent years due to the large amount of text information being produced and stored by electronic systems and the need to make this data information for organizations. In this context, the Court of Auditors of the State of Rio Grande do Norte (Tribunal de Contas do Rio Grande do Norte, TCE-RN) receives daily a large amount of electronic invoices containing data of product's purchases that need to be analyzed for the society's benefit. Still, these documents allow free filling, often erroneous, of some data by the sellers who issue the invoices. This way, the documents do not come to follow a pattern and make it possible to carry out analysis in a practical and efficient way through common tools for obtaining and filtering data. Therefore, there is a need for automated processing in order to standardize the data, make them available quickly and enable their use as information for audit purposes. So, this work presents a solution based on text mining and machine learning techniques for the problem of identifying commercialized products in the state of Rio Grande do Norte from the description field of Electronic Invoices as a way to enable the classification of these products into unique products

Keywords: Text mining; Electronic invoices; Data processing; Machine learning.

Resumen

El uso de técnicas de minería de textos se ha incrementado considerablemente en los últimos años debido a la gran cantidad de información textual que se produce y almacena por los sistemas electrónicos y la necesidad de convertir estos datos en información para las organizaciones. En este contexto, el Tribunal de Cuentas del Estado de Rio Grande do Norte (Tribunal de Contas do Rio Grande do Norte, TCE-RN) recibe diariamente una gran cantidad de Facturas Electrónicas que contienen información sobre compras de productos que necesitan ser analizados para uso interno y

externo la institución. Aún así, estos documentos permiten la cumplimentación libre, o muchas veces errónea, de algunos datos por parte de los vendedores que emiten las facturas. De esta forma, los documentos generados no siguen un patrón, lo que dificulta la realización de análisis de forma práctica y rápida mediante herramientas habituales de obtención y filtrado de datos. Por lo tanto, existe la necesidad de un procesamiento automatizado para estandarizar los datos, hacerlos disponibles rápidamente y permitir su uso como información para fines de auditoría. Así, este trabajo presenta una solución basada en técnicas de minería de texto y aprendizaje automático para el problema de identificación de productos vendidos en el estado de Rio Grande do Norte usando el campo de descripción para permitir la clasificación en productos únicos.

Palabras clave: Minería de textos; Facturas electrónicas; Procesamiento de datos; Aprendizaje automático.

1. Introdução

A Nota Fiscal eletrônica (NF-e) é um documento de existência apenas digital, emitido e armazenado eletronicamente, com o intuito de documentar para fins fiscais uma operação de circulação de mercadorias ou prestação de serviços ocorrida entre as partes. Sua validade jurídica é garantida pela assinatura digital do remetente (garantia de autoria e de integridade) e a Autorização de uso fornecida pelo Fisco, antes da ocorrência do fato gerador (Ministério da Fazenda [MF], 2020a).

Como afirma a Secretaria de Tributação do Rio Grande do Norte - SET-RN (2020), este documento, instituído pelo Ajuste SI-NIEF 07/05 de 30 de março de 2005, teve como objetivo a implantação de um modelo nacional de documento fiscal eletrônico visando a substituição dos modelos 1 e 1A em papel. Dentre os benefícios, veio para simplificar as obrigações acessórias dos contribuintes e permitir o acompanhamento em tempo quase imediato das operações comerciais pelo Fisco.

Partindo da obrigatoriedade da emissão das Notas em atividades econômicas envolvendo órgãos públicos (MF, 2020d), a SET-RN, responsável pelo recebimento e armazenamento das notas estaduais, firmou um acordo com o Tribunal de contas do Estado do Rio Grande do Norte (TCE-RN) em 2019 para repassar NF-e relativas aos órgãos jurisdicionados ao TCE-RN, permitindo análises de interesse interno.

Dentre as diversas informações presentes na NF-e (MF, 2020b), a motivação inicial do TCE-RN parte para a análise de preços realizados no dia a dia, em especial, o cálculo do preço de referência para os diversos produtos comercializados no estado. O preço de referência possibilita aos diversos órgãos a pesquisa de preços, tarefa que, como afirma o decreto federal nº 7.892/2013 inciso IV do artigo 5º, cabe a cada órgão realizar sua pesquisa de mercado periódica como procedimento prévio e indispensável para uma pretensão contratual (Brasil, 2013b). Ainda, como estabelecido no Acórdão 169/2013 do Tribunal de Contas da União, foi esclarecido que a ausência de pesquisa que represente adequadamente os preços de mercado, além de constituir afronta a jurisprudência do Tribunal de Contas, pode ser favorável a contratação de serviços ou aquisição de bens por preços superiores aos praticados pelo mercado, ferindo assim o princípio da economicidade, conforme entendimento constante do TCU 1785/2013 – Plenário (Brasil, 2013b; Silva, 2014).

Para permitir pesquisa de preços para produtos únicos, estes precisam poderem ser agrupados para que assim seus preços possam ser comparados. Através do código de barras, tecnologia responsável por categorização de produtos, idealmente todos os produtos teriam uma forma de serem unicamente identificados, entretanto, essa tarefa não é possível de forma devido a prática recorrente da emissão destes documentos preenchidos de forma negligente ou errônea (GS1, 2019). Dessa forma, a tarefa de identificação de produtos únicos recai sobre a interpretação do campo texto que descreve os produtos.

Tendo em vista também a grande quantidade de compras realizadas pelo estado periodicamente e o uso dessas informações serem voltados para painéis de *Business Intelligence*, a análise feita pelo órgão deve ser idealmente automatizada e próxima de tempo real. Logo, o processo de automatizar a identificação de produtos requer uma solução inteligente, que consiga aprender diferentes padrões baseado nas descrições dos próprios produtos.

Assim sendo, em virtude da limitação da interpretabilidade de descrições textuais automaticamente, este trabalho propõe uma metodologia para manuseio e controle destes textos como forma de agrupar produtos comercializados. Métodos de

processamento textual precisarão ser utilizados para extrair informações cruciais dos produtos essenciais para a etapa de classificação realizada por Comitês de Classificadores.

O restante deste artigo está dividido em 6 seções. A Seção II descreve alguns conceitos teóricos importantes enquanto que a seção III apresenta alguns trabalhos relacionados ao tema deste artigo. Na Seção IV, a metodologia utilizada no desenvolvimento deste trabalho é apresentada, enquanto que uma análise dos resultados é apresentada na Seção V. Finalmente, a Seção VI apresenta as considerações finais e trabalhos futuros.

2. Referencial Teórico

Essa seção aborda os conceitos mais relevantes para o desenvolvimento deste trabalho, partindo dos dados brutos, passando pelo processamento textual, para finalmente chegar na classificação dos produtos.

2.1 Definição dos dados

Dentre os campos que caracterizam os produtos em uma Nota Fiscal eletrônica, ilustrados na Tabela 1, há três campos relativos à identificação: “NCM”, “cEAN” e “xProd”. O primeiro é atrelado à tributação, agrupando produtos semelhantes a um mesmo código (MF, 2020c), enquanto que o segundo (código de barras) pode não existir para determinados produtos (GS1, 2020), sendo também de livre preenchimento e nulo em mais de 70% das notas analisadas. Por último, o campo “xProd”, que mesmo impreciso devido à livre digitação, descreve os produtos.

Tabela 1: Campos contendo características dos produtos comercializados em uma Nota Fiscal Eletrônica.

| Atributo | Descrição |
|----------|---|
| NCM | Nome Comum do Mercosul, código numérico único de oito dígitos que agrupa mercadorias comercializadas em operações de comércio exterior no Mercosul. |
| cEAN | GTIN (do inglês, <i>Global Trade Item Number</i>) do produto, antigo código EAN ou código de barras. |
| xProd | Descrição textual do produto ou serviço. |
| uCom | Unidade Comercial, informa a unidade de medida do produto, por exemplo, litros e quilos. |
| qCom | Quantidade Comercial, especifica quantos produtos estão sendo vendidos para cada item comercializado na nota. |
| vUnCom | Valor Unitário de Comercialização, indica em reais o valor unitário de cada produto descrito na nota. |
| vProd | Valor total bruto dos produtos ou serviços. |

Fonte: Ministério da Fazenda (2020b).

2.2 Processamento de Linguagem Natural (PLN)

É uma área de estudo responsável por manipular a linguagem humana, do inglês *Natural Language Understanding* (NLU) e dentre suas especialidades, possui ferramentas de PLN capazes de extrair informação de texto e possibilitar aprendizado destas. Técnicas de minimização de problemas com ruído, por exemplo, buscam remover termos textuais inúteis para o aprendizado, enquanto que a transformação de palavras em dados numéricos, usáveis em algoritmos, é feita através de codificadores de palavras.

Os codificadores de palavras são responsáveis pela transformação de termos linguísticos em números, possibilitando o uso por algoritmos. Dentre os tipos de representações para texto, destacam-se *Count Vectorizer* e *Word Embeddings*; a primeira é orientada pela ocorrência de palavras, sendo mais usada para classificação, enquanto que a segunda é orientada pelas

características textuais, capturando sentido semântico e mais usada na análise de sentimento e tradução.

O *Count Vectorizer*, além de contar ocorrências de palavras, pode também analisar sentenças através da contagem de sequências de n palavras (n -grama), ao invés de palavras isoladas, levando a um aprendizado de termos em conjunto. O *Term Frequency-Inverse Document Frequency* (TF-IDF), definido pela Fórmula 1, garante que termos como artigos e preposições, por exemplo, que não carregam informação de contexto e aparecem frequentemente em textos, se tornem irrelevantes comparados aos termos mais frequentes que carregam mais sentido específico às amostras. $Tf(t,d)$ sendo a frequência do termo t no documento d ; n , o número total de amostras; e $df(t)$, o número de documentos que contém t (Buitinck et al. 2013).

$$tf\ idf(t,d) = tf(t,d).idf(t) \quad (1)$$

Onde,

$$idf(t) = \log \frac{1+n}{1+df(t)} + 1 \quad (2)$$

2.3 Aprendizagem de máquina

Segundo Mitchell (1997), “A capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência”, define a classe de programas de computadores que conseguem aprender e compõem a área de Aprendizagem de Máquina. Esses algoritmos (também chamados de modelos) conseguem aprender a induzir uma função ou hipótese capaz de resolver um problema a partir de dados que representam instâncias do problemas a ser resolvido (Faceli et al., 2011).

Em AM o objetivo é fazer com que os algoritmos generalizem seu aprendizado, por meio de um conjunto de dados de treinamento, em dados de testes nunca antes visto — característica de que houve o aprendizado (Goodfellow et al., 2016). O problema de classificação pode ter soluções baseadas em diferentes formas de aprendizado, sendo agrupados em: métodos probabilísticos, baseados em distância, busca e em otimização (Faceli et al., 2011).

O aprendizado desses algoritmos pode ser medido através de diferentes métricas de desempenho, como acurácia, *f1-score* e curva ROC, por exemplo. Além da acurácia, mais intuitiva, há formas mais robustas de avaliar o desempenho de um modelo, levando em conta precisão e abrangência (do inglês, *recall*). A métrica denominada “f1-score” (ou *f-measure*), Equação 3, mede através dos positivos verdadeiros (TP), positivos falsos (FP) e negativos falsos (FN) a relação entre a porcentagem de amostras corretamente classificadas como sua classe, pela porcentagem da quantidade real de amostras que foram corretamente classificadas (Buitinck et al. 2013).

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (3)$$

Sabendo-se que não existe um algoritmo que seja melhor para todos os problemas de classificação, uma estratégia possível é agrupar conjuntos de preditores cujas decisões individuais são combinadas ou agregadas de alguma forma para prever novos exemplos. A ideia principal por trás desses métodos, portanto, é envolta no fato que diferentes técnicas de aprendizado de máquina exploram diferentes espaços de busca e funções de avaliação de hipóteses, possibilitando trabalharem juntos para atingir possíveis melhores resultados (Faceli et al., 2011).

Embora o processo de votação em comitês possa ocasionar perda de desempenho devido a presença de modelos com aprendizado insatisfatório dentre o grupo, há soluções baseadas em comitês com camadas, por exemplo, Pilha ou Cascata, em que o desempenho geral é garantido pelos melhores resultados. O método de Generalização em Pilha, por exemplo, possui uma arquitetura em camadas onde há níveis de predições, com diferentes classificadores base. Os classificadores de nível zero recebem como entrada os dados originais e cada classificador produz uma predição que servirá de entrada para as camadas imediatamente sucessivas. Um único classificador no nível mais alto produz a predição final, responsável por aprender a partir da combinação das predições imediatamente abaixo (Faceli et al., 2011).

Comitês de classificadores têm diversas arquiteturas possíveis com desempenho comprovadamente melhor se

comparada com classificadores base. *Bagging*, por exemplo, que treina modelos com modificações do conjunto de treinamento, é uma das formas mais eficientes para melhorar a precisão de classificadores tais como árvores de decisões e redes neurais. Contrariamente, Faceli et al., (2011) afirmam que a generalização em pilha ainda apresenta mistério sobre quais condições são melhores sucedidas.

O *Stacking*, também conhecido como generalização em pilha, é um tipo de comitê que objetiva minimizar o erro de predição a partir de classificadores nas camadas mais altas que aprendem o tipo de erro cometido dos classificadores imediatamente abaixo. A regra dos classificadores dos níveis mais altos, portanto, é aprender como classificadores anteriores cometem erros, em qual classe eles concordam ou discordam, e usar os seu conhecimento para fazer predições (Faceli et al., 2011).

Ainda, como Faceli et al., (2011) explicitam, comitês podem ser aplicados com métodos de votação e métodos de seriação. Dentre diversas variações existentes, a votação uniforme, por exemplo, leva em conta a opinião de todos os classificadores de base igualmente para a classificação final. De forma contrária, os métodos de seriação apresentam uma melhoria sob o método uniforme, pois cada classificador gera uma probabilidade de o exemplo pertencer a cada classe, ao invés de um rótulo apenas. Métodos de seriação se fazem úteis para predição dentre amostras de classes desconhecidas, pois o modelo não tentará atribuir uma classe a esta, ao invés, retornará uma baixa probabilidade para todas as classes conhecidas, levando a um possível descarte do exemplo.

Dessa forma, este trabalho conduzirá uma análise empírica sobre o desempenho de alguns classificadores base, como forma de escolher os componentes mais adequados para as diferentes arquiteturas de comitês que serão analisadas através do método de seriação.

3. Trabalhos Relacionados

O desafio de classificação de sentenças usando as descrições em Notas Fiscais eletrônicas tem sido investigado por órgãos públicos pelo Brasil, com apontam alguns trabalhos publicados na internet. Os relatos aqui possuem quantidade de informação limitada a disponível via pesquisa online. A pouca variedade e disponibilidade técnica dos projetos descritos nesta seção, impossibilitou uma análise profunda dos resultados e desafios ao medir se serviriam para o objetivo do TCE-RN.

O banco de preços público divulgado pelo TCE de Minas Gerais, por exemplo, usa notas de compras realizadas pelos órgãos públicos do estado para gerar análises de preços praticados. Nesta solução, ao invés de classificar as descrições em classes de produtos, agrupam-se todas as descrições do banco de dados baseando-se na similaridade destas com palavras-chave informadas pelo usuário, descrevendo o produto. O sistema web, aberto para uso público, exige que o usuário decida dentre todas as descrições retornadas, quais devem entrar ou não na sua análise de preços, realizando um processo de classificação em tempo de uso. Esta abordagem tem a vantagem de funcionar com taxa de erro dependendo somente do usuário e não haver limitação no número de produtos usados, porém a experiência do usuário do sistema pode ser exaustiva em caso de múltiplos produtos contendo a mesma palavra-chave (Tribunal de Contas do Estado de Minas Gerais [TCE-MG], 2020).

O Tribunal de Contas do Estado da Paraíba [TCE-PB] (2020a) também lançou um sistema web para análise de NF-es, intitulado "Painéis de Preços", relativo às compras realizadas pelas administrações municipais e estaduais de produtos alimentícios e combustíveis (cerca de vinte produtos no total). Baseado em algoritmos de AM, a equipe afirma que o sistema, ainda em fase inicial, minera milhões de descrições de produtos por meio de termos-chave de classificação e validação relacionados aos diferentes tipos de despesas com combustíveis e alimentos, porém, com uso do sistema foi visto que produtos com diferentes formas de comercialização não são diferenciados entre si, por exemplo, "Leite Integral em pó" possui um único preço (TCE-PB, 2020a).

Um outro sistema, feito pelo mesmo time atribuído ao do TCE-PB, chamado "Preço da Hora", usa outro tipo de Nota,

porém contendo informações semelhantes, Notas Fiscais de Consumidor Eletrônicas (NFC-e). Esta solução é mais voltada para resolver o problema de processamento de grandes quantidades de dados, buscando disponibilizar visualização de Notas online rapidamente e assim facilitar a compra de produtos locais pela população; não há ferramentas para análise de notas fiscais com objetivo de geração de *insights*. Essa abordagem, semelhante à do TCE-MG, possui o mesmo problema de listar todas as ocorrências das palavras-chave informadas, exigindo do usuário classificar os produtos retornados em tempo de uso (TCE-PB, 2020a).

No Tribunal de Contas do estado do Rio Grande do Sul (TCE-RS), de acordo com (Gandini, 2020), também houve a tentativa de criar um banco de preços para identificação de sobrepreço em processos licitatórios no Estado. Como relatado em seu repositório público, a abordagem para a classificação de descrições utilizou aprendizagem de máquina descritiva, buscando obter descrições de produtos agrupadas automaticamente pelo algoritmo. Mesmo afirmando ter chegado em bons resultados e que está sendo usado internamente, entretanto não são especificados detalhes do uso. A solução, muito interessante, foi testada e se mostrou inviável para uso do TCE-RN devido a alta granularidade das classes requeridas, sua instabilidade e por ter um processo de manutenção pesado (Gandini, 2020).

Por fim, esse problema foi também abordado em uma dissertação de mestrado como estudo de caso de uma plataforma distribuída de mineração de dados para *Big Data*. O autor aplica técnicas de aprendizagem preditivo, através de um conjunto de dados para classificar bebidas em água, cerveja, destilado, refrigerante, suco ou diversos. Foram utilizados nove atributos para treinar modelos de AM, dentre eles, “NCM”, “qCOM”, “vUnCom” e campos extraídos de “xProd” como Marca, Tipo, Embalagem e Volume. A engenharia de dados aplicada, possibilitou ter para cada amostra todas as características necessárias para caracterizar um produto único, como: “Marca: coca, Tipo: cola, Embalagem: lata, Volume: 350, Quantidade: 1”. Entretanto, o trabalho limitou-se a um conjunto de dados de 3 mil amostras e a classificar produtos não como “Coca-cola 350ml”, mas sim como um grande grupo refrigerante (dos Santos, 2018).

A solução proposta aqui tem o diferencial de buscar atingir um alto grau de precisão classificando automaticamente uma grande variedade e quantidade de produtos, que podem ter seus preços comparados. Utilizando apenas o campo de descrição, investigar viabilidade de ferramentas de AM para possibilitar uso integrado ao fluxo de dados automatizado, que parte do recebimento das notas via SET-RN, até chegar nos painéis interativos para inteligência de negócio.

4. Metodologia Experimental

Nesta seção descrevemos as etapas que constituem a metodologia utilizada para validar nossa hipótese de classificar descrições textuais a partir de processamento textual aplicado. Partindo para uma pesquisa experimental, foram escolhidas descrições de alguns produtos como variáveis para analisar o problema, construir as hipóteses e através de suas manipulações e diferentes modelos de AM, avaliar o resultado com a hipótese de que modelos são capazes de classificar os textos corretamente (Koche, J. C., 2011).

4.1 Pré-classificação de produtos

As notas fiscais disponíveis para este trabalho se constituem de dados que foram manipulados previamente do formato XML para o formato tabular, e que são referentes ao período entre Set/2019 e Nov/2020. A partir desses dados, possibilitou-se a criação de um subconjunto de produtos com classes (rótulos) analisadas singularmente a partir de consultas SQL feitas no banco de dados.

A escolha dos produtos que compõem esse subconjunto levou em consideração a grande quantidade de instâncias disponíveis (produtos mais comercializados) a partir de termos chaves compartilhados pelas descrições. Dessa forma, a atribuição de classes se deu a partir de uma análise das descrições originais e únicas, reduzindo drasticamente a quantidade de

registros para classificar.

Dentre todas características que um produto pode ter como fator de alteração no preço, foram utilizadas aquelas que independem de marca, cor, sabor ou outras perfumarias para um mesmo produto. O padrão buscado portanto consiste em produto base + possível característica + tamanho/embalagem, por exemplo, “Açúcar Refinado 1kg”, resultando em mais de vinte mil amostras rotuladas e mais de 400 produtos únicos.

A diferença de preços, muitas vezes, pode ser insignificante para produtos como “Açúcar Refinado 1kg” e “Açúcar Cristal 1kg”, porém, foram rotulados alguns produtos dessa forma mais granular para permitir uma análise mais ampla. Diferentemente, para outros produtos como “Arroz 1kg” e “Feijão 1kg” os tipos específicos foram mantidos como sendo pertencente a uma única classe.

Nesse processo é importante uma capacidade analítica acurada para evitar que o classificador conheça apenas algumas versões mais populares quando alguns produtos podem ser descritos de diferentes formas. O produto “Desodorizador de ar” pode ser chamado também de “Bom Ar” e “Odorizador de Ar”. Semelhantemente, produtos com álcool etílico podem ter sua porcentagem de álcool expressa com base em °INPM ou °GL para especificar o mesmo produto.

Por fim, descrições incompletas, que compõem grande parcela dos dados das NF- e como, por exemplo, “Álcool em Gel”, foram ignoradas. Mesmo que em casos fatores de alteração de preço como tamanho e/ou porcentagem de álcool, possibilitem inferências, estas amostras foram deixado de fora ciente da limitação causada no conhecimento dos comitês.

4.2 Pré-processamento das descrições de produtos

Pré-processar as descrições dos produtos envolve partir de uma sentença em português, em formato livre, deixando-o em um padrão que possa ser compreendido pelos modelos de AM. Neste processo, técnicas de limpeza textual padrões e específicas para este problema foram aplicadas em conjunto das técnicas de PLN, lematização, *stemming* e codificação.

4.2.1 Limpeza Textual

Sabendo que as descrições dos produtos possuem, devido sua natureza de escrita livre, infinitas possibilidades de serem representadas, a limpeza aplicada almejou reduzir ruído e ambiguidades em termos problemáticos gerais e tornar as descrições mais genéricas e informativas, mantendo o equilíbrio e integridade das classes. Abaixo são listadas as etapas de limpeza que foram aplicadas às descrições originais:

1. Remoção de palavras inúteis, zeros a esquerda, excesso de espaços em branco, acentos e símbolos - com exceção de números decimais;
2. Remoção de padrões específicos, como lote, endereço, validade, etc;
3. Aplicação de minúsculo em todas as letras;
4. Separação de número junto a letra;
5. Padronização de unidades de medida e outros termos de medição;
6. Junção de número com unidades de medida;
7. Junção de palavras compostas que descrevem produtos, como “Leite em pó” para “Leitepo”;
8. Remoção de dimensões de produtos e quantidades comprada;
9. Remoção de números a partir de 5 dígitos, números soltos sem unidades, caso tenha número com unidade na mesma descrição e números no início das descrições;
10. Remoção de letras soltas se não forem unidades;
11. Remoção de descrições de tamanho um ou vazias;
12. Limitação das descrições em 8 palavras.

Devido ao fluxo contínuo de dados, a etapa de limpeza dos mesmos precisa se manter em constante atualização e revisão, mesmo que não haja alteração nas classes dos produtos. Novas maneiras de descrevê-los podem surgir, levando a confusão entre classes, da mesma forma que as descrições também podem mudar devido surgimento de novas marcas e/ou

embalagens.

4.2.2 Codificação de Descrições

A codificação usada (*Count Vectorizer*) foi definida no intuito de contar ocorrências de palavras utilizando *n*-grama de intervalo (2,4), ou seja, o modelo interpreta sequências de duas até quatro palavras seguidas. Para este valor de intervalo foi buscado generalizar a quantidade de palavras capazes de identificar produtos únicos de forma geral, não levando em conta produtos com descrição de uma palavra. A escolha do *n*-grama ainda influencia diretamente nas limpezas textuais aplicadas às descrições, devendo-se buscar reduzir sequências repetitivas a partir de duas palavras que não caracterizam um produto único, exemplo de “1 litro”.

4.3. Escolha dos Produtos

Considerando a inviabilidade do uso de todos os produtos categorizados no banco de dados, decidiu-se por fazer uma seleção de uma quantidade limitada de produtos, visando testar a metodologia aqui apresentada. Dessa forma, orientou-se por selecionar produtos com quantidade e variabilidade suficientes para aplicar nos modelos de AM.

Também, devido aos parâmetros utilizados no *Count Vectorizer* (*n*-grama [2,4]), alguns produtos como frutas e legumes foram excluídos, por apresentarem padrão de descrições compostas por apenas “Produto base” + “Característica comum”. Por não apresentarem nos produtos em si características únicas, acabam por ter características compartilhadas por outras frutas e verduras ou descrições singulares.

Por fim, dentre os mais de quatrocentos produtos inicialmente rotulados, selecionou-se aqueles que mais se adequavam aos critérios já abordados. A Tabela 2 mostra os 74 produtos distintos que farão parte dos experimentos.

Tabela 2: Classes de produtos rotulados e selecionados para experimentos.

| | | | | | | | |
|----|--------------------------|----|-----------------------|----|-----------------------------|----|------------------------|
| 1 | Achocolatado em pó 400g | 20 | Chá 10g | 39 | Gás glp 13kg | 58 | Resma Papel A4 500fls |
| 2 | Adoçante 100ml | 21 | Cimento 50kg | 40 | Leite Condensado 395g | 59 | Sabão em Pó 500g |
| 3 | Aromatizador 360ml | 22 | Colorau 100g | 41 | Leite em pó 200g | 60 | Sabão em Barra 200g |
| 4 | Aromatizador 400ml | 23 | Copo Desc. 150ml | 42 | Leite em pó Integral 200g | 61 | Saco Plástico 100lt |
| 5 | Arroz 1kg | 24 | Copo Desc. 180ml | 43 | Leite Integral 1lt | 62 | Saco Plástico 150lt |
| 6 | Açúcar Cristal 1kg | 25 | Copo Desc. 50ml | 44 | Limpa Vidros 500ml | 63 | Saco Plástico 200lt |
| 7 | Açúcar Refinado 1kg | 26 | Costela Bovina | 45 | Lustra Móveis 200ml | 64 | Saco Plástico 30lt |
| 8 | Bebida Láctea 1lt | 27 | Coxa/Scoxa de Frango | 46 | Luva Cirúrgica Estéril | 65 | Saco Plástico 40lt |
| 9 | Bisc. Cream Cracker 400g | 28 | Creme de Leite 200g | 47 | Lápis Grafite | 66 | Saco Plástico 50lt |
| 10 | Bisc. Maizena 400g | 29 | Desinfetante 1lt | 48 | Macarrão Espaguete 500g | 67 | Saco Plástico 60lt |
| 11 | Bisc. Maria 400g | 30 | Desinfetante 2lt | 49 | Margarina 500g | 68 | Sal Refinado 1kg |
| 12 | Café em pó 250g | 31 | Desinfetante 500ml | 50 | Pano de Prato | 69 | Tijolo 8 furos |
| 13 | Cal Hidratada 5kg | 32 | Desinfetante 5lt | 51 | Papel Higie`nico c/4 rolos | 70 | Vinagre 500ml |
| 14 | Caneta Esferográfica | 33 | Detergente 500ml | 52 | Papel Higie`nico c/64 rolos | 71 | Água Mineral 20lt |
| 15 | Caneta Marca Texto | 34 | Esponja Dupla Face | 53 | Papel Toalha c/2 rolos | 72 | Água Sanitária 1lt |
| 16 | Carne Bovina | 35 | Farinha de milho 500g | 54 | Peito de Frango | 73 | Álcool etílico 70% 1lt |
| 17 | Carne de Charque | 36 | Farinha de Trigo 1kg | 55 | Polpa de Fruta 1kg | 74 | Óleo de Soja 900ml |
| 18 | Carne de Sol | 37 | Feijão 1kg | 56 | Polpa de Fruta 400g | - | - |
| 19 | Carne Moída | 38 | Frango Inteiro | 57 | Refrigerante 2lt | - | - |

Fonte: Autores.

4.4 Separação em conjuntos de treinamento e teste

A separação dos dados em conjuntos de treinamento e teste seguiu a abordagem de não replicação de dados (sem reamostragem), de forma que não houvesse intersecção entre os conjuntos. Portanto, foi proposta uma separação estratificada

aleatória de 75% de amostras para treinamento, 20% para testes e 5% para validação.

Por fim, ao conjunto de validação são agregadas também amostras ruidosas, restantes da rotulação manual. Devido à necessidade real da análise dos produtos envolver os comitês distinguirem produtos conhecidos de desconhecidos, foram aplicadas 5% do total de amostras não categorizadas, aleatórias, para simular o comportamento dos classificadores nessa distinção.

4.5 Modelos de classificação

A metodologia para compreender quais modelos são apropriados e funcionam melhor dentre tantos disponíveis, baseou-se na validação cruzada por *k-fold* com re-amostragem (do inglês, *k-fold cross validation*). Para os treinamentos foram aplicadas as descrições dos produtos em sua distribuição original (com repetições), *oversampling* e validação cruzada por *k-fold* estratificado.

A métrica *f1-score*, além de usada no processo de escolha dos modelos, foi usada também para avaliar modelos treinados com dados de teste e validação. Diferentemente, para a compreensão dos testes com ruídos, devido às amostras ruidosas não serem categorizadas, mas ainda sim puderem conter produtos conhecidos, foi feita uma contagem do percentual de erros e acertos — produtos com descrição incompleta preditos como um produto conhecido são considerados acertos, visto que por não apresentarem todas as características não pode ser refutado.

Tendo em vista que a separação em diferentes conjuntos de dados tem fator aleatório, pode ser que haja perda de informação para o treinamento ou criação de cenário favorável ao remover deste amostras contidas em teste e validação. Portanto, todas as etapas de treinamento, teste e validação foram realizadas cinco vezes, sobre diferentes dados por execução, e agrupados pela média dos resultados, minimizando qualquer tendência existente.

A análise dos modelos de classificação ocorreu em duas etapas, primeiro os classificadores base foram avaliados por validação cruzada com re-amostragem e depois os comitês de classificadores. Na primeira etapa foram explorados duas configurações diferentes para o algoritmo *k-NN* com $k=3$ e medidas de distância de Manhattan e euclidiana; Árvore de Decisão, versão otimizada do algoritmo CART com critério de separação “gini” e “entropia”; SVM com kernel “rbf” e gamma “scale” e “auto”; e, por fim, Naive Bayes com suavização 1 e 0,1. Para as redes neurais MLP (“adam”) foram exploradas diversas arquiteturas de neurônios que melhorassem o valor da função perda, resultando em duas camadas ocultas com (30,20) neurônios (Pedregosa et al., 2011).

Para os comitês, foram analisados os seguintes métodos: *Random Forest* composto por 100 árvores; *Bagging out-of-fold* com 10 MLP; *Stacking* utilizando os quatro melhores classificadores base (AD, NB, SVM e MLP), ficando a função de fusão das predições a cargo do algoritmo Regressão Logística — importante ressaltar que somente as melhores configurações foram consideradas na formação do *Stacking*. A escolha dos comitês orientou-se por modelos amplamente conhecidos pela literatura, com exceção dos comitês pelos métodos de *Boosting*, devido a testes iniciais com implementação do “sklearn” não mostrarem resultados satisfatórios.

Partindo do princípio que o classificador em ambiente real lidará com todas as descrições recebidas, todos os modelos implementados baseiam-se em predição probabilística, semelhante aos comitês por seriação explicados. Assim, a determinação se uma amostra pertencente a uma classe quando apresentada a um modelo é feita se o modelo obteve uma probabilidade acima de 50% para alguma das classes conhecidas, caso contrário são consideradas como não pertencentes aos produtos conhecidos — vale ressaltar que as probabilidades aqui mencionadas não são probabilidades no sentido real (em AM, probabilidades calibradas) e sim fatores de confiança do modelo.

5. Resultados Experimentais

Os resultados experimentais deste trabalho serão mostrados de acordo com as fases de classificação seguidas, iniciando com a análise dos classificadores bases, visando selecionar os modelos mais apropriados e terminando nos diferentes tipos de comitês aqui apresentados.

5.1 Análise dos classificadores base

Como forma de desempenhar uma análise robusta sobre o desempenho de diversas técnicas supervisionadas de AM para classificação de dados, utilizou-se a validação cruzada com k folds ($k = 10$). Além disso, foram feitas 5 execuções para cada técnica, permitindo, dessa forma, calcular a média da métrica $f1$ -score, utilizada como medida para avaliar o desempenho das mesmas.

Os resultados, na Tabela 3, permitem ter conhecimento de quais técnicas lidaram melhor com os dados fornecidos. Note que dentre elas, as Redes Neurais (MLP) obtiveram o melhor desempenho entre todas, seguida de perto pelo Naive Bayes (suavização =0,1) e AD (com “gini”). A técnica SVM (“auto”) teve um desempenho pouco expressivo, ficando apenas a frente do k -NN (distância euclidiana) que foi a técnica que apresentou o pior desempenho entre as cinco (5) técnicas.

Tabela 3: Média e Desvio Padrão dos valores de $f1$ -score para 5 execuções dos modelos base.

| Classificador | F1-score | Desvio Padrão |
|-------------------|-------------|---------------|
| Árvore de Decisão | 0,82 | 0,015 |
| k-NN | 0,65 | 0,019 |
| MLP | 0,89 | 0,011 |
| Naive Bayes | 0,88 | 0,012 |
| SVM | 0,73 | 0,017 |

Fonte: Autores.

5.2 Análise dos comitês de classificadores

Baseando-se nos resultados mostrados na Tabela 3, escolheu-se o algoritmo de Rede Neural (MLP) para compor a arquitetura do *Bagging*, principalmente por apresentar o melhor desempenho nas etapas de treinamento e teste, assim como por apresentar pouca sensibilidade a ruídos.

Já para compor a arquitetura do *Stacking*, escolheu-se os quatro classificadores base com melhor desempenho, ficando de fora o algoritmo k -NN que apresentou o pior desempenho entre todos os algoritmos analisados. Além dos componentes do comitê, escolheu-se o algoritmo de regressão logística com método de fusão.

A Tabela 4 mostra os resultados obtidos para os três (3) comitês analisados em nossos experimentos. Note que o *Bagging* apresentou o melhor desempenho ($f1$ -score médio) entre todos os comitês, ficando *Stacking* e *Random Forest* com desempenho inferior.

Tabela 4: Média e Desvio Padrão dos valores de $f1$ -score para 5 execuções dos comitês.

| Comitês | F1-score | Desvio Padrão |
|---------------|-------------|---------------|
| Bagging | 0,88 | 0,004 |
| Stacking | 0,87 | 0,004 |
| Random Forest | 0,87 | 0,004 |

Fonte: Autores.

Visando fazer uma análise mais robusta dos modelos reportados acima, decidi- se selecionar amostras que não haviam sido vistas pelos mesmos, ou seja, não estavam presentes nas etapas de treinamento e teste. Dessa forma, cerca de 293 (+/- 5%) amostras de validação foram utilizadas para avaliar os mesmos três comitês. Na Tabela 5 são ilustrados os resultados relativos ao desempenho dos modelos. Note o *Bagging* obteve novamente o melhor desempenho entre os três modelos, ficando o *Stacking* em segundo lugar, e o *Random Forest* em último.

Tabela 5: Média e Desvio Padrão dos valores de f1-score para 5 execuções com dados de validação.

| Comitês | F1-score | Desvio Padrão |
|---------------|-------------|---------------|
| Bagging | 0,89 | 0,003 |
| Stacking | 0,87 | 0,008 |
| Random Forest | 0,84 | 0,002 |

Fonte: Autores.

Além das amostras não vistas utilizadas na validação, decidiu-se por realizar uma análise mais profunda, utilizando pouco mais de 8 mil amostras não rotuladas, para compor junto das 293 amostras de validação o conjunto de validação com ruído. A Tabela 6 mostra a média do f1-score dentre as 5 execuções, com todos os comitês chegando no mesmo valor.

Tabela 6: Média e Desvio Padrão dos valores de f1-score para 5 execuções com dados de validação com ruído.

| Comitês | F1-score | Desvio Padrão |
|---------------|----------|---------------|
| Bagging | 0,97 | 0,000 |
| Stacking | 0,97 | 0,000 |
| Random Forest | 0,97 | 0,000 |

Fonte: Autores.

Sabendo que dentre as amostras não rotuladas pode haver amostras com características das classes conhecidas pelos modelos, é esperado que estes classifiquem mais de 293 amostras, pela predição probabilística. Portanto, dentre as amostras preditas como conhecidas misturadas com ruído, foram analisadas a quantidade correta, produto a produto, em cada execução de comitê, para analisar se os comitês ao predizerem mais amostras ocasionavam em redução na acurácia.

Os resultados da análise, Tabela 7, mostra o *Random Forest* com a melhor acurácia ao mesmo tempo que predisse menos amostras dentre os três comitês, enquanto que o *Bagging* e o *Stacking* apresentam resultados semelhante mesmo com número de amostras preditas pelo *Bagging* sendo superior. Dessa forma, é possível afirmar que os modelos que retornam mais amostras trazem consequentemente mais amostras errôneas.

Tabela 7: Média de acerto dentre 5 execuções com dados de validação com ruído.

| Comitês | Acurácia | Desvio Padrão |
|---------------|-------------|---------------|
| Bagging | 0,85 | 0,007 |
| Stacking | 0,86 | 0,009 |
| Random Forest | 0,88 | 0,002 |

Fonte: Autores.

Por fim, após percebido que os comitês encontraram dificuldades e facilidades semelhantes quanto a capacidade de reconhecer os produtos meio ruídos, foi feita uma última análise quanto ao aprendizado de cada produto. Como ilustrado pela Tabela 8, alguns produtos como, por exemplo, “Carne Bovina”, “Biscoito Cream Cracker” e “Peito de Frango” obtiveram apenas amostras corretas preditas pelos três comitês em todas as execuções; enquanto que, contrariamente, produtos como “Papel Higiênico 4 rolos”, “Papel Toalha 2 rolos” e “Copo Descartável 50ml” resultaram em uma baixa média de acerto.

Tabela 8: Média da acurácia, desvio padrão e quantidade de amostras preditas dentre as 5 execuções usando os 3 comitês sob dados de validação com ruído.

| Produtos | Acurácia | Desvio Padrão |
|---------------------------|----------|---------------|
| Bisc. Cream Cracker 400g | 1 | 0 |
| Carne Bovina | 1 | 0 |
| Peito de Frango | 1 | 0 |
| Copo Descartável 50ml | 0,59 | 0,036 |
| Papel Higiênico c/4 rolos | 0,51 | 0,161 |
| Papel Toalha c/2 rolos | 0,48 | 0 |

Fonte: Autores.

Ao comparar o padrão das amostras que compõem os produtos da Tabela 8, é perceptível que os produtos que funcionam perfeitamente têm descrições com um padrão mais constante, com poucas variações de característica e ausência de produtos desconhecidos pelo modelo com semelhança nas descrições. Por outro lado, produtos como papel higiênico e papel toalha apresentam por padrão descrições longas, com muitas características e compartilham padrões com outros produtos que o modelo não conhece.

6. Considerações Finais

Através dos testes realizados permitiu-se tirar conclusões sobre a metodologia implementada para a modelagem dos dados e sua aplicação em comitês de classificadores. Na limpeza dos dados esperava-se reduzir ruídos causadores de confusão para os modelos, na categorização usar produtos mais e menos granulares, enquanto que para os comitês obtenção de um aprendizado mais robusto com relação a classificadores em unidade.

A metodologia de limpeza se mostrou fundamental para melhorar a precisão dos modelos durante seu desenvolvimento, porém apresenta difícil rastreamento prévio. Apenas a partir de tentativa e erro implementando os dados nos modelos e investigando a forma como prediziam as amostras que foi possível rastrear padrões responsáveis por confusão e agir de acordo.

A escolha dos produtos se mostrou satisfatória, pois permitiu comprovar que os modelos conseguem apreender produtos muito semelhantes e que alguns produtos concentram a maior quantidade dos problemas de confusão. Diante da limitação de dados, produtos problemáticos que possuem características compartilhadas com produtos desconhecidos pelo modelo, em caso de não haver solução de limpeza, devem ser retirados temporariamente dos modelos.

Através dos experimentos realizados foi possível perceber a robustez do aprendizado dos comitês ao apresentarem pouca variação de precisão da validação sob diferentes dados de treinamento e mostrarem maior capacidade de aprendizado do que modelos singulares. Pelos resultados apresentados mesmo com a acurácia menor para o conjunto de validação com ruído, o comitê de *Bagging* se mostrou a melhor opção devido os *f1-score* apresentados e sua maior capacidade de predizer amostras novas.

Para o cenário exigido onde os classificadores não conhecem todos os produtos, mas devem distinguir

probabilisticamente, meio todas as Notas, os produtos conhecidos, não há como erradicar os erros de predição com amostras ambíguas devido a natureza da tecnologia de codificação de texto *Count Vectorizer*. Entretanto, a partir dos testes com ruído conclui-se que a metodologia apresentada consegue resolver o problema de agrupamento de produtos únicos a partir do campo textual de descrição. Seu uso pelo TCE-RN para a análise de preços se faz viável, porém orientado pelos produtos que apresentam melhor resultado para os dados possessos.

Por fim, tendo em vista as dificuldades encontradas nesta implementação, para os próximos passos deste trabalho, pode-se implementar diferentes técnicas de aprendizado de máquina como, por exemplo, arquiteturas baseadas em *Gradient Boosting*, comitês de modelos que favorecem modelos mais fracos e são popularmente utilizados para problemas de classificação tabular (Chen & Guestrin, 2016). Também, há espaço de experimentação com a metodologia utilizada para representar as descrições no domínio numérico, de forma a considerar o contexto das palavras descritas, como, por exemplo, *contextualized word embedding* (Reimers et al., 2019). Ainda, testes estatísticos devem ser implementados para tornar a compreensão dos diferentes aprendizados e a escolha dos melhores modelos melhor embasados.

Referências

- Brasil (2013a). *Acórdão 1785/2013, de 10 de julho de 2013*. Tribunal de Contas da União, Brasília, DF. https://pesquisa.apps.tcu.gov.br/#/documento/acordao-completo/*KEY%253AACORDAO-COMPLETO-1279889/DTRELEVANCIA%2520desc/0/sinonimos%253Dfalse
- Brasil (2013b). *Decreto 7.892, de 23 de janeiro de 2013. Regulamenta o Sistema de Registro de Preços previsto no art. 15 da Lei nº 8.666, de 21 de junho de 1993*. Diário oficial da República Federativa do Brasil. Poder Executivo, Brasília, DF.
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel ... & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, p. 108–122.
- Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *KDD*, São Francisco, CA. <https://arxiv.org/pdf/1603.02754.pdf>
- Koche, J. C. (2011). Fundamentos de metodologia científica. Petrópolis: Vozes. http://www.adm.ufrpe.br/sites/www4.deinfo.ufrpe.br/files/Fundamentos_de_Metodologia_Cienti%CC%81fica.pdf
- Ministério da Fazenda (2020a). *Conceito, uso e obrigatoriedade da nf-e (26 questões)*. <https://www.nfe.fazenda.gov.br/portal/perguntasFrequentes.aspx?tipoConteudo=E4+tmY+ODf4=>
- Ministério da Fazenda (2020b). *Manual de orientação do contribuinte - versão 6.00*. <https://www.nfe.fazenda.gov.br/portal/listaConteudo.aspx?tipoConteudo=33o15hhSYZk=>
- Ministério da Fazenda (2020c). *Ncm*. <https://receita.economia.gov.br/orientacao/aduaneira/classificacao-fiscal-de-mercadorias/ncm>
- Ministério da Fazenda (2020d). *Protocolo icms 42, de 3 de julho de 2009*. https://www.confaz.fazenda.gov.br/legislacao/protocolos/2009/pt042_09
- Tribunal de Contas do Estado da Paraíba (2020a). *Painéis preços*. <https://sagres.tce.pb.gov.br/paineis-precos/>
- Tribunal de Contas do Estado da Paraíba (2020b). *Preço da hora*. <https://precodahora.pb.gov.br/>
- Tribunal de Contas do Estado de Minas Gerais (2020). *Banco de preços tcemg*. <https://bancodepreco.tce.mg.gov.br/>
- Secretaria de Tributação do Rio Grande do Norte (2020). *Nota fiscal eletrônica*. http://www.set.rn.gov.br/content/Producao/Aplicacao/SET_v2/nfe/gerados/inicio.asp
- dos Santos, D. S. (2018). *Uma plataforma distribuída de mineração de dados para big data: um estudo de caso aplicado à secretaria de tributação do Rio Grande do Norte*. Dissertação (Mestrado em Engenharia de Software). Universidade Federal do Rio Grande do Norte, Natal, Brasil.
- Faceli, K., Lorena, A. C., Gama, J. & de Carvalho, A. C. P. L. F. (2011). *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*. Barueri, SP: LTC
- Gandini, A. (2020). *Banco de preços*. <https://github.com/alexgand/banco-de-precos>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press. <http://www.deeplearningbook.org>
- GS1 (2019). *Código EAN 13: entenda o que é, para que serve e como usar*. <https://blog.gs1br.org/codigo-ean-13-entenda-o-que-e-para-que-serve-e-como-usar/>
- GS1 (2020). *Gtin - número global do item comercial*. <https://www.gs1br.org/codigos-e-padroes/padroes-de-identificacao/gtin>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12: p. 2825–2830.

Reimers, N., Schiller, B., Beck, T., Daxenberger, J., Stab, C., Gurevych, I. (2019) *Classification and Clustering of Arguments with Contextualized Word Embeddings*. <https://arxiv.org/pdf/1906.09821.pdf>

Silva, D. S. (2014). *Manual de Orientação: pesquisa de preços*. Brasília, DF: Seção de Reprografia e Encadernação - Coordenadoria de Serviços Gerais. https://www.stj.jus.br/static_files/STJ/Licita%C3%A7%C3%B5es%20e%20contas%20p%C3%BAblicas/Manual%20de%20pesquisa%20de%20pre%C3%A7o/manual_de_orientacao_de_pesquisa_de_preços.pdf