IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 1, 2019, accepted October 27, 2019, date of publication November 4, 2019, date of current version November 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2951530

A Review on Arabic Sentiment Analysis: State-of-the-Art, Taxonomy and Open Research Challenges

MOHAMED ELHAG MOHAMED ABO [1], RAM GOPAL RAJ [1], AND ATIKA QAZI [2]

[1]Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia
[2]Centre for Lifelong Learning, Universiti Brunei Darussalam, Gadong BE1410, Brunei

Corresponding authors: Ram Gopal Raj (ramdr@um.edu.my)

This work was supported by the University of Malaya (UM), under Grant IIRG012C-2019.

ABSTRACT Due to the significant use of Arabic language in social media networks, the demand for Arabic sentiment analysis has increased rapidly. Although, numerous sentiment analysis techniques enable people to obtain valuable insights from the opinions shared on social media. However, these techniques are still in their infancy, and the Arabic sentiment analysis domain lacks a compressive survey. Therefore, this study focused on the various characteristics, State-of-the-Art, and the level of sentiment analysis along with the natural language processing applied in the Arabic sentiment analysis. Furthermore, this study also discussed the sentiment analysis of the modern standards and the dialects of Arabic languages along with various machine learning processes and a few popular algorithms. Moreover, this study adds values by critical analysis of two case studies, which displayed an extensive set of the various research communities in this field of sentiment analysis. Finally, open research challenges are investigated, with a focus on the shortage of lexicons; availability; use of Dialect Arabic (DA); lack of corpora and datasets; right to left reading and compound phrases and idioms.

INDEX TERMS Sentiment analysis, machine learning, supervised learning, support vector machines, modern standard Arabic, dialect Arabic.

I. INTRODUCTION

In the past few years, social media has become a popular and widespread network communication system, which is used by the people for sharing their comments, opinions and sentiments. These sentiments, which refer to opinions regarding products, places, news, movies or sports, have become essential daily reading for many people. There are a huge number of online reviews posted daily created by active users. Currently, greater than two billion people use social media, where they discuss, browse and post their opinions on the internet [1]. Due to the evolution of the data available through the internet and smartphones, social media network websites, it has become very important in people's lives. seen to play a vital role in the manner in which the people interact, communicate or share the data and debate with one another [2], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Huazhu Fu.

SA is vital for many Natural Language Processes (NLP) like information extraction and responsive systems [4]. Information Extraction (IE) was used for extracting the data which was important for a specific topic or for fulfilling the user's requirements. For instance, many people use social media for broadcasting their ideas and beliefs regarding certain issues, and in the process, they tend to use multiple social networks or exploit the forums. Many of the concepts are positive; for example, we are human, so let us live together in peace. On the other hand, others are violent in the manner and negative in content. SA occurs by accessing and analysing data obtained from Social Media Networks (SMN), aiming to meet customers' needs.

Thus, this topic includes many fields, like Machine Learning (ML), computational linguistics, and NLP. While most dialogue uses the English language, other languages are fast becoming an important part of the social network's population.

The Arabic language belongs to the Sami language family, which is spoken in many regions of the world like North

Africa and the Middle East. It is one of the 6 languages used by the United Nations [5]. Furthermore, it was seen to be the official language in 27 countries. It is also used by ≈295 million people who live in North Africa and the Middle East [6]. It is one of the few languages which is spoken by the Sami family members, along with other languages like Hebrew, Aramaic (including Syriac), and Amharic. These Sami languages showed a distinguished and long literary history and played a vital role in the Middle Eastern culture for greater than 4,000 years [7]. The Arabic language includes 28 alphabets, with 25 consonants and 3 long vowels. It is a very cursive language, and all the cursive letters are connected with one another. There are 3 kinds of the Arabic language, i.e., Classical Arabic (CA), Modern Standard Arabic (MSA), and Dialect Arabic (DA) [8].

- CA is the oldest form of Arabic e.g. "Like this" In English and in CA is "هكذا".
- MSA is a formal form of Arabic and is the official language of many Arabic countries. e.g. "Like this " In English and in MSA is "مثل هذا".
- DA is a colloquial language, which is different in every region of the Arabic countries e.g. "Like this" In English and in MSA is "وين" [9].

This literature review of Arabic SA focuses on sentiment analysis in social media networks; for example, especially Facebook and Twitter as they are the most well-known platforms used in both the standard Arabic language and in colloquial Arabic. Possibly, the most relevant related work appears in [10]–[12]. However, our study differs from the previous papers by providing a comprehensive and in-depth discussion of sentiment analysis of both modern and dialect Arabic.

In this study, we determined how the available SA processes can be used in the unstructured Arabic language and investigated whether the available machine language algorithms could be improved for their use in the unstructured Arabic language. Thereafter, metrics were used for evaluating the accuracy of this solution. Moreover, we discussed the earlier studies carried out in the field of SA, in addition to the various ML processes, with the main focus on the data sets and hyper algorithms.

The remaining paper is organized in the following manner. Section II described the techniques of the SA. In Section III, describes the taxonomy and State-of-the-art of ASA Section IV discussed SA of MSA. Section V discussed SA of Arabic dialect. VI. discussed Evaluation metrics, the advantage and disadvantage. Moreover, the case studies have been described in Section VII. The various research challenges and the issues discussed in earlier studies carried out in the field of Arabic SA presented in section VIII. Section IX has offered a future research direction while conclusions were presented in Section X.

II. SENTIMENT ANALYSIS TECHNIQUES

This section presents the sentiment analysis techniques and basic features of SA approaches, including levels and methods.

A. MACHINE LEARNING TECHNIQUES

Machine learning techniques include a group of mathematical algorithms which are used for classifying the feelings based on some features in specific categories [13]. The algorithms are trained using many examples and are also used as a guide for determining the categorised ones. Based on this method, the best match is used for determining the best combination which would present new data. This ML technique was applied, using a few predefined categories and was then adapted to the received data, based on the classification of the training data.

1) SUPERVISED TECHNIQUE

Supervised learning is a process of training a predictive model. The supervision does not refer to human involvement, however, but rather to the fact that the target values give way for the learner to know. Stated more formally, given a collection of information, a supervised learning algorithm attempts to optimise a function (the model) to search out the combination of feature values that lead to the target output [14]. Moreover, studies of supervised learning used in Arabic SA is showed in [15]–[18].

2) UNSUPERVISED TECHNIQUE

It is called unsupervised learning because there is no target to learn; in fact, the process of training uses a descriptive model. This technique can resolve the issue of field dependency and reduce the need for using expounded training data. A majority of the processes that are used for unsupervised SA are categorised into lexicon-based processes and the generative models [19]. Moreover, some studies of unsupervised learning used in Arabic SA is provided in [20], [21].

3) SEMI-SUPERVISED TECHNIQUE

The Semi-Supervised Learning (SSL) technique lies between supervised and unsupervised learning. Along with the unlabeled data, this algorithm includes supervision data, but, not for all the examples [22], [23].

B. LEXICON-BASED CLASSIFICATION

Lexicons which select the words that are annotated with the sentiment orientation are called the sentiment lexicons [24]. This classification was developed based on the belief that the significant indicators of sentiment in the natural language texts were called sentiment or opinion words. These words are generally used for determining positive or negative emotions. For instance, "amazing", "good", and "wonderful" were positive sentiments whereas "poor" and "terrible" were negative feeling words [25]. Moreover, studies of lexicon-based used in Arabic SA is presented in [15], [21], [26]–[29].

C. BASIC FEATURES OF SA APPROACHES

SA refers to the determination of the polarity of the sentiment that is exposed by the web users with regards to their interaction on the specific topic, which is then classified as positive or negative [30]. SA is also called appraisal extraction, sentiment mining, sentiment classification, review mining, opinion mining, polarity classification and decision analysis. Sometimes, SA includes a method for tackling the subjective sentiments and opinions at different points in a text [31]. The main idea is to understand the perspective or opinion of users concerning a precise topic or goal.

There are several manoeuvrings for SA. Initially, a lexicon is created using the data based on the positive or negative words [32]. For instance, SentiWordNet was an easily available lexical resource, wherein every WordNet synset was assigned 3 digital scores that described the objectives and the positive and negative terms in a synset [4]. Furthermore, the lexicon could be manually or automatically compiled. Generally, the glossary consisting of lexical elements or corpora is manually developed, and thereafter the classifiers are trained using a comprehensive feature set for classifying the words and phrases.

Method two focused on the sentence mining or mining of complete documents, instead of depending on the bag-of-words. This method was mainly used for the corpora of text documents. The main problem related to document classification was that it had to resolve the emotional characteristics of the text, while the sentiment could be described using a single word or sentence. In a few cases, the sentiment was implicitly uttered, which is difficult to classify and recognize. On the other hand, the context related to the 'hidden' sentiment can describe useful information required for classification. The SA was classified into 3 classes, i.e., sentence-level, word-level, and document-level, based on the above classification methodology.

SA consists of three major analysis levels, i.e., Document sentiment classification, which classifies the complete document as positive or negative [33]. During sentence subjectivity, the researchers identify the subjective statements from the objective statements and further categorise them into either positive or negative statements. It is mainly useful for basic sentences, however, is not useful for longer phrases. For instance, 'I like the book, but it is too long "انا أحب الكتاب ولكنه طويل جد". Furthermore, emotional analysis at the sentence and document levels was helpful but did not describe the things liked or disliked by people. Hence, it did not identify the target of extracted opinions. This restricted the use of all opinions. The opinion is based on its entity and functionality, i.e., 'The I-Phone is good, but has a bad battery life'.

After covering the SA level, the next important step is to focus on techniques to classify sentiments, each using ML algorithms or by simply classifying them, based on the semantics of the text, see Figure (1). Classification is responsible for separating various things into classes, such as into positive and negative N-Grams in this particular case.

D. ANALYSIS OF SENTIMENT ANALYSIS TECHNIQUES AND APPROACHES

Comparison of Arabic with the other languages indicated that very few studies have used the Arabic corpora for SA due to the challenges related to the use of many SA approaches in the Arabic language. Figure. 2 showed that out of the 1458 SA-related papers that were published in 4 different databases, i.e., Association for Computing Machinery (ACM), ScienceDirect (SD), IEEE Xplore (IEEE), and Web of Science (WoS), till May 2017, only 48 were related to Arabic SA. Figure. 2 indicated that the Arabic- based SA studies have been published since 2011. This data indicated that Arabic SA needs further investigation.

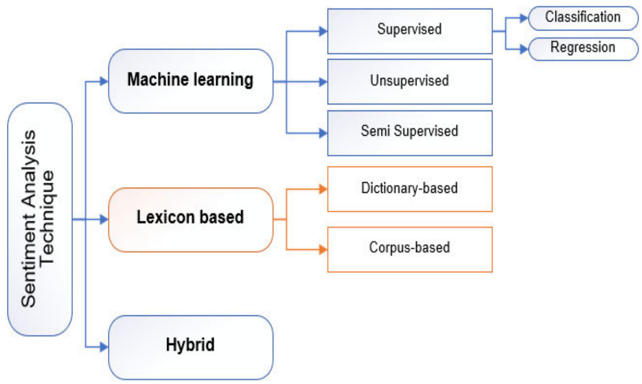

FIGURE 1. Sentiment analysis techniques.

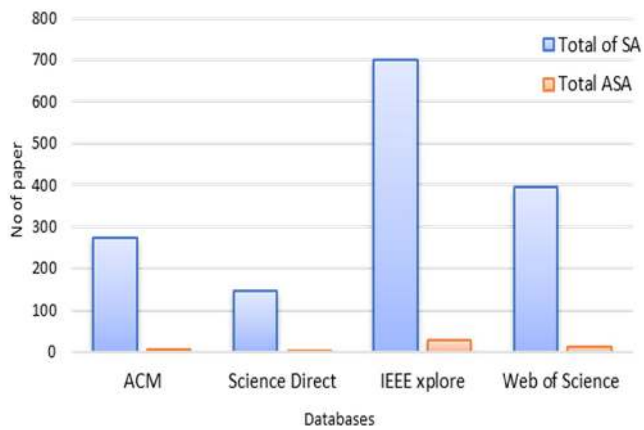

FIGURE 2. Number of Arabic SA articles compared with SA of other language articles.

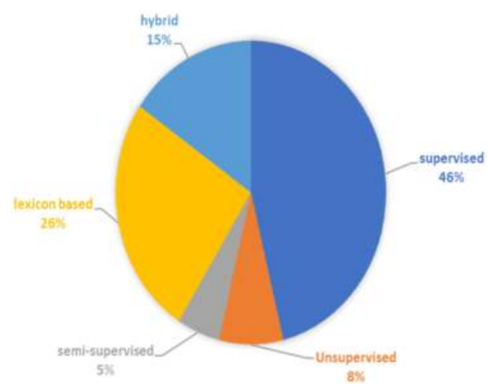

FIGURE 3. Summary of the SA approaches used for Arabic.

Figure. 3 indicated that supervised learning was more popular than the unsupervised learning technique 46% to 8% for unsupervised. In the past, researchers used different

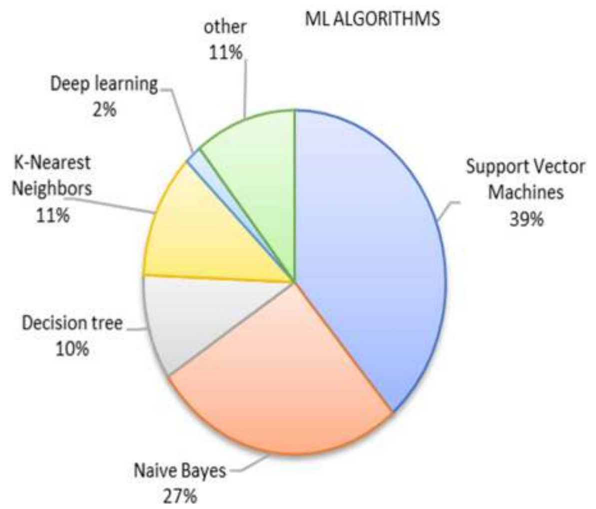

FIGURE 4. Frequency of ML algorithms in reviewed articles.

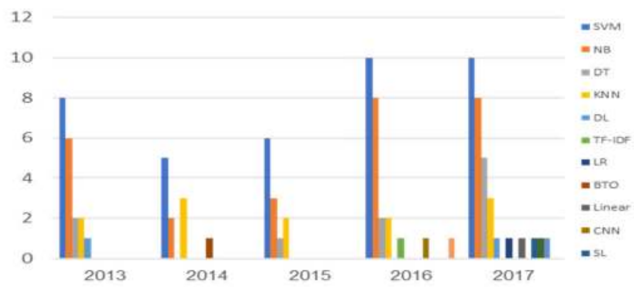

FIGURE 5. Studies using ML for Arabic sentiment analysis.

methods for carrying out SA for the Arabic language, where a few of the methods were more preferred. In this study, the authors investigated many methods used for MSA and DA. These included the supervised, semi-supervised, unsupervised, hybrid experiments and lexicon-based methods. However, the results showed that the hybrid and lexicon-based approaches were very popular amongst all the reviewed studies which lexicon-based approaches showed 26% and hybrid showed 15%.

Based on our analysis of the papers in Arabic research area, we identified 18 machine learning used for Arabic sentiment analysis which is depicted in Figure 4. We observed that 39% of the papers adopt Support Vector Machine (SVM), 27% are Naïve Bayes, 11% is K-Nearest Neighbors, Decision Tree showed 10%, and Deep learning comes with top five algorithms used of Arabic sentiment analysis which is presented 2%. However, the other thirteen machine learning algorithms used for Arabic sentiment analysis showed 11%. It is observed that the supervised approach is frequently used due to the lack of gold-standard Arabic data set online availability for research. Moreover, SVM and Naïve Bayes is an utmost used algorithm for sentiment analysis because of its simplicity and easy to apply. In comparison, the unsupervised and semi-supervised machine learning approach got attention in recent years because of the promising result for powerful ML algorithms such as deep learning. Furthermore, we believe deep learning needs a further drive with MSA and AD because of the variety of the Arabic language.

The most commonly-used algorithms include the Support Vector Machines (SVM), Naive Bayes (NB), Decision tree (DT), K-Nearest Neighbors (KNN), and Deep Learning (DL). Figure. 4 presents the algorithms used in various SA-based techniques. These ML algorithms have been used in the last five years.

Figure. 5 describes the analysis of all articles based on the years and the algorithms applied in the Arabic SA. When all results were analyzed, it was seen that the NB, SVM, and k-NN algorithms were used almost every year, though the DT algorithm became more popular in the past 3 years. Deep learning is also start used in leas year in Arabic SA.

III. TAXONOMY OF ARABIC SA

Here, we have described the taxonomy of Arabic SA process. This taxonomy included the algorithms, approaches, SA level and languages, as presented in Figure. 6.

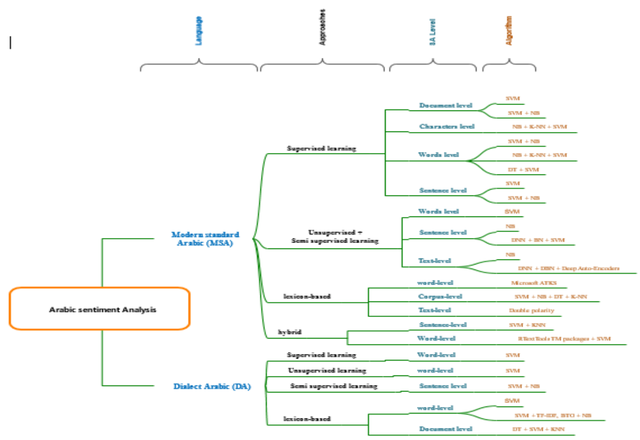

FIGURE 6. Taxonomy of Arabic Sentiment analysis.

Many studies have described SA techniques. Generally, these studies have used SA for English along with other languages like Turkish, Spanish and Greek. Several studies presented various methods for SA. On the other hand, very few researchers used the SA for the Arabic. The flowing two sections discuss important SA studies in MSA in section IV, following by a survey on Dialect Arabic (DA) SA studies in section V.

IV. SENTIMENT ANALYSIS IN MSA

A. SUPERVISED LEARNING

Comparing it with other approaches, Table 1 showed supervised learning was used more than other approaches. Many studies applied this approach with four levels of analysis; document, sentence, word and phrase level. Moreover, applied to Algorithm, Features, for MSA Languages, the experiments showed good result.

TABLE 1. Summary of the supervised approaches.

Ref	Year	Dataset	Level	Features	Algorithms	Language	Performance measures			
							Accuracy	Precision	Recall	F-measure
[34]	2014	Twitter	-	Unigrams	SVM, NB, K-NN	MSA	K-NN 59.99% NB 76.78% SVM 71.68%	N/A	N/A	N/A
[33]	2014	Arabic reviews	words and characters	N-gram	SVM, NB, K-NN	MSA + DA	K-NN 89.6 % NB 97.2 % SVM 93.8%	K-NN 96.6 % NB 99.62% SVM 89.71%	K-NN 84.2 % NB 94.8% SVM 99.6%	N/A
[35]	2014	Maktoob chat	sentence-level	POS + NEG adjective	SVM light	MSA	84.65	N/A	N/A	N/A
[36]	2015	Twitter	-	Feature correlation reduction, Word n-grams, Character n-grams	DT,NB, K-NN, SVM	MSA	N/A	DT 0.791 K-NN 0.822 NB 0.835 SVM 0.864	DT 0.693 K-NN 0.756 NB 0.763 SVM 0.787	DT 0.739 K-NN 0.788 NB 0.797 SVM 0.824
[37]	2015	Twitter	-	-	SVM and K-NN	MSA, Saudi DA	SVM 83.60% K-NN 74.80%	N/A	N/A	N/A
[38]	2015	different domains	sentence level	-	Manual classification	MSA +DA	97.00%	N/A	N/A	N/A
[39]	2015	BBN + Twitter	-	Word and character N-Grams	NRC-Canada, SVM	MSA	Accuracy = 87%	N/A	N/A	N/A
[40]	2016	Arabic Book Reviews	corpus-based approach	-	NB, DT, SVM	MSA	SVM 47.4% DT 47.6% NB 48.9% K-NN 57.8%	SVM 65% DT 66% NB 70% K-NN 70%	SVM 47% DT 47% NB 48% KNN 57%	SVM 55% DT 55% NB 57% K-NN 63%
[41]	2016	Multi-domain - Twitter	Document	Underlying word	NB, SVM	MSA	90.00%	N/A	N/A	N/A
[42]	2016	Twitter	word + phrase levels	Uni+ Bio, and trigrams	NB, SVM	MSA	SVM 89.5% NB 84.84%	SVM 0.817 NB 0.744	SVM 0.975 NB 0.968	SVM 0.885 NB 0.842
[43]	2016	Twitter	-	N-Grams	NB, SVM (TF-IDF, BTO) N-Grams	MSA+Gulf dialect)	NB 77.53 SVM 84.89	NB 82.84 SVM 59.00	NB 67.92 SVM 66.49	N/A
[44]	2016	Twitter	word	-	NB, SVM	Arabizi	N/A	NB 0.736 SVM 0.758	NB 0.823 SVM 0.869	N/A
[30]	2016	ecommerce		Bag of words	SVM, NB, K-NN	MSA	SVM 93.9% NB 93.87%	SVM 0.948 NB 0.946 K-NN 0.803	SVM 0.939 NB 0.939 K-NN 0.776	N/A
[45]	2016	Twitter			NB, DT, SVM	MSA	N/A	NB 0.782 DT 0.809 SVM 0.777	NB 0.752 DT 0.794 SVM 0.534	NB 0.750 DT 0.789 SVM 0.440
[46]	2016	different domains & dataset	sentence-level+ document-level		ML (SVM), 5fold cross validation, unigram, linear	MSA	Sentence-level = 89.3% Document-level = 93.4%	N/A	N/A	N/A
[47]	2017	Twitter		Sentic, PoS, Modification, Negation feature	K-NN, NB, DT	Iraqi, Egyptian + Lebanese DA	N/A	NB 86.76 K-NN 87.45 DT 89.53	NB 93.94 K-NN 88.77 DT 92.14	NB 87.97 K-NN 87.22 DT 86.78
[48]	2017	Twitter		Unigrams and bigrams	NB, LR, SVM, DNN, CNN TF-IDF	MSA	MNB 90.14% BNB 89.16% LR 88.32% SV 90.88% LSV 91.37% SGD 91.87% Nu-SV 87.82%	N/A	N/A	N/A
[49]	2017	different reviews	sentence level + document level	Bag-of-word	SVM, Brown Clustering Algorithm + 10-fold+ Scikit-Learning	MSA	Subjectivity 96% Polarity 90%	N/A	N/A	N/A

1) SUPPORT VECTOR MACHINES (SVM)

In 2017 and 2016, a study was carried out in MSA, which used SVM classifiers [49], [46]. This study was carried out by Alotaibi and Anderson [49], wherein they used the word clustering feature for classifying the Arabic sentiments. This method was based on the brown clustering method. The researchers clustered the words based on the context in the same data set and also used the SVM and Brown Clustering Algorithms with 10-fold cross-validation that used the Scikit-Learning and carried out a supervised SA of the Arabic language at the sentence and document level. All algorithms were used for the data set collected from various earlier reviews. An accuracy of 96% was noted. In 2016, Alotaibi and Khan [46] carried out another study where they used the SVM classifiers with 5-fold cross-validation, linear regression and unigrams, for carrying out a supervised SA on the Arabic language, at the sentence and document level. All algorithms were used for the data set collected from earlier reviews. Results indicated that the sentence level showed 89.3% of accuracy while the document-level showed 93.4% accuracy.

Salameh and Kiritchenko [39] used the SVM classifier and NRC-Canada for carrying out the supervised SA for the Arabic language at the word-level and also applied the character N-Grams feature. All algorithms were used for the data set derived from BBN and Twitter. 87% of accuracy was noted in the study.

Abdul-Mageed *et al.* [35] derived an annotated 11,918 sentences from various social media networks. The classification was carried out in 2 different stages. In Stage 1, the subjective and objective text was differentiated (i.e., Subjectivity Classification). In Stage 2, the positive and negative sentiments were also differentiated (i.e., Sentiment Classification). In their study, Abdul-Mageed et al. used the SVM classifiers as the learning algorithm along with the language-specific and some general features. The language-independent features included the domain, n-grams, unique or polarity lexicon features. They added a few Arabic-specific features for investigating the effect of the morphological data on the system performance. Results indicated that the POS tagging, lemmas and lexemes could be used for extracting the base form of the words, which could positively affect the sentiment and subjectivity classification. However, a number of a study done in 2015 and previous years the few works were done in modern standard Arabic [39], [35]. The studies were done by Salameh and another researcher. Using SVM with word-level and character and N-Grams applied in supervised approach yielded a better result than other using the same method with a dataset collected from Maktoob chat. Moreover, using word-level is show better result than other SA levels.

2) SUPPORT VECTOR MACHINES (SVM) AND NAÏVE BAYES (NB)

In 2016, a study was carried out where the researchers used the SVM and NB classifiers for investigating the MSA [41], [42], [43], [44]. Alabdullatif *et al.* [41] applied the SVM and NB classifiers for carrying out a supervised SA for the Arabic languages. These algorithms were used for classifying topics like religion, sports, politics, economy and technology, collected from Twitter. Results showed a 90% accuracy.

Hathlian and Hafezs [42] manually categorised 3700 tweets for developing their corpus. During classification, they noted that only 1550 tweets were related to a particular topic. The user names, hashtags, pictures, emoticons, URLs and non-Arabic words were deleted from these tweets. For overcoming all spelling mistakes and for standardising the word-writing formulae in the tweets, the researchers applied a normalization process. The extract features were based on the unigrams, bigrams and trigrams. They used 2 classifiers for testing the corpus, i.e., SVM and NB using the Weka Suite Software. Good results were noted for the SVM and NB classifiers in the unigram language model.

Al-Rubaiee *et al.* [43] developed a corpus of the Mubasher product reviews, with the help of a small program that they developed in C # and Twitter's API. They collected 2051 tweets for 57 days from the Twitter website of the Mubasher Company in Saudi Arabia, written in MSA and the local Saudi Arabic dialects. These tweets were classified by 2 experts into 3 classes (i.e., positive, negative or neutral), wherein the positive tweets were labelled as "1", negative tweets were labelled as "−1", while neutral as "0". They deleted the unrelated tweets, and the corpus consisted of 1331 tweets. During the normalization phase, the researchers assigned the set of codes, signs and Arabic words from different forms and English language, and replaced them with the standard Arabic equivalents, with the help of RapidMiner, Light stem, Removal stop word, Tokenization, and Filter token, by the length. The SVM and NB classifiers were applied using 2 feature selection schemes, i.e., TF-IDF and BTO (Binary-Term Occurrence), for creating the word vector. Best results were noted when they used the SVM classifier.

Duwairi *et al.* [44] applied the supervised SA on Arabizi, wherein the researchers converted the Arabizi text into Arabic, by using the rule-based process and a crowdsourcing tool. They used 2 algorithms, i.e., SVM and NB. They applied these algorithms on the Twitter dataset at a word-level. 86.9% recall was noted when they used the SVM algorithm. However, they noted better results when they eliminated the neutral data points from the dataset.

3) MULTI CLASSIFIERS (NB), (K-NN), AND (SVM)

Multi classifiers used between 2015 and 2017, the studies apply in MSA was done using SVM, NB, and K-NN classifier [47], [48], [30], [37]. Abdulkareem and Tiun [47] developed and used many POS tagging models like the NB, DT and the k-NN for many languages like Egyptian, Iraqi, and the Lebanese dialect Arabic. 87.9% of accuracy was noted. Alayba *et al.* [48] developed a Twitter dataset which consisted of many opinions regarding health services. They discussed the various data collection and filtration processes, along with the annotation and pre-processing processes. They carried out many classification experiments using many ML algorithms like CNN and DNN. The use of the DL approach showed good results, however, the use of SVM classifiers showed better results.

A supervised approach was applied by Sghaier and Zrigui [30], where they developed a new corpus for predicting the sentiment in the commercial sector. This dataset was manually derived from many web pages like jawal1232, reviewzat1 and jumia3. The data included reviews regarding 5 products, like notebook PC, camera, mobile phones, tablets and television. 3 experts manually classified the 250 reviews, and 125 positives and 125 negative reviews were noted. In this study, the researchers developed a symbol to word converter, which converted the emoticons and symbols to corresponding words. The special characters, stop words, numbers and non-Arabic words were deleted during the normalisation phase. They used unigram, bigrams, trigrams for extracting the features. They used 3 ML algorithms, i.e., SVM, NB and k-NN. The results indicated that the application of SVM and NB algorithms could help in better detection of the polarity of the opinions, compared to the KNN algorithm. Alhumoud *et al.* [37] manually carried out their annotation at the word-level. Furthermore, the annotators could handle issues like sarcasm, or a mix of opposite and sentimental words used in one sentence. The dataset was collected from the Twitter website. This was seen to be a rapid method, which could be easily processed by human annotators. 95% accuracy was noted after using this technique.

In 2014, a study was conducted where the researchers used a supervised approach for the MSA and applied the SVM, NB and the k-NN classifiers [34], [33]. Duwairi *et al.* [34] compiled 350,000 tweets after using the using Application Interface (API) on the Twitter website, after writing the PHP script. For improving the classification technique, they developed a tool for manually classifying all the tweets. This tool allowed the user to investigate each tweet and helped them to select the category, i.e., positive, negative or neutral and not applicable. This tweet filtering process depended on specific criteria like every tweet includes a minimum of 100 characters, did not include >4 hashtags, did not include any links or mentions and did not duplicate or retweet. In the final stage, they rated >25,000 tweets. They also developed a RapidMiner extension for matching the task of the words. This extension included the emoticon convertor, negation detection, repetition remover, and conversion of the dialect (i.e., Jordanian dialect) to MSA, Arabizi converter and a link remover. 3 ML algorithms were used for assessing this framework, i.e., SVM, NB and k-NN. Results indicated that the NB showed the best performance.

Duwairi and El-Orfali [33] studied the effects of feature correlation, stemming and n-gram models used in the Arabic texts on the SA process. They used 2 datasets, wherein Dataset 1 included 300 political tweets derived from the Aljazeera website, while the second dataset was the OCA. They carried out stemming and removed the stop words in the datasets. Furthermore, they also used the SVM, NB and the k-NN classifiers with various features. Results indicated that the stemming was ineffective, which was attributed to the use of a stemmer offered by RapidMiner that had higher error rates. Furthermore, the accuracy values were also improved when the features were decreased based on their correlation to 2 class labels and after using the word and character n-grams. The NB classifier showed a maximal accuracy of 96.6%.

Another Classifier: Ibrahim *et al.* [38] carried out a study and used the supervised approach and manual classification. They developed a MIKA corpus which included 4000 subjects of various forms of data (product reviews, tweets, hotel reservation and TV program comments) that were written in the Egyptian and MSA dialects. They applied the normalisation step. 3 Arabic experts classified all the data into 3 classes, i.e., positive (PO), negative (NG) and neutral (NU). Furthermore, in the classification step, they assigned a sentiment power to every topic, ranging between 1–10 for positive; −1 to −10 for negative and 0 for neutral.

B. UNSUPERVISED AND SEMI-SUPERVISED APPROACHES

In a popular ML approach used for SA, the input data was unlabelled and did not provide a good result. Table 2 describes the unsupervised learning. A few studies used this method for various SA levels.

TABLE 2. Summary of the unsupervised semi-supervised approaches.

Ref	Year	Dataset	Level	Features	Algorithms	Lang	Performance measures			
							Accuracy	Precision	Recall	F-measure
[50]	2014	Online game	sentence	Fine - grained	SVM	MSA	60.32 %	N/A	N/A	N/A
[51]	2015	The Linguistic Data Consortium Arabic Tree Bank	sentence	ArSenL	DNN, DBN and Deep Auto-Encoders	MSA	Linear SVM 6.1 DNN 55.5 DBN 57.5 Deep auto – DBN 60.4	N/A	N/A	Linear SVM 62.8 DNN 44.5 DBN 46.8 Deep auto –DBN 60.5

In 2015, several studies were carried out which used the unsupervised and semi-supervised approaches in the MSA [51], [50]. Al-Subaihin and Al-Khalif [50] proposed a new lexicon-based method for tackling the dialectal Arabic. This approach was very novel since it used an online game for developing a sentiment lexicon, using the human computational process. The researchers used the SVM algorithms for carrying out an unsupervised SA for the MSA. These algorithms were applied to the online game dataset consisting of a few fine-grained features. The best accuracy of 60.32% was noted. Al Sallab *et al.* [51] investigated many deep learning models like the Deep Neural Network (DNN), which applied the backpropagation to the conventional NN with many layers. The Deep Belief Networks (DBN) pre-trained the phases before feeding them to other steps, while the Deep AutoEncoder (DAE) decreased the dimensionality in the original models. When the researchers combined the

DAE with DBN and the Recursive AutoEncoder (RAE), they noted that the RAE constructed the raw words in the most appropriate order and then minimised the error of creating the sentence words in a similar order. Results indicated that the DAE model best represented the input sparse vector. RAE showed the best accuracy of 74% and showed a 9% better performance than the other models. However, the unsupervised approach can apply with different levels of SA, such as sentence level and text level with Deep Neural Network and yielded a better result.

C. LEXICON-BASED APPROACH FOR MSA

A lexicon can be defined as the vocabulary consisting of sentiment words with regards to their strength and sentiment polarity. Lexicon development includes many steps. In Step 1, a list of words, called seed words is created. This list is extended using the synonyms and antonyms of seed words. The common lexicon techniques require human annotation. Table 3 compared the various lexicon-based algorithms.

TABLE 3. A comparison of the lexicon-based algorithms.

Ref	Year	Dataset	Level	Features	Algorithms	Lang	Performance measures			
							Accuracy	Precision	Recall	F-measure
[52]	2016	LABR	-	(F1-F4)	SVM, K-NN	MSA	SVM 76.46 K-NN 72.22	N/A	N/A	N/A
[53]	2016	different domains	Feature and corpus	word	ATKS tools Microsoft	MSA	82.89%	80%	86.36%	83.05%

D. SVM, NB, DT, AND KNN

In 2016, the study on MSA was done in lexicon-based using SVM, NB, DT, K-NN and another classifier [52], [53]. Alqasemi *et al.* [52] used the SVM, NB and the k-NN algorithms, which have F1, F2, F3 and F4 features, for carrying out the lexicon-based SA for the Arabic language. These algorithms were used for the LABR (Large Arabic Book Review) dataset. SVM showed a maximal accuracy of 76.46%. Abd-Elhamid *et al.* [53] used a Part Of Speech (POS) tagging feature for carrying out an automatic extraction and weighting of sentiment features from the group of annotated reviews. All the collected features were organized into a tree structure, which highlights the relationship between the objects that are being reviewed and other components. Results showed that this proposed approach can automatically identify and extract the feature-sentiment expressions and display a higher accuracy of 82.89%.

Hybrid Approaches: The multi-algorithm used by the researchers is discussed in this section. The hybrid approach is explained in Table 4.

Between 2017 and 2015, the hybrid approaches done [55], [54]. Al-Moslmi *et al.* [55] developed an Arabic senti-lexicon for SA, wherein they created a Multi-domain Arabic Sentiment Corpus (MASC) using the supervised learning technique. The algorithms were used with a dataset collected from multidomain such as Google Play, Twitter, Facebook and Jeeran website. SVM and k-NN algorithms were used on a sentence-level. The experimental results showed accuracy varies between 40% and 68.7%.

TABLE 4. The hybrid approach.

Ref	Year	Dataset	Level	Features	Algorithms	Language	Performance measures			
							Accuracy	Precision	Recall	F-measure
[54]	2015	Twitter	-	unigram	SVM, (RText Tools tm, packages)	MSA	90.60 %	N/A	N/A	N/A
[55]	2017	Google Play, Twitter and Facebook + Jeeran website	sentence-level	-	NB K-NN, SVM, LLR and NEUNET	MSA	N/A	N/A	N/A	NB 80.52 KNN 82.36 SVM 83.16 LLR 79.44 NEUNET 81.59

TABLE 5. Summary of the top five algorithms used in arabic dialect (AD) SA.

Algorithms	Reference	Total # of paper
Support Vector Machines	[43] [33] [56] [44] [37]	5
Naive Bayes	[43] [57] [33] [44]	4
K nearest neighbors	[33] [37]	2
TF-IDF	[43]	1
Convolutional neural network	[43]	1

A hyper approach includes supervised and unsupervised methods and was applied by Alhumoud *et al.* [54]. The researchers manually annotated the *word level* having a unigram feature, where the annotators could potentially handle all the issues like sarcasm and included a mixture of various and opposing sentimental words in the sentence. This was also a very fast method and could be easily processed by human annotators. Also, the authors used the SVM algorithm in R software using RTextTools TM packages. The result of the experiment showed the best hybrid learning accuracy is up to 90.60%. However, the studies were done using the supervised and unsupervised approach with SVM and yielded better performance accuracy than the k-NN classifier.

V. SENTIMENT ANALYSIS IN ARABIC DIALECT

The Arabic language consists of many dialects. No study of the Arabic SA can be considered complete if one does not review the various dialects. Table 5 presents the overview of the top five algorithms which was used for studying the dialects in the reviewed studies. The results indicated that the MSA was used in all the studies, while the other dialects included Levantine, Saudi, Iraqi, Lebanese, Egyptian and Jordanian see table 6.

A. THE APPROACHES FOR DIALECT ARABIC (DA)

Refaee [58] applied SMOTE by using an imbalanced dataset consisting of 6,894 Arabic tweets. They used the word

TABLE 6. Summary of study used SA of Dialect Arabic (DA).

Ref	Year	Dataset	Level	Features	Algorithms	Languages	Performance measures			
							Accuracy	Precision	Recall	F-measure
[62]	2015	Twitter			Manual approach	Arabic dialects	0.46	0.70	0.46	N/A
[58]	2017	Twitter	word-level	word n-gram	SVM	Arabic & Dialectal	84.1	N/A	N/A	0.702
[56]	2015	Twitter			SVM, NB,	Jordanian Dialectical Arabic	N/A	SVM 0.903 NB 0.758	SVM 0.817 NB 0.784	N/A
[59]	2017	Twitter		Sentiment term length, Negatio, Supplication	WLBA	Saudi dialect	81%	90.4	78.9	85.2
[66]	2015	Twitter	character and word levels		Social Data Analytics (SDA)	Arabic dialects. Egyptian and Saudi dialects.	N/A	77.92%	71.48%	87.3
[43]	2016	Twitter		N-Grams	NB, SVM (TF-IDF, BTO) N-Grams	MSA+Gulf dialect)	NB 83.67 SVM 89.68	NB 78.25 SVM 98.62	NB 79.16 SVM 72.04	N/A
[47]	2017	Twitter		Sentic, PoS, Modification, Negation feature	K-NN, NB, DT	Iraqi, Egyptian + Lebanese DA	N/A	NB 93.94 K-NN 88.77 DT 92.14	NB 87.97 K-NN 87.22 DT 86.78	NB 86.76 K-NN 87.45 DT 89.53
[38]	2015	different domains	sentence level		Manual classification	MSA +DA	97.00%	N/A	N/A	N/A

unigrams and bi-grams features with the SVM classifier. It has been stated that use of SMOTE could significantly improve the F1 and accuracy of the results. However, poor results were noted when they applied SMOTE for a large dataset. This was attributed to the fact that a large dataset consisted of numerous features and was annotated automatically; hence, the expected noise was higher.

Assiri *et al.* [59] developed a Saudi dialect lexicon after integrating the various corpus-based and dictionary-based techniques. They expanded a list of Saudi seed words by applying the method described earlier [60], wherein they added the terms from the pre-developed lexicon [61] to a Saudi lexicon after they carried out a normalisation and cleaning process. Thereafter, they manually added a set of Saudi terms, which led to the development of a lexicon with 14,000 sentiment terms.

Duwairi *et al.* [62] the authors adopted the same scenario used in [63] to manually build an MSA sentiment lexicon of 2,376 words. Reference [64], stems and emotions. Reference [65] assumed that the sentiment terms appeared with other terms with similar polarity and developed an Egyptian lexicon that was built by using the corpus-based process. As a first step, they used a list of 380 sentiment seed words. Then following their hypothesis, the authors expanded this list by looking for patterns containing these seeds and their accompanied single terms.

The study using a supervised approach was done in 2015 using manual classification by Ibrahim *et al.* [38]. In this method, the researchers carried out a manual classification for conducting the supervised SA for the MSA and colloquial Arabic language. They classified the various domain datasets on a sentence level. 97% of accuracy was noted. Duwairi [56] developed a framework for SA based on the tweets written in the MSA or the Jordanian dialectical Arabic language. In their study, the researchers developed a dialect lexicon that mapped the dialectical words into their respective MSA words. They used a dataset with 22550 tweets and applied the NB and SVM classifiers for determining the polarity in the tweets.

Wang *et al.* [66] used the Social Data Analytics (SDA) software by developing a lexicon-based SA for the Saudi Arabic and Egyptian Arabic dialects, based on the word and character levels. They used a dataset from Twitter. The experiment result depends on Egyptian domain 1-TEL Precision = 92.23%, 2. GOV Recall = 91.29%, 3. EMP Precision = 87.56%. Moreover, in 2015 number of studies done in dialect Arabic, [62], [38], [56], [66]. The authors applied deferent approaches such as supervised, semi-supervised and lexicon-based with manual classification and ML algorithm classifier such as SVM. Moreover, the lexicon-based approach with SVM yielded is normally higher than other approaches.

Al-Rubaiee *et al.* [43] applied the SVM and NB classifiers with the N-Grams features for implementing the supervised SA for the standard Arabic and Saudi dialects. The SVM classifier showed the best performance, i.e., accuracy (96.06%), precision (95.80%) and recall (96.40%); followed by NB, i.e., accuracy (88.38%), precision (92.62%) and recall (84.99%). Experimental results indicated that the SVM was the best classifier for an unbalanced and large dataset. On the other hand, NB showed a better performance than SVM, if a feature selection technique was applied. Decreasing the feature correlation improved the independence between the features, which further enhanced the NB performance.

Abdulkareem and Tiun [47] developed a method that tagged the tweets written in the formal language, using proper spellings. They applied the k-NN, NB and the DT algorithms for conducting the supervised SA for all hyper languages like Egyptian, Iraqi and Lebanese dialect Arabic, having 10-fold cross-validation. They applied these algorithms to the dataset derived from Twitter. 87.9% of accuracy was noted.

However, ML and a manual classifier were used in a study done between 2015 and 2017 in Arabic dataset collected from Twitter using MSA and DA [43], [47]. The SVM result

TABLE 7. Evaluation metrics of Arabic sentiment analysis from 2013-2019.

References	Evaluation Metrics	Number of used
[6, 15-18, 21, 23, 26, 28, 29, 33, 34, 35, 42, 50, 51, 57, 60, 66, 67-96, 98]	Accuracy	49
[6, 15, 16, 20, 27-29, 33, 42, 60, 62, 67, 69, 72-74, 80, 82, 83, 85, 87, 90, 92, 96, 98-103]	Recall	30
[6, 15, 16, 27-29, 33, 42, 60, 62, 67, 69, 72-74, 80, 82, 83, 85, 87, 90, 92, 96, 98-103]	Precision	29
[6, 15, 16, 27, 42, 51, 60, 62, 67, 69, 72, 73, 82-85, 87, 90, 92, 98, 100-102, 104]	F-measure	24
[6, 35, 85]	False Positives (FP)	3
[69, 85, 90]	ROC-Area	3
[6, 85]	True Positives (TP)	2
[6, 85]	Receiver operating characteristic (ROC)	2
[72, 73]	GM	2
[35, 85]	False Negatives (FN)	2
[76, 86]	F1- Score	2
[65, 105, 106]	Bilingual evaluation understudy (BLEU)	2
[23, 67]	Avg-F1	2
[72, 73]	Area under Curve (AUC)	2
[84]	Gold standard (gold-std)	1
[84]	Cross-validation	1
[85]	True Negatives (TN)	1
[85]	Precision Recall Curve (PRC)	1
[106]	Out-Of-Vocabulary (OOV)	1
[72]	Matthews correlation coefficient (MCC)	1
[35]	FS	1
[35]	FO	1
[107]	Kappa	1

shows higher accuracy for the performance than another method, but using manual classifier also yielded a good result.

VI. PERFORMANCE EVALUATION

A. EVALUATION METRICS

To identify the most utilized evaluation metrics in the literature, we conducted a detailed analysis of Arabic sentiment analysis papers from 2013 – 2019 to extract the used evaluation metrics. Table 7 presents all the identified evaluation metrics in relation to the Arabic sentiment analysis techniques that utilized them. 23 evaluation metrics were identified, with Accuracy being the most utilized metric especially on machine learning techniques in Arabic sentiment analysis which is showing 30%. Flowed by recall and precision which is displayed 30 and 29 times respectively in the research area 18% for each. Moreover, the 3rd most use is F-measure evaluation metrics it's used more than 20 times is the Arabic sentiment analysis studies which shows 15%.

TABLE 8. Show Advantage and disadvantage of studies in Arabic sentiment analysis.

Ref	Year	Advantage	Disadvantage
[77]	2018	No need for feature extraction	Used only one classifier
[17]	2018	Huge Dataset Annotated for MSA and Dialect Arabic	Limited dialect Arabic
[108]	2018	Used a set of a linguistic feature Using MSA feature with Arabic dialect Used manual and auto-annotation	Could use more features
[58]	2017	An extensive set of features	No Bag-of-Word features used
[59]	2017	Used pre-weighted lexicons The new method developed for a large-scale Saudi dialect lexicon.	A limited set of classifiers and techniques
[47]	2017	A new methodology for POS tagging.	A limited set of data A limited set of features
[48]	2017	-	Could use more classifiers and more features
[49]	2017	Used a cluster label of the word features Improved the rules of the lexicon-based approaches	-
[40]	2016	Deals with the Hierarchical Classification	Limited to Bag-Of-Word feature
[110]	2016	Use of available English tools	No added value
[111]	2016	An extensive set of features was used	POS feature not used
[43]	2016	-	Requires more classifiers and features
[113]	2016	-	Did not use enough classifiers and features
[114]	2016	Corrects Spelling and eliminates proper nouns during the pre-processing step	Uses lesser features
[53]	2016	Tackles the Aspect level Included new methods for ontology and rule-based	-
[115]	2016	Uses POS tagging	Needs to use additional features
[116]	2015	It is a novel mathematical approach for computing feature weights	Is restricted to a single feature
[117]	2015	Deals with a 4-way classification	The data was not pre-processed
[62]	2015	Builds a polarity lexicon	Did not consider negation and Intensification
[56]	2015	Uses a lexicon dialect for translating the dialect words to MSA	Needs to use additional features and classifiers
[118]	2015	Deals with a Multi-way Classification	Needs to use additional features
[119]	2015	Uses lexicon and feature selection	-
[120]	2015	Evaluated the efficiency of Machine Translation	Is restricted to the English Sentiment Classifier
[121]	2015	Uses a large multi-domain dataset	The data was not pre-processed
[122]	2015	An extensive set of features was used	Did not use the Feature selection tool
[33]	2014	Use of feature correlation	Needs to use additional features
[123]	2014	-	Needs to use additional features and classifiers
[124]	2014	Deals with the intensity and polarity of sentiments	The weighting and extraction of features were manually done
[116]	2015	It is a novel mathematical approach for computing feature weights	Is restricted to a single feature

However, ROC-Area and False Positives (FP) displayed 3 times for each in the research area which is 2%. Another evaluation metric in the area is 18% of showed in table 8

TABLE 9. Summary of the case studies of SA.

Ref	Context	Objective	Approach	Result
[125]	A case study of Maghrebi Arabic	The study focuses on a corpus extricated from various Facebook pages the user using Maghrebi colloquial Arabic.	(a) Supervised approach to collect the sentiments (b) unsupervised approach to extract text. (c)A semi-supervised approach	The result of 55.05% can be considerably improved.
[122]	To enhance the discovery of sentiment polarity in Arabic sentences handling Arabic idioms/sayings/ phrases, a lexicon is critical.	Use of Arabic idioms/ sayings/ phrases is of e key value for increasing the detection of sentiment polarity in Arabic sentences. This approach was used, and some novel and rich sets of linguistically motivated features were employed (e.g. contextual Intensifiers, contextual Shifters and negation handling)	(a)Arabic sentiment words lexicon automatically extended with a semi-supervised approach s. (b)An ML tool and SVM classifier were used	Using SVM classifier technique, with resources, indicates high-performance, with accuracies of over 95%.

below in which 31 different metrics are used in Arabic sentiment analysis.

B. ADVANTAGE AND DISADVANTAGE

This section shows the advantage and disadvantage for the must study covered in this paper as shown in Table 8 below

VII. CASE STUDIES

Here, the researchers have used a case-study design and thoroughly analysed all the SA-based cases, like colloquial Arabic, HPV vaccines and Health. All the case studies have shown a SA deviation. two cases were selected which displayed an extensive set of the various research communities that used SA.

A. CASE STUDY (TOPIC AND SENTIMENT MODEL APPLIED TO COLLOQUIAL ARABIC: A CASE STUDY OF MAGHREBI ARABIC)

Zarra *et al.* [119] developed a corpus after extracting the various Facebook pages that used the Maghrebi colloquial Arabic language. The researchers used a supervised approach for extracting the sentiments, while an unsupervised approach was used for extracting the topic. They proposed a new semi-supervised approach for the Arabic language, which combined the sentiment and topic into one model, in order to connect every topic to the relative sentiment. A 55.05% improvement in the result was noted. For instance, blind classification helped in deriving good results, especially during the pre-processing steps. In this pre-processing step, all useless words or those having higher weight were eliminated. This decreased the corpus size and diversified the learning corpus.

B. CASE STUDY (SENTIMENT ANALYSIS FOR MSA AND COLLOQUIAL ARABIC):

Ibrahim *et al.* [116] used the Arabic idioms/sayings/phrases lexicon in order to improve the detection of the sentiment polarity in the Arabic sentences. In this approach, they used a few novels and rich set of linguistically-motivated features (like contextual Shifter, contextual Intensifiers, and negation handling). They also used a high-coverage Arabic sentiment word lexicon for the semi-supervised approach for SA and the sentiment Arabic idioms/sayings/phrases lexicon for improving the classification process, which used an ML algorithm and SVM classifier. Lastly, they used a few resources for improving the performance of the system, which improved the accuracy to $> 95\%$.

VIII. RESEARCH CHALLENGES

Though SA was accepted by numerous organizations, the research on the use of Arabic SA in the various social media networks has not been widely carried out. Many existing issues still have to be addressed. References [120]–[123] listed the most popular challenges.

- The Array of SA Tasks
- Sentiment of Words
- The sentiment of Phrases, Sentences, and Twitter: Sentiment Composition
- Challenges in Annotating for Sentiment
- Challenges in Multilingual SA.

Each day new challenges evolve due to the applications introduced by various organizations. In the following sections, a few research challenges like data quality, dataset availability, data pre-processing and other challenges related to the use of Arabic SA have been discussed.

A. DATASET AVAILABILITY

The accuracy of the SA method is based on the existence of large annotated corpora, which is a limited resource for the Arabic language. Moreover, a comparison of the available English and Arabic language datasets indicated that Arabic had a very small dataset [17], [68], [120], [124], [125].

TABLE 10. Challenges of SA in Arabic.

Challenges	Discourse
Lack of Lexicons	The MSA lexicons are inadequate in comparison to the English lexicons. There are no publicly-available lexicons for colloquial Arabic. A lexicon was proposed earlier [123], but it did not include all dialects [136]
Use of Dialectal Arabic (DA)	The task of developing analysers and POS taggers for the dialects is difficult because the dialects differ from the MSA syntactically and phonologically and exclude conventional orthographies. Furthermore, stop words are represented in numerous ways in the DAs and are seen to vary significantly [69, 110].
Lack of Corpora and Datasets	Accuracy of the SA system is based on the presence of annotated corpora, which did not include the Arabic dialects. Furthermore, the available datasets for Arabic are few than those available for the other languages. A comparison can be made about the accuracies between the new and existing systems [107]
Compound phrases and idioms	The Arabic speakers use popular compound idioms and slogans for sharing their opinions. For example, (لا يا شيخ !), which are used for expressing disbelief in what others are saying. Some words and phrases are used for expressing the sentiment, subject to usage, with an increasing number of new phrases daily [69, 137].
Use of Arabizi	The social media users generally switch between English and Arabic, which make it difficult to determine if the Latin letters were written in English or Arabizi. Arabizi is seen to be a novel trend in social media, where the users use Latin characters to express Arabic words [137]
Sarcasm	Sarcasm refers to a part of speech where a person says negative words though he means to say something positive and vice versa. It is not easy to detect sarcasm. Very few researchers have detected sarcasm using the ML approaches in the English language, while no studies have detected sarcasm in Arabic [25, 78]
Comparative Opinions	Mining of comparative opinions is a major issue in English, wherein the idea is expressed as a correlation between 2 entities. They are different from general opinions which present different semantic meanings and grammatical styles. The other languages are not exempted, including Arabic [138]
Spam Detection	Opinions include many spam messages, though this spam differs from internet spam and requires different detection methods. On the other hand, the Arabic opinion spam has not been investigated [137]
Co-reference resolution	Co-reference resolution takes place when multiple expressions in the text represent the same object or person. This problem occurs in a majority of NLP applications, which was also noted in SA [139]
Opinion Target	One must extract the opinion targets or speakers since the SA methods help in determining the polarity of the sentences [140]
Right to left	Arabic content is started from right to left and utilises new and exceptional shapes, imprints and vowels, and Arabic roots that are tri, quad or confined strictly. A large portion of these roots is trilateral. This makes it difficult to use an algorithm from another language with Arabic. [38, 57]

B. DATA QUALITY

Social media data for SA have a number of quality problems. Some examples of data quality problems are shown below [126].

1) NOISE AND OUTLIERS

- Noise refers to the modification of original values such as the distortion of a person's voice when talking on the phone while travelling along the road or talking in an industrial area where there is a lot of noise which changes the original values of noise quality.
- Outliers refer to data points with qualities which were not similar to other data points.

2) MISSING VALUES

During data collection, often the values are missing, which occurs due to several reasons, like women refusing to declare their weight or age. Such attributes differ from case to case. These missing value-related issues are handled by eliminating the data objects, ignoring all missing values during the data analysis, estimating missing values, or replacing them with other probable values [127].

C. DUPLICATE DATA

Data sets include the data objects which are duplicates or almost duplicate of other data points, during the data merging stage. For example, the same people with numerous addresses [128].

Data Pre-Processing: Data pre-processing is necessary because, in a real-world, the data often misses important values, is generally unclean, and includes many errors and outlier points. However, without any data pre-processing, these mistakes can affect the result [129]. Data pre-processing includes the tasks of filling the missing values, identifying and eliminating the noisy data, resolving redundancies and correcting all inconsistencies.

D. CHALLENGES OF SA

People often express their opinion in a complicated manner. Also, while expressing their opinions, the lexical content, in itself, is misleading. Also, negation, reversals and topic changes are very common in the sub-sentiment and intra-contextual analysis. Indirect expression of opinions is not easily discussed. Table 10 presents the most common challenges facing SA.

TABLE 11. Classification of open issues.

References	Type of language	Issues
[42] [141] [34] [96] [30] [53] [107] [45] [35] [15, 23, 26, 50]	MSA	Lexicon
[142] [51] [46] [18, 72]		Method
[110] [55] [142]		Build Arabic Lexicon and Wordlist
[38] [43] [33] [108, 143]	Mix MSA + DA	Corpus and SA levels
[66] [56] [144] [62] [47]	Arabic Dialect	Lexicon size
[145] [146] [147] [39] [148]	Other languages	Sentiment labels, Lexicons and Datasets

IX. FUTURE RESEARCH DIRECTION

This section will discuss several issues related to SA for the specific case of Modern Standard Arabic (MSA) and Dialect Arabic (DA). As shown in Table 11 these problems mainly lie in:

- Unavailability of dialect Arabic sentiment lexicons and word lists,
- Unavailability of a test dataset for Dialect Arabic and modern standard Arabic.

Hathlian *et al.* [42], [135], [34], [94], [30], [53], [103], [45], [35], [15], [23], [26], [50] addressed the issues associated with the lack of lexicons, in order to analyse and test the Arabic SA. The researchers collected a large dataset of MSA reviews and comments from social media networks like Twitter and Facebook. They also used popular classifiers like SVM, NB, DT and k-NN. Only MSA was used in all experiments.

Mountassir *et al.* [136], [51], [46], [49] focused only on the problems related to the open research, which were based on the method applied for SA (POS, cluster and analysis level) for MSA. The authors used many features like SVM, NB with 5-fold cross-validation, unigrams, and linear algorithm.

Rabab'ah *et al.* [105], [55] have addressed the problems associated with the development of the Arabic sentiment lexicon. They created and assessed the SentiStrength for including the scores; while they used Arabic SentiWordNet for the public and existing large-scale Arabic sentiment lexicon. They used only the MSA. Authors also used some classifiers (SVM), (k-NN) and (NB) in their experiments.

Ibrahim *et al.* [38], [43], [33], [137] addressed the issues related to the corpus and SA levels. Their corpus was flexible and included only reviews and comments expressed in MSA and colloquial Arabic language. It included 2 forms of languages (i.e., MSA and Dialect Arabic). During analysis, the researchers noted that a majority of the social media users preferred to use the MSA dialect.

Wang *et al.* [66], [56], [138], [62], [47] assessed the challenges, problems and open research-based issues associated with the Dialect Arabic (DA) SA. They aimed to develop this further by using a dialect lexicon, extending the texts of the tweet by using a list of synonyms and a vernacular-based Arabic sentiment lexicon and determining the computation of the semantic orientation in the Arabic language. The researchers used several techniques for computing the polarity using different algorithms like SVM, k-NN, DT and NB. Some researchers also used other ML methods for deriving good results.

Cheng *et al.* [139], [140], [141], [39], [142] addressed all problems related to the unlabelled data (i.e., unsupervised approach) and investigated the role played by the non-lexical and lexical features for testing and analysing the SA in various languages (like Turkish, Greek and English). The researchers used a dataset comprising of reviews and comments collected from social media networks like Facebook and Twitter. They also used general classifiers like SVM, DT, NB and k-NN.

X. CONCLUSION

The retrieval of Arabic dialect from social media network services such as Facebook and Tweeter is becoming a complex task. We conducted this literature review in the context of Arabic dialect sentiment analysis. First, we identified the State-of-the-art related to the ML-based Arabic SA and considered the existing SA techniques for the unstructured Arabic language. Moreover, the study indicated that the SA problem could be addressed by using classifiers like SVM, NB, DT, k-NN and a supervised approach. This was attributed to the higher estimation capability of the new ML processes. In addition, we also compared the results obtained for the modern standard Arabic (MSA) with Arabic dialect those derived from the Arabic language. However, we perceived on the basis of the extensive review that modern standard Arabic was the most commonly studied language for SA and was used for other research issues like information retrieval and book reviewer. Moreover, there has been a significant increase in the quality and number of studies carried out for the Arabic languages. Furthermore, we found it difficult to handle the diversity slang, due to its linguistic complexity. The results of the review could provide researchers with a comprehensive view of research trends and open challenges in Arabic language sentiment analysis.

REFERENCES

[1] L. E. Freund and W. Cellary, *Advances in The Human Side of Service Engineering*. San Jose, CA, USA: Springer, 2014.

[2] X. L. García, A. I. Rodríguez, and X. Soengas, "Estrategias de los diarios online de proximidad para las redes sociales: El caso de La Voz de Galicia Online local medias strategies for social networks: The case of La Voz de Galicia," in *Proc. 10th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Agueda, Portugal, 2015.

[3] M. Shinde-Pawar, "Formation of smart sentiment analysis technique for big data," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 2, no. 12, pp. 7481–7488, 2014.

[4] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[5] A. Eisele and Y. Chen, "MultiUN: A multilingual corpus from united nation documents," in *Proc. LREC*, 2010, pp. 2868–2872.
[6] 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 *Proc. 8th Int. Conf. Internet Technol. Secured Trans. (ICITST)*, Dec. 2013, pp. 101–106.
[7] R. Hetzron, *The Semitic Languages*. Evanston, IL, USA: Routledge, 2013.
[8] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions," *ACM Trans. Asian Lang. Inf. Process.*, vol. 8, no. 4, 2009, Art. no. 14.
[9] K. Versteegh, *The Arabic Language*. Edinburgh, U.K.: Edinburgh Univ. Press, 2014.
[10] N. Boudad, R. Faizi, R. O. H. Thami, and R. Chiheb, "Sentiment analysis in Arabic: A review of the literature," *Ain Shams Eng. J.*, vol. 9, no. 4, pp. 2479–2490, 2017.
[11] M. El-Masri, N. Altrabsheh, and H. Mansour, "Successes and challenges of Arabic sentiment analysis research: A literature review," *Social Netw. Anal. Mining*, vol. 7, no. 1, p. 54, 2017.
[12] A. Alnawas and N. Arıcı, "The corpus based approach to sentiment analysis in modern standard Arabic and Arabic dialects: A literature review," *Politeknik Dergisi*, vol. 21, no. 2, pp. 461–470, 2018.
[13] D. Librenza-Garcia, B. J. Kotzian, J. Yang, B. Mwangi, B. Cao, L. N. P. Lima, M. B. Bermudez, M. V. Boeira, F. Kapczinski, and I. C. Passos, "The impact of machine learning techniques in the study of bipolar disorder: A systematic review," *Neurosci. Biobehav. Rev.*, vol. 80, pp. 538–554, Sep. 2017.
[14] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2016.
[15] I. Touahri and A. Mazroui, "Studying the effect of characteristic vector alteration on Arabic sentiment classification," *J. King Saud Univ.-Comput. Inf. Sci.*, to be published.
[16] G. Jardaneh, H. Abdelhaq, M. Buzz, and D. Johnson, "Classifying Arabic tweets based on credibility using content and user features," in *Proc. IEEE Jordan Int. Joint Conf. Elect. Eng. Inf. Technol. (JEEIT)*, Apr. 2019, pp. 596–601.
[17] A. Elnagar, L. Lulu, and O. Einea, "An annotated huge dataset for standard and colloquial Arabic reviews for subjective sentiment analysis," *Procedia Comput. Sci.*, vol. 142, pp. 182–189, Jan. 2018.
[18] G. Alwakid, T. Osman, and T. Hughes-Roberts, "Challenges in sentiment analysis for Arabic social networks," *Procedia Comput. Sci.*, vol. 117, pp. 89–100, Jan. 2017.
[19] B. Lantz, *Machine Learning With R*. Birmingham, U.K.: Packt, 2015.
[20] N. Alami, M. Meknassi, and N. En-Nahnahi, "Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning," *Expert Syst. Appl.*, vol. 123, pp. 195–211, Jun. 2019.
[21] I. Guellil and F. Azouaou, "Arabic dialect identification with an unsupervised learning (based on a lexicon). Application case: ALGERIAN dialect," in *Proc. IEEE Int. Conf. Comput. Sci. Eng. (CSE) IEEE Int. Conf. Embedded Ubiquitous Comput. (EUC) 15th Int. Symp. Distrib. Comput. Appl. Bus. Eng. (DCABES)*, Aug. 2016, pp. 724–731.
[22] O. Chapelle and B. A. Scholkopf Zien, "Semi-supervised learning (Chapelle, O. et al., Eds.; 2006) [book reviews]," *IEEE Trans. Neural Netw.*, vol. 20, no. 3, p. 542, Mar. 2009.
[23] K. Elshakankery and M. F. Ahmed, "HILATSA: A hybrid incremental learning approach for Arabic tweets sentiment analysis," *Egyptian Inform. J.*, to be published.
[24] B. Ohana and B. Tierney, "Sentiment classification of reviews using SentiWordNet," in *Proc. 9th It T Conf.*, 2009, pp. 1–10.
[25] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
[26] M. Youssef and S. R. El-Beltagy, "MoArLex: An Arabic sentiment lexicon built through automatic lexicon expansion," *Procedia Comput. Sci.*, vol. 142, pp. 94–103, Jan. 2018.
[27] A. Al-Thubaity, Q. Alqahtani, and A. Aljandal, "Sentiment lexicon for sentiment analysis of Saudi dialect tweets," *Procedia Comput. Sci.*, vol. 142, pp. 301–307, Jan. 2018.
[28] M. El-Masri, N. Altrabsheh, H. Mansour, and A. Ramsay, "A Web-based tool for Arabic sentiment analysis," *Procedia Comput. Sci.*, vol. 117, pp. 38–45, Jan. 2017.
[29] H. Al-Hussaini and H. Al-Dossari, "A lexicon-based approach to build service provider reputation from Arabic tweets in Twitter," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 4, pp. 445–454, 2017.
[30] M. A. Sghaier and M. Zrigui, "Sentiment analysis for Arabic e-commerce Websites," in *Proc. Int. Conf. Eng. MIS (ICEMIS)*, Sep. 2016, pp. 1–7.
[31] D. M. El-Din, "Enhancement bag-of-words model for solving the challenges of sentiment analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 1, pp. 244–247, 2016.
[32] R. Feldman, "Techniques and applications for sentiment analysis," *Commun. ACM*, vol. 56, no. 4, pp. 82–89, 2013.
[33] R. Duwairi and M. El-Orfali, "A study of the effects of preprocessing strategies on sentiment analysis for Arabic text," *J. Inf. Sci.*, vol. 40, no. 4, pp. 501–513, 2014.
[34] R. M. Duwairi, R. Marji, N. Sha'ban, and S. Rushaidat, "Sentiment analysis in Arabic tweets," in *Proc. 5th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2014, pp. 1–6.
[35] M. Abdul-Mageed, M. Diab, and S. Kübler, "SAMAR: Subjectivity and sentiment analysis for Arabic social media," *Comput. Speech Lang.*, vol. 28, no. 1, pp. 20–37, 2014.
[36] W. Hadi, "Classification of Arabic social media data," *Adv. Comput. Sci. Technol.*, vol. 8, no. 1, pp. 29–34, 2015.
[37] S. Alhumoud, T. Albuhairi, and M. Altuwaijri, "Arabic sentiment analysis using WEKA a hybrid learning approach," in *Proc. 7th Int. Joint Conf. Knowl. Discovery, Knowl. Eng. Knowl. Manage. (IC3K)*, Nov. 2015, pp. 402–408.
[38] H. S. Ibrahim, S. M. Abdou, and M. Gheith, "MIKA: A tagged corpus for modern standard Arabic and colloquial sentiment analysis," in *Proc. IEEE 2nd Int. Conf. Recent Trends Inf. Syst. (ReTIS)*, Jul. 2015, pp. 353–358.
[39] M. Salameh, S. Mohammad, and S. Kiritchenko, "Sentiment after translation: A case-study on Arabic social media posts," in *Proc. HLT-NAACL*, 2015, pp. 767–777.
[40] M. Al-Ayyoub, A. Nuseir, G. Kanaan, and R. Al-Shalabi, "Hierarchical classifiers for multi-way sentiment analysis of Arabic reviews," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 2, pp. 531–539, 2016.
[41] A. Alabdullatif, B. Shahzad, and E. Alwagait, "Classification of Arabic Twitter users: A study based on user behaviour and interests," *Mobile Inf. Syst.*, vol. 2016, Jan. 2016, p. 11.
[42] N. F. B. Hathlian and A. M. Hafezs, "Sentiment—Subjective analysis framework for Arabic social media posts," in *Proc. 4th Saudi Int. Conf. Inf. Technol. (Big Data Anal.) (KACSTIT)*, Nov. 2016, pp. 1–6.
[43] H. Al-Rubaiee, R. Qiu, and D. Li, "Identifying Mubasher software products through sentiment analysis of Arabic tweets," in *Proc. Int. Conf. Ind. Inform. Comput. Syst. (CIICS)*, Mar. 2016, pp. 1–6.
[44] R. M. Duwairi, M. Alfaqeh, M. Wardat, and A. Alrabadi, "Sentiment analysis for Arabizi text," in *Proc. 7th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2016, pp. 127–132.
[45] M. M. Altawaier and S. Tiun, "Comparison of machine learning approaches on Arabic Twitter sentiment analysis," *Int. J. Adv. Sci., Eng. Inf. Technol.*, vol. 6, no. 6, pp. 1067–1073, 2016.
[46] S. AlOtaibi and M. B. Khan, "Discovering semantic and sentiment correlations using short informal Arabic language text," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 1, pp. 198–207, 2017.
[47] M. Abdulkareem and S. Tiun, "Comparative analysis of ML POS on Arabic tweets," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 2, p. 403, 2017.
[48] A. M. Alayba, V. Palade, M. England, and R. Iqbal, "Arabic language sentiment analysis on health services," 2017, *arXiv:1702.03197*. [Online]. Available: https://arxiv.org/abs/1702.03197
[49] S. Alotaibi and C. Anderson, "Word clustering as a feature for Arabic sentiment classification," *Int. J. Educ. Manage. Eng.*, vol. 7, no. 1, pp. 1–13, 2017.
[50] A. S. Al-Subaihin and H. S. Al-Khalifa, "A system for sentiment analysis of colloquial Arabic using human computation," *Sci. World J.*, vol. 2014, pp. 1–8, Apr. 2014.
[51] A. Al Sallab, H. Hajj, G. Badaro, R. Baly, W. El Hajj, and K. B. Shaban, "Deep learning models for sentiment analysis in Arabic," in *Proc. ANLP Workshop*, 2015, pp. 9–17.
[52] F. Alqasemi, A. Abdelwahab, and H. Abdelkader, "An enhanced feature extraction technique for improving sentiment analysis in Arabic language," in *Proc. 4th IEEE Int. Colloq. Inf. Sci. Technol. (CiSt)* Oct. 2016, pp. 380–384.

[53] L. Abd-Elhamid, D. Elzanfaly, and A. S. Eldin, "Feature-based sentiment analysis in online Arabic reviews," in *Proc. 11th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2016, pp. 260–265.
[54] S. Alhumoud, T. Albuhairi, and W. Alohaideb, "Hybrid sentiment analyser for Arabic tweets using R," in *Proc. 7th Int. Joint Conf. Knowl. Discovery, Knowl. Eng. Knowl. Manage. (IC3K)*, Nov. 2015, pp. 417–424.
[55] 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," *J. Inf. Sci.*, vol. 44, no. 3, pp. 345–362, 2017.
[56] R. M. Duwairi, "Sentiment analysis for dialectical Arabic," in *Proc. 6th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2015, pp. 166–170.
[57] S. AlOtaibi and M. B. Khan, "Sentiment analysis challenges of informal Arabic language," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 2, pp. 278–284, 2017.
[58] E. Refaee, "Sentiment analysis for micro-blogging platforms in Arabic," in *Proc. Int. Conf. Social Comput. Social Media*. Cham, Switzerland: Springer, 2017, pp. 275–294.
[59] A. Assiri, A. Emam, and H. Al-Dossari, "Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis," *J. Inf. Sci.*, vol. 44, no. 2, pp. 184–202, 2018.
[60] S. R. El-Beltagy and A. Ali, "Open issues in the sentiment analysis of Arabic social media: A case study," in *Proc. 9th Int. Conf. Innov. Inf. Technol. (IIT)*, Mar. 2013, pp. 215–220.
[61] G. Badaro, R. Baly, H. Hajj, N. Habash, and W. El-Hajj, "A large scale Arabic sentiment lexicon for Arabic opinion mining," in *Proc. EMNLP Workshop Arabic Natural Lang. Process. (ANLP)*, 2014, pp. 165–173.
[62] R. M. Duwairi, N. A. Ahmed, and S. Y. Al-Rifai, "Detecting sentiment embedded in Arabic social media—A lexicon-based approach," *J. Intell. Fuzzy Syst.*, vol. 29, no. 1, pp. 107–117, 2015.
[63] N. Abdulla, R. Majdalawi, S. Mohammed, M. Al-Ayyoub, and M. Al-Kabi, "Automatic lexicon construction for Arabic sentiment analysis," in *Proc. Int. Conf. Future Internet Things Cloud*, Aug. 2014, pp. 547–552.
[64] A. Reyes and P. Rosso, "On the difficulty of automatically detecting irony: Beyond a simple case of negation," *Knowl. Inf. Syst.*, vol. 40, no. 3, pp. 595–614, 2014.
[65] S. M. Mohammad, M. Salameh, and S. Kiritchenko, "How translation alters sentiment," *J. Artif. Intell. Res.*, vol. 55, pp. 95–130, Jan. 2016.
[66] H. Wang, A. Hanafy, M. Bahgat, S. Noeman, O. S. Emam, and V. R. Bommiredd, "A system for extracting sentiment from large-scale Arabic social data," in *Proc. 1st Int. Conf. Arabic Comput. Linguistics (ACLing)*, Apr. 2015, pp. 71–77.
[67] M. Abdul-Mageed, "Modeling Arabic subjectivity and sentiment in lexical space," *Inf. Process. Manage.*, vol. 56, no. 2, pp. 291–307, 2019.
[68] R. M. Alahmary H. Z. Al-Dossari, and A. Z. Emam, "Sentiment analysis of Saudi dialect using deep learning techniques," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Jan. 2019, pp. 1–6.
[69] J. O. Atoum and M. Nouman, "Sentiment analysis of Arabic jordanian dialect tweets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 2, pp. 256–262, 2019.
[70] M. Bekkali and A. Lachkar, "Arabic sentiment analysis based on topic modeling," in *Proc. New Challenges Data Sci., Acts 2nd Conf. Moroccan Classification Soc.*, Kenitra, Morocco, 2019, Art. no. 17.
[71] M. E. M. Abo, N. A. K. Shah, V. Balakrishnan, and A. Abdelaziz, "Sentiment analysis algorithms: Evaluation performance of the Arabic and English language," in *Proc. Int. Conf. Comput., Control, Elect., Electron. Eng. (ICCCEEE)*, Aug. 2018, pp. 1–5.
[72] S. Al-Azani and E. El-Alfy, "Emojis-based sentiment classification of Arabic Microblogs using deep recurrent neural networks," in *Proc. Int. Conf. Comput. Sci. Eng. (ICCSE)*, Mar. 2018, pp. 1–6.
[73] S. Al-Azani and E.-S. M. El-Alfy, "Imbalanced sentiment polarity detection using emoji-based features and bagging ensemble," in *Proc. 1st Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS)*, Apr. 2018, pp. 1–5.
[74] Y. AlMurtadha, "Mining trending hash tags for Arabic sentiment analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, pp. 189–194, 2018.
[75] A. Alwehaibi and K. Roy, "Comparison of pre-trained word vectors for arabic text classification using deep learning approach," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 1471–1474.
[76] M. Heikal, M. Torki, and N. El-Makky, "Sentiment analysis of Arabic tweets using deep learning," *Procedia Comput. Sci.*, vol. 142, pp. 114–122, Jan. 2018.
[77] L. Lulu and A. Elnagar, "Automatic Arabic dialect classification using deep learning models," *Procedia Comput. Sci.*, vol. 142, pp. 262–269, Jan. 2018.
[78] A. Soumeur, M. Mokdadi, A. Guessoum, and A. Daoud, "Sentiment analysis of users on social networks: Overcoming the challenge of the loose usages of the Algerian dialect," *Procedia Comput. Sci.*, vol. 142, pp. 26–37, Jan. 2018.
[79] M. F. Alhamid, S. Alsahli, M. Rawashdeh, and M. Alrashoud, "Detection and visualization of Arabic emotions on social emotion map," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2017, pp. 378–381.
[80] A. Altaher, "Hybrid approach for sentiment analysis of Arabic tweets based on deep learning model and features weighting," *Int. J. Adv. Appl. Sci.*, vol. 4, no. 8, pp. 43–49, 2017.
[81] R. Baly, G. El-Khoury, R. Moukalled, R. Aoun, H. Hajj, K. B. Shaban, and W. El-Hajj, "Comparative evaluation of sentiment analysis methods across Arabic dialects," *Procedia Comput. Sci.*, vol. 117, pp. 266–273, May 2017.
[82] N. F. B. Hathlian and A. M. Hafez, "Subjective text mining for Arabic social media," *Int. J. Semantic Web Inf. Syst.*, vol. 13, no. 2, pp. 1–13, 2017.
[83] N. El-Naggar Y. El-Sonbaty, and M. A. El-Nasr, "Sentiment analysis of modern standard Arabic and Egyptian dialectal Arabic tweets," in *Proc. Comput. Conf.*, Jul. 2017, pp. 880–887.
[84] E. S. Tellez, S. Miranda-Jiménez, M. Graff, D. Moctezuma, R. R. Suárez, and O. S. Siordia, "A simple approach to multilingual polarity classification in Twitter," *Pattern Recognit. Lett.*, vol. 94, pp. 68–74, Jul. 2017.
[85] M. N. Al-Kabi, H. A. Wahsheh, and I. M. Alsmadi, "Polarity classification of Arabic sentiments," *Int. J. Inf. Technol. Web Eng.*, vol. 11, no. 3, pp. 32–49, 2016.
[86] R. Bouchlaghem, A. Elkhelifi, and R. Faiz, "A machine learning approach for classifying sentiments in Arabic tweets," in *Proc. 6th Int. Conf. Web Intell., Mining Semantics*, Nîmes, France, 2016, pp. 1–6.
[87] A. M. Rabab'ah, M. Al-Ayyoub, Y. Jararweh, and M. N. Al-Kabi, "Evaluating SentiStrength for Arabic sentiment analysis," in *Proc. 7th Int. Conf. Comput. Sci. Inf. Technol. (CSIT)*, Jul. 2016, pp. 1–6.
[88] N. Al-Twairesh, H. Al-Khalifa, and A. Al-Salman, "Towards analyzing Saudi tweets," in *Proc. 1st Int. Conf. Arabic Comput. Linguistics (ACLing)*, Apr. 2015, pp. 114–117.
[89] R. T. Khasawneh, H. A. Wahsheh, I. M. Alsmadi, and M. N. Ai-Kabi, "Arabic sentiment polarity identification using a hybrid approach," in *Proc. 6th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2015, pp. 148–153.
[90] S. B. Hamouda and J. Akaichi, "Social networks' text mining for sentiment classification: The case Facebook' statuses updates 'Arabic Spring' era," *Int. J. Appl. Innov. Eng. Manage.*, vol. 2, no. 5, pp. 470–478, 2013.
[91] S. Ahmed, M. Pasquier, and G. Qadah, "Key issues in conducting sentiment analysis on Arabic social media text," in *Proc. 9th Int. Conf. Innov. Inf. Technol. (IIT)*, Mar. 2013, pp. 72–77.
[92] J. Akaichi, "Social networks' Facebook' statutes updates mining for sentiment classification," in *Proc. Int. Conf. Social Comput. (SocialCom)*, Sep. 2013, pp. 886–891.
[93] N. A. M. Zamani *et al.*, "Sentiment analysis: Determining people's emotions in Facebook," *Appl. Comput. Sci.*, vol. 20, pp. 978–960, 2013.
[94] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub, "Arabic sentiment analysis: Lexicon-based and corpus-based," in *Proc. IEEE Jordan Conf. Appl. Elect. Eng. Comput. Technol. (AEECT)*, Dec. 2013, pp. 1–6.
[95] G. S. Solakidis, K. N. Vavliakis, and P. A. Mitkas, "Multilingual sentiment analysis using emoticons and keywords," in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. (WI) Intell. Agent Technol. (IAT)*, Aug. 2014, pp. 102–109.
[96] 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 *Proc. 2nd Int. Conf. Natural Lang. Speech Process. (ICNLSP)*, Apr. 2018, pp. 1–6.
[97] 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 *Proc. 7th Int. Conf. Comput. Sci. Inf. Technol. (CSIT)*, Jul. 2016, pp. 1–6.
[98] M. Al-Smadi, M. Al-Ayyoub, H. Al-Sarhan, and Y. Jararweh, "Using aspect-based sentiment analysis to evaluate Arabic news affect on readers," in *Proc. 8th Int. Conf. Utility Cloud Comput.*, Limassol, Cyprus, Dec. 2015, pp. 436–441.

[99] A. Alnawas and N. Arici, "Sentiment analysis of Iraqi Arabic dialect on Facebook based on distributed representations of documents," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 18, no. 3, pp. 1–17, 2019.
[100] A. M. El-Halees, "Arabic opinion mining using distributed representations of documents," in *Proc. Palestinian Int. Conf. Inf. Commun. Technol. (PICICT)*, May 2017, pp. 28–33.
[101] A. Al-Shabi, A. Adel, N. Omar, and T. Al-Moslmi, "Cross-lingual sentiment classification from English to Arabic using machine translation," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 12, pp. 434–440, 2017.
[102] F. Mallek, B. Belainine, and F. Sadat, "Arabic social media analysis and translation," *Procedia Comput. Sci.*, vol. 117, pp. 298–303, Jan. 2017.
[103] E. Refaee and V. Rieser, "An Arabic Twitter corpus for subjectivity and sentiment analysis," in *Proc. LREC*, 2014, pp. 2268–2273.
[104] N. Al-Twairesh, R. Al-Matham, N. Madi, N. Almugren, A.-H. Al-Aljmi, S. Alshalan, R. Alshalan, N. Alrumayyan, S. Al-Manea, S. Bawazeer, N. Al-Mutlaq, N. Almanea, W. B. Huwaymil, D. Alqusair, R. Alotaibi, S. Al-Senaydi, and A. Alfutamani, "SUAR: Towards building a corpus for the Saudi dialect," *Procedia Comput. Sci.*, vol. 142, pp. 72–82, Jan. 2018.
[105] A. M. Rabab'ah, M. Al-Ayyoub, Y. Jararweh, and M. N. Al-Kabi, "Evaluating sentistrength for Arabic sentiment analysis," in *Proc. 7th Int. Conf. Comput. Sci. Inf. Technol. (CSIT)*, Jul. 2016, pp. 1–6.
[106] S. R. El-Beltagy, T. Khalil, A. Halaby, and M. Hammad, "Combining lexical features and a supervised learning approach for Arabic sentiment analysis," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*. Cham, Switzerland: Springer, 2016, pp. 307–319.
[107] R. M. Duwairi and I. Qarqaz, "A framework for Arabic sentiment analysis using supervised classification," *Int. J. Data Mining, Model. Manage.*, vol. 8, no. 4, pp. 369–381, 2016.
[108] L. Al-Horaibi and M. B. Khan, "Sentiment analysis of Arabic tweets using text mining techniques," *Proc. SPIE*, vol. 10011, Jul. 2016, Art. no. 100111F.
[109] M. Hammad and M. Al-awadi, "Sentiment analysis for Arabic reviews in social networks using machine learning," in *Proc. Inf. Technol., New Gener.* Cham, Switzerland: Springer, 2016, pp. 131–139.
[110] W. Cherif, A. Madani, and M. Kissi, "A new modeling approach for Arabic opinion mining recognition," in *Proc. Intell. Syst. Comput. Vis. (ISCV)*, Mar. 2015, pp. 1–6.
[111] M. Nabil, M. Aly, and A. Atiya, "ASTD: Arabic sentiment tweets dataset," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 2515–2519.
[112] B. Al Shboul, M. Al-Ayyoub, and Y. Jararweh, "Multi-way sentiment classification of Arabic reviews," in *Proc. 6th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2015, pp. 206–211.
[113] R. M. Elawady, S. Barakat, and N. M. Elrashidy, "Sentiment analyzer for Arabic comments," *Int. J. Inf. Sci. Intell. Syst.*, vol. 3, no. 4, pp. 73–86, 2014.
[114] E. Refaee and V. Rieser, "Benchmarking machine translated sentiment analysis for Arabic tweets," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Student Res. Workshop*, 2015, pp. 71–78.
[115] H. ElSahar and S. R. El-Beltagy, "Building large Arabic multi-domain resources for sentiment analysis," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*. Cham, Switzerland: Springer, 2015, pp. 23–34.
[116] H. S. Ibrahim, S. M. Abdou, and M. Gheith, "Sentiment analysis for modern standard Arabic and colloquial," 2015, *arXiv:1505.03105*. [Online]. Available: https://arxiv.org/abs/1505.03105
[117] R. M. Duwairi and I. Qarqaz, "Arabic sentiment analysis using supervised classification," in *Proc. Int. Conf. Future Internet Things Cloud (FiCloud)*, Aug. 2014, pp. 579–583.
[118] M. N. Al-Kabi, A. H. Gigieh, I. M. Alsmadi, H. A. Wahsheh, and M. M. Haidar, "Opinion mining and analysis for Arabic language," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 5, pp. 181–195, 2014.
[119] T. Zarra, R. Chiheb, R. Moumen, R. Faizi, and A. El Afia, "Topic and sentiment model applied to the colloquial Arabic: A case study of Maghrebi Arabic," in *Proc. Int. Conf. Smart Digit. Environ.*, 2017, pp. 174–181.
[120] M. Al-Ayyoub, S. B. Essa, and I. Alsmadi, "Lexicon-based sentiment analysis of Arabic tweets," *Int. J. Social Netw. Mining*, vol. 2, no. 2, pp. 101–114, 2015.
[121] A. Qazi, R. G. Raj, G. Hardaker, and C. Standing, "A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges," *Internet Res.*, vol. 27, no. 3, pp. 608–630, 2017.
[122] M. E. M. Abo, N. A. K. Shah, V. Balakrishnan, M. Kamal, A. Abdelaziz, and K. Haruna, "SSA-SDA: Subjectivity and sentiment analysis of sudanese dialect Arabic," in *Proc. Int. Conf. Comput. Inf. Sci. (ICCIS)*, Apr. 2019, pp. 1–5.
[123] A. Qazi, A. Tamjidyamcholo, R. G. Raj, G. Hardaker, and C. Standing, "Assessing consumers' satisfaction and expectations through online opinions: Expectation and disconfirmation approach," *Comput. Hum. Behav.*, vol. 75, pp. 450–460, Oct. 2017.
[124] N. Al-Twairesh, H. Al-Khalifa, and A. AlSalman, "AraSenTi: Large-scale Twitter-specific Arabic sentiment lexicons," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2016, pp. 697–705.
[125] A. Qazi, F. Hussain, N. A. Rahim, G. Hardaker, D. Alghazzawi, K. Shaban, and K. Haruna, "Towards sustainable energy: A systematic review of renewable energy sources, technologies, and public opinions," *IEEE Access*, vol. 7, pp. 63837–63851, 2019.
[126] A. Gani, A. Siddiqa, S. Shamshirband, and F. Hanum, "A survey on indexing techniques for big data: Taxonomy and performance evaluation," *Knowl. Inf. Syst.*, vol. 46, no. 2, pp. 241–284, 2016.
[127] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal texts," *J. Artif. Intell. Res.*, vol. 50, no. 1, pp. 723–762, 2014.
[128] Z. Li, *Study on Climate Change in Southwestern China*. Lanzhou, China: Springer, 2015, pp. 1–35.
[129] M. Durairaj and C. Vijitha, "Educational data mining for prediction of student performance using clustering algorithms," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 4, pp. 5987–5991, 2014.
[130] A. Mourad and K. Darwish, "Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs," in *Proc. WASSA NAACL-HLT*, 2013, pp. 55–64.
[131] N. Al-Twairesh, H. Al-Khalifa, and A. Al-Salman, "Subjectivity and sentiment analysis of Arabic: Trends and challenges," in *Proc. IEEE/ACS 11th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2014, pp. 148–155.
[132] A. El-Halees, "Opinion mining from Arabic comparative sentences," in *Proc. 13th Int. Arab Conf. Inf. Technol. (ACIT)*, Jordan, 2012, pp. 265–271.
[133] A. Farzindar and D. Inkpen, "Natural language processing for social media," *Synth. Lect. Hum. Lang. Technol.*, vol. 8, no. 2, pp. 1–166, 2015.
[134] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, 2015.
[135] M. N. Al-Kabi, N. A. Abdulla, and M. Al-Ayyoub, "An analytical study of Arabic sentiments: Maktoob case study," in *Proc. 8th Int. Conf. Internet Technol. Secured Trans. (ICITST)*, Dec. 2013, pp. 89–94.
[136] A. Mountassir, H. Benbrahim, and I. Berrada, "Some methods to address the problem of unbalanced sentiment classification in an Arabic context," in *Proc. Colloq. Inf. Sci. Technol. (CIST)*, Oct. 2012, pp. 43–48.
[137] M. Al-Kabi, M. Al-Ayyoub, I. Alsmadi, and H. Wahsheh, "A prototype for a standard arabic sentiment analysis corpus," *Int. Arab J. Inf. Technol.*, 13, no. 1A, pp. 163–170, 2016.
[138] S. Tartir and I. Abdul-Nabi, "Semantic sentiment analysis in Arabic social media," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 29, no. 2, pp. 229–233, 2017.
[139] K. Cheng, J. Li, J. Tang, and H. Liu, "Unsupervised sentiment analysis with signed social networks," in *Proc. AAAI*, 2017, pp. 3429–3435.
[140] V. Athanasiou and M. Maragoudakis, "A novel, gradient boosting framework for sentiment analysis in languages where NLP resources are not plentiful: A case study for modern Greek," *Algorithms*, vol. 10, no. 1, p. 34, 2017.
[141] S. A. Crossley, K. Kyle, and D. S. McNamara, "Sentiment analysis and social cognition engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis," *Behav. Res. Methods*, vol. 49, no. 3, pp. 803–821, 2017.
[142] C. Catal and M. Nangir, "A sentiment classification model based on multiple classifiers," *Appl. Soft Comput.*, vol. 50, pp. 135–141, Jan. 2017.

MOHAMED ELHAG MOHAMED ABO received the B.S. degree in business administrator from Omdurman Islamic University, the High Diploma degree in information technology from Neelain University, and the M.S. degree in information technology from the University of Malaya, Kuala Lumpur, Malaysia, where he is currently pursuing the Ph.D. degree. From 1997 to 2002, he was a Computer Technician with the Ministry of Agriculture North State. From 2002 to 2006, he was a Network Administrator with the Sudanese Standards and Metrology Organization, where he has been a Head of the Information Centre, since 2010. He is the author of one book. His research interests include data analysis in Arabic language, sentiment analysis, and opinion mining.

RAM GOPAL RAJ received the B.E. degree (Hons.) in electronics and computing, in 2004, the master's degree in information and multimedia technology, in 2006, and the Ph.D. degree from the University of Malaya, Kuala Lumpur, in 2011. He is currently a Computer Scientist and a Senior Lecturer with the Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya. He has authored various journal and conference papers related to his areas of expertise. From 1998 to 2005, he was a National Athlete: Singles Sculls Medalist in various National and International Rowing Regattas.

ATIKA QAZI received the Ph.D. degree from the University of Malaya, Kuala Lumpur, Malaysia. She was a Postdoctoral Research Fellow with University of Malaya. She is currently an Assistant Professor with Universiti Brunei Darussalam, Brunei. Her research interests include opinion mining, sentiment analysis, big data analytics, information systems, e-learning, ICT, and renewable energy. She was a recipient of the Bright Spark Scholarship for her Ph.D. degree. She is also actively involved as a Reviewer of IS/SCII journals.