EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Exploring Social Media Research Trends in Malaysia using Bibliometric Analysis and Topic Modelling

Rehan Tariq[1], Pradeep Isawasan[2], Lalitha Shamugam[3,*], Muhammad Akmal Hakim Ahmad Asmawi[2], Noramira Athirah Nor Azman[2], and Izzal Asnira Zolkepli[1]

[1] School of Communication, Universiti Sains Malaysia, Malaysia
[2] College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Perak Branch, Malaysia.
[3] Department of Business Analytics, Sunway Business School, Sunway University, Malaysia.

Abstract

This study explores the evolving dynamics of social media research in Malaysia. The main objective is to identify trends and patterns in research, specifically examining the volume and focus of scholarly articles over the last decade. Using bibliometric analysis and topic modelling, the study identifies major research clusters and key themes like digital marketing, political communication, and public health, and to map out collaborations among researchers. The findings show a significant increase in social media-related studies, highlighting a trend towards more varied and complex topics. This includes a greater emphasis on social media's role in political communication, consumer behaviour, and crisis management. Looking forward, this study suggests that future studies should explore the applications of emerging technologies such as artificial intelligence (AI) on social media practices, assess the spread and impact of negative information such as fake news and hate speech, and extend cross-disciplinary methodologies to fully understand the extensive effects of social media.

Keywords: social media, social networking sites, bibliometric analysis, topic modelling, Malaysia.

Received on 20 08 2024, accepted on 05 12 2024, published on 02 01 2025

doi: 10.4108/eetsis.7003

1. Introduction

Social media's global adoption has dramatically transformed interactions among individuals, communities, and society. Initially, it enabled users to create, share, and exchange content, promoting active participation and communication. [1]. The simplicity and utility of these platforms spurred widespread adoption, granting users greater control over information dissemination and consequently influencing their behaviors and attitudes [2]. Owing to these benefits, social media has been widely adopted by the Malaysian population and plays a pivotal role in reshaping communication and transforming social dynamics. According to Data Reportal's 2024 study [3], Malaysia's social media user base has grown to 28.68 million, comprising 83.1% of the total population. These statistics highlight social media as Malaysia's main platform for online engagement, showing significant changes in media consumption.

Recent studies have shown that social media significantly impacts various aspects of Malaysian society. For example, researchers found that Malaysians heavily relied on digital platforms such as websites, social media, and mobile applications during the COVID-19 pandemic for information and social support [4]. Similarly, studies have revealed that social media platforms play a major role in political engagement among Malaysians, influencing public opinion and participation [5]. Additionally, some researchers have highlighted that businesses in Malaysia are increasingly using social media marketing strategies to reach wider audiences,

*Corresponding author. Email: lalithas@sunway.edu.my

especially among the youth [6]. Furthermore, new directions in this field have explored the emergence of artificial intelligence in social media analytics [7], [8]. Other studies have investigated concerns regarding misinformation and hate speech on Malaysian social media platforms, emphasizing the need for effective strategies to address these issues [9], [10], [11]. The rapid increase in usage and research underscores the need for a bibliometric analysis to understand the changing patterns and effects of social media in Malaysia.

2. Objective

This bibliometric analysis aims to provide a comprehensive understanding of research trends in social media within Malaysia. The study's objectives are as follows:

- To analyse the growth and evolution of publications in social media research within Malaysia from 2011 to 2024, identifying key trends and patterns over this period.
- To map and evaluate the intellectual collaboration networks among Malaysian researchers in the social media domain, identifying key authors and their interconnections.
- To employ topic modeling to identify and categorize underlying themes and research gaps in Malaysian social media studies.

3. Study's Significance

This study is important because it provides a clear understanding of how social media research has grown and changed in Malaysia from 2011 to 2023. By analyzing the number of publications, the study highlights key trends and emerging areas such as artificial intelligence, misinformation, hate speech, and privacy concerns. Mapping the collaboration networks among Malaysian researchers shows how they work together, which can lead to stronger partnerships and better research outcomes. Using BERTopic modeling to identify themes and research gaps helps to uncover important topics that need more attention. The findings of this study have several benefits: for academia, it guides future research and helps develop relevant courses; for industry, it informs effective social media strategies and marketing efforts; and for policy-makers, it provides insights needed to create regulations that address issues like misinformation and privacy. Overall, this research contributes valuable information that supports the growth and improvement of social media practices and policies in Malaysia.

4. Methodology

Following the established bibliometric techniques outlined by [12], the methodology for this study is divided into performance analysis and science mapping. Performance analysis centers on measuring productivity and impact, utilizing metrics like publication counts and citation numbers to identify significant contributors within a field. In contrast, science mapping examines the relationships between research entities, employing technique such as author collaboration network. The topic modeling process, as described by [13], involves transforming text data, such as titles and abstracts from research papers, into coherent topic clusters. This process includes several key steps: embedding the documents to capture semantic meaning, reducing dimensionality for efficient processing, clustering similar documents, and applying a customized term-weighting method to emphasize important terms within each topic. Together, these steps allow the model to generate clear and interpretable topic representations, making it possible to identify meaningful themes within large sets of text data. The bibliometric methodology for this study is illustrated as Figure 1.

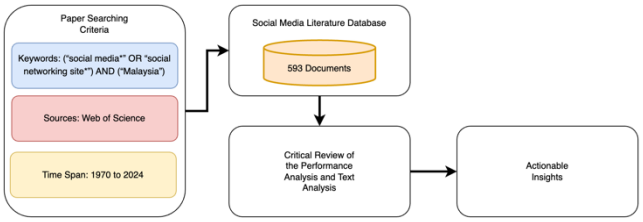

Figure 1. Bibliometric Methodology

4.1. Data Collection

Data was gathered from the Web of Science (WoS) database, recognized as a leading citation database [14]. The literature search was conducted on 19 May 2024 using the WoS Core Collection, including Science Citation Index Expanded (SCIE), the Social Science Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), and the Emerging Sources Citation Index (ESCI). The search focused on titles, abstracts, and keywords, capturing variations of social media and Malaysia-related terms, ensuring thorough retrieval of literature from 1970 to 2024. (see Table 1)

Table 1. Initial Search Criteria

Parameter	Details
Collection	Web of Science Core Collection
WoS Editions	Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Emerging Sources Citation Index (ESCI)
Search Field	Topic (titles, abstracts, keywords)
Search String	("social media*" OR "social networking site*") AND ("Malaysia")
Date Range	All years (1970-2024)
Date	19 May 2024

The search strategy initially yielded 766 records. After applying additional filters (see Table 2), only 'Article', 'Review Article', and 'Early Access' documents in English were included, excluding 'Proceeding Paper' and 'Data Paper'. This refined approach resulted in a focused dataset of 593 records for analysis.

Table 2. Inclusion and Exclusion Criteria

Parameter	Filter
Document Type	Include: Article, Review Article, Early Access Exclude: Proceeding Paper, Data Paper
Language	English
Country	Malaysia

4.2. Data Analysis

The study exports data as a ".bib" file from the Web of Science (WoS) repository, which is then analyzed using Python. Bibliometric analysis is performed using the pyBibX library [15] and topic modeling is performed with the BERTopic model from the bertopic library [13]. For the bibliometric analysis, basic data cleaning procedures are implemented, such as merging duplicate author names into a single record. The data preprocessing steps for the topic modeling are detailed in Section 7. Access to the ".bib" dataset and the list of generated topics with GPT interpretation are available via the following link [16]. The corresponding Python Notebook, which includes the code and related visualizations, can be provided upon request.

5. Performance Analysis

5.1. Exploratory Data Analysis (EDA)

Table 3 summarizes scholarly contributions from 2011 to 2024, covering 63 countries and 841 institutions with 593 documents in English (569 articles, 13 articles in press, 11 review articles). The 2,013 authors averaged 1.21 documents per author and 3.57 per institution, with an annual output of 42.36 documents. High collaboration is noted with 547 multi-authored documents and a collaboration index of 4.05. Total citations reached 5,861, averaging 9.88 per document, with a highest H-index of 6, reflecting a collaborative and impactful research environment.

Table 3. EDA Report

Parameters	Details	
Timespan	2011 - 2024	
Total Number of Sources	317	
Total Number of References	28894	
Total Number of Documents	593	
	Article	569
	Article in Press	13
	Review	11
Average Documents per Author	1.21	
Average Documents per Year	42.36	
Total Number of Authors	2013	
Total Number of Authors' Keywords	2026	
Total Number of Authors Keywords Plus	1114	
Total Single-Authored Documents	46	
Total Multi-Authored Documents	547	
Average Collaboration Index	4.05	
Max H-Index	6	
Total Number of Citations	5861	
Average Citations per Author	2.91	
Average Citations per Document	9.88	

5.2. Documents Per Year

Figure 2 illustrates the growth of social media research publications in Malaysia from 2011 to 2024, based on the WoS database. Starting with 2 documents in 2011, the number of publications steadily increased, reaching 14 in 2014, 23 in 2016, and 48 in 2018. Significant growth continued, peaking at 120 documents in 2022, likely driven by the digital transformation during the COVID-19 pandemic. There is a slight decrease to 104 documents in 2023, and 29 documents have been recorded as of May 2024. This recent data might not indicate a decline as the trend could increase. Overall, the trend shows a decade of increasing interest and activity in social media research in Malaysia, with potential for further growth.

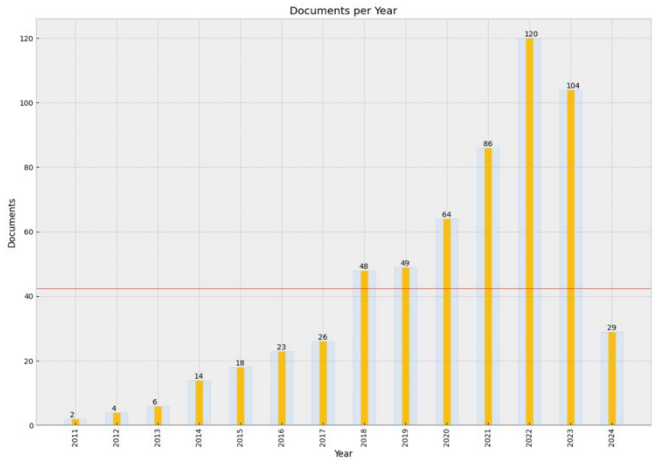

Figure 2. Documents per Year

5.3. Citations Per Year

Figure 3 shows citation data from 2011 to 2024. Citations began with 8 in 2011, rising sharply to 204 in 2012, indicating early interest. The numbers fluctuated, reaching 688 in 2015, dropping to 309 in 2016, then peaking at 849 in 2018. The dip to 599 citations in 2020 might reflect the COVID-19 pandemic's impact. Citations peaked at 877 in 2021, then declined to 533 in 2022 and 107 in 2023. The low count of 12 citations in 2024 likely reflects the data's cut-off in May, and the trend may increase as the year progresses. Overall, the data highlights increasing academic interest and evolving research impacts in social media.

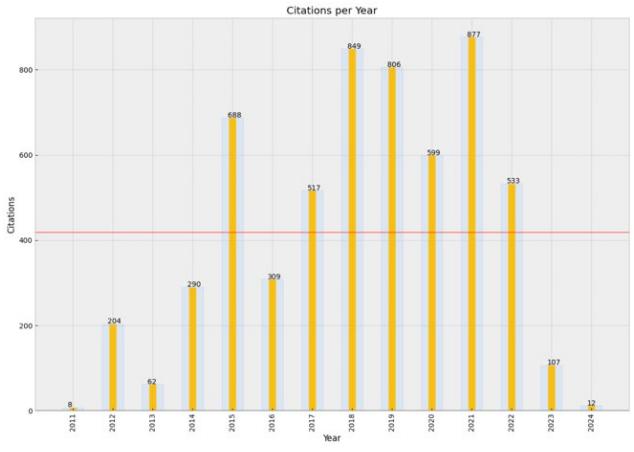

Figure 3. Citations per Year

6. Author Analysis

6.1. Authors per Document

The landscape of social media research in Malaysia, showcases a diverse array of scholars. Leading the field is Ramayah, T., with 10 documents, indicating a deep engagement with social media research. Following closely are Al Mamun, Abdullah, and Abrizah, A., each with 9 documents, contributing significantly to the field. Norazah Mohd Suki, Nicholas David, Al-Rahmi Waleed Mugahed, and Ab Rashid Radzuwan each have 6 documents. These scholars represent diverse research areas: Ramayah's extensive body of work, Al Mamun's exploration of economic, social, and educational implications, Abrizah's focus on information science and management, Suki's intersection with consumer behavior and marketing, David's interdisciplinary perspective, Al-Rahmi's emphasis on educational technology, and Radzuwan's study of communication practices. Collectively, their research spans economic effects, educational applications, communication practices, and marketing dynamics, significantly contributing to the understanding of social media in Malaysia. Analyzing their outputs provides valuable insights into how social media shapes various aspects of Malaysian life.

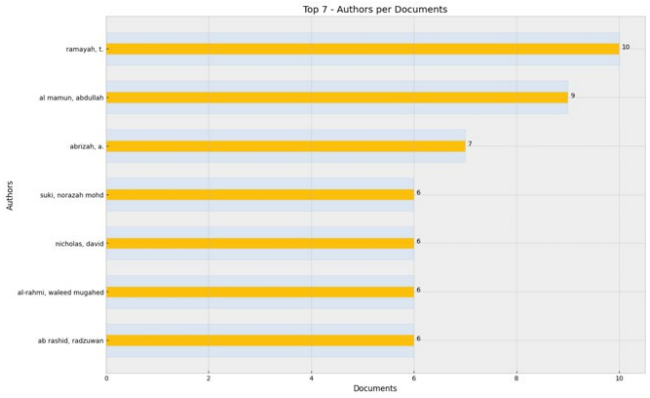

Figure 4. Authors per Document

6.2. Authors Productivity

Examining the productivity of several authors over a range of years reveals patterns that reflect their scholarly contributions and their possible academic career trajectories, as shown in Figure 5. Ramayah, T. has shown a steady increase in productivity, starting with one paper per year in early years, peaking at three papers in 2023. Al Mamun, Abdullah began publishing in 2017, with moderate increases, peaking at four papers in 2022 and one in 2023. Abrizah, A. has maintained a consistent output since 2013, with minor fluctuations between one and two papers annually. Suki, Norazah Mohd shows occasional output from 2011 to 2021, publishing one paper in select years with gaps in between. Nicholas, David

exhibited bursty patterns with peaks in 2017 and consistent output of one paper in other years. Al-Rahmi, Waleed Mugahed has shown steady productivity since 2018, with a consistent pace of one paper per year. Ab Rashid, Radzuwan had an initial steady output starting in 2014, with peaks in 2016 and 2018, and slower phases in between. Overall, the varying trends highlight different phases of academic careers and the nature of research projects.

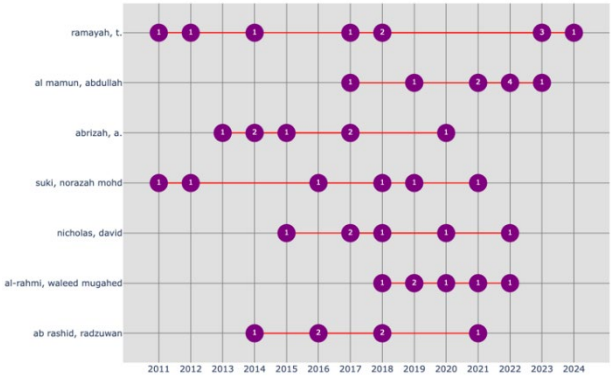

Figure 5. Authors Productivity

6.3. Authors Collaboration

The analysis of collaboration patterns among authors reveals distinct collaborative networks generated using the Girvan-Newman community algorithm, setting a threshold for relationships at a minimum of seven joint publications, as shown in Figure 6.

- Cluster 0: Ramayah, T. frequently appears in this cluster, marking him as a leading figure in this field.
- Cluster 1: Features Al Mamun, Abdullah, with multiple collaborations with Malarvizhi, Chinnasamy Agamudainambhi, Ahmad, Ismail, and others.
- Cluster 2: Centers around Siau, Ching Sin, collaborating with Fitriana, Mimi, Ab Rahman, Norny Syafinaz, and others.
- Cluster 3: Involves Nor, Nor Azlida Mohd, collaborating with Nazari, Nor Shafina Mohamed, Bahar, Aufa Dahlia, and others.
- Cluster 4: Centers around Cheah, Jun-Hwa, with connections to Sia, Bee-Chuan, Ameen, Nisreen, and others.
- Cluster 5: Led by Abrizah, A., with extensive collaborations including Badawi, Fathiah, Boukacem-Zeghmouri, Cherifa, and others.
- Cluster 6: Centers around Al-Rahmi, Waleed Mugahed, with broad connections to Al-Sharafi, Mohammed A., Alenazy, Wael M., and others.

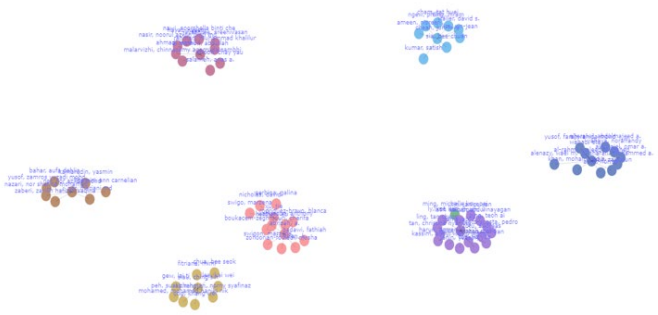

Figure 6. Collaboration Network between Authors

Refer to Table 4 for detailed cluster information. This table contains information from Figure 6, including the cluster number, the color representing each cluster, the author-defined label, and the author with the highest number of collaborators (n).

Table 4. Authors Cluster with their Label

Cluster	Color	Label*	Author (n)
0	Purple	Social Media Influence on Marketing, Education, and Behavioral Studies	Ramayah, T (16)
1	Dark Red	Social Media Influence on Business, Marketing, and Consumer Behavior	Al Mamun, Abdullah (9)
2	Yellow	Social Media Impact on Health Perceptions and Behaviors	Siau, Ching Sin (8)
3	Dark Brown	Social Media and Digital Health Communication in Dental Care	Nor, Nor Azlida Mohd (7)
4	Light Blue	Social Media Influence on Consumer Behaviour Trend and Digital Marketing	Cheah, Jun-Hwa (8)
5	Light Red	Social Media and Scholarly Communication in Academia	Abrizah, a. (9), Xu, Jie (8)
6	Dark Blue	Social Media and Technology Integration in Education	Al-Rahmi, Waleed Mugahed (10)

*Please note that the labels assigned to each cluster within this analysis, are derived from the authors interpretation of the common topics presented within the cluster's publications.

7. Keyword Analysis

7.1. Authors Keyword Importance

The keyword analysis of social media research in Malaysia, shown in Figure 7 and Table 5, highlights significant trends. “Covid” (0.2145) is the most frequent keyword, reflecting a strong focus on the pandemic's impact. “Health” (0.1652) examines public health information on social media, while “Communication” (0.1507) highlights social media's role in interpersonal and mass communication, especially during crises. “Networking” and “Facebook” (0.1246 each) underscore the importance of social networking platforms. “Marketing” (0.1130) and “Brand” (0.0928) indicate interest in social media's commercial uses. “Political” (0.0957) relates to political discourse, while “Education” (0.0812) focuses on e-learning and student engagement. “Sentiment” (0.0638) and “Vaccine” (0.058) reflect studies on public opinion during the pandemic. “Twitter” (0.0551) and “tourism” (0.0464) explore Twitter's role in tourism. “Environmental” and “addiction” (0.0377 each) highlight concerns about environmental issues and social media addiction. “Mental” (0.0319) and “Fake” (0.0348) address mental health and fake news, while “Privacy” (0.029) focuses on data privacy. Overall, the analysis shows that social media research in Malaysia covers public health, communication, marketing, political discourse, education, and mental health, along with privacy and misinformation concerns.

Figure 7. Authors Keyword Cloud

Table 5. Top 10 Keyword Importance

No.	Keyword	Importance (Probability)
1.	covid	0.2145
2.	health	0.1652
3.	communication	0.1507
4.	networking	0.1246
5.	facebook	0.1246
6.	marketing	0.1130
7.	political	0.0957
8.	brand	0.0928
9.	education	0.0812
10.	sentiment	0.0638

7.2. Authors Keyword Evolution

The bibliometric analysis of social media research in Malaysia over the past five years (2019-2024) reveals notable trends in the evolution of author keywords, indicating shifts in research focus and external influences, as shown in Figure 8.

- 2019 - Health Focus: Research focused on health-related topics and political communication, with keywords like “social networking sites” (4), “facebook” (3), “health promotion” (3), “political communication” (3), and “public health” (2). This suggests an interest in social media's impact on health and political discourse.
- 2020 - COVID-19 Pandemic Shift: The COVID-19 pandemic significantly influenced research, with “COVID-19” (5) emerging as a top keyword. Other keywords included “twitter” (4), “university students” (3), and “attitude” (3), indicating a focus on the pandemic's impact on different demographics and online behaviors.
- 2021 - Focus on COVID-19 and Its Impacts: The focus on COVID-19 (18) intensified. Keywords like “facebook” (7), “twitter” (5), and “sentiment analysis” (5) reflect ongoing studies on public sentiment and social media's role during the pandemic.
- 2022 - Emergence of New Concerns: While “COVID-19” (18) remained dominant, new concerns emerged. Keywords like “youth” (5), “vaccine hesitancy” (4), and “resilience” (4) indicate a growing interest in youth interactions with social media and public health communication.
- 2023 - Diversification and Social Media Platforms: Research on COVID-19 (15) continued but with a slight decline. Keywords like “instagram” (5), “social networking sites” (4), and “sentiment analysis” (4) suggest a broader examination across different platforms.
- 2024 - Transition to Broader Societal Issues: Research shifted towards broader societal issues, with keywords like “sustainable development” (2), “millennials” (2), and “hate speech” (1). This indicates a transition from pandemic-centered research to topics addressing long-term societal challenges.

Overall, the evolution of keywords over the past five years showcases a trajectory from health and political communication to pandemic-related issues, and eventually to broader social and sustainable development themes as the pandemic's influence decreases.

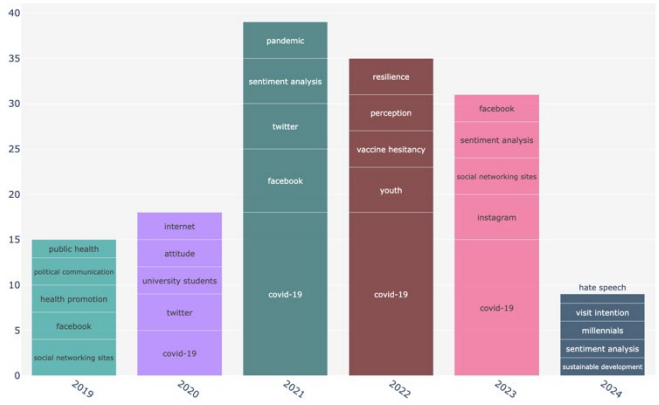

Figure 8. Top 5 Authors Keyword Evolution

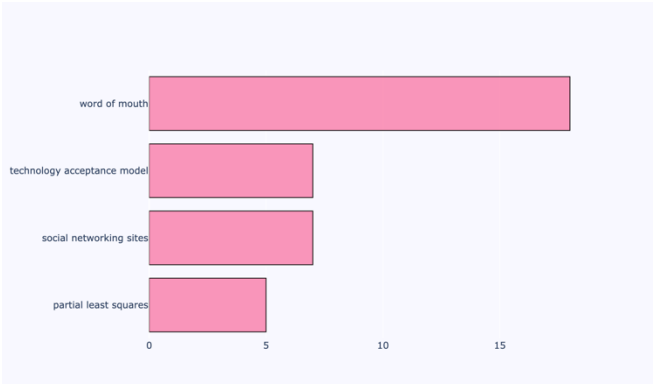

Figure 9. Top 4 Keyword Plus Tri-Grams

7.3. Keyword Plus Tri-Grams

Focusing on Keyword Plus from the WoS database and analyzing tri-grams provides valuable insights into prevalent themes and methodologies in social media research in Malaysia. Figure 9 illustrates the top four tri-grams, revealing key focal points in this research domain.

- "Word of mouth" appears 18 times, indicating its critical importance in Malaysian social media research. This term reflects a significant interest in how information and recommendations are shared on social media, influencing consumer behavior, trust, and brand loyalty.
- "Technology acceptance model", appearing 7 times, highlights its use as a theoretical framework to understand social media adoption and user engagement. The model helps researchers study factors driving the acceptance of social media technologies.
- "Social networking sites", also appearing 7 times, underscores the focus on platforms like Facebook, Twitter, and Instagram. This research explores user behavior, psychological impacts, privacy concerns, and the role of these platforms in shaping societal norms.
- "Partial least squares", appearing 5 times, indicates the use of this statistical technique for modeling complex relationships in social media research. PLS is valuable for testing theoretical models and analyzing constructs.

Overall, these top tri-grams reveal the thematic and methodological landscapes of social media research in Malaysia, emphasizing word of mouth dynamics, technology acceptance, the role of social networking sites, and the use of robust statistical techniques.

8. Topic Modelling

8.1. Data Preprocessing

In this study, a detailed four-step data preprocessing approach, as depicted in Figure 10 is used to improve text analysis. The first step is text cleaning, where all text is made lowercase to keep the dataset consistent and improve its accuracy without losing quality. During this step, non-letter characters like numbers and symbols, as well as very short words and common, less meaningful words known as stop words, are removed. This helps make the text clearer and more focused on important terms. Also, common terms that are not very important to the study are taken out manually [17]. Next, the process of creating embeddings transforms the cleaned text into numerical vectors using a model called SciBERT, which is specifically designed for scientific texts. The uncased version of SciBERT is chosen because it simplifies the model and increases its effectiveness by reducing the number of unique words it needs to handle [18]. Then, dimensionality reduction is done using two methods: KernelPCA and UMAP. KernelPCA first reduces the number of features in the dataset while keeping the complex patterns intact. UMAP further reduces these features while making sure that both the close and broader relationships in the data are maintained. This step helps manage the challenge of handling high-dimensional data, making the analysis more accurate and efficient [19]. Finally, the embeddings are clustered using a technique called HDBSCAN. This method groups similar data points together and identifies any outliers or unique points. This helps clearly define different topics in the text and makes the analysis more precise [20].

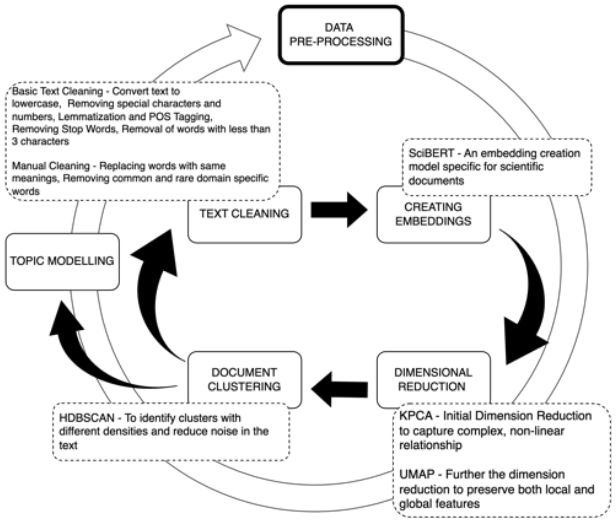

Figure 10. Data Pre-processing Lifecycle

8.2. Results

Topic Representation

The results from data preprocessing are then used to generate topic representations. In this study, the BERTopic model is implemented. BERTopic uses a class-based variation of the TF-IDF approach to identify the importance of a word to a document [21]. A total of 18 topics (Table 6) were generated, with each topic named based on the importance score of words within the corpus. The algorithm calculates this score by considering both the context and frequency of the words. High scores indicate words that are semantically significant, while lower scores are given to less contextually relevant words. The words selected for each topic name best describe the main themes or concepts associated with the corresponding corpus.

Table 6. Topics Generated from Topic Modelling

Topic	Count	Name
0	82	0_political_public_news_election
1	54	1_brand_marketing_consumer_relationship
2	38	2_usage_employee_academic_intention
3	32	3_student_learning_teaching_academic
4	30	4_library_academic_technology_mobile
5	30	5_tourism_chinese_destination_heritage
6	29	6_marketing_brand_customer_networking
7	27	7_sentiment_twitter_malay_machine
8	27	8_vaccine_covid_covid vaccine_hesitancy
9	26	9_intervention_exercise_awareness_healthy
10	25	10_waste_palm_food_sustainable
11	25	11_student_digital_professional_dental
12	23	12_adult_experience_pandemic_support
13	21	13_addiction_cyber_adult_youth
14	21	14_pharmacy_community pharmacy_healthy_product
15	21	15_muslim_healthy_food_intention
16	16	16_anxiety_depression_covid_symptom
17	12	17_privacy_scholarly_scholarly communication_asns

To aid in understanding the topic modeling results, each topic is interpreted by analyzing the cluster of documents it represents. For instance, Topic 0, labeled '0_political_public_new_election,' specifically pertains to political news and public elections, comprising 82 documents' abstract. This example demonstrates the methodology for interpreting a topic, where researchers can derive insights into the prevalence and focus of political communication within Malaysian social media. The complete list and detailed interpretation of all topics available here [16]

Topic Evaluation

Topic coherence is used to evaluate the quality of topics, focusing on how well the words within each topic represent a clear and unified concept. This coherence, indicative of semantic insight and interpretability, is exemplified by words such as "political_public_news_election" in Topic 0 [22]. Metrics such as *c_v*, *u_mass*, *c_uci*, and *c_npmi* are used to measure coherence, addressing word co-occurrence and semantic similarity [23]. A *c_v* score of 0.5042 indicates moderate to good coherence, suggesting well-formed topics. In contrast, a *c_npmi* score of -0.000 points to a neutral coherence level, neither strong nor random. The *u_mass* and *c_uci* scores further support a reasonable degree of coherence among the topics.

Table 7. Coherent Score for each Metrics

Metrics	c_v	u_mass	c_uci	c_npmi
Coherence Score	0.5042	-4.6462	-2.4770	-0.0000

9. Topic Interpretation

In this study, two methods were used to enhance the interpretability of the generated topics. First, the authors collaboratively analysed the themes and sub-themes to clearly define each topic. Then, GPT was used to summarize the abstracts for each topic, providing a brief overview and helping confirm the themes identified earlier. This approach combined detailed manual analysis with the efficiency of automated summarization to effectively interpret the topics.

The list of generated topics with GPT interpretation are available via the following link [16].

9.1. Themes

Thematic analysis [24] is an effective method for breaking down complex qualitative data into manageable steps. In this study, thematic analysis was applied to abstracts from each topic, treating them as qualitative data to systematically identify and interpret key themes within social media research in Malaysia. The authors reviewed all the abstracts within each topic to derive themes and sub-themes, as illustrated in Table 8.

Table 8. Themes for each Topics

Topic	Name	Themes
0	0_political_public_news_election	**Theme:** Social Media Influence on Political Campaigns **Sub-Themes:** - Political Marketing and Mobilization - Voter Information Dissemination - Impact on Election Outcomes
1	1_brand_marketing_consumer_relationship	**Theme:** Consumer Behavior and Purchase Intention **Sub-Themes:** - Social Media Influence - Customization and Personalization - Impact of Entertainment and Interaction
.		
.		
.		
10	10_waste_palm_food_sustainable	**Theme:** Food Waste Management **Sub-Themes:** - Patterns and causes of food waste in hospitality and food service - Sustainable household food waste management - Composting practices in rural communities
.		
.		
.		
17	17_privacy_scholarly_scholarly communication_asns	**Theme:** Scholarly Communication Behaviour **Sub-Themes:** - Diversity in Scholarly Communication Practices - Impact of Cultural Differences on Research Behavior - Use of Social Media for Scholarly Purposes

9.2. GPT Summarisation

The topic modelling analysis is enhanced by using a specific version of OpenAI's GPT-4 [25], known as GPT-4o, to generate summaries for each identified topic. GPT-4o is tasked with synthesizing the results into coherent summaries. The process involves familiarizing the model with the data, coding it, and guiding the model to compile the information into concise summaries. This approach, known as chain-of-thought [24], facilitates a more comprehensive and accurate output by enabling GPT to follow a structured reasoning process. Table 9 presents the summarization results.

Table 9. GPT Summarisation for each Topic

Topic	Name	Summarisation
0	0_political_public_news_election	The abstracts collectively shed light on various dimensions of social media usage and its multifaceted impact on Malaysian society and politics. They explore the role of social media in political campaigning, as exemplified by the opposition party's strategic use of social platforms to influence voter behavior. The studies delve into the sociocultural impacts of social media, such as the Islamisation phenomenon and the public discourse surrounding the hijab among Malaysian Muslim women. They also highlight the challenges and opportunities presented by social media in areas such as citizen journalism, political engagement, and the rise of new media influencers. Additionally, the research examines issues of digital safety, including cyberstalking and misinformation, and addresses the complex interplay between digital platforms, civic engagement, and environmental sustainability. Collectively, these abstracts underscore the profound ways social media is reshaping public spaces, cultural practices, and political dynamics in Malaysia.
1	1_brand_marketing_consumer_relationship	The collected abstracts primarily focus on the increasing role of social media marketing in various industries within Malaysia, examining factors influencing its adoption, effectiveness, and challenges. The studies cover a wide range of sectors such as small and medium enterprises (SMEs), property markets, higher education institutions, furniture, healthcare, and the food and beverage industry. Key themes include determinants of social media marketing adoption, its impact on customer purchase

		intentions and brand equity, the influence of social media influencers, and the role of digital marketing tools like eWOM, online communities, and interactive content. The abstracts also explore specific models and theories like the Technology Acceptance Model (TAM), Unified Theory of Acceptance and Use of Technology (UTAUT), and Stimulus-Organism-Response (S-O-R) model, highlighting the mediating roles of perceived usefulness, trust, and emotional engagement. Overall, the findings underscore the importance of social media in enhancing brand performance, customer engagement, and effective marketing strategies, particularly in a post-COVID-19 context.
. . .		
17	16_anxiety_depression_covid_symptom	The collected studies investigate various impacts of the COVID-19 pandemic in Malaysia, focusing on psychological well-being, social media influence, and behavioral adherence. Several studies highlight the role of risk perception and trust in government, showing that social media use significantly affects self-efficacy and preventive behavior adoption. Others examine mental health issues like anxiety, stress, and depression, particularly among specific groups such as university students and working adults. Factors such as perceived social support, counseling, and digital engagement strategies are crucial for mitigating negative emotional impacts. The effectiveness of specific therapies like CBT and ACT for depression in younger adults is explored, alongside the implications of mental health symptoms in broader Southeast Asian contexts. Overall, these studies underscore the complex interplay between digital information sources, mental health, and behavior during the pandemic.
18	17_privacy_scholarly_scholarly communication_asns	The selected abstracts cover various aspects of social media and online social networks, emphasizing their impact on public relations, privacy concerns, and academic usage across different countries, denominations, and user groups. Comparative studies reveal significant differences in social media practices and preferences between Malaysia and the US, as well as privacy concerns between Malaysia and the UK. The influence of social media on scholarly communication behaviors and privacy perceptions is explored in multiple abstracts, highlighting the universal popularity of platforms like Google Scholar and ResearchGate among early career researchers. In Malaysia, college students prioritize privacy when selecting social networking sites. While some studies illustrate the traditional methods of academic assessment, others emphasize the low usage rates and the perceived benefits and barriers of Academic Social Networking Sites among Malaysian researchers. Across the board, privacy concerns remain a central theme, with studies urging collective responsibility and better user education on privacy management across various social media platforms.

10. Making Sense the Analytics

The bibliometric analysis of social media research in Malaysia from 2011 to 2024 reveals several critical insights into the landscape of this research domain. The study highlights a substantial increase in scholarly output, particularly during and after the COVID-19 pandemic, which underscores the growing significance of social media as a research focus in Malaysia. Key thematic areas identified include social media's impact on marketing, education, health communication, and political discourse. Making sense of the analytics is crucial, and following the guidelines outlined in [26], the following study insights have been derived, along with proposed directions for future research.

10.1. Study Insights

The following are the insights derived from this study:

1. **Insight 1** ***(Objective 1)***

   The performance analysis revealed a significant growth in the number of publications related to social media research in Malaysia over the past decade. This trend underscores the increasing recognition of social media as a critical area of study, particularly highlighted by its instrumental role during the COVID-19 pandemic for crisis communication and public health initiatives.

2. **Insight 2** ***(Objective 2)***

   The author productivity analysis identified key researchers that form the foundation of social media research in Malaysia. The collaboration networks indicate an interconnectivity among top authors, facilitating interdisciplinary research and knowledge exchange. This collaborative ecosystem is pivotal for addressing complex social issues, as evidenced by joint studies in areas such as health communication and political engagement.

3. **Insight 3** ***(Objective 3)***

   The application of BERTopic modeling effectively categorized the research into distinct themes, including marketing, health communication,

education, political engagement, AI integration, misinformation, hate speech, and privacy concerns. This thematic segmentation not only highlights the diverse applications of social media but also illuminates existing research gaps, particularly in emerging areas like AI and privacy. The identification of these gaps provides a roadmap for future research endeavours to address pressing societal challenges and leverage social media's potential.

10.2. Future Direction

The authors also propose the following future directions for social media research in Malaysia:

1. **Exploration of AI and Social Media**
   The need for further research into AI in social media is supported by the Topic Modelling findings, where emerging topics like sentiment analysis and machine learning were identified. These topics are gaining prominence, indicating that deeper exploration into how advanced techniques from Natural Language Processing (NLP) can be used to analyse social media data. Studies can also investigate AI-mediated communication [27], as highlighted by the growing interest in AI-related topics.

2. **Negative Information**
   The Keyword Analysis and Topic Modelling identified misinformation, fake news and hate speech as critical issues in social media research. These forms of negative information can ruin public trust, polarize communities, and undermine democratic processes. Given the importance of these keywords, there is a pressing need to study how negative information spreads on social media, its impact on public opinion, and strategies to combat it. Additionally, the role of algorithms in either amplifying or mitigating negative information warrants further investigation.

3. **Privacy and Ethical Considerations**
   Findings from the Author Collaboration analysis and Topic Modelling showed that privacy concerns are a significant area of interest among researchers. This aligns with the growing emphasis on privacy issues in the Keyword Analysis. Future research should focus on understanding privacy concerns among Malaysian users, examining the ethical implications of data usage, and developing frameworks to enhance privacy protections.

4. **Interdisciplinary Research**
   The Performance Analysis, Author Collaboration and Topic Modelling findings revealed diverse research themes and significant collaboration across different domains, such as social media strategies and marketing, health communication, and political engagement. These findings underscore the need for a multidisciplinary research approach. Encouraging collaboration across fields such as communication, computer science, psychology, and sociology will be crucial in addressing the complex impact of social media.

5. **NLP for Non-English Social Media Content**
   There is a crucial need for more focused studies on NLP applications in non-English local languages and local dialects. These areas are currently under-researched, particularly in detecting and addressing negative social media content like hate speech, misinformation and cyberbullying in multilingual contexts. Enhancing NLP capabilities in these domains will improve content moderation and contribute to a safer online communication environment.

References

[1] A. Gutierrez *et al.*, "Retailers, don't ignore me on social media! The importance of consumer-brand interactions in raising purchase intention - Privacy the Achilles heel," *Journal of Retailing and Consumer Services*, vol. 72, p. 103272, May 2023, doi: 10.1016/j.jretconser.2023.103272.

[2] S. Kemp, "Digital 2024: Malaysia," Data Reportal. Accessed: Aug. 07, 2024. [Online]. Available: https://datareportal.com/reports/digital-2024-malaysia

[3] S. Kemp, "Digital 2024: Malaysia," Data Reportal. Accessed: Aug. 07, 2024. [Online]. Available: https://datareportal.com/reports/digital-2024-malaysia

[4] N. Mat Dawi, H. Namazi, H. J. Hwang, S. Ismail, P. Maresova, and O. Krejcar, "Attitude Toward Protective Behavior Engagement During COVID-19 Pandemic in Malaysia: The Role of E-government and Social Media," *Front Public Health*, vol. 9, Mar. 2021, doi: 10.3389/fpubh.2021.609716.

[5] A. Salman, M. A. Yusoff, M. A. Mohamad Salleh, and M. Y. Abdullah, "Political Engagement on Social Media as Antecedent for Political Support among Voters in Malaysia," *Jurnal Komunikasi, Malaysian Journal of Communication*, vol. 34, no. 2, pp. 152–165, Jun. 2018, doi: 10.17576/JKMJC-2018-3402-10.

[6] O. Y. Shien, N. S. Huei, and N. L. Yan, "The Impact of Social Media Marketing on Young Consumers' Purchase Intention in Malaysia: The Mediating Role of Consumer Engagement," *International Journal of Academic Research in Business and Social Sciences*, vol. 13, no.

1, Jan. 2023, doi: 10.6007/IJARBSS/v13-i1/15806.

[7] D. Camacho, M. V. Luzón, and E. Cambria, "New trends and applications in social media analytics," *Future Generation Computer Systems*, vol. 114, pp. 318–321, Jan. 2021, doi: 10.1016/j.future.2020.08.007.

[8] V. Nunavath and M. Goodwin, "The Role of Artificial Intelligence in Social Media Big data Analytics for Disaster Management -Initial Results of a Systematic Literature Review," in *2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, IEEE, Dec. 2018, pp. 1–4. doi: 10.1109/ICT-DM.2018.8636388.

[9] T. H. Davenport, "From analytics to artificial intelligence," *Journal of Business Analytics*, vol. 1, no. 2, pp. 73–80, Jul. 2018, doi: 10.1080/2573234X.2018.1543535.

[10] V. Balakrishnan, K. S. Ng, and H. A. Rahim, "To share or not to share – The underlying motives of sharing fake news amidst the COVID-19 pandemic in Malaysia," *Technol Soc*, vol. 66, p. 101676, Aug. 2021, doi: 10.1016/j.techsoc.2021.101676.

[11] N. F. Azman, N. A. K. Zamri, and M. N. Hassim, "Hate Speech on Social Media in Malaysia - A Thematic Review," *International Journal of Academic Research in Business and Social Sciences*, vol. 13, no. 6, Jun. 2023, doi: 10.6007/IJARBSS/v13-i6/17177.

[12] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J Bus Res*, vol. 133, pp. 285–296, Sep. 2021, doi: 10.1016/j.jbusres.2021.04.070.

[13] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.05794

[14] I. Vlase and T. Lähdesmäki, "A bibliometric analysis of cultural heritage research in the humanities: The Web of Science as a tool of knowledge management," *Humanit Soc Sci Commun*, vol. 10, no. 1, Dec. 2023, doi: 10.1057/s41599-023-01582-5.

[15] V. Pereira, M. P. Basilio, and C. H. T. Santos, "pyBibX -- A Python Library for Bibliometric and Scientometric Analysis Powered with Artificial Intelligence Tools," Apr. 2023.

[16] P. Isawasan and M. A. H. Ahmad Asmawi, "Dataset : Social Media Research Trends in Malaysia." Accessed: Aug. 13, 2024. [Online]. Available: https://www.kaggle.com/datasets/akma1xz/social-media-research-trends-in-malaysia/data

[17] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations," *Organ Res Methods*, vol. 25, no. 1, pp. 114–146, Jan. 2022, doi: 10.1177/1094428120971683.

[18] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A Pretrained Language Model for Scientific Text," Mar. 2019, [Online]. Available: http://arxiv.org/abs/1903.10676

[19] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, "Uniform Manifold Approximation and Projection (UMAP) and its Variants: Tutorial and Survey," Aug. 2021, [Online]. Available: http://arxiv.org/abs/2109.02508

[20] L. Scoccola and A. Rolle, "Persistable: persistent and stable clustering," *J Open Source Softw*, vol. 8, no. 83, p. 5022, Mar. 2023, doi: 10.21105/joss.05022.

[21] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.05794

[22] F. Bianchi, S. Terragni, and D. Hovy, "Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.03974

[23] J. M. Campagnolo, D. Duarte, and G. Dal Bianco, "Topic Coherence Metrics: How Sensitive Are They?," *Journal of Information and Data Management*, 2022.

[24] A. Turobov, D. Coyle, and V. Harding, "Using ChatGPT for Thematic Analysis," May 2024, [Online]. Available: http://arxiv.org/abs/2405.08828

[25] OpenAI *et al.*, "GPT-4 Technical Report," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.08774

[26] W. M. Lim and S. Kumar, "Guidelines for interpreting the results of bibliometric analysis: A sensemaking approach," *Global Business and Organizational Excellence*, vol. 43, no. 2, pp. 17–26, Jan. 2024, doi: 10.1002/joe.22229.

[27] J. T. Hancock, M. Naaman, and K. Levy, "AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations," *Journal of Computer-Mediated Communication*, vol. 25, no. 1, pp. 89–100, Mar. 2020, doi: 10.1093/jcmc/zmz022.