Decision Analytics Journal 3 (2022) 100073

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

A systematic review of social media-based sentiment analysis: Emerging trends and challenges ☆

Qianwen Ariel Xu [a], Victor Chang [b,*], Chrisina Jayne [a]

[a] School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, UK
[b] Department of Operations and Information Management, Aston Business School, Aston University, Birmingham, UK

ARTICLE INFO

Keywords:
Sentiment analysis
Social media
Challenges
Evaluation metrics
Systematic review

ABSTRACT

In the present information age, a wide and significant variety of social media platforms have been developed and become an important part of modern life. Massive amounts of user-generated data sourced from various social networking platforms also provide new insights for businesses and governments. However, it has become difficult to extract useful information from the vast amount of information effectively. Sentiment analysis provides an automated method of analyzing sentiment, emotion and opinion in written language to address this issue. In the existing literature, a large number of scholars have worked on improving the performance of various sentiment classifiers or applying them to various domains using data from social networking platforms. This paper explores the challenges that scholars have encountered and other potential problems in studying sentiment analysis in social media. It gives insights into the goals of the sentiment analysis task, the implementation process, and the ways in which it is utilized in various application domains. It also provides a comparison of different studies and highlights several challenges related to the datasets, text languages, analysis methods and evaluation metrics. The paper contributes to the research on sentiment analysis and can help practitioners select a suitable methodology for their applications.

Contents

☆ Acknowledgment: Prof. Chang's work is partly supported by VC Research (VCR 0000181).
* Corresponding author.
E-mail addresses: Q.Xu@tees.ac.uk, qianwen.ariel.xu@gmail.com (Q.A. Xu), v.chang1@aston.ac.uk, victorchang.research@gmail.com (V. Chang), C.Jayne@tees.ac.uk (C. Jayne).

https://doi.org/10.1016/j.dajour.2022.100073
Received 28 April 2022; Received in revised form 19 May 2022; Accepted 24 May 2022
Available online 1 June 2022

1. Introduction

1.1. Social media and sentiment analysis

Due to the rapid and enormous development of information technology, social media platforms such as Twitter, Instagram, and Facebook have become an important component of modern life. These platforms have grown extensively and have had a profound impact on people's daily life in recent years. A great number of users use them not just to meet new people and share their lives but also to express their opinions and sentiments related to various products, services and organizations through comments and posts [1,2]. Therefore, an infinite amount of information generated by users is produced [3]. It is critical for individuals, organizations and governments to extract and make use of the meaningful information user-generated information [4–7].

However, the rising popularity and proliferation of internet technologies, particularly social networking platforms, have created numerous challenges for data management [3]. As the volume of data continues to increase, how to extract useful information effectively and efficiently from the vast amount of data has become a critical issue. Researchers consider addressing this issue by utilizing big data approaches in terms of data storage, access, processing efficiency and security (N. C. [8].

Performing effective data analysis is another challenge that has been addressed by inventing automatic ways to extract sentiment or opinions. Sentiment Analysis can be defined as the technique for automatic extraction of opinions, emotions, and sentiments from written language [9]. The primary benefit of sentiment analysis is to effectively identify and classify the attitudes and sentiments (positive, negative, or neutral) of users in texts to ascertain their attitudes toward products, subjects, or services [1,10].

The application of sentiment analysis has covered a wide range of areas, including healthcare [3,11], films [12,13], products [14,15], politics [16], and more. All these applications across different domains prove that sentiment analysis is a valuable tool for gaining useful insights into public opinion on specific topics of interest [11].

In the existing literature, a large number of scholars have worked on improving the performance of various sentiment classifiers or applying them to various domains using data from social networking platforms. However, the variety of data, models and algorithms used in these studies and the wide range of domains involved make it difficult for interested parties to choose the most appropriate one for their research or application. This paper utilizes the Preferred Reporting Items for the Systematic Review and Meta-Analysis (PRISMA) framework to report the findings on reviewing literature related to sentiment analysis in social media. It identifies challenges and other potential problems that scholars have encountered and suggests potential solutions. This review paper can assist practitioners and researchers in addressing questions related to sentiment analysis in a more effective manner and provide a basis for planning future research.

1.2. Contributions

This paper makes several contributions to the literature by studying and evaluating the studies on sentiment analysis approaches and their applications using the systematic literature review method.

First of all, in the discussion section, this paper illustrates other aspects that researchers may encounter, including data collection and pre-processing, algorithms or models, and evaluation metrics.

Due to the popularity of social media platforms, the amount of user-generated data continues to expand exponentially, necessitating the need for increased people, time, and effort to deal with the growing volume and diverse types of data. Monitoring public sentiment in a timely manner, on the other hand, is extremely important for businesses and government organizations at present. However, most studies focus on the accuracy of the models and ignore their practical aspects, such as the execution time. In particular, this review paper argues that it is not sufficient to consider only the effectiveness of an algorithm, but also its versatility and efficiency.

Aside from presenting these challenges, we also present potential solutions for references. By referencing this paper, they may aim their efforts toward resolving these concerns and construct a more effective research framework.

Secondly, this review paper summarizes the existing studies in Table 1, which helps scholars in the construction of methodological aspects based on prior research findings. This table contributes to the determination of extremely significant information related to data, algorithms, and evaluation metrics, and domains. Every aspect included in Table 1 can assist researchers in addressing their practical questions related to sentiment analysis in a more effective manner and provide a basis for planning future research.

Section 2 introduces the systematic literature review protocol followed by this paper. Section 3 presents the general process of sentiment analysis and its applications. Section 4 describes the different types of methods employed for sentiment analysis tasks. Section 5 illustrates the languages of the reviews, comments, or posts analyzed by different sentiment classifiers and Section 6 discusses the challenges and issues relating to datasets, models, and evaluation metrics. Finally, Section 7 concludes this paper and underlines the contributions and limitations.

2. Systematic review protocol

According to Munn et al. [17], through systematic reviews, scholars are able to uncover international evidence, validate present practice, resolve any discrepancies, and identify new practices. Systematic reviews can also help identify and inform future research fields, investigate contradictory results, and guide decision-making. In this study, a systematic literature review was carried out to obtain a better and more comprehensive understanding of the research condition on sentiment analysis in social media.

The review process is divided into five stages (Fig. 1) which consist of the identification of research questions, search strategy, study selection, data extraction, and analysis and discussion [44,45].

2.1. The identification of research questions

Sentiment analysis techniques have been shown to enable individuals, organizations and governments to benefit from the wealth of meaningful information contained in the unstructured data of social media, and there has been a great deal of research devoted to the design of high-performance sentiment classifiers and their applications [1,4–7]. However, the variety of data, models and algorithms used in these studies and the wide range of domains involved make it difficult for interested parties to choose the most appropriate one for their research or application. Furthermore, to conduct future research related to sentiment analysis more helpful in solving real-world problems, it is necessary to analyze and discuss the challenges faced by scholars and identify potential problems. Therefore, to achieve these goals, we identified the following research questions for this study:

Table 1
Taxonomy of sentiment analysis in social media.

Author	Study language	Dataset_name/ source	Dataset_volume	Domain	approach type	Algorithm	Result
Chen et al. [18]	English	IMDB	2000	Movie	Hybrid	SentiWordNet/TF-IDF/word2vec/BERT + PCA/CHI/MI/CFRT + SVM	Bert-CFRT (Accuracy:0.758)
		Amazon product review	2000	Product			BERT-CHI (Accuracy: 0.744)
Agüero-Torales et al. [1]	English	TripAdvisor	33594	Tourism	Lexicon	Dictionary-based approach	-
Lombardo et al. [9]	Italian	Facebook	50000	Health	with other non-SA methods	Naive Bayes, Degree centrality (Social network analysis)	hierarchical classifier (Accuracy: 0.490)
Kastrati et al. [11]	Albanian	Facebook	10742	Health	Machine learning	DNN/CNN/BiLSTM+Attention, BERT, SVM, NB, DT, RF	BiLSTM+Attention (F-score: 0.721)
Khatoon et al. [19]	English	Amazon product review	1000	Product	Lexicon and ML	SVM, ME, FBS, SO-CAL, DIALS	DIALS (F-score: 0.926)
		IMDb	1000	Movie			DIALS (F-score: 0.926)
		Tweets_business	1000	Business			DIALS (F-score: 0.922)
Zheng and Zheng [20]	Chinese	Yelp Reviews full	150000	Movie	Machine learning	BiLSTM-CNN-attention(BRCAN), SVM, BOW,NB, CNN, RNN, LSTM	Att-CNN (Accuracy: 0.683)
		Yelp Reviews Polarity	300000				BRCAN (Accuracy: 0.968)
		Douban Movie ful	12500				BRCAN (Accuracy: 0.751)
		Douban Movie Polarity	100000				BRCAN (Accuracy: 0.963)
Rehman et al. [12]	English	IMDb	40000	Movie	Machine learning	CNN, LSTM, CNN-LSTM	CNN-LSTM (F-score: 0.880)
		Amazon Movie Review	2000				CNN-LSTM (F-score: 0.860)
Es-Sabery et al. [21]	English	sentiment140	1600000	Health	Machine learning	TF-IDF/FastText/word embedding (Word2Vec, GloVe)/Bag-Of-Words/N-grams+ ID3 Decision Tree + Chi-square/Gain Ratio/Information gain/Gini index	Fast-Text+ID3+information gain (Accuracy: 0.865)
		COVID-19_Sentiments	637978				Fast-Text+ID3+information gain (Accuracy: 0.888)
Jain et al. [22]	English	online airline reviews	165682	Tourism	Machine learning	KNN, SVM, MLP, LR, RF,Ensemble + RAkELo/Louvain partitioning	Ensemble+RAkELo partitioning (Accuracy: 0.827)
El-Affendi et al. [23]	Arabic	34 public datasets	275491	Multiple domains	Machine learning	LSTM-CNN-3 Embedding Channels(MPAN)	Tertiary classi cation case (Accuracy: 0.943); Binary classi cation case (Accuracy: 0.956)
Basiri et al. [24]	English	Tweets	1056049.00	Health	Machine learning	CNN, BiGRU, fastText, NBSVM, DistilBERT, and the proposed fusion model.	the proposed fusion model (Accuracy: 0.858; F-score: 0.858)
Salur and Aydin [10]	Turkish	Tweets	17289	Product	Machine learning	LSTM, GRU, BiLSTM, CNN	CNN-BiLSTM (Accuracy: 0.821)
Priyadarshini and Cotton [25]	English	Amazon product review	2880000	Product	Machine learning	KNN, NN, CNN, LSTM, CNN-LSTM, LSTM-CNN, LSTM-CNN-GRID SEARCH	LSTM-CNN-GS (Accuracy: 0.964; F-score: 0.981)
		IMDB	50000	Movie			LSTM-CNN-GS (Accuracy: 0.978; F-score: 0.972)
Shrivastava and Kumar [26]	Hindi	website	8604	Movie	Machine learning	SVM, NB, KNN, DT, CNN, LSTM, GA-GRU	GA-GRU model (Accuracy: 0.880)
Wang et al. [27]	Chinese	Douban Movie 1st	6179857	Movie	Lexicon	Corpus-based Approach	The recommendation model with lexicon sentiment analysis (F-score: 0.771)
Singla et al. [28]	English	sentiment140	1048576	Multiple domains	Machine learning	NB, SVM,Logistic Regression, CNN,BERT, LSTM, LSTM-CNN	LSTM-CNN (Accuracy: 0.856; F-score: 0.877)
Abd et al. [29]	English	IMDB	50000	Movie	Lexicon	Corpus-based Approach	Four experiments by building dictionaries with different sizes (Average Accuracy: 0.760)

(*continued on next page*)

Table 1 (*continued*).

Author	Study language	Dataset_name/ source	Dataset_volume	Domain	approach type	Algorithm	Result
Yan et al. [30]	English	Reddit	45303	Health	Machine learning	Random Forest	predict joy (RMSE: 0.083); sadness (RMSE: 0.068); anger (RMSE: 0.074); fear (RMSE: 0.079)
Ghorbani et al. [31]	English	Movie Reviews (MR) dataset	2000	Movie	Machine learning	CNN-BiLSTM-CNN	CNN-BiLSTM-CNN (Accuracy: 0.890)
Chandra and Krishna [32]	English	Senwave COVID-19 sentiment dataset	10000	Health	Machine learning	LSTM, BiLSTM, BERT	BERT (F1 macro: 0.530; F1 micro: 0.587)
Ibrahim et al. [3]	English	HTC smartphone	22977	Product	Lexicon	HTSM, rCRP, HASM, SSA, HUSTM	HTSM outperforms in terms of execution time
		COVID-19_Sentiments	33175	Health			
Kumar et al. [33]	English	Facebook	900	Book	Lexicon and ML	VADER, NB, ME, SVM,LSTM, CNN	CNN (Accuracy _without gender info: 0.770; Accuracy _female: 0.800; Accuracy _male: 0.700)
Bilro et al. [34]	English	Yelp Dataset	15000	business	Lexicon	Corpus-based Approach	-
Rahman and Islam [7]	English	Tweets	12000	Health	Machine learning	Decision Tree (DT), SVM, LR, Bagging Classifier, Stacking Classifier	Stacking Classifier (F-score: 0.835)
Ricard et al. [35]	English	Instagram	749	Health	Lexicon	Dictionary-based approach	-
Vashishtha and Susan [36]	English	Movie Reviews (MR) dataset	2000	Movie	Lexicon	SentiWordNet + unigram/bigram/trigram + fuzzy techniques + k-means clustering	Proposed system with all n-gram (Accuracy: 0.700; F-score: 0.701)
		IMDB	50000				Proposed system with all n-gram (Accuracy: 0.693; F-score: 0.691)
Zainuddin et al. [37]	English	Hate Crime Twitter Sentiment (HCTS) dataset	1078	hate crimes	Hybrid	SentiwordNet lexicon + PCALSA/RP + SVM/NB/RF/Extreme learning machine+ n gram	Sentiwordnet+PCA+SVM with POS Tags+Unigram features (Accuracy: 0.716; F-score: 0.647)
		Stanford Twitter Sentiment (STS) dataset	353	Multiple domains			Sentiwordnet+PCA+SVM with POS Tags (Accuracy: 0.766; F-score: 0.760)
		Sanders Twitter Corpus (STC) dataset	1091	business			Sentiwordnet+PCA+RF+POS Tags+Unigram (Accuracy: 0.742; F-score: 0.740)
Grljevic et al. [6]	Serbian	'Oceni Profesora'	3863	Education	Hybrid	Corpus-based Approach, corpus-NB/SVM/kNN	document level_positive: dictionary-based (F-score: 0.898); document level_negative: dictionary-based (F-score: 0.889) ; sentence level_positive: dictionary-based (F-score: 0.861); sentence level_negative: SVM (F-score: 0.861)
Jaidka et al. [16]	English	Tweets	3400000	Political	with other non-SA methods	SentiStrength lexicon, NB, Social network analysis	NB
Pai and Liu [38]	English	Tweets	6000000	Product	with other non-SA methods	SentiStrength lexicon, regression	-

(*continued on next page*)

Table 1 (*continued*).

Author	Study language	Dataset_name/source	Dataset_volume	Domain	approach type	Algorithm	Result
Dashtipour et al. [13]	Persian	Persian movie review (websites)	2010	Movie	Hybrid	LR, SVM, and multilayer perceptron (MLP) classifiers, CNN, LSTM, BiLSTM	Stacked-BiLSTM (Accuracy: 0.956; F-score: 0.960)
		Persian hotel review (websites)	–	Tourism			2D-CNN (Accuracy: 0.898; F-score: 0.890)
Daeli and Adiwijaya [39]	English	Polarity dataset v2.0	2000	Movie	Machine learning	KNN, NB, SVM, RF	KNN_k=3 + information gain (Accuracy: 0.968)
Thomas and Latha [40]	Malayalam	social media platforms	10000	Multiple domains	Machine learning	RNN-LSTM	RNN-LSTM (Accuracy: 0.815)
Chang et al. [41]	English	TripAdvisor	634277	Hotel	Machine learning	NB, SVR, SVM, LST, sentiment sensitive tree (SST)+convolution tree kernel (CTK) classification	
Al-Bakri et al. [2]	Iraqi	Facebook	14200	Tourism	Machine learning	Rough Set Theory (RST), Naïve Bayes (NB), K-Nearest Neighbors	SST+CTK (Accuracy: 0.893; F-score: 0.935)
Martin et al. [5]	English	Booking & TripAdvisor	12425	Tourism	Lexicon and ML	Corpus-based Approach, CNN, LSTM	embedding-LSTM (Accuracy: 0.892, training time:10,630); CNN-LSTM:(Accuracy: 0.883, training time:1602)
Dang et al. [15]	English	Multimodal Album Reviews Dataset (MARD)	263525	Music	Machine learning	DNN, LSTM, CNN	Hybrid models
		Amazon Movie Review	203967	Movie			Hybrid models
Kausar et al. [14]	English	Office product reviews	30842	Multiple domains	Lexicon and ML	Sentiword Net, NB, DT, RF, SVM, gradient boosting, Seq2Seq	DT, RF, SVM, gradient boosting classifier with Comparative adverbs (F-score: 0.960)
		Musical DVD reviews					
Jnoub et al. [42]	English	IMDB	50000	Movie	Machine learning	Proposed-CNN, Proposed SNN, SVM, KNN, NB, RF	Proposed SNN (Accuracy: 0.870; F-score: 0.870)
		Movie Reviews (Cornel)	10000	Movie			Proposed SNN (Accuracy: 0.820; F-score: 0.820)
		Amazon product review	1000	Product			Proposed SNN (Accuracy: 0.740; F-score: 0.740)
Ibrahim et al. [43]	English	imdb	–	Movie	Machine learning	LSTM-User Movie attention	LSTM-User Movie attention (Accuracy: 0.533)
		yelp13	–				LSTM-User Movie attention (Accuracy: 0.650)
		yelp14	–				LSTM-User Movie attention (Accuracy: 0.667)

(1) Which sentiment classifiers, datasets and evaluation metrics are most used when studying social networking media data in different domains?
(2) What are the challenges faced by authors when performing sentiment analysis on social media datasets?
(3) What are the potential issues in the existing literature on sentiment analysis of social media?

2.2. Search strategy

The literature search was conducted by following the PRISMA framework (Fig. 1), which includes the identification phase, screening phase and inclusion phase [46].

In the identification phase, Scopus and Web of Science (WoS), we searched two scientific digital databases for relevant articles. The publication period was restricted to between 2018 and 2021. Search terms used to acquire articles can be classified into three groups: sentiment analysis/opinion mining, comment/review, and social media. The Boolean operators (i.e., AND, OR) were employed on the search terms to make the inquiries. For example, the keywords "sentiment analysis" and "opinion mining" were combined by the OR logic, while "sentiment analysis" and "social media" were combined using the AND logic. In addition, the search indices were limited to the meta-data (i.e. Title, Abstract, and Keywords).

2.3. Study selection

In the identification phase, 128 records were identified from Scopus and 344 records were identified from the web of science, with a total number is 472 and 76 records duplicated. Therefore, the duplicates were removed, resulting in 396 records that would be processed in

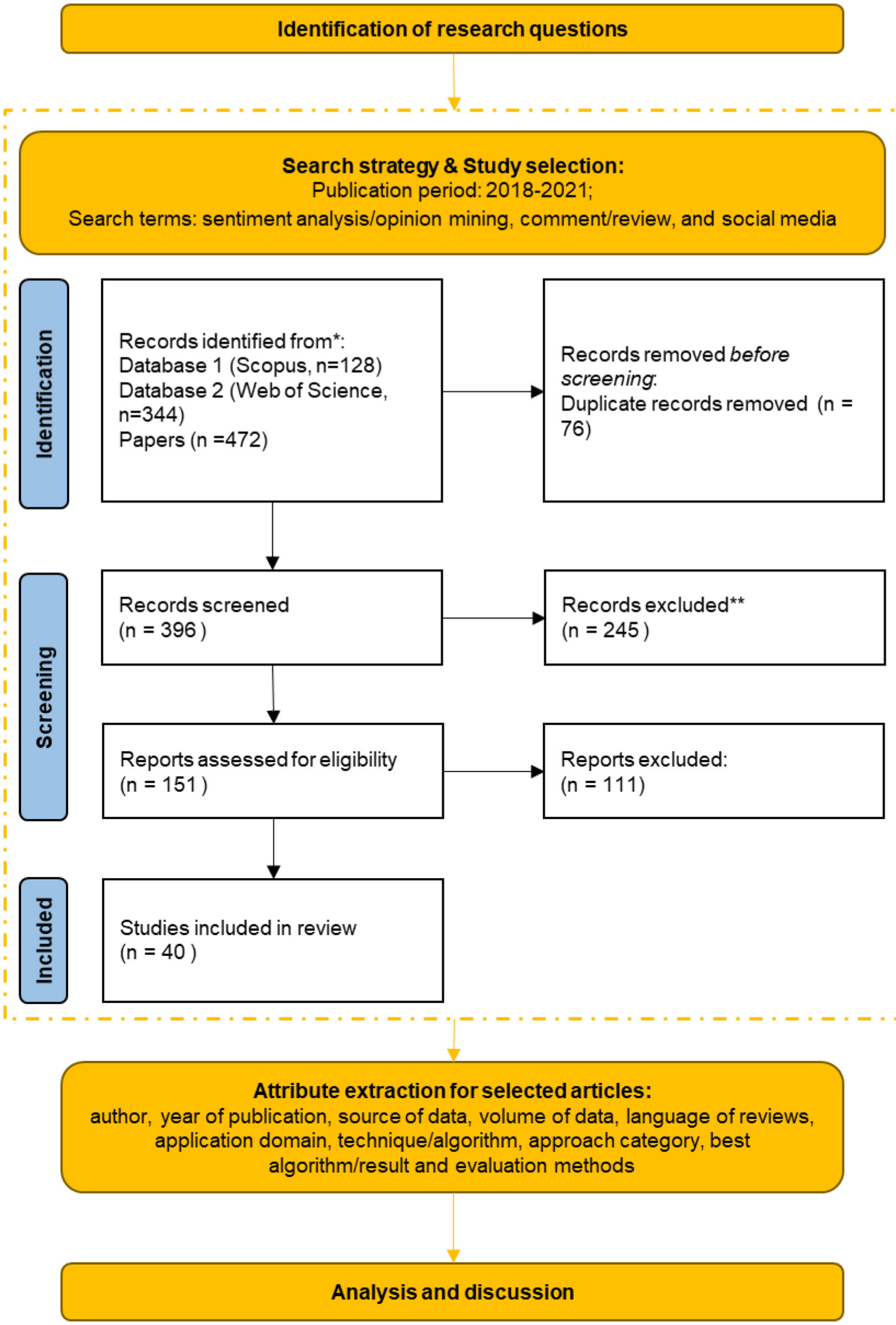

Fig. 1. The review protocol.

the next phase. The screening phase can be divided into two sub-phases, the title and abstract screening phase and the full-text screening phase. Each of the remained 396 records was firstly screened by reading its title and abstract so that whether it met the exclusion criteria could be determined. 151 remained in this sub-phase for the full-text screening phase. During the title and abstract screening phase and full-text screening phase, the following exclusion criteria were applied:

Social media. Records that concern the information distributed on independent social media platforms, not on corporate official websites, or e-news.

Methodology. Records that employ a qualitative method, or in which the sentiment analysis is not the research center, but a part must be conducted to study other research topics, were excluded.

Original studies. Review papers and editorials were excluded.

Dataset. Records that use the dataset in the form of photos, images, or audio, instead of texts (reviews or comments) were excluded.

Language. Records that were not written in English were excluded.

As a result, 40 articles were finally included in the inclusion phase of our review for further and detailed investigation.

2.4. Summary information for the selected articles

This stage aims to generate a piece of summary information from the articles in the inclusion phase [45]. The extracted information can identify research gaps and eventually help us to define research questions. We collected and extracted attributes from each article relating to authors, year of publication, sources of data, the volume of data, languages of reviews, application domain, technique/algorithm applied, approach category, best algorithm/result and evaluation metrics, which are considered to be the most significant perspectives for obtaining the major discussion points of systematic review [44]. The attributes were then listed and grouped into corresponding categories in a summary table (Table 1).

Each attribute was selected for a specific reason. The author and year of publication were indicated by citations for future referencing and for browsing by any interested readers. The next four attributes, namely sources of data, the volume of data, languages of reviews, and application domain, were extracted to discuss the data selection strategies used by scholars for sentiment analysis in different domains, for example, whether they chose to collect directly from social networking platforms or use data that had already been studied by previous researchers, as well as the amount of data collected. This discussion provides a guide for interested readers to select the most suitable data selection strategy for their research. In addition, the discussion of attributes, i.e., application domains, techniques/algorithms applied, approach category, best algorithm/result, evaluation metrics, provides a better understanding of where sentiment analysis can be applied, and what techniques scholars have used to implement these applications or enhance their usefulness and effectiveness as a reference for future research on these topics.

3. Sentiment analysis

3.1. The process of sentiment analysis

In this section, the general process of sentiment analysis is introduced. The specific steps scholars take to implement sentiment analysis vary depending on the research objectives or application requirements. However, all of them follow a general process, as shown in Fig. 2.

The first step is Data Collection. In this step, authors who study the sentiment analysis in low-resource language collect primary datasets from a variety of social media platforms, such as Twitter [3,10], Facebook [2], Reddit [30], TripAdvisor [1]. Low-resource languages are languages with none or very few data resources for machine learning algorithms [47]. Based on their research objectives, they search and acquire the datasets that meet their requirements by using particular search terms. Other researchers prefer to use secondary datasets, which are largely acknowledged and have played a significant role in a number of studies. They are often used to test the performance of models. The most used secondary datasets include IMDb (50, 000 reviews) [42], Amazon Product Review (2000 reviews) [18,19], and Yelp Dataset [34].

The second step is Data Preprocessing. In this step, the data samples are firstly cleaned by several procedures, such as converting the upper case to lower case, removing unnecessary and multiple pronunciations, removing hashtags and URLs., text correction, and removing stop words [28,29,48]. Subsequently, tokenization is conducted by extracting the words from the review text. When using machine learning-based approaches, the tokens need to be encoded to integer values [12,49].

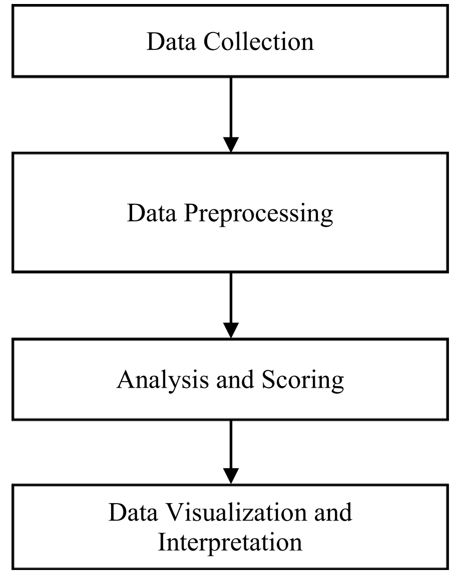

Fig. 2. The general process of sentiment analysis.

The third step is Analysis and Scoring. In this step, researchers employ different methods or models to process the dataset pre-processed in the second step. Based on the level of analysis, the analysis approaches can be divided into a document-level analysis [50], sentence-level analysis [51], phrase-level analysis [28], aspect-level analysis [52], and emotion-level analysis. From the technical perspective, the methods of sentiment analysis, there are three approaches to address the problem of sentiment analysis, including lexicon-based techniques [53], machine-learning-based techniques [54], and hybrid techniques [15].

The fourth step is Data Visualization and Interpretation. In this step, researchers visualize their results in eye-catching charts to gain insights that are easy to understand. There are many different types of charts and graphs, each expressing different information, such as word clouds showing the most frequently occurring words [3], heatmap analysis of type of aspects [41], as well as histogram plot [5], visual effect map [20] and confusion matrix [10] presenting the performance of the classifiers.

3.2. Application of sentiment analysis in different areas

Applications of sentiment analysis have covered a wide range of areas, including healthcare, film, products, travel, and politics, bringing a variety of benefits to businesses, governments, and even individual users.

First of all, sentiment analysis is a promising tool for governments and authorities to monitor online public opinion for better risk management and response, especially in the case of emergencies or major events. The COVID-19 pandemic, for example, has been affecting the lives of people around the world for more than two years now and has been mentioned on a large scale on social media. Posts, comments and tweets about the coronavirus have increased dramatically in a short space of time, reflecting people's thoughts and views about the pandemic [3,11]. A number of studies have used different techniques to apply sentiment analysis to show people's opinions about COVID-19 [3,7,21,24,30,32]. Additionally, recent research indicates that social media has become a major source of false information, particularly during the global pandemic crisis. For governments and public health authorities, learning about the information circulating among the public, as well as public opinion and sentiment through sentiment analysis, is very useful in clarifying rumors in a timely manner to appease the community, develop appropriate strategies and take timely action [4].

Secondly, the application of sentiment analysis in business has been extensively researched. For businesses, sentiment analysis helps determine how customers perceive their products, services, and even the business by analyzing social media data. Moreover, online reviews with ratings and textual information on a variety of aspects provide potential customers with vital information about the user experience of a product or service. Numerous businesses have begun to see the value of sentiment analysis and user ratings to improve existing goods and develop new ones [22].

Kausar et al. [14] proposed a sentiment polarity categorization technique for online product reviews with five sentiment classes, including strongly negative, negative, neutral, positive and strongly positive, and three polarity features, verb, adverb, and adjective. Rehman et al. [12] proposed a hybrid CNN-LSTM model to improve the performance of sentiment analysis on movie reviews. Dashtipour et al. [13] set their research target at movie reviews in Persian, which is the official language of Iran and Afghanistan. They compared the performance of shallow learning algorithms and deep learning algorithms, and the results demonstrate that the stacked-BiLSTM model outperformed all other methods.

Martin et al. [5] constructed eleven different deep learning models to deal with the issue of classifying the sentiments of online tourists' comments that the potential new tourists use to plan their trips. They collected their data from the websites of Booking[1] and Tripadvisor[2] and all the eleven models achieved an accuracy exceeding 87%. Analyzing sentiments from online posts or comments can also be employed to assess companies. Agüero-Torales et al. [1] developed an integrated software for the large-scale analysis of social media data regarding the restaurants in the Granada Province of Spain. Based on sentiment analysis, the restaurants can find what customers think about products and services and how they can improve their image, products and services.

Thirdly, individuals can also benefit from techniques of sentiment analysis to receive a personalized service. In recent years, sentiment analysis has been employed to improve the performance of recommendation systems for different services or products. Dang et al. [15] suggest approaches to improve the performance of recommender systems for streaming services by using sentiment analysis for a better understanding of user preferences. They tested and evaluated two different combinations of LSTM and CNN, LSTM-CNN and CNN-LSTM, on the Multimodal Album Reviews Dataset (MARD) and Amazon Movie Reviews. The baseline they used is the version of the recommendation system that does not include sentiment analysis and genres. The results show that their models outperform in the prediction of ratings and the evaluation of the top recommendation lists compared to the baseline. Wang et al. [27] also integrated sentiment analysis into their mobile movie recommender system and tested it on movie datasets in Chinese. Jain et al. [22] proposed a service recommendation system based on a multi-label ensemble classifier that predicts consumer recommendations in tourism. They collected online airline review data from an online platform and developed an ensemble classifier to evaluate the sentiments of the users.

4. Taxonomy

In this section, all the publications included in this review are classified into five categories according to the methods used for the social media data analysis: lexicon-based approaches, machine learning (ML) approaches, hybrid approaches that integrate the previous two categories, comparative approaches (Lexicon and ML) and approaches that combine the sentiment analysis with other analysis methods such as regression and social network analysis.

[1] http://booking.com

[2] http://tripadvisor.com

4.1. Lexicon-based approaches

According to Fig. 3, 17% of publications adopted lexicon-based approaches, 10% integrated the lexicon-based approaches and machine learning approaches, and 10% of them compared these two types of approaches. Lexicon-based approaches were the initial methods employed for sentiment analysis and include two types: dictionary-based approach and corpus-based approach. Dictionary-based sentiment classification relies on predefined dictionaries, such as WordNet and SentiWordNet [55]. However, corpus-based sentiment analysis is performed based on a statistical analysis of the documents' content, rather than predefined dictionaries [8].

Bilro et al. [34] applied text mining and both corpus and dictionary-based sentiment analysis to identify the sentimental drivers of online customer engagement by investigating the related concepts in online customer reviews (involvement, emotional states, experience and brand advocacy). They created their customer engagement dictionary based on the Yelp Dataset and then extended it based on WordNet 2.1. 59 topics were identified by MeaningCloud and 10 of them were selected for further investigation. Their sentiment results indicated that the cognitive processing dimension of engagement and hedonic experience might significantly influence consumers' review endeavors. Additionally, customers appear to be more interested in favorably promoting a company/brand than negatively.

In the study of Grljevic et al. [6], the lexicon-based sentiment study was applied to the area of higher education. They generated their HiEd-Sent corpus from a Serbian website containing the countries' teaching staff profiles. The spelling and grammar errors in the corpus were corrected by an online tool, Hascheck, and the ones omitted by the tool were manually corrected. To build sentiment dictionaries, a hierarchical annotation procedure for the aspects, sentiment polarity and intensity was conducted by four independent annotators, and the interrater agreement was evaluated. Five dictionaries that contain positive sentiment expressions, negative sentiment expressions, intensifiers, neutralizers and inverters were developed based on the HiEd-Sent corpus. In order to carry out the sentiment detection, they conducted a dictionary-based approach and a machine learning approach. Each sentence was first pre-processed and converted to tokens in the dictionary-based sentiment analysis. Then, the tokens were matched with words or phrases from the dictionaries and assigned corresponding labels and sentiment scores. Finally, the sentiment score of each sentence can be computed by cumulating the scores of all tokens in the sentence, which determines its polarity. The authors analyzed the performance of various combinations of dictionaries at the document and sentence classification levels, and the combination of the five dictionaries together achieved the best result.

Abd et al. [29] employed a phrase-level sentiment analysis using the corpus-based approach and tested their model on the IMDb dataset. Based on the training set, they built their sentiment dictionary by separating the lexicon into positive and negative words. In order to achieve the best result for their model, they conducted four experiments by building different sentiment dictionaries of 25,000, 30,000, 34,000, and 40,000 review terms and used them to classify the test set. On The IMDb dataset, although the second experiment achieved the best accuracy result of 76.585%, the variance between the four experiments was not significant, around 76%.

4.2. Machine learning-based approaches

A number of machine learning approaches have been utilized to deal with the task of classifying people's sentiments. Different from lexicon-based approaches, sentiment analysis based on machine learning approaches trains the models with sentimental features in the text to enable them to detect sentiments automatically. According to Fig. 3, 75% of included publications employed machine learning approaches (55% used only machine learning approaches, 10% used hybrid, and 10% compared the two types of approach).

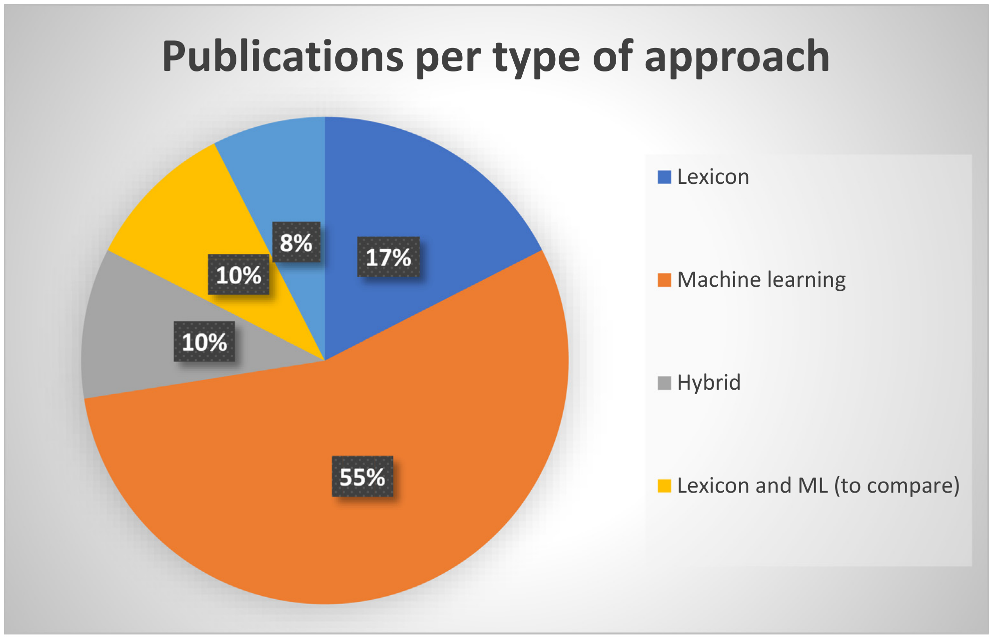

Fig. 3. Number of publications per type of approach.

K-Nearest Neighbors (KNN) is a simple ML approach, but it has an issue with noise features, leading to bad performance. In order to improve the performance of KNN in sentiment analysis, Daeli and Adiwijaya [39] attempted to employ information gain to conduct feature selection and evaluate its influence on the performance of KNN as well as other conventional machine learning (CML) models, including Naïve Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF). On a secondary movie dataset, named polarity dataset v2.0 with 2000 movie reviews, KNN, with the help of Information gain feature selection, becomes the best performing method with 96.8% accuracy while the optimum K is 3.

Al-Bakri et al. [2] employed CML algorithms to evaluate Iraqi e-tourism firms based on extracting sentiments from Iraqi dialect reviews on Facebook. The algorithms include KNN, NB and Rough Set Theory (RST), and their performances were evaluated through a confusion matrix. The results show that the best classifier is RST, with an accuracy value of nearly 95%. In the work of Kastrati et al. [11] that studies the public's views expressed on Facebook about the COVID-19 in low-resource languages, the third group of experiments is to analyze the performance of SVM, Decision Tree (DT), RF and NB models in the health domain. Their results show that RF outperformed all the CML classifier models, both in terms of count occurrence TF and in terms of term frequency-inverse document frequency TF-IDF as feature representations.

In recent years, deep learning models have shown excellent performance in sentiment analysis. Particularly, the Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) and their different combination forms have been studied by a number of scholars.

Rehman et al. [12] proposed a hybrid CNN-LSTM model to classify the IMDB dataset and Amazon movie review. The results show that their model improves the f-measure score up to 4%–8% when compared with CNN and LSTM individually. El-Affendi et al. [23] tested another combining form of the CNN and LSTM models, model. They compared the model's performance with other models', including KNN, NN, CNN, LSTM, and CNN-LSTM. The results show that the proposed model outperforms other models with an overall accuracy greater than 97%. As Rehman et al. [12] and El-Affendi et al. [23] both used the IMDb dataset and used the accuracy and F-measure to evaluate their models, we can compare their performance models. In terms of the CNN-LSTM model, the one of El-Affendi et al. (F-measure: 92%; Accuracy: 93.8%) is better than that of Rehman et al. (F-measure: 88%; Accuracy: 91%).

Bidirectional LSTM (BiLSTM) is designed to capture the semantics in both directions better to compensate for the inability of LSTM to encode information from back to front. Salur and Aydin [10] and Ghorbani et al. [31] tested models with the BiLSTM layer on Turkish tweets and IMDb dataset, respectively, and both of their results show that the BiLSTM layer can help sentiment classifiers achieve a better performance. However, the models with BiLSTM may not always outperform. For example, in the work of Kastrati et al. [11], the first group of experiments compares the performance of the CNN, BiLSTM, and the hybrid CNN-BiLSTM for sentiment analysis in the health domain. Unlike other scholars, their results show that the 1D CNN achieved the best performance in classifying the dataset they collected, outperforming the hybrid CNN-BiLSTM.

To improve deep learning models' performance better, several scholars attempt to apply the Attention Mechanism to extract high-level feature vectors. Zheng and Zheng [20] proposed a bidirectional recurrent convolutional neural network attention-based (BRCAN) model. Their evaluation results indicate that the attention mechanism improved the accuracy of all the deep learning models, and the BRCAN model achieved the best accuracy of 96.32%. In the work of Kastrati et al. [11], the first group of experiments also test the role of the attention mechanism in improving the performance of CNN, BiLSTM, and the hybrid CNN-BiLSTM models. The difference between the hybrid models in these two papers is the order of the CNN layer and the BiLSTM layer. Their results also show the effectiveness of the attention mechanism. However, in their study, the Bi-LSTM model with an attention mechanism achieved the best performance.

According to Devlin et al. [56], BERT is an attention-based architecture that shows significant performance on a variety of natural language processing tasks, including language inference, question answering, text classification, etc. The design of BERT co-regulates the left and right contexts in all layers, overcoming the unidirectional nature of the standard language model that limits the architectures available for pre-training. Kastrati et al. [11] tested the performance of the contextualized word embeddings and the Multilingual BERT classifier in their work. The mBERT encoder layer contains 12 successive transformer layers trained on Wikipedia pages. The performance of the Multilingual BERT classifier is also very competitive with the best model.

4.3. Hybrid approaches

Among the 40 papers, 10% of them adopted hybrid methods that integrate the lexicon-based and machine learning-based approaches.

Zainuddin et al. [37] conducted a finer-grained hybrid sentiment analysis on Twitter at the aspect level. The Association Rule Mining (ARM) and Stanford Dependency Parser (SDP) method were first used to identify the aspects in their model. Then, the dictionary-based method was employed to detect the sentiment words in the identified aspects to identify the significant aspects. Three feature selection approaches, including Principal Component Analysis (PCA), Random Projection (RP), and latent semantic analysis (LSA), were then conducted, respectively, to eliminate redundant and irrelevant features. After these steps of feature selection, the most relevant features were input to train the SVM classifier and conduct the sentiment classification. The authors test the hybrid models and other baseline modes on three different Twitter datasets representing different domains. The hybrid model achieved an accuracy of over 70%.

Gopi et al. [48] used the dictionary-based approach to conduct opinion mining and assigned sentiment scores of −5 to +5 to tweets using the WordNet lexicon. Their proposed model was built based on the existing SVM-RBF classifier. According to the authors, the existing system cannot accurately classify neutral opinions. Therefore, they intended to improve it by changing the gamma value and soft margin to realize the perfect scoring for neural sentences as well. Their results indicate that the proposed model outperformed the existing SVM-RBF classifier and other models, achieving an accuracy of 98.8%.

Dashtipour et al. [57] integrated the rule-based approach and deep neural networks to construct a hybrid sentiment analysis framework for the Persian language. By using the PerSent lexicon, the framework firstly detected the polarity of the reviews based on symbolic dependency relations. Then, the unclassified reviews from the rule-based approach are fed into the DL classifiers to conduct the polarity detection. The DL classifiers the authors used are CNN and LSTM. The two datasets they used are product reviews and hotel reviews, while the PerSent lexicon is built on product reviews. In terms of f-measure and accuracy, both the hybrid CNN and hybrid LSTM models outperform the rule-based approach as well as other classifiers. However, the DL classifiers improve the performance of the rule-based approach more significantly on the hotel review dataset than on the product review dataset.

Shrivastava and Kumar [26] proposed an unsupervised sentiment classifier based on a lexicon-based approach, fuzzy entropy and clustering technique. In each review, they identified the adjectives or adverbs as unigrams by the speech of part tagging and computed their sentiment scores and polarity scores based on the SentiWordNet lexicon [37]. After that, the phrases (bigrams and trigrams) were formed by using linguistic hedges, including concentrators, dilators and negators. The sentiment scores and polarity scores of the phrases were computed by predetermined rules depending on their linguistic hedges. Based on the fuzzy entropy, the authors then applied the k-means clustering technique to identify the key phrases used to determine the sentiment of reviews by computing and to add up their fuzzy products.

4.4. Combined with other analytic approaches

In this section, the studies that combine sentiment analysis with other analytic approaches are presented (8% of the 40 papers). Lombardo et al. [9] applied sentiment analysis and social network analysis techniques to support groups on Facebook for Italian patients who have Hidradenitis Suppurativa. They study how their mood was affected by various social network factors, for example, friendships and interactions in the group, over the years of observation. The dataset they used was collected from Facebook, which are 50000 posts and comments published from 2009 to 2017, and were manually annotated. They classified the posts or comments in a hierarchical way, and the sentiments were finally classified as love, joy, surprise, sadness, fear and anger. This study also constructed the interaction network that focused on the connections between review authors and repliers and the social network that focused on users and their friendship. The results indicate a significant correlation between the degree of a node and the prevalence of positive emotions, suggesting a possible positive effect in establishing stable social connections within support groups.

In the work of Jaidka et al. [16], they applied methods of social network analysis and sentiment analysis to predict the results of the general elections in Pakistan (2013), Malaysia (2013), and India (2014). Their method correctly predicted the election results in Pakistan and India but incorrectly predicted the election results in Malaysia. According to the authors, a combination of several methods performs better than independent methods. Twitter posts closer to the time of the election are more meaningful for election prediction.

Ricard et al. [35] employed an integrated method of sentiment analysis and an elastic-net regularized linear regression model to detect depression in social media content. Through sentiment analysis on the Instagram post captions and comments, several features were extracted and input into the regression model as a part of the variables. The features include sentiment scores regarding happiness, arousal and dominance, as well as emoji sentiment scores.

With the intention of improving the accuracy of the prediction model on monthly vehicle sales, Pai and Liu [38] combined the sentiment analysis technique with the statistical analysis models, including multivariate regression and time series forecasting models. The sentiment score obtained by the sentiment analysis on the social media comments was considered as an input in the statistical model. The findings confirmed that the integration of sentiment scores improved forecasting accuracy.

5. Languages and datasets

In this section, we present the languages of the datasets of reviews, comments, or posts for the reviewed papers included in this study. The languages of the datasets vary. As shown in Fig. 4, sentiment analysis has been applied to 11 languages: Albanian, Arabic, English, Hindi, Malayalam, Italian, Iraqi, Persian, Serbian and Turkish. Among the 40 papers investigated by this review paper, 29 of them use datasets for sentiment analysis in English. Research on sentiment analysis in English has yielded significant achievements, advancing not only to adapt the state-of-the-art theories in the fields of lexicon approaches [1,19,29] and ML approaches [15,35,39,41] but also at the application level, including different domains: movie [12,25], health [21], tourism [22], product [25], and election [16]. Although the sentiment analysis research for datasets in English has been extensive, the application domain may not use English as the major language [44]. For example, in the work of Jaidka et al. [16], they collected English tweets that expressed people's opinions on the general elections in Pakistan (2013), Malaysia (2013), and India (2014) to predict the election outcome based on sentiment analysis and social network analysis. The hybrid model successfully predicted the election results in Pakistan and India, but not in Malaysia. A potential reason is that only 23% of the tweets were written in English for the Malaysian election. The authors generated nearly 1.1 m tweets to predict the election in Malaysia (2013), and they removed a significant percentage of Malay tweets [16], possibly leading to bias in the classification results. Further research can be conducted on the construction of Malay lexicons and the use of effective ML classifiers.

In addition to English, sentiment analysis studies on other languages are increasing, such as Chinese [20,27], Albanian [11], Arabic [23], Hindi [26], Turkish [10], Iraqi [2], etc. Due to the unavailability of suitable lexicons, scholars who study sentiment analysis of non-English texts prefer to build their own dictionaries as opposed to sentiment analysis of English texts [2,6,27]. However, the usefulness of a model based on a self-built dictionary may not be ideal due to a variety of

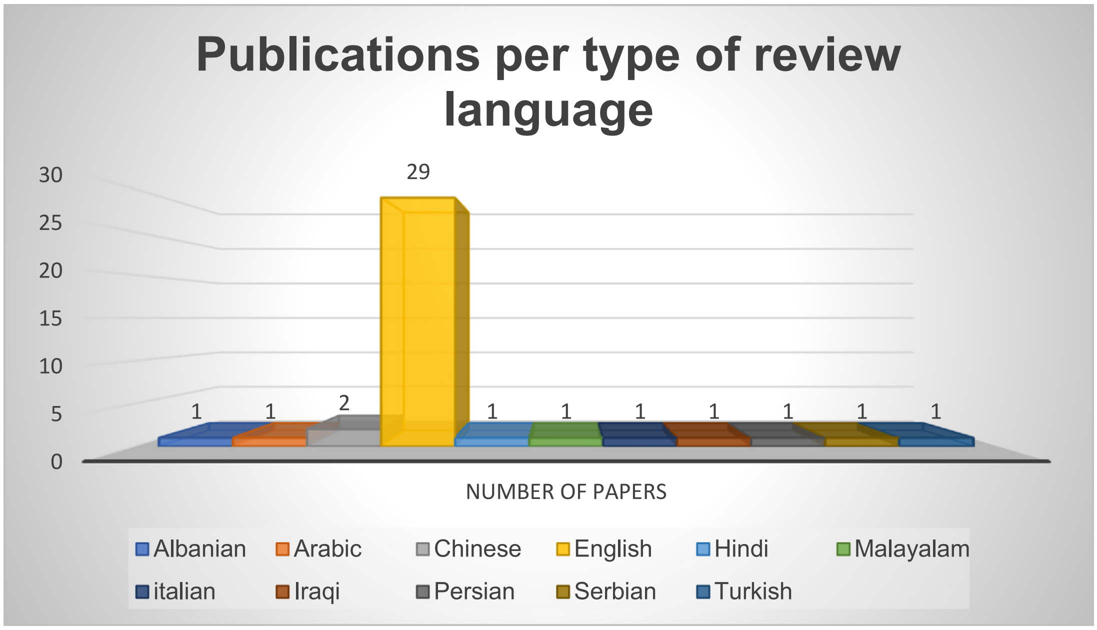

Fig. 4. Publications per type of review language.

factors in the self-built process. One potential reason for this is the source of the data. With the development of mobile technology, there is a wide range of platforms for social networking. Therefore, one lexicon-based data from one or a few social media platforms is not comprehensive. On the other hand, although larger corpora have been shown to improve the accuracy of classifiers (Rice and Zorn, 2021), acquiring such corpora requires a lot of time and effort.

Among the 40 studies included in the review, 22 of them employed the primary dataset, which was manually collected from various social media platforms, and 19 used the secondary datasets that have been studied by other authors. In the work of Khatoon et al. [19], to construct a domain-independent automatic labeling system, aside from the secondary dataset of movie reviews (IMDb) and product reviews (Amazon Product Review), they also collected reviews on business from Twitter. The domains of the datasets consist of tourism, health, movie, product, book, companies, education and crime. The most popular primary datasets are IMDb and Amazon product reviews, especially in studies where some authors design a novel sentiment classifier and use them to test the performance of the models constructed [18,19,25,29,36]. In addition, as shown in Fig. 5, only 18.33% of publications used datasets with less than 10,000 samples and 7.13% of datasets with more than 1 million samples.

6. Challenges and issues

In this section, we discuss the existing challenges faced by the researchers. These can be classified into two groups: challenges related to the datasets and challenges related to the methods. In addition, we also identify the potential issues relating to the evaluation metrics. The challenges and issues are summarized in Fig. 6 and discussed in detail in the following sections.

6.1. Data-related challenges

Massive amounts of data sourced from various social networking platforms provide new insights for individuals, businesses and governments [3]. The quality of the data is an important step toward effective analysis. The challenges of data samples can be further divided into two subcategories in relation to the data collection and the data pre-processing steps.

Data volume is one of the first challenges in the data collection process. On the one hand, building an adequate training set for sentiment analysis is crucial to improving the accuracy of the classifier [9,39]. A large dataset is a prerequisite for developing a model that is not prone to overfitting [11]. In some studies on sentiment analysis of non-English texts, scholars often decide to create their own training sets based on supervised learning and manual annotation due to the lack of useful datasets and to obtain sufficient data samples [9,11,13,22,24]. However, in addition to the advantages that come with large-scale data, researchers are overwhelmed by the volume and velocity of social media data [58]. Therefore, identifying and using an appropriate data collection method is crucial.

On the other hand, for scholars working with primary data, the imbalance of data [9,11] and comments of different lengths [11,12,20] are among the challenges often encountered, in addition to the challenges that come with large-scale data. Moreover, in order to use supervised methods, manual comment labeling is often performed after data collection, which may result in bias. This is particularly the case when annotators annotate comments that contain metaphorical language such as sarcasm and irony. Figurative language is very dependent on context, setting and topic, which creates difficulties for annotators to find the actual emotions people are expressing in their comments, and therefore a given comment may be annotated differently by the annotators [10,11].

After data collection, pre-processing of the data is often challenging as well due to the data variety. Data from social networking platforms is unstructured [22,25] and contains noisy [1,21], inefficient [10] and irrelevant [12] information that generally does not provide textual emotional content. Singla et al. [28] argue that pre-processing plays a major role in determining sentiments when it comes to a non-uniform stream of textual data. Therefore, a range of data pre-processing techniques is needed to extract useful information from the vast amount of data and turn text into a predictable and analyzable form to accomplish the required tasks [12,25].

When conducting the data pre-processing, aside from some common steps such as converting the upper case to lower case, removing unnecessary and multiple pronunciations, removing URLs, text correction, and removing stop words [28,29,48], some studies also exclude nonalphabetic terms [5,11]. However, emoticons are also a part of nonalphabetic terms and are significantly related to the users' sentiments. Removing them from the dataset may influence the performance of

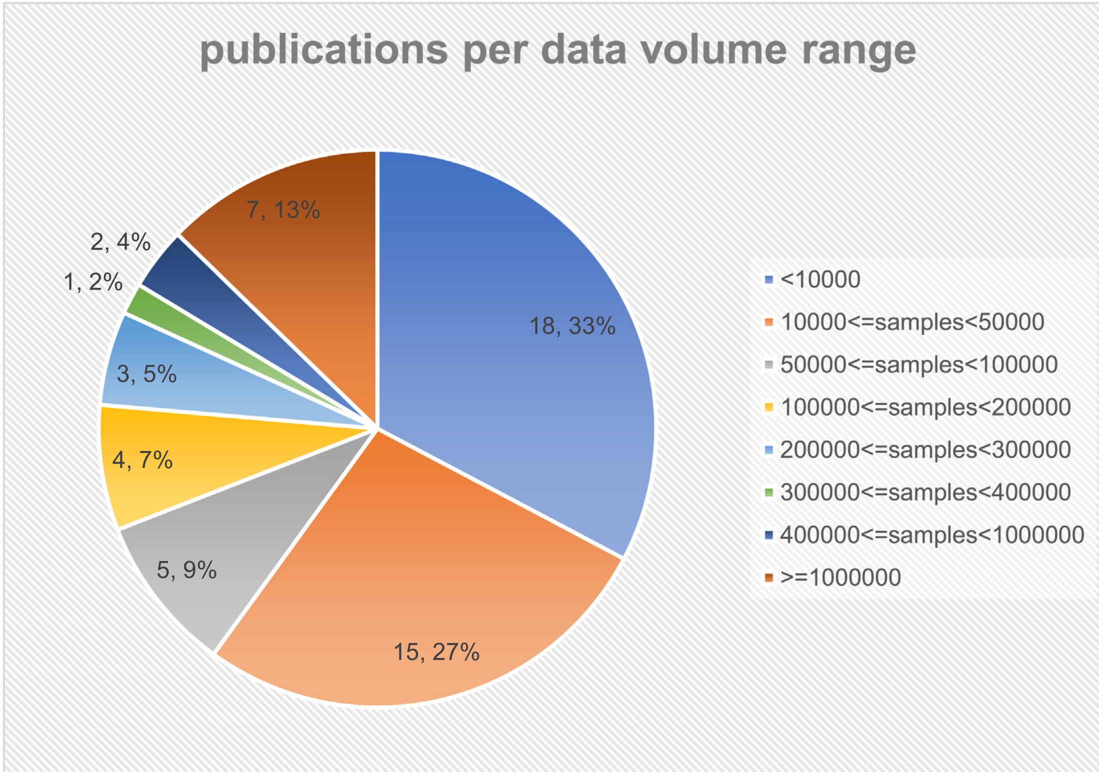

Fig. 5. The number of publications per data volume range.

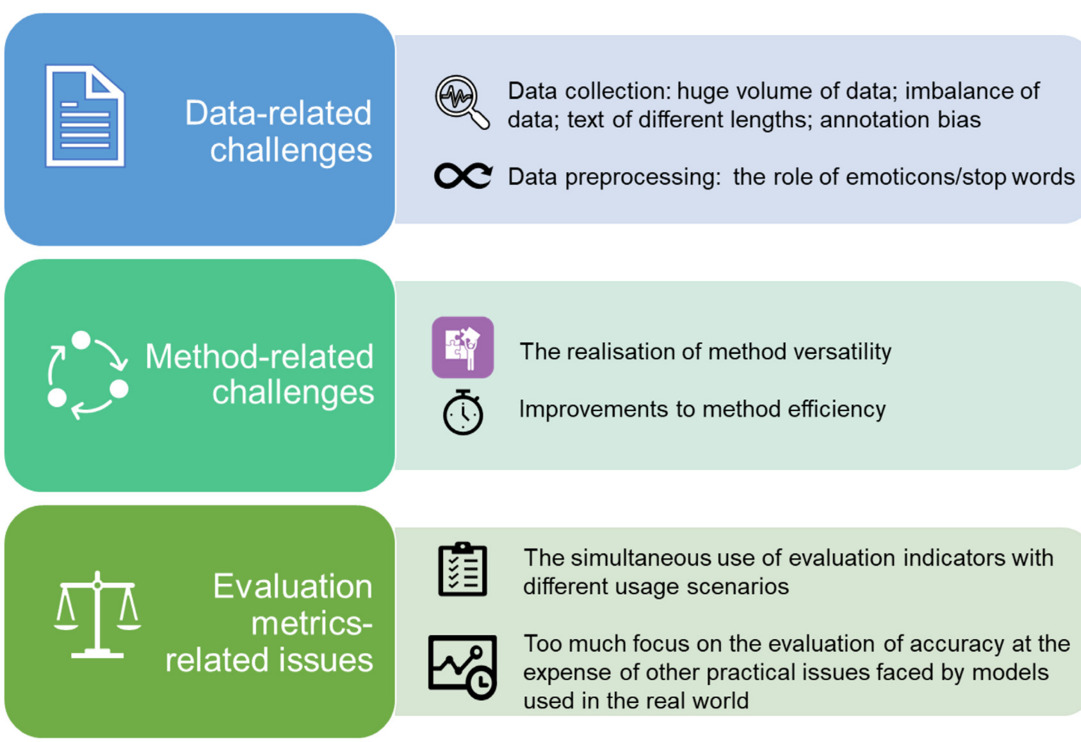

Fig. 6. Challenges and issues of existing sentiment analysis research.

the classifiers. In order to address this issue, Agüero-Torales et al. [1] employed an emoticons corpus, and Lombardo et al. [9] substituted smiles and emoticons with appropriate words. In the future, works can be conducted on constructing a comprehensive emoticons lexicon that incorporates the most used emoticons in social network platforms and effective and efficient ML classifiers, considering emoticons as they also imply people's attitudes or sentiments toward a given topic.

In addition, the treatment of stop words varies from study to study. Some studies removed stopwords [9,18,19], while others retained them [1]. Removing stop words can reduce the size of the text corpus, thus improving the robustness and performance of models. However, it can also change the meaning of the sentences, thus affecting the accuracy of the model. HaCohen-Kerner et al. [59] investigated the impact of different types of text pre-processing methods on model performance. One of their findings is that the impact of removing stop words varies between datasets, improving or decreasing the performance of the model on some data. Therefore, to optimize sentiment classifiers for different domains, scholars need to test and adapt them to the project.

6.2. Method-related challenges

As discussed in Sections 5 and 3.2, sentiment analysis has been applied in a variety of domains, including tourism, health, movie, product, book, companies, education and crime. A common thread in all these cases across different application domains suggests that sentiment analysis is invaluable for gaining accurate insight into public opinion on specific topics of interest [11]. However, the realization of a general and efficient approach to sentiment analysis can be challenging for scholars.

One major challenge is the versatility of methods. Sentiment analysis is usually conducted by two approaches: lexicon-based approach and machine learning approach. The lexicon-based approach depends on the presence of predefined words in dictionaries. Therefore, its accuracy is influenced by the richness of the dictionaries, for example, usage patterns, rules, and language constructs [19]. Moreover, the sentiment direction of a word in one dictionary may be opposite to that in another dictionary. The absence of a comprehensive dictionary makes it difficult to realize the versatility of lexicon-based methods. In a machine learning-based approach, a number of studies employ supervised learning methods, which train the sentiment classifiers using labeled data samples from a specific domain. The topic of the data samples significantly impacts the application domain of the sentiment classifiers.

Although these methods achieve great application in the studied domain, they are not absolutely great in other domains. For example, in the work of Dashtipour et al. [57], it can be seen that due to the PerSent lexicon being more related to products, the lexicon-based approach performed better than the LSTM and CNN classifiers on the product review dataset, while less on hotel review dataset. In addition, while the hybrid models outperformed other classifiers on both datasets, the hybrid CNN model performed better on the product review dataset. In contrast, the hybrid LSTM model performed better on the hotel review dataset. In order to address the domain-dependent issue of sentiment analysis, Khatoon et al. [19] constructed a DIALS system and compared its performance with lexicon-based and ML-based approaches in three datasets with different domains. According to their results, the performance of ML approaches decreased significantly when they were trained by movie reviews and tested on product reviews and business tweets. In terms of the lexicon approaches, SO-CAL outperformed FBS on product reviews and tweets but underperformed on movie reviews. Therefore, it is crucial to design a general sentiment analysis method to make sentiment analysis more practical. In addition, building a comprehensive lexicon or dataset with various topics often requires a lot of time and manpower for manual annotation [60]. A few scholars have already started to work on this area [23,28,37]. For example, Khatoon et al. [19] developed an automatic system to label extensive textual data in a domain-independent, unsupervised, and scalable manner. The future research direction should gradually shift from the application to a single domain to a multi-domain application.

For the application of sentiment analysis in the real world, aside from the versatility of methods, the efficiency of the methods is another challenge. One reason is that the user-generated data grow significantly due to the popularity and complexity of social media platforms, leading to more manpower, time, and effort to cope with increasing volume and heterogeneous types of data [61]. Another reason is that monitoring public sentiments in a timely manner for businesses and government is of great significance. For businesses, they have to respond quickly to their customer's opinions about their products, and services [4], timely assess engagement, satisfaction, and culture in the workplace [62], or constantly monitor their stock price [63,64]. For governments, an efficient application of the sentiment analysis is able to help them to understand and monitor the public's opinions, and sentiments toward public events, such as COVID-19 [3,7,21,24,30,32], election [16], welfare [65], etc., so that they can take corresponding actions or make appropriate policies in time. Therefore, the execution time of the sentiment analysis is a key factor in their practical implementation in the real world. While the various methods in a number of studies show great performance in terms of accuracy, F1-score, or other statistical evaluation metrics, few of them evaluate their models on the basis of running time.

6.3. Evaluation metrics related issues

After investigating the papers included in this review paper, it can be found that there are a variety of measures that have been used to evaluate the sentiment classifiers, including accuracy, F1-score, recall, precision, processing time, space employing, ROC curve, root mean square error (RMSE), BCE Loss, Hamming Loss, Jaccard Score, LRAP Score, Mean Absolute Error (MAE), etc. The most frequent group of evaluation metrics for these papers and reports is accuracy, f-measure, recall, and precision, which can be computed by the confusion matrix, which visualizes and summarizes the performance of a classification algorithm. However, the most discussed metric is accuracy, followed by F-measure. F-measure is defined as a harmonic mean of precision and recall to deal with the issue that there is a trade-off between precision and recall. Aside from the accuracy and F-measure, Shrivastava and Kumar [26] used ROC curve to evaluate the performance of their proposed GA-GRU model in distinguishing one class from another. Jaidka et al. [16] and Yan et al. [30] employed root mean square error (RMSE) to compare model outputs with true values. In the work of Chandra and Krishna [32], BCE Loss, Hamming Loss, Jaccard Score and LRAP Score were used as metrics to evaluate the models for dealing with multi-label classification issues. In addition to RMSE, Jaidka et al. [16] also employed Mean Absolute Error (MAE) to conduct the model evaluation.

Moreover, a collated summary of model performance evaluation metrics from these studies reveals that accuracy and F-score are often reported simultaneously. However, while both assume that the distribution of categories in the test set is meaningful, they assume different use cases. Accuracy assumes that the cost difference between correctly classifying an instance and incorrectly classifying an instance is the same in each category, while the F-score additionally assumes that true negatives do not add value [66].

In addition to evaluating the performance of sentiment classifiers using the statistical metrics discussed above, it is also important to analyze their practical value to realize the use of these classifiers in the real world. Machine learning, including sentiment analysis, is a tool for solving government, business, and personal problems. These methods have emerged to help solve the problem of processing large amounts of data in the information age and therefore need to be efficient rather than complicating the problem [9].

While the various sentiment classifiers in these studies perform great in terms of accuracy, F1-score, or other statistical evaluation metrics, few of them were evaluated from a practical perspective, such as execution time [3,21,27]. For example, in the work of Dang et al. [15], the performance of the recommendation system with the two-hybrid DL sentiment classifiers (CNN-LSTM and LSTM-CNN model) both perform well and the result is similar. However, from the total number of parameters of these two models, it can be seen that CNN-LSTM has about 5.4 million parameters fewer than LSTM-CNN, which means the training time of CNN-LSTM may be lesser and makes it outperform the LSTM-CNN in terms of application value. Martin et al. [5] constructed eleven

Table 2
Suggestions for future research.

Challenges	Suggestions for future research
Data-related challenges	• Create and use a unified dictionary or dataset containing a variety of subject terms and emoticons from various social media platforms; • Test and adapt pre-processing methods to the project
Method-related challenges	• Design generic sentiment classifiers to realize multi-domain applications.
Evaluation measure-related issues	• Pay attention to the use of each evaluation metric; • Evaluate sentiment classifiers in a more holistic way, considering the efficiency, scalability and extensibility of models

different deep learning models to deal with the sentiment classification problem in the domain of tourism. They evaluated their model using accuracy and concluded that the proposed LSTM model had the best performance, with an accuracy of 89.19%. Although the authors did not discuss the training time, they presented the results for each model allowing us to evaluate the models from a practical perspective. The combination of CNN and LSTM performed the best, taking into account the training time as a part of the performance evaluation. The reason is that the accuracy of the proposed LSTM is only 0.9% higher than the CNN-LSTM model, but the training time is more than six times.

As discussed in Section 6.2, the efficiency of a sentiment analysis model is crucial to dealing with real-world problems, especially in the areas where data flow is high and answers need to be taken in real-time [67]. Therefore, future works should evaluate the sentiment classifiers in a more comprehensive way, including the practical use of the methods such as the efficiency, extensibility and scalability.

6.4. Suggestions for future research on sentiment analysis in social media

To provide a reference for researchers to address these issues in future studies, this paper proposes the following suggestions or research directions (Table 2):

First, for both sentiment analysis tasks on English and non-English texts, it is recommended to use a unified dictionary or dataset containing a variety of subject terms and emoticons. Although this task is time and labor-intensive, having a unified lexicon or dataset allows researchers to focus on improving model performance and does not have to spend additional effort on the data. Moreover, model performance can be compared across studies. Due to the diversity of social media platforms, this unified lexicon or dataset should cover as many social media platforms as possible to improve comprehensiveness. In addition, to optimize sentiment classifiers for different domains, scholars need to test and adapt their pre-processing methods to the project.

Second, in order to make sentiment analysis more practical, future research should gradually move from applications in a single domain to multi-domain applications by designing generic sentiment classifiers.

Third, scholars should pay attention to the use of each evaluation metric when evaluating the performance of their models. While there are a variety of evaluation metrics, it is not necessary to report many evaluation metrics at the same time, as they may assume different use cases.

Fourth, the efficiency of sentiment analysis models is crucial for dealing with real-world problems, especially in areas with high data traffic where answers need to be made in real-time. Therefore, future work should evaluate sentiment classifiers in a more holistic way that includes the practical use of the methods, such as efficiency, scalability and extensibility.

7. Conclusion

This paper provides a systematic review and analysis of the literature on sentiment analysis in social media. In addition to gaining a comprehensive understanding of the application of sentiment analysis to user-generated data, the paper identifies the challenges and issues in the existing sentiment analysis research. Using the PRISMA framework, the paper reports the objectives of sentiment analysis tasks, the general implementation process, the algorithms adopted and how they are used in different domains. Afterward, by comparing aspects of different studies, the paper presents several challenges and issues related to datasets, languages of the review text, analysis methods and evaluation metrics in the existing literature.

Our work also has several limitations. As our aim is to investigate the most recent studies on sentiment analysis in social media, the publication period was set to be between 2018 to 2021. In addition, this review paper only included the publications written in English, which may lead to an inadequate understanding of the sentiment analysis of non-English texts, for example, the development of lexicons of non-English texts, as well as methods for non-English data extraction, cleaning, tokenization and analysis. In our future work, the timeframe will be expanded to capture more literature varieties and understand more sentiment analysis techniques. Researchers from non-English speaking countries or multilingual scholars will be invited to work together to understand the current state of global sentiment analysis research better.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M.M. Agüero-Torales, M.J. Cobo, E. Herrera-Viedma, A.G. López-Herrera, A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor, Procedia Comput. Sci. 162 (2019) 392–399, http://dx.doi.org/10.1016/j.procs.2019.12.002.

[2] N.F. Al-Bakri, J. Yonan, A.T. Sadiq, A. Abid, Tourism companies assessment via social media using sentiment analysis, Baghdad Sci. J. (2021).

[3] F.A. Ibrahim, M. Hassaballah, A.A. Ali, Y. Nam, A.I. Ibrahim, COVID19 outbreak: A hierarchical framework for user sentiment analysis, Comput. Mater. Continua 70 (2) (2022) 2507–2524, http://dx.doi.org/10.32604/cmc.2022.018131.

[4] A. Alamsyah, D.M. Ginting, Analyzing employee voice using real-time feedback, in: Proceedings - 2018 4th International Conference on Science and Technology, ICST 2018, 2018, http://dx.doi.org/10.1109/ICSTC.2018.8528569.

[5] C.A. Martin, J.M. Torres, R.M. Aguilar, S. Diaz, Using deep learning to predict sentiments: case study in tourism, Complexity (2018) http://dx.doi.org/10.1155/2018/7408431.

[6] O. Grljevic, Z. Bosnjak, A. Kovacevic, Opinion mining in higher education: A corpus-based approach, Enterprise Inf. Syst. (2020) http://dx.doi.org/10.1080/17517575.2020.1773542.

[7] Md.M. Rahman, M.N. Islam, Exploring the performance of ensemble machine learning classifiers for sentiment analysis of COVID-19 tweets, in: S. Shakya, V.E. Balas, S. Kamolphiwong, K.-L. Du (Eds.), Sentimental Analysis and Deep Learning, Vol. 1408, Springer, Singapore, 2022, pp. 383–396, http://dx.doi.org/10.1007/978-981-16-5157-1_30.

[8] N.C. Dang, M.N. Moreno-García, F. De la Prieta, Sentiment analysis based on deep learning: A comparative study, Electronics 9 (3) (2020) 483, http://dx.doi.org/10.3390/electronics9030483.

[9] G. Lombardo, P. Fornacciari, M. Mordonini, L. Sani, M. Tomaiuolo, A combined approach for the analysis of support groups on facebook—The case of patients of hidradenitis suppurativa, Multimedia Tools Appl. 78 (3) (2019) 3321–3339, http://dx.doi.org/10.1007/s11042-018-6512-5.

[10] M.U. Salur, I. Aydin, A novel hybrid deep learning model for sentiment classification, IEEE Access 8 (2020) 58080–58093, http://dx.doi.org/10.1109/ACCESS.2020.2982538.

[11] Z. Kastrati, L. Ahmedi, A. Kurti, F. Kadriu, D. Murtezaj, F. Gashi, A deep learning sentiment analyser for social media comments in low-resource languages, Electronics 10 (10) (2021) http://dx.doi.org/10.3390/electronics10101133.

[12] A.U. Rehman, A.K. Malik, B. Raza, W. Ali, A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis, Multimedia Tools Appl. 78 (18) (2019) 26597–26613, http://dx.doi.org/10.1007/s11042-019-07788-7.

[13] K. Dashtipour, M. Gogate, A. Adeel, H. Larijani, A. Hussain, Sentiment analysis of Persian movie reviews using deep learning, ENTROPY 23 (5) (2021) http://dx.doi.org/10.3390/e23050596.

[14] S. Kausar, H. Xu, W. Ahmad, M.Y. Shabir, A sentiment polarity categorization technique for online product reviews, IEEE Access 8 (2020) 3594–3605, http://dx.doi.org/10.1109/ACCESS.2019.2963020.

[15] C.N. Dang, M.N. Moreno-Garcia, F. De la Prieta, Using hybrid deep learning models of sentiment analysis and item genres in recommender systems for streaming services, Electronics 10 (20) (2021) http://dx.doi.org/10.3390/electronics10202459.

[16] K. Jaidka, S. Ahmed, M. Skoric, M. Hilbert, Predicting elections from social media: A three-country, three-method comparative study, Asian J. Commun. 29 (3) (2019) 252–273, http://dx.doi.org/10.1080/01292986.2018.1453849.

[17] Z. Munn, M.D.J. Peters, C. Stern, C. Tufanaru, A. McArthur, E. Aromataris, Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach, BMC Med. Res. Methodol. 18 (1) (2018) 143, http://dx.doi.org/10.1186/s12874-018-0611-x.

[18] J. Chen, Y. Chen, Y. He, Y. Xu, S. Zhao, Y. Zhang, A classified feature representation three-way decision model for sentiment analysis, Appl. intell. (2021) http://dx.doi.org/10.1007/s10489-021-02809-1.

[19] S. Khatoon, L. Abu Romman, M.M. Hasan, A domain-independent automatic labeling system for large-scale social data annotation using lexicon and web-based augmentation, Inf. Technol. Control 49 (1) (2020) 36–54, http://dx.doi.org/10.5755/j01.itc.49.1.23769.

[20] J. Zheng, L. Zheng, A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification, IEEE Access 7 (2019) 106673–106685, http://dx.doi.org/10.1109/ACCESS.2019.2932619.

[21] F. Es-Sabery, K. Es-Sabery, J. Qadir, B. Sainz-De-Abajo, A. Hair, B. Garcia-Zapirain, I. de la Torre-diez, A MapReduce opinion mining for COVID-19-related tweets classification using enhanced ID3 decision tree classifier, IEEE Access 9 (2021) 58706–58739, http://dx.doi.org/10.1109/ACCESS.2021.3073215.

[22] P.K. Jain, R. Pamula, E.A. Yekun, A multi-label ensemble predicting model to service recommendation from social media contents, J. Supercomput. (2021) http://dx.doi.org/10.1007/s11227-021-04087-7.

[23] M.A. El-Affendi, K. Alrajhi, A. Hussain, A novel deep learning-based multilevel parallel attention neural (MPAN) model for multidomain arabic sentiment analysis, IEEE Access 9 (2021) 7508–7518, http://dx.doi.org/10.1109/ACCESS.2021.3049626.

[24] M.E. Basiri, S. Nemati, M. Abdar, S. Asadi, U.R. Acharrya, A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets, Knowl.-Based Syst. 228 (2021) 107242, http://dx.doi.org/10.1016/j.knosys.2021.107242.

[25] I. Priyadarshini, C. Cotton, A novel LSTM-CNN-grid search-based deep neural network for sentiment analysis, J. Supercomput. 77 (12) (2021) 13911–13932, http://dx.doi.org/10.1007/s11227-021-03838-w.

[26] K. Shrivastava, S. Kumar, A sentiment analysis system for the hindi language by integrating gated recurrent unit with genetic algorithm, Int. Arab J. Inf. Technol. 17 (6) (2020) 954–964, http://dx.doi.org/10.34028/iajit/17/6/14.

[27] Y. Wang, M. Wang, W. Xu, A sentiment-enhanced hybrid recommender system for movie recommendation: A big data analytics framework, Wirel. Commun. Mob. Comput. (2018) http://dx.doi.org/10.1155/2018/8263704.

[28] C. Singla, N.F. Al-Wesabi, Y. Singh Pathania, B. Sulaiman Alfurhood, A. Mustafa Hilal, M. Rizwanullah, M. Ahmed Hamza, M. Mahzari, An optimized deep learning model for emotion classification in tweets, Comput. Mater. Continua 70 (3) (2022) 6365–6380, http://dx.doi.org/10.32604/cmc.2022.020480.

[29] D.H. Abd, A.R. Abbas, A.T. Sadiq, Analyzing sentiment system to specify polarity by lexicon-based, Bull. Electr. Eng. Inf. 10 (1) (2021) 283–289, http://dx.doi.org/10.11591/eei.v10i1.2471.

[30] C. Yan, M. Law, S. Nguyen, J. Cheung, J. Kong, Comparing public sentiment toward COVID-19 vaccines across Canadian cities: Analysis of comments on reddit, J. Med. Internet Res. 23 (9) (2021) http://dx.doi.org/10.2196/32685.

[31] M. Ghorbani, M. Bahaghighat, Q. Xin, F. Ozen, ConvLSTMConv network: A deep learning approach for sentiment analysis in cloud computing, J. Cloud Comput. Adv. Syst. Appl. 9 (1) (2020) http://dx.doi.org/10.1186/s13677-020-00162-1.

[32] R. Chandra, A. Krishna, COVID-19 sentiment analysis via deep learning during the rise of novel cases, PLoS One 16 (8) (2021) http://dx.doi.org/10.1371/journal.pone.0255615.

[33] S. Kumar, M. Gahalawat, P.P. Roy, D.P. Dogra, B.-G. Kim, Exploring impact of age and gender on sentiment analysis using machine learning, Electronics 9 (2) (2020) http://dx.doi.org/10.3390/electronics9020374.

[34] R.G. Bilro, S.M. Correia Loureiro, J. Guerreiro, Exploring online customer engagement with hospitality products and its relationship with involvement, emotional states, experience and brand advocacy, J. Hosp. Mark. Manag. 28 (2) (2019) 147–171, http://dx.doi.org/10.1080/19368623.2018.1506375.

[35] B.J. Ricard, L.A. Marsch, B. Crosier, S. Hassanpour, Exploring the utility of community-generated social media content for detecting depression: An analytical study on instagram, J. Med. Internet Res. 20 (12) (2018) http://dx.doi.org/10.2196/11817.

[36] S. Vashishtha, S. Susan, Highlighting keyphrases using senti-scoring and fuzzy entropy for unsupervised sentiment analysis, Expert Syst. Appl. 169 (2021) 114323, http://dx.doi.org/10.1016/j.eswa.2020.114323.

[37] N. Zainuddin, A. Selamat, R. Ibrahim, Hybrid sentiment classification on twitter aspect-based sentiment analysis, Appl. Intell. 48 (5) (2018) 1218–1232, http://dx.doi.org/10.1007/s10489-017-1098-6, SI.

[38] P.-F. Pai, C.-H. Liu, Predicting vehicle sales by sentiment analysis of Twitter data and stock market values, IEEE Access 6 (2018) 57655–57662, http://dx.doi.org/10.1109/ACCESS.2018.2873730.

[39] N.O.F. Daeli, A. Adiwijaya, Sentiment analysis on movie reviews using information gain and K-nearest neighbor, J. Data Sci. Appl. (2020) 1–7, http://dx.doi.org/10.34818/JDSA.2020.3.22.

[40] M. Thomas, C.A. Latha, Sentimental analysis of transliterated text in malayalam using recurrent neural networks, J. Ambient Intell. Humaniz. Comput. 12 (6) (2021) 6773–6780, http://dx.doi.org/10.1007/s12652-020-02305-3.

[41] Y.-C. Chang, C.-H. Ku, C.-H. Chen, Social media analytics: Extracting and visualizing hilton hotel ratings and reviews from TripAdvisor, Int. J. Inf. Manage. 48 (2019) 263–279, http://dx.doi.org/10.1016/j.ijinfomgt.2017.11.001.

[42] N. Jnoub, F. Al Machot, W. Klas, A domain-independent classification model for sentiment analysis using neural models, Appl. Sci. Basel 10 (18) (2020) http://dx.doi.org/10.3390/app10186221.

[43] M. Ibrahim, I.S. Bajwa, R. Ul-Amin, B. Kasi, A neural network-inspired approach for improved and true movie recommendations, Comput. Intell. Neurosci. (2019) http://dx.doi.org/10.1155/2019/4589060.

[44] A.H. Alamoodi, B.B. Zaidan, A.A. Zaidan, O.S. Albahri, K.I. Mohammed, R.Q. Malik, E.M. Almahdi, M.A. Chyad, Z. Tareq, A.S. Albahri, H. Hameed, M. Alaa, Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review, Expert Syst. Appl. 167 (2021) 114155, http://dx.doi.org/10.1016/j.eswa.2020.114155.

[45] A. Kumar, G. Garg, Systematic literature review on context-based sentiment analysis in social multimedia, Multimedia Tools Appl. 79 (1534) (2020) http://dx.doi.org/10.1007/s11042-019-7346-5, 21–229–15380.

[46] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J.M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, et al., The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, BMJ (2021) http://dx.doi.org/10.1136/bmj.n71, n71.

[47] S. Chauhan, S. Saxena, P. Daniel, Monolingual and parallel corpora for kangri low resource language, 2021, http://dx.doi.org/10.48550/ARXIV.2103.11596.

[48] A.P. Gopi, R.N.S. Jyothi, V.L. Narayana, K.S. Sandeep, Classification of tweets data based on polarity using improved RBF kernel of SVM, Int. J. Inf. Technol. (Singapore) (2020) http://dx.doi.org/10.1007/s41870-019-00409-4.

[49] Z. Hameed, B. Garcia-Zapirain, Sentiment classification using a single-layered BiLSTM model, IEEE Access 8 (2020) 73992–74001, http://dx.doi.org/10.1109/ACCESS.2020.2988550.

[50] G. Choi, S. Oh, H. Kim, Improving document-level sentiment classification using importance of sentences, ENTROPY 22 (12) (2020) http://dx.doi.org/10.3390/e22121336.

[51] L. Mai, B. Le, Joint sentence and aspect-level sentiment analysis of product comments, Ann. Oper. Res. 300 (2) (2021) 493–513, http://dx.doi.org/10.1007/s10479-020-03534-7, SI.

[52] W. Ansar, S. Goswami, A. Chakrabarti, B. Chakraborty, An efficient methodology for aspect-based sentiment analysis using BERT through refined aspect extraction, J. Intell. Fuzzy Syst. 40 (5) (2021) 9627–9644, http://dx.doi.org/10.3233/JIFS-202140.

[53] Md.S. Hossen, N.R. Dev, An improved lexicon based model for efficient sentiment analysis on movie review data, Wireless Personal Commun. 120 (1) (2021) 535–544, http://dx.doi.org/10.1007/s11277-021-08474-4.

[54] G.S. Budhi, R. Chiong, I. Pranata, Z. Hu, Using machine learning to predict the sentiment of online reviews: a new framework for comparative analysis, in: Archives of Computational Methods in Engineering, Vol. 28, (4) SPRINGER, 2021, pp. 2543–2566, http://dx.doi.org/10.1007/s11831-020-09464-8.

[55] S. Baccianella, A. Esuli, F. Sebastiani, SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: Proceedings of the Seventh International Conference on Language Resources and Evaluation, LREC'10, 2010, http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf.

[56] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, 2019, [Cs] arXiv:1810.04805. http://arxiv.org/abs/1810.04805.

[57] K. Dashtipour, M. Gogate, J. Li, F. Jiang, B. Kong, A. Hussain, A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks, Neurocomputing 380 (2020) 1–10, http://dx.doi.org/10.1016/j.neucom.2019.10.009.

[58] S. Stieglitz, M. Mirbabaie, B. Ross, C. Neuberger, Social media analytics – challenges in topic discovery, data collection, and data preparation, Int. J. Inf. Manage. 39 (2018) 156–168, http://dx.doi.org/10.1016/j.ijinfomgt.2017.12.002.

[59] Y. HaCohen-Kerner, D. Miller, Y. Yigal, The influence of pre-processing on text classification using a bag-of-words representation, PLoS One 15 (5) (2020) e0232525, http://dx.doi.org/10.1371/journal.pone.0232525.

[60] C. Diamantini, A. Mircoli, D. Potena, E. Storti, Social information discovery enhanced by sentiment analysis techniques, Future Gener. Comput. Syst. 95 (2019) 816–828, http://dx.doi.org/10.1016/j.future.2018.01.051.

[61] S. Sagnika, B.S.P. Mishra, S.K. Meher, Improved method of word embedding for efficient analysis of human sentiments, Multimedia Tools Appl. 79 (3238) (2020) http://dx.doi.org/10.1007/s11042-020-09632-9, 43–449–32413.

[62] A. Alamsyah, W. Rizkika, D.D.A. Nugroho, F. Renaldi, S. Saadah, Dynamic large scale data on Twitter using sentiment analysis and topic modeling, in: 2018 6th International Conference on Information and Communication Technology, ICoICT, 2018, pp. 254–258, http://dx.doi.org/10.1109/ICoICT.2018.8528776.

[63] S. Das, R.K. Behera, M. kumar, S.K. Rath, Real-time sentiment analysis of Twitter streaming data for stock prediction, Procedia Comput. Sci. 132 (2018) 956–964, http://dx.doi.org/10.1016/j.procs.2018.05.111.

[64] V. Sharma, R. Khemnar, R. Kumari, B.R. Mohan, Time series with sentiment analysis for stock price prediction, in: 2019 2nd International Conference on Intelligent Communication and Computational Techniques, ICCT, 2019, pp. 178–181, http://dx.doi.org/10.1109/ICCT46177.2019.8969060.

[65] P. Singh, Y.K. Dwivedi, K.S. Kahlon, R.S. Sawhney, A.A. Alalwan, N.P. Rana, Smart monitoring and controlling of government policies using social media and cloud computing, Inf. Syst. Front. (2019) http://dx.doi.org/10.1007/s10796-019-09916-y.

[66] P. Flach, Performance evaluation in machine learning: The good, the bad, the ugly, and the way forward, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 9808–9814, http://dx.doi.org/10.1609/aaai.v33i01.33019808.

[67] A.K. Sahoo, C. Pradhan, H. Das, Performance evaluation of different machine learning methods and deep-learning based convolutional neural network for health decision making, in: M. Rout, J.K. Rout, H. Das (Eds.), Nature Inspired Computing for Data Science, Vol. 871, Springer International Publishing, 2020, pp. 201–212, http://dx.doi.org/10.1007/978-3-030-33820-6_8.