

Tópicos e tendências no ensino de física utilizando inteligência artificial

Topics and trends in physics teaching using artificial intelligence

Érick Ghuron^{1*}, Daniel Trugillo Martins Fontes², André Machado Rodrigues³

¹ Licenciando em Física, Universidade de São Paulo, Rua do Matão 1371, CEP 05508-090 – São Paulo, SP, Brasil.

² Programa de Pós-Graduação Interunidades em Ensino de Ciências da USP, Universidade de São Paulo, Rua do Matão 1371, CEP 05508-090 - São Paulo, SP, Brasil.

³ Instituto de Física, Universidade de São Paulo, Rua do Matão 1371, CEP 05508-090 - São Paulo, SP, Brasil.

*E-mail: ghuron@usp.br

Resumo

Este trabalho investiga a aplicação do aprendizado de máquina, particularmente utilizando o algoritmo Latent Dirichlet Allocation, para identificar tópicos e tendências em revistas acadêmicas de ensino de Física no contexto latino-americano. As revistas analisadas foram a Revista Brasileira de Ensino de Física (RBEF) e a Revista de Enseñanza de la Física (REF), abrangendo o período de 2001 a 2022. Foram coletados 1664 artigos da RBEF e 885 da REF, representando 79% e 85% dos artigos publicados nesses períodos, respectivamente. Os resultados indicaram tópicos dominantes em cada revista e suas respectivas tendências ao longo do tempo. Enquanto a RBEF mostrou um declínio em tópicos relacionados à Educação e um aumento nas publicações sobre Física Geral, a REF apresentou uma predominância de tópicos relacionados à Educação, com um aumento significativo em publicações sobre Laboratório Virtual e Ensino. Estas descobertas proporcionam informações valiosas sobre a evolução dos temas de pesquisa em ensino de Física nos contextos brasileiro e argentino e quais são as potencialidades de se utilizar o aprendizado de máquina em pesquisas de ensino de Física.

Palavras-chave: Aprendizado de máquina; Tecnologia da informação e comunicação; Análise quantitativa; Revisão bibliográfica.

Abstract

This study investigates the application of machine learning, specifically using the Latent Dirichlet Allocation algorithm, to identify topics and trends in academic journals on Physics education in the Latin America context. The journals analyzed were the Revista Brasileira de Ensino de Física (RBEF) and the Revista de Enseñanza de la Física (REF), covering the period from 2001 to 2022. A total of 1664 articles from RBEF and 885 from REF were collected, representing 79% and 85% of the articles published in these periods, respectively. The results indicated dominant topics in each journal and their respective trends over time. For instance, RBEF showed a decline in topics related to Physics Education and an increase in publications on General Physics. In contrast, REF displayed a predominance of topics related to Teaching, with a significant rise in publications about Virtual Laboratory and Teaching. These findings provide valuable insights into the evolution of research themes in Physics education in the Brazilian and Argentine contexts and the potential of using machine learning in Physics education research.

Keywords: Machine learning; Information and communication technology; Quantitative analysis; Literature review.

I. INTRODUÇÃO

Atualmente, pode-se afirmar que a tecnologia da informação e comunicação, especialmente a inteligência artificial em suas múltiplas formas e aplicações, tem ganhado espaço como suporte à pesquisa em Educação em Ciências. Com as mudanças ocorridas nesse campo de pesquisas, não se pode ignorar o impacto dos recursos tecnológicos tanto na pesquisa quanto no ensino de Física e Ciências (Goulart, Pastorio e Vidmar, 2023). Nesse contexto, destaca-se o

trabalho de Chen et al. (2020), que aplicaram conceitos de aprendizagem de máquinas para identificar tópicos e tendências em tecnologias educacionais, em uma análise retrospectiva que cobriu quatro décadas de volumes da revista *Computers & Education*. Da mesma forma, Odden, Marin e Rudolph (2021) empregaram o aprendizado de máquina para analisar as transformações ocorridas nos trabalhos dos últimos 100 anos da *Science Education*. No campo específico do Ensino de Física, Odden, Marin e Caballero (2020) utilizaram a inteligência artificial para avaliar cerca de 1300 trabalhos do *Physics Education Research Conference Proceedings* entre 2001 e 2018, apontando as principais tendências da área no contexto internacional.

Diante desse movimento internacional, o presente trabalho se propõe a realizar uma análise semelhante no contexto latino-americano, reiterando a importância e o potencial de tais pesquisas para ajudar a compreender trabalhos no ensino de física. Mais importante, este estudo busca realizar uma análise focada na Revista Brasileira de Ensino de Física (RBEF) e na Revista de Enseñanza de la Física (REF), duas revistas importantes em Ensino de Física em seus respectivos contextos e abrangência internacional. Para isso, será utilizado o algoritmo *Latent Dirichlet Allocation* (LDA), o mesmo algoritmo de aprendizado de máquina usado nas pesquisas citadas anteriormente. Nosso objetivo é identificar os principais tópicos e tendências presentes na RBEF e REF, no período de 2001 a 2022, por meio da aplicação do aprendizado de máquina.

II. FUNDAMENTAÇÃO TEÓRICA

O aprendizado de máquina é um subcampo da Inteligência Artificial (IA) que envolve o desenvolvimento de algoritmos permitindo que os computadores aprendam e realizem tarefas a partir de dados (Mitchell, 1997). Em vez de serem programados explicitamente para executar uma tarefa, os computadores são treinados usando conjuntos de dados e algoritmos que lhes conferem a capacidade de aprender padrões estatísticos dos dados e, assim, realizar previsões ou tomar decisões (Chollet, 2018). O aprendizado de máquina pode ser dividido em três formas básicas: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. No aprendizado supervisionado, o algoritmo recebe um conjunto de dados rotulados, ou seja, cada dado possui uma resposta esperada. Nesse caso, o objetivo é aprender uma função que mapeie os dados de entrada para as saídas desejadas. No aprendizado não supervisionado, o algoritmo recebe um conjunto de dados não rotulados, ou seja, sem respostas esperadas. O objetivo é encontrar padrões, agrupamentos ou estruturas nos dados sem nenhuma orientação prévia. Por fim, no aprendizado por reforço, o algoritmo interage com um ambiente dinâmico e recebe recompensas ou punições por suas ações. O objetivo é aprender uma política ótima que maximize a recompensa acumulada (Chollet, 2018).

O LDA (Latent Dirichlet Allocation) é um modelo probabilístico generativo que faz parte da classe dos modelos de tópicos. Os modelos de tópicos são uma forma de aprendizado não supervisionado que têm como objetivo descobrir os tópicos latentes presentes em uma coleção de documentos (Blei, Andrew e Michael, 2003). Cada documento é representado como uma mistura de tópicos, e cada tópico é representado como uma distribuição de palavras. Nesse modelo insere-se o número de tópicos que se deseja gerar e ajustam-se os hiperparâmetros α e η que controlam o grau de dispersão das distribuições de tópicos e palavras, respectivamente. Quanto menor o valor de α , mais concentrada será a distribuição de tópicos por documento, ou seja, cada documento terá poucos tópicos dominantes. Quanto menor o valor de η , mais concentrada será a distribuição de palavras por tópico, ou seja, cada tópico terá poucas palavras relevantes.

Para avaliar a qualidade dos tópicos gerados, utiliza-se a métrica de coerência. Esta métrica verifica o quanto as palavras dentro de um tópico se relacionam entre si. Uma das formas mais populares de medir a coerência é através da métrica C_v . Ela considera a frequência com que pares de palavras coexistem em um mesmo documento e a frequência individual de cada palavra. Tópicos que agrupam palavras frequentemente encontradas juntas em textos tendem a ser mais coerentes e, conseqüentemente, mais fáceis de interpretar (Röder, Both e Hinneburg, 2015).

III. METODOLOGIA

A obtenção dos documentos analisados (artigos) foi realizada através da técnica *web scraping*, que consiste no processo de extrair informações de sites da internet. Para isso utiliza-se um código (frequentemente em linguagens de programação como Python) que acessa o site, analisa seu conteúdo e extrai os dados desejados (Mitchell, 2018). Os artigos da RBEF foram obtidos no indexador da SciELO e os artigos da REF foram obtidos no site da própria revista.

A escolha de analisar os artigos em português da RBEF e os em espanhol da REF se deu devido às particularidades linguísticas de cada idioma que são aceitos para publicação. Essa decisão foi baseada no idioma de publicação predominante em cada revista, que, embora aceite submissões em inglês, concentra artigos nas línguas nativas de seus respectivos autores, corroborando a tendência apontada por Fontes e Rodrigues (2022). Ao todo, foram obtidos 885

artigos da REF, o que representa 85% de todos os artigos publicados no período em questão. Para a RBEF, foram obtidos 1664 artigos, equivalentes a 79% dos artigos publicados no mesmo intervalo de tempo. Essas porcentagens alcançadas são robustas, conferindo confiança à amostragem e assegurando que a análise realizada não perde em generalidade.

Sobre os formatos dos dados, a RBEF disponibiliza os artigos em formato HTML, o que facilitou significativamente o processo de pré-processamento. Enquanto isso, a REF fornece os artigos em formato PDF, padrão adotado para todos os documentos dessa revista. No que diz respeito ao pré-processamento, as etapas foram às mesmas para ambas as revistas. Utilizou-se o Python (biblioteca *Spacy*) para realizar a leitura dos textos de forma automática. As palavras foram filtradas de acordo com sua classe gramatical, mantendo apenas os substantivos, adjetivos e verbos, devido ao seu profundo valor semântico em relação às outras classes. Após essa etapa, as palavras restantes foram lematizadas e alocadas em um dicionário próprio contendo todas as palavras presentes no texto. Palavras que apareciam em mais de 50% dos artigos foram excluídas para garantir a formação de tópicos realmente distintos.

Finalmente, foram inseridos nos modelos LDA de cada revista o dicionário criado e os documentos previamente processados. A quantidade de tópicos gerados para os modelos foi escolhida em 12, levando em consideração o número de diferentes linhas temáticas comumente presente em eventos da área do Ensino de Física. O atual XXIII Reunión de Educación conta com 12 diferentes linhas temáticas, e o último XIX Encontro de Pesquisas em Ensino de Física contou com 11 linhas temáticas. Vale destacar que os demais hiperparâmetros α e η do LDA foram ajustados automaticamente pela biblioteca *Gensim*, otimizando os resultados obtidos. Uma vez definidos os tópicos para ambas as revistas, passa-se à análise de tendências, com o intuito de compreender as principais temáticas e sua evolução ao longo do tempo. As linhas de tendências foram traçadas a partir do método dos mínimos quadrados.

IV. RESULTADOS

Os modelos da RBEF e da REF apresentaram bons resultados, com coerências de 0,61 e 0,58, respectivamente. Esses valores estão dentro da faixa esperada para trabalhos da área, que costumam ter coerências entre 0,4 e 0,6, conforme indicado por Odden, Marin, e Caballero (2020) e Syed e Spruit (2018).

As tabelas I e II mostram os tópicos latentes identificados para cada modelo, bem como a quantidade de artigos que se enquadram em cada tópico. Além disso, as tabelas também exibem as 10 palavras que têm a maior probabilidade de representar cada tópico, de acordo com o modelo. É importante ressaltar que a ordem das palavras na lista é significativa, pois reflete o grau de relevância de cada termo dentro do respectivo tópico. Além disso, para ambas as revistas, os tópicos latentes foram nomeados por nós após análise das palavras que melhor caracterizavam o tópico. Essa foi a primeira ação humana no modelo. As figuras 1 e 2 ilustram a variação de cada tópico ao longo do período analisado.

TABELA I. Tópicos latentes para a Revista Brasileira de Ensino de Física entre 2001 e 2022.

Tópico Latente	Palavras	Quantidade
Ensino	aluno, professor, estudante, atividade, aula, científico, ciência, pesquisa, conteúdo, curso	244 (14,7%)
Mecânica Quântica	partícula, quântico, potencial, vetor, operador, onda, interação, simetria, matriz, clássico	228 (13,7%)
Mecânica Clássica	ar, corpo, atrito, aceleração, esfera, coeficiente, altura, pressão, eixo, superfície	185 (11,1%)
História da Ciência	ciência, corpo, científico, livro, século, natureza, publicar, matéria, época, história	159 (9,6%)
Circuito Elétrico	circuito, corrente, tensão, sinal, frequência, elétrico, placa, sensor, capacitor, porta	136 (8,2%)
Termodinâmica	temperatura, radiação, onda, átomo, calor, partícula, frequência, solar, superfície, corda	135 (8,1%)
Física dos Materiais	distribuição, rede, parâmetro, fase, comportamento, simulação, probabilidade, escala, numérico, estrutura	117 (7,0%)
Física Moderna	partícula, corpo, quântico, mecânico, referencial, clássico, equilíbrio, relatividade, coordenada, entropia	112 (6,7%)
Ótica e Luz	luz, onda, imagem, raio, cor, lente, feixe, refração, espelho, objeto	102 (6,1%)
Eletromagnetismo	magnético, elétrico, carga, corrente, campo magnético, eletromagnético, campo elétrico, condutor, fio, potencial	96 (5,8%)
Astronomia	estrela, sol, planeta, universo, órbita, solar, lua, observação, raio, centro	79 (4,7%)
Metodologia	aluno, estudante, aula, pêndulo, curso, disciplina, item, atividade, gráfico, software	71 (4,3%)

TABELA II. Tópicos latentes para Revista de Enseñanza de la Física entre 2001 e 2022.

Tópico Latente	Palavras	Quantidade
Educação e Sociedade	curricular, país, profesional, taller, comunidad, institución, profesorado, académico, sociedad, problemática	220 (24,9%)
Laboratório Virtual	virtual, simulación, video, remoto, tic, digital, experimento, dispositivo, software, presencial	111 (12,5%)
Ensino	competencia, habilidad, informe, secuencia, escolar, escrito, categoría, pensamiento, argumentación, capacidad	105 (11,9%)
Metodologia	entrevista, epistemológico, discurso, sujeto, categoría, visión, cognitivo, noción, rol, pensamiento	102 (11,5%)
Mecânica Clássica	fuerza, velocidad, aceleración, masa, partícula, constante, mecánico, ecuación, rozamiento, angular	65 (7,3%)
Experimento	velocidad, posición, masa, experimento, medición, aire, medir, relatividad, cálculo, distancia	57 (6,4%)
Avaliação	opción, activo, prueba, cuestionario, correcto, visión, test, instrucción, porcentaje, par	51 (5,8%)
Física Moderna	energía, partícula, electrón, estado, calor, cuántico, radiación, átomo, temperatura, conservación	45 (5,1%)
Ótica e Luz	onda, luz, frecuencia, longitud, ecuación, ángulo, imagen, óptica, rayo, fase	40 (4,5%)
Matemática	gráfico, vector, lenguaje, matemática, registro, significado, posición, cálculo, imagen, semiótico	35 (4,0%)
Astronomia	tierra, astronomía, sol, solar, agua, luna, día, historia, planeta, universo	32 (3,6%)
Eletromagnetismo	eléctrico, circuito, magnético, corriente, carga, color, energía, luz, resistencia, inducción	22 (2,5%)



FIGURA 1. Frequência e tendência anual de tópicos da Revista Brasileira de Ensino de Física entre 2001 e 2022.

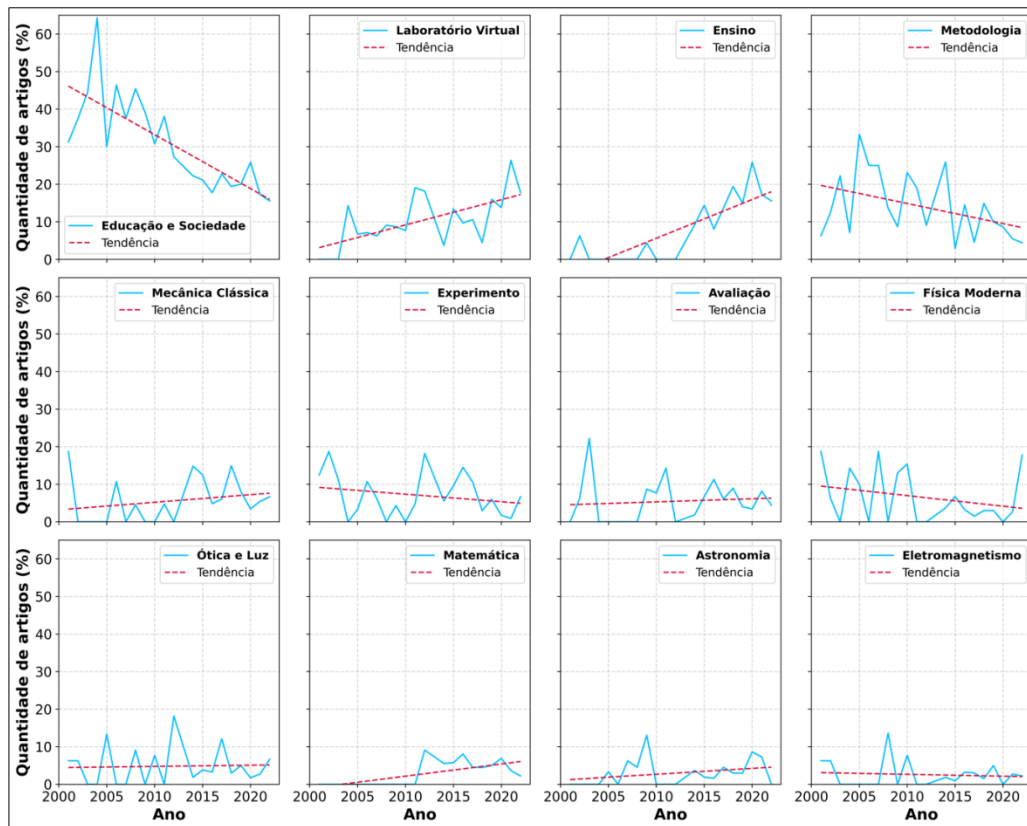


FIGURA 2. Frequência e tendência anual de tópicos da Revista de Enseñanza de la Física entre 2001 e 2022.

V. DISCUSSÃO

A figura 1 apresenta as tendências dos tópicos para o modelo da RBEF. Observa-se certa estabilidade na frequência do tópico Ensino, em torno de 15%, embora esse tópico apresente grandes variações em suas publicações anuais, chegando a valores inferiores a 5% dos artigos publicados. Retomando o objetivo da pesquisa, temos que tópicos da Física Geral (Mecânica Quântica, Mecânica Clássica, Circuito elétrico, Termodinâmica, Física dos Materiais, Física Moderna, Ótica, Eletromagnetismo, Astronomia) ocupam espaço relevante (aproximadamente 70% da Revista), enquanto os tópicos mais relacionados à Educação (Ensino, Metodologia e História da Ciência) são menos expressivos. Além disso, os tópicos da Física Geral estão bem distribuídos entre si, sem uma predominância clara. Em relação às tendências, percebe-se que tópicos como História da Ciência, Física Moderna e Eletromagnetismo parecem ocupar um menor espaço relativo de pesquisa, enquanto Mecânica Quântica, Mecânica Clássica, Física dos Materiais, Astronomia e Termodinâmica ganham visibilidade na pesquisa em Ensino de Física retratada pela RBEF. É importante ressaltar que um pico de publicações sobre Metodologia em 2002 pode ser atribuído a um número especial da Revista que abordou a física computacional e suas aplicações na pesquisa e no ensino. Portanto, a análise desse tópico latente ao longo do tempo deve ser interpretada com cautela.

Já no modelo da REF (figura 2), nota-se uma grande diminuição de publicações nos tópicos de Metodologia, Educação e Sociedade, enquanto houve um aumento significativo nas publicações sobre Laboratório Virtual e Ensino. Ao contrário da RBEF, na REF os tópicos relacionados ao Ensino ocupam um espaço de predominância, representando aproximadamente 60% dos artigos da revista. Quanto aos tópicos não diretamente relacionados ao ensino, destaca-se uma grande presença de artigos sobre Mecânica Clássica, com uma tendência de crescimento, enquanto o tópico de Física Moderna parece indicar uma tendência inversa, com um baixo número de publicações entre 2010 e 2020, período em que o tópico de Mecânica Clássica teve um grande número de publicações. Nesse mesmo intervalo de tempo, surge o tópico de Matemática com uma tendência positiva, podendo ser objeto de estudo em pesquisas futuras.

É importante notar que as revistas analisadas possuem características diferentes, mesmo sendo revistas que compartilham o mesmo nicho do Ensino de Física. A RBEF possui um maior número percentual de artigos sobre Física Geral, algo que não ocorre na REF, onde é possível ver mais publicações direcionadas às temáticas de Educação. No entanto, ambos os periódicos apresentam algumas tendências semelhantes, como o tópico de Física Moderna, que

está diminuindo ao longo do tempo. Outro ponto a se destacar é a ausência de um tópico relacionado à História da Ciência no modelo da REF. Esse resultado não significa a inexistência de artigos sobre essa temática na revista, mas sim que o modelo pode ter classificado esses artigos em outra categoria devido a palavras muito semelhantes em outros tópicos. Em geral, essa classificação em tópicos próximos ocorre quando um tópico dominante (que possui uma maior quantidade de artigos) possui palavras semelhantes a outro tópico menos dominante, e, portanto, o algoritmo entende que fazem parte da mesma categoria. Em outras palavras, é possível que artigos sobre a História da Ciência, que abordam a história da ótica ou da mecânica, tenham sido classificados dentro dos tópicos específicos de Ótica ou Mecânica Clássica, por exemplo.

A análise desse conjunto de dados sugere que os tradicionais periódicos latino-americanos avaliados e que agregam a pesquisa em Ensino de Física se distanciaram ao longo do tempo. Enquanto os artigos publicados na RBEF ainda se concentram mais em tópicos tradicionais da Física, a REF apresenta temas mais diversos. Desse modo, podemos dizer que, em certo sentido, a pesquisa em Ensino de Física, tal como retratada pela REF, está se alinhando com uma tendência internacional mais ampla da pesquisa em educação em ciências, que vem ampliando suas agendas de pesquisa para incluir discussões que englobam a sociedade e a diversidade cultural (Carter, 2008).

VI. CONCLUSÃO

O presente trabalho contribui para o entendimento da pesquisa em ensino de física no contexto latino-americano, considerando 2.549 artigos em dois tradicionais periódicos da área. O objetivo era identificar os principais tópicos e tendências presentes na RBEF e REF nos últimos anos, por meio da aplicação de um algoritmo de aprendizado de máquina não supervisionado. Ao final, algumas lições foram aprendidas.

Primeiro, é importante ressaltar certas limitações ao considerar o LDA como um mecanismo de classificação não supervisionado. Por exemplo, alguns artigos disponibilizados pela REF estão apenas no formato PDF, o que pode dificultar a extração de texto em fragmentos devido à estrutura original do documento. No entanto, apesar dessas limitações, o uso de modelos de inteligência artificial se mostra como um recurso interessante para apontar tendências em uma área de pesquisa, especialmente quando se avaliam milhares de trabalhos. Esses resultados reforçam a importância de acompanhar e compreender as mudanças nas tendências de pesquisa e ensino para orientar futuros estudos e práticas no campo da física. Pesquisas futuras podem abranger mais revistas da área, além de avaliar a congruência entre o trabalho realizado pelo LDA e o realizado por seres humanos a partir de uma amostra. Com a crescente disseminação das práticas que envolvem o aprendizado de máquina entre os pesquisadores, acredita-se que haverá uma significativa economia de tempo e esforço em estudos do tipo estado da arte.

Em segundo lugar, apesar de ambas as revistas consideradas serem homólogas, atualmente elas apresentam significativas diferenças nos principais tópicos, sugerindo perfis distintos. Enquanto na RBEF predominam temáticas da Física, na REF destacam-se tópicos ligados a métodos, experimentação, avaliação, educação e sociedade, caracterizando o ensino de física em sua complexidade. No fim, não parece haver uma força geral orientadora ou dominante que permeie a pesquisa em ensino de física latino-americana. Nesse contexto, mesmo que a pesquisa em ensino de física seja comumente tratada como uma unidade do campo da educação em ciências e que tenha alcançado maturidade institucional (Fontes e Rodrigues, 2022), é importante ressaltar a sua diversidade e pluralidade dentro dela própria.

AGRADECIMENTOS

O segundo autor agradece a bolsa de pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – Brasil, nº 140901/2022-1.

REFERÊNCIAS

- Bird, S. (2016). *Natural language processing with python*. O'Reilly Media.
- Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 3, 993–1022.
- Carter, L. (2008). Sociocultural influences on science education: innovation for contemporary times. *Science Education*, 92(1), 165-181. DOI: 10.1002/sci.20228

- Chen, X., Zou, D., Cheng, G. & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of *Computers & Education*. *Computers & Education*, 151, 103855. DOI: 10.1016/j.compedu.2020.103855
- Chollet, F. (2018). *Deep learning with Python*. Manning Publications.
- Fontes, D. T. M. y Rodrigues, A. M. (2022). Tendencias de investigación en la enseñanza de la física en revistas académicas iberoamericanas. *Revista De Enseñanza De La Física*, 34(2), 33–45. DOI: 10.55767/2451.6007.v34.n2.39481
- Goulart, B. N. K., Pastorio, D. P. e Vidmar, M. P. (2023). O papel do professor diante das tecnologias digitais de informação e comunicação no contexto do ensino remoto emergencial de Física e Ciências. *Revista De Enseñanza De La Física*, 35(1), 17–26. DOI: 10.55767/2451.6007.v35.n1.41387
- Mitchell, R. E. (2018). *Web scraping with Python: Collecting more data from the modern web* (Second edition). O’Reilly Media.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Odden, T. O. B., Marin, A. & Caballero, M. D. (2020). Thematic analysis of 18 years of physics education research conference proceedings using natural language processing. *Physical Review Physics Education Research*, 16(1), 010142. DOI: 10.1103/PhysRevPhysEducRes.16.010142
- Odden, T. O. B., Marin, A. & Rudolph, J. L. (2021). How has Science Education changed over the last 100 years? An analysis using natural language processing. *Science Education*, 105(4), 653–680. DOI: 10.1002/sce.21623
- Röder, M., Both, A. & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408. DOI: 10.1145/2684822.2685324
- Syed, S. & Spruit, M. (2017). Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation. 2017 *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 165–174. DOI: 10.1109/DSAA.2017.61