

# Mineração de dados aplicada à gestão de pessoas

## *Data mining applied to people management*

Valéria Simas Schultz<sup>a</sup>

<sup>a</sup> vsschultz@yahoo.com.br

**Resumo:** Este artigo tem por objetivo revisar pesquisas sobre mineração de dados aplicada à gestão de pessoas. Aborda ainda uma análise bibliométrica em torno do tema, explorando as suas relações e relevância e, neste contexto, descreve as aplicações de mineração de dados encontradas durante a revisão. Esta pesquisa considera que a aplicação de *data mining* à gestão de pessoas tornar-se-á uma ferramenta de apoio e poderá gerar resultados relevantes para a atuação dos gestores, desde que algumas precauções sejam adotadas para mitigar impactos negativos.

**Palavras chave:** Gestão de Pessoas; Seleção de Pessoas; Desempenho no Trabalho; Mineração de Dados.

**Abstract:** This article aims to review research on data mining applied to people management. It also presents a bibliometric analysis on the theme, exploring its relationships and relevance and, in this context, describing the data mining applications found during the review. This research considers that the application of data mining to people management will become a support tool and may generate relevant results for managers, provided that some precautions are taken to mitigate negative impacts.

**Keywords:** Human Resources Management; Employee Selection; Work Performance; Data Mining.

## 1. Introdução

A pesquisa é baseada na premissa de que a aplicação de *data mining* à gestão de pessoas permite aprimoramento de processos, maior transparência nos critérios de decisão e a atuação dos gestores de forma mais assertiva, pontual e customizada, a partir do profundo entendimento dos fatores preponderantes nas mais diversas facetas do processo de potencializar resultados com pessoas.

A tomada de decisões dos líderes baseia-se cada vez mais na análise inteligente de *big data* (van Knippenberg, Dahlander, & Haas, 2015) em vez de sua própria experiência e intuição (McAfee & Brynjolfsson, 2012).

Líderes também recebem maior demanda de trabalho e o aumento da pressão pela inovação força-os, assim como a seus funcionários, a agir com maior velocidade e a tomar decisões rápidas (Schwarz Müller et al., 2018).

Nesse contexto, surge um crescente de pesquisas para apoiar a adoção prática da mineração de dados em Gestão de Pessoas - GP. São várias as aplicações às atividades e processos de GP, tais como a seleção de funcionários, prevenção à rotatividade de pessoal, competências

dos empregados em desenvolvimento ou predição e avaliação do desempenho. Todo o espectro de métodos de mineração de dados, como árvores de classificação, agrupamento, análise de associação, máquinas de vetores de suporte ou redes neurais são empregados, enquanto avanços metódicos e personalizações também são apresentadas (Piazza & Strohmeier, 2012).

## 2. Metodologia

Na pesquisa bibliográfica, com enfoque em bibliometria, realizada em 02/10/19, na base Scopus da Elsevier, utilizou-se a expressão "*data mining*", com recuperação de 160.683 documentos.

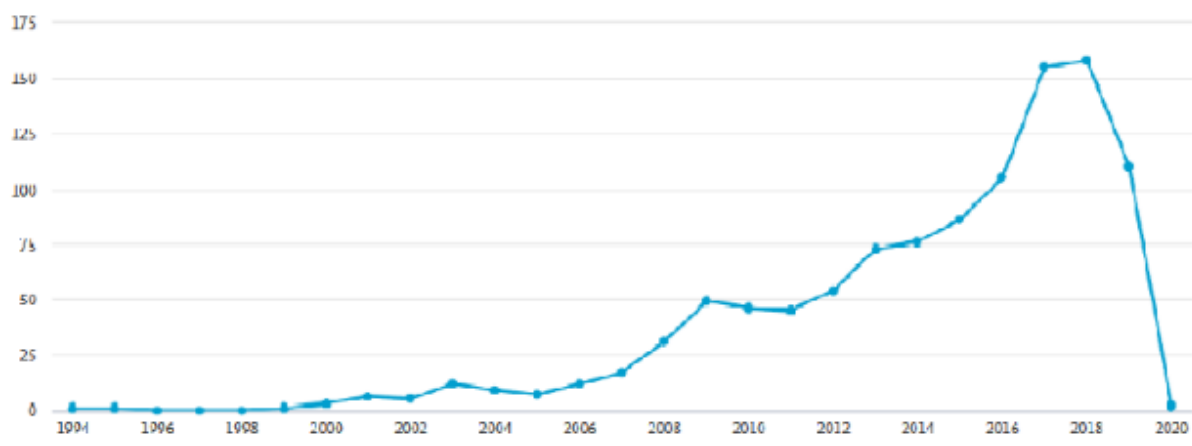
O refinamento da busca deu-se a partir das palavras chave "HRM", "*employee selection*", "*human resource*", "*management recruitment and selection*", "*work performance*" e "*organization reward*", juntamente com a exclusão de documentos do tipo editorial, *note*, *letter*, *book chapter*, *book*, *erratum*, *retracted*, e, ainda, por definição de escopo, limitando os resultados a artigos em português, inglês e espanhol, o que resultou em 1.064 artigos.

A pesquisa apresentou um aumento de publicações a partir de 2008, com crescimento acentuado e constante a partir de 2016 e ápice em 2018/19, o que denota que se trata de questão relevante e atual, conforme mostrado na Figura 1.

Para identificar os termos mais relevantes, a partir da importação dos dados da pesquisa, foi utilizado o *software* VOSviewer para *Co-ocurrence* onde o *Threshold* foi configurado para o mínimo de 10, gerando a representação gráfica apresentada na Figura 2.

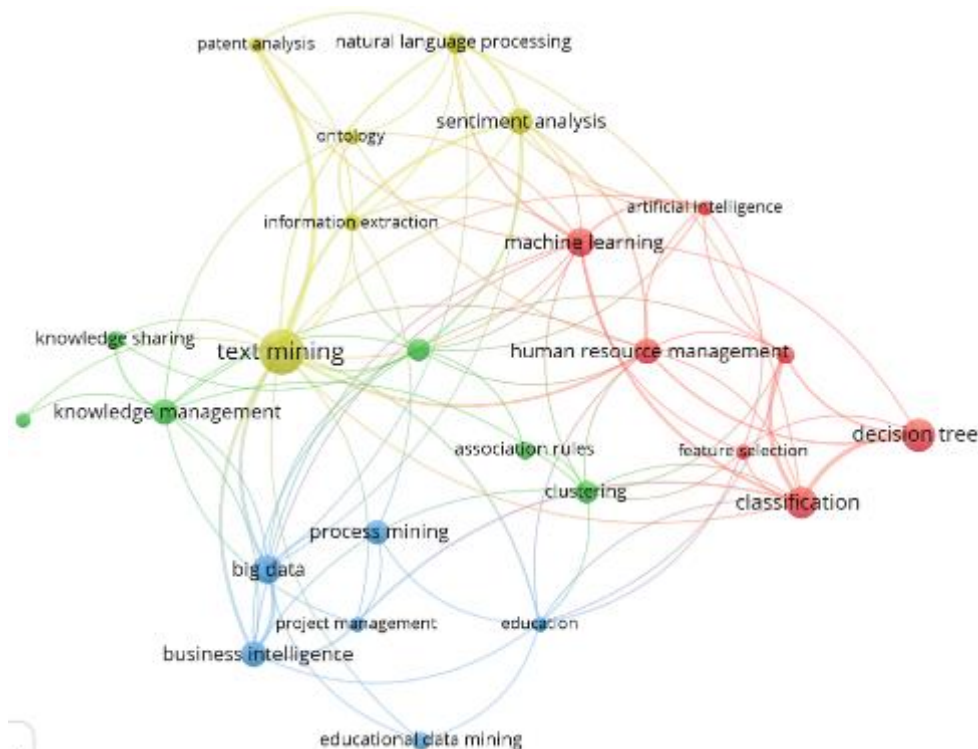
Do mapa de termos, destacam-se quatro agrupamentos relevantes. O verde, o menor deles, agrega temas relacionados à gestão e compartilhamento de conhecimento. O grupamento vermelho que se relaciona com o processo de seleção, agrega inteligência artificial e aprendizado de máquinas. Os temas agrupados em azul tendem à exploração de *big data*, inteligência de negócios e educação. O grupamento amarelo é o que reúne expressões relacionadas à mineração de textos e análise de sentimentos. Todos os agrupamentos apresentam assuntos relevantes para este trabalho.

**Figura 1.** Quantidade de publicações por ano



Fonte: a autora (2020).

**Figura 2.** Análise de coocorrência



Fonte: a autora (2020).

Dos 1.064 artigos recuperados na busca, são destacados a seguir alguns que figuram dentre os mais relevantes.

Strohmeier e Piazza (2008) consideram que além da mera automação de tarefas, um grande potencial dos sistemas de gestão de recursos humanos (SGRH) é informar a Gestão. O trabalho apresenta uma avaliação conceitual dos potenciais da mineração de dados para fornecer informações à Gestão de Pessoas, discutindo e avaliando a mineração de dados com uma nova abordagem com potenciais interessantes para gerar informações explicativas e prognósticas.

Em outro trabalho, Strohmeier e Piazza, (2012) afirmam que a GRH constitui um novo e notável campo da pesquisa de mineração de dados. No entanto, requisitos específicos, como avaliar o sucesso do domínio ou cumprir os padrões legais, frequentemente não são reconhecidos ou considerados nas pesquisas atuais.

Para Tai e Hsu (2006), o recrutamento é uma questão importante da GRH. O principal objetivo do recrutamento é obter talentos internos e externos da organização. O artigo trata de uma ferramenta de seleção de pessoal baseada no método de mineração de dados para ajudar os gerentes de negócios a encontrar talentos com mais eficiência.

Mendes (2013) aplica técnicas de Mineração de Dados para detectar padrões de comportamento relacionados à transferência de pessoas entre as áreas organizacionais (rotatividade interna), a fim de oferecer às organizações um modelo para avaliar o impacto dessas transferências sobre as suas políticas de gestão de pessoas.

Blasko, Rabcan e Vaclavkova (2017) abordam o recrutamento e a seleção de novos funcionários no setor segurança privada da República Eslovaca, propondo-se a resolver um dos problemas mais frequentes nas organizações de segurança privada: rotatividade.

Barroso et al. (2018) desenvolveram um estudo de caso usando técnicas de análise de dados pertinentes à gestão de pessoas em um processo de uma universidade federal, com mineração de texto para identificar automaticamente a correlação entre a atividade exercida, o cargo e o curso objeto de incentivo. A abordagem proposta contribui para a velocidade do processo de julgamento e aprimora a transparência das informações rastreando e divulgando todas as etapas.

Todos estes artigos forneceram insumos para

este trabalho.

### 3. Discussão

Pessoas não são vistas apenas como uma força de trabalho, mas são reconhecidas como um ativo valioso. Sendo assim, uma das preocupações da gestão de pessoas é reter os melhores talentos. Nesse contexto, merece atenção a rotatividade de pessoal (em inglês, *employee turnover*) que é a dinâmica de entrada e saída de empregados de uma empresa em um determinado período. Porém, a rotatividade também engloba as transferências de um indivíduo de uma função, ou de uma área, para outra função, ou ainda a transferência entre áreas dentro da mesma organização (Mendes, 2013).

O uso de mineração de dados contribui com o entendimento e qualificação da rotatividade interna de pessoas e, conseqüentemente, o seu impacto.

Mendes (2013) descreveu um modelo que aplica técnicas de agrupamento e regras de associação para descobrir conhecimento sobre transferências de funcionários entre as unidades de uma organização financeira brasileira, com mais de quatro mil unidades espalhadas por todo o país e mais de 95 mil funcionários.

Os achados mais relevantes denotaram que as transferências de homens da geração Y, com graduação, dentro do subsistema negocial são maioria, que há um grupo de transferências de homens sem graduação e com cinco anos de empresa e outro de mulheres da geração X que são promovidas geralmente para função de chefia. Pode-se dizer, com 90% de acerto, que empregados do subsistema negocial que não possuem experiência externa se transferem para unidades do próprio subsistema negocial. Concluindo que a rotatividade interna é um fenômeno de grande interesse que requer uma solução científica e tecnológica para o seu entendimento, de forma a embasar a tomada de decisão, principalmente no que se refere às políticas de promoção interna (Mendes, 2013).

A maioria dos processos organizacionais consome e gera dados, principalmente em formato textual. Por exemplo, um histórico escolar que reúne informações do currículo do aluno, poderá gerar um documento com a relação de disciplinas e os respectivos desempenhos, bem como, um processo eletrônico de ações judiciais que agrupa os atos e acumula informações conforme os documentos são inseridos. Da

mesma forma, um processo de gerenciamento de pessoas que lida com informações sobre e qualificação do funcionário também pode gerar dados a partir de pareceres técnicos (Barroso et al., 2018).

Uma Universidade Federal, ao realizar a avaliação de correlação entre o programa do curso objeto da solicitação da concessão de incentivo, o cargo e as atividades exercidas pelo empregado, adota processos com análise subjetiva.

É nesse cenário que se aplicam as técnicas de mineração de texto, que consiste em um processo de conhecimento no qual um usuário interage com uma coleção de documentos ao longo do tempo usando um conjunto de ferramentas de análise para identificar automaticamente a correlação através de abordagens de representação e modelagem sintática e semântica, que recuperam termos e dimensionam seus respectivos significados.

Para obter o grau de semelhança entre os respectivos documentos, foi utilizada a medida da similaridade de cossenos. Os resultados mostraram que os documentos avaliados como correlacionados pelo especialista apresentaram um grau de similaridade consistente com a avaliação automática. Para os casos não correlacionados, percebeu-se que o grau de alta similaridade foi influenciado pela abrangência do ambiente organizacional comum a todos os documentos (Barros et al., 2017).

A abordagem contribui para dar velocidade ao processo de julgamento, bem para promover formulações de críticas sobre o conteúdo das qualificações, aprimorando e conferindo maior transparência aos processos.

A capacidade de selecionar, atrair e reter os funcionários certos é crucial para o sucesso de qualquer organização. A seleção de candidatos adequados, com base em classificação de mineração de dados, é utilizada em muitas áreas, por exemplo, nos cuidados de saúde, *marketing*, finanças e muitos outros.

No campo da segurança privada há uma luta contra a rotatividade de pessoal, causada, dentre vários fatores, por baixas recompensas financeiras e alta responsabilidade. Além disso, os funcionários estão armados e são permanentemente envolvidos em várias situações de conflito, o que exige um profundo conhecimento especializado. Portanto, é importante executar corretamente o processo de seleção de novo funcionário e julgar minuciosamente o perfil

das competências e habilidades dos candidatos (Rabcan, Vaclavkova & Blasko, 2017).

O processo de seleção de candidatos a um tipo de cargo, em serviços de segurança, consiste em três fases básicas: a preliminar, contendo a especificação e descrição da posição do tipo; definição de adequação da funcionários e recrutamento. A fase de seleção, que inclui a aquisição de antecedentes de candidatos em potencial. E a fase de avaliação na qual o processo termina com a seleção dos candidatos mais adequados. A fase de avaliação do processo possui quatro etapas: avaliação da seleção; entrevista final seletiva; decisão de selecionar um proponente específico e a comunicação ao proponente da seleção (Rabcan, Vaclavkova & Blasko, 2017).

A ferramenta de decisão C4.5 foi aplicada nas etapas de seleção e avaliação do processo para encontrar os candidatos mais adequados. A vantagem das árvores de decisão está em suas interpretações e a capacidade de identificar os principais atributos do conjunto de dados. Porém, a decisão final deve ser sempre confirmada por um humano (Rabcan, Vaclavkova & Blasko, 2017).

O protótipo do sistema foi elaborado com base na árvore de decisão, de acordo com o algoritmo C4.5. Os dados dessa árvore consideraram o conjunto de dados público “Recursos Humanos Análise” do repositório Kaggle (<https://www.kaggle.com>). A aplicação permitiu avaliar a possibilidade de usar o sistema de tomada de decisão para currículo. A ferramenta apresentou precisão igual a 97,27%. O sistema e o método propostos podem ser usados não apenas em serviços de segurança, sendo necessária apenas a modificação dos requisitos e da coleta de dados iniciais para desenvolver um sistema semelhante em outras áreas (Rabcan, Vaclavkova & Blasko, 2017).

A gestão de recursos humanos constitui um bem estabelecido domínio corporativo para muitas informações e decisões de RH. Em consequência, qualquer nova solução de mineração de dados compete claramente com métodos convencionais e, portanto, somente serão aceitos em prática se mostrarem vantagens demonstráveis em eficácia (maior qualidade da informação e apoio à decisão) e / ou eficiência (menor esforço de oferecer informações apoio à decisão).

Atualmente há pouco conhecimento sobre os

dados de RH disponíveis de maneira que os dados podem constituir uma obstrução grave é até questionável se o tipo de dados disponíveis nos recursos humanos atuais Sistemas de Informação (HRIS) é adequado para suporte à decisão. Esses dados mapeiam principalmente políticas explícitas anteriores de RH e como consequência obrigatória, os padrões minerados apenas espelham decisões e processos de RH bem conhecidos do passado, e, portanto, têm um potencial bastante limitado para fornecer novas percepções e apoio à decisão (Strohmeier & Piazza, 2012).

Além disto, há uma dimensão ética e legal a ser considerada. A mineração de dados no domínio RH objetiva apoiar decisões que frequentemente afetam diretamente os seres humanos de maneiras positivas e negativas como, por exemplo, uma contratação. Esta decisão implica que um determinado indivíduo seja aceito ou recusado para a posição desejada. Além disso, dependendo do contexto nacional, existem regulamentos legais que, às vezes, apresentam sérias limitações à mineração de dados em domínio RH e, portanto, devem ser respeitadas (Strohmeier & Piazza, 2012).

Aspectos éticos e legais relevantes referem-se a proteger a privacidade ou a evitar a discriminação, que pode ser óbvia ou oculta. Se, por exemplo, uma regra para seleção de candidatos aconselha a não contratar aprendizes de uma determinada escola, isso pode não ser discriminatório se esta escola mostrar uma qualidade limitada de educação e, portanto, dos candidatos. No entanto, haverá discriminação se esta escola estiver em uma determinada área com população principalmente negra (Strohmeier & Piazza, 2012).

Portanto, os parâmetros devem ser transparentes e verificáveis, e, neste sentido, a aplicação de algoritmos de árvore de decisão apresenta padrões altamente compreensíveis, o que é bastante apropriado. No entanto, apesar da falta de transparência das Redes Neurais ainda são aplicados em muitos casos. Apenas uma minoria de publicações fornece informações sobre ajustes necessários ou mesmo o desenvolvimento de novos métodos de mineração de dados ajustados às peculiaridades da Gestão de pessoas (Strohmeier & Piazza, 2012).

Nesse contexto tem-se que pesquisas sobre mineração de dados orientada à gestão de pessoas precisa (Strohmeier & Piazza, 2012):

- comprovar a relevância do problema de RH;
- demonstrar a eficácia e eficiência dos dados de mineração em comparação com os métodos concorrentes (não mineração de dados);
- elaborar os dados necessários e sua provisão;
- avaliar os métodos de mineração existentes e, se necessário, personalizar os existentes ou desenvolver novos métodos;
- minimizar e dar suporte às tarefas relacionadas ao usuário final;
- incorporar a solução oferecida em específico ou geral HRIS;
- cumprir padrões éticos e regulamentações legais.

#### 4. Conclusão

Como já afirmado, pessoas não são apenas força de trabalho, são um ativo valioso para as organizações. Neste momento do mundo, em que a velocidade nas decisões toma relevância, estas baseiam-se cada vez mais em dados inteligentes e menos na experiência e intuição dos gestores. É neste contexto que a mineração de dados surge com soluções práticas para acelerar os processos e incrementar a assertividade das decisões relacionadas à gestão de pessoas.

Um ponto de atenção comum aos estudiosos deste assunto é a fonte e adequação dos dados que serão utilizados para mineração, pois, o resultado decorrente apenas de dados históricos de uma organização tende a exibir as políticas adotadas e não ofertará praticamente nenhum valor quanto a novas percepções para apoiar decisões.

Por outro lado, a mineração de dados traz à tona informações inéditas, antes indisponíveis, ou confere um grau de profundidade aos dados já conhecidos, porém que não se correlacionavam com outros. São estes fatores que levam aos necessários questionamentos éticos quanto as decisões que serão adotadas a partir da análise dos resultados da mineração de dados.

O cuidado com a transparência dos parâmetros e com a fonte dos dados é preponderante para prevenir vieses indesejados nos resultados e, por consequência, nas decisões. É unânime também entre os autores, a despeito do nível de assertividade que seja oferecida pelos algoritmos, que a decisão final é, necessária e criticamente, de um humano.

## Referências

- Barroso Jr, J., Cappelli, C., Revoredo, K., Nunes, V. (2018). Text Mining as a Transparency Enabler to Support Decision - Making in a People Management Process. *dg.o '18: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*.
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.
- Mendes, A. S. (2013). *Aplicação de Técnicas de Mineração de Dados na Caracterização da Rotatividade Interna de Pessoal* (Dissertação de mestrado em Engenharia Elétrica), Brasília: Universidade de Brasília.
- Rabcan, J., Vaclavkova, M., Blasko, R. (2017). Selection of Appropriate Candidates for a Type Position Using C4.5 Decision tree. *IDT: International Conference on Information and Digital Technologies*, p. 332-338.
- Strohmeier, S., Piazza, F. (2008). Informating HRM through Data Mining? A Conceptual Evaluation. *Proceedings of the 2nd International Workshop on Human Resource Information Systems*, p. 51-62.
- Strohmeier, S., Piazza, F. (2012). Domain driven data mining in human resource management: a review of current research. *Expert Systems With Applications*, 40(7), 2410-2420.
- Schwarz Müller, T., Brosi, P., Duman, D., Welpe, I. M. (2018). How Does the Digital Transformation Affect Organizations? Key Themes of Change in Work Design and Leadership. *Management Revue*, 29(2), 114-138.
- Tai, W.-S., Hsu, C.-C., (2006) A realistic personnel selection tool based on fuzzy data mining method. *JCIS: Proceedings of the 2006 Joint Conference on Information Sciences*.
- van Knippenberg, D., Dahlander, L., Haas, M. R., & George, G. (2015). Information, attention, and decision making. *Academy of Management Journal*, 58, 649-657.

## Sobre a autora

### **Valéria Simas Schultz**

Bacharel em Direito pelo Centro Universitário de Brasília (UNICEUB), Especialista em Gestão Estratégica do Setor Público pelo Centro Universitário do Distrito Federal (UNIDF) e Mestre em Governança, Tecnologia e Inovação pela Universidade Católica de Brasília (UCB). É funcionária da Caixa Econômica Federal desde 1999 e participa de Projeto de pesquisa, desenvolvimento e inovação para as áreas estratégicas da Procuradoria-Geral da Fazenda Nacional (PGFN).