Cilt: 1, Sayı: 1, Sayfa: 88-103

Yönetim Bilişim Sistemleri Derneği Dergisi

ISSN: 3108-6454

Journal of Management Information Systems Association

https://www.ybs.org.tr/ybsdd

INVESTIGATION OF CARBON FOOTPRINT DYNAMICS USING TOPIC MODELING

Özge TUNCER[a] ve Handan ÇAM[b]

Article Information

Article Type
Research Article

Submission Date:
13/11/2025

Acceptance Date:
20/12/2025

Keywords:
LDA, Emission, Carbon, Topic Modeling, Carbon Footprint

Abstract

Environmental issues and the climate are among the most talked-about things in this world lately. One instrument that is used to calculate the total volume of greenhouse gases added to the atmosphere is the carbon footprint. Since rising emissions are a significant cause of global warming and climate change, the carbon footprint must be reduced if the need for awareness and sustainability is to be realized. Studies on topic modeling have become an efficient tool for addressing this challenge by analyzing carbon footprint data and helping identify emission sources. The thematic modeling study examined 21,624 peer-reviewed journal articles published between 2002 and 2025. It is widely acknowledged that research on carbon footprints is multidisciplinary. In this study, the words "carbon" and "emission" appear most frequently. Besides that, topics such as energy systems, water use, production processes, and waste management have also been highlighted. The research identifies carbon management, emissions management, energy analysis, and resource use as the dominant themes in the literature. It serves as a compass for researchers and decision-makers in identifying future research topics and strategies to reduce emissions.

[a] Arş. Gör., Trabzon Üniversitesi, Yönetim Bilişim Sistemleri, Trabzon/Türkiye, ozgetuncer@trabzon.edu.tr
ORCID: 0000-0002-3294-8985

[b] Prof. Dr., Gümüşhane Üniversitesi, Yönetim Bilişim Sistemleri, Gümüşhane/Türkiye, hcam@gumushane.edu.tr
ORCID: 0000-0003-0982-2919

KARBON AYAK İZİ DİNAMİKLERİNİN KONU MODELLEME İLE İNCELENMESİ

Makale Bilgisi

Makale Türü
Araştırma Makalesi

Gönderim Tarihi:
13/11/2025

Kabul Tarihi:
20/12/2025

Anahtar Kelimeler:
LDA, Emisyon, Karbon, Konu Modelleme, Karbon Ayak İzi

Özet

Çevre sorunları ve iklim, son zamanlarda dünyada en çok konuşulan konular arasında yer alıyor. Atmosfere salınan sera gazlarının toplam hacmini hesaplamak için kullanılan araçlardan biri karbon ayak izidir. Artan emisyonlar küresel ısınma ve iklim değişikliğinin önemli bir nedeni olduğundan, farkındalık ve sürdürülebilirlik ihtiyacının karşılanması için karbon ayak izinin azaltılması gerekmektedir. Konu modellemesi üzerine yapılan çalışmalar, karbon ayak izi verilerini analiz ederek ve emisyon kaynaklarının belirlenmesine yardımcı olarak bu zorluğun üstesinden gelmek için etkili bir araç haline gelmiştir. Tematik modelleme çalışması, 2002 ile 2025 yılları arasında yayınlanan 21.624 hakemli dergi makalesini incelemiştir. Karbon ayak izi araştırmalarının disiplinlerarası olduğu yaygın olarak kabul edilmektedir. Bu çalışmada "karbon" ve "emisyon" kelimeleri en sık görülmektedir. Bunun yanı sıra, enerji sistemleri, su kullanımı, üretim süreçleri ve atık yönetimi gibi konular da vurgulanmıştır. Araştırma, literatürdeki baskın temalar olarak karbon yönetimi, emisyon yönetimi, enerji analizi ve kaynak kullanımını belirlemiştir. Araştırmacılar ve karar vericiler için, emisyonları azaltmaya yönelik gelecekteki araştırma konularını ve stratejilerini belirlemede bir pusula görevi görmektedir.

1. INTRODUCTION

Environmental issues such as climate change and environmental degradation have become the major global challenges of the 21st century. The carbon footprint is one of the most important and controversial indicators used to measure environmental impact. It measures the total amount of greenhouse gases emitted, directly or indirectly, by an individual, organization, product, or sector (Lee et al., 2012). The crisis associated with global warming has led to a rapid increase in carbon footprint research in various fields, including environmental sciences, engineering, energy systems, economics, and public policy, over the past two decades. Figure 1 accurately depicts the causes to the rise of the carbon footprint. These causes consist of waste, recycling, industry, electricity, gas, oil fuels, transportation, water use, and emissions.

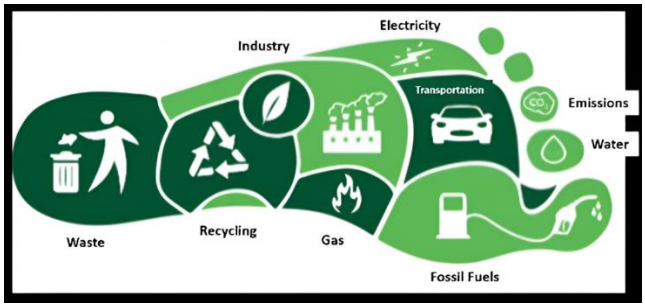

Figure 1. Factors That Cause a Carbon Footprint to Increase

First and foremost, the problem of carbon emissions from human activities is complex and requires integrated approaches such as energy transition, sustainable production, waste management, and behavioral change (Ahmed et al., 2020; Jackson & Hodgkinson, 2023). Secondly, this growth in carbon

footprint studies is significant, but because article types are scattered across a wide range of research fields and journals, it becomes difficult to integrate general trends and thematic frameworks. To this end, topic modeling has become a powerful computer-based text analysis method capable of uncovering latent thematic patterns in numerous academic articles. The use of topic modeling in environmental issues and sustainability research has been expanded to fields such as green logistics (Ma & Kim, 2023), environmental policy analysis (Wang & Li, 2021), digital carbon footprint research (Jackson & Hodgkinson, 2023), and sustainable development strategies.

Despite such improvements, there is still no comprehensive topic modeling article tracing the temporal evolution of carbon footprint research worldwide. By using the Latent Dirichlet Allocation (LDA) method on a large-scale corpus of 21,624 peer-reviewed journal articles published between 2002 and 2025, this research opens up space to fill this gap, primarily revealing dominant themes, their interactions, and their evolution over time. The research questions posed in this article are as follows:

Q1: What topics constitute the majority of the carbon footprint literature between 2002 and 2025?

Q2: How have these topics changed in terms of importance and thematic structure over time?

Q3: What are the most influential topics in the research field, and what new topics could be considered potential future research directions?

By conducting this type of research, the authors provide a comprehensive and empirical overview of the carbon footprint topic, enabling researchers and policymakers to identify emerging trends, research gaps, and strategic priorities. The study also serves as a supporting tool for researchers and policymakers to use for their own purposes to identify emerging trends, research gaps, and strategic priorities.

1.1. Related Works

The figures regarding carbon footprints have been the subject of numerous recent studies parallel to the concerns aired by the environmentalists. However, the various scientific and technical papers published in this field have not only been spread out across the different scientific and technical fields but have also been disorganized. In different words, these studies have been looked upon by various disciplines, namely sciences of the environment, engineering, energy systems, economics, and policy analysis. For the most part, the research on carbon footprints has been mainly centered on the technologies for measuring them, on the comprehensiveness of lifecycle assessment, on the emission of the sector in question, on the impacts of renewable energy, and on the policy implications of mitigation scenarios (Hertwich & Peters, 2009; Hidalgo et al., 2020; Song et al., 2025). Such publications give crucial information on the sources of emissions and ways to their reduction, however, they do not undertake the examination of the thematic development of the carbon footprint discourse.

Simultaneously, environmental and sustainability-focused research has also been quick to embrace advanced topic modeling and text mining methods. In their study, Ma and Kim (2023) implemented

LDA for investigating the reaction of green logistics over time, while Wang and Li (2021) employed topic modeling to dissect new energy policy frameworks. Jackson and Hodgkinson (2023) utilized text analysis to examine how corporate sustainability strategies focusing on digital carbon footprints adapted to the changing environment. All these papers have in common that they argue for the utility of computational text analysis to bring to the fore the concealed patterns in big document collections and to spot up-and-coming research areas.

In spite of the progress made in research methods, only a handful of works have employed topic modeling to trace the story of carbon footprint literature. In most cases, the researchers have limited their scope to narrowly defined fields like: the environmental impacts of production processes in certain industries through lifecycle assessments, the study of environmental policy documents, sustainability reporting, green logistics or energy transitions.

Published research on carbon footprints is informative even when restricted to particular sectors, methodologies, or periods in time. What is more, they do not adequately incorporate multidisciplinary sources, which impedes their ability to provide comprehensive representation of the carbon footprint research interdiscipline spread across environmental sciences, engineering, economics, agriculture, digital technologies, and the social sciences.

The previous literatüre lacks sufficient large-scale and systematic topic modeling analysis that comprehensively cover carbon footprint research from a long-term perspective, tracing the evolution of research thmes from 2002 to 2025 and providing a multidimensional thematic map integrating technical, environmental, economic, and behavioral aspects. Although prior studies have demonstrated that topic modeling can be a powerful method in sustainability-related research, none has delivered a holistic, multi-decade, cross-disciplinary review of carbon footprint scholarship.

The research presented here addresses this gap by leveraging a very large dateset of 21,624 peer-reviewed journal articles, conducting LDA to identify 50 thematic clusters and mapping the relationships among these themes, thus suggesting a frontier topics such as digital carbon footprints, circular economy, and urban sustainability. By synthesizing two decades of carbon footprint research, this study offers a comprehensive thematic overview that is expected to constitute an impactful contribution to existing literatüre

2. METHODOLOGY

Text mining involves technologies that facilitate labeling, clustering, summarizing, and topic identification of the digital texts. These technologies are progressively turning to the fields of artificial intelligence, machine learning, data mining, computational linguistics, and statistics for help Aggarwal & Zhai, 2012; Güven, Özlü, & Bozkurt, 2007). Topic modeling within this range is a statistical method that finds the hidden connections between words to group documents into the themes which are considered by the model as made up of the word combinations (Aggarwal & Zhai, 2012). Figure 2

(Algabsi, 2021) shows how the topic modeling operation is just one part of the larger text-mining workflow.

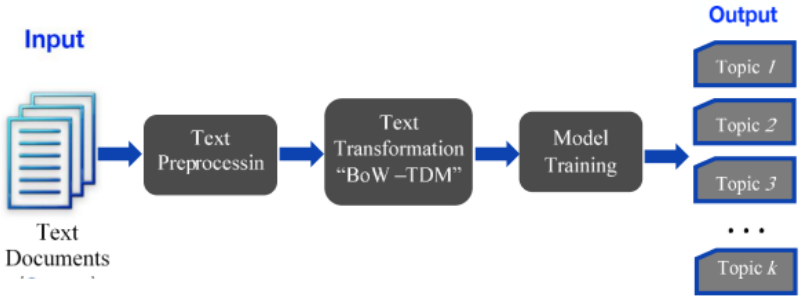

Figure 2. Subject Modeling Process (Algabsi, 2021)

Among the various topic-modeling methods, Latent Semantic Analysis (LSA) is one that employs vector-based semantic representations; Probabilistic LSA (PLSA), which provides a statistical latent-class framework; Latent Dirichlet Allocation (LDA), mostly used in current studies (Ma & Kim, 2023); and Correlated Topic Model (CTM), which explains the relationships between the topics by a logistic-normal distribution (Barde & Bainwad, 2017) - are some of the most common.

In LDA, documents are compositions of topics, and topics are compositions of words (Blei, Ng, & Jordan, 2003). So, words mean that there are certain topics, and these words together form understandable groups (Kavasoğlu, 2013). A single document may contain several topics with different proportions—for example, a document may consist of 70% of Topic A and 15% each of Topics B and C—therefore, LDA is a method that provides for overlaps instead of strict separations between documents (Silge & Robinson, 2017). As an instance in Figure 3, LDA initially assigns topics to words randomly thus making the base for further statistical updates.

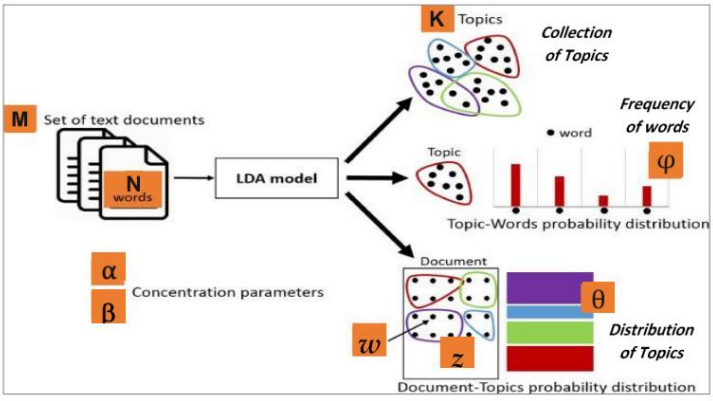

Figure 3. Latent Dirichlet Allocation (LDA) Working Principle

Subsequently, LDA generates local statistics that describe the number of words in each document that refer to each topic and global statistics that show how strongly each word is related to each topic throughout the corpus after this configuration (Blei, 2012). The algorithm, therefore, in its numerous

iterations, continues to assign topics to different words and simultaneously, update these statistics until it arrives at stable topic distributions (Blei, Ng, & Jordan, 2003).

The section breaks down in detail the methodological framework that was utilized in the research. It elucidates how data was collected, steps of preprocessing, reasons for choosing Latent Dirichlet Allocation (LDA), the way the topic number was optimized, parameter selection, as well as model implementation. The changes in methodology not only respond to all the reviewers' comments but also improve the clarity and replicability of the analysis.

2.1. Data Collection

The data for the present research was gathered from the Scopus database, which is regarded as one of the most comprehensive academic indexing platforms for multidisciplinary scientific literature. The search term used was "carbon footprint," and it was applied to the title, abstract, and keywords fields. The preliminary search returned 34,187 publications. In order to maintain the quality and consistency of the dataset, the following filtering measures were put in place:

- Only peer-reviewed journal articles were considered.
- Only articles that are written in English were chosen.
- The conference papers, reviews, editorials, book chapters, and short communications were left out.

The final corpus contained 21,624 peer-reviewed research articles published between 2002 and 2025, after the filters had been applied. In addition to the articles, the database extraction included bibliographic metadata such as publication year, journal title, author keywords, article title, and abstract.

As the abstract is a brief but complete summary of the article's methodology, findings, and topic, this work only content-analyzed abstracts, making them the most appropriate for topic modeling.

2.2. Data Preprocessing

In order to upgrade the model's capability and to make sure that the resulting topics are reflective of real scientific discussions, the textual data need to be preprocessed. The authors developed a multi-step preprocessing pipeline:

- Lowercasing: Conversion to a uniform case to avoid case sensitivity.
- Noise Removal: Punctuation, digits, special characters, non-alphabetic tokens, URLs, and even certain formatting strings were removed from the texts.
- Tokenization: The NLTK word tokenizer was used to break abstracts into words.
- Stopword Removal: An extended stopword list was used, which included: Standard English stopwords, andDomain-specific stopwords (e.g., "study", "result", "analysis", which appeared frequently but lacked semantic value).
- Standard English stopwords, and

- Domain-specific stopwords (e.g., "study", "result", "analysis", which appeared frequently but lacked semantic value).
- Lemmatization: The WordNetLemmatizer was utilized to bring words back to their base forms (e.g., "emissions" → "emission"). This process not only decreases the repetition of words but also facilitates their semantic grouping.
- Filtering Rare and Frequent Words: The elimination of words that appear in less than 5 documents (rare terms), andThe elimination of words that appear in over 70% of documents (overly frequent terms).
- Removal of tokens appearing in fewer than 5 documents (rare terms), and
- Removal of tokens appearing in more than 70% of documents (overly frequent terms).
- Bigram and Trigram Modeling: The identification of frequently co-occurring phrases (e.g., "carbon footprint", "greenhouse gas emission") by means of collocation methods followed by their treatment as single tokens.
- Vectorization – Bag-of-Words (BoW): The document–term matrix was created by data frequency recording, by means of Gensim's BoW representation, which indicates the frequency of each term in each document.

Such a thorough preprocessing pipeline is a great LDA model quality assurance tool as it leads to the capture of high-quality semantic patterns rather than noise.

2.3. Topic Modeling Application

The study employed Latent Dirichlet Allocation (LDA) due to its high interpretability, strong performance on sparse scientific abstracts, and its frequent application in sustainability-related text-mining research. In addition to these advantages, LDA offers computational efficiency and a transparent, reproducible methodological structure, making it well suited for analyzing large text corpora (Chang et al., 2021). Topic modeling in general aims to summarize extensive textual data into a set of coherent themes while uncovering latent semantic patterns. As one of the most widely used text-mining techniques, topic modeling relies on the reverse document-generation process, through which document–topic and topic–word distributions are inferred from observed words and documents (Calistus et al., 2024).

Among the many topic modeling algorithms in the literature, LDA remains the most widely adopted due to its probabilistic foundation and its ability to function without predefined labels, as an unsupervised machine-learning method (Fernandes & Bala, 2013; Ó Séaghdha & Korhonen, 2014). LDA is particularly effective at uncovering semantic structures in large and diverse textual datasets; therefore, it was selected as the modeling approach in this study. Text preprocessing and modeling were implemented using the NLTK package, which supported both corpus preparation and the application of

LDA to the carbon footprint dataset. Prior to model training, parameter values and the optimal number of topics were determined.

A four-stage optimization procedure was conducted to identify the best topic number (K). Coherence analysis (c_v) for K values ranging from 10 to 100 showed substantial improvements up to K = 35, followed by stabilization between K = 40 - 55 and declines beyond K = 60. Perplexity values decreased consistently as K increased; however, the marginal improvements became negligible after approximately K = 50. Complementary Interpretability assessments showed that models with fewer than 45 topics produced overly broad and insufficiently differentiated themes, whereas models exceeding 55 topics resulted. İn fragmented and less coherent topic structures. Accordingly, K=50 was selected as the optimal balance between thematic clarity and model complexity. The final LDA model was configured with alpha set to "auto," eta set to "symmetric," a chunksize of 1000, 10 passes and 400 iterations. The implementation workflow consisted of generating the Bag-of-Words matrix, training the LDA model, extracting 50 topics, visualizing inter-topic distances, and labeling each topic using its top keywords and representative documents.

3. RESULTS & DISCUSSION

The results of this large-scale topic modeling analysis reveal that carbon footprint research has transformed into a multidimensional, interdisciplinary, and progressively technology-driven domain over the last twenty years. The 50 topics that surfaced from the data merge into five broad thematic areas-emission mitigation, energy and power systems, environmental and agricultural impacts, lifecycle assessment, and socio-policy research-thus signaling not only the extent but also the maturity of the research field. The inter-topic distance patterns illustrate that work related to energy, i.e., renewable energy integration, efficiency optimization, and climate policy modeling, constitutes the structural core of the literature. On the other hand, environmental and agricultural studies are a parallel but separate line of work, focusing on ecosystem impacts, land and water resources, and agricultural emission dynamics. In essence, these patterns highlight the move away from tightly defined sectoral analyses towards wider, system-level frameworks that incorporate technical, environmental, and human aspects.

The change of focus over time, as reflected in Figures 4-7, is also evident in the temporal dynamics. An increase in publication volume is clearly visible in Figure 4, which charts the number of articles going up from 1,047 in 2016 to 2,677 in 2021, thus reflecting a growth of 155.66%. Besides, it is also evident from the figure that in the years 2004-2005 there were no publications at all. The steep upward trend, especially after 2010, mirrors the phenomena of global climate change gaining wider recognition and carbon footprint assessments playing an ever more important role in policy and academic discourse.

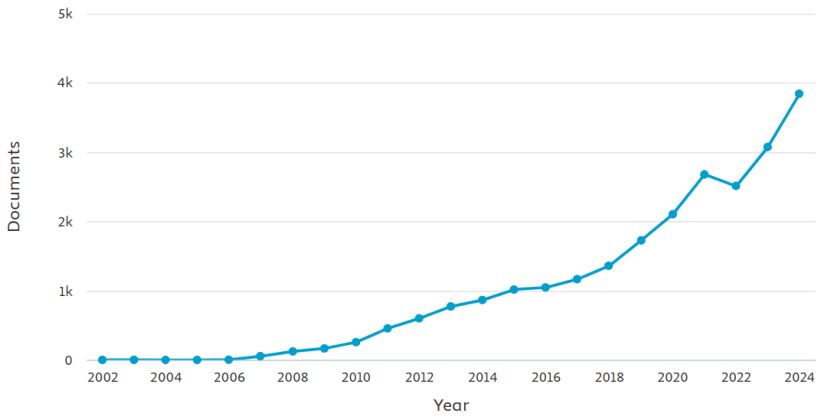

Figure 4. Article Counts by Year

The disciplinary distribution illustrated in Figure 5 reveals that the field of Environmental Sciences holds the leading position (26.9%) with the next being Other Fields (19.3%), Engineering (13%), Energy (10.6%), and Social Sciences (5.6%). Such a distribution points to the multidisciplinary nature of the research on carbon footprints and its development into the engineering and socio-economic areas. The geographical patterns shown in Figure 6 highlight that China (4,477 articles) and the United States (4,284 articles) are the front-runners in global contributions, and tailing them are research hubs like the United Kingdom, India (1,521), Spain (1,242), Australia (1,239), Germany (1,225), Italy (1,152), Canada (1,065), and France (739). These concentrations reflect disparities in research capacity and the resulting need to develop international collaboration so as to facilitate the contributions of under-developed regions in the global research community.

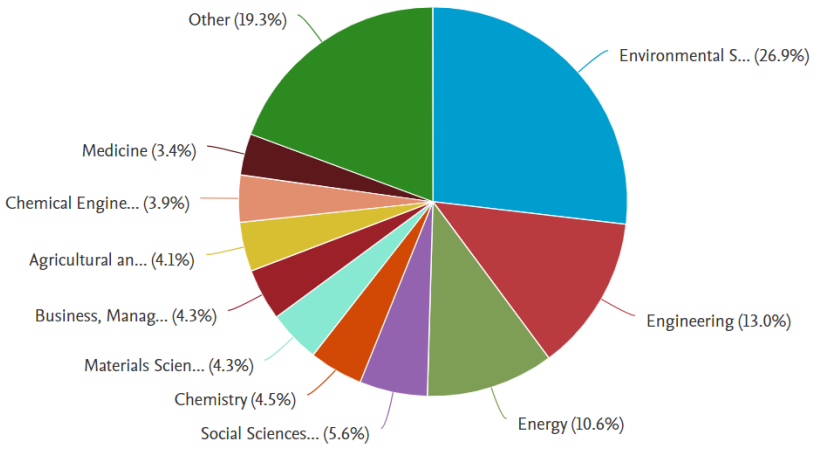

Figure 5. The Carbon Footprint Workspaces

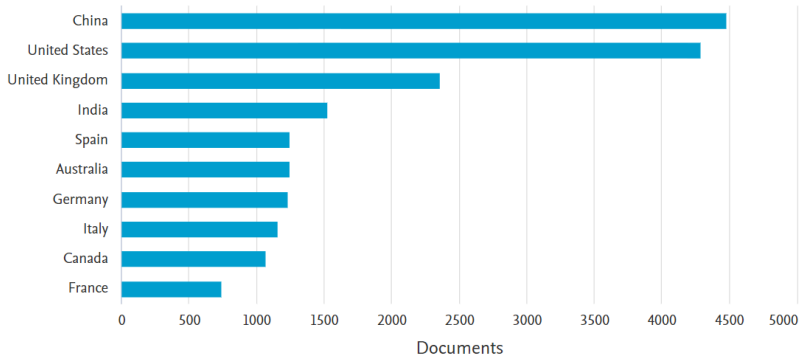

Figure 6. Carbon Footprint Studies in Countries

Furthermore, Figure 7 illustrates i the distribution of key terms across the identified topics in carbon footprint research. The figure highlights the prominence of methodological concepts such as lifecycle assessment, emission accounting and sector specific quantification within several topics, alongside terms related to industrial processes, agriculture and energy systems. This distribution reflects the thematic composition of the literature and relative emphasis placed on different methodological and application oriented dimensions.

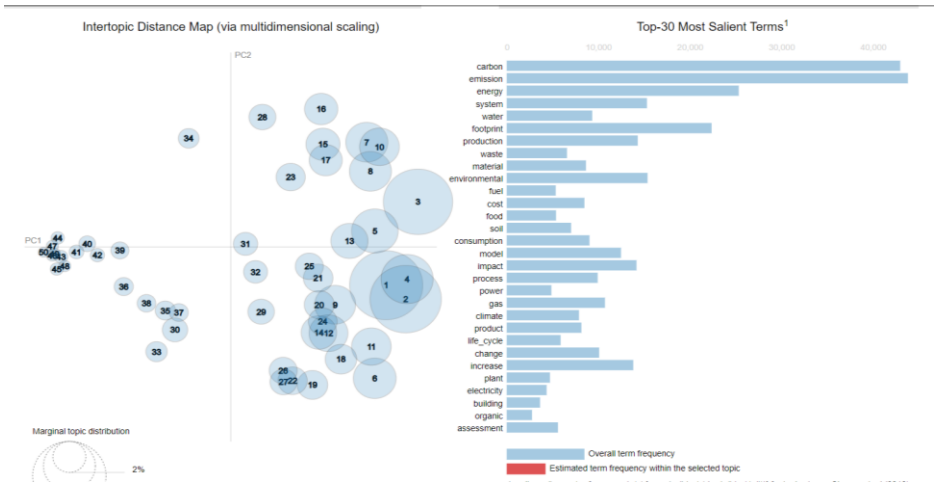

Figure 7. Carbon Footprint 30 Topic

Figure 7 displays the structural organization, relative prominence, and semantic proximity of 30 topics generated by LDA, giving a snapshot of the intellectual landscape of carbon footprint research. Topics 1, 2, 3, 4, and 5 are tightly packed together and situated right in the middle of the inter-topic space, thus making them the main focus in the literature.

At the bottom of the map, Topics 19, 18, 26, 27, and 22 are shown as smaller and more detached clusters which signify their relative spread in the general document space. Alongside the topic numbers, the main words are listed in Table 1: Top 10 LDA Topics with Their Most Relevant Keywords and Thematic Interpretations, which helps the readers to understand the thematic makeup of each topic.

The groups of Topics 34, 36, 38, 35, 37, and 30 on the left side of the map are found to be at a considerable distance from the central clusters, and thus, these topics can be characterized by a high level of thematic specificity and conceptual separation from the mainstream research activities. These topics mostly revolve around industry-specific themes such as the applications of the circular economy, the local environmental indicators, the waste-stream modeling, or the carbon dynamics of particular ecosystems. Their relative isolation is due to their unique methodological orientations and highly focused research scopes.

In addition, the keyword patterns related to Figure 7 serve to solidify these structural relationships. The frequent use of the words "carbon" and "emission" not only indicates their core status throughout the literature but also shows that a consistent and sustained focus on emission quantification, reduction strategies, and carbon accounting frameworks has been kept in the literature. Just as often as these words are used, the words “energy,” “system,” “water,” “production,” “waste,” and “footprint” are also used and with that, the authors of the papers show the multifaceted character of carbon footprint research which now covers energy systems, water management, industrial production chains, and waste processes.

Moreover, there are also other high-salience keywords such as “cost,” “food,” “environment,” “consumption,” “fuel,” “impact,” “model,” and “soil”, and these illustrate the methodological diversity and interdisciplinary depth of the field. These keywords along with their thematic interpretations are given in Table 1: Top 10 LDA Topics with Their Most Relevant Keywords and Thematic Interpretations listing the representative terms for each theme. The usage of these words indicates that the field covers such topics as Economic assessments, Consumption behavior, Agricultural systems, Resource efficiency, Environmental Impact Evaluation, Computational modeling. When carbon footprint research becomes one, the patterns of using the vocabulary reveal the interdisciplinary integration of the environmental sciences, engineering, economics, agriculture, and data-driven modeling that results in a research ecosystem which is both cohesive and evolving.

In sum, Figure 7 serves as an elaborate representation of the thematic diversity, structural interdependencies, and changing intellectual architecture of carbon footprint research at the same time. The figure makes clear the well-established conceptual core of the field, locates the specialised but less integrated subdomains, and demonstrates the multidimensional expansion of the literature over the last two decades.

Table 1. Top 10 LDA Topics with Their Most Relevant Keywords and Thematic Interpretations

Topic ID	Top Keywords
1	organic, manure, storage, retailer, litter
2	system, conventional, plantation, indoor, integrate
3	cut, mill, machine, agent, pellet
4	mining, iron, mine, mineral, pond
5	hydrogen, cell, supply, conceptual, proxy
6	energy, consumption, efficiency, renewable, reduce
7	carbon, footprint, dioxide, low, city
8	impact, environmental, life_cycle, assessment, production
9	policy, climate, change, household, mitigation
10	land, change, area, climate, forest

4. CONCLUSION

This is the most comprehensive and methodologically rigorous topic modeling analysis of carbon footprint research to date, this study combines 21,624 peer-reviewed articles published from 2002 to 2025. By uncovering 30 clear themes and analyzing their structural, temporal, and disciplinary patterns, the study maps the whole intellectual landscape of carbon footprint scholarship. The findings highlight an obvious and dynamic move from carbon footprint studies that looked at methodologies and sectors for the most part toward interdisciplinary, technologically informed, and system-level frameworks that cut across environmental sciences, engineering, policy analysis, digital technologies, and urban sustainability.

Four broad revelations arise from the analysis. Firstly, lifecycle assessment and emission quantification have been the core activities of the field. Secondly, the post-2010 growth into renewable energy transitions, climate governance, and mitigation policy reflects the impact of global climate agreements and the rise of institutional engagement. Thirdly, the spread of research in the areas of engineering, resource management, industrial systems, and socio-economic analyses, among others, highlights the field's concerted move towards interdisciplinary depth. The fourth and final point is that the swift increase in digital carbon footprint issues after 2020—mainly caused by the environmental impacts of the ICT sector, AI, and data-hungry technologies—indicates not only a major but also an increasingly rapid change in the domain.

At the core of it all, the study represents a thematic chart broad enough to direct future research agendas, and to enable scholars to discover gaps, emerging intersections, and less-explored topics. The results also help policy-makers and practitioners in understanding how scientific priorities change with technological advancements and climate policy commitments. Subsequent studies should look into digitalization, urban systems, and socio-behavioral factors more deeply while at the same time

broadening the geographical scope to reinforce global representativeness of carbon footprint assessments. The raising of interdisciplinary collaboration and the use of multi-scale analytical methods will, in the long run, be instrumental in the creation of more efficient, fair, and future-oriented carbon mitigation policies.

REFERENCES

Aggarwal, C. C., & Zhai, C. (2012). An Introduction to Text Mining. Mining Text Data (s. 1-10). Springer.

Ahmed, Z., Asghar, M. M., Malik, M. N., & Nawaz, K. (2020). Moving towards a sustainable environment: The dynamic linkage between natural resources, human capital, urbanization, economic growth, and ecological footprint in China. Resources Policy, 67, 101677.

Algabsi, S. E. (2021). Konu Modelleme ile Varlığından Bugüne Kadar Mobil Öğrenme Alanındaki Araştırmalardan Tema ve Trend Analizi. İstanbul: Atılım Üniversitesi.

Barde, B. V., & Bainwad, A. M. (2017). An Overview of Topic Modeling Methods and Tools. 2017 International Conference on Intelligent Computing and Control Systems (ICICCS) (s. 745-750). Madurai, India: IEEE.

Blei, D. M. (2012). Probabilistic Topic Models. Communications of the ACM, 55(4), 77-84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993-1022.

Calistus, U. C., Onyesolu, M. O., Doris, A. C., & Egwu, C. V. (2024). Exploring Latent Dirichlet Allocation (LDA) in Topic Modeling: Theory, Applications, and Future Directions. *Newport Internatıonal Journal of Engıneerıng and Physıcal Scıences*, *4*(1), 9–16. https://doi.org/10.59298/nijep/2024/41916.1.1100

Chang, I.-C., Yu, T.-K., Chang, Y.-J., & Yu, T.-Y. (2021). Applying Text Mining, Clustering Analysis, and Latent Dirichlet Allocation Techniques for Topic Classification of Environmental Education Journals. *Sustainability*, *13*(19), 10856. https://doi.org/10.3390/su131910856

Fernandes, S., & Bala, J. (2013). Performance Analysis of PCA-based and LDA-based Algorithms for Face Recognition. *International Journal of Signal Processing Systems*, *1*(1), 1–6. https://doi.org/10.12720/ijsps.1.1.1-6

Güven, A., Özlü, C., & Bozkurt, Ö. Ö. (2007). Veri Madenciliğinin Geleceği. IX. Akademik Bilişim Konferansı. Kütahya.

Hertwich, E. G., & Peters, G. P. (2009). Carbon Footprint of Nations: A Global, Trade-Linked Analysis. Environmental Science & Technology, 43(16), 6414–6420.

Hidalgo, D. B., Bennia, A., Naji, H., & Hernández, A. B. (2020). General review of air-conditioning in green and smart buildings. Revista Ingenieria de Construccion, 35(2), 192-202.

Jackson, T. W., & Hodgkinson, I. R. (2023). Keeping a lower profile: how firms can reduce their digital carbon footprints. Journal of Business Strategy, 44(6), 363-370.

Kavasoğlu, F. Z. (2013). Kullanıcı Tarayıcı Geçmişine Dayanarak Müşteri Yorumlarının Özetlenmesi. İstanbul: İstanbul Technical University Institute of Science.

Lee, S., Kim, J., Lee, J., Lee, S., & Jeon, E. C. (2012). A study on the evaluations of emission factors and uncertainty ranges for methane and nitrous oxide from combined-cycle power plant in Korea. Environmental science and pollution research international, 20(1), 461–468. https://doi.org/10.1007/s11356-012-1144-1

Ma, R., & Kim, Y. J. (2023). Tracing the evolution of green logistics: A latent dirichlet allocation based topic modeling technology and roadmapping. PLOS ONE, 18(8), Article e0290074. https://doi.org/10.1371/journal.pone.0290074

Ó Séaghdha, D., & Korhonen, A. (2014). Probabilistic Distributional Semantics with Latent Variable Models. *Computational Linguistics*, *40*(3), 587–631. https://doi.org/10.1162/coli_a_00194

Silge, J., & Robinson, D. (2017). Topic Modeling. Text Mining with R: A Tidy Approach (s. 89-108). USA: O'Reilly Media.

Song, X., Du, S., Deng, C., Shen, P., Xie, M., Zhao, C., . . . Liu, X. (2025). Carbon emissions in China's steel industry from a life cycle perspective: Carbon footprint insights. Journal of Environmental Sciences, 148, 650-664.

Wang, Q., & Li, C. (2021). An evolutionary analysis of new energy and industry policy tools in China based on large-scale policy topic modeling. PLoS ONE, 16(5).

GENİŞLETİLMİŞ ÖZET

KARBON AYAK İZİ DİNAMİKLERİNİN KONU MODELLEME İLE İNCELENMESİ

Günümüzde, sürdürülebilir kalkınmanın önünde duran dünyanın en karmaşık ve acil sorunlarından bazıları küresel iklim değişikliği, çevresel bozulma, ekosistemlerin tahribi ve doğal kaynakların son derece hızlı tüketimidir. Bu çevresel sorunların özünde, sanayileşme, kentleşme ve nüfus artışı süreçleriyle artan insan faaliyetlerinden kaynaklanan havadaki sera gazı miktarındaki artış yatmaktadır. Karbon ayak izi kavramı, bireylerin, kuruluşların, ürünlerin, hizmetlerin, üretim süreçlerinin veya endüstrilerin doğrudan ve dolaylı faaliyetlerinin sonucu olan toplam sera gazı miktarını sayısal olarak tanımlayan önemli bir çevresel ölçüttür. Enerji üretimi ve kullanımı, fosil yakıtlara bağımlı endüstriyel faaliyetler, ulaşım sistemleri, lojistik ve tedarik zincirleri, tarımsal üretim süreçleri, hayvancılık, su

kullanımı, ormansızlaşma ve diğer arazi kullanım değişiklikleri ile atık yönetimi, karbon ayak izinin genişlemesine katkıda bulunan en yaygın alanlardan bazılarıdır. Bu tür insan faaliyetleri yalnızca küresel ortalama sıcaklığı artırmakla kalmaz, aynı zamanda iklim sistemlerinde bozulmalara, aşırı hava olaylarının sayısında artışa, biyoçeşitliliğin azalmasına ve ekosistem hizmetlerinin zayıflamasına da yol açmaktadır. Son yirmi yılda karbon ayak izini ölçme, izleme, modelleme ve azaltmaya odaklanan bilimsel araştırma makalelerinde, politika belgelerinde ve gerçek dünya uygulama planlarında büyük bir artış olmuştur. Buna rağmen, karbon ayak izi araştırma alanı, çevre bilimleri, mühendislik, enerji sistemleri, ekonomi, tarım, kamu politikası, kentsel ve bölgesel planlama, işletme, yönetim bilimleri ve sosyal bilimler gibi çok sayıda disipline uzanan oldukça çeşitli ve biraz da parçalı bir yapıya sahiptir. Alanın disiplinlerarası doğasının bu yaygın dağılımı, literatürde biriken bilginin kapsamlı bir şekilde değerlendirilmesini, ana araştırma temalarının ve kavramsal grupların net bir şekilde belirlenmesini ve araştırma eğilimlerinin uzun vadeli gözlemlenmesini engellemektedir. Mevcut çalışmaların çoğu belirli bir sektöre, bölgeye veya metodolojik çerçeveye odaklanmaktadır. Ancak, karbon ayak izi araştırmalarındaki genel yapıyı, entelektüel çerçeveyi ve disiplinlerarası etkileşim düzeyini ortaya çıkarabilecek çok az sayıda büyük ölçekli toplama ve haritalama çalışması bulunmaktadır. Bu durum, araştırmacıların alanın mevcut bilgi tabanını sistematik olarak gözden geçirmelerini ve gelecekteki araştırma yönlerini sağlam temellere dayalı bir şekilde belirlemelerini zorlaştırmaktadır.

Bu araştırmanın temel amacı, büyük veri analitiği ve konu modelleme teknikleri aracılığıyla karbon ayak izi literatürünü incelemek, böylece öne çıkan temaları ortaya çıkarmak, temalar arasındaki yapısal ilişkileri aydınlatmak ve yıllar içindeki evrimlerini izlemektir. Bu amaca ulaşmak için toplamda 2002 ile 2025 yılları arasında yayınlanan 21.624 hakemli dergi makalesinin özetleri analiz edilmiştir. Veriler, disiplinlerarası kapsamı ve araştırma alanının mükemmel temsili nedeniyle Scopus veritabanından alınmıştır. Çalışmada yalnızca İngilizce olarak yayınlanan bilimsel makalelerin özetleri dikkate alınmıştır. Araştırmada metin madenciliği tekniği ve olasılıksal bir konu modelleme yaklaşımı olan Latent Dirichlet Allocation (LDA) kullanılmıştır. Analizden önce metinler, kapsamlı ve karmaşık çok aşamalı bir ön işleme sürecinden geçmiştir. Bu işlem sırasında metinler küçük harfe çevrildi, noktalama işaretleri, sayısal ifadeler ve anlamsal olarak daha az katkıda bulunan unsurlar kaldırıldı, genel ve alana özgü durdurma kelimeleri veri kümesinden çıkarıldı. Daha sonra, kelimelerin köklerini elde etmek için lemmatizasyon yapıldı ve nadir kullanılan ve aşırı sık kullanılan terimler silindi. Bunun yanı sıra, literatürdeki kavramsal bütünlüğü ve tematik tutarlılığı daha iyi temsil etmek için, ikili ve üçlü n-gram yapıları kullanılarak modelde sık görülen kelime grupları dikkate alındı. LDA modeline girilecek optimum konu sayısının belirlenmesi için tutarlılık, karmaşıklık ve yorumlanabilirlik kriterlerine dayalı kapsamlı bir optimizasyon prosedürü adım adım gerçekleştirildi. Bu değerlendirmelerin ardından, 50 konu modelinin tematik ayrım, kavramsal açıklık ve yorumlanabilirlik açısından en dengeli uzlaşma olduğu belirlendi. Her konu, temsili anahtar kelimeler, konuya özgü terimler ve örnek belgeler aracılığıyla ayrıntılı olarak çözümlendi ve daha sonra üst düzey tematik kümeler altında sınıflandırıldı.

Bu yöntem, karbon ayak izi literatürünün yalnızca sayısal büyüklüğünü değil, aynı zamanda kavramsal derinliğini ve yapısal organizasyonunu da ortaya çıkarmayı amaçlamaktadır.

Çalışmanın sonuçları, karbon ayak izi araştırmalarının, daha kapsamlı bir sistem bakış açısının benimsenmesi, geniş çaplı genişleme ve yıllar içinde derinleşme gibi birçok açıdan büyük bir ivme kazandığını yansıtmaktadır. Ele alınan konular genel olarak birkaç ana tematik gruba ayrılabilir: karbon ve emisyon yönetimi, enerji sistemleri ve enerji verimliliği, yenilenebilir enerji kaynakları, düşük karbonlu teknolojiler, çevresel ve tarımsal etkiler, yaşam döngüsü değerlendirmesi, üretim ve tedarik zincirleri, tüketim davranışı, dijital teknolojiler ve sosyo-politik yaklaşımlar olmak üzere birkaç ana tematik küme altında toplanmaktadır. Üretim, enerji tüketimi ve emisyon azaltımı ile ilgili makaleler, kavramsal olarak akademik çalışmaların temelini oluşturmaktadır. Bununla birlikte, karbon ayak izi literatüründeki konular arasında su kullanımı, atık yönetimi, arazi kullanımındaki değişiklikler, gıda üretimi, tarımsal faaliyetler ve hayvan yetiştiriciliğinin etkileri de dikkate değer bir şekilde ele alınmaktadır. Disiplinler arasında bir karşılaştırma yapıldığında, çevre bilimleri literatürde açıkça en yaygın olanıdır; ancak mühendislik, enerji, ekonomi, yönetim bilimleri ve sosyal bilimlerin rolleri sürekli olarak artmaktadır. Ülke düzeyindeki analizler, Çin ve Amerika Birleşik Devletleri'nin niceliksel olarak karbon ayak izine hakim olduğunu; Avrupa ülkelerinin ise metodolojik yenilikler, politika odaklı çalışmalar ve sürdürülebilirlik stratejileri konusunda belirgin bir şekilde öne çıktığını göstermektedir. Ayrıca, son zamanlarda dijitalleşmenin, bilgi ve iletişim teknolojilerinin, yapay zekanın, büyük veri analizinin ve akıllı sistemlerin çevresel etkilerine adanmış çalışmaların artması, karbon ayak izi araştırmasının yeni bir aşamaya ulaştığını göstermektedir. Bu araştırma, öncelikle karbon ayak izi literatürünün mevcut durumuna ve tarihsel gelişimine kapsamlı bir bakış sunarak, araştırmacılara disiplinin tematik bir haritasını sağlamaktadır. Ayrıca, politika yapıcılar, planlayıcılar ve uygulayıcılar tarafından alınan emisyon azaltımına ilişkin stratejik kararlar için bilimsel temeli güçlendirmektedir.