

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA DE CIÊNCIAS EXATAS E DA COMPUTAÇÃO
CURSO DE CIÊNCIA DA COMPUTAÇÃO



MINERAÇÃO DE DADOS APLICADA A COMPORTAMENTO DE LICITAÇÕES

FELYPE SANTOS MORAES

GOIÂNIA

2021

FELYPE SANTOS MORAES

MINERAÇÃO DE DADOS APLICADA A COMPORTAMENTO DE LICITAÇÕES

Trabalho de Conclusão de Curso apresentado à Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Sibelius Lellis Vieira

Banca examinadora:

Prof. Me. Fernando Abadia

Prof. Me. Gustavo Siqueira Vinhal

GOIÂNIA

2021

FELYPE SANTOS MORAES

MINERAÇÃO DE DADOS APLICADA A COMPORTAMENTO DE LICITAÇÕES

Trabalho de Conclusão de Curso aprovado em sua forma parcial pela Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, para obtenção do título de Bacharel em Ciência da Computação, em ____/____/____.

Orientador: Prof. Dr. Sibélius Lellis Vieira

Prof. Me. Fernando Abadia

Prof. Me. Gustavo Siqueira Vinhal

GOIÂNIA

2021

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus por ter me dado forças para realizar este trabalho, à minha família, em especial aos meus pais Valdoveu e Rosangela e minha irmã Elisangela, que me apoiaram e incentivaram em meus estudos.

Agradeço ao meu amigo Murillo Mesquita pelos conselhos e incentivos.

Agradeço também aos colegas e professores do curso pelo convívio e solidariedade.

Agradeço ao professor Dr. Sibelius Lellis Vieira, pela oportunidade e encorajamento ao estudo científico, pelo apoio e incentivo que me foi dedicado, os meus mais sinceros e profundos votos de admiração e respeito. Por fim, gostaria de agradecer ainda os meus amigos pelo carinho e pela compreensão nos momentos em que a dedicação aos estudos foi única. A todos que colaboraram direta ou indiretamente para que este trabalho fosse realizado fica registrado os meus mais singelos agradecimentos.

RESUMO

Este estudo tem o propósito de aplicar técnicas de mineração de dados no âmbito da Controladoria Geral do Estado de Goiás com o objetivo de relacionar possíveis indícios de irregularidades em processos licitatórios, auxiliando assim em demonstrar sua aplicação em processos de fiscalização, a fim de proporcionar um ganho de produtividade no processo de fiscalização das contas públicas. Para melhor compreensão deste trabalho foi realizada uma revisão bibliográfica para delimitar os fatos e fenômenos relacionados a licitações públicas e o processo de descoberta de indícios de fraudes e irregularidades através do uso de mineração de dados. São abordados conceitos como a Descoberta de Conhecimentos em bases de Dados (*Knowledge Discovery in Databases*), *data science* e técnicas de mineração de dados. Como resultado foram encontradas regras que podem auxiliar no apoio à decisão e definição de estratégias para fiscalizar o processo licitatório, pois geram informações que permitem maior assertividade em indicar possíveis irregularidades, permitindo assim a redução dos gastos do recurso público.

Palavras-chave: licitações, cartelização, ciência de dados, descoberta de conhecimento, conluio.

ABSTRACT

This study aims to apply data mining techniques within the scope of the General Controllorship of the State of Goiás in order to relate possible indications of irregularities in bidding processes, thus helping to demonstrate its viability in inspection processes, in order to verify whether there is a productivity gain in the process of auditing public accounts. For a better understanding of this work, a bibliographic review was carried out in order to know the facts and phenomena related to public tenders and the process of discovering evidence of fraud and irregularities through the use of data mining. Concepts such as Knowledge Discovery in Databases (Knowledge Discovery in Databases), data science and data mining techniques are addressed. It is expected that with the items presented, it generates a potential for the development of decision support systems and the definition of strategies to supervise the bidding process, as they generate information that allows greater assertiveness in indicating possible irregularities, thus allowing the reduction of resource costs. public.

Keywords: bids, cartelization, data science, knowledge discovery, collusion.

LISTA DE FIGURAS

Figura 1. Processo <i>Knowledge Discovery in Databases</i> (KDD).....	21
Figura 2. Fluxograma de execução do <i>K-means</i>	25
Figura 3. Fluxograma da abordagem proposta.....	27
Figura 4. Valor adjudicado no ano de 2019.....	31
Figura 5. Licitações e empresas.....	31
Figura 6. Número de propostas e número de licitações por ano (2016-2019).....	32
Figura 7. Área de processos/operadores da ferramenta <i>RapidMiner</i>	33
Figura 8. Parâmetros do operador <i>log to session</i>	35
Figura 9. Operador FP-Growth e seus parâmetros.....	36
Figura 10. Operador Create Association Rules e seus parâmetros.....	36
Figura 11. Processo para realizar a clusterização.....	38
Figura 12. Processo para aplicar a Regra de associação em cada <i>cluster</i>	40

TABELAS

Tabela 1. Número de licitações, valor estimado e valor total por modalidade.....	30
Tabela 2. Licitações e participantes.....	34
Tabela 3. Matriz esparsas de licitações e participantes.....	35
Tabela 4. Resultado do processamento dos dados minerados.....	37
Tabela 5. Dataset para realizar a clusterização.....	38
Tabela 6. Saída do WriteCSV com os <i>clusters</i> gerados.....	39
Tabela 7. Execução do algoritmo de associação FP-Growth em cada <i>cluster</i>	40
Tabela 8. Tabela das regras de associações obtidas no <i>cluster 4</i>	41
Tabela 9. Tabela das regras de associações obtidas no <i>cluster 3</i>	42
Tabela 10. Tabela das regras de associações obtidas no <i>cluster 1</i>	43

LISTA DE SIGLAS

BI – *Business Intelligence*

CGE/GO – Controladoria Geral do Estado de Goiás

CSV – *Comma-Separated Values*

CGU – Controladoria de Contas da União

DCDB – Descoberta de Conhecimento em Banco de dados

FP-tree – *Frequent Pattern Tree*

KDD – *Knowledge Discovery in Databases*

OCDE – Organização para a Cooperação e Desenvolvimento Econômico

PIB– Produto Interno Bruto

RDC – Regime Diferenciado de Contratações Públicas

SRP – Pregão por Sistema de Registro de Preços

TCU – Tribunal de Contas da União

XML – *eXtensible Markup Language*

SUMÁRIO

1. INTRODUÇÃO	10
1.1 Contextualização	10
1.2 Justificativa	11
1.3 Objetivo	12
1.3.1 Objetivo geral	12
1.3.2 Objetivos específicos	12
1.4 Estrutura do trabalho	12
2. REFERENCIAL TEÓRICO	14
2.1 Licitações	14
2.1.1 Modalidades	15
2.1.2 Etapas de uma licitação	16
2.1.3 Fraudes em licitações	17
2.1.4 Formação de cartel e conluio	18
2.2 Descoberta de conhecimento em bases de dados	19
2.2.1 Fases do DCDB	20
2.2.2 Mineração de dados	21
2.2.3 Tarefas e técnicas de mineração de dados	22
2.2.4 Associação e FP-Growth	22
2.2.5 Agrupamento e K-means	24
2.3 Estudos correlatos	25
3. PROCEDIMENTO METODOLOGICO	27
4. RESULTADOS	30
4.1 Análise descritiva	30
4.2 Análise preditiva	33
5 CONCLUSÃO	42
5.1 Trabalhos futuros	42
REFERÊNCIAS	43
ANEXO I	48

1. INTRODUÇÃO

1.1 Contextualização

Para adquirir serviços ou produtos, os órgãos da administração pública em todos os níveis devem, via de regra, realizar uma licitação, processo administrativo pelo qual busca-se selecionar a proposta mais vantajosa e proporcionar igual oportunidade aos participantes. Este processo é regulado pela lei ordinária número 8.666/93 (BRASIL, 1993), que define, entre outras coisas, as seguintes modalidades de licitação: Concorrência, Tomada de Preços, Leilão, Concurso e Convite. Em 2002, a lei número 10.520/02 (BRASIL, 2002) instituiu o pregão no ordenamento jurídico brasileiro para aquisição de bens e serviços comuns em processos licitatórios.

A redução da corrupção em licitações requer fiscalização e controle das atividades realizadas por funcionários, órgãos governamentais e fornecedores. A observação dos padrões de comportamento dos diversos participantes da licitação pode identificar situações anormais que podem estar relacionadas a comportamentos ilegais. No Brasil, esse controle, quando externo ao órgão licitante, é exercido pelo Tribunal de Contas da União (TCU) e pela Controladoria de Contas da União (CGU), no nível federal. Essas agências têm autonomia para supervisionar o processo de compras governamentais, punir fraudadores e prevenir fraudes (ALVARENGA, 2011).

A Controladoria Geral do Estado de Goiás (CGE/GO) é o órgão central dos sistemas de controle externo, correição, transparência e ouvidoria, integrante da Administração Direta do Poder Executivo do Estado de Goiás, na forma da Lei nº 20.491/2019, de 25 de junho de 2019.

De acordo com a legislação, o CGE/GO disponibiliza um portal transparente, uma plataforma *online* onde as pessoas podem consultar as seguintes informações: despesas; receitas; convênios e repasses; gestão fiscal e governamental; pessoal; e planejamento e orçamento de todos os órgãos do estado. Portanto, o CGE/GO disponibiliza mais um mecanismo de fiscalização.

Neste trabalho será abordado como a área de ciência de dados, mais especificamente mineração de dados, pode lidar com grandes volumes de informação relacionadas à licitação. A utilização dessa técnica tem-se mostrado de grande valia na obtenção de informações potencialmente úteis. Este estudo pretende contribuir para o

campo de investigação realizar um cotejo em relação aos principais trabalhos de referência, a saber Moraes (2016), Silva e Ralha (2011) e Grilo Júnior (2010).

Mineração de dados é o processo de explorar grandes quantidades de dados para identificação de padrões que podem auxiliar na descoberta de conhecimento. Este conhecimento pode ser apresentado de diversas formas como: agrupamentos, regras de associação, hipóteses e grafos. Essa área se desenvolveu devido ao crescimento exorbitante do volume de dados disponível, o que inviabiliza a análise humana e fortalece o conceito de que o computador pode ser ágil e útil para detectar novos padrões e conceitos (ALVARENGA, 2011).

1.2 Justificativa

Como a licitação é a porta de entrada mais comum para realização da despesa pública, as atividades de auditoria governamental dão especial atenção à análise dos processos de licitação e contratos dela decorrentes. Esta preocupação se dá pelo fato de que o envolvimento de recursos financeiros possibilita e até mesmo incita a criação de esquemas ilícitos para manobrar a Lei, com finalidades diversas.

O caráter competitivo da licitação é um dos fundamentos mais importantes desse ato, pois possibilita à administração pública realizar suas compras a preços justos. A falta de competitividade num processo de licitação permite que uma empresa contrate com a administração a preços superfaturados, ou que privilegie um ou outro participante “conhecido”, mesmo que o preço seja justo.

Segundo Araújo e Martinez (2010), os dados da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) mostraram que os cartéis em licitações geram um sobrepreço mínimo de 20% comparado ao preço em um mercado competitivo, causando perdas anuais de centenas de bilhões de reais ao Erário e contribuintes, que deixam de se beneficiar de outros serviços e produtos que poderiam ser contratados com o pagamento de seus impostos que foram ilicitamente desviados. Países desenvolvidos gastam em média 15% do seu PIB em compras públicas, sendo que esse percentual é maior em países em desenvolvimento.

Com vários casos de corrupção em licitações no Brasil e no estado de Goiás, em que os esforços de garantir a melhor opção de negócio para a administração pública, com a consequente economia para os cofres públicos, esta pesquisa justifica-se por relacionar os possíveis indícios de fraude em processos licitatórios com a implantação

de técnicas de mineração de dados em órgãos da administração pública. Focando no processo analítico das operações eletronicamente armazenadas, o acúmulo de informação e o seu potencial para análise e tomada de decisão. Sendo assim, uma pergunta pode ser feita: a mineração de dados pode contribuir para a análise do processo de fiscalização em licitações?

1.3 Objetivo

1.3.1 Objetivo geral

Aplicar técnicas de mineração de dados no âmbito dos dados de competência da Controladoria Geral do Estado de Goiás com vistas a relacionar possíveis indícios de fraudes em processos licitatórios, utilizando mecanismos de associação e agrupamento entre participantes destes processos

1.3.2 Objetivos específicos

Para se atingir o objetivo geral, propõem-se os seguintes objetivos específicos:

- Desenvolver uma investigação através de uma pesquisa exploratória para conhecer os fatos e fenômenos relacionado a licitações públicas;
- Explorar os conceitos de ciência de dados e técnicas de mineração de dados;
- Determinar e analisar a aplicabilidade das técnicas de mineração de dados no processo de fiscalização de órgãos governamentais, especialmente para CGE/GO;
- Identificar entidades com base na repetitividade do grupo suspeito, levando em consideração tendências de cartelização, concentração em órgãos específicos e existência de simulações por meio de concorrentes fictícios, em geral via associação e agrupamento de participantes.

1.4 Estrutura do trabalho

Esse estudo apresenta-se estruturado em 5 (cinco) capítulos, sendo o capítulo 1 (um) referente a esta introdução. O capítulo 2 (dois) apresenta o referencial teórico,

abordando uma revisão bibliográfica sobre conceitos e procedimentos das licitações, mineração de dados, técnicas, algoritmos e trabalhos correlatos. O capítulo 3 (três) descreve o procedimento metodológico utilizado nesse estudo. O capítulo 4 (quatro) descreve e discute os resultados obtidos, análise descritiva e análise preditiva. O capítulo 5 (cinco) apresenta a conclusão e trabalhos futuros.

2. REFERENCIAL TEÓRICO

2.1 Licitações

Segundo o Tribunal de Contas da União (TCU), o processo licitatório é um procedimento administrativo formal em que a administração pública convoca, por meio de condições estabelecidas em ato próprio (edital ou convite), empresas interessadas na apresentação de propostas mais vantajosa para celebração de contrato (TCU, BRASIL, 2010). No Brasil, a licitação busca escolher qual proposta é mais vantajosa para o Estado, tanto em qualidade como em preço, no intuito de reduzir custos de aquisição de bens públicos e minimizar desperdícios de recursos. Segundo Amorim (2017), o procedimento licitatório tem por objetivos:

- a observância do princípio constitucional da isonomia: assegura aos administrados interessados a oportunidade de contratar com o Estado tendo por base as regras previamente estipuladas e aplicáveis, de forma indistinta, a todos os eventuais interessados;
- a seleção da proposta mais vantajosa para a Administração Pública: a competição que se estabelece entre os interessados que preenchem os atributos e requisitos necessários para contratar resulta na obtenção da melhor proposta para a Administração;
- a promoção do desenvolvimento nacional sustentável: inserido na redação do artigo 3º da Lei nº 8.999/1993 por força da Lei nº 12.349/2010.

Como se trata a licitação de ser um ato público, tem que ser totalmente transparente perante a sociedade e acessível a todos os potenciais interessados, favorecendo a ampla participação, o ingresso democrático e a concorrência de todos, de forma igualitária. O processo licitatório deve afastar qualquer suspeita de favorecimento e garantir que o dinheiro público seja utilizado com cautela e eficiência. A licitação é a forma mais clara de se atender aos princípios das atividades da Administração Pública (SOUZA, 1997).

A Lei Federal nº 8.666/1993 que rege as licitações e contratos da Administração Pública no Brasil apresenta, em seu artigo 3º, diz o seguinte:

Art. 3º - A licitação destina-se a garantir a observância do princípio constitucional da isonomia, a seleção da proposta mais vantajosa para a administração e a promoção do desenvolvimento nacional sustentável e será processada e julgada em estrita conformidade com os princípios básicos da legalidade, da

impessoalidade, da moralidade, da igualdade, da publicidade, da probidade administrativa, da vinculação ao instrumento convocatório, do julgamento objetivo e dos que lhes são correlatos (BRASIL, 2010).

2.1.1 Modalidades

Existem várias modalidades de licitações, sendo que três delas são definidas em razão do valor do contrato: a concorrência, a tomada de preço e o convite.

A concorrência é adequada para contratos de alto valor que apoiam a participação das partes interessadas, mas que preencham as condições do aviso, publicadas pelo menos trinta dias, com extensa divulgação da autoridade oficial. Alguns contratos específicos requerem o uso desta categoria. Também é usado para comprar e vender bens públicos. Apesar de permitir uma ampla participação, eles definem regras bastante exigentes que eliminam parte dos concorrentes, deixando apenas as empresas habilitadas para real (ROGOSKI, 2019).

A tomada de preços é realizada com as partes interessadas devidamente registradas e as condições necessárias para o registro no momento de três dias antes do recebimento das propostas. As propostas são julgadas por um comitê composto por três membros, que também são apresentadas sob o modelo competitivo, mas a diferença é que os licitantes podem ser qualificados antecipadamente por meio dos documentos cadastrais, de modo que as habilitações são resumidas no processo de revisão. Os dados contidos no certificado de registro do licitante e se o licitante possui as capacidades operacionais e financeiras exigidas no anúncio de atividades ou fornecimento de bens e serviços. Após o registro e análise dos documentos enviados, será emitido um certificado. É este certificado que permite a participação de uma empresa nesta modalidade.

O convite é uma modalidade que atende ao contrato de menor valor. De forma simples, pode ser usado para mercadorias que precisam ser adquiridas rapidamente. Assim, o convite é enviado às três empresas convidadas a participar do evento, e o prazo para que essas empresas apresentem propostas no menor espaço de tempo possível.

E outras três são definidas em razão do objeto a ser contratado: o concurso, o leilão e o pregão.

O concurso é utilizado para selecionar e premiar obras de natureza técnica, científica ou artística. As normas de licitação devem ser seguidas na notificação, sendo

importante ressaltar que, desta forma, a intenção não é a aquisição de bens ou serviços, não sendo exigida forma específica de concorrência.

O leilão é um tipo de licitação utilizado para a compra e venda de bens móveis, quando os bens móveis são confiscados, sem valor ou penhorados em juízo, e para a compra de bens doados para quitar dívidas de instituições públicas e obter litígios de forma judicial. Desta forma, o licitante com maior lance sempre ganha a licitação, e os critérios devem ser analisados no edital.

O pregão serve para aquisição de bens e serviços e são muito comuns, o pregão garante um processo de licitação maior e só aceita o menor preço. Este método é dividido em dois tipos: o presencial e o eletrônico.

No pregão presencial o pregoeiro define a participação da empresa e durante a ação pública são dados lances verbais, além das propostas que já foram encaminhadas ao pregoeiro e apresentadas por pessoas físicas legalmente habilitadas a participar da licitação.

Em termos de análise documental, os pregões eletrônicos seguem o mesmo modelo dos pregões presenciais, mas o licitante vencedor enviará o documento em até 02 horas após o término do pregão, podendo ser realizado por meio eletrônico (e-mail) ou direto na entidade que licita.

2.1.2 Etapas de uma licitação

Tendo em vista que as instituições precisam comprar, vender, transferir, arrendar ou assinar os produtos ou serviços contratados, o processo licitatório inicia-se na fase interna:

- Abertura: o responsável deve lançar um edital com as regras de licitação para que todas as empresas aptas a concorrer possam tomar conhecimento.
- Habilitação: É a etapa de verificação dos documentos apresentados pelos participantes. Leva em consideração requisitos pessoais, reconhecimento de habilitações legais, conformidade tributária, habilitações técnicas, econômicas e financeiras. Este estágio é eliminatório, ou seja, os licitantes que deixarem de apresentar os documentos apropriados são considerados desabilitados e impedidos de participar da próxima etapa da licitação.
- Abertura das propostas: Esta é a fase de lances. Nela, é identificada a proposta mais adequada e satisfatória para o órgão público.

- Julgamento e classificação: No processo de julgamento, é verificado se os produtos ou serviços da proposta selecionada atendem aos padrões exigidos no edital e a classificação relaciona as condições de cada participante na ordem que é benéfica para a gestão pública.
- Homologação: Esta etapa ocorre após a verificação e comprovação de que todo o processo licitatório foi realizado de acordo com as leis, normas e regras vigentes, e conforme as normas do edital, a homologação é a aprovação da licitação.
- Adjudicação: É a última etapa do processo licitatório, quando o objeto da licitação é atribuído ao vencedor.

2.1.3 Fraudes em Licitações

A fraude tem crescido assustadoramente em todos os setores da nossa sociedade e não é um privilégio da atualidade. As fraudes que importam o desvio de verba pública para o interesse particular, já não são novidades para a realidade da corrupção no Brasil.

Segundo Castro (2010), os órgãos governamentais que possuem como responsabilidade o combate às fraudes, tais como por exemplo, a Controladoria-Geral da União, a Polícia Federal e o Ministério Público, têm identificado que o principal tipo de corrupção na execução dos recursos públicos é a fraude em licitações e contratos, em geral com o uso de empresas inidôneas.

O mercado de licitações públicas é bastante vulnerável e as empresas que praticam comportamentos anticompetitivos nesses mercados costumam articular estratégias de preços, simular concorrência, interagir com agentes públicos no intuito de obter vantagens econômicas. Segundo relatório da OCDE (2009), destacam-se os seguintes casos de corrupção:

- Apresentação de propostas fictícias ou de cobertura;
- Direcionamento de licitação;
- Divisão de mercado;
- Fixação de preços;
- Rodízio de vencedores ou ofertantes;
- Subcontratação de empresas do grupo ou entes públicos;
- Supressão de propostas;
- Consórcios entre concorrentes - reunião de duas ou mais empresas que se unem com a finalidade de participar de determinada licitação.

2.1.4 - Formação de cartel e conluio

A Secretaria de Defesa da Economia, órgão vinculado ao Ministério da Fazenda define o cartel como:

Acordos ou práticas concertadas entre concorrentes para a fixação de preços, a divisão de mercados, o estabelecimento de quotas ou a restrição da produção e a adoção de posturas pré-combinadas em licitação pública. Os cartéis “clássicos”, por implicarem aumentos de preços e restrição de oferta, de um lado, e nenhum benefício econômico compensatório, de outro, causam graves prejuízos aos consumidores tornando bens e serviços completamente inacessíveis a alguns e desnecessariamente caros para outros. Por isso, essa conduta anticoncorrencial é considerada, universalmente, a mais grave infração à ordem econômica existente (BRASIL, 2017).

De acordo com a Organização para a Cooperação e Desenvolvimento Econômico (OCDE, 2020), tais conluios retiram recursos dos adquirentes e contribuintes, diminuem a confiança do público no processo competitivo e enfraquecem os benefícios de um mercado competitivo.

Ainda segundo a OCDE, os contratos públicos constituem aproximadamente 15% do PIB e, em países não-membros, os números podem ser ainda maiores. No Brasil, em 2018, apenas o governo federal realizou mais de 100 mil processos de contratação, envolvendo aproximadamente 48 bilhões de reais, valor que demonstra o impacto que cartéis em licitação podem causar ao erário.

Os cartéis são altamente prejudiciais à concorrência, pois geram dois efeitos imediatos ao consumidor: o aumento de preço e a redução da oferta. Assim, “quanto maior a duração do cartel, maiores os danos aos consumidores”. Decorre ainda de sua atuação o monopólio num dado mercado relevante às empresas cartelizadas, de forma que o consumidor deixa de experimentar inovações e evoluções no mesmo segmento.

Uma das formas mais comuns de cartelização é a prática conhecida como rodízio (*bid rotation*), em que os licitantes procuram alternar, a cada licitação, o detentor da melhor proposta. Com isso, cria-se uma aparência de competitividade nos certames, dificultando a descoberta de acordo falsos (CECCATO, 2017).

O *cover bidding* é outra forma de cartelização, na qual as empresas rivais combinam a oferta de propostas, com preços mais elevados, ou sob condições desarrazoadas, com o propósito de beneficiar a empresa escolhida pelo grupo para ganhar a licitação. As empresas também podem combinar submeter propostas inferiores,

mas que não atendam às especificações técnicas requeridas pelo órgão contratante, de modo a fazer crer que se trata de um processo concorrencial em condições normais e manter, ao menos, a aparência de concorrência entre as empresas, evitando, assim, a descoberta do acordo prévio. Esse tipo de participação provoca nos gestores públicos a sensação de acirrada competição.

O *bid suppression* ocorre quando uma ou mais empresas concorrentes decidem, de forma orquestrada, não apresentar propostas em uma licitação ou, então, submeter propostas e retirarem-se do certame, a fim de beneficiar determinada empresa que foi eleita pelo grupo para tornar-se a vencedora;

Os cartéis possuem duas classificações sendo elas difusos e clássicos: os difusos de acordo com a OCDE caracterizam-se pelo ajuste eventual entre empresas, numa simples troca de informações sensíveis entre os seus representantes. Os clássicos são os mais comuns, sendo um acordo explícito, anticompetitivo e concentrado, entre concorrentes para fixar preços, estabelecer restrições na oferta ou quotas, dividir mercados por meio de alocação de clientes, fornecedores, área geográfica ou linhas de comércio.

Para mitigar os efeitos de cartelização ou de quaisquer outras irregularidades, a administração pública conta com vários órgãos de controle externo, sendo os mais importantes a Controladoria-Geral da União, Polícia Federal, Ministério Público, Secretaria de Transparência e Prevenção da Corrupção (STPC), para a fiscalização de qualquer irregularidade nas licitações.

2.2 Descoberta de conhecimento em base de dados (DCBD)

O *Knowledge Discovery in Databases* (KDD), ou descoberta de conhecimento em bancos de dados, em português, é o processo não trivial de identificar em dados novos padrões que sejam válidos, que antes eram desconhecidos, e que são potencialmente úteis, possibilitando assim melhor entendimento de um problema ou procedimento de tomada de decisão (SILVA *et al.*, 2002).

O objetivo da descoberta de conhecimento em um banco de dados não é apenas encontrar padrões e relacionamentos na grande quantidade de informações disponíveis, mas também extrair conhecimento que seja compreensível e possa ser usado imediatamente para apoiar a tomada de decisões.

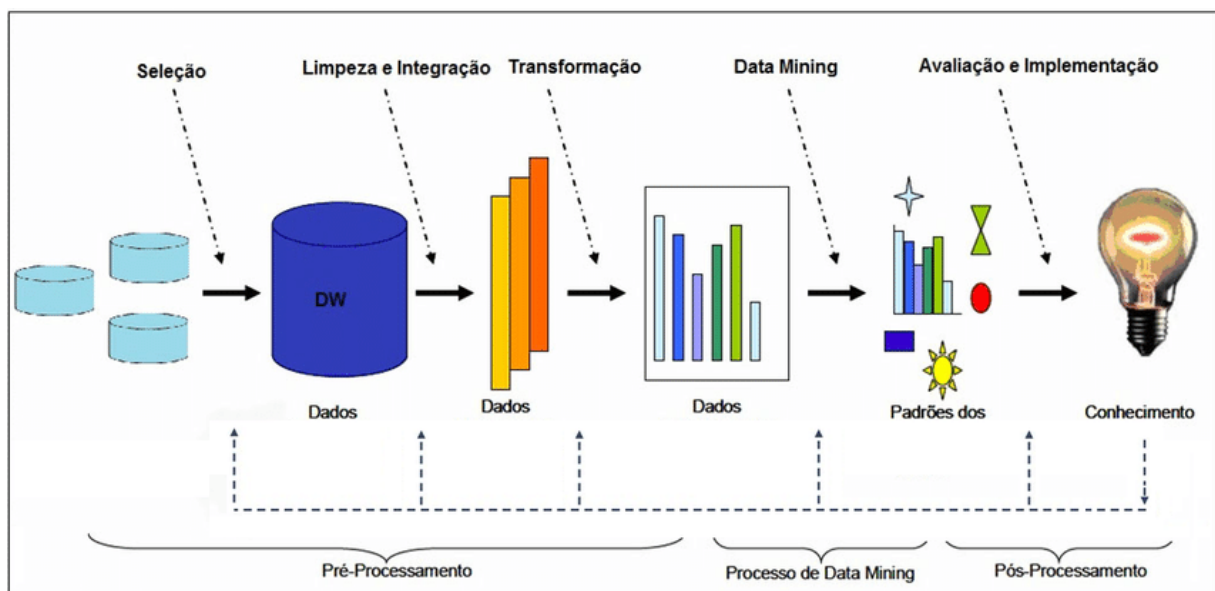
As diferentes fontes de dados a serem utilizadas, coletadas em diferentes momentos e em diferentes locais, proporcionam um esforço inicial para integrar e agrupar todas as informações, que se tornarão a base do processo. É muito importante investigar as inconsistências e problemas causados por diferentes escalas, assim como o tratamento de valores fora da normalidade (*outliers*) e observações erradas. Depois de realizar essas tarefas iniciais de tornar os dados processáveis e homogêneos, as tarefas de descoberta em si, proporcionadas pela mineração de dados podem ser iniciadas para procurar padrões e relacionamentos que sejam significativos e úteis para resolver o problema ou o objetivo a ser alcançado.

2.2.1 Fases do DCDB

O processo de descoberta de conhecimento em base de dados é composto por cinco fases, sendo elas a seleção de dados, pré-processamento dos dados, transformação, mineração e a interpretação/avaliação.

O KDD é uma sequência iterativa de etapas, que são ilustrados, na Figura 1.

Figura 1. Processo Knowledge Discovery in Databases (KDD)



Fonte: Adaptado de Santos, Menezes e da Hora (2014).

A primeira etapa, seleção de dados, é a etapa dedicada à obtenção dos dados. Esta etapa pode-se tornar muito complexa, pelo fato de os dados podem se originar de diversas fontes, como planilhas e sistemas legados, além de possuírem os mais diversos formatos (FAYYAD *et al.*, 1996). A segunda etapa é o tratamento dos dados, que também

é complexa pelo fato de possuir muitos *outliers*. A terceira etapa é a de mineração dos dados, que é o processo de explorar os dados à procura de padrões consistentes. E a última etapa, que é a análise dos dados, é o processo de transformar os dados em conhecimento e *insights* relevantes.

2.2.2 Mineração de dados

O termo mineração de dados (*data mining*) define a exploração e análise de grandes quantidades de dados com o objetivo de encontrar padrões, regras e relações interessantes e significativas para algum fim (BERR, 2000). Esta definição cobre as mais amplas áreas de conhecimento e é eficaz, por exemplo, no entendimento dos clientes da empresa e no desenvolvimento de novas vacinas para certas doenças.

Existem dois tipos de mineração de dados: a direta e a indireta. A mineração de dados direta tenta explicar ou categorizar uma variável alvo definida, como receita proveniente de algum esforço de vendas ou a resposta a uma campanha de *marketing*. A mineração de dados indireta, por sua vez, procura encontrar padrões ou similaridades entre grupos de registros de uma base de dados sem o uso de um alvo particular ou de alguma coleção de classes pré-definida.

2.2.3 Tarefas e técnicas de mineração de dados

O conceito de mineração de dados inclui muitas tecnologias, algoritmos e tipos de modelo, e seu objetivo é sempre pesquisar informações úteis nos dados. O processo de mineração de dados sempre envolve uma das técnicas ou tarefas de mineração listada a seguir:

- Classificação;
- Agrupamento por afinidade ou regras de associação;
- *Clustering*;
- Estimação;
- Predição;
- Descrição e visualização.

A classificação, estimação e predição são exemplos de mineração de dados direta, enquanto o *clustering*, agrupamento por afinidade ou regras de associação e descrição e visualização são exemplos de mineração de dados indireta.

2.2.4 Associação e *FP-Growth*

Uma tarefa em mineração de dados consiste na especificação do que se pretende buscar nos dados, que tipo de regularidades ou categoria de padrões deseja-se encontrar. Já a técnica de mineração consiste na especificação de métodos que garantam como descobrir os padrões que são relevantes. E o algoritmo é o procedimento passo a passo para determinar a solução do problema (SILVA; RALHA, 2011).

O objetivo é encontrar em grandes bases de dados tendências que ajudem a compreender padrões. Os algoritmos buscam encontrar relações entre os itens, verificando os eventos que ocorrem simultaneamente, possibilitando o entendimento de novos modelos e assim atingindo melhores resultados (SERRA, 2002).

A tarefa de associação tem o intuito de identificar associações entre registros de dados que, de alguma maneira, estão ou devem estar relacionados. Sua premissa básica é encontrar elementos que implicam na presença de outros em uma mesma transação, alguns algoritmos que utilizam os conceitos desta tarefa são as regras de associação, ou seja:

Conforme destacado por Han e Kamber (2006), uma regra de associação pode ser formalmente definida como: Seja $I = \{I_1, I_2, \dots, I_n\}$ um conjunto de itens, D um conjunto de dados onde cada transação T é um subconjunto do conjunto de itens $T \subseteq I$. Cada transação é identificada por uma variável-chave. Logo, se A é um conjunto de itens, diz-se que uma transação T contém A se, e somente se, $A \subseteq T$. Por sua vez, uma regra de associação é uma implicação na forma $A \Rightarrow B$, onde $A \subset I$, $B \subset I$ e $A \cap B = \emptyset$. O percentual de transações em D que contém o conjunto união $A \cup B$ é chamado de suporte da regra $A \Rightarrow B$, ou seja, a probabilidade incondicional $p(A \cup B)$. Por outro lado, o percentual de transações em D contidas em A e que também contém B informa a estatística de confiança, isto é, a probabilidade condicionada $p(B|A)$. A partir desses dados é possível calcular valores de suporte para várias combinações de conjuntos de empresas.

O algoritmo *FP-Growth* utiliza uma estrutura compacta e funcional, a *FP-Tree* (*Frequent Pattern Tree*), para armazenar o conjunto dos *itemsets* frequentes e evitar varreduras repetitivas da base de dados. O *FP-Growth* gera o conjunto dos *itemsets* frequentes, depois é feito o armazenamento destes *itemsets* em uma estrutura

denominada *FP-Tree*, se várias transações compartilham um conjunto de itens, eles podem ser agrupados nos mesmos nós da árvore *FP-Tree*, que registram junto ao nó o número de ocorrências em uma variável count.

Para o funcionamento do algoritmo é necessário informar ao menos dois parâmetros de entrada, o suporte e a confiança. Suporte é a métrica utilizada para encontrar todos os N *itemsets*. O suporte de uma regra de associação $X, A \Rightarrow B$, corresponde a frequência com que A e B ocorrem em uma base de dados. Confiança é responsável pela força da regra, dessa forma, a confiança corresponde a frequência com que B ocorre, dentre os eventos que contêm A.

Com o conjunto de N *itemsets* frequentes encontrados através da estrutura *FP-Tree*, o algoritmo *FP-Growth* gera as regras de associação. A partir deste momento começa a fase de pós-processamento das regras geradas, objetivando a seleção das regras úteis e fortes. Somente as regras de associação que satisfizerem todas as medidas de qualidade são consideradas relevantes e úteis. (HAMILTON, 2007).

O algoritmo *FP-Growth* irá primeiro calcular a confiança, em seguida o *Lift*, que é utilizado para avaliar dependências, depois o *Leverage* será calculado, é utilizado para descobrir o valor da diferença entre o suporte real e o esperado de uma regra de associação. E por último a Convicção é calculada, ela avalia a regra de associação como uma implicação do antecedente no consequente.

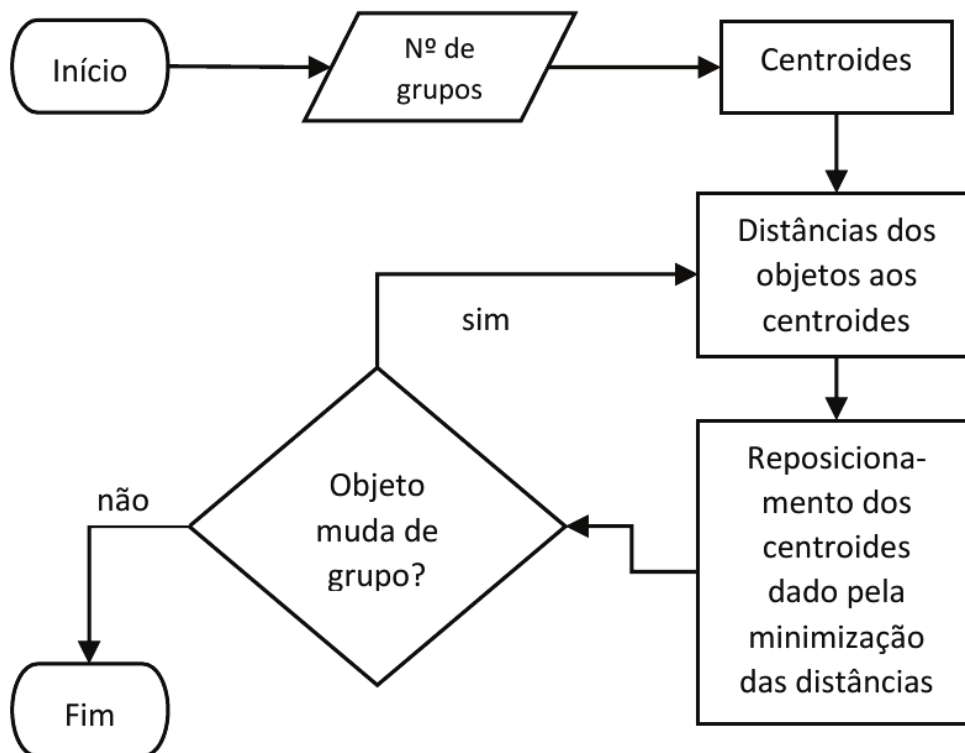
2.2.5 Agrupamento e K-means

A tarefa de clusterização (agrupamento) agrupa um conjunto de elementos em subconjuntos de elementos ou partes, observando para isso características conforme critério apropriado. Esses subconjuntos são chamados de *clusters*, de forma que os objetos pertencentes a um mesmo cluster possuam semelhanças entre si e, ao mesmo tempo, os objetos pertencentes a *clusters* diferentes apresentem alta dissimilaridade (SOARES, 2004).

A técnica de agrupamento em métodos de particionamento é uma das técnicas possíveis para a realização de clusterização, que tem como objetivo encontrar a melhor partição dos n objetos em K grupos. Normalmente os K grupos encontrados possuem mais qualidade comparados com K grupos produzidos pelos métodos hierárquicos. Um dos algoritmos mais relevantes para essa técnica é o algoritmo denominado como K-means (RODRIGUES; 2009).

A ideia do algoritmo K-means é escolher k objetos (aleatoriamente ou com alguma heurística) que formam a base de cada grupo (denominados centróides), ou seja, o centro de cada grupo é denominado centróide, e os demais objetos são associados ao centróide mais próximo. A cada passo os centróides são recalculados dentre os objetos do seu próprio grupo e os objetos são realocados para o centróide mais próximo, e este procedimento é repetido até que o nível de convergência seja satisfatório de acordo com alguma heurística estabelecida. A Figura 2 é um exemplo de fluxograma do funcionamento do algoritmo.

Figura 2. Fluxograma de execução do K-means.



Fonte: Adaptado de SILVA (2015).

2.3 Estudos correlatos

Existem vários trabalhos relevantes disponíveis relacionados à mineração de dados com licitações públicas, sendo um deles apresentado em MORAIS (2016). Neste trabalho, é realizado um estudo utilizando indicadores para investigação de licitações. Esses indicadores foram obtidos a partir da análise do comportamento da empresa vencedora, empresa perdedoras e do órgão licitante, além dos dados da própria licitação,

por meio das técnicas de reconhecimento de padrões estatísticos e mineração de dados, os indicadores utilizados para a descoberta de formação de cartéis nas licitações, distribuição de preços, associação de empresas, número de participantes de licitações, número de empresas vencedoras por órgãos licitantes, utilizou as técnicas de regressão linear, associação e clusterização. Os algoritmos utilizados foram o Apriori e o K-means.

Silva e Ralha (2011) fizeram um estudo e abordaram os conceitos sobre tecnologia de *Business Intelligence* (BI) e dos seus principais componentes, bem como os conceitos de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases*), e uma comparação das funcionalidades presentes nas ferramentas de mineração de dados, cujo objetivo é a detecção de cartéis em licitações públicas utilizando as tarefas de associação e clusterização.

Outro trabalho relevante foi o de SILVA (2020), abordou os conceitos de ciência de dados, Descoberta de Conhecimento em Bases de Dados, e utilizou as técnicas de mineração de dados agrupamento e associação, os dados que foram utilizados no estudo de SILVA (2020) foram oriundos da CGE/GO, tratou os dados em datasets e aplicou as técnicas de mineração de dados utilizando os algoritmos FP-Growth e o K-means

3. PROCEDIMENTO METODOLÓGICO

Define-se pesquisa como o procedimento racional e sistemático que busca proporcionar respostas aos problemas que são propostos (SEVERINO, 2017).

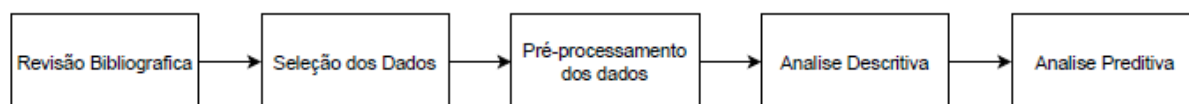
As pesquisas científicas podem ser classificadas de acordo com a natureza, abordagem, objetivos e procedimentos técnicos. Quanto à natureza, esta pesquisa é considerada aplicada, dedicada à geração de conhecimento para solução de problemas específicos, e dirigida à busca da verdade para determinada aplicação prática em situação particular (NASCIMENTO; SOUSA, 2015).

As atividades deste trabalho incluem métodos e técnicas disponíveis nas literaturas voltadas à pesquisa científica em computação e ciência de dados. O projeto é classificado de acordo com seus objetivos de natureza exploratória e descritiva, pois foi desenvolvido com o objetivo principal de relacionar os principais aspectos e elementos necessários à compreensão da relação de ciência de dados com técnicas de mineração de dados e sua aplicação em licitações públicas, buscando anomalias que antes eram desconhecidas, descrevendo a eficiência da mineração de dados para essa situação problema (WAZLAWICK, 2014).

Quanto à abordagem, essa pesquisa é do tipo quantitativa, pois utiliza métodos que empregam medidas padronizadas e sistemáticas, reunindo respostas pré-determinadas, facilitando a comparação e a análise de medidas estatísticas de dados.

Para a realização do trabalho, foram definidas etapas que englobam desde a seleção dos dados até a análise dos dados, ilustrada na Figura 3.

Figura 3. Fluxograma da abordagem proposta



Fonte: Elaborado pelo autor.

De acordo com os meios utilizados para investigação, esse estudo inicia com uma pesquisa bibliográfica e experimental, utilizando-se de livros, artigos acadêmicos, revistas, leis, dissertações, relacionado aos temas: ciência de dados, descoberta de conhecimento em base de dados, mineração de dados e licitações públicas (WAZLAWICK, 2014).

Foi realizada uma revisão bibliográfica de estudos relacionados à ciência de dados, tecnologia de mineração de dados e licitações públicas, conceituação e exploração das técnicas e como é realizado o processo licitatório, especialmente no Estado de Goiás. Foi realizado um estudo com o objetivo de identificar empresas fraudulentas em órgãos governamentais, com o objetivo de analisar e complementar as variáveis utilizadas na obtenção dos indicadores.

Os dados foram provenientes do site [\(http://www.transparencia.go.gov.br/portaldatransparencia/\)](http://www.transparencia.go.gov.br/portaldatransparencia/) da CGE (Controladoria geral do Estado de Goiás), e no portal estão disponíveis informações divididas em seis categorias, sendo elas: despesas; receitas; convênios e repasses; pessoal; planejamento e orçamento; gestão fiscal e governamental. Dentro da categoria despesas existem informações de advogados dativos, contratos, empenhos e pagamentos, execução orçamentária, fornecedores, gastos governamentais, gastos com publicidades e propaganda, ordem cronológica de pagamentos e informações sobre licitações, foco deste estudo.

O portal apresenta duas opções para obter os dados sobre licitações, sendo a primeira acessando através da categoria despesas e a seguinte escolhendo licitações. A segunda opção é através da seção de dados abertos, seção na qual se permite a todos os interessados obterem as informações que são apresentadas no painel analítico do portal.

Os dados foram extraídos em formatos *comma-separated values* (CSV), tipo de formato de arquivo de valores separados por vírgula amplamente utilizado para armazenar dados tabulares.

A realização do processamento dos dados foi executada de maneira manual, sendo posteriormente efetuado um tratamento dos dados, realizando a remoção de dados impuros.

A análise descritiva consiste em descrever as principais tendências nos dados existentes e em observar situações que levam a novos fatos. Foi realizada uma análise descritiva sobre os dados coletados, e em cada certame foram coletadas as empresas participantes e a empresa vencedora do certame.

A análise preditiva usa dados e algoritmos estatísticos para identificar a probabilidade de resultados futuros a partir de dados históricos. Foi realizada uma análise preditiva dividida em três etapas: a primeira etapa é a aplicação de regras de associação, a segunda etapa é a aplicação da clusterização e a terceira etapa é a aplicação de regras

de associação. A análise é baseada na pesquisa realizada por Silva e Ralha (2011), também utilizado por Silva (2021). Com o auxílio do algoritmo FP-Growth, as regras de associação são aplicadas ao *dataframe* de propostas de licitação. Utilizando o mesmo *dataframe*, procura-se encontrar o grupo de participantes que reunissem comportamentos e características semelhantes utilizando a tarefa de mineração de dados de clusterização, com técnica de agrupamento em métodos de particionamento e o algoritmo K-means. A etapa final é associada ao *cluster* obtido, selecionando as melhores regras e analisando as características dos participantes com o apoio do *software* RapidMiner.

4 RESULTADOS

Este capítulo apresenta os experimentos e os principais resultados encontrados. A Seção 4.1 refere-se à análise descritiva e a Seção 4.2 a análise preditiva. Ao final, são discutidas algumas conclusões dos experimentos de análise preditiva.

4.1 Análise descritiva

A análise descritiva realizada neste trabalho visa explorar o comportamento das licitações exibindo dados obtidos através de gráficos estatísticos, ou seja, traduzir diversas informações constantes novos arquivos em relatórios e ilustrações de forma visual.

A Tabela 1 exibe as modalidades com a quantidade de licitações, o valor estimado e o valor total referente ao ano de 2019. Observa-se que a modalidade com maior quantidade de licitações é a de Participante e Registro de Preço tendo 1640 licitações no total de 2254. O Pregão Eletrônico por Sistema de Registro de Preço (SRP) é a segunda maior com 454 licitações, o Pregão Eletrônico com 139 licitações, a Concorrência com 11 licitações, o Pregão Presencial por Regime Diferenciado de Contratação Pública (RDC) com 3, e as demais Concorrência SRP, Convite, Tomada de preço com o número de 1 licitação.

Tabela 1. Número de licitações, valor estimado e valor total por modalidade.

MODALIDADE	NUMERO DE LICITAÇÕES	VALOR ESTIMADO	VALOR TOTAL
Participante Registro de Preço	1640	4.636.699.100,48	644.830.856,01
Pregão Eletrônico SRP	454	19.413.720.525,40	2.025.371.517,24
Pregão Eletrônico	139	656.000.327,87	49.150.125,60
Concorrência	11	124.352.577,86	104.352.577,86
Pregão Presencial	4	60.850.541,79	10.183.008,12
Pregão Presencial RDC	3	17.077.012,01	17.077.012,01
Concorrência SRP	1	63.105.700,00	3.712.100,00
Convite	1	435.000,00	87.000,00
Tomada de Preço	1	73.440.872,00	4.590.054,50
TOTAL	2254	25.045.681.657,41	2.859.354.251,34

Fonte: Elaborado pelo autor.

A Figura 4 mostra o valor adjudicado no ano de 2019, assim é possível observar um gasto muito elevado, totalizando R\$ 3,77 Bilhões no ano.

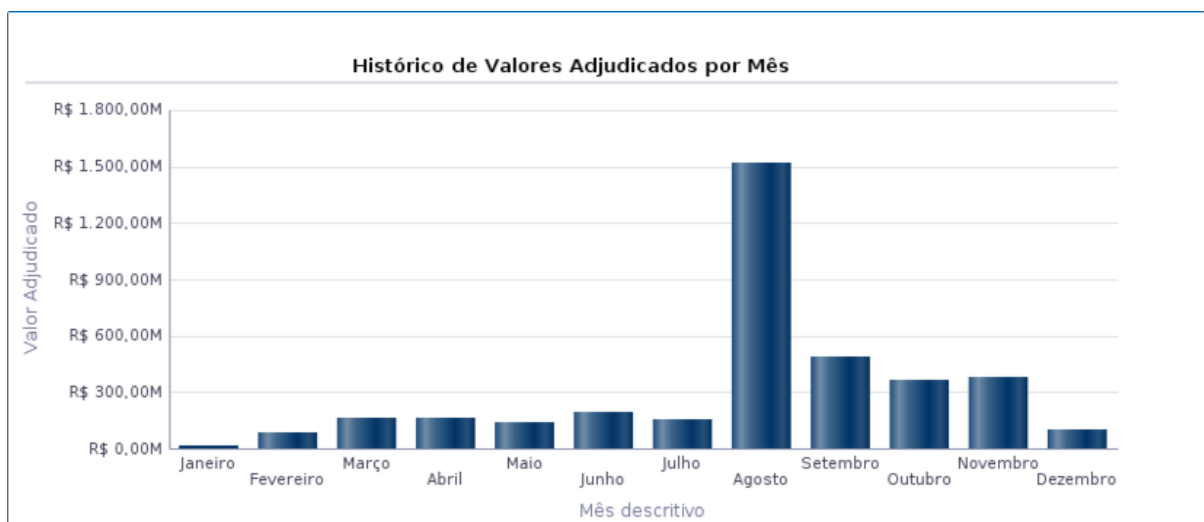
Figura 4. Valor adjudicado no ano de 2019.

	Valor Adjudicado
Compra Direta - Dispensa de Licitação	
Compra Direta - Obras e Serviços de Engenharia	R\$ 176.743,63
Concorrência	R\$ 509.707.995,31
Convite	R\$ 346.967,48
Inexigibilidade	R\$ 156.998.237,98
Outras Dispensas	R\$ 1.468.340.396,25
Participante Registro de Preço	R\$ 278.949.172,55
Pregão Eletrônico	R\$ 154.899.327,28
Pregão Eletrônico SRP	R\$ 609.890.107,42
Pregão Presencial	R\$ 589.929.170,78
Pregão Presencial RDC	R\$ 560.584,21
Pregão Presencial SRP	R\$ 300.000,00
Tomada de Preço	R\$ 3.744.343,13
Total Geral	R\$ 3.777.032.026,09

Fonte: Controladoria Geral do Estado de Goiás

A Figura 5 mostra o valor adjudicado por mês no decorrer do ano de 2019, sendo que o mês de agosto possui o valor mais alto do ano, de aproximadamente 1.5 milhões de reais.

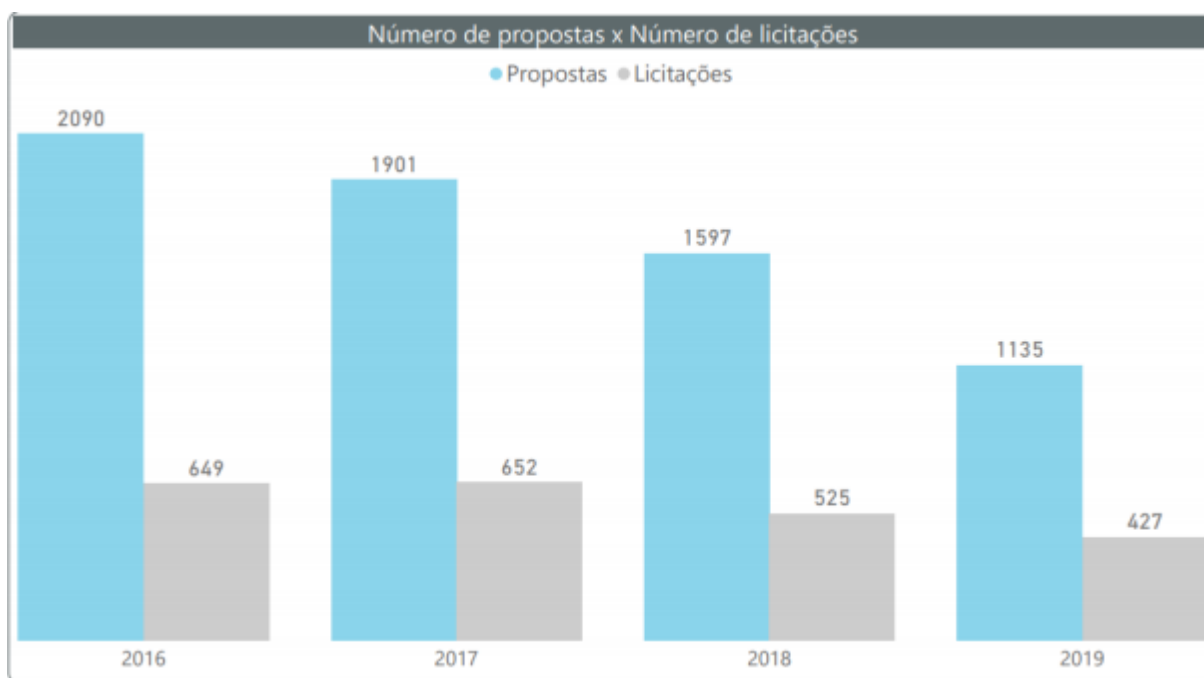
Figura 5. Valor adjudicado por mês no ano de 2019.



Fonte: Controladoria Geral do Estado de Goiás

Na Figura 6 é exibido o número de licitações e o número de propostas no período de 2016 a 2019. No período de 2016-2019 observa-se uma queda média de 4,6% de propostas ao longo dos anos, 2016 obteve 31%, 2017 obteve 28%, 2018 obteve 24% e 2019 obteve 17%.

Figura 6. Número de propostas e número de licitações por ano (2016-2019).



Fonte: Adaptado de SILVA (2021).

4.2 Análise preditiva

Uma vez concluída a análise descritiva, considerada de grande relevância, pois pode permitir compreender melhor o problema e visualizar as variáveis disponíveis, a análise preditiva pode ser iniciada.

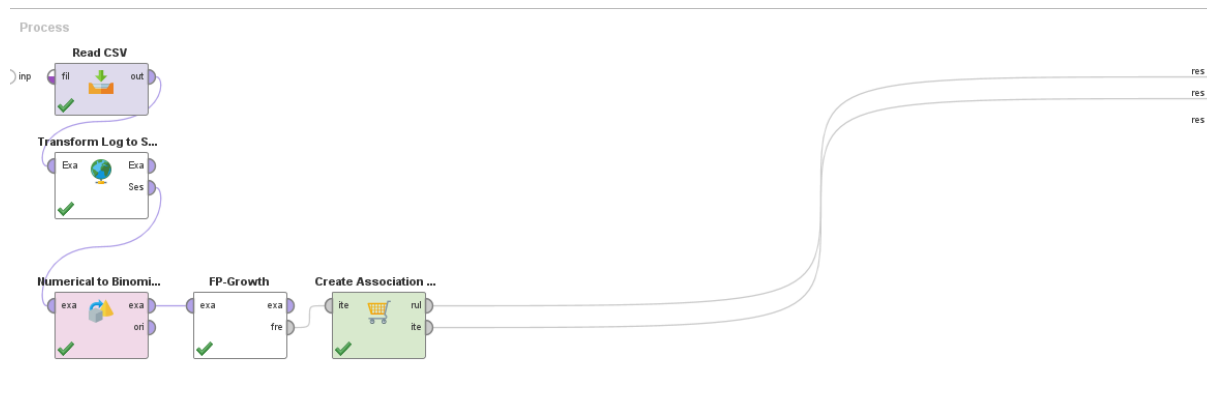
Para essa análise não foram utilizados todos os indicadores, e sim os indicadores que representassem os participantes, vencedores e licitações, utilizando a ferramenta de mineração de dados RapidMiner.

Conforme Grilo Júnior (2010), as operações no RapidMiner funcionam por meio de processos, ou seja, para realizar qualquer tarefa de mineração é preciso incluir operadores para cada processo que antecede a tarefa final desejada, sendo que alguns operadores possuem parâmetros a serem ajustados, necessitando que sejam informados pelo usuário. Os operadores, sequência de execução e parâmetros são armazenados em um arquivo do tipo *eXtensible Markup Language (XML)*, permitindo editar para ajustes ou para salvar as configurações do experimento realizado para futura reutilização.

A execução do processo conforme ilustrado na Figura 7, com os dados das licitações e empresas participantes composto por 5556 (cinco mil quinhentos e cinquenta e seis) linhas, com 2254 (duas mil, duzentas e cinquenta e quatro) licitações e 884

(oitocentas e oitenta e quatro) empresas participantes, sendo que cada linha representa uma licitação.

Figura 7. Área de processos/operadores da ferramenta RapidMiner



Fonte: Elaborado pelo autor.

Em licitações que contém mais de uma empresa participante, o número da licitação se repete, porém com o indicador da empresa é diferente, representando uma nova empresa para a licitação, conforme ilustrado na Tabela 2. Isso ocorre porque em uma mesma licitação, pode haver vencedores e/ou participantes por lote da licitação.

Tabela 2. Licitações e participantes.

LICITAÇÃO	PARTICIPANTE
LIC_01	PAR_01
LIC_02	PAR_02
LIC_03	PAR_03
LIC_04	PAR_04
...	...
LIC_N	PAR_N

Fonte: Elaborado pelo autor.

Após realizar a relação das licitações e empresas participantes, o arquivo foi salvo no formato CSV. Para realizar a leitura do arquivo no *RapidMiner* foi utilizado o operador *ReadCSV*, conforme a Figura 7. Este operador lê um arquivo CSV e gera um conjunto de treinamento, uma tabela com duas colunas e 5556 linhas que representam as licitações e as propostas.

A tabela criada a partir do operador *ReadCSV* pode ser acessada por qualquer outro operador através da linha *out* que é a saída do operador. O operador que foi utilizado em seguida transforma a tabela gerada pelo operador *ReadCSV* em uma matriz

esparsa, denominado *Transform log to session*. Para este operador é preciso configurar dois parâmetros, o *session attribute* e *resource attribute*. O *session attribute* recebe o nome da coluna que contém as licitações e *resource attribute* o nome da coluna que contém os participantes, conforme Figura 8.

Figura 8. Parâmetros do operador log to session



Fonte: Elaborado pelo autor.

O operador *Transform Log to Session* possui duas saídas, uma com o resultado do processamento e outra com os dados originais. É utilizada a saída com os dados processados. Após obter a matriz esparsa, os dados precisam ser transformados do tipo numérico para dados binário, para utilizar o operador *FP-Growth* que é responsável por identificar os conjuntos de itens frequentes. O operador que realiza esse procedimento é o *Numerical to Binominal*, gerando os dados conforme apresentado na Tabela 3.

Tabela 3. Matriz esparsas de licitações e participantes, considere 1 *true* e 0 *false*.

LICITAÇÕES	PAR_01	PAR_02	PAR_03	PAR_4	PAR_N
LIC_01	1	1	0	0	...
LIC_02	1	1	1	1	...
LIC_03	0	1	1	1	...
LIC_04	0	1	0	1	...
LIC_N

Fonte: Elaborado pelo autor.

A Tabela 3 ilustra o código do certame e as empresas participantes do certame, sendo que cada linha da Tabela 3 corresponde a um certame, por exemplo utilizando as regras de associação, as licitações LIC_002, LIC_003, LIC_004 observa-se que o conjunto de participantes (PAR_02, PAR_04), isto é, os participantes 02 e 04 concorrem em três licitações juntos.

Após o uso do operador *Numerical to Binominal* foi possível fazer uso do operador *FP-Growth* para a identificação dos itens que são frequentes. Na figura 9, o único parâmetro ajustável do operador *FP-Growth* é o *min support*, que é o valor do suporte mínimo a ser considerado.

Figura 9. Operador FP-Growth e seus parâmetros



Fonte: Elaborado pelo autor.

Com base nos itens frequentes e no valor mínimo de suporte gerados pelo operador *FP-Growth* é possível utilizar o operador *Create Association Rules* para gerar as regras de associação. No operador *Create Association Rules* o único parâmetro ajustável é o *min confidence*, ou seja, o valor mínimo para a confiança, como é demonstrado na Figura 10.

Figura 10. Operador Create Association Rules e seus parâmetros.



Fonte: Elaborado pelo autor.

No RapidMiner, o processo de mineração de itens frequentes é dividido nesses dois operadores, sendo que o *FP-Growth* implementa a geração de itens frequentes, e o *Create Association Rules* implementa as regras de associações a partir do conjunto do operador anterior.

No primeiro experimento realizado foi utilizado o suporte mínimo de 0,01 (1%) e confiança mínima de 0.50 (50%). Com estes parâmetros foram encontradas 22 (vinte e duas) regras de associação e 84 (oitenta e quatro) conjuntos de itens frequentes, e a

Tabela 4 representa os resultados do processamento dos dados nesse primeiro experimento.

Tabela 4. Resultado do processamento dos dados minerados.

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
6	119.0, 155.0	131.0	0.016	0.556	0.988	-0.040	0.010	2.959	1.828
7	313.0	131.0	0.013	0.558	0.990	-0.033	0.009	2.970	1.836
8	101.0	135.0	0.025	0.560	0.981	-0.064	0.022	7.886	2.111
9	101.0, 123.0	135.0	0.012	0.562	0.991	-0.031	0.010	7.921	2.123
10	204.0	135.0	0.016	0.581	0.989	-0.039	0.014	8.176	2.215
11	131.0, 111.0	119.0	0.016	0.590	0.989	-0.038	0.012	4.155	2.093
12	157.0	131.0	0.012	0.605	0.993	-0.027	0.008	3.221	2.055
13	123.0	135.0	0.025	0.606	0.984	-0.058	0.022	8.539	2.360
14	155.0	119.0	0.028	0.643	0.985	-0.059	0.022	4.526	2.402
15	166.0	119.0	0.011	0.658	0.994	-0.023	0.009	4.632	2.508
16	131.0, 142.0	119.0	0.018	0.667	0.991	-0.036	0.014	4.694	2.574
17	119.0, 142.0	131.0	0.018	0.667	0.991	-0.036	0.013	3.551	2.437
18	119.0, 313.0	131.0	0.011	0.667	0.995	-0.021	0.008	3.551	2.437
19	119.0, 313.0	155.0	0.011	0.667	0.995	-0.021	0.010	15.327	2.870
20	131.0, 155.0	119.0	0.016	0.686	0.993	-0.030	0.012	4.832	2.735
21	313.0	119.0	0.016	0.692	0.993	-0.030	0.013	4.874	2.788
22	111.0, 142.0	131.0	0.010	0.697	0.996	-0.019	0.007	3.712	2.680
23	111.0, 155.0	131.0	0.010	0.719	0.996	-0.018	0.008	3.828	2.888
24	111.0, 142.0	119.0	0.011	0.758	0.997	-0.018	0.009	5.334	3.539
25	111.0, 155.0	119.0	0.011	0.781	0.997	-0.017	0.009	5.500	3.922
26	131.0, 313.0	119.0	0.011	0.828	0.998	-0.015	0.009	5.827	4.976
27	155.0, 313.0	119.0	0.011	0.857	0.998	-0.014	0.009	6.035	6.006

Fonte: Elaborado pelo autor.

Os resultados não foram muito satisfatórios, pois foram geradas poucas regras de associação. Além disso, as regras obtidas não tinham o suporte e a confiança tão altas. A confiança máxima obtida foi de 0.857 (85%) e o suporte máximo foi de 0.028 (2,8%). No caso dos cartéis, mais especificamente do rodízio de licitações, um alto valor de confiança pode garantir que a frequência de ocorrência dos participantes no lado antecessor da regra seja próxima à frequência dos participantes no lado sucessor da regra.

O segundo experimento realizado foi com o valor do suporte em 0,001 com o objetivo de se obter mais regras de associação. Foram geradas 28045 regras de associação e 4390 conjuntos de itens frequentes. Mesmo com essa grande quantidade de regras, algumas regras importantes poderiam ficar de fora, devido a sua ocorrência ser menos frequente. Desta forma, não se garante a inclusão de algumas regras úteis, mas não frequentes com esta metodologia.

Uma solução foi utilizar a tarefa de clusterização, e um novo *dataset* foi preparado para realizar a execução desta tarefa para que se encontre os prováveis grupos que estão simbolizando mercados de licitações, dividindo desta forma o espaço de soluções e obtendo mais precisão nos *insights*

Este *dataset* foi criado com base no primeiro *dataset* utilizado nos experimentos anteriores contendo 5556 registros, ou seja, cada linha representando licitações e os itens. A Tabela 5 apresenta como foi estruturado o *dataset*.

Tabela 5. *Dataset* para realizar a clusterização

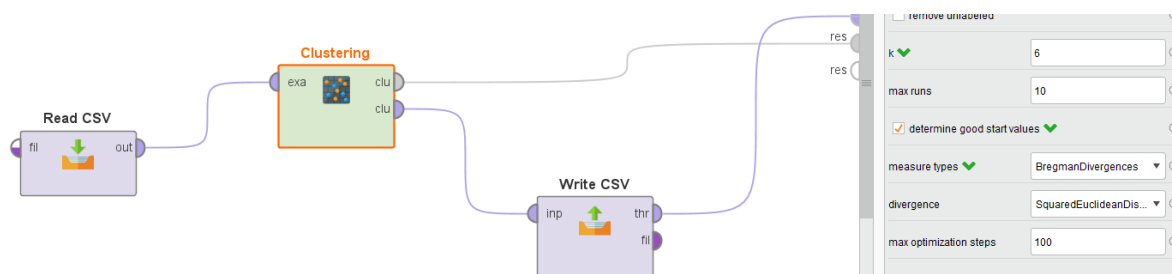
LICITAÇÃO	PARTICIPANTES	ÓRGÃO
LIC_01	PAR_01	ORG_01
LIC_02	PAR_02	ORG_02
LIC_03	PAR_03	ORG_03
LIC_04	PAR_04	ORG_04
...
LIC_N	PAR_N	ORG_N

Fonte: Elaborado pelo autor.

O *dataset* foi salvo em formato CSV, inserido no RapidMiner e foi lido pelo operador ReadCSV, sendo que a saída do operador ReadCSV é conectada na entrada do operador *Clustering*, operador responsável por realizar o modelo de clusterização. Em seguida, a primeira saída do operador *Clustering*, que é responsável por gerar o modelo é conectada na entrada do resultado do processamento. A segunda saída do operador *Clustering* possibilita visualizar o cluster gerado correspondente a cada registro da base de dados. Para finalizar o processo, a segunda saída é conectada no operador WriteCSV, operador que permite exportar os dados gerados em um arquivo CSV.

Esse experimento foi realizado utilizando a técnica de agrupamento, com o uso do algoritmo *K-means*, sendo que a quantidade de *clusters* resultantes no parâmetro do algoritmo foi 6, conforme mostra a Figura 11.

Figura 11. Processo para realizar a Clusterização.



Fonte: Elaborado pelo autor.

Na Tabela 6 é demonstrada a saída do operador WriteCSV com os *clusters* gerados, no qual mostra licitação, participante e órgão respectivos de cada *cluster*

Tabela 6. Saída do *WriteCSV* com os *clusters* gerados.

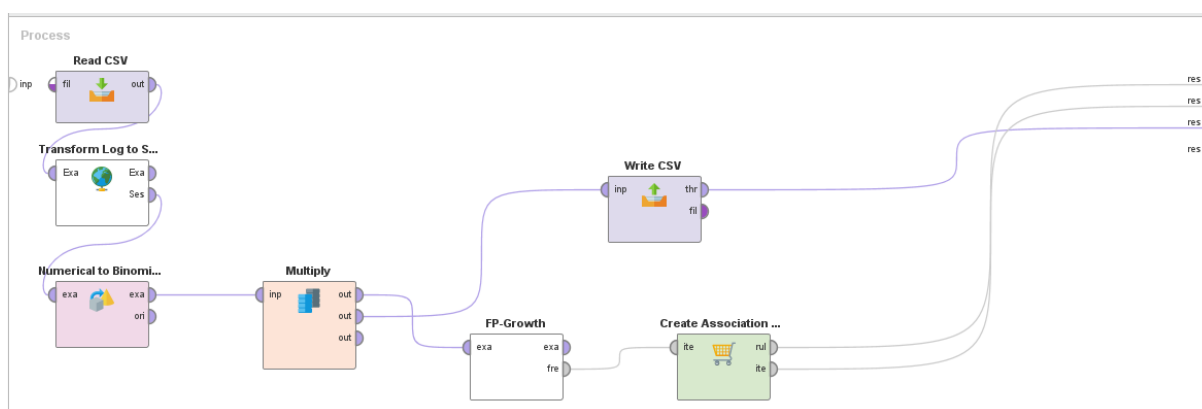
LICITAÇÃO	PARTICIPANTE	ÓRGÃO	CLUSTER
LIC_01	PAR_01	ORG_01	CLU_01
LIC_02	PAR_02	ORG_02	CLU_02
LIC_03	PAR_03	ORG_03	CLU_02
LIC_04	PAR_04	ORG_04	CLU_03
...
LIC_N	PAR_N	ORG_N	CLU_N

Fonte: Elaborado pelo autor.

O *cluster* 0 (zero), teve a quantidade de itens igual à 88, o *cluster* 1 (um) teve 1431 itens, o *cluster* 2 (dois) 1366 itens, o *cluster* 3 (três) 1012 itens, o *cluster* 4 (quatro) 1546 itens, o *cluster* 5 (cinco) 113 itens, totalizando 5556 itens.

Após o processo de clusterização foi realizada novamente a aplicação das regras de associação em cada *cluster* conforme é demonstrado na Figura 12, na tentativa de identificar grupos de participantes associadas atuando especificamente nos novos grupos definidos. Além disso, com menos registros, foi possível obter mais regras em cada grupo, com a confiança e suporte maiores, proporcionando mais qualidade na escolha das regras e mais precisão na geração de insights. A Tabela 7 demonstra os *clusters* gerados, a quantidade de licitações, a quantidade de atributos, o suporte mínimo, a confiança mínima, a quantidade de itens frequentes e a quantidade de regras por *cluster*.

Figura 12. Processo para aplicar a Regra de Associação em cada *cluster*.



Fonte: Elaborado pelo autor.

Tabela 7. Execução do algoritmo de associação FP-Growth em cada cluster.

CLUSTERS	LICITAÇÕES	ATRIBUTOS	SUORTE MIN.	CONFIANÇA MIN.	ITENS FREQUENTES	REGRAS
CLU_0	71	65	1%	50%	64	34
CLU_1	380	341	1%	50%	225	651
CLU_2	405	376	1%	50%	187	228
CLU_3	266	198	1%	50%	138	97
CLU_4	544	477	1%	50%	433	1854
CLU_5	94	53	1%	50%	84	47

Fonte: Elaborado pelo autor.

Com esses experimentos realizados foi possível escolher as melhores regras de associação gerada por cada *cluster* com base nos parâmetros fornecidos, como confiança, suporte, *lift* e convicção, para analisar o comportamento nas participações das licitações. Em busca de irregularidades como cartel e rodízio, foram escolhidas 3 (três) regras com o intuito de apresentar os resultados alcançados.

Uma regra presente no “CLUSTER_4” que possui o valor de suporte de 0.029 (2,9%) e confiança 95%, informa que nas ocorrências em que a PAR_131 está presente, a PAR_119 também participa em 57,9% dos casos. Essas duas empresas participaram de 15 licitações. Das 15 licitações que participaram juntas, a PAR_131 venceu sozinha em apenas 2 licitações, e a PAR_119 venceu sozinha em 1 licitação. A PAR_131 contém 40 participações e a PAR_119 contém 15 participações nesse cluster, e na base toda a PAR_131 possui 193 participações e a PAR_119 possui 218 participações em licitações, essa regra do *cluster* 4 é representada na Tabela 8.

Tabela 8. Tabela das regras de associações obtidas no *cluster* 4

REGRA ESCOLHIDA DO CLUSTER 4
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
[131] --> [119] (confidence: 0.950)
TOTAL = 15

Fonte: Elaborado pelo autor.

Uma regra no “CLUSTER_1” com valor de suporte 0.038 e confiança de 100% permitiu estabelecer a relação entre duas empresas, PAR_135 e PAR_144, que possuem participações individuais em licitações, sendo que, entre essas, elas participaram juntas de 25 licitações e ganharam juntas 13 licitações. Vale ressaltar que, das 30 licitações que a PAR_135 participou, a PAR_144 não participou somente de 6 licitações. A regra do cluster 1 é representada na Tabela 9.

Tabela 10. Tabela das regras de associações obtidas no *cluster 3*

REGRA ESCOLHIDA DO CLUSTER 3
[189] --> [205] (confidence: 0.905)
[189] --> [205] (confidence: 0.905)
[189] --> [205] (confidence: 0.905)
[189] --> [205] (confidence: 0.905)
[189] --> [205] (confidence: 0.905)
[189] --> [205] (confidence: 0.905)
[189] --> [205] (confidence: 0.905)
[189] --> [205] (confidence: 0.905)
[189] --> [205] (confidence: 0.905)
[189] --> [205] (confidence: 0.905)
TOTAL = 10

Fonte: Elaborado pelo autor.

5 CONCLUSÃO

A partir dos dados obtidos através do portal CGE/GO foi possível a obtenção de indicadores para realizar a investigação de licitações utilizando os conceitos de ciência de dados e as técnicas de mineração de dados.

Este trabalho corresponde a algumas das possibilidades que a ciência de dados e a mineração de dados podem fornecer, de modo que utilizando as técnicas adequadas, *insights* de grandes valores podem surgir. As regras encontradas neste experimento não significam que de fato existem irregularidades ou formação de cartel nas participações e vitórias conjuntas, mas abrange a visão dos profissionais que fazem a análise dos processos licitatórios. Espera-se que com o resultado de técnicas de mineração de dados aplicada ao âmbito de dados como é disponibilizado no portal da CGE/GO um ganho de produtividade e rapidez no processo de fiscalização pela área responsável.

Conclui-se que com a utilização dos métodos adequados que a mineração de dados oferece pode encontrar-se situações anômalas que podem contribuir na análise e identificação das possíveis irregularidades nos processos licitatórios analisados pelo órgão fiscalizador, auxiliando no desenvolvimento de sistemas de apoio à decisão e definição de estratégias, gerando informações que permitem maior assertividade, garantindo um aumento na eficiência do uso dos recursos públicos.

5.1 Trabalhos futuros

- Testar a metodologia em outros órgãos que realizam o controle de fiscalização, a fim de verificar se a mineração de dados contribui positivamente.
- Desenvolver um método para que a escolha das regras de associação seja mais eficiente.
- Aplicação de novos métodos utilizando novas variáveis.

REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules. **Proc. of the 20th Int'l Conference on Very Large Databases**. Santiago, Chile, set. 1994.

AMORIM, Victor Aguiar Jardim de. **Licitações e contratos administrativos: teoria e jurisprudência**. Brasília: Senado Federal, 2017.

BATISTA, Gustavo Enrique de Almeida Prado Alves. **Pré-processamento de Dados em Aprendizado de Máquina Supervisionado**. 2003. 232 f. Tese (Doutorado) - Curso de Ciências - Ciências de Computação e Matemática Computacional, Universidade de São Paulo, São Carlos, 2003.

BRAGA, Tereza Cristine Almeida. CADE, Cartéis e Licitações: um Novo Nicho da Política Antitruste Brasileira. **Revista de Defesa da Concorrência**, Brasília, v. 3, n. 1, p. 108-132, maio 2015.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal: Centro Gráfico, 1988. 292 p.

BRASIL. Constituição (1988). Lei Complementar nº101, de 4 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. **Diário Oficial da União**. Brasília, DF, 4 de maio de 2000.

BRASIL. Constituição (1988). Lei Nº 8.666, de 21 de Junho de 1993. Regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da Administração Pública e dá outras providências. **Diário Oficial da União**. Brasília, DF, 21 de junho de 1993.

BRASIL. Constituição (1988). Lei Nº 10.520, de 17 de Julho De 2002. Institui, no âmbito da União, Estados, Distrito Federal e Municípios, nos termos do art. 37, inciso XXI, da Constituição Federal, modalidade de licitação denominada pregão, para aquisição de bens e serviços comuns, e dá outras providências. **Diário Oficial da União**. Brasília, DF, 17 de julho de 2002.

CABENA P. *et al.* **Discovering data mining: from concept to implementation.** Englewood Cliffs: Prentice Hall, 1998.

CAMPOS, Francisco. As práticas de conluio nas licitações públicas à luz da teoria dos jogos. **Revista Análise Econômica**, Porto Alegre, v. 1, n. 50, p. 185-206, set. 2008.

CARVALHO, Lorena Moreira Seal. **O Cartel na Licitação Pública.** 2018. 115 f. Dissertação (Mestrado) - Curso de Ciências Jurídico-políticas Com Menção em Direito Administrativo, Universidade de Coimbra, Coimbra, 2018.

CÔRTEZ, Sérgio da Costa; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. **Mineração de Dados - Funcionalidades, Técnicas e Abordagens.** 2002. Disponível em: ftp://obaluae.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf. Acesso em: 26 maio 2020.

DI PIETRO, Maria Sylvia Zanella. **Direito Administrativo.** 32. ed. Rio de Janeiro: Forense, 2019.

FAYYAD, U.; SHAPIRO, G. P.; SMYTH P. **From data mining to knowledge discovery in databases.** AI Magazine, 17(3): 37-54, Fall 1996.

FIESP. **Corrupção: custos econômicos e propostas de combate.** São Paulo: Fiesp, 2010.

FOLHA DE S. PAULO. **Operação Lava Jato.** Disponível em: <http://arte.folha.uol.com.br/poder/operacao-lava-jato/>. Acesso em: 15 maio 2020.

FORTINI, Cristiana; MOTTA, Fabrício. Corrupção nas licitações e contratações públicas: sinais de alerta segundo a Transparência Internacional. **Revista de Direito Administrativo & Constitucional**, Belo Horizonte, v. 1, n. 64, p. 93-113, jun. 2016.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa.** 6. ed. São Paulo: Atlas, 2017.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel; BEZERRA, Eduardo. **Data Mining: conceitos, técnicas, algoritmos, orientações e aplicações**. 2. ed. Rio de Janeiro: Elsevier, 2015.

GONÇALVES, Lóren Pinto Ferreira. **Mineração de dados em supermercados: o caso do supermercado "tal"**. 1999. 36 f. Dissertação (Mestrado) - Curso de Administração, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1999.

GRILO JÚNIOR, Tarcísio Ferreira. **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) - Curso de Engenharia de Produção, Centro de Tecnologia, Universidade Federal da Paraíba, João Pessoa, 2010.

HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. New York: Morgan Kaufmann, 2000.

KOTU, Vijay; DESHPANDE, Bala. **Data Science: concepts and practice**. 2. ed. Cambridge: Elsevier, 2019.

MORAIS, Cirilo Max Macedo de. **Proposição De Indicadores Para Investigação De Licitações Por Meio De Técnicas De Reconhecimento De Padrões Estatísticos E Mineração De Dados**. 2016. 146 f. Dissertação (Mestrado) - Curso de Informática Forense e Segurança da Informação, Universidade de Brasília, Brasília, 2016.

MOURÃO, Roberto Nunes. **Mineração de Dados para Previsão de Renda de Clientes com Contas-Correntes Digitais**. 2018. 70 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade de Brasília, Brasília, 2018.

OCDE. **Corruption raises the cost of business, undermines public trust and hampers growth**. 2016. Disponível em: <https://www.oecd.org/corruption-integrity/>. Acesso em: 22 maio 2020.

OCDE. **Diretrizes para combater o conluio entre concorrentes em contratações públicas**. Brasília: Oecd, 2009.

PINHEIRO, André Alves; MELO, Telma Maria Escóssio. **O Portal De Licitações Dos Municípios Do TCM/CE Sob A Luz Do Princípio Da Transparência**. Brasília, 2016.

QUEIROZ, Altamira de Souza. **Algoritmos de inteligência computacional utilizados na detecção de fraudes nas redes de distribuição de energia elétrica**. 2016. 86 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica e Computação, Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, 2016.

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business Intelligence, Analytics and Data Science: a managerial perspective**. 4. ed. Boston: Pearson, 2018.

SILVA, Carlos Vinícius Sarmiento; RALHA, Célia Ghedini. **Detecção de cartéis em licitações públicas com agentes de mineração de dados**. Revista Eletrônica de Sistemas de Informação, Curitiba, v. 10, n. 1, 2011. Disponível em: <https://repositorio.unb.br/bitstream/10482/12315/1/ARTIGO_Deteccao%20Carteis%20Licitacoes.pdf> Acesso: 15 mai. 2020. doi:10.5329/RESI.2011.1001008

SILVA, Marcelino Pereira dos Santos. **skdq!:** uma linguagem declarativa de especificação de consultas e processos para descoberta de conhecimento em bancos de dados e sua implementação. 2002. 113 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco, Recife, 2002.

SOARES JUNIOR, Jair Sampaio; QUINTELLA, Rogério Hermida. Descoberta de conhecimento em bases de dados públicas: uma proposta de estruturação metodológica. **Revista de Administração Pública (RAP)**, Rio de Janeiro, v. 1, n. 1, p. 1077-1107, out. 2005.

SOUZA, Fatima Regina de. **Manual básico de licitação**. São Paulo: Nobel, 1997.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao Data Mining**. Rio de Janeiro: Editora Ciência Moderna, 2009.

TATSCH, Dante Rodrigo Delibero. **Caracterização dos Cartéis no Brasil**. 2012. 36 f. Monografia (Especialização) - Curso de Economia, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2012.

TRANSPARÊNCIA INTERNACIONAL. **Índice de Percepção da Corrupção 2019**. 2019. Disponível em: <https://transparenciainternacional.org.br/ipc/>. Acesso em: 22 maio 2020.

WAZLAWICK, Raul Sidnei. **Metodologia de Pesquisa para Ciência da Computação**. 2. ed. Rio de Janeiro: Elsevier, 2014.

WITTEN, Ian H. *et al.* **Data Mining: practical machine learning tools and techniques**. 4. ed. Cambridge: Elsevier, 2017.

ANEXO I
APÊNDICE ao TCC

Termo de autorização de publicação de produção acadêmica

O estudante Felype Santos Moraes do Curso de Ciência da Computação, matrícula 2016.1.0028.0088-5, telefone: 62 992066089, e-mail felypes56@gmail.com, na qualidade de titular dos direitos autorais, em consonância com a Lei nº 9.610/98 (Lei dos Direitos do Autor), autoriza a Pontifícia Universidade Católica de Goiás (PUC Goiás) a disponibilizar o Trabalho de Conclusão de Curso intitulado Mineração de dados aplicada a comportamento de licitações gratuitamente, sem ressarcimento dos direitos autorais, por 5 (cinco) anos, conforme permissões do documento, em meio eletrônico, na rede mundial de computadores, no formato especificado (Texto(PDF); Imagem (GIF ou JPEG); Som (WAVE, MPEG, AIFF, SND); Vídeo (MPEG, MWV, AVI, QT); outros), específicos da área para fins de leitura e/ou impressão pela internet, a título de divulgação da produção científica gerada nos cursos de graduação da PUC Goiás.

Goiânia, 15 de junho de 2021

Assinatura do autor:

Felype Santos Moraes

Nome completo do autor: Felype Santos Moraes

Assinatura do professor – orientador:

Sibelius Lellis Vieira

Nome completo do professor – orientador: Sibelius Lellis Vieira