

A UTILIZAÇÃO DE MINERAÇÃO DE DADOS NO CCIMAR COMO AUXÍLIO AO PROCESSO DE AUDITORIA NAS LICITAÇÕES DA MARINHA DO BRASIL.

Ewerton Ferreira Gonçalves

Comando do 3º Distrito Naval

Rua Coronel Flamínio, 100, Santos Reis, Natal/RN ewerton@marinha.mil.br

George Hamilton Andrade Costa

Centro de Instrução Almirante Newton Braga Av. Brasil, 10.500, Penha, Rio de Janeiro/RJ
georgehamiltoncosta@gmail.com

RESUMO

O objetivo deste trabalho é verificar a viabilidade da utilização de técnicas de Mineração de Dados nas auditorias conduzidas pelo Centro de Controle Interno da Marinha (CCIMAR) na área de licitações. Para tal, foi necessário conceituar Mineração de Dados e seus termos associados, bem como sua conexão com auditoria contínua e licitações para, em seguida, propor um ambiente próprio e adequado ao uso de Mineração de Dados, a fim de utilizar esta ferramenta em prol da organização como um apoio adicional às auditorias em licitações da Marinha.

Palavra-chave: Auditoria; *Data Warehouse*; *KDD*; Licitações; Mineração de Dados.

ABSTRACT

This study intends to verify the viability of using Data Mining techniques in the Navy bid audits conducted by the Brazilian Navy Internal Control Center (CCIMAR). In order to achieve that goal, it was necessary to conceptualize Data Mining and its associated terms, as well as its connection with continuous auditing and bidding, in order to propose a suitable environment for Data Mining, as an additional support for the Brazilian Navy bid audits.

Keywords: Auditing; Data Warehouse; KDD; Government Biddings; Data Mining.

Como Citar:

GONÇALVES, Ewerton; COSTA, Hamilton. A Utilização de Mineração de Dados no CCIMAR como auxílio ao processo de Auditoria nas Licitações da Marinha do Brasil. *In: SIMPÓSIO DE PESQUISA OPERACIONAL E LOGÍSTICA DA MARINHA*, 19., 2019, Rio de Janeiro, RJ. *Anais* [...]. Rio de Janeiro: Centro de Análises de Sistemas Navais, 2019.

1. INTRODUÇÃO

Dentre vários direitos conquistados pelo cidadão brasileiro por meio da atual Constituição Federal [6], destaca-se o incentivo à participação popular no processo de elaboração e discussão das leis orçamentárias e orçamentos, por meio de instrumentos de fortalecimento da participação do cidadão.

Um desses instrumentos é a Lei de Responsabilidade Fiscal [7], que dentre outras disposições, dedica uma Seção específica à transparência no gasto público. Em decorrência disso, a maioria dos processos governamentais são suportados por sistemas informatizados, que mantém detalhados registros sobre a execução dos recursos públicos.

Nesse contexto, o Centro de Controle Interno da Marinha (CCIMAR), como um órgão de controle interno¹, é responsável por fiscalizar os diversos atos da Marinha do Brasil (MB) perante a sociedade. Para esse fim, o órgão utiliza a auditoria como instrumento de análise da “legalidade e a legitimidade dos atos de gestão dos responsáveis pela aplicação de recursos públicos, quanto aos aspectos contábil, financeiro, orçamentário e patrimonial” [8].

De acordo com Vasarhelyi e Halper [40], os avanços da tecnologia da informação tornam viável a prática da auditoria contínua, que segundo Alles et al. [1], estendem os métodos analíticos da auditoria tradicional ao examinar fluxos contínuos de dados, a fim de estabelecer expectativas para o seu conteúdo. Este tipo de auditoria, ao contrário do modelo tradicional que normalmente baseia-se em dados amostrais, ocorre em tempo real e em toda

¹ O Controle interno compreende as atividades de fiscalização contábil, financeira, orçamentária, patrimonial, administrativa e operacional das unidades dos poderes da União. É exercido internamente por órgão de controle inerente ao âmbito das organizações e tem por finalidade, além das atinentes às atividades já descritas, apoiar o Controle Externo no exercício de sua missão [13].

uma população de registros. Nesse contexto, aumenta-se a relevância sobre uma ferramenta de auditoria contínua chamada Mineração de Dados (MD) ou *Data Mining* (DM).

A Mineração de Dados, segundo Fayyad, Piatetsky-Shapiro e Smyth [16] é uma etapa de um processo conhecido como Descoberta de Conhecimento em Bases de Dados ou *Knowledge Discovery in Databases* (KDD). Trata-se da aplicação de uma série de técnicas e algoritmos para análise e descoberta de conhecimento ou para produção de padrões ou modelos que podem produzir conhecimento inteligível. Como forma de nortear a pesquisa, este estudo baseou-se na seguinte pergunta: é possível utilizar Mineração de Dados nas auditorias de licitações do CCIMAR?

Nesse sentido, o objetivo principal do trabalho é estudar a aplicação da ferramenta Mineração de Dados nas atividades de auditoria de licitações do CCIMAR, por meio de uma pesquisa descritiva, bibliográfica e documental.

A proposta do trabalho será o levantamento dos aspectos necessários à criação de um ambiente de MD de licitações para uso do Centro. Dessa forma, o estudo ganha relevância ao relacionar a Mineração de Dados como um passo em direção ao modelo de auditoria contínua, visando uma oportunidade de melhoria nos atuais processos de auditoria daquele órgão.

Para atendimento desse fim, os seguintes objetivos específicos foram traçados: (i) identificar os requisitos necessários para utilização da mineração de dados; (ii) identificar se essa ferramenta pode ser utilizada na área de auditoria de licitações; e (iii) verificar a capacidade do CCIMAR em aplicar essa ferramenta em seus processos de auditoria.

Este artigo é organizado em seis seções. Esta dedica-se a uma breve contextualização e apresentação do problema de pesquisa bem como os objetivos a serem alcançados de forma a responder à pergunta do autor. Na segunda, será apresentada a base teórica sob a qual se reveste este estudo. Na seção seguinte, será exposta a metodologia da qual se serviu o autor para conduzir seu estudo e atingir uma conclusão. Na quarta seção, será realizada uma breve análise, bem como a apresentação da proposta de solução. As considerações finais, limitações e sugestões para estudos futuros encontram-se na quinta seção. A sexta e última seção destina-se a identificar o referencial teórico utilizado.

2. REFERENCIAL TEÓRICO

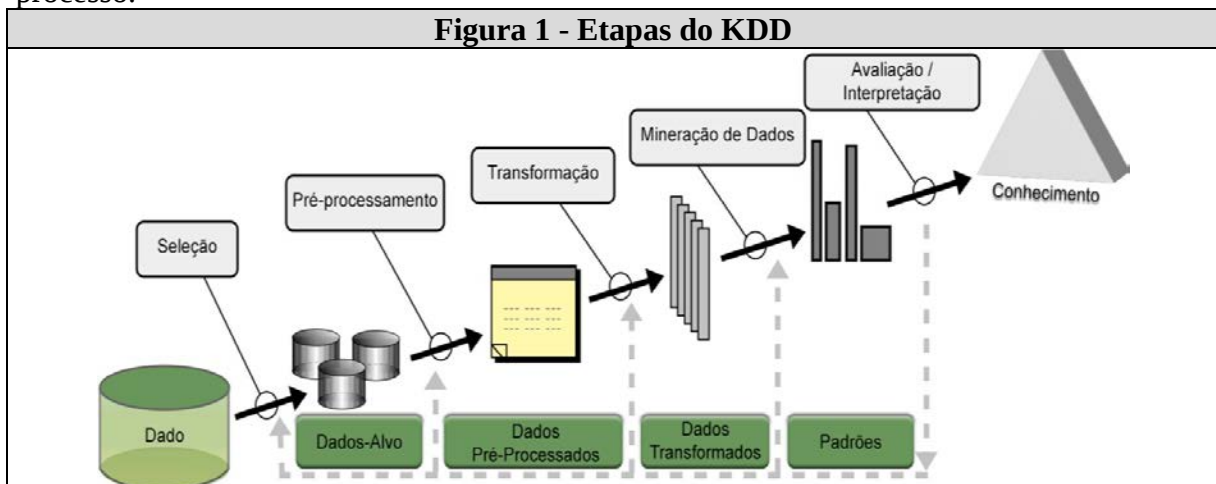
2.1. DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS (DCBD)

Segundo Goldschmidt, Passos e Bezerra [21], os avanços na área de Tecnologia da Informação (TI) resultam no armazenamento de grandes volumes de dados, que podem

apresentar tendências, correlações, ou padrões com potencial para melhorar os processos de uma organização.

Por sua vez, a análise de todo esse montante de dados torna-se inviável sem a utilização de instrumentos computacionais que auxiliem suas interpretações e correlações. A fim de preencher essa lacuna, uma área conhecida como KDD ou *Descoberta de Conhecimento em Bases de Dados (DCBD)* concentra o interesse de diversos setores, como Pesquisa e Desenvolvimento, Marketing, Indústria, Bancos e Governo.

De acordo com Piatetsky-Shapiro [31] esse termo foi concebido em 1989 e faz referência a um conceito amplo de extrair conhecimento a partir de uma base de dados. Nesse sentido, Fayyad, Piatetsky-Shapiro e Smyth [19] ampliam esse conceito, definindo-o como um processo não trivial de identificação de novos padrões válidos, úteis e compreensíveis, incluindo a Mineração de Dados como uma das várias etapas deste processo.



Fonte: Adaptado de: Fayyad, Piatetsky-Shapiro e Smyth (1996).

Complementando a ilustração da Figura 1, Han, Pei e Kamber [23] explicam as etapas do KDD da seguinte forma:

- a) Seleção e limpeza: etapa na qual ocorre a remoção de dados irrelevantes;
- b) Pré-processamento: etapa na qual múltiplas fontes de dados podem ser combinadas para posterior transformação;
- c) Transformação: etapa em que os dados são transformados ou consolidados nas formas apropriadas para mineração, de acordo com a técnica a ser utilizada;
- d) **Mineração de Dados:** um processo essencial no qual métodos inteligentes

são aplicados no intuito de extrair padrões a partir dos dados;

e) Avaliação dos padrões: consiste na identificação dos padrões obtidos que foram realmente relevantes, dadas as medidas ou funções de avaliação; e

f) Interpretação e avaliação do conhecimento: fase final na qual técnicas de representação e visualização são utilizadas para apresentar o conhecimento minerado ao usuário.

Com base nesse modelo, pressupõe-se que para que seja possível a utilização de MD na geração de conhecimento, devem existir dados à disposição da organização que atendam às suas necessidades. Dessa forma, serão apresentados os conceitos de Banco de Dados relevantes ao entendimento do KDD e da Mineração de Dados.

2.1.1. Banco de Dados, Data Warehouse e Data Mart

De acordo com Elmasri e Navathe [18], um Banco de Dados (BD) é uma coleção de dados relacionados que podem ser registrados e possuem um significado implícito. Nesse contexto, conforme destacam Silberschatz, Korth e Sudarshan [37], as organizações, ao longo do tempo, tendem a gerar grandes volumes de BD, contendo inúmeros tipos de registros como a carteira de clientes da organização, as entradas e saídas de estoque, os históricos de vendas e o cadastro de seus fornecedores.

Essa coleção de registros pode ser agregada em um Armazém de Dados ou *Data Warehouse (DW)*, que Inmon [25] define como uma “coleção de dados orientada a assunto, integrada, não volátil e variável em relação ao tempo, de forma a apoiar as decisões gerenciais.” Assim, os DW podem permitir o acesso a dados para análise complexa, descoberta de conhecimento e tomada de decisão.

De acordo com a necessidade da organização, os DW podem ser substituídos por um ou mais *Data Marts (DM)*, que são *Data Warehouses* setoriais ou por assunto, e que apesar de possuir uma capacidade menor, contém as mesmas características de um DW normal. São utilizados para atender a um departamento ou uma unidade organizacional, a um custo de implementação e tempo de resposta menores. Dessa forma, os DM podem ser usados para um projeto limitado a uma área de negócio, até mesmo como oportunidade de aprendizado, visando futura integração em um projeto global [32].

Adicionalmente, Shams e Farishta [36] resumem a arquitetura de funcionamento de um DW a um processo no qual os dados são extraídos inicialmente de fontes externas ou de seus sistemas operacionais, com posterior transformação ou integração ao modelo preexistente, para depois serem carregados no *Data Warehouse*.

Kimball [26] complementa essa definição ao tratar o DW como um banco de dados que exerce Processamento Analítico de Dados (*Online Analytical Processing* – OLAP), cujas informações não são atualizadas tempestivamente, mas sim armazenadas como um registro histórico, temporal, e por isso os dados adquirem as características de serem somente de leitura e não-voláteis². As ferramentas OLAP, diferentemente dos BD transacionais, permitem a realização de consultas gerenciais e apresentação dos dados com maior flexibilidade para o usuário.

2.1.2. Bancos de Dados Tradicionais x Data Warehouse

Segundo Correia [17], os Bancos de Dados Transacionais (*Online Transaction Processing* – OLTP) são responsáveis por inserir, apagar ou atualizar as informações necessárias para o dia a dia de uma organização. Eles são utilizados para registrar e executar operações pré-definidas, e seus dados podem sofrer constantes alterações. Exemplos desses BD transacionais são o SIASG³, SINGRA⁴, QUAESTOR⁵ e SISPAG2⁶.

De acordo com Hokama et al. [24], devido ao fato dos Bancos de Dados Transacionais possuírem estruturas normalizadas, ou seja, não-redundantes, e por realizarem muitas transações de inclusões, alterações e exclusões de registros de dados em suas tabelas, esses sistemas não são os ideais e recomendados para consultas gerenciais ou tomadas de decisões, uma vez que para essas consultas, as modelagens e implementações dos sistemas irão exigir muito consumo de processamento dos servidores.

Em contraponto, um DW utiliza uma estrutura desnormalizada, ou seja, com registros muitas vezes redundantes, armazenando dados provenientes dos OLTP, que não podem mais ser mais alterados ou excluídos, facilitando as consultas gerenciais ou analíticas, através de agregações, sumarizações e outras operações mais complexas, sendo assim ideais para Mineração de Dados e suporte à tomada de decisões [24]. Um DW armazena

² Armazenamento não-volátil é aquele cujos dados são conservados ao se retirar a fonte de energia.

³ O SIASG - Sistema Integrado de Administração de Serviços Gerais - é um sistema do Governo que mantém registro das compras e contratações firmadas pelo poder executivo da Administração Pública Federal.

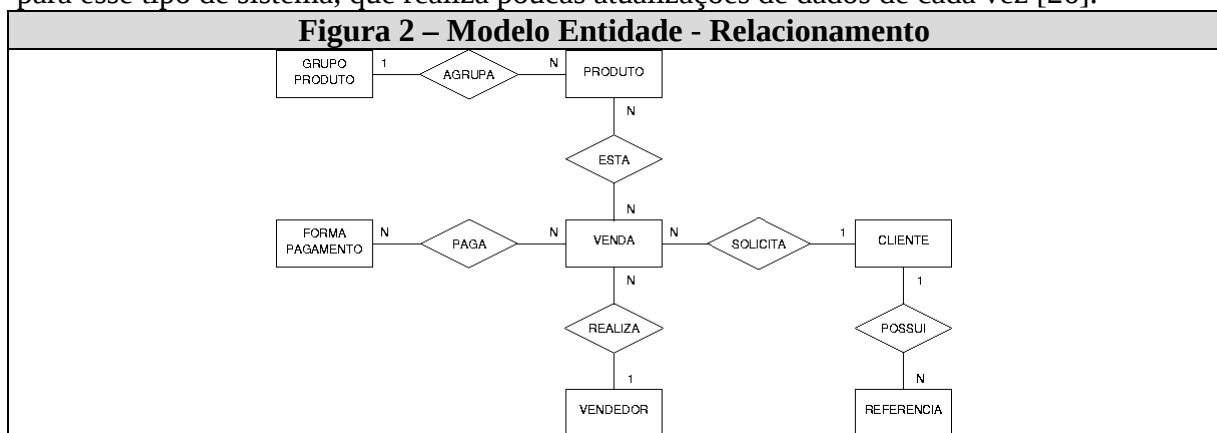
⁴ O SINGRA - Sistema de Informações Gerenciais de Abastecimento, é um software da Marinha do Brasil que além de outras funcionalidades, é responsável por efetuar a gestão de estoques e a previsão de demanda de sobressalentes da Força.

⁵ O QUAESTOR é um software de gerenciamento de estoques e recursos relacionados à alimentação dos servidores da Marinha do Brasil.

⁶ O SISPAG2 - Sistema de Pagamento da MB, é o software da Marinha do Brasil responsável pelo processamento da folha de pagamento de seus servidores.

informações históricas de muitos anos e, por isso, deve possuir uma grande capacidade de processamento e de armazenamento.

É relevante mencionar que existe diferença entre os modelos de dados utilizados nos dois sistemas, que impactam na capacidade de realizar MD. O modelo entidade-relacionamento (figura 2) aplica-se perfeitamente aos OLTP, pois o número reduzido de redundâncias aumenta a velocidade de processamento das informações, o que é desejável para esse tipo de sistema, que realiza poucas atualizações de dados de cada vez [26].



Fonte: Ataides [4].

Em contrapartida, ainda de acordo com Kimball [26], os sistemas OLAP, devido à sua natureza de processar grandes volumes de registros simultaneamente, requerem um modelo dimensional, no qual os dados são representados como um cubo de várias dimensões, com capacidade de ser cortado em quaisquer dessas dimensões, de forma a visualizar as informações de forma mais simples e flexível.



Fonte: Adaptado de: Kimball [26].

Esse modelo, conhecido como cubo OLAP ou modelo estrela (*star schema*), combina uma ou mais tabelas centrais em que os dados históricos do negócio (tabelas de fatos), são ligados a tabelas descritivas (ou tabelas de dimensões). A figura 3 facilita essa compreensão, mostrando as diferentes representações de um modelo dimensional.

Ressaltada a diferença entre modelos, Han, Pei e Kamber [23] explicam que a Mineração de Dados deve ser realizada com o uso de ferramentas OLAP em um DW, cujos dados devem ser provenientes, dentre outras fontes, das informações extraídas e transformadas de um sistema OLTP.



Fonte: Autor. Adaptado de: Shams e Farishta [36] e Han, Pei e Kamber [23].

2.2. MINERAÇÃO DE DADOS

Dentre as diferentes definições existentes de MD, destacam-se as de Han, Pei e Kamber [23], que a conceituam como o processo de descoberta de padrões interessantes e conhecimento a partir de grandes quantidades de dados, Goldschmidt, Passos e Bezerra [21], que de forma breve definem MD como a principal etapa do processo de KDD, na qual ocorre a busca efetiva por conhecimentos novos e úteis a partir dos dados, e por fim, Fayyad, Piatetsky-Shapiro e Smyth [19] que afirmam que Mineração de Dados é etapa do KDD que consiste na aplicação de algoritmos específicos para extração de padrões de dados.

Portanto, existe um consenso ao definir a Mineração de Dados como a busca efetiva de conhecimento útil nas Bases de Dados, e de acordo com Han, Pei e Kamber [23], essas fontes de dados podem incluir Bancos de Dados, DW/DM, a Web, ou outros repositórios de informação, após a devida transformação em um modelo dimensional.

Apesar das definições nos levar a crer que todo o processo é realizado automaticamente, sabe-se hoje que, de fato, isso não é verdade [27]. Embora existam diversas ferramentas que nos auxiliam na execução dos algoritmos de mineração, os resultados ainda carecem de análise humana. Porém, ainda assim, a MD contribui de forma significativa no processo de descoberta de conhecimento, permitindo aos especialistas concentrarem esforços apenas em partes mais significativas dos dados disponíveis.

2.2.1. Tarefas de Mineração de Dados

De acordo com Amo [3], as tarefas de MD consistem na especificação do que se pretende fazer com os dados, ou seja, na definição do objetivo do processo em si. Larose [27], classifica essas tarefas da seguinte forma:

a) **Descrição:** É a tarefa utilizada para descrever tendências ou padrões revelados pelos dados. Geralmente oferece uma possível interpretação, como a influência de certas variáveis no resultado obtido. Um exemplo prático é o posicionamento de determinados itens nas prateleiras de supermercado, aproximando produtos com as mesmas tendências de consumo, baseando-se na análise dos padrões de consumo dos clientes;

b) **Classificação:** a classificação visa identificar a qual classe ou categoria um determinado registro pertence. Nesta tarefa, o modelo analisa uma base de dados inicial, com cada registro já contendo a indicação à qual classe pertence, a fim de “descobrir a regra” de classificação de um novo registro. Dessa forma, os futuros registros podem, após análise, ser automaticamente enquadrados em suas classes. Essa tarefa pode ser aplicada, por exemplo, na classificação do porte de uma empresa de acordo com seu número de empregados ou seu faturamento bruto anual, a taxa de juros para empréstimos bancários ou precificação de seguros conforme a análise de risco do cliente, ou quando uma transação em cartão de crédito pode ser considerada fraudulenta;

c) **Estimação (ou Regressão):** A estimação é similar à classificação, porém, é usada quando o registro é identificado por um valor numérico e não um categórico. Consiste no estabelecimento de uma função que mais se ajusta ao comportamento de variáveis dependentes. Assim, pode-se prever um valor futuro a partir de dados históricos de uma base de dados. Por exemplo, pode-se estimar aumento da demanda por um determinado produto em função dos gastos com marketing, ou apontar a necessidade de empréstimo em função da renda e dívidas acumuladas [19];

d) **Predição:** tarefa semelhante às anteriores, porém, visa descobrir o valor futuro de um determinado atributo, como o preço estimado de uma determinada ação em cinco meses ou o possível vencedor de uma licitação de acordo com as estatísticas dos participantes; e

e) **Clusterização:** visa identificar e aproximar os registros similares, porém, diferentemente da classificação, os registros não necessitam estar previamente classificados. Portanto, apenas identifica os grupos de dados similares. Um cluster é uma coleção de registros semelhantes entre si, mas diferentes dos demais clusters. Essa análise permite verificar a existência de diferentes grupos dentro de um determinado conjunto de dados, e

em caso de sua existência, determinar quais são eles [35]. Possui aplicações bastantes visíveis em Marketing, como agrupar compradores de comportamento similar, de forma a direcionar as futuras vendas a esses clientes. Silva [38] utiliza-se dessa técnica para mostrar indícios de formação de cartéis em licitações.

Essas técnicas podem ser utilizadas individualmente ou de forma combinada, dependendo da natureza do negócio. De acordo com a tarefa a ser realizada, diferentes tipos de algoritmos, técnicas ou métodos podem ser utilizados, como Redes Neurais, Algoritmos Genéticos, Árvore de Decisão, Lógica *Fuzzy*, Método *k-means*, dentre outros, que não serão abordados neste estudo por não fazerem parte de seu escopo, pois a definição das técnicas e métodos a ser utilizados é fruto do trabalho conjunto entre Analistas de Dados, profissionais que possuem habilidade de coletar, processar e aplicar algoritmos estatísticos aos dados estruturados, a fim de elaborar análises que interpretam os registros disponíveis [28;34], e analistas do negócio, representados neste trabalho pelos auditores do CCIMAR.

Estabelecidos os principais conceitos necessários ao entendimento geral sobre a Mineração de Dados, será abordada sua contextualização na auditoria dos órgãos governamentais.

2.3. AUDITORIA GOVERNAMENTAL

De acordo com as Normas Internacionais das Entidades Fiscalizadoras Superiores [30], a auditoria governamental pode ser definida como “um processo sistemático de obter e avaliar objetivamente evidência para determinar se a informação ou as condições reais de um objeto estão de acordo com critérios estabelecidos”. Ainda conforme essa norma, a auditoria pode ser classificada em três tipos principais: Financeira, Operacional e Auditoria de Conformidade.

A auditoria financeira foca em determinar se a informação financeira de uma entidade é apresentada em conformidade com a estrutura de relatório financeiro e o marco regulatório aplicável.

A auditoria operacional foca em determinar se intervenções, programas e instituições estão operando em conformidade com os princípios de economicidade, eficiência e efetividade, visando responder a questões-chave de auditoria e apresentar recomendações para aperfeiçoamento.

Auditoria de conformidade foca em determinar se um particular objeto está em conformidade com normas estabelecidas. A auditoria de conformidade é realizada para avaliar se as regras, leis, regulamentos, resoluções orçamentárias, políticas, códigos estabelecidos, acordos ou os princípios gerais são cumpridos agentes públicos. [30]

Esse tipo de abordagem, de acordo com Chan e Vasarhelyi [16], é típico do modelo de auditoria tradicional, que vem perdendo espaço no atual contexto mundial devido à dificuldade em prover informação em tempo real, uma vez que a maioria dos processos de auditoria são intensivos em tempo e mão de obra.

Ainda segundo os mesmos autores, todo esse esforço e tempo podem ser reduzidos com o uso da tecnologia e automação. Esse tipo de ferramenta tecnológica, aliada ao processo de auditoria, é conhecida como Auditoria Contínua, a qual utiliza-se de técnicas de Mineração de Dados de forma a analisar o comportamento de toda uma população, tornando o processo de auditoria mais eficiente em contraste com o modelo tradicional, cujas diferenças principais são apresentadas no quadro 1:

Quadro 1 - Comparativo entre os modelos de auditoria Tradicional e Contínua.		
	Auditoria Tradicional	Auditoria Contínua
Frequência	Periódica	Contínua ou mais frequente
Abordagem	Reativa	Proativa
Procedimentos	Manuais	Automatizados
Natureza dos testes	Amostragem	População Inteira
Base dos testes	Baseado no julgamento do auditor	Baseado em análise de dados
Relatórios	Periódicos	Contínuos

Fonte: Adaptado de: Chan e Vasarhelyi [16].

Conforme a definição de Yu, Yu e Chou (2000, *apud* SILVA; ALMEIDA JR.) [39]:

O processo de Auditoria Contínua compreende as funções do processo de auditoria periódica, usando um sistema de supervisão de transações em tempo real, com software que monitora continuamente as transações e compara as suas características com os resultados esperados (p. 10).

Um benefício potencial da auditoria contínua é que sua prática permite superar as limitações de recursos que normalmente levam as organizações a adotar controles a posteriori e por amostragem, mesmo nos casos em que seria pertinente realizar um controle concomitante de todas as transações [29].

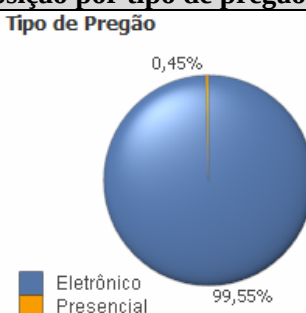
2.3.1. Licitações e Pregão Eletrônico

Um dos aspectos intrínsecos à atividade governamental é a subordinação das compras públicas aos ditames da Lei nº 8.666 [14]. Tal normativo dispõe sobre normas gerais de licitações e contratos e prevê, em seu artigo 22, cinco modalidades de certames, a

saber: concorrência, tomada de preços, convite, concurso e leilão.

Em paralelo, a lei nº 10.520 [15] instituiu uma modalidade chamada Pregão, em que a disputa por fornecimento de bens ou serviços comuns é feita através de um sistema de lances, presenciais ou eletrônicos. No ano de 2005, a expedição do Decreto nº 5.450 [12] regulamentou a forma eletrônica do pregão, cujo ambiente operacional é o Portal de Compras do Governo Federal (COMPRASNET). No ano de 2017, essa modalidade correspondeu a mais de 99,5% do total de pregões realizados.

Tabela 1 - Valores das compras por modalidade licitatória e composição por tipo de pregão.		
Modalidade	Valor das Compras (R\$)	Percentual
Pregão	18.177.009.022,38	87,91 %
Concorrência	2.352.425.357,46	11,38 %
Tomada de Preços	126.181.117,83	0,61 %
Concorrência Internacional	16.328.736,71	0,08 %
Convite	4.099.378,04	0,02 %
Concurso	1.065.163,00	0,01 %
Total	20.677.108.775,42	100,00 %



Fonte: Painel de Compras do Governo Federal⁷(2017).

De acordo com a tabela 1, pode-se pressupor que o acompanhamento das licitações da modalidade pregão eletrônico pelos órgãos de auditoria é fator de relevância, pois: (i) essa modalidade concentra o maior percentual de gastos realizados pelo governo; e (ii) é possível utilizar técnicas de MD para as verificações de todos os pregões eletrônicos conduzidos por um órgão governamental, uma vez que as bases de dados do COMPRASNET são disponíveis para extração, transformação e carregamento em um DM dotado de ferramentas OLAP.

2.4. OBRAS RELACIONADAS

No Brasil, diversos estudos e trabalhos acadêmicos sobre Mineração de Dados são publicados, no entanto, uma quantidade reduzida se concentra nas atividades governamentais ou de auditoria. Nesse contexto, Guimarães [22] abordou o uso de ferramentas de MD aplicadas a BD de instituições públicas permanentes, utilizando-as em uma base de dados do Ministério Público de Rondônia, demonstrando a existência de conhecimento útil que

⁷ Dados referentes ao ano de 2017, sem considerar afastamentos de licitação. Disponível em: www.paineldecompras.planejamento.gov.br. Acesso em 15/10/2018.

pagariam o retorno do investimento na utilização de ferramentas de MD integradas aos seus sistemas de tomada de decisão.

No mesmo sentido, Balaniuk [5] apresenta uma metodologia de MD adaptada à realidade do Tribunal de Contas da União, enfatizando a necessidade de capacitação do Corpo Técnico daquele Tribunal em áreas de conhecimento relacionadas à TI.

Utilizando metodologia similar, Silva [38] desenvolveu um trabalho propondo o uso de MD no domínio da auditoria governamental. Estabelecendo diversas regras de associação, de forma a detectar indícios de formação de cartéis em licitações, concluiu que o seu uso é eficaz, uma vez que, de posse desses indícios, os órgãos de controle podem desenvolver mecanismos específicos nas regiões onde o risco de ocorrência desse tipo de irregularidade for maior.

Por fim, Alvarenga [2], fazendo uso da ferramenta de mineração WEKA⁸ em dados coletados do DW do Ministério do Planejamento, Desenvolvimento e Gestão, responsável por armazenar informações de compras efetuadas pelos órgãos do Governo Federal, utilizando-se de determinados indicadores, buscou aferir o grau de competitividade dos pregões eletrônicos, conseguindo caracterizar grupos de pregões com diferentes graus de competitividade.

3. METODOLOGIA

3.1. TIPO DE PESQUISA

De acordo com Gil [20], as pesquisas podem ser classificadas, com base nos objetivos e procedimentos técnicos utilizados, em três grandes grupos: exploratórias, descritivas e explicativas. Este trabalho consiste em uma pesquisa descritiva, pois foi desenvolvida com o objetivo primordial de relacionar os principais aspectos e elementos necessários à compreensão do tema Mineração de Dados e sua aplicação em auditoria.

Ainda segundo Gil [20], e de acordo com os meios utilizados para investigação, este estudo emprega uma pesquisa bibliográfica e documental, utilizando-se de livros, artigos de revistas, leis, dissertações, normas internas da Marinha, e outras fontes secundárias de pesquisa.

Também foi realizada, de acordo com Prodanov e Freitas [33], uma pesquisa de

⁸ O WEKA é uma coleção de algoritmos de Data Mining escritos pela Universidade de Waikato, Nova Zelândia.

campo, que serviu como passo para verificar em que ponto se encontra o atual nível de conhecimento sobre o objeto de estudo.

3.2. COLETA E TRATAMENTO DOS DADOS

Na análise dos dados coletados, dentro de uma abordagem qualitativa, na qual, segundo Prodanov e Freitas [33], não há preocupação em comprovar hipóteses previamente estabelecidas, mas sem eliminar a existência de um quadro teórico que direcione a coleta, a análise e a interpretação dos dados, foram examinados artigos e livros voltados aos temas: Descoberta de Conhecimento, Bancos de Dados, Mineração de Dados, Auditoria contínua e Auditoria Governamental.

No mesmo sentido, foram realizadas visitas às instalações do CCIMAR e entrevistas não estruturadas com auditores, de forma a verificar as possibilidades de aplicação da teoria em um caso prático. Após levantar a base necessária de dados e informações relevantes ao estudo, este artigo propôs um ambiente de Mineração de Dados que resultará em um passo na direção do modelo de Auditoria Contínua, resultando em ganhos de eficiência no processo, cuja mensuração não está contemplada no escopo do trabalho.

4. APRESENTAÇÃO DOS RESULTADOS

4.1. ESTUDO DE CASO DO CCIMAR

O CCIMAR é o órgão responsável pelo controle interno no âmbito da Marinha do Brasil. Segundo as Normas sobre Auditoria, Análise e Apresentação de Contas na Marinha (SGM-601), o Centro realiza, de modo integrado, as Auditorias de Acompanhamento, Especial, Operacional, de Atos de Gestão de Pessoal, Conformidade Documental e de Avaliação da Gestão [8], em alinhamento com os tipos de auditoria definidos pela ISSAI [30].

As entrevistas não estruturadas conduzidas com auditores do CCIMAR mostram que o Centro realiza as auditorias de acordo com um Plano Anual de Auditoria, no qual são elencadas as Unidades Jurisdicionadas (UJ) a serem auditadas durante o ano. O resultado dessa verificação é consolidado em um relatório que é encaminhado para o órgão de controle externo (Tribunal de Contas da União - TCU) e para o órgão auditado.

Esse relatório é separado por áreas de auditoria (Execução Financeira e Suprimento de Fundos, Licitações e Acordos Administrativos, Caixa de Economias, Pagamento de Pessoal, Material, Municimento e Gestorias Especiais) e as constatações para cada área

devem ser respondidas pelo órgão auditado, discriminando a providência adotada.

O processo descrito leva a conclusão que, atualmente, a auditoria é realizada *a posteriori*, ou seja, com um certo tempo transcorrido após a ocorrência dos atos praticados pelo órgão auditado. Nesse contexto, a Mineração de Dados pode ser vista como uma oportunidade de melhoria, visto que seu uso para monitoramento dos processos licitatórios, por meio dos dados provenientes dos sistemas de informação, representam uma convergência a um modelo de auditoria contínua, permitindo uma abordagem mais proativa e ampla, já que toda a base de dados é analisada de uma só vez, e não apenas por amostragem [16].

Ainda como resultado das entrevistas, percebeu-se que o CCIMAR se encontra em uma fase embrionária no conhecimento sobre a utilização de técnicas de Mineração de Dados em suas auditorias, com estudos nas áreas de Pagamento de Pessoal e Licitações e Contratos.

Nessa última área, o órgão tem buscado desenvolver técnicas de auditoria contínua, contando com o apoio do professor e pesquisador da Universidade americana de Rutgers, Miklos Vasarhelyi, referência mundial no assunto.

Nesse mesmo sentido, o Centro tem procurado firmar parcerias junto a órgãos que desenvolvem atividades de Mineração de Dados, como: (i) a Controladoria Geral do Município do Rio de Janeiro (CGM-RJ), que possui um laboratório específico para esse fim; e (ii) o TCU, que faz uso de seu agente de Mineração ALICE⁹, capaz de coletar informações do Diário Oficial da União e do COMPRASNET na busca por anomalias em Licitações e Contratos, podendo verificar, por exemplo, se uma licitação está prestes a entregar uma contratação para alguma empresa impedida de contratar junto à administração pública.

Por fim, observa-se que a atual estrutura de TI do CCIMAR, em aspectos materiais e de força de trabalho, é montada para suporte às estações de trabalho, manutenção de rede e sistemas corporativos, não possuindo participação ativa nos atuais processos de auditoria.

4.2. PROPOSTA DE UM AMBIENTE DE MD NO CCIMAR:

Em consonância com o levantamento bibliográfico realizado, e de acordo com as informações obtidas nas entrevistas, o trabalho propõe que o CCIMAR possua um ambiente próprio e adequado para utilização de Mineração de Dados em apoio às auditorias de licitações.

⁹ 9 ALICE é um acrônimo para Análise de Licitações e Editais.

Cabe destacar que a MB normatiza a política de TI por meio da Diretoria de Comunicações e Tecnologia da Informação da Marinha (DCTIM). As Normas de Tecnologia da Informação da Marinha [10] descrevem a utilização e a padronização de sistemas para emprego na força, além de apoiar os processos de homologação de sistemas de TI, definindo a melhor arquitetura e recomendando (ou não) a hospedagem no Centro de Dados da Marinha do Brasil (CD-MB).

O CD-MB é um ambiente, gerenciado pelo Centro de Tecnologia da Informação da Marinha (CTIM), que hospeda os principais sistemas da MB. Esse Centro de Dados foi identificado como ideal para criação de um *Data Mart* como ambiente de MD, pois além de dispor dos requisitos de segurança e confiabilidade necessários, dispensam que o CCIMAR disponha de uma infraestrutura de TI própria para manutenção desse sistema em suas dependências.

O estudo apontou que os pregões eletrônicos podem ser utilizados como ponto de partida, tanto pela sua relevância no percentual de processos de compra, quanto pelo fato de suas Bases e modelos de Dados se encontrarem disponíveis para extração no Portal Brasileiro de Dados Abertos (www.dados.gov.br). A figura 5 mostra um exemplo de modelo de dados do módulo "Licitações" do SIASG.



Fonte: Portal de Dados abertos do Governo Federal.

De acordo com Hokama et al. [24], essa modelagem, típica de um sistema OLTP não é adequada para Mineração de Dados. Dessa forma, após a extração das bases de dados

dos diversos OLTP (SIASG, SIAFI e outros), essas bases devem ser transformadas para um modelo adequado.

Com base nas literaturas analisadas, a modelagem dos dados contidos no DM deve ser realizada no modelo estrela, de forma a permitir a utilização de ferramentas OLAP para Mineração de Dados [26]. Esse trabalho deve ser realizado por uma equipe composta de Analistas de Dados, que farão a extração e transformação das bases disponíveis, e por auditores do CCIMAR, que conhecem o negócio e construirão um modelo adequado às necessidades.

A Marinha padroniza, por meio de boletim técnico [9], as tecnologias, linguagens e ferramentas para o desenvolvimento de sistemas digitais (SD), destacando os softwares *Oracle* (mediante licenciamento) e *PostgreSQL* (software livre) como soluções de SGBD¹⁰ possíveis para uso.

Por fim, o trabalho coaduna com o de Balaniuk [5], ao afirmar que “Equipes multidisciplinares de análise, compostas por perfis de negócio e TI, poderão obter melhores resultados em mais curto espaço de tempo”. Para que o conhecimento sobre MD seja internalizado no CCIMAR, torna-se imprescindível investir na captação ou capacitação de servidores em Análise de Dados, a fim de, juntamente com os auditores, conhecer a fundo as regras de negócio pertinentes às auditorias de licitações, de forma a produzir melhores regras de análise dos dados disponíveis.

5. CONSIDERAÇÕES FINAIS

Este trabalho intencionou responder a seguinte pergunta de pesquisa: é possível utilizar Mineração de Dados nas auditorias de licitações do CCIMAR?

Para responder a essa pergunta, e a fim de atender os objetivos do trabalho, inicialmente foi necessário verificar a relevância do estudo em questão. Em entrevista não estruturada com auditores do CCIMAR, observou-se que os processos de auditoria conduzidos pelo Centro são realizados com um certo tempo transcorrido após a ocorrência dos atos praticados pelo órgão auditado. A fim de preencher essa lacuna, o órgão tem se dedicado a desenvolver um modelo de auditoria contínua, baseada em análise de dados, mas ainda se situa nos estágios iniciais dessa pesquisa.

A área de licitações do CCIMAR, em que pese a existência de uma enorme

¹⁰ SGBD - Sistema de Gerenciamento de Banco de Dados, é um conjunto de softwares que têm como principal objetivo controlar o acesso, a persistência, a manipulação e a organização dos dados.

quantidade de registros provenientes dos sistemas de informação da Administração Pública, realiza a maioria de suas verificações sem o apoio de ferramentas de análise de dados. Dessa forma, essa área foi escolhida por apresentar um potencial de melhoria nos processos.

O passo seguinte constituiu-se no levantamento bibliográfico e documental sobre os assuntos que culminaram no entendimento sobre o tema, tomando como ponto de partida o KDD, conceito que faz referência a um sentido amplo de extrair conhecimento a partir de uma Base de Dados. Dessa forma, foi necessário conceituar Bancos de Dados e a diferença na utilização de sistemas OLAP e OLTP na construção de um *Data Warehouse* capaz de realizar consultas analíticas.

De posse dessas informações, foi possível trazer o entendimento sobre a Mineração de Dados contextualizado a um processo de descoberta de conhecimento em bases de dados. A classificação em diferentes tarefas foi apresentada a fim de destacar as diferentes aplicações dessa ferramenta, que podem ser utilizadas individualmente ou em conjunto, dependendo do tipo de conhecimento a ser extraído. Assim, o primeiro objetivo específico, que era identificar os requisitos necessários para utilização da Mineração de Dados, foi atingido.

Dessa maneira, foi possível abordar a auditoria contínua, vista como uma evolução do processo de auditoria tradicional, no que refere a frequência e a forma de análise dos dados, destacando o papel da MD na análise proativa de todos os registros disponíveis como sendo um benefício potencial na realização das auditorias de licitações. Também foram descritos outros trabalhos referentes a MD, a fim de apresentar diferentes aplicações práticas dessa ferramenta nos órgãos governamentais, atingindo assim, o segundo objetivo específico, que consistia em identificar se a MD pode ser utilizada na área de auditoria de licitações.

Encerrado o levantamento do referencial teórico que ampara o tema, foi realizado um estudo de caso no CCIMAR, que constatou, dentre outras observações, que a área de licitações se encontra numa fase embrionária de estudos sobre auditoria contínua e Mineração de Dados, no entanto, já foram iniciadas parcerias para preencher essa lacuna, e que há carência de pessoal com capacitação em TI dedicado às atividades de auditoria.

Com isso, o último objetivo específico foi alcançado e a pergunta de pesquisa é respondida com a proposta de hospedar um *Data Mart* no Centro de Dados da Marinha, utilizando uma modelagem do tipo estrela como ambiente de Mineração de Dados próprio e adequado às necessidades do CCIMAR. A pesquisa enfatiza a capacitação de servidores em Análise de Dados como forma de conhecer e desenvolver as regras de negócio em conjunto com os auditores do Centro, permitindo uma melhor utilização dos dados disponíveis.

Acredita-se que o resultado dessa pesquisa possa fomentar a discussão sobre a Mineração de Dados. O trabalho foi conduzido visando aproximar esse tema, que é típico da área de TI, ao público geral, pois a compreensão sobre esse assunto e a abrangência das possibilidades de utilização permeiam diversas áreas, como saúde, finanças e logística. Portanto, este estudo não pretende esgotar a abordagem do assunto, podendo servir como ponto de partida para futuros trabalhos.

Cabe registrar que, por limitações de escopo, não foram realizadas experiências ou testes de Mineração de Dados em licitações, pois essa análise requer a formulação de determinadas regras e a criação de um modelo de dados em formato estrela, um trabalho a ser realizado conjuntamente por analistas e auditores de acordo com as necessidades da tarefa. Por esse motivo, sugere-se como pesquisa futura o uso de técnicas de MD para auditoria em áreas nas quais a Marinha já possui bases de dados em um *Data Mart* em modelo estrela, como o SINGRA ou SISPAG2.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ALLES, Michael G. et al. Continuous auditing: the USA experience and considerations for its implementation in Brazil. **Journal of Information Systems and Technology Management**. v. 3, n. 2, pp. 211-224, 2006.
- [2] ALVARENGA, Ricardo Akl Lasmar de. **Caracterização de competitividade de pregões eletrônicos por meio de Mineração de Dados**. 2011. 76 f. Dissertação (Mestrado) - Curso de Gestão do Conhecimento e da Tecnologia da Informação, Universidade Católica de Brasília, Brasília, 2011.
- [3] AMO, Sandra de. **Técnicas de Mineração de Dados**. In: Congresso da Sociedade Brasileira de Computação. Jornada de atualização em Informática, 24, 2004, Salvador.
- [4] ATAIDES, A. P. **Desenvolvimento de sistemas em Linux: uma análise das ferramentas de desenvolvimento para Linux**. 2004. 125 p. Monografia (Especialização em Administração em Redes Linux) – Universidade Federal de Lavras, Lavras, 2004.
- [5] BALANIUK, Remis. A Mineração de Dados como apoio ao Controle Externo. **Revista do TCU**, Brasília, v. 42, n. 117, p.77-84, 2010.
- [6] BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal: Centro Gráfico, 1988. 292 p.
- [7] _____. Lei Complementar nº101, de 4 de maio de 2000. Estabelece normas de

finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. **Diário Oficial [da] União**. Brasília, DF, 4 de maio de 2000.

- [8] _____. Marinha do Brasil, Secretaria-Geral da Marinha. SGM-601. **Normas sobre Auditoria, Análise e Apresentação de Contas na Marinha**. 5ª Rev. Brasília, DF, 2014b.
- [9] _____. Marinha do Brasil. Diretoria de Comunicações e Tecnologia da Informação da Marinha. DCIMTBOTEC 30/009/2016: **Padronização de tecnologias, linguagens e ferramentas para o desenvolvimento de Sistemas Digitais (SD)**. Rio de Janeiro, 2016.
- [10] _____. Marinha do Brasil. Diretoria-Geral do Material da Marinha. DGMM-0540: **Normas de tecnologia da Informação da Marinha**. 2ª Rev. Rio de Janeiro, 2017.
- [11] _____. Ministério do Planejamento e Controladoria-Geral da União. **Manual do Usuário: Sistema de Trilhas de Auditoria - Auditores e Gestores da CGU**. Brasília, DF, 2014a.
- [12] _____. Decreto nº5.450, de 31 de maio de 2005. Regulamenta o pregão, na forma eletrônica, para aquisição de bens e serviços comuns, e dá outras providências. **Diário Oficial [da] União**. Brasília, DF, 31 de maio de 2005.
- [13] _____. Lei nº4.320, de 17 de março de 1964. Estatui Normas Gerais de Direito Financeiro para elaboração e controle dos orçamentos e balanços da União, dos Estados, dos Municípios e do Distrito Federal. **Diário Oficial [da] União**. Brasília, DF, 17 de março de 1964.
- [14] _____. Lei nº8.666, de 21 de junho de 1993. Regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da Administração Pública e dá outras providências. **Diário Oficial [da] União**. Brasília, DF, 21 de junho de 1993.
- [15] _____. Lei nº10.520, de 17 de julho de 2002. Institui, no âmbito da União, Estados, Distrito Federal e Municípios, nos termos do art. 37, inciso XXI, da Constituição Federal, modalidade de licitação denominada pregão, para aquisição de bens e serviços comuns, e dá outras providências. **Diário Oficial [da] União**. Brasília, DF, 17 de julho de 2002.
- [16] CHAN, David; VASARHELYI, Miklos. **Innovation and Practice of Continuous**

- Auditing**. p.271-283, 2018. Emerald Publishing Limited.
- [17] CORREIA, Marcelo. **Armazenamento e Mineração de Dados**. Dakar: Universidade Virtual Africana, 2018. 92 f.
- [18] ELMASRI, Ramez; NAVATHE, Shamkant. **Sistemas de Banco de Dados**. 6. ed. São Paulo: Pearson, 2011. 770 p.
- [19] FAYYAD, Usama; PIATETSKY-SHAPIO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. **Ai Magazine**, Providence, v. 17, n. 3, p.37-54, 1996.
- [20] GIL, Antônio Carlos. **Métodos e Técnicas de Pesquisa Social**. 6ª. ed. São Paulo: Atlas S.A, 2008.
- [21] GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel; BEZERRA, Eduardo. **Data Mining: conceitos, técnicas, algoritmos, orientações e aplicações**. 2. ed. Rio de Janeiro: Elsevier, 2015.
- [22] GUIMARÃES, William Sérgio Azevêdo. **Data Mining aplicado ao serviço público, extração de conhecimento das ações do Ministério Público Brasileiro**. 2000. 152 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2000.
- [23] HAN, Jiawei; PEI, Jian; KAMBER, Micheline. **Data Mining: Concepts and Techniques**. 3. ed. San Francisco, CA, USA: Elsevier, 2011. 744 p.
- [24] HOKAMA, Daniele; CAMARGO, Denis; FUJITA, Francine; FOGLIENE, João. **Modelagem de Dados no Ambiente Data Warehouse**. 2004. 121 f. Monografia (Especialização) - Curso de Sistemas de Informação, Universidade Presbiteriana Mackenzie, São Paulo, 2004.
- [25] INMON, William H. **Como Construir o Data Warehouse**. 2a. Ed., Editora Campus, Brasil, 1997.
- [26] KIMBALL, Ralph. **Data Warehouse Toolkit**. Makron Books. São Paulo. 1998.
- [27] LAROSE, D. T. **Discovering Knowledge in Data: An Introduction to Data Mining**. Hoboken, New Jersey, USA: John Wiley and Sons Inc, 2005. 214p.
- [28] LOH, Stanley. **BI na era do Big Data para Cientistas de Dados - indo além de cubos e dashboards na busca pelos porquês, explicações e padrões**. Porto Alegre, 2014.

- [29] MURCIA, F.; SOUZA, F.; BORBA, J. Continuous Auditing: A Literature Review. **Revista Organizações em Contexto**, v. 4, n. 7, p.1-17, 30 jun. 2008. Instituto Metodista de Ensino Superior.
- [30] NORMAS INTERNACIONAIS DAS ENTIDADES FISCALIZADORAS SUPERIORES. **ISSAI 100: Princípios Fundamentais de Auditoria do Setor Público**. Viena, Áustria: Traduzido pelo Tribunal de Contas da União, 2013.
- [31] PIATETSKY-SHAPIO, Gregory. Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop. **Ai Magazine**, Providence, v. 11, n. 4, p.68-70, 1990.
- [32] POE, Vidette. **Building a Data Warehouse for Decision Support**. New Jersey, USA: Prentice Hall PTR, 1996.
- [33] PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. Novo Hamburgo: Feevale, 2013.
- [34] RAJPUROHIT, Anmol. **Infographic - Data Scientist or Business Analyst? Knowing the Difference is Key**. 2015.
- [35] SFERRA, Heloisa Helena; CORRÊA, Ângela M. C. Jorge. Conceitos e Aplicações de Data Mining. **Revista de Ciência & Tecnologia**, Piracicaba, v. 11, n. 22, p.19-34, 2003.
- [36] SHAMS, K.; FARISHTA, M. (2001), Data Warehousing: Toward knowledge Management, **Topics in Health Information Management**, v. 21, n. 3, p. 24-32.
- [37] SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. **Sistema de banco de dados**. 5. ed. Rio de Janeiro: Elsevier, 2006. 781 p.
- [38] SILVA, Carlos Vinícius Sarmiento. **Agentes de Mineração e sua Aplicação no Domínio de Auditoria Governamental**. 2011. 122 f. Dissertação (Mestrado) - Curso de Informática, Universidade de Brasília, Brasília, 2011.
- [39] SILVA. W. L.; ALMEIDA J.R. **Auditoria Contínua de dados como Instrumento de Automação do Controle Empresarial**. Escola Politécnica da Universidade de São Paulo. São Paulo. 2012.
- [40] VASARHELYI, Miklos. A.; HALPER, F.B. **The continuous audit of online systems**. Rutgers University, New Jersey. 1990.

- [1] As Referências Bibliográficas devem seguir as normas da ABNT/NBR 6023.