

## Review Article

# Arabic Sentiment Analysis: A Systematic Literature Review

Abdullatif Ghallab , Abdulqader Mohsen , and Yousef Ali

University of Science and Technology, Sana'a, Yemen

Correspondence should be addressed to Abdullatif Ghallab; [ghallab@ust.edu](mailto:ghallab@ust.edu)

Received 10 July 2019; Revised 21 December 2019; Accepted 7 January 2020; Published 29 January 2020

Academic Editor: Christian W. Dawson

Copyright © 2020 Abdullatif Ghallab et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the recently grown attention from different research communities for opinion mining, there is an evolving body of work on Arabic Sentiment Analysis (ASA). This paper introduces a systematic review of the existing literature relevant to ASA. The main goals of the review are to support research, to propose further areas for future studies in ASA, and to smoothen the progress of other researchers' search for related studies. The findings of the review propose a taxonomy for sentiment classification methods. Furthermore, the limitations of existing approaches are highlighted in the preprocessing step, feature generation, and sentiment classification methods. Some likely trends for future research with ASA are suggested in both practical and theoretical aspects.

## 1. Introduction

Nowadays, Sentiment Analysis (SA) as well as opinion mining is broadly investigated research areas [1, 2]. It is an application of natural language processing (NLP), computational linguistics, and text mining to extract people's opinions or emotions towards an event, product, or others [3, 4]. In general, SA imposes identifying four elements comprising entity, its aspect, opinion holder, and his sentiment [5]. The extracted opinions can be classified to either objective or subjective text. The subjective text also can be classified to positive or negative sentiments [6].

Most studies undertaken in SA have been carried out on natural languages, such as English, Chinese, and Arabic. NLP in Arabic is still at the beginning stages [7]. It lacks the resources and tools. Therefore, Arabic still meets challenges in NLP tasks due to its structure complexity, history, and different cultures [8, 9].

A large number of tools and approaches, in the literature, are utilized to conduct the SA task. Most of them are designed to manage SA in English which is the science language [9]. These approaches are either the semantic approach or the machine learning (ML) approaches. The semantic approach extracts the sentiment words and calculates its polarities based on a sentiment lexicon [10]. On

the contrary, to build a new model, ML classifiers train annotated data, after converting to feature vectors, to conclude specific features used in a particular class. Finally, the new model can be used to predict the class of the new data [4]. It is worthy to note that these approaches can be adapted to others languages, such as Arabic [9].

Arabic language has received less efforts compared with other languages [11]; however, hundreds of studies have been proposed for ASA. Since its introduction since a decade, ASA has become one of the most popular forms of information extraction from the reviews. These reviews contributed in many benefits, such as showing the product brand or service valuable insights [12–15], identifying potential product advocates or social media influencers [16–19], and detecting e-mail spam [20]. Consequently, ASA has been studied in various contexts, and a large number of studies has been published on the topic. To our best knowledge, there has not been a systematic review that would synthesize the results obtained on ASA.

This review aims to introduce SLR of the research efforts on ASA. This SLR starts with determining the basic requirements, and subsequently, a total of 191 papers have been initially considered to be relevant. This number has been reduced to 140 papers after reviewing the abstracts of the studies. Finally, only a total of 108 papers are completely

reviewed. Through careful study and analysis of these papers, the desired information has been extracted.

It is sought in this SLR to give an outline of major research themes and techniques and suggestions for future research. The first goal of this study is to review articles to know the research current state. The second goal is to discuss the major issues influencing SA in Arabic based on the reviewed studies. In addition to theory development in the field, this article has fourfold contributions. Firstly, systematic literature review of the research efforts on ASA will be provided. Secondly, an enhanced taxonomy of ASA methods will be introduced. Thirdly, we make an attempt at creating such synthesis while placing particular emphasis on preprocessing step, feature generation, and sentiment classification methods. Finally, new trends in ASA will be suggested, and the implications for future research and practice will be highlighted.

The remainder of this paper is organized as follows. Section 2 provides the methodology. Findings are illustrated in Section 3. Section 4 presents the analysis and discussion of the existing research to identify the research gaps and makes implications for future research. Section 5 draws up the conclusion.

## 2. Review Methodology

SLR was developed as an approach to identify and review SA in Arabic. A systematic review is carried out using a systematic, explicit, and rigorous standard, aiming not only to summarize current research on the topic but also to involve an element of analytical criticism. It presents eight major steps which are essential for any review to be scientifically rigorous [21]. This study followed guidelines of [21–23] in the collection and analysis process. The details are presented in the following subsections.

*2.1. Research Questions.* Identifying the research questions is the first step of a systematic review. This step has to be concise and clear. In the context of this study, the research questions are stated as follows:

- (i) RQ1. What is the current state of research? Who has published? When?
- (ii) RQ2. What are the most effective techniques used in ASA?
- (iii) RQ3. Which are the most significant gaps and limitations in the reviewed studies?
- (iv) RQ4. What are directions of future research on ASA?

*2.2. Searching the Literature.* The targeted strategy of this review included determining the population, selecting resources, deriving search strings, and the inclusion and exclusion criteria. The literature search process of this review involved querying reputed journals and conferences dealing with ASA indexed in Scopus, including several databases, such as Springer, Elsevier, and IGI Global. The temporal

range for the review was set for the articles published from January 2013 until end of November 2018.

According to the research questions, the search was conducted with the permutations of keywords. The search query was (Arabic AND (Sentiment Analysis, Opinion Mining, SA, OR OM)) AND (Classification, classifier, Prediction, OR Polarity). Consequently, the investigated studies yielded in a total number of 191 publications. In total, it is sought to review 32 journals, 81 proceedings from 23 databases. The selected conferences, journals, and databases are illustrated in Table 1.

These publications varied in scope as well as type. In details, SLR included journal articles, conference papers, and book chapters. Referring to Figure 1, the most used source is conference papers, which covers out almost 70% of the total sources. However, reviews and journal articles are 27%. On the contrary, the least used sources are book chapters, which contribute only 3%.

*2.3. Inclusion and Exclusion Criteria.* The reviewed articles were filtered utilizing multilevel criteria which were for exclusion and inclusion, as depicted in Table 2. The exclusion process is called quality appraisal. while the inclusion process is called practical screening.

*2.4. Data Collection.* The data collected from each article to conduct the review of ASA were identified to include the following:

- (i) The source whether conference or journal and the full reference
- (ii) The article authors and their institutions
- (iii) The article title, publication year, and publisher
- (iv) The type of SA tasks conducted
- (v) The dataset or lexicon and its size, domain, and source
- (vi) SA classification level
- (vii) SA approach
- (viii) SA algorithms and its accuracy
- (ix) Arabic language type
- (x) Preprocessing process
- (xi) The feature selection and generation process
- (xii) Associative tools and applications to perform the SA process

This SLR was carried out from January 2013 to the end of November 2018 resulting in 191 articles. Figure 2 depicts the steps of prescreening process of articles resulting in the conducted review. First, a pilot study by searching the journals and conferences was conducted. The purpose of this pilot study was to test the search parameters. This step resulted in 142 articles. Second, the initial pool of articles was reduced to 91 through selecting the relevant articles. This step involved removing duplicated papers and filtering unrelated articles by reviewing titles and abstracts. Third, the pilot study was conducted again for new articles at the end of

TABLE 1: Conference proceedings, journals, and databases searched by the conducted study.

Conference(s)	Journal	Databases
(1) International Conference on Arabic Computational Linguistics	Journal of Computer Science	Springer
(2) International Conference on Information and Communication Systems	Journal of Information Science	Elsevier B.V.
(3) International Conference on Advanced Intelligent Systems and Informatics	International Arab Journal of Information Technology	IGI Global
(4) IEEE/ACS International Conference of Computer Systems and Applications	Journal of Theoretical and Applied Information Technology	SAGE Publications Ltd
(5) International Conference for Internet Technology and Secured Transactions	ACM Transactions on Asian and Low-Resource Language Information Processing	Asian Research Publishing Network (ARPN)
(6) International Conference on Future Internet of Things and Cloud	Cognitive Computation	Science Publications
(7) International Conference on Model and Data Engineering	Information Processing and Management	Zarka Private University
(8) International Conference on Social Computing and Social Media	International Journal of Advanced Computer Science and Applications	Association for Computing Machinery (ACM)
(9) Annual Meeting of the Association for Computational Linguistics	International Journal of Information Technology and Web Engineering	Insight Society
(10) International Conference on Brain-Inspired Cognitive Systems	International Journal on Advanced Science, Engineering and Information Technology	Instituto Politecnico Nacional
(11) International Conference on Language Resources and Evaluation	Journal of King Saud University-Computer and Information Sciences	King Saud Bin Abdulaziz University
(12) Conference on Language Processing and Knowledge Management	Studies in Computational Intelligence	Science and Information Organization
(13) European Network Intelligence Conference		
(14) IEEE International Conference on Big Data		
(15) International Conference of Cloud Computing Technologies and Applications		
(16) International Conference on Ambient Systems Networks and Technologies		
(17) International Conference on Applications of Natural Language to Information Systems		
(18) International Conference on Computational Linguistics and Intelligent Text Processing		
(19) International Conference on Intelligent Systems and Computer Vision		
(20) International Conference on Natural Language and Speech Processing		
(21) International Conference on Neural Information Processing		
(22) International Conference on New Trend in Computing Sciences		

November 2018 yielding 49 articles. Finally, filtering and selecting relevant articles were again done with reading full-text. Only 108 relevant articles were identified across the journals and conferences.

**2.5. Data Extraction.** This stage showed consideration for extracting data from the reviewed papers to answer the research questions. The required data to conduct the review of ASA were identified, including the type of SA tasks, the dataset, domain, and source, SA classification level, SA approach, algorithms, Arabic language type, preprocessing process, and the feature selection and generation process.

Table 3 presents the methods used in preprocessing step of the literature. These preprocessing methods include text cleaning, normalization, stemming, tokenization, part of

speech (POS) tagging, negation detection, segmentation, stop words removal, lemmatization, irony detection, and named entity recognition.

In a similar context, for datasets utilized in ASA, Table 4 presents the most datasets with their size and source extracted from the reviewed papers. These datasets were represented by antecedent's category, which ranges from D01 to D25, as shown in the first column. The second column shows the name of datasets used in the literature, such as LABR, OCA, ASTD, and BRAD. Dataset size and its trends are introduced, respectively, from the third to the sixth column. The dataset sizes ranged from 147 to 5615943 sentences. In addition, the data trends varied in positive, negative, and neutral. Finally, the last column presents the dataset sources which involved different areas, including Twitter, Facebook, and the others.

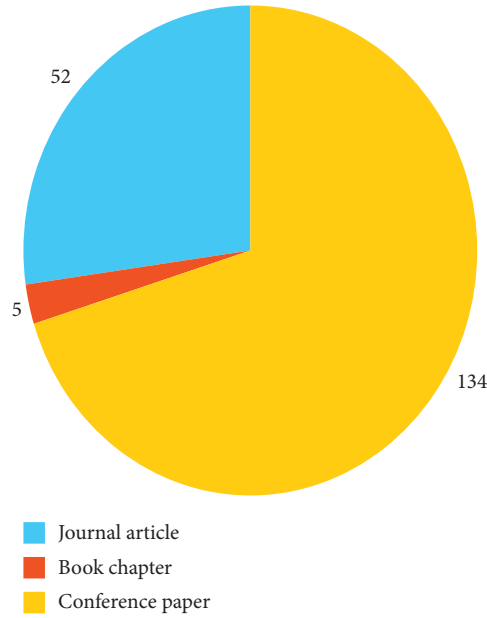


FIGURE 1: Type of searched studies.

TABLE 2: Inclusion and exclusion criteria.

Inclusion	Exclusion
Articles published between January 2013 and November 2018	Nonacademic articles and gray literature, including working papers and technical reports
Using relative keywords	Articles with weak writing and analysis
Prestige and academic papers and articles	Articles about natural language processing (NLP) and its applications, such as speech recognition, fact-checking systems, handwritten words recognition, and named entity recognition
The research focuses on sentiment analysis classification	Articles about other classifications, such as emotions classification and identifying a dialect
The research was related to Arabic language	Articles address only building a resource for SA, such as lexicon or corpus
Articles written in English	Duplicated and review articles Articles presenting research in non-Arabic language, such as Arabizi, English, and Urdu Papers that did not meet the practical screening

Table 5 presents the information extracted from 108 studies to help in exploring ASA. As shown in Table 5, SA tasks were categorized into five tasks, including aspect detection (AD), building resource (BR), sentiment classification (SC), subjectivity classification (Subj C), and aspect-based sentiment analysis (Aspect SA). Arabic language has three types, including Classical Arabic (CA), Modern Standard Arabic (MSA), and Dialect Arabic (DA).

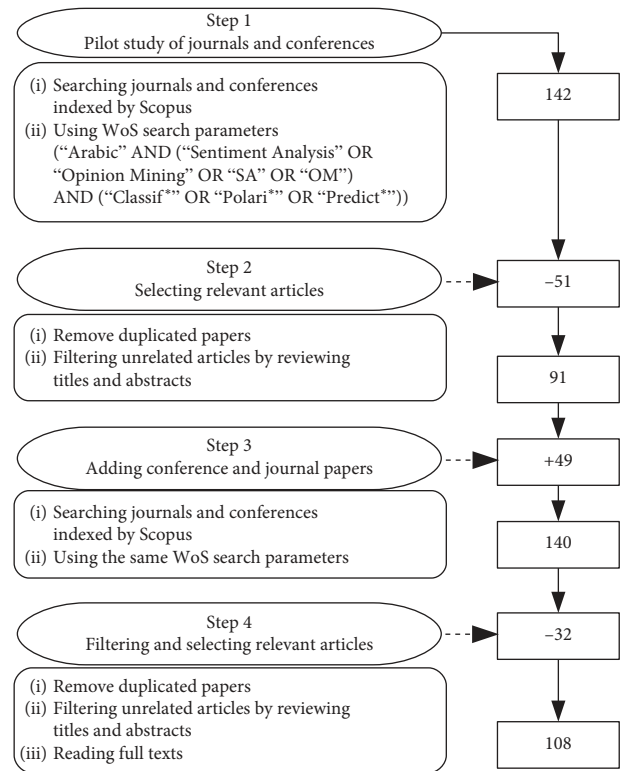


FIGURE 2: Prescreening process of the papers.

Moreover, to give more interpretation in case of recording different accuracies with multiple algorithms and multiple datasets, three typographical emphases inspired by studies [66, 108] were exploited. These typographical emphases include underline, bold, and italic. For example, in row no. 6, there are three algorithms with three datasets and three different results. To distinguish them, bold indicates that K-nearest neighbors (KNNs) algorithm

TABLE 3: Preprocessing steps conducted in the reviewed articles.

Reference	Preprocessing
[17]	Normalization, POS tagging
[24–27]	Stemming
[28–33]	Text cleaning
[34–39]	Normalization, stemming, stop words removal
[40–42]	Text cleaning, normalization, stemming, stop words removal
[43–45]	Normalization
[16–48]	Text cleaning, normalization, tokenization, stemming, stop words removal
[49–52]	Normalization, tokenization
[53, 54]	Text cleaning, normalization, tokenization
[55, 56]	Normalization, tokenization, POS tagging
[13, 57–64]	Normalization, tokenization, stemming, stop words removal
[65, 66]	Normalization, tokenization, stemming, lemmatization
[67, 68]	Text cleaning, normalization, tokenization, stemming
[69]	Text cleaning, tokenization, stemming, negation detection
[70–78]	Tokenization, stemming, stop words removal
[79]	Normalization, stop words removal
[80]	Tokenization, POS tagging, stemming
[81]	Stemming, stop words removal, irony detection
[15, 82]	Tokenization, stop words removal
[83]	Normalization, tokenization, POS tagging, stemming, stop words removal
[84, 85]	Normalization, POS tagging, stemming, stop words removal
[14, 86, 87]	Text cleaning, stemming
[88, 89]	Tokenization, stemming
[90]	Text cleaning, normalization
[91, 92]	Text cleaning, tokenization, stemming, stop words removal
[93]	POS tagging, lemmatization
[94]	Normalization, segmentation, stemming, stop words removal
[95, 96]	Text cleaning, normalization, stop words removal
[18, 97, 98]	Text cleaning, normalization, tokenization, POS tagging
[99]	POS tagging
[100]	POS tagging, stemming, stop words removal
[101]	Tokenization, stop words removal, named entity recognition
[102]	Normalization, tokenization, stop words removal
[103]	Negation detection
[104]	Text cleaning, stemming, stop words removal, negation detection
[105]	Segmentation, stemming
[106]	Text cleaning, POS tagging, named entity recognition
[107]	Text cleaning, normalization, tokenization, stop words removal

TABLE 4: The public datasets in the reviewed articles.

Key ID	Dataset (DS)	DS size	Positive	Negative	Neutral	Source
D01	Arabic Health Services (AHS)	2,026	628	1,398		Twitter
D02	Arabic Dataset	22,550	8,529	7,021	7,000	Twitter
D03	ArTwitter	1,951	993	958		Twitter
D04	ATB	1,834				Newswire
D05	ASTD	10,006	799	1,684	7,523	Twitter
D06	BHA	8,224	4,112	4,112		TripAdvisor.com., booking.com, agoda.ae
D07	BRAD	510,598				Goodreads
D08	CNN	5,070				
D09	LABR	51,056	42,832	8,224		Goodreads
D10	ACOM (DS1)	594	184	284	106	Aljazeera's site
D11	ACOM (DS2)	1,492	486	517	391	Aljazeera's site
D12	ACOM (DS3)	611	149	462		Aljazeera's site
D13	HARD	94,052	46,968	47,084		Booking.com
D14	PatientJo	1,228	227	951		Jordanian hospitals
D15	TA	2,000	1,000	1,000		Twitter
D16	OCA	500	250	250		Arabic movies reviews
D17	QCRI	754	377	377		Twitter
D18	RR	2,817	876	1,941		Twitter

TABLE 4: Continued.

Key ID	Dataset (DS)	DS size	Positive	Negative	Neutral	Source
D19	SemEval-2016	15,562				Booki006Eg.com, TripAdvisor.com
D20	SemEval-2017	3,355	743	1,142	1,470	Twitter
D21	SIAAC	147	32	91	24	Echorouk newspaper
D22	SWN	30,828	14,543	16,285		
D23	Syria tweets	2,000	448	1,350	202	Twitter
D24	TEAD	5,615,943	3,122,615	2,115,325	378,003	Twitter
D25	AraSenti tweet	10,133	4,329	5,804		Twitter

TABLE 5: Overview of the data extracted from related articles in ASA.

Ref	SA task	Ap	Algorithm	SA level	DS/size	Domain	Language	Features	Accuracy (%)
[17]	AD	Un	Syntax-based approach	Aspect	15000, 15000	Hotel, products	MSA, DA		65.32
[28]	BR&SC	Su	SVM	Aspect	5K	Airline	MSA, DA	Skip-gram, FastText	89
[34]	BR&SC	Su	SVM		363		MSA, DA		76.09
[25]	BR&SC	Su	KNN, NB, SVM		3015	Food, sports, weather	MSA, DA		76.33
[109]	SC	Su	Bagging, NB, KNN, DT		<u>D10, D11, D12</u>		MSA, DA	SMOTE	<u>74.37; 74.9, 84.02</u>
[40]	BR&SC	LB	Corpus and LB approach	Word	15,274	13 domains	MSA, DA, CA	TF-IDF, N-gram	93
[43]	BR&SC	Su	SVM	Sentence	D23, D05, D03, D17, D20		MSA, DA	TF-IDF, LSA, CBOW, SG	83.02
[110]	BR&SC	Su	DCNN	Sentence	2390	News	MSA	Word2ve, N-gram	69.9
[44]	BR&SC	Su	SVM, LGR		183531	Products	MSA	BoW, TF-IDF	Bal:77.76 Unbal:91.21
[16]	BR&SC	Su	NB, SVM, RFT		10254	Elections	MSA, DA	N-gram, TF, TF-IDF	77
[49]	BR&SC	Su	SVM, KNN, BNB		5986		MSA, DA	N-gram, TF, TF-IDF	F-m = 88.8
[45]	SC	Su	RFT, GNB, LGR, SGD		D09, <u>D05</u>		MSA, DA	CBOW, SG	87.10
[53]	SC	Hb	Bagging, SVM, RFT, NB		<u>D10, D11</u>		MSA, DA	Unigram, bigram	<u>90.4; 90</u>
[55]	SC	LB	SVM, ME, Bagging, Boosting, RFT, NNET, DT, NB		D16, D03, lex = 5376				95.98
[111]	SC	Su	NB, NB-MLP		2154; 13420; 1353; 3962; 8522	Attraction, hotel, movies, products, restaurants	MSA, DA		99.8; 85.1; 95.4; 97.3; 93.1
[50]	BR&SC	Su	NB, SVM		48	Stock market	MSA, DA	N-gram, TF-IDF, BTO	83.58
[51]	BR&SC	Hb	LB, LSTM, CNN, SVM, LGR, NB, DT, RFT		<u>D24, D05</u>		MSA, DA	TF-IDF, CBOW	87.5; 81
[57]	Subj C	Su	DL		1100		MSA, DA	TF-IDF	92.96
[65]	SC	Su	NB, SVM		D09	Books	MSA, DA	TF-IDF, N-gram	90.98
[67]	BR&SC	Su	NB, SVM			News for multidomains	MSA, DA	Unigram, bigram, TF, TF-IDF	84,56
[112]	SC	Su	CNN-LSTM	Character, Ch5Gram, Word	D01, D03, D05	Health	MSA	Trigram, ReLU	<u>95.68; 77.62, 94.24; 88.10</u>
[69]	BR&SC	Hb	SO-NB	Sentence	1200		MSA	N-gram models, semantic features	90

TABLE 5: Continued.

Ref	SA task	Ap	Algorithm	SA level	DS/size	Domain	Language	Features	Accuracy (%)
[79]	BR&SC	Hb	LGR, PAG, SVM, PRN, RFT, ABT, LB	Review	D13	Hotel	MSA, DA	Unigram, bigram, semantic features, BOW	94 to 97
[113]	Aspect SA	Su	NB, BYN, DT, KNN, SVM	Sentence	D19	Hotel	MSA	Morphological, syntactic, semantic features	95.4
[56]	BR&SC	Su	NB, SVM, DT, RFT		1543347; 1462	Politics	MSA	Syntactic, surface-form, sentiment features	71.95
[47]	BR&SC	Su	NB	Sentence	18278	Elections	MSA, DA	<u>Unigram</u> , bigram, <u>IG</u>	93.13
[59]	BR&SC	Su	SVM, NB	Document	D21	News, politics, sports, culture	MSA, DA	Unigram, bigram	90.20
[46]	BR&SC	Hb	SO, SVM, NB	Sentence	1520		DA	N-gram, CountVectorizer Glove, SG, CBOW,	92.98
[114]	SC	Su	CNN	Sentence	<u>D05</u> , <u>D20</u>		MSA, DA	random word vectors	<u>72.1463</u>
[60]	BR&SC	Su	SVM, NB, KNN	Sentence	996	Multidomain	MSA, DA	Unigram, <u>bigram</u> , <u>BTO</u> , TF-IDF	78
[73]	SC	Su	DT	Document	<u>D16</u> , <u>D06</u>	Movies, hotel	MSA, DA		<u>93.83</u> ; <u>90.63</u>
[80]	Subj C, &SC	Hb	DT, NB, KNN, <u>Ontology Baseline</u> , LB	Aspect	D09, 2000	<u>Books</u> , <u>Hotel</u>	MSA	Domain features	<i>f</i> -m = <u>79.18</u> ; 78.83
[81]	BR&SC	Hb	SVM, NB, KNN	Sentence	3476	Movies, economy, sports, history, politics	MSA, DA		97.44
[61]	BR&SC	Su	SVM, KNN, NB, DT, <u>LEM2</u>	Document	4812		DA	Rough set method	74
[66]	BR&SC	Su	RNTN	Phrase, sentence	1177		MSA	Orthographic, morphological features	80
[35]	SC	Su	SVM, NB, <u>MNB</u> , SGD, DT	Review	D09	Books	MSA, DA	<u>Unigram</u> , bigram, <u>TF-IDF</u> , <u>genetic algorithm</u>	94
[32]	BR&SC	Un	LB	Document	1000, 1000	News, arts	DA		73–96
[15]	BR&SC	Su	SVM	Sentence, document	D19	Hotel	MSA	Unigram	76.42
[83]	BR&SC	Hb	KNN, SVM, <u>LLR</u> , NB, NEUNET	Document	8861	15 domains	DA	Polarity scores, dialects, synsets, inflected forms	97.8
[84]	Aspect SA	Un	LB	Review	200		MSA	Sentiment features	92.15
[86]	BR&SC	Hb	SVM, NB, LB	Sentence	64342		DA	N-gram	70
[115]	SC	Su	<u>Stacking</u> , SGD, RFT, LGR, GNB	Sentence	1350		DA	Word2vec, <u>SMOTE</u>	85.28
[54]	SC	Su	SVM, NB	Sentence	9096		DA	<u>BOW</u> , bigram, trigram	<i>f</i> -m = 73
[68]	BR&SC	Su	SVM, NB	Sentence	1800		MSA, DA	N-gram models, TF-IDF	88.72
[88]	SC	Su	MLP, <u>LGR</u>	Sentence	D09		MSA, DA	Doc2vec	32.38
[90]	SC	Su	CNN, <u>LSTM</u> , RNN	Document	<u>D05</u> , <u>D03</u>		DA	<u>CBOW</u> , SG, <u>mul</u> , <u>CONC</u>	<u>81.63</u> ; <u>87.27</u>
[82]	SC	Su	SVM, DT, NB, KNN, <u>HCS</u>	Document	D09		MSA, DA	BOW, correlation analysis	72.64
[48]	BR&SC	Su	DT, RFT, SVM	Sentence	10254	Politics	MSA, DA	N-gram models, TF-IDF, TF	81

TABLE 5: Continued.

Ref	SA task	Ap	Algorithm	SA level	DS/size	Domain	Language	Features	Accuracy (%)
[91]	SC	Su	DNN (LSTM), RNN	Sentence	D09		MSA, DA	Word embedding, BOW	71
[14]	SC	Su	RNN, <u>SVM</u>	Sentence	D19		MSA	word2vec, lexical, morphological, semantic features, N-gram	95.4
[116]	Subj C&SC	Su	SVM	Sentence	D04		MSA		69.37
[104]	BR&SC	Su	<u>SVM</u> , NB	Sentence	1121	Education	MSA	N-gram	84.62
[71]	SC	Su	<u>SVM</u> , NB	Review	D16	Movies	MSA	N-gram models, <u>TF-IDF</u> , BTO, <u>TF</u>	96.67
[29]	SC	Su	SVM, <u>LGD</u>	Review	<u>Unbal</u> , bal (D09)	Books	MSA, DA	BOW, N-gram,	<u>88.51</u> , 78.14
[30]	BR&SC	Su	LGR, PAG, <u>SVM</u> , PRN	Review	<u>Unbal</u> , bale (D07)	Books	MSA, DA	Unigram, bigram	0.744–0.911; 0.847–0.85
[13]	BR&SC	Su	<u>SVM</u> , NB	Document	1331	Products	MSA, DA	N-gram, BTO; <u>TF-IDF</u>	89.68
[89]	SC	Su	<u>SVM</u> .NB, KNN, ANN	Sentence	500	Hotel	MSA		$f$ -M = 92
[36]	BR&SC, Subj C	Su	<u>SVM</u> , NB	Sentence	3700		MSA, DA	N-gram models	89.55
[58]	SC	Su	AR		D16	Movies	MSA, DA	IG, <u>chi-square</u> , GI Position, bigram,	86.81
[93]	BR&SC	Su	Lexical semantic, <u>CRF</u>	Sentence	381	News	MSA	trigram, morphological	$f$ -m = 84.93
[94]	BR&SC	Su	<u>SVM</u> , NB, KNN	Document	250	Products	MSA, DA	N-gram models	94
[95]	BR&SC	Su	SVM, <u>MNB</u>	Sentence	134194		MSA, DA	<i>Automatic labeling</i>	75.7
[117]	SC	Su	Rule-based approach	Sentence	D16, <u>2000</u>	Movies, arts, <u>politics</u>	MSA, DA		85.6, 93.9
[97]	BR&SC, Subj C	Su	SVM, NB, <u>DT</u> , RFT	Sentence		Politics	MSA	N-gram, TF-IDF	$P$ = 70.97
[98]	SC	Su	SVM, KNN, <u>NB</u> , DT, RFT	Sentence		Terrorism	MSA	Lexical, surface- form, N-gram	$P$ = 71.76
[99]	SC	Su	SVM	Aspect	Restaurant reviews, <u>1000</u> , D16,	Novels, products, movies, sports, hotels, restaurants	MSA	N-gram	85.35 <u>90.60</u> <u>96.00</u>
[100]	BR&SC	Su	<u>SVM</u> , BPNN, NB, DT	Sentence	2000		MSA, DA		96.06
[62]	BR&SC	Su	SVM, <u>NB</u>	Sentence	18278	Politics	DA	N-gram, TF-IDF	$P$ = 92.5
[118]	SC	Su	SVM, <u>CNB</u> , MNB	Sentence	D10, D11, D12		MSA	N-gram, TF, <u>IDF</u> , TF-IDF, IG	77.34
[37]	BR&SC	Hb	SO-SVM-NB	Sentence	4800	News	DA	N-gram models	80.9
[92]	SC	Su	<u>SVM</u> , NB, KNN	Sentence	3073	Politics, arts	MSA, DA	N-gram	68.69
[74]	BR&SC	Su	<u>SVM</u> , NB, KNN	Sentence	2,591	Education, sports, politics	MSA	BTO, TF-IDF, <u>TF</u>	69.97
[63]	BR&SC	LB	LB	Sentence, document	D09, D14	Health, book	MSA, DA		71
[38]	BR&SC	Su	NB, SVM, <u>ME</u>	Document	28576		DA	BoW, <u>N-gram</u> <u>models</u>	86.75
[41]	SC	Su	SVM, NB, <u>DT</u>	Sentence	2000	Politics, arts	MSA		$P$ = 80.9
[101]	BR&SC	Hb	LB, NB	Sentence	2590	Restaurants	DA	Sentiment words	90.54
[102]	BR&SC	Su	<u>NB</u> , DT	Sentence	2000		MSA		64.85
[33]	BR,&SC, Subj C	Su	RFT, GNB, <u>SVM</u> , LGR, SGD	Word	D05, D03, D17, D09	Quran, customers	MSA, DA	CBOW	80.21–81.69
[119]	SC	LB	Rule-based, LB		D16		MSA		89.6



TABLE 5: Continued.

Ref	SA task	Ap	Algorithm	SA level	DS/size	Domain	Language	Features	Accuracy (%)
[18]	BR&SC	Su	CRF, DT, NB, KNN	Sentence	2265	News	MSA	NER, N-gram	86.5
[42]	BR&SC	Hb	SO, SVM	Sentence	1103		MSA, DA	N-gram models	84.01
[103]	BR&SC	LB	LB	Word	<u>D25</u> , D05, D18, Lex: <u>D25-PMI</u>		DA		89.58
[87]	BR&SC	Su	SVM	Sentence	625	Multidomain	MSA, DA		83.5
[96]	BR&SC	Su	<u>Bagging</u> , boosting	Sentence	1500	Sports, news, Economics	MSA, DA		85.95
[72]	BR&SC	Su	SVM, <u>NB</u>	Document	D02		MSA, DA		M-P = 90.5
[31]	BR&SC	Su	SVM-KNN	Review	625	Hotel	DA		97
[105]	SC	Su	DT, Dtable, SVM, MNB, voting (KNN, DT, and NB)	Sentence	Unbal, bal: D09		MSA	BOW	42.7–46.4
[120]	BR,&SC, Subj C	Su	<u>SVM</u> , MNB, BNB, PAG, SGD, LGR, PRN, KNN	Sentence	D05		DA	N-gram models	69.1
[121]	BR&SC	Su	Voting (RSS and SVM)	Sentence	800	News	MSA, DA		98
[64]	SC	Su	<u>Voting (ME, SVM, and ANN)</u> , bagging, boosting, stacking	Document	D16, D11-D13		MSA	Stylistic, morpholexical, tigram, bigram, TF-IDF, TF, BTO	F1 = 85.06
[122]	BR&SC	Ss	<u>SVM</u> , LGR, BNB, KNN, SGD	Document	D09	Movies, hotels, restaurants, products	MSA	Unigram, bigram, TF-IDF, word count	82.4
[106]	BR&SC	Su	SVM, NB, <u>KNN</u>	Sentence	500		MSA		<i>f</i> -m = 91.5
[70]	BR&SC	Su	NB, KNN	Document	2591	Education, sports, politics	MSA, DA	Bigram	M-P = 75.25
[26]	BR&SC	Su	SVM, <u>NB</u> , KNN	Sentence	D16, 164	Politics	MSA, DA	Correlation, N-gram models	96.6
[107]	BR&SC	Su	<u>NB</u> , KNN	Sentence	300		MSA		63.79
[75]	BR&SC	Su	SVM, NB	Sentence	300, 250		DA	N-gram models	0.75.64
[123]	SC	Su	<u>SVM</u> , NB	Document	5070	News, entertainment, sports, science, business	MSA	Chi-square, correlation, GSS coefficient, IG, relief F	M-F = 95.1
[76]	BR&SC	Su	NB, <u>voting</u>	Document	4812		DA	Rough set method, <u>genetic</u>	57
[52]	BR&SC	Su	SVM, NB	Document	D22, D16, 7400		MSA	BTO, TF, TF-IDF, score	94.88–97.81
[124]	BR&SC	Ss	Pattern matching, <u>majority with entities</u>	Sentence	5000	Restaurants	DA		60.5
[24]	Subj C, and SC	Su	<u>SVM</u> , NB		<i>Obama</i> , <i>Messi</i> , iPhone, shia	Politics, social, products	MSA	N-gram, <i>IDF</i> , <u>TF-IDF</u>	68.05; 87.43
[85]	BR&SC	Su	<u>SVM</u> , MNB	Sentence	260		DA	N-gram	72.78
[39]	BR&SC	Su	SVM	Sentence	1350	News	DA		<i>P</i> = 88.63
[77]	SC	Hb	Voting (SVM, NB, and SO)	Sentence	D16, D22		MSA	N-gram models	F1 = 90.74
[78]	SC	Su	<u>NB</u> , SVM, <u>KNN</u>	Document	<u>D16</u> , <i>D10</i> , <i>D11</i>		MSA	TF-IDF, TF, IDF, N-gram	<u>93.0</u> , 78.2

TABLE 5: Continued.

Ref	SA task	Ap	Algorithm	SA level	DS/size	Domain	Language	Features	Accuracy (%)
[125]	SC	Su	MNB, <i>BNB</i> , <u>SVM</u>	Sentence	<u>Unbal</u> , bal: D09		MSA, DA		<u>91.0</u> , 82.7
[27]	BR&SC	Su	<u>SVM</u> , NB	Sentence	<u>Unbal</u> , bal: 4625	Arts, politics, science, technology, social	MSA, DA		68.2, 61.4
[12]	BR&SC	Su	KNN, <u>NB</u> , DT, <u>SVM</u>	Sentence	1000		MSA	<u>Social mention</u> , <i>Senti. strength</i>	<u>95.59</u> , 93.29

with DS2 dataset producing the accuracy of 74.9%. Similarly, underlining is used to say that algorithm of Decision Tree (DT) with dataset DS1 producing the accuracy of 74.37%. In the same vein, italic is utilized to denote that KNN with dataset DS3 and feature technique of SMOTE produced the accuracy of 84.02%.

In addition, the precedent category of datasets is utilized in Table 5; otherwise, the dataset size was presented if the dataset was built in-house. Also, SA approaches were contracted in the third column of Table 5 to give more details of the reviewed articles. These approaches involved supervised (Su), unsupervised (Un), hybrid (Hb), lexicon-based (LB), and semisupervised (Ss).

### 3. Findings

This section introduces the findings of SLR and contributes to answer the two first research questions. The sections below are relevant to the obtained results from SLR. ASA has been needed since Arabic audience, who use Internet and applications, has recently gone up a lot [9]. A total of 191 papers were found during the period from January 2013 until end of November 2018. It is clearly noticed from Figure 3 that the number of research on ASA has increased gradually in the last five years; i.e., it reached 53 articles in 2018. It is evident that ASA is recently an absolutely hot research topic. Therefore, it has witnessed a speedy growth and an increased interest from researchers. Thus, it is a fascinating research field.

In addition, Figure 4 provides information about the most tasks used in ASA. According to the reviewed studies, most tasks are broken down into sentiment classification, building resource, and building resource and sentiment classification. Moreover, it is apparent from Figure 4 that the task of building resource and sentiment classification dominated over all other tasks. Indeed, Arabic language still lacks resources and tools that can be employed to support classifying sentiments.

As known, Arabic language is usually varied in MSA and DA. Figure 5 illustrates the number of articles targeting Arabic language types. Overall, it is worthy to note that studying both MSA and DA is more trend rather than studying only MSA. It is due to the abundant presence of DA in social media and microblogging channels.

Furthermore, the sources of datasets for opinions on any topic vary from social media platforms to websites that introduce products or services. Figure 6 depicts the datasets sources used in ASA.

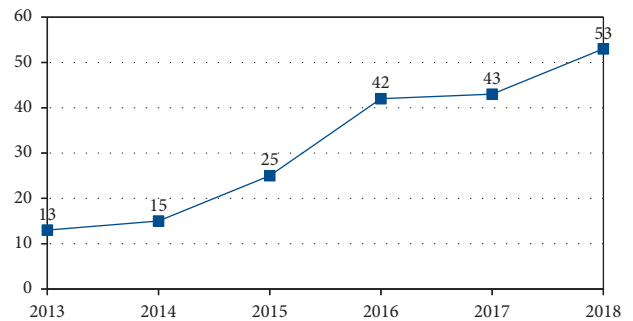


FIGURE 3: Distribution of ASA reviewed articles over years.

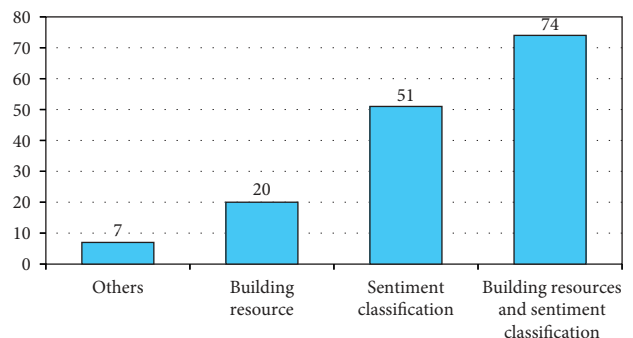


FIGURE 4: Numbers of articles targeting SA tasks.

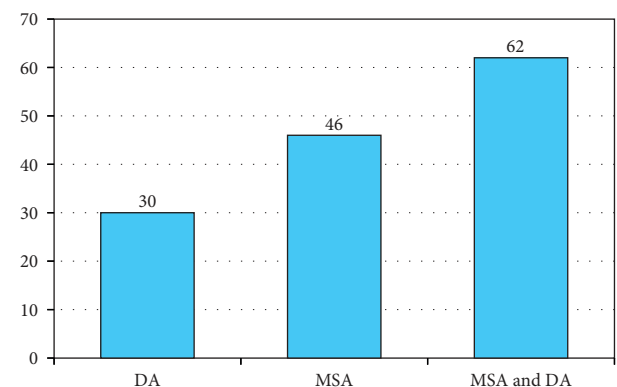


FIGURE 5: Arabic language types addressed in the articles.

It is clearly noticed that Twitter, i.e., 50% of the data sources used in articles, is the most frequently application of social media used in the reviewed articles. It has a great potential of exploring people lives and their potentials,

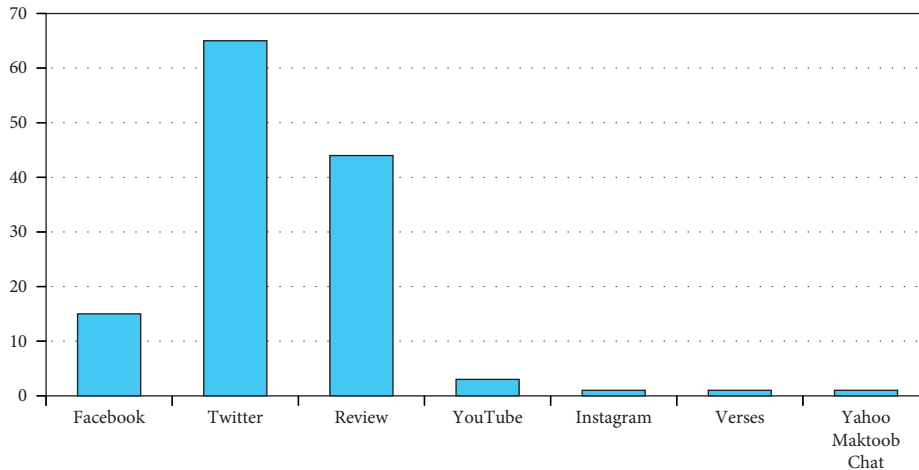


FIGURE 6: The sources of dataset used in Sentiment Analysis.

opinions, and interests. It is restricted to a very short message called tweets which are often written using a lot of Arabic slang.

In general, machine learning- (ML-) based approaches for ASA involves the following stages, including data preprocessing, feature generation, and selection, and ML methods. In literature, there are several techniques proposed in every stage of SA to improve the performance better. Arabic research performed in each of these stages is addressed in the following sections.

**3.1. Preprocessing Arabic Text.** Arabic is a rich language that is a challenging linguistic domain for NLP. It has morphological complexities and dialectal varieties which require advanced preprocessing [126, 127]. Another complexity is expressing the sentiment and feeling in dialect rather than MSA [126]. Preprocessing and analysis of Arabic raw texts extremely reduce the noise and improve the efficiency. Unfortunately, most studies concentrated on preprocessing English text; nevertheless, few studies focused on Arabic text [69].

Figure 7 shows the information about the most preprocessing strategies used in ASA.

In general, it is clearly seen from Figure 7 that the most studies went through the preprocessing phase, including text cleaning, normalization, tokenization, stop words removal, and stemming. It is found that 52% of the articles considered stemming as an important step. Meanwhile, text cleaning, normalization, stop word removal, and tokenization were used in 30%, 47%, 44%, and 45%, respectively.

**3.2. Common Features in Arabic Sentiment Analysis.** Machine learning presents several algorithms for sentiment classification. Nevertheless, the challenge of capturing sentiment in the written context is to select the best features to be utilized [9]. Features provide comprehensive summarization of the outcomes with more precise analysis of the sentiments [128]. Figure 8 reveals the most frequent features that are used in ASA. It shows that, per the features used in

the reviewed articles, n-gram models are the highest and semantic and lexical features are the lowest.

**3.3. Methods Used in Arabic Sentiment Analysis.** Reviewed studies have introduced a wide set of methods and techniques to solve the ASA problem. Figure 9 illustrates the most methods used in ASA.

It is clearly noticed that SVM and NB are the highly used methods in articles, while voting, boosting, and semantic orientation (SO) are the lowest. SVM has been adopted in several previous sentiment classification works resulting in 74 papers out of 108 papers, while NB was used in 71 papers. It is worthy to note that applying SVM classifier in the previous studies has been superior or comparable to others classifiers, such as NB.

**3.4. Arabic Sentiment Analysis Applications.** In recent years, ASA has gained considerable attention, and its applications have spread to almost every possible domain. Figure 10 shows the most domains were targeted to apply sentiment classification. It is clearly observed from Figure 10 that many researchers inclined to apply SA in business and economy and social and politics domains numbering in 40 and 39 articles, respectively. On the contrary, the lowest domains addressed in ASA were education, health, and travel, and tourism numbering in 3 studies for each domain.

Most studies focused on ASA applications in a limited set of domains, such as politics [15, 48, 62, 89], hotel [79, 113], business and economy [12, 20, 129], arts and books [29, 32, 92], entertainment and movies [71, 73, 99], and sport [81, 96]

Several papers were published [12, 74, 83] to study ASA for several purposes such as building Arabic senti-lexicon, designing a framework for ASA, and comparing two free online SA tools that support Arabic. These studies involved collecting small datasets with size less than 3000 tweets that are relevant to several domains such as education, sports, and politics. In [83], little tweets were dedicated separately to each domain, including education. Nevertheless, the

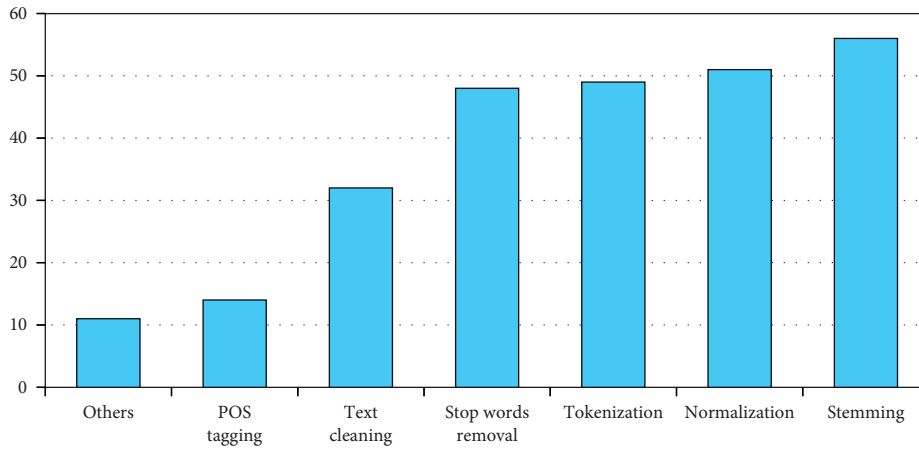


FIGURE 7: The most preprocessing strategies used in ASA.

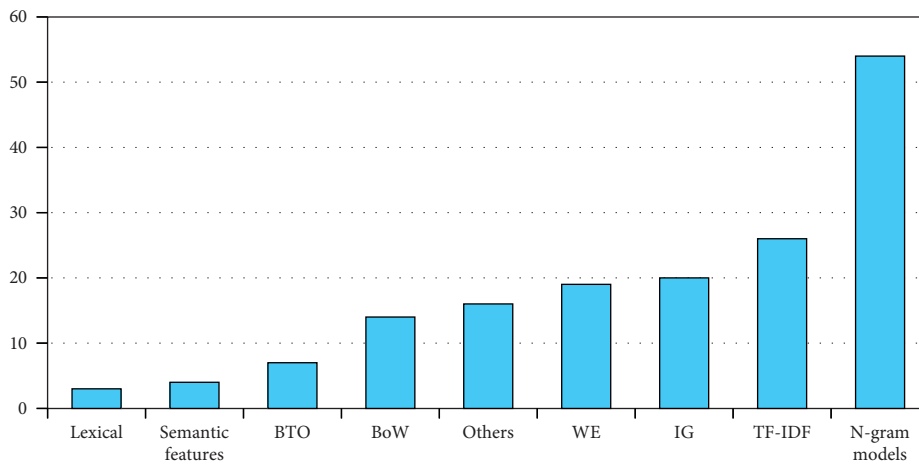


FIGURE 8: The most frequent features used in ASA.

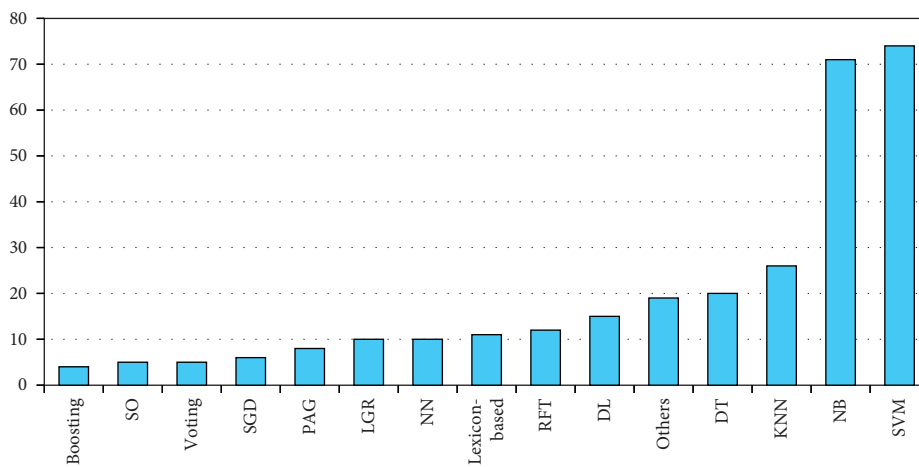


FIGURE 9: The most methods used in Arabic sentiment classification.

drawback of this study was that results for each domain was neglected and did not show up [83]. In addition, Al-Horaib et al. [102] studied ASA in e-learning using traditional ML

algorithms, such as SVM and NB. However, the volume of dataset was 2000 tweets and related only to King Abdul-Aziz University.

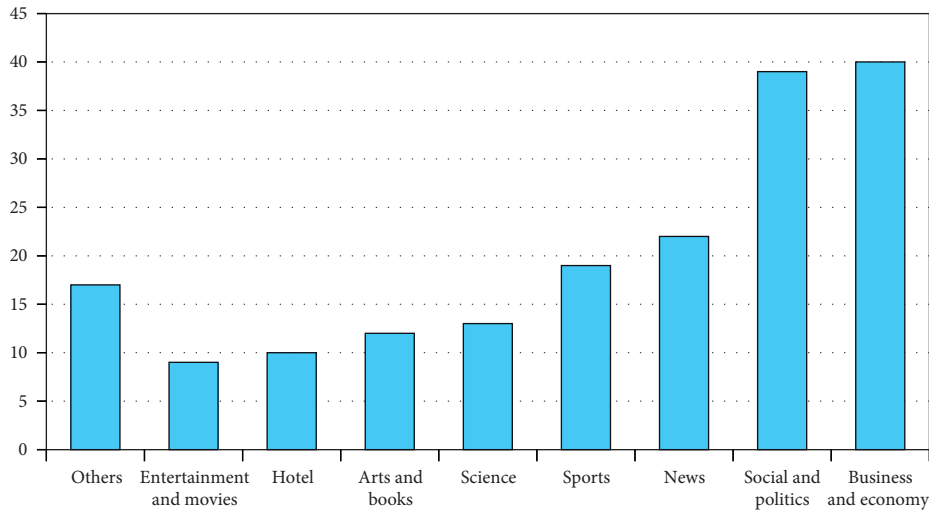


FIGURE 10: The most targeted domains in Arabic sentiment classification.

In conclusion, little work has looked at using ASA to classify the sentiments in the education domain. Nevertheless, there was no intended orientation from researchers to study the domain in particular. Moreover, the collected data were small, and the results of classification, that was related to the education domain, were not particularly highlighted and clarified in the discussion.

**3.5. Taxonomy of Arabic Sentiment Analysis Methods.** To fulfill ASA, several methods of sentiment classification were proposed. An enhanced taxonomy of Arabic sentiment classification methods adopted from [130, 131] was proposed in Figure 11.

This enhanced taxonomy on classification of the SA methods, which are reviewed and discussed previously, is presented in Figure 11. As shown in the figure, the applied methods can be classified based on the approaches using machine learning. ML approaches can be supervised, unsupervised, or semisupervised. The supervised ML approaches can be divided into probabilistic and non-probabilistic. On the one hand, NB, maximum entropy (ME), conditional random field (CRF), Bayesian network (BN), logistic regression (LR) are examples of the probabilistic classification method. On the other hand, SVM, KNN, DT, rule-based, and NN, including deep learning and the traditional NN can be classified as non-probabilistic classification methods. The unsupervised ML approach, such as genetic and clustering algorithms, is exploited when it is hard to find labeled documents. All ensemble approaches, such as Random forest, voting, bagging, boosting, and stacking, can be categorized as semisupervised approaches. Moreover, ASA methods differ in the point of using a lexicon. These methods can be based on dictionary, corpus, or ontology. In addition, the applied methods can be a hybrid of ML and semantic orientation approaches. The modification done in Figure 11 involves adding some methods, such as DL, traditional neural network, genetic algorithm, clustering, LR,

CRF, all ensemble methods, and ontology-based approaches.

#### 4. Discussion and Future Research Avenues

This part sought to discuss the obtained results from SLR and to give the answer to the third and fourth research questions. A total of 108 articles have been reviewed in ASA to capture the current state and achieve the research aims. The aims were summarizing the most effective techniques used in ASA, revealing the gaps and limitations in the reviewed studies, and highlighting the directions of future research on ASA.

It is obvious that ASA has been studied from three important perspectives. The first is the preprocessing strategies that strongly impact upon the obtained results of SA classification. The second is the process of feature generation and selection that play a significant role to build the vectors and, accordingly, improve the results. The last process is the classification method which in role receives the vectors outputted from features generation to classify the sentiments.

As shown in the literature, there are still some challenges in ASA that have to be addressed. These challenges varied into preprocessing strategies, feature selection, classification method, and the targeted domain.

Arabic sentences greatly involve noisy, missing, and inconsistent data that need to be preprocessed to improve Arabic sentiment classification. Not using preprocessing, such as eliminating insignificant comments and repetition of letters, may lead to ignore important words. Applying a wide set of preprocessing strategies, such as normalization, tokenization, word removal, and stemming, will enhance the sentiment classification.

Al-Rubaiee et al. [13] explored the preprocessing steps within RapidMiner; normalization, tokenization, stop word removal, and stemming. It was demonstrated that text preprocessing is a key factor in sentiment classification and shows different levels of accuracy by creating N-grams term of tokens.

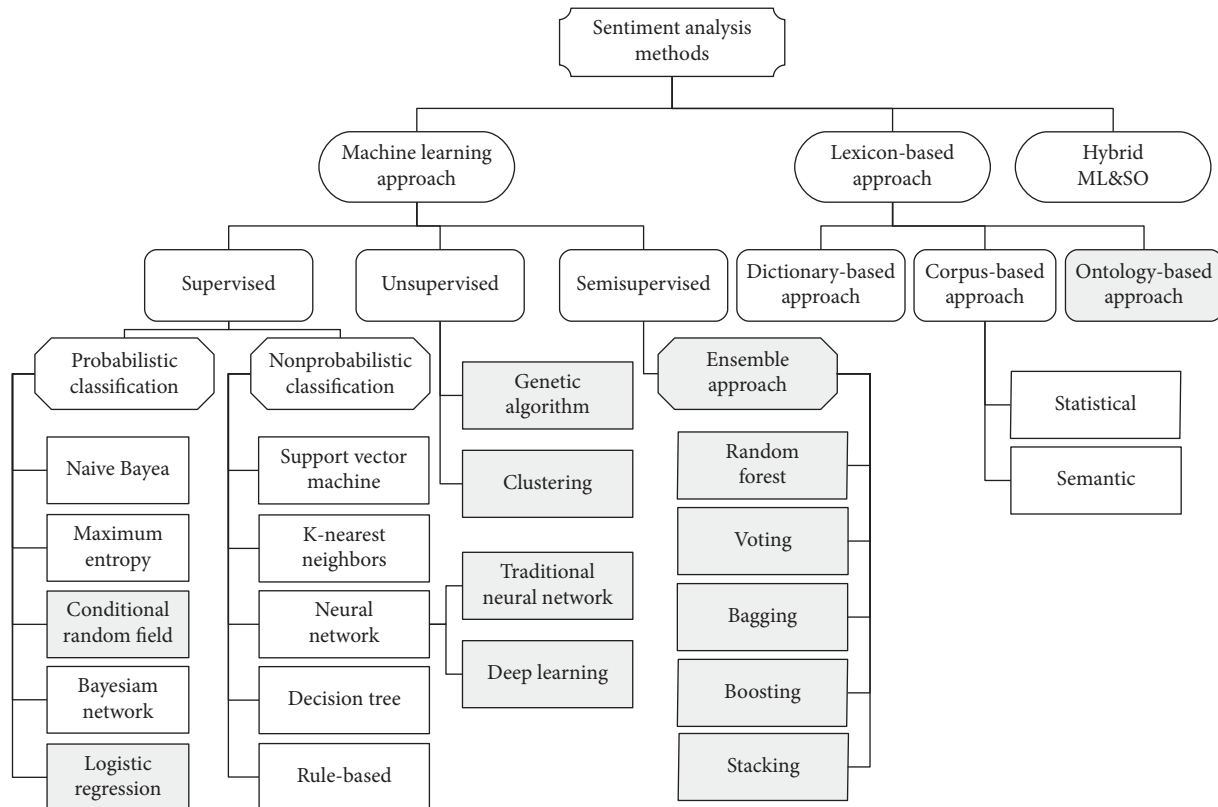


FIGURE 11: A new taxonomy of ASA methods.

The preprocessing effect on ASA was addressed, especially on Egyptian presidential elections in 2012. It resulted that employing information gain for selecting features with N-gram, stemming, and normalization improved the accuracy of Arabic text classification [47].

In addition, Alomari et al. [68] investigated several preprocessing strategies, including stemming, stop words, N-grams, and different weighting schemes employing several scenarios for each. In details, N-gram models were employed with both stemming types and without stemming as well as with weighting schemes, including TF-IDF or TF. An Arabic stop word removal was excluded on the ground that it reduces performance in all scenarios. Moreover, the experimental results indicated that SVM classifier using TF-IDF with stemming through bigrams feature outperformed the best scenario performance resulting from the Naive Bayesian classifier.

The authors of [92] studied the impact of the stemming on the Arabic sentiment classification problem. They performed experiments on two datasets: “2000 tweets” collected by Abdulla et al. [132] and “BBN tweets” by Salameh et al. [133]. They implemented Arabic root stemmer of Khoja and light stemmer that are integrated in RapidMiner. The results showed that light stemming is preferable than the root stemming and trigrams of characters combined with tokens of the text given; therefore, the best results for sentiment classification were obtained.

Duwairi and El-Orfali [26] performed a study on SA for Arabic text. The authors used two datasets: one prepared in-house related to politics domain and the other prepared by

Rushdi-Saleh et al. [8] related to movie domain. Their objective was to investigate data representations and preprocessing strategies on ASA. The results showed that using stemming and light stemming combined with stop words removal adversely affected the classification performance of the dataset related to movie domain, while it improved slightly the classification for the other datasets.

Even though the preprocessing phase is a significant step in SA for text mining, it is still underestimated and not extensively covered in the literature. In addition, the best preprocessing techniques, that play a decisive role and effective in improving ASA, are still an open field to study and experiment.

A good preprocessing leads to select a suitable feature. Feature representation includes a semantic representation that is still a challenging task in NLP. Thus, capturing word semantic is possible with distributional semantic models. Merging word embedding with combination of N-gram models will improve the results of SA.

In ASA, N-grams models have been largely used as features. Some studies exposed that unigrams resulted in a better performance than bigrams and trigrams [35, 54]. This behavior was due to the fact that BOW can give a good data coverage, whereas bigrams and trigrams tend to be very sparse.

The features, such as count vector of unigrams, bigrams, and trigrams, were experimented separately. Diverse combinations of several N-gram models were attempted and resulted that combinations improved the classification process [37, 69, 94].

Alomari et al. [68] examined the use of several N-grams (unigram, bigrams, and trigrams) with various weighting schemes, including TF-IDF and TF, and found that the bigrams model with TF-IDF weighting scheme outperformed others.

Also, POS tagging feature has been utilized in analyzing Arabic textual contents. For example, Al-Moslmi et al. and Mohammad et al. [18, 83] exploited POS tagging features, including nouns, adverbs, and adjectives to investigate sentiments in Arabic text. Alhazmi et al. [106] carried out two sets of experiments, with POS and without POS, to assess the POS pattern effectiveness as features in sentiment classification. The experiments showed that using POS patterns did not make big improvements and that might be due to the fact that Arabic dialect is commonly used in twitter.

TF-IDF and BTO were widely used as a weighting scheme to create the word vector [13, 60, 71, 73]. The performance was comparable. It relied on the word vector model generated and supervised the ML algorithm used.

Moreover, feature representation that includes a semantic representation is still a challenging task in NLP. Thus, capturing words semantic is possible with distributional semantic models which involve mainly word embedding. Word embedding is an alternative approach for such hand-crafted features in ASA. Several recent studies have exploited this technique [14, 33, 91, 134].

It is noticed that using the word embedding with DL models helped improving the results over the linear models, such as SVM, as it is suitable for large datasets and can be computationally efficient [33, 90].

There are many methods that have been proposed to deal with the Arabic sentiment classification problem. However, the accuracy of these methods is varied due to the dimensionality, large datasets, and features. Employing the DL model in ASA will participate greatly to solve problems implicated in other common methods, such as SVM and NB.

Several studies have employed SVM and NB together to investigate the Arabic sentiment classification problem. In many of these studies, it is noticed that there is a strong competition in achieving higher accuracy of SVM and NB. There are 22 studies in which NB accuracy outperforms SVM [52, 62, 83, 106, 118]. NB has the ability to classify sentiments using a small training set. It utilizes the statistics to accomplish probabilities classification and is very effective in classifying documents. Basically, its purpose is to analyze the absence and presence of a specific feature using probabilities to classify features independently. It was highly effective when dealing with words that have the probabilities to be sentiment or not, such as adverbs or adjectives [126].

In contrast, SVM has been successfully used for general classification as well as regression, and it has proven its effectiveness in Arabic sentiment classification. It has the ability to model several sources of data, the highest obtained accuracy, and flexibility in handling high-dimensionality data. Moreover, to avoid incorrect classification, it utilizes a larger margin. Therefore, the performance of SMV outperformed NB performance in 29 studies [13, 27, 36, 59, 71, 102].

In [123], several experiments were conducted using SVM and NB on different feature sizes to examine the performance of frequently used features selection techniques. It was noticed that the accuracy of SVM and NB decreased when the features size increased. Consequently, SVMs were superior for relatively small datasets and features with fewer outliers.

Deep neural network has been successfully adopted to extract features. It has a big advantage over other ML methods. Deep contextual features about words are extracted in a lower dimensional space rather than requiring any features engineering for learning continuous text representation from data. Furthermore, DL models are the most adequate with very large datasets, large number of features, and complex classification tasks. Consequently, the DL model is a promising way to solve Arabic sentiment classification.

Recently, many research studies have significantly exploited DL in SA, as depicted in Figure 12, according to a survey that we have conducted. The line graph shows a steady increase over a six-year period. It is clearly observed from the line graph that the numbers of articles that implemented DL in SA soared slightly at the first and climbed steeply in the last two years. Nevertheless, implementing DL in ASA has received little efforts.

It is noticed that out of 209 articles is only six Arabic articles utilizing DL.

Alayba et al. [112] integrated convolutional neural network (CNN) and long short-term memory (LSTM) methods to investigate their benefits on ASA. As a result, the obtained accuracy for ASA was improved on several datasets. In addition, DL method for ASA was presented in [90]. The authors investigated several combinations of skip-gram and CBOW, including CNN and LSTM evaluated on two publicly available datasets. Using the combined LSTMs introduced, the highest results in terms of accuracy and other performance measures. Al-Azani et al. [135] used LSTM and its simplified variant gated recurrent unit (GRU) to detect sentiment polarity of Arabic microblogs. They compared the performance of DL to baseline traditional machine learning methods. The results showed that models based on LSTM and GRU outperformed other classifiers.

In general, using a combination of DL models for ASA is a promising alternative to traditional machine learning techniques, and it assists to increase the accuracy. The main idea of DL techniques is to use the deep neural network algorithms to learn complex features extracted from large raw data without relying on prior knowledge of predictors. These algorithms automatically learn new complex features instead of passing features created manually. To perform well, DL approaches need large amounts of data. Thus, the two main factors affecting the performance of DL techniques are automatic feature extraction and availability of resources. They are very important when comparing the DL techniques and traditional machine learning techniques.

*4.1. Implications.* Based on the reviewed work, several trends are noticed in the ASA area. It is clear that the reviewed

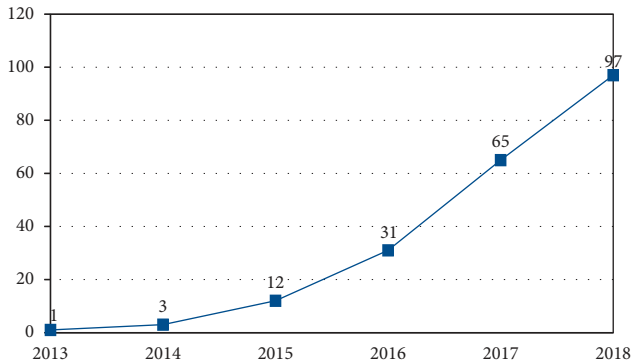


FIGURE 12: The number of articles using deep learning in SA.

literature covered ASA from the viewpoint of the classification methods and building resources related to specific domains. It is worthwhile to apply SA in many domains not targeted with the hottest methods.

However, many issues are still not sufficiently discussed and solved in ASA. These issues include shortcomings and gaps in the reviewed work to point out the implications exposed from reviewing ASA in several articles. These implications are involving two perspectives: for future research and for practice.

**4.1.1. Implications for Future Research.** As SLR focused on contributions of the existing literature relevant to ASA, almost implications, that are for future research, are discussed below:

- (i) Applying deep learning techniques to classify Arabic sentiments was conducted by some studies. However, it is not applied in many domains such as education.
- (ii) A comprehensive paradigm that expresses all the details of the preprocessing process in various situations have to be developed to conclude the appropriate processes that meet the characteristics of Arabic language.
- (iii) Building an Arabic lexicon is an open field. Most researchers built many lexicons that are either of limited size or not publicly available. Thus, oriented lexicons for each domain should be built since there are very few freely accessible Arabic corpora and lexicons for SA.
- (iv) It is clear from Table 5 that many researchers claimed to have achieved a high accuracy of sentiment classification, while they applied their classifiers on not standardized datasets. Thus, this can be unreliable. Therefore, the results of studies cannot be generalized, unless there is a standardization for DS as a benchmark in different domains.
- (v) Most current feature representations for Arabic language are borrowed from other languages, such as English. Therefore, developing new feature representation suiting Arabic language characteristics will help to improve the classification results.

**4.1.2. Implications for Practice.** ASA still needs applicable systems. These systems should consider the following:

- (i) Incomplete solutions were introduced to classify sentiments or opinions and predict events result. Thus, there is a need to develop recommendation systems in many fields, for instance, economy, business intelligence, politics, sports, education, and so on.
- (ii) An enhanced framework for ASA in different domains will contribute broadly to improve the performance of several industries. This will enhance the mental image of an organization through improving their services and products, therefore, customer satisfaction, and revenue.

## 5. Conclusion

In this SLR, the research articles on ASA were systematically reviewed. The contributions were analyzed with respect to specific research questions. It provides a systematic overview of existing research in ASA. After filtering, 108 published research, in 11 journals and 22 conference proceedings, have been analyzed.

ASA became an important issue in terms of talking about preprocessing process, features selection, and classification methods. The state of the art of ASA has shown the various and widespread works from different viewpoints. SLR highlights the frequent preprocessing strategies and the most methods used in feature selection. Furthermore, it presents a taxonomy of sentiment classification methods. This taxonomy was constructed to answer the research question: What are the most effective techniques used in ASA?.

Through SLR, it is obvious that ASA still needs more research. It contributes to implications for future research and implications for practice. The review shows that there is limited research on building standardized datasets and applying promising classification methods. Moreover, the review also reveals a lack of research from developing a new feature representation that suits Arabic language characteristics. Furthermore, avenues for future research also exist within developing recommendation systems in many fields and an enhanced framework for ASA in different domains. Researchers are encouraged to join this current research area.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] T. T. Thet, J.-C. Na, and C. S. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," *Journal of Information Science*, vol. 36, no. 6, pp. 823–848, 2010.
- [2] H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences," in *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pp. 129–136, Association for Computational Linguistics, Sapporo, Japan, July 2003.



- [3] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," *Knowledge-Based Systems*, vol. 89, pp. 14–46, 2015.
- [4] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [5] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*, pp. 415–463, Springer, Berlin, Germany, 2012.
- [6] A. Abbasi, H. Chen, and A. Salem, "Sentiment analysis in multiple languages: feature selection for opinion classification in web forums," *ACM Transactions on Information Systems (TOIS)*, vol. 26, no. 3, p. 12, 2008.
- [7] A. Farghaly and K. Shaalan, "Arabic natural language processing: challenges and solutions," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 8, no. 4, p. 14, 2009.
- [8] M. Rushdi-Saleh, M. T. Martín-Valdivia, L. A. Ureña-López, and J. M. Perea-Ortega, "OCA: opinion corpus for Arabic," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 10, pp. 2045–2054, 2011.
- [9] S. S. Alotaibi, "Sentiment analysis in the Arabic language using machine learning," Ph. D. dissertation, Colorado State University, Fort Collins, CO, USA, 2015.
- [10] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 417–424, Association for Computational Linguistics, Philadelphia, PA, USA, July 2002.
- [11] R. Tsarfaty, D. Seddah, Y. Goldberg et al., "Statistical parsing of morphologically rich languages (SPMRL): what, how and whither," in *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, pp. 1–12, Association for Computational Linguistics, Los Angeles, CA, USA, June 2010.
- [12] R. T. Khasawneh, H. A. Wahsheh, M. N. Al-Kabi, and I. M. Alsmadi, "Sentiment analysis of Arabic social media content: a comparative study," in *2013 8th International Conference for Internet Technology and Secured Transactions (ICITST)*, pp. 101–106, IEEE, Los Angeles, CA, USA, June 2013.
- [13] H. Al-Rubaiee, R. Qiu, and D. Li, "Identifying mubasher software products through sentiment analysis of Arabic tweets," in *Proceedings of the 2016 International Conference on Industrial Informatics and Computer Systems (CIICS)*, pp. 1–6, IEEE, Sharjah-Dubai, UAE, March 2016.
- [14] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews," *Journal of Computational Science*, vol. 27, pp. 386–393, 2018.
- [15] A.-S. Mohammad, O. Qwasmeh, B. Talafha, M. Al-Ayyoub, Y. Jararweh, and E. Benkhelifa, "An enhanced framework for aspect-based sentiment analysis of hotels' reviews: Arabic reviews case study," in *Proceedings of the 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)*, pp. 98–103, IEEE, Barcelona, Spain, December 2016.
- [16] A. Elouardighi, M. Maghfour, H. Hammia, and F.-z. Aazi, "A machine learning approach for sentiment analysis in the standard or dialectal Arabic facebook comments," in *Proceedings of the 2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech)*, pp. 1–8, IEEE, Rabat, Morocco, October 2017.
- [17] M. Mataoui, T. E. B. Hacine, I. Tellache, A. Bakhtouchi, and O. Zelmati, "A new syntax-based aspect detection approach for sentiment analysis in Arabic reviews," in *Proceedings of the 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP)*, pp. 1–6, IEEE, Trento, Italy, April 2018.
- [18] A.-S. Mohammad, M. Al-Ayyoub, H. N. Al-Sarhan, and Y. Jararweh, "An aspect-based sentiment analysis approach to evaluating Arabic news affect on readers," *Journal of Universal Computer Science*, vol. 22, no. 5, pp. 630–649, 2016.
- [19] Z. Kechaou, A. Wali, M. Ben Ammar, H. Karray, and A. M. Alimi, "A novel system for video news' sentiment analysis," *Journal of Systems and Information Technology*, vol. 15, no. 1, pp. 24–44, 2013.
- [20] A. Hammad and A. El-Halees, "An approach for detecting spam in Arabic opinion reviews," *The International Arab Journal of Information Technology*, vol. 12, 2013.
- [21] C. Okoli and K. Schabram, "A guide to conducting a systematic literature review of information systems research," *Sprouts: Working Papers on Information Systems*, vol. 10, no. 26, 2010.
- [22] D. Tranfield, D. Denyer, and P. Smart, "Towards a methodology for developing evidence-informed management knowledge by means of systematic review," *British Journal of Management*, vol. 14, no. 3, pp. 207–222, 2003.
- [23] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," EBSE Technical Report (EBSE-2007-01), Keele University, Keele, UK, 2007.
- [24] S. Ahmed, M. Pasquier, and G. Qadah, "Key issues in conducting sentiment analysis on Arabic social media text," in *Proceedings of the 2013 9th International Conference on Innovations in Information Technology (IIT)*, pp. 72–77, IEEE, Vancouver, BC, Canada, July 2013.
- [25] M. N. Al-Kabi, I. M. Alsmadi, R. T. Khasawneh, and H. A. Wahsheh, "Evaluating social context in Arabic opinion mining," *The International Arab Journal of Information Technology*, vol. 15, no. 6, pp. 974–982, 2018.
- [26] R. Duwairi and M. El-Orfali, "A study of the effects of preprocessing strategies on sentiment analysis for Arabic text," *Journal of Information Science*, vol. 40, no. 4, pp. 501–513, 2014.
- [27] M. N. Al-Kabi, N. A. Abdulla, and M. Al-Ayyoub, "An analytical study of Arabic sentiments: maktoob case study," in *Proceedings of the 2013 8th International Conference for Internet Technology and Secured Transactions (ICITST)*, pp. 89–94, IEEE, London, UK, December 2013.
- [28] M. M. Ashi, M. A. Siddiqui, and F. Nadeem, "Pre-trained word embeddings for Arabic aspect-based sentiment analysis of airline tweets," in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*, pp. 241–251, Springer, Cairo, Egypt, September 2018.
- [29] A. Elnagar, "Investigation on sentiment analysis for Arabic reviews," in *Proceedings of the 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, pp. 1–7, IEEE, Agadir, Morocco, November 2016.
- [30] A. Elnagar and O. Einea, "Brad 1.0: book reviews in Arabic dataset," in *Proceedings of the 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, pp. 1–8, IEEE, Agadir, Morocco, November 2016.
- [31] W. Cherif, A. Madani, and M. Kissi, "Towards an efficient opinion measurement in Arabic comments," *Procedia Computer Science*, vol. 73, pp. 122–129, 2015.
- [32] M. Itani, C. Roast, and S. Al-Khayatt, "Corpora for sentiment analysis of Arabic text in social media," in *Proceedings of the*

- 2017 8th International Conference on Information and Communication Systems (ICICS), pp. 64–69, IEEE, Irbid, Jordan, April 2017.
- [33] A. A. Altowayan and L. Tao, “Word embeddings for Arabic sentiment analysis,” in *Proceedings of the 2016 IEEE International Conference on Big Data (Big Data)*, pp. 3820–3825, IEEE, Washington, DC, USA, December 2016.
- [34] A. S. Alqarafi, A. Adeel, M. Gogate, K. Dashitpour, A. Hussain, and T. Durrani, “Toward’s Arabic multi-modal sentiment analysis,” in *Proceedings of the International Conference in Communications, Signal Processing, and Systems*, pp. 2378–2386, Springer, Harbin, China, July 2017.
- [35] A. Aliane, H. Aliane, M. Ziane, and N. Bensaou, “A genetic algorithm feature selection based approach for Arabic sentiment classification,” in *Proceedings of the ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, pp. 1–6, IEEE, Agadir, Morocco, November 2016.
- [36] N. F. B. Hathlian and A. M. Hafezs, “Sentiment-subjective analysis framework for Arabic social media posts,” in *Proceedings of the Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT)*, pp. 1–6, IEEE, Riyadh, Saudi Arabia, November 2016.
- [37] A. Shoukry and A. Rafea, “A hybrid approach for sentiment classification of egyptian dialect tweets,” in *Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing)*, pp. 78–85, IEEE, Cairo, Egypt, 2015.
- [38] A. Y. Al-Obaidi and V. W. Samawi, “Opinion mining: analysis of comments written in Arabic colloquial,” in *Proceedings of the World Congress on Engineering and Computer Science*, vol. 1, London, UK, November 2016.
- [39] T. H. A. Soliman, M. A. M. A. R. Hedar, M. Ali, and M. Doss, “Mining social networks’ Arabic slang comments,” in *Proceedings of IADIS European Conference on Data Mining*, vol. 22, p. 24, Prague, Czech Republic, July 2013.
- [40] A. Hamdi, K. Shaban, and A. Zainal, “A class-specific sentiment analysis framework,” *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 17, no. 4, p. 32, 2018.
- [41] M. M. Altawaier and S. Tiun, “Comparison of machine learning approaches on Arabic twitter sentiment analysis,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, no. 6, pp. 1067–1073, 2016.
- [42] H. K. Aldayel and A. M. Azmi, “Arabic tweets sentiment analysis-a hybrid scheme,” *Journal of Information Science*, vol. 42, no. 6, pp. 782–797, 2016.
- [43] S. Al-Azani and E.-S. M. El-Alfy, “Combining emojis with Arabic textual features for sentiment classification,” in *Proceedings of the 2018 9th International Conference on Information and Communication Systems (ICICS)*, pp. 139–144, IEEE, Irbid, Jordan, April 2018.
- [44] A. Elnagar, O. Einea, and L. Lulu, “Comparative study of sentiment classification for automated translated Latin reviews into Arabic,” in *Proceedings of the 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, pp. 443–448, IEEE, Hammamet, Tunisia, October 2017.
- [45] A. A. Altowayan and A. Elnagar, “Improving Arabic sentiment analysis with sentiment-specific embeddings,” in *Proceedings of the 2017 IEEE International Conference on Big Data (Big Data)*, pp. 4314–4320, IEEE, Boston, MA, USA, December 2017.
- [46] S. Abuelenin, S. Elmougy, and E. Naguib, “Twitter sentiment analysis for Arabic tweets,” in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*, pp. 467–476, Springer, Cairo, Egypt, September 2017.
- [47] A. Mahmoud and T. Elghazaly, “Using twitter to monitor political sentiment for Arabic slang,” in *Intelligent Natural Language Processing: Trends and Applications*, pp. 53–66, Springer, Berlin, Germany, 2018.
- [48] A. Elouardighi, M. Maghfour, and H. Hammia, “Collecting and processing Arabic facebook comments for sentiment analysis,” in *Proceedings of the International Conference on Model and Data Engineering*, pp. 262–274, Springer, Barcelona, Spain, October 2017.
- [49] A.-K. Al-Tamimi, A. Shatnawi, and E. Bani-Issa, “Arabic sentiment analysis of youtube comments,” in *Proceedings of the 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1–6, IEEE, Amman, Jordan, October 2017.
- [50] H. AL-Rubaiee, R. Qiu, K. Alomar, and D. Li, “Techniques for improving the labelling process of sentiment analysis in the saudi stock market,” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 3, pp. 34–43, 2018.
- [51] H. Abdellaoui and M. Zrigui, “Using tweets and emojis to build tead: an Arabic dataset for sentiment analysis,” *Computación Y Sistemas*, vol. 22, no. 3, 2018.
- [52] F. H. H. Mahyoub, M. A. Siddiqui, and M. Y. Dahab, “Building an Arabic sentiment lexicon using semi-supervised learning,” *Journal of King Saud University-Computer and Information Sciences*, vol. 26, no. 4, pp. 417–424, 2014.
- [53] N. El-Naggar, Y. El-Sonbaty, and M. A. El-Nasr, “Sentiment analysis of modern standard Arabic and egyptian dialectal Arabic tweets,” in *Proceedings of the Computing Conference*, pp. 880–887, IEEE, London, UK, July 2017.
- [54] G. Alwakid, T. Osman, and T. Hughes-Roberts, “Challenges in sentiment analysis for Arabic social networks,” *Procedia Computer Science*, vol. 117, pp. 89–100, 2017.
- [55] M. Biltawi, G. Al-Naymat, and S. Tedmori, “Arabic sentiment classification: a hybrid approach,” in *Proceedings of the 2017 International Conference on New Trends in Computing Sciences (ICTCS)*, pp. 104–108, IEEE, Amman, Jordan, October 2017.
- [56] R. Bouchlaghem, A. Elkhelifi, and R. Faiz, “Sentiment analysis in Arabic twitter posts using supervised methods with combined features,” in *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 320–334, Springer, Konya, Turkey, April 2016.
- [57] Y. AlMurtadha, “Mining trending hash tags for Arabic sentiment analysis,” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 2, pp. 189–194, 2018.
- [58] A. AL-Saffar, B. Sabri, H. Tao, S. Awang, M. Abdul Majid, and W. Al Saiagh, “Sentiment analysis in Arabic social media using association rule mining,” *Journal of Engineering and Applied Sciences*, vol. 11, pp. 3239–3247, 2016.
- [59] H. Rahab, A. Zitouni, and M. Djoudi, “SIAAC: sentiment polarity identification on Arabic algerian newspaper comments,” in *Proceedings of the Computational Methods in Systems and Software*, pp. 139–149, Springer, Szczecin, Poland, September 2017.
- [60] N. Boudad, R. Faizi, R. Thami, and R. Chiheb, “Sentiment classification of Arabic tweets: a supervised approach,” *Journal of Mobile Multimedia*, vol. 13, no. 3-4, pp. 233–243, 2017.

- [61] Q. A. Al-Radaideh and G. Y. Al-Qudah, "Application of rough set-based feature selection for Arabic sentiment analysis," *Cognitive Computation*, vol. 9, no. 4, pp. 436–445, 2017.
- [62] T. Elghazaly, A. Mahmoud, and H. A. Hefny, "Political sentiment analysis using twitter data," in *Proceedings of the International Conference on Internet of things and Cloud Computing*, p. 11, ACM, Cambridge, UK, March 2016.
- [63] H. Awwad and A. Alpkocak, "Performance comparison of different lexicons for sentiment analysis in Arabic," in *Proceedings of the 2016 Third European Network Intelligence Conference (ENIC)*, pp. 127–133, IEEE, Wroclaw, Poland, September 2016.
- [64] A. Bayoudhi, H. Ghorbel, and L. H. Belguith, "Sentiment classification of Arabic documents: experiments with multi-type features and ensemble algorithms," in *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, pp. 196–205, Shanghai, China, October 2015.
- [65] O. El Ariss and L. M. Alnemer, "Morphology based Arabic sentiment analysis of book reviews," in *Proceedings of the International Conference on Computational Linguistics and Intelligent Text Processing*, pp. 115–128, Springer, Budapest, Hungary, April 2017.
- [66] R. Baly, H. Hajj, N. Habash, K. B. Shaban, and W. El-Hajj, "A sentiment treebank and morphologically enriched recursive deep models for effective sentiment analysis in Arabic," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 16, no. 4, p. 23, 2017.
- [67] M. Maghfour and A. Elouardighi, "Standard and dialectal Arabic text classification for sentiment analysis," in *Proceedings of the International Conference on Model and Data Engineering*, pp. 282–291, Springer, Marrakesh, Morocco, October 2018.
- [68] K. M. Alomari, H. M. ElSherif, and K. Shaalan, "Arabic tweets sentimental analysis using machine learning," in *Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 602–610, Springer, Arras, France, October 2017.
- [69] S. Sabih, A. Sallam, and G. S. El-Taweel, "Manipulating sentiment analysis challenges in morphological rich languages," in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*, pp. 429–439, Springer, Cairo, Egypt, September 2017.
- [70] R. M. Duwairi and I. Qarqaz, "Arabic sentiment analysis using supervised classification," in *Proceedings of the 2014 International Conference on Future Internet of Things and Cloud (FiCloud)*, pp. 579–583, IEEE, Rome, Italy, August 2014.
- [71] S. Atia and K. Shaalan, "Increasing the accuracy of opinion mining in Arabic," in *Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing)*, pp. 106–113, IEEE, Konya, Turkey, April 2015.
- [72] R. M. Duwairi, "Sentiment analysis for dialectal Arabic," in *Proceedings of the 2015 6th International Conference on Information and Communication Systems (ICICS)*, pp. 166–170, IEEE, Amman, Jordan, April 2015.
- [73] W. A. Ahmed and A. M. El-Halees, "Arabic opinion mining using parallel decision trees," in *Proceedings of the 2017 Palestinian International Conference on Information and Communication Technology (PICICT)*, pp. 46–52, IEEE, Gaza, Palestinian, May 2017.
- [74] R. M. Duwairi and I. Qarqaz, "A framework for Arabic sentiment analysis using supervised classification," *International Journal of Data Mining, Modelling and Management*, vol. 8, no. 4, pp. 369–381, 2016.
- [75] J. Akaichi, "Sentiment classification at the time of the tunisian uprising: machine learning techniques applied to a new corpus for Arabic language," in *2014 European Network Intelligence Conference (ENIC)*, pp. 38–45, IEEE, Wroclaw, Poland, September 2014.
- [76] Q. A. Al-Radaideh and L. M. Twaiq, "Rough set theory for Arabic sentiment classification," in *Proceedings of the 2014 International Conference on Future Internet of Things and Cloud (FiCloud)*, pp. 559–564, IEEE, Rome, Italy, August 2014.
- [77] J. M. Perea-Ortega, M. T. Martín-Valdivia, L. A. Ureña-López, and E. Martínez-Cámara, "Improving polarity classification of bilingual parallel corpora combining machine learning and semantic orientation approaches," *Journal of the American Society for Information Science and Technology*, vol. 64, no. 9, pp. 1864–1877, 2013.
- [78] A. Mountassir, H. Benbrahim, and I. Berraba, "Sentiment classification on Arabic corpora. A preliminary cross-study," *Document Numérique*, vol. 16, no. 1, pp. 73–96, 2013.
- [79] A. Elnagar, Y. S. Khalifa, and A. Einea, "Hotel Arabic-reviews dataset construction for sentiment analysis applications," in *Intelligent Natural Language Processing: Trends and Applications*, pp. 35–52, Springer, Berlin, Germany, 2018.
- [80] A. El-Halees and A. Al-Asmar, "Ontology based Arabic opinion mining," *Journal of Information & Knowledge Management*, vol. 16, no. 3, p. 1750028, 2017.
- [81] A. M. Mostafa, "An automatic lexicon with exceptional-negation algorithm for Arabic sentiments using supervised classification," *Journal of Theoretical & Applied Information Technology*, vol. 95, no. 15, 2017.
- [82] A. Nuseir, M. Al-Ayyoub, M. Al-Kabi, G. Kanaan, and R. Al-Shalabi, "Improved hierarchical classifiers for multi-way sentiment analysis," *International Arab Journal of Information Technology (IAJIT)*, vol. 14, 2017.
- [83] T. Al-Moslmi, M. Albared, A. Al-Shabi, N. Omar, and S. Abdullah, "Arabic senti-lexicon: constructing publicly available language resources for Arabic sentiment analysis," *Journal of Information Science*, vol. 44, no. 3, pp. 345–362, 2018.
- [84] L. Abd-Elhamid, D. Elzanfaly, and A. S. Eldin, "Feature-based sentiment analysis in online Arabic reviews," in *Proceedings of the 2016 11th International Conference on Computer Engineering & Systems (ICCES)*, pp. 260–265, IEEE, Cairo, Egypt, December 2016.
- [85] A. R. Hedar and M. Doss, "Mining social networks Arabic slang comments," in *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, Singapore, April 2013.
- [86] M. El-Masri, N. Altrabsheh, H. Mansour, and A. Ramsay, "A web-based tool for Arabic sentiment analysis," *Procedia Computer Science*, vol. 117, pp. 38–45, 2017.
- [87] W. Cherif, A. Madani, and M. Kissi, "A new modeling approach for Arabic opinion mining recognition," in *Proceedings of the 2015 Intelligent Systems and Computer Vision (ISCV)*, pp. 1–6, IEEE, Fez, Morocco, March 2015.
- [88] A. Barhoumi, Y. E. C. Aloulou, and L. H. Belguith, "Document embeddings for Arabic sentiment analysis," in *Proceedings of the Conference on Language Processing and Knowledge Management*, Sfax, Tunisia, September 2017.
- [89] W. Cherif, A. Madani, and M. Kissi, "A combination of low-level light stemming and support vector machines for the classification of Arabic opinions," in *Proceedings of the 2016*

- 11th International Conference on Intelligent Systems: Theories and Applications (SITA)*, pp. 1–5, IEEE, Mohammedia, Morocco, October 2016.
- [90] S. Al-Azani and E.-S. M. El-Alfy, “Hybrid deep learning for sentiment polarity determination of Arabic microblogs,” in *Proceedings of the International Conference on Neural Information Processing*, pp. 491–500, Springer, Guangzhou, China, November 2017.
- [91] M. Abbes, Z. Kechaou, and A. M. Alimi, “Enhanced deep learning models for sentiment analysis in arab social media,” in *Proceedings of the International Conference on Neural Information Processing*, pp. 667–676, Springer, Guangzhou, China, November 2017.
- [92] B. Brahimi, M. Touahria, and A. Tari, “Data and text mining techniques for classifying Arabic tweet polarity,” *Journal of Digital Information Management*, vol. 14, no. 1, 2016.
- [93] I. Touati, M. Graja, M. Ellouze, and L. H. Belguith, “Towards Arabic semantic opinion mining: identifying opinion, polarity and intensity,” in *Proceedings of the Mediterranean Conference on Pattern Recognition and Artificial Intelligence*, pp. 131–136, ACM, Tebessa, Algeria, November 2016.
- [94] M. A. Sghaier and M. Zrigui, “Sentiment analysis for Arabic e-commerce websites,” in *Proceedings of the International Conference on Engineering & MIS (ICEMIS)*, pp. 1–7, IEEE, Agadir, Morocco, 2016.
- [95] W. A. Hussien, Y. M. Tashtoush, M. Al-Ayyoub, and M. N. Al-Kabi, “Are emoticons good enough to train emotion classifiers of Arabic tweets?,” in *Proceedings of the 7th International Conference on Computer Science and Information Technology (CSIT)*, pp. 1–6, IEEE, Helsinki, Finland, June 2016.
- [96] R. T. Khasawneh, H. A. Wahsheh, I. M. Alsmadi, and M. N. Al-Kabi, “Arabic sentiment polarity identification using a hybrid approach,” in *Proceedings of the 2015 6th International Conference on Information and Communication Systems (ICICS)*, pp. 148–153, IEEE, Amman, Jordan, April 2015.
- [97] R. Bouchlaghem, A. Elkhelifi, and R. Faiz, “SVM based approach for opinion classification in Arabic written tweets,” in *Proceedings of the 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, pp. 1–4, IEEE, Marrakech, Morocco, November 2015.
- [98] A. E. Rihab, B. Faiz, and Rim, “A machine learning approach for classifying sentiments in Arabic tweets,” in *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, p. 24, ACM, Nîmes, France, June 2016.
- [99] S. Ismail, A. Alsammak, and T. Elshishtawy, “A generic approach for extracting aspects and opinions of Arabic reviews,” in *Proceedings of the 10th International Conference on Informatics and Systems*, pp. 173–179, ACM, Cairo, Egypt, May 2016.
- [100] M. Hammad and M. Al-awadi, “Sentiment analysis for Arabic reviews in social networks using machine learning,” in *Information Technology: New Generations*, pp. 131–139, Springer, Berlin, Germany, 2016.
- [101] W. Adouane and R. Johansson, “Gulf Arabic linguistic resource building for sentiment analysis,” in *Proceedings of the 2016 International Conference on Language Resources and Evaluation*, Portorož, Slovenia, May 2016.
- [102] L. Al-Horaibi and M. B. Khan, “Sentiment analysis of Arabic tweets using text mining techniques,” in *Proceedings of the First International Workshop on Pattern Recognition*, vol. 10011, International Society for Optics and Photonics, Tokyo, Japan, May 2016.
- [103] N. Al-Twairesh, H. Al-Khalifa, and A. AlSalman, “Arasenti: large-scale twitter-specific Arabic sentiment lexicons,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 697–705, Berlin, Germany, August 2016.
- [104] H. Al-Rubaiee, R. Qiu, K. Alomar, and D. Li, “Sentiment analysis of Arabic tweets in e-learning,” *Journal of Computer Science*, vol. 12, no. 11, pp. 553–563, 2016.
- [105] B. Al Shboul, M. Al-Ayyoub, and Y. Jararweh, “Multi-way sentiment classification of Arabic reviews,” in *Proceedings of the 2015 6th International Conference on Information and Communication Systems (ICICS)*, pp. 206–211, IEEE, Amman, Jordan, April 2015.
- [106] M. Alhazmi and N. Salim, “Arabic opinion target extraction from tweets,” *ARNP Journal of Engineering and Applied Sciences*, vol. 10, no. 3, pp. 1023–1026, 2015.
- [107] S. Al-Osaimi and K. M. Badruddin, “Role of emotion icons in sentiment classification of Arabic tweets,” in *Proceedings of the 6th International Conference on Management of Emergent Digital Ecosystems*, pp. 167–171, ACM, Buraidah Al Qassim, Saudi Arabia, September 2014.
- [108] E. Refaee, “Sentiment analysis for micro-blogging platforms in Arabic,” in *Proceedings of the International Conference on Social Computing and Social Media*, pp. 275–294, Springer, Vancouver, Canada, July 2017.
- [109] S. Al-Azani and E.-S. M. El-Alfy, “Imbalanced sentiment polarity detection using emoji-based features and bagging ensemble,” in *Proceedings of the 2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1–5, IEEE, Riyadh, Saudi Arabia, April 2018.
- [110] A. El Ali, T. C. Stratmann, S. Park, J. Schöning, W. Heuten, and S. C. Boll, “Measuring, understanding, and classifying news media sympathy on twitter after crisis events,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 556, ACM, Montreal, Canada, April 2018.
- [111] M. Al-Batah, S. Mrayyen, and M. Alzaqebah, “Investigation of naive bayes combined with multilayer perceptron for Arabic sentiment analysis and opinion mining,” *Journal of Computer Science*, vol. 14, pp. 1104–1114, 01 2018.
- [112] A. M. Alayba, V. Palade, M. England, and R. Iqbal, “A combined CNN and LSTM model for Arabic sentiment analysis,” in *Proceedings of the International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, pp. 179–191, Springer, Hamburg, Germany, August 2018.
- [113] M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, and O. Qawasmeh, “Enhancing aspect-based sentiment analysis of Arabic hotels’ reviews using morphological, syntactic and semantic features,” *Information Processing & Management*, Springer, Berlin, Germany, 2018.
- [114] M. Gridach, H. Haddad, and H. Mulki, “Empirical evaluation of word representations on Arabic sentiment analysis,” in *Proceedings of the International Conference on Arabic Language Processing*, pp. 147–158, Springer, Fez, Morocco, October 2017.
- [115] S. Al-Azani and E.-S. M. El-Alfy, “Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short Arabic text,” *Procedia Computer Science*, vol. 109, pp. 359–366, 2017.
- [116] M. Abdul-Mageed, “Modeling arabic subjectivity and sentiment in lexical space,” *Information Processing & Management*, vol. 56, no. 2, pp. 291–307, 2017.
- [117] S. Siddiqui, A. A. Monem, and K. Shaalan, “Towards improving sentiment analysis in Arabic,” in *Proceedings of the*

- International Conference on Advanced Intelligent Systems and Informatics*, pp. 114–123, Springer, Cairo, Egypt, October 2016.
- [118] T. Khalil, A. Halaby, M. Hammad, and S. R. El-Beltagy, “Which configuration works best? an experimental study on supervised Arabic twitter sentiment analysis,” in *Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing)*, pp. 86–93, IEEE, Cairo, Egypt, April 2015.
- [119] S. Siddiqui, A. A. Monem, and K. Shaalan, “Sentiment analysis in Arabic,” in *Proceedings of the International Conference on Applications of Natural Language to Information Systems*, pp. 409–414, Springer, Salford, UK, June 2016.
- [120] M. Nabil, M. Aly, and A. Atiya, “ASTD: Arabic sentiment tweets dataset,” in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2515–2519, Lisbon, Portugal, September 2015.
- [121] A. Ziani, N. Azizi, and Y. T. Guiyassa, “Combining random sub space algorithm and support vector machines classifier for Arabic opinions analysis,” in *Advanced Computational Methods for Knowledge Engineering*, pp. 175–184, Springer, Berlin, Germany, 2015.
- [122] H. ElSahar and S. R. El-Beltagy, “Building large Arabic multi-domain resources for sentiment analysis,” in *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 23–34, Springer, Cairo, Egypt, April 2015.
- [123] N. Omar, M. Albared, T. Al-Moslmi, and A. Al-Shabi, “A comparative study of feature selection and machine learning algorithms for Arabic sentiment classification,” in *Proceedings of the Asia information retrieval symposium*, pp. 429–443, Springer, Sarawak, Malaysia, June 2014.
- [124] A. S. Al-Subaihini and H. S. Al-Khalifa, “A system for sentiment analysis of colloquial Arabic using human computation,” *The Scientific World Journal*, vol. 2014, Article ID 631394, 8 pages, 2014.
- [125] M. Aly and A. Atiya, “LABR: a large scale Arabic book reviews dataset,” in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, vol. 2, pp. 494–498, Sofia, Bulgaria, August 2013.
- [126] K. Khalifa and N. Omar, “A hybrid method using lexicon-based approach and naive bayes classifier for Arabic opinion question answering,” *Journal of Computer Science*, vol. 10, no. 10, pp. 1961–1968, 2014.
- [127] S. O. Alhumoud, M. I. Altuwajri, T. M. Albuhairi, and W. M. Alohaideb, “Survey on Arabic sentiment analysis in twitter,” *International Science Index*, vol. 9, no. 1, pp. 364–368, 2015.
- [128] Y. Yoshida, T. Hirao, T. Iwata, M. Nagata, and Y. Matsumoto, “Transfer learning for multiple-domain sentiment analysis-identifying domain dependent/independent word polarity,” in *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence AAAI*, San Francisco, CA, USA, February 2011.
- [129] S. Abdennadher, H. Ayman, C. Sabty, R. Salem, N. Tarhouny, and S. Zohny, “Building a corpus to categorize Arabic short text using games with a purpose,” in *Proceedings of the 13th International Conference WWW/Internet (ICWI)*, Porto, Portugal, 2014.
- [130] U. Kumar and P. Jaiswal, “Comparative Study on Sentiment Analysis and Opinion Mining,” *International Journal of Engineering and Technology, IJET*, vol. 8, no. 2, 2016.
- [131] F. Hemmatian and M. K. Sohrabi, “a survey on classification techniques for opinion mining and sentiment analysis,” *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1–51, 2017.
- [132] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub, “Arabic sentiment analysis: lexicon-based and corpus-based,” in *Proceedings of the 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1–6, IEEE, Amman, Jordan, 2013.
- [133] M. Salameh, S. Mohammad, and S. Kiritchenko, “Sentiment after translation: a case-study on Arabic social media posts,” in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 767–777, Denver, CO, USA, June 2015.
- [134] N. Farra and K. McKeown, “Smarties: sentiment models for Arabic target entities,” 2017, <https://arxiv.org/abs/1701.03434>.
- [135] S. Al-Azani and E.-S. El-Alfy, “Emojis-based sentiment classification of Arabic microblogs using deep recurrent neural networks,” in *Proceedings of the 2018 International Conference on Computing Sciences and Engineering (ICCSE)*, pp. 1–6, IEEE, Kuwait City, Kuwait, March 2018.