Journal of Theoretical and Applied Information Technology
15th July 2025. Vol.103. No.13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EXPLORING ENHANCEMENT IN SENTIMENT ANALYSIS USING DESCRIPTIVE-SEMANTIC TECHNIQUES: A SYSTEMATIC LITERATURE REVIEW

NURAIN BATRISYIA BIDIN[1], KHYRINA AIRIN FARIZA ABU SAMAH[2], NOR AFIRDAUS BINTI ZAINAL ABIDIN[3], LALA SEPTEM RIZA[4]

[1,2,3] Universiti Teknologi MARA Melaka Branch, Jasin Campus Melaka, Malaysia

E-mail: [1]nurainbatrisyiabidin@gmail.com, [2]khyrina783@uitm.edu.my*, [3]afirdaus@uitm.edu.my, [4]lala.s.riza@upi.edu

ABSTRACT

Sentiment analysis (SA) is a critical component of natural language processing (NLP) that enables the extraction of subjective information from large volumes of textual data. However, achieving high accuracy and contextual understanding remains challenging due to the unstructured nature of text and the complexity of linguistic patterns. To address this issue, this study investigates applying descriptive and semantic analysis (DSA) techniques to improve sentiment reliability and contextual accuracy. This study conducts a systematic literature review (SLR) of 28 publications from 2019 to 2024 using the PRISMA methodology, focusing on works indexed in Electrical and Electronics Engineers (IEEE) and Science Direct (SD). The review identifies Exploratory Data Analysis, Term Frequency-Inverse Document Frequency, and Latent Dirichlet Allocation as the most prevalent DSA techniques to enhance SA. Four thematic areas emerged from the analysis: sector, purpose, algorithm, and method employed. The study concludes that while DSA techniques contribute significantly to addressing contextual and semantic limitations in SA, there remains a need for more integrated approaches to handle complex sentiment structures. The findings highlight key challenges, recent advancements, and directions for future research in this evolving field.

Keywords: *Sentiment Analysis, Lexicon, Machine Learning, Descriptive-Semantic Techniques, Systematic Literature Review*

1. INTRODUCTION

Sentiment analysis (SA) is a subfield of Natural Language Processing (NLP) and machine learning (ML) that aids in comprehending the context of any text and identifying the emotions expressed in a sentence [1]. As a classification problem, SA categorizes opinions, attitudes and sentiments into positive or negative polarities [2]. The importance of SA lies in its wide range of applications, from analyzing customer feedback to tracking public opinion, which can provide valuable insights for businesses, governments, and researchers.

The growing reliance on SA stems from human communication's inherent affective states and personality traits [3]. Hybrid models that combine lexicon-based and ML-based approaches are commonly used to enhance the sentiment's accuracy classification. However, these models struggle with contextual ambiguity and polysemy, where words have multiple meanings depending on context [4].

To address these limitations, researchers have introduced Descriptive and Semantic Analysis (DSA) techniques to enhance the understanding of contextual and semantic nuances in textual data. Descriptive analysis summarizes the main features of datasets, such as frequency distributions and statistical measures to infer sentiment polarity [5][6][7]. For instance, the frequency of positive and negative words in product reviews can indicate overall sentiment, while advanced methods like n-grams and part-of-speech tagging uncover deeper relationships among terms [8].

Semantic analysis, in contrast, focuses on the context and conceptual relationships between words to improve interpretation. It recognizes that sentiment meaning varies depending on usage and helps avoid misclassification by capturing the overall tone more accurately [9][10]. Despite these strengths, both descriptive and semantic approaches can face challenges such as dealing with noisy data, domain-specific variations, and inconsistent application in real-world scenarios.

The motivation for this study stems from the lack of a comprehensive review that evaluates the effectiveness of DSA techniques in SA, particularly across diverse applications such as social media. With the growing number of publications in this field, identifying relevant and high-quality studies has become increasingly challenging [11]. Although SA techniques have advanced, inconsistencies in the application of DSA have limited their overall impact on classification accuracy. A unified review is needed to categorize and compare DSA approaches based on their application sectors, algorithms, and methods. Therefore, a structured and Systematic Literature Review (SLR) is needed to examine current trends, identify research gaps, and assess the state of DSA implementation in SA. This study is guided by the following research questions:

1) What are the current trends in the use of DSA techniques for SA?
2) How do different algorithms and methods contribute to the accuracy and contextual understanding in SA?
3) What are the major challenges and gaps in current DSA studies?

This study addresses this need by conducting a SLR using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [12]. By applying clear inclusion and exclusion criteria, the review ensures the selection of relevant and high-quality studies. The SLR provides a structured synthesis of DSA practices, highlights trends, and identifies research gaps. This approach offers a solid foundation for understanding how DSA techniques affect sentiment classification and supports future improvements in SA research.

This paper is divided into five sections such as Introduction, Methods, Results, Discussions and Conclusion, which includes Section 1 for introduction, Section 2 presenting the methods, Section 3 summarizing the literature, Section 4 providing extensive discussions and lastly Section 5 presenting the conclusions as well as research gaps and recommendations for future applications.

2. METHODS

The main objective of the SLR is to explore, synthesize and assess current studies on a particular topic. Classifying relevant papers finds patterns, limitations and research gaps. SLR can also reduce bias and improve dependability by ensuring results are derived from outstanding and reliable studies. According to [13], the SLR process involves six key steps. The process starts with identifying the research question, followed by data sources and search strategy, study selection, inclusion and exclusion criteria, and quality assessment.

This study adopts SLR as its primary research design in accordance with the PRISMA methodology. Researchers widely recognize SLR as a rigorous and transparent approach for synthesizing existing knowledge, identifying research gaps, and guiding future investigations. Similar SLR methodologies have been successfully applied in various disciplines, such as healthcare, education, and business analytics [14][15], demonstrating flexibility across domains and geographic contexts. By applying this structured approach within the SA and DSA techniques field, this study contributes a consolidated and evidence-based perspective tailored to the field of NLP.

Following the PRISMA framework, the study used a SLR approach to identify, analyze and compile previous research on SA implementing DSA techniques. This offers a thorough and reliable overview of SA techniques while emphasizing emerging advances, limitations and future prospects for related domains. Figure 1 shows that the PRISMA framework encompasses four primary stages: 1) identification, 2) screening, 3) eligibility, and 4) data abstraction.

2.1 Identification

In the PRISMA framework, identification is the first phase, where 478 articles were evaluated as part of the literature search process using predetermined selection criteria. This study conducts a comprehensive search across two main primary research databases, Electrical and Electronics Engineers (IEEE) Xplore and Science Direct (SD), consisting of material published between 2019 and 2024. The first identification phase yielded 442 articles from SD and 36 from IEEE. The study selected these two databases over Scopus and Web of Science (WoS), particularly due to their wide range of published papers in computer science, technological advances, and DSA contexts. In addition, SD and IEEE are suitable for exploring developments in DSA techniques to provide information pertinent to this research's technical and methodological components.

In order to ensure a consistent application of uniform criteria throughout all sources, this evaluation was conducted using the filtering tools accessible within SD and IEEE databases. However, due to the large amount of published research, it continues to be quite challenging for researchers to comprehensively evaluate all the accessible studies. Hence, certain selection criteria were established in order to overcome this issue. Selection criteria consist of important review articles and published research that included keywords like *descriptive statistics, data distribution, exploratory data analysis, trend analysis, semantic analysis, topic modeling, named entity recognition, lexical analysis* and *sentiment analysis*. Following that, the study applied Boolean operators like "OR" and "AND" to filter the SLR results based on well-defined search criteria and keywords.

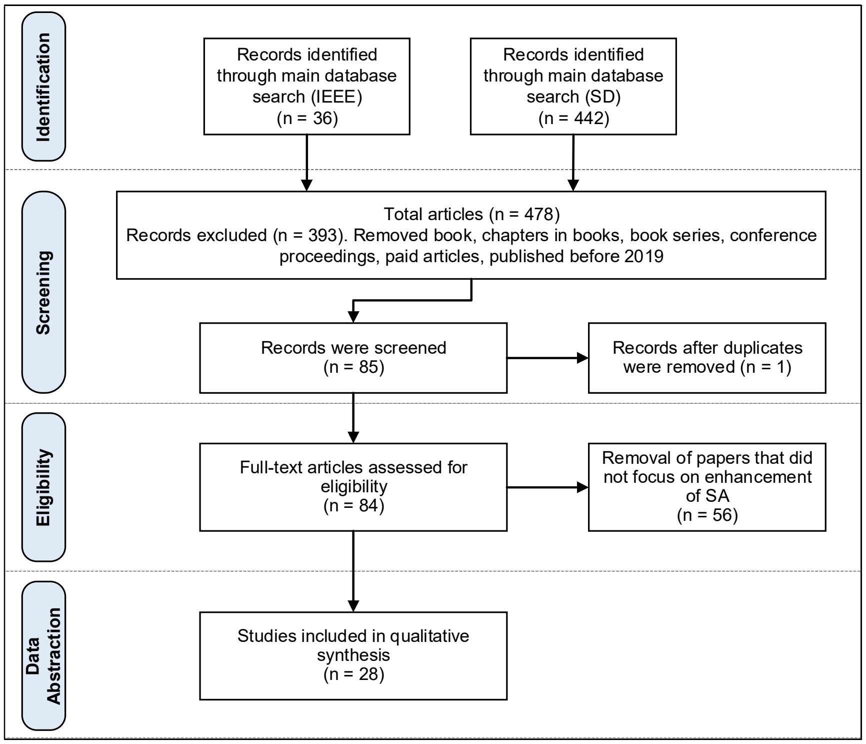

Figure 1: PRISMA Flow Diagram (Online Database's Identification of Studies)

2.2 Screening

The second phase, the screening procedure involving an analysis of search results from the chosen databases, showed a significant rise in research related to the enhancement of SA that began in 2019 and extended until 2024. This led to establishing the 2019–2024 timeframe as an inclusion criterion. Furthermore, in order to preserve the article's quality and reliability, only research articles with theoretical data and those published in credible journals were considered. A total of 393 articles were removed because they failed to satisfy the established inclusion criteria, while 85 articles qualified for further analysis were retained, with one article for duplicate studies.

In other words, applying inclusion and exclusion criteria, evaluating titles and abstracts for relevancy and removing duplicate studies collected from several databases are all part of the screening phase in SLR. These phases optimize the completed dataset's credibility, consistency and comprehensiveness while offering a balanced and fact-based overview of the findings of previous studies.

Hence, Table 1 shows the inclusion and exclusion criteria used in the selection. In order to ensure accessibility and uniformity in the analysis, the study analyzed research articles written in English. It then specifically focused on articles that advanced SA. The study excluded on-research publications, including books, websites, conference proceedings and review articles, to maintain thoroughness and relevance.

Table 1: The Criteria for Inclusion and Exclusion.

Criteria	Inclusion	Exclusion
Publication time	2019-2024	2018 and earlier
Document Type	Journal (research articles)	Other than journal
Language	English	Non-English

Focus of Research	Enhancement in SA	Not related to enhancement in SA
Country	Worldwide	(N/A)

2.3 Eligibility

During the eligibility phase which identified as the third stage of the selection process, 84 filtered articles are thoroughly evaluated corresponding to the predetermined inclusion criteria. Moreover, when relevance remains uncertain, full-text evaluations are conducted, titles and abstracts are reviewed and redundant entries across databases are removed. As a result, 56 articles are excluded due to factors such as duplication, non-English language, lack of focus on primarily on enhancing SA along with classification as review papers rather than primary research. The final selection for analysis consisted of 28 articles that fulfilled the eligibility requirements.

2.4 Data Abstraction

The final phase of the SLR which is data abstraction involves extracting and analyzing relevant data from the selected studies. A total of 28 balanced research papers are reviewed to align with the study's goals and objective as well as relevant topics identification by extracting significant information from the titles, abstracts, and full texts. Themes are then examined to identify patterns and trends, particularly those related to DSA techniques. This process highlights critical issues, as well as similarities and differences across studies. Overall, data abstraction serves as a crucial process in the SLR, enabling the synthesis of findings and the derivation of meaningful insights. A well-executed data abstraction process also contributes to future studies by offering a detailed, unbiased, and evidence-based overview of the current literature.

3. RESULTS

Figure 2 presents the publication trend of 28 research articles on DSA techniques from 2019 to 2024 The number of studies increased steadily over the years, indicating a growing interest in this area. In 2019, only one article was published, followed by three in 2020. A notable increase occurred in 2021 with six publications, highlighting the rising relevance of DSA. Although there was a minor descrease to five publications in 2022, the level of interest remained consistent. The number of publication peaked in 2023 with eight studies, but decreased to five in 2024. This overall trend underscores the the growing focus on implementing DSA techniques in recent research.

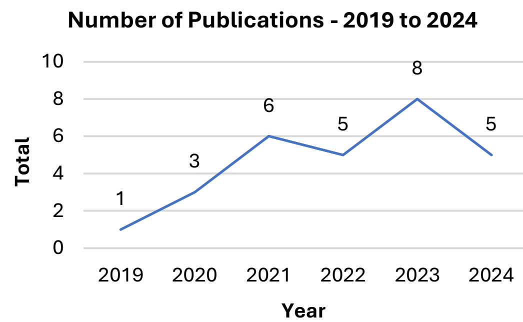

Figure 2: Number of Publications from 2019 to 2024 Related to DSA Studies

Table 2 provides an overview of 28 research studies conducted between 2019 and 2024 across various countries, including the United States of America (USA), India, China, Bangladesh, Indonesia, Pakistan, Canada, France, Italy, the United Kingdom (UK), and South Africa. This table also categorizes the major aspects of DSA into four key areas: 1) sectors, 2) purposes, 3) algorithms and 4) methods. This methodical approach demonstrates the numerous methods and applications of DSA across different domains. The primary objectives include analyzing user behaviour, enhancing sentiment classification, predicting sentiment, and classifying sentiment polarity. Additionally, various algorithms and techniques are applied within DSA to improve the accuracy of SA.

Figure 3 presents a bar chart illustrating the distribution of articles grouped by main themes and subthemes across four categories: sector, purpose, algorithm, and method. Social media emerges as the most researched sector which featured in 15 articles. Within the purpose category, the most frequent objective involves analyzing user behavior for 9 articles. Following that, support vector machine (SVM) stands out as the most commonly used algorithm, appearing in 11 studies. In the methods category, analytical approaches dominate with exploratory data analysis (EDA) identified as the most widely applied technique in 18 articles. Overall, the chart reveals the key focus areas and major trends in DSA studies.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Matrix Table on SLR Research Study From 2019 to 2024

	Authors	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[36]	[37]	[38]	[39]	[40]	[41]	[42]	[43]
	Country	USA	USA	India	China	India	Bangladesh	USA	Indonesia	India	Indonesia	USA	Pakistan	China	Canada	India	France	Indonesia	Indonesia	Italy	UK	India	India	India	India	South Africa	India	India	India
Sector	OC	√	√						√	√																			
	AB	√																											
	SM		√	√	√	√				√			√	√	√	√			√		√	√		√		√	√		
	HL		√		√	√		√					√																√
	PO		√																	√		√							
	AI			√																	√								
	SP		√				√																						
	NW						√	√						√			√			√				√	√				
	BU									√							√						√	√			√		
	FN										√																		
	OP										√							√					√				√	√	
	IN											√												√					
	ED											√																	
Purpose	AUB			√				√				√		√		√			√		√			√			√		
	PS	√				√	√				√		√		√					√								√	
	CSP				√																								√
	RT		√						√																				
	PCT									√							√					√				√			
	ESC																	√					√		√				
Algorithm	XGBC	√																√					√				√		
	LR	√					√						√					√		√			√		√		√	√	
	SVM	√				√	√						√		√					√		√	√		√		√		√
	RF	√					√											√		√		√	√		√		√		
	GB	√																											
	LA		√																										
	NN		√																					√					
	FC		√																			√							
	ANOVA			√																									
	NB					√	√								√							√			√		√	√	
	LSTM					√														√					√				
	DT						√						√							√		√	√		√			√	
	KNN						√													√		√							
	VADER	√				√								√							√								
	BERT				√			√									√				√				√		√		√
	KMC												√																
	KCC															√													
	ROUGE								√																				
	BLEU								√																				
	TRANS																			√									
	RNN																			√					√				
Metho	TFIDF	√			√	√	√	√	√	√			√	√	√		√	√		√	√		√		√		√		

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	TMA	√							√																				
	EDA	√		√				√			√	√	√	√		√		√	√	√		√	√	√		√	√	√	√
	LDA		√	√		√		√				√		√		√		√	√		√			√		√	√	√	√
	TA		√		√				√						√														
	TM		√							√																			
	ABL		√															√											
	CTM			√																								√	
	THA				√	√		√				√															√		
	TSA					√								√							√						√		
	NER									√																			√
	LB										√				√	√	√					√							
	NMF													√							√								√
	CS																√												
	LSA																			√								√	
	BOW																	√			√				√				
	TB							√		√			√	√	√														√
	SMOTE																	√											
	HT																	√	√										

Theme 1: Sector Using DSA	**Theme 2: Purpose of Using DSA**	**Theme 3: Algorithm Used in DSA**	**Theme 4: Methods Used in DSA**
OC = Online Community AB = Abuse SM = Social Media HL = Health PO = Politics AI = Artificial Intelligence SP = Sport NW = News BU = Business FN = Finance OP = Online	AUB = Analyze User Behavior PS = Predict Sentiment CSP = Classify Sentiment Polarity RT = Research Trends PCT = Propose Concepts or Techniques ESC = Enhance Sentiment Classification	XGBC = XGBoost Classifier LR = Logistic Regression SVM = Support Vector Machine RF = Random Forest GB = Gradient Boosting LA = Linear Algebra NN = Neural Network FC = Fuzzy Clustering ANOVA = Analysis of Variance NB = Naïve Bayes LSTM = Long Term Short Memory DT = Decision Tree KNN = K-Nearest Neighbor VADER = Valence Aware Dictionary and Sentiment Reasoner BERT = Bidirectional Encoder Representation from Transformer KMC = K-Means Clustering KCC = K-Centering Clustering ROUGE = Recall-Oriented Understudy for Gisting Evaluation BLEU = Bilingual Evaluation Understudy TRANS = Transformer RNN = Recurrent Neural Network	TFIDF = Term Frequency-Inverse Document Frequency TMA = Topic Modeling Approach EDA = Exploratory Data Analysis LDA = Latent Dirichlet Allocation TA = Trend Analysis TM = Text Mining ABL = Aspect-Based Label CTM = Correlated Topic Models THA = Thematic Analysis TSA = Time Series Analysis NER = Named Entity Recognition LB = Lexicon-based Sentiment Analysis NMF = Non-negative Matrix Factorization CS = Cosine Similarity LSA = Latent Semantic Analysis BOW = Bag-of-Words TB = TextBlob SMOTE = Synthetic Minority Over-Sampling Technique HT = Hyperparameter Tuning

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Platform IN = Interview ED = Education			

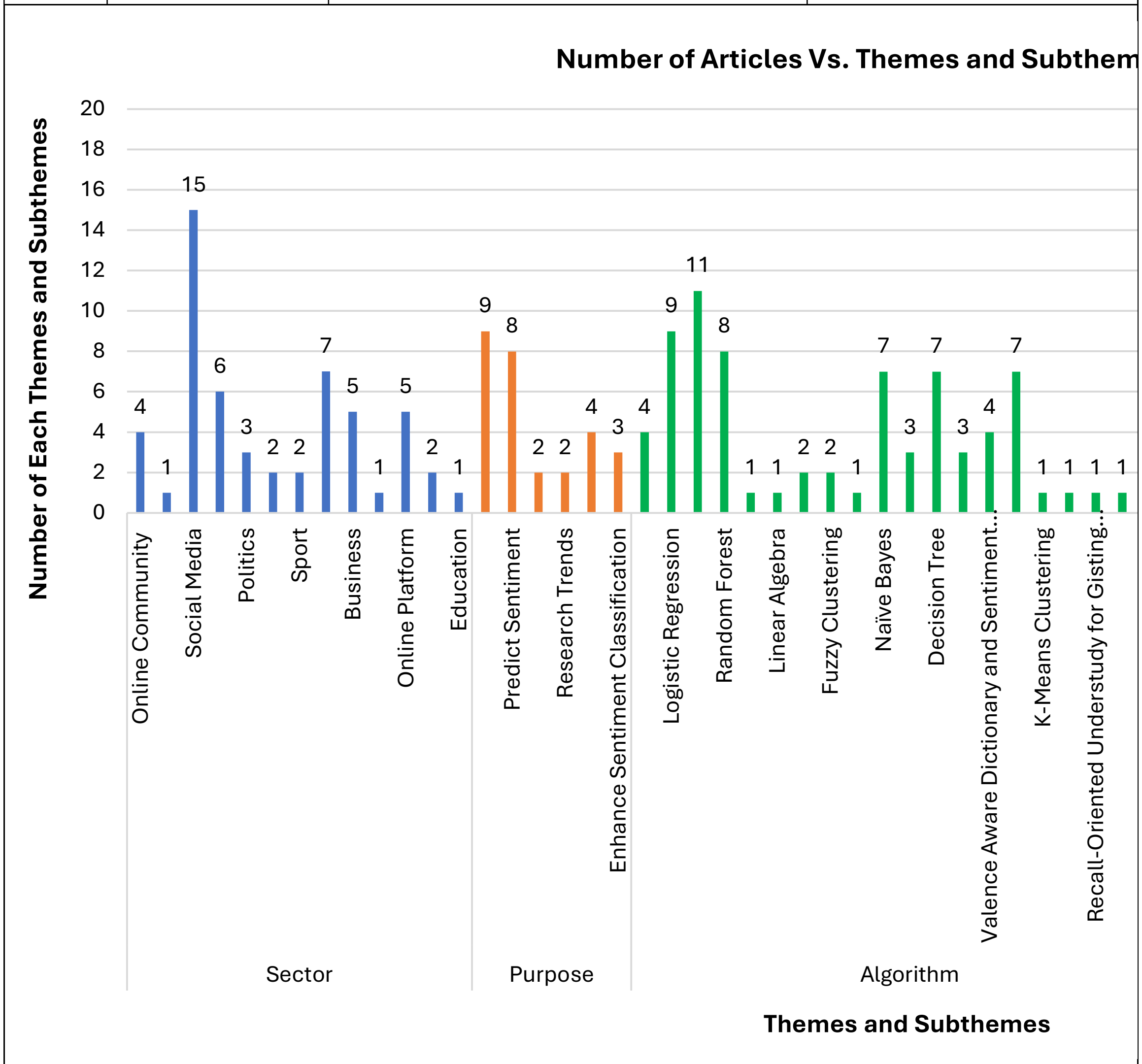

Figure 3: Themes and Subthemes Based on the Articles

Based on a total of 255 subthemes, Figure 4 presents a pie chart illustrating the percentage distribution of research article themes. Method-related themes account for the largest share, comprising 38% (96 subthemes), followed by algorithm themes at 30% (77 subthemes). Sector themes represent 21% (54 subthemes), reflecting the numerous application areas explored in the studies, while purpose-related themes constitute 11% (28 subthemes).

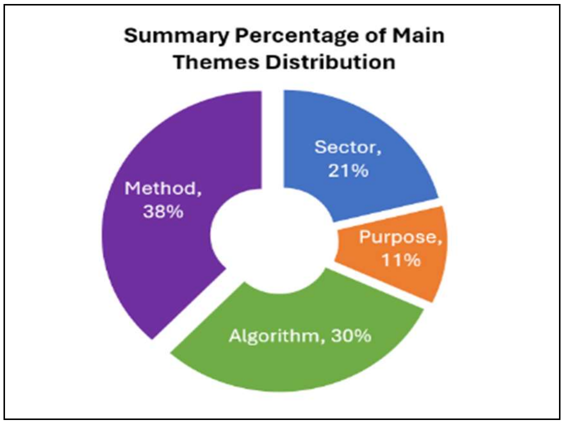

Figure 4: Summary Percentage of Themes Distribution

4. DISCUSSION

The significance of text representation and contextual comprehension in SA remains a critical challenge, particularly as digital communication becomes more informal, sarcastic, and domain-specific. Conventional sentiment classification techniques often rely on shallow keyword matching or static lexicons which fail to handle evolving linguistic trends, ambiguity, sarcasm, and context-dependent meaning. These limitations can lead to misinterpretation of sentiment, especially in complex and high-volume environments such as social media, customer feedback, and opinion mining.

This study emphasizes the need to incorporate DSA techniques as a more robust framework for addressing these challenges. From a theoretical perspective, descriptive analysis enables structured understanding of data distribution and linguistic features, while semantic analysis enhances context-sensitive interpretation through deeper language modeling. Integrating both approaches can offer a more comprehensive and accurate sentiment classification strategy that captures both surface-level and conceptual nuances.

Despite the growing interest in DSA techniques, their application across SA studies remains inconsistent. This inconsistency creates uncertainty regarding their effectiveness and the identification of best practices. Therefore, this study addresses the lack of a unified, evidence-based understanding of how DSA techniques influence sentiment accuracy across various domains, algorithms, and implementations.

To address this, this study sets the following hypothesis: *A systematic synthesis and classification of DSA applications in SA will reveal effective patterns, reduce methodological inconsistencies, and highlight opportunities for future improvements in sentiment modeling.* Furthermore, this study aims to investigate the major factors contributing to the use of DSA techniques by analyzing 28 selected articles, categorized into four thematic areas: 1) sector using DSA, 2) purpose of DSA implementation, 3) algorithms used in DSA, and 4) methods used in DSA. The findings aim to offer both conceptual and empirical insights that support the advancement of SA through more context-aware techniques.

4.1 Sector Using DSA

Table 3 outlines the number of DSA-related studies across various sectors and categorized by publication source: IEEE and SD. These data also are visualized in Figure 5, which highlights the sectoral distribution and comparative contributions from both databases. The figure shows that IEEE dominates in terms of publication volume, particularly in key sectors such as social media, health, business, and online platforms.

Table 3: Number of Studies in IEEE and SD Based on Sector

Sector	IEEE	SD	Total
Social Media	12	3	15
News	3	4	7
Health	5	1	6
Business	4	1	5
Online Platform	4	1	5
Online Community	3	1	4
Politics	3	0	3
Interview	2	0	2
Artificial Intelligence	1	1	2
Sport	1	1	2
Abuse	1	0	1
Finance	1	0	1
Education	1	0	1

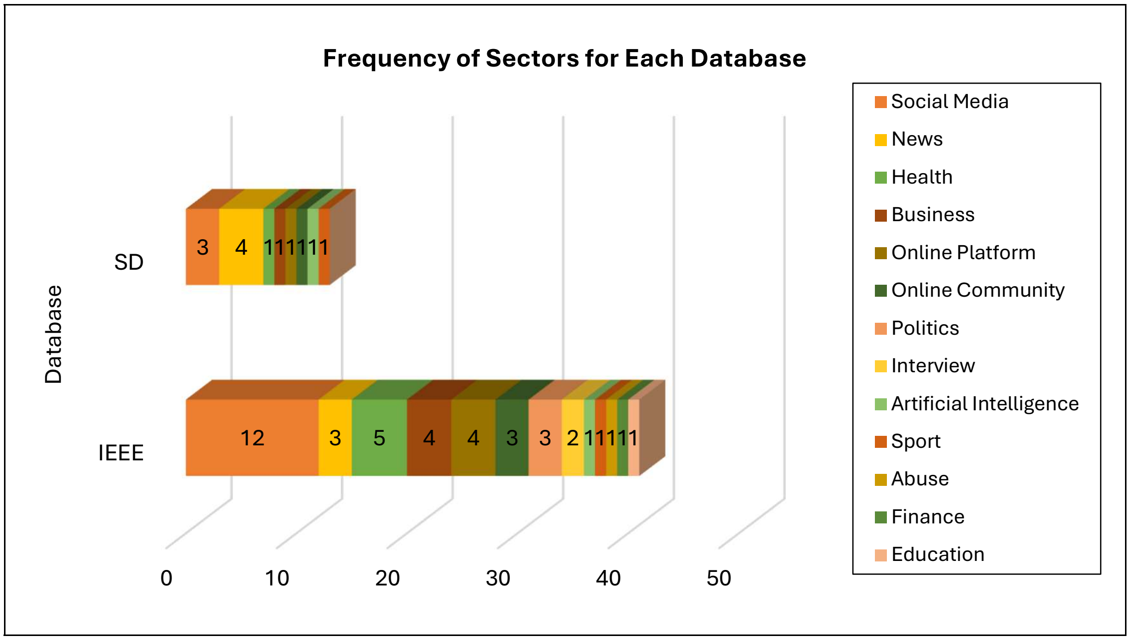

Figure 5: Frequency of Sectors for Each Database of DSA Studies

Social media stands out as the most studied sector, accounting for 15 articles, followed by news (7), health (6) as well as both business and online platforms (5 each). Sectors such as artificial intelligence and sports show equal representation from both IEEE and SD, although with fewer total studies. In contrast, interviews, abuse, finance, and education are among the least represented sectors.

This distribution indicates that DSA research tends to concentrate in domains that produce high volumes of opinion-based textual data, which explains the strong focus on social media, news, and online communities. The limited attention to areas such as education, finance, and abuse indicates potential research gaps and opportunities for broader application of DSA techniques. Furthermore, the higher publication output from IEEE across most sectors reflects its strong emphasis on technical and applied research within the field of SA.

Moreover, Table 4 presents the total number of DSA-related studies conducted across different countries, aligning with patterns observed in previous research. India leads with 11 studies, indicating a strong research focus and growing academic engagement in DSA techniques. The United States and Indonesia follow with four studies each, representing moderate contributions and suggesting the presence of established or emerging research communities in NLP. China contributes two studies, while Bangladesh, Pakistan, Canada, France, Italy, the United Kingdom, and South Africa each contribute one study, showing minimal but notable interest. This distribution suggests that Asia is actively driving DSA research which particularly India, while contributions from Europe and Africa remain limited. The findings may also reflect regional priorities, resource availability, or the integration of SA technologies in academic and industry settings. These disparities highlight opportunities for increased collaboration and research expansion in underrepresented regions.

Table 4: Number of Studies in Each Country in Previous Research

Country	Number of Studies
India	11
United States of America	4
Indonesia	4
China	2
Bangladesh	1
Pakistan	1
Canada	1
France	1
Italy	1
United Kingdom	1
South Africa	1

Figure 6 illustrates the distribution of studies by country in the form of a pie chart. India accounts for the largest share at 39%, followed by the United States and Indonesia, each contributing 14%. China represents around 7%, while other countries make up

Journal of Theoretical and Applied Information Technology
15th July 2025. Vol.103. No.13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

between 3% and 4%. This visual representation highlights the leading contributors and underscores the relatively smaller input from other regions, offering a clearer understanding of the global research landscape.

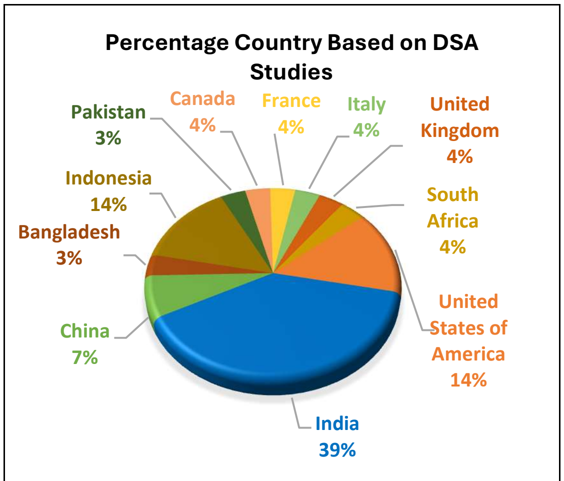

Figure 6: Percentage Country Based on Reviewed DSA Studies

4.2 Purpose of Using DSA

Table 5 outlines the purposes of 28 previous studies included in the SLR framework. Based on the table, the purpose of studies consisting of: 1) analyze user behavior [10-33], 2) predict sentiment [8-34], 3) propose concepts or techniques [16-32], 4) enhance sentiment classification [24-31], 5) classify sentiment polarity [11,35] and lastly 6) research trends [9,15]. Overall, DSA techniques serve multiple purposes within NLP, contributing to enhance sentiment classification accuracy along with the identification of evolving research trends that reflect changes in user perception.

Table 5: List of Previous Studies with Its Purpose

#	Studies	Purpose	Total
1	[18]	Analyze user behavior	9
2	[22]		
3	[26]		
4	[28]		
5	[30]		
6	[33]		
7	[35]		
8	[38]		
9	[41]		
10	[16]	Predict sentiment	8
11	[20]		
12	[21]		
13	[25]		
14	[27]		
15	[29]		
16	[34]		
17	[42]		
18	[24]	Propose concepts or techniques	4
19	[31]		
20	[36]		
21	[40]		
22	[32]	Enhance sentiment classification	3
23	[37]		
24	[39]		
25	[19]	Classify sentiment polarity	2
26	[43]		
27	[17]	Research trends	2
28	[23]		

On the other hand, Figure 7 chart reveals the distribution of DSA study purposes. A total of 32% of the studies focus on analyzing user behavior while 29% concentrate on predicting sentiment. These two are dedicated to understand user interactions and develop more accurate predictive models. Additionally, 14% of studies propose new concepts or techniques which reflect ongoing efforts to refine methodologies. Another 11% aim to improve sentiment classification accuracy. In contrast, only a small proportion of studies explore research trends and sentiment polarity classification. Overall, the findings suggest that DSA research is evolving beyond simple classification, with a growing focus on user behavior analysis and predictive modeling to enhance practical applications.

4.3 Algorithm Used in DSA

Table 6 outlines the various ML, deep learning (DL), lexicon-based (LB) and other analytical algorithms used in DSA studies. For ML algorithms, Support Vector Machine (SVM) is the most commonly used algorithm for dealing with high dimensional data and text classification tasks [3], followed by Logistic Regression (LR), which is an effective algorithm for its interpretability in binary classification issues [6]. Moreover, Random Forest (RF) is an ensemble learning that augments the Decision Tree (DT) approach for classification accuracy [44], while Naïve Bayes (NB) is preferred due to its speed and ease of usage. Other algorithms include XGBoost Classifier, K-Nearest Neighbour (KNN), Neural Networks, Gradient Boosting and Analysis of Variance.

In terms of DL algorithms, Bidirectional Encoder Representations from Transformers (BERT) is a growing use of DL models for enhanced contextual understanding [45]. Long Short-Term Memory (LSTM) and Recurrent Neural Networks

(RNN) are also used in DL algorithms. For lexicon-based algorithms, Value Aware Dictionary and Sentiment Reasoner (VADER) is widely used to determine sentiment through polarity score calculation [46]. Besides, clustering and other algorithms used include Fuzzy Clustering, K-Means Clustering, K-Centering, Linear Algebra, Recall-Oriented Understudy for Gisting Evaluation (ROUGE), Bilingual Evaluation Understudy and Transformer.

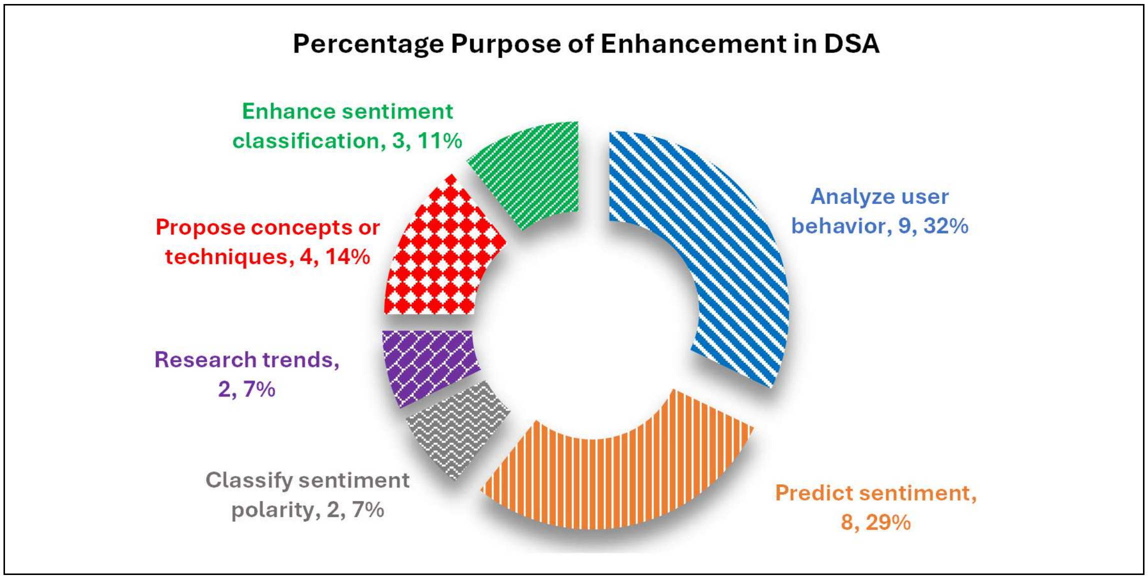

Figure 7: Percentage Purpose of DSA Studies

For instance, the previous study [21] compares the performance of six common ML algorithms, such as SVM, LR, RF, NB, DT, and KNN, to identify Bengali sports news. The experiment results showed that the SVM model outperformed the other algorithms by acquiring the highest weighted F1-score of 97.60%. Another study [37] extensively evaluates Japanese product reviews through various methods and algorithms. The reviews from three separate translators are classified as either positive or negative using five ML algorithms: SVM, LR, RF, DT and XGBoost. This methodical approach enables a comprehensive assessment of both translator and algorithm effectiveness in SA on Japanese product reviews.

Besides that, a previous study [35] analyzes threads on Twitter about privacy and cyber security using two algorithms. Both algorithms for BERT and VADER perform better than conventional ML algorithms, leading to more accurate predictions by comprehending contextual subtleties of language in tweets. Furthermore, another study [20] emphasizes the efficiency of ML algorithms in analyzing sentiments and extracting topics from tweets on the COVID-19 vaccine. SVM, NB, LSTM and VADER are among the algorithms used in this study. As a result, the LSTM model outperformed SVM and NB in sentiment classification, while VADER offered insights into sentiment patterns. The algorithms also uncover important issues in the vaccination debate and classify tweets into positive, negative, and neutral categories. Overall, these studies underscore the growing importance of advanced NLP techniques in capturing nuanced emotions and context, offering deeper insights into public opinion on critical issues.

Table 6: List of Algorithms Used in Each Study of the Previous Works

Studies	Algorithm Used	Total
[16]	Support Vector Machine	11
[20]		
[21]		
[27]		
[29]		
[34]		
[36]		
[37]		
[39]		
[41]		
[43]		
[16]	Logistic Regression	9
[21]		
[27]		

Journal of Theoretical and Applied Information Technology
15th July 2025. Vol.103. No.13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[32]		
[34]		
[37]		
[39]		
[41]		
[42]		
[16]	Random Forest	8
[21]		
[32]		
[34]		
[36]		
[37]		
[39]		
[41]		
[20]	Naïve Bayes	7
[21]		
[29]		
[36]		
[39]		
[41]		
[42]		
[21]	Decision Tree	7
[27]		
[34]		
[36]		
[37]		
[39]		
[42]		
[19]	Bidirectional Encoder Representation from Transformer	7
[22]		
[31]		
[35]		
[39]		
[41]		
[43]		
[16]	Valence Aware Dictionary and Sentiment Reasoner	4
[20]		
[28]		
[35]		
[16]	XGBoost Classifier	4
[32]		
[37]		
[41]		
[20]	Long Term Short Memory	3
[34]		
[39]		
[21]	K-Nearest Neighbor	3
[34]		
[36]		
[17]	Neural Network	2
[38]		
[17]	Fuzzy Clustering	2
[36]		
[34]	Recurrent Neural Network	2
[39]		
[16]	Gradient Boosting	1
[17]	Linear Algebra	1
[18]	Analysis of Variance	1
[27]	K-Means Clustering	1
[30]	K-Centering Clustering	1
[23]	Recall-Oriented Understudy for Gisting Evaluation	1
[23]	Bilingual Evaluation Understudy	1
[34]	Transformer	1

4.4 Methods Used in DSA

Table 7 outlines the methods used within DSA across different studies. Descriptive techniques are necessary for converting unprocessed data into numerical and analytic representations. The two most popular feature extraction techniques in descriptive analysis are Exploratory Data Analysis (EDA) and Term Frequency-Inverse Document Frequency (TF-IDF). Meanwhile, in order to detect changes in sentiment over time, other popular methods include the Bag-of-Words model, Aspect-Based Labeling, Trend Analysis, and Time Series Analysis. In contrast, supporting methods such as Synthetic Minority Over-Sampling Technique (SMOTE), Cosine Similarity, and Hyperparameter Tuning are used to enhance feature diversity and classification performance to ensure the robustness and dependability of descriptive sentiment classification frameworks.

On the other hand, semantic techniques have been employed to comprehend the text's contextual and thematic significance where the most popular technique for revealing hidden themes is Latent Dirichlet Allocation (LDA). Furthermore, additional techniques such as Non-negative Matrix Factorization (NMF) and Latent Semantic Analysis (LSA) retrieve latent features from document-term matrices. In semantic processing, Lexicon-based (LB) techniques and Thematic Analysis (THA) are also used to classify sentiment using predetermined dictionaries and recognizing recurring themes. In the meantime, TextBlob, Named Entity Recognition (NER), and Topic Modeling Approaches are tools used to strengthen classification accuracy and enhance the conceptual value of the analysis.

For instance, a case study [28] employs EDA to analyze Twitter data by examining word frequency, sentiment trends and time-series patterns along with identifying significant terms using TF-IDF and topic modelling using LDA. This analysis provides insights into public sentiment and key discussion topics, particularly in conversations about the COVID-19 pandemic and Biden's presidency. Another study [29] employs lexicon-based and ML techniques to identify trends and perform SA. It classifies tweets as positive or negative and assigns sentiment scores with terms using the NRC Word-Emotion Association Lexicon, which contains

27,000 words with 10 sentiment values. Moreover, this study also used Twitter's Trending API to detect ongoing topics by analyzing tweets in real time before conducting SA.

Furthermore, a study [22] examines COVID-19 to uncover specific themes and changing trends. It conducts EDA on 57,921 abstracts using descriptive features such as word count, sentence count, and character count. TF-IDF is employed to extract relevant terms, while LDA organizes abstracts into topic clusters based on dominant keywords. Moreover, BERT embeddings are then generated for each abstract using a sentence transformer, followed by dimensionality reduction to trace topic trends over time.

In conclusion, these studies demonstrate how DSA techniques have advanced in their ability to derive meaningful insights from textual data. Overall, descriptive techniques provide essential insights, while semantic approaches giving detail by revealing underlying themes and contextual sentiment. When paired together, these methods allow for a more thorough and insightful analysis of essential occurrences and discussions for various sectors. This integration ultimately enhances the capacity of SA to capture the diversity and complexity of human expression.

Table 7: List of Methods Used in Each Study

Studies	Methods Used	Total
[16] [18] [22] [25] [26] [27] [28] [30] [32] [33] [34] [36] [37] [38] [40] [41] [42] [43]	Exploratory Data Analysis	18
[16] [19] [20] [21] [22] [23] [24] [27] [28] [29] [31] [32] [34] [35] [37] [39] [41]	Term Frequency-Inverse Document Frequency	17
[17] [18] [20] [22] [26] [28] [30] [32] [33] [35] [38] [40] [41] [42] [43]	Latent Dirichlet Allocation	15
[22] [24] [27] [28] [29] [43]	TextBlob	6
[19] [20] [22] [26] [41]	Thematic Analysis	5
[25] [29] [30] [31] [36]	Lexicon-based Sentiment Analysis	5
[17] [19] [23] [29]	Trend Analysis	4
[20] [28] [35] [41]	Time Series Analysis	4
[28] [35] [43]	Non-negative Matrix Factorization	3
[32] [35] [39]	Bag-of-Words	3
[16] [23]	Topic Modeling Approach	2
[17] [24]	Text Mining	2
[17]	Aspect-Based Label	2

[32]		
[18] [42]	Correlated Topic Models	2
[24] [43]	Named Entity Recognition	2
[34] [42]	Latent Semantic Analysis	2
[32] [33]	Hyperparameter Tuning	2
[31]	Cosine Similarity	1
[32]	Synthetic Minority Over-Sampling Technique	1

4.5 Outcome from the Systematic Review

The systematic review explores the DSA studies to enhance sentiment classification accuracy and contextual understanding. The research comprehensively evaluates the current literature to identify trends, challenges and enhancements in SA methodology. Moreover, it also covers research studies published between 2019 and 2024, focusing on articles retrieved from the IEEE Xplore and SD databases. This review classifies the research studies into four main themes: sector, purpose, algorithm and method.

Compared to earlier literature reviews that focused on either SA methods and domain-specific applications, this study provides a more structured synthesis by categorizing DSA techniques across four major themes. Unlike previous surveys that often mix descriptive and semantic methods without clear differentiation, this review distinctly analyzes them, offering insights into how their integration influences sentiment classification performance. Moreover, while prior studies tend to overlook the sectoral diversity or are limited to single-language analyses, this review maps global research efforts across sectors and highlights underrepresented domains such as education and finance. This broader scope enables a more inclusive understanding of DSA adoption.

Sectors accounted for 21% of the research focus, with social media, news, health, business and online platforms being the most studied areas. The findings reveal that DSA has been extensively applied across various sectors, with social media emerging as the dominant domain. Among the 28 reviewed studies, 15 (53.6%) focused on social media platforms such as Twitter and YouTube, demonstrating the significance of user-generated content in DSA. These platforms offer large amount of text data, making them ideal for analyzing social movements, brand perception, and public opinion.

Other major sectors identified in DSA studies include online communities, politics, interviews, artificial intelligence, sport, abuse, finance and education. While DSA has been widely applied across domains, some areas remain underexplored, highlighting opportunities for future research. The research findings also demonstrate a stable increase in research interest over the years, peaking in 2023. Geographically, India contributed the highest number of studies (39%), followed by the United States and Indonesia with 14%. This leads to a growing global interest in the implementation of DSA methodologies.

The research studies primarily involved analyzing user behavior, predicting sentiment, proposing concepts and enhancing sentiment classification, which collectively comprised 11% of the focus areas. The research identifies the main objectives behind DSA applications, with user behavior analysis being the most common (32%) in comprehending user preferences. Other applications include predicting sentiment (29%), enhancing sentiment classification models (11%) and proposing new methodologies (14%). Less frequently, studies focused on sentiment polarity classification (7%) and trend analysis (7%), which focused on categorizing sentiment polarity and evaluating the DSA studies trends.

A significant part of the 30% of research focused on algorithms used within DSA, indicating the comprehensive use of ML, DL and LB algorithms. SVM emerged as the most widely used ML algorithm, featured in 11 studies (39%), followed by LR in 9 studies (32%) and RF in 8 studies (28%), which are known for their simplicity and robustness. Among DL algorithms, BERT appeared in 7 studies (25%) and LSTM in 3 studies (11%) as both demonstrating higher contextual understanding in sentiment classification tasks. Additionally, LB models like VADER was used in 4 studies (14%) providing relevance in real-time sentiment classification. These findings reveal that hybrid approaches combining ML, DL and LB methods can provide better sentiment classification accuracy compared to conventional standalone models.

On the other hand, 38% of the research focuses on the methodology, which points out the relevant methodology in DSA studies, resulting in EDA being the most commonly applied method among 18 studies (64%). Following that, TF-IDF is used in 17 studies (61%) to enhance feature extraction and sentiment classification performance. Another widely applied method is LDA, appeared in 15 studies (54%) for topic modeling and uncovering hidden sentiment themes. These results emphasize the value of combining DSA techniques to enhance the analytical depth of SA.

Besides that, TextBlob was employed in 6 studies (21%) and THA in 5 studies (18%) to extract meaningful patterns and sentiment trends. Techniques such as hyperparameter tuning and SMOTE were also used to enhance model accuracy and address class imbalance in sentiment datasets. These diverse methodologies collectively provide more robust and comprehensive SA by addressing the inherent complexities of textual data.

Furthermore, this research emphasizes the effectiveness of combining DSA to address challenges such as complex phrases and contextual ambiguity in sentiment classification. By integrating statistical trends from descriptive analysis with qualitative data theories from semantic analysis for DSA implementations, researchers observed enhancements in both the accuracy and reliability of sentiment models. Additionally, the hybrid approach such as combination of lexicon-based and ML enables a more nuanced understanding of textual data, particularly in context-sensitive domains.

Despite recent developments, significant research gaps remain in the implementation of DSA. Particularly, limited research studies address underrepresented sectors such as education, finance, and sports. Moreover, there is a lack of research integrating multilingual SA with DSA techniques, especially involving low-resource languages contrary to the advancements in semantic modeling. This study also emphasizes the importance of methodological improvements and the development of cross-domain adaptable frameworks to facilitate the scalable and context-aware DSA methods. Consequently, future research should focus on creating more sophisticated models that address these constraints while enhancing overall accuracy and implementation of DSA across diverse domains.

In conclusion, the research demonstrates the evolution and growing impact of DSA techniques in various sectors. For more complex sentiment classification, the research emphasizes the application of hybrid model approaches. Therefore, developing more flexible models, improving sentiment classification in low-resource languages, and extending DSA applications beyond common sectors should be among the primary targets of future research. Overall, these findings align directly with the research objective of identifying trends and evaluating the effectiveness of DSA techniques. By analyzing sector distribution, algorithm use, and methodology patterns, the study aims to classify the current research landscape and identify areas for improvement.

4.6 Gaps and Future Research Directions

Although DSA has made significant progress, several research gaps must be addressed to improve its effectiveness. One major challenge is the lack of multilingual and cross-lingual capabilities, as most models are designed for English-based datasets. This restricts their applicability in low-resource language contexts. Future research should focus on cross-lingual transfer learning techniques to improve adaptability across diverse languages. Addressing this issue can contribute to DSA performing effectively across diverse linguistic contexts, which could be a promising area for future research that assists in developing low-resource languages.

Another critical gap is the lack of domain-specific models in underrepresented sectors such as education, finance, sports and interviews. General-purpose models frequently unable in capturing the distinctive language traits and sentiment drivers encountered in these domains. Future research should focus on developing domain-specific DSA applications would provide for more precise and meaningful insights which enabling better decision-making based on sector-specific data.

Following that, difficulty in handling contextual ambiguity and sarcasm can obscure intended meaning and leads to misclassification in SA [46]. Integrating DL and LB approaches may offer a viable solution to improve sentiment classification accuracy and interpretability. Moreover, LB can adjust polarity scores based on linguistic rules, allowing for improved handling of complex phrases. This approach improves sentiment classification accuracy while maintaining interpretability and offers a robust solution for DSA in diverse textual contexts.

Additionally, there is a lack of comprehensive hybrid models that effectively combine ML, DL and LB methods. For instance, combining methods like transformer-based DL architectures with LB sentiment features enhances classification and demonstrates that external linguistic knowledge can support DL models in making more transparent predictions [47]. Consequently, development of hybrid SA models that leverage the complementary strengths of ML, DL and LB approaches are increasingly required to achieve both improved sentiment classification accuracy and interpretability.

While the study met its goal of mapping out key patterns in DSA applications, it also revealed some limitations. For instance, although several algorithms and methods were identified, there was limited comparative analysis in existing studies about their relative effectiveness across domains. Additionally, the lack of benchmark datasets and evaluation metrics across the studies makes it difficult to objectively determine the superiority of one technique over another. These insights form the basis for proposing a more standardized and comparative framework in future research.

Due to the focus on IEEE and Science Direct (SD)-indexed articles and a publication timeframe limited to 2019–2024, this study may not have captured the most recent research or studies published in other relevant databases such as Scopus or WoS. Additionally, only English-language articles were included, which may have excluded valuable insights from non-English publications. Furthermore, this review did not conduct a quantitative performance benchmarking of algorithms across studies, as such analysis would require access to raw data and standardized evaluation metrics. These aspects are acknowledged as limitations and present opportunities for future research to provide a more comprehensive and comparative evaluation of DSA techniques.

4.7 Research Contribution

Furthermore, this research provides a comprehensive overview of DSA techniques. It conducts the SLR review using the PRISMA framework in order to identify key trends, methodologies and challenges in DSA. This research also evaluates various algorithms and methods used in DSA studies and emphasizes the importance of combining multiple techniques to enhance sentiment classification accuracy and reliability.

In contrast to prior literature that tends to focus narrowly on either ML-based or lexicon-based methods, this study combines descriptive and semantic techniques into a unified framework, allowing a deeper understanding of both statistical and contextual aspects of sentiment. It also differs from earlier works by applying a theme-based classification (sector, purpose, algorithm, method), which offers a more granular view of how DSA is operationalized in practice.

This research also identifies key gaps, including the lack of multilingual approaches, a limited focus on sectors, and the absence of comprehensive hybrid models to improve sentiment classification performance. Overall, this research provides both practical guidance and theoretical insights essential for DSA studies. Additionally, the research contributes to the field by proposing a roadmap for future research, emphasizing the need for more robust, scalable, and domain-specific SA frameworks. By addressing these gaps, this study aims to advance the field of DSA and support more accurate, actionable insights across diverse applications and industries.

5. CONCLUSION

This SLR analyzes the integration of DSA from 2019 to 2024, revealing a growing interest in the field. It implements the PRISMA framework to ensure a transparent selection process and classifies existing approaches into four main themes. The analysis covers that 255 articles with subthemes distributed as follows: sector themes (21%, 54 subthemes), research purposes (11%, 28 subthemes), algorithm themes (30%, 77 subthemes), and method themes (38%, 96 subthemes), with methods being the most frequently analyzed area for improving sentiment classification models.

The research also evaluates the method of DSA, emphasizing a strong reliance on ML algorithms such as SVM and the increasing utilization of DL models like BERT. Feature extraction techniques like TF-IDF and LDA play a critical role in text representation, while LB approaches remain relevant and are often integrated into hybrid models to enhance performance. Overall, the findings suggest that DSA research is evolving toward more advanced and hybridized techniques, although gaps remain in certain sectors, indicating opportunities for future exploration.

Therefore, this research points out the research gaps in current approaches, guiding future research, bridging existing gaps, refining methodologies and exploring innovative techniques using DSA across various sectors. This review extends the current state-of-the-art by providing a multidimensional analysis of DSA techniques, identifying thematic trends and research gaps often missed in earlier studies. By distinguishing between descriptive and semantic contributions to SA, the study proposes a more integrated direction for future research aimed at improving both classification accuracy and interpretability.

FUNDING STATEMENT

The authors would like to thank Universiti Teknologi MARA for supporting the research. The corresponding author can be contacted at *khyrina783@uitm.edu.my.

REFERENCES:

[1] B. Dutta, "Semantic analysis: Working and techniques," *Analytic Steps*, 2021. [Online]. Available: https://www.analyticssteps.com/blogs/semantic-analysis-working-and-techniques. [Accessed: 23-Jul-2021].

[2] F. Aftab *et al.*, "A comprehensive survey on sentiment analysis techniques," *Int. J. Technol.*, vol. 14, no. 6, pp. 1288–1298, 2023.

[3] A. P. Kumar, A. Nayak, and M. S. K, "A statistical approach to evaluate the efficiency and effectiveness of the machine learning algorithms analyzing sentiments," *2019 IEEE Int. Conf. Distrib. Comput. VLSI, Electr. Circuits Robot. Discov. 2019 - Proc.*, pp. 1–6, 2019.

[4] E. Kleppen, "What is sentiment analysis?," *builtin.com*, 2023. [Online]. Available: https://builtin.com/machine-learning/sentiment-analysis. [Accessed: 03-Mar-2023].

[5] M. Wankhade, A. C. S. Rao, and C. Kulkarni, *A survey on sentiment analysis methods, applications, and challenges*, vol. 55, no. 7. Springer Netherlands, 2022.

[6] F. A. Ramadhan and P. H. Gunawan, "Sentiment analysis of 2024 presidential candidates in Indonesia: Statistical descriptive and logistic regression approach," *2023 Int. Conf. Data Sci. Its Appl. ICoDSA 2023*, pp. 327–332, 2023.

[7] P. S. Nunez, M. J. Cobo, C. D. L. H. Pedrosa, J. I. Pelaez, and E. H. Viedma, "Opinion mining, sentiment analysis and emotion understanding in advertising: A bibliometric analysis," *IEEE Access*, vol. 8, no. July, pp. 134563–134576, 2020.

[8] D. N. Dwivedi and G. Mahanty, "Decoding the UK's stance on AI: A deep dive into sentiment and topics in regulations BT - communication and intelligent systems," 2024, pp. 123–140.

[9] V. Ramanathan, H. Al Hajri, and A. Ruth, "Conceptual level semantic sentiment analysis using Twitter data," *2024 Int. Conf. Adv. Data Eng. Intell. Comput. Syst. ADICS 2024*, pp. 1–8, 2024.

[10] J. Li, L. Jiang, G. Huang, and J. Zhang, "A sentiment analysis method for COVID-19 network comments integrated with semantic concept," *Eng. Appl. Artif. Intell.*, vol. 128, no. October 2023, p. 107511, 2024.

[11] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102048, 2024.

[12] A. Trifu, E. Smîdu, D. O. Badea, E. Bulboacă, and V. Haralambie, "Applying the PRISMA method for obtaining systematic reviews of occupational safety issues in literature search," *MATEC Web Conf.*, vol. 354, p. 00052, 2022.

[13] N. A. S. Abdullah and N. I. A. Rusli, "Multilingual sentiment analysis: A systematic literature review," *Pertanika J. Sci. Technol.*, vol. 29, no. 1, pp. 445–470, 2021.

[14] L. Huang, R. Said, H. C. Goh, and Y. Cao, "The residential environment and health and well-being of chinese migrant populations: A systematic review," *Int. J. Environ. Res. Public Health*, vol. 20, no. 4, 2023.

[15] L. Tang and Y. S. Su, "Ethical implications and principles of using Artificial Intelligence models in the classroom: A systematic literature review," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 8, no. 5, pp. 25–36, 2024.

[16] C. Rawat, A. Sarkar, S. Singh, R. Alvarado, and L. Rasberry, "Automatic detection of online abuse and analysis of problematic users in Wikipedia," in *2019 Systems and Information Engineering Design Symposium (SIEDS)*, 2019, pp. 1–6.

[17] A. Karami, M. Lundy, F. Webb, and Y. K. Dwivedi, "Twitter and research: A systematic literature review through text mining," *IEEE Access*, vol. 8, pp. 67698–67717, 2020.

[18] P. Sarin, A. K. Kar, K. Kewat, and P. V. Ilavarasan, "Factors affecting future of work: Insights from social media analytics," *Procedia Comput. Sci.*, vol. 167, pp. 1880–1888, 2020.

[19] T. Wang, K. Lu, K. P. Chow, and Q. Zhu, "COVID-19 sensing: Negative sentiment analysis on social media in China via BERT model," *IEEE Access*, vol. 8, pp. 138162–138169, 2020.

[20] T. Baldha, M. Mungalpara, P. Goradia, and S. Bharti, "Covid-19 vaccine tweets sentiment analysis and topic modelling for public opinion mining," in *2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)*, 2021, pp. 1–6.

[21] A. Barua, O. Sharif, and M. M. Hoque, "Multi-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

class sports news categorization using machine learning techniques: Resource creation and evaluation," *Procedia Comput. Sci.*, vol. 193, pp. 112–121, 2021.

[22] C. Dornick, A. Kumar, S. Seidenberger, E. Seidle, and P. Mukherjee, "Analysis of patterns and trends in COVID-19 research," *Procedia Comput. Sci.*, vol. 185, pp. 302–310, 2021.

[23] G. N. Harywanto, R. Siautama, A. C. I. A., and D. Suhartono, "Extractive hotel review summarization based on TF/IDF and adjective-noun pairing by considering annual sentiment trends," *Procedia Comput. Sci.*, vol. 179, pp. 558–565, 2021.

[24] S. V. S., "Pre-processing techniques of text mining using computational linguistics and Python libraries," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 2021, pp. 879–884.

[25] A. D. Widiantoro, A. Wibowo, and B. Harnadi, "User sentiment analysis in the Fintech OVO review based on the lexicon method," in *2021 Sixth International Conference on Informatics and Computing (ICIC)*, 2021, pp. 1–4.

[26] B. A. Gronneberg and K. E. Nygard, "NLP text analysis on influential factors of persistence for female computer science undergraduates: A comparative study," in *2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, 2022, pp. 72–77.

[27] M. W. Karam, A. Imran, A. Alzahrani, K. M. A. Alheeti, A. Abbas, and A. A. N. Al-Aloosy, "Dramatic increase in fear-related discussion on Twitter during COVID-19: Analysis, topic modeling and tweets classification," in *2022 International Conference on Electrical Engineering and Sustainable Technologies (ICEEST)*, 2022, pp. 1–8.

[28] C.-Y. Lau, M.-C. Yuen, W.-F. Cheng, M.-Y. Lau, W.-F. Chan, and H.-C. Wong, "Intellectual social scanning and analytics platform," in *2022 IEEE 16th International Conference on Big Data Science and Engineering (BigDataSE)*, 2022, pp. 13–20.

[29] A. Motz, E. Ranta, A. S. Calderon, Q. Adam, F. Alzhouri, and D. Ebrahimi, "Live sentiment analysis using multiple machine learning and text processing algorithms," *Procedia Comput. Sci.*, vol. 203, pp. 165–172, 2022.

[30] D. R. Raman, S. Jayalakshmi, K. Arumugam, A. V. A. Raj, D. Balaji, and R. Brightsingh, "Implementation of data analysis and document summarization in social media data using R and Python," in *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2022, pp. 1457–1464.

[31] M. Arslan and C. Cruz, "Leveraging NLP approaches to define and implement text relevance hierarchy framework for business news classification," *Procedia Comput. Sci.*, vol. 225, pp. 317–326, 2023.

[32] M. F. Herqutanto, R. P. Zatari, and R. Sutoyo, "Topic modeling using LDA-based and machine learning for aspect sentiment analysis," in *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, 2023, pp. 142–148.

[33] M. A. Khadija, I. S. D. Jayanti, F. U. Ni'Mah, and H. Kartikasari, "Big data analysis of social media: A case study of citizens comments of integrated dynamic record information system application (SRIKANDI)," in *2023 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2023, pp. 116–121.

[34] S. Mondal, P. Srinivasan, and M. E. Ahammed, "Fake News Detection: A Comparative Study of Machine Learning Techniques," in *2023 International Conference on Computational Intelligence, Networks and Security (ICCINS)*, 2023, pp. 1–6.

[35] N. Pattnaik, S. Li, and J. R. C. Nurse, "Perspectives of non-expert users on cyber security and privacy: An analysis of online discussions on Twitter," *Comput. Secur.*, vol. 125, p. 103008, 2023.

[36] A. Puri, G. Agrawal, A. Dukare, and M. Jawale, "A survey and analysis of textual content based on exploratory data analysis technique and opinion analysis," in *2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, 2023, pp. 1–6.

[37] R. T, S. D. Pradeep, and V. K. R, "Japanese language review mining using translators, word embedding and ML techniques," in *2023 International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS)*, 2023, pp. 171–176.

[38] C. Vipin, S. Harshit, K. L. S. Kumar, K. Vijay, and M. Zaid, "Checking the truthfulness of news channels using NLP techniques," in *2023 International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM)*, 2023, pp. 1–5.

[39] K. S. S. Datta, G. N. Naidu, S. Abhishek, and A. T, "Enhancing veracity: Empirical evaluation of fake news detection techniques," *Procedia Comput. Sci.*, vol. 233, pp. 97–107, 2024.
[40] M. Mbooi, M. R. Rangata, and T. J. Sefara, "Topic modelling of short texts in the health domain using LDA and bard," in *2024 Conference on Information Communications Technology and Society (ICTAS)*, 2024, pp. 82–87.
[41] S. T. Kakde, S. Roychowdhury, A. T. Bhosale, and J. Maiti, "CPAN chart: A novel customer perception analysis system using natural language processing and attribute control charting," *IEEE Trans. Eng. Manag.*, vol. 71, pp. 7609–7622, 2024.
[42] J. Krishnanjaykrishnan, B. Bhattacharjee, M. Pratap, J. K. Yadav, and M. Maiti, "Survival strategies for family-run homestays: Analyzing user reviews through text mining," *Data Sci. Manag.*, 2024.
[43] A. Yenkikar, M. Bali, R. R. Patil, R. Mirajkar, and T. Ara, "Biomedical named entity recognition through spaCy: A visual exploration," *Proc. - 2nd Int. Conf. Adv. Comput. Comput. Technol. InCACCT 2024*, pp. 17–22, 2024.
[44] Q. A. B. K. Zaman, W. N. S. B. W. Yusoff, and Q. B. B. A. Shah, "Sentiment analysis on the place of interest in Malaysia," *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 43, no. 1, pp. 54–65, 2024.
[45] S. Sinha, A. Jayan, and R. Kumar, "An analysis and comparison of deep-learning techniques and hybrid model for sentiment analysis for movie review," *2022 3rd Int. Conf. Emerg. Technol. INCET 2022*, pp. 1–5, 2022.
[46] N. M. Sham and A. Mohamed, "Climate change sentiment analysis using lexicon, machine learning and hybrid approaches," *Sustain.*, vol. 14, no. 8, pp. 1–28, 2022.
[47] M. Polignano, V. Basile, P. Basile, G. Gabrieli, M. Vassallo, and C. Bosco, "A hybrid lexicon-based and neural approach for explainable polarity detection," *Inf. Process. Manag.*, vol. 59, no. 5, 2022.