ARTIFICIAL INTELLIGENCE–BASED ANALYSIS OF DISCOURSES ON SUSTAINABLE AGRICULTURE
/
SÜRDÜRÜLEBILIR TARIM SÖYLEMLERININ YAPAY ZEKA TABANLI ANALIZI

Mehmet KAYAKUŞ[1)], Onder KABAS [2,*)], Valentin VLADUT [3,*)], Aylin KABAS [4)]

[1)] Department of Biomedical Engineering, Faculty of Engineering, Akdeniz University, 07058 Antalya, Türkiye; mehmetkayakus@akdeniz.edu.tr

[2)] Department of Machine, Technical Science Vocational School, Akdeniz University, 07070 Antalya, Türkiye; okabas@akdeniz.edu.tr

[3)] National Institute of Research—Development for Machines and Installations Designed for Agriculture and Food Industry—INMA Bucharest, 013813 Bucharest, Romania

[4)] Department of Organic Farming, Manavgat Vocational School, Akdeniz University, 07070 Antalya, Türkiye

Tel: +90-505-514-59-61; E-mail: okabas@akdeniz.edu.tr (O.K.)

Tel: +40-748-283-409; E-mail: vladut@inma.ro (V.V.)

DOI: https://doi.org/10.35633/inmateh-79-56

Keywords: *sustainable agriculture; social media analysis; artificial intelligence; text mining; sentiment analysis; topic modeling*

ABSTRACT

This study aims to investigate how sustainable agriculture is represented in the digital public sphere by examining technology-driven agricultural systems through artificial intelligence–based text analytics. A dataset of 13,354 English posts collected from the X platform during January 2026 was used, with 10,782 posts retained after preprocessing. The methodology integrates text mining, TF-IDF keyword extraction, BERT-based sentiment analysis, and LDA topic modelling. The optimal number of LDA topics was determined using coherence score evaluation to ensure statistical robustness and thematic interpretability. This study integrates text mining, BERT-based sentiment analysis, and LDA topic modelling within a unified artificial intelligence framework to provide a comprehensive analysis of sustainable agriculture discourse on social media. The results show predominantly neutral discourse (65.27%), reflecting informational content, while positive discourse (33.17%) highlights smart farming and innovation. Negative discourse (1.56%) primarily addresses structural challenges. Three dominant themes emerged: climate-oriented sustainability, community-based practices, and technology-driven agriculture, emphasizing the role of digital technologies in shaping sustainable agricultural systems. The findings provide practical insights for policymakers, agricultural stakeholders, and researchers by supporting evidence-based communication strategies and technology-oriented sustainability policies.

ÖZET

Bu çalışma, yapay zekâ tabanlı metin analitiği yöntemlerini kullanarak sürdürülebilir tarımın dijital kamusal alanda nasıl temsil edildiğini ve teknoloji odaklı tarımsal sistemlere ilişkin söylemin temel duygu ve konu yapılarını ortaya koymayı amaçlamaktadır. Ocak 2026 dönemine ait X platformundan 13.354 İngilizce paylaşım toplanmış, ön işleme sonrası 10.782 paylaşım analiz edilmiştir. Yöntemde metin madenciliği, TF-IDF anahtar kelime analizi, BERT tabanlı duygu analizi ve LDA konu modelleme kullanılmıştır. LDA için en uygun konu sayısı, istatistiksel güvenilirlik ve tematik yorumlanabilirlik dikkate alınarak coherence (uyumluluk) skoru yardımıyla belirlenmiştir. Bu çalışma, metin madenciliği, BERT tabanlı duygu analizi ve LDA konu modellemeyi bütünleşik bir yapay zekâ çerçevesinde bir araya getirerek sürdürülebilir tarım söylemine kapsamlı bir bakış sunmaktadır. Bulgular, söylemin ağırlıklı olarak nötr (%65,27) olduğunu, pozitif söylemin (%33,17) akıllı tarım ve inovasyonu vurguladığını göstermektedir. Negatif söylem (%1,56) ise ağırlıklı olarak yapısal sorunlara odaklanmaktadır. Analiz sonucunda üç temel tema belirlenmiştir: iklim odaklı sürdürülebilirlik, topluluk temelli uygulamalar ve teknoloji odaklı tarım. Elde edilen bulgular, sürdürülebilir tarım politikalarının geliştirilmesi, kanıta dayalı dijital iletişim stratejilerinin oluşturulması ve teknoloji odaklı sürdürülebilirlik uygulamalarının desteklenmesi açısından politika yapıcılara, sektör paydaşlarına ve araştırmacılara uygulamaya yönelik katkılar sunmaktadır.

INTRODUCTION

Sustainable agriculture has emerged as a critical framework for addressing the complex and interrelated challenges of global food security, climate change, and natural resource depletion. As agricultural systems are increasingly expected to produce more with fewer inputs, the transition toward sustainability requires not only ecological and economic adjustments but also the integration of innovative technologies and data-driven approaches. In this regard, the transformation of agriculture through digitalization and artificial intelligence has become a defining feature of contemporary agri-food systems.

In recent years, the convergence of sustainability principles with smart farming technologies—such as precision agriculture, artificial intelligence, Internet of Things (IoT), and data analytics—has reshaped the way agricultural production is managed and optimized. These developments align closely with the thematic focus of this special issue, which emphasizes the role of technological innovation in enhancing agricultural sustainability, efficiency, and resilience. Smart agriculture systems enable real-time monitoring, predictive analytics, and optimized resource allocation, thereby contributing to reduced environmental impact and improved productivity.

However, while the technological dimension of sustainable agriculture has been extensively explored, there remains a critical need to understand how these transformations are perceived, communicated, and socially constructed within the digital public sphere. Public discourse plays a vital role in shaping the adoption, legitimacy, and policy direction of sustainable agricultural practices. In this context, social media platforms function as dynamic environments where diverse stakeholders—including farmers, institutions, companies, and individuals—engage in discussions, share knowledge and construct narratives around sustainability and agricultural innovation.

The X platform provides a large-scale, real-time, and user-generated data environment for analyzing public discourse on sustainable agriculture. Unlike traditional data sources, social media data reflect immediate reactions, emerging trends, and multi-actor interactions, making them highly valuable for analyzing the societal dimension of agricultural transformation. Despite this potential, the integration of artificial intelligence–based text analytics into the study of sustainable agriculture discourse remains relatively underdeveloped in the literature.

Addressing this gap, the present study adopts a comprehensive artificial intelligence–based analytical framework to investigate social media discourse on sustainable agriculture. By combining text mining techniques, TF-IDF-based word frequency analysis, BERT-based sentiment analysis, and Latent Dirichlet Allocation (LDA) topic modeling, the study systematically examines large-scale Twitter data to uncover dominant sentiment patterns and thematic structures. This approach not only aligns with the technological orientation of the special issue but also extends it by incorporating the socio-discursive dimension of sustainability.

The contribution of this study is twofold. First, it provides empirical insights into how sustainable agriculture is framed in the digital public sphere, revealing the interplay between technology, sustainability narratives, and stakeholder perceptions. Second, it bridges the gap between artificial intelligence applications and sustainability research by demonstrating how advanced natural language processing methods can be employed to analyze complex, unstructured social media data in the agricultural domain.

LITERATURE REVIEW

Sustainable agriculture has evolved into a comprehensive and multidimensional framework that integrates environmental sustainability, economic viability, and social equity within agricultural systems. Early conceptualizations defined sustainable agriculture as a system capable of maintaining long-term productivity while minimizing environmental degradation and ensuring efficient use of natural resources *(Hobbs et al., 2008; Reganold et al., 1990*). Over time, this perspective has been expanded to include broader ecological and socio-economic dimensions, emphasizing the importance of soil health, biodiversity conservation, and integrated resource management *(Botero-Valencia et al., 2025; Brodt et al., 2011; Kumari et al., 2025; Velten et al., 2015*). Despite these advancements, the practical implementation of sustainable agriculture remains complex, as it requires balancing productivity with environmental protection and socio-economic inclusiveness across diverse regional and institutional contexts.

In recent years, sustainable agriculture has undergone a profound transformation driven by rapid technological advancements. The emergence of artificial intelligence (AI), machine learning, and digital agriculture has redefined traditional farming practices and introduced new opportunities for enhancing efficiency and sustainability. Studies indicate that AI-based systems can significantly improve decision-making processes, optimize resource allocation, and support climate-resilient agricultural practices *(Rejeb et al., 2025; Xu et al., 2024*). In addition, machine learning applications have been widely adopted in agricultural supply chains, enabling predictive analytics, yield optimization, and risk management (*Anggoro & Sari, 2025*). Bibliometric analyses further confirm that technological innovation has become a central theme in contemporary sustainable agriculture research, reflecting a shift toward data-driven and precision-based farming systems (*Liu & Wan, 2024; Tucciarone et al., 2024*).

However, while existing literature strongly emphasizes the technological dimension of sustainable agriculture, it often adopts a predominantly system-oriented or engineering-focused perspective. These studies typically prioritize efficiency, productivity, and optimization, while paying limited attention to how sustainability is perceived, communicated, and socially constructed by different stakeholders. This indicates a critical gap in literature, as the success of sustainable agriculture depends not only on technological advancement but also on societal acceptance, awareness, and engagement.

Parallel to these technological developments, social media has emerged as a significant platform for information exchange, knowledge dissemination, and stakeholder interaction in agricultural systems. Social media platforms enable farmers, policymakers, institutions, and the public to share information, express opinions, and construct narratives about agricultural practices and sustainability. Studies suggest that social media plays an important role in agricultural development by facilitating knowledge sharing and influencing farmers' decision-making processes (*Chowdhury et al., 2024*). Furthermore, research focusing on precision agriculture highlights that social media discourse reflects both opportunities and challenges associated with digital transformation, including issues related to accessibility, adoption barriers, and institutional constraints (*Sharma et al., 2023*).

Nevertheless, existing studies on social media in agriculture tend to focus primarily on functional or adoption-related aspects rather than examining the discursive and perceptual dimensions of sustainability. In other words, while these studies explore how social media contributes to agricultural development, they do not sufficiently address how sustainable agriculture is framed, interpreted, and communicated within digital discourse. This limitation restricts the understanding of the broader socio-cultural dynamics that shape sustainability narratives in the agricultural domain.

In addition to social media research, computational approaches such as text mining, bibliometric analysis, and topic modeling have gained increasing attention in recent years. These methods allow researchers to analyze large-scale textual datasets and identify underlying patterns, themes, and research trends (*Chiangnangam et al., 2024; Panda & Sinha, 2023*). Topic modeling techniques, particularly Latent Dirichlet Allocation (LDA), have been widely used to uncover latent thematic structures within agricultural and sustainability-related datasets (*Ofori & El-Gayar, 2021*). These approaches provide valuable insights into the evolution of research topics and thematic priorities, contributing to a more systematic understanding of agricultural knowledge production.

However, despite their analytical strengths, most computational studies remain limited in scope, as they often focus on either thematic structures or research trends without incorporating the emotional and perceptual dimensions of discourse. In particular, the integration of advanced sentiment analysis techniques with topic modeling remains underexplored in agricultural research. While sentiment analysis has been applied in various domains to capture user attitudes and emotional responses (*Driscoll & Walker, 2014*), its combined use with topic modeling to analyze agricultural discourse—especially in the context of social media—remains relatively limited.

This gap becomes particularly significant in the context of sustainable agriculture, where public perception, stakeholder attitudes, and narrative framing play a crucial role in shaping policy development, technological adoption, and long-term sustainability outcomes. Understanding not only what is being discussed but also how it is perceived and emotionally framed is essential for developing a comprehensive understanding of agricultural transformation in the digital age.

Although previous studies have investigated sustainable agriculture, social media, or artificial intelligence separately, limited research has integrated text mining, transformer-based sentiment analysis, and topic modelling within a unified analytical framework to examine sustainable agriculture discourse. Consequently, there remains a need for a comprehensive artificial intelligence–based approach capable of simultaneously identifying both the thematic structure and sentiment dynamics of digital sustainability discussions.

Addressing this gap, the present study adopts an integrated analytical framework that combines text mining, TF-IDF-based keyword analysis, BERT-based sentiment analysis, and LDA topic modeling. This approach enables a comprehensive examination of both the thematic and emotional dimensions of sustainable agriculture discourse on social media. Unlike previous studies, which typically focus on isolated analytical dimensions, this research provides a multidimensional and data-driven perspective on how sustainable agriculture is constructed, interpreted, and communicated within the digital public sphere.

The originality of this study lies in integrating text mining, BERT-based sentiment analysis, and LDA topic modelling within a unified artificial intelligence framework to investigate sustainable agriculture discourse on social media. Methodologically, the study demonstrates the applicability of advanced natural language processing techniques to agricultural communication research. Empirically, it provides evidence on how sustainable agriculture is represented, interpreted, and discussed in the digital public sphere, thereby contributing to both artificial intelligence applications and sustainability research.

Based on the identified research gap, this study addresses the following research questions:

RQ1. How is sustainable agriculture represented in social media discourse on the X platform?

RQ2. What are the dominant sentiment patterns associated with sustainable agriculture discourse?

RQ3. What are the major thematic structures underlying sustainable agriculture discussions identified through LDA topic modelling?

Given the exploratory nature of this study and its objective of identifying thematic and sentiment patterns rather than testing causal relationships, no formal research hypotheses were proposed.

MATERIALS AND METHODS

This study utilized a dataset comprising 13,354 English posts exchanged between January 1 and January 31, 2026, to analyze social media discourse on sustainable agriculture. After cleaning the data and preparing the text (removing duplicate content, cleaning URLs and special characters, changing everything to lowercase, removing stop words, and lemmatizing), 10,782 posts were ready for analysis. The study utilized TF-IDF-based word frequency analysis as a text mining technique, employing a pre-trained BERT model for sentiment classification. Additionally, Latent Dirichlet Allocation (LDA) was used for topic modelling to find hidden themes in the conversation about sustainable agriculture. Word clouds, density maps, and keyword co-occurrence networks were used to show the results. The Python programming language was employed for all data processing and analysis tasks.

All analyses were conducted using Python 3.14.6. The main libraries included pandas 3.0.5, NumPy 2.5, NLTK 3.13, scikit-learn 1.9, transformers 5.14.1, gensim 4.4.0, matplotlib 3.11, NetworkX 3.6.1, and WordCloud 1.9.6.

Dataset

The dataset for this study was generated using the keyword 'sustainable agriculture' to examine the contemporary digital discourse surrounding sustainable agriculture. This keyword was selected as it represents the most prevalent and all-encompassing articulation of the concept of sustainable agriculture in the literature, thereby facilitating access to a wide array of content that includes environmental, economic, and social dimensions (*Liu & Wan, 2024; Sharma et al., 2023*). The X platform was selected as the data source due to its widespread use in academic research. It shows real-time public reactions, policy discussions, project announcements, and individual opinions in an open data format (*Driscoll & Walker, 2014; Naskar et al., 2020*). X's official Application Programming Interface (API) infrastructure was utilized to retrieve the data. This approach ensures the scientific reliability of the data in terms of accuracy, integrity, and reproducibility *(Das et al., 2024*).

The data collection process lasts from January 1 to January 31, 2026. This time frame was chosen because it reflects current discourse, minimizes temporal fluctuations by providing a homogeneous single-month period, and allows for the observation of intensive content production on the topic of sustainable agriculture. During the specified time, 13,354 posts were collected. After text preprocessing and cleaning, 10,782 of these posts were used for analysis. In social media–based text mining studies, datasets exceeding ten thousand observations are generally considered sufficient to reliably model thematic structures and sentiment distributions. In this context, the acquired dataset is sufficiently large for statistical and analytical examination of the thematic and sentimental framework of the discourse on sustainable agriculture.

Text Mining

This study employed text mining techniques to facilitate the systematic and quantifiable examination of social media content related to sustainable agriculture. Text mining is an analytical method based on natural language processing (NLP) that tries to find useful patterns, themes, and statistical relationships in unstructured text data (*Lan, 2026*). Because social media data is inconsistent, short, and changes depending on the situation, it needs to be standardized before it can be analyzed (*Zucco et al., 2020*).

In this case, text preprocessing was done first. Text preprocessing was performed to improve data quality and ensure reliable subsequent analyses (*Chai, 2023*). First, duplicate posts were removed to eliminate repeated content. URLs, user mentions, hashtags, punctuation marks, numbers, and other non-textual characters were then removed to reduce noise. Subsequently, all text was converted to lowercase to ensure consistent word representation. Common stop words that did not contribute meaningful semantic information were removed using the NLTK stop-word list (*Mubarak et al., 2023*). Finally, lemmatization was applied to reduce words to their base forms while preserving their semantic meaning. These preprocessing steps reduced the dataset from 13,354 to 10,782 posts, resulting in a cleaner and more consistent corpus for text mining, sentiment analysis, and topic modelling.

Word frequency analysis is one of the most important methods used in text mining. Word frequency refers to the number of occurrences of a term within a dataset. It is used to figure out which ideas are most important in the conversation (*Yeomans et al., 2023*). Nevertheless, raw frequency values may underscore excessively repetitive general concepts, resulting in words with low distinctiveness appearing predominant in the analysis. This study also used the Term Frequency–Inverse Document Frequency (TF-IDF) method for this reason.

TF-IDF is a weighting scheme that quantifies the importance of a term within a document relative to the entire corpus (*Liu et al., 2022*). "Term Frequency (TF)" counts how many times a word shows up in a single document (*Mohammed & Rashid, 2023*), while "Inverse Document Frequency (IDF)" counts how many times the word shows up in all the documents (*Yulita et al., 2023*). The IDF value of a word goes down if it is used a lot in many documents (*Mee et al., 2021*). If it is only used in a few documents, the IDF value goes up. The TF-IDF score, which is the product of these two numbers, helps you find more meaningful keywords by looking at both their local frequency and their global distinctiveness (*Xiao et al., 2024*). Therefore, ideas that are different from each other in the discussion of sustainable agriculture have come to the fore.

Sentiment Analysis

Sentiment analysis is a way to use natural language processing to figure out what someone's attitude and sentimental state (positive, negative, or neutral) is based on text data (*Nandwani & Verma, 2021*). It is commonly employed to methodically categorize users' opinions, expectations, and criticisms based on social media data(*Chakraborty et al., 2020*). This study used a supervised machine learning methodology and employed a pre-trained deep learning model to analyze the sentiment distribution of posts regarding sustainable agriculture.

The pretrained transformer model cardiffnlp/twitter-roberta-base-sentiment-latest, based on the RoBERTa architecture, was used for sentiment classification (*Acheampong et al., 2021*). RoBERTa is an optimized variant of BERT that builds on the Transformer architecture and has demonstrated high performance in sentiment analysis tasks, particularly for short and context-dependent texts such as social media posts *(Pota et al., 2020; Mars, 2022*). Its bidirectional contextual representation enables more accurate interpretation of word meanings by considering both preceding and following contexts, thereby improving classification performance.

The BERT model is trained on big text datasets ahead of time, and it can be fine-tuned to work on a specific task (*Qasim et al., 2022*). The model was used in this study to categorize Twitter texts into positive, negative, and neutral classifications (*Acheampong et al., 2021*). The model breaks the text down into tokens, turning each word into a vector representation. It then learns how words are related to each other through Transformer layers and gives a probabilistic classification output in the last layer *(Rodrawangpai & Daungjaiboon, 2022*). This made it possible to find sentiment orientation in conversations about sustainable agriculture with outstanding accuracy.

Topic Modelling

Topic modelling is a statistical method for analyzing text that tries to find hidden topics in large sets of text (*Vayansky & Kumar, 2020*). This method uses probabilities to model how documents are spread out around certain topic clusters and what words make up each topic (*Churchill & Singh, 2022*). In unstructured datasets with a wide range of content, like social media, topic modelling techniques are a keyway to systematically show the thematic structure (*Laureate et al., 2023*). In this research, topic modelling was used to discern the prevailing themes in the discourse surrounding sustainable agriculture.

In this study, the Latent Dirichlet Allocation (LDA) algorithm was used for topic modelling. The primary reason for selecting LDA is its ability to model texts as encompassing multiple topics, delineating each topic through a distinct word probability distribution (*Li et al., 2024*). Social media texts usually have multidimensional content, LDA's flexible and probabilistic structure makes it a good way to analyze these kinds of data *(Neijzen & Lunter, 2023*). LDA was selected due to its prevalence in social science literature and its capacity to generate interpretable results.

The LDA model posits that each document (in this study, each post) comprises a blend of various topics in designated proportions (*Qiang et al., 2020*). The model works by using two main distributions: (1) the word distribution for each topic and (2) the topic distribution for each document (*Chauhan & Shah, 2021*). The algorithm figures out how likely it is that words belong to certain topics by repeating the process repeatedly (*Churchill & Singh, 2022*). Over time, it finds the most consistent topic-word structures. Therefore, keywords that frequently appear together represent each topic. This has made it possible to systematically and statistically find the main themes of sustainable agriculture conversation.

RESULTS

This part shows the results of the analysis for the dataset that was collected from January 1 to January 31, 2026. After the text pre-processing stage, it was cut down to 10,782 posts. The researchers employed text mining, word frequency analysis, BERT-based sentiment analysis, LDA topic modelling, and diverse data visualization techniques. The Python programming language was employed for all analysis stages. Libraries such as pandas, NLTK, scikit-learn, transformers, gensim, and matplotlib were utilized for data cleaning, natural language processing, modeling, and visualization. The results show, through a comprehensive approach, the themes and sentiment frameworks that shape the conversation about sustainable agriculture on social media. Table 1 shows the framework of sentiment that social media users are using to talk about agriculture.

Table 1

Distribution of sentiment categories in sustainable agriculture discourse on the X platform

Positive	Negative	Neutral
3.576	168	7.038

As presented in Table 1, of the 10,782 posts analyzed, 3,576 (33.17%) were classified as positive, 168 (1.56%) as negative, and 7,038 (65.27%) as neutral. The results suggest that neutral and informative content primarily influences the conversation about sustainable agriculture on social media. The fact that most of the content is neutral shows that the topic is mostly talked about in the context of policy announcements, project sharing, corporate statements, and research-based content.

The fact that 33.17% of posts are positive compared to negative content shows that sustainable agriculture is usually linked to a vision of the future that is hopeful, innovative, and solution orientated. On the other hand, the very low level of negative content (1.56%) shows that criticism is limited but focused on specific implementation problems and structural deficiencies.

These findings indicate that sustainable agriculture doesn't see it as a controversial crisis area but rather as an informative and somewhat positive topic of change and growth.

Qualitative Evaluation of Sample Posts

In this part, a qualitative evaluation was done by picking the three posts with the most likes, reposts, and comments in each sentiment category from the posts that BERT-based sentiment analysis said were positive, negative, or neutral. The goal is to go beyond just looking at the numbers and find out which narrative frameworks, thematic emphasis, and social sensitivities shape the conversation about sustainable agriculture.

The chosen posts exemplify the sentiment category and illustrate the discourse formats that yield the highest visibility and engagement among social media users. In this context, the analysis seeks to elucidate the perceptual and discursive aspects of sustainable agriculture within the digital public sphere with greater depth.

Positive Post Examples

Post 1: "Innovative sustainable farming systems are the future of global food security. Supporting farmers with smart technology will drive real growth and long-term resilience.

This post frames sustainable agriculture as a future-oriented and innovation-driven process, emphasizing the role of smart technologies in enhancing productivity and resilience. The reference to supporting farmers highlights the central role of producers in sustainable transformation.

Post 2: "New agricultural projects are empowering rural communities through sustainable solutions and economic development."

This post presents sustainable agriculture as a multidimensional development tool, linking environmental sustainability with economic growth and rural empowerment. It reflects a project-based and solution-oriented perspective.

Post 3: "Smart agriculture and innovative crop systems are transforming food production for a more sustainable and secure future."

This post emphasizes transformation and modernization, framing sustainable agriculture as a pathway toward a secure and resilient future through technological innovation.

Negative Post Examples

Post 1: "Excessive pesticide use and poor land management are harming sustainable farming efforts."

This post highlights the gap between sustainability discourse and actual practices, emphasizing environmental risks associated with excessive chemical use and poor management.

Post 2: "Small farmers still struggle with limited land and low production despite sustainable farming promises."

This post reflects structural inequalities in agricultural systems, particularly the challenges faced by small-scale farmers in achieving sustainable production.

Post 3: "Food waste and inefficient management systems undermine sustainable agriculture goals."

This post underscores inefficiencies in resource management, indicating that sustainability goals are undermined by waste and systemic limitations.

Neutral Post Examples

Post 1: "National research on sustainable agriculture focuses on food security, rural development and climate resilience."

This post reflects a policy-oriented and research-driven perspective, focusing on food security and climate resilience without evaluative judgment.

Post 2: "Sustainable agriculture initiatives support rural communities and strengthen global food systems."

This post presents sustainable agriculture within an institutional and systems-based framework, emphasizing its role in strengthening global food systems.

Post 3: "Agricultural development strategies address soil health, water management and long-term security."

This post highlights a technical and strategic perspective, focusing on resource management and long-term agricultural planning.

Word Frequency Analysis

This part did a word frequency analysis to find the main ideas that were talked about in the conversation about sustainable agriculture. The analysis process used both raw frequency values and the TF-IDF (Term Frequency–Inverse Document Frequency) method. The primary rationale for selecting TF-IDF is to mitigate the influence of common words that are excessively repetitive yet possess low distinctiveness (e.g., agriculture' or 'sustainable') within the dataset, thereby highlighting more unique and elucidative concepts related to specific sentiment categories or text clusters. This method evaluates both the frequency of terms and their distinctiveness across the corpus. The analysis aims to reveal a more substantial and thematically cohesive conceptual framework, free from superficial redundancies in the discourse on sustainable agriculture. Table 2 shows the words and frequency distributions that stand out by sentiment category.

Table 2

TF-IDF-based keyword distribution across sentiment categories

Positive	Negative	Neutral
Future	Waste	Agriculture
Innovative	Management	Sustainable
Growth	Small	Development
Project	Less	Security
Systems	Land	Rural
Support	Production	Research
Supporting	Nutrition	National
Solutions	Sustainable farming	World
Technology	Need	Health
Economic	Kerala	Resilience

The results indicate that positive discourse is strongly associated with future-oriented, innovation-driven, and solution-focused narratives. Keywords such as “innovation”, “technology”, and “growth” suggest that sustainable agriculture is framed as a driver of economic development and technological transformation, emphasizing systemic change rather than individual practices.

Negative discourse is primarily associated with structural and operational challenges, including resource inefficiency, inadequate management practices, and constraints faced by small-scale producers. These findings highlight a discrepancy between sustainability narratives and their practical implementation.

Neutral discourse is predominantly informational and institutional, characterized by terms related to policy, research, and development. This indicates that sustainable agriculture is largely framed within a technical and governance-oriented context rather than an evaluative one.

Topic Modelling Analysis

This section employed topic modelling analysis to uncover the underlying thematic structure of the discourse on sustainable agriculture (*Hussain et al., 2026*). The Latent Dirichlet Allocation (LDA) algorithm was used for the topical modelling process. The primary rationale for selecting LDA is its assumption that documents within extensive text datasets may probabilistically pertain to multiple themes, modelling each theme according to the distributions of co-occurring words (*De La Hoz-M et al., 2026*). This is a strong and well-known way to find hidden themes in datasets with multidimensional and mixed content structures, like social media. In addition, LDA makes it easier to analyze social science studies by giving them topical keyword outputs that are easy to understand (*Cheng et al., 2023*).

To identify the optimal topic structure, candidate LDA models with topic numbers ranging from 2 to 10 were evaluated using coherence scores. The corresponding coherence values are presented in Table 3.

Table 3

Coherence scores of candidate LDA models

Number of Topics	Coherence Score
2	0.3462
3	0.4710
4	0.3618
5	0.4526
6	0.4212
7	0.4315
8	0.4403
9	0.3376
10	0.4682

As shown in Table 3, the coherence scores ranged from 0.3376 to 0.4710 across the tested models. The three-topic model achieved the highest coherence score (0.4710), while the ten-topic model produced a very similar value (0.4682). However, qualitative examination indicated that the ten-topic solution generated fragmented and overlapping themes, whereas the three-topic model provided clearer thematic separation and greater interpretability. Models with fewer topics produced overly broad themes, while models with more topics resulted in increasingly fragmented topic structures. Therefore, considering both the quantitative coherence scores and the qualitative interpretability of the extracted topics, the three-topic model was selected for the subsequent analyses.

The findings of the topic modelling analysis are summarized in Table 4, which presents the extracted topics and their corresponding keyword distributions.

Table 4

Identified topics, prevalence, and representative keywords from LDA topic modelling

Topic N	Mean Topic Probability	Prevalence	Topmost co-occurring words
1	0.3900	40.18%	food, farming, systems, 2026, growth, security, innovation, support, farmers, future, development, today, climate, global, health, organic, agricultural, real, new
3	0.3102	30.11%	project, food, pulses, farmers, water, farming, security, innovative, Veera, on Veera, future, soil, communities, irrigation, world, practices, livelihoods, women, day, approach
2	0.2998	29.72%	farmers, farming, development, new, growth, energy, solutions, water, tech, production, land, food, ai, global, climate, farm, 2026, Kenya, soil

The table presents the dominant thematic structures identified through LDA, including topic prevalence and the most representative co-occurring terms.

Topic 1: The Future Vision of Sustainable Agriculture in the Context of Food Security and Climate

The fact that Topic 1 is the most common topic (40.18% of the time, with a mean topic probability of 0.3900) shows that posts about sustainable agriculture are most often about food security and future-orientated change. The words "food", "security", "climate", "global", "health", "organic", "innovation", "farmers", "future", and "development" often appear together. This shows that sustainable agriculture is linked to more than just production efficiency; it is also linked to a larger social agenda that includes healthy eating, fighting the climate crisis, global risks, and organic farming. The focus on "systems" and "innovation" shows that users see agriculture as more than just one thing; they see it as an agriculture-food system. The words "support" and "farmers" together suggest that producers are at the heart of the change. In terms of discourse, this topic sets up a framework that sees sustainable agriculture to build the future and deal with crises. It gives people hope and direction by saying, "Today's problem is tomorrow's solution.

Topic 3: Water, Soil and Community-Based Resilience: Innovative Projects and Sustainable Practices

Topic 3 (Prevalence: 30.11%; Mean Topic Probability: 0.3102) has a narrative that is more focused on projects and applications. Words like "project", "innovative", "practices", and "approach" suggest that sustainable agriculture is being talked about in terms of field applications and projects rather than as an abstract ideal. The three words "water, irrigation, and soil" are the main ideas of this theme. They stress that managing water and keeping soil healthy are important parts of sustainability. Words like "communities", "livelihoods", and "women" show that the conversation isn't just about the technical side of things. It also talks about sustainable agriculture in the context of rural livelihoods, social inclusion, and especially the role of women in agriculture. The language used in this topic supports the ideas of "solution-building" and "transformation". It shows that sustainable agriculture is seen as a practical way to develop that makes communities stronger, gives them more power, and helps them manage their resources better.

Topic 2: Efficient Production with Technology and Artificial Intelligence: Smart Agriculture in the Water-Energy-Food Nexus

Topic 2 (Prevalence: 29.72%; Mean Topic Probability: 0.2998) clearly moves the sustainable agriculture agenda toward technology and digitalization. The words "tech", "solutions", "AI", "energy", and "production" show that people are talking about sustainability goals through smart farming, AI-supported decision systems, and technology-based solutions. The inclusion of "energy" alongside "water, land, soil" signifies a growing public comprehension of sustainable agriculture through the lens of the water-energy-food (WEF nexus) paradigm, emphasizing the interrelations among resources. Also, the words "Kenya" and "global" show that the conversation has moved beyond being about a local production problem and is now being talked about in a way that compares it to examples from around the world and different country contexts. This topic frames sustainable agriculture as "producing smarter and more resource-efficiently" instead of "producing more". It focuses on the language of efficiency, optimization, and innovation, and it also connects these ideas to climate and food security issues, making technology an "accelerator" of sustainability.

Data visualization

This section employs diverse data visualization techniques to enhance the clarity and interpretability of findings pertaining to discourse on sustainable agriculture. Word clouds, density maps, and keyword co-occurrence networks were used to illustrate the prominent concepts based on sentiment categories and the underlying thematic clustering structures. The visualization process was done with the Python programming language, and the structural relationships of the text mining results were shown in a graph. This method backs up the results of the quantitative analysis and gives us a chance to look at the main ideas and the relational network structure of the sustainable agriculture discourse in a more complete way.

Word cloud analysis

The word cloud was made with the WordCloud library in Python. This library automatically looks at how often words appear in the text and makes images that are proportional to how often they occur *(Trejo, 2025*). The pandas and NLTK libraries were used for data pre-processing, and the matplotlib library was used to organize the graphical output. This combination made it possible to show word density both in numbers and in pictures. Figure 1 shows the positive word cloud, which shows how the ideas in the positive sentiment category are spread out visually.

Fig. 1 - Positive word cloud

The word cloud analysis indicates that positive discourse on sustainable agriculture is dominated by terms related to innovation, productivity, and sustainability. Keywords such as “innovation”, “technology”, and “development” highlight a strong association with future-oriented and solution-driven narratives. In addition, the presence of terms such as “climate”, “water”, and “soil” reflects the integration of environmental sustainability within this positive framing. The prominence of words such as “farmers” and “communities” further suggests a producer-centered and socially inclusive perspective. Overall, the results demonstrate that sustainable agriculture is constructed in positive discourse as a transformative and development-oriented process.

Fig. 2 - Negative word cloud

The negative word cloud highlights that critical discourse is primarily centered on environmental and structural challenges. Terms such as “pesticide”, “waste”, and “management” indicate concerns related to resource inefficiency and unsustainable practices. Additionally, the presence of words associated with climate and resource constraints reflects the perceived risks and limitations in the implementation of sustainable agriculture. Overall, negative discourse emphasizes the gap between sustainability goals and practical challenges.

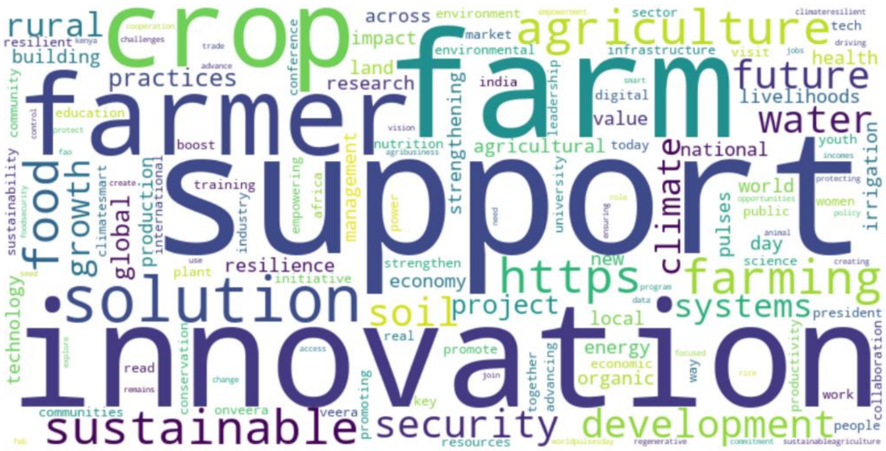

Fig. 3 - Neutral word cloud

The neutral word cloud reflects an informational and institutional discourse structure, dominated by terms related to policy, research, and development. Keywords such as “agriculture”, “development”, and “research” suggest that sustainable agriculture is primarily discussed within a technical and governance-oriented framework. This indicates that a significant portion of the discourse is descriptive rather than evaluative.

Density visualization

The density visualization was made to show how closely related and clustered concepts are in the discourse of sustainable agriculture (*Dabare et al., 2025*). For this analysis, words that often appeared together were placed on a two-dimensional plane to find areas of conceptual density. This made it possible to see which themes were at the center and which concepts were more strongly connected. This method has given us a more complete view of the main themes and environmental groups in sustainable agriculture conversation.

Figure 4 shows the density map that shows how concepts are grouped based on how similar their meanings are.

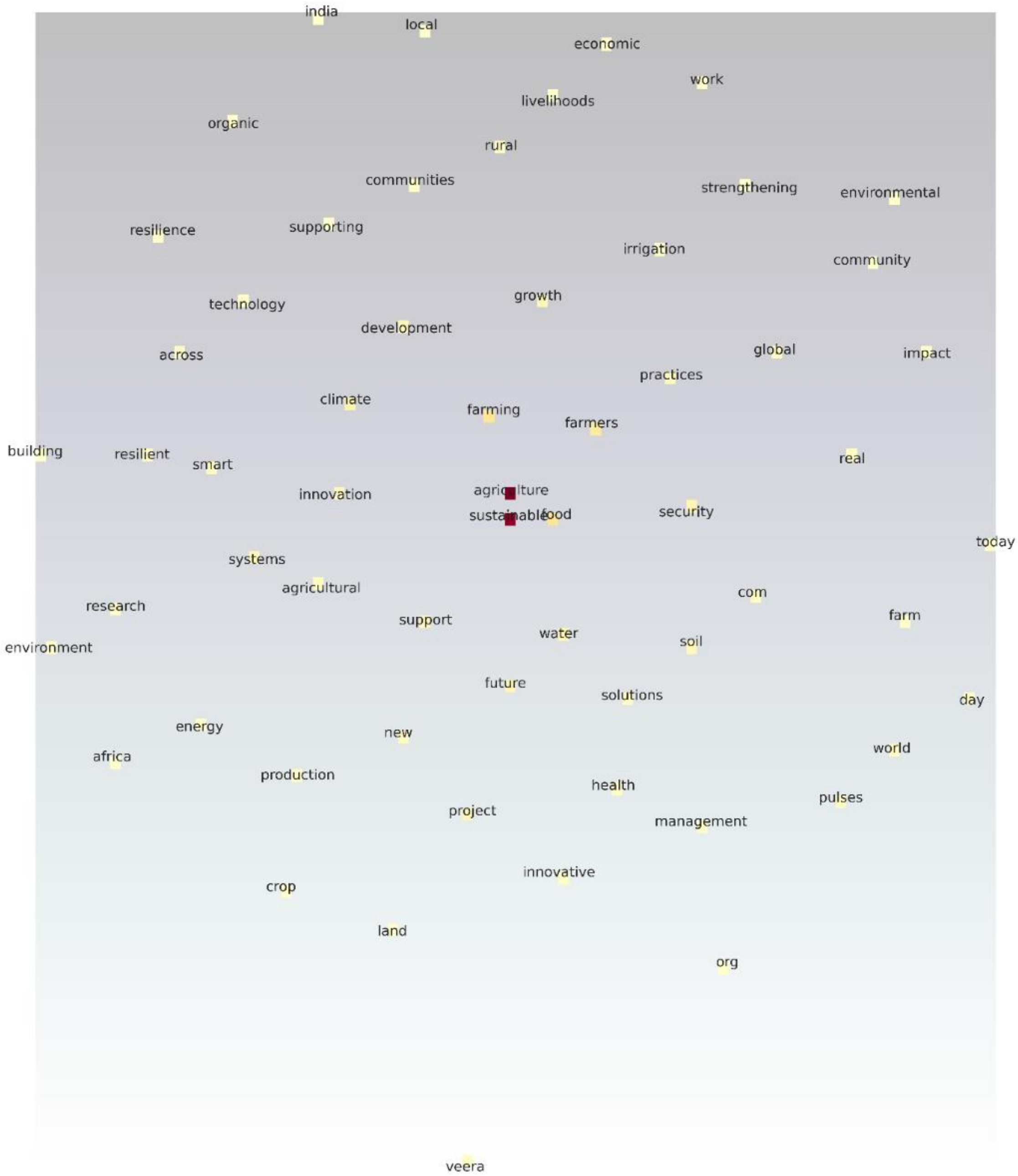

Fig. 4 - Density visualization

The density visualization reveals a clustered conceptual structure centered around key terms such as agriculture, food, and sustainability. This pattern reflects a multi-dimensional discourse shaped by environmental, economic, and social factors.

Keyword Co-occurrence Network

Keyword co-occurrence network analysis was conducted to examine how the concepts used together in the discourse on sustainable agriculture are interconnected. This approach used a node and edge structure to model words that co-occurred in the same post, facilitating an analysis of the connection density between concepts (*Yuan et al., 2022*). In this way, the main ideas that were talked about and the subthemes that went along with them were looked at in a visual network structure (*Yang et al., 2023*). The analysis took place in Python. The word 'co-occurrence matrix' was made first, then the NetworkX library was used to model the network structure, and finally Matplotlib was used to show the results. This method shows not only the ideas that shape the conversation about sustainable agriculture but also how these ideas fit together to make a whole. Figure 5 shows the keyword co-occurrence network, which demonstrates how keywords are related to each other and how the network is structured visually.

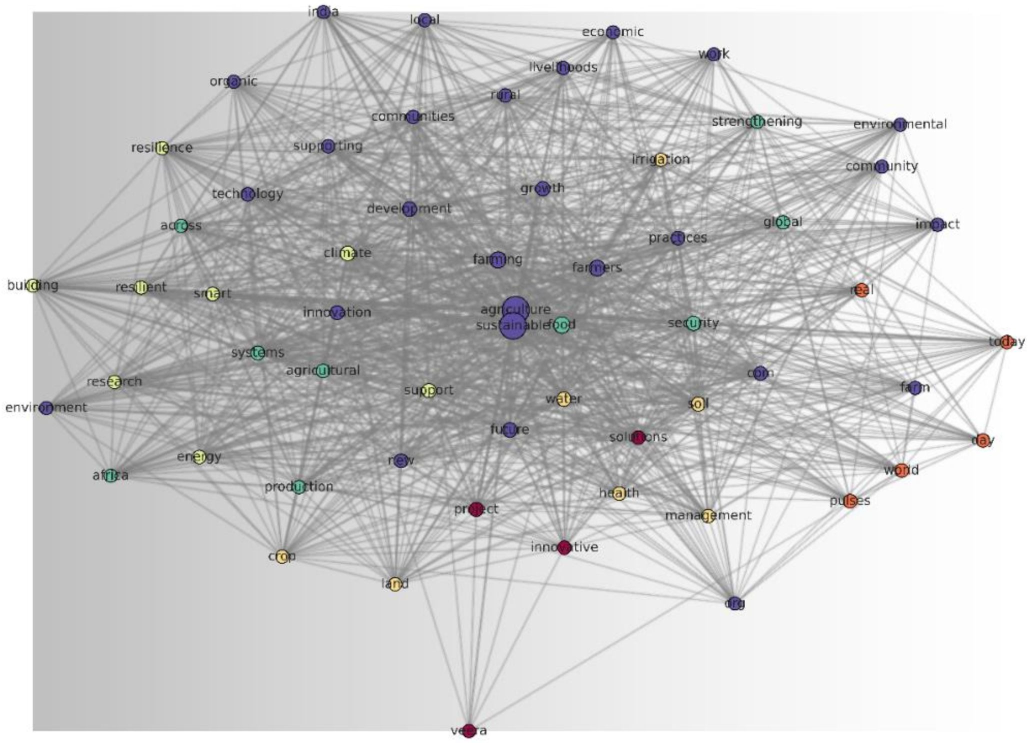

Fig. 5 - Keyword co-occurrence network

The keyword co-occurrence network demonstrates strong interconnections between sustainability, food systems, and technological innovation, indicating a highly integrated and systemic discourse structure.

CONCLUSIONS

This study provides a comprehensive analysis of sustainable agriculture discourse in the digital public sphere by integrating artificial intelligence–based methods with large-scale social media data. The findings reveal that discussions on sustainable agriculture are predominantly neutral and informational, largely shaped by institutional communication, policy-oriented content, and project-based narratives. At the same time, the significant presence of positive sentiment indicates that sustainable agriculture is strongly associated with innovation, technological advancement, and future-oriented development. In contrast, negative discourse remains limited and is primarily related to structural challenges such as resource inefficiency, management limitations, and the constraints faced by small-scale producers.

These findings are consistent with existing literature emphasizing the transformative role of artificial intelligence and digital technologies in sustainable agriculture (*Botero-Valencia et al., 2025; Kumari et al., 2025*). The prominence of themes such as food security, climate resilience, and smart farming aligns with studies highlighting the integration of technological innovation into sustainability frameworks (*Xu et al., 2024*). Furthermore, the dominance of informational discourse supports prior research suggesting that social media functions as both a communication and knowledge dissemination platform in agricultural systems *(Ofori & El-Gayar, 2021; Panda & Sinha, 2023*). However, unlike previous studies that primarily focus on technological applications or bibliometric patterns, this study demonstrates that sustainable agriculture is framed in digital discourse more as a narrative of transformation and opportunity rather than crisis. This shift toward solution-oriented and future-focused narratives represents a notable contribution to literature.

The originality of this study lies in its integration of artificial intelligence–based analytical methods with social media discourse analysis to examine sustainable agriculture from a socio-discursive perspective. By combining text mining, TF-IDF, BERT-based sentiment analysis, and LDA topic modelling, the study provides a multidimensional understanding of both the emotional and thematic structures of digital discourse. This integrative approach extends existing research by bridging the gap between technological innovation and public perception in the context of sustainable agriculture.

Despite these contributions, the study has several limitations. First, the analysis is based solely on data collected from the X platform, which may limit the generalizability of the findings across different digital environments. Second, the dataset was generated using the single keyword "sustainable agriculture," which may have excluded relevant discussions expressed through alternative keywords or hashtags. Third, only English-language posts were analyzed, limiting the representation of multilingual perspectives on sustainable agriculture. In addition, although standard text preprocessing techniques were applied to improve data quality, these procedures may have removed contextual information and linguistic nuances. Furthermore, despite the use of a pretrained transformer-based sentiment classification model, contextual ambiguity and occasional misclassification may still occur, particularly in short and informal social media posts. Therefore, the findings should be interpreted within the scope of the analyzed dataset and should not be generalized to all online discussions on sustainable agriculture.

Future research can build on this study by incorporating multilingual datasets and cross-platform data sources, including Facebook, Instagram, Reddit, and YouTube, to achieve a more comprehensive understanding of digital agricultural discourse. Longitudinal analyses could also be conducted to examine how perceptions of sustainable agriculture evolve over time. Furthermore, multimodal approaches integrating text, images, and videos may provide richer insights into online discourse. Future studies may also compare alternative topic modelling techniques, such as BERTopic, Non-negative Matrix Factorization (NMF), and Dynamic Topic Models, together with different sentiment analysis models, to further improve the robustness and interpretability of artificial intelligence–based discourse analysis.

This study highlights the importance of integrating artificial intelligence and social media analytics into sustainability research, offering new insights into how agricultural transformation is perceived, constructed, and communicated in the digital age. These findings provide valuable implications for policymakers, practitioners, and researchers, emphasizing the need for inclusive communication strategies, technology-driven solutions, and policy frameworks that address both innovation and structural challenges in sustainable agriculture.

REFERENCES

[1] Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review*, 54(8), 5789-5829.

[2] Anggoro, M. R. L., & Sari, C. A. (2025). Harnessing artificial intelligence and machine learning for sustainable agriculture in Asia: insights from a bibliometric review. *2025 IEEE Conference on Technologies for Sustainability (SusTech)*, Santa Ana, California, United States.

[3] Botero-Valencia, J., García-Pineda, V., Valencia-Arias, A., Valencia, J., Reyes-Vera, E., Mejia-Herrera, M., & Hernández-García, R. (2025). Machine learning in sustainable agriculture: systematic review and research perspectives. *Agriculture*, 15(4), 377.

[4] Brodt, S., Six, J., Feenstra, G., Ingels, C., & Campbell, D. (2011). *Sustainable Agriculture, Nature Education Knowledge*, 2. In.

[5] Chai, C. P. (2023). Comparison of text preprocessing methods. Natural language engineering, 29(3), 509-553.

[6] Chakraborty, K., Bhattacharyya, S., & Bag, R. (2020). A survey of sentiment analysis from social media data. *IEEE Transactions on Computational Social Systems*, 7(2), 450-464.

[7] Chauhan, U., & Shah, A. (2021). Topic modeling using latent Dirichlet allocation: A survey. *ACM Computing Surveys (CSUR)*, 54(7), 1-35.

[8] Cheng, X., Zhao, Y., & Liao, S. S. (2023). Key topics in social science research on COVID-19: An automated literature analysis. *Health Information & Libraries Journal*, 40(4), 343-358.

[9] Chiangnangam, S., Manorom, P., & Chansanam, W. (2024). Analyzing Research Trends in Organic Agriculture and Entrepreneurship: A Topic Modeling Approach Using Latent Dirichlet Allocation (LDA). *Journal of Computational Analysis & Applications*, 33(5).

[10] Chowdhury, S. R., Ghosh, D., & Rao, T. J. (2024). Sentiment and success potential of farmers' producer organizations: A systematic literature review. *Local Economy*, 39(1-2), 92-104.

[11] Churchill, R., & Singh, L. (2022). The evolution of topic modeling. *ACM Computing Surveys*, 54(10s), 1-35.

[12] Dabare, S., Rajapaksha, S., & Munaweera, I. (2025). Exploring the intersection of innovation management and nanotechnology in sustainable agriculture. *International Journal of Innovation Science*, 1-37.

[13] Das, N., Sadhukhan, B., Bhakta, S. S., & Chakrabarti, S. (2024). Integrating EEMD and ensemble CNN with X (Twitter) sentiment for enhanced stock price predictions. *Social Network Analysis and Mining*, 14(1), 29.

[14] De La Hoz-M, J., Montes-Escobar, K., Salas-Macias, C. A., Fors, M., & Ballaz, S. J. (2026). Using Latent Dirichlet Allocation Topic Modeling to Uncover Latent Research Topics and Trends in Renal Cell Carcinoma: Bibliometric Review. *JMIR Cancer*, 12(1), e78797.

[15] Driscoll, K., & Walker, S. (2014). Big data, big questions| working within a black box: Transparency in the collection and production of big twitter data. *International Journal of Communication*, 8, 20-20.

[16] Hobbs, P. R., Sayre, K., & Gupta, R. (2008). The role of conservation agriculture in sustainable agriculture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1491), 543-555.

[17] Hussain, M. R., Hannoun, O., Kovács, S., Mátyás, S., Orbán, S., & Szabados, G. N. (2026). Topic modelling and thematic analysis of research gaps and methodological innovations in rural sociology: A systematic review. *Social Sciences & Humanities Open*, 13, 102354.

[18] Kumari, K., Mirzakhani Nafchi, A., Mirzaee, S., & Abdalla, A. (2025). AI-driven future farming: achieving climate-smart and sustainable agriculture. *AgriEngineering*, 7(3), 89.

[19] Lan, M. (2026). Natural language processing and text mining. *In Artificial Intelligence for Drug Design* (pp. 189-215). Springer.

[20] Laureate, C. D. P., Buntine, W., & Linger, H. (2023). A systematic review of the use of topic models for short text social media analysis. *Artificial Intelligence Review*, 56(12), 14223-14255.

[21] Li, D., Wu, K., & Lei, V. L. (2024). Applying topic modeling to literary analysis: A review. *Digital Studies in Language and Literature*, 1(1-2), 113-141.

[22] Liu, H., Chen, X., & Liu, X. (2022). A study of the application of weight distributing method combining sentiment dictionary and TF-IDF for text sentiment analysis. *IEEE Access*, 10, 32280-32289.

[23] Liu, Y., & Wan, F. (2024). Unveiling temporal and spatial research trends in precision agriculture: A BERTopic text mining approach. *Heliyon*, 10(17).

[24] Mars, M. (2022). From word embeddings to pre-trained language models: A state-of-the-art walkthrough. *Applied Sciences*, 12(17), 8805.

[25] Mee, A., Homapour, E., Chiclana, F., & Engel, O. (2021). Sentiment analysis using TF–IDF weighting of UK MPs' tweets on Brexit. *Knowledge-Based Systems*, 228, 107238.

[26] Mohammed, M. T., & Rashid, O. F. (2023). Document retrieval using term frequency inverse sentence frequency weighting scheme. *Indonesian Journal of Electrical Engineering and Computer Science*, 31(3), 1478.

[27] Mubarak, H., Abdaljalil, S., Nassar, A., & Alam, F. (2023). Detecting and identifying the reasons for deleted tweets before they are posted. *Frontiers in Artificial Intelligence*, 6, 1219767.

[28] Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 81.

[29] Naskar, D., Singh, S. R., Kumar, D., Nandi, S., & Rivaherrera, E. O. d. l. (2020). Emotion dynamics of public opinions on twitter. *ACM Transactions on Information Systems (TOIS)*, 38(2), 1-24.

[30] Neijzen, D., & Lunter, G. (2023). Unsupervised learning for medical data: A review of probabilistic factorization methods. *Statistics in Medicine*, 42(30), 5541-5554.

[31] Ofori, M., & El-Gayar, O. (2021). Drivers and challenges of precision agriculture: a social media perspective. *Precision Agriculture*, 22(3), 1019-1044.

[32] Panda, R., & Sinha, B. (2023). An exploratory study on impact of social media on agriculture development from farmers growth perspective. *The Online Journal of Distance Education and e-Learning*, 11(2), 1966-1973.

[33] Pota, M., Ventura, M., Catelli, R., & Esposito, M. (2020). An effective BERT-based pipeline for Twitter sentiment analysis: A case study in Italian. *Sensors*, 21(1), 133.

[34] Qasim, R., Bangyal, W. H., Alqarni, M. A., & Ali Almazroi, A. (2022). A fine-tuned BERT-based transfer learning approach for text classification. *Journal of Healthcare Engineering*, 2022(1), 3498123.

[35] Qiang, J., Qian, Z., Li, Y., Yuan, Y., & Wu, X. (2020). Short text topic modeling techniques, applications, and performance: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(3), 1427-1445.

[36] Reganold, J. P., Papendick, R. I., & Parr, J. F. (1990). Sustainable agriculture. *Scientific American*, 262(6), 112-121.

[37] Rejeb, A., Rejeb, K., & Hassoun, A. (2025). The impact of machine learning applications in agricultural supply chain: a topic modeling-based review. *Discover Food*, 5(1), 141.
[38] Rodrawangpai, B., & Daungjaiboon, W. (2022). Improving text classification with transformers and layer normalization. *Machine Learning with Applications*, 10, 100403.
[39] Sharma, S., Sharma, C., Asenso, E., & Sharma, K. (2023). Research constituents and trends in smart farming: an analytical retrospection from the lens of text mining. *Journal of Sensors*, 2023(1), 6916213.
[40] Trejo, J. A. O. (2025). Basic Uses of Python for Research in Our Laboratory: Examples of Data Analysis, and Processing of Histological Samples Based on the Special Graduate Lecture for Postgraduate Students Presented on January 2025. *Juntendo Medical Journal*, 71(4), 231-237.
[41] Tucciarone, I., Secci, G., Contiero, B., & Parisi, G. (2024). Sustainable aquaculture over the last 30 years: An analysis of the scientific literature by the Text Mining approach. *Reviews in Aquaculture*, 16(4), 2064-2076.
[42] Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, 94, 101582.
[43] Velten, S., Leventon, J., Jager, N., & Newig, J. (2015). What is sustainable agriculture? A systematic review. *Sustainability*, 7(6), 7833-7865.
[44] Xiao, L., Li, Q., Ma, Q., Shen, J., Yang, Y., & Li, D. (2024). Text classification algorithm of tourist attractions subcategories with modified TF-IDF and Word2Vec. *PloS One*, 19(10), e0305095.
[45] Xu, J., Li, Y., Zhang, M., & Zhang, S. (2024). Sustainable agriculture in the digital era: Past, present, and future trends by bibliometric analysis. *Heliyon*, 10(14).
[46] Yang, K., Lu, J., Duan, L., Tang, H., Bao, Z., Liu, Y., & Jiang, X. (2023). Research hotspots and theme trends in post-traumatic growth: A co-word analysis based on keywords. *International Journal of Nursing Sciences*, 10(2), 268-275.
[47] Yeomans, M., Boland, F. K., Collins, H. K., Abi-Esber, N., & Brooks, A. W. (2023). A practical guide to conversation research: How to study what people say to each other. *Advances in Methods and Practices in Psychological Science*, 6(4), 25152459231183919.
[48] Yuan, C., Li, G., Kamarthi, S., Jin, X., & Moghaddam, M. (2022). Trends in intelligent manufacturing research: a keyword co-occurrence network based review. *Journal of intelligent manufacturing*, 33(2), 425-439.
[49] Yulita, W., Untoro, M. C., Praseptiawan, M., Ashari, I. F., Afriansyah, A., & Pee, A. B. C. (2023). Automatic Scoring Using Term Frequency Inverse Document Frequency Document Frequency and Cosine Similarity. *Sci. J. Informatics*, 10(2), 93-104.
[50] Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H., & Cannataro, M. (2020). Sentiment analysis for mining texts and social networks data: Methods and tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(1), e1333.