

MINERAÇÃO DE DADOS: uma revisão da literatura em Administração

Cátia Fabíola Parreira de Avelar¹

Thiago Augusto Hernandes Rocha²

Flávia Juliesse Soares Cruz³

RESUMO

O objetivo do artigo foi realizar uma revisão da literatura sobre a técnica de mineração de dados (*Data Mining*) em bases de dados que concentram as publicações voltadas para Administração: EBSCO, SCIELO e SPELL. Foram identificados 238 artigos de diversas áreas do conhecimento e após a exclusão dos trabalhos que não pertenciam a Administração, restaram 24 artigos. Não foi adotada limitação temporal quanto à data de publicação dos trabalhos. Os artigos foram analisados quanto: a qualidade do veículo de publicação (webqualis), o objetivo, as técnicas utilizadas, a subárea de concentração dentro da Administração, ao país de origem e principais conceitos articulados. Observou-se que apesar do grande número de artigos que utilizam a técnica em diversas áreas do conhecimento, a publicação científica em Administração é incipiente. As publicações concentram-se

¹ Doutoranda, mestre e especialista pelo Centro de Pós-Graduação e Pesquisas em Administração da Universidade Federal de Minas Gerais – CEPEAD/UFMG- Email : catiaavelar@yahoo.com.br

² Doutorando e mestre pelo Centro de Pós-Graduação e Pesquisas em Administração da Universidade Federal de Minas Gerais – CEPEAD/UFMG - e mail : rochahernandes285@yahoo.com.br

³ Bacharel em Administração pela Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais FACE/UFMG- email: flavia.jsacruz@gmail.com

nas subáreas de Finanças e *Marketing* e em sua maioria são estudos empíricos aplicados. Os países que mais publicaram foram Taiwan e Brasil.

PALAVRAS-CHAVE: DATA MINING. MINERAÇÃO DE DADOS. REVISÃO. ADMINISTRAÇÃO.

INTRODUÇÃO

Os avanços tecnológicos na área da informática que tornaram o uso global da internet parte do cotidiano de consumidores e organizações; o aumento do espaço de armazenamento de dados, da velocidade e da capacidade de análise de computadores, além da sofisticação de softwares disponíveis para transformação de dados em informação; possibilitou o acesso a um gigantesco volume de dados que pode ter várias aplicações, como por exemplo, fornecer informações para auxiliar a solução de problemas dentro de uma organização. Atualmente, a maioria das operações e atividades de instituições, é registrada em computadores e armazenada em grandes bases de dados (GALVÃO; MARIN, 2009). Porém, apesar da existência de grandes bancos de dados com muitas informações sobre consumidores, fornecedores e operações da própria organização, existem dificuldades na descoberta de conhecimento baseado nessas informações, fazendo com que recursos computacionais ainda não sejam usados na tomada de decisão em várias organizações (DIAS, 2002).

As técnicas de mineração de dados possibilitam uma série de aplicações na Administração pública ou de empresas privadas, seja como processo de verificação ou como processo de descoberta. Desde classificar pedidos de crédito, esclarecer pedidos de seguros fraudulentos, identificar a melhor forma de tratamento de um paciente, estimar o número de filhos ou a renda total de uma família, estimar o valor

em tempo de vida de um cliente, prever a demanda de um consumidor para um novo produto; até determinar quais produtos tendem a ser comprado junto, agrupar clientes por região ou comportamento de compra similar, análise de usuários da internet e reputação de marca via material postado em redes sociais. A disponibilidade de dados como: aqueles oriundos de banco de dados das empresas, do governo, conteúdos em páginas da internet, atualizações em redes sociais, áudio, vídeo, imagens, check-ins e praticamente todo o conteúdo indexável de redes sociais, associados à disponibilidade de pacotes de programa de mineração de dados (SPSS[®], SAS[®], Kdnuggets[®], Angoss[®], Megaputer[®], para citar alguns dentre vários) e de ferramentas para monitoramento e gerenciamento de páginas da internet (Google Alertas[®], Feedly[®], TwitterAdvancedSearch[®], Tagboard[®], BrandWatch[®], Socialbakers[®], dentre vários) torna o uso do método cada vez mais viável e atraente para tomada de decisões e planejamento estratégico das organizações.

Apesar das várias possibilidades de aplicação da técnica de mineração de dados na Administração, ainda são encontradas dificuldades na descoberta de conhecimento baseada nessas informações. Essas dificuldades podem estar relacionadas a vários fatores: falta de conhecimento da existência de técnicas de mineração de dados; complexidade na aplicação de uma técnica de mineração de dados; falta de ferramentas adequadas; alto custo das ferramentas disponíveis no mercado; falta de parâmetros de referência na escolha da técnica e da ferramenta mais adequada para cada problema a ser solucionado (DIAS, 2002).

Considerando as perspectivas abertas por este novo tipo de tecnologia e as possíveis dificuldades para aplicação e uso da técnica, o presente artigo tem por objetivo revisar a literatura sobre o tema. Várias possibilidades de aplicação de técnicas inerentes à mineração de dados foram analisadas para fornecer um panorama da utilização de ferramentas dessa natureza, considerando aplicações junto à ciência administrativa de modo geral, sem a priorização de subáreas.

A organização do artigo seguiu a seguinte ordem: na próxima seção será apresentado o referencial teórico acerca do tema; a seguir serão expostos os procedimentos metodológicos para realização do trabalho de revisão. Na seção seguinte são apresentados os dados da categorização dos trabalhos levantados e discutidas as tendências em termos da utilização de técnicas de mineração de dados. Na última parte foram discutidas as perspectivas da área, considerando-se os avanços em termos de disponibilização de informações e quais os desafios se colocam para os administradores incorporarem esse tipo de técnica em suas atividades de planejamento e gestão.

1 REFERENCIAL TEÓRICO

A técnica da mineração de dados– do Inglês *Data Mining* – é o processo de descoberta de padrões interessantes, inovadores e desconhecidos assim como de modelos descritivos, compreensíveis e preditivos a partir de dados em grande escala (ZAKY; MEIRA JUNIOR., 2014). A técnica pode ser considerada uma alternativa para extrair conhecimento a partir de grandes volumes de dados, pois possibilita a descoberta de padrões e relações ocultas possibilitando a geração de regras para prever e correlacionar dados, com o intuito de tornar a tomada de decisão mais rápida e/ou confiável. Malhotra e Birks (2007, p. 138) definem a técnica como um processo de descoberta de correlações significativas, padrões e tendências através de grandes quantidades de dados armazenados em repositórios, usando reconhecimento de padrões, estatística e técnicas matemáticas. Assim, dentre os principais objetivos da mineração de dados, pode-se destacar: (1) a descoberta de relacionamentos entre dados e (2) fornecer subsídios para que possa ser feita uma previsão de tendências futuras baseadas no passado.

Além de auxiliar na tomada de decisão, os resultados obtidos a partir da aplicação de técnicas de mineração de dados podem ser utilizados no gerenciamento de informação, processamento de pedidos de informação, controle de processos, criação de modelos preditivos e pesquisa, dentre várias outras aplicações. Segundo Dias (2002) diversas áreas do conhecimento já utilizam técnicas de mineração de dados, destacando-se: o Marketing, que faz o uso de técnicas de mineração de dados para descobrir preferências do consumidor, padrões de compras para segmentação e personalização de programas de marketing; Finanças, que usa as técnicas para descobrir padrões de fraude e desenvolver modelos de previsão e de inadimplência e o Controle de processos e de qualidade, que usam algumas técnicas para auxiliar no planejamento estratégico de linhas de produção e para buscar padrões de condições físicas na embalagem e armazenamento de produtos.

A técnica de mineração de dados pode também ser usada na Medicina para caracterizar comportamentos de pacientes, identificar terapias médicas de sucesso para diferentes doenças, buscar por padrões de novas doenças e para previsão de epidemias. Em instituições governamentais, a técnica pode ser usada para descobrir padrões para melhorar as coletas de taxas e impostos ou descobrir fraudes; e nas Ciências ajudando cientistas em suas pesquisas, seja na descoberta de novos padrões como na previsão de comportamentos de partículas, moléculas e populações, como na previsão de comportamentos. A autora (Dias, 2002) ainda sugere o uso de técnicas de mineração de dados por instituições financeiras para detectar padrões do uso de cartão de crédito, identificar clientes “leais” e encontrar correlações não óbvias entre diferentes indicadores financeiros; o uso no transporte, para determinar as escalas de distribuição entre distribuidores e analisar padrões de carga; e em áreas de Ciências e Tecnologia, na avaliação de grupos de pesquisas do país.

A mineração de dados pode ser aplicada como um processo de verificação em que hipóteses sobre a relação entre os dados são formuladas e testadas a partir de análises estatísticas e multidimensionais de bancos de dados e também pode ser aplicada como um processo de descoberta para descobrir padrões não óbvios a partir de bancos de dados.

As principais tarefas de mineração de dados são descritas a seguir (segundo DIAS, 2002; HAIR JUNIOR; BUSH; ORTINAU, 2003, p.164; GALVÃO; MARIN, 2009): (i) Classificação (construção de um modelo que possa ser aplicado a dados não classificados, a fim de categorizá-los em classes, com o objetivo de descobrir um relacionamento entre um atributo meta e um conjunto de atributos de previsão); (ii) Estimativa ou regressão (usada para definir um valor para alguma variável contínua desconhecida como, por exemplo, estimar o número de filhos ou a renda total de uma família ou o valor em tempo de vida de um cliente); (iii) Associação (usada para determinar quais itens tendem a ser adquiridos juntos em uma mesma transação); (iv) Segmentação (divisão de uma população heterogênea em subgrupos ou grupos mais homogêneos); (v) Sumarização (envolve métodos para encontrar uma descrição compacta para um subconjunto de dados) e (vi) Séries temporais (similarmente operacionalizado como a regressão, mas utilizando medidas baseadas no tempo explorando relações que podem existir em períodos específicos de tempo).

Dentre as várias técnicas de mineração de dados existentes, destacam-se: Descoberta de Regras de Associação; Árvores de Decisão; Raciocínio Baseado em Casos ou MBR; Algoritmos Genéticos; Redes Neurais Artificiais; Modelagem Preditiva; Indução de Regras e Lógica Difusa ou nebulosa (*fuzzylogic*). Para uma descrição dessas técnicas e exemplos de utilização é sugerida a leitura de Dias (2002), Hair Jr., Bush e Ortinau (2003, p.165), Galvão e Marin (2009).

2 MÉTODOS E TÉCNICAS ADOTADOS NA PESQUISA

Para cumprir o propósito do presente trabalho, foram realizadas buscas junto a três bases de dados que concentram as publicações voltadas à ciência administrativa: EBSCO, SCIELO e SPELL. Inicialmente a base de dados SCOPUS também seria incluída, mas a mesma apresentava restrições de acesso via portal de periódicos da CAPES, o que impossibilitou o exame de trabalhos publicados junto a essa base de dados.

Junto à base de dados da EBSCO foi realizada a busca considerando os termos: (*data mining techniques* OR *data mining method*) AND (*systematicreview* OR *meta-analysis*). Foram procurados trabalhos que utilizassem esses conceitos considerando a sua veiculação no título do documento. Além disso, foi aplicado o filtro relativo à apresentação apenas de trabalhos completos, em qualquer idioma. Ao todo nessa base foram encontrados 171 artigos relativos a campos diversos dentre os quais podemos citar: Biologia, Engenharia, Computação, Geografia, Medicina, entre outros. Com a intenção de selecionar apenas trabalhos que relatassem experiências de aplicação de técnicas de mineração de dados na Administração, todos os resumos foram lidos e após a exclusão dos trabalhos que atendessem a este critério foi possível obter um total de 15 artigos.

Junto à base dados do SCIELO a expressão de busca foi apenas: “*data mining* OR mineração de dados”. Essa divergência se deu em função da inexistência de artigos publicados que atendessem aos mesmos critérios de busca adotados junto à base da EBSCO. Quanto ao campo de busca do termo, foi feita a opção por considerar apenas aqueles trabalhos que veiculassem o conceito no título do documento. Ao todo foram encontrados 64 trabalhos que foram analisados a partir de seus respectivos resumos. Após exame para seleção daqueles que se dedicassem as aplicações no campo da Administração, foi possível obter 6 artigos.

Por último, foi feita uma consulta à base de dados SPELL. O termo de busca utilizado foi “*data mining* OR mineração de dados”. A busca se restringiu exclusivamente a artigos que contivessem a palavra-chave no seu respectivo título e que tratassem sobre a Administração. Assim, nessa base foram encontrados apenas 3 trabalhos e todos foram considerados para análise.

Levando-se em conta os três repositórios pesquisados foram identificados um total 238 artigos dos quais 24 atenderam aos critérios de seleção para o exame mais aprofundado. Não foi adotada limitação temporal quanto à data de publicação dos trabalhos selecionados, sendo a busca realizada em dezembro de 2015. Os artigos foram categorizados, para que fosse possível caracterizar como as técnicas de *data mining* têm sido utilizadas no contexto da ciência administrativa, bem como a finalidade de utilização das mesmas. Foi analisada a qualidade do veículo de publicação dos trabalhos segundo o webqualis, objetivo, técnicas utilizadas, subárea de concentração dentro da administração, país de origem do trabalho e principais conceitos articulados na publicação.

3 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

O quadro 1 sumariza os trabalhos que foram selecionados de acordo com o fluxograma de seleção e análise definido na Figura 1.

Figura 1 - Etapas para busca e seleção dos artigos sobre mineração de dados na administração.



Fonte: Elaborada pelos autores.

Quadro 1 - Sumário dos trabalhos examinados.

Autores	Qualis	Objetivo	Tipo de estudo	Subárea	País de origem	Conceitos trabalhados
(Astudillo, Bardeen & Cerpa, 2014)	Não avaliado	Apresentar possibilidades de aplicação de técnicas de mineração de dados em comércio eletrônico	Bibliométrico	Marketing	Chile	Classificação, Clusterização, Associação, Regressão, Detecção de outlier,
(Bae & Kim, 2011)	A2	Identificação de parâmetros preconizados por potenciais clientes para o desenvolvimento de produtos	Empírico aplicado	Marketing	Coréia	Árvores de decisão, Associação, Detecção de padrões, Clusterização, Visualização de dados,
(Blanco, Ortega, Vanti & Johann, 2013)	B2	Identificação automática de indicadores que pudessem caracterizar a imagem da organização segundo as metáforas organizacionais de Morgan.	Empírico aplicado	Cultura organizacional	Espanha	Classificação, Clusterização, Detecção de padrões
(S. C. Chen & Huang, 2011)	A2	Conceber um modelo preditivo que instrumentalizasse a oferta de crédito de qualidade de acordo com o modelo de mineração de dados CRISP.	Empírico aplicado	Finanças	Taiwan	Redes Neurais Artificiais, Árvores de decisão, Detecção de padrões, Predição
(W.-S. Chen & Du, 2009)	A2	Conceber um modelo preditivo de risco financeiro atrelado a companhias que ofertam ações na bolsa de Taiwan	Empírico aplicado	Finanças	Taiwan	Redes Neurais Artificiais, Análise Fatorial, Classificação, Clusterização, Predição
(Cruz & Cortez, 2009)	Não avaliado	Comparação entre dois conjuntos de técnicas relativas à mineração de dados	Revisão sistemática	Métodos	Portugal	Redes Neurais Artificiais, Máquinas de vetores de suporte, Classificação, Regressão
(D'Haen, Van Den Poel, & Thorleuchter, 2013)	A2	Conceber um modelo preditivo de lucratividade de clientes que articule as melhoras técnicas de mineração de dados frente a diferentes tipos	Empírico aplicado	Finanças	Alemanha	Árvores de decisão, Regressão, Predição

Autores	Qualis	Objetivo	Tipo de estudo	Subárea	País de origem	Conceitos trabalhados
		de fonte de informações				
(Duran, Pamukcu, &Bozkurt, 2006)	Não avaliado	Determinar grupos ótimos para o balizamento de campanhas de marketing com maior probabilidade de sucesso	Empírico aplicado	Marketing	Turquia	Árvores de decisão, Redes Neurais Artificiais, Regressão
(Gepp, Wilson, &Kumar, 2012)	B4	Realizar a análise comparativa de diferentes técnicas de detecção de fraude em seguros para estimativa de ferramentas mais eficazes para controle da questão	Empírico aplicado	Finanças	Estados Unidos	Árvores de decisão, Análise de sobrevivência, Modelos híbridos, Regressão, Redes Neurais Artificiais
(Gordillo-ruiz, Martínez-miranda, &Stephens, 2012)	B3	Desenvolver metodologia que permita a inferência, com base em dados de mercado, se agentes com desempenho similar estão utilizados estratégias de negociação diferenciadas	Empírico aplicado	Finanças	México	Classificação, Inferência bayesiana
(C.-J. Huang, Chen, & Pan, 2011)	B1	Conceber um modelo preditivo capaz de apreender as mudanças de tendência no mercado de ações de Taiwan	Empírico aplicado	Finanças	Taiwan	Predição, Redes Neurais Artificiais, Algoritmos genéticos
(S.-C. Huang, Chang, & Wu, 2009)	A2	Identificar a melhor técnica de clusterização para a análise dos padrões de compra de um determinado revendedor	Empírico aplicado	Marketing	Taiwan	Clusterização, Lógica fuzzy
(Liao, Chu, &Hsiao, 2012)	A2	Fornecer um panorama do desenvolvimento das técnicas de mineração de dados ao longo do período entre 2000 e 2011	Revisão sistemática	Métodos	Taiwan	Redes Neurais Artificiais, Predição, Algoritmos genéticos, Sistemas especialistas, Clusterização, Máquinas de vetores de suporte, Classificação, Associação
(C.-C. Lin, Chiu, Huang, & Yen, 2015)	B3	Examinar os aspectos inerentes ao triângulo da fraude com base em técnicas de mineração de dados e confrontar os	Empírico aplicado	Finanças	Taiwan	Redes Neurais Artificiais, Árvores de decisão

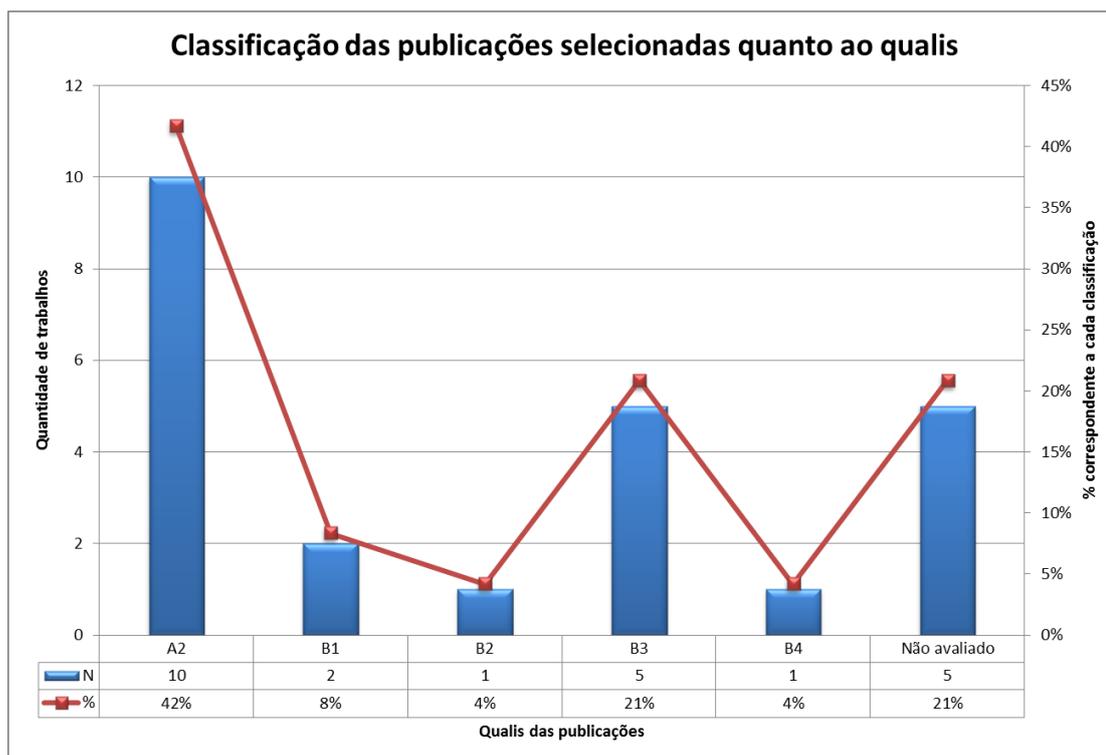
Autores	Qualis	Objetivo	Tipo de estudo	Subárea	País de origem	Conceitos trabalhados
		achados com as percepções de especialistas para identificar padrões de concordância				
(W. T. Lin, Wang, Wu, &Ye, 2011)	A2	Identificar informações não óbvias que pudessem impactar no planejamento e desenho de cursos de treinamento	Empírico aplicado	Recursos Humanos	Taiwan	Clusterização, Redes Neurais Artificiais
(Ngai, Xiu, &Chau, 2009)	A2	Identificar tendências na literatura quanto ao gerenciamento do relacionamento com o cliente	Revisão sistemática	Marketing	China	Classificação, Clusterização, Associação, Regressão, Detecção de padrões, Predição
(Pinho, 2008)	B3	Identificar perfis de clientes, por meio da aplicação de mapas de Kohonen, como estratégia de diferenciação no marketing empresarial	Empírico aplicado	Marketing	Brasil	Clusterização, Classificação
(Safer, 2003)	B1	Comparação entre dois conjuntos de técnicas relativas à mineração de dados quanto à acurácia em prever retornos anormais de ativos de mercado	Revisão sistemática	Finanças	Estados Unidos	Redes Neurais Artificiais, Regressão
(Silva, 2011)	B3	Identificação de cartéis participantes de licitações com o uso de técnicas de mineração de dados, para delineamento de estratégias de combate à corrupção	Empírico aplicado	Gestão pública	Brasil	Classificação, Clusterização, Associação
(Sousa & Figueiredo, 2014)	B3	Desenvolvimento de modelo preditivo quanto à capacidade de pagamento dos cooperados vinculados a programas de concessão de crédito de uma cooperativa de crédito	Empírico aplicado	Finanças	Brasil	Redes Neurais Artificiais, Árvores de decisão
(Wei, Lee, Chen, & Wu, 2013)	A2	Concepção de modelo preditivo capaz de orientar as estratégias de marketing de um salão de beleza de Taiwan	Empírico aplicado	Marketing	Taiwan	Clusterização

Autores	Qualis	Objetivo	Tipo de estudo	Subárea	País de origem	Conceitos trabalhados
(R. S. Wu, Ou, Lin, Chang, &Yen, 2012)	A2	Conceber um modelo capaz de melhorar o desempenho de ações para detecção de sonegação de impostos	Empírico aplicado	Finanças	Taiwan	Associação
(R.-C. Wu, Chen, &Chian, 2006)	Não avaliado	Identificação das causas de não conformidade na produção de embalagens para monitores de cristal líquido	Empírico aplicado	Organização e métodos	Taiwan	Redes Neurais Artificiais, Árvores de decisão
(Aular& Pereira, 2007)	Não avaliado	Examinar técnicas e ferramentas com potencial de embasar a tomada de decisão empresarial	Bibliométrico	Estratégia	México	Business Intelligence, Online Analyticalprocessing - OLAP, Aprendizagem das máquinas

Fonte: Elaborado pelos autores.

Levando-se em conta a avaliação da qualidade dos veículos de divulgação dos artigos selecionados, pode-se perceber uma polarização de revistas da área de Tecnologia da Informação, publicadas no contexto internacional, com mais de 35% dos trabalhos sendo veiculados em revistas com esse perfil, conforme ilustra o gráfico Gráfico 1. As publicações de maior qualidade podem ser agrupadas nas categorias A2 e B1, sendo que ambas corresponderam a 50% dos trabalhos publicados. Apesar disso há inúmeros trabalhos interessantes que apresentam novas possibilidades de aplicação de técnicas de mineração de dados em diferentes campos e que foram publicados em veículos de menor impacto acadêmico, mas que, nem por isso, deixam de destacar potencialidades relevantes.

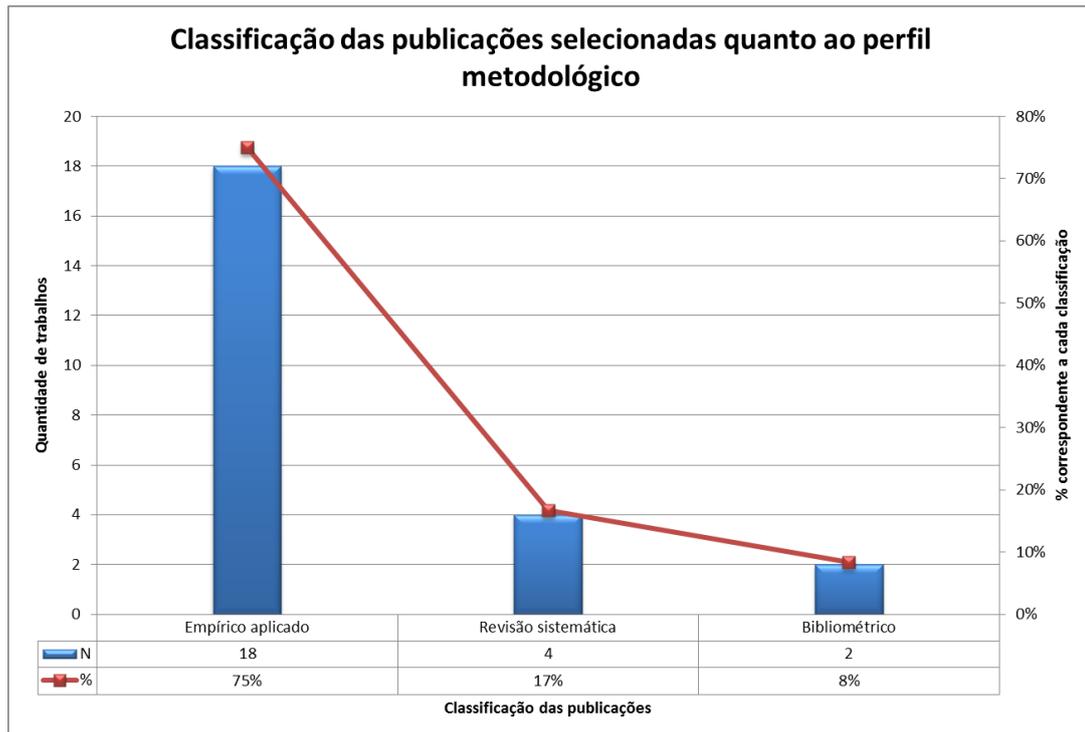
Gráfico 1 - Classificação no webqualis das revistas consideradas na análise.



Fonte: Elaborado pelos autores.

Uma característica vinculada aos estudos sobre mineração de dados é a grande ocorrência de trabalhos empíricos aplicados. A disponibilidade de bases de dados de diferentes fontes e o crescimento da comunidade de analistas com domínio das técnicas referentes a este campo, têm fomentado a aplicação de procedimentos analíticos de forma pragmática, como pode ser observado junto ao Gráfico 2, no qual 75% dos trabalhos analisados foram referentes ao uso de técnicas para a abordagem de problemas concretos. Em segundo lugar aparecem os estudos de revisão sistemática, que são essenciais para o fornecimento de um panorama de evolução dos trabalhos dedicados à mineração de dados junto à ciência administrativa.

Gráfico 2 Classificação metodológica dos estudos de mineração de dados na Administração.

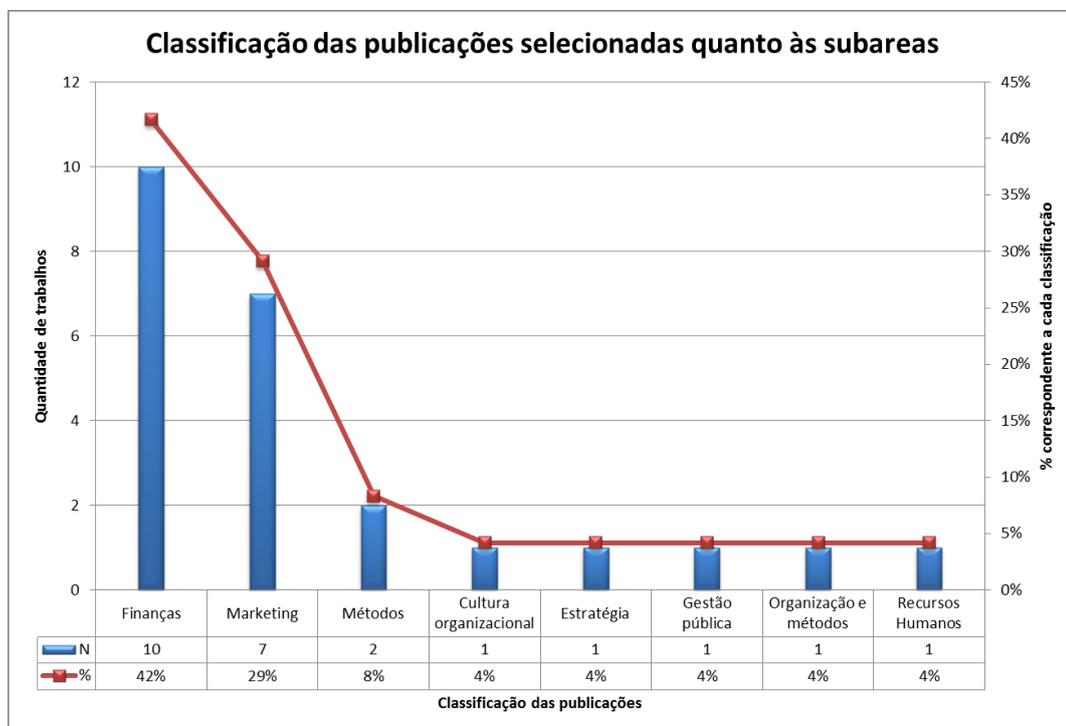


Fonte: Elaborado pelos autores.

Quanto às subáreas da Administração que mais desenvolveram estudos de mineração de dados, podemos destacar Finanças e *Marketing*. Essas duas áreas tradicionalmente se valem de metodologias estatísticas complexas para a abordagem de problemas. Assim, pode-se inferir que há uma relativa similaridade entre as técnicas dedicadas à mineração de dados e aquelas ferramentas já consolidadas junto aos trabalhos de pesquisa usualmente desenvolvidos nessas duas áreas. As demais áreas que apresentaram artigos publicados com a utilização de ferramentas de mineração de dados evidenciaram um padrão incipiente de uso

de técnicas dessa natureza, fato esse que acena para inúmeras possibilidades ainda inexploradas, passíveis de abordagem pelos profissionais do campo.

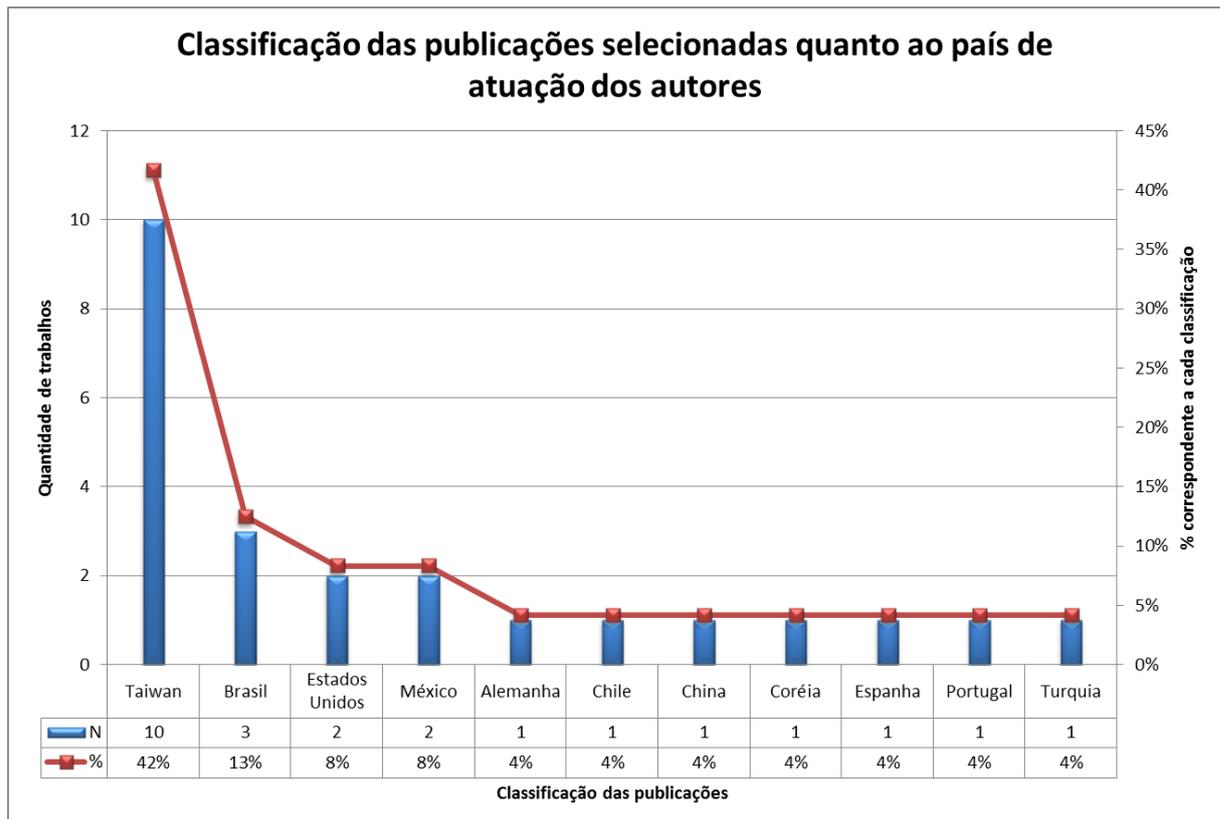
Gráfico 3 - Subáreas da Administração nas quais foram desenvolvidos os estudos.



Fonte: Elaborado pelos autores.

O Gráfico 4 ilustra o país de vinculação profissional dos autores que publicaram os trabalhos considerados nas análises. Pode-se observar uma concentração de autores de origem asiática, seguido por brasileiros, americanos e mexicanos. No levantamento realizado, diversos trabalhos feitos para examinar problemas de diversas áreas, especialmente vinculados à perspectiva financeira, foram conduzidos em Taiwan.

Gráfico 4 - País de atuação profissional dos autores dos trabalhos analisados.



Fonte: Elaborado pelos autores.

O último elemento analisado remeteu aos conceitos trabalhados nos artigos investigados. A nuvem de palavras da Figura 2 sumariza aqueles conceitos mais frequentemente trabalhados nos artigos. Na figura, quanto maior a palavra, maior é a frequência com que a mesma foi encontrada nos diferentes trabalhos. A mineração de dados pode ser utilizada para diferentes finalidades e nos trabalhos examinados, as ações mais recorrentes eram aquelas dedicadas à clusterização (segmentação) e classificação de padrões, em menor grau podem-se observar alguns trabalhos dedicados à associação de elementos. As técnicas mais utilizadas foram: Redes Neurais Artificiais e Árvores de Decisão. Essas duas técnicas são as mais simples e

com maior literatura instrumental. Técnicas mais sofisticadas como Análise de Sobrevivência e técnicas dedicadas ao exame de fenômenos longitudinais e de inferência *bayesiana* foram apenas tangenciadas pelos trabalhos analisados.

Figura 2- Nuvem dos conceitos mais frequentemente articulados nos trabalhos analisados.



Fonte: Elaborada pelos autores.

CONCLUSÕES

No contexto tecnológico atual, o conteúdo digital disponível no mundo inteiro, seja dentro das organizações, seja na internet, apresenta dados que, se forem devidamente trabalhados, tornam-se informações valiosas para organizações, seja para o planejamento, controle, previsão ou tomada de decisão. Todo volume de informações, acumulado dia após dia, gera uma quantidade de dados disponíveis que permite a extração de informação e construção de modelos relevantes para diversas áreas da Administração, mas que também apresentam uma dificuldade de análise devido ao grande volume. A mineração de dados é uma técnica que sobrepõe à dificuldade de manipulação e transformação de dados em informação relevante. Porém os resultados do presente trabalho mostram que ainda existem poucos artigos acadêmicos publicados vinculados a área de Administração (24 trabalhos). Tal achado pode ser uma possível explicação para a afirmação de Dias (2002) de que muitas organizações ainda não utilizam a técnica. Percebe-se assim a necessidade do desenvolvimento de mais pesquisas na área da Administração que mostrem a aplicabilidade da técnica para descoberta de relações não óbvias entre variáveis e construção de modelos preditivos baseado em dados disponíveis.

Dentre os artigos analisados, a predominância das áreas de *Marketing* e *Finanças*, além de trabalhos cuja finalidade foi a segmentação e classificação de padrões, confirma a aplicabilidade da técnica para finalidades típicas das áreas, ou seja, classificar e segmentar pessoas e comportamentos. Outra explicação possível para tais resultados é a prática recorrente do uso de métodos quantitativos nas duas áreas, o que facilita o uso de técnicas de mineração de dados, que exige conhecimento e/ou familiaridade com técnicas matemáticas e estatísticas. Esse resultado também indica que existe um grande campo de estudos para mineração de dados nas demais áreas da Administração.

Neste artigo procurou-se fornecer informações sobre a mineração de dados em Administração. Mesmo com uma grande potencialidade de aplicação da técnica para geração de informações relevantes para diversas áreas da Administração, a baixa quantidade de artigos acadêmicos publicados e a concentração de artigos nas áreas de *Marketing* e Finanças, apresentam um grande campo a ser explorado por outras áreas da Administração. Como a produção acadêmica está relacionada com o conhecimento produzido nos cursos de pós-graduação, abre-se a oportunidade do ensino e pesquisa de técnicas de mineração de dados nos centros de pós-graduação em Administração do país.

DATA MINING: a literature review in the field of bussiness

ABSTRACT

The aim of this paper was to make a literature review about the Data Mining Technique in scientific databases that concentrate papers related to Business Science: EBSCO, SCIELO and SPELL. The search on EBSCO database was conducted using the words: data mining techniques or data mining method and systematic review ou meta-analysis. The literature search using the words data mining and/or “mineração de dados” (in Portuguese), was conducted on SCIELO and SPELL databases. The exclusion criteria were to have the key word in the publication’s title and to belong to the Business field. Were identified 238 papers from several knowledge fields, all abstracts have been analyzed and the exclusion criteria applied, leaving 24 papers. The papers were separated by the quality of the journal according to WebQualis, by it’s main objective, the techniques used, the business’s subarea concentration, the country’s publication and main concepts articulated. The findings showed that despite the large number of published articles in several fields,

the scientific publications related to data mining in Business field is still scarce. Most of the paper were concentrated Finance and Marketing and mostly are empirical or applied studies. Taiwan and Brazil published the large amount of papers.

KEYWORDS: DATA MINING. LITERATURE REVIEW. BUSINESS.

REFERÊNCIAS

AUSTILLO,C.; BARDEEN, M.; CERPA, N. Data Mining in eletronic Commerce – support vs. Confidence. **Journal of Theoretical and Applied EletronicCommerce research**, 9(1), 1-2, 2014.Disponívelem: <<http://doi.org/10.4067/S0718-18762014000100001>>. Acesso em: 10 dez. 2015.

AULAR, Y.;JOSEFINA, M.; PEREIRA, R. T. Minería de datos como soporte a la toma de decisiones empresariales. **Opción**, 23(52), 1–6, 2007.

BAE, J. K.; KIM, J. Product development with data mining techniques: A case on design of digital camera. **Expert Systems with Applications**,38(8), 9274–9280, 2011.Disponívelem: <<http://doi.org/10.1016/j.eswa.2011.01.030>>. Acesso em: 10 dez. 2015.

BLANCO, E. R.; ORTEGA, A. C.; VANTI, A.; JOHANN, S. Imágenesorganizacionalesem empresas brasileiras:deteccion y análisiscon técnicas de minería de datos. **Revista de Ciências da Administração**, 15(37), 105–120, 2013. Disponível em <<http://doi.org/10.5007/21758077.2013v15n37p105>>. Acesso em: 10 dez. 2015.

CHEN, S. C.; HUANG, M. Y. Constructing credit auditing and control & management model with data mining technique. **Expert Systems with Applications**, 38(5), 5359–5365, 2011. Disponível em: <<http://doi.org/10.1016/j.eswa.2010.10.020>> Acesso em: 10 dez. 2015.

CHEN, W.-S.; DU, Y.-K. Using neural networks and data mining techniques for the financial distress prediction model. **Expert Systems with Applications**, 36(2), 4075–4086, 2009. Disponível em: <<http://doi.org/10.1016/j.eswa.2008.03.020>> Acesso em: 10 dez. 2015.

CRUZ, A.; CORTEZ, P. Data Mining via redes neuronais artificiais e máquinas de vetores de suporte. **Revista de Estudos Politécnicos**, VII(12), 099–118, 2009.

D'HAEN, J.; VAN DEN POEL, D.; THORLEUCHTER, D. Predicting customer profitability during acquisition: Finding the optimal combination of data source and data mining technique. **Expert Systems with Applications**, 40(6), 2007–2012, 2013. Disponível em: <<http://doi.org/10.1016/j.eswa.2012.10.023>>. Acesso em: 10 dez. 2015.

DIAS, M.M. Parâmetros na escolha de técnicas e ferramentas de mineração de dados. **Acta Scientiarum**, 24(6), 1715-1725, 2002.

DURAN, E. A.; PAMUKCU, A.; BOZKURT, H. Comparison of data mining techniques for direct marketing campaigns. **Journal of Engineering and Natural Sciences**, 32, 142–152, 2006.

GALVÃO, N.D.; MARIN, H.F. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem**, 22(5):686-90, 2009.

GEPP, A.; WILSON, J. H.; KUMAR, K. A Comparative analysis of decision trees vis-à-vis other computational data mining techniques in automotive insurance fraud detection. **Journal of Data Science**, 10, 537–561, 2012.

GORDILLO-RUIZ, J. L.; MARTÍNEZ-MIRANDA, E.; STEPHENS, C. R. Develandoestrategias de mercado : minería de datos aplicada al análisis de mercados financieros. **ComputaciónySistemas**, 16(2), 221–231, 2012.

HAIRJR., J. F.; BUSH, R. P.; ORTINAU, D. J. **Marketing Research within a changing information environment**. 2. ed. Boston: McGraw-Hill/Irwin series in marketing. 2003.

HUANG, C.J.; CHEN, P.-W.; PAN, W.-T.Using multi-stage data mining technique to build forecast model for Taiwan stocks.**Neural ComputingandApplications**, 21(8), 2057–2063, 2011. Disponível em: <<http://doi.org/10.1007/s00521-011-0628-0>>. Acesso em: 10 dez. 2015.

HUANG, S.C.; CHANG, E.-C.; WU, H.-H.A case study of applying data mining techniques in an outfitter's customer value analysis.**Expert Systems with Applications**,36(3), 5909–5915, 2009.Disponívelem: <<http://doi.org/10.1016/j.eswa.2008.07.027>>Acessoem: 10 dez. 2015.

LIAO, S. H.; CHU, P. H.; HSIAO, P. Y. Data mining techniques and applications - A decade review from 2000 to 2011. **Expert Systems with Applications**, 39(12), 11303–11311, 2012.Disponível em:<<http://doi.org/10.1016/j.eswa.2012.02.063>>. Acessoem: dez. 2015.

LIN, C.-C.; CHIU, A.-A.; HUANG, S. Y.;YEN, D. C. Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. **Knowledge-Based Systems**, 89, 459–470, 2015.Disponívelem:<<http://doi.org/10.1016/j.knosys.2015.08.011>>. Acessoem: 10 dez. 2015.

LIN, W. T.; WANG, S. J.; WU, Y. C.; YE, T. C.An empirical analysis on auto corporation training program planning by data mining techniques.**Expert Systems with Applications**,,38, 5841–5850, 2011. Disponívelem: <<http://doi.org/10.1016/j.eswa.2010.11.038>>. Acessoem: 10 dez. 2015.

MALHOTRA, N. K.; BIRKS, D. F. **Marketing Research: an applied approach**. 3ed. Harlow: Prentice Hall, 2007.

NGAI, E. W. T.; XIU, L.; CHAU, D. C. K. Application of data mining techniques in customer relationship management: A literature review and classification. **Expert Systems with Applications**, 36(2), 2592–2602, 2009. Disponível em: <<http://doi.org/10.1016/j.eswa.2008.02.021>>. Acesso em: 10 dez. 2015.

PINHO, A. G. Mineração de dados com mapas de Kohonen: uma abordagem no setor financeiro. **Pensamento Contemporâneo em Administração**, 2(2), 39–49, 2008.

SAFER, A. A comparison of two data mining techniques to predict abnormal stock market returns. **Intelligent Data Analysis**, 7, 3–13, 2003. Disponível em: <<http://dl.acm.org/citation.cfm?id=1293921>>. Acesso em: 10 dez. 2015.

SILVA, C. V. S. Detecção de cartéis em licitações públicas com agentes de mineração de dados. **Revista Eletrônica de Sistemas de Informação**, 10(1), 1–19, 2011. Disponível em: <<http://doi.org/10.5329/RESI.2011.1001008>>. Acesso em: 10 dez. 2015.

SOUZA, M. de M.; FIGUEIREDO, R. S. Credit analysis using data mining: application in the case of a credit union. **Journal of Information Systems and Technology Management**, 11(2), 379–396, 2014. Disponível em: <<http://doi.org/10.4301/S1807-17752014000200009>>. Acesso em 10 dez. 2015.

WEI, J.T.; LEE, M.-C.; CHEN, H.-K.; WU, H.-H. Customer relationship management in the hairdressing industry: an application of data mining techniques. **Expert Systems with Applications**, 40(18), 7513–7518, 2013. Disponível em: <<http://doi.org/10.1016/j.eswa.2013.07.053>>. Acesso em: 10 dez. 2015.



WU, R. S.; OU, C. S.; LIN, H. Y.; CHANG, S. I.; YEN, D. C. Using data mining technique to enhance tax evasion detection performance. **Expert Systems with Applications**, 39, 8769–8777, 2012. Disponível em: <<http://doi.org/10.1016/j.eswa.2012.01.204>>. Acesso em: 10 dez. 2015.

WU, R. C.; CHEN, R.-S.; CHIAN, S. S. Design of a product quality control system based on the use of data mining techniques. **IIE Transactions**, 39–51, 2006. Disponível em: <http://doi.org/10.1080/07408170500346402>. Acesso em: 10 dez. 2015.

ZAKI, M., MEIRA JR. W. **Data mining and analysis: fundamental concepts and algorithms**. New York: Cambridge University Press, 2014.