



## DATA MINING TO COMBAT CARTELS: A CASE STUDY

## MINERAÇÃO DE DADOS NO COMBATE AOS CARTÉIS: ESTUDO DE CASO

**Pedro Sant'ana Bastos da Silva**

Mestre em Pesquisa Operacional e Inteligência Computacional – UCAM

<https://orcid.org/0009-0000-2154-7783>

[pedrosantanabastos@gmail.com](mailto:pedrosantanabastos@gmail.com)

**Ítalo de Oliveira Matias**

Doutor em Engenharia Civil - COPPE

<https://orcid.org/0000-0001-6683-5287>

[itombr@gmail.com](mailto:itombr@gmail.com)

**Aldo Shimoya**

Doutor em Genética e Melhoramento - UFV

<https://orcid.org/0000-0003-1098-9580>

[aldoshimoya@yahoo.com.br](mailto:aldoshimoya@yahoo.com.br)

**Fábio Freitas da Silva**

Doutor em Planejamento Regional e Gestão da Cidade – UCAM

<https://orcid.org/0000-0002-3712-8265>

[fabio1\\_freitas@hotmail.com](mailto:fabio1_freitas@hotmail.com)

**Fabiano de Oliveira Prado**

Professor do Instituto Federal Fluminense – IFF

<https://orcid.org/0009-0000-7349-4900>

[foprado@gmail.com](mailto:foprado@gmail.com)

### ABSTRACT

Open tendering is a purchasing process intended to give equal treatment to suppliers, creating a fair environment of competition between them and encouraging a drop in the prices paid by the government. When purchasing goods or contracting services, one of the main problems governments face worldwide is the so-called cartels in open tendering procedures. It involves a prior agreement by a group of companies, which take turns winning the bid, thus eliminating competition and increasing prices for the public coffers. This article focuses on the application of Data Mining (DM) in the fight

against cartel formation, using the Association Task, a methodology proposed by Silva (2011), implemented with the changes described in this article. The case study was conducted using an open database, available on the website of the Office of the Comptroller General (CGU, acronym in Portuguese), considering a specific object among the goods and services purchased by the public administration. As it is not the aim of this work to report to the CGU but rather to test the proposed methodology on an experimental basis, the Brazilian National Registry of Legal Entities (CNPJ, acronym in Portuguese) of the participating companies was replaced by random four-digit numbers, preserving the identity of the companies involved. Having done this, the method was presented and tested with a case study applied to a real database of tenders, available on the website of the Office of the Comptroller General (CGU), and the results proved promising.

**Keywords:** tenders; Office of the Comptroller General; public administration.

## RESUMO

As licitações públicas são processos de compra que visam dar tratamento isonômico aos fornecedores, de modo a criar um ambiente justo de competição entre eles, que estimule a queda dos preços pagos pelo governo. Um dos principais problemas enfrentados por governos do mundo todo no que tange a compras de bens ou contratação de serviços são os chamados cartéis em licitações públicas. Trata-se de um acordo prévio realizado por um grupo de empresas, que se revezam na posição vencedora, eliminando assim a competição e aumentando os preços para os cofres públicos. O objetivo deste artigo é aplicar Mineração de Dados (MD) no combate à formação de cartéis, por meio da Tarefa de Associação, metodologia proposta por Silva (2011), aplicada com alterações apresentadas neste artigo. O estudo de caso foi realizado com uma base de dados aberta, disponibilizada no sítio eletrônico da Controladoria Geral da União (CGU), adotando um objeto específico dentre os bens e serviços adquiridos pela administração pública. Como não é objetivo deste trabalho oferecer uma denúncia à CGU, e sim testar a metodologia proposta em caráter experimental, o CNPJ das empresas participantes foi substituído por números aleatórios de quatro dígitos, preservando, assim, em sigilo a identidade das empresas envolvidas. Feito isso, o método foi apresentado e testado com um estudo de caso aplicado a uma base de dados real de licitações, disponível no sítio eletrônico da Controladoria Geral da União (CGU), e os resultados se mostraram promissores.

**Palavras-chave:** licitações; controladoria geral da união; administração pública.

## 1 INTRODUÇÃO

Para que o governo realize suas compras com Legalidade, Impessoalidade, Moralidade, Publicidade e Eficiência, este realiza um processo administrativo

denominado de licitação, conforme ressalta Pietro (2019). Trata-se de um conjunto de procedimentos definidos em lei, de modo a garantir que os fornecedores possam competir de maneira justa entre si, garantindo ao mesmo tempo, o não favorecimento de nenhum particular, e também os melhores preços aos cofres públicos (NIEBUHR; NIEBUHR, 2018).

Contudo, há uma fraude que assombra os governos de todo o mundo no que diz respeito a compras públicas, denominada cartel. Trata-se de um acordo prévio entre os fornecedores participantes do edital, que se revezam na posição vencedora fazendo um rodízio, eliminando a competição, e aumentando assim os preços pagos pelo governo (SANTOS; SOUZA, 2018).

Por essa motivação, esse artigo visa aplicar a Descoberta de Conhecimento em Base de Dados (DCBD) a uma base de dados aberta de licitações, disponibilizada pela Controladoria Geral da União (CGU) em seu sítio eletrônico, visando testar uma metodologia capaz de revelar a formação de cartéis. Essa proposta foi inspirada em Silva (2011), com algumas alterações que serão apresentadas neste artigo.

## 2 REVISÃO BIBLIOGRÁFICA

Como explicam Sferra e Corrêa (2003), com o grande volume de dados digitais gerados pelos sistemas informatizados, o homem percebeu a oportunidade de extrair conhecimento útil dos mesmos, e a ferramenta capaz de fazê-lo ficou conhecida como Descoberta de Conhecimento em Base de Dados (DCBD). De acordo com Fayyad et al. (1996), trata-se de um processo não trivial, estruturado em cinco etapas, interativo e iterativo, que busca extrair padrões ocultos nos bancos de dados, que sejam novos e potencialmente úteis para a tomada de decisões.

Como pode ser observada na Figura 1, adaptada de Fayyad et al. (1996), as cinco etapas da DCBD são: Seleção, Pré-Processamento, Formatação, Mineração de Dados (MD) e Interpretação de Resultados. Conforme explicam Sumathi e Sivanandam (2006), elas podem ser traduzidas como:

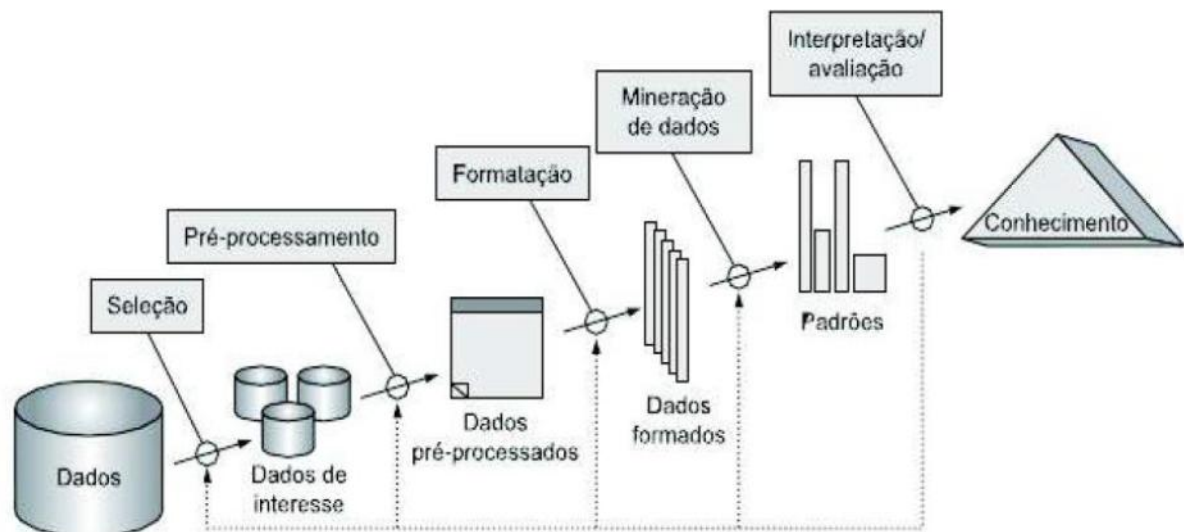
- Seleção: Etapa em que os dados úteis ao estudo são selecionados, enquanto aqueles que não possuem importância são descartados;

- Pré-processamento e Formatação: Etapas de preparo da base de dados para a etapa seguinte. Compreende a adequação de itens incorretos, em branco, remoção de ruídos, formatação a padrões específicos de cada algoritmo de mineração, entre outros;

- Mineração de Dados (MD): Etapa mais importante do processo, onde algoritmos são executados na busca por padrões ocultos na base de dados. Dentre as muitas Tarefas da mineração, existe a Classificação, a Clusterização, a Associação, entre outras;

- Interpretação e Avaliação dos Resultados: Etapa conclusiva, onde se avalia se os padrões obtidos são úteis ou não para tomada de decisões. Cabe destacar que a análise deste produto final pode sugerir no retorno a alguma etapa anterior, a fim de se obter melhorias.

Figura 1- Input, Etapas (Seleção, Pré-Processamento, Formatação, Mineração de Dados e Interpretação de Resultados) e Output da DCBD



Fonte: Própria, adaptada de Fayyad et al. (1996).

Como mencionado, uma das mais importantes tarefas de Mineração de Dados (MD) é a chamada Associação, que visa identificar a relação entre variáveis com regras de causa e consequência (SE... ENTÃO...), conforme explicam Camilo e Silva (2009).

Exemplificando com este estudo de caso, poderíamos encontrar regras do tipo: SE a empresa “A” e a empresa “B” participam de uma licitação, ENTÃO a empresa “C” também participa. Observe que esta é uma situação em que as três empresas tendem a concorrer juntas, logo, há certa suspeita de formação de cartel. Reescrevendo a regra no padrão formal, ficaria: “A = Sim”, “B = Sim” → “C = Sim”.

Como explica Larose (2005), essas regras possuem medidas de qualidade, chamadas de Suporte e Confiança. A primeira diz respeito ao percentual de vezes que a “causa” (lado esquerdo da regra) aparece na base de dados, já a segunda indica o percentual de vezes que a “consequência” (lado direito da regra) é verdadeira, caso ocorra a “causa”. Para ilustrar, basta supor que a base de dados possui 10.000 registros, e que as empresas “A” e “B” apareçam juntas por 20 vezes, logo, o Suporte dessa regra é obtido dividindo 20 por 10.000 (0,2%). Se, dentre essas 20 aparições conjuntas de “A” e “B”, a empresa “C” aparecesse 15 vezes, então a Confiança seria obtida dividindo 15 por 20 (75%).

Entendidos esses conceitos iniciais, pode-se entender as ferramentas usadas na Associação. De acordo com Goldschmidt e Passos (2005), os algoritmos que produzem as chamadas Regras de Associação (RA) recebem como input valores mínimos de Suporte e Confiança, a fim de gerar regras que se enquadrem nesses dois requisitos. Como explica Tan, Steinbach e Kumar (2009), quando a base de dados é muito volumosa, é inviável calcular esses dois parâmetros para todas as regras possíveis, já que a complexidade é de  $3d - 2d + 1$ , onde “d” é o número de itens da base de dados.

Nesse contexto, em 1994 foi proposto por Agrawal e Srikant (1994) o algoritmo Apriori. Esta ferramenta usa de estratégias que evitam o cálculo exaustivo de todas as regras, reduzindo assim consideravelmente o tempo de processamento. O funcionamento do algoritmo em mais detalhes pode ser encontrado em Tan, Steinbach e Kumar (2009), e será brevemente explicado a seguir.

#### -PODA BASEADA NO SUPORTE:

O Primeiro artifício usado pelo algoritmo Apriori que reduz consideravelmente o tempo de processamento das regras se baseia na seguinte premissa: se um

conjunto é frequente, então todos os seus subconjuntos também serão, e analogamente, se um subconjunto não é frequente, então seus superconjuntos também não serão.

A partir dessa afirmação, se um subconjunto  $\{a, b\}$  não é frequente, o algoritmo elimina todas as regras contendo seus superconjuntos  $\{a, b, \dots\}$ , sem necessidade de mais cálculos. É importante perceber que essa estratégia é capaz de reduzir consideravelmente o volume de cálculos executados pelo computador.

#### -PODA BASEADA NA CONFIANÇA:

Além disso, o algoritmo possui uma estratégia inteligente para realizar a poda pelo valor mínimo de Confiança. Primeiramente o Apriori se utiliza de regras com apenas um elemento consequente, como exemplo:  $\{a, c, d\} \rightarrow \{b\}$  e  $\{a, b, d\} \rightarrow \{c\}$ . Caso elas apresentem Confiança satisfatória, ocorre a fusão das mesmas, gerando uma nova regra:  $\{a, d\} \rightarrow \{b, c\}$ . A nova Confiança é calculada, e se atendido o valor mínimo, ela é mantida.

De maneira análoga, se a referida regra possui Confiança não satisfatória, então todas as regras contendo aquele consequente também são descartadas, reduzindo consideravelmente a complexidade do problema. Em outras palavras, se  $\{b, c, d\} \rightarrow \{a\}$  tem baixa Confiança ela será eliminada, e conseqüentemente, as regras  $\{c, d\} \rightarrow \{a, b\}$  e  $\{b, d\} \rightarrow \{a, c\}$  também serão.

### 3 METODOLOGIA

A base de dados utilizada está disponível no portal transparência da CGU, visto em CGU (2019b), acessado em 01 de julho de 2019, com licitações abrangendo um período de cinco anos. Na tabela 1, adaptada da base original com a omissão de certas informações, podem ser observados os atributos: números dos processos, informações sobre órgãos licitantes, unidades de destino, bens ou serviços adquiridos, fornecedores participantes e vencedores.

Tabela 1 - Exemplo adaptado da base de dados utilizada

Núm. Proc.	Cód. Órgão	Nome Órgão	Cód. UG	Nome UG	Cód. Item	Descrição Item	CNPJ	Nome Particip.	Vencedor
A	B	C	D	E	1103...	MAN/REF PREDIAL	...	...	NÃO
A	B	C	D	E	1103...	MAN/REF PREDIAL	...	...	NÃO
A	B	C	D	E	1103...	MAN/REF PREDIAL	...	...	NÃO
A	B	C	D	E	1103...	MAN/REF PREDIAL	...	...	NÃO
A	B	C	D	E	1103...	MAN/REF PREDIAL	...	...	NÃO
A	B	C	D	E	1103...	MAN/REF PREDIAL	...	...	NÃO
A	B	C	D	E	1103...	MAN/REF PREDIAL	...	...	NÃO
A	B	C	D	E	1103...	MAN/REF PREDIAL	...	...	NÃO
A	B	C	D	E	1103...	MAN/REF PREDIAL	...	...	NÃO
A	B	C	D	E	1103...	MAN/REF PREDIAL	...	...	NÃO
F	G	H	I	J	9266...	MAT P/ VNI...	...	...	NÃO
F	G	H	I	J	9266...	MAR P/ VNI...	...	...	NÃO
F	G	H	I	J	9266...	MAR P/ VNI...	...	...	SIM
F	G	H	I	J	9266...	MAR P/ VNI...	...	...	NÃO
F	G	H	I	J	9266...	MAR P/ VNI...	...	...	NÃO
F	G	H	I	J	9266...	MAR P/ VNI...	...	...	NÃO
F	G	H	I	J	9266...	MAR P/ VNI...	...	...	NÃO
F	G	H	I	J	9266...	MAR P/ VNI...	...	...	NÃO
F	G	H	I	J	9266...	MAR P/ VNI...	...	...	SIM
F	G	H	I	J	9266...	MAR P/ VNI...	...	...	NÃO

Fonte: Dados da pesquisa.

Com o intuito de preparar a base de dados para aplicação do algoritmo Apriori, foi necessário excluir certos atributos, conforme metodologia descrita em Silva (2011):

1) Número do processo: Este atributo é omitido para execução do algoritmo de mineração, porém permanece na base para fins de identificação da licitação.

2) Nome e código do órgão licitante e do órgão de destino: Nada impede que os cartéis participem em licitações nos mais variados órgãos, e por essa razão, essa informação não é relevante, e foi excluída.

3) Nome e código do item comprado: Conforme descrito no resumo, este estudo de caso foi feito adotando um objeto específico de compra. Sendo assim, todas as licitações envolvendo bens ou serviços distintos a este, foram excluídas.

4) Nome fantasia e CNPJ do participante: Devido ao caráter experimental da pesquisa, que visa testar a metodologia e não oferecer uma denúncia propriamente dita aos órgãos de controle, o nome fantasia e o CNPJ das empresas foram excluídos, e substituídos por um número de identificação aleatório de quatro dígitos.



5) Identificação do Vencedor: Este atributo não participa das Regras de Associação (RA), porém após a mineração ele será útil para verificar se os cartéis formados realmente obtiveram sucesso.

### 3.1 Pré-processamento

Conforme descrito em Silva (2011), foi gerado um algoritmo capaz de transformar a base de dados apresentada nos itens 4.4.1 e 4.4.2 em uma matriz, onde cada linha ou instância representa uma licitação, e cada coluna ou atributo representa uma empresa. Dessa forma, quando uma empresa participa de determinada licitação, a interseção entre essas linhas e colunas recebe a variável booleana "SIM", e caso contrário, "NÃO".

Entretanto, percebeu-se que essa matriz não era satisfatória, pelo fato de o algoritmo de Associação gerar, em sua grande maioria, regras com a variável "NÃO". Estas regras não trazem nenhuma evidência da formação de cartéis, logo, para solucionar este problema, o "NÃO" foi substituído pelo caractere "?". Isso foi feito porque, na linguagem do software utilizado, isso significa uma variável não informada ou missing, e conseqüentemente, são geradas apenas as regras com a variável "SIM".

Feito isso, ainda há três alterações que o algoritmo deve fazer de modo a preparar a base para a Mineração de Dados (MD): exclusão das linhas repetidas; exclusão de empresas com apenas uma licitação; exclusão de licitações dispensadas ou inexigíveis.

### 3.2 Mineração de dados

Para realização da Mineração de Dados (MD), etapa em que definitivamente os algoritmos são executados na busca por padrões ocultos, foi utilizado o pacote de software WEKA (Waikato Environment for Knowledge Analysis). Este software é de uso gratuito, possui código aberto, e foi desenvolvido inicialmente pela Universidade de Waikato, na Nova Zelândia. A escolha por este aplicativo se deu pelo fato de que esta ferramenta, além de possuir uma interface de fácil compreensão e uso, também contempla às necessidades do trabalho, abrangendo diversos algoritmos com a tarefa de Associação, dentre eles o Apriori (WAIKATO, 2019).



Como mencionado no referencial teórico, para que sejam geradas as Regras de Associação (RA), é necessário o fornecimento de dois parâmetros de entrada: o Suporte Mínimo e a Confiança Mínima. Conforme explicado por Silva (2011), altos valores de Suporte Mínimo irão excluir regras importantes, sendo assim, foi adotado 0,1% para esse parâmetro. Em consequência do tamanho da base de dados, este percentual corresponde a 14 aparições do conjunto de empresas que formam a “causa” da regra.

Com relação à Confiança Mínima, conforme definido em Silva (2011), foi adotado o valor de 90%, considerado por especialistas um percentual razoável. Este índice, pela dimensão da base de dados usada, equivale a um mínimo de 12 licitações conjuntas das empresas da figuram a “consequência” da Associação.

“A = Sim”; “B = Sim” → “C = Sim”. (Suporte = 0,1% e Confiança = 90%)

Com o baixo valor de Suporte Mínimo adotado, é natural que haja a geração de um grande volume de Regras de Associação (RA), que neste caso foi de 26.821. Sendo assim, conforme explica Silva (2011), é necessário algum critério capaz de ranquear as melhores regras a fim de tornar viável a análise futura pela ação humana. Neste ponto, este artigo altera a metodologia original, propondo um novo Algoritmo de Seleção de Regras.

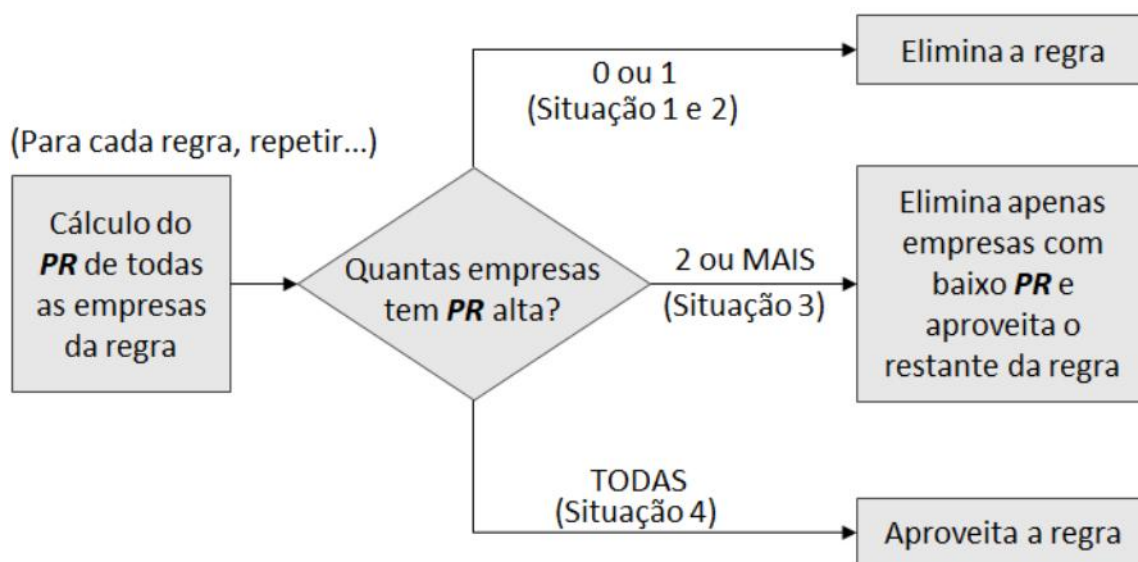
O algoritmo aqui proposto se baseia em quatro situações distintas, a fim de eliminar totalmente ou parcialmente regras que não atendam seus requisitos. Para exemplificar essas quatro situações, pode-se imaginar que as empresas “A”, “B”, “C” estão presentes no lado esquerdo da regra, com uma frequência de 15.

A palavra-chave para entender o método criado chama-se Participação Relativa (PR), que varia de 0 a 100%, dada pela razão entre o número de participações em conjunto, e o número total de participações na base de dados, conforme equação:  $PR = 100 \times (N^{\circ} \text{ Participações Conjuntas} / N^{\circ} \text{ Participações Total})$ .

A título de exemplo, caso a empresa “A” tivesse participado de 140 licitações na base de dados, esta teria uma PR no valor de 10%. Em contrapartida, caso a mesma empresa tivesse participado de apenas 15 licitações, teria uma PR de 93%.

Na Figura 2 pode-se observar o fluxograma do Algoritmo de Seleção de Regras.

Figura 2- Algoritmo proposto para seleção de regras



Fonte: Próprio autor.

Resta agora definir qual é o valor de corte da PR, de modo a classifica-la como alta ou baixa na aplicação do método. Neste estudo de caso foram feitos experimentos com diversos valores de Participação Relativa Mínima (PRM), onde o número de regras remanescentes após execução do algoritmo varia sensivelmente de acordo com PRM adotada.

Cabe ressaltar que o objetivo deste artigo não é fixar um valor de PRM universal válido para qualquer situação. Na verdade, a proposta é deixar esse valor em aberto como um parâmetro de entrada, a ser variado caso a caso pelo especialista encarregado da auditoria. Neste estudo, por exemplo, foi adotado um PRM de 50% inicialmente, e avaliadas as 11 regras remanescentes, como as principais suspeitas do estudo. Em seguida, foi adotado um PRM de 40% e avaliadas as 66 regras resultantes, que conseqüentemente apresentam indícios menos relevantes do que os anteriores.

#### 4 RESULTADOS E DISCUSSÃO

Após execução da Mineração de Dados (MD) por intermédio do algoritmo Apriori, com Suporte Mínimo de 0,1% e Confiança Mínima de 90, foram obtidas um total de 26.821

Regras de Associação (RA). Feito isso, essas regras foram alvo do Algoritmo de Seleção de Regras, primeiramente com um PRM de 50%, gerando 11 regras, e em seguida com o PRM de 40%, gerando 66 regras remanescentes.

O CNPJ das empresas foi substituído por um número de identificação aleatório, devido ao caráter experimental do estudo, que visa apenas avaliar a metodologia, e não oferecer denúncia contra as companhias. Por esta razão, cabe deixar claro que não há nenhuma rastreabilidade entre o ID apresentado nas suspeitas, e o CNPJ real das empresas. Segue abaixo alguns resultados em caráter exemplificativo.

SUSPEITA 1: Um grupo formado por quatro fornecedores identificados pelos números 4982, 5681, 6167 e 3804 participou em conjunto de 18 licitações, com Participações Relativas (PR) acima de 50% para todas as empresas. Indo mais a fundo, percebeu-se que este grupo venceu dez, das 18 licitações disputadas, o que confere a estas empresas um forte indício de formação de cartel.

SUSPEITA 2: Um grupo formado por quatro fornecedores identificados pelos números 1634, 5023, 5082 e 2286 participou em conjunto de 42 licitações, porém com PR baixa para os dois últimos, o que os exclui da suspeita. Analisando melhor a relação entre os dois primeiros, percebe-se que as empresa 1634 e 5023 possuem PR de 82% e 51% respectivamente, o que indica chances razoáveis da formação de cartel entre elas. Apesar de o grupo ter vencido apenas sete licitações das 42 disputadas, o alto valor de PR da primeira chama a atenção para este possível conluio.

SUSPEITA 3: Um grupo formado por três fornecedores identificados pelos números 1721, 4600 e 3226 participou em conjunto de 19 licitações, com alta PR para os dois primeiros, e baixa para o último, excluído então da suspeita. Analisando as empresas 1721 e 4600, percebe-se que além dessas empresas terem vencido cinco das 19 licitações, elas apresentam PR de 53% e 51%, o que indica suspeita de cartel.

SUSPEITA 4: As empresas identificadas pelos números 1720 e 5023 participaram juntas de 47 licitações, o que representa uma PR de 68% e 53%, respectivamente. Apesar do elevado valor de licitações conjuntas, percebeu-se que este grupo não venceu nenhuma das referidas licitações, o que torna essa suspeita menos importante.

SUSPEITA 5: As empresas identificadas pelos números 3106 e 2599 participaram juntas de 19 licitações, o que representa uma PR de 76% e 63%, respectivamente. Este alto valor de PR, aliado ao fato de que este grupo venceu sete das 19 licitações, torna essa suspeita muito importante, trazendo à luz forte indício de formação de cartel entre esses dois fornecedores.

SUSPEITA 6: As empresas identificadas pelos números 1163 e 2437 participaram juntas de 24 licitações, o que se traduz em altos valores de PR, atingindo 92% para a primeira, e 69% para a segunda. Apesar de o grupo ter vencido apenas duas das 24 licitações, há forte indício da formação de cartel pelo altíssimo valor da Participação Relativa (PR) do fornecedor 1163.

## 5 CONCLUSÃO

A formação de cartéis por empresas privadas inidôneas visando lesar os cofres públicos é um problema vivido por governos de todo o mundo, como explicado na Introdução deste artigo. Neste contexto, surge como forte aliada nos processos de auditoria governamental, a Ciência de Dados, trazendo mais eficiência a essa fiscalização. É notório que os resultados obtidos por este artigo, que apontam empresas com mais chances de atuação ilícita, são de grande valia no planejamento dos auditores, que podem orientar seus esforços da maneira mais racional possível.

Sobre a etapa de Mineração de Dados (MD) também cabe algumas reflexões importantes. A primeira é que a tarefa de Associação se mostrou ideal na descoberta da relação entre as empresas que tendem a concorrer juntas, logo, conclui-se que esta Tarefa atendeu bem aos objetivos deste estudo. Outro ponto de discussão é sobre o parâmetro de entrada do algoritmo Apriori, Suporte Mínimo, que como explica Silva (2011), deve ter baixo valor a fim de não se perder regras valiosas. Isto é comprovado na análise de resultados, onde se pode observar que ótimas suspeitas possuem Suporte de apenas 0,1%. Por fim, conclui-se que a ferramenta WEKA, utilizada na MD, se mostrou bastante eficiente, de fácil operação e compreensão, e capaz de produzir resultados coerentes, de forma rápida, e de fácil interpretação.

Na etapa de Interpretação de resultados, tem-se o ponto mais importante deste artigo, pelo fato de que a metodologia original de Silva (2011) foi alterada, dando origem a um novo Algoritmo de Seleção de Regras. Esta ferramenta pós-mineração é essencial, pois o número de regras gerado é enorme, e atingiu mais de 26 mil neste estudo de caso. A palavra chave ao entendimento dessa proposta é a Participação Relativa (PR), parâmetro que se mostrou eficiente na seleção de regras, e trouxe bons resultados finais. Cabe lembrar que este artigo não fixou a Participação Relativa Mínima (PRM), deixando esta como um input do algoritmo, a ser analisado caso-a-caso pelos auditores de acordo com a situação.

Por fim, conclui-se que as suspeitas levantadas neste estudo são de grande utilidade para que, durante o planejamento de uma auditoria governamental, essas empresas com indícios de fraude sejam auditadas de maneira prioritária, por apresentarem mais chances de formação de cartel. Note que tal recurso provavelmente tornará o processo mais assertivo, diminuindo assim a ocorrência deste conluio, e economizando recursos públicos com compras e contratações governamentais.

## REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. Fast Algorithms for Mining Association Rules in Large Databases. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994. **Proceedings** [...]. Hove, East Sussex: Morgan Kaufmann, 1994. p. 487-499.

CAMILO, C.; SILVA, J. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Goiânia: Universidade Federal de Goiás, 2009.

CGU. **PORTAL TRANSPARÊNCIA**, 2019b. Disponível em: <http://transparencia.gov.br/download-de-dados/licitacoes>. Acesso em: 1 jul. 2019.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **Advances in knowledge discovery and data mining**. Menlo Park, Calif.: AAAI Press, USA, 1996.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de Pesquisa**. Porto Alegre: UFRGS, 2009.

GOLDSCHMIDT, R.; PASSOS, E. **Data mining**: um guia prático, conceitos, técnicas, ferramentas, orientações e aplicações. São Paulo: Elsevier, 2005.

JUNG, C. F. **Metodologia Científica**: ênfase em Pesquisa Tecnológica. 3. ed. Porto Alegre: Penso, 2003.

LAKATOS, E. M.; MARCONI, M. de A. **Metodologia científica**. 7. ed. São Paulo: Atlas, 2017.

LAROSE, D. T. **Discovering knowledge in data**: an introduction to data mining. Hoboken, N. J. Wiley-Interscience, 2005.

PIETRO, M. S. Z. D. **Direito Administrativo**. 32. ed. São Paulo: Forense, 2019.

SANTOS, F. B.; SOUZA, K. R. de. **Como combater a corrupção em licitações**: detecção e prevenção de fraudes. 3. ed. Belo Horizonte: Fórum, 2020.

SFERRA, H. H.; CORRÊA, A. C. J. Conceitos e Aplicações de Data Mining. **Revista de Ciência & Tecnologia**, Piracicaba, v. 11, n. 22, p. 19-34, 2003.

SILVA, C. V. S. **Agentes de Mineração e sua Aplicação no Domínio da Auditoria Governamental**. Brasília: Universidade de Brasília, 2011.

SUMATHI, S.; SIVANANDAM, S. N. **Introduction to data mining and its applications**. Berlin; New York: Springer, 2006.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introdução ao datamining**: mineração de dados. Rio de Janeiro: Ciência Moderna, 2009.

WAIKATO, U. **Weka 3**: Machine Learning Software in Java. 2019. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>. Acesso em: 21 out. 2019.

YIN, R. K. **Case study research**: design and methods. 5th. ed. Los Angeles: SAGE, 2014.

NIEBUHR, J. de M.; NIEBUHR, P. de M. **Licitações e contratos das estatais**. Belo Horizonte: Fórum, 2018.