Journal of
Information Systems Engineering
and Business Intelligence

Vol.11, No.2, June 2025
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

A Systematic Literature Review of Topic Modeling Techniques in User Reviews

Ilham Zharif Mustaqim [1)*], Ryan Randy Suryono[2)]

[1)]*Faculty of Computer Science, Universitas Indonesia, Jakarta, Indonesia*
[1)]ilham.zharif@ui.ac.id

[2)]*Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia*
[2)]ryan@teknokrat.ac.id

Abstract

Background: The escalating volume of user review data is necessitating automated methods for extracting valuable insights. Topic modeling was a vital method for understanding key discussions and user opinions. However, there was no comprehensive analysis of the scientific work specifically on topic modeling applied to user review datasets, including its main applications and a comparative analysis of the strengths and limitations of identified methods. This study addressed the gap by characterizing the scientific discussion, identifying potential directions, and exploring currently underutilized application areas within the context of user review analysis.
Objective: This study aimed to recognize the implementation trend of topic modeling in various areas and to comprehend the methodology that could be applied to the user review dataset.
Methods: A systematic literature review (SLR) was adopted by implementing Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines within six-year spans, narrowing 1746 to 28 selected primary studies.
Results: The underlying insight was that user reviews had been critical as the primary data for topic modeling in analyzing various applications. Digital banking and transportation applications were the sectors that received the greatest attention. In this context, Latent Dirichlet Allocation (LDA) was the most extensively used method, with a focus on overcoming its limitations by incorporating additional strategies into LDA-based models.
Conclusion: The bibliometric analysis and mapping study practically contributed as a reference when assessing the dominant topic in similar app categories and topic modeling algorithms. Furthermore, this study comprehensively analyzed various topic modeling algorithms, presenting both the strengths and weaknesses of informed selection in relevant applications. Considering the keywords cluster analysis, service quality could be adopted based on the output of the topic modeling.

Keywords: Topic modeling, User review, Systematic literature review, Bibliometric analysis

Article history: Received 10 March 2025, first decision 15 May 2025, accepted 17 June 2025, available online 22 July 2025

I. INTRODUCTION

Topic modeling is a method for summarizing a large collection of documents by applying probabilistic word distributions [1], [2]. This method is used in knowledge discovery to reveal the hidden themes in a collection of documents. Various methods were categorized into four, namely Algebraic, Fuzzy, Probabilistic, and Neural [1]. The four primary methods exhibit distinctions in the underlying mathematical principles, treatment of uncertainty, level of interpretability, and computational complexity [1]. Latent Dirichlet Allocation (LDA) stands out as a particularly significant algorithm among the diverse methods developed for topic modeling [3]. According to previous studies, LDA is widely used for clustering topics but the method cannot be directly implemented on user reviews and other short texts due to data sparsity issues and a lack of co-occurrence patterns [4], [5]. A key factor influencing topic model selection is the input dataset, as its characteristics can significantly impact performance [1].

User-generated ratings and reviews constitute a form of feedback that can contribute to the enhancement of software quality and the identification of desired application features [6]. As unstructured data, user reviews can be used to discover valuable knowledge in decision-making mechanisms for the organization. Reviews serve as a crucial resource for requirements engineering, facilitating the evolution and improvement of applications grounded in actual user needs and sentiments [7], [8]. The development of automated tools and machine learning methods aimed to efficiently

[*] Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.11.2.238-253

categorize, analyze, and extract actionable insights from extensive review datasets. This process mitigates bias and enhances the precision of app evaluations from the user's standpoint [6], [7].

Topic modeling has been widely used to uncover hidden patterns in large datasets. The growing accessibility of extensive textual datasets from digital media has led communication scholars and social scientists to use topic modeling as an automated analytical method for text [9]. The application of topic modeling to user reviews provides a robust methodology for deriving actionable insights, informing app development strategies, and improving user satisfaction through the systematic identification of salient themes within extensive feedback datasets [10], [11]. Wulandari and Hidayanto [12] adopted topic modeling to assess the service quality of contract tracing in a healthcare application based on user reviews using LDA. Çallı [13] used LDA as a topic modeling method to evaluate customers' reviews on a digital banking application. The LDA method is a well-suited method for long documents and large corpora, considering the co-occurrence patterns of words within a document. However, its ineffectiveness on short texts due to data sparsity and the lack of co-occurrence patterns has led to LDA adaptation for short texts as an observable study topic. A strategy for overcoming data sparsity, as found in the literature, is aggregating small texts to generate pseudo-long studies [4]. The use of NMF-based methods represents another efficacious method for topic mining in short texts, complementing the more prevalent LDA-based, as shown with tweet datasets [14].

Although previous studies have explored topic modeling, a gap exists in the literature concerning a focused synthesis of the method specifically applied to user review datasets and conducted within the framework of systematic literature review guidelines. This study was conducted following the need for a comprehensive characterization of scientific production around topic modeling, as well as the identification of its leading applications and evolutions. Considering the current state of the literature, this study was conducted to acknowledge the potential areas within topic modeling based on user review datasets and explore the untapped app category. The mapping study practically contributes as a reference when assessing the dominant topic in the same app categories. Therefore, this study aimed to (1) recognize the implementation trend of topic modeling in various areas and (2) comprehend the topic modeling methodology that can be applied to the user review dataset.

The structure of this study consists of five sections and the first section elaborates on the background and motivation. The second section delves into the related work concerning topic modeling methods and the overview of user reviews. The third section provides a detailed explanation of the proposed methodology. The fourth offers the presentation of the results covering bibliometric analysis and mapping, while the last section outlines the conclusions and provides insights into future study directions.

II. LITERATURE REVIEW

A. Topic Modeling Techniques

Topic modeling represents an unsupervised learning method within the field of natural language processing (NLP) that facilitates the clustering of textual documents based on the underlying latent semantic structure [15]. This method is a mathematical model used in the domain of machine learning, enabling the discovery of recurring word patterns within textual datasets [16]. The field has garnered attention and investigation for numerous decades, resulting in the proposal of various methodologies. Abdelrazek et al. [1] classified the method into four, namely Algebraic, Fuzzy, Probabilistic, and Neural. Each category possesses its unique mechanism, advantages, and disadvantages. For instance, Bayesian probabilistics operates by delineating a generative process using a Bayesian graphical model. The subsequent inference includes working backward and the strengths lie in simplicity, intuitiveness, extensibility, and interpretability. However, the challenges arise when the model complexity escalates, leading to increased complications in the inference process.

B. App User Review

App user reviews available in application stores offer crucially valuable insights to aid software engineers in comprehending user needs and in the creation, debugging, and enhancement of software products [17]. These app reviews consist of textual feedback coupled with a star rating, allowing app users to share experiences with others and developers. The rating scores that are complemented with user reviews serve as a clear indicator of the satisfaction levels with the provided services [8], [18].. Typically, reviews are concise, spanning up to 675 characters, and cover a range of topics such as feature requests, bug reports, or user opinions. For requirements engineering purposes, the analysis of app reviews proves beneficial, assisting software engineers in extracting new insights regarding desired app features [17].

Haggag et al. [19] carried out an extensive study on health applications, scrutinizing 278 mHealth app reviews, where 5 million were extracted and translated to gain insights into the significant issues. The study showed that among all mHealth subcategories analyzed, fitness activity tracking was the least favorably rated app subcategory, attributed to diverse issues and challenges across multiple investigated aspects. More than half of the users who reported problems leading to uninstallation in the reviews assigned a 1-star rating to the app.

C. Reviews of Secondary Studies

Previous studies have summarized topic modeling methods as shown in Table 1. Silva et al. [20] discussed the implementation of topic modeling in software engineering studies. Grisales et al. [2] investigated the evolution of the topic modeling method through bibliometric analysis. Abdelrazek et al. [1] outlined study insufficiencies, showed similarities and dissimilarities between topic models and model categories, reviewed extant benchmark datasets in topic modeling, as well as provided an overview of diverse tools and applications. Laureate et al. [21] examined the application of topic modeling to social media analysis with a particular focus on the users, the underlying motivations, and the methodologies used. However, none of the previous studies specifically summarized the method of topic modeling using datasets as well as adopted systematic literature review guidelines.

TABLE 1
SECONDARY STUDIES

Reference	Objective	Study Gap
[1]	Identified gaps in the existing study, to compare topic modeling methods, and addressed the inconsistency in evaluation metrics across studies by proposing a standardized set of metrics.	Comparative analysis of topic modeling methodologies, evaluating their respective advantages and disadvantages at a categorical level
[2]	Investigated the developmental trajectory of topic modeling, its principal field of application, and the selection of optimal models for distinct data modalities.	A multidisciplinary nature was observed, not analyzing topic modeling to define topics in specific domains.
[20]	Examined the application of topic modeling, showing its strengths and weaknesses, addresses the selection of suitable methods, the critical role of data preparation, the challenges of interpreting and labeling topics, and the necessity of contextual awareness in topic naming.	A comprehensive review of topic modeling's evolution within software engineering, and an analysis of its application and implementation domains remain absent.
[21]	Explored the application of topic modeling in social media analysis, focusing on the users, motivations, and methodologies	Focused on social media data, not including the comprehensive picture of topic modeling applications

III. METHODS

A systematic literature review (SLR) was conducted to investigate the field of topic modeling using user review datasets. The primary objective of an SLR was to pinpoint, assess, and interpret all pertinent previous studies. In this section, the SLR protocol was elaborated, which adopted Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA). This methodology was regarded as exceptionally efficient, trustworthy, and all-encompassing, rendering it an essential tool for conducting rigorous and transparent SLRs across various disciplines [22]. This method covered three phases, namely identification, screening, and inclusion.

A. Identification Phase

The initial step included the development of a comprehensive strategy to proceed with the review process. This strategy included selecting electronic databases, building search queries, defining inclusion criteria, and outlining quality assessment criteria. Following the study objectives, two questions were formulated, namely (1) What were the main areas of topic modeling implementation based on the user review dataset? (2) What was the recent trend of methods for topic modeling based on user review datasets in the proposed literature?

1) Selection of Digital Repository

The Institute of Electrical and Electronics Engineers (IEEE) and the Association for Computing Machinery (ACM) were selected as the electronic databases, due to the comprehensive coverage of a wide range of reputable journals and proceedings in this field [23]. To extend the literature scope, some other sources were used, namely Emerald Insight, ScienceDirect, Scopus, Taylor & Francis, SpringerLink, and Sage Journal.

2) Search String

The search string was constructed to accommodate some categories, such as topic modeling, data type, and sources. The final search string was shown as *("Topic Modeling" OR "Topic Model") AND (User Review OR "User Comment") AND ("Google Play Store" OR "Apple Store") AND ("Method" OR "Technique").*

3) Inclusion and Exclusion Criteria

The inclusion criteria consisted of a series of specified qualifications that the publications must satisfy. Meanwhile, the exclusion criteria represented the unwanted characteristics, thereby simplifying the process of rejecting undesired publications [24], as shown in Table 2.

TABLE 2
INCLUSION AND EXCLUSION CRITERIA

Phase	Inclusion Criteria	Exclusion Criteria
Identification	Published in the last 6 years: 2018-2024 Written in English	Not published between 2018-2024
Screening	Focus on discussing topic-modeling themes. Published in an international journal or proceeding through a peer-reviewed paper process	SLR paper/Literature River/Conference Notes/ Speaker Notes Focus on discussing other than topic modeling themes
Eligibility	-	Paper can be downloaded as a complete document (full text)

4) Quality Assessment

The quality of studies was evaluated using a quality checklist (QC).This procedure aimed to determine the relevance of the chosen studies in the context of the objectives. A set of questions outlined in Table 3 was used to validate the quality of the selected studies. Based on these checklist questions, selected studies were assessed and marked accordingly. Any article that provides a clear response to a checklist question received a score of 1. However, when any article that failed to address the quality checklist questions was marked with a score of 0. When the selected study partially addressed the checklist questions, a score of 0.5 was assigned [25], [26]. The study established a threshold score of 3 to pass this assessment process.

TABLE 3
QUALITY CHECKLIST QUESTIONS

Code	Quality Checklist Questions
QC1	Is the outcome and analysis of the study aligned with the proposed study questions?
QC2	Do the findings of the chosen study rely on primary data?
QC3	Does the literature explore topic modeling techniques?
QC4	Does the study employ user reviews or comments collected from the Google Play Store or Apple App Store?

B. Screening Phase

The constructed search string resulted in 1746 studies from eight electronic databases. This was followed by a screening phase based on the inclusion and exclusion criteria, which resulted in 868 studies. The title and abstract were reviewed to ensure that the candidate studies were consistent with the objectives. A total of 71 studies passed the review and were followed by the quality assessment process and 28 were selected as the scores were above the threshold. Figure 1 shows the complete flow of the screening phase.

C. Reporting Phase

This study used a bibliometric analysis of the attributes of identified articles. Bibliometric analysis constituted a quantitative study evaluation methodology, appraising previous relevant scholarly works through the lens of quantitative indicators [27]. The increasing adoption of this method across numerous professional disciplines in recent times served to visually represent the current state of knowledge, key attributes, developmental progression, and nascent trends. Consequently, this method enabled analysts, including those new to a field, to acquire a thorough comprehension of the existing literature [28]. Key data were extracted to be analyzed based on keyword, country, and published year for further analysis.

The next step was capturing and enlisting the statements that focused on the applied area of topic modeling and the trend method that was used in the selected studies. Open coding was conducted using NVIVO 12 Pro to categorize and analyze the data.

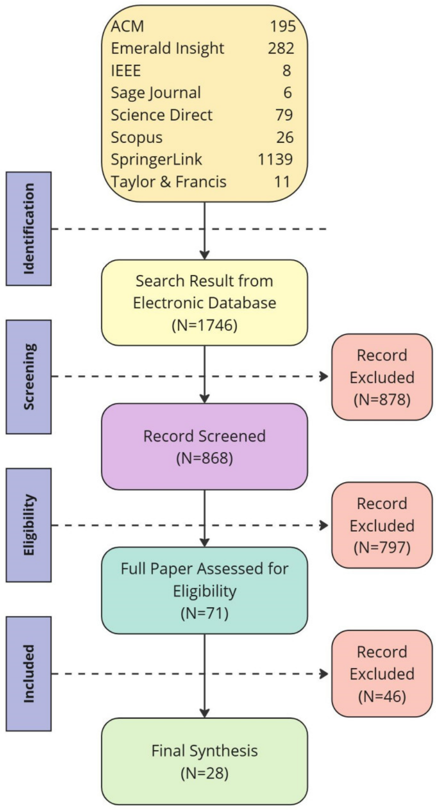

Fig. 1 Flow of SLR

IV. Results

In this section, the demographic of selected studies was discussed based on several attributes. The results were elucidated, including the trend of topic modelling in various fields and the topic modelling methodology that could be adapted to user-review datasets from the Google Play Store and Apple App Store.

A. Bibliometric Analysis

As expounded in the preceding section, 28 primary studies were designated as the foundation. These selected studies were classified according to the keyword clusters, publication year, and the home country of the initial author. The VOSViewer software was used to present trends by clustering keywords and creating a visualization [29].

1) Keyword Clusters

The keywords were analyzed using the VOSViewer program, which was able to show trends with the publication year by grouping terms from selected studies. Figure 2 showed the constructed diagram. Starting in 2019 (see the blue cluster), topic modeling using the LDA method was frequently used to study mobile multimedia applications and application categories. During that period, other works used word embedding and bag-of-words as the basis for topic modeling. After the pandemic, some studies concentrated on discussing COVID-19 applications in 2021-2022 (note the green cluster), using sentiment analysis and topic modeling methodologies. Furthermore, based on user feedback, privacy concerns were investigated.

The study trend of topic modeling shifted to service quality and customer satisfaction based on online customer reviews after 2023 (see the yellow cluster). The investigation was stretched to various fields of application, such as mobile banking, contact tracing applications, grocery mobile apps, platform management, mobile gaming, and the metaverse. Surprisingly, the LDA method was used in topic modeling studies in recent years for a variety of subjects. This network provided insight into potential study areas connected to topic modeling. For instance, none of the evaluations of e-government apps, insurance apps, and education platform apps were explored. Furthermore, service quality was one of the primary trends in 2024, and there was potential for further exploration based on each app type, considering the distinctive nature inherent to each app category.

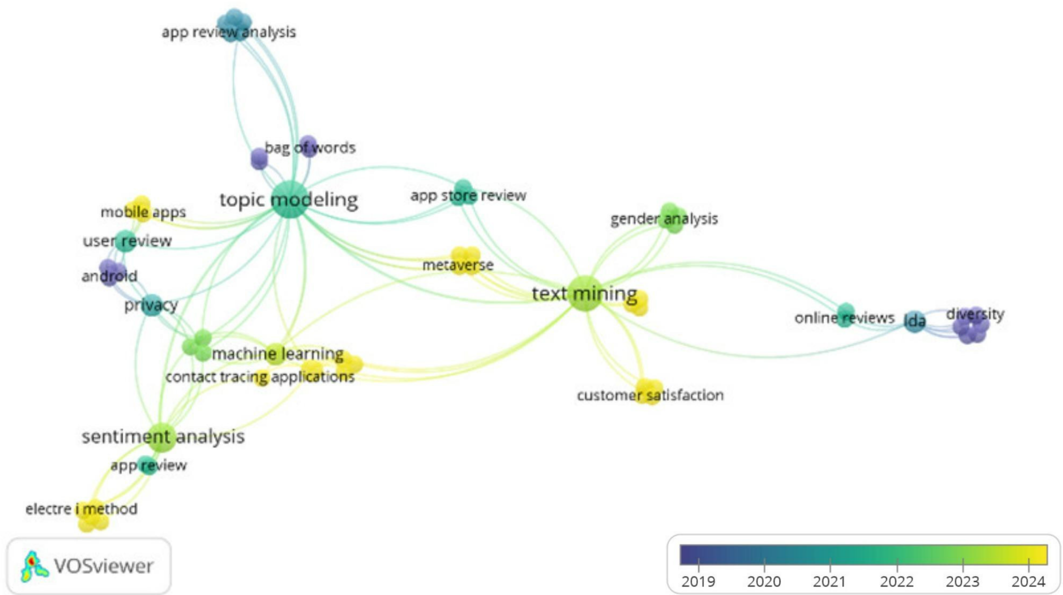

Fig. 2 Keywords clusters based on co-occurrence from 2019 to 2023

2) Country Analysis

The studies were assessed based on the geographic distribution of the first author who published an article related to topic modeling. Examining the distribution of countries allowed for the exploration of the articles' spatial and geographic distribution [28]. The first author of selected studies originated from 13 countries spread across six continents. Figure 4 shows the detailed spread of authorship geographic distribution with several related publications from those countries. In addition, it can reflect trends occurring in the region. China and Indonesia produced more publications related to topic modeling over the years. The examination of the sources of publication and the geographical dispersion of authorship was shown in Figure 3.

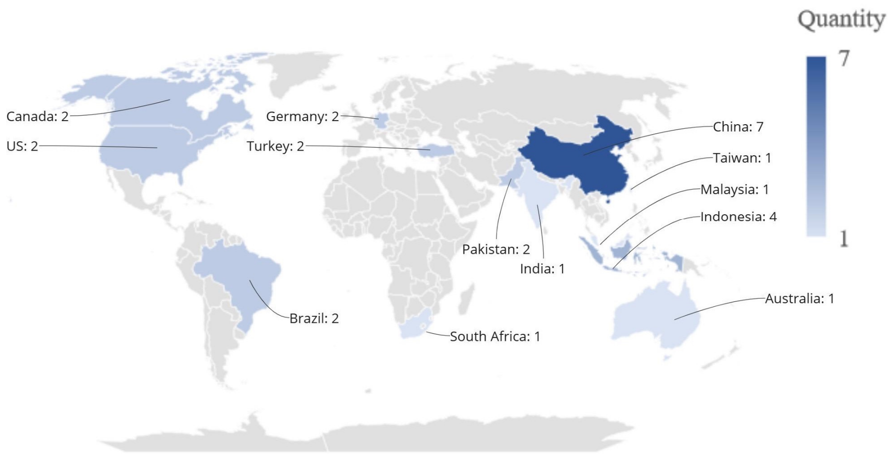

Fig. 3 Country distribution of primary studies

3) Published Year Analysis

Examining the publication year distribution of studies could help identify trends in topics and areas over time, as well as show the growth rate of studies in a particular field or discipline. This process increased the understanding of the subjects that gained or lost prominence. Figure 4 shows the publication trend of topic modeling study. The study in the topic modeling area increased over the years and peaked in 2023, which had a total of 10. This result showed that topic modeling using the user-generated data field had significant growth and interest. The study included the analysis of a single study published in 2024. The limited inclusion was due to the study's initiation in the same year, thereby precluding comprehensive coverage of all relevant publications for that entire year.

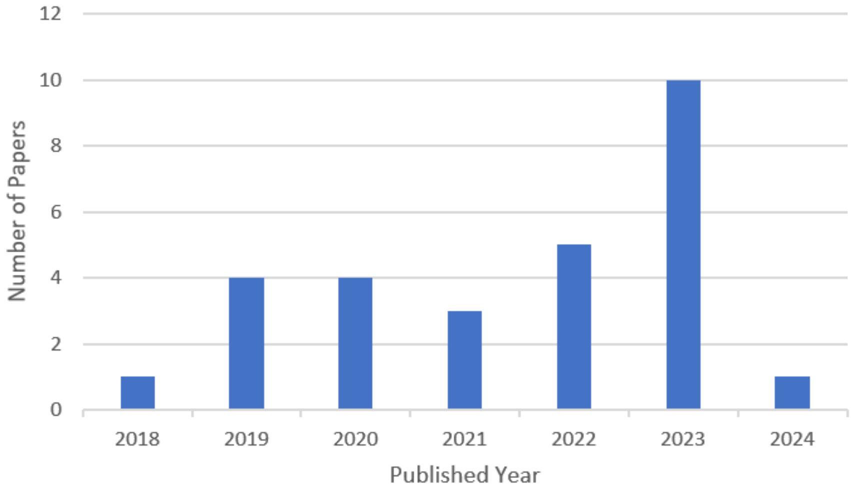

Fig. 4 Year-wise of the selected primary studies

B. RQ1: What are the main areas of topic modeling implementation based on the user review dataset?

In this section, the implementation of topic modeling in various mobile applications from the 28 selected studies was summarised. In general, topic modeling was applied to analyze the online customer reviews of mobile applications in 10 specific categories. However, some studies did not specifically narrow the implications to certain application categories, which were classified as general. Table 4 shows the list of implementation areas, as well as the dominant topic and the referred study.

TABLE 4
IMPLEMENTATION AREA OF TOPIC MODELLING

Applied Application Area	Dominant Topics	References
Digital Banking App	Perceived Usefulness; Efficiency and Accessibility; Ease of Interaction; Cost Factors	[13], [30], [31], [32], [33], [34]
Games App	Computing Resources; Price Fairness; Authorization	[35]
Grocery App	Offers and Discounts; Customer Phone Support Delivery; Delivery Commitment	[36]
Health Care App	System Performance and Reliability; Perceived Value and Benefits; Usability	[12], [37], [38]
Social Media App	Algorithm; Knowledge Base; Biased Content Curation	[39], [40]
Sport App	Quality of Content; Up-todateness; Return on Investment	[41]
Streaming App	Algorithm; Knowledge Base	[39]
Transportation App	Trip Attributes; Service and Functionality; User Experience; Transactional Aspects	[18], [39], [40], [42], [43]
Tourism App	Design; Fulfillment; Functionality	[44]
General App	-	[45], [46], [47]

Topic modeling was used extensively in the finance and transportation industries to show hidden insights from customers. With escalating competition within finance and transportation, customer satisfaction assumed a significant role in service enhancement and sustains market advantage. The frequent, often daily, use of mobile applications within these fields generated significant volumes of user-generated reviews, providing a rich dataset for analysis. Furthermore, the inherent criticality of financial transactions and transportation logistics in users' daily lives results in high review activity, particularly in instances of service disruption or dissatisfaction. This strategy aimed to cultivate a loyal customer base by prioritizing satisfaction, a significant metric for sustained usage in information systems, indicative of post-adoption behavior [13], [43].

C. RQ2: What is the recent trend of techniques for topic modeling based on user review datasets in the proposed literature?

Topic modeling was a complex field that offered insights into the structure of electronic media data and enabled the identification of sentiments using user review datasets. The method evolved, as shown in Table 5, which summarized the topic modeling method used and developed by the selected primary studies.

TABLE 5
THE TREND OF TOPIC MODELING METHODS

Topic Modeling Methods	Detailed Methods	References
Latent Dirichlet Allocation (LDA)	LDA	[12], [13], [18], [30], [32], [35], [36], [38], [39], [41], [42], [43], [44], [48], [49], [50], [51], [52]
LDA-Based	LDA + keyATM	[34]
	Hierarchical Dirichlet Process (HDP)	[38]
	Adaptively Online LDA (AOLDA)	[47], [53]
	Online Biterm Topic Model (OBTM)	[53]
	Adaptive Online Biterm Sentiment-Topic Model (AOBST)	[54]
Non-Negative Matrix Factorization (NMF)	NMF	[37], [46]
Other Techniques	Security-Related Review Miner (SRR-Miner)	[45]
	Mobile App Reviews Summarization (MARS)	[55]
	CluWords	[40]

1) Latent Dirichlet Allocation

LDA was a Bayesian probabilistic topic generation [50] and unsupervised machine learning method developed by Blei et al. [56]. The plate notation of LDA was presented to show the generative process for each document in the corpus within the framework of LDA, illustrated in Figure 6.

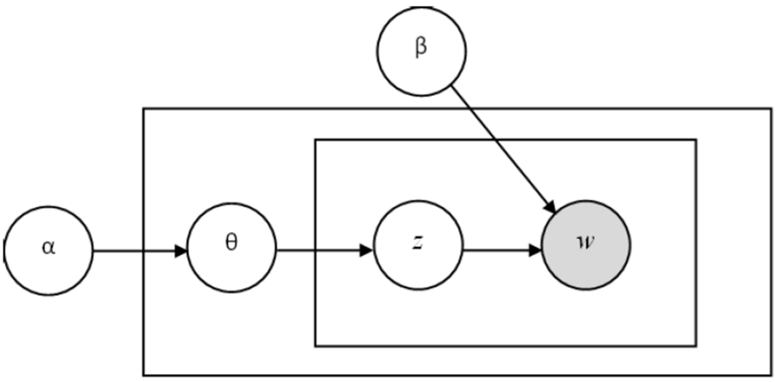

Fig. 5 LDA model with plate notation [36], [56]

The LDA method was shown in a previous study to contain three levels. Parameters (α) and (β) represented the topic distribution at the corpus level, which was a collection of M documents. α and β were used to determine the topic and word distribution within documents, respectively. The variable θ represented the topic distribution for a specific document. Z and W were word-level variables, where Z represented the topic of a particular word in a document. Meanwhile, W represented words related to a specific topic in the document [56].

Following LDA's advent, probabilistic models were the subject of intense studies for several years [1]. LDA was a powerful topic modeling method that outstripped other models in its ability to show hidden semantic relationships between words. These included words not used together in a document and the method identified documents that covered multiple topics [35]. The structure comprised several topics, each containing similar terms that were correlated with a certain probability [30]. A total of 18 of 28 studies applied LDA methods for the topic modeling, as shown in Figure 5.

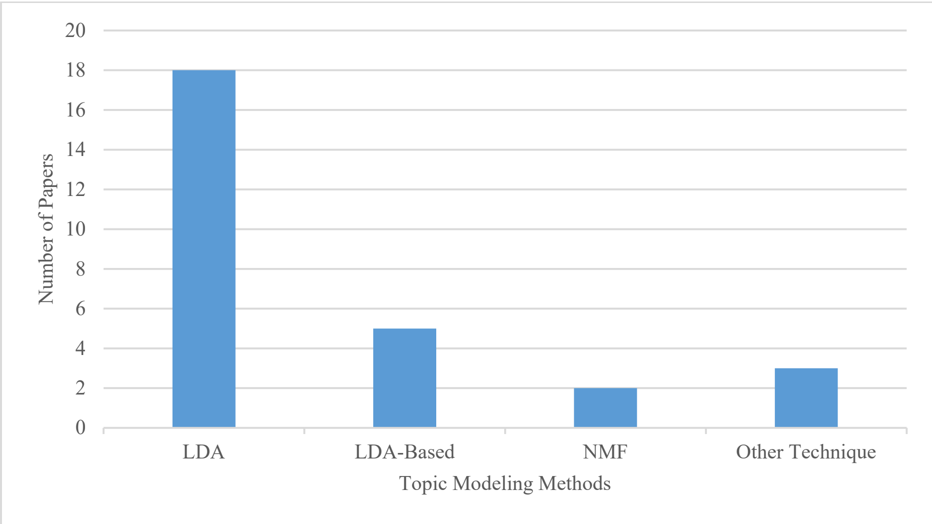

Fig. 6 Distribution of implemented topic modeling methods

2) LDA-Based Model

LDA had some challenges for certain conditions despite the numerous benefits. This method could be susceptible to data sparsity and may generate incoherent topics [46]. Therefore, this method was not considered to be well-suited for short texts, such as user reviews [54]. In response to the challenges, studies proposed novel methods, such as keyword-assisted topic modeling (keyATM) and the Hierarchical Dirichlet Process (HDP). KeyATM addressed the susceptibility of LDA to data sparsity and the generation of incoherent topics by incorporating user-defined seed words for topic-word distributions. This model outperformed LDA in both the coherence of its generated topics and its capacity to extract previously underrepresented themes from a corpus characterized by limited semantic scope [34]. In contrast to the requirement of LDA for a predefined number of topics, HDP was a nonparametric topic modeling method that automatically inferred the number of topics in the data [38]. Referring to the experimental setup conducted by Abdelrazek et al. [1], HDP had the least favorable results compared to other models based on coherence and stability metrics.

Gao et al. [57] introduced Adaptively Online Latent Dirichlet Allocation (AOLDA), a novel method to generate topic distributions designed for each version by intelligently combining information from past topics. In the context of online review analysis, AOLDA leverages previously discovered themes to develop topic models for new versions, offering a dynamic understanding of evolving customer sentiment. AOLDA surpassed the performance of its predecessor, Online Latent Dirichlet Allocation (OLDA), by adaptively incorporating insights from previous app iterations [57]. The adaptive online algorithm of LDA was able to tackle varying version contributions. However, these still inherited the susceptibility of LDA to data sparsity when dealing with short texts [58]. TOUR, a Customizable and Automatic Tool for Dynamic Topic and Sentiment Analysis of User Reviews, proposed by Yang et al. [47], aimed to empower app developers by automatically tracking changes in user sentiment and topics surrounding specific app features across different versions. This tool offered a unique "glimpse" into user reviews on a topic level, showing both individual phrases and complete sentences. Additionally, TOUR presented interactive sentiment analysis results, facilitating deeper understanding. The core innovation was in using customizable opinion words for lightweight sentiment analysis of emerging issues, removing the dependence on external tools [47].

Biterm Topic Model (BTM) was another generative probabilistic model and an extension of LDA that treated each co-occurring word pair as a single term, constructing the model on a set rather than documents [1], [53]. In pursuit of accelerating BTM inference on voluminous datasets, Cheng et al. [59] proposed online algorithms specifically designed for BTM, namely "online BTM" (OBTM). The online algorithm excelled in tackling both data sparsity [60] and dynamic data streams effectively. Its efficient data storage for updates ensured continuous model relevancy while minimizing memory and processing time [59]. Current online algorithms failed to consider the varying importance of different versions. Textual data from different times or versions significantly differed in its relevance to the latest, yet existing algorithms neglected this nuance [58].

Gao et al. [54] developed an improved issue detection method (MERIT) which was built on the Adaptive Online Biterm Sentiment-Topic Model (AOBST), addressing the sentiment and topic evaluation in app reviews. This novel unsupervised model leveraged the Biterm Sentiment-Topic (BST) model, specifically adapted to handle the brevity and sentiment-rich nature of app reviews. Recognizing the dynamic nature of app iterations, AOBST incorporated online adaptation to capture topic variations across consecutive versions. By intelligently connecting sentiment-topic word distributions from older versions, AOBST guided the previous for the current version's word distribution. This enabled the detection of developing topics and the automatic interpretation [54].

3) Non-Negative Matrix Factorization

Non-negative matrix factorization (NMF) was a popular algebraic model in NLP that decomposed a high-dimensional vector into a lower-dimensional representation, with the constraint that all the resultant representations were non-negative. As a tool for extracting high-dimensional data, NMF extracted sparse and meaningful attributes from non-negative data vectors by finding non-negative low-rank matrices [1], [37]. NMF effortlessly excelled at extracting interpretable factors with sparsity. This originated from its use of factor analysis, which inherently downplayed the influence of weakly coherent words. This characteristic showed its suitability, especially in scenarios with numerous ambiguous attributes [61]. The presented models outperformed peers in both topic coherence and accuracy, showcasing exceptional performance. As discovered by Chen et al. [14], NMF outperformed LDA when analyzing short text, such as tweets [9]. However, the implemented algorithm's reliance on random initialization introduced a degree of instability that warrants consideration [1]. Viegas et al. [40] developed the optimal NMF configuration, FV+NMF+ASToC, which consistently outperformed the recently proposed state-of-the-art topic modeling method on all explored datasets.

4) Other Techniques

Antecedent studies have proposed topic modeling methods to enrich the knowledge of topic modeling as well as focus on certain aspects or improvements. Tao et al. [45] introduced the Security-Related Review Miner (SRR-Miner), an innovative method for summarizing security concerns and user sentiments including misbehavior, aspect, and opinion words. SRR-Miner used a keyword-driven strategy to extract sentences about security from reviews, and then through an examination of the sentence structures [45]. Hatamian et al. [55] introduced Mobile App Reviews Summarization (MARS), a novel model applying machine learning, NLP, and sentiment analysis techniques, which functioned as a summarization tool designed to ease the comprehension of privacy-related statements within app user reviews. This model targeted high data protection quality by showing high efficacy in detecting privacy practices, users' concerns, and perceptions. However, it depended on a manually curated privacy threat catalog and expert-driven validation of automatically generated labels [55]. Viegas et al. [40] presented a novel document representation method and individual words were replaced with semantically related clusters ("CluWords"). However, due to CluWords' inherent nature, standard TF-IDF weighting did not fully capture the informative value. Therefore, the study proposed a new weighting scheme that integrated the strengths of TF-IDF with the semantic information embedded within CluWords. The model's strength was in capturing both syntactic and semantic nuances. A key limitation of the majority of dictionary-based methods was the reliance on manual development, typically for a specific application. This inherent rigidity hindered the scalability and adaptability to novel contexts and evolving use cases [40].

V. Discussion

A. RQ1: What are the main areas of topic modeling implementation based on the user review dataset?

Online customer reviews are a valuable data source for marketing mobile payment services due to perceived affordability, easy accessibility, and dynamic nature [48]. Unstructured reviews prove especially beneficial in comprehending the customer experience in swiftly evolving services, such as mobile banking, as this format mirrors the current preferences of customers [32]. The dominant topic identified in the reviews revolves around perceived usefulness, including factors, such as ease of use, swift processing, and user-friendliness [13].

In the health sector, the utilization of topic modeling extends to the examination of user feedback on mobile applications for diabetes [37] and contact tracing [12]. This process unveils user preferences and identifies aspects of design and functionality that require enhancement. Armed with this information, developers craft impactful applications that positively influence the health of individuals with diabetes [37]. The dominant topic was the efficiency of the system, covering several elements, such as accessibility, service duration, simplicity, and structure. The discoveries of the studies offer crucial insights for both application developers and policymakers, signaling the necessity for specific improvements to substantially enhance the quality of service provided by contact tracing applications [12].

Intelligent systems, such as topic modeling, can be used to analyze transportation user feedback and identify common themes and concerns [39]. An extensively discussed subject is related to trip attributes, particularly the schedule. This subject delves into the precision and dependability of public transit schedules presented within the application. The information can then be used to develop policies and strategies to improve user satisfaction and enhance the travel experience [18]. Topic modeling can be used to identify specific aspects of the service that users find unsatisfactory, such as safety, travel fees, convenience, or system reliability [18], [42]. After identification, targeted interventions can be implemented to address these areas. Integrating topic modeling with sentiment analysis enables a more distinct evaluation of service attributes, distinguishing between areas of strength and those requiring improvement. For example, a study examining user reviews of the ride-hailing application Gojek identified customer satisfaction regarding customer experience, service delivery, and event management. However, areas necessitating enhancement were observed in payment processing, application functionalities, and task management [43].

Topic modeling is a ubiquitous method in daily applications for understanding user concerns, due to its potential to enhance the identification of problems, directions, and satisfaction requirements [35]. During the COVID-19 pandemic, everyday app usage experienced unprecedented growth, as evidenced by the significant increase in food delivery [34] and online grocery market services [36]. This field also included the sports app industry, which was significantly broader and captured areas, such as fitness apps, sports tools, and dedicated social media platforms for sports [41]. Stakeholders can identify the relative importance of determinants of customer satisfaction, as well as determinants that lead to satisfaction and determinants that lead to dissatisfaction, for future enhancement.

TABLE 6
SUMMARY OF TOPIC MODELING METHOD WITH PROS AND CONS

Topic Modeling Methods	Detailed Methods	Pros	Cons
Latent Dirichlet Allocation (LDA)	LDA	Simple, rapid processing speed, robust convergence behavior, and high level of resilience	Challenges with data sparsity could produce incoherent topics, and an explicit number of topics is necessary.
LDA-Based	LDA + keyATM	Generates more coherent topics and can uncover undetected thematic elements.	Heavily depends on a comprehensive selection of representative seed words, derived from the corpus of reviews.
	Hierarchical Dirichlet Process (HDP)	Not require a predetermined number of topics and presents specific keywords that facilitate the identification of each topic's characteristics.	Performed poorly in both coherence and stability evaluations.
	Adaptively Online LDA (AOLDA)	Perform well by adaptively leveraging topics from previous app versions.	The performance of short texts diminishes due to the inherent data sparsity issue.
	Online Biterm Topic Model (OBTM)	Mitigates the problem of data sparsity and manages dynamic data streams effectively.	Textual data from different times or versions may significantly differ from the latest, impacting their relevance to understanding the present.
	Adaptive Online Biterm Sentiment-Topic Model (AOBST)	Excels at identifying and interpreting negative topics through coherent and representative labels.	The model might neglect infrequent emerging issues and changelogs might miss resolved issues, hindering complete understanding.
Non-Negative Matrix Factorization (NMF)	NMF	Exceptional performance concerning topic coherence and accuracy.	Displays instability as a result of the random initialization utilized in the implemented algorithm
Other Techniques	Security-Related Review Miner (SRR-Miner)	Exhibits proficiency in both identifying security-related reviews and synthesizing the underlying security issues and user sentiment with clarity and accuracy.	Requires a hand-crafted lexicon of security-specific keywords to function effectively, necessitating extensive manual effort
	Mobile App Reviews Summarization (MARS)	Effective in uncovering privacy practices, user concerns, and improving data protection.	Dependence on a manually curated privacy threat catalog and expert-driven validation of automatically generated labels
	CluWords	The ability to capture both syntactic and semantic information while employing a flexible weighting scheme by integrating manual semantic insights with automated large-scale embeddings.	Inflexibility due to the manual construction for specific applications limits the ability to scale effectively or adapt to new contexts.

B. RQ2: What is the recent trend of techniques for topic modeling based on user review datasets in the proposed literature?

A synthesized overview of the topic modeling method, including the respective advantages and limitations was presented in this study. LDA, as the most used method, has some advantages, namely fast speed, strong convergence behavior, and good robustness. These advantages made LDA a widely adopted unsupervised method in several fields of machine learning and topic mining [49].

Referring to its diverse implementation, LDA has been used across a broad spectrum of applications, as shown in Table 3, stating its effectiveness and ease of application. More recently developed methodologies, such as SRR-Miner, AOLDA, OBTM, AOBST, and MARS tend to exhibit suitability for a wider array of general applications, rather than being constrained to particular domains. This observation suggests a potential avenue for future scholarly inquiry into the experimental application of these newer methods within more specific implementation contexts. Enhancements, such as LDA-based methods with keyATM and HDP strive to improve topic coherence and automate topic number determination, respectively. These enhancements introduce complexities, such as reliance on seed words or diminished coherence and stability. Furthermore, online adaptations, such as AOLDA and OBTM, along with sentiment-focused models (AOBST), address data sparsity and dynamic data streams. Despite these potentials, these adaptations may grapple with short texts, data relevance over time, and the oversight of emerging issues. NMF excels in topic coherence and accuracy but may exhibit instability due to random initialization. Lastly, specialized methods, such as SRR-Miner, MARS, and CluWords show proficiency in specific fields, including security, privacy, and semantic analysis. These methods have limitations regarding manual effort, reliance on curated catalogs, and inflexibility in adapting to new contexts. Table 6 summarizes the advantages and limitations of each topic modeling method.

Based on the extracted publication year and the applied methods, this section analyzes the development and temporal popularity of various topic modeling implementations within user review datasets. Figure 7 visually shows the shifts in dominant topic modeling methods across the examined years. The application of LDA to user-generated datasets commenced in 2019, despite its first introduction in 2015 [56], and experienced consistent implementation, leading to its peak in 2023. Methods derived from the LDA model showed a trend of increasing adoption starting in 2020, with a significant presence observed in 2022. This suggests a growing increment towards building upon and extending the original LDA framework by potentially addressing inherent limitations within the foundational LDA method. Alternative methodologies categorized as "Other Methods" were developed in the earlier part of the examined period (2019-2020) but are absent in subsequent years, potentially signifying a transition towards the more established or recently developed LDA-centric approaches.

There is a significant increase in the implementation of LDA and LDA-based methods for analyzing user-generated review data that occurred during the 2022-2023 period. This increase may be attributed to the high trend in mobile application usage, potentially influenced by the COVID-19 pandemic. Therefore, studies predominantly focused on examining mobile applications within the health sector using LDA [12], [38], [62]. Furthermore, indirect effects of increased mobile app usage in areas, such as mobile banking [13], [32] and transportation [43], [49] were also analyzed employing LDA and LDA-based methodologies during that period.

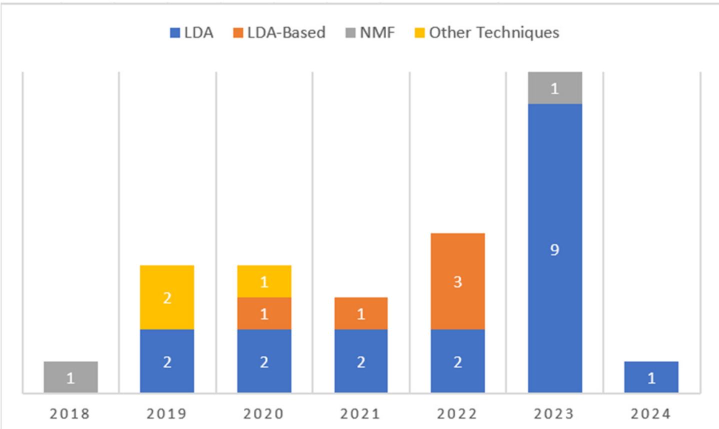

Fig. 7 Topic modeling technique progression over the years

This study acknowledges the subjective nature of the qualitative content analysis. Moreover, there is a need to show the relatively confined scope of the preceding studies, covering the period from 2018 to 2024. This may not fully capture the endeavors related to the analysis of user reviews through topic modeling methods, including newly proposed methodologies for conducting topic modeling. The academic implication of this study originates from enhancing the existing literature by providing an overview and insights into the field of topic modeling implementation, along with introducing novel methods applicable to user review datasets. The study also helps in pinpointing gaps within the current body of knowledge and can serve as a foundational reference for future endeavors. Furthermore, the practical contribution is in its role as a reference for assessing the prevalent topic within comparable app categories, facilitated by the mapping study of dominant topics. Leveraging the comprehensive analysis of various topic modeling methods, as well as the unique strengths and weaknesses documented in this study, practitioners and analysts can select based on the specific application context and objectives.

Drawing from the discoveries, topic modeling has found extensive application across various fields. However, none of the chosen studies delved into the use of topic modeling in e-government applications. The use of topic modeling can serve as a foundational tool to evaluate and comprehend public opinions, whether satisfaction or dissatisfaction, regarding the services offered through mobile applications for public services. Considering the keywords cluster analysis, service quality can be adopted based on the output of the topic modeling. Furthermore, other categories have not been explored, such as insurance and education platform apps. The state-of-the-art LDA-based models are actively developed to address the weaknesses inherent in LDA. The study intends to implement an LDA-based model that demonstrates efficacy within user review datasets for its application in public services applications.

VI. CONCLUSIONS

In conclusion, the exploration of topic modeling has been a dynamic focus of scholarly investigation over the last two decades. This study aimed to identify the prevailing trend of implementing topic modeling across diverse fields and grasp the methodology applicable to user review datasets. The PRISMA protocol led to 28 studies which were selected as the baseline. A bibliometric analysis was carried out to examine the demographics of the primary studies, including publication sources, publication years, and country-based analysis. The overarching observation is that user reviews have been significant as primary data for reviewing diverse applications through topic modeling methods. Digital banking and transportation apps have become the most extensively studied areas. Furthermore, LDA remains the most widely used method, with a concentration on addressing its limitations.

The limitations of this study included the subjective nature of qualitative analysis and its focus on publications from 2018 to 2024. However, the study offered academic value by surveying topic modeling's application to user reviews, pointing out new methods and existing knowledge gaps. From a practical standpoint, it assisted in understanding dominant topics across app categories and in choosing suitable topic modeling methods based on the characteristics. A key result was the under-exploration of topic modeling in e-government, insurance, and education applications. Consequently, future studies should investigate the application of robust LDA-based models to user review data in e-government to measure public sentiment on mobile services, as well as explore other currently neglected app categories.

Author Contributions: *Mustaqim*: Conceptualization, Formal analysis, Methodology, Writing – Original Draft, Writing – Review & Editing. *Suryono*: Formal analysis, Methodology, Writing – Review & Editing.

All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from any funding agency Conflicts

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Informed Consent: There were no human subjects.

Institutional Review Board Statement: Not applicable.

Animal Subjects: There were no animal subjects

ORCID:
Ilham Zharif Mustaqim: https://orcid.org/0009-0003-5085-2260
Ryan Randy Suryono: https://orcid.org/0000-0001-9378-8148

REFERENCES

[1] A. Abdelrazek, Y. Eid, E. Gawish, W. Medhat, and A. Hassan, "Topic modeling algorithms and applications: A survey," *Inf Syst*, vol. 112, p. 102131, 2023, doi: 10.1016/j.is.2022.102131.

[2] A. M. Grisales, S. Robledo, and M. Zuluaga, "Topic Modeling: Perspectives From a Literature Review," *IEEE Access*, vol. 11, no. November 2022, pp. 4066–4078, 2023, doi: 10.1109/ACCESS.2022.3232939.

[3] D. Li, K. Wu, and V. L. C. Lei, "Applying Topic Modeling to Literary Analysis : A Review," *Digital Studies in Language and Literature*, vol. 1, pp. 113–141, 2024, doi: 10.1515/dsll-2024-0010.

[4] B. Ozyurt and M. A. Akcayol, "A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA," *Expert Syst Appl*, vol. 168, p. 114231, 2021, doi: 10.1016/j.eswa.2020.114231.

[5] L. George and P. Sumathy, "An integrated clustering and BERT framework for improved topic modeling," *International Journal of Information Technology (Singapore)*, vol. 15, no. 4, pp. 2187–2195, 2023, doi: 10.1007/s41870-023-01268-w.

[6] N. Genc-Nayebi and A. Abran, "A systematic literature review: Opinion mining studies from mobile app store user reviews," *Journal of Systems and Software*, vol. 125, pp. 207–219, Mar. 2017, doi: 10.1016/J.JSS.2016.11.027.

[7] N. Jha and A. Mahmoud, "Mining non-functional requirements from App store reviews," *Empir Softw Eng*, vol. 24, no. 6, pp. 3659–3695, 2019, doi: 10.1007/s10664-019-09716-7.

[8] J. Dąbrowski, E. Letier, A. Perini, and A. Susi, "Mining and searching app reviews for requirements engineering: Evaluation and replication studies," *Inf Syst*, vol. 114, p. 102181, 2023, doi: 10.1016/j.is.2023.102181.

[9] Y. Chen, Z. Peng, S. H. Kim, and C. W. Choi, "What We Can Do and Cannot Do with Topic Modeling: A Systematic Review," *Commun Methods Meas*, vol. 17, no. 2, pp. 111–130, 2023, doi: 10.1080/19312458.2023.2167965.

[10] O. Grljević, M. Marić, and R. Božić, "Exploring Mobile Application User Experience Through Topic Modeling," *Sustainability (Switzerland)*, vol. 17, no. 3, 2025, doi: 10.3390/su17031109.

[11] S. Mankad, S. Hu, and A. Gopal, "Single stage prediction with embedded topic modeling of online reviews for mobile app management," *Annals of Applied Statistics*, vol. 12, no. 4, pp. 2279–2311, 2018, doi: 10.1214/18-AOAS1152.

[12] R. Wulandari and A. N. Hidayanto, "Measuring contact tracing service quality using sentiment analysis: a case study of PeduliLindungi Indonesia," *Qual Quant*, 2023, doi: 10.1007/s11135-023-01695-8.

[13] L. Çallı, "Exploring mobile banking adoption and service quality features through user-generated content: the application of a topic modeling approach to Google Play Store reviews," *International Journal of Bank Marketing*, vol. 41, no. 2, pp. 428–454, 2023, doi: 10.1108/IJBM-08-2022-0351.

[14] Y. Chen, H. Zhang, R. Liu, Z. Ye, and J. Lin, "Experimental explorations on short text topic mining between LDA and NMF based Schemes," *Knowl Based Syst*, vol. 163, pp. 1–13, Jan. 2019, doi: 10.1016/J.KNOSYS.2018.08.011.

[15] W. Chen, F. Rabhi, W. Liao, and I. Al-Qudah, "Leveraging State-of-the-Art Topic Modeling for News Impact Analysis on Financial Markets: A Comparative Study," *Electronics (Switzerland)*, vol. 12, no. 12, p. 2605, 2023, doi: 10.3390/electronics12122605.

[16] G. Papadia, M. Pacella, M. Perrone, and V. Giliberti, "A Comparison of Different Topic Modeling Methods through a Real Case Study of Italian Customer Care," *Algorithms*, vol. 16, no. 2, pp. 1–19, 2023, doi: 10.3390/a16020094.

[17] J. Dąbrowski, E. Letier, A. Perini, and A. Susi, "Analysing app reviews for software engineering: a systematic literature review," *Empirical Software Engineering*, vol. 27, no. 2, 2022, doi: 10.1007/s10664-021-10065-7.

[18] J. J. C. Aman and J. Smith-Colin, "Application of crowdsourced data to infer user satisfaction with Mobility as a Service (MaaS)," *Transportation Research Interdisciplinary Perspectives*, vol. 15, p. 100672, 2022, doi: https://doi.org/10.1016/j.trip.2022.100672.

[19] O. Haggag, J. Grundy, M. Abdelrazek, and S. Haggag, "A large scale analysis of mHealth app user reviews," *Empirical Software Engineering*, vol. 27, no. 7, p. 196, 2022, doi: 10.1007/s10664-022-10222-6.

[20] C. C. Silva, M. Galster, and F. Gilson, "Topic modeling in software engineering research," *Empirical Software Engineering*, vol. 26, no. 6, p. 120, 2021, doi: 10.1007/s10664-021-10026-0.

[21] C. D. P. Laureate, W. Buntine, and H. Linger, "A systematic review of the use of topic models for short text social media analysis," *Artificial Intelligence Review*, vol. 56, no. 12, pp. 14223–14255, 2023, doi: 10.1007/s10462-023-10471-x.

[22] W. S. Wibowo and S. Yazid, "Unveiling Roadblocks and Mapping Solutions for Blockchain Adoption by Governments: A Systematic Literature Review," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 18, no. June, pp. 547–581, 2023, doi: 10.28945/5186.

[23] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Information and Software Technology*, vol. 55, no. 12, pp. 2049–2075, 2013, doi: 10.1016/j.infsof.2013.07.010.

[24] S. Castillo and P. Grbovic, "The APISSER Methodology for Systematic Literature Reviews in Engineering," *IEEE Access*, vol. 10, pp. 23700–23707, 2022, doi: 10.1109/ACCESS.2022.3148206.

[25] M. A. W. P. Rahmadhan, D. I. Sensuse, R. R. Suryono, and Kautsarina, "Trends and Applications of Gamification in E-Commerce : A Systematic Literature Review," *Journal of Information Systems Engineering and Business Intelligence*, vol. 9, no. 1, pp. 28–37, 2023, doi: 10.20473/jisebi.9.1.28-37.

[26] N. Z. Maharani, S. S. Kurniawan, D. I. Sensuse, I. Eitiveni, E. H. Purwaningsih, and D. S. Hidayat, "Motivations and Potential Solutions in Developing a Knowledge Management System for Organization at Higher Education : A Systematic Literature Review," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 2, pp. 270–289, 2024.

[27] H. Ahmad, M. Ahtazaz, A. Adnan, and N. Mian, "Trends in publishing blockchain surveys : a bibliometric perspective," *International Journal of Information Security*, vol. 22, no. 2, pp. 511–523, 2023, doi: 10.1007/s10207-022-00653-z.

[28] Y. M. Guo *et al.*, "A bibliometric analysis and visualization of blockchain," *Future Generation Computer Systems*, vol. 116, pp. 316–332, 2021, doi: 10.1016/j.future.2020.10.023.

[29] I. Budi and R. R. Suryono, "Application of named entity recognition method for Indonesian datasets : a review," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 969–978, 2023, doi: 10.11591/eei.v12i2.4529.

[30] L. C. Cheng and L. R. Sharmayne, "Analysing Digital Banking Reviews Using Text Mining," in *Proceedings of the 12th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, in ASONAM '20. IEEE Press, 2021, pp. 914–918. doi: 10.1109/ASONAM49781.2020.9381429.

[31] A. P. Darko, D. Liang, Z. Xu, K. Agbodah, and S. Obiora, "A novel multi-attribute decision-making for ranking mobile payment services using online consumer reviews," *Expert Systems with Applications*, vol. 213, no. PC, p. 119262, 2023, doi: 10.1016/j.eswa.2022.119262.

[32] A. Hussain, A. Hannan, and M. Shafiq, "Exploring mobile banking service quality dimensions in Pakistan: a text mining approach," *International Journal of Bank Marketing*, vol. 41, no. 3, pp. 601–618, Jan. 2023, doi: 10.1108/IJBM-08-2022-0379.

[33] G. Shams, M. A. Rehman, S. Samad, and E. L. Oikarinen, "Exploring customer's mobile banking experiences and expectations among generations X, Y and Z," *Journal of Financial Services Marketing*, vol. 25, no. 1–2, pp. 1–13, 2020, doi: 10.1057/s41264-020-00071-z.

[34] M. Tushev, F. Ebrahimi, and A. Mahmoud, "Domain-Specific Analysis of Mobile App Reviews Using Keyword-Assisted Topic Models," in *Proceedings of the 44th International Conference on Software Engineering*, in ICSE '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 762–773. doi: 10.1145/3510003.3510201.

[35] B. A. Çallı and Ç. Ediz, "Top concerns of user experiences in Metaverse games: A text-mining based approach," *Entertainment Computing*, vol. 46, no. February, 2023, doi: 10.1016/j.entcom.2023.100576.

[36] A. Kumar, S. Chakraborty, and P. K. Bala, "Text mining approach to explore determinants of grocery mobile app satisfaction using online customer reviews," *Journal of Retailing and Consumer Services*, vol. 73, p. 103363, 2023, doi: https://doi.org/10.1016/j.jretconser.2023.103363.

[37] C. I. Ossai and N. Wickramasinghe, "Automatic user sentiments extraction from diabetes mobile apps – An evaluation of reviews with machine learning," *Informatics for Health and Social Care*, vol. 48, no. 3, pp. 211–230, Jul. 2023, doi: 10.1080/17538157.2022.2097083.

[38] U. Yaqub, T. Saleem, and S. Zaman, "Analysis of COVID-19 Gov PK app user reviews to determine online privacy concerns of Pakistani citizens," *Global Knowledge, Memory and Communication*, Nov. 2022, doi: 10.1108/GKMC-10-2022-0230.

[39] M. Eiband, S. T. Völkel, D. Buschek, S. Cook, and H. Hussmann, "A Method and Analysis to Elicit User-Reported Problems in Intelligent Everyday Applications," *ACM Trans. Interact. Intell. Syst.*, vol. 10, no. 4, Nov. 2020, doi: 10.1145/3370927.

[40] F. Viegas *et al.*, "CluWords: Exploiting Semantic Word Clustering Representation for Enhanced Topic Modeling," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, in WSDM '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 753–761. doi: 10.1145/3289600.3291032.

[41] B. Chembakottu, H. Li, and F. Khomh, "A large-scale exploratory study of android sports apps in the google play store," *Information and Software Technology*, vol. 164, p. 107321, 2023, doi: 10.1016/j.infsof.2023.107321.

[42] Y. Zhang, X. Shi, Z. Abdul-Hamid, D. Li, X. Zhang, and Z. Shen, "Factors influencing crowdsourcing riders' satisfaction based on online comments on real-time logistics platform," *Transportation Letters*, vol. 15, no. 5, pp. 363–374, May 2023, doi: 10.1080/19427867.2022.2052643.

[43] V. H. Pranatawijaya, N. N. K. Sari, R. A. Rahman, E. Christian, and S. Geges, "Unveiling User Sentiment: Aspect-Based Analysis and Topic Modeling of Ride-Hailing and Google Play App Reviews," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 3, pp. 328–339, 2024, doi: 10.20473/jisebi.10.3.328-339.

[44] R. A. Masrury, Fannisa, and A. Alamsyah, "Analyzing Tourism Mobile Applications Perceived Quality using Sentiment Analysis and Topic Modeling," in *2019 7th International Conference on Information and Communication Technology (ICoICT)*, 2019, pp. 1–6. doi: 10.1109/ICoICT.2019.8835255.

[45] C. Tao, H. Guo, and Z. Huang, "Identifying security issues for mobile applications based on user review summarization," *Information and Software Technology*, vol. 122, p. 106290, 2020, doi: https://doi.org/10.1016/j.infsof.2020.106290.

[46] F. Viegas *et al.*, "Semantically-Enhanced Topic Modeling," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, in CIKM '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 893–902. doi: 10.1145/3269206.3271797.

[47] T. Yang, C. Gao, J. Zang, D. Lo, and M. Lyu, "TOUR: Dynamic Topic and Sentiment Analysis of User Reviews for Assisting App Release," in *Companion Proceedings of the Web Conference 2021*, in WWW '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 708–712. doi: 10.1145/3442442.3458612.

[48] A. P. Darko, D. Liang, Z. Xu, K. Agbodah, and S. Obiora, "A novel multi-attribute decision-making for ranking mobile payment services using online consumer reviews," *Expert Systems with Applications*, vol. 213, p. 119262, 2023, doi: https://doi.org/10.1016/j.eswa.2022.119262.

[49] X. Shi and J. Li, "Key factors for instant logistics platforms to attract and retain couriers: An analysis based on online comments," *Research in Transportation Business & Management*, vol. 50, p. 101031, Oct. 2023, doi: 10.1016/j.rtbm.2023.101031.

[50] J. J. C. Aman, J. Smith-Colin, and W. Zhang, "Listen to E-scooter riders: Mining rider satisfaction factors from app store reviews," *Transportation Research Part D: Transport and Environment*, vol. 95, p. 102856, 2021, doi: 10.1016/j.trd.2021.102856.

[51] S. F. Verkijika and B. N. Neneh, "Standing up for or against: A text-mining study on the recommendation of mobile payment apps," *Journal of Retailing and Consumer Services*, vol. 63, p. 102743, 2021, doi: https://doi.org/10.1016/j.jretconser.2021.102743.

[52] D. R. K. Raja and S. Pushpa, "Diversifying personalized mobile multimedia application recommendations through the Latent Dirichlet Allocation and clustering optimization," *Multimedia Tools and Applications*, vol. 78, no. 17, pp. 24047–24066, 2019, doi: 10.1007/s11042-019-7190-7.

[53] M. A. Hadi and F. H. Fard, "ReviewViz: Assisting Developers Perform Empirical Study on Energy Consumption Related Reviews for Mobile Applications," in *2020 IEEE/ACM 7th International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, 2020, pp. 27–30. doi: 10.1145/3387905.3388605.

[54] C. Gao *et al.*, "Emerging App Issue Identification via Online Joint Sentiment-Topic Tracing," *IEEE Transactions on Software Engineering*, vol. 48, no. 8, pp. 3025–3043, 2022, doi: 10.1109/TSE.2021.3076179.

[55] M. Hatamian, J. Serna, and K. Rannenberg, "Revealing the unrevealed: Mining smartphone users privacy perception on app markets," *Computers & Security*, vol. 83, pp. 332–353, 2019, doi: https://doi.org/10.1016/j.cose.2019.02.010.

[56] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data," *The Art and Science of Analyzing Software Data*, vol. 3, pp. 139–159, 2015, doi: 10.1016/B978-0-12-411519-4.00006-9.

[57] C. Gao, J. Zeng, M. R. Lyu, and I. King, "Online App Review Analysis for Identifying Emerging Issues," *Proceedings - International Conference on Software Engineering*, vol. 2018-Janua, pp. 48–58, 2018, doi: 10.1145/3180155.3180218.

[58] M. A. Hadi and F. H. Fard, "AOBTM: Adaptive Online Biterm Topic Modeling for Version Sensitive Short-texts Analysis," *Proceedings - 2020 IEEE International Conference on Software Maintenance and Evolution, ICSME 2020*, pp. 593–604, 2020, doi: 10.1109/ICSME46990.2020.00062.

[59] X. Cheng, X. Yan, Y. Lan, and J. Guo, "BTM: Topic modeling over short texts," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 2928–2941, 2014, doi: 10.1109/TKDE.2014.2313872.

[60] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15169–15211, 2019, doi: 10.1007/s11042-018-6894-4.

[61] A. Gupta and H. Fatima, "Topic Modeling in Healthcare: A Survey Study," *NeuroQuantology*, vol. 20, no. 11, pp. 6214–6221, 2022, doi: 10.14704/NQ.2022.20.11.NQ66619.

[62] T. N. Saifana and N. Wilantika, "User Perception towards Telemedicine Before and During COVID-19," in *2022 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2022, pp. 123–130. doi: 10.1109/ICACSIS56558.2022.9923475.