

Caracterização de competitividade de pregões eletrônicos por meio de mineração de dados

Characterization of competitiveness in electronic auctions by means of data mining

Ricardo Akl Lasmar de Alvarenga^a, Remis Balaniuk^b,
Hercules Antonio do Prado^c, Edilson Ferneda^d

^a ricardoakl@gmail.com

^b remis_balaniuk@yahoo.com

^c prado.hercules@gmail.com

^d eferneda@gmail.com

Resumo: Para adquirir serviços ou produtos os órgãos do Governo Federal brasileiro devem realizar uma licitação. Dentre as modalidades de compra contempladas pela licitação está o pregão eletrônico que, assim como as demais, está sujeito a fraudes. Os comportamentos dos fraudadores modificam-se constantemente e detectá-los é uma tarefa complexa. A presente pesquisa tem como objetivo aferir o grau de competitividade dos pregões eletrônicos por meio da utilização de padrões obtidos pela aplicação de técnicas de Mineração de Dados (MD) no Data Warehouse (DW) das compras governamentais brasileiras. O principal resultado alcançado foi a caracterização de grupos de pregões com diferentes graus de competitividade, a partir de indicadores especificados para este fim, com base no conhecimento de especialistas do ramo.

Palavras chave: Compras Governamentais, Pregão Eletrônico, Competitividade, Descoberta de Conhecimento em Base de Dados, Mineração de Dados.

Abstract: For purchasing services or products, agencies of the Brazilian Federal Government must perform a bidding. Among the purchase modalities considered a bidding is the electronic auction, which, like any other modality, is subject to fraud. The behavior of fraudsters is constantly changing and detecting them is a complex task. This research aims to assess the degree of competitiveness of electronic auctions using standards obtained by applying Data Mining (MD) techniques in the Data Warehouse (DW) of Brazilian government purchases. The main result achieved was the characterization of trading groups with different degrees of competitiveness, based on indicators specified for this purpose, based on the knowledge of specialists in the field.

Keywords: Government Procurement, Electronic Auction, Competitiveness, Knowledge Discovery in Database, Data Mining.

1. Introdução

Para adquirir serviços ou produtos, os órgãos do Governo Federal brasileiro devem realizar uma licitação, processo administrativo pelo qual se busca selecionar a proposta mais vantajosa e proporcionar igual oportunidade aos fornecedores (licitantes), sejam eles pessoas físicas ou jurídicas. Este processo é regulado pela lei ordinária brasileira nº 8.666 (Brasil, 1993) que define as seguintes modalidades de licitação: Concorrência, Tomada de Preços, Leilão, Concurso e Convite. Em 2002, a lei 10.520 (Brasil, 2002) instituiu o pregão no ordenamento jurídico brasileiro para aquisição de bens e

serviços comuns em processos licitatórios.

Apesar de todo cuidado que cerca a realização de uma licitação, desde a elaboração do edital até a adjudicação dos resultados, com frequência a imprensa divulga casos de fraude. Em um cenário em que recursos financeiros dos órgãos governamentais são desviados de forma criminosa por indivíduos ou grupos organizados, áreas importantes como transporte, saúde e educação são privadas de investimentos, deixando à população serviços públicos de baixa qualidade. Os gastos do governo, com dinheiro público, precisam estar claros e analisá-los é um trabalho de fundo cívico essencial.

Com todo rigor que caracteriza a legislação

sobre o assunto, práticas ilícitas sempre ocorrem, requerendo da parte do poder público o aprimoramento constante do processo de compras governamentais.

Para evitar corrupção em licitações, é necessário fiscalizar e controlar as atividades exercidas por pessoas, órgãos governamentais e fornecedores para não ocorrer desvios das normas pré-estabelecidas, pois dar chance ampla e igual a todos os fornecedores, além de ser uma questão de justiça, é uma questão econômica que permite aquisições pelo menor preço. A observação de padrões de comportamento dos vários atores que participam de uma licitação permite a identificação de situações anômalas possivelmente associadas a atos ilícitos.

O controle governamental no Brasil é exercido pelos Tribunais de Contas (TC) e pela Controladoria Geral da União (CGU). Estes órgãos possuem autonomia para fiscalizar os processos de compras do governo, punir os fraudadores e prevenir a ocorrência de fraudes.

Os Tribunais de Contas exercem jurisdição administrativa, julgando contas dos administradores de recursos públicos, e são organizados da seguinte maneira:

- Um Tribunal de Contas da União (TCU), com sede no Distrito Federal e representação em todas as Unidades da Federação;
- Vinte e seis Tribunais de Contas Estaduais (TCE), sendo um em cada Unidade da Federação;
- Quatro Tribunais de Contas dos Municípios, localizados nos Estados da Bahia, Ceará, Pará e Goiás; e
- Dois Tribunais de Contas Municipais, localizados nos Municípios de São Paulo e Rio de Janeiro.

Vale ressaltar que o Tribunal de Contas dos Municípios é distinto do Tribunal de Contas Municipal, pois o primeiro fiscaliza as contas relacionadas aos municípios pertencentes a um estado e o segundo fiscaliza as contas de um município específico.

De forma complementar, a Controladoria Geral da União (CGU) é responsável por assistir a Presidência da República quanto a assuntos relativos à defesa do patrimônio público e ao incremento da transparência da gestão por meio de atividades de controle interno, auditoria, correição, prevenção, combate à corrupção e ouvidoria.

Dentre as modalidades de compras

contempladas pela licitação, aborda-se o Pregão Eletrônico, a terceira mais utilizada pelos órgãos para aquisição de materiais e serviços.

Esta pesquisa visa, por meio da exploração de técnicas de Mineração de Dados, analisar a competitividade de pregões eletrônicos para identificação de compras forjadas ou de baixa concorrência. A análise foi realizada sobre os dados de compras do Sistema Integrado de Administração de Serviços gerais - SIASG (www.comprasnet.gov.br/gerencial/menu_siasg.htm), gerido pelo Ministério do Planejamento, Orçamento e Gestão.

Após análise da estrutura destes dados, foram identificadas entidades relevantes - compras, órgãos, pregoeiros, fornecedores e funcionários - que podem ter comportamentos irregulares, dentre estes comportamentos está o conluio, prática que tem reflexo na competitividade do processo licitatório.

O conluio é a caracterização de ações combinadas entre entidades que visam, irregularmente, obter vantagem perante o mercado. Em licitações isto é muito prejudicial aos órgãos governamentais, pois pode privilegiar um fornecedor, ferindo a lei da livre concorrência ou fazer com que os órgãos adquiram produtos e serviços com preços acima do praticado no mercado. Essa prática pode ser realizada por meio de diversas combinações de entidades como - órgão-fornecedor, pregoeiro-fornecedor, fornecedor-fornecedor e órgão-pregoeiro - e resultam na diminuição da concorrência.

O propósito central deste trabalho é identificar se os pregões eletrônicos realizados pelos órgãos do Governo Federal brasileiro estão de acordo com os padrões de competitividade para possível identificação de conluios.

2. Pregão eletrônico

O Pregão Eletrônico é uma modalidade de licitação adotada pelo Governo Federal brasileiro para contratação de produtos ou serviços que foi criada por meio da Lei Federal 10.520 (Brasil, 2002).

Esta modalidade aumenta a competitividade das aquisições, pois facilita a participação de fornecedores localizados em regiões geográficas distintas da contratante, uma vez que dispensa a presença física no local do leilão porque todo o certame é realizado e acompanhado pela internet. Para participar de um pregão eletrônico basta que os fornecedores tenham uma

infraestrutura de TI com acesso à rede mundial de computadores.

Para conduzir um pregão eletrônico os órgãos da administração federal devem utilizar o ComprasNet, um dos subsistemas do SIASG. Os participantes e envolvidos no processo devem estar devidamente cadastrados no ComprasNet e todas as transações são efetuadas de modo seguro, uma vez que os acessos são concedidos mediante autenticação e os dados trafegam de forma criptografada.

O pregão ocorre em uma sessão pública com horário marcado e semelhante a um *chat* onde os fornecedores apresentam as suas propostas e o pregoeiro - pessoa designada pelo órgão contratante para conduzir o certame - instiga os participantes a diminuírem os valores - a identidade dos autores dos lances não é divulgada aos concorrentes - até que chegue a um ponto onde não há mais redução. Funciona como um leilão às avessas, pois no leilão tradicional o vencedor é quem propõe o maior lance já, no pregão, é quem lança o menor preço.

Finalizado o processo, o fornecedor vencedor deverá comprovar a sua habilitação, mediante encaminhamento da documentação via fax, com posterior envio dos originais, ou cópias autenticadas. Se a proposta de menor valor não for aceitável, ou se o licitante não atender às exigências de habilitação, o pregoeiro examinará a proposta subsequente, procedendo à sua habilitação, na ordem de classificação, sucessivamente, até a apuração de um que atenda ao edital.

Devido a estas facilidades, este processo proporciona agilidade, garante a transparência, otimização de recursos e redução de custos

Embora haja transparência no processo, diversas práticas conhecidas são empregadas para burlar o princípio da livre concorrência. Editais confusos ou direcionados servem para inibir a participação de alguns fornecedores. Empresas mal-intencionadas, muitas vezes agindo em grupo, se especializaram em confundir outras empresas durante os leilões, buscando fazê-las desistir do certame. Acordos prévios entre comprador e vendedor ou entre concorrentes podem transformar o pregão num jogo de cartas marcadas.

Os dados visíveis do sistema de pregões eletrônicos não permitem detectar a maioria dessas práticas ilegais. Só a análise manual de toda a documentação do processo de compra permite, em alguns casos, provar que houve

prática ilícita. Mas dado o número de pregões que ocorrem diariamente, é impossível aos órgãos de controle analisar em detalhe todos os processos. A única solução é fazer uma pré-seleção, baseada em critérios razoavelmente sólidos, que indique quais compras apresentam riscos e devem ser analisadas em detalhe. Um bom critério é o da competitividade aparente do pregão.

É de se esperar que em pregões reais haja uma intensa competição entre fornecedores. Um pregão de baixa competitividade indica um alto risco de irregularidade ou de má gestão por parte do comprador, ainda mais quando um mesmo órgão é recorrente em casos de pregões pouco competitivos quando comparado com seus congêneres.

As questões envolvidas nesta pesquisa são: Como medir a competitividade de um pregão? Quais as escalas e quando é que um pregão pode ser considerado como de risco? O objetivo desse trabalho é tentar encontrar respostas para essas perguntas através da mineração de dados.

3. Revisão da literatura

Poucas publicações foram encontradas em fontes acadêmicas envolvendo as palavras-chave desta pesquisa (Descoberta de conhecimento em base de dados; Conluio; Detecção de fraudes; Compras governamentais; Pregão eletrônico), retratando que o tema é pouco explorado no Brasil. O mesmo ocorre no âmbito internacional. Dentre os trabalhos encontrados, os que mais se aproximam do contexto desta pesquisa abordam a detecção de fraudes por meio da utilização de técnicas de mineração de dados (MD).

Como contexto para a discussão da literatura pertinente, foi utilizado o trabalho de Phua et al. (2005), que expõe uma visão geral da aplicação das técnicas de MD para detecção de fraudes em diversos campos. Os autores sintetizaram diversos estudos, descrevendo os métodos e técnicas de MD usualmente aplicados para a detecção de fraudes e as respectivas aplicações.

Inicialmente, os autores abordaram a importância que as empresas dão, atualmente, à detecção de fraudes devido ao grande prejuízo que os fraudadores têm causado. Para melhor identificação dos perfis dos fraudadores os autores classificam-nos em: (i) internos, gerentes e funcionários ligados à instituição inerente a

fraude, e (ii) externos, clientes e fornecedores que podem ter os perfis de fraudadores habituais, criminosos e crime organizado.

Após enfatizar a importância da mitigação e detecção de fraudes, os autores julgam que as várias ações tomadas nesse sentido, quando operacionais, são geralmente caras, lentas e com baixa eficácia. Nesse contexto, as técnicas de MD ganham importância, pois são capazes de otimizar o processo por meio de atividades que proporcionam automatização de processos, redução de custos e diminuição ou mesmo eliminação de esforço manual.

A utilização de técnicas de MD pode ser compensadora para detectar transações fraudulentas e apontar comportamentos suspeitos de forma *online* ou *offline*. Entretanto, apesar dessa detecção ser eficiente em alguns casos, não se pode acomodar com a identificação de um único padrão, já que os fraudadores mudam constantemente de comportamento visando desenvolver novas formas de transgredir as barreiras desenvolvidas.

Apesar das vantagens em utilizar MD serem óbvias, há dois fatores que dificultam o amadurecimento dessa disciplina na detecção de fraudes. A falta de dados reais para realização dos experimentos é um deles. Devido à dificuldade de obter dados reais, muitos pesquisadores criam dados sintéticos para testar seus modelos. Barse, Kvarnstrom e Jonsson (2003) admitem que os dados sintéticos podem ser úteis no treinamento de bases para identificação de comportamentos fraudulentos, mas ressaltam que estes comportamentos devem ser, posteriormente, buscados em bases reais. Além disso, há resistência de se publicar os estudos, pois assim os fraudadores tomariam conhecimento dos comportamentos mapeados e produziram novos tipos de ações fraudulentas.

Após relatar, de modo genérico, a utilização de MD para a detecção de fraudes, os autores aprofundam o assunto e discutem os tipos de aprendizagem de máquina que podem ser aplicados ao tema:

- A aprendizagem supervisionada examina todas as ações “classificadas” anteriormente (legais e ilegais) para caracterizar, matematicamente, quão fraudulenta é a transação. Este tipo de aprendizagem é parecido com pontuação de riscos e requer treinamento para identificação de padrões fraudulentos em dados históricos e detecta apenas fraudes que ocorreram previamente.

- A aprendizagem não-supervisionada utiliza dados não classificados e detecta fraudes que não foram identificadas anteriormente por meio da identificação de padrões anômalos.
- A aprendizagem semi-supervisionada mescla a aprendizagem supervisionada com a não-supervisionada.

Após analisar as técnicas utilizadas para cada tipo de aprendizagem, os autores concluem que técnicas mais complexas como Redes Neurais e *Support Vector Machine* (SVM) são mais utilizadas, porém, alternativas menos complexas como Redes Bayesianas e métodos estatísticos podem produzir resultados tão bons quanto aqueles, e se os dados analisados precisam ser examinados em tempo real, a utilização de aprendizagem supervisionada pode ser ineficaz.

Fawcett (1997) cita uma ideia muito utilizada para detecção de fraudes originada da aplicação de MD em serviços de *antispam*. Nessa abordagem, busca-se entender a natureza temporal da irregularidade incluída em uma *blacklist* visando a configuração do comportamento do motor de varredura para automatizar o processo de remoção dos *spams*. Além disso, o autor enfatiza a existência de barreiras que prejudicam a identificação de fraudes com MD, como o grande volume de dados, os diversos modelos de fraudes, a mudança das leis e dos comportamentos dos fraudadores.

Abbott, Matkovsky e Elder (1998) apresentam um estudo que compara as ferramentas de MD do mercado utilizadas para detecção de fraudes, visando prover aos desenvolvedores de aplicações subsídios para a escolha da ferramenta mais adequada a uma determinada aplicação.

Para Weatherford (2002), a disciplina de MD possui dois desafios relacionados à detecção de Fraudes: (i) elaborar algoritmos que possam aprender sobre os diversos cenários de fraude e identificar e prever novos cenários e (ii) elaborar sistemas que trabalhem rápido o suficiente para detectar atividades fraudulentas em tempo de execução. Para a autora, a utilização de redes neurais é vantajosa por proporcionar flexibilidade, velocidade e adaptatividade.

Kou et al. (2004) abordaram a verificação de comportamentos suspeitos para detectar, estimar e impedir fraudes relacionadas a cartão de crédito, telecomunicação e detecção de intrusão em ambientes computacionais pela aplicação de técnicas de Inteligência Artificial, MD e

Estatística, analisando bases de dados, *logs* de sistemas e o comportamento dos usuários. Outro ponto explorado pelos autores é a utilização de indicadores de desempenho para avaliar a eficácia da técnica aplicada, pois o objetivo principal na detecção de fraudes é maximizar as previsões corretas e minimizar as incorretas. Por exemplo, no caso de se construir classificadores (aprendizagem supervisionada), poder-se-ia utilizar como métricas de qualidade a taxa de erro do modelo, a média de falso positivo/negativo e a rapidez da detecção. Os autores citam apenas as abordagens supervisionada e não-supervisionada e, para cada classe de estudo, são utilizadas abordagens específicas. Assim, (i) para fraudes com cartões de crédito, sugerem o uso de Redes Neurais; (ii) para detecção de intrusão em ambientes computacionais, indicam o uso de Sistemas Especialistas, Redes Neurais, Modelos Baseados em Raciocínio, Modelos de Associação/Classificação, Análise de Transição de Estado ou Algoritmos Genéticos; (iii) no contexto das telecomunicações, sugerem Redes Neurais e Métodos de Visualização.

Chen et al. (2004) focam em investigação policial. Os autores afirmam que as técnicas de MD são vantajosas para as investigações, pois podem analisar bases de dados extensas de forma rápida e eficiente. Os custos destes softwares são menores do que o treinamento de recursos humanos e os computadores são menos propensos a falhas do que os seres humanos, uma vez que as investigações precisam de horas contínuas de dedicação e análise. Dentre as técnicas citadas pelos autores estão: Extração de Entidades, Análise de Agrupamentos, Regras de Associação, Detecção de Desvios, Classificação, Comparação de *Strings* e Análise de Redes Sociais. Um estudo de caso mostra como a utilização de MD, por meio de análise de vínculos, desvendou 16 membros-chaves de gangues, bem como os seus grupos e subgrupos.

Apparao et al. (2009) apresenta um *framework* para a aplicação de MD em extensas bases de dados com o objetivo de detectar fraudes. O *framework* segue o fluxo tradicional de informação para MD: coleta de dados, integração de dados, processamento de dados, mineração de dados e avaliação de padrões. Os dados utilizados são classificados como oriundos de “companhias com fraude e sem fraude”, “auditoria” e “governança corporativa”. Classificação, análise de agrupamentos e a

identificação de séries temporais foram as abordagens de detecção citada se as técnicas utilizadas foram: regressão, redes neurais, árvore de decisão, redes bayesianas e SVM.

Nos estudos analisados, verificou-se ausência de abordagens voltadas para a identificação de situações anômalas em licitações, justificando, assim, a pertinência ao presente estudo.

4. Mineração de Dados, CRISP-DM e Indicadores

Mineração de Dados é o processo de explorar grandes quantidades de dados para identificação de padrões que podem auxiliar a descoberta de conhecimento. Este conhecimento pode ser apresentado de diversas formas como: agrupamentos, regras, hipóteses e grafos. Essa disciplina se desenvolveu devido ao crescimento exorbitante do volume de dados disponíveis, o que inviabiliza a análise humana e fortalece o conceito de que o computador pode ser ágil e útil para detectar novos padrões e conceitos. Para aplicar mineração de dados em informações relacionadas às compras governamentais, neste estudo, foi utilizada a metodologia CRISP-DM (Chapman et al., 2000), que divide em fases as atividades do projeto, as quais foram adaptadas para abranger o contexto desta pesquisa:

- Entendimento do negócio - estudo dos principais conceitos relacionados a compras governamentais (foco em pregão eletrônico) e comportamentos suspeitos.
- Entendimento dos dados - estudo de como os dados estão distribuídos para serem capturados e combinados, de forma que possam ser aplicados no contexto da pesquisa.
- Preparação dos dados - construção de um conjunto de dados derivado da base inicial. Nesta fase os dados úteis à pesquisa são selecionados, transformados e limpos para serem utilizados.
- Modelagem - seleção, aplicação e calibragem da técnica de mineração utilizada no projeto.
- Avaliação - avaliação dos resultados obtidos para averiguar se os objetivos do negócio foram atingidos.
- Aplicação - final do projeto. Apresentação dos resultados obtidos às partes interessadas e demonstração de como utilizá-los.

Para estabelecer medidas relevantes na

perspectiva da pesquisa e viabilizar a mineração de dados, foram definidos indicadores.

Não há uma metodologia ou processo padrão para definição de indicadores, porém, a Secretaria de Gestão do Ministério do Planejamento elaborou, a partir da revisão das principais experiências de construção de indicadores para o setor público e privado, um roteiro para assegurar a coerência e objetividade na formulação de indicadores (Brasil, 2009), o qual foi utilizado neste trabalho.

5. Aplicação

5.1. Compreensão do negócio

Por meio de fiscalização e controle dos processos licitatórios é possível identificar eventuais irregularidades como o desvio de recursos públicos. Geralmente, estes desvios visam beneficiar fornecedores e/ou pessoas e são caracterizados como corrupção.

Os órgãos controladores do Governo Federal Brasileiro trabalham constantemente com métodos e ferramentas para detecção de irregularidades, porém, à medida que os fraudadores descobrem como as irregularidades são detectadas, criam novos comportamentos para burlar os meios de prevenção. Essa constante mudança de comportamento leva estas entidades a um trabalho contínuo de Inteligência.

Esta pesquisa propõe um método para auxiliar os órgãos controladores a medirem o grau de competitividade de pregões eletrônicos, o que pode levar a detecção de irregularidades em licitações.

Um pregão eletrônico competitivo tem como característica a participação de diversos fornecedores que estipulam lances para baixar o preço da venda, gerando economia para o contratante. No final do certame espera-se que o preço estipulado pelo vencedor do processo licitatório seja inferior ao estimado no início do pregão e que tenha ocorrido competição entre os fornecedores.

Analisando o cenário de um pregão eletrônico competitivo, foram levantadas hipóteses que poderiam caracterizar comportamentos de processos com pouca competitividade para uma compreensão mais ampla do fenômeno de conluio em compras governamentais, com foco em pregão eletrônico, e atestar a viabilidade desta pesquisa. Aqui estão algumas delas:

- *Empresas Irmãs* possuem vínculos e

participam dos pregões juntas para terem vantagens perante os lances. São exemplos de vínculo: (i) empresas distintas com o mesmo endereço e (ii) empresas com sócios em comum.

- *Empresa “coelho”* é uma instituição que participa de pregões eletrônicos para beneficiar um dos outros participantes e prejudicar os demais. Este tipo de empresa age lançando preços baixos, sucessivamente, forçando os demais concorrentes a também abaixarem o lance até que haja desistência e/ou desclassificação. Enquanto todas as empresas são eliminadas, o fornecedor que age em conluio com a empresa coelho mantém a cautela, aumentando a sua chance de vencer o pregão eletrônico.
- *Cartel* é a união de duas ou mais empresas para se beneficiarem perante a concorrência. Como exemplos de ações de cartel em pregões eletrônicos pode-se citar (i) grupos de empresas que participam constantemente das mesmas concorrências e que revezam, com certa frequência, a vitória no processo seletivo e (ii) empresas que vendem produtos de um mesmo fabricante e que participam das mesmas concorrências, porém, a combinação vencedor-órgão é a mesma.
- Pode-se caracterizar uma *relação pregoeiro-empresa* quando determinado pregoeiro declara a mesma empresa como vitoriosa nos pregões eletrônicos repetidas vezes.
- Uma *relação órgão-empresa* pode ser caracterizada quando uma empresa ganha muitas concorrências em um mesmo órgão.
- A *vitória de empresas novas em pregões de grande porte* pode ser tendenciosa, além de este tipo de companhia poder ser “fantasma” (servir de fachada para justificar determinado custo).
- Dá-se a denominação de *Laranja* a um dono da empresa que participa da concorrência ganha pouco oficialmente ou possui um cargo para pessoas com baixo grau de instrução com o objetivo de disfarçar os órgãos fiscalizadores.
- Dá-se o nome de *apoio amigo* quando o órgão que abriu o pregão possui algum colaborador que possa beneficiar determinado fornecedor. Por exemplo, quando o dono da empresa vencedora já trabalhou ou trabalha no órgão ou quando o dono ou algum colaborador da empresa vencedora do pregão possui parente

no órgão.

- *Dono insistente* são os donos de empresas que ganham pregões, vão à falência, mas abrem outras e repetem a ação.

Considerando as hipóteses supracitadas e a viabilidade de mineração nos dados de compras governamentais dos órgãos do Governo Federal brasileiro, o objetivo desta pesquisa é identificar: (i) pregões eletrônicos com grau de competitividade irregular e (ii) entidades suspeitas de praticarem comportamentos irregulares.

5.2. Compreensão dos dados

O principal recurso de dados utilizado no projeto é o DW - Compras, Orçamento e Gestão. A Figura 1 representa a constelação (estrutura) do DW. Analisando a imagem é possível identificar as entidades envolvidas nos processos licitatórios e pertencentes à base de dados explorada.

Há várias entidades relacionadas neste DW e nem todas contemplam o escopo do projeto. Além disto, a base é volumosa, o que pode tornar lenta a recuperação dos dados.

Para tratar estes contratempos foi necessário

apontar as entidades relevantes para o projeto e definir uma estratégia para mitigar os riscos ocasionados pelo alto volume de dados. As entidades relevantes para o projeto e que foram mais exploradas são: (i) compra, (ii) contrato, (iii) fornecedor, (iv) itens da compra (materiais/serviços) e (v) órgão.

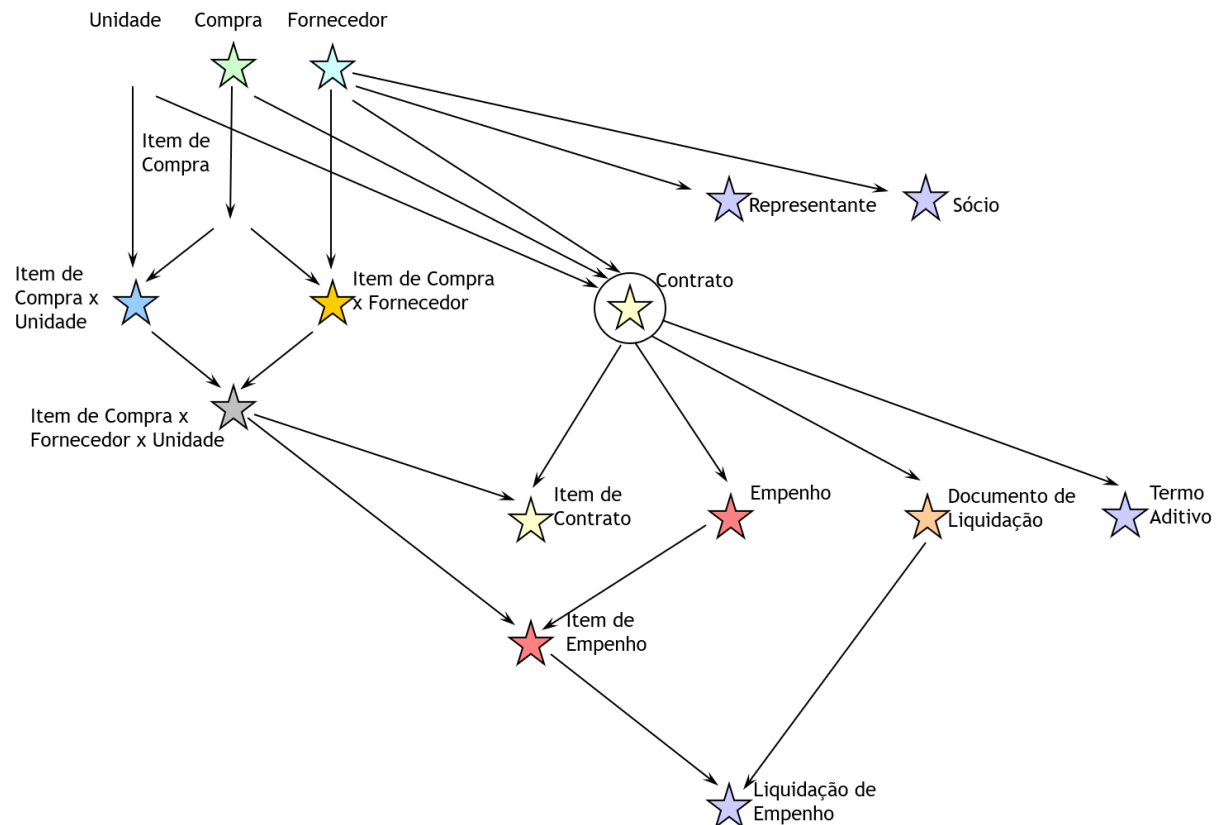
Para diminuir o volume de dados e agilizar a exploração destes foram definidos indicadores para serem alimentados com informações extraídas do DW, os quais foram organizados em níveis e dimensões.

Os níveis foram definidos a partir das entidades e classificados como: (i) Pregão, (ii) Contrato, (iii) Fornecedor, (iv) Itens e (v) Órgão.

As dimensões foram definidas a partir da análise das características de um processo licitatório, fatores que podem indicar irregularidades e comportamentos comuns a todos os níveis: (i) capacidade, (ii) competitividade, (iii) confiabilidade, (iv) conformidade e (v) economicidade. Analisando esta estrutura (níveis e dimensões) é possível definir uma série de indicadores para cada nível e agrupá-los em dimensões.

Além da definição dos indicadores, foram analisados apenas os pregões eletrônicos

Figura 1. Constelação do DW.



Fonte: Os autores.

organizados por órgãos com sede no Distrito Federal. Este modelo de definição de indicadores e a restrição da Unidade da Federação ajudaram a organizar os dados relevantes para o escopo da pesquisa e a diminuir o volume do DW.

Após aplicação das restrições supracitadas, a base contemplou 11.623 (onze mil, seiscentos e vinte e três) pregões, os quais foram organizados por 228 órgãos e tiveram a participação de 13.053 (treze mil e cinquenta e três) fornecedores.

5.3. Preparação dos dados

Com os níveis e as dimensões bem definidos, muitos indicadores foram elencados.

Para se chegar a uma definição adequada destes indicadores foram necessárias diversas reuniões com especialistas do negócio para validação, pois nem todos eram relevantes ou viáveis de serem extraídos da base.

Este processo de validação ocorreu seguindo os seguintes passos:

1. Validação dos níveis;
2. Validação das dimensões;
3. Definição de indicadores para cada nível/dimensão;
4. Verificação de viabilidade de extração dos indicadores;
5. Verificação de relevância dos indicadores para a pesquisa;
6. Eliminação dos indicadores inviáveis ou irrelevantes;
7. Adequação de alguns indicadores;
8. Elaboração da tabela final com a descrição dos indicadores.

Após conclusão do processo de validação, foram definidos noventa e três indicadores, os quais foram alimentados com dados extraídos do DW. O volume ainda foi considerado alto e chegou-se ao consenso de que a redução seria pertinente, de modo que fossem utilizados os indicadores com maior relevância e gerados novos consequentes da combinação de outros.

Para restringir o escopo da pesquisa foram contemplados apenas os indicadores do nível pregão. Após análise dos indicadores pré-definidos com especialistas do negócio foram definidos sete, considerados como mais relevantes para utilização na mineração. Destes, três já estavam listados e quatro foram resultantes da combinação de dois ou mais indicadores. Essa combinação visou reduzir ainda mais o volume

de dados e produzir informações com alto grau de relevância.

A utilização de sete indicadores foi baseada no número mágico de Miller, “7 mais ou menos 2” (Miller, 1956). Abaixo estão os sete indicadores escolhidos:

- *Número de participantes*, ou *concorrentes* (n_p) - Um pequeno n_p em um pregão indica baixa competitividade;
- *Distribuição de vencedores* - caracteriza a distribuição dos itens entre os fornecedores. Quanto menor a relação entre o número de vencedores distintos (n_v) e o número de itens (n_i) de um pregão, maior a possibilidade de favorecimento a determinado fornecedor.

$$\text{Distribuição de vencedores} = \frac{n_v}{n_i}$$

- *Tempo de pregão* (T) - Um pregão muito curto aponta um processo sumário de licitação.
- *Competitividade* - caracteriza a redução no valor dos preços propostos durante o pregão, pois indica o quanto os competidores estavam dispostos a baixar seus preços. Ela é calculada pela fórmula abaixo, onde $l_{p,i}^p$ e $l_{p,i}^u$ representam, respectivamente, o primeiro e o último lance do participante p para o item i . Quanto mais próximo de zero, maior a competitividade do pregão eletrônico.

Competitividade =

$$\frac{\sum_{i=1}^{n_i} [(\sum_{p=1}^{n_p} l_{p,i}^u / \sum_{p=1}^{n_p} l_{p,i}^p) / n_p]}{n_i}$$

- *Economia* - indica o grau de economia gerada pelo pregão e é calculada pela fórmula abaixo, onde l_i^v representa o lance vencedor do item i . Quanto mais próximo de zero, maior a economia do pregão eletrônico.

$$\text{Economia} = \frac{\sum_{i=1}^{n_i} (l_i^v / \min(l_i^p))}{n_i}$$

- *Acurácia da estimativa* - Indica se a proposta ganhadora estava abaixo ou acima do valor estimado para aquele item. É calculada pela fórmula abaixo, onde v_i^H representa o valor homologado para o item i e v_i^E representa o valor estimado para o item i . Se o valor for maior que um é ruim, pois foi gasto mais que o estimado, caso seja muito próximo a zero também é ruim porque o preço foi estimado erroneamente. O ideal é que o resultado seja menor e próximo a um.

$$\text{Acurácia da estimativa} = \frac{\sum_{i=1}^{n_I} (v_i^H / v_i^E)}{n_I}$$

- Número de lances (n_L) - O número de lances de um pregão indica o interesse dos participantes em ganhar a concorrência para o item.

O primeiro passo após a definição final dos indicadores foi alimentá-los com dados extraídos do DW. Estes dados foram armazenados em arquivos CSV para facilitar futuras integrações, uma vez que este formato é um padrão comportado pela maioria das ferramentas de manipulação de dados.

5.4. Modelagem

Existem diversas técnicas de modelagem implementando soluções para as tarefas de Mineração de Dados. Em consequência, ao se iniciar um trabalho de mineração é preciso analisar a tarefa em questão e selecionar a técnica ou técnicas mais adequadas.

Muitas vezes, as técnicas que são julgadas mais adequadas podem não ser utilizadas devido a restrições em dados e de recursos. Após escolher a técnica a ser aplicada, é selecionada a ferramenta para mineração dos dados, pois estas podem ser limitadas, atendendo um grupo específico de técnicas.

Ao analisar o contexto desta pesquisa foram levantadas diversas situações de fraude. Entretanto, não haviam disponíveis instâncias rotuladas que permitissem a construção de um classificador. Dessa forma, a aprendizagem não-supervisionada tornou-se a abordagem de MD mais adequada.

Dentre as opções relacionadas ao modelo de aprendizagem não-supervisionada foi escolhida a segmentação, partindo da premissa de que os pregões forjados ou com baixa concorrência saíam do padrão de comportamento da maioria se fossem definidos indicadores relevantes para aplicação da mineração.

Com a técnica definida foram pesquisadas algumas ferramentas de mineração de dados que contemplassem algoritmos de análise de agrupamentos, sendo selecionada a WEKA por se tratar de uma ferramenta gratuita e utilizada na disciplina Descoberta de Conhecimento em Base Dados do programa de mestrado em Gestão de Conhecimento e da Tecnologia da UCB.

O primeiro passo da construção do modelo foi consolidar os indicadores em um único arquivo CSV com diversas tuplas para viabilizar a

importação dos dados no WEKA. Cada tupla referencia um pregão e seus respectivos indicadores.

Com os dados importados iniciou-se a mineração. Vale ressaltar que, dentre os atributos selecionados para a mineração, foram excluídos os que faziam referência a código de identificação e descrição e que a técnica utilizada foi à segmentação. O WEKA possui diversos algoritmos de análise de agrupamentos e para melhor produtividade do experimento foi necessário escolher um. Foram testados os algoritmos *Cobweb*, *DBScan*, *EM* (*Expectation Maximization*), *Farthest First*, *Filtered Clusterer*, *Hierarchical Clusterer*, *Make Density Based Clusterer*, *OPTICS* e *Simple K-Means*. No primeiro momento os algoritmos *EM* e *Fatherst First* foram os que geraram agrupamentos com melhor distribuição, porém, apenas um deveria ser escolhido e *EM* foi selecionado por ser mais conhecido pelos envolvidos na pesquisa. *EM* é um algoritmo de geração de estimativas de máxima verossimilhança surgiu da unificação de uma série de trabalhos sem relação e apresentado por Dempster et al. (1977).

Com o algoritmo selecionado foram gerados diversos resultados, com quatro, cinco e seis agrupamentos, utilizando os seguintes parâmetros: *maxIterations* = 100, *minStdDev* = 1.0E-6 e *Seed* = 100. Os parâmetros foram definidos configurando as variáveis com diversos valores e analisando os resultados obtidos até chegar em números que geraram agrupamentos interessantes.

A cada geração procurou-se analisar a coerência dos resultados obtidos. Essa análise foi feita interpretando os centroides dos agrupamentos segundo regras de negócio do pregão. Os resultados mais coerentes foram com 5 e 6 agrupamentos.

Antes de descrever os resultados, é coerente definir a palavra *outlier*, pois possui diversos significados. Na disciplina de Mineração de Dados, *outlier* é a caracterização de um agrupamento, em um conjunto de dados, que apresenta grande afastamento dos demais grupos da série, podendo ser classificado como uma exceção.

A seguir, os resultados gerados pelas duas aplicações que produziram *outliers* e que foram escolhidas para análise aprofundada.

O resultado com seis agrupamentos (1ª aplicação) é apresentado na Tabela 1.

Tabela 1. Distribuição dos agrupamentos na primeira análise.

ID	Membros	%
0	368	3
1	6893	65
2	3318	29
3	567	5
4	131	1
5	346	3

Após análise dos grupos gerados, foram apontados quatro com um número reduzido de pregões (grifados em vermelho na Tabela 1) para uma análise aprofundada, porém, logo na primeira verificação foram detectados atributos com números discrepantes que levaram a identificação de sujeiras na base. Para limpá-la, foram removidas todas as compras pertencentes a estes grupos. Esta identificação justifica o pequeno número de pregões contemplados por estes agrupamentos.

Este resultado pode ser considerado um subproduto da mineração, pois permitiu a identificação de registros incoerentes e pode ser utilizado pelos órgãos de controle para auditoria de bases de sistemas importantes na administração pública.

Nesta análise foram utilizados as médias e os desvios padrão de cada indicador para cada um dos grupos gerados. No entanto, houve atributos que destoaram. Após análise detalhada dos *outliers*, foram identificadas outras compras que não pertenciam aos grupos suspeitos que também foram removidas, devido ao conhecimento gerado pelo primeiro resultado.

A segunda aplicação ocorreu após a realização da limpeza da base, resultante da primeira análise de agrupamentos. O resultado desta nova aplicação é apresentado na Tabela 2.

Tabela 2. Distribuição dos agrupamentos na segunda análise

ID	Membros	%
0	2156	22%
1	2563	26%
2	2561	26%
3	2015	21%
4	526	5%
0	2156	22%

O grupo identificado pelo número 4 (grifado

em vermelho) teve o menor número de membros e, por isso, foi escolhido para ser detalhado, pois caracterizou comportamento particular e destoante dos demais.

Para gerar os agrupamentos citados acima, o algoritmo utilizou todos os indicadores e, para cada grupo, definiu médias de caracterização. As Tabelas 3 e 4 contêm a média e o desvio padrão de cada indicador para cada um dos grupos gerados.

Sob a ótica de uma avaliação técnica, conclui-se que na primeira análise foi utilizado o resultado que gerou seis agrupamentos. Ao estudar estes grupos, foram detectados quatro com número reduzido de ocorrências (compras). Após análise aprofundada concluiu-se que as incoerências ocorreram devido a sujeiras na base. O ponto que ocasionou essa discrepância foi que em diversos registros o valor inicial estava relacionado ao preço unitário do item e o valor homologado ao preço global, com isso a divisão do valor vencedor pelo inicial gerou resultados absurdos. Devido a essa dedução foram removidas todas as tuplas pertencentes a esses quatro grupos para uma nova aplicação de MD.

Com a base limpa, após análise do primeiro resultado, foi realizada uma segunda aplicação que resultou em agrupamentos coerentes e interessantes. Este resultado foi avaliado detalhadamente conforme descrito no próximo capítulo.

5.5. Avaliação

Depois da aplicação dos algoritmos de agrupamentos nos indicadores foram realizadas reuniões com especialistas do negócio para análise dos resultados. Essa análise foi efetuada de maneira subjetiva, tendo como base o entendimento e os objetivos do negócio e técnicos.

Os principais insumos da análise foram às tabelas que descrevem a média de cada indicador em cada agrupamento (Tabela 3). Abaixo estão as descrições para cada um dos agrupamentos:

- *Econômico*, o grupo 0 possui compras com boa economia, pois o valor homologado foi bem abaixo do estimado e a diferença média entre o primeiro e o último lance dos participantes, bem como do valor vencedor e do menor inicial foi alta. O número de lances foi elevado, o que indicou competitividade. Mas ao reparar a divisão de vencedores pelo número de itens do pregão chega-se à

Tabela 5. Resultados das médias da segunda análise de agrupamentos

Cluster	Economia	Número de participantes	Distribuição de Vencedores	Tempo de pregão	Competitividade	Acurácia da estimativa	Número de lances
0	0,1940	18,978	0,2197	261,5194	0,0460	0,0616	895,8765
1	0,0847	13,1949	1	244,3753	0,3981	0,5507	205,1775
2	0,1838	8,8528	1	236,7066	0,8296	0,8967	95,9899
3	0,1500	17,176	0,5372	257,4562	0,1897	0,2540	343,4837
4	0,9083	1,1351	0,8111	215,1179	0,7253	0,7573	4,5215
5	0,1940	18,978	0,2197	261,5194	0,0460	0,0616	895,8765

Tabela 6. Resultados dos desvios padrões da segunda análise de agrupamentos

Cluster	Economia	Número de participantes	Distribuição de Vencedores	Tempo de pregão	Competitividade	Acurácia da estimativa	Número de lances
0	0,1637	18,2252	0,1217	156,1805	0,0467	0,0567	916,4206
1	0,0642	10,3807	0,3388	147,9668	0,1648	0,2210	142,2468
2	0,1401	10,1563	0,3388	147,8689	0,1516	0,1469	92,0120
3	0,1311	15,5387	0,1508	155,8844	0,1355	0,1780	320,1314
4	0,1373	0,4048	0,2868	143,49	0,3214	0,2890	5,8123
5	0,1637	18,2252	0,1217	156,1805	0,0467	0,0567	916,4206

conclusão que apesar da competitividade os ganhadores foram concentrados.

- *Competitivo*, o grupo 1 é caracterizado como competitivo, pois teve compras com boa quantidade de participantes e de lances, a distribuição de vencedores por item foi boa e a diferença da média dos primeiros lances pelos últimos foi alta. Apesar da competitividade, não foi tão econômico como o Grupo 0, pois a diferença do homologado para o estimado e do valor vencedor e do menor inicial não foram altas.
- *Discreto*, o grupo 2 apresentou pouca movimentação e teve baixa economia e competitividade. Apesar de ter tido os vencedores bem distribuídos pelos números de itens, teve poucos lances e participantes e as diferenças entre os preços homologado e estimado, bem como entre a média do último lance e do menor inicial foram baixas.
- *Equilibrado*, o grupo 3 pode ser considerado o grupo mais equilibrado, pois teve bom número de participantes e lances e a diferença entre os preços homologado e estimado, bem como dos valores dos lances foi boa. Muitos arriscariam dizer que este comportamento seria dos pregões ideais, pois além do equilíbrio dos fatores citados acima, a distribuição de vencedores pelos itens foi boa.
- *Suspeito*, o grupo 4 teve comportamento suspeito, pois analisando os dados deste grupo,

chega-se à conclusão que foram pregões com baixa competitividade e que podem indicar irregularidades. Poucos participantes e lances indicam falta de competitividade e a pequena diferença entre os valores estimado e homologado, bem como dos lances indicam pouca economia.

Analisando os comportamentos de cada grupo, pode-se afirmar que:

- O número de participantes e de lances está diretamente relacionado ao grau de competitividade do pregão. Pregões com mais participantes e lances tiveram maior economia.
- A duração do pregão não foi um fator tão relevante na distribuição dos agrupamentos, pois teve valores equilibrados nos diferentes grupos.
- Ao analisar o indicador que verifica se houver muita diferença entre o valor estimado e o homologado, é coerente afirmar que foi determinante para o outlier, pois pouca diferença indica pouca economia.
- As diferenças da média do último lance dividido pelo primeiro, pode indicar a economia das compras.
- A diferença de vencedores pelo número de itens foi a grande surpresa, pois os pregões competitivos e com muita economia ficaram concentrados em determinados fornecedores, o que pode indicar que este seletivo grupo

é formado por empresas com alta capacidade.

Após as análises descritas acima chegou-se à conclusão que o grupo suspeito (grupo 4) foi formado por pregões com baixa concorrência, o que era um dos objetivos desta pesquisa.

Para melhor investigação, foi definido que o grupo discrepante deveria ser analisado com profundidade para produzir insumos que auxiliariam a atingir outro objetivo deste projeto: identificação de pregões forjados e, para isto, foram gerados gráficos para relatar os envolvidos e as características chaves.

Vale reforçar que não é possível afirmar se determinado pregão foi forjado apenas com o resultado desta pesquisa, é necessária uma análise mais aprofundada.

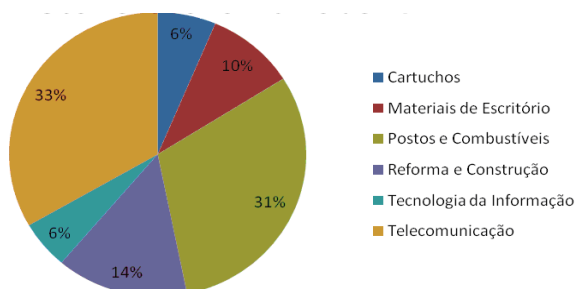
A geração dos gráficos, listados abaixo, seguiu duas vertentes: (i) geração de dados das compras como um todo e (ii) geração de dados referentes a cada item das diversas compras.

1. Ramos de atividades dos treze fornecedores com maior número de ocorrências

Os gráficos relacionados neste tópico visam identificar características das treze empresas que mais apareceram nas compras pertencentes ao *outlier*.

O resultado apresentado na Figura 2 mostra que as empresas que mais participaram das compras do *outlier* são dos ramos de atividade: “Cartuchos”, “Materiais de Escritório”, “Postos e Combustíveis”, “Reforma e Construção”, “Tecnologia da Informação” e “Telecomunicação”. As compras que envolveram a aquisição de produtos dos segmentos “Cartuchos”, “Materiais de Escritório”, “Reforma e Construção” e “Tecnologia da Informação” merecem uma investigação aprofundada, pois são ramos com muitas empresas no mercado e com produtos homogêneos, consequentemente, estes pregões deveriam ter se enquadrado em um dos grupos com alta concorrência.

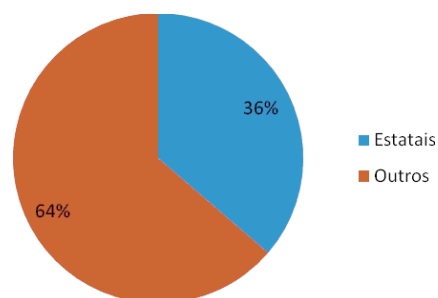
Figura 2. Percentual de Fornecedores por ramo de atividade.



Os ramos “Postos e Combustíveis” e “Telecomunicação” reforçam o conceito de que o *outlier* foi formado por pregões com baixa concorrência, pois são compostos por empresas estatais, como a Petrobras, ou com baixo número de concorrentes como o segmento de telecomunicação.

O gráfico da Figura 3 comprova o que foi descrito no parágrafo anterior: há um alto número de empresas estatais dentre as ocorrências.

Figura 3. Empresas Estatais x Empresas de Outros Tipos



2. Os onze órgãos com maior número de ocorrências

Aqui, identificam-se os Órgãos com mais ocorrências nas compras pertencentes ao grupo de pregões com baixa concorrência, conforme mostrado na Figura 4.

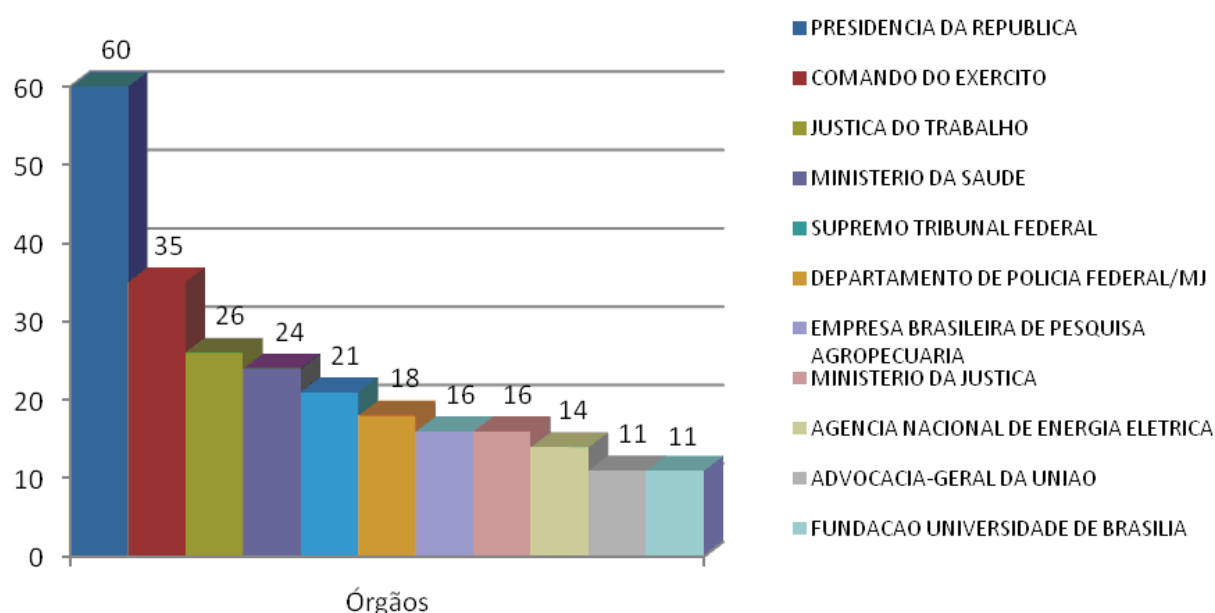
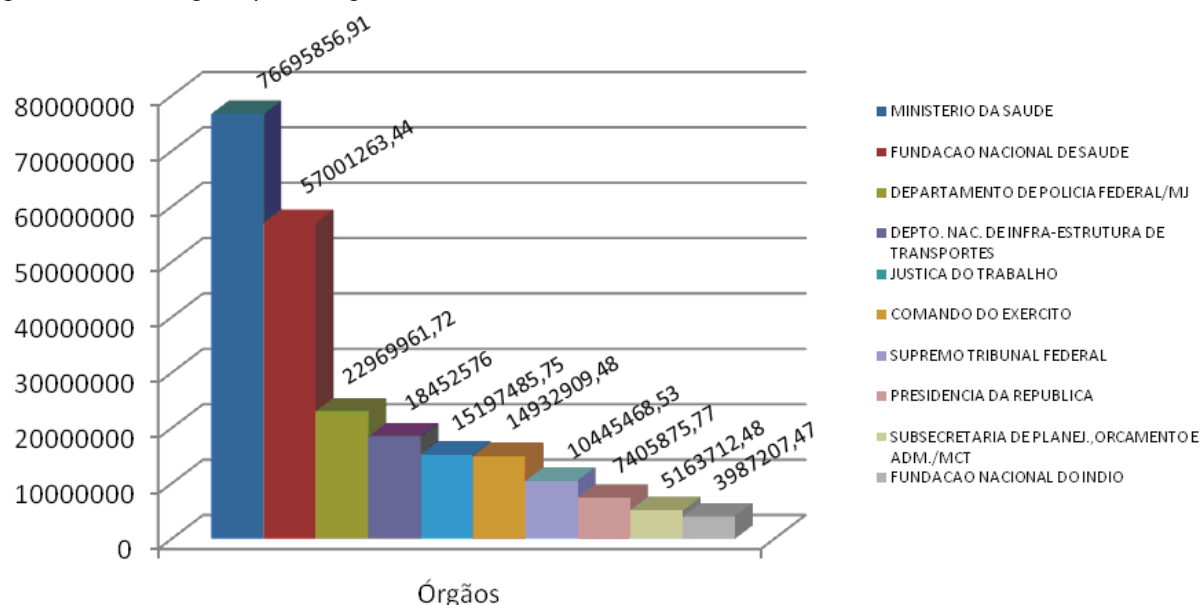
3. Os dez órgãos que mais gastaram

Aqui, visa-se complementar o “TOP 10 - Órgãos (Ocorrências)” descrevendo quais foram os dez órgãos pertencentes ao *outlier* que mais gastaram no conjunto de compras do agrupamento, como mostrado na Figura 5.

O Ministério da Saúde, a Fundação Nacional de Saúde e o Departamento nacional de Infraestrutura de Transportes tiveram o gasto bem acima dos demais, provavelmente, porque o custo de itens associados a saúde e transporte são mais elevados.

Dentre os órgãos relacionados nos gráficos acima (Figuras 3 e 4), destacam-se: Presidência da República, Comando do Exército, Justiça do Trabalho, Ministério da Saúde, Supremo Tribunal Federal e Departamento de Polícia Federal porque foram comuns aos dois resultados. Isto pode agravar uma eventual suspeita de irregularidade, pois além dos pregões eletrônicos efetuados por estes órgãos terem baixa concorrência, resultaram em um gasto muito alto para os cofres públicos.

Apesar da afirmação do parágrafo acima, os

Figura 4. Os onze órgãos com mais participação nos pregões**Figura 5.** Os dez órgãos que mais gastaram.

demais órgãos não devem ser menosprezados porque a quantidade de pregões realizados pode estar diretamente ligada ao porte da instituição. A aparição da Fundação Nacional do Índio e da Fundação Universidade de Brasília nestes gráficos é incoerente, pois são organizações de porte bem inferior às demais.

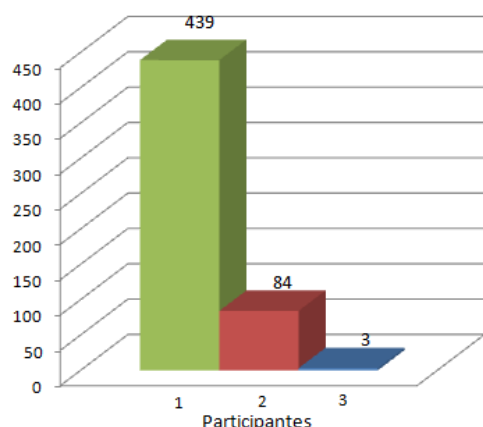
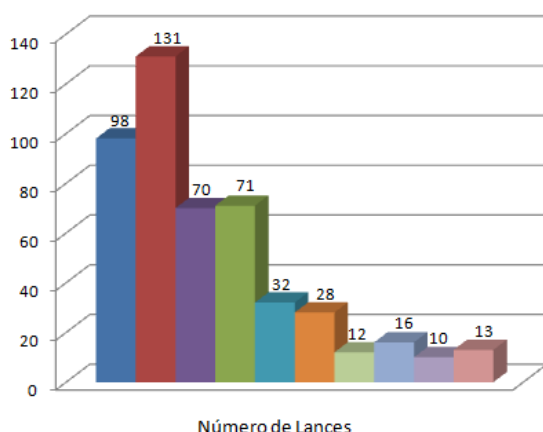
4. Número médio de participantes

Identifica-se aqui, como mostrado na Figura 6, a faixa de número de participantes dos pregões eletrônicos do *outlier*. Este gráfico confirma que o grupo analisado se trata de pregões eletrônicos de baixa concorrência, pois na

maioria dos casos os pregões tiveram apenas um participante. Não faz sentido a ocorrência de pregões com um número tão baixo de participantes.

5. Os dez números de lances com maior ocorrência

Aqui, identifica-se a faixa de número de lances que ocorreram nos pregões pertencentes ao outlier (Figura 7). Este gráfico, assim como o anterior, reforça a teoria de que houve baixa concorrência nestes pregões, pois o número de lances foi muito baixo, tendo dois lances na maioria dos casos.

Figura 6. Número médio de participantes.**Figura 7.** Os dez números de lances dos pregões com maior ocorrência

6. Os dez materiais/serviços que foram mais adquiridos

No gráfico da Figura 8, exibe-se quais foram os dez materiais e serviços que mais apareceram nas compras pertencentes ao grupo suspeito.

Este gráfico é um reflexo do que foi descrito na Figura 2, pois o alto número de compras com produtos e serviços relacionados a telefonia e a petróleo está ligado à existência de pouca concorrência nestes setores, porém, os demais itens devem ser analisados para identificação de possíveis irregularidades.

7. As dez médias da divisão do valor homologado pelo estimado com maior ocorrência

Aqui, são identificadas as médias da divisão do valor homologado pelo estimado que mais ocorreram nas compras pertencentes ao *outlier* (Figura 9). Este gráfico indica que na maioria dos casos o valor estimado foi exatamente igual ao homologado o que pode ser um forte indício de “combinação” de preço e de ausência de concorrência. Um outro ponto interessante é

que houveram quinze compras com o valor homologado maior que o estimado, o que reforça o indício de irregularidade dos pregões pertencentes a este grupo.

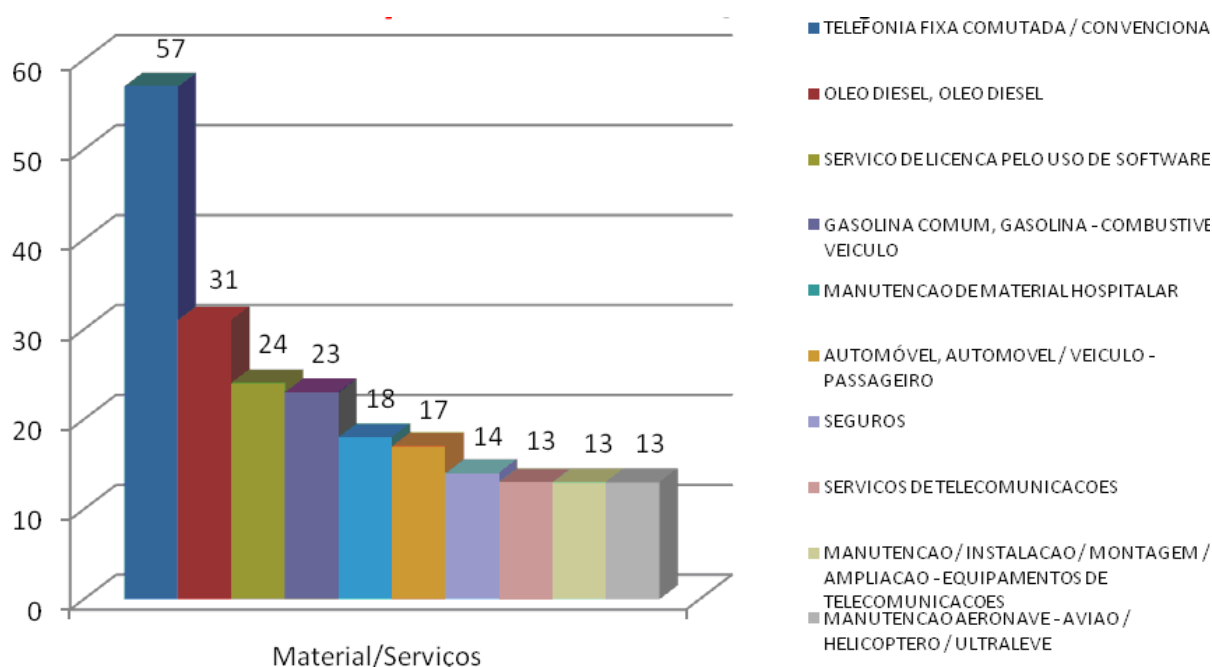
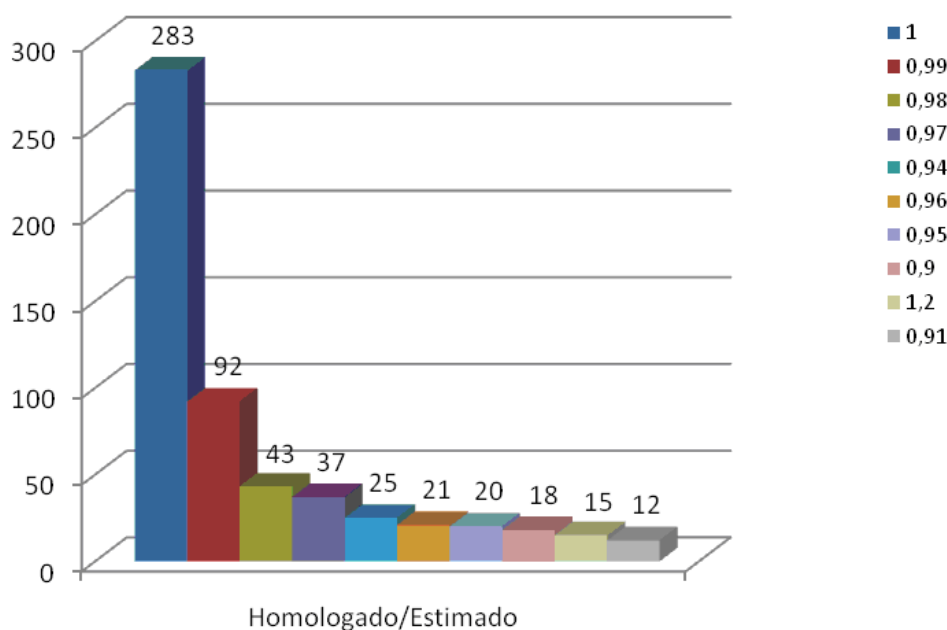
Após analisar todos os gráficos gerados a partir de dados das compras pertencentes ao *outlier* fica clara a afirmação de que o grupo foi formado por pregões eletrônicos de baixa concorrência.

5.6. Aplicação

A aplicação, apesar de ser uma das fases do CRISP-DM, está fora do escopo deste projeto, uma vez que o controle deve ser exercido por órgãos reguladores do governo federal brasileiro.

De qualquer forma, vale rever as principais experiências obtidas no decorrer do projeto. Estas experiências são descritas abaixo conforme as etapas da MD:

- **Definição do escopo do projeto:** inicialmente, não havia um cliente/especialista disponível para definição clara do escopo do projeto. Houve uma certa demora para essa definição e diversos temas foram pesquisados até o fechamento do escopo. Essa foi a tarefa mais demorada e que consumiu a maior parte do tempo.
- **Entendimento do negócio:** a falta de um cliente/especialista disponível ainda dificultou o esclarecimento de eventuais dúvidas surgidas no decorrer da pesquisa. Para ter um entendimento claro do negócio foi necessário recorrer a diversas fontes alternativas de informação como internet, livros e artigos.
- **Definição de indicadores:** a metodologia proposta pelo Ministério do Planejamento foi muito útil para definição dos indicadores. Este foi um ponto positivo neste projeto e pode servir de base para trabalhos futuros que terão a necessidade de definir indicadores.
- **Base de dados:** a base de dados utilizada para a pesquisa demorou para ser definida e disponibilizada. Após tê-la disponível foi necessário realizar diversas ações para deixá-la viável para exploração. Essa preparação da base foi complexa e demorada e este trabalho não deve ser desperdiçado, pois pode poupar tempo de futuras pesquisas que utilizarão o DW.
- **Ferramenta:** foi necessário algum tempo para aprender a utilizar a ferramenta WEKA.

Figura 8. Os dez materiais/serviços que foram mais adquiridos.**Figura 9.** As dez médias da divisão do valor homologado pelo estimado com maior ocorrência

Devido à escassez de recursos e de tempo não foi possível testar outras ferramentas. A utilização de ferramentas mais robustas poderia ter enriquecido o conteúdo da pesquisa.

- **Alocação de máquina:** o volume de dados da pesquisa e a técnica de mineração de dados exigiram muito recurso tecnológico. O ideal seria que houvessem servidores disponíveis para desenvolvimento do projeto. A UCB tentou adquirir máquinas, mas não foi possível. O notebook utilizado no início do projeto não era avançado tecnologicamente e foi

necessário utilizar outro computador para prosseguir com a pesquisa.

- **Análise dos resultados:** foi possível gerar conhecimento a cada análise de resultado. Se houvesse tempo para a geração de mais resultados poderia ter sido gerado ainda mais conhecimento.

6. Conclusões e trabalhos futuros

Os processos licitatórios dos órgãos do Governo Federal brasileiro, apesar de serem bem

definidos e burocráticos, são sujeitos a irregularidades. Isto causa prejuízos à população. Órgãos, fornecedores e pregoeiros agem em conluio para beneficiar um grupo de pessoas.

Os fraudadores do processo licitatório procuram formular diversas ações para burlar o controle. É constante essa atividade, pois à medida que ações que levam a irregularidades são detectadas, há a necessidade de inovação de comportamento para continuarem burlando a lei.

A mineração de dados já é utilizada para a detecção de fraudes em outros segmentos como compras com cartão de crédito, intrusão computacional e análise de vínculos. Este campo já obteve sucesso nestas áreas, porém não foi muito aplicado em compras governamentais. Este campo tem muito a ser explorado neste contexto e os órgãos controladores do Governo Federal brasileiro podem tornar o trabalho mais efetivo com a utilização de técnicas de mineração de dados.

A definição de indicadores úteis para controlar os processos licitatórios foi algo positivo. O conteúdo descrito neste trabalho pode servir de base para a formulação de uma metodologia de definição de indicadores com foco em compras governamentais ou, até mesmo, fraudes.

Os resultados alcançados com relação aos modelos gerados podem ser utilizados por analistas dos órgãos de controle. Os indicadores, as consultas SQL e o DW deste trabalho podem ser reaproveitados para aplicação em outras pesquisas ou, até mesmo, para continuidade deste projeto.

Apesar de se basear no CRISP-DM, é possível formular uma metodologia de mineração de dados com foco em exploração de indicadores.

Nesta pesquisa foi utilizada apenas uma técnica de mineração de dados. Para obter melhores resultados, o ideal seria utilizar diversas técnicas para obtenção de insumos úteis à medição do grau de eficiência da técnica para o negócio em questão.

A elaboração de um “robô” que alerte o pregoeiro em tempo real se o pregão está sendo competitivo ou não, tendo como base os valores dos indicadores pertencentes ao outlier gerado nesta pesquisa pode ser elaborado e implantado no sistema que auxilia o pregoeiro na condução dos pregões eletrônicos.

Os insumos desta pesquisa são passíveis de investigação e deve haver uma análise aprofundada antes da indicação de qualquer suspeita.

Referências

- Abbott, D., Matkovsky, P., & Elder, J. (1998). An evaluation of high-end data mining tools for fraud detection. *IEEE SMC98 - Systems, Man, and Cybernetics Society*, p. 12-14.
- Apparao, G., Singh, A., Rao, G. S., Bhavani, B. L., Eswar, K., & Rajani, D. (2009). Financial statement fraud detection by data mining. *International Journal of Advanced Networking and Applications*, 1. 159-163.
- Barse, E. L., Kvarnstrom, H., & Jonsson, E. Synthesizing Test Data for Fraud Detection Systems. *19th Annual Computer Security Applications Conference*, p. 384-395, 2003.
- Brasil (1993). *Lei n. 8.666, de 21 de junho de 1993*. Regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da Administração Pública e dá outras providências. Brasília, DF.
- Brasil (2002). *Lei n. 10.520, de 17 de julho de 2002*. Institui, no âmbito da União, Estados, Distrito Federal e Municípios, nos termos do art. 37, inciso XXI, da Constituição Federal, modalidade de licitação denominada pregão, para aquisição de bens e serviços comuns, e dá outras providências. Brasília, DF.
- Brasil (2009). *Melhoria da gestão pública por meio da definição de um guia referencial para medição do desempenho da gestão, e controle para o gerenciamento dos indicadores de eficiência, eficácia e de resultados do programa nacional de gestão pública e desburocratização*, Brasília: Ministério do Planejamento.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*, SPSS.
- Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., Chau, M. (2004). Crime data mining: a general framework and some examples. *Computer*, 37(4), 50-56.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1-38.
- Fawcett, T. (1998). *AI approaches to fraud detection and risk management*. Papers from the 1997 AAAI Workshop. Technical Report WS-97-07. AAAI Press.
- Kou, Y., Lu, C., Sinvongwattana, S., & Huang, Y. (2004). Survey of Fraud Detection Techniques. *International Conference on Networking, Sensing & Control*, Taipei, Taiwan.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97.

Phua, C., Lee, V., Smith, K., & Gayler, R. (2005). A comprehensive survey of Data Mining-based fraud detection research. <https://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf>

Weatherford, M. (2002). Mining for Fraud. *IEEE Intelligent Systems*, 17(4), 4-6.

Sobre os autores

Ricardo Akl Lasmar de Alvarenga

Graduado em Computação pelo Centro Universitário de Brasília - UniCEUB (2007) e Mestre em Gestão do Conhecimento e Tecnologia da Informação pela Universidade Católica de Brasília - UCB (2010). Atua em auditoria, gestão e TI. Atualmente é Auditor Federal de Controle Externo no TCU, lotado na Secretaria de Fiscalização de Tecnologia da Informação (Sefti). É professor das disciplinas de Ciência de Dados e de Inteligência Artificial no UniCEUB. É palestrante nas áreas de auditoria, governança e gestão, licitações e contratos, e TI. Prêmio Reconhe-Ser 2018 por ter coordenado auditoria destaque no TCU, que realizou análise integrada de dados governamentais para detecção de fraudes e de irregularidades.

Remis Balaniuk

Tecnólogo em Processamento de Dados pela Universidade de Brasília (1986), Mestre em Ciências da Computação pela Universidade Federal do Rio Grande do Sul (1989), Doutor em Informática pelo Institut National Polytechnique de Grenoble - França (1996), com pós-doutorado pelo Institut National pour la Recherche en Informatique et Automatique - INRIA, França (2000 e 2009). Foi pesquisador associado da Universidade de Stanford, EUA e pesquisador visitante da Universidade de Oxford, no Reino Unido. Atualmente é professor e pesquisador da Universidade Católica de Brasília, auditor do Tribunal de Contas da União, onde é especialista sênior na detecção de fraudes e avaliação de políticas públicas utilizando Mineração de Dados, Geoprocessamento e *Business Intelligence*, e coordena a Especialização em Análise de Dados para o Controle. Na pesquisa atua principalmente na área de Ciência da Computação, com ênfase em métodos de aprendizagem de máquina e aplicações de Inteligência Artificial à Mineração de Dados e ao geoprocessamento.

Hércules Antônio do Prado

Graduado em Processamento de Dados pela Universidade Federal de São Carlos (1976), Mestre em Engenharia de Sistemas e Computação pela COPPE - Universidade Federal do Rio de Janeiro (1989) e Doutor em Ciências da Computação pela Universidade Federal do Rio Grande do Sul (2001). Foi professor visitante da School of Computing and Information da Universidade de Pittsburgh, EUA em 1999. É professor da Universidade Católica de Brasília, onde atua no Mestrado em Gestão do Conhecimento e Tecnologia da Informação (MGCTI). Seus interesses incluem Inteligência Artificial, Gestão do Conhecimento, Gestão Organizacional, Planejamento Estratégico e Aprendizagem Organizacional.

Edilson Fereda

Graduado em Tecnologia de Computação pelo Instituto Tecnológico de Aeronáutica (1979), Mestre em Sistemas e Computação pela Universidade Federal da Paraíba (1988) e Doutor em Ciência da Computação pelo *Laboratoire d'Informatique, Robotique et de Microélectronique de Montpellier* - LIRMM/CNRS, França (1992). Entre 1986 e 2000, foi professor do Departamento de Sistemas e Computação da UFPB (atual Universidade Federal de Campina Grande), tendo atuado nos cursos de Bacharelado em Ciência da Computação, Mestrado em Informática e Doutorado em Engenharia Elétrica. Desde 2001 é professor titular da Universidade Católica de Brasília, onde atua no Curso de Bacharelado em Ciência da Computação e no Mestrado em Gestão, Tecnologia e Inovação (antigo Mestrado em Gestão do Conhecimento e Tecnologia da Informação). Seus interesses incluem Inteligência Artificial e Gestão do Conhecimento.