Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4, Issue 2 (2024)

Machine Learning Applications in Social Networking: Bibliometric Analysis and Future Avenues for Research

Priya Sangwan[1], Tanu Khasa[2], Poonam[3], Dr. Sushil Sharma[4]

[1,2,3,4]Kurukshetra University, Kurukshetra, India.

Abstract

This research paper provides a comprehensive descriptive analysis of the application of machine learning techniques within the realm of social networks, focusing on scholarly contributions and emerging trends over a 21-year period. Through an examination of prominent nations, authors, journals, and thematic clusters, key findings underscore the central role of China and the USA in machine learning research applied to social networks. Furthermore, the study identifies IEEE Access and Journal of Machine Learning Research as leading journals in the field, while highlighting Lise Getoor as a preeminent author within the domain. The emergent themes of sentiment analysis, feature extraction, natural language processing, deep learning, and social media analysis underscore the evolving landscape of machine learning in social network research. Finally, a conceptual framework derived from thematic mapping offers insights into future research directions, highlighting applications of machine learning techniques such as neural networks, deep learning, and support vector machines in predicting online customer engagement, detecting fake accounts, and addressing other pertinent issues in the digital marketing landscape. This research encapsulates the key findings and implications of the research, serving as a guide for scholars and practitioners in the field.

Keywords: Artificial intelligence, Bibliometric analysis, Machine Learning, Sentiment Analysis, Social Network.

Introduction

The fundamental concept of machine learning is that the actions and decisions made by a machine, or computer should not be solely determined by a programmer. Rather, the computer should acquire knowledge from current observed data using algorithms and utilize that knowledge to process (e.g., categorize) previously unfamiliar data. This mirrors how individuals can navigate novel and unfamiliar circumstances by drawing upon past encounters. Mahesh (2020) defined "Machine learning as the field of study that gives computers the ability to learn without being explicitly programmed". Frank Rosenblatt from Cornell University is frequently credited with the inception of contemporary machine learning. In the late 1950s, he and his group developed the perceptron classifier, a machine-learning algorithm used to identify and classify letters of the alphabet Rosenblatt, (1958, 1960). Machine learning is used to teach machines how to handle the data more efficiently. Researchers developed techniques and methodologies to analyze extensive observed data by training machine learning models, which acquired patterns from the data Jordan & Mitchell (2015).

Currently, the notion of machine learning has become essential in social networking and other areas (Ballestar et al., 2019; Ben Jabeur et al., 2023; Nilashi et al., 2019). Social networking extensively employs the notion to evaluate user sentiments, detect hate speech, enhance lead time, and provide friend recommendations. (Cheng et al., 2019; Dang et al., 2020; Kilroy et al., 2022).

There is a dearth of bibliometric studies that have been undertaken on the topic of machine learning.. Ben Jabeur et al.(2023) "investigated the use of AI and machine learning in fake review detection. Kenger & Özceylan (2023) evaluated the development of 'Fuzzy min-max neural network' from 1992 to 2022. Bonkra et al. (2024) applied bibliometric technique based on a dataset of 109 publications sourced from the Scopus database, spanning the years 2011 to 2022 to analyze the impact of unexpected weather on agricultural output and evaluate the effectiveness of AI-based machine learning and deep learning algorithms to accurately detect and classify diseases affecting apple leaves". Upon reviewing the available literature, we have discovered that there are only a few bibliometric studies that focus on the notion of machine learning. However, there is a dearth of studies that specifically explore the application of ML in social networking to influence online purchase intention. The objective of the study is to address this research gap by providing answers to the following inquiries:

1. What is the current trend for utilizing machine learning in social networking?
2. What is the current state of thematic progression and global research collaboration for this topic?
3. What are the emergent subjects that can be incorporated under this overarching theme?

This study aims to evaluate the current state of machine learning within the realm of social networking. This research contributes to existing knowledge by utilizing conceptual evaluation, thematic evolution, and bibliometric analysis to explore the application of ML in social networks. Spanning from 2003 to 2024, the analysis comprehensively reviews the literature on this topic. The findings are expected to assist future researchers in identifying key authors, countries, and publication sources relevant to this area of study. Additionally, it will provide insights for businesses in leveraging machine learning to address issues on social media platforms. The subsequent section provides a summary of the current literature, followed by a discussion of the methodology. Furthermore, the findings and interpretations are shown in the fourth section, while the fifth section covers the study's outcomes, conceptual framework, contributions, limitations, and prospects. Lastly, the sixth section outlines the final findings and conclusions derived from this research endeavor.

Related Literature

The literature review encompasses a wide array of research contributions in the field of using machine learning in social networking. Yang et al.(2011) "propose a dynamic stochastic block model for identifying communities in evolving social networks. Zheng et al.(2015) introduce a supervised ML-based solution for spammer detection on social networks, utilizing feature extraction and an SVM-based algorithm". Tabassum et al. (2018) emphasize the significance of social network analysis in understanding relationships, behaviors, and network impacts.

Chen et al. (2010) introduce a game-theoretic framework for community detection in social networks. Sadeh et al.(2009) explore privacy attitudes and technologies like PEOPLEFINDER for location sharing. Kosinski et al. (2014) “establish links between user personalities, website preferences, and Facebook profile features. Buettner (2017) proposes a personality-based product recommender framework utilizing social media data. Rossetti et al. (2017) propose TILES for community discovery in dynamic social networks. Al-Zoubi et al.(2018) propose a hybrid ML model for spam detection. Meo et al.(2015) introduce a quantitative measure considering similarity and trustworthiness among social media users”. Guimarães et al.(2017) analyze age group behavior using machine learning algorithms. Liu et al.(2006) explore taste capture and analysis in online social networks. Wu et al. (2020) propose deep learning techniques for community detection. Varshney et al. (2017) predict message diffusion probabilities through Bayesian networks. Javari & Jalili (2014) propose a method for sign prediction in networks with positive and negative links. Ballestar et al. (2019) present a predictive model for personalized financial incentives. Aljarah et al.(2020) focus on hate speech detection using NLP and ML on Twitter. Z. Li et al. (2017) recommend links based on cost, value, and linkage likelihood. Bostani & Sheikhan (2017) propose a modified algorithm for intrusion detection. Y. Li et al. (2017) developed a machine learning-based user identification solution. Su et al. (2019) introduce a method for detecting locker ransomware. Cheng et al.(2019) propose a scalable friend recommendation framework. X. Li et al.(2014) explores social network-based recommender systems for e-commerce. Verbeke et al. (2014) use social network information for customer churn prediction. Sahoo & Gupta, (2019) classify recent OSN attacks and analyze defenses. . Fang & Hu (2018) predict top persuaders in social networks. Fu et al.(2018) improve spam detection by leveraging users' temporal evolution patterns. Sarna & Bhatia (2017) categorize and address bullying messages. Wei-dong et al.(2018) propose a sentiment analysis method on micro-blog platforms. Nilashi et al.(2019) developed a recommender system for hotel recommendations. Wanda & Jie, (2020) propose DeepProfile for addressing fake accounts. Shuai et al. (2018) introduce a machine-learning algorithm for early detection of social network mental disorders. Ding et al.(2016) predict unknown users' attributes using social network attributes. Barushka & Hajek (2020) used deep neural networks for spam detection. These studies collectively contribute to advancing the understanding and application of machine learning in social networking across various domains.

Bibliometric Analysis

The current study employed a distinctive approach across different phases, including planning and data extraction as well as analysis and dissemination of research results.

Data extraction

In the data collection phase, researchers opted to gather information from the (WoS) Web of Science database, aiming to acquire research publications spanning 21 years. On April 3, 2024, a document-based search was conducted within the WoS database using the search term 'Machine Learning' AND ('Social Network*' OR 'Online Purchase Intention'). This query yielded a total of 1,954 publications since 2003, with no filters applied initially. Subsequently, a refined search query, 'Machine Learning AND ('Social Network*' OR 'Online Purchase Intention') (Topic) and English (Language) and or Computer Science Information System, Business or Management or Computer Science Artificial Intelligence (WoS Categories)', was executed, resulting in the identification of 1,102 records. Following the exclusion of paper letters, early access papers, proceedings, and other non-full text publications. A final dataset of 1,098 articles was compiled. To understand the application of machine learning in business, management, and advanced computing, a subset of 1,058 research publications from the period 2003 to 2024 was subjected to bibliometric analysis and network visualization.

Framework for Analysing the Data

The analysis involved a detailed examination of key elements, including author source analysis, nation analysis, author analysis, and theme analysis. A preliminary investigation was conducted on a collection of 1,098 scholarly documents spanning the years 2003 to 2024. Specifically, the analysis focused on identifying the top countries, authors, sources, and themes within the dataset. The study utilized Microsoft Excel and the 'Biblioshiny', developed by Aria & Cuccurullo (2017), which operates within the 'R' programming language framework. Additionally, the network clustering and visualization process was facilitated by the bibliometric program VOSviewer 1.6.18, developed by van Eck and Waltman in 2010. These tools enabled comprehensive analysis and visualization of the research data, providing insights into the trends and patterns within the literature on machine learning in social networks.

Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4, Issue 2 (2024)

Figure 1 depicts the study framework, which consists of several extraction and analysis phases specifically designed for bibliometric visualization.

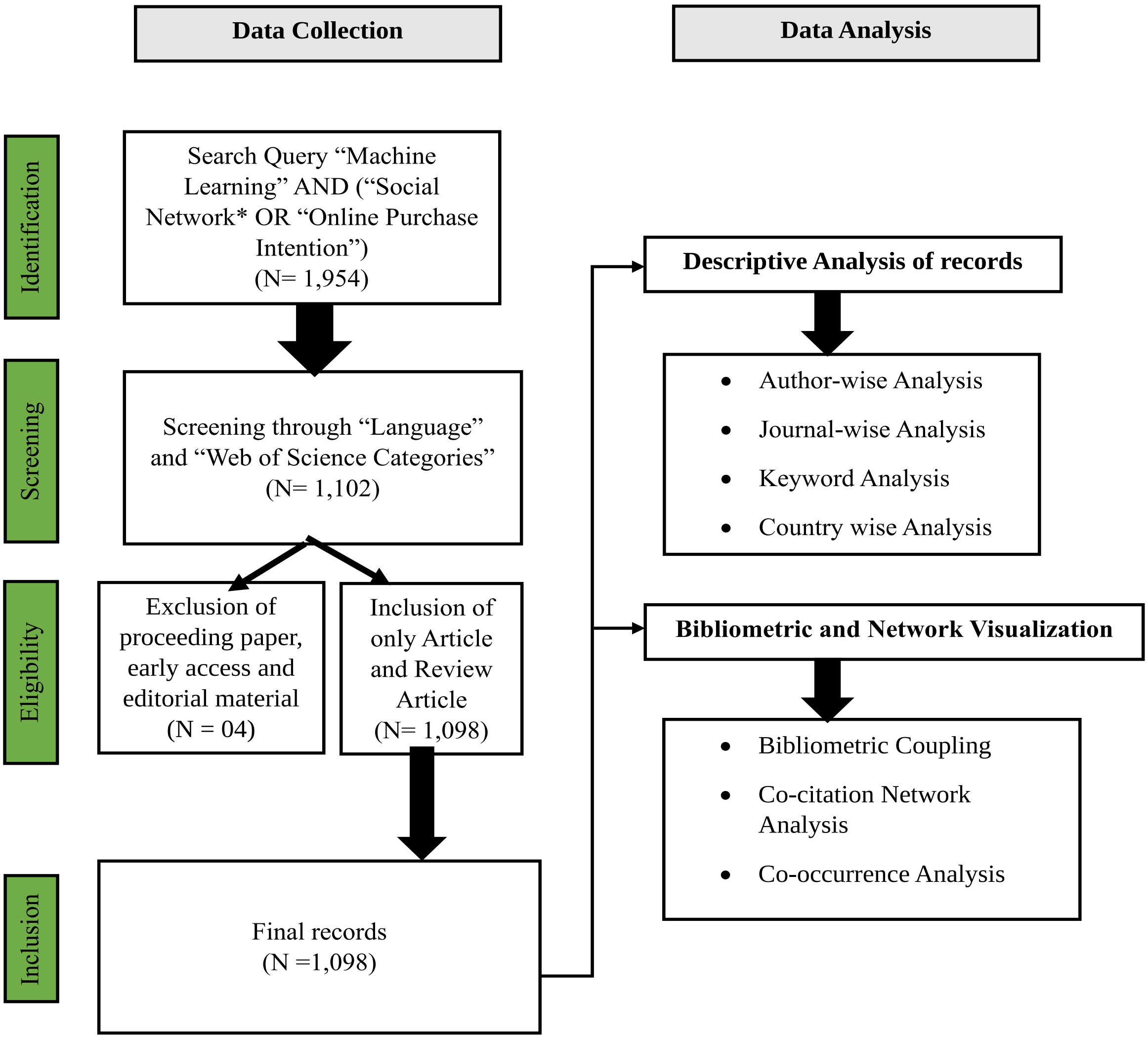

Figure 1 Research Structure

Results and Discussion

Results

The data presented in this section encompasses an extensive examination of scholarly documents about the intersection of machine learning and social networking, with a focus on their influence on online purchase intention.

Detailed Analysis of Countries, Journals, Authors and Keywords

Across the period of 2003 to 2024, a total of 1098 documents were analyzed, drawn from various sources including journals, books, and other scholarly outlets, amounting to 207 distinct sources in total. Notably, the average citation per document stood at 22.3, indicating a substantial level of academic engagement with the subject matter.

The collaborative nature of research within this field is evident, with a total of 3,380 authors contributing to the corpus of literature. Among these, 24 authors produced single-authored documents, while the majority engaged in collaborative efforts, resulting in a high average of 3.89 co-authors per document. This collaborative ethos is further underscored by the average of 3.078 authors per document.

Moreover, a diverse range of keywords was identified within the documents, with 1,073 instances of Keywords Plus (ID) and 3,572 instances of Author's Keywords (DE), reflecting the multifaceted nature of research inquiries and topics explored within this domain.

Overall, these findings provide valuable insights into the scholarly landscape surrounding machine learning implementation in social networking contexts and its implications for online purchase intention, underscoring the collaborative and interdisciplinary nature of academic inquiry in this area.

Table 1 Descriptive data related to the Research Theme

Description	Results
Total Documents	1098
Period	2003:2024
Sources (Journals, Books and Others)	207
Authors	3,380
Document per author	0.32485207
Authors per document	3.07832423
Co-authors per documents	3.89
Average citation per document	22.3
Authors of single-authored documents	24
Authors of multi-authored documents	1,074
Keywords plus (ID)	1,073
Author’s Keywords (DE)	3,572

Figure 2 depicts the distribution of articles related to the specified subject matter across the years 2003 to 2024. Notably, there is a discernible trend of increasing scholarly output over time, with a few fluctuations evident. From 2003 to 2008, there is a relatively low number of articles, with only sporadic contributions in some years. It is due to technological advancements and breakthroughs in related fields, such as artificial intelligence, big data, and machine learning, that have not yet reached a level of maturity or prominence to spur significant research activity. However, starting from 2009, there is a noticeable uptick in output, with the number of articles gradually increasing. A significant surge in publication activity is observed from 2016 onwards, with the number of articles more than doubling compared to previous years. This trend continues into the subsequent years, peaking in 2023 with 223 articles. Overall, the data reflects a growing interest and engagement within the academic community regarding the subject matter, with a particularly pronounced increase in research output in recent years.

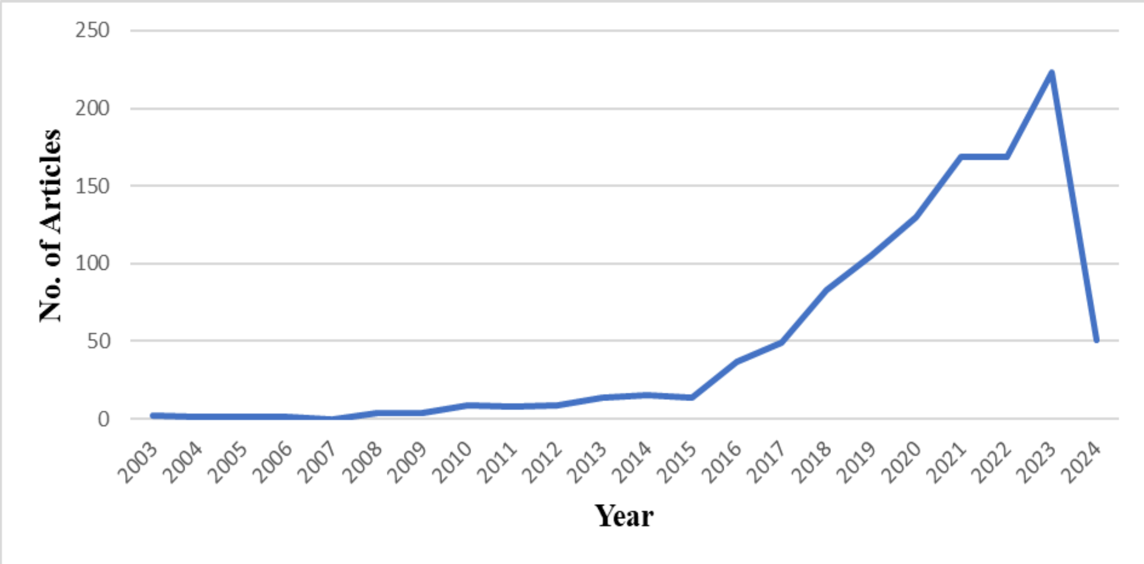

Figure 2 Year-wise Publication trend.

Country-wise Annotation

The data presented in Table 2 offers a nuanced portrayal of the scholarly landscape across the top 10 countries, delineating their respective contributions based on citation per paper and total citation metrics. Citation per paper metrics help to assess the significance and reach of research output in a particular field (Zhang et al., 2021). The United States emerges as a formidable leader, commanding a substantial share of both total citations (6220) and papers (457), alongside an impressive citation per paper (13.6105). This underscores the profound influence of American research endeavors on global academic discourse. China, on the other hand, exhibits remarkable scholarly output in terms of total papers (966), yet its citation impact paper (3.505176) is comparatively moderate, indicative of a burgeoning research landscape with evolving impact potential. India, while displaying commendable performance in total citations (1894) and papers (420), maintains a moderate citation per paper (4.509524), reflecting a balance between quantity and impact in its scholarly contributions. Among other nations such as the UK, Germany, Korea, Australia, Spain, Italy, and Malaysia, a diverse array of scholarly productivity and impact profiles is discernible, contributing to the rich tapestry of global academic discourse. This analysis illuminates the multifaceted dynamics underlying international research contributions and underscores the imperative of considering both quantitative and qualitative metrics in evaluating scholarly impact.

Table 2 Ranking of the top countries based on their per paper and total citations.

Country	TC	% of TC	Citation Per Paper
USA	6220	33.11	13.6105
China	3386	18.02	3.505176
India	1894	10.08	4.509524
UK	1687	8.98	14.66957
Germany	1209	6.43	16.79167
Korea	1162	6.18	7.958904
Australia	1052	5.60	8.840336
Spain	971	5.17	5.137566
Italy	647	3.44	5.436975
Malaysia	560	2.98	6.511628
Total	18788	100	6.986984

Figure 3 presents a comprehensive overview of the total publication output across several countries. Notably, China emerges as the predominant contributor, boasting the highest number of publications. This finding underscores China's increasingly prominent role in the global research landscape and its significant contributions to the field of using machine learning in social networking to influence online purchase intention. Following China, the United States and India exhibit considerable publication numbers, indicative of their substantial research activity and scholarly productivity. Overall, the data underscores the heterogeneous distribution of research output across different countries, offering valuable insights into the global scholarly landscape and the varying levels of research activity among nations.

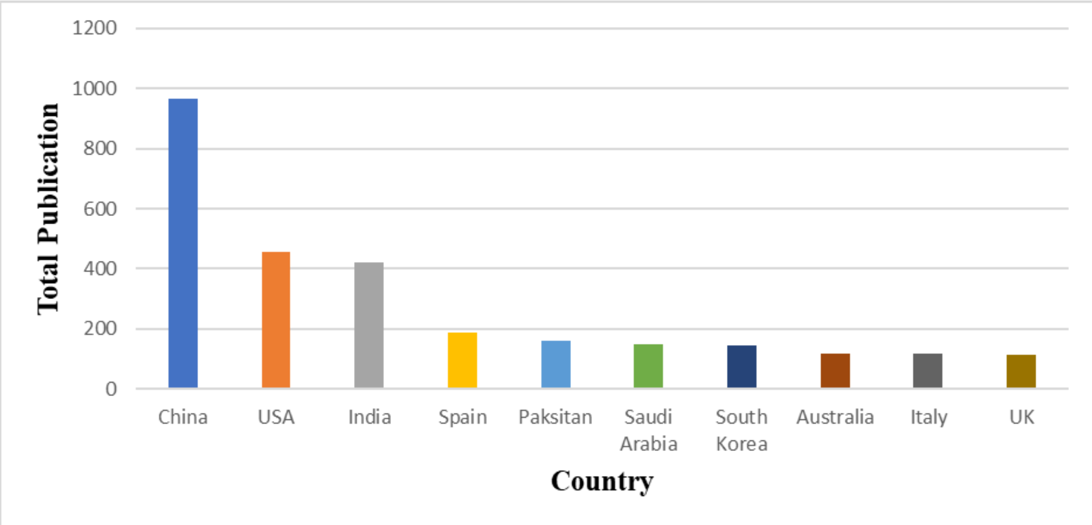

Figure 3 Top Countries based on Total Production.

Figure 4 illustrates the single-country publication (SCP) and multiple-country publications (MCP) for the top 10 countries. Single-country publications refer to publications where all authors are from the same country, indicating intra-country collaboration. On the other hand, multiple country publications involve authors from different countries, representing inter-country collaboration, also known as international collaboration (Sweileh et al., 2016). Notably, China emerges as a frontrunner in both SCP and MCP, showcasing extensive international collaboration and active participation in intra-country research projects. Following closely, India demonstrates a substantial SCP count, albeit with a relatively lower MCP count. This analysis underscores the diverse landscape of global scholarly cooperation and knowledge dissemination, shaped by the varying degrees of international collaboration across different nations.

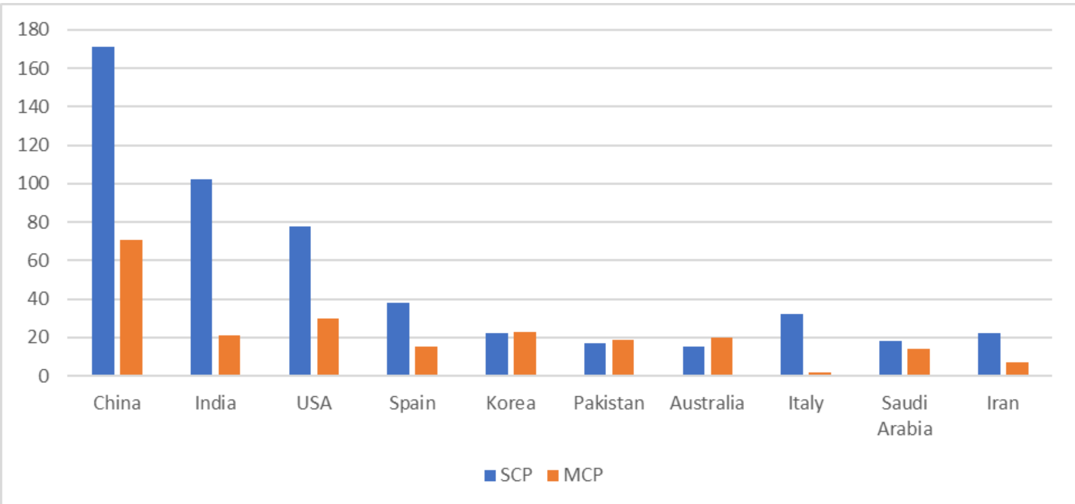

Figure 4 Ranking of the Top 10 Countries Based on SCP and MCP

Source-wise Elucidation

Table 3 provides an insightful glimpse into the scholarly landscape surrounding machine learning applications in social networking with a focus on influencing online purchase intention. Among all the journals, *IEEE Access* stands out with a substantial total citation count of 3048 and a commendable number of total publications, suggesting its significant influence and prolific publication output in this niche domain. Similarly, the *Journal of Machine Learning* research demonstrates notable citation impact with a total citation count of 3,222 indicating its substantial contribution to the intersection of machine learning and social networking. These findings underscore the growing importance of machine learning techniques in leveraging social networking data to understand and influence online purchase intention. Further research and exploration in this area, as evidenced by the scholarly output in these journals, hold promise for advancing our understanding of consumer behavior in the digital realm and in forming effective strategies for online marketing and commerce.

Table 3 Top Journal based on Total Citation and Total Production

Sources	TC	TP
IEEE Access	3048	217
Journal of Machine Learning Research	3222	33
Expert Systems with Applications	1068	36
IEEE Transactions on Knowledge and Data Engineering	508	29
Information Sciences	415	19
Knowledge-Based Systems	785	15
Multimedia Tools and Applications	276	38
IEEE Transactions on Computational Social Systems	382	41
Machine Learning	613	24
Applied Soft Computing”	312	11

Author-wise Elucidation

For Table 4, A thorough investigation was conducted to get an understanding of the researcher’s collaborations, their relevance in terms of citation, and publications in different disciplines. Mustafa Bilgic, Tina Eliassi-Rad, Brian Galligher, and Galileo Namata, all from the USA, are recognized as the most influential authors. They have achieved the highest citation-to-publication ratio (C/P) of 1633. Each of these authors has only published a single paper. Lise Getoor is the author with the highest publications (3 publications) among all the most influential authors, based on TCs and publications. She is exceptionally proficient in the field of ML. Authors that have a high average citation rate (C/P) do not necessarily have many publications. This suggests that the topic of utilizing machine learning in social networking to affect purchase intention is now popular. Emerging authors are extensively consulting publications in these fields to guide their future research.

Table 4 Authors ranked according to their TC and Overall Contribution

Authors	Countries	TC	TP	C/P	m index	g index	h index
Lise Getoor	USA	1713	3	571	0.091	3	2
Prithviraj Sen	USA	1635	2	817.5	0.118	2	2
Mustafa Bilgic	USA	1633	1	1633	0.059	1	1
Tina Eliassi-Rad	USA	1633	1	1633	0.059	1	1
Brian Galligher	USA	1633	1	1633	0.059	1	1
Galileo Namata	USA	1633	1	1633	0.059	1	1
Georgios Paltoglou	UK	1135	2	567.5	0.133	2	2
Mike Thelwall	UK	1135	2	567.5	0.133	2	2
Jure Leskovec	USA	1119	2	559.5	0.133	2	2
Di Cai	UK	1096	2	548	0.1	2	2

Theme Analysis

To grasp the evolving and prominent research themes in the area of machine learning in social networking, a keyword-based analysis was conducted. Figure 5 shows a word cloud showcasing the top 30 terms extracted from the research corpus. Amid these themes, sentiment analysis, deep learning, and NLP emerge as the most prevalent topics in the intersection of machine learning and social networking (Birjali et al., 2021; Dang et al., 2020; Meel & Vishwakarma, 2020). Additionally, feature extraction, social network analysis, Twitter, task analysis, and classification stand out as significant keywords applicable in the realms of business management and computer science research.

Figure 5 Word Cloud

Themes via Structural Maps and Dendrogram Plots

The research domain underwent bibliographic clustering, which considered mutual collaboration, prospective subjects, and the extent of research. Conceptual structure analysis was carried out utilizing a dendrogram plot and multiple correspondence analysis (Figures 6 and 7, respectively). The conceptual structure graph, a two-dimensional diagram, illustrates current research areas on the y-axis and potential future research areas on the x-axis. Based on these characteristics, the corpus of 1,098 research materials was divided into four broad clusters corresponding to the overarching themes identified. The density of clusters reflects the concentration of study themes, with densely populated clusters indicating extensively studied areas with potential for future development, while sparse clusters suggest emerging topics. The dendrogram plot presents a hierarchical tree of topics, depicting relationships and similarities between theme clusters using U-shaped figures to indicate closure. The vertical distance of U-shaped lines between neighboring themes indicates correlation and separation. Initially, the study themes were categorized into four distinct groups, which are outlined below:

Cluster 1: Machine Learning (M.L.) and Artificial Intelligence (A.I.)

The research studies that discuss ML in the field of social networking are concentrated in the most densely populated clusters, which are represented by the color red in Figure 6 and the color violet in Figure 7. Based on the graph's dimensions, this theme group exhibits a high level of saturation, indicating that a large number of research articles on this theme have been published in prestigious journals. Furthermore, the majority of the interconnected topics possess the potential for future emergence and expansion. This cluster encompasses approximately 538 research publications that are related to these specific issues. (Birjali et al., 2021) the paper provides a detailed study of sentiment analysis. It talks about the challenges of sentiment analysis and its future directions for the days to come. (Xu et al., 2020) reviews and analyses the multi-output learning paradigm, focusing on the four Vs: volume, velocity, variety, and veracity. The algorithm presented by (Kilroy et al., 2022) aims to extract keyphrases from multiple documents, predicting future customer needs based on user's social media posts on Reddit. This innovative approach highlights the potential of utilizing social media data for the proactive understanding of customer preferences and emerging trends. In terms of research cluster points, neural networks, feature extraction, algorithms, and sentiment analysis emerge as prominent areas poised to continue evolving across various application domains. These clusters represent critical aspects of ML and NLP, indicating ongoing interest and innovation in leveraging advanced techniques for tasks such as predictive analytics, information extraction, and sentiment understanding. As these areas continue to advance, they hold promise for addressing diverse challenges and opportunities in fields ranging from product development, marketing, customer service, and beyond.

Cluster 2: Sentiment Analysis in Social Networking

Figures 6 and 7 display research on sentiment analysis in social networking, highlighted in blue. These study issues are considered under-researched since there are not many research publications on them and the identification of closely related themes. Themes such as social media networking and sentiment analysis are currently lacking in the study, indicating that these topics are relatively new in the field of research. Future studies will focus on exploring these issues in more depth. Publications falling in this cluster applied machine learning in several ways, for instance, (Lifang et al., 2022) applied multiple supervised ML and topic modeling methods to characterize situational information types for crisis management. (Antonakaki et al., 2021) summarized research topics on Twitter, focusing on sentiment analysis, social graph structure, threats like bots, spam, fake news, and hate speech.

Cluster 3 Network Dynamics

The third largest cluster themes are indicated in violet in Figure 6 and in red in Figure 7. The graph dimensions indicate that the cluster is situated in an area with a high level of research activity and potential for future growth. This indicates that academics widely choose research subjects in this category for their studies. Moreover, these possess immense potential for future study by incorporating embedded topics with fundamental areas. (Bhat & Abulaish, 2013) presented a framework to identify spammers in Online Social Networks.

Cluster 4: Digital Ecosystem

Publications and research that focus on the interconnected digital landscape and the dynamic processes within it are found to be obsolete due to the negligible scope in the present times (see green cluster in Figure 6 and Figure 7). (Borkar & Reiffers-Masson, 2022) examines a DeGroot model of opinion transmission through random interactions in their study. (Kumar et al., 2022) The research aims to improve link prediction accuracy in dynamic networks by proposing a new feature called the Path Weight Aggregation Feature.

Cluster 1 is characterized by its high density and includes well-researched and chosen areas that are expected to undergo further development in the future. Moreover, Clusters 2 and 3 represent the most extensive regions characterized by under-studied and extensively investigated themes, respectively. However, Cluster 4 is a sparsely populated region of research topics that lack current trends and have no potential for future development.

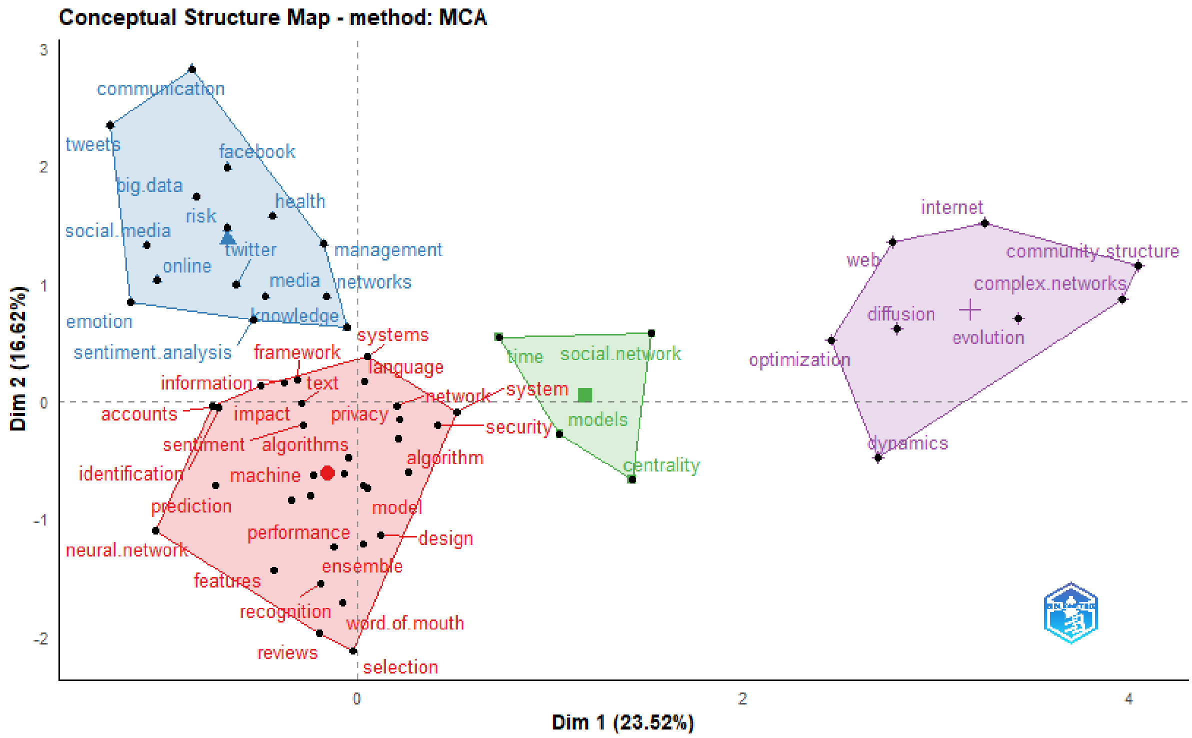

Figure 6 Conceptual Structure Map organized into Four Distinct Clusters

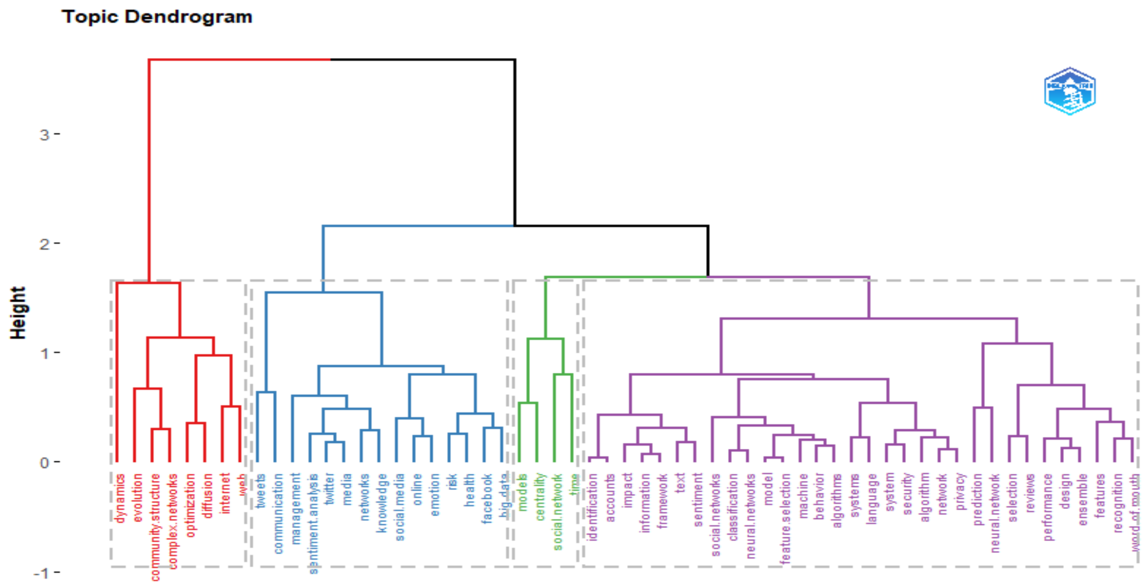

Figure 7 Dendrogram Graph organized into Four Distinct Clusters.

Bibliographic Coupling

To analyze the interconnection and cooperation between different entities, a bibliometric coupling analysis was performed. This analysis reveals the shared citations utilized by authors or countries. Figure 8 illustrates the interconnection between prominent authors, indicating that Pan Hui, Yang Chen, Qingyuan Gong, Xin Wang, and Yu Xiao have strong connections as a result of mutual citations. Pan Hui, Yang Chen, and Qingyuan Gong are the most prominent authors with significant bibliometric connections. Xin Wang and Yu Xiao belong to the same cluster.

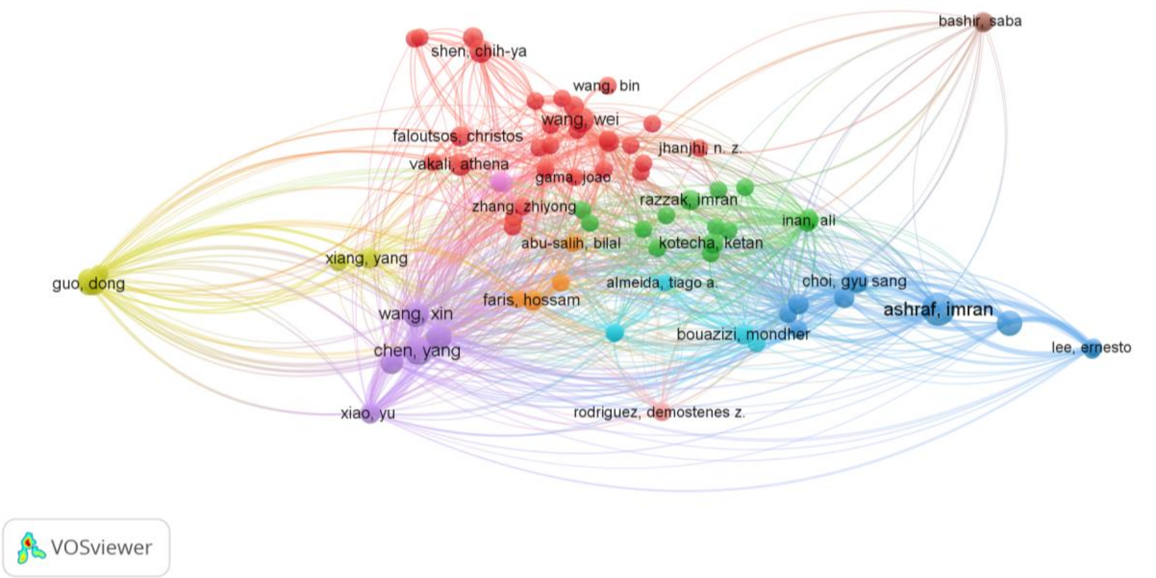

Figure 8 Authors Bibliometric Coupling.

Table 5 List of Top Authors based on Bibliometric Coupling

Authors	Total Link Strength
Hui, Pan	2,697
Chen, Yang	2,533
Gong, Qingyuan	2,195
Wang, Xin	2,171
Xiao, Yu	1,820
Ashraf, Imran	1,742
Mehmood, Arif	1,112
Rustam, Furqan	1,087
Umer, Muhammad	1,031
Choi, Gyu Sang	882

China has the highest level of bibliographic coupling with other nations, as indicated by its link strength. The United States and India follow closely behind in terms of this measure (Table 6).

Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4, Issue 2 (2024)

It indicates that articles published in other nations receive a high number of citations compared to publications from those countries. China, the USA, India, Saudi Arabia, and Pakistan exhibit significant collaboration in terms of publications on the topic of ML in social networking (Figure 9).

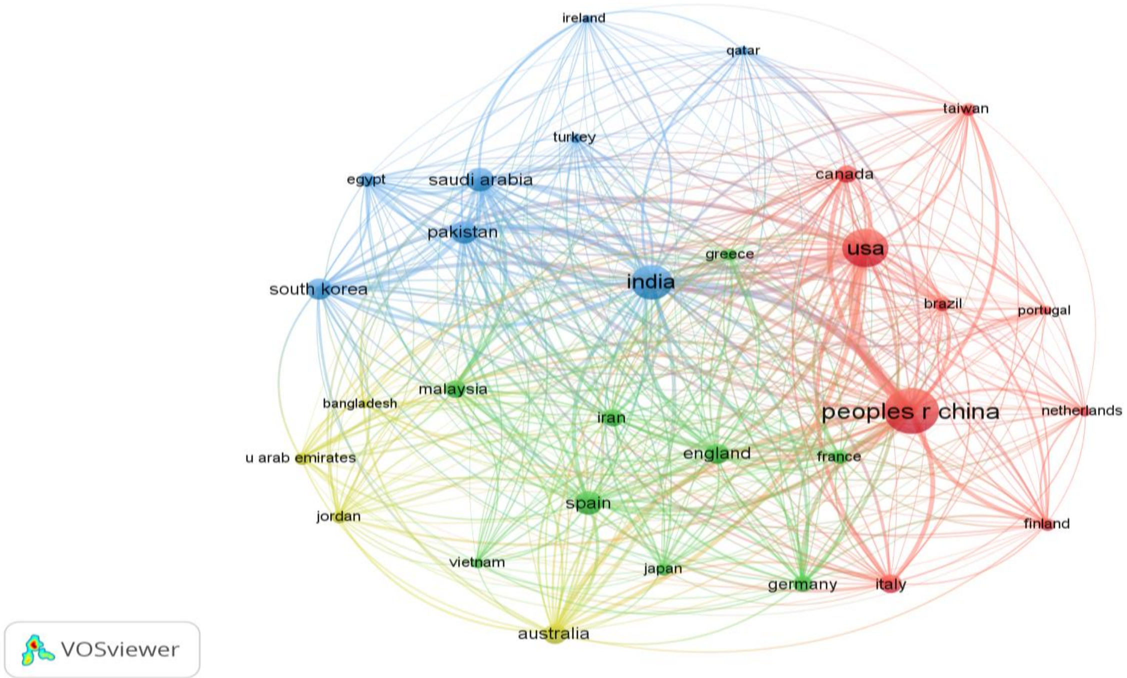

Figure 9 Bibliometric Coupling based on Countries.

Table 6 List of Top Countries based on Bibliometric Coupling

Countries	Total Link Strength
China	41,432
USA	31,796
India	27,937
Saudi Arabia	16,608
Pakistan	15,505
South Korea	13,673
Spain	13,572
Australia	12,578
England	10,785
Canada	9,699

According to the descriptive study, a total of 1,074 authors have co-authored the article with other authors. Imran Ashraf has the most extensive co-authorship network, collaborating with 23 authors from various nations. Figure 10 illustrates the collaborative research efforts of Imran Ashraf with Furqan Rustam, Patrick Bernard Washington, Ernesto Lee, Fatima El Barakaz, Vaibhav Rupapara, and Wajdi Aljedaan. Their study focuses on several aspects of machine learning, employing multiple sophisticated algorithmic models. These researchers are mostly associated with research on sentiment analysis and ML, employing advanced techniques and applying them to various domains such as marketing analysis, online shopping, and fraudulent marketing (Aslam et al., 2022; Birjali et al., 2021; Lee et al., 2022; Rupapara et al., 2021; Umer et al., 2021).

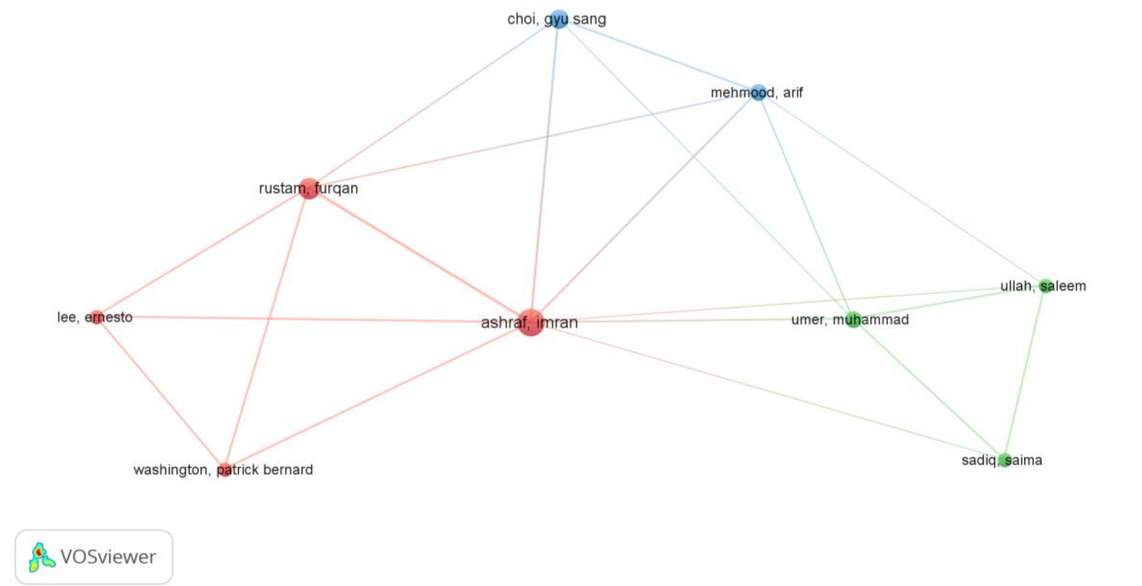

Figure 10 Author's Co-authorship analysis.

Co-citation analysis of research publications helps to understand the interconnectedness between authors, and research documents. To assess the strength of citations, a co-citation network analysis was conducted on the most influential research documents, using references as the basis. Please refer to Figure 11 for more details. Among top references, Grover & Leskovec (2016) have been cited in the majority of the documents. The research proposes an algorithm called node2vec for learning feature representation. The algorithm aims to capture the diversity of connectivity patterns observed in networks by mapping nodes to a low-dimensional space of features. This work offers a more efficient way to learn high-quality representations in complex networks. Grover & Leskovec, (2016) related to node2vec have been cited by 81 times in other documents. Airoldi et al. (2008) have the least citation links.

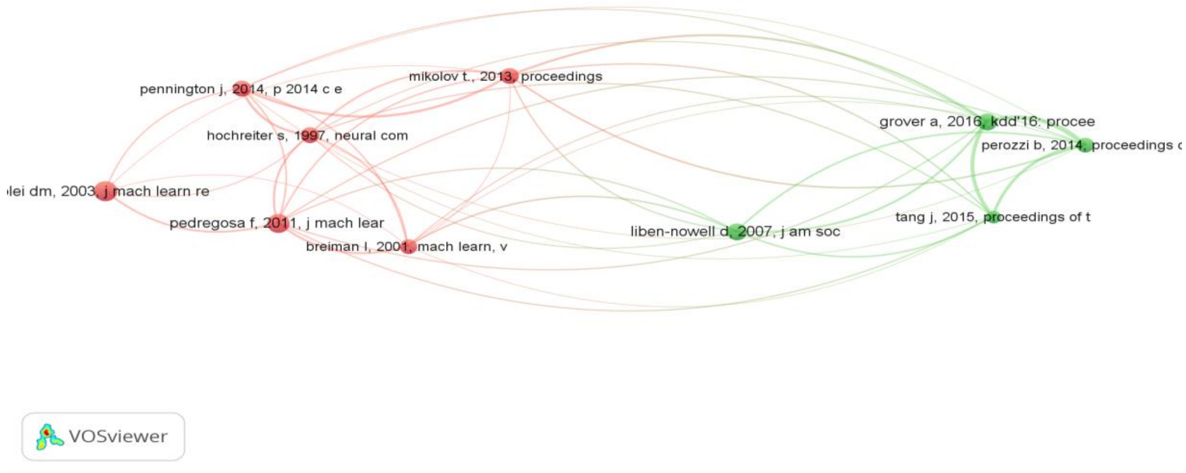

Figure 11 References-based Co-Citation Analysis.

Keyword co-occurrence analysis generates a network that shows the keywords that are used together, based on the strength of their connections. The size of the bubbles representing keywords indicates the level of co-occurrence and the connections they have with other keywords (refer to Figure 12). The network illustrates the association of the term "machine learning" with NLP, sentiment analysis, text mining, and social network analysis (represented by red bubbles), as well as feature extraction and support vector machine (represented by a blue circle). The said topics are distinguished by bubbles of the same color. The red circle explores the utilization of NLP, deep learning, sentiment analysis, and tweets in the field of ML. The blue color network represents the advanced computing techniques, including feature extraction, classification algorithms, and support vector machine, in the realm of social media. In addition, the yellow network highlights developing areas such as artificial intelligence and deep learning.

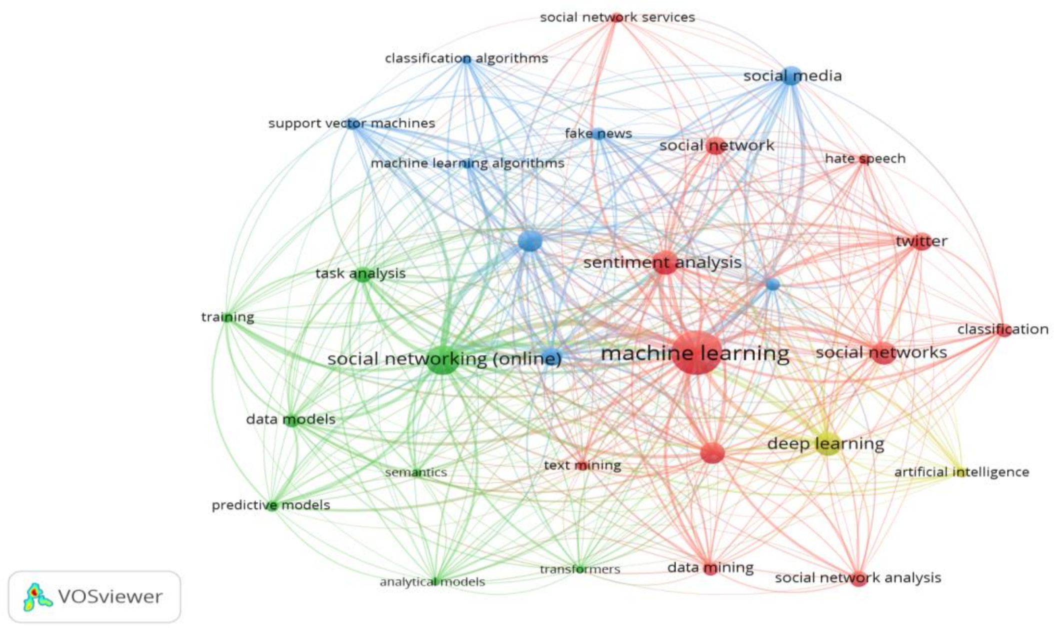

Figure 12 Authors keyword-based Co-occurrence network.

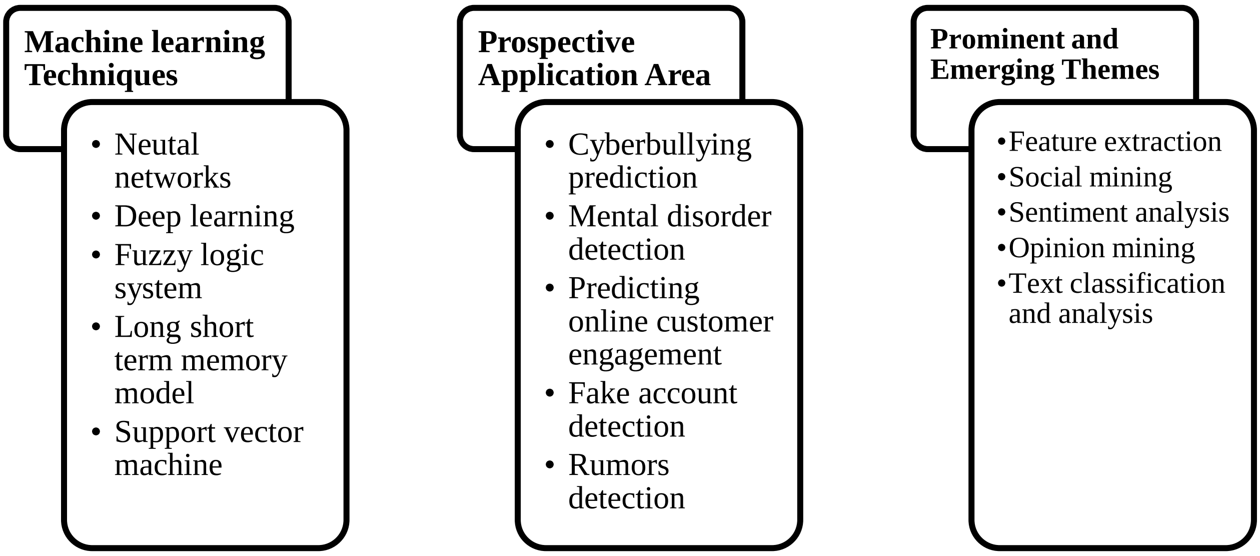

Figure 13 Framework for Emerging Themes

Discussions

This section provides the main results of the bibliometric study and identifies emergent opportunities for future research that address all of the research questions provided at the beginning.

Major Findings

A descriptive analysis was undertaken to address the first question, and the findings are shown in the following points. The nations where researchers have made significant contributions in applying ML to social networks are China and the USA. Over 21 years, these countries became a center of study in the application of machine learning techniques to social networks. Other countries' scholars frequently quote and utilize these studies for application in various marketing fields. Regarding TCs, the United States, China, and India are the most prominent countries, whereas Germany is the most frequently mentioned country in terms of Citation Per Paper (refer to Table 2). China, India, and the USA have the highest number of multi-country publications in terms of national collaboration (refer to Figure 4). According to Table 3, the journals IEEE Access and Journal of Machine Learning Research are ranked highest among all specialized journals in artificial intelligence, machine learning, and other computer approaches. They are recognized as the most prolific and most cited sources in the field. Regarding the most frequently referenced and productive authors in the examined subject domain, Lise Getoor emerges as the leading author. Getoor holds a professorship at the University of California and is a distinguished researcher specializing in machine learning, graph and network analysis, and information integration (refer to Table 4). The research findings indicate that sentiment analysis, feature extraction, NLP, deep learning, and social media analysis are the most prominent and emerging themes identified in Figure 5. To address the second research question of the study, the major findings are presented below in the form of sub-points. When considering the overall strength of links for bibliographies and citations, Pan Hui is the top author and China is the top country. This information can be found in Tables 5 and 6. Based on Figure 10, Imran Ashraf emerges as the preeminent author with the greatest level of collaboration with other authors in the domain of ML and social networks. The author is located within the cluster that exhibits the greatest number of connections among all authors. Grover & Leskovec (2016) have the highest citations that link to other documents, forming mutual citations. (Perozzi et al. (2014) was ranked as the second most interconnected and referenced document, as shown in Figure 11. The author keywords with the highest co-occurrence and linkages are machine learning, social networking, sentiment analysis, feature extraction, blogs, Twitter, social media, deep learning, natural language processing, and task analysis (Figure 11). According to Figure 6, the research topic labeled 'Machine learning and Artificial Intelligence' is the most concentrated with a large number of research publications (red cluster). Conversely, the second most concentrated group, known as the 'Sentiment analysis in social media' cluster, is recognised as a study topic with significant potential for future investigation (blue cluster). The conceptual framework in this study was constructed based on the findings of the thematic map, co-occurrence network, and conceptual structure plot, as seen in the third answer. The framework identifies promising areas where future researchers can apply machine learning and associated concepts. Figure 13 depicts the conceptual framework for the central theme of machine learning and its associated sub-themes. The framework demonstrates that ML techniques such as neural networks, deep learning, long short-term memory networks and support vector machines can be applied to different areas of marketing in the digital age. These areas include predicting online customer engagement, predicting customer churn, predicting personality based on user behaviour, detecting fake accounts, and detecting fake news. Future research projects can integrate these topics to generate optimal and cutting-edge outcomes for researchers and organisations.

Theoretical and Practical Implication

This study will contribute to the existing knowledge and assist researchers in identifying and implementing new themes regarding the use of machine learning technology in social networking sites to impact customers' online purchase intention. In addition to providing thematic support, it offers scholars information about the leading countries where machine learning studies are commonly conducted. This will allow them to cooperate with the researchers from these exceptionally productive nations. Moreover, the paper presents the most efficient sources that encompass a substantial amount of research in the field of ML and social networking. In addition, it offers insights into the utilization of machine learning in several fields, such as identifying fake news and fake accounts, predicting customer churn and engagement, and detecting emotions in social media networks. Companies can utilize machine learning and artificial intelligence-powered sentiment models to generate the expressed opinions of those who have an interest or involvement in the company. The application of machine learning-based research can assist organizations in using machine learning models to forecast customer behavior in the digital realm.

Limitations

This bibliometric study offers valuable information, but it also has certain drawbacks that can be addressed in future research. Initially, the data extraction was exclusively conducted using the WoS database. Furthermore, a specific search query was utilized to retrieve the corpus. Additional keywords may be included in future research to obtain the synthesis of other interconnected disciplines. Furthermore, the study does not address the systematic examination of literature, which could be a potential avenue for future research.

Future Direction for the Theme

The potential scope of ML in social networking is presented in the below-mentioned points.

1. Personality Prediction: By leveraging advanced algorithms and large-scale data analytics, researchers and practitioners can uncover patterns and insights that elucidate an individual's personality traits from their online interactions.
2. Customer Churn Detection: The predictive analytics technique utilizes ML algorithms to analyze various customer-related data and identify patterns or indicators that precede churn.
3. Customer Online Engagement: Predictive online customer engagement involves using machine learning and predictive analytics techniques to anticipate and optimize customer interactions in online environments. By analyzing various data sources and customer behaviors, businesses can predict how customers are likely to engage with their online platforms and content.

Conclusion

This article presents a unique study that uses bibliometric analysis and network visualization to examine the use of machine learning models in social networking from 2003 to 2024. The study demonstrates that China is the leading country globally in terms of research output, having produced 966 articles on the subject. This conclusion is based on an analysis of the most influential countries, authors, journals, and themes. Furthermore, the authors from China have engaged in substantial collaboration with researchers from several other countries. Furthermore, the United States of America is the country that places the highest emphasis on citations for the research theme. Moreover, the popular subject of 'machine learning and artificial intelligence' indicates the future emergence alongside the current expansion. This study provides valuable insights for academics by allowing them to identify the nations and authors that are actively engaged in machine learning within the social networking domain. In addition, it offers the potential application of machine learning for detecting false accounts, predicting personality traits, forecasting online customer churn, predicting customer turnover, and identifying instances of cyberbullying. Future academics can utilize machine learning to integrate various domains and offer assistance to corporations in making strategic decisions.

References

[1] Airoldi, E. M., Blei, D., Fienberg, S., & Xing, E. (2008). Mixed Membership Stochastic Blockmodels. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems* (Vol. 21). Curran Associates, Inc.

https://proceedings.neurips.cc/paper_files/paper/2008/file/8613985ec49eb8f757ae6439e879bb2a-Paper.pdf
[2] Al-Zoubi, A. M., Faris, H., Alqatawna, J., & Hassonah, M. A. (2018). Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts. *Knowledge-Based Systems*, *153*, 91–104. https://doi.org/https://doi.org/10.1016/j.knosys.2018.04.025
[3] Aljarah, I., Habib, M., Hijazi, N., Faris, H., Qaddoura, R., Hammo, B., Abushariah, M., & Alfawareh, M. (2020). Intelligent detection of hate speech in Arabic social network: A machine learning approach. *Journal of Information Science*, *47*(4), 483–501. https://doi.org/10.1177/0165551520917651
[4] Antonakaki, D., Fragopoulou, P., & Ioannidis, S. (2021). A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks ☆. In *Expert Systems with Applications* (Vol. 164). https://doi.org/10.1016/j.eswa.2020.114006
[5] Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/https://doi.org/10.1016/j.joi.2017.08.007
[6] Aslam, N., Rustam, F., Lee, E., Washington, P. B., & Ashraf, I. (2022). Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model. *IEEE Access*, *10*, 39313–39324. https://doi.org/10.1109/ACCESS.2022.3165621
[7] Ballestar, M. T., Grau-Carles, P., & Sainz, J. (2019). Predicting customer quality in e-commerce social networks: a machine learning approach. *Review of Managerial Science*, *13*(3), 589–603. https://doi.org/10.1007/s11846-018-0316-x
[8] Barushka, A., & Hajek, P. (2020). Spam detection on social networks using cost-sensitive feature selection and ensemble-based regularized deep neural networks. *Neural Computing and Applications*, *32*(9), 4239–4257. https://doi.org/10.1007/s00521-019-04331-5
[9] Ben Jabeur, S., Ballouk, H., Ben Arfi, W., & Sahut, J.-M. (2023). Artificial intelligence applications in fake review detection: Bibliometric analysis and future avenues for research. *Journal of Business Research*, *158*, 113631. https://doi.org/https://doi.org/10.1016/j.jbusres.2022.113631
[10] Bhat, S. Y., & Abulaish, M. (2013). Community-based features for identifying spammers in online social networks. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 100–107.
[11] Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134. https://doi.org/https://doi.org/10.1016/j.knosys.2021.107134
[12] Bonkra, A., Pathak, S., Kaur, A., & Shah, M. A. (2024). Exploring the trend of recognizing apple leaf disease detection through machine learning: a comprehensive analysis using bibliometric techniques. *Artificial Intelligence Review*, *57*(2), 21. https://doi.org/10.1007/s10462-023-10628-8
[13] Borkar, V. S., & Reiffers-Masson, A. (2022). Opinion Shaping in Social Networks Using Reinforcement Learning. *IEEE Transactions on Control of Network Systems*, *9*(3),

1305–1316. https://doi.org/10.1109/TCNS.2021.3117231

[14] Bostani, H., & Sheikhan, M. (2017). Modification of supervised OPF-based intrusion detection systems using unsupervised learning and social network concept. *Pattern Recognition*, *62*, 56–72. https://doi.org/https://doi.org/10.1016/j.patcog.2016.08.027

[15] Buettner, R. (2017). Predicting user behavior in electronic markets based on personality-mining in large online social networks. *Electronic Markets*, *27*(3), 247–265. https://doi.org/10.1007/s12525-016-0228-z

[16] Chen, W., Liu, Z., Sun, X., & Wang, Y. (2010). A game-theoretic framework to identify overlapping communities in social networks. *Data Mining and Knowledge Discovery*, *21*(2), 224–240. https://doi.org/10.1007/s10618-010-0186-6

[17] Cheng, S., Zhang, B., Zou, G., Huang, M., & Zhang, Z. (2019). Friend recommendation in social networks based on multi-source information fusion. *International Journal of Machine Learning and Cybernetics*, *10*(5), 1003–1024. https://doi.org/10.1007/s13042-017-0778-1

[18] Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment Analysis Based on Deep Learning: A Comparative Study. In *Electronics* (Vol. 9, Issue 3). https://doi.org/10.3390/electronics9030483

[19] Ding, Y., Yan, S., Zhang, Y., Dai, W., & Dong, L. (2016). Predicting the attributes of social network users using a graph-based machine learning method. *Computer Communications*, *73*, 3–11. https://doi.org/https://doi.org/10.1016/j.comcom.2015.07.007

[20] Fang, X., & Hu, P. (2018). Top Persuader Prediction for Social Networks. *MIS Quarterly*, *42*, 63–82. https://doi.org/10.25300/MISQ/2018/13211

[21] Fu, Q., Feng, B., Guo, D., & Li, Q. (2018). Combating the evolving spammers in online social networks. *Computers & Security*, *72*, 60–73. https://doi.org/https://doi.org/10.1016/j.cose.2017.08.014

[22] Grover, A., & Leskovec, J. (2016). node2vec: Scalable Feature Learning for Networks. *KDD : Proceedings. International Conference on Knowledge Discovery & Data Mining*, *2016*, 855–864. https://doi.org/10.1145/2939672.2939754

[23] Guimarães, R. G., Rosa, R. L., Gaetano, D. De, Rodríguez, D. Z., & Bressan, G. (2017). Age Groups Classification in Social Network Using Deep Learning. *IEEE Access*, *5*, 10805–10816. https://doi.org/10.1109/ACCESS.2017.2706674

[24] Javari, A., & Jalili, M. (2014). Cluster-Based Collaborative Filtering for Sign Prediction in Social Networks with Positive and Negative Links. *ACM Trans. Intell. Syst. Technol.*, *5*(2). https://doi.org/10.1145/2501977

[25] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

[26] Kenger, Ö. N., & Özceylan, E. (2023). Fuzzy min–max neural networks: a bibliometric and social network analysis. *Neural Computing and Applications*, *35*(7), 5081–5111. https://doi.org/10.1007/s00521-023-08267-9

[27] Kilroy, D., Healy, G., & Caton, S. (2022). Using Machine Learning to Improve Lead Times in the Identification of Emerging Customer Needs. *IEEE Access*, *10*, 37774–37795. https://doi.org/10.1109/ACCESS.2022.3165043

[28] Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014).

Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, *95*(3), 357–380. https://doi.org/10.1007/s10994-013-5415-y

[29] Kumar, M., Mishra, S., & Biswas, B. (2022). PWAF : Path Weight Aggregation Feature for link prediction in dynamic networks. *Computer Communications*, *191*, 438–458. https://doi.org/https://doi.org/10.1016/j.comcom.2022.05.019

[30] Lee, E., Rustam, F., Washington, P. B., Barakaz, F. E., Aljedaani, W., & Ashraf, I. (2022). Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets Using Stacked Ensemble GCR-NN Model. *IEEE Access*, *10*, 9717–9728. https://doi.org/10.1109/ACCESS.2022.3144266

[31] Li, X., Wang, M., & Liang, T.-P. (2014). A multi-theoretical kernel-based approach to social network-based recommendation. *Decision Support Systems*, *65*, 95–104. https://doi.org/https://doi.org/10.1016/j.dss.2014.05.006

[32] Li, Y., Peng, Y., Ji, W., Zhang, Z., & Xu, Q. (2017). User Identification Based on Display Names Across Online Social Networks. *IEEE Access*, *5*, 17342–17353. https://doi.org/10.1109/ACCESS.2017.2744646

[33] Li, Z., Fang, X., Bai, X., & Sheng, O. R. L. (2017). Utility-based link recommendation for online social networks. *Management Science*, *63*(6), 1938–1952.

[34] Lifang, L., Wen, H., & Zhang, Q. (2022). Characterizing the role of Weibo and WeChat in sharing original information in a crisis. *Journal of Contingencies and Crisis Management*, *31*, n/a-n/a. https://doi.org/10.1111/1468-5973.12433

[35] Liu, H., Maes, P., & Davenport, G. (2006). Unraveling the Taste Fabric of Social Networks. *Int. J. Semantic Web Inf. Syst.*, *2*, 42–71. https://doi.org/10.4018/jswis.2006010102

[36] Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*(1), 381–386.

[37] Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, *153*, 112986. https://doi.org/https://doi.org/10.1016/j.eswa.2019.112986

[38] Meo, P. De, Ferrara, E., Rosaci, D., & Sarné, G. M. L. (2015). Trust and Compactness in Social Network Groups. *IEEE Transactions on Cybernetics*, *45*(2), 205–216. https://doi.org/10.1109/TCYB.2014.2323892

[39] Nilashi, M., Yadegaridehkordi, E., Ibrahim, O., Samad, S., Ahani, A., & Sanzogni, L. (2019). Analysis of Travellers' Online Reviews in Social Networking Sites Using Fuzzy Logic Approach. *International Journal of Fuzzy Systems*, *21*(5), 1367–1378. https://doi.org/10.1007/s40815-019-00630-0

[40] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: online learning of social representations. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 701–710. https://doi.org/10.1145/2623330.2623732

[41] Rosenblatt, F. (1958). *Two theorems of statistical separability in the perceptron*. United States Department of Commerce Washington, DC, USA.

[42] Rosenblatt, F. (1960). Perceptron simulation experiments. *Proceedings of the IRE*, *48*(3), 301–309.

[43] Rossetti, G., Pappalardo, L., Pedreschi, D., & Giannotti, F. (2017). Tiles: an online algorithm for community discovery in dynamic social networks. *Machine Learning*, *106*(8), 1213–1241. https://doi.org/10.1007/s10994-016-5582-8
[44] Rupapara, V., Rustam, F., Shahzad, H. F., Mehmood, A., Ashraf, I., & Choi, G. S. (2021). Impact of SMOTE on Imbalanced Text Features for Toxic Comments Classification Using RVVC Model. *IEEE Access*, *9*, 78621–78634. https://doi.org/10.1109/ACCESS.2021.3083638
[45] Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M., & Rao, J. (2009). Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing*, *13*(6), 401–412. https://doi.org/10.1007/s00779-008-0214-3
[46] Sahoo, S. R., & Gupta, B. B. (2019). Classification of various attacks and their defence mechanism in online social networks: a survey. *Enterprise Information Systems*, *13*(6), 832–864. https://doi.org/10.1080/17517575.2019.1605542
[47] Sarna, G., & Bhatia, M. P. S. (2017). Content based approach to find the credibility of user in social networks: an application of cyberbullying. *International Journal of Machine Learning and Cybernetics*, *8*(2), 677–689. https://doi.org/10.1007/s13042-015-0463-1
[48] Shuai, H.-H., Shen, C.-Y., Yang, D.-N., Lan, Y.-F. C., Lee, W.-C., Yu, P. S., & Chen, M.-S. (2018). A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining. *IEEE Transactions on Knowledge and Data Engineering*, *30*(7), 1212–1225. https://doi.org/10.1109/TKDE.2017.2786695
[49] Su, D., Liu, J., Wang, X., & Wang, W. (2019). Detecting Android Locker-Ransomware on Chinese Social Networks. *IEEE Access*, *7*, 20381–20393. https://doi.org/10.1109/ACCESS.2018.2888568
[50] Sweileh, W. M., AbuTaha, A. S., Sawalha, A. F., Al-Khalil, S., Al-Jabi, S. W., & Zyoud, S. H. (2016). Bibliometric analysis of worldwide publications on multi-, extensively, and totally drug - resistant tuberculosis (2006-2015). *Multidisciplinary Respiratory Medicine*, *11*, 45. https://doi.org/10.1186/s40248-016-0081-0
[51] Tabassum, S., Pereira, F., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*, e1256. https://doi.org/10.1002/widm.1256
[52] Umer, M., Ashraf, I., Mehmood, A., Kumari, S., Ullah, S., & Sang Choi, G. (2021). Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model. *Computational Intelligence*, *37*(1), 409–434. https://doi.org/https://doi.org/10.1111/coin.12415
[53] Varshney, D., Kumar, S., & Gupta, V. (2017). Predicting information diffusion probabilities in social networks: A Bayesian networks based approach. *Knowledge-Based Systems*, *133*, 66–76. https://doi.org/https://doi.org/10.1016/j.knosys.2017.07.003
[54] Verbeke, W., Martens, D., & Baesens, B. (2014). Social network analysis for customer churn prediction. *Applied Soft Computing*, *14*, 431–446. https://doi.org/https://doi.org/10.1016/j.asoc.2013.09.017
[55] Wanda, P., & Jie, H. J. (2020). DeepProfile: Finding fake profile in online social

network using dynamic CNN. *Journal of Information Security and Applications*, *52*, 102465. https://doi.org/https://doi.org/10.1016/j.jisa.2020.102465
[56] Wei-dong, H., Qian, W., & Jie, C. (2018). Tracing Public Opinion Propagation and Emotional Evolution Based on Public Emergencies in Social Networks. *International Journal of Computers Communications & Control*, *13*, 129. https://doi.org/10.15837/ijccc.2018.1.3176
[57] Wu, L., Zhang, Q., Chen, C.-H., Guo, K., & Wang, D. (2020). Deep Learning Techniques for Community Detection in Social Networks. *IEEE Access*, *8*, 96016–96026. https://doi.org/10.1109/ACCESS.2020.2996001
[58] Xu, D., Shi, Y., Tsang, I. W., Ong, Y.-S., Gong, C., & Shen, X. (2020). Survey on Multi-Output Learning. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(7), 2409–2429. https://doi.org/10.1109/TNNLS.2019.2945133
[59] Yang, T., Chi, Y., Zhu, S., Gong, Y., & Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Machine Learning*, *82*(2), 157–189. https://doi.org/10.1007/s10994-010-5214-7
[60] Zhang, J., Srivastava, D. P., Sharma, V. D., & Eachempati, P. (2021). Big Data Analytics and Machine Learning: A Retrospective Overview and Bibliometric Analysis. *Expert Systems with Applications*, *184*. https://doi.org/10.1016/j.eswa.2021.115561
[61] Zheng, X., Zeng, Z., Chen, Z., Yu, Y., & Rong, C. (2015). Detecting spammers on social networks. *Neurocomputing*, *159*, 27–34. https://doi.org/https://doi.org/10.1016/j.neucom.2015.02.047