UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE EDUCAÇÃO / CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO NAS ORGANIZAÇÕES APRENDENTES

HUGO MEDEIROS SOUTO

MINERAÇÃO DE DADOS ABERTOS: UMA ANÁLISE DO USO DE BOTS EM PREGÕES ELETRÔNICOS

JOÃO PESSOA
2019

HUGO MEDEIROS SOUTO

MINERAÇÃO DE DADOS ABERTOS: UMA ANÁLISE DO USO DE BOTS EM PREGÕES ELETRÔNICOS

Dissertação apresentada ao Programa de Pós-Graduação em Gestão nas Organizações Aprendentes em cumprimento às exigências para conclusão do Mestrado Profissional em Gestão nas Organizações Aprendentes.

Orientador: Prof. Dr. Wagner Junqueira de Araújo.

JOÃO PESSOA

2019

Catalogação na publicação
Seção de Catalogação e Classificação

S728m Souto, Hugo Medeiros.
Mineração de dados abertos: uma análise do uso de bots em pregões eletrônicos / Hugo Medeiros Souto. - João Pessoa, 2019.
100 f.

Dissertação (Mestrado) - UFPB/CE/CCSA.

1. Governo Aberto. 2. Mineração de Dados. 3. Pregão Eletrônico. 4. Bots. 5. Gestão da Segurança da Informação. I. Título

UFPB/BC

HUGO MEDEIROS SOUTO

MINERAÇÃO DE DADOS ABERTOS: UMA ANÁLISE DO USO DE BOTS EM PREGÕES ELETRÔNICOS

Dissertação apresentada ao Programa de Pós-Graduação em Gestão nas Organizações Aprendentes em cumprimento às exigências para conclusão do Mestrado Profissional em Gestão nas Organizações Aprendentes.

Aprovada em: 28/06/2019.

BANCA EXAMINADORA

Prof. Dr. Wagner Junqueira de Araújo (UFPB)
Orientador

Prof.ª Dra. Sandra Albuquerque Siebra (UFPE)
Membro Externo

Prof. Dr. Guilherme Ataíde Dias (UFPB)
Membro Interno

Prof. Dr. Miguel Maurício Isoni (UFPB)
Membro Suplente

Prof.ª Dra. Julianne Teixeira e Silva (UFPB)
Membro Suplente

AGRADECIMENTOS

Este trabalho não teria sido realizado com tanto empenho e satisfação sem a contribuição de algumas pessoas em especial. Dentre elas e em destaque está a minha família, mãe, pai e irmã, que deram todo o suporte e incentivo possível ao longo de todo o processo. Mas sou especialmente grato à Prof. Dr. Cláudia Maria Ramos Medeiros, que além de seu inestimável papel de mãe, cumpriu novamente com encanto o papel de me ensinar a dar os primeiros passos, agora como pesquisador.

Ao Prof. Dr. Wagner Junqueira de Araújo agradeço pela verdadeira orientação que me deu durante todo o processo. Sua resiliência me ensinou, através do exemplo, a aceitar algumas dificuldades. Sua visão possibilitou abordarmos o que hoje considero o melhor possível. Graças a isso, o entusiasmo com a pesquisa continua crescendo. Espero que esta etapa seja apenas um começo.

Agradeço também aos bem mais que colegas de trabalho, com quem tenho a sorte do conviver na Superintendência Federal de Agricultura, Pecuária e Abastecimento na Paraíba. Vivenciaram, vibraram e me apoiaram a cada passo. Agradeço especialmente a Danilo Losada Japiassú, que deu a segurança de que não haveria prejuízo ao trabalho durante a realização do mestrado. Agradeço também ao Superintendente, Lúcio Aurélio Braga Matos, pela concessão do afastamento parcial das atividades como chefe da Seção de Compras e Contratos.

Sou grato a toda a família Isoni, que contribuiu de diversas formas. Além de acadêmicos e empreendedores inspiradores, tornaram possível a realização de dois cursos em ciência de dados no *software* RapidMiner. Agradeço também ao atencioso professor Jefferson Barbosa e a Augusto Benvenuto, ambos da área de tecnologia da informação Ministério Público da Paraíba, que me ajudaram a superar barreiras que pareciam intransponíveis.

Agradeço também aos demais membros das bancas de qualificação e avaliação final do trabalho. Os comentários minuciosos do Prof. Dr. Guilherme Ataíde Dias contribuíram muito para a melhoria da pesquisa, desde a qualificação até sua conclusão, enquanto a Prof.ª Dra. Sandra Albuquerque Siebra, da UFPE, enriqueceu o trabalho trazendo à avaliação final seu olhar clínico e sua experiência como docente e pesquisadora em ciências da informação e da computação.

Aos demais que participaram de alguma forma e aos leitores, muito obrigado.

RESUMO

O governo federal tem buscado adequar os procedimentos de contratação às evoluções e tendências das tecnologias da informação e comunicação. O pregão eletrônico foi um dos produtos desses esforços, caracterizando-se como uma modalidade que apresentou soluções estruturais para melhorar a eficiência das aquisições de bens e serviços comuns e que representa mais de 94% das licitações ocorridas no país. Apesar dos benefícios decorrentes do formato eletrônico, esse ambiente traz desafios, como o de lidar com a utilização de *bots*, *software* de lançamento automático de lances (ou robôs eletrônicos). Enquanto não há lei que proíba sua utilização, acórdãos do Tribunal de Contas da União afirmam que seu uso propicia vantagem competitiva aos fornecedores que detêm essa tecnologia em questão sobre os demais licitantes, caracterizando uma afronta ao princípio da isonomia. Também no sentido de modernização das compras públicas está a ampliação da transparência através das políticas de dados abertos, como parte do contexto de Governo Aberto e de transformação digital. Este estudo tem como objetivo principal analisar a situação do uso de *bots* em pregões eletrônicos através da mineração de dados abertos. Foram analisados os pregões eletrônicos realizados no âmbito do Ministério da Agricultura, Pecuária e Abastecimento em 2017. Os dados foram obtidos por meio de solicitação pelo Sistema Eletrônico do Serviço de Informações ao Cidadão (e-SIC), tendo sido adotada como metodologia a descoberta de conhecimento em bases de dados. Os resultados apontam que uso de *bots* em pregões eletrônicos em 2017 que representou uma vantagem de mais de 5% em sucesso nos itens de disputa observados para apenas 1,99% dos licitantes da amostra, indicados como suspeitos do uso. O indicador mais relevante para classificar licitantes como suspeitos foi a alta quantidade de lances emitidos em relação ao comportamento observado na amostra. Espera-se que os resultados que fomentem a discussão acerca dos efeitos do uso de *bots* em pregões eletrônicos e que evidencie a necessidade de desenvolvimento das políticas de dados abertos para que a mineração de dados seja um meio cada vez mais eficaz para avaliar anomalias e aumentar a integridade das licitações realizadas pelo Portal de Compras do Governo Federal.

Palavras-chave: Governo Aberto. Mineração de Dados. Pregão Eletrônico. *Bots*. Gestão da Segurança da Informação.

ABSTRACT

Brazil's federal government has sought to match procurement procedures to trends in information and communication technologies. The electronic reverse auction was one of the products of these efforts, being characterized as a modality that presented structural solutions to improve the efficiency of purchases of common goods and services and that represents more than 94% of the bids that occurred in the country. Despite the benefits of electronic format, this environment brings challenges, such as dealing with the use of bots, which works by automatically bidding. While there is no law prohibiting its use, judgments of the Federal Court of Auditors state that its use provides a competitive advantage to suppliers holding this technology in question over other bidders, characterizing an affront to the principle of isonomy. Also in the sense of modernizing public procurement is increasing transparency through open data policies, as part of the context of Open Government and digital transformation. This study aims to analyze the situation of bot use in electronic reverse auctions through open data mining. Electronic reverse auctions held at the Ministry of Agriculture, Livestock and Supply in 2017 were analyzed. Data were obtained by request by the Electronic Information System for Citizen Information (e-SIC), having been adopted as methodology the knowledge discovery in databases. The results indicate that bot use in electronic reverse auctions in 2017 represented a more than 5% advantage in successful bid items observed for only 1.99% of the sample bidders, indicated as suspected use. The most relevant indicator for classifying bidders as suspects was the high number of bids issued in relation to the behavior observed in the sample. Results are expected to foster discussion of the effects of bot use on e-trading and to highlight the need for open data policy development for data mining to be an increasingly effective means to assess anomalies and increase the integrity of the bids made by the Federal Government Procurement Portal.

Keywords: Open Government. Data Mining. Electronic reverse auction. Bots. Information Security Management.

LISTA DE QUADROS

Quadro 1 - Fase externa do pregão eletrônico 31
Quadro 2 - Variáveis nas configurações dos leilões 33
Quadro 3 - Classificação PEAS 36
Quadro 4 - Cinco estratégias comuns de lances 39
Quadro 5 - Formas de fraude em leilões eletrônicos 40
Quadro 6 - Classificação PEAS para *bots* licitantes 42
Quadro 7 - Definições de Governo Eletrônico 47
Quadro 8 - Principais marcos legais relacionados a dados abertos 49
Quadro 9 - Evolução da tecnologia de sistemas de Banco de Dados 53
Quadro 10 - Modelo de processo *The Two Crows* 60
Quadro 11 - Correspondências entre KDD, SEMMA e CRISP-DM 61
Quadro 12 - Dicionário dos dados 64
Quadro 13 - Atributos excluídos e motivos 69
Quadro 14 - Funções para criação de grupos Fibonacci 78
Quadro 15 - Análise descritiva dos atributos finais 82
Quadro 16 - Critérios dos grupos Fibonacci 86

LISTA DE TABELAS

Tabela 1 - Dados sobre pregões realizados entre 2014 e 2018 30
Tabela 2 - Gastos do governo federal por modalidade de licitação entre 2014 e 2018 30
Tabela 3 - Dados do gráfico da Figura 18 88

LISTA DE FIGURAS

Figura 1 - Gastos com compras do governo federal por modalidade de licitação em 2018..... 15
Figura 2 - Etapas do pré-processamento dos dados..... 66
Figura 3 - Subprocesso “Unificar tipos” 68
Figura 4 - Subprocesso “Tratamento inicial” 69
Figura 5 - Subprocesso “Extração horas”..... 71
Figura 6 - Subprocesso “Separa data e hora” 71
Figura 7 - Subprocesso “Decimal para horas” 72
Figura 8 - Subprocesso “lanData_time final”..... 73
Figura 9 - Subprocesso “Extração datas” 74
Figura 10 - Subprocesso “Consolidação tempo”..... 75
Figura 11 - Subprocesso “Diferenciações”..... 77
Figura 12 - Subprocesso “Agregações”..... 79
Figura 13 - Subprocesso “CNPJ” 80
Figura 14 - Subprocesso “Indicador” 81
Figura 15 - Subprocesso “Organização final”..... 82
Figura 16 - Gráfico de Pareto com distribuição de lances por licitante 84
Figura 17 - Gráfico de bolhas com relação entre licitante, indicador e número de lances..... 85
Figura 18 - Agrupamento de intervalo de tempo entre lances nos itens por *indBot* 86
Figura 19 - Gráfico de barras do agrupamento em pregões com todos os licitantes..... 88
Figura 20 - Gráfico de barras empilhadas da relação entre *indBot* e aceitação da proposta 89

LISTA DE SIGLAS E ABREVIATURAS

CG	Compras Governamentais
CGU	Controladoria Geral da União
CNPJ	Cadastro Nacional de Pessoa Jurídica
Comprasnet	Portal de Compras Governamentais
CRISP-DM	*Cross-Industry Standard Process for Data Mining*
DBMS	*Database Management System*
E-gov	*E-government*
e-SIC	Serviço de Informação ao Cidadão
EGD	Estratégia do Governança Digital
ENAP	Escola Nacional de Administração Pública
IA	Inteligência Artificial
INDA	Infraestrutura Nacional de Dados Abertos
INTOSAI	*International Organization of Supreme Audit Institutions*
KDD	*Knowledge Discovery in Databases*
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MBA	*Master in Business Administration*
MPDG	Ministério do Planejamento, Desenvolvimento e Gestão
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
OGP	*Open Government Partnership*
PEAS	*Performance, Environment, Actuators, Sensors*
PF	Polícia Federal
PPA	Plano Plurianual
RDC	Regime Diferenciado de Contratações Públicas
Sefti	Secretaria de Fiscalização de Tecnologia da Informação
SEMMA	*Sample, Explore, Modify, Model, Assess*
SISG	Sistema de Serviços Gerais
SFA	Superintendência Federal de Agricultura
SLTI	Secretaria de Logística e Tecnologia da Informação
TCU	Tribunal de Contas da União
TIC	Tecnologias de Informação e Comunicação
UASG	Unidade Administrativa de Serviços Gerais

SUMÁRIO

1 INTRODUÇÃO 11
1.1 APRESENTAÇÃO DA RELEVÂNCIA DO PROBLEMA 11
1.2 DELIMITAÇÃO DO ASSUNTO E QUESTÃO DE PESQUISA 16
1.3 JUSTIFICATIVA DE ESCOLHA DO ASSUNTO 20
1.4 OBJETIVOS E HIPÓTESES 22
1.5 ESTRUTURA 23

2 PROCEDIMENTOS METODOLÓGICOS 24
2.1 CARACTERIZAÇÃO DA PESQUISA 24
2.2 CAMPO DA PESQUISA 24
2.3 PROCEDIMENTOS DE COLETA E ANÁLISE DOS DADOS 25

3 COMPRAS GOVERNAMENTAIS 27
3.1 LICITAÇÕES TRADICIONAIS 27
3.2 PREGÃO ELETRÔNICO 29
3.3 LEILÕES 32

4 INTELIGÊNCIA ARTIFICAL 34
4.1 INTELIGÊNCIA ARTIFICIAL 34
4.2 BOTS 37
4.3 O USO DE BOTS EM LEILÕES 38
4.4 O USO DE BOTS EM PREGÕES ELETRÔNICOS 40

5 DADOS ABERTOS E TRANSPARÊNCIA GOVERNAMENTAL 47

6 MINERAÇÃO DE DADOS 52
6.1 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS 56
6.2 OUTRAS ABORDAGENS DE EXPLORAÇÃO DE DADOS 60

7 RESULTADOS 63
7.1 BASE DE DADOS 63
7.2 PRÉ-PROCESSAMENTO DOS DADOS 66
7.3 MINERAÇÃO DOS DADOS 75
7.4 AVALIAÇÃO DO CONHECIMENTO (ANÁLISE DOS RESULTADOS) 84

8 CONSIDERAÇÕES FINAIS 90

REFERÊNCIAS 93

1 INTRODUÇÃO

O governo federal tem buscado aprimorar seus procedimentos competitivos de compras e contratações, conhecidos como licitações. São notáveis os esforços, tanto normativos quanto tecnológicos, em busca de atualizar seus modelos de seleção de fornecedores através de inovações nas regras de funcionamento e nos sistemas de informação em que são operados. A criação do pregão (BRASIL, 2002) foi uma das inovações normativas, caracterizando-se como uma modalidade de licitação que apresentou soluções estruturais para aumentar a eficiência dos processos para a aquisição de bens e contratação serviços ditos comuns. Por sua vez, a instituição da versão eletrônica do pregão (BRASIL, 2005a) representou um relevante salto tecnológico, contribuindo para o aumento da participação de fornecedores e a consequente ampliação da competitividade, pois reduziu e até eliminou, na maioria dos casos, a necessidade da presença física dos fornecedores nos órgãos interessados.

Segundo um estudo recente realizado pelo Banco Mundial (OLIVEIRA, FABREGAS MASLLOVET, FAZEKAS, 2019), em 2018, os leilões reversos (realizados predominantemente na forma do pregão eletrônico) foram de longe a forma de compra competitiva (ou licitação) mais utilizada pelos órgãos do governo federal no Brasil. Só o pregão eletrônico representou mais de 94% das licitações ocorridas no país entre 2014 e 2018, somando mais de 93% do volume de recursos financeiros gastos por compras competitivas no período (PAINEL DE COMPRAS, 2019).

Apesar dos benefícios decorrentes do formato eletrônico, esse ambiente traz alguns desafios, como o de lidar com a utilização de *bots*, *software* de lançamento automático de lances (ou robôs eletrônicos). Enquanto não há lei que proíba sua utilização, o Tribunal de Contas da União (TCU), órgão de controle externo do governo federal, afirmou que “o uso de dispositivos de inserção automática de lances, por alguns licitantes, pode comprometer a isonomia entre os participantes do certame” (TCU, 2010a, p. 50). Mais recentemente, o estudo do Banco Mundial (OLIVEIRA, FABREGAS MASLLOVET, FAZEKAS, 2019) ressaltou o aumento da ocorrência de lances de alta frequência como ameaça ao pregão eletrônico, que quase sempre diminuem o preço apenas em centavos, indicando o uso de *bots* como a possível origem do problema.

Ao mesmo tempo em que são observadas inovações tecnológicas relacionadas às compras governamentais, há uma busca pela ampliação da transparência. Segundo a Cartilha Técnica para Publicação de Dados Abertos no Brasil (BRASIL, 2012), desde 2009 o governo brasileiro vem realizando ações para o desenvolvimento de uma política de disseminação de dados e informações governamentais para o livre uso pela sociedade, alinhadas com um crescente movimento global para democratização do acesso a dados e informações no paradigma de dados abertos. Em relação às compras governamentais, a transparência se materializa na forma da disponibilização de informações e dados relacionados a licitações e contratos, incluindo os pregões eletrônicos.

Nesse contexto, o campo desta pesquisa são as compras governamentais, com foco no pregão eletrônico, tendo como objetivo geral analisar a situação do uso de *bots* nos pregões eletrônicos. Para isso, será utilizada a técnica de mineração de dados abertos disponibilizados pelo próprio governo federal, obtidos através de solicitação feita pelo Sistema Eletrônico do Serviço de Informações ao Cidadão (e-SIC), de acesso amplo e gratuito. O *software* utilizado para a mineração foi o RapidMiner, versão 9.3, sob licença educacional.

Serão apresentados a seguir um aprofundamento da contextualização do estudo, a justificativa para sua realização, seus objetivos específicos, os procedimentos metodológicos utilizados e a estrutura da dissertação.

1.1 APRESENTAÇÃO DA RELEVÂNCIA DO PROBLEMA

As compras governamentais afetam a sociedade de várias formas, tais como através da qualidade percebida nos serviços públicos prestados ou na infraestrutura oferecida à sociedade. São determinantes para uma alocação eficiente de recursos públicos, pois representam o momento de negociação dos preços que serão pagos. São também um meio de distribuição de recursos financeiros no mercado através de pagamentos aos fornecedores contratados, principalmente a micro e pequenas empresas. O impacto financeiro das compras governamentais é representativo. De acordo com informações extraídas do Boletim de Compras Públicas (BRASIL, 2018a, p. 6), “no período de janeiro de 2007 a dezembro de

2017, as compras do governo federal de entes que utilizaram o Sistema Comprasnet[1], movimentaram mais de R$ 510 bilhões". Segundo o boletim (BRASIL, 2018a, p. 6), "quando passa-se a comparar as compras [públicas] com o PIB brasileiro, observou-se que, em média, elas representaram 0,96%".

Segundo Amorim (2017, p. 2), o termo "compras governamentais" abrange todas as formas de contratação possíveis à Administração Pública, tais como aquisição de bens, contratação de serviços, alienação de bens móveis e imóveis, concessões de serviços públicos, permissões de uso de bem público, entre outras. Está estritamente relacionado a licitação, que consiste em "certame que as entidades governamentais devem promover e no qual abrem disputa entre os interessados em com elas travar determinadas relações de conteúdo patrimonial, para escolher a proposta mais vantajosa às conveniências públicas" (MELLO, 2015, p. 536). Mello (2015, p. 536) ressalta ainda que "estriba-se na ideia de competição, a ser travada isonomicamente entre os que preencham os atributos e aptidões necessários ao bom cumprimento das obrigações que se propõem assumir". Tem-se então a isonomia, a igualdade nas condições de competição e a impessoalidade como pilares das licitações. Não por acaso, são alguns dos princípios citados no art. 3º da Lei de Licitações (BRASIL, 1993) em que toda o conjunto normativo pertinente a licitações e contratos se baseia e que orientam também a conduta esperada dos agentes públicos no dia a dia.

Peculiaridades inerentes às compras governamentais trazem complexidades operacionais a mais quando comparadas a processos de compras em organizações privadas. Demandam a aplicação rigorosa de procedimentos preestabelecidos por lei nos processos de aquisição de bens e serviços, "ao contrário dos particulares, que dispõem de ampla liberdade quando pretendem adquirir, alienar, locar bens, contratar a execução de obras ou serviços" (MELLO, 2015, p. 536). Tais procedimentos não buscam apenas garantir a integridade das disputas e a eficácia dos resultados, mas objetivam também promover o desenvolvimento sustentável do país, como dito no art. 3º da Lei nº 8.666, de 1993 (BRASIL, 1993), tornando a licitação um meio para alcance de políticas públicas de fomento a economias locais, sustentabilidade ambiental, inclusão social, regularização fiscal, entre outras.

Justen Filho (2012) ressalta outros aspectos da relevância do contrato administrativo, como sendo este o instrumento de formalização dos procedimentos de compras

[1] Portal de Compras do Governo Federal. Disponível em: https://www.comprasgovernamentais.gov.br.

governamentais. O autor traz a noção de que há íntima relação entre a democratização do Estado e a ampliação da utilização do contrato administrativo através da observância de certos procedimentos e atuação dentro de limites específicos, em detrimento à requisição compulsória de bens e serviços por um Estado autoritário, por exemplo (JUSTEN FILHO, 2012). Traz também, sob a égide do princípio da especialização, sua relevância econômica por permitirem que, "o Estado se dedique às atividades essenciais a ele atribuídas pela Nação e promova a contratação de terceiros para obter os bens e serviços de que necessita" (JUSTEN FILHO, 2012, p. 12).

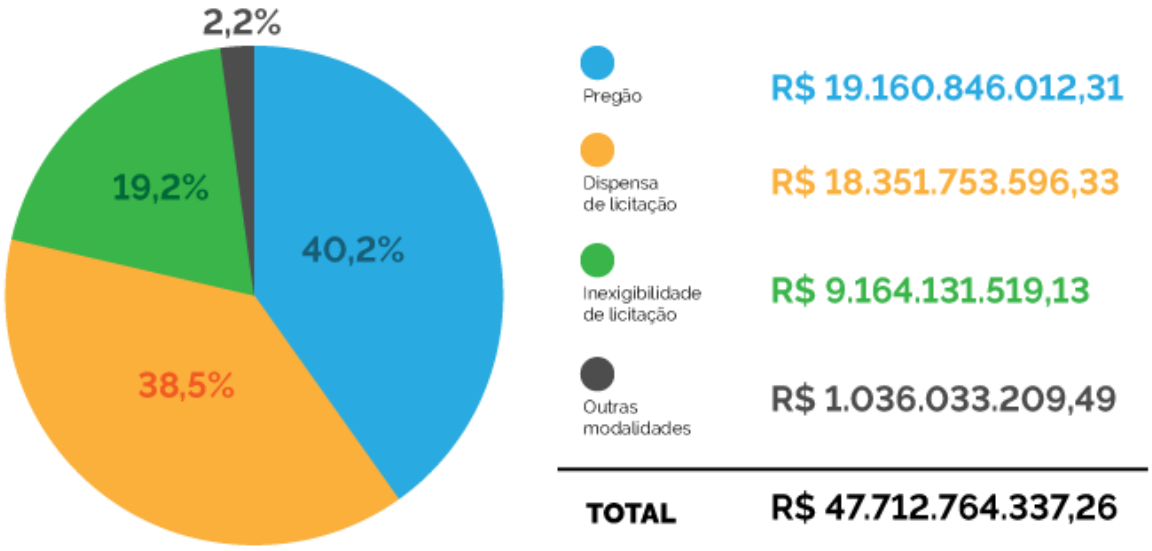
O pregão é uma modalidade de licitação mais recente, criada quase uma década após a publicação da Lei de Licitações (BRASIL, 1993) através da Lei nº 10.520, de 17 de julho de 2002 (BRASIL, 2002), conhecida como Lei do Pregão. Surgiu como uma alternativa de processo licitatório mais eficiente na aquisição de serviços e bens, inicialmente sendo realizada apenas em sua versão presencial. Teve a versão eletrônica regulamentada em 2005 pelo Decreto nº 5.450, de 31 de maio de 2005 (BRASIL, 2005a), quando foi instituída sua obrigatoriedade em licitações que tivessem como objeto a aquisição de bens ou contratação de serviços comuns.

Quanto a caracterização de uma contratação "comum", o art. 1º da Lei do Pregão (BRASIL, 2002) diz que são bens e serviços "cujos padrões de desempenho e qualidade admitam definição objetiva no edital, por meio de especificações usuais de mercado". Segundo Amorim (2017), o que caracteriza os bens e serviços comuns é a sua possibilidade de padronização, devendo ser substituíveis por outros com o mesmo padrão de qualidade e eficiência, com a possibilidade de serem objetivamente definidos em edital por meio de especificações usuais no mercado. Complementa dizendo que o enquadramento de um bem ou serviço como comum dependerá do caso concreto, não se tratando de uma lista taxativa, e conclui que não se aplica às contratações de obras de engenharia (com exceção das caracterizadas como comuns), bem como às locações imobiliárias e alienações em geral (AMORIM, 2017).

Segundo dados divulgados no Portal de Compras (BRASIL, 2019f), dos gastos do governo federal realizados através de licitações, as realizadas via pregão, presencial e eletrônico, corresponderam a R$ 19,1 bilhões (Figura 1), sendo que o formato eletrônico representou 99,71% do valor total (BRASIL, 2019f). No entanto, cabe ressaltar que a licitação, apesar de ser a regra, representou um pouco mais de 42% do gasto do governo

federal com compras governamentais no geral em 2018, que foi de quase R$ 48 bilhões no total (BRASIL, 2019f; PAINEL DE PREÇOS, 2019). Isso porque quase 58% desse total foi realizado através de compras diretas, denominadas de dispensas ou inexigibilidades de licitação (BRASIL, 1993), que consistem em situações especiais de contratação em que a disputa não é possível ou não seria vantajosa devido à própria natureza do objeto a ser contratado. Apesar da representatividade das compras diretas, não serão objeto desse estudo não por serem conhecidamente afetadas por uso de *bots*. Ainda, todas as outras modalidades de licitação somaram apenas pouco mais de 2% do gasto total (Figura 1), o que realça do pregão, especialmente no formato eletrônico, diante das demais em relação a seu impacto na sociedade.

Figura 1 - Gastos com compras do governo federal por modalidade de licitação em 2018



Fonte: Brasil, 2019.

Sendo uma modalidade recente, nascida em 2005 no formato eletrônico, apresentou um crescimento bastante representativo, mas seu sucesso não decorre apenas de seu caráter obrigatório; decorre também dos benefícios e facilidades advindos da operacionalização através da internet. A participação dos licitantes em um pregão presencial (assim como em qualquer outra modalidade não eletrônica) depende da presença física de um representante legal e da entrega de toda a documentação no horário determinado para a realização da sessão, restringindo a disputa a fornecedores locais ou aos de outras localidades que estejam dispostos a pagar para estarem presentes na sessão. Já o pregão eletrônico possibilita a participação de fornecedores de qualquer localidade através de credenciamento prévio perante o provedor do

sistema eletrônico, quando lhe é atribuída chave de identificação e senha para acesso ao sistema eletrônico, conforme o Decreto nº 5.450, de 2005 (BRASIL, 2005a).

Além de facilitar a participação dos fornecedores, o pregão eletrônico mitiga algumas das práticas tradicionais de conluio e fraudes em licitações presenciais, contribuindo para o aumento da confiabilidade das disputas nos interessados. No entanto, o ambiente eletrônico traz novas possibilidades de desequilíbrio da disputa relacionadas à tecnologia utilizada no Comprasnet. Uma das questões tecnológicas conhecidas é o uso de *bots* pelos fornecedores na fase de lances, a etapa competitiva de um pregão eletrônico. Os *bots*, no contexto da internet, são sistemas de computador programados para realizarem um conjunto de tarefas automatizadas (MONAHAN, 2018). Na literatura nacional, o termo "robô eletrônico" é amplamente utilizado como um sinônimo de *bot*. Será adotado o termo original em inglês por trazer consigo significado mais específico, podendo o "robô eletrônico" ser utilizado eventualmente, sem distinção de significado.

Bots são criados e utilizados com diversas finalidades, podendo ter uma atuação funcional benigna e transparente ou uma atuação maliciosa. Distil Networks (2019) afirma que "robôs bons" garantem que as empresas virtuais e seus produtos possam ser encontrados pelos clientes em potencial. Os exemplos incluem rastreadores de mecanismos de pesquisa que, por meio da indexação de conteúdo, ajudam as pessoas correspondendo suas consultas aos conjuntos de sites mais relevantes. Assim, fazem parte do rol de atuação benigna de *bots* executar tarefas repetitivas e enfadonhas a seres humanos, mas em velocidades e volumes maiores do que um homem seria capaz de lidar (DISTIL NETWORKS, 2019). Por outro lado, *bots* malignos atuam na capturam de dados de sites sem permissão a fim de reutilizá-los para obtenção vantagem competitiva, ou ainda realizando atividades criminosas, como fraudes e roubo de dados de valor. Mas e aplicado ao pregão eletrônico, como funciona e quais os possíveis efeitos do uso de bots? É o que vamos discutir a seguir.

1.2 DELIMITAÇÃO DO ASSUNTO E QUESTÃO DE PESQUISA

No pregão eletrônico, os *bots* atuam geralmente na emissão de lances de acordo com a programação definida, como se fossem os próprios representantes das empresas. A questão principal relacionada à sua utilização é que podem tornar a disputa desequilibrada,

favorecendo quem os utiliza em detrimento da capacidade humana limitada de resposta durante a disputa de lances. Ao mesmo tempo, não há norma que a proíba. No entanto, parágrafo único do art. 5º da Lei do Pregão diz que "as normas disciplinadoras da licitação serão sempre interpretadas em favor da ampliação da disputa entre os interessados" (BRASIL, 2005a). Com base nisso, o TCU afirma que a vantagem obtida através do uso de *bots* é considerada como uma afronta ao princípio da isonomia e da igualdade na livre disputa de um pregão eletrônico (TCU, 2011), pois sua adoção não é homogênea, o que afeta diretamente a lógica de funcionamento do pregão eletrônico.

Ao longo dos anos, decorrente das provocações e orientações emitidas pelo TCU, o Comprasnet sofreu implementação de mecanismos que buscam inibir os efeitos do uso de *bots*, mas não foram suficientes para a impedir completamente a prática. Uma das barreiras ao desenvolvimento de mecanismos mais eficazes contra a utilização de *bots* é a dificuldade de detecção, ou seja, de diferenciar lances emitidos por pessoas reais e por robôs. Os lances são indistinguíveis do ponto de vista do registro no sistema, não havendo um atributo que indique se um lance foi emitido por um humano ou um robô, gerando a necessidade de se investigar os indícios para que seja possível diferenciá-los. Essa é a forma a qual o próprio TCU vem utilizando em diversos processos, traduzindo-se na análise dos intervalos de tempo e valor entre lances caso a caso. Quando o padrão é extraordinariamente repetitivo ou aparentemente impossível a uma pessoa real, conclui-se que houve a utilização do *bot*. Tal forma pode ser considerada frágil diante da possibilidade de programação dos *bots* para que simulem a oscilação inerente ao comportamento humano (como a emissão de lances em intervalos aleatórios dentro de um período de tempo predeterminado), o que pode tornar a investigação por indícios ainda mais ineficaz.

O TCU tem atuado como um dos principais atores responsáveis por trazer à tona a questão da amenização de desequilíbrios na competitividade decorrentes do uso de *bots*. No entanto, a revisão bibliográfica feita neste estudo demonstra que as auditorias realizadas pelo órgão são feitas lance a lance e caso a caso, em que são analisados os intervalos de tempo e valor entre lances, como já dito. Uma segunda característica encontrada é que as auditorias são realizadas de forma reativa, decorrentes de denúncias feitas por fornecedores que sentem prejudicados e tem o esclarecimento necessário para que haja suspeita de que o desequilíbrio ocorreu por uso de *bots*. Apesar de eficaz para responder às demandas das auditorias, essa forma de investigação pontual restringe o monitoramento do uso de *bots* a casos específicos e

não consegue entregar uma visão ampla de como essa prática afeta as licitações de forma geral. Assim, encontrar meios que possibilitem um controle mais amplo da situação parece ser importante para podermos enxergar o cenário e quantificar impactos reais

Mais recentemente, nove anos após os primeiros posicionamentos do TCU relacionados aos efeitos negativos do uso de *bots* (TCU, 2010; TCU, 2011), o já citado estudo realizado pelo Banco Mundial (OLIVEIRA, FABREGAS MASLLOVET, FAZEKAS, 2019) ressaltou como ameaça ao pregão eletrônico o aumento da ocorrência de lances de alta frequência, que quase sempre diminuem o preço apenas em centavos. Os autores afirmam que se acredita amplamente que os lances de alta frequência são emitidos por *bots* e que há fortes preocupações de que eles estejam impactando não apenas os preços nos contratos públicos, mas também a competitividade dos certames, por representarem vantagem sobre os licitantes que usam pessoas para enviar suas ofertas. Apontam ainda que o amplo uso de *bots* sobrecarrega a infraestrutura do sistema em que são operados os pregões, dificultando seu funcionamento normal.

Uma forma de investigar essa questão é através do uso da mineração de dados. Segundo Lin (2005), a mineração de dados é um processo que desenha padrões úteis de dados. Busca expor quatro questionamentos fundamentais: (1) quais os dados, (2) quais os padrões, (3) qual a lógica para desenhar os padrões dos dados e (4) como os padrões estão relacionados ao mundo real (sua utilidade). Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996a), à medida que avançamos na digitalização da informação, nossa capacidade de analisar e compreender conjuntos enormes de dados enormes vai ficando muito atrás da capacidade atual de coleta e armazenamento de dados. Uma nova geração de técnicas e ferramentas computacionais é necessária para apoiar a extração de conhecimento útil dos volumes de dados em rápido crescimento. Algumas dessas técnicas e ferramentas fazem parte do campo da mineração de dados e da descoberta de conhecimento em bancos de dados

Assim, a mineração de dados apresenta-se como um meio interessante para a investigação desse tipo de problema por alguns motivos. Primeiro, devido ao Comprasnet ser sistema estruturador da administração pública federal que gera grande parte das compras governamentais em âmbito nacional, o que garante um grande volume de dados com a mesma estrutura. Segundo, devido a essa base de dados ser abrangida por políticas governamentais de dados abertos, o que possibilita amplo acesso por cidadãos interessados. Terceiro, a identificação do uso de *bots* é um problema detectável através da análise fenômenos que

podem ser extraídos de dados, como intervalos de tempo e valor entre lances e a quantidade de lances emitidos por licitante, sendo então adequado à utilização de técnicas e ferramentas de mineração de dados.

Nesse sentido, a utilização da análise dos dados contidos na base do governo federal pode trazer benefícios. O programa de desenvolvimento e implementação de estratégias de governo digital da OCDE (Organização para a Cooperação e Desenvolvimento Econômico) apresenta 12 princípios fundamentais para o êxito, dentre os quais está a criação de uma cultura orientada a dados no setor público (OCDE, 2018). Aponta que os governos devem desenvolver uma cultura de análise e uso de dados públicos que ajudem a prever novas necessidades e tendências e entender como melhorar os processos e dinâmicas existentes.

No Brasil, isso tem sido reforçado através de políticas de dados abertos no âmbito do *e-government (e-gov)*, ou governo eletrônico. Segundo definição da *Open Knowledge Internacional*, organização global sem fins lucrativos focada em fomentar a cultura de dados abertos, dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, exigências que visem preservar sua proveniência e sua abertura (OPEN KNOWLEDGE INTERNACIONAL, 2019). Na prática, esses critérios têm sido satisfeitos pelas entidades governamentais através da criação de portais que disponibilizam conjuntos de dados estruturados (linhas de dados dispostos em colunas pré-determinadas) e organizados sob uma licença aberta.

O governo federal, através do agora denominado Ministério da Economia, vem expandindo e aprimorando suas estratégias de governança digital, incluindo ações voltadas à ampliação da transparência e à abertura dos dados. Publicou em 2016 a Estratégia do Governança Digital (EGD) (BRASIL, 2016a) para o período de 2016 a 2019, alinhado com a Política de Governança Digital, instituída pelo Decreto nº 8.638, de 15 de janeiro de 2016 (BRASIL, 2016b) e com o Plano Plurianual (PPA) 2016-2019 (BRASIL, 2016a). O documento é composto por três eixos, dez objetivos e 51 iniciativas que visam integrar a atuação governamental na vida das pessoas através da tecnologia, sendo apontado como um dos desafios a serem enfrentados o de "mitigar as vulnerabilidades de segurança nos sistemas de informação governamentais" (BRASIL, 2016a, p. 12).

A transparência, a segurança e a integridade das informações que se encontram nos sistemas do governo são aspectos fundamentais para que todos os esforços em governança

digital se transformem em confiabilidade para a sociedade. Destacam-se tanto que uma das iniciativas da EGD 2016-2019 (a IE.03.06) é "promover cooperação nacional e internacional com setor produtivo e academia, visando à troca de experiências e o fortalecimento dos temas de Segurança da Informação e Comunicação e de Segurança Cibernética" (BRASIL, 2016, p. 29), indicando ser um dever de todos, como cidadãos, contribuir para a promoção da consolidação do Governo Eletrônico.

Diante do exposto, os dados abertos podem e devem se tornar fonte de análise e pesquisa em compras governamentais, pois apresenta alto potencial de controle social na aplicação de grandes volumes financeiros em despesas públicas. Partindo desse pressuposto, abordando especificamente o pregão eletrônico devido a sua relevância no âmbito das compras governamentais e diante do cenário de possível ofensa à isonomia das licitações, nasceu a seguinte questão, que norteou o presente estudo: **é possível identificar os efeitos do uso de *bots* em pregões eletrônicos utilizando dados abertos?**

1.3 JUSTIFICATIVA DE ESCOLHA DO ASSUNTO

A escolha do assunto tem como motivação pessoal do pesquisador a intenção de participar e de contribuir com a modernização da área de licitações e contratos administrativos através da aplicação de técnicas e de tecnologias que possam melhorar a integridade do sistema Comprasnet e a qualidade das despesas realizadas através das licitações. Atua como pregoeiro no governo federal há mais de seis anos e possui formação como *Master Business Administration* em inovação, pontos que convergem os interesses nas compras governamentais e no uso de tecnologias inovativas, como a mineração de dados, a aprendizagem de máquina e a inteligência artificial. A motivação do orientador converge sua experiência de 15 anos trabalhando no desenvolvimento de sistemas no SERPRO com o projeto de pesquisa cujo tema é "Implicações na aplicação da gestão da segurança da informação nas organizações", em que atua como coordenador.

Traz como inovação a aplicação e disponibilização de método replicável para a identificação de padrões e anomalias em pregões eletrônicos com dados fornecidos pelo governo federal. Traz como benefícios à sociedade evidenciar que é possível para cidadãos comuns (com acesso a recursos tecnológicos similares e conhecimento suficiente) exercer

controle social de forma autônoma, viabilizando a investigação contínua sobre os efeitos e impactos reais do uso *bots* e de outros fenômenos observados no Comprasnet.

É também relevante ressaltar o momento de transformação em que a pesquisa é realizada. Segundo as informações divulgadas no Portal de Compras Governamentais (BRASIL, 2018b), o sistema Comprasnet vem sendo modernizado através de um investimento de aproximadamente 11 milhões de reais. Será adotado o uso de novas tecnologias, como a "aprendizagem de máquina (*machine learning*), que consiste na estruturação de algoritmos capazes de tomar a melhor decisão possível em uma dada situação a partir da análise de dados" (BRASIL, 2018b).

Mais recentemente, o Decreto nº 10.024, de 20 de setembro de 2019 (BRASIL, 2019a) revogou o Decreto nº 5.450, de 31 de maio de 2005 (BRASIL, 2005), trazendo inovações resultantes de ampla discussão com as partes envolvidas, incluindo os próprios fornecedores licitantes e os agentes do governo que atuam diretamente na realização dos certames, sobre os quase 15 anos de experiência com o pregão eletrônico (ENAP, 2019). No discurso de abertura do workshop realizado pela ENAP (Escola Nacional de Administração Pública) sobre o novo decreto, é mencionado que a norma já nasce com expectativa de vida bastante curta devido à expectativa de que seja aprovada em breve a nova lei de licitações, fato que ensejaria a edição de outra regulamentação do pregão eletrônico, alinhada às novas regras (ENAP, 2019).

Uma das principais mudanças nas regras de funcionamento do que podemos chamar de "novo pregão eletrônico" foi a alteração do encerramento aleatório, momento da sessão do certame em que os *bots* mais se mostravam úteis aos licitantes que utilizassem essa tecnologia. O encerramento aleatório, que antes estava presente em todos as sessões de pregão eletrônico (pois existia apenas uma forma de realização), agora estará presente em apenas um dos dois modos de disputa (BRASIL, 2019a) e não será mais a última etapa da sessão, pois será sucedida de um último momento em que, de modo geral, serão convocados pelo sistema apenas o autor da oferta mais vantajosa e aqueles com valores até 10% superiores a essa oferta para que ofertem um lance final e fechado (BRASIL, 2019e). No entanto, não se espera que as novas regras eliminem o uso de *bots* (ENAP, 2019), sendo então relevante que as pesquisas relacionadas ao assunto continuem para que sirvam de base para o delineamento de novas regras que podem vir a surgir.

Diante desse contexto, este estudo busca contribuir, direta ou indiretamente, com a discussão relacionada à evolução das normas e do ambiente tecnológico em busca do

desenvolvimento da integridade do pregão eletrônico através da: a) publicação dos resultados obtidos; b) da disponibilização e divulgação dos procedimentos utilizados no estudo, para que a abrangência do estudo possa ser ampliado através da replicação por qualquer interessado na sociedade; e c) da continuidade e aprofundamento em novas pesquisas.

1.4 OBJETIVOS E HIPÓTESES

Este estudo buscou então investigar a questão de pesquisa com o objetivo geral de **analisar a situação do uso de *bots* em pregões eletrônicos através da mineração de dados abertos.** A investigação teve como **objetivos específicos**: a) identificar comportamentos anômalos que possam indicar o uso de *bots*; b) analisar se o uso de *bots* em pregões eletrônicos oferece vantagem competitiva; c) avaliar a execução das políticas de dados abertos observadas.

As hipóteses que nortearam a busca por evidências se basearam no pressuposto de que seria possível identificar a utilização de *bots* e seus possíveis efeitos através da identificação dos seguintes comportamentos anômalos, que consistiram em:

a) **alta quantidade de participação em pregões eletrônicos**, como hipótese de consequência da redução do esforço para emissão de lances em relação aos licitantes que não utilizam *bots*;
b) **alta frequência de lances emitidos por item que participa**, como hipótese de consequência da rapidez e da reatividade autônoma a outros lances;
c) **alta frequência de participação simultânea em vários itens de um mesmo pregão**, como efeito da ampliação da capacidade de operação simultânea em relação aos licitantes que não utilizam *bots*;
d) **habitualidade no registro de lances com intervalos similares de tempo e/ou valor em relação a novos menores lances de cada item**, como hipótese de consequência da programação para automação do *software*;
e) **alto índice de sucesso nas disputas em relação ao número de itens que participa**, como hipótese de consequência de suposta vantagem competitiva em relação aos licitantes que não utilizam *bots*.

1.5 ESTRUTURA

A primeira seção constituiu-se da Introdução, compondo-se pela contextualização, justificativa e objetivos geral e específicos.

A segunda seção traz a metodologia utilizada, com a caracterização da pesquisa, campo da pesquisa, procedimentos de coleta e análise dos dados e as hipóteses a serem abordadas.

A terceira seção discorre sobre a temática das compras governamentais, discute sobre a relevância da manutenção dos princípios que as regem, apresentando a legislação base e sua evolução até o Pregão Eletrônico, modalidade que será objeto da investigação. Abordará também brevemente os leilões, pois é um campo relacionado que concentra estudos relacionados a *bots*.

A quarta seção aborda a temática da inteligência artificial e dos *bots*, apresentando conceituações, histórico e sua relação com leilões e o pregão eletrônico, discutindo a ética de sua utilização nas disputas.

A quinta seção apresenta aspectos relacionados a Governo Eletrônico e Governo Aberto, trazendo o percurso histórico até o cenário atual, que possibilita a realização dessa pesquisa através da disponibilização de dados abertos referentes às compras governamentais.

A sexta seção conceitua a mineração de dados e justifica a escolha da metodologia da descoberta de conhecimento em bases de dados.

A sétima seção aborda a execução da descoberta de conhecimento na base de dados obtida, apresentando os procedimentos realizados, análise dos dados e a avaliação dos resultados.

Por fim, as considerações finais e as referências utilizadas no trabalho.

2 PROCEDIMENTOS METODOLÓGICOS

Nesse capítulo será apresentada a metodologia utilizada na pesquisa, incluindo as técnicas e ferramentas que adotadas no tratamento e na análise dos dados.

2.1 CARACTERIZAÇÃO DA PESQUISA

Matias Pereira (2016) afirma que a metodologia científica empírico-analítica de pesquisa ocorre quando há utilização em comum de técnicas de coleta, tratamento e análise de dados quantitativos. Segundo o autor,

> Possui como característica privilegiar os estudos práticos, visto que suas propostas possuem caráter técnico, restaurador, incrementalista e forte preocupação com a relação causal entre variáveis. O processo de validação da prova científica é feito por meio de testes dos instrumentos, graus de significância e sistematização das definições operacionais. (MATIAS-PEREIRA, 2016, p. 85-92).

A pesquisa apresenta as características apontadas pelo autor. Utiliza como fonte dados quantitativos estruturados na busca de relações entre as variáveis disponíveis. Trata-se de uma pesquisa de natureza descritiva, pois busca revelar a situação atual do uso de *bots* em pregões eletrônicos de uma forma objetiva. É aplicada a um estudo de caso, pois aborda uma situação real, situada nos pregões eletrônicos realizados no âmbito do MAPA (Ministério da Agricultura, Pecuária e Abastecimento) no ano de 2017.

2.2 CAMPO DA PESQUISA

Segundo dados divulgados no Portal de Compras (BRASIL, 2019f), dos gastos do governo federal realizados através de licitações, as realizadas via pregão, presencial e eletrônico, corresponderam a R$ 19,1 bilhões (Figura 1), sendo que o formato eletrônico representou 99,71% do valor total (BRASIL, 2019f). No entanto, cabe ressaltar que a licitação, apesar de ser a regra, representou um pouco mais de 42% do gasto do governo federal com compras governamentais no geral em 2018, que foi de quase R$ 48 bilhões no total (BRASIL, 2019f; PAINEL DE PREÇOS, 2019). Isso porque quase 58% desse total foi realizado através de compras diretas, denominadas de dispensas ou inexigibilidades de

licitação (BRASIL, 1993), que consistem em situações especiais de contratação em que a disputa não é possível ou não seria vantajosa devido à própria natureza do objeto a ser contratado. Apesar da representatividade das compras diretas, não serão objeto desse estudo não por serem conhecidamente afetadas por uso de *bots*. Ainda, todas as outras modalidades de licitação somaram apenas pouco mais de 2% do gasto total (Figura 1), o que realça do pregão, especialmente no formato eletrônico, diante das demais em relação a seu impacto na sociedade.

O estudo se restringiu aos pregões eletrônicos da esfera federal realizados no âmbito do MAPA em 2017, realizados na plataforma Comprasnet. A escolha do pregão eletrônico em detrimento das outras modalidades se justifica por ser a mais utilizada atualmente e pelo volume de recursos por ela destinados, representando 99,71% do valor total gasto através de licitações do governo federal (BRASIL, 2019f). A restrição à esfera federal se justifica pela centralização dos dados em um sistema único, o Comprasnet. A escolha do período se justifica por ter sido 2017 o último ano completo até do início da pesquisa, quando foi feita a coleta de dados. A escolha do MAPA se justifica por ser o órgão de trabalho do pesquisador, cujo incentivo dado à presente qualificação torna profícuo o retorno à instituição, sendo uma importante pasta do governo federal, cuja estrutura e volume de licitações pode representar a de uma realidade nacional. Foi o nono ministério na quantidade de pregões e o décimo sexto no volume de gastos realizados por pregões eletrônicos em 2017 (PAINEL DE COMPRAS, 2019).

2.3 PROCEDIMENTOS DE COLETA E ANÁLISE DOS DADOS

A base de dados que será utilizada para a realização dos estudos propostos foi obtida através de solicitação dos dados dos lances dos pregões eletrônicos realizados no âmbito do MAPA em 2017 (administração direta, autarquias e fundações). Os dados têm como fonte a plataforma digital Comprasnet, módulo do SIASG, onde são divulgados e operacionalizados os pregões, demais licitações e divulgadas as contratações diretas.

A solicitação foi realizada através do e-SIC (Sistema Eletrônico do Serviço de Informação ao Cidadão), mantido pelo Ministério da Transparência e CGU (Controladoria Geral da União). Foi registrada sob o número de protocolo 03950000211201839 em 17/01/2018, tendo sido respondida a tempo em 26/01/2018, com o envio de planilha

eletrônica gravada em mídia ótica. O arquivo contendo a base de dados utilizados na pesquisa, denominada originalmente por “DadosLanceMapa.csv”, apresenta tamanho de 2,12 GB, com 6.124.274 linhas, correspondendo aos registros de lance emitido em pregões eletrônicos realizados no âmbito do MAPA em 2017.

A técnica específica de análise dos dados será a mineração dos dados abertos disponibilizados pelo próprio governo federal, que foram manipulados e processados através do *software* RapidMiner Studio, versão 9.3.001 (*rev.*: 289b20, *plataform*: WIN64), *Educational Edition*, registrado com licença educacional de 1 ano. Para a visualização dos dados e elaboração de gráficos foi também utilizado o *software* Tableau Desktop, *Professional Edition*, versão 2019.2.1 (compilação 20192.19.0621.1547, 64 bits), também registrado com licença educacional de 1 ano. O ambiente computacional foi composto pelas seguintes características: Sistema Operacional Microsoft Windows 10 *Home Single Language*, sistema PC baseado em X64, processador Intel Core i5-5200U (CPU em 2.20GHz, 2201 Mhz, 2 núcleos, 4 processadores lógicos), com 8,00 GB de memória física (RAM) instalada.

Foi realizada revisão teórica com os objetivos de: a) fundamentar os conceitos abordados na pesquisa; b) contextualizar o problema abordado; c) encontrar os caminhos iniciados por outros autores, tanto na literatura comum quanto na científica. Inicialmente aborda os temas de governo eletrônico e de governo aberto, por representarem ações que permitem que essa pesquisa e outras possam ser implementadas através da utilização de dados abertos como fonte. Em seguida, o pregão eletrônico é posicionado no contexto das licitações e em sua importância para a sociedade. A terceira seção traz um panorama da utilização de *bots* em leilões, onde é historicamente um problema conhecido e amplamente debatido e combatido, e em pregões eletrônicos, indicando as trilhas de investigação que guiarão o estudo.

3 COMPRAS GOVERNAMENTAIS

As compras governamentais definem o escopo onde será realizada essa pesquisa. Da perspectiva gerencial são um conjunto de ferramentas da área de logística dos órgãos públicos utilizadas para adquirir os bens e serviços necessários ao funcionamento dos órgãos. No entanto, diferentemente da logística aplicada às organizações privadas, onde o critério de escolha de cada solução depende de apenas da decisão de cada organização, a pública deve obedecer rigorosamente às normas e princípios que a rege. A fim dar maior precisão ao objeto de estudo, o termo "Compras Governamentais" (ou CG) poderá ser utilizado para se referir apenas às aquisições e contratações de bens e serviços comuns, espécies mais usuais e as únicas permitidas para serem licitadas por pregão eletrônico.

3.1 LICITAÇÕES TRADICIONAIS

As licitações têm origem jurídica na Constituição Federal de 1988, em seu art. 37, inciso XXI, onde diz que:

> XXI - Ressalvados os casos especificados na legislação, as obras, serviços, compras e alienações serão contratados mediante processo de licitação pública que assegure igualdade de condições a todos os concorrentes, com cláusulas que estabeleçam obrigações de pagamento, mantidas as condições efetivas da proposta, nos termos da lei, o qual somente permitirá as exigências de qualificação técnica e econômica indispensáveis à garantia do cumprimento das obrigações. (BRASIL, 1988).

Tal dispositivo é regulamentado pela Lei nº 8.666, de 21 de junho de 1993 (BRASIL, 1993), que estabelece as regras gerais sobre a obrigatoriedade da licitação em seu art. 2º:

> Art. 2º As obras, serviços, inclusive de publicidade, compras, alienações, concessões, permissões e locações da Administração Pública, quando contratadas com terceiros, serão necessariamente precedidas de licitação, ressalvadas as hipóteses previstas nesta Lei.
>
> Parágrafo único. Para os fins desta Lei, considera-se contrato todo e qualquer ajuste entre órgãos ou entidades da Administração Pública e particulares, em que haja um acordo de vontades para a formação de vínculo e a estipulação de obrigações recíprocas, seja qual for a denominação utilizada.

Ressalvadas as hipóteses legais, é através das licitações que são selecionados os fornecedores para as contratações na esfera pública, seja para aquisição de bens, fornecimento de serviços ou realização de obras. Os recursos que custeiam essas despesas são públicos e

executados através de agentes de governo, por isso é necessário que sejam estabelecidos princípios e regras que busquem garantir a integridade e finalidade desses processos.

O art. 3º da Lei nº 8.666/93 expressa alguns dos princípios das licitações:

> 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, 1993, grifos nossos).

Esse dispositivo estabelece o funcionamento ideal das licitações. A base fundamental é o princípio da legalidade, pois este, segundo Justen Filho (2012, p. 72), "disciplina integralmente a atividade administrativa". O autor afirma que "o princípio da legalidade significa ser vedado à autoridade administrativa adotar qualquer providência ou instituir qualquer restrição sem autorização legislativa" (JUSTEN FILHO, 2012, p. 72). Traduz-se na afirmação de que a administração pública só pode (ou deve) fazer o que está expresso em lei.

Os princípios da impessoalidade, da moralidade, da igualdade e da publicidade estão estreitamente relacionados ao da isonomia, fundamento que percorre todos os aspectos das licitações e que sua defesa se caracteriza como o objetivo maior deste estudo. A impessoalidade tem a função principal de evitar direcionamentos e desvios na aplicação dos recursos públicos. Em sua função mais notória, busca dissociar os resultados das licitações das subjetividades dos agentes públicos envolvidos, tendo ou não má fé envolvida.

Aprofundando a discussão sobre os princípios da isonomia e da igualdade, Justen Filho (2012) ressalta que é inválida a discriminação que não retrate uma diferença efetiva no mundo real. Mello (2015, *apud* JUSTEN FILHO, 2012, p. 59) afirma que "o princípio da igualdade consiste em assegurar regramento uniforme às pessoas que não sejam entre si diferenciáveis por razões lógicas e substancialmente (isto é, em face da Constituição) afinadas com eventual disparidade de tratamento".

A Lei nº 8.666/1993 (BRASIL, 1993) instituiu cinco modalidades, concorrência, tomada de preços, convite, concurso e leilão, sendo o pregão instituído pela Lei nº 10.520/2002 (BRASIL, 2002). Segundo Justen Filho (2012), na Lei de Licitações (BRASIL, 1993) há três modalidades "comuns", a concorrência, a tomada de preços e o convite, que podem ser utilizadas em qualquer tipo de contratação e que possuem estrutura de procedimentos definida. Já o concurso e o leilão utilizados apenas em situações especiais,

sendo adaptados a cada situação, tais como leilão de veículos, de joias, ou um concurso literário ou para escolha de um projeto de engenharia para a construção de um edifício público, por exemplo.

O critério de escolha na utilização das modalidades comuns é o valor total da contratação. O art. 23 da Lei nº 8.666/1993 (BRASIL, 1993) institui os limites máximos para a utilização de cada modalidade, não havendo limite mínimo. Assim, uma modalidade mais complexa, como a concorrência ou tomada de preços, pode ser utilizada na contratação de um objeto cujo valor esteja enquadrado no limite para convite, ou a concorrência para um que se enquadre na tomada de preços, mas não o contrário.

Nesse sentido, o pregão se diferencia por não haver limitação quanto ao valor da contratação, bastando apenas que os bens e serviços sejam comuns, ou seja, que "seus padrões de desempenho e qualidade possam ser objetivamente definidos pelo edital, por meio de especificações usuais no mercado", como diz o art. 1º, *caput* e § 1º, Lei nº 10.520/2002 (BRASIL, 2002).

Existe também o Regime Diferenciado de Contratações Públicas (RDC), instituído pela Lei nº 12.462, de 4 de agosto de 2011 (BRASIL, 2011a). Trata-se de um modelo de contratação oferecem uma variedade de configurações, propiciando flexibilidade para o gestor buscar atingir mais eficiência e qualidade do que os regimes tradicionais.

Modalidade não deve ser confundido com tipo de licitação, que pode ser menor preço, melhor técnica, técnica e preço e maior lance ou oferta – exceto para a modalidade concurso, que é operado sob uma espécie de melhor técnica específica, mais objetiva.

3.2 PREGÃO ELETRÔNICO

O pregão é uma modalidade de licitação resultado de uma inovação bem posterior à publicação da Lei de Licitações (BRASIL, 1993). Instituído pela Lei nº 10.520, de 17 de julho de 2002 (BRASIL, 2002), teve como objetivo aumentar a eficiência das licitações nas aquisições de bens e serviços comuns, que "poderiam ter seus padrões e desempenho e qualidade objetivamente definidos em edital" (BRASIL, 2002), inicialmente realizado apenas da forma presencial. Entretanto, foi a forma eletrônica operacionalizada via internet, regulamentada pelo Decreto nº 5.450, de 31 de maio de 2005 (BRASIL, 2005a), que trouxe uma popularização expressiva da modalidade.

Entre 2014 e 2018, segundo dados obtidos através do Painel de Compras do Governo Federal (2019), o formato eletrônico representou 99,58% (110.667) dos 111.132 pregões realizados no período, representando 98,83% (R$ 120,9 bilhões) dos R$ 122,3 bilhões registrados (Tabela 1).

Tabela 1 - Dados sobre pregões realizados entre 2014 e 2018

Modalidade de Compras	Quantidade	%	Valor das Compras	%
Pregão Eletrônico	110.667	99,58%	120.901.701.431,63	98,83%
Pregão Presencial	465	0,42%	1.431.032.143,58	1,17%
Pregão	**111.132**	**100,00%**	**122.332.733.575,21**	**100,00%**

Fonte: Painel de Compras Governamentais, 2019.

Quando comparado com as demais modalidades, o pregão eletrônico representou 95,14% do total, somando um total de R$ 120, 9 bilhões (95,14%) de todo o gasto do governo federal através de licitações no mesmo período, que foi de R$ 131,7 bilhões, conforme demonstrado na Tabela 2.

Tabela 2 - Gastos do governo federal por modalidade de licitação entre 2014 e 2018

Modalidade de Compras	Quantidade	%	Valor das Compras	%
Pregão Eletrônico	110.667	95,14%	120.901.701.431,63	91,80%
Pregão Tradicional	465	4,86%	1.431.032.143,58	8,20%
Tomada de Preços	2.495	2,5%	955.440.134,05	13,2%
Concorrência	1.947	2,6%	8.084.638.026,13	0,8%
Convite	514	0,1%	30.434.548,10	0,2%
Concorrência Internacional	125	0,6%	275.589.810,18	0,0%
Concurso	109	0,1%	26.626.676,78	0,0%
Total	**116.322**	**100,00%**	**131.705.462.770,45**	**100,00%**

Fonte: Painel de Compras Governamentais, 2019.

Considerando ainda a soma dos gastos dos estados, municípios e distrital, as licitações de fato têm grande importância na qualidade do uso dos recursos públicos. Essa qualidade depende de alguns fatores, dos quais podemos citar alguns: capacitação dos agentes de compras e alocação de adequada de pessoal, fluidez da rede legislativa e normativa, estrutura disponibilizada pelos órgãos e a usabilidade dos sistemas. Além desses fatores diretamente relacionados aos processos, a segurança e integridade das informações são especialmente importantes no meio de licitações, historicamente afetadas por fraudes e corrupção.

O procedimento do pregão eletrônico, regulamentado pelo Decreto nº 5.450 de 2005 (BRASIL, 2005a), é composto por duas fases, interna e externa. A fase interna corresponde à

fase preparatória e abrange a identificação da demanda, elaboração de todos os documentos necessários ao detalhamento da contratação e das regras do pregão propriamente dito. A fase externa tem como núcleo em si, que ocorre na chamada sessão pública, liderada pelo pregoeiro e realizada em conjunto com uma equipe de apoio, todos devidamente nomeados por alguma autoridade regimentalmente competente. É então iniciado um processo de várias etapas, conforme descrito no Quadro 1, que podem simultaneamente ou não, pois os itens, em regra, são independentes entre si.

Quadro 1 - Fase externa do Pregão Eletrônico

Módulo/Fase	Procedimentos e Ações Possíveis
Operar	**Início**: A partir do horário previsto no edital, tem início a Sessão Pública, quando é disponibilizado chat de comunicação com os fornecedores para mensagem de abertura da Sessão Pública, que permanece disponível ao Pregoeiro por toda a sessão; **Abrir item para lances:** é realizada a análise, classificação e desclassificação das propostas para cada item. Em seguida o item é aberto para lances. **Suspender Item**: opção disponível na etapa de lances e poderá ser utilizada em situações de dúvida sobre um determinado item. A suspensão do item impedirá o envio de lances enquanto perdurar tal situação; **Excluir Lance**: tal procedimento não impedirá a continuidade do envio de lances pelos fornecedores; **Encerrar o Item para Lances**: um item encerrado não pode ser reaberto. Ao clicar em encerrar item, o pregoeiro deve informar o tempo de iminência, entre 1 e 60 minutos, em formato numérico. Após esse tempo, os itens entram no horário de encerramento aleatório do sistema no período de 1 a 30 minutos. Após o encerramento aleatório, os itens passam a constar na situação de "encerrado", devendo o pregoeiro ir para a fase de aceitação.
Aceitar	Nesta fase, o pregoeiro analisa as propostas de melhor lance, relativamente ao objeto e o valor, decidindo por sua aceitação ou recusa.
Habilitar	Destina-se a fornecedores que tiveram suas propostas aceitas para o item.
Fechar/Reabrir Prazo de Intenção de Recurso	Nesta funcionalidade o pregoeiro informa o prazo para os fornecedores manifestarem a intenção de recurso.
Juízo de Admissibilidade	Nesta fase o pregoeiro deve acatar ou recusar a intenção de recurso do licitante. Havendo intenção de recurso aceita, o sistema disponibiliza tela própria para que o pregoeiro informe os prazos para registro da razão, da contrarrazão e da decisão do recurso.
Encerrar Sessão	O sistema eletrônico gera a Ata da Sessão Pública do pregão, com indicação do lance vencedor e demais informações relativas à Sessão Pública.
Decidir Recurso	Caso tenha sido aceito alguma intenção de recurso, findos os prazos legais para registro das razões do recurso e das contrarrazões dos licitantes, o pregoeiro informa à autoridade competente da sua decisão, que acatará ou não, emitindo sua decisão final.
Adjudicar	Decididos os recursos e constatada a regularidade dos atos praticados, a autoridade competente adjudicará o procedimento licitatório.
Homologar	É um procedimento da autoridade competente que equivale à decisão final e/ou à ratificação dos atos do pregoeiro.

Fonte: Adaptado de Brasil, 2005b.

3.3 LEILÕES

Este estudo busca investigar fraudes em pregão eletrônico, modalidade de licitação que pode ser considerada como um leilão do tipo reverso. Existe uma trilha de décadas de estudos sobre leilões abordando o funcionamento e os comportamentos indevidos, com recentes e importantes contribuições sobre fraudes em leilões eletrônicos em plataformas globais populares como o Ebay, criado em 1995 e que, segundo Hubbard (2016), continua sendo o leilão dominante da internet. Enquanto isso, no pregão eletrônico do governo brasileiro, por ter sido legalmente criado em 2005, as fraudes começaram a ser estudados apenas nos últimos anos. Por isso as bases teóricas da investigação serão predominantemente os estudos relacionados a leilões adaptados à sistemática e ao cenário da ferramenta em foco.

O leilão tem sido usado desde a antiguidade para a venda diversas coisas, havendo relatos de que era utilizado já na Babilônia, em 500 A.C. (KRISHNA, 2009). Os estudos relacionados à sua dinâmica tiveram origem com William Vickrey (1961) na teoria dos jogos, especificamente sobre incentivos sob informação assimétrica, trabalho que rendeu o reconhecimento da Academia Real Sueca de Ciências com o Prêmio Nobel em Ciências Econômicas de 1996 (NOBEL FOUNDATION, 2018).

Hubbard (2016) afirma que sua antiguidade reflete a robustez e a flexibilidade desse mecanismo de venda. Segundo o autor, sua robustez é garantida sistema de regras em que funciona:

> Para que o leilão funcione bem, as regras que determinam o vencedor, bem como quanto esse licitante irá pagar, devem ser estabelecidas com antecedência. Além disso, os potenciais licitantes devem estar cientes dessas regras e entendê-las. Mais importante, o leiloeiro se compromete a não alterar nenhuma das regras. O leiloeiro é um intermediário que coleta informações dos licitantes na forma de lances que refletem sua disposição em pagar pelo objeto à venda. (HUBBARD, 2016, p. 4, tradução nossa)

A flexibilidade do sistema se deve em muito à diversidade dos tipos de leilão, que se moldam às necessidades do mercado, seja atendendo primariamente aos vendedores ou aos compradores. Ainda segundo o autor, quatro tipos de configuração (formados pelas combinações de formato das propostas e regras de precificação) e suas variáveis (combinações das regras de pagamento, formato de realização e da plataforma) são responsáveis pela maioria dos leilões encontrados hoje no mercado (Quadro 2).

Quadro 2 - Variáveis nas configurações dos leilões

Configuração		**Variáveis**	
Formato das Propostas	**Regra de Precificação**	**Regra para Pagamento**	**Formato de Realização e da Plataforma**
Aberto	Preço ascendente (tipo inglês)	Primeiro preço	Presencial (Tradicional ou Eletrônico)
Lance fechado	Preço descendente (tipo holandês)	Segundo preço	Virtual (Eletrônico)

Fonte: Adaptado de Hubbard, 2016.

Existem basicamente dois formatos de leilão, o aberto, em que as alterações de preço são visíveis para todos os participantes durante a sessão, e o lance fechado, também chamado de proposta fechada, em que cada participantes apresenta lance único. Outra variável é a regra de precificação estabelecida, que pode ser ascendente, que corresponde ao modelo mais tradicional de leilão e é chamado de inglês, ou de preço descendente, conhecido como tipo holandês. Pode variar ainda em outros aspectos, como se o valor que a proposta vencedora pagará o valor de seu lance ou da segunda melhor proposta.

No entanto, as variações podem ser muitas, como a que ocorre no Mercado de Peixes de Portland, nos Estados Unidos, em que o leilão começa com um preço de abertura determinado pelo leiloeiro e inicialmente cai. Quando um comprador então indica interesse em um certo tipo de peixe, o preço sobe em incrementos de um centavo até que ninguém esteja disposto a aumentar o preço. O motivo da vasta gama de possibilidades de configuração de um leilão é que cada sistema apresenta intrinsicamente incentivos (ou barreiras) que influenciam profundamente a dinâmica na participação e proposição de lances, o que afeta diretamente o negócio a ser transacionado, objetivo maior de um leilão.

4 INTELIGÊNCIA ARTIFICAL E BOTS

Este capítulo busca posicionar os *bots* como uma ferramenta tecnológica amplamente utilizada atualmente, classificando-os e descriminando seu funcionamento em pregões eletrônicos. Aproveitaremos alguns estudos relativos a fraudes em leilões eletrônicos devido ao leilão eletrônico uma ferramenta pioneira em compra e venda via internet enquanto que o pregão eletrônico é uma modalidade de licitação específica e relativamente nova, ao passo que há similaridade no modo de funcionamento dos *bots* em ambos os ambientes.

4.1 INTELIGÊNCIA ARTIFICIAL

Poole e Mackworth (2010) definem a inteligência artificial, ou IA, como o campo da ciência que estuda a síntese e análise de agentes computacionais que atuam de forma inteligente. Russell e Norvig (2016) afirmam que é um dos campos mais novos da ciência e da engenharia e que busca não apenas entender como os humanos pensam, o que seria o objeto de interesse da inteligência comum, mas também construir entidades inteligentes. Russell e Norvig (2016) trazem 4 perspectivas nas quais outros autores abordam a IA podendo partir do pressuposto de que ela: pensa humanamente, age humanamente, pensa racionalmente ou atua racionalmente.

Autores da primeira categoria definem IA como sendo um conjunto de esforços voltados para fazer máquinas pensarem como os serem humanos, associando a atividades como tomada de decisões, resolução de problemas e aprendizado. Essa abordagem de modelagem cognitiva transita no campo interdisciplinar da ciência cognitiva, que "reúne modelos computacionais da IA e técnicas experimentais da psicologia para construir teorias precisas e testáveis da mente humana" (RUSSELL; NORVIG, 2016, p. 3, tradução nossa).

Autores da segunda perspectiva abordam a IA como sendo "a arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas" (KURZWEIL, 1990 *apud* RUSSELL; NORVIG, 2016, p. 2, tradução nossa) e relacionam essa abordagem ao Teste de Turing, proposto por Alan Turing em 1950. Projetado para fornecer uma definição satisfatória de inteligência, o teste consiste em um sistema em que um computador é aprovado se um interrogador humano, após realizar algumas perguntas, não souber dizer se as respostas vêm de uma pessoa ou de um computador. Para tal, o computador

precisaria dominar processamento de linguagem natural (para se comunicar fluentemente), representação do conhecimento (para armazenar antigas e novas informações), raciocínio automatizado (para utilizar a informação armazenada para tirar conclusões e responder) e aprendizagem de máquina (para se adaptar a novas circunstâncias, detectar padrões e extrapolá-los quando necessário).

A terceira perspectiva se refere ao pensar racionalmente, que os autores descrevem como sendo a abordagem das leis do pensamento e que tem como núcleo a lógica para construir sistemas inteligentes. A quarta perspectiva tem como base a figura do agente racional, ou que atua racionalmente. Para ilustrar essa perspectiva, Russell e Norvig (2016, p. 4, tradução nossa) diferenciam o agente computacional do *software* computacional comum, pois espera-se do agente que: "operem de forma autônoma, percebam seu ambiente, persistam por um período de tempo prolongado, adaptem-se a mudanças e criem e busquem metas". Enquanto a abordagem das "leis do pensamento" colocava ênfase estava nas inferências corretas, tal precisão é apenas uma parte da atuação do agente racional. Os autores comentam ainda que:

> Em algumas situações, não há nenhuma coisa comprovadamente correta a ser feita, mas que, mesmo assim, algo ainda precisa ser feito. Há também maneiras de agir racionalmente que não podem ser consideradas como envolvendo inferência. Por exemplo, recuar de um fogão quente é uma ação reflexa que geralmente é mais bem-sucedida do que uma ação mais lenta tomada após deliberação cuidadosa. (RUSSELL; NORVIG, 2016, p. 4, tradução nossa).

A abordagem do agente racional será a adotada neste estudo, pois explica melhor os fenômenos investigados. Segundo Russel e Norvig (2016, p. 34, tradução nossa), no campo da IA, "um agente é qualquer coisa que possa ser vista percebendo seu ambiente através de sensores e agindo sobre esse ambiente através de atuadores". Uma forma de caracterizar um agente através de sua atuação, segundo Russel e Norvig (2016, p. 40), é a metodologia denominada pelo acrônimo *PEAS* (do inglês, *Performance*, *Environment*, *Actuators*, *Sensors*), composto pelos componentes Performance, Ambiente, Atuadores e Sensores, exemplificado no Quadro 3.

Quadro 3 - Classificação PEAS

Tipo do Agente	Medida de Performance	Ambiente	Atuadores	Sensores
Motorista de taxi	Segurança, rapidez, legalidade, conforto, viagem, maximização dos lucros	Estradas, tráfego de outros veículos, pedestres, clientes	Direção, acelerador, freios, setas, iluminação, buzina, *display*	Câmeras, sonar, velocímetro, GPS, odômetro, acelerômetro, sensores do motor, teclado
Sistema de diagnóstico médico	Paciente saudável, custos reduzidos	Paciente, hospital, equipe	Investigar o paciente, fazer testes, diagnósticos, tratamentos, referências	Teclado para entrada de sintomas, descobertas, respostas do paciente
Sistema de análise de imagens de satélite	Categorização correta da imagem	Conexão com satélite em órbita	Exibição de categorização de cena	Matrizes coloridas de pixels
Robô de coleta de partes	Porcentagem de peças em caixas corretas	Correia transportadora com peças, caixas	Braço articulado e mão	Câmeras, sensores de ângulo de articulação
Controlador de refinaria	Pureza, rendimento, segurança	Refinaria, operadores	Válvulas, bombas, aquecedores, displays	Temperatura, pressão, sensores químicos
Professor de inglês interativo	Pontuação do aluno no teste	Conjunto de estudantes, agência de testes	Exibição de exercícios, sugestões, correções	Entrada de teclado

Fonte: Russell e Norvig, 2016, p. 40-42 (Tradução nossa)

Já do ponto de vista interno, Russel e Norvig (2016) descrevem que um agente de IA é composto por uma arquitetura (*hardware*, sensores, físicos e atuadores) e por um programa. Trazem então uma tipologia para diferenciar agentes de acordo seu programa, resumida da seguinte forma:

> Os **agentes de reflexos simples** respondem diretamente às percepções, enquanto os **agentes reflexivos baseados em modelos** mantêm o estado interno para rastrear aspectos do mundo que não são evidentes na percepção atual. **Agentes baseados em metas** agem para atingir seus objetivos, e **agentes baseados em utilidade** tentam maximizar sua própria "felicidade" esperada. Todos os agentes podem melhorar seu desempenho por meio do **aprendizado**. (RUSSEL; NORVIG, 2016, p. 59, tradução nossa, grifos do autor).

Russell e Norvig (2016), trazem ainda uma distinção mais genérica que contribui com a compreensão da amplitude da IA:

> Um **agente humano** tem olhos, ouvidos e outros órgãos como sensores e mãos, pernas, voz e assim por diante como atuadores. Um **agente robótico** pode ter câmeras e rastreadores infravermelhos como sensores e vários motores como atuadores. Um **agente de *software*** recebe pressionamentos de teclas, conteúdos de arquivos e pacotes de rede como entradas sensoriais e age no ambiente exibindo na tela, gravando arquivos e enviando pacotes de rede. (RUSSELL E NORVIG, 2016, p. 34, tradução nossa, grifo nosso).

Poole e Mackworth (2010) complementam dizendo que o agente de *software* é aquele que atua em um ambiente puramente computacional, ou seja, sem habilidades e sensações tipicamente humanas, como visão, audição ou movimento, restringindo-se a tarefas essencialmente intelectuais. Um termo utilizado para descrever um agente de *software* é o *bot*, ou robô eletrônico, que pode ter atuações diversas e está cada vez mais presente na sociedade atual.

4.2 BOTS

Bot é o termo atual utilizado para descrever entidades inteligentes de atuações diversas. Tem origem no termo *softbot* (LIEBERMAN, 1997), abreviação em inglês dos termos "*software robot*" (ETZIONI; WELD, 1994, p. 72). No Brasil, é também conhecido como robô eletrônico. Recentemente o Gabinete de Segurança Institucional do governo federal, através do Glossário de Segurança da Informação, definiu *bot* como sendo um "tipo de código malicioso" (BRASIL, 2019d). No entanto, em um contexto geral, o *bot* não é necessariamente malicioso, podendo agir também como ferramentas de automação para empresas e pessoas.

Um estudo realizado por empresa de mitigação dos efeitos de *bots* estimou que 42,2% de todo o tráfego da internet em 2017 não foi humano, sendo 21,8% feito por *bots* maliciosos, o que representou um aumento de 9,5% em comparação a 2016 (DISTIL NETWORKS, 2018). Outro estudo aponta que a maioria do tráfego na internet em 2016 já era feito por *bots*. Esse estudo analisou dados de 90 dias entre julho e outubro de 2015, com dados de tráfego de 100.000 domínios escolhidos aleatoriamente na rede, somando com mais de 16,7 bilhões de visitas de humanos e robôs e cobrindo 249 países e territórios, e seus resultados apontaram que apenas 48,2% dos acessos foram feitos por humanos contra 51,8% feitos por *bots*, sendo 28,9% do total feitos por *bots* maliciosos (INCAPSULA, 2016).

Neste estudo busca-se especificamente identificar os efeitos e fenômenos decorrentes da utilização de *bots* em pregões eletrônicos. No entanto, esse tipo de *bot*, que emite lances em disputas, é realidade em leilões eletrônicos em todo o mundo e vem sendo estudado nesse ambiente há bastante tempo. Por esse motivo, serão apresentados abaixo um levantamento teórico sobre a atuação de *bots* em leilões, visando o aproveitamento análogo do conhecimento e de técnicas utilizadas nos estudos dos efeitos relacionados.

4.3 O USO DE BOTS EM LEILÕES

Houve uma grande popularização do leilão com o advento do formato eletrônico. A plataforma o eBay, seu expoente mais bem-sucedido, criado em 1995 e considerada como uma das 500 maiores empresas do mundo pela revista Fortune, movimenta um volume bruto de mercadorias de mais de 22,7 bilhões de dólares atualmente (EBAYINC.COM, 2018). Esse volume de transações vem gerando grande interesse no combate a fraudes pela literatura e pela academia.

Rebouças (2011) reúne relevantes pesquisas realizadas sobre figurantes em leilões eletrônicos, como Wang, Hidvégi e Whinston (2001), Shah, Joshi e Wurman (2002), Bhargava, Jenamani e Zhong (2005), Trevathan e Read (2006), Dong, Shatz e Xu (2009) e Ford, Xu e Valova (2010). O autor resume a perspectiva de alguns:

> Trevathan e Read (2006) definem figurantes em leilões como atores que dão lances falsos com a finalidade de aumentar artificialmente o preço de um determinado item. Para Dong, Shatz e Xu (2009a), figurante é uma pessoa que se posiciona como sendo um legítimo comprador que finge entusiasticamente através de lances, se fazendo de cúmplice para o vendedor, no intuito de que o item do leilão aumente cada vez mais o seu preço. Bhargava, Jenamani e Zhong (2005) definem o figurante como um cúmplice que se comporta como um participante entusiasta que incentiva outros participantes (compradores e jogadores) a continuar no leilão. (REBOUÇAS, 2011, p. 32).

É evidente a preocupação em como a participação de figurantes afeta a dinâmica do leilão. Apesar de não ser a única forma de fraudar leilões eletrônicos, a utilização de *bots* é um problema relevante nesse meio. Em uma curta definição, são sistemas de computador que são programados para um conjunto de tarefas automatizadas (MONAHAN, 2018). Segundo o autor, existem *bots* bons e ruins, ou maliciosos. Os bons são criados por empresas para executar tarefas repetitivas, como examinar e coletar informações em sites, a uma taxa muito

maior do que a capacidade humana seria capaz. Já os maliciosos são criados para obter indevidamente controle de dispositivos (móveis ou não), dados (pessoais ou corporativos), cometer fraudes e outros crimes, prejudicar sites de concorrentes, entre outras inúmeras finalidades. Por não estarem associados ao atacante, automatizáveis, distribuídos globalmente e, por fim, descartáveis, os *bots* são especialmente eficazes em muitos desses objetivos.

Os estudos pioneiros sobre a atuação de figurantes surgiram devido à popularização dos leilões online, que tem como expoente global o Ebay e em nível nacional o Mercado Livre. O formato online favoreceu o aumento de participantes que, por sua vez, atraiu um número maior de vendedores. Entretanto, devido aos participantes não estarem fisicamente presentes, cresceu também a ocorrência de participantes que buscam influenciar os resultados de forma irregular.

Os *bots* atuam em leilões e licitações se utilizando dessas características tecnológicas aplicando uma estratégia comum de lances, agindo como pessoas, mas com tempo de resposta e capacidade de carga de trabalho incomparáveis. Segundo Dong, Shatz e Xu (2009), uma estratégia de lances é um plano de ação projetado para atingir metas específicas através de lances em leilões, podendo-se adotar uma única ou usar estratégias mistas. No Quadro 4 estão listadas algumas estratégias de licitação representativas que são comumente usadas em mercados de leilão:

Quadro 4 - Cinco estratégias comuns de lances

Tipo	Comportamento
Skeptico (Cético)	Dá vários lances, mas aumentando o valor o mínimo possível
Proxy bidding (Licitante por procuração)	Dá quantos lances forem necessários até atingir o máximo previamente especificado
Sniping (Atirador de precisão)	Dá lance nos últimos segundos, não dando tempo para outros cobrirem
Unmasking (Desmascarador)	Dá vários lances em um curto espaço de tempo com o objetivo de expor o lance máximo ou os participantes mais competitivos
Evaluator (Avaliador)	Dá apenas um lance no início com um valor alto

Fonte: Dong, Shatz e Xu, 2009 (Tradução nossa)

Os mesmos autores listam algumas formas de fraudes possíveis em leilões eletrônicos sob a perspectiva criminal do *Investigation's Internet Crime Complaint Center (IC3)* do *FBI*, apresentadas no Quadro 5.

Quadro 5 – Formas de fraude em leilões eletrônicos

Forma de fraude	Descrição
Deturpação de um produto para venda	Utiliza descrição falsa de características ou condições do produto
Não entrega de mercadorias ou serviços vendidos	Burla os mecanismos de proteção das plataformas intermediárias para receber o pagamento sem que haja produto ou serviço real a ser comercializado
Vendendo bens do mercado negro	Comercialização de bens ilícitos ou com venda proibida
Triangulação	Os fraudadores compram itens usando um cartão de crédito roubado, vendendo os itens a compradores não iniciados, retendo assim o dinheiro e transferindo o risco de apreensão para o destinatário final
Cobrança de taxas	Cobrando dinheiro extra após o término de um leilão
Lances múltiplos	Os compradores aumentam os preços usando pseudônimos, o que frustra os concorrentes e, no último momento, os lances altos são retirados para garantir um lance baixo
Figurante (*shill bidding*)	Vendedores ou seus associados fazem lances em seus próprios leilões para fins fraudulentos

Fonte: Adaptado de Dong; Shatz e Xu, 2009 (Tradução nossa)

Aplicada à realidade das licitações brasileiras, mais especificamente à dos pregões eletrônicos, as duas últimas formas listadas no Quadro 5 são as mais recorrentes, tendo em vista que não se trata de um mecanismo de venda e que: a) o comprador é sempre o governo (federal, estadual, municipal ou distrital); b) a plataforma é o Portal de Compras Governamentais (Comprasnet); c) exige um procedimento de cadastramento prévio Sistema de Cadastramento Unificado de Fornecedores – SICAF, que interliga alguns sistemas governamentais; e d) todas as condições que regem cada licitação são previamente divulgadas através dos editais, sendo disponibilizado o tempo legalmente previsto para sanear dúvidas e demais questões sobre o procedimento e sobre o objeto a ser licitado.

4.4 O USO DE BOTS EM PREGÕES ELETRÔNICOS

O uso de *bots* é uma questão conhecida em leilões eletrônicos e se tornou realidade também em pregões eletrônicos. Os órgãos de controle e as plataformas de operacionalização de pregões eletrônicos vêm constantemente desenvolvendo barreiras para mitigar os efeitos de seu uso, visando evitar que afetem negativamente as disputas em desfavor de quem não os utiliza. No entanto, o governo federal, como gestor da maior plataforma de operacionalização de pregões eletrônicos do país, o Comprasnet, parece encontrar dificuldades em superar

barreiras já identificadas. As soluções até então implementadas não surgiram por iniciativa do SERPRO (Serviço Federal de Processamento de Dados), desenvolvedor da plataforma, mas de forma compulsória por reiteradas determinações do Tribunal de Contas da União (TCU), tendo se mostrado insuficientes para sanar o problema do desequilíbrio na disputa.

O pregoeiro não encerra imediatamente a fase de lances. Seu poder quanto ao encerramento da disputa se restringe a estipular um prazo de iminência (de 1 a 60 minutos) (BRASIL, 2019b) para que o item avance para a fase de encerramento aleatório. Decorrido o prazo, o encerramento se dá de forma aleatória no intervalo de 1 a 30 minutos após o fim do prazo de iminência (BRASIL, 2019b). Essa sistemática visa, entre outros objetivos, evitar fraudes relacionadas ao encerramento em horário ou valor específicos em benefício de algum licitante, por exemplo, pois o vencedor é aquele que tiver ofertado o menor valor (ou maior desconto, de acordo com as regras estipuladas em edital), assim como é definida a classificação do segundo colocado em diante.

É nesse contexto que se localizam os *bots* para emissão de lances automáticos. Atuam mediante programação para cobrir o melhor lance atual, tornando-se então o melhor lance, buscando estar o máximo de tempo possível nessa condição durante a fase de encerramento aleatório, aumentando suas chances de ser declarado o vencedor. Em tese, licitantes que os utilizam apresentam vantagem competitiva diante de outros que não os utilizam, pois são muito mais rápidos que um ser humano no processo que consiste em: colher informações de todos os itens simultâneos que estiver participando (que podem ser dezenas ou centenas); identificação de um novo melhor lance para tal item; pensar no valor do novo lance; digitar o valor do lance; pressionar o botão para emitir o lance.

Podemos caracterizar o *bot* como um agente racional de *software*, de interface e de reflexos simples, pois funcionam basicamente reagindo imediatamente a lances emitidos por outros licitantes. Quanto a seu ambiente de tarefas, podemos classificar esse o *bot* utilizado em licitações na forma do Quadro 6.

Quadro 6 - Classificação PEAS para *bots* licitantes

Tipo do Agente	Medida de Performance	Ambiente	Atuadores	Sensores
Licitante	Licitações vencidas, valor dos lances, intervalos entre lances	Portal de Compras Governamentais, outros licitantes, agentes do governo, órgãos de controle, sociedade	Emissão de lances	Valor do menor lance atual, horário do menor lance atual, horário do último lance emitido por ele próprio, itens em encerramento aleatório

F**onte:** Elaborado pelo Autor de acordo com Russell e Norvig, 2016.

Quanto ao aspecto da autonomia na atuação, Franklin e Graesser (1997) trazem uma classificação, dizendo que qualquer agente autônomo apresenta pelo menos quatro características: reatividade, autonomia, orientação a objetivos e continuidade. Pode ainda apresentar outras, como comunicabilidade (com outros agentes, incluindo pessoas), capacidade de aprendizado, mobilidade, flexibilidade e até personalidade. Os autores resumem da seguinte forma: "um agente autônomo é um sistema situado dentro e parte de um ambiente que percebe esse ambiente e age sobre ele, ao longo do tempo, em busca de sua própria agenda e de modo a afetar o que ele percebe no futuro" (FRANKLIN E GRAESSER, 1997, p. 5, tradução nossa).

Lieberman (1997) apresenta duas ramificações de tipos de agentes de *software*, que são os de interface, que ajudam ativamente um usuário a operar uma interface interativa, e os autônomos, que executam ações sem intervenção do usuário. Enquanto os primeiros atuam com manipulação direta de interfaces gráficas, perceptíveis pelos usuários humanos, e dependem da intervenção humana, os do segundo tipo atuam em plano de fundo, de forma contínua. No entanto, traz um terceiro tipo, o agente de interface autônomo, que mescla características descritas acima, dizendo que

> Para que um agente seja considerado um agente de interface e autônomo, deve haver alguma parte da interface que o agente deve operar de maneira autônoma. O usuário deve poder observar diretamente as ações autônomas do agente e o agente deve poder observar as ações tomadas de forma autônoma pelo usuário na interface. Concretamente, o usuário verá os elementos da interface que parecem "mover-se sozinhos" em resposta à entrada que o agente pareceu "ver por si mesmo" em vez de ter sido explicitamente instruído. (LIEBERMAN, 1997, p. 69, tradução nossa).

A discussão sobre a existência desse fenômeno ganhou relevância entre os anos de 2009 e 2012, sendo objeto de projetos de lei, processos judiciais e determinações do Tribunal de Contas da União (TCU). Um dos fatos de destaque foi o Relatório de Levantamento de Auditoria, que resultou no Acórdão nº 1.647/2010 – TCU – Plenário, o consequente Acórdão

nº 2.601/2011 – TCU – Plenário e outros (Acórdão nº 656/2011 – TCU – Plenário, Acórdão nº 837/2011 – TCU – Plenário, Acórdão nº 1.216/2014 – Plenário, Acórdão nº 485/2015 – TCU – Plenário).

No primeiro, resultado de trabalho iniciados em 2009, decorreu de denúncia feita diretamente ao TCU através de manifestação de ouvidoria por "licitantes que se sentiram prejudicados por terem identificado a resposta extremamente rápida de outros licitantes, numa agilidade impossível de ser acompanhada por operador humano" (TCU, 2010a, p. 49). Na época da denúncia, bem anterior à época da análise em foco do TCU, a antiga Secretaria de Logística e Tecnologia da Informação (SLTI), responsável pela manutenção e desenvolvimento do então denominado Comprasnet, reagiu criando em 2006 um mecanismo 'antirrobô', que solicitava um código de confirmação (tecnologia *captcha*) para os licitantes que apresentassem lance em período inferior a 6 segundos em relação a seu lance anterior. No entanto, o Ministro-Relator afirmou haver tecnologias capazes de interpretar o código de confirmação, tornando a regra inútil. Ou mais, ampliando ainda mais a diferença de tempo de resposta entre máquina e ser humano, pois além do tempo de reação para a digitação havia o tempo para interpretar o código. O relator conclui:

> **Conclusão**
> 5.4.18 O uso de dispositivos de inserção automática de lances, por alguns licitantes, pode comprometer a isonomia entre os participantes do certame (item 5.4.2).
> 5.4.9. A atual regra "antirrobô" do Comprasnet não é suficiente para atingir o objetivo de impedir vantagem competitiva, por meio do uso de dispositivo robô (item 5.4.5).
>
> **Proposta de Encaminhamento**
> 5.4.10. Recomendar à Secretaria de Logística e Tecnologia da Informação/MP que:
> 5.4.10.1. Estude meios de prover isonomia entre os licitantes do pregão eletrônico, em relação a possível vantagem competitiva que alguns licitantes podem obter ao utilizar dispositivos de envio automático de lances (TCU, 2010a, p. 50).

Por fim, o Acórdão determina:

> 9.1.13. no prazo de noventa dias adote meios de prover isonomia entre os licitantes do pregão eletrônico, em relação a possível vantagem competitiva que alguns licitantes podem obter ao utilizar dispositivos de envio automático de lances (robôs) (TCU, 2010a, p. 58).

Tal publicação teve repercussão na mídia, sendo objeto de uma matéria na edição de 22/12/2010 da revista Veja, intitulada "Matrix federal", de reportagem no Jornal Valor Econômico de 13 de setembro de 2011, intitulada "Empresas questionam uso de *software* em

licitações", além de várias outras publicações pela internet, que questionaram a integridade do Comprasnet. Como resposta à sociedade, o governo federal divulgou uma notícia intitulada "Melhoria do Sistema amplia segurança em Pregões eletrônicos", onde a então secretária da SLTI disse que "a utilização desses programas não garante vitória no pregão, pois nosso sistema é preparado para rejeitar lances em intervalos menores dos que são alcançados por uma pessoa" (BRASIL, 2010).

Posteriormente, o próprio TCU abriu processo para monitoramento específico do subitem 9.1.13 transcrito acima, onde não identifica qualquer implementação no sistema Comprasnet que estivesse em conformidade com o que foi determinado no acórdão de 2010. Entretanto, traz à luz análises importantes tratadas pela Secretaria de Fiscalização de Tecnologia da Informação (Sefti), do TCU:

> 4.10. No caso concreto ora em análise, as informações prestadas pela representante, de fato, fornecem robustos indícios de utilização de programa dessa natureza pela empresa STOQUE Soluções Tecnológicas Ltda. Dos documentos apresentados verifica-se que os 16 últimos lances postados pela ora representante foram imediatamente seguidos de lances inferiores da empresa vencedora do certame. Com efeito, a fase de lances foi encerrada às 15:15:56:770, sendo que a empresa RICALL havia apresentado lance de R$ 1.634.000,00, às 15:15:55:710, no entanto, a empresa STOQUE, apenas 1,06 segundos depois, foi considerada vencedora pelo menor preço de R$ 1.633.990,49 (diferença de R$ 9,51). Além disso, o tempo total em que a empresa RICALL permaneceu como detentora da melhor proposta foi de 16,52s, pois de forma quase que concomitante surgia o lance da outra concorrente, que permaneceu 358,71s com a melhor proposta (fl. 135). Nesse caso, a probabilidade de o encerramento da fase de lances se dar em favor da empresa que dispõe desses robôs foi de 95,60%. Assim, corrobora-se entendimento da representante de que as chances de se sagrar vencedora eram remotas, por não ser viável competir manualmente com recurso de tecnologia da informação, que pode colocar em vantagem competitiva a licitante que dele dispõe.
>
> 4.11. Verifica-se que, de fato, a utilização de *software* de lançamento automático de lances (robôs) demonstra vantagem competitiva dos fornecedores que detêm a tecnologia em questão sobre os demais licitantes, o que demonstra real afronta ao princípio da isonomia, insculpido no artigo 3º, caput, da Lei n.º 8.666/1993 e no artigo 5º, caput, do Decreto n.º 5.450/2005' (TCU, 2011, p. 4, grifos originais).

À frente, traz mais definições importantes à discussão sobre a existência de ilegalidade ou não no uso de robôs quando apresenta uma solução normativa para o problema. Conforme o item 5.9 do relatório do qual decorre o acórdão, a fundamentação não depende necessariamente de norma específica que reprimisse o uso dessa ferramenta, pois o próprio Decreto nº 5.450/2005 afirma, em consonância com a já mencionado art. 3º da Lei nº 8.666/1993:

> Art. 5º A licitação na modalidade de pregão é condicionada aos princípios básicos da legalidade, impessoalidade, moralidade, igualdade, publicidade, eficiência, probidade administrativa, vinculação ao instrumento convocatório e do julgamento objetivo, bem como aos princípios correlatos da razoabilidade, competitividade e proporcionalidade.
>
> Parágrafo único. **As normas disciplinadoras da licitação serão sempre interpretadas em favor da ampliação da disputa entre os interessados**, desde que não comprometam o interesse da administração, o princípio da isonomia, a finalidade e a segurança da contratação.' (BRASIL, 2005, grifos nosso)

Por fim, conclui:

> 5.11. Assim, entende-se que o MP deve definir a forma pela qual será perseguido o princípio da isonomia no pregão eletrônico, o que poderá envolver **adoção de medidas que diminuam a vantagem competitiva atualmente existente a favor dos licitantes que utilizam robôs, em detrimento dos demais que não utilizam esse recurso**. Tal definição pode vir a ser regulamentada por instrução complementar, nos termos do art. 31 do Decreto 5.450/2005, que definiria como a SLTI entende que garantirá a isonomia entre todos os licitantes e introduziria regras adicionais para a inibição ou limitação do uso dos robôs.
>
> 5.12. **Essa regulamentação também propiciaria, por via indireta, a persecução da competitividade do certame, pela possibilidade de um operador humano competir com o robô**. Na via finalística da licitação, também ajudaria na busca pelo melhor preço para a administração pública, na medida em que aumentaria a incerteza do licitante que utiliza robô, atualmente seguro em apenas cobrir discretamente os demais lances ofertados (TCU, 2011, grifos nossos).

No mesmo ano, foi publicada a matéria "Golpe no pregão eletrônico" pela ISTOÉ Independente, em que o então secretário de logística e Tecnologia da Informação do Planejamento, Delfino de Souza, afirma: "a gente tenta coibir, mas é difícil". O diretor de Serviços do ministério, Carlos Henrique Moreira completa: "É uma corrida atrás do rabo o tempo todo, mas não há ilegalidade porque (o uso de robôs) não está proibido por lei" (ISTOÉ, 2018).

Sendo a ausência de regulamentação identificada como barreira ao enfrentamento efetivo dessa prática e como consequência das determinações do TCU, é publicada então pela SLTI a Instrução Normativa nº 3, de 16 de dezembro de 2011 (BRASIL, 2011b), cujo objetivo foi determinar que o sistema Comprasnet não mais aceitaria lances enviados pelo mesmo licitante em intervalo inferior a vinte (20) segundos. Posteriormente veio a Instrução Normativa nº 3, de 4 de outubro de 2013, alterando a IN/SLTI nº 3/2011 para a inclusão da regra de que o intervalo entre lances de fornecedores diferentes não poderia ser inferior a três (3) segundos (BRASIL, 2011b).

Essa o instrumento buscou atenuar a vantagem obtida pelo uso dos robôs, mas não extinguiu seu uso. Em março de 2017 foi deflagrada a denominada Operação Licitante Fantasma pelo Ministério da Transparência, Fiscalização e Controladoria-Geral da União (CGU) e a Polícia Federal (PF), com o objetivo de desarticular uma organização criminosa responsável por fraudes em pregões eletrônicos e presenciais do governo federal (BRASIL, 2017b). Segundo a PF (BRASIL, 2017b), em três anos de investigação ficou constatado que a quadrilha movimentou R$ 60 milhões entre 2011 e 2015. A estimativa é de que o prejuízo ao erário seja de R$ 25 milhões, referente ao superfaturamento de itens em que foram vencedores através de práticas criminosas realizadas por um grupo de empresários por meio de conluio para acerto de preço, mas também por fraudes nos lances efetuados através de *bots*, afetando diversos órgãos federais, incluindo a Superintendência Federal de Agricultura do Mato Grosso do Sul (SFA/MS), do MAPA. Foi solicitado acesso ao relatório através do protocolo e-SIC nº 08850.004651/2018-15, porém foi indeferido devido a sigilo, inerente às investigações policiais em curso.

Isso demonstra que regras unidimensionais são insuficientes para que o pregão eletrônico se torne íntegro, e que ações relacionadas à prevenção de fraudes trazem grandes benefícios e impactam o orçamento federal significantemente devido às proporções dessa modalidade. Este estudo busca trazer à luz um método de verificação dos efeitos da utilização desses *bots*, utilizando, para isso, dados governamentais abertos.

5 DADOS ABERTOS E TRANSPARÊNCIA GOVERNAMENTAL

O governo federal vem passando por um processo de transformação digital. Tem elaborado estratégias de governança digital para atender às novas legislações e às novas demandas de serviços e transparência de informações. A emissão de certidões, licenças e consultas a dados pessoais constantes dos cadastros governamentais pela internet, a exposição da destinação de recursos públicos, a desburocratização e a simplificação de serviços e as licitações eletrônicas são alguns exemplos do novo cenário governamental.

Mas não apenas esses fatores demandaram mudanças profundas na forma de armazenamento, processamento, organização e publicação dos dados. Medeiros (2004) aponta que outros aspectos incentivaram essa transformação, como a necessidade de as Administrações aumentarem sua arrecadação e melhorarem a eficiência de seus processos internos e as "pressões da sociedade para que o governo otimizasse seus gastos e atuasse, cada vez mais, com transparência, qualidade e de modo universal na oferta de serviços e provimento de informações aos cidadãos e organizações em geral" (MEDEIROS, 2004, p. 29).

Com a evolução da ideia de governo eletrônico, hoje o termo não se traduz apenas como a simples utilização de tecnologias de informação no setor público (DINIZ, 2009). Segundo levantamento feito pelo autor, pode estar vinculada à modernização da administração pública por meio do uso de tecnologias de informação e comunicação (TICs), mas associada também à melhoria da eficiência dos processos operacionais e administrativos dos governos e ao uso de internet no setor público para prestação de serviços públicos. Nesse sentido, a OCDE (2003) diz que o *e-government* é mais sobre governo do que sobre "e" (fazendo alusão ao termo *eletrônico*). Afirma que seu impacto é um melhor governo por possibilitar melhores resultados das políticas públicas, serviços de mais qualidade, maior engajamento dos cidadãos, entre outros benefícios. Medeiros (2004) traz alguns conceitos do termo no Quadro 7:

Quadro 7 – Definições de Governo Eletrônico

Definição	Fonte
Governo Eletrônico refere-se ao uso, por agências governamentais, de tecnologias de informação (como redes de longa distância, Internet e computação móvel) capazes de transformar as relações com cidadãos, empresas e outras unidades do governo. Essas tecnologias podem servir a diferentes fins, como: melhor prestação de serviços aos cidadãos, interações mais eficazes com	Banco Mundial

empresas e a indústria, empowerment do cidadão por meio do acesso a informações ou mais eficiência na administração governamental.	
Governo eletrônico ou e-governo significa prover acesso público via Internet a informações sobre os serviços oferecidos pelos departamentos centrais do governo e suas agências, habilitando o público à condução e conclusão de transações para tais serviços.	National Audit Office
O termo *'e-government'* [utilizado em inglês para denominar o governo eletrônico] aborda o uso de novas tecnologias de informação e comunicação pelos governos, aplicadas a todas as suas funções.	OCDE
E-governo é definido como: utilização da Internet e da web para ofertar informações e serviços governamentais aos cidadãos.	Nações Unidas
O termo governo eletrônico é definido como a oferta e troca de informações e serviços governamentais on-line para cidadãos, empresas e outras agências governamentais.	INTOSAI
Governo eletrônico refere-se a processos e estruturas relativos ao fornecimento eletrônico de serviços governamentais ao público.	Okot-Uma

Fonte: Medeiros, 2004.

A abertura dos dados ao público é etapa fundamental nesse processo. É através dessa publicação de todas as informações que não sejam impedidas por lei de serem divulgadas que a mudança cultural é possível, ampliando as possibilidades para que a sociedade exerça a fiscalização social. Mais que isso, a abertura de dados significa abrir portas para a inovação, pois dá informação como matéria prima para a geração de novas ideias, produtos, serviços e mecanismos de controle nunca previstos.

Cabe, neste ponto, diferenciar os conceitos de Governo Eletrônico e Governo Aberto. Tem como propulsor no país a Parceria para Governo Aberto (ou *Open Government Partnership – OGP*), iniciativa internacional criada em 2011 para difundir e incentivar e que conta com a adesão de mais de 70 países e mais 15 governos subnacionais (CGU, 2018, p. 5). O Brasil foi um dos 8 países fundadores. Segundo a CGU (2018), na divulgação do 4º Plano de Ação Nacional em Governo Aberto, este é um conceito que:

> [...] se refere a uma nova visão da Administração Pública, que promove projetos e ações voltados ao aumento da transparência, à luta contra a corrupção, ao incentivo à participação social e ao desenvolvimento de novas tecnologias que tornem os governos mais responsáveis por suas ações e preparados para atender às necessidades dos cidadãos. (CGU, 2018, p. 6).

Como efeito, busca trazer ferramentas que possam ser utilizadas pelo cidadão em uma perspectiva de participação ativa, trazendo-o ao foco das políticas públicas. Considerando o cenário histórico político nacional, representa uma profunda quebra de paradigmas. Para tal, foram instituídos os seguintes princípios para o Governo Aberto:

> a. Accountability (prestação de contas e responsabilização) – um governo responsivo estabelece normas, regulamentos e mecanismos que obrigam atores governamentais a justificar ações, agir de acordo com as críticas ou exigências colocadas e aceitar a responsabilidade no cumprimento de seus deveres. b. Participação social – um governo participativo promove o a participação ativa da sociedade nos processos de formulação das políticas públicas e a criação de novos espaços de interlocução que favoreçam o protagonismo e o envolvimento dos cidadãos. Além disso, viabiliza que as administrações públicas se beneficiem do conhecimento, das ideias e da experiência dos cidadãos. c. Transparência – um governo transparente proporciona informações sobre o que está fazendo, sobre seus planos de ação, suas fontes de dados e suas atribuições perante a sociedade. Além disso, estimula a prestação de contas da administração e as ações permanentes de controle social. d. Tecnologia e Inovação – um governo inovador entende o papel das novas tecnologias na inovação e busca assegurar aos cidadãos a capacidade de utilizarem novas ferramentas de desenvolvimento disponíveis e o acesso a elas. (CGU, 2018, p. 6).

As políticas relacionadas a dados abertos têm como principais marcos legais as normas constantes do Quadro 8.

Quadro 8 - Principais marcos legais relacionados a dados abertos

Legislação	Descrição	Trechos relevantes e contribuições
Lei Complementar nº 101/2000	Lei de Responsabilidade Fiscal	Criou o paradigma da transparência na gestão pública brasileira e deu início a um processo de abertura que levou o país a promover uma série de mudanças legais e administrativas (CGU, 2018).
Lei Complementar nº 131/2009	Lei da Transparência	Expande a transparência e a responsabilidade na gestão fiscal para todos os entes da federação.
Decreto s/nº de 15 de setembro de 2011	Plano de Ação Nacional sobre Governo Aberto	Passo relevante na participação do Brasil como protagonista na OGP (Parceria de Governo Aberto, tradução da sigla em inglês).
Lei nº 12.527, de 18 de novembro de 2011	Lei de Acesso à Informação	Amplia o acesso à informação pela sociedade.
Lei nº 12.965, de 23 de abril de 2014	Marco Civil da Internet	"Art. 24. Constituem diretrizes para a atuação da União, dos Estados, do Distrito Federal e dos Municípios no desenvolvimento da internet no Brasil: V - adoção preferencial de tecnologias, padrões e formatos abertos e livres; VI - publicidade e disseminação de dados e informações públicos, de forma aberta e estruturada; VII - otimização da infraestrutura das redes e estímulo à implantação de centros de armazenamento, gerenciamento e disseminação de dados no País, promovendo a qualidade técnica, a inovação e a difusão das aplicações de internet, sem prejuízo à abertura, à neutralidade e à natureza participativa".
Decreto nº 8.243, de 23 de maio de 2014	Política Nacional de Participação Social e Sistema Nacional de Participação Social	Fortalece a cultura de participação social no governo federal: "Art. 18. Na criação de ambientes virtuais de participação social devem ser observadas, no mínimo, as seguintes diretrizes: [...] VIII - priorização da exportação de dados em formatos

		abertos e legíveis por máquinas".
Decreto nº 8.638 de 15 de janeiro de 2016	Política de Governança Digital no âmbito dos órgãos e das entidades da administração pública federal	"Art. 2º Para os fins deste Decreto, considera-se: [...] II - dados em formato aberto - dados representados em meio digital em um formato sobre o qual nenhuma organização tenha controle exclusivo, passíveis de utilização por qualquer pessoa".
Decreto nº 8.777 de 11 de maio de 2016	Política de Dados Abertos do Poder Executivo federal	"Art. 2º Para os fins deste Decreto, entende-se por: I - dado - sequência de símbolos ou valores, representados em qualquer meio, produzidos como resultado de um processo natural ou artificial; II - dado acessível ao público - qualquer dado gerado ou acumulado pelo Governo que não esteja sob sigilo ou sob restrição de acesso nos termos da Lei nº 12.527, de 18 de novembro de 2011; III - dados abertos - dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte; IV - formato aberto - formato de arquivo não proprietário, cuja especificação esteja documentada publicamente e seja de livre conhecimento e implementação, livre de patentes ou qualquer outra restrição legal quanto à sua utilização;"

Fonte: Elaborado pelo Autor, 2019.

Um levantamento feito pelo TCU (2014a, p. 75), ressalta a "intensificação no país de ações para a disponibilização de informações governamentais por meio dos dados abertos, de forma a promover a transparência, a prestação de contas e a participação popular na administração pública." Exemplos disso são a criação do Portal da Transparência, (www.transparência.gov.br), os avanços com a Parceria para Governo Aberto e a criação da INDA (Infraestrutura Nacional de Dados Abertos) e do Portal Brasileiro de Dados Abertos (www.dados.gov.br). No entanto, destaca que "as ações ainda não se mostraram suficientes para a implementação e popularização do governo aberto no âmbito federal".

Ainda sobre o estudo realizado pelo TCU, cabe trazer o seguinte trecho do relatório:

> Em um cenário em que a abertura de dados públicos se populariza, aumentam as possibilidades de criação de soluções pelo governo ou pela própria sociedade, o que favorece o aparecimento de novas oportunidades de mercado para aqueles que desejem elaborar alternativas inovadoras para melhoria da prestação de serviços públicos. Portanto, a disponibilização de dados públicos de interesse coletivo ou geral tem grande potencial para agregar valor ao negócio da administração pública e à própria sociedade, com incremento da participação social e empoderamento do cidadão, que se coadunam com os objetivos estratégicos constantes do plano

> estratégico do TCU, quais sejam ‘contribuir para a melhoria do desempenho da Administração Pública’ e ‘contribuir para a transparência da Administração Pública’. (TCU, 2014a, p. 76).

Uma próxima etapa realizada foi a elaboração e publicação, em 2017, da Estratégia de Governança Digital do Governo Federal (EGD), que traz o objetivo de “fomentar a disponibilização e o uso de dados abertos”. Segue sua descrição:

> Fomentar a abertura dos dados dos órgãos e entidades públicas, respeitando o sigilo dos dados pessoais do cidadão, e promover o uso dos dados em formato aberto pela sociedade não somente para fins de controle social e transparência, mas também para o surgimento de iniciativas não governamentais de prestação de serviços por meio de aplicativos, de pesquisa científica e mercadológica, dentre outros (BRASIL, 2016, p. 16).

Diante dos aspectos e iniciativas apresentadas, são inequívocos o interesse e a necessidade de fomentar e de desenvolver os mecanismos de transparência das compras públicas. No entanto, da perspectiva do controle social, a efetivação das políticas de dados abertos não são um fim, mas um meio. Para que se torne real, é necessário que a sociedade seja capaz de utilizar os dados para retirar informações que respondam a questões específicas. Esse estudo busca cumprir esse papel através da exploração de dados abertos de compras governamentais e da mineração desses dados, por meio de técnicas e programas específicos, visando responder a questões relacionadas ao uso de *bots* em pregões eletrônicos. No próximo capítulo serão a fundamentação teórica relacionada a mineração de dados e descoberta de conhecimentos em base de dados.

6 MINERAÇÃO DE DADOS

As últimas décadas foram marcadas por grandes evoluções tecnológicas que vêm transformando a sociedade e que resultaram no aumento exponencial do volume de dados gerados, armazenados e transacionados. Han, Kamber e Pei (2012, p. 2) afirmam que esse crescimento explosivo do volume de dados disponível é resultado da informatização de nossa sociedade e do rápido desenvolvimento de poderosas ferramentas de coleta e armazenamento de dados. Segundo os autores, empresas de todo o mundo estão gerando conjuntos de dados gigantescos como nunca antes havia acontecido, incluindo transações de vendas, registros de negociação de ações, descrições de produtos, promoções de vendas, perfis e desempenho de empresas e feedback dos clientes. Ethem (2016, l. 193) ressalta que algumas das maiores transformações em nossas vidas nos últimos 50 anos são decorrentes da computação e tecnologia digital e que as ferramentas, dispositivos e serviços tradicionais têm sido substituídos por suas versões eletrônicas cada vez mais, ao mesmo tempo em que continuamente nos adaptamos a esse novo ambiente digital.

Como consequência, os dados têm ganhado cada vez mais valor nas empresas, assim como em governos, para pesquisadores e para a sociedade em geral. Saber extrair informações significativas de grandes volumes de dados se tornou uma das habilidades mais desejadas pelas empresas em busca do aumento da eficiência e da redução de custos em processos internos, de ações mais eficazes em *marketing* e vendas, da criação de novos produtos e da identificação de novos mercados. Em outubro de 2012 a *Harvard Business Review* publicou um artigo de Davenport e Patil (2012) intitulado *"Data Scientist: The Sexiest Job of the 21st Century"*, (Cientista de dados: o profissional mais cobiçado do século 21, tradução pela Harvard Business Review Brasil), enfatizando que o termo Cientista de Dados havia sido criado muito recentemente (apenas em 2008) e que em pouco tempo a atividade análise de dados havia crescido de forma exponencial.

Termos como "Economia dos Dados" e "Indústria 4.0" têm sido utilizados de forma generalizada para descrever esse fenômeno. "Nós vivemos na era da informação" é um dito popular, segundo Han, Kamber e Pei (2012, p. 1), enquanto que a revista The Economist (2017) se referiu aos dados como sendo o "novo petróleo", tratando-os como *commodities* e comparando os efeitos de seu domínio no mercado com os do petróleo no século passado. Ao mesmo tempo, a matéria expressa a preocupação com os efeitos do "oligopólio de dados" por

empresas de tecnologia dominantes e indica a abertura e o compartilhamento de dados como uma medida para manter a privacidade e a confiança dos usuários.

Apesar da popularização recente, o armazenamento e o acesso aos bancos de dados tem sido um dos principais objetivos de desenvolvedores, de organizações, de pesquisadores e demais usuários desde os primeiros sistemas computacionais. Han, Kamber e Pei (2012, p. 2) afirmam que a "mineração de dados pode ser vista como o resultado da evolução natural da tecnologia da informação". Os autores fazem um apanhado histórico da evolução da tecnologia de sistemas de bancos de dados (Quadro 9) e explicam que

> A indústria de banco de dados e do gerenciamento de dados evoluiu no desenvolvimento de várias funcionalidades críticas: coleta de dados e criação de banco de dados, gerenciamento de dados (incluindo armazenamento e recuperação de dados e processamento de transações de banco de dados) e análise avançada de dados (envolvendo *Data Warehousing* e *Data Mining*). O desenvolvimento inicial dos mecanismos de coleta de dados e de criação de bancos de dados serviu como um pré-requisito para o desenvolvimento posterior de mecanismos eficazes de armazenamento e recuperação de dados, bem como processamento de consultas e transações. Atualmente, vários sistemas de bancos de dados oferecem processamento de consultas e transações como prática comum. A análise avançada de dados tornou-se naturalmente o próximo passo. (HAN, KAMBER e PEI, 2012, p. 2-3, tradução nossa).

O progresso descrito pelos autores é detalhado no quadro abaixo:

Quadro 9 - Evolução da tecnologia de sistemas de Banco de Dados

Coleta de Dados e Criação de Banco de Dados (1960 e mais cedo)	▪ Processamento primitivo de arquivos
Sistemas de Gerenciamento de Banco de Dados (de 1970 a começo dos anos 80)	▪ Sistemas de bancos de dados hierárquicos e de rede ▪ Sistemas de banco de dados relacional ▪ Modelagem de dados: modelos de entidade-relacionamento, etc. ▪ Métodos de indexação e acesso ▪ Linguagens de consulta: SQL, etc. ▪ Interfaces de usuário, formulários e relatórios ▪ Processamento e otimização de consultas ▪ Transações, controle de concorrência e recuperação ▪ Processamento de Transações Online (OLTP)
Sistemas avançados de banco de dados (meados dos anos 80 até o presente)	▪ Modelos de dados avançados: estendido-relacional, objeto relacional, dedutivo, etc. ▪ Gerenciamento de dados complexos: espacial, temporal, multimídia, sequência e estruturado, científica, engenharia, objetos em movimento etc.

	▪ Fluxos de dados e sistemas de dados ciberfísicos ▪ Bancos de dados baseados na Web (XML, web semântica) ▪ Gerenciando dados incertos e limpeza de dados ▪ Integração de fontes heterogêneas ▪ Sistemas de banco de dados de texto e integração com recuperação de informação ▪ Gerenciamento de dados extremamente grandes ▪ Sistemas de ajuste de banco de dados e sistemas adaptativos ▪ Consultas avançadas ▪ Computação em nuvem e processamento de dados paralelo ▪ Questões de privacidade e segurança de dados
Análise avançada de dados (final dos anos 80 até o presente)	▪ *Data Warehouse* e Processamento Analítico *Online* (OLAP) ▪ Mineração de dados e descoberta de conhecimento: classificação, agrupamento, análise outlier, associação e correlação, comparativo resumo, análise de discriminação, padrão descoberta, análise de tendência e desvio, etc. ▪ Mineração de tipos complexos de dados: fluxos, sequência, texto, espacial, temporal, multimídia, Web, redes, etc. ▪ Aplicações de mineração de dados: negócios, sociedade, varejo, bancário, telecomunicações, ciência e engenharia, blogs, vida cotidiana etc. ▪ Mineração de dados e sociedade: dados invisíveis mineração, mineração de dados com preservação da privacidade, redes sociais e de informação de mineração, sistemas de recomendação etc.

Fonte: Adaptado de Han, Kamber e Pei, 2012 (Tradução nossa)

A mineração de dados surge então dos desafios que o tratamento de grandes volumes de dados trouxe às organizações. Segundo Dhar (2013), as técnicas tradicionais de exploração de dados não são mais adequadas para tratar a grande maioria dos repositórios de dados e, devido a isso, foi proposta, no final da década de 80, a Mineração de Dados (em inglês *Data Mining)*. O autor a descreve como o estudo da extração generalizada de conhecimento a partir de dados, um esforço sistemático que constrói e organiza conhecimento na forma de explicações e previsões testáveis. De Castro e Ferrari (2017, l. 610) trazem uma perspectiva mais ampla: “em tempos exponenciais, encontrar e acessar fontes de informação, pessoas, produtos e serviços não é mais problema; na verdade, o desafio atual é gerenciar, armazenar, processar e extrair conhecimento a partir dessa quantidade quase ilimitada de dados”.

No entanto, diante de diversas termos e tecnologias que lidam com dados, cabe pontuar que a mineração se relaciona, mas não se confundem com *Data Warehouse* (armazém de dados), *Business Intelligence* (inteligência empresarial), Inteligência Artificial, *Machine*

Learning (aprendizado de máquina), *Big Data*, *Data Science* (ciência de dados) e *Internet Of Things* (internet das coisas).

Então, como podemos definir a Mineração de Dados? Hand, Mannila e Smyth (2001) a definem como a análise de conjuntos de dados (geralmente grandes) para encontrar relações insuspeitas e resumir os dados de maneiras novas que sejam compreensíveis e úteis para o proprietário dos dados. Edelstein (1999) diz que a mineração de dados é um processo que usa uma variedade de ferramentas de análise de dados para descobrir padrões e relacionamentos em dados que podem ser usados para fazer previsões válidas.

De acordo com De Castro e Ferrari (2017), trata-se de um processo sistemático, interativo e iterativo, de preparação e extração de conhecimentos a partir de grandes bases de dados. Segundo os autores, o termo foi cunhado como alusão ao processo de mineração, uma vez que se explora uma base de dados (mina) usando algoritmos (ferramentas) adequados para obter conhecimento (minerais preciosos). Silva, Peres e Boscarioli (2017) a definem como um processo automatizado de explorar analiticamente grandes bases de dados, com a finalidade de descobrir padrões relevantes que ocorrem nos dados e que sejam importantes para embasar a assimilação de informação importante, suportando a geração de conhecimento.

Silva, Peres e Boscarioli (2017) afirmam ainda que não se trata então de um processamento bruto, meramente calculista, mas com características semânticas, em que os algoritmos são as ferramentas que possibilitam a transformação de dados brutos em conhecimento útil. Para tal, faz uso de conceitos provenientes de diversas áreas, Inteligência Artificial, Aprendizado de Máquina, Estatística, Banco de Dados (SILVA; PERES; BOSCARIOLI, 2017; MCKINSEY, 2011), Engenharia, Matemática, Sistemas de Informação (DE CASTRO e FERRARI, 2017), além de exigir integração de outras áreas como Ciências da Computação, Linguística, Econometria, Sociologia e qualquer outra necessária para que as interpretações façam sentido (DHAR, 2013).

A noção de processo é indissociável da mineração de dados, que pode ser utilizada para diversas finalidades. Neste estudo, a mineração será utilizada para identificar padrões repetitivos que caracterizem um comportamento não humano em pregões eletrônicos através do processo sistemático chamado Descoberta de Conhecimento em Bases de Dados (do inglês *Knowledge Discovery in Databases*, ou KDD).

6.1 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

Hand, Mannila e Smyth (2001) comentam que não é fácil definir os limites conceituais do que é mineração de dados do processo KDD. Afirmam que, para muitas pessoas, a transformação de dados (limpeza de dados, verificação de dados e definição de variáveis) é uma parte intrínseca da mineração de dados, enquanto que eles abordam o tema sob uma visão mais estrita, como sendo uma etapa do processo macro de descoberta de conhecimento, focando na modelagem de dados e na construção de processos algorítmicos.

De Castro e Ferrari (2017) trazem a existência dessa controvérsia afirmando que o termo mineração de dados é comumente utilizado para denominar o processo de transformação por completo, mas que se trata mais precisamente de uma das etapas do KDD. Segundo os autores, o KDD é um processo que compreende: a) seleção e integração das bases de dados, b) limpeza da base, seleção e transformação dos dados, c) mineração dos dados e d) avaliação dos dados.

Há ainda autores como Kum, Paulsen e Wang (2001) que definem o processo de KDD como *Knowledge Discovery and Data mining* (ou Descoberta de Conhecimento e Mineração de Dados, em tradução nossa). Para este estudo, o os termos mineração de dados e KDD serão tratados como sinônimos (exceto quando discriminado e no capítulo 7.3). Essa visão mais ampla da mineração de dados é trazida por Han, Kamber e Pei (2012): a mineração de dados é o processo de descobrir padrões e conhecimentos interessantes a partir de grandes quantidades de dados, que podem ter origens diversas, como bancos de dados, *Data Warehouse*, a *web* e outros repositórios de informações ou dados que são transmitidos dinamicamente para o sistema.

Em relação às aplicações, Han, Kamber e Pei (2012) dizem que, como tecnologia geral, a mineração de dados pode ser aplicada a qualquer tipo de dados, desde que sejam significativos para uma aplicação de destino. As formas mais básicas de dados para aplicativos de mineração são dados do banco de dados, dados do *Data Warehouse* e dados transacionais.

Nesse ponto, cabe distinguir dado, informação e conhecimento, três conceitos elementares da Ciência da Informação. Zins (2007) afirma que esses três conceitos, o de dados, de informação e de conhecimento, são fundamentais no contexto da Ciência da Informação, sendo frequentemente considerados como os blocos de construção básicos do campo. O autor define que, no domínio objetivo (em oposição à perspectiva subjetiva, que

tem o homem como elemento central), os dados são conjuntos de signos que representam estímulos empíricos ou percepções, a informação é um conjunto de signos, que representam conhecimento empírico, e conhecimento é um conjunto de signos que representam o significado (ou o conteúdo) dos pensamentos que o indivíduo acredita justificadamente que são verdadeiras.

Silva, Peres e Boscarioli (2017) trazem uma perspectiva linear, considerando os três conceitos como parte de um processo:

> O **dado** é um fato, um valor documentado ou um resultado de medição. Quando um sentido semântico ou um significado é atribuído aos dados, gera-se **informação**. Quando estes significados se tornam familiares, ou seja, quando um agente os aprende, este se torna consciente e capaz de tomar decisões a partir deles, e surge o **conhecimento** (SILVA; PERES; BOSCARIOLI, 2017, l. 384).

Zins (2007) comenta que, assim como Silva, Peres e Boscarioli (2017), muitos estudiosos afirmam que dados, informações e conhecimento são parte de uma ordem sequencial. Tratam os dados como a matéria-prima para a informação, e a informação como a matéria-prima para o conhecimento. Zins critica esse raciocínio ao dizer que se fosse assim, a Ciência da Informação não exploraria o conhecimento, pois seria uma entidade de ordem superior, o que parece não ocorrer devido à existência de subcampos como a organização do conhecimento e a gestão do conhecimento. No entanto, para este estudo, será adotada a lógica sequencial por colaborar com o raciocínio processual da mineração de dados, em que há a seleção e limpeza de dados brutos, transformação e processamento dos dados para a obtenção de informações que formem conhecimentos e embasem ações e tomadas de decisão.

Em relação às etapas do processo, Fayyad, Piatetsky-Shapiro e Smyth (1996a, p. 30-31, tradução nossa, grifo nosso), traz em detalhes um processo de KDD constituído em 9 etapas:

> 1) **Aprendendo o domínio da aplicação**: inclui conhecimento prévio relevante e os objetivos da aplicação.
> 2) **Criando um conjunto de dados de destino**: inclui a seleção de um conjunto de dados ou focar em um subconjunto de variáveis ou amostras de dados no qual a descoberta deve ser realizada.
> 3) **Limpeza de dados e pré-processamento**: inclui operações básicas, como a remoção de ruídos ou *outliers* [do inglês, valores atípicos] se apropriado, coletando as informações necessárias para modelar ou contabilizar o ruído, decidindo sobre estratégias para lidar com campos de dados perdidos e respondendo por informações de sequência de tempo e mudanças conhecidas, bem como decidindo sobre problemas de DBMS [do inglês, Sistema de Gerenciamento de Banco de Dados], como tipos de dados, esquema e mapeamento de valores ausentes e desconhecidos

> 4) **Redução e projeção de dados**: inclui encontrar recursos úteis para representar os dados, dependendo do objetivo da tarefa, e usar métodos de redução ou transformação de dimensionalidade para reduzir a número efetivo de variáveis sob consideração ou para encontrar representações invariantes para os dados.
> 5) **Função de mineração de dados**: inclui a decisão da finalidade do modelo derivado pelo algoritmo de mineração de dados (por exemplo, sumarização, classificação, regressão e *clustering* [ou análise de agrupamento de dados]).
> 6) **Escolhendo o(s) algoritmo(s) de mineração de dados**: inclui selecionar método (s) a ser usado para pesquisar padrões nos dados, como decidir quais modelos e parâmetros podem ser apropriados (por exemplo, modelos para dados categóricos são diferentes de modelos em vetores sobre reais) e corresponder um método de mineração de dados específico com os critérios gerais do processo KDD (por exemplo, o usuário pode estar mais interessado em entender o modelo do que em suas capacidades preditivas.
> 7) **Mineração de dados**: inclui a busca por padrões de interesse em uma forma representacional específica ou um conjunto de tais representações, incluindo regras de classificação ou árvores, regressão, agrupamento, modelagem de sequência, dependência e análise de linha.
> 8) **Interpretação**: inclui interpretar os padrões descobertos e possivelmente retornar a qualquer uma das etapas anteriores, bem como a possível visualização dos padrões extraídos, removendo padrões redundantes ou irrelevantes e traduzindo os padrões úteis em termos compreensíveis pelos usuários.
> 9) **Usando o conhecimento descoberto**: inclui incorporar esse conhecimento ao sistema de desempenho, tomando ações baseadas no conhecimento, ou simplesmente documentá-lo e relatá-lo às partes interessadas, bem como verificar e resolver potenciais conflitos com conhecimento previamente acreditado (ou extraído).

Han, Kamber e Pei (2012, p. 6/8, grifo do autor, tradução nossa, grifo do autor) descrevem de forma mais breve o processo de descoberta de conhecimento como uma sequência interativa das seguintes etapas:

> 1) **Limpeza de dados**: para remoção de ruído e dados inconsistentes;
> 2) **Integração de dados**: onde várias fontes de dados podem ser combinadas;
> 3) **Seleção de dados**: onde os dados relevantes para a tarefa de análise são recuperados do banco de dados;
> 4) **Transformação de dados**: em que os dados são transformados e consolidados em formulários apropriados para mineração, executando operações de resumo ou agregação;
> 5) **Mineração de dados**: processo essencial em que métodos inteligentes são aplicados para extrair padrões de dados;
> 6) **Avaliação de padrões**: para identificar os padrões verdadeiramente interessantes que representam o conhecimento com base em medidas de interesse); e
> 7) **Apresentação do conhecimento**: onde técnicas de visualização e representação de conhecimento são usadas para apresentar conhecimento minado aos usuários.

De Castro e Ferrari (2017, l. 660-673, grifos do autor) descrevem o processo de KDD como um processo de quatro etapas:

> 1) **Base de dados**: coleção organizada de dados, ou seja, valores quantitativos ou qualitativos referentes a um conjunto de itens, que permite uma recuperação eficiente dos dados. Conceitualmente, os dados podem ser entendidos como o nível

> mais básico de abstração a partir do qual a informação e, depois, os conhecimentos podem ser extraídos [...];
> 2) **Preparação ou pré-processamento de dados**: são etapas anteriores à mineração que visam preparar os dados para uma análise eficiente e eficaz. Essa etapa inclui a limpeza (remoção de ruídos e dados inconsistentes), a integração (combinação de dados obtidos a partir de múltiplas fontes), a seleção ou redução (escolha dos dados relevantes à análise) e a transformação (transformação ou consolidação dos dados em formatos apropriados para a mineração);
> 3) **Mineração de dados**: essa etapa do processo corresponde à aplicação de algoritmos capazes de extrair conhecimentos a partir dos dados pré-processados. [...] [Inclui] técnicas de análise descritiva (medidas de distribuição, tendência central e variância, e métodos de visualização), agrupamento (segmentação de bases de dados), predição (classificação e estimação), associação (determinação de atributos que coocorrem) e detecção de anomalias; e
> 4) **Avaliação ou validação do conhecimento**: avaliação dos resultados da mineração objetivando identificar conhecimentos verdadeiramente úteis e não triviais.

Silva, Peres e Boscarioli (2017, l. 518) adotam uma visão semelhante ao descrever o processo de KDD em quatro partes: Obtenção de Dados, Pré-Processamento, Mineração de Dados e Pós-Processamento, descrevendo-o da seguinte forma:

> O processo KDD inicia com a organização das bases de dados que contêm os dados da área de interesse, a partir dos quais se deseja descobrir algum tipo de conhecimento útil. Esses dados passam por uma série de procedimentos de pré-processamento que inclui, mas não se limita a [...]: organização em um repositório único (por exemplo, em um *Data Warehouse*); tratamento para eliminação de instâncias repetidas e ou valores discrepantes; procedimentos de seleção de dados que permitem que se escolha apenas os exemplares e/ ou os atributos relevantes para a mineração de dados; normalizações para que os valores dos atributos estejam em uma mesma escala e sejam considerados com a mesma relevância pelos algoritmos de análise. Com os dados pré-processados, inicia-se a tarefa de mineração de dados propriamente dita, envolvendo a resolução de tarefas como predição, agrupamento ou associação. Por fim, os resultados obtidos são validados, avaliados e formatados em gráficos, tabelas e relatórios estruturados. Esse processo pode ser iterativo e interativo, ou seja, cada fase pode ser executada mais de uma vez, na sequência usual ou fora dela, a depender do conjunto de dados original e/ ou de decisões tomadas pelo analista (especialista no domínio dos dados).

Para este estudo, será adotada a metodologia descrita por De Castro e Ferrari (2017), tendo em vista que atende aos requisitos de simplicidade, flexibilidade e objetividade na aplicação em trabalhos acadêmicos. É simples e flexível por ser composta por apenas quatro macroetapas, permitindo que se molde às necessidades de cada aplicação. Torna-se objetiva por não abranger etapas de cunho estratégico e organizacional ou de interesse mercadológico (concernentes a aplicações empresariais), partindo já de uma base de dados organizada, como são os repositórios de dados abertos. Os requisitos adotados visam reduzir a complexidade metodológica, propiciando condições favoráveis para que o trabalho seja testado, reutilizado, aprofundado ou modificado por outros pesquisadores e cidadãos de interesse.

6.2 OUTRAS ABORDAGENS DE EXPLORAÇÃO DE DADOS

Existem diversas outras metodologias ou processos para abordar a mineração de dados, como a CRISP–DM (*Cross-Industry Standard Process for Data Mining*), a SEMMA (*Sample, Explore, Modify, Model, Assess*), entre outras. Shearer (2000, p. 13) diz que o CRISP-DM foi concebido no final de 1996 por um consórcio entre quatro líderes do mercado de mineração de dados na época: Daimler-Benz (agora DaimlerChrysler), Integral Solutions Ltd. (ISL), NCR e OHRA Shearer. Foi projetado para ser neutro em termos de indústria, ferramentas e aplicativos, por isso recebeu informações de uma ampla gama de profissionais e outros (como fornecedores de *data warehouse* e consultores de gerenciamento) com interesse na mineração de dados. "O CRISP-DM organiza o processo de mineração de dados em seis fases: compreensão de negócios, compreensão de dados, preparação de dados, modelagem, avaliação e implantação" (SHEARER, 2000, p. 13). SEMMA é um acrônimo usado para descrever o processo de mineração de dados da empresa SAS *Inc*. e significa, em tradução nossa, Amostra, Explorar, Modificar, Modelar e Avaliar.

O processo descrito por Eldelstein (1999, p. 22-33, tradução nossa) ilustra como os processos para descoberta de conhecimento podem se moldar às necessidades. Foi desenvolvido por uma consultoria (a *The Two Crows*) e é mais amplo que o KDD, pois envolve etapas de concepção dos problemas que se pretende abordar e toda a preparação para a construção da base de dados, com influência da metodologia CRISP-DM, como demonstrado no Quadro 10:

Quadro 10 - Modelo de processo *The Two Crows*

Etapa	Descrição
Definição do problema de negócio	Em primeiro lugar, o pré-requisito para a descoberta de conhecimento é entender seus dados e seus negócios. Sem esse entendimento, nenhum algoritmo, independentemente de sofisticação, fornecerá um resultado no qual você deve ter confiança. Sem esse plano de fundo, você não poderá identificar os problemas que está tentando resolver, preparar os dados para mineração ou interpretar corretamente os resultados. Para fazer o melhor uso da mineração de dados, você deve fazer uma declaração clara de seus objetivos. [...]
Construção do banco de dados	Este passo, juntamente com os próximos dois, constitui o núcleo da preparação de dados. Juntos, eles levam mais tempo e esforço do que todos os outros passos combinados. Pode haver repetidas iterações das etapas de preparação de dados e criação de modelo à medida que você aprende algo com base no modelo que sugere a modificação dos dados. Essas etapas de preparação de dados podem levar de 50% a 90% do tempo e esforço de todo o processo de descoberta de conhecimento. [...] As tarefas na criação de um banco de dados de mineração de dados são: a) coleção de dados; b) descrição de dados; c) seleção; d) avaliação da qualidade de dados e limpeza de dados; e) consolidação e integração; f) construção de metadados; g) carregar o banco de dados de mineração de dados; e h) manter o banco de dados de mineração de dados. [...]
Exploração dos	O objetivo é identificar os campos mais importantes na previsão de um resultado e

dados	determinar quais valores derivados podem ser úteis. Em um conjunto de dados com centenas ou até milhares de colunas, a exploração dos dados pode ser tão demorada e trabalhosa quanto é esclarecedora.
Preparação dos dados para modelagem	Esta é a etapa final de preparação de dados antes de criar modelos. Existem quatro partes principais para esta etapa: a) selecionar as variáveis; b) selecionar linhas; c) construir novas variáveis; e d) transformar variáveis. [...]
Construção do modelo	A coisa mais importante a lembrar sobre a construção de modelos é que é um processo iterativo. Você precisará explorar modelos alternativos para encontrar aquele que é mais útil na solução de seu problema de negócios. O que você aprende na busca de um bom modelo pode levá-lo a voltar e fazer algumas alterações nos dados que você está usando ou até mesmo modificar sua declaração de problema. [...]
Avaliação do modelo	**Validação do modelo**. Depois de construir um modelo, você deve avaliar seus resultados e interpretar seu significado. Lembre-se de que a taxa de precisão encontrada durante o teste se aplica apenas aos dados em que o modelo foi criado. Na prática, a precisão pode variar se os dados aos quais o modelo é aplicado diferirem de maneiras importantes e desconhecidas dos dados originais. Mais importante, a precisão por si só não é necessariamente a métrica certa para selecionar o melhor modelo. Você precisa saber mais sobre o tipo de erros e os custos associados a eles. [...] **Validação externa**. Como apontado acima, não importa quão boa seja a precisão de um modelo, não há garantia de que ele reflita o mundo real. Um modelo válido não é necessariamente um modelo correto. Uma das principais razões para esse problema é que sempre há suposições implícitas no modelo. Por exemplo, a taxa de inflação pode não ter sido incluída como uma variável em um modelo que prevê a propensão de um indivíduo comprar, mas um salto na inflação de 3% para 17% certamente afetará o comportamento das pessoas. Além disso, os dados usados para construir o modelo podem não corresponder ao mundo real de alguma forma desconhecida, levando a um modelo incorreto. Portanto, é importante testar um modelo no mundo real. [...]
Implantação modelo e resultados	Uma vez que um modelo de mineração de dados é construído e validado, ele pode ser usado de duas maneiras principais. A primeira maneira é que um analista recomende ações com base na simples visualização do modelo e seus resultados. [...] A segunda maneira é aplicar o modelo a diferentes conjuntos de dados. O modelo pode ser usado para sinalizar registros com base em sua classificação ou atribuir uma pontuação, como a probabilidade de uma ação [...].

Fonte: Adaptado de Eldelstein, 1999 (Tradução nossa)

Azevedo e Santos (2008) fazem uma comparação entre as metodologias KDD, CRISP-DM e SEMMA e concluem que todas são equivalentes e consideram que a SEMMA e a CRISP-DM podem ser vistas como uma implementação do processo de KDD descrito por Fayyad, Piatetsky-Shapiro e Smyth (1996b), conforme sumarizado no Quadro 11.

Quadro 11 - Correspondências entre KDD, SEMMA e CRISP-DM

KDD	SEMMA	CRISP-DM
Pré-KDD	-	Compreensão de Negócios
Seleção	Amostra	Compreensão de Dados
Pré-processamento	Explorar	
Transformação	Modificar	Preparação de Dados
Mineração de Dados	Modelar	Modelagem
Interpretação/Avaliação	Avaliar	Avaliação
Pós-KDD	-	Implantação

Fonte: Azevedo e Santos, 2008.

A adoção do KDD como metodologia de exploração de dados se dá por ser a mais abordada e utilizada no meio acadêmico, não havendo diferenças significativas que justifiquem a adoção da CRISP-DM, da SEMMA ou de outra mais específica. No próximo capítulo será apresenta a realização da descoberta de conhecimentos em base de dados, divido em quatro seções representando cada uma das etapas do KDD proposto por De Castro e Ferrari (2017).

7 RESULTADOS

Os resultados serão apresentados de acordo com a estrutura da Descoberta de Conhecimentos em Base de Dados descrita por De Castro e Ferrari (2017), com a terminologia e processo elaborado no RapidMiner (RM). Serão inicialmente descritas as características da base de dados e como foi feita a importação para o RM. Depois serão detalhadas as etapas de pré-processamento e de mineração de dados. Em seguida serão apresentados observações e gráficos relevantes à pesquisa. Buscou-se criar um processo completo de tratamento dos dados a partir do arquivo original fornecido pelo governo federal, sendo também objeto de análise alguns aspectos referentes à execução da política de dados abertos observada. O intuito é auxiliar na replicação dos procedimentos aos interessados, pesquisadores, agentes de governo, fornecedores ou cidadãos no simples exercício do controle social.

7.1 BASE DE DADOS

O arquivo contendo a base de dados foi obtida sob o nome de "DadosLanceMapa.csv", apresentando tamanho de 2,12 GB, com 6.124.274 linhas, correspondendo aos registros de lance emitido em pregões eletrônicos realizados no âmbito do MAPA em 2017. Trata-se de um conjunto de dados (*dataset*) estruturados, dispostos em 33 colunas (ou atributos), cujas funções identificadas no dicionário de dados disponibilizado pelo SERPRO (2019) estão descritas no Quadro 12.

O RapidMiner eliminou automaticamente linhas com erros detectados na importação dos dados, apresentando como resultado um total de 6.117.559 linhas, o que representa uma redução de 6.715 linhas, aproximadamente 1,1% da amostra original. O arquivo foi importado sob codificação UTF-8, utilizando o símbolo ponto e vírgula (;) como delimitador. Todos os atributos foram inicialmente tratados como nominais (*polynominal*), possibilitando verificar análises descritivas disponibilizadas na função *Turbo Prep* do RM para itens não numéricos (*category*), como a contagem de cada valor e seus respectivos percentuais em cada atributo. Importar todos os atributos do *dataset* como *polynominal* para o RM é indispensável para que o processo seja reproduzido.

Quadro 12 - Dicionário dos dados

Atributo	Descrição
ch_lance	Não identificado em SERPRO (2019), nem similares. Não foi possível interpretar esse atributo
prpCod	Código da proposta a que este item de proposta está ligado
coduasg	Código da UASG do pregão
numprp	Número do pregão
codItemCompra	Não identificado em SERPRO (2019). Interpretou-se como representação de condições especiais do item
Data_Compra	Não identificado em SERPRO (2019)
Unidade_solicitante	Não identificado em SERPRO (2019). Interpretou-se como código da UASG solicitante de pregões em sistema de registro de preços
UASG_unidade	Não identificado em SERPRO (2019). Interpretou-se como código da UASG do pregão
Orgao_superior	Não identificado em SERPRO (2019). Interpretou-se como código do órgão ao qual a UASG pertence
CMPJ_Empresa	Não identificado em SERPRO (2019). Interpretou-se como CNPJ do fornecedor que incluiu a proposta
Nome_Empresa	Não identificado em SERPRO (2019). Interpretou-se como razão social do fornecedor que incluiu a proposta
Porte_Empresa	Identificado por aproximação ao atributo *PorteEmpresa* em SERPRO (2019). Indicativo do porte da empresa. 0: grande empresa; 1: ME ou EPP; N: CNPJ não cadastrado na base da Receita Federal
prpIndParticiparConvocacao	Indica se a ME/EPP deseja participar da fase de convocação. N: não; S: sim
ippIndHabilitado	Indicador de que a proposta está habilitada. Nulo: não avaliada; H: habilitada; I: inabilitada
ippIndAceito	Indicador de que a proposta está aceita. Nulo: não avaliada; A: aceita; R: recusada
ippIndAdjudicado	Indicador de que a proposta está adjudicada. Nulo: Não avaliada; S: adjudicada; N: recusada
ippObsAceitacao	Qualquer informação que necessite ser adicionada durante a fase de aceitação
ippObsAdjudicacao	Qualquer informação que necessite ser adicionada durante a fase de adjudicação
ippDeclaracaoProdNacional	Indicativo se o fornecedor possui sócios e/ou dirigentes em comum com outros fornecedores que apresentaram proposta para o mesmo item da licitação
ippIndTP	Indicador de processo produtivo básico (usado no desempate do Decreto 7.174/2010). S: sim; N: não
ippOrdemConvocacao	Ordem de convocação da ME/EPP que satisfaz ao critério de proposta superior até 5% do preço vencedor
ippLanceMin	Lance mínimo feito pelo fornecedor para este item
lanCod	Código de identificação único do lance gerado na gravação do registro
ipgCod	Código que identifica o item de um determinado pregão
lanData	Data e hora que o lance foi realizado
lanValor	Valor do lance efetuado pelo fornecedor
lanStatus	Indica se o lance foi excluído ou não. V: lance válido (padrão); *: lance excluído pelo pregoeiro; N: lance negociado
lanClassif	Este campo armazena a mesma informação da coluna *lanValor* para fins de classificação, porém, para itens por maior desconto, este valor

	será negativo, e para pregão por ICMS, este valor será equalizado, permanecendo o mesmo valor para os demais casos
lanIndConvocacao	Indicador de Convocação. Quando for dado um lance na fase de convocação (desempate), é atribuído o valor (S).
lanDataInsHist	Código de identificação único do lance gerado na gravação do registro
DataCadastro	Data e hora da inclusão do cadastro do fornecedor
CNPJ	CNPJ do fornecedor que incluiu a proposta
RazaoSocial	Razão social da empresa

Fonte: Elaborado pelo Autor com base em SERPRO (2019).

Cabe aqui ressaltar aspectos relacionados à política de dados abertos observados no processo de obtenção dos dados. O primeiro é que o arquivo foi disponibilizado sem dicionário de dados para a interpretação dos códigos que denominam os atributos, que poderiam, ser enviados pelo e-mail cadastrado na requisição formal, por um arquivo separado na mídia física em que os dados foram enviados ou na resposta à requisição formal. Enquanto alguns nomes são facilmente identificáveis, outros não. A ausência de um dicionário foi um obstáculo inicial que causou demora e erros iniciais devido a interpretações equivocadas.

Foi localizado o recurso "Metadados - Comprasnet" no portal da empresa SERPRO (2019), desenvolvedora e mantenedora da plataforma ComprasNet. O recurso apresenta 1248 atributos utilizados no Comprasnet, descrevendo a base de dados e a tabela originária, o nome da coluna, sua descrição e o tipo de dados (podendo ser *bigdecimal*, *boolean*, *byte*, *char*, *clob*, *integer*, *short*, *string* ou *timestamp*). Dentre as 33 colunas originais, apenas a denominada "ch_lance" não constava no dicionário nem foram identificados atributos similares, motivo pelo qual foi descartada durante o tratamento dos dados.

Outra dificuldade relacionada à política de dados abertos é que o pedido é feito de forma única e unilateral, sem diálogo para refinamento do pedido. A seleção dentre os mais de 1200 atributos disponíveis depende da compreensão que o agente que realiza a extração dos dados tem sobre a finalidade do pedido, por isso um canal aberto e dinâmico pode suprir deficiências das requisições causadas por incompreensão de aspectos técnicos, por exemplo. Existe uma API[2] (do inglês, Interface de Programação de Aplicativos) para que qualquer pessoa possa fazer extrações de dados do Comprasnet, mas que apresenta um espectro limitado de informações. A ampliação dos dados disponíveis para autoextração também poderia contribuir com a melhoria da transparência dos dados das compras governamentais federais.

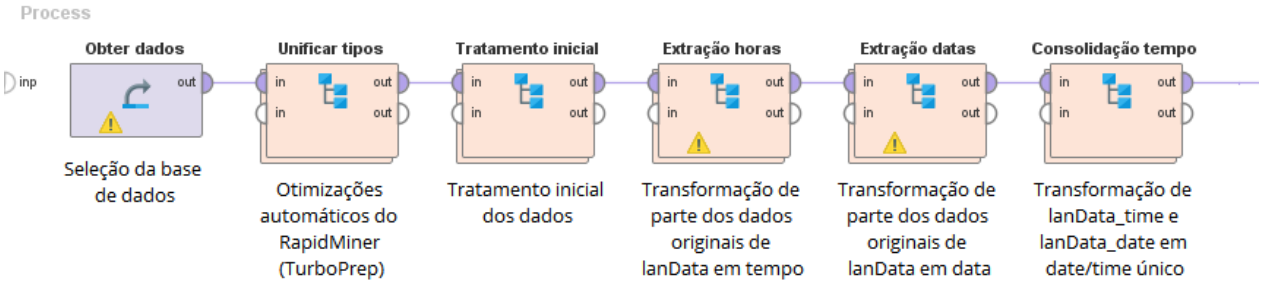
[2] Disponível em http://compras.dados.gov.br/docs/home.html.

7.2 PRÉ-PROCESSAMENTO DOS DADOS

Os procedimentos relacionados ao pré-processamento de dados envolvem a limpeza, redução e transformação de valores no sentido de ajustá-los para que pudessem receber manipulações relacionadas à mineração de dados propriamente dita, estando o processo no RapidMiner representado na Figura 2. O processo é representado por operadores, retângulos coloridos conectados através de portas por conexões entre si e para saídas. Os operadores mais comuns desta etapa são os do tipo *Blending* (que manipulam nomes, tipos e valores dos atributos), e o operador processual de criação de *Subprocess* (que permite organizar o fluxo do processo e é representado por dois retângulos sobrepostos). Foi utilizado também o *Retrieve* (recuperação), que traz ao processo a base de dados selecionada.

Cabe ressaltar que todos os alertas do RM, representados por ponto de exclamação em ícone amarelo, foram analisados e tratados quando considerados prejudiciais ao estudo. Portanto, os remanescentes são considerados irrelevantes. Ainda, é necessário o número de dígitos fracionários no RM esteja configurado para pelo menos 7 devido às características do conjunto de dados. Abaixo disso, os resultados da amostra são afetados. A configuração original do RM é 3 e foi alterada através do menu (em *Settings/Preferences/General/Fractions digits os numbers)*.

Figura 2 - Etapas de pré-processamento dos dados



Fonte: Elaborado pelo Autor, 2019.

Segundo De Castro e Ferrari (2017), são etapas anteriores à mineração que visam preparar os dados para uma análise eficiente e eficaz. Essa etapa inclui a limpeza (remoção de ruídos e dados inconsistentes), a seleção ou redução (escolha dos dados relevantes à análise) e a transformação (transformação ou consolidação dos dados em formatos apropriados para a mineração). O objetivo desta etapa é preparar os dados brutos para serem analisados. Os autores afirmam ainda que esta etapa normalmente consome muito tempo e demandam

bastante trabalho para a maioria das bases de dados reais, mas que o sucesso da mineração depende da qualidade dos resultados dessa atividade. Segundo De Castro e Ferrari (2017), ferramentas simples como histogramas ou gráficos, que possibilitem a visualização da distribuição de valores de um atributo ou que permita a comparação entre vários, são bastante úteis para um entendimento inicial da base de dados.

a) Obter dados

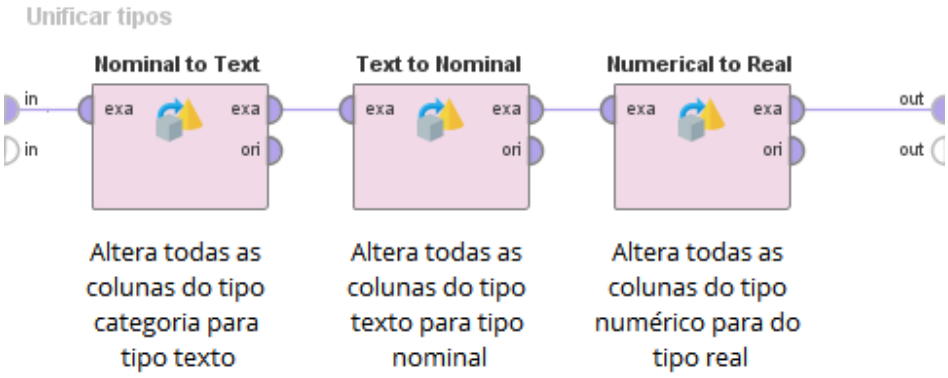
A primeira etapa do pré-processamento foi realizar a importação de todos as colunas como nominais, em que os dados são interpretados pelo *software* como nome, possibilitando categorizar aqueles dados. Apesar de computacionalmente mais custoso, esse procedimento permitiu a visualização de algumas informações relacionadas a distribuição, quantidade e frequência dos valores, pois colunas do tipo numéricas apresentam apenas métricas aritméticas como média aritmética, mínimo, máximo, média, entre outros. Essa análise inicial de distribuição dos valores permitiu uma melhor compreensão do que cada atributo representa e qual sua utilidade para a pesquisa, além de permitir a eliminação de atributos com dados duplicados.

Foram realizadas extrações de amostras de 10% (611.755 linhas) e de 1% (61.175 linhas) do total de linhas originais para facilitar a manipulação dos dados, utilizadas durante a elaboração das etapas de pré-processamento e de mineração dos dados. Com isso, obteve-se uma redução da demanda de recursos computacionais e consequentemente do tempo de processamento de cada teste. O processo de extração de amostras foi realizado pelo próprio RM através da funcionalidade de tratamento automatizado de dados denominada *Turbo Prep*.

b) Unificar tipos

As operações de pré-processamento realizadas pela função *Turbo Prep* geram automaticamente processos contendo as seguintes otimizações iniciais (Figura 3), que visam padronizar tipos de atributos similares: *Nominal to Text*, que altera todas as colunas do tipo categoria para tipo texto; *Text to Nominal*, que altera todas as colunas do tipo texto para tipo nominal; e *Numerical to Real*, que altera todas as colunas do tipo numérico para do tipo real.

Figura 3 - Subprocesso “Unificar tipos”



Fonte: Extraído do RapidMiner Studio.

c) Tratamento inicial

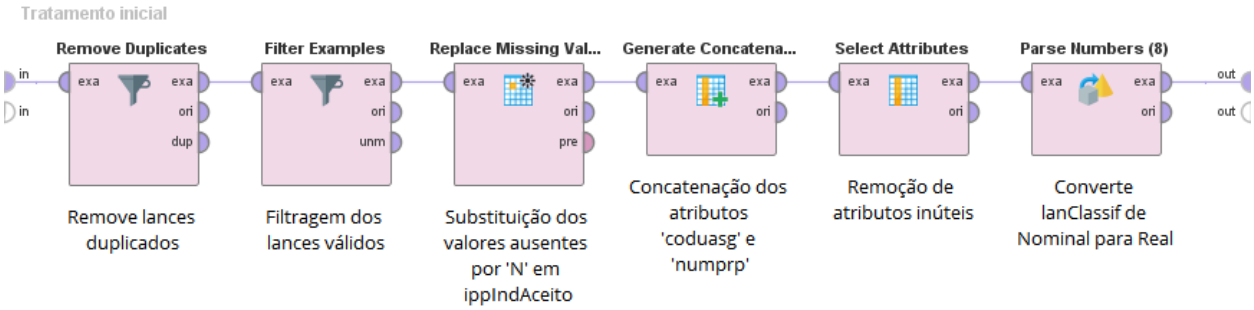
Inicialmente foi realizada uma redução (limpeza) através da utilização do operador “*Remove duplicates”* (remover duplicações) no atributo *lanCod*, que representa individualmente cada lance através de um código, como mostra a Figura 4. O resultado foi a redução para 76.372 linhas, aproximadamente 1,25% da quantidade original. Em análise, foi constatado que a coluna *codItemCompra* é a origem da grande quantidade de lances multiplicados, por representar as várias cotas de participação exclusiva de microempresas e empresas de pequeno porte em itens que utilizem tal benefício, instituído pelo artigo 48, inciso III, da Lei Complementar nº 123/2006 (BRASIL, 2006). Como efeito, alguns lances foram multiplicados pelo número de cotas de cada item que utilizavam esse benefício, o que geraria distorções nos resultados, motivo pelo qual a redução foi mantida.

O atributo *lanStatus*, segundo o dicionário disponibilizado por SERPRO (2019), indica se o lance foi excluído ou não pelo pregoeiro no momento da sessão pública, tendo como valores o “V” para lances válidos, “*” (asterisco) para lances excluídos e ainda “N” para lances negociados. No conjunto de dados original, não há incidência de valores “N” e apenas 0,08% de valores “*”, que foram eliminados para que a amostra representasse apenas os lances válidos.

O atributo *ippIndAceito* indica se a proposta foi aceita pelo pregoeiro, apresentando valor ausente quando não avaliado, “A” quando aceita e “R” quando recusada. Trata-se de um indicador relacionado à proposta de cada fornecedor e não a cada lance individualmente. Decorrente disso, por exemplo, todos os lances de um licitante que teve seu menor lance

aceito terão valor “A”, incluindo a proposta inicial, os lances intermediários e o seu menor lance.

Figura 4 - Subprocesso “Tratamento inicial”



Fonte: Extraído do RapidMiner Studio.

Os atributos *coduasg* e *numprp* representam respectivamente o código da UASG (Unidade Administrativa de Serviços Gerais) do pregão e o número do pregão. Optou-se pela concatenação dos valores desses dois atributos para reduzir o número de coluna no final e para que fosse entregue um valor único para cada pregão, tendo em vista que um número de pregão se repete em vários órgãos (a exemplo de 1/2017) e isso poderia afetar as análises. Como resultado, obtivemos o atributo “coduasg numprp”.

Com base na análise de distribuição dos valores realizada na etapa de Base de Dados do KDD, optou-se por excluir, do conjunto de dados, os atributos demonstrados no Quadro 13, pelos motivos especificados. Por fim, *lanClassif* é convertido de texto para número.

Quadro 13 - Atributos excluídos e motivos

Atributos	Motivo da exclusão
ch_lance	Atributo de significado não identificado. Não consta no dicionário de dados
CMPJ_Empresa	Correspondência de 100% com o atributo "CNPJ", que será mantido
codItemCompra	Sem valor para a pesquisa. Representa cotas de participação exclusiva de ME/EPP
coduasg	Foi concatenado a *numprp*
Data_Compra	Irrelevante ao estudo
DataCadastro	Relacionado ao licitante, sendo indesejável ao estudo
ipgCod	Irrelevante ao estudo
ippDeclaracaoProdNacional	Irrelevante ao estudo
ippIndAdjudicado	Valores semelhantes a *ippIndAceito*, sendo este mais relevante ao estudo
ippIndHabilitado	Valores semelhantes a *ippIndAceito*, sendo este mais relevante ao estudo
ippIndTP	Atributo com valor nulo em todas as linhas
ippLanceMin	Irrelevante ao estudo
ippObsAceitacao	Valores textuais, não considerados no estudo
ippObsAdjudicacao	Valores textuais, não considerados no estudo
ippOrdemConvocacao	Irrelevante ao estudo

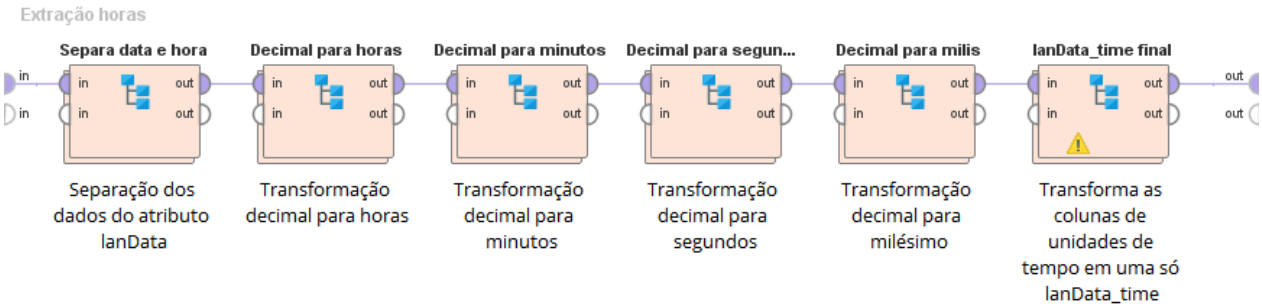
lanCod	Irrelevante ao estudo
lanDataInsHist	Relacionado ao licitante, sendo indesejável ao estudo
lanIndConvocacao	Irrelevante ao estudo
lanStatus	Já utilizado para filtragem de lances válidos, sem mais utilidade
lanValor	Valores semelhantes a *lanClassif*, sendo este mais relevante ao estudo
Nome_Empresa	Correspondência de 98,25% com o atributo "RazaoSocial", sendo este mais atualizado
numprp	Foi concatenado a *coduasg*
Orgao_superior	Continha o mesmo valor ("MINIST. DA AGRICUL. PECUARIA E ABASTECIMENTO") em 100% das linhas
Porte_Empresa	Relacionado ao licitante, sendo indesejável ao estudo
prpIndParticiparConvocacao	Irrelevante ao estudo
RazaoSocial	Relacionado ao licitante, sendo indesejável ao estudo
UASG_unidade	Correspondência de 100% com o atributo *coduasg*
Unidade_solicitante	Irrelevante ao estudo

Fonte: Elaborado pelo Autor, 2019.

d) Extração de horas

O atributo *lanData*, segundo o dicionário de dados, mostra a data e hora que o lance foi realizado, tendo como valor um *timestamp*, ou seja, uma representação numérica do tempo. No entanto existem diversas formas de representação e o dicionário de dados não dispõe desse detalhe. No caso do conjunto de dados em estudo, *lanData*, *lanDataInsHist* e *DataCadastro* são representados pela mesma estrutura de 5 dígitos numéricos iniciais separados por vírgula de um outro conjunto de caracteres de tamanho variado, ultrapassando 9 dígitos em sua maioria. Pesquisas e testes revelaram que a primeira parte representa a data, correspondendo à contagem linear de dias a partir do dia 1º de janeiro de 1900. A segunda parte corresponde à fração de um dia, com precisão superior à de milésimos de segundos. Ter campos com as variáveis de datas e horas padronizadas e que permitissem o tratamento dos dados de forma uniforme e consistente foi uma etapa fundamental para validar as hipóteses propostas nesta pesquisa.

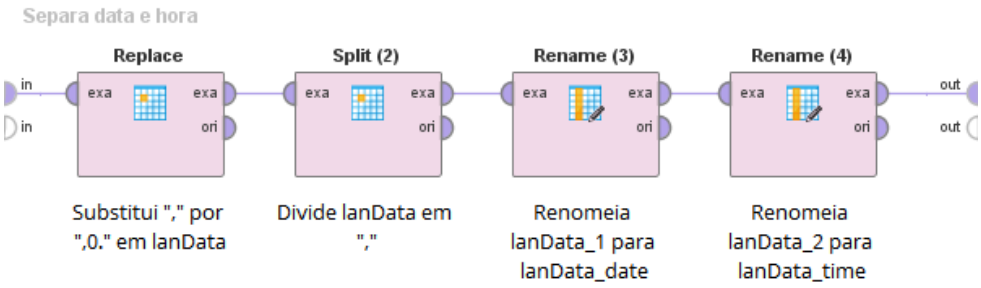
Figura 5 – Subprocesso “Extração horas”



Fonte: Extraído do RapidMiner Studio.

Para realizar a conversão dos valores em *timestamp* da coluna *lanData* (Figura 5), foi feita a separação na vírgula, tendo como resultado as colunas temporárias *lanData_time e lanData_date*, posteriormente concatenadas para uma nova coluna *lanData* com valor de data em calendário gregoriano e de tempo em horas, minutos, segundos e milésimos. O subprocesso de separação, denominado “*Separa data e hora”* e demonstrado na Figura 6, foi realizado através dos seguintes operadores: *Replace*, para substituir "," por ",0." em *lanData; Split*, para dividir *lanData* em ","; e *Rename*, para renomear *lanData_1* para *lanData_date* e *lanData_2* para *lanData_time* (o RapidMiner nomeia sequencialmente as colunas resultantes da divisão devido ao parâmetro *ordered_split* em *split mode*). Como exemplo, um dado *42760,472301389* no *lanData* original resultou em *42775* em *lanData_date* e *0.472301389 em lanData_time.*

Figura 6 – Subprocesso “Separa data e hora”



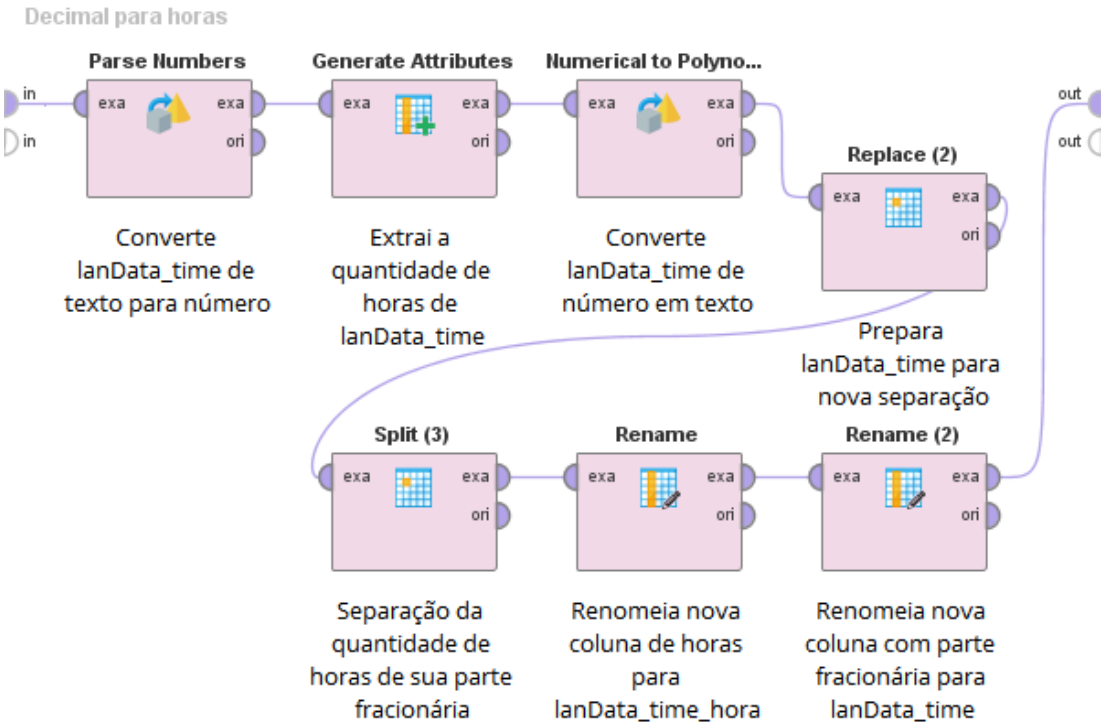
Fonte: Extraído do RapidMiner Studio.

Foram criados quatro subprocessos após a separação do *lanData* original para a conversão do tempo em formato de número decimal na nova coluna *lanData_time*, cada um contendo basicamente a mesma estrutura de etapas. Esclarecendo o formato de tempo utilizado na base de dados, a fração decimal de um dia implica dizer que *0.5* é equivalente a

12:00 horas, por exemplo. No exemplo já utilizado, *42760,472301389* resulta em *0.472301389* que equivale a 11:20:06.840 em horas, minutos, segundos e milissegundos.

Os subprocessos da transformação podem ser observados na Figura 7. Inicialmente é realizada a transformação do tipo da categoria de texto para número. Em seguida, é gerado um novo atributo com a extração das horas representada pelo número inteiro e restante da representação do tempo na parte decimal. Para tal, utiliza-se a expressão de função "*lanData_time* * 24". No exemplo já utilizado, *0.472301389* que equivale é transformado em *11.335233336*, em que 11 equivale a hora do dia em que o lance foi emitido. Optou-se por manter o padrão do RapidMiner de se utilizar o ponto (.) como separador decimal, pois a vírgula está sendo utilizada como caractere de separação de valores. Então é realizado um procedimento similar ao já descrito no subprocesso "*Split date and time"*, mas com algumas alterações, através dos seguintes operadores: *Replace*, para substituir dessa vez "." por ",0."*; Split*, para dividir os valores ","; *Rename*, para renomear *lanData_time_1* para *lanData_time_hora* e *lanData_2* para *lanData_time* (preparando essa última coluna para a próxima transformação).

Figura 7 - Subprocesso "Decimal para horas"

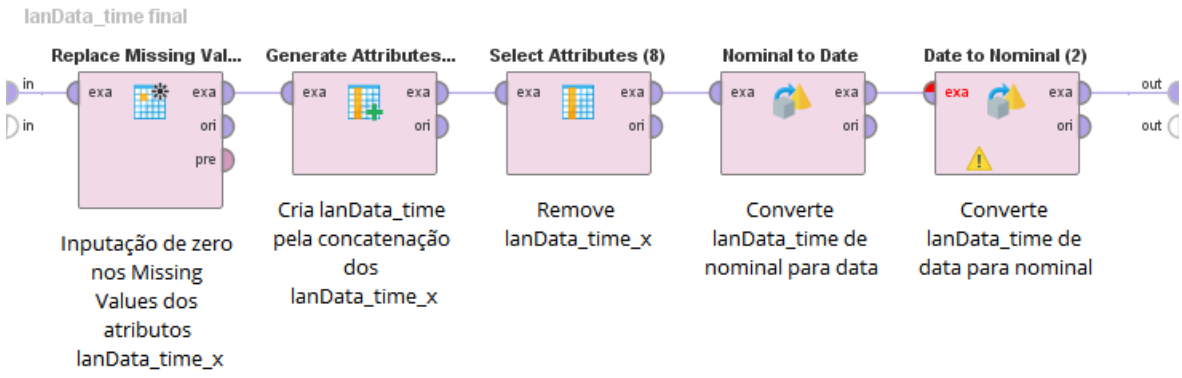


Fonte: Extraído do RapidMiner Studio.

A estrutura apresentada na Figura 7 foi utilizada para extrair também os minutos (subprocesso *Decimal to minutes*), os segundos (*Decimal to seconds*) e os milissegundos (*Decimal to millis*). Utilizou-se a expressão "*lanData_time* * 60" para minutos e segundos e "*lanData_time* * 1000" para os milissegundos. Como resultado, foram criadas as colunas *lanData_time_minutos*, *lanData_time_segundos* e *lanData_time_milis*. Por fim, em *Decimal to millis* foi utilizado o operador *Select Attributes* (com seleção invertida) para remover lanData_3 (coluna gerada automaticamente durante o uso do operador *Split*, sem valor para o estudo).

Após esses procedimentos, ocorreram algumas linhas com valor nulo (nos casos em que os lances foram registrados em unidades de tempo inteiras, a exemplo de quando os milissegundos ou milissegundos e segundos eram zero). Foi então realizada a imputação de valor "0" em todos os casos. Por fim, foi gerada uma nova coluna denominada *lanData_time* através da concatenação dos valores das colunas com horas, minutos, segundos e milissegundos pela utilização da expressão "concat([lanData_time_hora],":",[lanData_time_ minutos],":", [lanData_time_segundos],":",[lanData_time_milis])", com a posterior remoção dos atributos utilizados. Esse subprocesso, demonstrado na Figura 8, é finalizado com uma dupla conversão de *lanData_time* do tipo nominal para data e de data para nominal. A função da primeira conversão é padronizar todos os valores em três dígitos nos milissegundos, e da segunda, preparar para a futura concatenação com a coluna *lanData_date*, trabalhada em seguida.

Figura 8 - Subprocesso "lanData_time final"

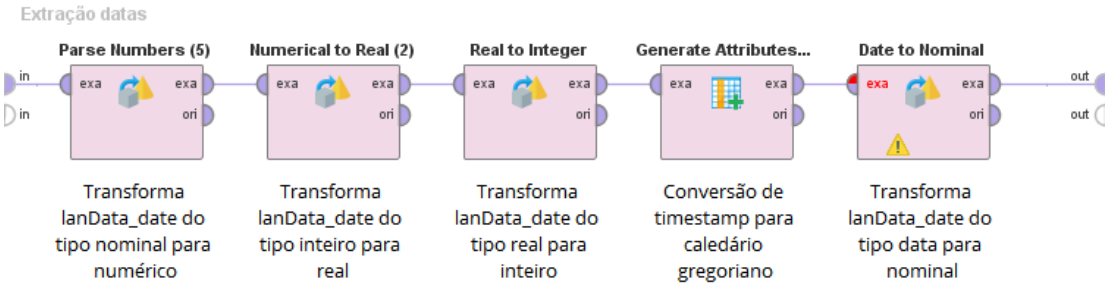


Fonte: Extraído do RapidMiner Studio.

e) Extração de datas

O tratamento de *lanData_date*, ilustrado na Figura 9, gira em torno da criação de um novo atributo (de mesmo nome) através da expressão “date_add(date_parse("01/01/1900"), lanData_date-2,DATE_UNIT_DAY)”, que resolve a conversão da parte do *timestamp* da coluna *lanData* original correspondente à contagem linear de dias a partir do dia 1 de janeiro de 1900. Tal função só aceita valores do tipo inteiro, motivo pelo qual foram realizadas previamente as transformações do tipo nominal para numérico, de numérico para real e de real para inteiro. A aparente redundância ocorre devido ao RM não dispor de operador mais direto, nominal para inteiro. Por fim, é realizada a conversão de *lanData_date* do tipo data para nominal, também com a função de preparar para a futura concatenação com a coluna *lanData_time*.

Figura 9 - Subprocesso “Extração datas”



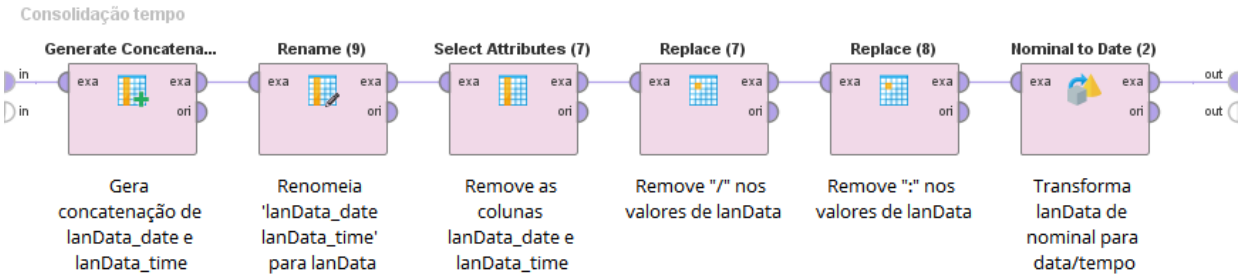
Fonte: Extraído do RapidMiner Studio.

f) Consolidação do tempo

O último subprocesso do tratamento dos dados é a concatenação das colunas *lanData_date* e *lanData_time* em uma *lanData* convertida em unidades de tempo oficiais no formato *date/time*, demonstrado na Figura 10. Após a concatenação propriamente dita, são realizados os seguintes ajustes: renomeação da coluna *lanData_date lanData_time* (nomeada automaticamente) para *lanData*; remoção das colunas *lanData_date* e *lanData_time*; remoção das barras (/) e dois pontos (:) nos valores da nova coluna *lanData* e a sua transformação definitiva do tipo nominal para tipo data/tempo. A conversão foi realizada sob o formato *MMddyyyyHHmmssSSS*, onde que *MM* é o dia do mês, *dd* é o dia do mês abreviado, *yyyy* é o ano, *HH* é a hora no formato de 24 horas, *mm* e *ss* são os minutos e segundos respectivamente e SSS, os milissegundos. Em todos os casos, o resultado refletirá o número de dígitos do

código. Essa padronização elimina unidades de tempo inferiores ao milésimo de segundo e proporciona precisão suficiente para investigações relacionadas às diferenças de tempo entre lances.

Figura 10 - Subprocesso “Consolidação tempo”



Fonte: extraído do RapidMiner Studio.

O subprocesso *lanData final* encerra a etapa de pré-processamento do KDD descrita por Leandro e De Ferrari (2017), onde foram realizadas operações de limpeza, redução, transformação e padronização dos dados. A partir desse ponto no processo serão gerados novos valores interessantes ao estudo, visando extrair informações inicialmente ocultas a partir da interação entre atributos.

7.3 MINERAÇÃO DOS DADOS

Essa etapa do processo corresponde à aplicação de algoritmos capazes de extrair conhecimentos a partir dos dados pré-processados. O RM facilita o processo ao embutir esses algoritmos nos operadores visuais, permitindo sua realização mesmo sem domínio de linguagens de programação pelo usuário. Das técnicas descritas por De Castro e Ferrari (2017), foram aplicadas principalmente as de análise descritiva (métodos de visualização), de agrupamento (segmentação de bases de dados) e de associação (determinação de atributos que correlacionados). Outras técnicas descritas pelos autores, como as de detecção de anomalias e de predição (classificação e estimação), podem constituir trabalhos futuros.

a) Diferenciações

A primeira criação de valor efetuada é a extração da diferença de tempo entre cada lance em relação ao lance imediatamente anterior visando a exploração relacionada a possíveis comportamentos anômalos de licitante em relação à rapidez na resposta e outros lances, possibilitando a análise de possíveis efeitos desses comportamentos. Para tanto, foi criada uma coluna com as diferenças ocorridas dentro de cada item, cenário em que há intervenção do governo federal para reduzir os efeitos positivos do uso de *bots* através das regras criadas pela da Instrução Normativa/SLTI nº 3, de 16 de dezembro de 2011 (BRASIL, 2011b), que dizem que o intervalo entre os lances enviados pelo mesmo licitante não poderá ser inferior a vinte (20) segundos e o intervalo entre lances quaisquer não poderá ser inferior a três (3) segundos, já abordadas no referencial teórico.

No entanto, foi criada também uma coluna com a diferença entre lances dentro de cada pregão, extrapolando assim o item para o qual o lance foi emitido. Essa intervenção busca investigar possíveis efeitos benéficos do uso de *bots* em pregões de vários itens, situações em que é comum os licitantes disputarem vários lances simultaneamente, sem regras que protejam licitantes que não utilizem *bots* na emissão dos lances.

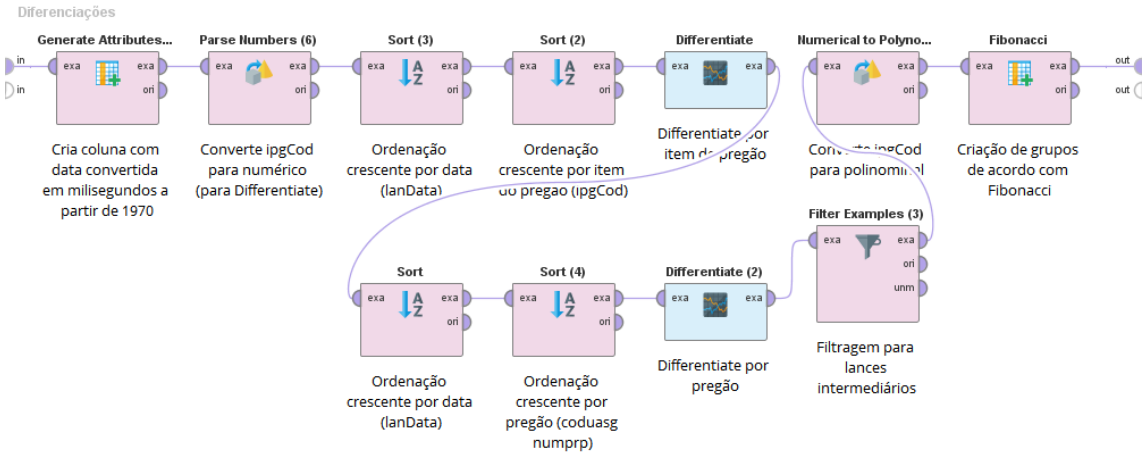
A primeira etapa para a criação dessas colunas foi a geração de novo atributo contendo os valores de *lanData* em número de milissegundos a partir de 1 de janeiro de 1970 00:00:00 GMT através da utilização da expressão "*date_millis*(*lanData*)" no operador *Generate Attribute*, demonstrada na Figura 11, tendo como resultado a coluna *milis*. Em seguida, a coluna *ipgCod*, que identifica o item do pregão para o qual foi emitido o lance, foi convertida do tipo polinomial para numérico por exigência do operador *Differentiate*, que será utilizado em seguida e funciona calculando a diferença de cada valor em relação ao da linha superior.

Para a criação das colunas de diferenciação no âmbito dos itens dos pregões, foi realizada uma dupla ordenação de valores, sendo uma por data e hora de cada lance em ordem crescente e outra pelo número do item, também em ordem crescente. Foi adicionado o "(diff)" ao nome das colunas que foram objeto do *Differentiate*. Os itens ficaram então ordenados em tempo e agrupados por item, tendo como resultado os atributos *ipgCod(diff), lanClassif(diff)* e *milis(diff)*. Para a criação da coluna de diferenciação relativa ao pregão, independentemente do item para o qual o lance foi emitido, utilizada apenas a coluna *milis*, com a adição do sufixo "(diff_preg)". Foi realizada uma ordenação de valores por data e hora de cada lance em ordem crescente seguida de uma ordenação pelo número do pregão, ficando assim ordenados

em tempo e agrupados por pregão na coluna *milis*(diff_preg). Os resultados dessas diferenciações são de grande importância para análise do uso de *bots* por fornecerem evidências para comparações entre intervalos de tempo e valores dos lances.

Permitiu também realizar a filtragem dos lances intermediários, tipo em que os *bots* atuam, eliminando os lances iniciais. Enquanto os lances intermediários são os emitidos durante a sessão pública, que consiste na fase competitiva do pregão, os lances iniciais são incluídos durante os dias anteriores, período em que o edital está recebendo as propostas dos fornecedores interessados. Todos os lances iniciais são registrados no sistema no exato momento da abertura da sessão, por isso a diferença de tempo entre eles é zero. A filtragem foi realizada através do operador *Filter Examples* usado os seguintes critérios: *milis(diff)* $\neq 0$; *milis(diff_preg)* $\neq 0$ e *ipgCod(diff)* $= 0$. Esse último critério foi utilizado para eliminar os primeiros lances de cada item de pregão. A filtragem representou uma redução de 23,47%, resultando em um total de 58.441 linhas. Em seguida, a coluna *ipgCod* foi convertida de volta ao tipo polinomial.

Figura 11 - Subprocesso "Diferenciações"



Fonte: Extraído do RapidMiner Studio.

Os resultados das diferenças de tempo entre lances foram agrupados, visando facilitar o aparecimento de padrões de comportamento entre os licitantes. Utilizou-se como critério a sequência de números Fibonacci que, segundo Ramos (2013), consiste em sequência (como 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, ...) onde cada termo, a partir do terceiro, é igual a soma dos dois termos imediatamente anteriores. Adotou-se a sequência Fibonacci por apresentar uma progressão que interessa ao estudo devido a principal vantagem do uso do *bot* decorrer da rapidez na emissão dos lances. Assim, quanto mais rápido, mais vantajoso o uso do *bot* em

relação a pessoas reais, enquanto não seria útil agrupar em pequenos incrementos lances emitidos após vários segundos de diferença.

Foi feita uma exceção à sequência Fibonacci no grupo "E" de *group(milis)*, em que o limite de 21.000 foi adaptado para 20.000 para representar o relevante comportamento decorrente da regra do intervalo mínimo de 20 segundos da IN/SLTI 3/2011. Foram criadas duas colunas através do operador *Generate Attribute*, de acordo com as expressões de função do Quadro 14. Destaca-se que, enquanto a coluna *group(milis_preg)* foi agrupada tendo como limite do primeiro grupo 1.000 milissegundos, para *group(milis)* foi utilizado o limite inicial de 3.000 milissegundos devido à regra dos 3 segundos, já comentada, que reduz a incidência de registros abaixo desse tempo. Manteve-se os agrupamentos abaixo dos 3 segundos em *group(milis_preg)* para que fosse possível investigar possível vantagem no uso de *bots* em pregões de vários itens.

Quadro 14 - Funções para criação de grupos Fibonacci

Nova coluna	Expressão de função

group(milis)	if([milis(diff)]<3000,"A",
	if([milis(diff)]<5000,"B",
	if([milis(diff)]<8000,"C",
	if([milis(diff)]<13000,"D",
	if([milis(diff)]<20000,"E",
	if([milis(diff)]<34000,"F",
	if([milis(diff)]<55000,"G",
	if([milis(diff)]<88000,"H",
	if([milis(diff)]<144000,"I","J"))))))))

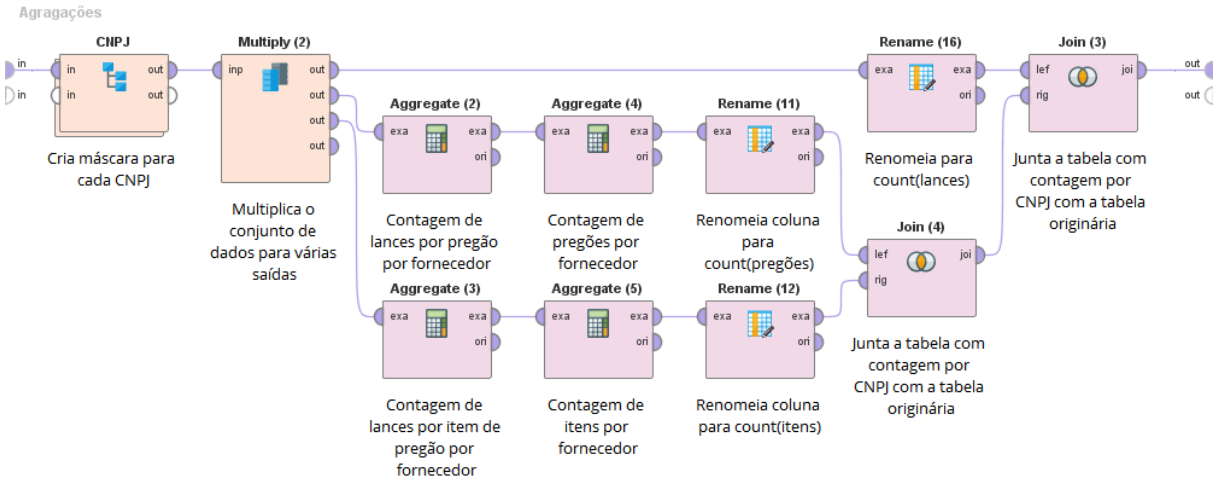
group(milis_preg)	if([milis(diff_preg)]<1000,"A",
	if([milis(diff_preg)]<2000,"B",
	if([milis(diff_preg)]<3000,"C",
	if([milis(diff_preg)]<5000,"D",
	if([milis(diff_preg)]<8000,"E",
	if([milis(diff_preg)]<13000,"F","G"))))))

Fonte: Elaborado pelo Autor, 2019.

b) Agregações

O conjunto de tratamentos seguintes foram realizadas em torno do operador *Aggregate* (Figura 12), que tem como função extrair, em novas tabelas, agrupamentos de acordo com critério escolhidos. A primeira etapa foi a criação de máscara a coluna CNPJ (que contém os números reais do Cadastro Nacional da Pessoa Jurídica de cada licitante), pois a partir desse ponto serão considerados os indícios que sugerem uso dos *bots*, evitando conflitos éticos durante a análise dos resultados. Em seguida, foram criadas algumas colunas visando a extração de valores antes ocultos, com a subsequente consolidação no conjunto de dados principal.

Figura 12 - Subprocesso "Agregações"

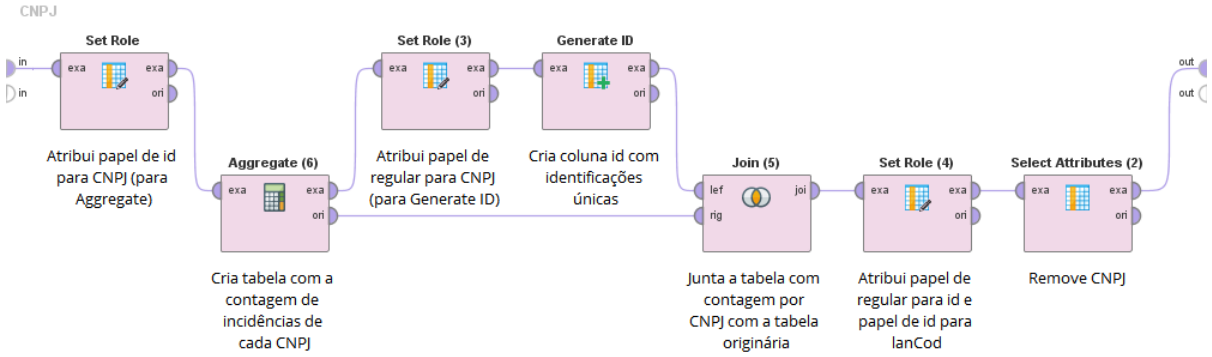


Fonte: Extraído do RapidMiner Studio.

O subprocesso referente à criação de máscara para os CNPJ, demonstrado na Figura 13, é iniciado com o operador *Set Role*. Segundo a documentação do RM, o papel de um atributo descreve como os operadores vão reconhecer e manipulá-lo. A função padrão é a regular, enquanto que as outras funções são classificadas como especiais, sendo que cada função especial só pode aparecer uma vez em conjunto de dados. Os atributos com função regular são usados como variáveis de entrada para tarefas de aprendizado de máquina. As funções especiais que serão utilizadas neste estudo são o *id* e o *label*. Um atributo com papel de *id* atua como identificador para cada linha, devendo ser exclusivo para cada incidência e é bastante utilizado em operadores de mesclagem. Um *label*, por sua vez, funciona como rótulo

e age como um destino para os operadores de aprendizado. É também chamado de "variável de destino" ou "classe".

Figura 13 - Subprocesso “CNPJ”



Fonte: Extraído do RapidMiner Studio.

No subprocesso em foco, o *Set Role* inicial foi realizado na coluna *CNPJ* para *id* por exigência do operador *Aggregate* subsequente, configurado para agregar a contagem (função *count*) das incidências de cada CNPJ agrupadas por CNPJ. O resultado é uma nova tabela composta por duas colunas, uma contendo os CNPJ em incidência única, e outra, denominada *count(CNPJ)*, contendo o número de vezes que cada CNPJ aparece na tabela principal, correspondendo ao número de lances emitidos por cada licitante. Após esse procedimento, é utilizado o operador *Set Role* para atribuir o papel à coluna *CNPJ* de volta para *regular*, por exigência do operador *Generate ID*, utilizado em seguida para gerar uma nova coluna com nomes sequenciais (a partir de id_1001) em cada linha. O procedimento principal de mesclagem da nova tabela com a principal é realizado pelo operador *Join*, utilizando como parâmetro *join type right* e como atributos chave o *CNPJ* (para as tabelas ambas as tabelas). O operador *Set Role* é utilizado novamente para atribuir o papel de *regular* à coluna *Id* e, como parâmetro de adicional (em *set additional role*), papel de *id* para a coluna *lanCod*, pois essa coluna é que representa a identificação única de cada linha. Esse procedimento é realizado nesse momento porque não será utilizado mais nenhum operador que dependa do papel *id*, sendo então uma operação definitiva. Por fim, a coluna *CNPJ* é removida da tabela consolidada.

Avançando nas etapas de agregação, conforme Figura 12, é utilizado o operador *Multiply*, que multiplica a tabela para operações simultâneas. Foram então realizados os seguintes procedimentos: 1) agregação para extração da contagem de lances por licitante em

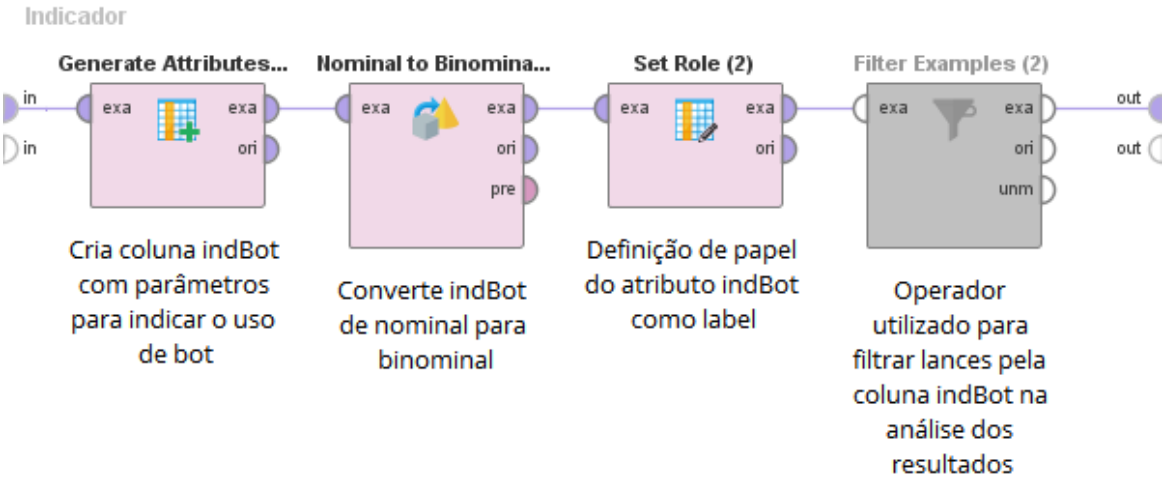
cada pregão; 2) agregação para extração da contagem de pregões que cada licitante participou; 3) renomeação da coluna resultante para *count(pregões)*. Paralelamente, foram realizados os seguintes procedimentos: 1) agregação para extração da contagem de lances por licitante em cada item de pregão; 2) agregação para extração da contagem de itens que cada licitante participou; 3) renomeação da coluna resultante para *count(itens)*.

Por fim, é realizada a mesclagem das tabelas de contagem por pregão e por item, seguida da mesclagem de seu resultado com a tabela principal, precedida da renomeação da coluna *count(CNPJ)*, utilizada anteriormente para a criação de máscara, para *count(lances)*, pois representa o número total de lances registrados para cada licitante.

c) Indicador

Foi criado um indicador de uso de *bots* com base na quantidade de lances enviadas por fornecedor. Na análise, buscou-se comportamentos anômalos, representado por alto número de lances emitidos no total ou em itens em relação ao apresentado pelos demais licitantes da amostra trabalhada. O subprocesso, demonstrado na Figura 14, consiste na criação de uma coluna denominada *indBot* através do operador Generate Attribute com as expressão de função "if([count(lances)]>350,"S","N")", com a conversão para o tipo binominal e a definição de papel para label, para permitir que seja alvo de aprendizado de máquina. A quantidade utilizada como parâmetro será explicada na análise descritiva dos dados (seção 6.3.5).

Figura 14 - Subprocesso "Indicador"



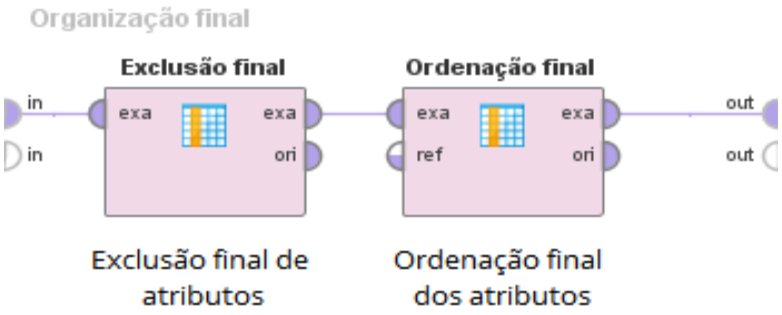
Fonte: Extraído do RapidMiner Studio.

Os números utilizados como parâmetros para os requisitos de indicação de uso de *bot* foram definidos através da análise descritiva dos dados, que será demonstrada a seguir. A função gera valor “S” (de Sim), caso atenda a qualquer um dos requisitos, e valor “N” (de Não), caso não atenda a nenhum. Por fim, foi colocado um operador de filtragem de valores (“S” ou “N”) para que seja possível a realização de análises elaboração de gráficos referentes aos lances sinalizados como suspeitos e como não suspeitos separadamente.

d) Organização final

Por fim, o último subprocesso consiste na exclusão final de atributos (caso necessário) e a definição da ordem de apresentação, conforme Figura 15.

Figura 15 - Subprocesso “Organização final”



Fonte: Extraído do RapidMiner Studio.

e) Análise descritiva dos dados

Após o pré-processamento, diferenciações, associações e criação do indicador, a tabela apresentou 58.441 linhas e 17 colunas, sendo dois atributos especiais (*lanCod* como *id* e *indBot* como *label*) e 15 regulares, como demonstrado no Quadro 15.

Quadro 15 - Análise descritiva dos atributos finais

Atributo	Descrição	Tipo	Valores (contagem, %)
lanCod	Código de identificação única do lance	Nominal	Valores únicos; menor: 419846029; maior: 455083289
indBot	Indicador de uso de *bot*	Binominal	N (41170, 70,45%); S (17271, 29,55%)
ippIndAceito	Indica se o licitante teve sua proposta aceita no item	Nominal	N (32384, 55%); A (14990, 26%); R (11067, 19%)
count(lances)	Número de lances registrados pelo	Inteiro	Mínimo: 1; máximo: 1220; média:

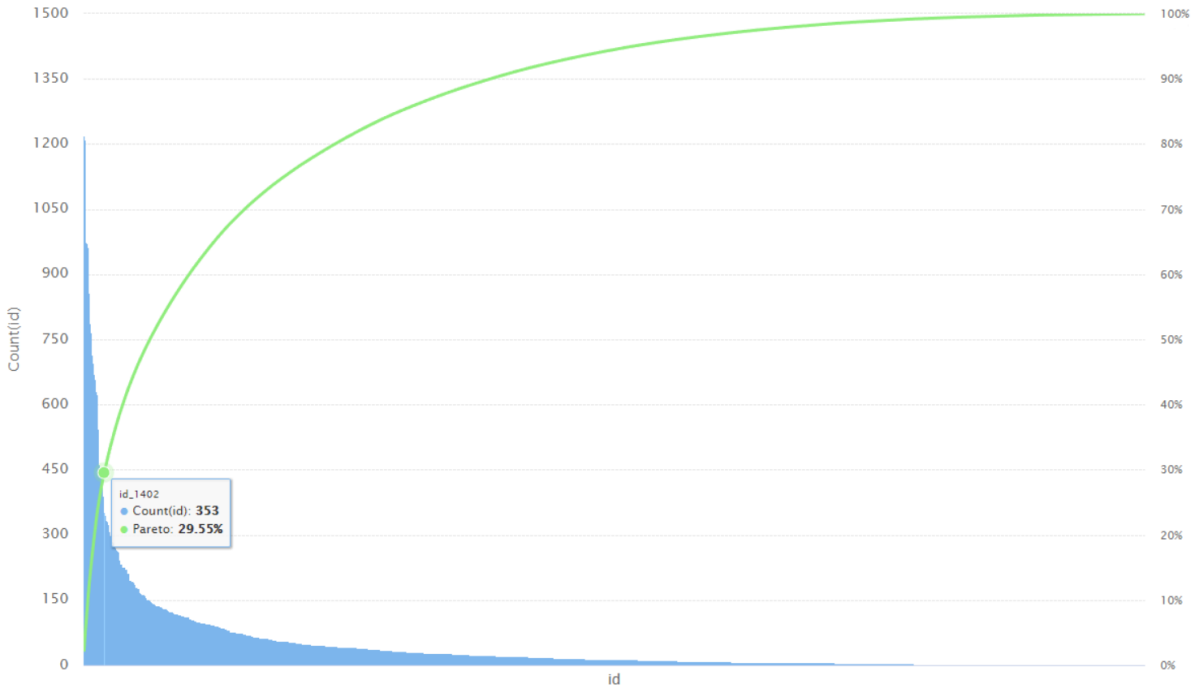
	licitante no *dataset*		310,55; desvio padrão: 342,92
count(itens)	Número de itens em que o licitante disputou	Inteiro	Mínimo: 1; máximo: 280; média: 44,11; desvio padrão: 54,60
count(pregões)	Número de pregões em que o licitante disputou	Inteiro	Mínimo: 1; máximo: 12; média: 3,01; desvio padrão: 2,73
lanClassif(diff)	Diferença de valor para o lance imediatamente anterior	Numérico	Mínimo: -90; máximo: 7500000; média: 28291.82; desvio padrão: 226351.42
milis(diff)	Diferença em milissegundos para o lance imediatamente anterior dentro de cada item	Numérico	Mínimo: 2; máximo: 4843300979; média: 570566; desvio padrão: 25933831
milis(diff_preg)	Diferença em milissegundos para o lance imediatamente anterior dentro de cada pregão	Numérico	Mínimo: 2; máximo: 483300979; média: 283415; desvio padrão: 25166899
lanClassif	Valor do lance	Numérico	Mínimo: -90; máximo: 7500000; média: 28291,82; desvio padrão: 226351,42
group(milis)	Agrupamento Fibonacci para diferença em milissegundos para o lance imediatamente anterior dentro de cada item	Nominal	G (17336, 29,66%); E (9557, 16,35%); D (7841, 13,41%); B (7013, 12,00%); F (6970, 11,93%); C (5862, 10,03%); A (3862, 6,61%)
group(milis_preg)	Agrupamento Fibonacci para diferença em milissegundos para o lance imediatamente anterior dentro de cada pregão	Nominal	A (13751, 23,53%); D (9590, 16,41%); B (8744, 14,96%); E (7369, 12,61%); G (7230, 12,37%); C (6229, 10,66%); F (5528, 9,46%)
group(valor)	Agrupamento Fibonacci para diferença de valor para o lance imediatamente anterior dentro de cada item	Nominal	A (22467, 38,44%); G (21609, 36,98%); B (420, 7,19%); F (3188, 5,46%); C (2403, 4,11%); E (2296, 3,93%); D (2275, 3,89%)
lanData	Data e hora de registro do lance no sistema SIASG	Data e hora	Mais antigo: 17/01/2017, às 9:05; mais recente: 31/10/2017, às 9:17; lapso temporal: 287 dias
id	Identificação do licitante	Nominal	Valores distintos: 1.256; primeiro: Id_1001; último: id_2256
coduasg numprp	Código da unidade administrativa que realizou o pregão, número e ano do pregão	Nominal	Valores distintos: 166; compostos por código UASG (até 6 dígitos), número do pregão (até 2 dígitos), ano do pregão (4 dígitos)
ipgCod	Código de identificação do item de pregão		Valores distintos: 2.476; mínimo: 17984568; máximo: 19314499

Fonte: Elaborado pelo Autor, 2019.

7.4 AVALIAÇÃO DO CONHECIMENTO (ANÁLISE DOS RESULTADOS)

O critério utilizado para que a coluna *indBot* indique “S” (de Sim para suspeito de uso de *bot*) é que o valor da coluna de contagem de lances totais no conjunto de dados (*count(lances)*) maior que 350. A determinação desses limites vem da análise do gráfico de Pareto elaborado no RapidMiner (Figura 16), tendo sido adotado como critério de escolha a proporção de 30% dos licitantes com os maiores números nas colunas indicadas (círculo verde no gráfico, em que o primeiro licitante que atende ao critério contém 353 lances registrados e representa 29,55% dos lances). Como resultado, foram obtidos 17.271 lances de licitantes sinalizados com “S” (suspeitos) (29,55% dos lances), emitidos por 25 dos 1.256 da amostra (1,99% dos licitantes) e 41.170 lances de licitantes sinalizados com “N” (não suspeitos) (70,45% dos lances), emitidos por 1.231 dos 1.256 (98,01% dos licitantes).

Figura 16 - Gráfico de Pareto com distribuição de lances por licitante

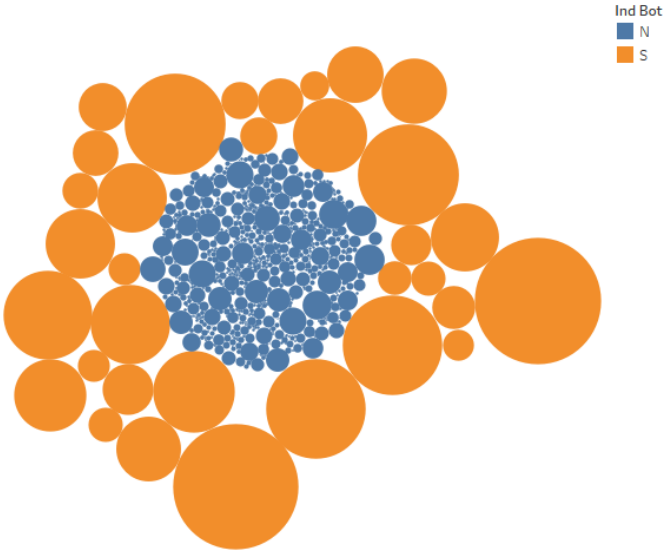


Fonte: Extraído do RapidMiner Studio.

A análise do resultado da aplicação do filtro pode ser visualizada na Figura 17, em que cada círculo é um licitante e que métrica utilizada para determinar o tamanho dos círculos é a soma de todos os valores de *count(lances)*, método utilizado para ampliar as diferenças. Os

1,99% dos licitantes que foram indicados como suspeitos de utilizarem *bots* (em laranja) ficam sobressalentes em relação aos não suspeitos, em azul.

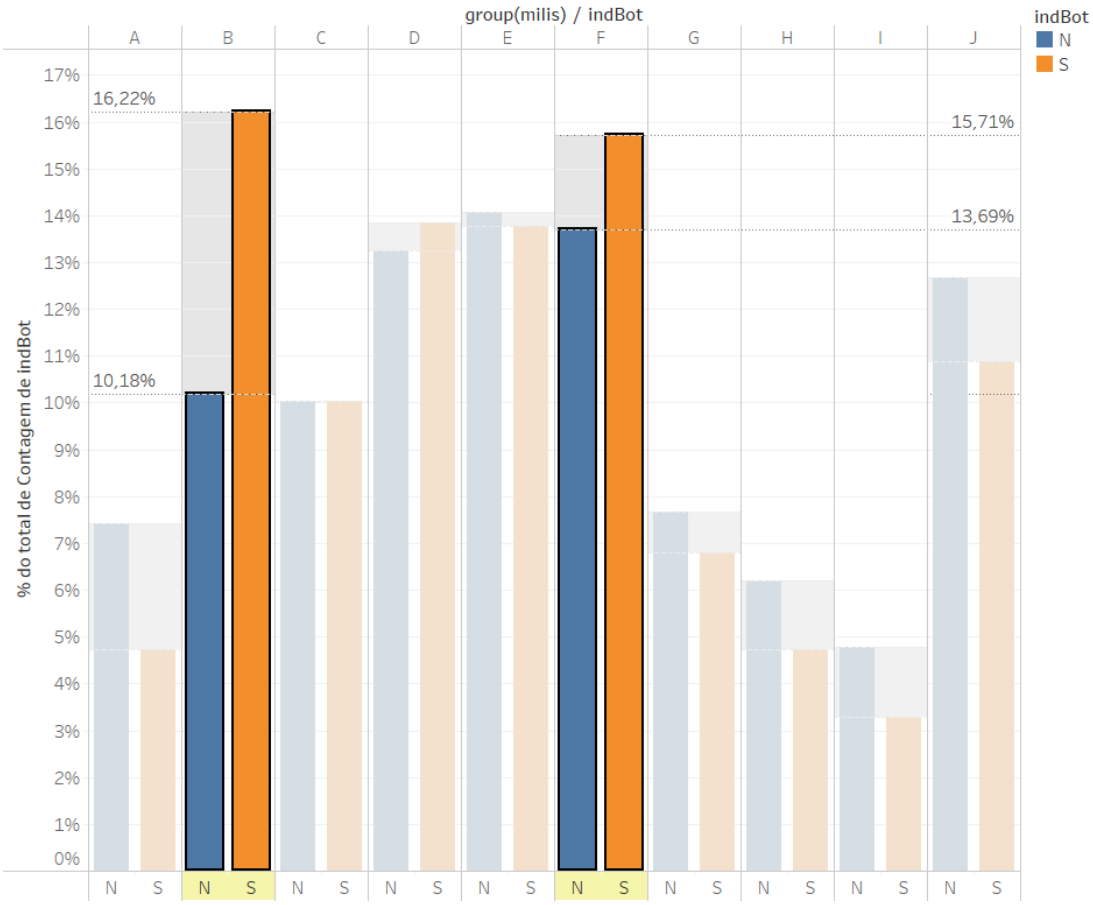
Figura 17 - Gráfico de bolhas com relação entre licitante, indicador e número de lances



Fonte: Elaborado pelo Autor, 2019.

A diferença no comportamento de reação entre lances de licitantes indicados como suspeitos e não suspeitos pode ser visualizado na Figura 18, que relaciona os dados do atributo *group(millis)* com o *IndBot*. Os painéis do gráfico indicados de A a J na Figura 18 refletem o agrupamento demonstrado no Quadro 14, simplificado no Quadro 16 abaixo. Ressaltamos que os lances constantes do grupo A da coluna *group(milis)* apresenta a peculiaridade de serem teoricamente irregulares, segundo a regra dos 3 segundos, criada pela IN/SLTI 3/2011 (BRASIL, 2011b). De acordo com a norma, os lances em desacordo devem ser identificados e anulados pelo pregoeiro, devendo a ocorrência ser comunicada imediatamente a Secretaria de Logística e Tecnologia da Informação (BRASIL, 2011b).

Figura 18 - Agrupamento de intervalo de tempo entre lances nos itens por *indBot*



Fonte: Elaborado pelo Autor, 2019.

O gráfico da Figura 18 expõe o percentual, em cada grupo, dos lances de licitantes suspeitos e não suspeitos. Os percentuais são relativos a cada situação (suspeitos e não suspeitos). A relativização permite a comparação de comportamento em situações em cenários com proporções diferentes (17.271 lances com "S" e 41.170 com "N"). Para facilitar a análise, ficaram destacados os grupos B e F, em que "S" apresentou superioridade significativa.

Quadro 16 - Critérios dos grupos Fibonacci

Coluna	Grupo	Faixa de diferença entre lances (em milissegundos)
group(milis)	A	Menor que 3.000
	B	De 3.000 a 4.999
	C	De 5.000 a 7.999
	D	De 8.000 a 12.999
	E	De 13.000 a 19.999
	F	De 20.000 a 33.999
	G	De 34.000 a 54.999

	H	De 55.000 a 87.999
	I	De 88.000 a 143.999
	J	A partir de 144.000
group(milis_preg)	A	Menor que 1.000
	B	De 1.000 a 1.999
	C	De 2.000 a 2.999
	D	De 3.000 a 4.999
	E	De 5.000 a 7.999
	F	De 8.000 a 12.999
	G	A partir de 13.000

Fonte: Elaborado pelo Autor, 2019.

O fato de os grupos A, C, D, E, G, H e J apresentarem comportamento de retração relativa aos grupos B e F foi considerado como validação de eficácia do processo e da separação por grupos na detecção de lances emitidos *bots*. No grupo B, que representa os lances enviados logo após os 3 segundos da regra de intervalo mínimo entre lances de licitantes diferentes (Quadro 16), o gráfico de "S" se apresentou com uma diferença a maior de 6,04% em relação ao de "N". No grupo F, que representa os lances enviados logo após os 20 segundos da regra de intervalo mínimo entre lances do mesmo licitante (Quadro 16), o gráfico de "S" se apresentou com uma diferença a maior de 2,02% em relação ao de "N".

No entanto, apesar de se considerar evidenciado o uso de *bot*, é possível inferir que a classificação de *indBot* é imprecisa, pois a amostra recebeu uma quantidade significativa de lances indicados como sendo de licitantes suspeitos, mas que apresentam comportamento semelhante aos classificados como emitidos por humanos. A contagem de lances em cada grupo está demonstrada na Tabela 3.

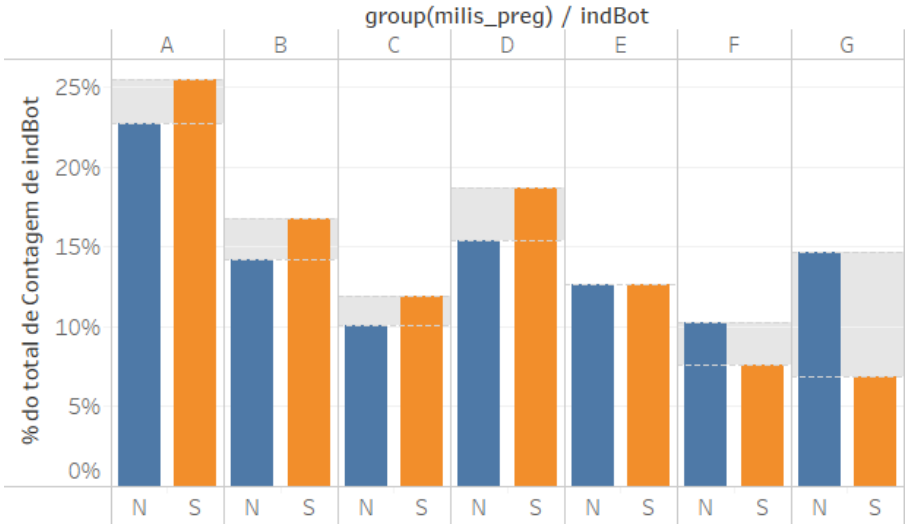
Tabela 3 - Dados do gráfico da Figura 18

group(milis)	Contagem de *indBot*		Percentil de Contagem de *indBot*	
	N	**S**	**N**	**S**
A	3.055	817	7,42%	4,73%
B	4.190	2.802	10,18%	16,22%
C	4.132	1.734	10,04%	10,04%
D	5.459	2.390	13,26%	13,84%
E	5.797	2.379	14,08%	13,77%
F	5.637	2.713	13,69%	15,71%
G	3.159	1.174	7,67%	6,80%
H	2.553	816	6,20%	4,72%
I	1.964	566	4,77%	3,28%
J	5.224	1.880	12,69%	10,89%
Total	**41.170**	**17.271**	**100,00%**	**100,00%**

Fonte: Elaborado pelo Autor, 2019.

Quanto ao comportamento no âmbito nos pregões (coluna *group(milis_preg)*), os gráficos da Figura 19 evidenciaram que os grupos indicados com "S" (em laranja) se sobressaem em lances mais rápidos, com intervalos de tempo menores entre lances (grupos A, B, C e D) (Quadro 17). Os percentuais do Quadro 17 também são relativos a cada situação (suspeitos e não suspeitos). Concluiu-se que o agrupamento por diferença entre lances fora do âmbito de cada item (*milis_preg*) evidenciou a expectativa inicial de que a rapidez na reação para emissão do lance é coerente com o alto número de lances (critério utilizado para a classificação em *indBot*).

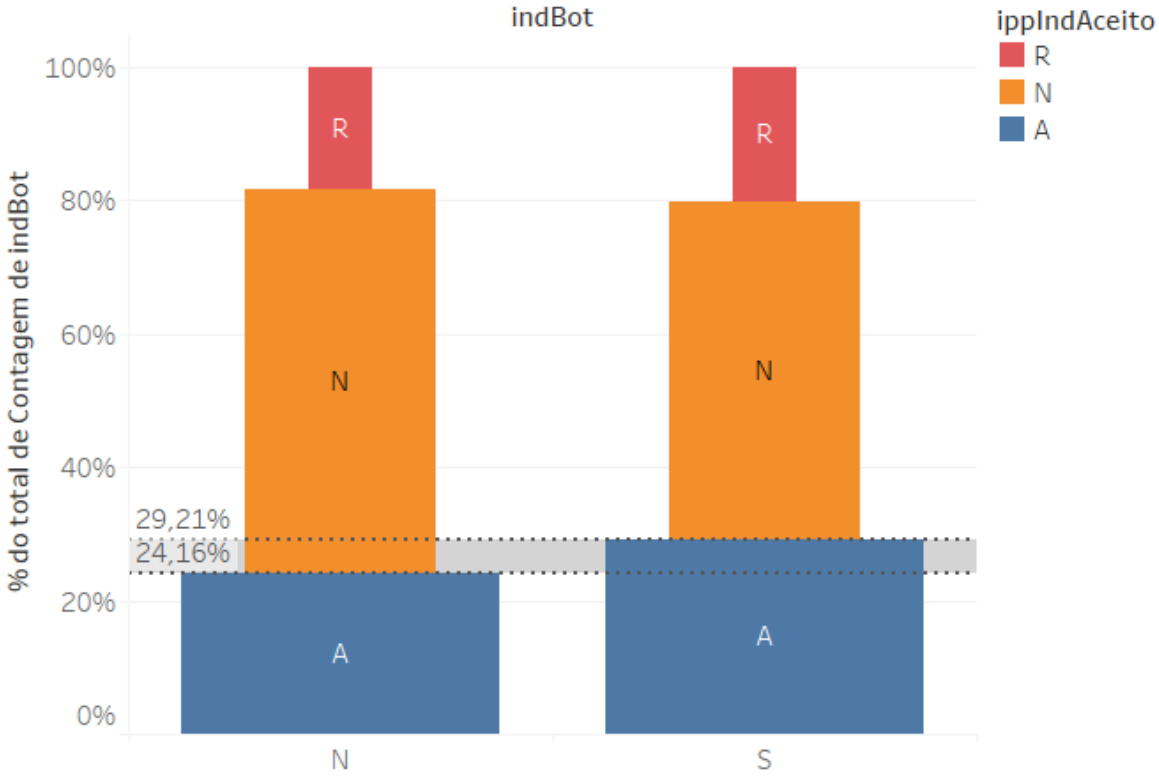
Figura 19 - Gráfico de barras do agrupamento em pregões com todos os licitantes



Fonte: Elaborado pelo Autor, 2019.

Por fim, cabe avaliar o sucesso obtido pelos licitantes suspeitos de usar *bot*. A Figura 20 apresenta a distribuição relativa aos não suspeitos (à esquerda) e aos suspeitos (à direita). Enquanto o grupo dos indicados como não suspeitos teve um aproveitamento de 24,15% na aceitação de suas propostas, o do grupo dos indicados como suspeitos foi de 29,21%, o que representa um ganho de 5,06%.

Figura 20 - Gráfico de barras empilhadas da relação entre *indBot* e aceitação da proposta



Fonte: Elaborado pelo Autor, 2019.

Apesar de evidenciar um ganho, é notável que essa vantagem não é proporcional à quantidade de lances emitida, ressaltada na Figura 16. No entanto, inserida no contexto das licitações, em que há um cenário de alta competitividade e que movimentam bilhões de reais por ano, foi considerado que persistia, em 2017, vantagem competitiva no uso de *bots* para emissão de lances e que os comportamentos evidenciados neste estudo indicam que seu uso ainda é restrito (1,99% da amostra analisada).

8 CONSIDERAÇÕES FINAIS

A avaliação dos resultados da mineração permite identificar conhecimentos úteis e não triviais. As análises geradas por este estudo são decorrentes principalmente do uso de um indicador hipotético. No entanto, quando analisados em conjunto com outros comportamentos observados, podemos afirmar que foi possível identificar e observar efeitos do uso de *bots* em pregões eletrônicos através da mineração de dados abertos, objetivo geral do estudo.

Foi evidenciada uma vantagem de mais de 5% em sucesso nos itens de disputa observados para apenas 1,99% dos licitantes da amostra, indicados como suspeitos do uso de *bots*, segundo critério de alta quantidade de lances emitidos em relação ao comportamento observado na amostra. Estes indicadores podem ser complementados com uma análise qualitativa/quantitativa sobre o que estes 1,99% representam em termos de valores ou volumes de compras, pois tais avaliações não são possíveis de serem realizadas com o *dataset* utilizado neste trabalho.

Em conjunto com a análise de outros comportamentos, descritos abaixo, os resultados apontam que persiste uma vantagem do uso de *bots* em pregões eletrônicos na amostra analisada (2017). Foram ainda observados indícios de falhas no registro de lances no Comprasnet.

Foi possível abordar todas as hipóteses propostas para o trabalho. A **alta frequência de lances emitidos por item que participa** em relação aos demais licitantes foi identificada como sendo o comportamento anômalo predominante para a indicação de uso de *bots*. Na amostra analisada, 1,99% do total de licitantes (25 dos 1.256) foram responsáveis por 29,55% do total de lances registrados, conforme ilustrado na Figura 16. A **alta quantidade de participação em pregões eletrônicos** se mostrou relevante, mas secundária, tendo em vista que o número de pregões não foi significante para as análises (máximo de 12 para um licitante) e que a hipótese predominante (alta frequência de lances emitidos por item que participa) o absorve, pois os fornecedores com maiores quantidades de lances emitidos coincidiram em ambas as situações. É possível que a utilização de amostras com maior volume de dados permita validar tal hipótese.

Alta frequência de participação simultânea em vários itens de um mesmo pregão se mostrou irrelevante na amostra analisada, possivelmente devido ao corte vertical da amostra, o que reduz a ocorrência de pregões simultâneos de objetos similares. Uma

suposição é a de que esse comportamento possa ser evidenciado em amostras com cortes horizontais (pregões de vários órgãos relacionados a um tipo de material específico em determinado período, por exemplo), o que evidenciaria a incidência de pregões simultâneos do mesmo ramo, em que fornecedores participem em mais de um pregão ao mesmo tempo. Nesse cenário, em que os licitantes necessitam dividir seus recursos escassos, o uso de *bots* pode ter sua vantagem competitiva ampliada.

A **habitualidade no registro de lances com intervalos similares de tempo e/ou valor em relação a novos menores lances de cada item** foi a hipótese trabalhada através da criação dos grupos Fibonacci. A criação de grupos decorreu da impossibilidade de se trabalhar com valores precisos de diferença entre os lances. No caso de intervalos de tempo, a impossibilidade se deu por motivo técnico: o atraso (*lag*) que ocorre entre o envio do lance pelo licitante, a transmissão dos dados através da infraestrutura de internet e o registro do lance nos servidores do Comprasnet. Assim, mesmo que um *bot* esteja configurado para cobrir lances em um mesmo intervalo de tempo, o registro do lance no sistema não reflete com exatidão o tempo configurado. Ainda, *bots* podem ser configurados para emitir lances aleatórios dentro de uma faixa de tempo pré-estabelecida, visando simular o comportamento humano.

Apesar de não ter sido possível trabalhar com diferenças exatas, como originalmente previsto, o agrupamento em intervalos de tempo se mostrou mais satisfatório para o trabalho, pois possibilitou identificar aglomerações de lances registrados imediatamente após os intervalos estipulados por normativos (de 3 e 20 segundo). Tais comportamentos apresentaram alta correlação com o comportamento anômalo predominante observados (alta frequência de lances emitidos por item que participa).

Por fim, o **alto índice de sucesso nas disputas em relação ao número de itens que participa** não se mostrou um comportamento indicador do uso de *bots*, possivelmente pelo caráter altamente competitivo dos pregões. Hipoteticamente, um *bot* emitirá muitos lances apenas em itens com alta competitividade pois necessita desse gatilho para reagir, casos em que outros fornecedores ou também estão utilizando *bots* ou estão conseguindo competir com o *bot* utilizando pessoas reais. Esse fenômeno pode explicar por que o alto volume de lances não se mostrou proporcional a um alto índice de vitórias. Outra explicação possível é que as regras de 3 e 20 segundos podem estar cumprindo sua função de proteção ao desequilíbrio causado pelo uso de *bots*, mesmo que não integralmente. Pesquisas futuras podem buscar

confirmar essa hipótese, cujos resultados podem contribuir com estudos de viabilidade relacionados à implantação da parametrização do envio de lances no Comprasnet, disponibilizado para todos os licitantes, como forma de trazer isonomia à utilização do *bot*.

Sobre a avaliação das políticas de dados abertos observadas durante a fase de obtenção dos dados, foi percebida prematuridade no modelo adotado, derivado de uma extração de dados de baixa qualidade, com falhas que resultaram em uma redução significativa durante a limpeza realizada na fase de pré-processamento (de 6.117.559 linhas para 76.506 após a eliminação de lances duplicados). A ausência do fornecimento de um dicionário de dados pode indicar uma característica de que serviço não foi pensado para atender a um usuário comum da sociedade civil, pois foi priorizado o cumprimento das necessidades mínimas de requisitos computacionais ao invés da busca ao atendimento pleno da finalidade informada na solicitação e com linguagem facilitada. Na realização de novos estudos relacionados que demandem dados abertos, é recomendado que seja feita uma seleção de atributos mínimos necessários para a realização da investigação através de consulta prévia a um dicionário de dados, incluindo-os na solicitação dos dados. A ampliação do período a qual os dados se referem e a ampliação ou afunilamento do campo (que neste foi limitado ao MAPA) podem também contribuir para a qualidade dos resultados.

Este estudo se deparou ainda com outras limitações. Não foi possível trabalhar com valores dos lances por envolver uma maior complexidade em relação a se trabalhar apenas com tempo. Enquanto o tempo é uma medida universal, ou seja, tem o mesmo peso em todas as situações analisadas, as dimensões de valor são relativas a cada tipo de item. Exemplificando, cinquenta centavos podem representar uma grande diferença em um item de baixo valor (como caneta esferográfica), mas é irrelevante em contratações de maior porte (serviços terceirizados de limpeza e conservação, por exemplo). Esse comportamento dificultou a análise dos padrões em intervalos de valor. Outro fator limitante relacionado a valores foi a ausência de atributo no *dataset* que representasse a quantidade licitada de cada item, o que inviabilizou sua multiplicação pelo valor dos lances para a obtenção dos valores totais e, consequentemente, a análise do potencial impacto financeiro da utilização dos *bots*.

Ainda, o atributo utilizado para indicar a suspeita de uso de *bot* classificou o licitante e consequentemente todas as suas participações, partindo do pressuposto de que o *software* adquirido ou desenvolvido se torna um ativo do licitante, o que pode gerar distorções. Sugerimos, em trabalhos futuros, verificar se é viável a adoção da quantidade relativa de

lances dentro de cada sessão ou de uma combinação, com a utilização de pesos diferentes para o histórico do licitante e para seu comportamento em cada sessão, como base para uma classificação de suspeição.

Um ponto relevante, que já vem sendo debatido pelos envolvidos nos projetos de desenvolvimento de sistemas de compras da APF (ENAP, 2019), surgiu durante a realização deste trabalho. Tanto órgãos de controle como responsáveis pelo desenvolvimento desses sistemas vêm criando regras e mecanismos computacionais de forma continuada para tentar inibir o uso dos *bots*. Contudo, diante do que foi verificado nesta pesquisa, chegou-se à seguinte reflexão: por que não institucionalizar o mecanismo de parametrização dos lances ao invés de buscar inibir o uso de algo que pode, de certa forma, beneficiar os licitantes através de uma operacionalização mais fácil (automatizada) e reduzir o preço final dos itens que estão sendo licitados? Se, no sentido contrário da isonomia, a utilização de um *bot* pode gerar uma vantagem indevida a um determinado fornecedor por este ter acesso a esse tipo de tecnologia que outros não têm e considerando que as restrições a seu uso, identificadas na revisão das normas e regulamentações que versam sobre o tema, não foram efetivas para estabelecer a igualdade nas condições de competição, como pode ser verificado nos resultados apurados, então a questão acima parece ser de alta relevância mesmo no contexto de novos formatos de pregão eletrônico que mantenham disputa aberta e encerramento aleatório.

Uma amostra do *dataset*[3] utilizado e o processo[4] desenvolvido estarão disponíveis para *download*. A amostra contém 10.000 linhas, que é o limite máximo atual da versão gratuita do RapidMiner, e foi extraída apenas dos lances válidos através de sua função *Turbo Prep*. Para sua utilização, é necessário importar o arquivo com o nome de "DadosLanceMapa" e os atributos sob o tipo "*polynominal"*. O *dataset* completo pode ser solicitado aos pesquisadores.

[3] Disponível em: https://doi.org/10.6084/m9.figshare.11322986.
[4] Disponível em: https://doi.org/10.6084/m9.figshare.11323028.

REFERÊNCIAS

AMORIM, V. A. J. **Licitações e contratos administrativos**: teoria e jurisprudência. Brasília: Senado Federal, Coordenação de Edições Técnicas, 2017.

AZEVEDO, Ana; SANTOS, Manuel Filipe. KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW. **IADIS European Conference Data Mining**, 2008.

BHARGAVA, B.; JENAMANI, M.; ZHONG, Y. Counteracting shill bidding in online English auction. CiteSeerX. **International Journal of Cooperative Information Systems**, v. 14, n. 2-3, p. 245-263. 2005.

BRASIL. Constituição da República Federativa do Brasil de 1988. **Constituição Federal**. Brasília. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Acesso em: 14 jan. 2018.

BRASIL. Decreto nº 1.094, de 23 de março de 1994. **Dispões sobre o Sistema de Serviços Gerais**. Brasília. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/Antigos/D1094.htm. Acesso em: 9 nov. 2019.

BRASIL. Decreto nº 5.450, de 31 de maio de 2005. **Regulamento do Pregão Eletrônico**. Brasília. 2005a. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/decreto/d5450.htm. Acesso em: 15 jan. 2018.

BRASIL. Decreto nº 8.638, de 15 de janeiro de 2016. **Institui a Política de Governança Digital**. Brasília. 2016b. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Decreto/D8638.htm. Acesso em: 9 nov. 2019.

BRASIL. Decreto nº 10.024, de 20 de setembro de 2019. **Institui a Política de Governança Digital**. Brasília, 2019a. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D10024.htm. Acesso em: 15 nov. 2019.

BRASIL. Lei Complementar nº 123/2006. **Estatuto Nacional da Microempresa e da Empresa de Pequeno Porte**. Brasília. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp123.htm. Acesso em 16 nov. 2019.

BRASIL. Lei nº 8.666, de 21 de junho de 1993. **Lei de Licitações**. Brasília. Disponível em: http://www.planalto.gov.br/ccivil_03/Leis/L8666compilado.htm. Acesso em: 14 jan. 2018.

BRASIL. Lei nº 10.520, de 17 de julho de 2002. **Lei do Pregão**. Brasília. Disponível em: http://www.planalto.gov.br/CCivil_03/leis/2002/L10520.htm. Acesso em: 14 jan. 2018.

BRASIL. Lei nº 12.462, de 4 de agosto de 2011. **Lei do RDC**. Brasília, 2011a. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12462.htm. Acesso em: 8 nov. 2019.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Pregão Eletrônico: manual do pregoeiro**. Brasília: MP, 2005b.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. **Boletim de Compras Públicas.** 2018a. Disponível em: https://www.comprasgovernamentais.gov.br/images/conteudo/ArquivosCGIAL/Boletim_de_Compras_Publicas.pdf. Acesso em: 9 ago. 2019.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. **Mais de R$ 11 milhões serão investidos na modernização do sistema de compras públicas**. 2018b. Disponível em: https://www.comprasgovernamentais.gov.br/index.php/noticias/1033-mais-de-r-11-milhoes-serao-investidos-na-modernizacao-do-sistema-de-compras-publicas. Acesso em: 18 nov. 2018.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. **Melhoria do sistema amplia segurança em pregões eletrônicos**. Brasília, 23 dez. 2010. Disponível em: https://www.governoeletronico.gov.br/noticias/melhoria-do-sistema-amplia-seguranca-em-pregoes-eletronicos. Acesso em: 15 jan. 2018.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **Cartilha Técnica para Publicação de Dados Abertos no Brasil**. v. 1.0. Brasília, 2012. Disponível em: http://dados.gov.br/pagina/cartilha-publicacao-dados-abertos. Acesso em: 10 nov. 2019.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **Estratégia de Governança Digital da Administração Pública Federal 2016-19**. Brasília: MP, 2016a. Disponível em: https://www.governodigital.gov.br/documentos-e-arquivos/Estrategia-de-Governanca-Digital.pdf. Acesso em: 9 nov. 2019.

BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. **Referencial de Governança e Gestão do Sistema de Serviços Gerais – SISG** / Secretaria de Gestão. Brasília, 2017a.

BRASIL. Polícia Federal do Brasil. **PF desarticula organização criminosa que fraudava licitações no MS**. 2017b. Disponível em: http://www.pf.gov.br/agencia/noticias/2017/03/pf-desarticula-organizacao-criminosa-que-fraudava-licitacoes-no-ms. Acesso em 29 out. 2018.

BRASIL. Portal de Compras. **Decreto aprimora regras do pregão eletrônico**. 2019f. Disponível em: https://www.comprasgovernamentais.gov.br/index.php/noticias/1170-decreto-aprimora-regras-do-pregao-eletronico. Acesso em: 15 nov. 2019.

BRASIL. Portal de Compras. **Fluxo didático**. In: Aplicação dos modos de disputa aberto e aberto/fechado - Decreto nº 10.024, de 2019. 2019e. Disponível em: https://www.comprasgovernamentais.gov.br/index.php/noticias/1208-aplicacao-modos-disputa-pregao-eletronico. Acesso em: 15 nov. 2019.

BRASIL. Portal de Compras. **Perguntas Frequentes - Pregão Eletrônico**. 2019b. Disponível em: https://www.comprasgovernamentais.gov.br/index.php/pregaoeletronico-faq. Acesso em: 9 ago. 2019.

BRASIL. Portal do Servidor. **Pregoeiro virtual ajudará no combate a fraudes em licitações**. 2019c. Disponível em: https://www.servidor.gov.br/noticias/2019/pregoeiro-virtual-ajudara-no-combate-a-fraudes-em-licitacoes. Acesso em 1 ago. 2019.

BRASIL. Presidência da República. Gabinete de Segurança Institucional. Portaria nº 93, de 26 de setembro de 2019. **Glossário de Segurança da Informação**. Brasília, 2019d. Disponível em: http://www.in.gov.br/web/dou/-/portaria-n-93-de-26-de-setembro-de-2019-219115663. Acesso em: 15 nov. 2019.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **Instrução Normativa n° 3, de 16 de dezembro de 2011**. Brasília, 2011b. Disponível em: https://www.comprasgovernamentais.gov.br/index.php/legislacao/instrucoes-normativas/403-instrucao-normativa-n-3-de-16-de-dezembro-de-2011. Acesso em: 18 nov. 2019. Alterada pela pela Instrução Normativa nº 3, de 4 de outubro de 2013.

CGU. Ministério da Transparência e Controladoria-Geral da União. **4º Plano de Ação Nacional em Governo Aberto**. Brasília, 2018. Disponível em: http://governoaberto.cgu.gov.br/no-brasil/planos-de-acao-1/copy_of_3o-plano-de-acao-brasileiro/4o-plano-de-acao-nacional_portugues.pdf. Acesso em: 3 dez. 2018.

DAVENPORT, T.; PATIL, D. J. Data Scientist: The Sexiest Job of the 21st Century. **Harvard business review**. 90. 70-6, 128. 2012.

DE CASTRO, L.N.; FERRARI, D.G. **Introdução a Mineração de Dados**. São Paulo: Saraiva. Edição do Kindle, 2017.

DHAR, V. Data Science and Prediction. **Communications of the ACM**. v. 56. n. 12, dez. 2013.

DINIZ, E.H; BARBOSA, A.F.; JUNQUEIRA, A.R.B.; PRADO, O. O governo eletrônico no Brasil: perspectiva histórica a partir de um modelo estruturado de análise. **RAP**, Rio de Janeiro v. 43, n. 1, p. 23-48, jan./fev. 2009.

DISTIL NETWORKS, Inc. **2018 Bad Bot Report Executive Summary**: The Year Bad Bots Went Mainstream. [S.l.], 2018. Disponível em: https://resources.distilnetworks.com/whitepapers/2018-bad-bot-report. Acesso em: 4 nov. 2018.

DISTIL NETWORKS, Inc. **2019 Bad Bot Report**: The Bots Arms Race Continues. [S.l.], 2019. Disponível em: https://resources.distilnetworks.com/white-paper-reports/bad-bot-report-2019. Acesso em: 10 nov. 2019.

DONG, F.; SHATZ, S. M.; XU, H. Combating Online In-Auction Fraud: Clues, Techniques and Challenges. **Computer Science Review**, v. 3, n. 4, p. 245-258. University of Illinois at Chicago, Chicago, U.S. National Science Foundation. 2009.

EBAY. eBay Inc. **Reports Third Quarter 2018 Results**. Disponível em: https://www.ebayinc.com/stories/news/ebay-q3-2018-results/. Acesso em: 10 nov. 2018.

EDELSTEIN, Herbert A**. Introduction to Data Mining and Knowledge Discovery**. 3. ed. Potomac: Two Crows Corporation, 1999.

ENAP. **Workshop sobre o novo decreto do Pregão Eletrônico - ME**. (3h06m51s). Canal da Enap no YouTube. 2019. Disponível em: https://www.youtube.com/watch?v=iMIBkeBdHpY. Acesso em: 15 nov. 2019.

ETHEM, A. **Machine Learning** (MIT Press Essential Knowledge series). The MIT Press. 2016. Edição do Kindle. Paginação irregular.

ETZIONI, Oren; WELD, Daniel. A Softbot-Based Interface to the Internet. **Communications of the ACM**, v. 37, n. 7, p. 72-76, jul. 1994.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. Data mining and knowledge discovery: making sense out of data. **IEEE Expert**, v. 11, n. 5, p. 20-25, 1996a.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. The KDD Process for Extracting Useful Knowledge from Volumes of Data. **Communications of the ACM**, v. 39, n. 11, p. 27-34, nov. 1996b.

FORD, B. J.; XU, H.; VALOVA, I. Identifying Suspicious Bidders Utilizing Hierarchical Clustering and Decision Trees. **International Conf. Artificial Intelligence**, ICAI'10. 2010.

FRANKLIN, S.; GRAESSER, A. **Is It an Agent, or Just a Program?**: A Taxonomy for Autonomous Agents. Institute for Intelligent Systems, University of Memphis, Memphis. 1997.

GOOGLE TRENDS. In: WIKIPEDIA: the free encyclopedia. [San Francisco, CA: Wikimedia Foundation, 2019]. Disponível em: https://pt.wikipedia.org/wiki/Google_Trends. Acesso em: 4 nov. 2019.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining**: Concepts and Techniques. 3. ed. Waltham: Elsevier, 2012.

HAND, David; MANNILA, Heikki; SMYTH, Padhraic. **Principles of Data Mining**. Cambridge, The MIT Press, 2001.

HUBBARD, T.P. **Auctions** (The MIT Press Essential Knowledge series). The MIT Press. Edição do Kindle, 2016.

INCAPSULA. **Bot Traffic Report 2016**. 2016. Disponível em: https://www.incapsula.com/blog/bot-traffic-report-2016.html. Acesso em: 4 nov. 2018.

ISTOÉ. **Golpe no Pregão Eletrônico**. Disponível em: https://istoe.com.br/139247_GOLPE+NO+PREGAO+ELETRONICO. Acesso em: 29 out. 2018.

JUSTEN FILHO, M. **Comentários à lei de licitações e contratos administrativos** / Marçal Justen Filho. 15. ed. São Paulo: Dialética, 2012.

KRISHNA, V. **Auction Theory**. Elsevier Science. Edição do Kindle. 2009.

KUM, Hye-Chung (Monica); PAULSEN, Susan; WANG, Wei. Comparative Study of Sequential Pattern Mining Models. *In:* LIN, Tsau Young *et al*. (ed.). **Foundations of Data Mining and Knowledge Discovery**. Berlim: Springer, 2005. (Studies in Computational Intelligence, v. 6).

LIEBERMAN, Henry. Autonomous interface agents. **ACM**, New York, 1997, p. 67-74. DOI: http://dx.doi.org/10.1145/258549.258592.

LIN, Tsau Young. Mathematical Foundation of Association Rules – Mining Associations by Solving Integral Linear Inequalities. *In:* LIN, Tsau Young *et al*. (ed.). **Foundations of Data Mining and Knowledge Discovery**. Berlim: Springer, 2005. (Studies in Computational Intelligence, v. 6).

MATIAS-PEREIRA, José. **Manual de metodologia da pesquisa científica**. 4. ed. São Paulo: Atlas, 2016. Versão digital.

MCKINSEY GLOBAL INSTITUTE. Big data: The next frontier for innovation, competition, and productivity. **McKinsey Global Institute June 2011 Report**. [S.l.], 2011.

MEDEIROS, P.H.R. **Governo Eletrônico no Brasil**: aspectos institucionais e reflexos na governança. Orientador: Tomás de Aquino Guimarães. 2004. Dissertação (Mestrado em Administração) – Faculdade DE Economia, Administração, Contabilidade e Ciência da Informação e Documentação. Universidade de Brasília, 2004.

MELLO, C. A. B. **Curso de direito administrativo**. 32. ed. São Paulo: Malheiros, 2015.

MONAHAN, D. Bot Defense: Insights Into Basic And Advanced Techniques For Thwarting Automated Threats. **Enterprise Management Associates**. 2016. Disponível em: https://www.enterprisemanagement.com/research/asset.php/3317/Bot-Defense:-Insights-Into-Basic-and-Advanced-Techniques-for-Thwarting-Automated--Threats. Acesso em: 4 nov. 2018.

NOBEL FOUNDATION. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1996. **NobelPrize.org**. [S.l.]: Nobel Media AB, 2018. Disponível em: https://www.nobelprize.org/prizes/economic-sciences/1996/press-release/. Acesso em: 10 nov. 2018.

OCDE. **OECD Digital Government Toolkit**: 12 principles. 2018. Disponível em: http://www.oecd.org/governance/digital-government/toolkit/12principles/. Acesso em: 25 nov. 2018.

OCDE. **The E-government Imperative**. Paris: OCDE, 2003.

OLIVEIRA, Alexandre Borges De; FABREGAS MASLLOVET, Abdoulaye; FAZEKAS, Mihaly.Auction Length and Prices: Evidence from Random Auction Closing in Brazil (English). **Policy Research Working Paper**. n. WPS8828. Washington, D.C.: World Bank Group, 2019. Disponível em: http://documents.worldbank.org/curated/en/663441555954348673/Auction-Length-and-Prices-Evidence-from-Random-Auction-Closing-in-Brazil. Acesso em: 13 nov. 2019.

OPEN KNOWLEDGE INTERNACIONAL. **What is Open?**, Disponível em: https://okfn.org/about/. Aceso em: 29 mar. 2019.

PAINEL DE COMPRAS. Ministério do Planejamento, Orçamento e Gestão. **Painel de Compras do Governo Federal.** Disponível em: https://paineldecompras.planejamento.gov.br Acesso em: 14 nov. 2019.

POOLE, David L.; MACKWORTH, Alan K. **Artificial Intelligence**: Foundations of Computational Agents. New York: Cambridge University Press, 2010. Versão digital.

PORTAL DA TRANSPARÊNCIA. 2018. Disponível em: www.transparência.gov.br. Acesso em: 14 jan. 2018.

RAMOS, Marcos Gertrudes Oliveira. **A Sequência de Fibonacci e o Número de Ouro**. Orientado: Romenique da Rocha Silva. 2013. Dissertação (Mestrado Profissional em Matemática) - Universidade Estadual de Santa Cruz, Ilhéus, 2013. Disponível em: http://www.biblioteca.uesc.br/biblioteca/bdtd/201160277d.pdf. Acesso em: 26 jul. 2019.

REBOUÇAS, R. R. **Detecção de figurantes em pregões eletrônicos do governo federal brasileiro**. Orientador: Hércules Antônio do Prado. 2011. Dissertação (Mestrado em Gestão do Conhecimento e da Tecnologia da Informação) - Universidade Católica de Brasília, Brasília, 2011.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial Intelligence**: A Modern Approach. 3. ed. England: Pearson Education Limited, 2016.

SANTOS, R., **Conceitos de mineração de dados na web**. *In:* XV SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB, VI SIMPÓSIO BRASILEIRO DE SISTEMAS COLABORATIVOS, 15., 2008, [s.l] **Anais [...].** 2008. p. 81-124.

SERPRO. **Metadados – Comprasnet**. 2019. Disponível em: http://catalogo.quartzo.serpro.gov.br/dataset/portal-de-compras-do-governo-federal-comprasnet/resource/86272f80-697b-451d-8238-ae9a63de5614. Acesso em: 3 jul. 2019.

SHAH, H. S.; JOSHI, N. R.; WURMAN, P. R. **Mining for Bidding Strategies on eBay**. Lecture Notes on Artificial Intelligence. 2002.

SHEARER, C. The CRISP-DM Model: The New Blueprint for Data Mining. **Journal of Data Warehousing**, v. 5, n. 4, 2000, p. 13-22.

SILVA, Leandro Augusto; PERES, Sarajane Marques; BOSCARIOLI, Clodis. **Introdução à mineração de dados**: Com Aplicações em R. Editora Elsevier. Edição do Kindle. 2017. Paginação irregular.

SYMANTEC CORPORATION. **O que são bots?** Disponível em: https://br.norton.com/internetsecurity-malware-what-are-bots.html. Acesso em: 16 fev. 2019.

TCU. **Acórdão nº 1.216/2014**. Plenário. Relator: Ana Arraes. Sessão de 14/5/2014. 2014b. Disponível em: https://contas.tcu.gov.br/sagas/SvlVisualizarRelVotoAcRtf?codFiltro=SAGAS-SESSAO-ENCERRADA&seOcultaPagina=S&item0=496742. Acesso em: 27 out. 2018.

TCU. **Acórdão nº 2.569/2014**. Plenário. Relator: Ministro Benjamin Zymler. Sessão de 1/10/2014. 2014a.

TCU. **Acórdão nº 1.647/2010**. Plenário. Relator: Ministro Valmir Campelo. Sessão de 14/7/2010. Disponível em: http://www.tcu.gov.br/Consultas/Juris/Docs/judoc/Acord/20100721/AC_1647_25_10_P.doc. Acesso em: 15 jan. 2018.

TCU. **Acórdão nº 2.601/2011**. Plenário. Relator: Ministro Valmir Campelo. Sessão de 28/9/2011. Disponível em: http://www.tcu.gov.br/Consultas/Juris/Docs/judoc/Acord/20111003/AC_2601_40_11_P.doc. Acesso em: 15 jan. 2018.

THE ECONOMIST. **The world's most valuable resource is no longer oil, but data**. The Economist, Londes, 6 mai. 2017. Disponível em: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data. Acesso em: 11 abr. 2019.

TREVATHAN, J.; READ, W. Detecting Shill Bidding in Online English Auctinos. **ACM Transactions on Computacional Logic**, v. 5, mai. 2006.

WANG, W.; HIDVÉGI, Z.; WHINSTON, A. B. **Shill Bidding in English Auctions**. Austin, 2001.

ZINS, C. Conceptual approaches for defining data, information, and knowledge. **J. Am. Soc. Inf. Sci**., [s. l.], v. 58, n. 4, p. 479-493, jan. 2007. DOI: 10.1002/asi.20508. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1002/asi.20508. Acesso em: 15 abr. 2019.