

UNIVERSIDADE FUMEC
FACULDADE DE CIÊNCIAS EMPRESARIAIS – FACE
CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Tiago Silveira Gontijo

***DATA SCIENCE* E ENGENHARIA DE PRODUÇÃO:
UMA PESQUISA BIBLIOMÉTRICA SOBRE AS PUBLICAÇÕES
CIENTÍFICAS REALIZADAS ENTRE 1987 E 2022**

Belo Horizonte, MG
2022

Tiago Silveira Gontijo

DATA SCIENCE E ENGENHARIA DE PRODUÇÃO:
UMA PESQUISA BIBLIOMÉTRICA SOBRE AS PUBLICAÇÕES CIENTÍFICAS
REALIZADAS ENTRE 1987 E 2022

Trabalho de conclusão de curso, apresentado ao Curso de Engenharia de Produção, da Universidade FUMEC – Campus Cruzeiro, como requisito parcial para obtenção do título de **Engenheiro de Produção**.

Orientador: Prof. Dr. Thiago Henrique Martins Pereira

Belo Horizonte, MG
2022

Tiago Silveira Gontijo

***DATA SCIENCE E ENGENHARIA DE PRODUÇÃO:
UMA PESQUISA BIBLIOMÉTRICA SOBRE AS PUBLICAÇÕES CIENTÍFICAS
REALIZADAS ENTRE 1987 E 2022***

Trabalho de conclusão de curso, apresentado ao Curso de Engenharia de Produção, da Universidade FUMEC – Campus Cruzeiro, como requisito parcial para obtenção do título de **Engenheiro de Produção**.

Aprovado em 03 de junho de 2022:

Thiago Henrique Martins Pereira (FUMEC)
(Presidente/Orientador)

Rodrigo Barbosa de Santis (PPGEP-UFMG)

Belo Horizonte, MG
2022



FACULDADE DE CIÊNCIAS EMPRESARIAIS DA
UNIVERSIDADE FUMEC

Belo Horizonte, 03 de junho de 2022.

ATA DA DEFESA DE TTC – TRABALHO DE CONCLUSÃO DE CURSO DO
CURSO

Às **14h30**, no dia **03 de junho** de 2022, instalou-se a comissão indicada pelo professor orientador do trabalho de conclusão de curso, Dr. Thiago Henrique Martins Pereira para a avaliação final do TCC do aluno **Tiago Silveira Gontijo** da Faculdade de Ciências Empresariais da Universidade FUMEC. O trabalho apresentado corresponde ao requisito final para obtenção do Grau Acadêmico de Bacharel (a) em Engenharia de Produção, sob o título de: “Engenheiro de Produção”. O Senhor Presidente da Comissão, Orientadora, Prof. Dr. Thiago Henrique Martins Pereira, formalizou a abertura da sessão pública, deu conhecimento aos presentes do regulamento protocolar da sessão, passando a palavra ao candidato. Após a apresentação do trabalho pelo candidato, passou-se a palavra aos membros da Comissão Examinadora e procederam-se as arguições e argumentações. Ao término da arguição a Comissão se reuniu, sem as presenças do candidato e do público, para julgamento e expedição do parecer final conclusivo, sendo comunicado publicamente o resultado pela senhora presidente de:


1. Trabalho reprovado - Nota:	()
2. Analisar e reenviar para nova revisão. Prazo: dias	()
3. Analisar e enviar em versão final para a secretaria. Nota:	()
4. Aprovado como está, enviar para Secretaria em versão final. Nota: 60	(X)

Nada mais havendo a tratar, o(a) senhor(a) presidente encerrou a sessão e lavrou a presente ATA, assinada, nesse ato, por todos os membros da Comissão Examinadora.

Comissão Examinadora

Thiago Henrique Martins Pereira
Assinado de forma digital por
Thiago Henrique Martins Pereira
Dados: 2022.06.04 11:52:42 -03'00'

Prof. Dr. Thiago Henrique Martins Pereira – Orientador FUMEC;


Rodrigo Barbosa de Santis (PPGEP-UFMG)

Saudade dos tempos de outrora, que não voltam mais.

$2 = 3 - 2x$
 $4x + 5 = 2x + 3$
 $2x = -2$
 $x = -1$

$\lim_{x \rightarrow 2} \frac{3x^2 + 4x + 3}{x^2 + 1} = \frac{8 + 4 + 3}{2 + 1} = \frac{15}{3} = 5$

$14x + 5 = 2x + 3$
 $12x = -2$
 $x = -\frac{1}{6}$

$3x^2 = 27$
 $x^2 = 9$
 $x = \pm 3$

$2x + 5 = x - 0.3$
 $x = -5.3$

$7 > 2x$
 $x < 3.5$

$7 > -2x$
 $x > -3.5$

$0 < x < 3.5$

$13x + 4 = 8x - 5$
 $5x = -9$
 $x = -1.8$

$13x + 4 = 4|2x - 3|$
 $13x + 4 = 8x - 12$
 $5x = -16$
 $x = -3.2$

$16 + 6 = 14$
 22

$12x - 5 < 3$
 $12x < 8$
 $x < \frac{2}{3}$

$3x - 2 \leq 10$
 $3x \leq 12$
 $x \leq 4$

$4 < 3x - 2$
 $6 < 3x$
 $2 < x$

$4 < 3x - 2 \leq 10$
 $2 < x \leq 4$

$12x - 5 < 3$
 $12x < 8$
 $x < \frac{2}{3}$

$\int \frac{1}{x^2} dx = -\frac{1}{x} + C$

$\int \frac{1}{x^3} dx = -\frac{1}{2x^2} + C$

$\int \frac{1}{x^4} dx = -\frac{1}{3x^3} + C$

$\int \frac{1}{x^5} dx = -\frac{1}{4x^4} + C$

$\int \frac{1}{x^6} dx = -\frac{1}{5x^5} + C$

$\int \frac{1}{x^7} dx = -\frac{1}{6x^6} + C$

$\int \frac{1}{x^8} dx = -\frac{1}{7x^7} + C$

$\int \frac{1}{x^9} dx = -\frac{1}{8x^8} + C$

$\int \frac{1}{x^{10}} dx = -\frac{1}{9x^9} + C$

$\int \frac{1}{x^{11}} dx = -\frac{1}{10x^{10}} + C$

$\int \frac{1}{x^{12}} dx = -\frac{1}{11x^{11}} + C$

$\int \frac{1}{x^{13}} dx = -\frac{1}{12x^{12}} + C$

$\int \frac{1}{x^{14}} dx = -\frac{1}{13x^{13}} + C$

$\int \frac{1}{x^{15}} dx = -\frac{1}{14x^{14}} + C$

$\int \frac{1}{x^{16}} dx = -\frac{1}{15x^{15}} + C$

$\int \frac{1}{x^{17}} dx = -\frac{1}{16x^{16}} + C$

$\int \frac{1}{x^{18}} dx = -\frac{1}{17x^{17}} + C$

$\int \frac{1}{x^{19}} dx = -\frac{1}{18x^{18}} + C$

$\int \frac{1}{x^{20}} dx = -\frac{1}{19x^{19}} + C$

**DATA SCIENCE E ENGENHARIA DE PRODUÇÃO:
UMA PESQUISA BIBLIOMÉTRICA SOBRE AS PUBLICAÇÕES CIENTÍFICAS
REALIZADAS ENTRE 1987 E 2022**

AUTOR: Tiago Silveira Gontijo

ORIENTADOR: Dr. Thiago Henrique Martins Pereira

Publicações científicas sobre *Data Science* e Engenharia de Produção tem despertado a atenção da literatura, dos governos locais e do setor empresarial. Desta forma esta pesquisa realizou um estudo bibliométrico das publicações indexadas junto a *Web of Science*, que adereçaram o tema proposto, na janela temporal compreendendo os anos de 1987 a 2022. Verificou-se que as publicações sobre o tema cresceram a uma taxa anual da ordem de 8,35%. Alguns países notabilizaram-se como aqueles que mais publicam sobre o assunto. São eles: China, Estados Unidos e Alemanha. Juntos, estes países respondem por aproximadamente 44% das publicações sobre o tema. Em acréscimo, destaca-se que o Brasil ocupa a 15ª posição no *ranking* global de nacionalidade dos países que mais publicam sobre este assunto, logo é necessário fomentar a realização de novas pesquisas locais. Conforme apontado por esta pesquisa, o *journal Computers & Industrial Engineering* foi aquele que possui mais publicações sobre o assunto. Esta pesquisa verificou ainda que os professores George Belgiu, Cristian Gheorghe Turc e Constantin Carausu, são aqueles que mais publicaram sobre o tema proposto. Por outro lado, os professores Dirk Schaefer e Abdulazeez Abdulraheem notabilizaram-se como os principais pesquisadores em termos de citações, ou seja, de impacto científico. Esta pesquisa verificou ainda que as principais *keywords* utilizadas na atualidade em pesquisas sobre *Data Science* e Engenharia de Produção, são: *Cyber-Physical Systems*, *Education*, *Feature Extraction* e *Machine Learning*. Destaca-se que o uso destas expressões passou a ser frequente após o ano de 2020, logo, estas são algumas das áreas quentes sobre a temática proposta e, portanto, merecem atenção especial.

Palavras-chave: Bibliometria. *Big Data*. *Data Science*. Engenharia de Produção. *Industry 4.0*.

ABSTRACT**DATA SCIENCE AND INDUSTRIAL ENGINEERING: A BIBLIOMETRIC RESEARCH ON SCIENTIFIC PUBLICATIONS CARRIED OUT BETWEEN 1987 AND 2022**

AUTHOR: Tiago Silveira Gontijo
ADVISOR: Dr. Thiago Henrique Martins Pereira

Scientific publications on Data Science and Production Engineering have attracted the attention of the literature, local governments and the business sector. In this way, this research carried out a bibliometric study of the publications indexed by the Web of Science, which addressed the proposed theme, in the time window comprising the years 1987 to 2022. It was found that publications on the subject grew at an annual rate of the order of 8.35%. Some countries have become notable as those that publish the most on the subject. They are: China, United States and Germany. Together, these countries account for approximately 44% of publications on the topic. In addition, it is noteworthy that Brazil occupies the 15th position in the global ranking of nationality of the countries that publish the most on this subject, so it is necessary to encourage the conduction of new local research. As pointed out by this research, the journal *Computers & Industrial Engineering* was the one with the most publications on the subject. This research also verified that professors George Belgiu, Cristian Gheorghe Turc and Constantin Carausu are the ones who published the most on the proposed topic. On the other hand, professors Dirk Schaefer and Abdulazeez Abdulraheem were noted as the main researchers in terms of citations, that is, scientific impact. This research also verified that the main keywords currently used in research on Data Science and Production Engineering are: Cyber-Physical Systems, Education, Feature Extraction and Machine Learning. It is noteworthy that the use of these expressions became frequent after the year 2020, logos, these are some of the hot areas on the proposed theme and, therefore, deserve special attention.

Keywords: Bibliometrics. Big Data. Data Science. Industrial Engineering. Industry 4.0.

LISTA DE FIGURAS

Figura 1 - Evolução anual das publicações sobre Data Science e Engenharia de Produção	5
Figura 2 - Principais países que publicaram sobre o tema, com base no autor correspondente	6
Figura 3 - Principais autores mundiais sobre <i>Data Science</i> e Engenharia de Produção	8
Figura 4 - Evolução temporal das palavras-chaves das publicações sobre <i>Data Science</i> e Engenharia de Produção	10

LISTA DE QUADROS

Quadro 1 - Descrição dos processos metodológicos adotados.....	3
Quadro 2 - Os cinco artigos globais sobre <i>Data Science</i> e Engenharia de Produção que mais vezes foram citados	9

LISTA DE TABELAS

Tabela 1 - Estatísticas descritivas das publicações sobre <i>Data Science</i>	4
Tabela 2 - Principais fontes de publicação sobre <i>Data Science</i>	7

SUMÁRIO

1.INTRODUÇÃO	1
1.1 OBJETIVO GERAL	2
1.2 OBJETIVOS ESPECÍFICOS	2
2. MATERIAIS E MÉTODOS	3
3. RESULTADOS E DISCUSSÃO	4
4. CONCLUSÕES.....	11
REFERÊNCIAS	13

1. INTRODUÇÃO

Com o advento da Indústria 4.0 e a crescente evolução computacional associada a ela, diversos temas de pesquisa tem despertado a atenção de pesquisadores nacionais e internacionais, como por exemplo, pode-se citar: (i) *Big Data* (SAGIROGLU e SINANC, 2013); (ii) *Internet of Things* (WEBER e WEBER, 2010); (iii) *Cloud Computing* (SRIVASTAVA e KHAN, 2018), etc. Neste contexto, dentre os principais vetores de interesse acadêmico e profissional, destaca-se a *Data Science*, ou a “Ciência dos Dados”, que aglutina diversos conhecimentos, que vão desde a computação, estatística e matemática (CAO, 2017; GRUS, 2019; SARKER, 2021; EGGER e YU, 2022).

A relevância da Ciência dos Dados extrapola o universo acadêmico e já uma realidade dentro de grandes organizações. Como prova, cita-se os requisitos de processos seletivos, tanto para estagiários, quanto para *trainees*, que passaram a exigir recentemente, domínio de ferramentas como linguagens de programação (Python e R) e demais ferramentas como o *Microsoft Power BI* e o *Tableau* (HALWANI et al., 2021). Em acréscimo, as ferramentas de Ciência de dados também têm sido utilizadas pelas próprias empresas de recrutamento e seleção, como forma de mineração de informações textuais e seleção de perfis (FALLUCCHI et al., 2020; KHAN et al., 2020).

Neste contexto, pode-se dizer que a produção de conhecimento científico sobre *Data Science* é vasta, portanto, analisar àquelas publicações relacionadas ao campo da Engenharia de Produção constitui trabalho inovador, uma vez que se trata de uma publicação inédita. Desta forma esta pesquisa se justifica em algumas perspectivas. Em primeiro lugar, trata-se de um tema recente na literatura, que irá fornecer substratos e novas habilidades valoradas no mercado de trabalho para o autor desta pesquisa. Para a literatura, esta pesquisa contribui ao gerar uma sistematização de todas as pesquisas já publicadas sobre o tema, fornecendo um guia a possíveis interessados sobre o assunto. Por fim, para a Universidade FUMEC, este Trabalho de Conclusão de Curso (TCC) contribui no sentido de aproximar a instituição da vanguarda de pesquisas acadêmicas sobre *Data Science*.

Diante do exposto, formula-se a pergunta de pesquisa: quais são os principais autores, fontes de publicação, os países e as áreas quentes e frias de pesquisa sobre *Data Science* e Engenharia de Produção no mundo? De modo a responder esta pergunta, em seguida apresentam-se os objetivos de pesquisa, geral e específico.

1.1 Objetivo Geral

O objetivo geral desta pesquisa é mapear a publicação científica sobre *Data Science* e suas interfaces, no contexto da Engenharia de Produção.

1.2 Objetivos Específicos

Especificamente, pretende-se:

- i. Realizar um estudo Bibliométrico, com base nas publicações nacionais e internacionais da *Web of Science*;
- ii. Mapear os principais autores, países e temas de pesquisa na área;
- iii. Definir as áreas quentes e frias de pesquisa sobre o assunto em questão;
- iv. Estabelecer uma agenda de pesquisa aos interessados em *Data Science* e Engenharia de Produção.

2. MATERIAIS E MÉTODOS

Esta pesquisa fez um levantamento de publicações sobre *Data Science* e suas interfaces no contexto da Engenharia de Produção, publicadas até 29/05/2022. Assim, examinaram-se todas as publicações indexadas junto à *Web of Science* (WOS). A escolha deste repositório de pesquisa se deu, pois, a WOS é considerada um dos melhores repositórios globais de ciência (LEYDESDORFF, CARLEY e RAFOLS, 2013; XIE et al., 2020). A plataforma de pesquisa de da WOS possui diversas funcionalidades. Nela é possível diversos metadados: *abstract*, *cited references*, *times cited*, *authors*, *institutions* e *countries*. Após a escolha da WOS, definiram-se os critérios de chaveamento e o conjunto de descritores para adereçar o tema (Quadro 1). Note que algumas das *keywords* utilizadas possuem um caráter especial “*”, que auxilia na busca por variações de um dado termo (expressões no singular e plural) (KASSAI, 2006). Ao todo, selecionou-se um total de as 277 publicações realizadas entre os anos de 1987 (data da primeira publicação sobre o tema) a 2022 (último ano com informações disponíveis), indexadas junto a base da WOS e que se enquadraram no chaveamento de pesquisa apresentado.

Quadro 1 - Descrição dos processos metodológicos adotados na busca por descritores

Repositório	Descritores	Fontes	Universo	Registros
Web of Science (WoS)	(("Data Science") OR ("Ciência dos Dados") OR ("Data Mining") OR ("Text Mining") OR ("Natural Language Processing") OR ("Feature Engineering") OR ("Predictive Model*") OR ("Analytics") OR ("Big Data") OR ("Machine Learning") OR ("Test* and Train*") OR ("Supervised Model") OR ("Unsupervised Model") OR ("Reinforcement Learning") OR ("Deep Learning") OR ("Neural Networks") OR ("Hybrid Model*") OR ("Clustering") OR ("Statistical Model*") OR ("Data Analysis") OR ("Data Cleaning") OR ("Data Preprocessing") OR ("Artificial Intelligence") OR ("Internet of Things")) AND (("Industrial Engineering") OR ("Production Engineering") OR ("Engenharia de Produção"))	Publicações Científicas (Artigos de Conferências, Periódicos, Livros e Materiais Técnicos)	1987 a 2022	277

Fonte: Elaborado pelo autor (2022).

Por fim, depois da definição dos critérios de busca e extração dos metadados, foi feita uma análise exploratória dos dados relacionados àquelas publicações selecionadas. Desta forma, esta pesquisa fez uso do Pacote “Bibliometrix”, escrito na linguagem R e elaborado por Aria e Cuccurullo (2017). Destaca-se que este pacote tem sido amplamente utilizado em pesquisas científicas (MORAL-MUÑOZ, 2020; DONTU et al., 2021) Esta pesquisa também fez uso do *Software* VOSviewer (ECK e WALTMAN, 2010).

3. RESULTADOS E DISCUSSÃO

A presente pesquisa investigou um total de 277 publicações científicas relacionadas a *Data Science* e suas interfaces, no contexto da Engenharia de Produção. Para tal, todas as publicações, independente do formato, realizadas entre os anos de 1987 e 2022 foram selecionadas. Ao todo, estas publicações foram veiculadas em 277 fontes distintas (congressos, *journals* e livros) (Tabela 1). Verifica-se que as publicações selecionadas possuem um total de 832 autores distintos, entretanto, alguns deles possuem mais de uma publicação, logo, o número de aparições de autores foi igual a 893.

Tabela 1 - Estatísticas descritivas das publicações sobre Data Science e Engenharia de Produção

Descrição	Resultados
Anos das publicações	1987-2022
Fontes de publicação	215
Número total de publicações	277
Número de <i>keywords</i>	1057
Número total de autores	832
Número de aparições de autores	893
Número total de referências bibliográficas	8632
Número médio de documentos por autor	0,333
Número médio de autores por publicações	3
Número médio de citações por publicação	6,087

Fonte: Elaborado pelo autor (2022).

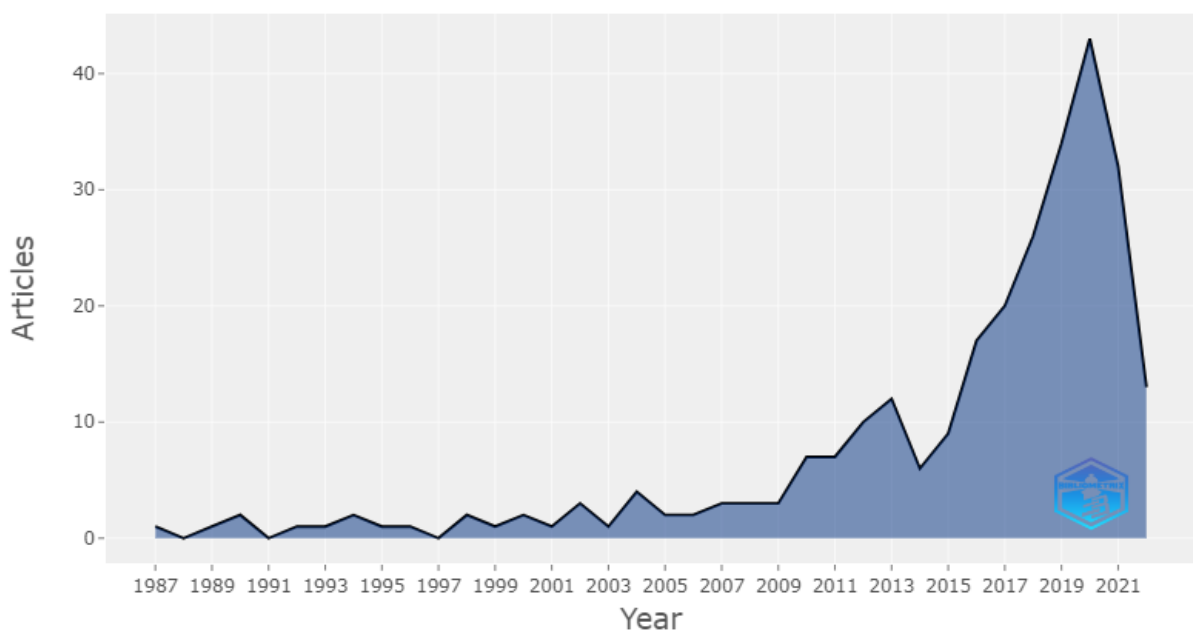
De acordo com a Tabela 1 verificou-se que ao todo, os registros utilizados analisados nesta pesquisa possuem no total, 1057 *keywords* distintas. Em geral, cada publicação sobre o tema possui uma média de 6,087 citações por artigo, o que demonstra que, apesar de ser um tema de pesquisa recente, há um grande interesse por parte das pesquisas sobre *Data Science* e Engenharia de Produção.

As pesquisas sobre *Data Science* e suas interfaces, no contexto da Engenharia de Produção se iniciaram com o trabalho desenvolvido por Aleksander (1987). Neste trabalho, o autor descreve os princípios das técnicas avançadas de programação, comumente chamadas de “Inteligência Artificial”. O foco da publicação é o debate sobre como as novas tecnologias podem ter impacto no processo de tomada de decisões, sobretudo aquelas relacionadas a automatização de processos. O autor sustenta que a adoção de robôs no ambiente fabril pode auxiliar no planejamento de uma sequência de ações com base em sinais obtidos de mudanças nas condições de seu ambiente. Esses sinais podem, de fato, ser bastante complexos, por exemplo, a entrada de informações visuais de uma câmera, por exemplo.

Por sua vez, o artigo mais recente, publicado até a data de realização desta pesquisa, foi proposto por Gröger (2022). Nesta publicação o autor sustenta que a transformação digital gera enormes quantidades de dados heterogêneos em toda a cadeia de valor industrial, desde dados de simulação em engenharia, dados de sensores na fabricação até dados de telemetria sobre o uso do produto. Conforme apontado por Gröger (2022), extrair *insights* desses dados constitui um fator crítico de sucesso para empresas industriais, por exemplo, para se otimizar processos e aprimorar os recursos do produto. Isso é conhecido como análise industrial, ou alternativamente, análise de dados para criação de valor industrial. A análise industrial é uma área de assunto interdisciplinar entre ciência de dados e engenharia industrial e está no centro da Indústria 4.0. O objetivo desta publicação foi a construção de uma estrutura conceitual para a análise industrial.

De modo a ilustrar a evolução temporal das pesquisas sobre *Data Science* e Engenharia de Produção, a Figura 1 apresenta o número de publicações realizadas em cada um dos anos analisados. Percebe-se que as publicações da amostra indicam um crescimento exponencial no número de registros, sobretudo após o ano de 2014. Note que o montante de publicações sobre a temática desta pesquisa saltou de 1 (1987) para 35 em 2021, último ano com informações completas disponíveis¹. Por fim, tem-se que a taxa de crescimento anual das publicações científicas foi igual a de 8.35%, o que indica um substancial interesse sobre esta área.

Figura 1 - Evolução anual das publicações sobre Data Science e Engenharia de Produção



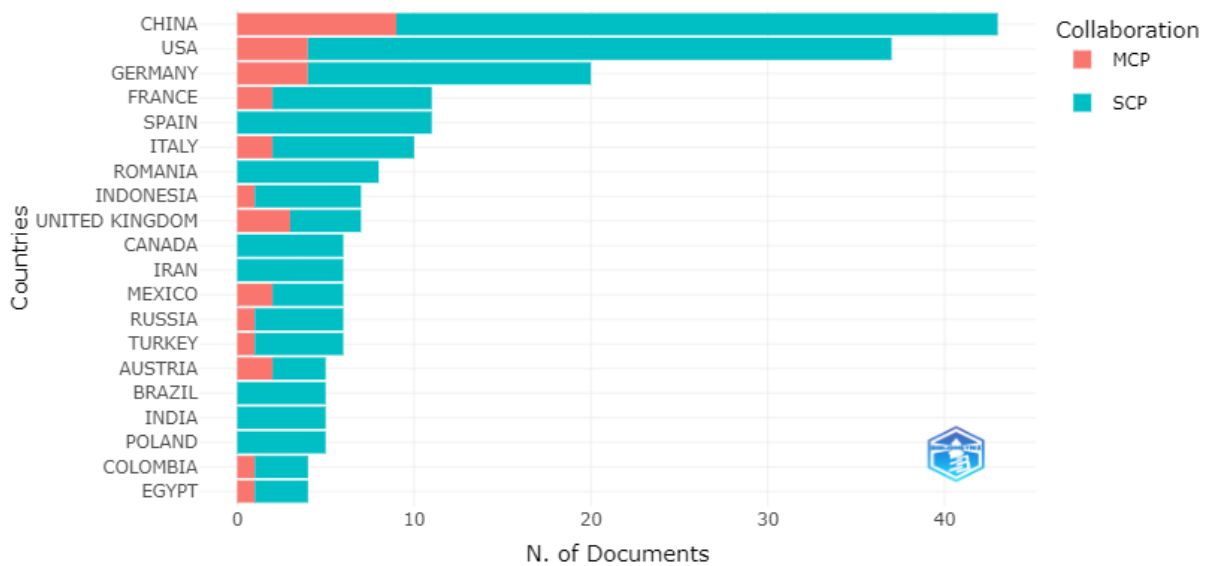
Fonte: Elaborado pelo autor (2022).

¹ Observação: o ano de 2022 ainda está incompleto, mas na data de 29/05/2022, há 11 registros de publicações sobre o tema na base da *Web of Science*.

Levando-se em consideração a origem do autor correspondente das 277 publicações selecionadas sobre *Data Science* e suas interfaces no contexto da Engenharia de Produção, percebe-se que autores de 55 países realizaram publicações sobre o tema proposto. A Figura 2 indica os 20 países que mais pesquisaram na área. As cores das barras indicam a relação de trabalhos feito em regime de parceria com outros países: (i) MCP: *Multiple Country Publication* e (ii) SCP: *Single Country Publication*). Por se tratar de um tema estratégico, as publicações em regime de parceria são minoritárias. Percebe-se ainda que a China (43 registros), Estados Unidos (37 registros) e Alemanha (20) respondem por 44 % das publicações sobre o assunto em questão.

O Brasil possui 5 publicações sobre o tema e ocupa a 15ª posição, junto com a Índia e a Polônia. Percebe-se ainda que todas as publicações realizadas pelo Brasil são relativas ao sistema *Single Country Publication*, ou seja, são publicações que não envolvem colaboradores de outros países, logo, recomenda-se o fomento de parcerias entre os autores brasileiros e as instituições de pesquisa do exterior.

Figura 2 - Principais países que publicaram sobre o tema, com base no autor correspondente



Fonte: Elaborado pelo autor (2022).

Após a análise dos principais países que desenvolveram pesquisas sobre a temática proposta, procedeu-se o estudo das principais fontes de publicação destes trabalhos (Tabela 2). Desta forma, é possível ver o *ranking* contendo as três primeiras colocações, no que diz respeito ao número de artigos veiculados. O Fator de Impacto de cada uma destas fontes é apresentado na figura do *Journal Citation Reports* (JCR) (LEYDESDORFF, 2004).

Tabela 2 - Principais fontes de publicação sobre *Data Science* e Engenharia de Produção

Fonte	JCR	Website	Publicações
<i>Computers & Industrial Engineering</i>	5,431	https://www.journals.elsevier.com/computers-and-industrial-engineering	10
<i>Heliyon</i>	2,85	https://www.sciencedirect.com/journal/heliyon	6
<i>IEEE Access</i>	3,557	https://ieeaccess.ieee.org/	6
<i>Jordan Journal of Mechanical and Industrial Engineering</i>	0,672	http://jjmie.hu.edu.jo/	5
<i>Journal of Intelligent Manufacturing</i>	6,485	https://www.springer.com/journal/10845	5

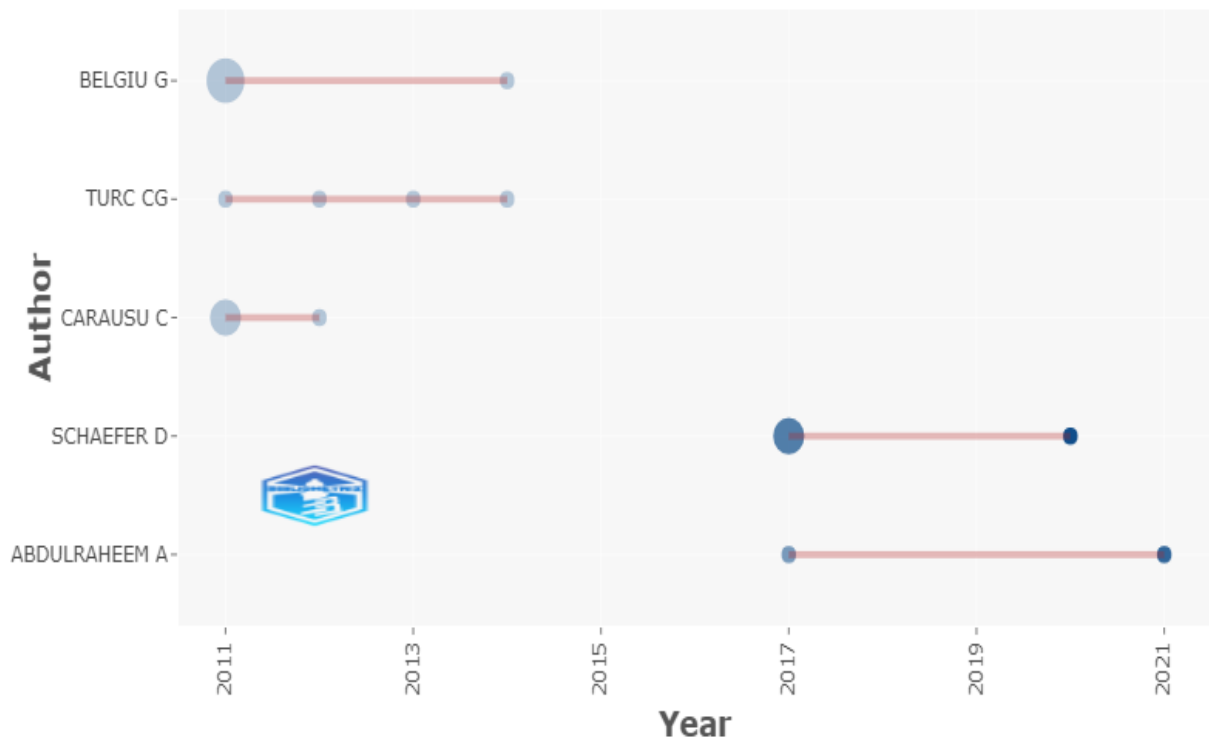
Fonte: Elaborado pelo autor (2022).

Após a análise das principais fontes de publicações, procedeu-se o estudo da longevidade e do impacto de publicação, em termos de citações, dos principais autores no campo de pesquisa analisado. Conforme apontado na Figura 3, os cinco principais autores sobre *Data Science* e Engenharia de Produção, são, respectivamente: (i) George Belgiu (Politehnica University of Timișoara), que possui quatro publicações; (ii) Cristian Gheorghe Turc (Politehnica University of Timișoara), que possui quatro publicações; (iii) Constantin Carausu (Gheorghe Asachi Technical University of Iasi), que possui três publicações; (iv) Dirk Schaefer (University of Liverpool), que possui três publicações e (v) Abdulazeez Abdulraheem (King Fahd University of Petroleum & Minerals), que possui duas publicações.

Note que, de acordo com a Figura 3, cada um dos nós representa as publicações de um autor, em um ano específico. Assim, o tamanho da circunferência de cada nó representa a quantidade de artigos que cada autor publicou em um ano. De modo complementar, a cor dos nós indica o número de citações obtidas pelo autor no ano em questão (cores mais escuras denotam os anos com mais citações de um dado autor).

Ainda, conforme apresentado pela Figura 3, os autores mais representativos quanto ao número de citações são: (i) Dirk Schaefer (FLYNN, DANCE e SCHAEFER, 2017; MOURAD et al., 2017; MOURAD et al., 2020), com um total de 33 citações *Web of Science* e em segundo lugar: (ii) Abdulazeez Abdulraheem (RAMMAY e ABDULRAHEEM, 2017; TARIQ, MAHMOUD e ABDULRAHEEM, 2021), com um total de 19 citações *Web of Science*. Assim, estes são, na atualidade, os pesquisadores mais proeminentes sobre *Data Science* e Engenharia de Produção tema na atualidade.

Figura 3 - Principais autores mundiais sobre *Data Science* e Engenharia de Produção



Fonte: Elaborado pelo autor (2022).

Após o estudo dos principais autores, tanto em número de publicações, quanto no impacto de suas pesquisas (citações), procedeu-se a análise dos trabalhos seminais sobre *Data Science* e Engenharia de Produção. Para tal, realizou-se aqui uma etapa importante dos estudos bibliométricos, que é a análise das referências bibliográficas de todas as 277 publicações selecionadas nesta pesquisa.

Desta forma, mediante um modelo classificado como “*bag of words*” (MATSUBARA, MARTINS e MONARD, 2003; ZHANG, JIN e ZHOU, 2010), compilou-se todas as referências bibliográficas dos 277 metadados da amostra. Estes registros foram classificados quanto ao número de citações do tipo *Web of Science*. O número total de referências analisadas foi igual a 8632 registros. Desta forma, foi possível analisar em detalhes quais foram aquelas publicações científicas mais relevantes sobre *Data Science* e suas interfaces, no contexto da Engenharia de Produção. O Quadro 2 descreve quais foram as cinco principais publicações científicas sobre o tema proposto, isto é, aqueles trabalhos seminais que mais vezes foram citados pelas pesquisas analisadas neste trabalho.

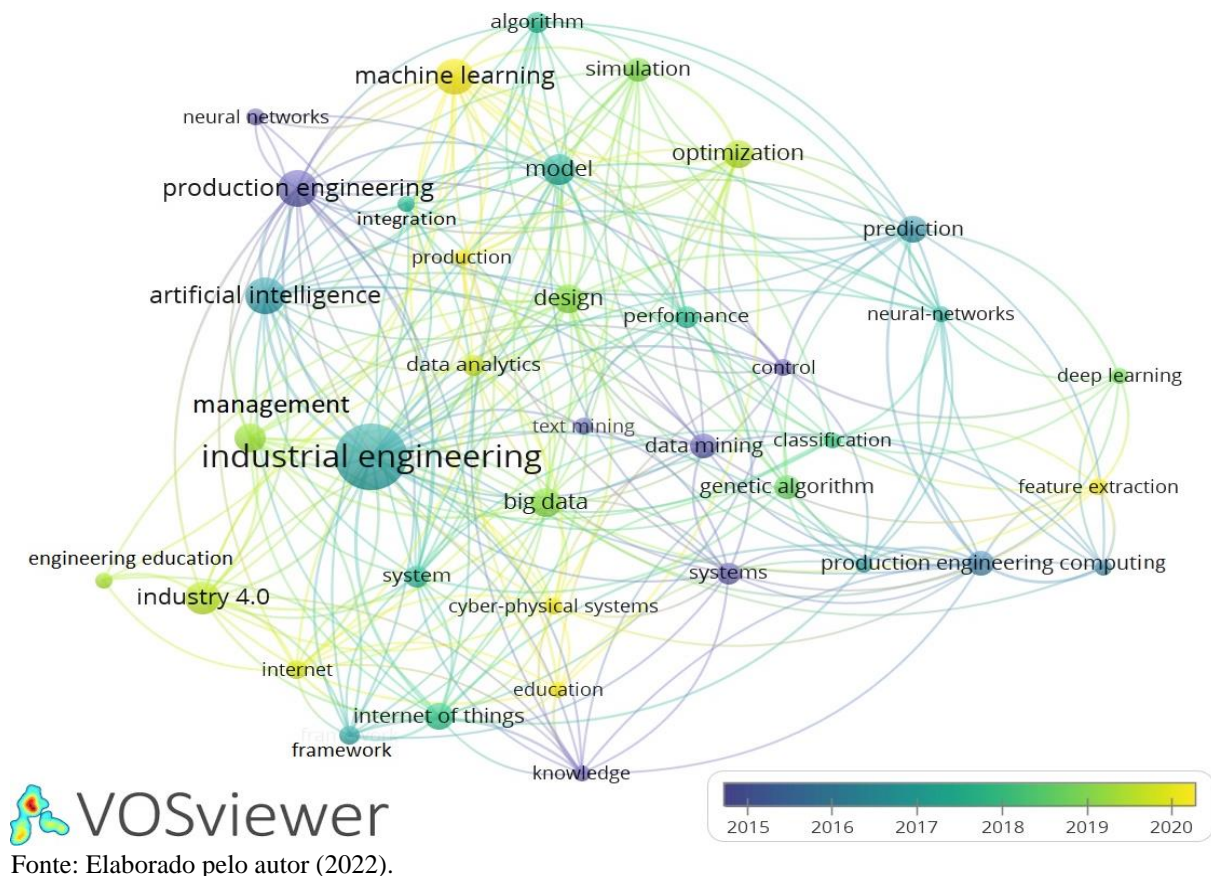
Quadro 2 - Os cinco artigos globais sobre *Data Science* e Engenharia de Produção que mais vezes foram citados

Título completo da publicação	Autores	Nº Citações	Objeto de investigação da publicação analisada
<i>The impact of industry 4.0 on procurement and supply management: A conceptual and qualitative analysis</i>	Glas e Kleemann (2016)	182	Este trabalho aborda as potencialidades da Indústria 4.0 para as funções de gestão de compras, fornecimento e distribuição. Os resultados indicaram que as tecnologias da Indústria 4.0 legitimam o próximo nível de maturidade em compras (Gestão de Compras e Suprimentos 4.0).
<i>Big data and predictive analytics for supply chain sustainability: A theory-driven research agenda</i>	Hazen et al. (2016)	126	Os autores sustentaram que a <i>Big data</i> e a Análise Preditiva (BDPA) afetam o desempenho financeiro/estratégico no <i>Supply Chain Management</i> (SCM). Esta pesquisa indicou ainda que o BDPA também pode ser usado para aprimorar e permitir um SCM sustentável. Os autores propuseram uma agenda de pesquisa baseada em teorias.
<i>OR-methods for coping with the ripple effect in supply chains during COVID-19 pandemic: Managerial insights and research implications</i>	Ivanov e Dolgui (2021)	84	Neste artigo os autores conceituam o estado atual e as direções de pesquisa futuras sobre o efeito cascata para o contexto pandêmico. Para tal, examinaram-se estudos existentes de OR (<i>Operational Research</i>) publicados em revistas internacionais que tratam da propagação de disrupção e dinâmica estrutural em <i>Supply Chains</i> (SCs).
<i>Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study</i>	Qi e Tang (2018)	83	Os autores comparam seis abordagens integradas de inteligência artificial (IA) para a previsão de <i>slope stability</i> com base em metaheurísticas e algoritmos de <i>Machine Learning</i> (ML). Seis algoritmos de ML, incluindo regressão logística, árvore de decisão, floresta aleatória, <i>gradient boosting machine</i> , <i>support vector machine</i> e rede neural perceptron multicamada, foram usados para a modelagem de relacionamento e o algoritmo firefly (FA) foi usado para o ajuste de hiperparâmetros.
<i>Intelligent laser welding through representation, prediction, and control learning: An architecture with deep neural networks and reinforcement learning</i>	Günther et al. (2016)	75	Como primeira contribuição deste trabalho, mostrou-se como uma rede neural profunda e auto codificada é capaz de extrair características salientes e de baixa dimensão de dados reais de soldagem a laser de alta dimensão.

Fonte: Elaborado pelo autor (2022).

A Figura 4 apresenta as 38 principais *keywords* utilizadas nas pesquisas sobre *Data Science* e Engenharia de Produção. Como critério de seleção, cada *keyword* selecionada apareceu pelo menos 6 vezes em publicações distintas. O tamanho de cada nó denota a importância relativa da *keyword*. As *keywords* estão conectadas por arestas, que denotam relações de coocorrência. A proximidade entre os nós indica a força da relação entre cada uma das *keywords*. Por fim é possível medir a tendência no uso das *keywords* ao longo dos anos (1987 a 2022). Desta forma, classificou-se aquelas *keywords* mais utilizadas no passado (“áreas frias de pesquisa” – tons roxos) e aquelas mais recorrentes em trabalhos recentes (“áreas quentes de pesquisa” – tons amarelos) sobre *Data Science* e Engenharia de Produção.

Figura 4 - Evolução temporal das palavras-chaves das publicações sobre *Data Science* e Engenharia de Produção



De acordo com a Figura 4, percebe-se que após o ano de 2020, houve uma concentração de publicações científicas sobre *Cyber-Physical Systems* (KIM e LEE, 2021), *Education* (WOSCHANK, 2021), *Feature Extraction* (FATANI et al., 2021) e *Machine Learning* (ELHEDHLI, OMAR e BOUABID, 2021), logo, estas são áreas promissoras de pesquisa, uma vez que se tratam das principais *keywords* utilizadas na atualidade sobre o tema abordado nesta pesquisa.

4. CONCLUSÕES

Esta pesquisa realizou um procedimento bibliométrico para investigar a produção do conhecimento científico sobre *Data Science* e suas interfaces, no contexto da Engenharia de Produção. Para tal, analisaram-se todas as publicações indexadas junto a *Web of Science*, que adereçaram o tema proposto, na janela temporal compreendendo os anos de 1987 a 2022. Em função da 4ª Revolução Industrial, notadamente na figura da Indústria 4.0, o tema em questão assume relevância acadêmica, empresarial e governamental.

Em função da importância da temática em voga, esta pesquisa verificou que as publicações sobre o tema cresceram a uma taxa anual da ordem de 8,35%, no período entre os anos de 1987 e 2022. Destaca-se, entretanto, que esta é uma estimativa, dado que esta pesquisa foi finalizada em maio de 2022, logo, o crescimento anual deve ser maior do que o calculado até aqui. Destaca-se que a produção científica sobre *Data Science* e Engenharia de Produção está concentrada em apenas três países, a saber: China (43 publicações), Estados Unidos (37 publicações) e Alemanha (20 publicações). Juntos, estes países respondem por aproximadamente 44% das publicações sobre o tema. Em acréscimo, destaca-se que o Brasil ocupa a 15ª posição no *ranking* global de nacionalidade dos autores correspondentes que mais publicam sobre este assunto, logo é necessário fomentar a realização de novas pesquisas locais.

Conforme apontado por esta pesquisa, o *journal Computers & Industrial Engineering* foi aquele que possui mais publicações sobre o assunto. Alternativamente, o *Journal of Intelligent Manufacturing* notabilizou-se como a principal fonte de publicação no tocante ao fator de impacto.

Esta pesquisa verificou que os professores George Belgiu (Politehnica University of Timișoara), Cristian Gheorghe Turc (Politehnica University of Timișoara) e Constantin Carausu (Gheorghe Asachi Technical University of Iasi), são aqueles que mais publicaram sobre o tema proposto. Por outro lado, os professores Dirk Schaefer (University of Liverpool) e Abdulazeez Abdulraheem (King Fahd University of Petroleum & Minerals) notabilizaram-se como os principais pesquisadores em termos de citações, ou seja, de impacto científico.

Destaca-se que as 277 publicações analisadas nesta pesquisa, utilizaram 8632 referências bibliográficas. Neste sentido, destaca-se que a principal publicação citada nestes trabalhos foi proposta por Glas e Kleemann (2016), intitula-se: *The impact of industry 4.0 on procurement and supply management: A conceptual and qualitative analysis* e contabilizou 182 citações *Web of Science*. A supracitada publicação abordou as potencialidades da Indústria 4.0 para as funções de gestão de compras, fornecimento e distribuição.

Esta pesquisa verificou ainda que as principais *keywords* utilizadas na atualidade em pesquisas sobre *Data Science* e Engenharia de Produção, são: *Cyber-Physical Systems*, *Education*, *Feature Extraction* e *Machine Learning*. Destaca-se que o uso destas expressões passou a ser frequente após o ano de 2020, logo, estas são algumas das áreas quentes sobre a temática proposta e, portanto, merecem atenção especial.

Finalmente, esta pesquisa possui algumas limitações, uma vez que utilizou apenas uma base de pesquisa, a saber, *Web of Science*. Assim, para adereçar esta lacuna, recomenda-se em trabalhos futuros, a busca de publicações em outros repositórios de pesquisa, de notada reputação, como é o caso do IEEE Xplore e da *Scopus*. Recomenda-se ainda a construção de uma Revisão Sistemática de Literatura (RSL) sobre o *Data Science* e Engenharia de Produção. Desta forma será possível analisar em detalhes a contribuição de cada uma das pesquisas da amostra.

REFERÊNCIAS

- ALEKSANDER, Igor. Artificial intelligence for production engineering: a historical approach. **Robotica**, v. 5, n. 2, p. 99-110, 1987.
- ARIA, Massimo; CUCCURULLO, Corrado. bibliometrix: An R-tool for comprehensive science mapping analysis. **Journal of informetrics**, v. 11, n. 4, p. 959-975, 2017.
- CAO, Longbing. Data science: challenges and directions. **Communications of the ACM**, v. 60, n. 8, p. 59-68, 2017.
- DONTHU, Naveen et al. How to conduct a bibliometric analysis: An overview and guidelines. **Journal of Business Research**, v. 133, p. 285-296, 2021.
- EGGER, Roman; YU, Joanne. Data science and interdisciplinarity. **Applied Data Science in Tourism**, p. 35-49, 2022.
- ELHEDHLI, Samir; OMAR, Mohammed; BOUABID, Ali. Data analytics-is it industrial engineering reborn?. **Global Journal of Engineering Education**, v. 23, n. 1, 2021.
- FALLUCCHI, Francesca et al. Predicting employee attrition using machine learning techniques. **Computers**, v. 9, n. 4, p. 86, 2020.
- FATANI, Abdulaziz et al. Advanced feature extraction and selection approach using deep learning and Aquila optimizer for IoT intrusion detection system. **Sensors**, v. 22, n. 1, p. 140, 2021.
- FLYNN, Joseph; DANCE, Steven; SCHAEFER, Dirk. Industry 4.0 and its Potential Impact on Employment Demographics in the UK. **Advances in Transdisciplinary Engineering**, v. 6, p. 239-244, 2017.
- GLAS, Andreas H.; KLEEMANN, Florian C. The impact of industry 4.0 on procurement and supply management: A conceptual and qualitative analysis. **International Journal of Business and Management Inventon**, v. 5, n. 6, p. 55-66, 2016.
- GRÖGER, Christoph. Industrial analytics—An overview. **it-Information Technology**, 2022.
- GRUS, Joel. **Data Science do zero: Primeiras regras com o Python**. Alta books, 2019.
- GÜNTHER, Johannes et al. Intelligent laser welding through representation, prediction, and control learning: An architecture with deep neural networks and reinforcement learning. **Mechatronics**, v. 34, p. 1-11, 2016.
- HALWANI, Marwah Ahmed et al. Job qualifications study for data science and big data professions. **Information Technology & People**, 2021.
- HAZEN, Benjamin T. et al. Big data and predictive analytics for supply chain sustainability: A theory-driven research agenda. **Computers & Industrial Engineering**, v. 101, p. 592-598, 2016.

IVANOV, Dmitry; DOLGUI, Alexandre. OR-methods for coping with the ripple effect in supply chains during COVID-19 pandemic: Managerial insights and research implications. **International Journal of Production Economics**, v. 232, p. 107921, 2021.

KASSAI, Tibor. The impact on database searching arising from inconsistency in the nomenclature of parasitic diseases. **Veterinary parasitology**, v. 138, n. 3-4, p. 358-361, 2006.

KHAN, Alam Sher et al. Personality classification from online text using machine learning approach. **International Journal of Advanced Computer Science and Applications**, v. 11, n. 3, p. 460-476, 2020.

KIM, Jun; LEE, Ju Yeon. Server-Edge dualized closed-loop data analytics system for cyber-physical system application. **Robotics and computer-integrated manufacturing**, v. 67, p. 102040, 2021.

LEYDESDORFF, Loet; CARLEY, Stephen; RAFOLS, Ismael. Global maps of science based on the new Web-of-Science categories. **Scientometrics**, v. 94, n. 2, p. 589-593, 2013.

MATSUBARA, Edson Takashi; MARTINS, Claudia Aparecida; MONARD, Maria Carolina. Pretext: Uma ferramenta para pré-processamento de textos utilizando a abordagem bag-of-words. **Technical Report**, v. 209, n. 4, p. 10-11, 2003.

MORAL-MUÑOZ, José A. et al. Software tools for conducting bibliometric analysis in science: An up-to-date review. **Profesional de la Información**, v. 29, n. 1, 2020.

MOURAD, Mohamed et al. C-MARS-ABM: A deployment approach for cloud manufacturing. **Proceedings of the Advances in Transdisciplinary Engineering; IOS Press BV: Amsterdam, The Netherlands**, v. 6, p. 213-218, 2017.

MOURAD, Mohamed H. et al. Assessment of interoperability in cloud manufacturing. **Robotics and Computer-Integrated Manufacturing**, v. 61, p. 101832, 2020.

QI, Chongchong; TANG, Xiaolin. Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study. **Computers & Industrial Engineering**, v. 118, p. 112-122, 2018.

RAMMAY, Muzammil Hussain; ABDULRAHEEM, Abdulazeez. PVT correlations for Pakistani crude oils using artificial neural network. **Journal of Petroleum Exploration and Production Technology**, v. 7, n. 1, p. 217-233, 2017.

SAGIROGLU, Seref; SINANC, Duygu. Big data: A review. In: **2013 international conference on collaboration technologies and systems (CTS)**. IEEE, 2013. p. 42-47.

SARKER, Iqbal H. Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. **SN Computer Science**, v. 2, n. 5, p. 1-22, 2021.

SRIVASTAVA, Priyanshu; KHAN, Rizwan. A review paper on cloud computing. **International Journal of Advanced Research in Computer Science and Software Engineering**, v. 8, n. 6, p. 17-20, 2018.

TARIQ, Zeeshan; MAHMOUD, Mohamed; ABDULRAHEEM, Abdulazeez. Machine learning-based improved pressure–volume–temperature correlations for black oil reservoirs. **Journal of Energy Resources Technology**, v. 143, n. 11, 2021.

VAN ECK, Nees; WALTMAN, Ludo. Software survey: VOSviewer, a computer program for bibliometric mapping. **scientometrics**, v. 84, n. 2, p. 523-538, 2010.

WEBER, Rolf H.; WEBER, Romana. **Internet of things**. Heidelberg: Springer, 2010.

WOSCHANK, Manuel et al. The Usage of Challenge-Based Learning in Industrial Engineering Education. In: **International Conference on Interactive Collaborative Learning**. Springer, Cham, 2021. p. 869-878.

XIE, Lin et al. Bibliometric and visualized analysis of scientific publications on atlantoaxial spine surgery based on Web of Science and VOSviewer. **World neurosurgery**, v. 137, p. 435-442. e4, 2020.

ZHANG, Yin; JIN, Rong; ZHOU, Zhi-Hua. Understanding bag-of-words model: a statistical framework. **International Journal of Machine Learning and Cybernetics**, v. 1, n. 1, p. 43-52, 2010.