



## AVALIAÇÃO DE DESEMPENHO SUSTENTÁVEL USANDO MINERAÇÃO DE DADOS: UMA ANÁLISE BIBLIOMÉTRICA

### SUSTAINABLE PERFORMANCE ASSESSMENT USING DATA MINING: A BIBLIOMETRICA ANALYSIS

Jovani Taveira de Souza, Universidade Tecnológica Federal do Paraná, Brasil, jovanisouza5@gmail.com  
Antonio Carlos de Francisco, Universidade Tecnológica Federal do Paraná, Brasil, acfrancisco@utfpr.edu.br

#### Resumo

O rápido crescimento da população e dos processos que envolvem indústrias trouxe consigo alto consumo de recursos naturais e efeitos negativos sobre o meio ambiente. Entre tais efeitos, destaca-se o uso e consumo de energia, que é o principal contribuinte para a mudança climática global. No entanto, existem dificuldades em medir e avaliar o desempenho sustentável das empresas. Em alguns casos, a mineração de dados é usada pois promete extrair informações potencialmente úteis em bancos de dados. Portanto, o objetivo deste trabalho foi realizar uma análise bibliométrica sobre o uso da análise de dados por meio da mineração de dados em estudos que trabalhem com avaliação de desempenho sustentável. Utilizou-se o método de revisão sistemática denominado Methodi Ordinatio e os softwares: EndNote®, Jabref® e Microsoft Excel®. Os dez artigos com o maior nível de importância encontrados na literatura seguindo o procedimento metodológico adotado foram apresentados, bem como os métodos aplicados pelos autores, enfocando principalmente na abordagem de avaliação. Nesta perspectiva, este trabalho promove o incentivo e a motivação para que as empresas gerenciem adequadamente seu desempenho em um contexto sustentável, através de ferramentas computacionais.

**Palavras-chave:** Análise bibliométrica; Mineração de dados; Desempenho Sustentável

#### Abstract

The rapid growth of population and processes involving industries has brought with it high consumption of natural resources and negative effects on the environment. Among these effects, we highlight the use and consumption of energy, which is the main contributor to global climate change. However, there are difficulties in measuring and evaluating the sustainable performance of companies. In some cases, data mining is used because it promises to extract potentially useful information in databases. Therefore, the objective of this work was to perform a bibliometric analysis on the use of data analysis through data mining in studies that work with sustainable performance evaluation. The systematic review method called Methodi Ordinatio and the software EndNote®, Jabref® and Microsoft Excel® were used. The ten articles with the highest level of importance found in the literature following the methodological procedure adopted were presented, as well as the methods applied by the authors, focusing mainly on the evaluation approach. In this perspective, this work promotes the incentive and motivation for companies to adequately manage their performance in a sustainable context through computational tools.

**Keywords:** Bibliometric analysis; Data Mining; Sustainable performance.



## 1. INTRODUÇÃO

O desenvolvimento sustentável tornou-se um campo de pesquisa muito conhecido na última década, devido a mudanças sociais, deterioração ambiental e acompanhamento do interesse público, sendo um tema em destaque para a comunidade científica e acadêmica (BÜYÜKÖZKAN; KARABULUT, 2018). Porém, analisar o desempenho e a sustentabilidade de uma determinada empresa é um desafio (SCHÖGGL; FRITZ; BAUMGARTNER, 2016). Muitas vezes, é necessário a utilização de métodos e ferramentas capazes de avaliar, mensurar e facilitar o processo de avaliação do desempenho sustentável dessas empresas, devido a presença de grande quantidade de informações alocadas em bases de dados.

De acordo com Büyüközkan e Karabulut (2018), o desempenho sustentável é resultado agregado negativo ou positivo dos impactos econômicos, sociais e ambientais, a partir de uma base definida. Segundo os autores, não se pode focar apenas no ambiente e ignorar as outras dimensões. Portanto, é necessário trabalhar com as diferentes perspectivas neste âmbito.

Uma das abordagens que podem ser utilizadas para análise de desempenho sustentável é através das técnicas de mineração de dados, que segundo Kamber, Han e Pei (2012), encontram padrões, regras e dados ocultos em grandes quantidades de dados.

Segundo Zhong et al. (2015), para a análise de um conjunto de dados, é preciso extrair dados brutos das várias fontes, executar um trabalho de limpeza e unificar as informações em apenas uma base. Em relação a qualidade das informações e análise final, métodos estatísticos, aprendizado de máquina e mineração de dados são utilizados, pois fornecem resultados mais precisos, auxiliando no processo de tomada de decisão.

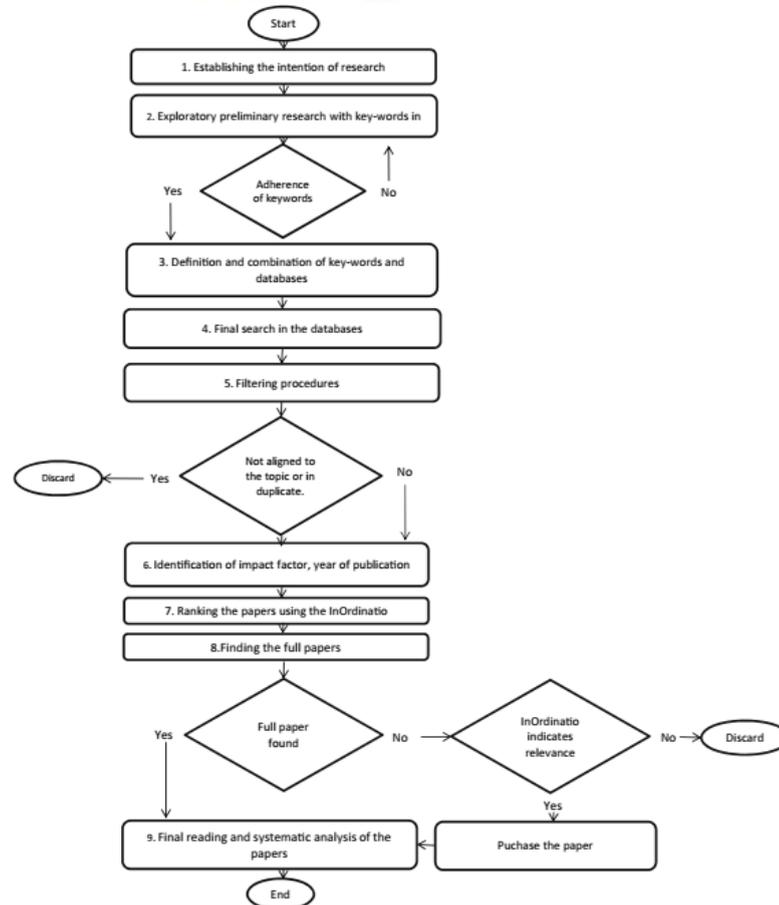
Schuch et al. (2010), afirmam que as técnicas de mineração de dados contribuem significativamente em empresas e organizações, devido produzir conhecimento, identificando medidas significativas que possam auxiliar em nível estratégico qualquer empresa.

Diante da importância de identificar técnicas que auxiliem nesse processo, este artigo tem como objetivo apresentar uma análise bibliométrica de artigos de revistas altamente conceituadas listados na categoria *Web of Science* (Coleção Principal), com o intuito de revisar a literatura sobre o uso da análise de dados por mineração de dados em estudos que trabalhem ou que corroborem com a avaliação de desempenho sustentável.

Este trabalho está organizado da seguinte maneira: a seção 2 apresenta a metodologia da pesquisa. Os resultados e discussões são documentadas na seção 3 e a seção 4 apresenta a conclusão do artigo.

## 2. METODOLOGIA

Este artigo apresenta uma análise bibliométrica que busca ampliar o conhecimento sobre a temática em questão. Para isso, foi realizada uma revisão de literatura estrutura de acordo com o *Methodi Ordinatio* (PAGANI; KOVALESKI; RESENDE, 2015). A metodologia é constituída de nove etapas, conforme Figura 1.



Fonte: Pagani; Kovaleski e Resende (2015)

Figura 1 – Etapas da metodologia

As combinações de palavras utilizadas para a construção da análise bibliométrica foram: “*data mining*” e “*sustainability indicators*” ou “*indices*”; “*data mining*” e “*sustainable consumption*”; “*data mining*” e “*sustainable performance*”. Os artigos foram selecionados da base *Web of Science* (Coleção Principal).

A coleção principal da *Web of Science* foi escolhida por sua abrangência, ser multidisciplinar e indexar somente os periódicos mais citados em suas respectivas áreas. Atualmente existem mais de 12.000 periódicos de alto impacto indexados (CAPES, 2018).

Os softwares escolhidos para gerenciar o material encontrado nas bases de dados foram o EndNote®, Jabref® e Microsoft Excel®. O EndNote® foi utilizado para examinar e organizar os artigos encontrados, como a exclusão de artigos duplicados. Para manipulação dos dados foi utilizado o software JabRef® que permite a importação de vários formatos, cria e customiza formatos de exportação e possibilita traduções automáticas. Os dados organizados foram então exportados para um arquivo do Microsoft Excel®. Por fim, foi utilizado o Excel para aplicação da fórmula do Methodi Ordinatio.

Com isso, foi possível definir os principais autores, periódicos, além de elencar os artigos mais citados sobre o tema.



### 3. RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados após a análise bibliométrica. Serão apresentados os 10 artigos com o maior nível de importância dos 36 que foram encontrados seguindo o procedimento da Figura 1 (Etapa 1 a 9). O Quadro 1 apresenta detalhes dos artigos encontrados.

Título	Autores	Título do periódico	Ano
Global water trends and future scenarios for sustainable development: The case of Russia	Proskuryakova; Saritas; Sivaev	Journal of Cleaner Production	2018
Monitoring soil lead and zinc contents via combination of spectroscopy with extreme learning machine and other data mining methods	Khosravi et al.	Geoderma	2018
Identification of key energy efficiency drivers through global city benchmarking: A data driven approach	Wang et al.	Applied Energy	2017
Determinants and development of a web-based child mortality prediction model in resource-limited settings: A data mining approach	Tesfaye et al.	Computer Methods and Programs in Biomedicine	2017
Decision-making framework with double-loop learning through interpretable black-box machine learning models	Bohanec; Robnik-Sikonja; Borstnar	Industrial Management & Data Systems	2017
Identifying services for short-term load forecasting using data driven models in a Smart City platform	Massana et al.	Sustainable Cities and Society	2017
A differential multi-criteria analysis for the assessment of sustainability performance of European countries: Beyond country ranking	Antanasijevic et al.	Journal of Cleaner Production	2017
The dynamics of public safety in cities: A case study of Shanghai from 2010 to 2025	Yu and Fang	Habitat International	2017
Understanding the effects of environmental factors on building energy efficiency designs and credits Case studies using data mining and real-time data	Kim et al.	Journal of Engineering Design and Technology	2017
Intelligent data analysis for sustainable smart grids using hybrid classification by genetic algorithm based discretization	Panda	Intelligent Decision Technologies-Netherlands	2017

Quadro 1 – Resultados da análise bibliométrica



Após a seleção dos artigos a partir da análise bibliométrica, foram discutidas as principais contribuições dos trabalhos encontrados, evidenciando principalmente nas técnicas de mineração de dados.

O estudo de Proskuryakova, Saritas e Sivaev (2018) realizado na Rússia apresentou o desenvolvimento de cenários alternativos de apoio e decisão, com o intuito de mitigar as consequências indesejadas que podem ocorrer no setor de recursos hídricos. Para isso, uma variedade de métodos foi utilizada, incluindo a mineração de dados, que participou na escolha das principais vertentes (áreas prioritárias) relacionadas ao setor, que são: irrigação, qualidade da água, água doce, gestão da água, abastecimento de água, seca, precipitação, produtividade, entre outros. A análise foi feita em uma amostra de mais de 10 milhões de publicações de textos completos, 200 mil relatórios e 1,8 milhões de trabalhos de pesquisa de acesso aberto.

O estudo realizado por Khosravi et al. (2018) investigou o controle de chumbo (Pb) e zinco (Zn) no solo na maior mina de cobre do Irã, com o objetivo de limitar o risco de poluição e desenvolver estratégias adequadas de remediação, a partir de monitoramento rápido e sustentável. Para este fim, foi aplicado um novo algoritmo de mineração de dados, denominado de *Extreme Learning Machine* (ELM), que foi avaliado e comparado com outras técnicas frequentemente usadas para prever concentrações de Pb e Zn, as quais sejam: *back-propagation neural networks* (BPNN), *support vector machine* (SVM) e *partial least squares regression* (PLSR). Os resultados apontaram que o ELM apresentou o melhor desempenho de previsão em relação aos outros métodos, devido a resultados mais precisos, menor tempo de execução e menor complexidade.

O trabalho feito por Wang et al. (2017) retrata um estudo sistemático para identificação dos principais impulsionadores em termos de eficiência energética urbana. Os dados de 25 cidades localizadas em três continentes (Ásia, Europa e América) foram coletados do relatório *Global Power City Index* (GPCI), além de outras fontes de dados. Posteriormente, foram realizadas análises de dados pelas seguintes técnicas: análise envoltória de dados, *clustering* e árvore de decisão, com a finalidade de calcular e comparar a eficiência energética dessas cidades. Como resultado observou-se que as técnicas utilizadas permitiram avaliar de forma adequada os diferentes objetivos do trabalho e, também, encontrar abordagens alternativas para agrupamentos.

Destaca-se também a pesquisa desenvolvida por Tesfaye et al. (2017) que trabalhou com mineração de dados, mais precisamente as técnicas de árvore de decisão (algoritmo J48) e de indução de regra (algoritmo PART), em dados demográficos e de pesquisa de saúde da Etiópia. O objetivo foi apresentar as áreas determinantes relacionados a mortalidade infantil, além de desenvolver um modelo de previsão que é acessado através da internet na língua local etíope.

A proposta feita por Bohanec, Robnik-Sikonja e Borstnar (2017) teve como propósito relatar o problema da fraca aceitação dos modelos de aprendizado de máquina (*machine learning* - ML) em diferentes negócios. Para tal, os autores estruturam um framework integrado a modelos de ML. O framework proposto foi aplicado em um caso do mundo real, onde foi apresentado todos os passos, resultados e cuidados para sua aplicação. Os resultados experimentais mostraram melhora no desempenho do negócio, o que corrobora na redução de erros aos tomadores de decisão e cria uma base para o aprendizado organizacional sustentável.



O trabalho de Massana et al. (2017) abrange a incorporação de vários serviços em uma arquitetura *Smart City*, em particular a eletricidade, identificando serviços que podem ajudar a melhorar a eficiência energética em infraestruturas urbanas. Um caso de estudo foi relatado para previsão de carga de curto prazo em edifícios não residenciais. Para isso, o processo contemplou as etapas de seleção de dados, pré-processamento, seleção dos modelos, processo de treinamento, avaliação de modelos e exploração dos resultados, utilizando-se o modelo autorregressivo (AR) para as previsões.

O estudo desenvolvido por Antanasijevic et al. (2017) fornece uma análise do desempenho sustentável no contexto da Estratégia de Desenvolvimento Sustentável da União Europeia (EU *Sustainable Development Strategy*). Os indicadores de desenvolvimento sustentável foram analisados usando métodos de mineração de dados multivariadas: PROMETHEE e DMCA (*differential multi criteria analysis*), a fim de quantificar as diferenças entre os países quanto aos progressos em cada tema estudado, bem como o progresso geral no período abrangido, e avaliar os progressos na redução do desenvolvimento desigual entre os membros da União Europeia.

Yu and Fang (2017) investigaram e simularam a dinâmica de segurança pública urbana de Xangai, uma das maiores cidades da China, a partir de um conjunto de indicadores coletados que descrevem a infraestrutura, desenvolvimento, população, crime, habitabilidade e desastre durante a última década (2000-2009) em Xangai. Foram utilizadas técnicas de mineração de dados e análises estatísticas, tais como: regressão multivariada e ajuste de curvas, para construir circuitos de feedback do sistema entre eles, para imitar o circuito real da segurança pública da cidade.

No estudo de Kim et al. (2017) foi explorado a relação entre fatores ambientais e a construção de consumo de energia de três edifícios certificados pela Liderança em Energia e Design Ambiental (LEED) na Universidade Estadual do Arizona, estabelecendo as relações entre a temperatura atmosférica externa e a energia consumida no prédio usando dados em tempo real. A análise de cluster foi utilizada para calibrar e eliminar influências ou fatores indesejados de um conjunto de dados em tempo real de consumo de edifício. Posteriormente, foi utilizado o qui-quadrado para verificação dos resultados.

O trabalho de Panda (2017) discutiu algumas das técnicas de mineração de dados em conjunto de dados de agregação de energia disponíveis publicamente. Foi apresentado as informações referentes ao comportamento dos consumidores em termos de consumo de eletricidade. As técnicas aplicadas foram: *radial basis function (RBF) kernel*, *best first tree (BFTree)* e *ordered weighted average fuzzy-rough K-nearest neighbor (OWAKNN)* com *equal width (EWD)* e *Genetic algorithm based Discretization (GAD)*. Ambas as técnicas foram utilizadas para prever o comportamento do consumidor no consumo de eletricidade para a sustentabilidade.

Contextualizando ambos os trabalhos apresentados, pode-se enfatizar a importância das técnicas de mineração de dados em ambos os contextos. As técnicas foram fatores fundamentais nos estudos, principalmente nas questões de previsões, avaliações, filtros de indicadores, cálculos de eficiência, dentre outros. É fundamental, antes de tudo, que a escolha do método seja escolhida criteriosamente, pois a qualidade das informações vai depender da assertividade com relação a essa questão.



Por fim, serão apresentadas as principais conclusões do trabalho realizado.

#### 4. CONCLUSÃO

Este artigo investigou a partir de uma análise bibliométrica, trabalhos em destaque na literatura, mais especificamente da coleção *Web of Science* (Coleção Principal), voltados a análise de desempenho sustentável por meio de técnicas de mineração de dados.

Diante do estudo realizado, foi possível verificar que existem diferentes maneiras de se utilizar as técnicas, como: calcular a eficiência energética, fazer previsões, filtrar os atributos essenciais, avaliar progressos, prever comportamentos, dentre outros aspectos. Porém, é preciso que a escolha do método seja assertiva, para que assim possa auxiliar na análise final, garantindo a qualidade das informações e, também, corroborando para o processo de tomada de decisão.

O estudo pode demonstrar em uma análise rápida de dez artigos selecionados, que é essencial a utilização de uma metodologia para avaliar de forma mais confiável e compressível as diferentes problemáticas que são encontradas na avaliação de desempenho sustentável, indiferentemente do setor estudado.

Neste contexto, este trabalho incentiva empresas e organizações, através de alguns exemplos relatados, a gerenciarem adequadamente seu desempenho em um contexto sustentável, através de ferramentas computacionais (técnicas de mineração de dados).

Sugere-se, para estudos futuros, discutir de maneira mais profunda as técnicas de mineração de dados empregados nos artigos listados.

#### REFERÊNCIAS

- Antanasijević, D., Pocajt, V., Ristić, M., Perić-grujić, A. (2017). A differential multi-criteria analysis for the assessment of sustainability performance of European countries: Beyond country ranking, *Journal of Cleaner Production*, 165, 213-220.
- Bohanec, M., Robnik-šikonja, M., Kljajić Borštnar, M. (2017). Decision-making framework with double-loop learning through interpretable black-box machine learning models, *Industrial Management & Data Systems*, 117, 7, 1389-1406.
- Büyüközkan, G., Karabulut, Y. (2018). Sustainability performance evaluation: Literature review and future directions, *Journal of environmental management*, 217, 253-267.
- Capes. Disponível em: < <http://www.capes.gov.br/>>. Acesso em: 27 fev. 2018.
- Kamber, M., Han, J., Pei, J. (2012). *Data mining: Concepts and techniques*, Elsevier.
- Khosravi, V., Ardejani, F. D., Yousefi, S., Aryafar, A. (2018). Monitoring soil lead and zinc contents via combination of spectroscopy with extreme learning machine and other data mining methods, *Geoderma*, 318, 29-41.
- Kim, J., Hyun, J-Y., Chong, W. K., Ariaratnam, S. (2017). Understanding the effects of environmental factors on building energy efficiency designs and credits: Case studies using data mining and real-time data, *Journal of Engineering, Design and Technology*, 15, 3, 270-285.



- Massana, J., Pous, C., Burgas, L., Melendez, J., Colomer, J. (2017). Identifying services for short-term load forecasting using data driven models in a Smart City platform, *Sustainable cities and society*, 28, 108-117.
- Pagani, R. N., Kovaleski, J. L., Resende, L. M. (2015). Methodi Ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication. *Scientometrics*, 105, 3, 2109-2135.
- Panda, M. (2017). Intelligent data analysis for sustainable smart grids using hybrid classification by genetic algorithm based discretization, *Intelligent Decision Technologies*, 11, 2, 137-151.
- Proskuryakova, L. N., Saritas, O., Sivaev, S. (2018). Global water trends and future scenarios for sustainable development: The case of Russia, *Journal of Cleaner Production*, 170, 867-879.
- Schöggel, J-P., Fritz, M. M., Baumgartner, R. J. (2016). Toward supply chain-wide sustainability assessment: A conceptual framework and an aggregation method to assess supply chain performance, *Journal of Cleaner Production*, 131, 822-835.
- Schuch, R., Dill, S. L., Suasen, P. S., Padoin, E. L., de Campos, M. (2010). Mineração de dados em uma subestação de energia elétrica. In: Proceedings of the 9<sup>th</sup> Brazilian Conference on Dynamics, Control and Their Applications – dincon'10. Anais... Serra Negra, 804-810.
- Tesfaye, B., Atique, S., Elias, N., Dibaba, L., Shabbir, S-A., Kebede, M. K. (2017). Determinants and development of a web-based child mortality prediction model in resource-limited settings: A data mining approach. *Computer methods and programs in biomedicine*, 140, 45-51.
- Wang, X., Li, Z., Meng, H., Wu, J. (2017). Identification of key energy efficiency drivers through global city benchmarking: A data driven approach, *Applied energy*, 190, 18-28.
- Yu, D., Fang, C. (2017). The dynamics of public safety in cities: A case study of Shanghai from 2010 to 2025, *Habitat International*, 69, 104-113.
- Zhong, M., Schuetter, J., Mishra, S., & Lafollette, R. F. (2015). Do data mining methods matter?: A Wolfcamp Shale case study. In SPE Hydraulic Fracturing Technology Conference, *Society of Petroleum Engineers*.