



CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS  
CURSO DE ENGENHARIA DE COMPUTAÇÃO

# **AVALIAÇÃO DA APLICABILIDADE DA MINERAÇÃO DE DADOS PARA O AUXÍLIO NA DETECÇÃO DE INDÍCIOS DE FRAUDES EM COMPRAS LICITADAS PELO GOVERNO DE MINAS GERAIS**

**LUÍS FERNANDO DE QUADROS GUADAGNIN**

Orientador: Ismael Santana Silva  
Centro Federal de Educação Tecnológica de Minas Gerais

BELO HORIZONTE  
JUNHO DE 2018

**LUÍS FERNANDO DE QUADROS GUADAGNIN**

**AVALIAÇÃO DA APLICABILIDADE DA MINERAÇÃO DE  
DADOS PARA O AUXÍLIO NA DETECÇÃO DE INDÍCIOS  
DE FRAUDES EM COMPRAS LICITADAS PELO  
GOVERNO DE MINAS GERAIS**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador:     Ismael Santana Silva  
                          Centro Federal de Educação Tecnológica de  
                          Minas Gerais

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS  
CURSO DE ENGENHARIA DE COMPUTAÇÃO  
BELO HORIZONTE  
JUNHO DE 2018

Esta folha deverá ser substituída pela cópia digitalizada da folha de aprovação fornecida.

# Resumo

As compras realizadas por órgãos públicos, como é o caso de um governo estadual, passam pelo processo de licitação. A licitação é um processo administrativo, isonômico, no qual a administração pública seleciona a proposta mais vantajosa, menos onerosa e com melhor qualidade possível, para a contratação de um serviço, compra de um produto, locação ou alienação. Ela deve ser pública e acessível a qualquer cidadão. No entanto, estratégias têm sido usadas para burlar esse processo, sendo uma dessas estratégias a formação de cartel. Cartel é um acordo informal entre empresas objetivando o aumento de preços dos produtos e/ou serviços oferecidos. No caso de licitações públicas, trata-se de um acordo entre competidores para manipular o vencedor de determinada compra. O Portal de Compras de Minas Gerais é o sistema pelo qual todas as compras do estado são realizadas, porém os dados de processos de compra não são disponibilizados de uma forma legível para programas automatizados, o que torna difícil a análise dessas informações. Devido a isso, nunca foi realizada uma análise desses dados em busca de indícios de cartéis. O presente trabalho coletou dados relacionados a compras públicas realizadas no Portal de Compras de Minas Gerais e por meio de técnicas de mineração de dados analisou os mesmos com o objetivo de identificar grupos de fornecedores que realizam compras conjuntas recorrentemente. Por meio das técnicas de mineração de dados, observou-se que dois fornecedores participaram de 89,53% das licitações em conjunto, sendo que um ou outro foi vencedor de 76,62% dessas licitações, o que pode ser considerado um indício de formação de cartel. Com os resultados obtidos, foi desenvolvida uma função de avaliação dos grupos de fornecedores minerados que apresentam indícios relevantes de formação de cartéis e em seguida foi feita uma avaliação com dois especialistas. Ambos ressaltaram a importância dos procedimentos licitatórios como meio de aquisição de bens e serviços, mas reconheceram a fragilidade do sistema. Segundo eles, as técnicas utilizadas neste trabalho se mostraram úteis para dirigir o olhar no caso de uma investigação. Também apontaram para outras possibilidades de mineração de dados complementares à usada neste trabalho. Portanto a mineração de dados se mostrou auxiliar na detecção de indícios de fraudes em compras licitadas pelo governo de Minas Gerais.

**Palavras-chave:** Licitações públicas. Mineração de dados. Cartel. Compras governamentais.

# Lista de Figuras

Figura 1 – Categorização dos itens de compras no CATMAS . . . . .	16
Figura 2 – Exemplo de classificação de um item de compra . . . . .	17
Figura 3 – Quantidade de itens de material ou serviço por grupo . . . . .	18
Figura 4 – Quantidade de fornecedores por grupo de fornecimento . . . . .	18
Figura 5 – Quantidade de processos de compra concluídos por ano . . . . .	19
Figura 6 – Quantidade de tipos de processos de compras concluídos por ano . . . .	20
Figura 7 – Exemplo de relacionamento entre pregão, lotes do pregão e fornecedores	21
Figura 8 – Modelo físico do relacionamento entre lotes do pregão e fornecedores	22
Figura 9 – Exemplo de dados inseridos na tabela de relacionamento entre lotes de pregões e fornecedores . . . . .	22
Figura 10 – Tabela de relacionamento gerada para servir como entrada de softwares de mineração . . . . .	23
Figura 11 – Exemplo de arquivo ARFF de entrada do Weka . . . . .	23
Figura 12 – Tabela pré-processada de participações de fornecedores em lotes de pregão . . . . .	24
Figura 13 – Algoritmo para gerar regras de associação com apenas um elemento antecedente . . . . .	25
Figura 14 – Execução do novo algoritmo, considerando F1 como antecedente, para um grupo de 4 fornecedores . . . . .	26
Figura 15 – Regras de associação geradas . . . . .	27

# Lista de Abreviaturas e Siglas

CAGEF	Cadastro Geral de Fornecedores
CATMAS	Catálogo de Materiais e Serviços
COTEP	Cotação Eletrônica de Preços
DGA	Dados Governamentais Abertos
EM	Expectation-Maximization
FAPEMIG	Fundação de Amparo a Pesquisa do Estado de Minas Gerais
KDD	Knowledge Discovery in Database
MPE	Microempresas e Empresas de Pequeno Porte
NF-e	Nota Fiscal Eletrônica
SAGRES	Sistema de Acompanhamento da Gestão dos Recursos da Sociedade
SEPLAG	Secretaria de Planejamento e Gestão do Estado de Minas Gerais
SIRP	Sistema Integrado de Registro de Preços
SISGOP	Sistema de Gestão de Ocorrências Policiais
TIC	Tecnologia de Informação e Comunicação

# Sumário

<b>1 – Introdução</b>	<b>1</b>
<b>2 – Referencial Teórico</b>	<b>3</b>
2.1 Legislação Envolvendo Licitação e Transparência	3
2.2 Dados Governamentais Abertos	4
2.3 Descoberta de Conhecimento em Banco de Dados	6
2.4 Dados de Compras Governamentais	9
<b>3 – Trabalhos Relacionados</b>	<b>12</b>
<b>4 – Metodologia</b>	<b>14</b>
<b>5 – Caracterização dos Dados de Compras Governamentais</b>	<b>16</b>
5.1 Itens de Compra	16
5.2 Fornecedores	17
5.3 Compras	19
<b>6 – Mineração dos Dados: Técnica e Resultado</b>	<b>21</b>
6.1 Extração dos Dados	21
6.2 Pré-processamento dos Dados	22
6.3 Avaliação de Técnicas e Novo Algoritmo	24
6.4 Regras de Associação Geradas	26
6.5 Classificação dos Grupos de Fornecedores	26
<b>7 – Avaliação dos Especialistas</b>	<b>30</b>
7.1 Entrevista com o Especialista 1	30
7.2 Entrevista com o Especialista 2	31
<b>8 – Considerações Finais</b>	<b>34</b>
<b>Referências</b>	<b>36</b>

# 1 Introdução

A quantidade de informação disponibilizada nos dias de hoje é enorme e vem crescendo exponencialmente. Isso torna possível a realização de tarefas que previamente não poderiam ser executadas, como o reconhecimento de tendências de negócios, a prevenção de doenças, o combate ao crime, entre muitas outras ([ECONOMIST, 2010](#)). Pessoas e organizações produzem e coletam diferentes tipos de dados para auxiliar na realização de suas tarefas. Essa grande massa de dados está além das capacidades de captura, armazenamento, gerência e análise. Esse grande volume de dados com essas características é comumente referenciado como Big Data ([MANYIKA et al., 2011](#)).

Órgãos públicos têm uma importância ainda maior nesse aspecto, não só pela quantidade e pela centralização dos dados que coletam, mas também porque a maior parte dos dados governamentais devem ser públicos por lei e, por isso, devem ser abertos e disponibilizados para uso público. Quando esses dados são indexados na web, estão disponíveis em formato compreensível e não possuem qualquer impedimento legal para uso e replicação, eles são considerados dados governamentais abertos (DGA) ([EAVES, 2009](#)).

Em anos recentes, um número de movimentos a favor da organização e disponibilização de dados abertos surgiu ao redor do mundo, sendo a transparência e o reuso de dados os principais objetivos ([ATTARD et al., 2015](#)). Como resultado desses movimentos, pode-se citar exemplos como o Portal da Transparência do Governo Federal, o Tribunal de Contas do Estado da Paraíba e o Portal do Governo Aberto de São Paulo. Porém, apesar dessa tendência de abertura dos dados governamentais estar apenas crescendo, o número de dados abertos ainda é baixo se comparado aos dados não abertos, devido aos custos e à falta de interesse cultural ([ATTARD et al., 2015](#)).

Alguns trabalhos já fizeram a análise de DGA e chegaram a conclusões muito úteis para a tomada de decisões pelas autoridades de órgãos públicos. Como exemplo, pode-se citar [Braz et al. \(2009\)](#), que utilizou dados de boletins de ocorrência da Polícia Militar para auxiliar ações estratégicas da polícia baseadas no comportamento dos criminosos e vítimas; [Reis \(2014\)](#), que utilizou dados da Polícia Rodoviária Federal para identificar fatores de contribuição de acidentes rodoviários na BR-381; [Silva e Ralha \(2011\)](#) e [Grilo Júnior \(2010\)](#), que utilizaram respectivamente dados da Controladoria Geral da União e do Tribunal de Contas do Estado da Paraíba para identificar cartéis em licitações públicas. Com isso, pode-se ter uma ideia da relevância da análise de DGA como um mecanismo de auxílio na administração pública.

Nesse contexto está inserido o Portal de Compras do Estado de Minas Gerais. O Portal é um sistema gerenciado pela Secretaria de Planejamento e Gestão do Estado



de Minas Gerais no qual empreendedores podem vender produtos e fornecer serviços ao Governo de Minas Gerais, por meio de processos licitatórios ou por dispensa e inexigibilidade de licitação. A transparência de todos os processos de compra é definida pela Lei de Acesso à Informação Nº 12.527 de 2011, que regulamenta o direito constitucional de acesso dos cidadãos às informações públicas, o que significa que todos os dados de compra do Estado de Minas Gerais são públicos e estão disponíveis, porém, a maneira como são disponibilizados impossibilita a leitura dos dados automatizada por programas.

Devido à falta de disponibilidade da base de dados estruturada do Portal de Compras de Minas Gerais, nunca foi feita uma análise desses dados. O objetivo desse trabalho é realizar uma caracterização dos dados de compras e licitações do governo de Minas Gerais com o intuito de prover insumos para tomadas de decisões e demonstrar, a partir do estudo de caso, as técnicas aplicáveis para extração desses insumos.

Este presente trabalho visa utilizar técnicas de mineração e análise de dados para encontrar indícios de formação de cartéis em licitações públicas do tipo pregão eletrônico nas compras do estado de Minas Gerais. A descoberta de padrões baseada na participação conjunta de fornecedores em licitações pode otimizar o trabalho de auditoria pública.

Para isso, foi feita inicialmente a caracterização dos dados de compras governamentais. Para a mineração dos dados, o estudo foi dividido em cinco partes: extração dos dados, pré-processamento dos dados, avaliação de técnicas de mineração e criação de um novo algoritmo, geração de regras de associação e classificação dos grupos de fornecedores.

As técnicas de mineração de dados permitiram a criação de uma tabela com fornecedores que participam de licitações em conjunto e a respectiva métrica de vitórias de um ou outro componentes desses grupos. Os dados encontrados podem ser utilizados para investigar possíveis formações de cartéis caso houver esse interesse.

Por fim, foram realizadas entrevistas com dois especialistas, para os quais foram apresentados os resultados deste trabalho para serem avaliados. Ambos ressaltaram a importância dos procedimentos licitatórios, mas reconheceram a fragilidade do sistema. Ao avaliarem os resultados obtidos por meio das técnicas utilizadas neste trabalho, argumentaram serem úteis para dirigir o olhar no caso de uma investigação. Além disso, sugeriram a adição de outros dados complementares, como os preços dos itens de compra e a razão de vitória de cada fornecedor dentro de um grupo.

## 2 Referencial Teórico

Para o entendimento do processo de análise dos dados do Portal de Compras, neste capítulo serão abordados o contexto e as características dos dados governamentais abertos, relacionando-os ao processo de compras públicas. Em seguida, será explicada brevemente a descoberta de conhecimento em banco de dados com ênfase nas técnicas de mineração de dados.

### 2.1 Legislação Envolvendo Licitação e Transparência

O processo licitatório no Brasil, parece ter ganho notoriedade a partir da publicação do Decreto-Lei nº 2.300/86, que foi atualizado por decretos posteriores (Decretos–Lei nº 2.348 e 2.360). Com eles se institui, pela primeira vez, o “Estatuto das Licitações e Contratos Administrativos”, o que representou um conjunto de normas gerais e específicas envolvendo a licitação.

Porém, foi com a Constituição de 1988, por meio do artigo 37, inciso XXI, que a licitação recebeu status de princípio constitucional, de observância obrigatória pela Administração Pública direta e indireta de todos os entes da Federação, o que inclui a União, os Estados, o Distrito Federal e os Municípios. Em 1993 a União promulgou a Lei nº 8.666/93, que instituiu as normas gerais para licitações e para os contratos da Administração Pública, em todos os âmbitos. A Medida Provisória nº 2.026/00 instituiu, por sua vez, a modalidade licitatória chamada “Pregão”.

O pregão pode ser resumido como um leilão inverso e é esperado que, por meio dele, o licitante ofereça lances para baixo, propiciando à administração uma compra mais vantajosa. Essa nova modalidade visa garantir a aquisição de bens e serviços e permite maior agilidade nas aquisições. À medida que se percebem deficiências, novas adaptações vão sendo feitas, por meio de medida provisória e/ou similar.

Em 2016, foi publicada a Lei 13.303/16, também conhecida como “Lei das Estatais”. Ela estabelece uma série de normas de transparência e governança e define como é o procedimento de realização de licitações e contratos, substituindo os procedimentos especificados na Lei 8.666/93. O pregão passou a ser o método preferencial para a aquisição de bens e serviços comuns.

Meirelles, Azevedo e Prendes (1991, p. 19) afirmam que “Licitação é o procedimento administrativo mediante o qual a Administração Pública seleciona a proposta mais vantajosa para o contrato de seu interesse”. O principal objetivo desse processo é garantir o princípio constitucional da isonomia, selecionando a proposta mais vantajosa para administração,

assegurando oportunidades a todos os interessados em participar desse processo.

Mesmo com a legislação procurando garantir a isonomia, é possível encontrar evidências na mídia ([FONSÊCA, 2006](#)) acerca de fraudes e corrupções no âmbito das licitações públicas. Portanto, nem mesmo o processo licitatório pode ser considerado 100% confiável.

## 2.2 Dados Governamentais Abertos

As tecnologias de informação e comunicação (TICs) estão evoluindo constantemente. Com esse desenvolvimento, veio junto a modernização do setor governamental, que possibilita o uso das tecnologias para promover a transparência da administração pública. A possibilidade de uso dessas ferramentas é apresentada constantemente na literatura, a exemplo de [Pinho \(2008\)](#), que investigou portais de governos estaduais no Brasil, de [Welp \(2007\)](#), que identificou as iniciativas de uso das TICs pelos governos federais da América Latina, e de [Prado e Loureiro \(2008\)](#), que analisaram o papel do governo eletrônico na transparência dos atos públicos.

Dados Governamentais Abertos (DGA) é um conceito que se refere à disponibilização, reutilização e manipulação de dados públicos. Essa disponibilidade só é possível graças à evolução das TICs, principalmente a Internet, que vieram a permitir e facilitar tanto a disponibilização quanto o acesso a essas informações, por meio das quais os governos buscam maior aproximação com a sociedade. Tem-se como premissa, no caso dos DGA, que os dados brutos referentes a processos públicos fiquem disponíveis para que interessados tenham livre acesso, respeitando as exigências legais, para manipular e, com isso, gerar novos produtos e/ou serviços. Para [Agune, Filho e Bolliger \(2010\)](#), o conceito de governo aberto remete à disponibilização, por meio da Internet, de informações e dados governamentais de domínio público para a livre utilização pela sociedade. Vários autores ([AGUNE; FILHO; BOLLIGER, 2010](#); [DEMCHAK; FRIIS; PORTE, 2000](#); [DINIZ, 2010](#)) defendem que os dados públicos são pertencentes ao cidadão e que esse deveria ter acesso irrestrito às informações governamentais.

Em 2007, o OpenGovData, grupo de trabalho vinculado ao Open Knowledge Foundation, desenvolveu os oito princípios dos dados governamentais abertos ([OPEN, 2010](#)). Esses princípios estabelecem como os DGA devem ser:

- **Completo.** Todos os dados públicos estão disponíveis. Dado público é o dado que não está sujeito a limitações válidas de privacidade, segurança ou controle de acesso.
- **Primários.** Os dados são apresentados tais como os coletados na fonte, com o maior nível possível de granularidade e sem agregação ou modificação.
- **Atuais.** Os dados são disponibilizados tão rapidamente quanto necessário à preservação do seu valor.

- Acessíveis. Os dados são disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades.
- Compreensíveis por máquinas. Os dados são razoavelmente estruturados de modo a possibilitar processamento automatizado.
- Não discriminatórios. Os dados são disponíveis para todos, sem exigência de requerimento ou cadastro.
- Não proprietários. Os dados são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo.
- Livres de licenças. Os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. Restrições sensatas relacionadas a privacidade, segurança e privilégios de acesso são permitidas.

Mais adiante, [Eaves \(2009\)](#) apresentou as três leis dos dados governamentais abertos:

- Se o dado não for encontrado e indexado na web, ele não existe;
- Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser aproveitado;
- Se algum dispositivo legal não permitir sua replicação, ele é inútil.

Segundo [Vaz, Ribeiro e Matheus \(2010\)](#), os benefícios da adoção dos DGA no campo da transparência e do controle social são evidentes. A oferta de DGA tende a contribuir para o aumento da transparência do governo, criando melhores possibilidades de controle social das ações governamentais. No entanto, dada a relativa novidade do tema, ainda não se dispõe de pesquisas que demonstrem a totalidade desta possibilidade. Porém ainda há a possibilidade de criação de novas informações e aplicativos a partir dos dados governamentais abertos. Nesse caso, não somente a transparência é fomentada, mas também novos serviços podem se originar da interação entre o governo e sociedade por meio da exploração dos DGA. Esses serviços podem ser gerados a partir de novas formas de atuação participativa e colaborativa, uma vez que o conceito de DGA permite superar a visão do cidadão como simples receptor da informação pública, pois ele ainda pode, ao processar livremente os dados governamentais, criar conteúdo a partir da reutilização dos dados.

O conceito de DGA também encontra fundamento político, de acordo com [Bobbio et al. \(1987\)](#). Para ele, a república democrática exige que o poder seja visível. É essencial à democracia o exercício dos vários direitos de liberdade, que permitem a formação da opinião pública e asseguram que as ações dos governantes não sejam ocultas. Como a maioria dos Estados adotam o princípio da publicidade, grande parte do mundo vive em um contexto em que o acesso aos dados públicos é uma exigência.

[Silva \(2010\)](#) e [Agune, Filho e Bolliger \(2010\)](#) ressaltam a importância dos dados governamentais abertos na criação de governos mais transparentes. O Brasil tem feito cons-

tantes esforços para promover a transparência dos seus dados em âmbito federal e estadual. Segundo esses autores, os portais de transparência possuem inúmeras bases que contêm dados e informações de caráter público (não sigilosos) e cujo acesso, quando disponível, normalmente é feito de forma parcial e por caminhos extremamente burocratizados e pouco transparentes.

O modelo atual no Brasil, que não adota em sua maioria a política de dados abertos, “impede e dificulta que o interessado possa trabalhar, analisar, cruzar e integrar os dados e informações segundo foco e interesses próprios” (AGUNE; FILHO; BOLLIGER, 2010, p. 8). Assim, os dados abertos, por suas características de acesso irrestrito aos documentos, dados e informações no âmbito da administração pública, também garantiriam maior transparência aos dados públicos.

Diversos países já disponibilizam seus dados a partir das orientações de DGA. Entre os países que contam com políticas de dados abertos pode-se citar: Estados Unidos, Reino Unido, Austrália e Nova Zelândia (AGUNE; FILHO; BOLLIGER, 2010). Esses países criaram políticas nacionais de acesso aos dados públicos na web e criaram portais que disponibilizam as informações públicas em diferentes formatos. Além disso, muitas vezes eles incentivam o cidadão a criar novos aplicativos e modos de utilizar a informação publicada em seus portais.

De acordo com vários autores (AGUNE; FILHO; BOLLIGER, 2010; DINIZ, 2010; SILVA, 2010), no Brasil, até 2010, não existia qualquer experiência de portal de transparência, nacional ou estadual, que disponibilizasse os seus dados integralmente, adotando os formatos abertos. Contudo, hoje já existem experiências isoladas e parciais de disponibilização de dados públicos por alguns órgãos governamentais pontuais e legislações que incentivam a adoção dos dados abertos pelos governos. Como exemplos de abertura de dados, pode-se citar o Tribunal de Contas dos Municípios do Ceará, o Portal da Transparência do Governo Federal e o Portal do Governo Aberto de São Paulo.

## 2.3 Descoberta de Conhecimento em Banco de Dados

Constantes avanços na área de tecnologia da informação têm possibilitado o armazenamento de grandes bases de dados. A análise de grandes quantidades de dados pelo homem é inviável sem o auxílio de ferramentas computacionais adequadas, tornando-se necessário o desenvolvimento de ferramentas que auxiliam de forma automática e inteligente, na tarefa de analisar, interpretar e relacionar esses dados a fim de desenvolver e selecionar estratégias de ação em cada contexto da aplicação. Diante dessa necessidade, surge uma área denominada descoberta de conhecimentos em bases de dados, do inglês Knowledge Discovery in Databases (KDD) (FAYYAD; PIATETSKY-SHAPIO; SMYTH, 1996).

Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), o processo de KDD é iterativo

e iterativo, envolvendo numerosos passos com muitas decisões tomadas pelo usuário. Os autores enumeram alguns passos básicos do processo de descoberta de conhecimento:

- Desenvolvimento da compreensão do domínio da aplicação e do conhecimento relevante prévio, além de identificar o objeto do processo de KDD do ponto de vista do cliente;
- Seleção de uma coleção de dados na qual as técnicas de descoberta de conhecimento serão aplicadas;
- Limpeza e pré-processamento dos dados: operações básicas que incluem remoção de ruídos, coleta das informações necessárias que modelam e representam ruídos, decisão de estratégias para tratar campos faltantes nos dados e representação da cronologia da informação e das alterações conhecidas;
- Redução e projeção dos dados: encontrar funcionalidade úteis para representar os dados dependendo do objetivo da tarefa. Com redução da dimensionalidade ou com métodos de transformação, o número de variáveis efetivas consideradas pode ser reduzido ou representações imutáveis para os dados podem ser encontradas;
- Associação de objetivos do processo de KDD definidos no primeiro passo com um método específico de mineração de dados, como por exemplo sumarização, classificação, regressão, agrupamento, entre outros;
- Análise e modelo exploratórios e seleção de hipóteses: escolha do algoritmo de mineração de dados e seleção do método para ser usado na busca por padrões de dados. Esse processo inclui a decisão de quais modelos e parâmetros podem ser apropriados e a associação de um método específico de mineração de dados aos critérios gerais do processo de KDD;
- Mineração de dados: busca por padrões de interesse em uma forma de representação específica ou um conjunto de tais representações, o que inclui as regras ou árvores de classificação, regressão e agrupamento. O usuário pode auxiliar significativamente o método de mineração de dados ao garantir que as etapas anteriores foram executadas corretamente.
- Interpretação dos padrões encontrados, possivelmente retornando a qualquer passo anterior para uma próxima iteração. Esse passo também pode envolver a visualização dos padrões e modelos extraídos ou visualização dos próprios dados a partir dos modelos extraídos.
- Ação sobre o conhecimento descoberto: usar diretamente o conhecimento, incorporando-o em outro sistema para ações futuras, ou simplesmente documentá-lo e informar as partes interessadas no resultado. Esse processo inclui checar e resolver potenciais conflitos com conhecimentos ou crenças prévios.

A mineração de dados é um dos principais passos que resultam na descoberta de padrões previamente desconhecidos, mas ainda assim potencialmente úteis (CARVALHO, 2001). Tradicionalmente, a tomada de decisões é baseada na informação superficialmente

estudada, em lições aprendidas com decisões passadas e nas limitações de orçamento, porém técnicas de mineração de dados podem ser usadas para determinar padrões significativos e desenvolver soluções estratégicas (PANWAR, 2008).

De acordo com Tan et al. (2006), as tarefas de mineração de dados são geralmente divididas em duas categorias principais: tarefas de previsão e tarefas descritivas. As tarefas de previsão objetivam prever o valor de um determinado atributo baseados em valores de outros atributos. Já as tarefas descritivas objetivam derivar padrões (correlações, tendências, grupos, trajetórias e anomalias) que resumem os relacionamentos adjacentes de dados. Para os autores, existem quatro tarefas centrais de mineração de dados: modelagem de previsão, análise de associação, análise de agrupamentos e detecção de anomalias. A seguir são descritas cada uma destas tarefas:

- **Modelagem de previsão:** se refere à tarefa de construir um modelo para a variável alvo como uma função das variáveis explicativas. Nesse processo há dois tipos de tarefas: (1) classificação, usada para variáveis alvo discretas, e (2) regressão, que é usada para variáveis alvo contínuas. O objetivo de ambas as tarefas é aprender um modelo que minimize o erro entre os valores previsto e real da variável alvo. Um exemplo de uso é a identificação de clientes que responderão a uma campanha de vendas ou avaliar se um paciente possui uma determinada doença baseado nos resultados de exames médicos.
- **Análise de associação:** usada para descobrir padrões que descrevem características altamente associativas dentro dos dados. Os padrões descobertos são normalmente apresentados na forma de regras de implicação ou subconjuntos de dados. Um exemplo de uso inclui a descoberta de genes que possuem funcionalidade associada.
- **Análise de agrupamentos ou *clustering*:** procura encontrar grupos de observações intimamente relacionadas, de modo que as observações que pertençam a um mesmo grupo sejam mais semelhantes entre si do que com as que pertençam a outros grupos. Um exemplo de uso é a identificação de conjuntos de clientes com as mesmas afinidades.
- **Detecção de anomalias:** tarefa de descobrir grupos utilizando a similaridade dos valores de seus atributos cujas características sejam significativamente diferentes do resto dos dados. O objetivo de um algoritmo de detecção de anomalias é identificar anomalias reais e evitar falsos-positivos, que são situações normais identificadas como anomalias. Um exemplo de uso é a detecção de compras irregulares em cartões de crédito.

Os algoritmos de Regras de Associação descrevem padrões de relacionamento entre itens de um banco de dados, como definido por Gonçalves (2004). Uma de suas típicas aplicações é a análise de transações de compras. Este processo examina padrões de compras de consumidores para determinar produtos que costumam ser adquiridos em



conjunto. Da mesma forma que podemos identificar padrões em cestas de compras, também podemos identificar padrões constantes em processos licitatórios.

Existem duas medidas de interesse que devem ser levadas em consideração ao aplicar algoritmos de regras de associação: valor de suporte e valor de confiança. De acordo com [Tan et al. \(2006\)](#), o fator de suporte determina a frequência na qual uma regra é aplicável a um determinado conjunto de dados, enquanto o fator de confiança, em uma regra  $X \Rightarrow Y$ , determina a frequência na qual os itens em  $Y$  aparecem em transações que contenham  $X$ .

O modelo típico para mineração de regras de associação em bases de dados consiste em encontrar todas as regras que possuem suporte e confiança maiores ou iguais, respectivamente, a um suporte mínimo (SupMin) e uma confiança mínima (ConfMin), especificados pelo usuário. Por este motivo, o modelo costuma ser referenciado na literatura como Modelo Suporte/Confiança ([GONÇALVES, 2004](#)). Para que uma regra seja considerada forte, contendo informação interessante, é necessário que ela apresente bons valores de suporte e confiança. A decisão sobre quais regras devem ser mantidas e quais deverão ser descartadas durante o processo de mineração é baseada nos valores desses dois índices.

## 2.4 Dados de Compras Governamentais

O Portal de Compras de Minas Gerais disponibiliza uma cartilha ([SEPLAG, 2014](#)) para auxiliar os fornecedores a entenderem o mercado das compras públicas e as principais características das compras realizadas pelo Governo de Minas Gerais. Esta parte do texto traz algumas definições que foram baseadas nessa cartilha.

Quando o governo necessita adquirir um produto ou contratar um serviço, ele realiza um processo de compra. As compras podem ser realizadas por meio de licitação, dispensa ou inexigibilidade de licitação. A maior parte dos procedimentos de compras são processos licitatórios.

A licitação pode ser entendida como a maneira pela qual a Administração Pública realiza a escolha do fornecedor dos produtos e serviços de que ela necessita. É um procedimento formal, em que ficam registrados, de forma objetiva e clara, quais são as características do que está sendo comprado (objeto da licitação), além dos critérios e dos processos para a apresentação de propostas. Assim, a licitação é o procedimento administrativo no qual vários interessados, dentre eles: sociedades, associações, empresas individuais de responsabilidade limitada, microempreendedores individuais, entre outros, participam de uma competição para vender algum produto, material ou serviço ao governo. Aquele que apresentar as melhores condições de negociação com o menor preço vence, firma um contrato e torna-se fornecedor por um período de tempo predefinido em edital.



As licitações possuem regras, impostas por leis e regulamentos, com a finalidade de padronizar os procedimentos de compras de todos os governos e de proporcionar oportunidades para que o maior número de interessados possa participar das competições. Elas podem ocorrer de forma presencial ou eletrônica. As licitações presenciais ocorrerem nas seguintes modalidades: concorrência, tomada de preços, convite, concurso ou pregão presencial. As compras eletrônicas são aquelas em que o governo utiliza a Internet para realizar a disputa e a seleção dos interessados em fornecer os materiais ou serviços e podem ser realizadas por meio de pregão eletrônico ou por meio de Cotação Eletrônica de Preços (COTEP).

Pregão é a modalidade de licitação para aquisição de bens e serviços comuns, qualquer que seja o valor estimado da contratação, em que a disputa pelo fornecimento é feita por meio de propostas e lances em sessão pública, de forma presencial ou eletrônica, nas quais os lances deverão ser sucessivos e decrescentes. O pregão eletrônico é conduzido por um pregoeiro, auxiliado por uma equipe de apoio, em que os fornecedores encaminham as suas propostas por meio de um sistema de compras eletrônico e, posteriormente, ocorre uma competição na qual cada um deles envia seus lances de preços eletronicamente. Toda a comunicação com o governo é feita em um chat nesse sistema eletrônico.

Apesar de a grande maioria das compras dos governos serem realizadas por meio de licitações, a lei prevê situações excepcionais em que não é exigida a realização de um processo licitatório. Nesses casos, é realizada uma dispensa ou uma inexigibilidade de licitação. Para elas, a lei prevê a possibilidade de realização de procedimentos mais simples e rápidos, nos quais a Administração Pública contrata diretamente o fornecedor, sem realizar uma licitação.

Existe uma forma de contratação denominada Dispensa de Licitação por Limite de Valor (compras de pequeno valor). Nesses casos, o governo coleta alguns orçamentos de preços para os produtos ou serviços e contrata o fornecedor que ofereceu o menor preço. Como não há licitação, o procedimento recebe o nome de Compra Direta. Porém, no caso do governo de MG, na grande maioria das vezes, esse procedimento é realizado por meio da Internet, em que a coleta (disputa) de preços é feita com o envio de lances pelos fornecedores diretamente no Portal de Compras. Nesse caso, o procedimento de compra recebe o nome de COTEP.

A COTEP é um procedimento utilizado para compras de pequeno valor (até quinze mil reais). O Portal notifica com antecedência os fornecedores cadastrados para participarem enviando os seus lances eletronicamente, porém sem a necessidade do envio prévio de propostas como acontece no pregão. A sessão de lances da COTEP fica aberta por, no mínimo, oito horas no horário comercial, ou seja, em dias úteis de segunda à sexta-feira, das 08:00h às 18:00h. Com isso, o procedimento amplia a disputa, a transparência e a publicidade.

Entre as diferenças dos processos de compra COTEP e pregão eletrônico, pode-se citar que o pregão necessita de edital, enquanto a COTEP possui limite de valor da compra, sendo este no ano de 2017 correspondente a oito mil reais para materiais e serviços e quinze mil reais para serviços de engenharia. O pregão dura em média 20 dias úteis, enquanto a COTEP pode durar apenas um.

Os fornecedores são gerenciados pelo Cadastro Geral de Fornecedores (CAGEF), que é um cadastro com a identificação e o controle dos documentos de todos os fornecedores do governo de Minas Gerais. Esses documentos são relacionados à regularidade jurídica e fiscal das empresas, e também à sua qualificação econômico-financeira. O Cadastro permite que o fornecedor participe de compras eletrônicas com um representante credenciado no CAGEF. Durante o processo de cadastramento, o representante obtém uma senha para participar de licitações de compra.

O representante da empresa cadastrado no CAGEF passa a receber avisos de licitação informando que haverá um processo de compra relacionado ao produto que o fornecedor comercializa ou ao serviço por ele prestado. Esses avisos são enviados de acordo com as linhas de fornecimento elencadas pelos fornecedores no seu cadastro.

A linha de fornecimento é uma forma de classificação utilizada pelo Portal de Compras para agrupar e diferenciar várias categorias de materiais e de serviços, por seguimento de mercado. Escolhidas durante o processo de inscrição no CAGEF, elas podem ser alteradas a qualquer tempo de acordo com as respectivas atividades econômicas que a empresa pratique.

As Microempresas e as Empresas de Pequeno Porte, referenciadas como MPE, possuem tratamento diferenciado no processo de compras governamentais. Pelos dados de 2017, nas compras públicas do Governo de Minas Gerais, as MPE têm benefícios exclusivos em relação aos demais fornecedores:

- Empate ficto: em caso de empate nas licitações, a preferência pela contratação é de uma pequena empresa;
- Participação de licitações com pendências fiscais: a pequena empresa deve comprovar a regularidade fiscal apenas para efeitos de assinatura do contrato;
- Compras exclusivas: todas as compras de até R\$ 80.000,00 contam com a participação exclusiva de MPE;
- Reserva de cota: quando o objeto licitado é divisível, o gestor da compra poderá reservar cota de até 25% para disputa exclusiva de MPE;
- Subcontratação: o governo pode exigir dos grandes licitantes a subcontratação de pequenas empresas para auxiliar na consecução de até 30% do objeto contratual.

### 3 Trabalhos Relacionados

A abertura dos dados governamentais é um processo em ascensão, porém ainda pouco desenvolvido. A publicidade dos dados está prevista em lei, porém a falta de costume e o pouco preparo humano para essa atividade podem criar um imaginário de que há falta de interesse em estruturar esses dados para que possam ser usados em estudos ou analisados pelo cidadão comum. São raríssimos os órgãos públicos que disponibilizam as bases de dados e informações de forma eletrônica e por meio de padrões abertos (AGUNE; FILHO; BOLLIGER, 2010). Muitas vezes, como é o caso do Portal de Compras do Estado de Minas Gerais, os dados de cada processo licitatório individual são disponibilizados na plataforma, porém não é publicado nenhum relatório em formato aberto com os dados estruturados de todos os processos licitatórios de um período.

Segundo Agune, Filho e Bolliger (2010), muitos dados e informações estão disponíveis nos portais e páginas que os órgãos e entidades mantêm na Internet, mas este acesso normalmente é parcial e fragmentado, o que impede e dificulta que o interessado possa trabalhar, analisar, cruzar e integrar os dados e informações segundo foco e interesses próprios. Como isso é um caso recorrente, consequentemente são poucos os trabalhos que realizaram estudos de caracterização de dados em bases governamentais.

Para descobrir informações que auxiliem ações estratégicas da polícia, baseadas no comportamento de criminosos e vítimas, Braz et al. (2009) utilizaram técnicas de mineração sobre o sistema SISGOP, um banco de dados da Polícia Militar do Estado de Alagoas que armazena os boletins das ocorrências de algumas cidades do estado. O software Weka<sup>1</sup> foi utilizado na mineração dos dados e nos experimentos foi aplicado o algoritmo SimpleKMeans para realizar o agrupamento dos dados. Apesar da base de dados incompleta, esses autores conseguiram extrair informações relevantes que ressaltam o potencial da análise de dados dos boletins de ocorrência, como por exemplo “mulheres são vítimas de roubo e ameaça no Jacintinho entre as 14hs e 20hs no dia de domingo” (p. 1486).

Reis (2014) utilizou um banco de dados fornecido pela Polícia Rodoviária Federal para identificar fatores de contribuição de acidentes rodoviários, no período de 2008 a 2012, em segmentos críticos da rodovia BR-381. Também utilizou a ferramenta Weka, porém aplicou o algoritmo Apriori para a geração de regras de associação. Assim, foram apresentadas as associações e a probabilidade de cada uma, obtendo padrões como “Das 117 ocorrências que possuíam a causa do acidente = não guardar distância de segurança, com a fase do dia = pleno dia, com a condição meteorológica = céu claro, com o traçado da via = reta, na cidade de Sabará e com veículos do tipo automóvel tiveram como

<sup>1</sup> Waikato Environment for Knowledge Analysis (Weka) é um software de aprendizado de máquina escrito em Java, desenvolvido na Universidade de Waikato, Nova Zelândia.

consequência o tipo de acidente igual a colisão traseira” (p. 69).

Em relação às compras governamentais, [Silva e Ralha \(2011\)](#) extraíram dados do sistema ComprasNet, o qual realiza pregões eletrônicos do Governo Federal. Inicialmente, com auxílio do Weka, foram gerados dois *datasets* a partir de execuções do algoritmo Apriori, sendo que o primeiro gerou 294 regras com a confiança mínima de 70% e o segundo gerou 145 regras com a confiança mínima de 80%. Depois foi utilizado o algoritmo Expectation-Maximization (EM), uma extensão do algoritmo K-Means, para mapear os grupos comuns de atuação dos fornecedores. Com esse resultado, foi novamente aplicado o algoritmo Apriori para detectar grupos suspeitos de praticar cartéis em regiões específicas. As melhores regras foram apresentadas a especialistas para verificação e três delas tiveram fortes indícios de cartelização, com destaque a uma empresa que no ano de 2008 ganhou nove licitações concorrendo com outra empresa que não ganhou nenhuma, sendo que esta outra empresa só participou dessas mesmas nove licitações nesse período, o que, segundo os autores, indicou simulação de concorrência.

Na esfera estadual, [Grilo Júnior \(2010\)](#) utilizou os dados do sistema SAGRES, que é o portal de dados abertos do Tribunal de Contas do Estado da Paraíba, com o objetivo de encontrar possíveis indícios de fraudes em processos licitatórios. Diferentemente dos trabalhos anteriores, o software utilizado foi o RapidMiner. A partir dele, utilizou-se o algoritmo FP-Growth para minerar as regras de associação. Foram necessários ajustes nos valores SupMin e ConfMin de um ano para o outro, devido ao aumento do interesse dos fornecedores nas compras governamentais, e com isso foram geradas 491 regras de associação. Analisando-as, foi encontrado um fornecedor que apresentava correlações com outros fornecedores desde 2005, o que é um indício de fraude.

Este presente trabalho se diferencia dos demais uma vez que realiza um estudo de caso utilizando dados do Portal de Compras do Estado de Minas Gerais, fornecidos pela Secretaria de Estado de Planejamento e Gestão (SEPLAG), com o objetivo de caracterizar os dados de compras governamentais. Os dados do Portal de Compras não são disponibilizados publicamente de maneira estruturada. Para que este trabalho pudesse acontecer, os dados estruturados foram fornecidos pela SEPLAG. Portanto, pode-se inferir que este tipo de trabalho nunca havia sido realizado.

## 4 Metodologia

Neste capítulo são apresentados os procedimentos realizados para encontrar indícios de ação de cartéis, do tipo rodízio de licitações, nas compras realizadas pelo estado de Minas Gerais via pregões eletrônicos no ano de 2015. Para tal, o trabalho foi dividido em sete partes:

- Caracterização dos dados de compras governamentais do Portal de Compras de Minas Gerais;
- Extração dos dados para análise;
- Pré-processamento dos dados;
- Avaliação de técnicas de mineração de dados e desenvolvimento de um novo algoritmo;
- Geração de regras de associação;
- Classificação dos grupos de fornecedores;
- Avaliação realizada por especialistas.

A extração dos dados visa analisar como o banco de dados do Portal está estruturado e, a partir dele, extrair os dados relevantes para o estudo. Para contextualizar e entender com que tipos de dados este trabalho lida, foi feita uma caracterização dessas informações, explicando como são organizados os dados de fornecedores, itens e compras.

Antes de fazer a análise desses dados, foi necessário um pré-processamento dos dados para aumentar sua qualidade e colocá-los no formato de entrada dos algoritmo de mineração de dados, sendo gerada nesse processo uma tabela de fornecedores e lotes de pregão com milhares de linhas e colunas, representando, respectivamente, os lotes e os fornecedores.

Com os dados estruturados, foi realizada uma avaliação das técnicas de mineração mais utilizadas e, tendo sido constatado que nenhuma delas cumpriria a necessidade específica deste estudo com eficiência, um novo algoritmo baseado no algoritmo Apriori foi desenvolvido. A saída desse algoritmo foi analisada e se mostrou viável para este estudo. Ao aplicar o algoritmo, foi gerada uma lista de regras de associação, a qual foi transformada em uma tabela de grupos de fornecedores, apresentando as métricas de participação e vitória.

Dois especialistas foram convidados a analisar os resultados. Um desses especialistas atua na Secretaria de Estado de Planejamento e Gestão de Minas Gerais - SEPLAG, e o outro atua junto ao setor de compras da Fundação de Amparo a Pesquisa do Estado de Minas Gerais - FAPEMIG. Apesar de não serem responsáveis diretos pelas compras realizadas, ambos acompanham os processos de licitação e, portanto, têm pleno conhecimento

desse processo. Com o primeiro, a entrevista foi realizada presencialmente e os resultados foram apresentados no momento da entrevista. Com o segundo, a entrevista aconteceu por meio de videoconferência e os dados foram enviados previamente, possibilitando uma análise mais consistente.

Para ambos foi explicada a técnica de mineração de dados utilizada e a atenção se deu sobre a [Tabela 1](#), que identifica os grupos de fornecedores e suas respectivas métricas. Além de apresentar as métricas, as questões semi-estruturadas envolveram a utilidade desse tipo de resultado, a sugestão de novas métricas e, enfim, a viabilidade de uso dessa técnica para detecção de fraudes.

## 5 Caracterização dos Dados de Compras Governamentais

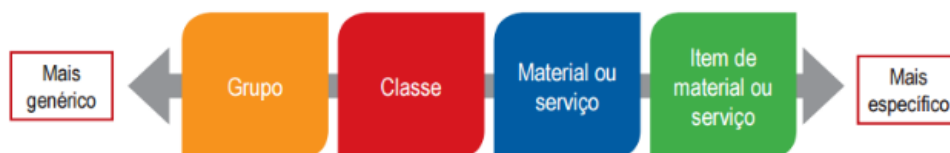
O Portal de Compras do Estado de Minas Gerais está em operação desde 2009, quando substituiu o sistema LicitaNet, o que significa que todos os dados foram inseridos no período de 2009 a 2018, apesar de alguns dados mais antigos do LicitaNet terem sido importados para o novo sistema. Desde então, apesar de inúmeras alterações pequenas, o processo de compras não mudou drasticamente. Com isso infere-se que nenhum dado desde 2009 é ultrapassado ou irrelevante.

O Portal de Compras possui diversos módulos interligados, porém, para facilitar a compreensão dos dados, eles foram divididos em três categorias principais: fornecedores, itens e compras. Nenhuma dessas categorias é independente, pois em um processo de compra são negociados itens providos por fornecedores. Com isso, torna-se importante ressaltar que será feita uma descrição mais geral de todo esse processo.

### 5.1 Itens de Compra

Os objetos de uma compra governamental podem ser materiais, serviços, obras ou uma mescla de todos esses tipos. No entanto, no caso do governo mineiro, um produto é adquirido ou um serviço é contratado somente se o mesmo estiver previamente registrado no Catálogo de Materiais e Serviços (CATMAS)<sup>1</sup>. No catálogo existem cerca de 75.000 itens ativos passíveis de serem comprados pelos diversos órgãos e entidades estaduais. Eles são classificados em grupos, classes e materiais/serviços de acordo com suas características. Essa relação está representada na Figura 1.

Figura 1 – Categorização dos itens de compras no CATMAS



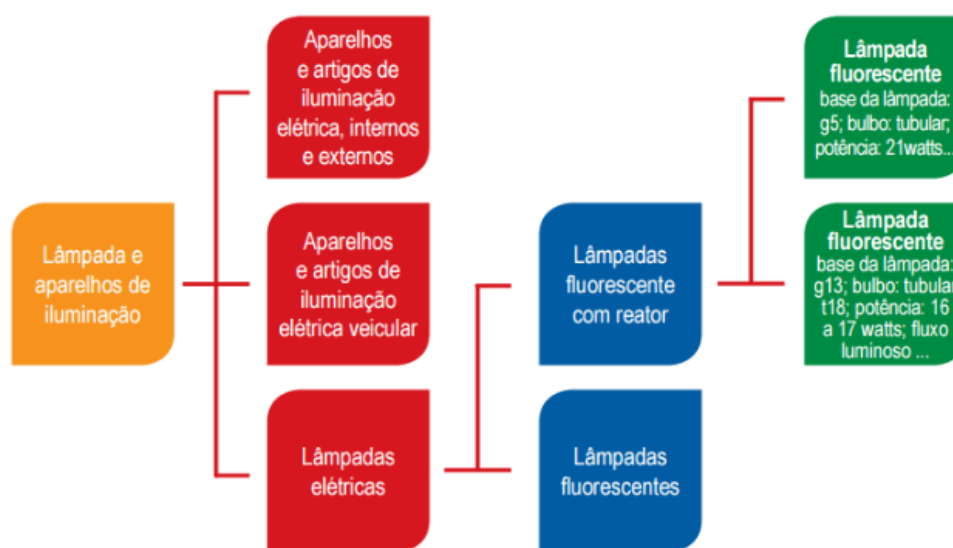
Fonte: SEPLAG (2014)

Como exemplo, tem-se o grupo Lâmpadas e aparelhos de iluminação, composto de quatro classes: (1) aparelhos e artigos de iluminação elétrica, internos e externos, (2)

<sup>1</sup> O CATMAS pode ser acessado no link <https://www1.compras.mg.gov.br/catalogo/consultaMateriaisCadastrados.html>

aparelhos e artigos de iluminação elétrica veicular, (3) aparelhos e artigos de iluminação elétrica, portáteis e manuais e (4) lâmpadas elétricas. Cada classe possui diversos materiais. Por exemplo, a classe lâmpadas elétricas possui, além de outros, os materiais lâmpadas fluorescentes e lâmpadas fluorescentes com reator. Por sua vez, o material lâmpadas fluorescentes possui, entre outros, o item de material Lâmpada fluorescente - base da lâmpada: g5; bulbo: tubular; potência: 21 watts; fluxo luminoso min: 1900 lúmens; eficiência luminosa min: 90 lm/w; vida útil mínima: 19000 h; temperatura cor min: 3100 k; irc mínimo: 85%; norma aplicável: inmetro. Esse exemplo é ilustrado na [Figura 2](#).

Figura 2 – Exemplo de classificação de um item de compra



Fonte: [SEPLAG \(2014\)](#)

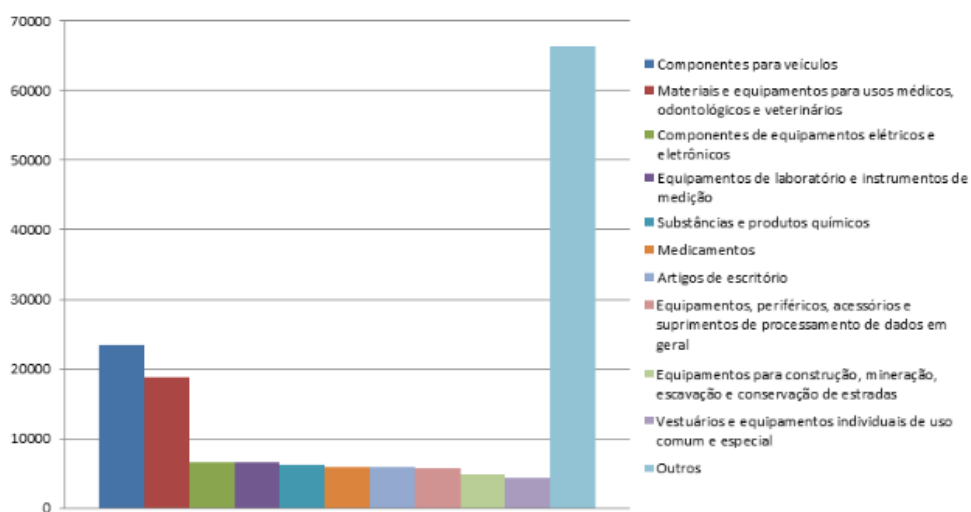
Essa divisão entre tantas categorias se faz necessária pois a quantidade de itens de material ou serviço no catálogo, incluindo itens que não estão mais disponíveis para compra, é de 154851. A categoria mais externa, que é o grupo de material ou serviço, possui 86 registros, resultando em uma média aproximada de 1801 itens por grupo. A quantidade de itens nos grupos mais populosos é mostrada na [Figura 3](#).

## 5.2 Fornecedores

No total, são 52240 fornecedores encontrados no banco de dados do Portal de Compras. As ligações entre os fornecedores e os itens de materiais ou serviços são os grupos de fornecimento. Um fornecedor pode trabalhar simultaneamente com itens de diferentes grupos. Por exemplo, uma farmácia de manipulação pode vender ao governo itens pertencentes ao grupo de medicamentos e também itens pertencentes ao grupo de substâncias e produtos químicos. Há 64 registros de grupos de fornecimento no Portal e



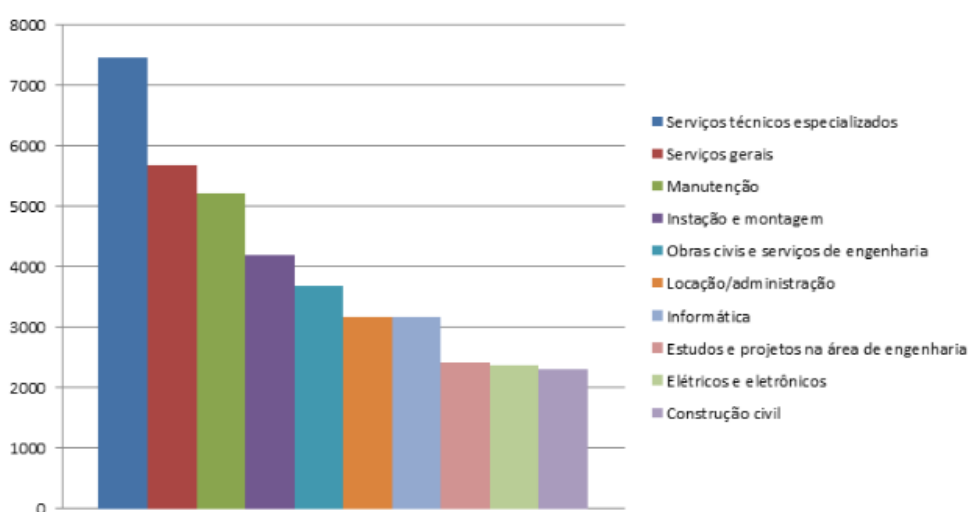
Figura 3 – Quantidade de itens de material ou serviço por grupo



Fonte: O autor

uma média de 1283 fornecedores por grupo. A quantidade de fornecedores para os 10 maiores grupo de fornecimento é mostrada na Figura 4.

Figura 4 – Quantidade de fornecedores por grupo de fornecimento



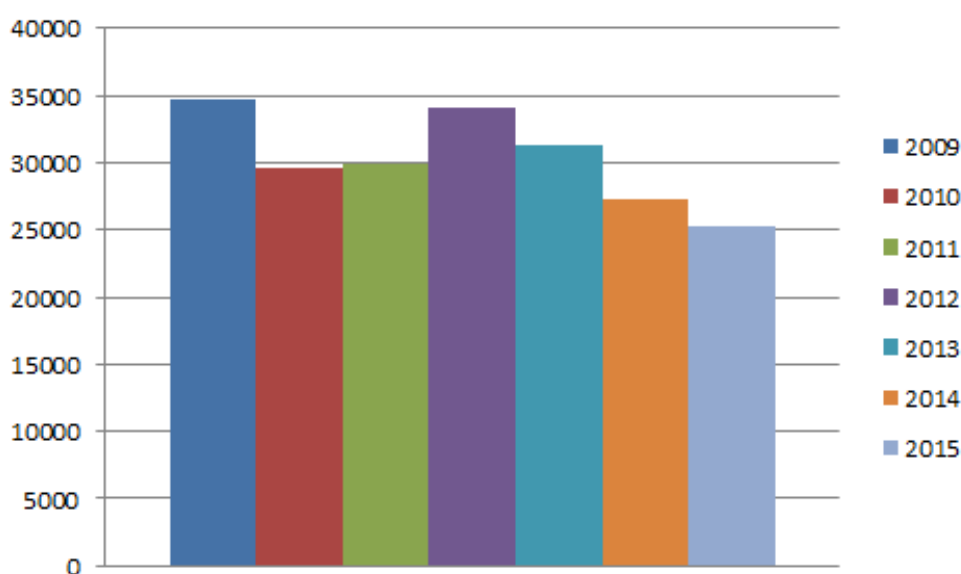
Fonte: O autor

Por meio dos dados presentes na Figura 4, é possível perceber a enorme quantidade de fornecedores registrados no Portal de Compras. Esse dado é importante quando se trata de acompanhar possíveis organizações desses fornecedores, pois se trata de encontrar, algumas vezes, uma informação pequena dentro de um grande universo de dados.

### 5.3 Compras

Todas as compras governamentais passam pelo sistema do Portal de Compras do Estado de Minas Gerais. Desde as mais simples, como a compra periódica de papel higiênico para um órgão público, até as mais complexas, como a construção de presídios, estão registradas no banco de dados. Devido a isso, a quantidade de processos de compra realizados por ano é relativamente alta. A [Figura 5](#) apresenta a quantidade de processos de compra concluídos de cada ano, desde 2009 a 2015. Vale ressaltar que esses dados, mesmo sendo de anos anteriores, ainda são dinâmicos, visto que processos iniciados em um determinado ano podem demorar vários anos para serem concluídos.

Figura 5 – Quantidade de processos de compra concluídos por ano



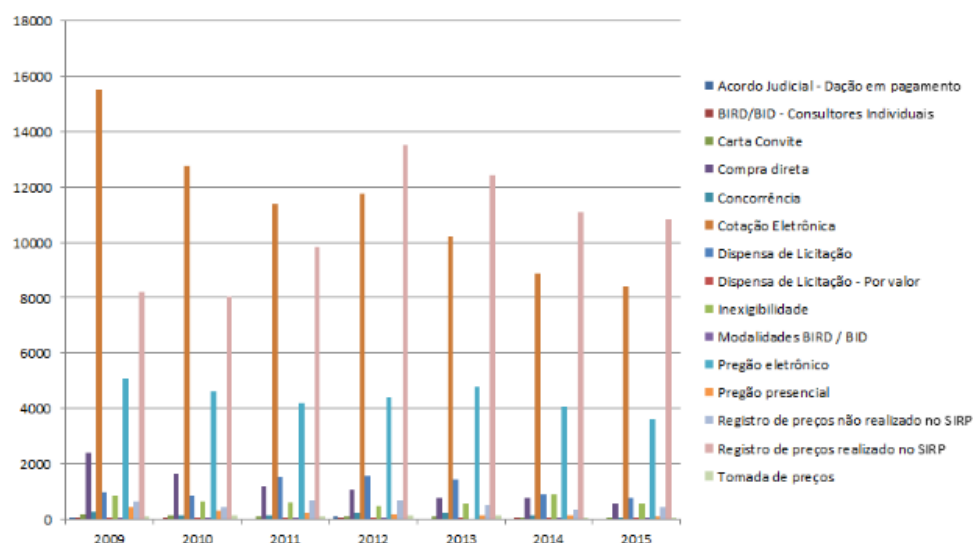
Fonte: O autor

Os processos de compra são divididos em 15 categorias de procedimento de contratação, sendo algumas delas licitatórias e outras não. Entre as licitatórias, estão: concorrência, tomada de preços, carta convite, pregão eletrônico e pregão presencial. Entre as não licitatórias, estão: compra direta, cotação eletrônica, dispensa de licitação, inexigibilidade de licitação e registro de preços. A [Figura 6](#) ilustra a quantidade de compras de cada categoria por ano.

Segundo [SEPLAG \(2014\)](#), em 2013, 88% dos procedimentos de contratação foram compras eletrônicas e 95% de toda a variedade de itens de materiais e serviços que o governo comprou foi contratada via procedimentos eletrônicos. Essa informação pode ser confirmada pelo gráfico anterior, pois os procedimentos Registro de preços no SIRP, Cotação eletrônica e Pregão eletrônico foram definitivamente os mais utilizados durante todos os anos.

Para o entendimento das próximas caracterizações dos dados, é importante destacar

Figura 6 – Quantidade de tipos de processos de compras concluídos por ano



Fonte: O autor

que, tanto na COTEP quanto no Pregão, os itens a serem comprados/contratados são agrupados em lotes para facilitar a competição entre os licitantes. Num mesmo processo eletrônico, pode haver vários lotes em disputa. O fornecedor decide de qual(is) deseja participar.

A partir de dados de pregões eletrônicos realizados em 2015, extraídos do banco de dados fornecido pelo SEPLAG, constatou-se que:

- Foram realizados 3663 pregões eletrônicos em 2015;
- Os pregões foram divididos, em média, em 3,82 lotes;
- Em cada lote participaram, em média, 2,31 fornecedores.

Considerando esses números, é possível perceber cerca de 32 mil participações no ano de 2015. Com isso, fazer um acompanhamento manual de fornecedores se mostra uma tarefa inviável. Nesse sentido, as técnicas de mineração de dados podem auxiliar na descoberta de padrões.

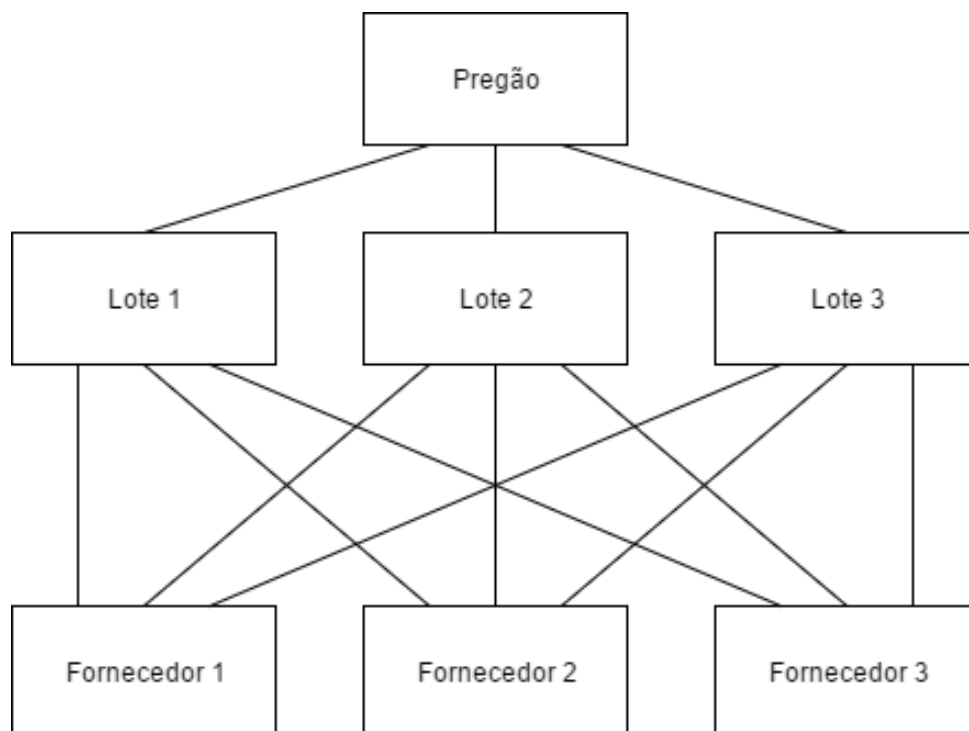
## 6 Mineração dos Dados: Técnica e Resultado

Esta parte do trabalho foi iniciada pela extração dos dados presentes no Portal de Compras e, após, foi feito um pré-processamento desses dados, com a intenção de tornar esses dados legíveis para programas automatizados. Foi feita então a avaliação de técnicas disponíveis de mineração e desenvolvido um novo algoritmo. Com isso, foram geradas regras de associação entre fornecedores que permitiram a classificação desses fornecedores em grupos. Cada uma dessas etapas está descrita a seguir.

### 6.1 Extração dos Dados

Os dados do Portal de Compras de Minas Gerais estão armazenados em um banco de dados relacional. Ou seja, o banco de dados é dividido em tabelas, sendo que cada tabela representa um tipo de registro e cada linha da tabela representa o registro em si. O objetivo dessa parte do trabalho era extrair os relacionamentos entre fornecedores e lotes de pregão.

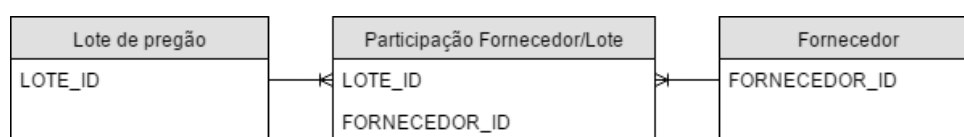
Figura 7 – Exemplo de relacionamento entre pregão, lotes do pregão e fornecedores



Fonte: O autor

Como pode ser visto na [Figura 7](#), cada fornecedor pode participar de diversos lotes de pregão, assim como cada lote de pregão pode conter vários fornecedores participantes. A esse tipo de relacionamento é dado o nome de muitos para muitos (M para N). Em um banco de dados relacional, a forma ideal de representar esse relacionamento é uma tabela que possui as chaves primárias envolvidas. Nesse caso, cada registro dessa tabela possui uma chave estrangeira representando o fornecedor envolvido no relacionamento e uma chave estrangeira representando o lote, como foi ilustrado na [Figura 8](#). Um exemplo de uma base de dados utilizando essa tabela pode ser visto na [Figura 9](#).

Figura 8 – Modelo físico do relacionamento entre lotes do pregão e fornecedores



Fonte: O autor

A partir dessa tabela é possível extrair toda a informação necessária para encontrar padrões utilizando algoritmos de mineração de dados. Porém, a informação ainda não está em formato de entrada dos algoritmos. Desse modo, foi necessário o pré-processamento dos dados para só depois realmente ser realizada uma análise.

Figura 9 – Exemplo de dados inseridos na tabela de relacionamento entre lotes de pregões e fornecedores

	LOTE_ID	FORNECEDOR_ID
1	L1	F1
2	L1	F2
3	L1	F3
4	L2	F2
5	L2	F3
6	L2	F4
7	L3	F3
8	L4	F1
9	L4	F4
10	L4	F5

Fonte: O autor

## 6.2 Pré-processamento dos Dados

Uma das maneiras encontradas para transformar os dados para um formato de entrada dos algoritmos de mineração de dados foi representar o relacionamento entre lotes de pregão e fornecedores com uma tabela de participações, na qual cada linha da tabela

representa um lote de pregão, cada coluna da tabela representa um fornecedor e o valor de cada célula indica se o respectivo fornecedor participou daquele lote.

Para realizar a conversão, inicialmente foi necessário extrair todos os lotes e todos os fornecedores da tabela do banco de dados relacional. Ou seja, caso a tabela possuísse as colunas (FORNECEDOR\_ID, LOTE\_ID) e os valores (L1, F1), (L1, F2), (L1, F3) e (L2, F2), ela possuiria os fornecedores F1, F2 e F3 e os lote L1 e L2, totalizando 6 células na nova tabela, conforme representado na [Figura 10](#).

Figura 10 – Tabela de relacionamento gerada para servir como entrada de softwares de mineração

	F1	F2	F3
L1	1	1	1
L2	0	1	0

Fonte: O autor

A partir dessa tabela, torna-se relativamente fácil converter os dados para qualquer software minerador de dados. O Weka, por exemplo, aceita como entrada arquivos ARFF, no qual cada célula da tabela é separada por vírgula e os fornecedores são representados como meta-dados. Tal exemplo pode ser visto na [Figura 11](#).

Figura 11 – Exemplo de arquivo ARFF de entrada do Weka

```

1 @relation Participacoes
2
3 @attribute F1 {0,1}
4 @attribute F2 {0,1}
5 @attribute F3 {0,1}
6
7 @data
8 1,1,1
9 0,1,0

```

Fonte: O autor

O pré-processamento teve como base uma tabela de relacionamentos que possui uma chave estrangeira para a tabela de fornecedores e uma chave estrangeira para a tabela de lotes de pregão. Com tais dados, foi desenvolvido um aplicativo para gerar uma nova tabela, na qual cada coluna representa um fornecedor, cada linha representa um lote de pregão e cada célula possui um valor indicando se houve ou não participação do respectivo fornecedor no respectivo lote.

Foi constatado que em 2015 foram realizadas 7.686 sessões de lotes de pregão, com um total de 3.464 fornecedores participantes. Assim, foi gerada uma tabela com 7.686 linhas e 3.464 colunas, totalizando 26.624.304 células. Essa quantidade de dados torna completamente inviável uma análise sem as ferramentas adequadas. Uma pequena amostra desses dados pode ser vista na [Figura 12](#).



dessa maneira, decidiu-se implementar um. A nova técnica é uma modificação do algoritmo Apriori para considerar somente um registro à esquerda da regra de associação e realizar uma busca baseada no suporte mínimo por grupo de fornecedores à direita da regra de associação. Tal algoritmo pode ser visto na [Figura 13](#).

Figura 13 – Algoritmo para gerar regras de associação com apenas um elemento antecedente

```

suporteMinimo <- 0.01
confiancaMinima <- 0.7

início
  regrasGeradas <- {}
  para cada fornecedor:
    antecedente <- fornecedor
    consequentes <- {}
    chamar processarSubRegra(antecedente, consequentes)
  fim para
fim

procedimento processarSubRegra(antecedente, consequentes):
  regra <- nova Regra(antecedente, consequentes)
  se regra.suporte >= suporteMinimo e regra.confianca >= confiancaMinima então:
    regrasGeradas <- concatenar regrasGeradas e regra
    para cada fornecedor: if (fornecedor != antecedente)
      novosConsequentes <- concatenar consequentes e fornecedor
      chamar processarSubRegra(antecedente, novosConsequentes)
    fim para
  fim se
fim procedimento

```

Fonte: O autor

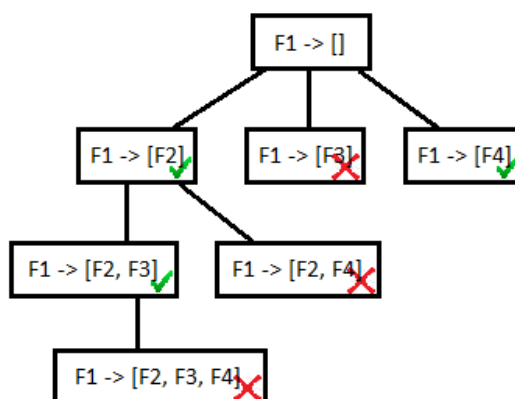
Baseado em recursão e poda, o algoritmo inicia considerando os seguintes candidatos a regra de associação: F1 -> [], F2 -> [], F3 -> [], etc. No primeiro momento, são criados todos os candidatos nos quais a participação de um fornecedor implica na participação de um único outro fornecedor, como por exemplo os seguintes candidatos: F1 -> [F2], F1 -> [F3], F1 -> [F4], etc. Então esses candidatos são avaliados de acordo com a confiança e suporte. Se, nesse exemplo, apenas F1 -> [F3] obteve suporte e confiança mínimos, ele já é considerado uma regra de associação. O próximo passo é gerar os candidatos que possuem F1 como antecedente e F3 como consequente, por exemplo F1 -> [F3, F4], F1 -> [F3, F5], etc. Nota-se que não é necessário gerar o candidato F1 -> [F3, F2], pois F1 -> [F2] já foi analisado anteriormente, podendo ter sido desclassificado ou virado regra, que, nesse caso, avaliaria o candidato F1 -> [F2, F3] em outro momento.

Os novos candidatos são avaliados e, se classificados, são considerados regras de associação, além de terem seus sucessores avaliados. Se o candidato F1 -> [F3, F5] for classificado, são gerados os candidatos sucessores F1 -> [F3, F5, F6], F1 -> [F3, F5, F7], F1 -> [F3, F5, F8] e assim por diante, recursivamente. Por fim, todas as regras geradas são



apresentadas ao usuário, assim como seus valores de suporte e confiança. Uma execução desse algoritmo para um grupo de quatro fornecedores está representada na [Figura 14](#).

Figura 14 – Execução do novo algoritmo, considerando F1 como antecedente, para um grupo de 4 fornecedores



Fonte: O autor

## 6.4 Regras de Associação Geradas

Para gerar essas regras de associação foi desenvolvido um algoritmo explicado com detalhes na [Seção 6.3](#). Esse algoritmo gera os elementos consequentes da regra de associação, ou seja, os grupos de fornecedores, a partir de podas baseadas no suporte e na confiança mínimos. Com base no trabalho de [Silva e Ralha \(2011\)](#), foram escolhidos os valores de suporte mínimo 1% e confiança mínima 70%. Esse valor de suporte indica que os fornecedores participaram juntos em pelo menos 76 lotes de pregão.

As regras de associação geradas podem ser vistas parcialmente na [Figura 15](#). Foram encontradas 429 regras de associação, sendo que muitas dessas regras são subgrupos de regras com uma confiança igual ou maior. Por exemplo, a regra de associação 249274 -> 34139, 82191 (confiança = 85%) é um subgrupo da regra 249274 -> 34139 (confiança = 88%).

Pode-se perceber que, mesmo com vários estreitamentos, ainda é uma grande quantidade de dados para analisar. Daí surgiu a necessidade de classificar essas regras de associação para direcionar melhor uma possível investigação manual.

## 6.5 Classificação dos Grupos de Fornecedores

Com as regras de associação geradas, era necessário classificar as regras por probabilidade de fraude. Para isso, inicialmente fez-se a conversão de regra de associação em grupo de fornecedores. Isso significa que uma regra de associação F1 -> (F2, F3) passou

Figura 15 – Regras de associação geradas

```

438 -> (383) - Suporte: 0.05 - Confiança: 0.73
4912 -> (1005) - Suporte: 0.07 - Confiança: 0.74
4912 -> (34139) - Suporte: 0.07 - Confiança: 0.75
4912 -> (34139, 82191) - Suporte: 0.07 - Confiança: 0.70
4912 -> (82191) - Suporte: 0.07 - Confiança: 0.74
1005 -> (4912) - Suporte: 0.05 - Confiança: 0.94
1005 -> (4912, 34139) - Suporte: 0.05 - Confiança: 0.87
1005 -> (4912, 34139, 82191) - Suporte: 0.05 - Confiança: 0.82
1005 -> (4912, 82191) - Suporte: 0.05 - Confiança: 0.86
1005 -> (4912, 82191, 5895) - Suporte: 0.05 - Confiança: 0.70
1005 -> (4912, 5895) - Suporte: 0.05 - Confiança: 0.74
1005 -> (34139) - Suporte: 0.05 - Confiança: 0.87
1005 -> (34139, 82191) - Suporte: 0.05 - Confiança: 0.82
1005 -> (82191) - Suporte: 0.05 - Confiança: 0.86
1005 -> (82191, 5895) - Suporte: 0.05 - Confiança: 0.70
1005 -> (5895) - Suporte: 0.05 - Confiança: 0.78
34139 -> (4912) - Suporte: 0.05 - Confiança: 1.00
34139 -> (4912, 1005) - Suporte: 0.05 - Confiança: 0.92
34139 -> (4912, 1005, 82191) - Suporte: 0.05 - Confiança: 0.86
34139 -> (4912, 1005, 5895) - Suporte: 0.05 - Confiança: 0.72

```

Fonte: O autor

a ser tratada como o grupo de fornecedores (F1, F2, F3). Cada grupo de fornecedores foi avaliado com duas métricas, denominadas métrica de participações e métrica de vitórias.

A métrica de participações é a média da proporção do número de participações em grupo e do número de participações individuais de cada fornecedor. Supondo o grupo acima, se F1 participou de um total de 200 lotes de pregões e o grupo participou de um total de 150 lotes, a sua métrica de participação individual é 75%. Se F1 foi avaliado com 75%, F2 com 55% e F3 com 50%, a métrica de participações do grupo é a média aritmética, que no caso é 60%.

A métrica de vitórias é a proporção de vitórias de um dos fornecedores do grupo quando o grupo inteiro participou e o total de participações do grupo. Isso é, se o grupo participou de um total de 150 lotes, mas em apenas 50 lotes um dos fornecedores do grupo saiu vitorioso, então a métrica de vitórias do grupo é 33%.

A partir desses dois valores, foi feita uma média aritmética das duas métricas, o que foi chamada de métrica global. Os grupos de fornecedores gerados foram ordenados decrescentemente pela sua avaliação a partir da métrica global e foi considerado que, quanto maior a posição do grupo de fornecedores nessa lista, maior a probabilidade de fraude.

Os grupos de fornecedores, assim como as avaliações dos grupos, podem ser vistas na [Tabela 1](#). Se o grupo de fornecedores está mais acima na tabela, maior o indício de ter ocorrido uma fraude do tipo rodízio de licitações, e, quanto mais abaixo, menor o indício. Apesar desse resultado não garantir fraude alguma, o que provavelmente nenhuma análise

Tabela 1 – Grupos de fornecedores e suas respectivas métricas

	Grupo de fornecedores	Métrica de participações	Métrica de vitórias	Métrica global
1	473, 880	89,53%	76,62%	83,08%
2	1005, 4912, 34139	69,74%	87,60%	78,67%
3	4461, 247994	89,07%	41,07%	65,07%
4	34139, 82191	81,46%	40,92%	61,19%
5	4912, 82191	65,34%	52,75%	59,05%
6	1005, 4912	64,77%	52,27%	58,52%
7	1005, 34139, 82191	74,72%	41,10%	57,91%
8	4912, 5895	60,00%	54,17%	57,09%
9	82191, 249274	59,81%	52,03%	55,92%
10	1005, 34139	80,57%	22,74%	51,66%
11	4912, 34139, 82191	73,97%	21,77%	47,87%
12	4912, 34139	70,84%	22,73%	46,79%
13	1005, 249274	60,32%	32,70%	46,51%
14	5895, 249274	58,78%	32,05%	45,42%
15	380, 273592	70,13%	16,01%	43,07%
16	1005, 82191	66,51%	15,97%	41,24%
17	34139, 249274	62,21%	18,79%	40,50%
18	1005, 5895	61,48%	18,68%	40,08%
19	1005, 4912, 82191	62,00%	16,92%	39,46%
20	5895, 82191	60,00%	11,37%	35,69%

Fonte: O autor

automática fará, a ordenação por métrica global é uma boa forma de ordenar a prioridade de investigações manuais por fraudes.

Nesse caso específico, se uma investigação fosse ocorrer, ela deveria iniciar pelos fornecedores 473, 880, 1005, 4912 e 34139, já que eles apresentaram uma métrica global de 83,08% ou 78,67%. Isso significa que esses fornecedores participaram frequentemente nas mesmas licitações e, dentre essas licitações, venceram um bom número. Esses dados apresentam indícios de que os valores podem ter sido combinados entre eles, favorecendo ora um ora outro nas licitações.

Para contrapor os dados dos dois primeiros grupos da [Tabela 1](#), foi utilizado como exemplo o grupo com os fornecedores 1005 e 34139, listados na linha 10 da figura. De todas as licitações em que eles participaram, 80,57% foram em conjunto, porém os dados mostram que um ou outro foi vencedor de apenas 22,74% delas. Com isso, parece lógico que não houve formação de cartel nesse caso ou, se houve, não há necessidade de investigação. Assim como [Silva e Ralha \(2011\)](#) e [Grilo Júnior \(2010\)](#), que identificaram cartéis em licitações públicas, por meio da mineração de dados da Controladoria Geral da União e do Tribunal de Contas do Estado da Paraíba, os dados encontrados nesse trabalho também poderiam ser usados para investigar possíveis formações de cartéis.

Portanto os dados obtidos por meio do processo de mineração se mostram úteis para

indicar por onde uma investigação deveria ser iniciada caso fosse identificada essa necessidade. Eles não indicam a presença de fraudes ou a identificação de cartéis, mas podem ser usados tanto para novas minerações mais específicas quanto para o direcionamento de um acompanhamento manual de um grupo menor de fornecedores.

## 7 Avaliação dos Especialistas

A partir dos resultados alcançados, foi realizada uma entrevista com dois especialistas ligados ao setor de compras do governo do estado para avaliar a utilidade dos padrões encontrados. Um deles atua na SEPLAG e o outro na FAPEMIG. Para eles foi explicado como as regras de associação foram geradas e como os grupos de fornecedores foram avaliados de acordo com as métricas utilizadas. Depois dessa explicação, os especialistas teceram comentários tanto sobre os dados quanto sobre o processo licitatório.

### 7.1 Entrevista com o Especialista 1

A partir da entrevista com o primeiro especialista (SEPLAG), foram produzidas algumas notas, descritas e comentadas a seguir:

- a. O objetivo de um cartel em licitações públicas é sempre aumentar o preço do produto ou serviço oferecido ao estado. Por isso, uma boa forma de analisar o indício de cartéis é verificar o preço do produto quando o grupo de fornecedores participa da sessão de pregão e comparar com o preço do mesmo produto quando esse grupo de fornecedores não participa.
- b. A métrica de participações é muito boa para encontrar indícios de rodízios de licitações, porém a métrica de vitórias pode não ser relevante, visto que é muito comum fornecedores externos ao grupo também participarem de uma sessão de pregão, inviabilizando a fraude naquele momento. Inclusive a existência de uma empresa que sempre envia propostas mas nunca sai vitoriosa da licitação também pode ser um indício de fraude.
- c. Outro fato a ser observado é quando empresas enviam lances com valores inferiores aos do mercado e posteriormente desistem da venda, dando a vitória ao segundo menor lance. Essa é uma forma de causar a desistência de um licitante externo ao grupo de fornecedores participantes do rodízio.
- d. Os pregões eletrônicos são os procedimentos de contratação de licitações públicas com menor probabilidade de irregularidades, mas obviamente não estão imunes. Os resultados obtidos pelo trabalho têm potencial para inibir grande parte dessas fraudes.

É possível perceber, por meio da análise feita pelo especialista, a complexidade envolvida em um processo licitatório, como é o caso das compras realizadas pelo governo do Estado de Minas Gerais. Ao mesmo tempo em que o especialista enfatiza a importância dos procedimentos usados nas licitações para diminuir as irregularidades (item d), também reconhece a fragilidade do sistema.

Para ele, qualquer empresa ou pessoa física que participa de uma licitação oferecendo seus produtos com um custo/benefício vantajoso, pode inviabilizar a formação de cartéis nas compras do governo, ao vencer a licitação. Nos dados encontrados por meio deste trabalho, foram identificadas empresas que participam em conjunto de licitações e que vencem muitas delas. É claro que, se houver licitações ganhas por meio da formação de cartéis, é muito provável que tentativas em número muitas vezes maior que as bem sucedidas aconteçam e que, por um ou outro motivo, seus participantes não tenham sido vencedores.

O especialista alerta para outros tipos de fraude nos processos licitatórios de compra do governo do estado de Minas Gerais e que devem ser comuns a qualquer outro órgão da administração pública que adquira produtos por meio de pregão ou licitação. O que pareceu já ser do conhecimento do especialista, são empresas que vencem uma licitação, oferecendo produtos com valor menor que o de mercado. Elas desistem da venda após vencer a licitação, dando o direito de entrega dos produtos ao segundo classificado, que provavelmente terá lucros consideráveis com a venda. Segundo o especialista, ao oferecer produtos abaixo do valor de mercado, esse concorrente acaba provocando a desistência de outras empresas, provavelmente reforçando a presença no processo apenas das empresas que combinaram ou negociaram essa estratégia. Nesse caso, outro tipo de técnica de mineração de dados deveria ser realizado.

Mais especificamente sobre a métrica usada neste trabalho, o especialista elogiou o potencial dos dados de participação das empresas nas diversas licitações. Porém, ele alerta que uma empresa que sempre participa das licitações sem nunca vencer também deveria ser identificada, pois ela pode estar sendo usada para maquiar dados e dificultar a identificação de fraudes.

No item “a”, o especialista apontou para uma forma de analisar a presença de cartéis, por meio do acompanhamento do preço ofertado pelas empresas, comparando-os com o preço de mercado. No entanto, não apontou quem deveria fazê-lo ou quando deveria ser feito.

## 7.2 Entrevista com o Especialista 2

Esse especialista (FAPEMIG) afirmou não ser responsável pela execução de compras, mas está diretamente relacionado a elas por fazer solicitações de compras e receber produtos resultantes dessas licitações. Portanto, ele acompanha os processos. Foram sintetizados, a seguir, os comentários feitos por ele.

- a. Há uma preocupação com a formação de cartéis já que eles visam aumentar o preço do produto ou serviço oferecido ao estado. Se o preço é condizente com o mercado e o produto tiver qualidade, não interessa qualquer tipo de investigação. Assim sendo,

verificar o preço do produto vencedor da licitação seria uma boa estratégia. Porém o número de compras é razoavelmente grande, o que dificulta esse acompanhamento manual.

- b. Sobre os dados presentes na [Tabela 1](#), esse especialista afirma serem relevantes. Diferente do primeiro, este afirma que a métrica de vitórias é extremamente interessante. Ele reforça que o estado não tem condição de acompanhar 100% dos fornecedores e a métrica pode ser usada para dirigir o olhar. Ressalta, ainda, que nos casos em que os grupos de fornecedores tiveram poucas vitórias, isto está dentro da lógica da livre concorrência.
- c. Para uma investigação sobre preço, por exemplo, esse especialista afirma ser necessário delimitar fornecedores. Nesse sentido, a métrica usada nesse trabalho pode auxiliar nessa delimitação. Ele afirma ainda que tudo que está fora da “média” merece uma atenção. No caso de um grupo pequeno de empresas participar de muitas licitações, pode significar que o mecanismo de atração do estado não esteja sendo eficiente.
- d. Como alternativa de continuidade de análise, esse especialista aponta caminhos possíveis: verificar a quantidade de empresas que participaram das licitações em que o primeiro grupo (473, 880) se saiu vitorioso; verificar, no caso dos grupos presentes nas linhas 1 e 2 da [Tabela 1](#), o índice de vitórias de cada um dos fornecedores individualmente.
- e. Ao ser questionado se esse tipo de análise já é feita, esse especialista citou a Nota Fiscal Eletrônica (NF-e). Segundo ele, com a NF-e, é possível, por exemplo, comparar os preços dos produtos fornecidos por uma empresa nas vendas para o setor público e para o setor privado. No entanto, ele desconhece a eficácia dessa estratégia e se ela é usada atualmente. Portanto há um interesse na inserção de TICs no monitoramento de fornecedores, sendo que, em Minas Gerais, esse processo é muito inicial.
- f. Os dados presentes na [Tabela 1](#) mostram que os fornecedores se conhecem e podem se organizar. Segundo ele, isso exige a organização do outro lado, ou seja, do estado.

Pode-se perceber que esse segundo especialista apontou que a técnica de mineração de dados utilizada para obter padrões nos dados do Portal de Compras do governo do estado pode representar uma maneira eficaz do estado identificar caminhos de investigação. No caso de grupos de empresas vencedores de um grande número de licitações, como foi identificado na linha 1 da [Tabela 1](#), esse especialista sugere o acompanhamento dessas empresas nos processos licitatórios futuros. Nesse sentido, os dados se mostram extremamente úteis.

Os caminhos apontados no item “d” não foram realizados neste trabalho em função de não ser esse o objetivo. No entanto, são etapas posteriores ao trabalho feito que poderiam ser realizadas com certa facilidade. Isso indica que, se o estado fizer uso das TICs com mais propriedade, será possível acompanhar os processos licitatórios com mais eficiência.

Mesmo com a NF-e, que representa um bom instrumento para monitorar as empresas participantes de licitações, é absolutamente necessário fazer uso das TICs e das técnicas de mineração de dados para facilitar a obtenção de padrões que mostram caminhos para investigações. Portanto, o uso das TICs e das técnicas de mineração de dados se mostram extremamente úteis na detecção de indícios de fraudes em compras públicas.



## 8 Considerações Finais

Este trabalho foi proposto com o objetivo de realizar uma caracterização dos dados de compras e licitações do governo de Minas Gerais e, com isso, prover insumos para tomadas de decisões e demonstrar, a partir do estudo de caso, as técnicas aplicáveis para extração desses insumos. Tais técnicas se mostraram úteis para priorizar a investigação de fraudes pelos órgãos que realizam auditoria dos dados públicos.

Tratam-se, como alerta [Carvalho \(2001\)](#) de descobrir padrões previamente desconhecidos, mas que podem se apresentar como potencialmente úteis ou, como informa [Panwar \(2008\)](#), a mineração de dados pode determinar padrões significativos que auxiliam no desenvolvimento de soluções estratégicas. Considerando que os dados disponíveis no portal investigado não estão totalmente disponíveis ou estão em um formato que não facilita o acesso a informação, algumas adaptações precisaram ser feitas. Considerando, ainda, a quantidade extraordinariamente grande de dados encontrados em uma primeira mineração no portal de compras do governo do estado de Minas Gerais, fez-se necessário diminuir a amostra usando, para isso, dados parciais.

Mesmo sendo um trabalho inicial de mineração de dados, a [Tabela 1](#) sintetiza os grupos de empresas que participam em conjunto de inúmeros processos licitatórios e a porcentagem de vezes em que uma dessas empresas sai vencedora e, portanto, é a fornecedora do estado. Todos os dados constantes nessa tabela síntese não podem ser usados para identificar formação de cartéis, mas fornecem bons indícios de, no caso de uma investigação, por quais empresas ela poderia começar. Os principais resultados deste trabalho para a área foram, portanto, a criação de um novo algoritmo para gerar regras de associação com um único elemento antecedente e a aplicação de métricas de avaliação de resultados.

Como já foi descrito, alguns trabalhos de mineração de dados se mostraram importantes para auxiliar a polícia a definir estratégias ([BRAZ et al., 2009](#)) e para identificar fatores que contribuem para acidentes ([REIS, 2014](#)). Outros mais relacionados ao trabalho ora realizado identificou uma empresa que em um ano venceu nove licitações ao concorrer com outra que não foi vencedora em nenhuma das concorrências ([SILVA; RALHA, 2011](#)) e um fornecedor com altos indícios de fraude ([GRILO JÚNIOR, 2010](#)). Neste trabalho, por meio dos dados obtidos, cinco empresas surgiram com forte concorrência, por terem vencido várias licitações quando concorreram em conjunto. Duas delas mostraram índices razoavelmente satisfatório em termos de sucesso nas licitações. É possível que não haja formação de cartéis nesses casos, mas é razoável que uma investigação mais específica envolvendo essas empresas e as que concorreram com elas nos mesmos processos fosse

realizada futuramente.

Como apontado pelos especialistas, novas técnicas de mineração de dados poderiam ser aplicadas envolvendo essas empresas e outras que concorreram com elas, por exemplo. Ao citar a NF-e, o segundo especialista também mostrou possibilidades relevantes para que novas técnicas sejam usadas. O objetivo de um cartel é sempre vender mais caro, então é possível que fazer uma análise dos preços em conjunto com a busca por grupos de fornecedores recorrentes seja a melhor maneira de encontrar possíveis fraudes.

No caso da NF-e, é possível comparar preços de um mesmo produto fornecido ao estado e a empresas privadas. No entanto, as notas fiscais formam uma base de dados enorme, inviabilizando qualquer tipo de busca manual. A mineração de dados se mostra um campo rico para se trabalhar com esses dados. Portanto, a principal contribuição deste estudo está na comprovação do potencial das técnicas de mineração de dados.

Este trabalho apresenta dados iniciais sobre a mineração de dados do Portal de Compras do Estado de Minas Gerais. Apesar de não apresentar dados contundentes, este trabalho aponta para novas técnicas e fornece bons indícios de onde uma investigação poderia ser iniciada e em que caminhos essa investigação poderia prosseguir, no caso de haver interesse em fazê-lo.

# Referências

AGUNE, R. M.; FILHO, A. S. G.; BOLLIGER, S. P. Governos abertos: disponibilização de bases de dados e informações em formato aberto. In: **III Congresso CONSAD de Gestão Pública**. [S.l.: s.n.], 2010. Citado 4 vezes nas páginas 4, 5, 6 e 12.

ATTARD, J. et al. A systematic review of open government data initiatives. **Government Information Quarterly**, Elsevier, v. 32, n. 4, p. 399–418, 2015. Citado na página 1.

BOBBIO, N. et al. **Estado, governo, sociedade: para uma teoria geral da política**. [S.l.]: Paz e Terra Rio de Janeiro, 1987. v. 7. Citado na página 5.

BRAZ, L. M. et al. Aplicando mineração de dados para apoiar a tomada de decisão na segurança pública do estado de Alagoas. In: **XXIX Congresso da Sociedade Brasileira de Computação**. [S.l.: s.n.], 2009. Citado 3 vezes nas páginas 1, 12 e 34.

CARVALHO, L. A. V. de. **Datamining: a mineração de dados no marketing, medicina, economia, engenharia e administração**. [S.l.]: Érica, 2001. Citado 2 vezes nas páginas 7 e 34.

DEMCHAK, C. C.; FRIIS, C.; PORTE, T. M. L. Webbing governance: National differences in constructing the face of public organisations. **GARSON, G. David. Handbook of public information system**. Nova York: Marcel Dekker, 2000. Citado na página 4.

DINIZ, V. Como conseguir dados governamentais abertos. In: **III Congresso CONSAD de Gestão Pública**. [S.l.: s.n.], 2010. Citado 2 vezes nas páginas 4 e 6.

EAVES, D. **The three laws of open government data**. 2009. Disponível em: <<http://eaves.ca/2009/09/30/three-law-of-open-government-data/>>. Citado 2 vezes nas páginas 1 e 5.

ECONOMIST, T. A special report on managing information: Data, data everywhere. **The Economist**, 2010. Citado na página 1.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996. Citado na página 6.

FONSÊCA, M. A. R. Pregão eletrônico: uma análise de sua evolução histórico-legislativa e das inovações decorrentes do decreto nº 5.450/2005. **Jus Navigandi**, 2006. Citado na página 4.

GONÇALVES, E. C. Regras de associações e suas medidas de interesse objetivas e subjetivas. **INFOCOMP Journal of Computer Science**, v. 4, n. 1, p. 26–35, 2004. Citado 2 vezes nas páginas 8 e 9.

GRILO JÚNIOR, T. F. **Aplicação de técnicas de Data Mining para auxiliar no processo de fiscalização no âmbito do Tribunal de Contas do Estado da Paraíba**. 2010. 103 f. Dissertação (Mestrado em Engenharia de Produção), 2010. Citado 4 vezes nas páginas 1, 13, 28 e 34.

- MANYIKA, J. et al. **Big data: The next frontier for innovation, competition, and productivity**. [S.l.], 2011. Disponível em: <<https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>>. Acesso em: 20 de junho de 2018. Citado na página 1.
- MEIRELLES, H. L.; AZEVEDO, E. de A.; PRENDES, C. M. **Licitação e contrato administrativo**. [S.l.]: Ed. Revista dos Tribunais, 1991. Citado na página 3.
- OPEN, G. D. **Eight principles of open government data**. [S.l.], 2010. Disponível em: <[https://public.resource.org/8\\_principles.html](https://public.resource.org/8_principles.html)>. Acesso em: 20 de junho de 2018. Citado na página 4.
- PANWAR, V. Role of data warehousing & data mining in e-governance. **Review of Professional Management - A Journal of New Delhi Institute of Management**, v. 6, p. 41, 2008. Citado 2 vezes nas páginas 8 e 34.
- PINHO, J. d. Investigando portais de governo eletrônico de estados no brasil: muita tecnologia, pouca democracia. **Revista de Administração Pública**, SciELO Brasil, v. 42, n. 3, p. 471–493, 2008. Citado na página 4.
- PRADO, O.; LOUREIRO, M. R. G. Governo eletrônico e transparência: avaliação da publicização das contas públicas das capitais brasileiras. **Revista Alcance**, v. 13, n. 3, p. 355–372, 2008. Citado na página 4.
- REIS, C. V. R. O uso da descoberta de conhecimento em banco de dados nos acidentes da br-381. **Projetos e Dissertações em Sistemas de Informação e Gestão do Conhecimento**, v. 3, n. 1, 2014. Citado 3 vezes nas páginas 1, 12 e 34.
- SEPLAG. **Como vender para o governo de Minas Gerais**. 2014. Disponível em: <[http://www.compras.mg.gov.br/images/stories/Fornecedores/2014\\_demais\\_arquivos/dclc-cartilha-compras-governamentais-2014-com-logo.pdf](http://www.compras.mg.gov.br/images/stories/Fornecedores/2014_demais_arquivos/dclc-cartilha-compras-governamentais-2014-com-logo.pdf)>. Citado 4 vezes nas páginas 9, 16, 17 e 19.
- SILVA, C. V. S.; RALHA, C. G. Detecção de cartéis em licitações públicas com agentes de mineração de dados. **Revista Eletrônica de Sistemas de Informação ISSN 1677-3071 doi: 10.5329/RESI**, v. 10, n. 1, 2011. Citado 5 vezes nas páginas 1, 13, 26, 28 e 34.
- SILVA, D. B. da. **Transparência na esfera pública interconectada**. Tese (Doutorado) — Dissertação de mestrado, Faculdade Cásper Líbero, 2010. Citado 2 vezes nas páginas 5 e 6.
- TAN, P.-N. et al. **Introduction to data mining**. [S.l.]: Pearson Education India, 2006. Citado 2 vezes nas páginas 8 e 9.
- VAZ, J. C.; RIBEIRO, M. M.; MATHEUS, R. Dados governamentais abertos e seus impactos sobre os conceitos e práticas de transparência no brasil. **Cadernos ppg-au/ufba**, v. 9, n. 1, 2010. Citado na página 5.
- WELP, Y. Democracia y fractura digital en américa latina. In: **Conferencia Internacional Democracia Directa en América Latina**. [S.l.: s.n.], 2007. p. 14–15. Citado na página 4.