Heliyon 10 (2024) e39786

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Review article

Advancements and challenges in Arabic sentiment analysis: A decade of methodologies, applications, and resource development

Amani A. Aladeemy [a,b], Ali Alzahrani [c], Mohammad H. Algarni [d], Saleh Nagi Alsubari [a], Theyazn H.H. Aldhyani [e,*], Sachin N. Deshmukh [a], Osamah Ibrahim Khalaf [f], Wing-Keung Wong [g], Sameer Aqburi [h]

[a] *Dept. of Computer Science & Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India*
[b] *Faculty of Computer Sciences and Information Technology, Taiz University, Taiz, Yemen*
[c] *Department of Computer Engineering, King Faisal University, Al-Alahsa, 31982, Saudi Arabia*
[d] *Computer Science Department Al-Baha University, Al-Baha, 65779, Saudi Arabia*
[e] *Applied College in Abqaiq, King Faisal University, Al-Ahsa, 31982, Saudi Arabia*
[f] *Department of Solar, Al-Nahrain Research Center for Renewable Energy, Al-Nahrain University, Jadriya, Baghdad, Iraq*
[g] *Asia University, Taiwan*
[h] *Al-Kitab University College of Engineering Techniques, Kirkuk, Iraq*

A R T I C L E I N F O

Keywords:
Arabic sentiment analysis
Machine learning
And lexical tools

A B S T R A C T

The exponential growth of digital information, particularly user-generated content on social media and blogging platforms, has underscored the importance of sentiment analysis (SA). Arabic language sentiment analysis (ASA) involves identifying the orientation of ideas, feelings, emotions, and attitudes within Arabic text to determine whether they convey a positive, negative, or neutral sentiment. This paper presents a comprehensive review of the past decade, focusing on the utilization of SA in the Arabic language. It examines various applications, methodologies, and challenges associated with ASA, highlighting gaps and limitations in existing approaches, lexicons, and annotated datasets. The primary objective of this review is to assist researchers in identifying these gaps and limitations while offering accessible annotated datasets, preprocessing techniques, and procedures. We defined specific criteria for selecting research publications from the last 10 years, including 150 papers in our review process, while excluding earlier publications. The review utilized multiple databases, including Google Scholar, Scopus, and Web of Science. The inherent complexity of the Arabic language, due to its unique traits and diverse dialects, presents significant challenges in ASA. Moreover, the lack of annotated datasets, lexicon resources, and programming tools further complicates sentiment analysis in Arabic. The morphological variations within Arabic make it linguistically challenging. To address these issues, it is crucial to develop additional resources and construct new Arabic sentiment lexicons that account for the various dialects within Modern Standard Arabic (MSA). Our findings reveal that there is no standard public lexicon that adequately enhances the calculation of ASA across different domains, such as e-commerce, politics, public health, and marketing.

* Corresponding author.
E-mail addresses: amani.aladeemy@gmail.com (A.A. Aladeemy), aalzahrani@kfu.edu.sa (A. Alzahrani), malgarni@bu.edu.sa (M.H. Algarni), salehalsubri2018@gmail.com (S.N. Alsubari), taldhyani@kfu.edu.sa (T.H.H. Aldhyani), Sachin.csit@bamu.ac.in.com (S.N. Deshmukh), usama81818@nahrainuniv.edu.iq (O.I. Khalaf), wong@asia.edu.tw (W.-K. Wong), sameer.algburi@uoalkitab.edu.iq (S. Aqburi).

https://doi.org/10.1016/j.heliyon.2024.e39786
Received 11 January 2024; Received in revised form 16 October 2024; Accepted 23 October 2024
Available online 24 October 2024
2405-8440/

1. Introduction

Sentiment analysis (SA) utilizes tools from the fields of text mining, computational linguistics, and natural language processing (NLP) to analyses user input. The first academic publication devoted to English SA was published in 2008 by Pang et al. Opinionated texts like status updates, comments, and debates are the mainstay of the vast quantities of new digital data supplied by the proliferation of social media, online forums, and online reviews. Thus, the study of SA expanded rapidly to accommodate the storage and analysis of this mountain of data. SA has been one of the most popular NLP research areas since the early 2000s (NLP). English has been the primary focus of this field's research because of the widespread availability of relevant data sets [1].

In recent years, with the surge in sentiments expressed on online marketing and social networking sites, customers or users are frequently sharing their thoughts and experiences about products and news. These reviews serve as a resource for new consumers, manufacturers, and sales teams, providing them with valuable insight into the quality of products and helping them make informed decisions about buying, producing, or selling. The challenge with these analyses is that they are typically in text format, requiring proper processing to extract meaningful information. SA addresses this challenge by processing the reviews and categorizing or clustering them based on user requirements [2].

The primary purpose of SA is to determine whether or not a piece of writing is objective or subjective. An objective text does not contain any opinion, while a subjective text carries an opinion. Based on the feelings conveyed, a subjective writing can be further divided into broad groups, including positive, negative, or neutral. To ascertain the mood or emotion expressed in the text, SA includes evaluating text data. This can be helpful for a variety of things, such figuring out the overall tone of a document, identifying the sentiment of specific entities or subjects within the text, and determining the sentiment of specific passages or lines within the text.

As a native of the Arabian Peninsula, Arabic is a Semitic language that has its origins dating back to the 6th century and spread to Africa and Asia as the Muslim world expanded. The choice to use Arabic as a working language was made for multiple reasons. Firstly, because of its morphology and complicated nature, there are limited resources available for ASA [3]. However, ASA is becoming increasingly relevant due to the size of its audience. Secondly, the Arabic language holds significant value due to its rich history, strategic importance of its people, their region, and cultural heritage, making it both challenging and fascinating. The complexities of the language, including its various dialects and limited resources, pose numerous challenges for ASA.

The Arabic language presents distinctive difficulties for sentiment analysis as a result of its intricate linguistics. The complex morphology of Arabic, which includes numerous inflections and derivations, might obscure the underlying sentiment polarity. Furthermore, the absence of diacritics and the pervasive ambiguity surrounding voxelizations further hinders precise interpretation. One important factor to consider is the huge diversity in dialects across the Arab globe. This means that the way people express their sentiments might range greatly between different places. As a result, specific strategies need to be customized to accommodate these differences [4].

Existing methods often struggle to handle Arabic's nuanced sentiment phenomena robustly. Pervasive phenomena like code-switching between Arabic and other languages, widespread use of sarcasm/irony, and lack of full context can easily mislead models. Furthermore, the compositionality of multi-word expressions and idioms frequently deviates from their constituent words' sentiment in challenging ways. Finally, the sentiment polarity of words/phrases is often domain and dialect dependent, necessitating dynamic context-aware modeling [5].

The following are some of the key reasons that motivated us to conduct this review.

- Arabic language has a rich literary and cultural background and has recently seen a surge in online content. (SA) is crucial for Arabic data given its limited resources, complex structure and the presence of various dialects. This study aims to provide justifications for SA in Arabic based on the availability of datasets, preprocessed lexical and linguistic sources, and frequency of usage techniques for different Arabic dialects.
- After conducting a comprehensive analysis, we found that a systematic review was necessary to summarize the existing studies on ASA. This study provides an overview of the available research on the subject, based on thorough and methodical research methods.

Arabic Sentiment Analysis (ASA) assumes a pivotal role in the preservation and utilization of cultural heritage in several significant ways: Preservation of Language and Traditions: ASA fosters a deeper grasp of language intricacies, idiomatic expressions, and cultural context, achieved through the analysis of sentiment in both classical and contemporary Arabic texts. Unveiling Cultural Insights: ASA unveils profound insights into the emotions, beliefs, and values woven into Arabic literature and culture. It sheds light on the emotional and cultural dimensions of classical Arabic poetry, prose, and historical texts. Literary Analysis: ASA involves the critical analysis of classical Arabic literary works, enhancing our comprehension of their emotional and cultural components. Facilitating Cross-Cultural Exchange: ASA eases cross-cultural comprehension by examining sentiment within Arabic content and making it accessible to a global audience. This bridging of cultural divides fosters dialogue and cultural exchange. Museum Curation and Exhibition: ASA finds application in the meticulous curation of exhibitions and museum presentations centered on Arabic culture and literature. It aids in the selection and presentation of content that effectively conveys the emotional depth and cultural significance of diverse artifacts and texts. Educational Enrichment: ASA seamlessly integrates into educational tools and platforms, enriching students' understanding of Arabic literature and culture. It proves invaluable in imparting the emotional and cultural dimensions of the language.

In summary, Arabic Sentiment Analysis serves as a cornerstone in preserving and unraveling Arabic cultural heritage, empowering language appreciation, cultural exploration, literary analysis, intercultural communication, museum presentation, and educational enhancement.

1.1. Introduction to Arabic language types

Arabic, which belongs to the Semitic language family, is the official language of 27 nations and is spoken by more than 400 million people. According to research conducted by Boudad et al. (2018) and Guellil et al. (2018b), Arabic is currently ranked as the fourth most prevalent language utilized on the internet. Alongside languages like Hebrew, Aramaic (including Syriac), and Amharic, all part of the Semitic language family, Arabic is celebrated for its extensive and rich literary tradition, which has played a pivotal role in shaping Middle Eastern culture for more than 4000 years. The Arabic language encompasses a total of 28 alphabets, including 25 consonants and 3 long vowels. It is worth noting that Arabic employs a highly cursive script, where all cursive letters intricately connect with each other. Classical Arabic, Modern Standard Arabic, and Speaking Arabic are the three primary categories under which the language of Arabic is commonly subdivided when discussing it in written form as shown in Fig. 1. The language is spoken in a diverse range of countries and regions [6,7].

Classical Arabic: Arabic in its classical form is the version of the language that was utilized in the writing of the Qur'an and other works of classical literature. It is the basis for Modern Standard Arabic and is still used as a literary and formal language in some contexts [8].

Modern Standard Arabic: This is the standardized form of the language that is used in formal settings such as media, education, and government. It is based on Classical Arabic but has evolved over time to include many modern words and phrases.

Spoken Arabic or Dialectal Arabic (DA): This refers to the various spoken dialects of Arabic that are used in everyday communication and conversational level [6]. These dialects can vary significantly from region to region and often differ from Modern Standard Arabic in terms of vocabulary, grammar, and pronunciation. Some common spoken Arabic dialects include Egyptian Arabic, Gulf Arabic, and Levantine Arabic. It is worth noting that these categories are not mutually exclusive and that many speakers of Arabic are fluent in more than one variety of the language.

There is a great diversity of Arabic dialects, with varying degrees of mutual intelligibility, and there are significant differences in vocabulary, phonetics, and grammar. Arabic dialects are classified into several dialect families according to region, which are the dialects of the **Arabian Peninsula, Levantine, Maghreb, Nilotic, and Iraqi.**

The dialects of the **Arabian Peninsula** are Arabic tongues spoken by the inhabitants of the Arabian Peninsula. These tongues share some phonological and grammatical characteristics, and several dialects branch out from each tongue. Its branches: Gulf dialect, Hijazi dialect, Najdi dialect, Bahraini dialect, Yemeni dialect.

Levantine dialects are usually divided into two large groups:

Northern Levantine dialects (in western Syria and Lebanon in general but not specifically). Southern Levantine dialects (in northern Palestine and northwestern Jordan in general but not specifically).

Maghreb dialects, or "colloquial" dialects, are the total Arabic dialects spoken in the countries of the Maghreb region. These dialects include Moroccan dialect, Algerian dialect, Tunisian dialect, and Libyan dialect.

These are the dialects spoken in the Arab Nile Basin countries only. It can be divided into:

Egyptian (a. Saidi dialect b. Alexandrian dialect.), Sudanese, Chadian, Baqara.

The **Iraqi dialect** is a continuum of mutually intelligible varieties of Arabic dialects indigenous to the Mesopotamian basin of Iraq as well as extending into southeastern Turkey, Iran, Syria and Kuwait, and is spoken in Iraqi diaspora communities.

Sfor example, “جملة " أريد ان اشتري سيارة the sentence "I want to buy a car" is followed by Arabic dialects as showing in Table 1.

This one of the most challenges in ASA dialectal Arabic as showing in Table 1 to solve this we try create a lexicon include many dialects and convert it Modern Standard Arabic MSA mapping.

1.2. Evolution of ASA

The development of sentiment analysis tools for Arabic has gained increasing attention in recent years. In SA, opinions, attitudes, and emotions are identified and extracted from text. Arabic is a complex language with a wide variety of dialects, making sentiment analysis tools difficult to develop. A number of SA tools are now available for Arabic due to significant progress in this area. For analyzing text and identifying sentiment, these tools use machine learning and natural language processing.

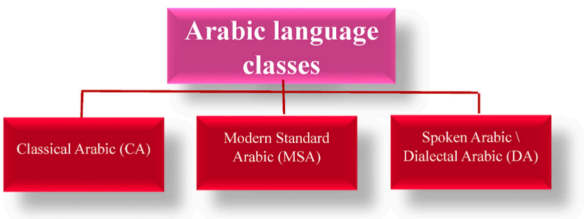

Fig. 1. Classification of Arabic language.

Table 1
One sentence with different dialects.

Arabic dialects "اللهجات العربية"	Its corresponding sentence
Gulf dialect	"ابغى اشتري سيارة "
Egyptian dialect	" عايز اشتري سيارة "
Levantine dialect	" بدي اشتري سيارة "
Moroccan dialect	"بغيت نشري سيارة "

1.3. Differentiating Arabic from English sentiment analysis

SA in English and Arabic differs in a number of ways which are summarized and enlisted as follows.

1. Script: Arabic is written in a script that reads from right to left, while English is written in a script that reads from left to right. This can affect the way text is analyzed and the tools that are used to perform the analysis.
2. Vocabulary: English and Arabic have different vocabularies, with many words having different meanings or connotations in each language. This can affect the way sentiment is expressed and interpreted.
3. Syntax: The syntax of Arabic and English are also different, with Arabic having a more complex system of verb conjugation and noun-adjective agreement. This can affect the way text is analyzed and the tools that are used to perform the analysis.
4. Emoticons: Emoticons are commonly used in English to convey emotion, but they are not as widely used in Arabic. This means that sentiment in Arabic may be more difficult to detect based on the presence or absence of emoticons.
5. Cultural differences: There are also cultural differences between English and Arabic speaking communities that can affect the way sentiment is expressed and interpreted. For example, some emotions may be more commonly expressed or considered more appropriate in one culture than in another.

Inclusively, ASA can be more challenging due to the differences in script, vocabulary, syntax, and cultural norms. It is important to carefully consider these differences when performing sentiment analysis on Arabic text.

Here is a summary of the contribution made by the survey.

1. The systematic review methodology was applied to gather relevant research studies on ASA. The number of publications was tallied annually and data was extracted from conferences, journals, and workshops.
2. The analyzed Arabic language was grouped according to their historical development, as well as online resources for linguistic and vocabulary research, annotated datasets, and various domains.
3. A thorough examination of major SA methods for Arabic text was conducted using available resources.
4. The challenges faced by researchers in evaluating sentiments in Arabic writings were also discussed.
5. The volume of studies conducted on different emotional levels, classifications, and areas for SA were introduced.
6. The final section also discussed potential future avenues for SA research for Arabic text.

The remaining parts of the paper can be broken down into the following categories:

Part 2 gives related works also, SA Progress As process of SA, and includes levels of ASA. Part 3 covers the methodology used for the survey and, covers the preliminary steps for ASA, includes sentiment lexicons, pre-processing language resources, and public annotated datasets. Part 4 talks about various ASA methods and evaluation metrics. Part 5 presents the different applications of ASA. Part 6 covers the challenges in ASA. Part 7 presents the findings of the survey. Part 8 gives conclusion of the survey and future directions.

2. Literature review

Arabic sentiment analysis (ASA) has garnered significant attention due to the complexity of the Arabic language, which includes a wide range of dialects, rich morphology, and linguistic nuances.

In recent years, Arabic sentiment analysis has emerged as a critical area of research within the field of Natural Language Processing (NLP). This surge in interest can be attributed to the proliferation of Arabic content on social media platforms and the growing need to understand public opinion in Arabic-speaking regions (Abdul-Mageed et al.) [9]. This literature review aims to provide a comprehensive overview of the key developments and influential studies that have shaped the field of Arabic sentiment analysis.

The initial phase of Arabic sentiment analysis research was characterized by efforts to address the unique challenges posed by the Arabic language. Maamouri et al. (2004) [10] highlighted the complexities inherent in processing Arabic text, particularly its rich morphological structure and the prevalence of dialectal variations. These challenges necessitated the development of specialized approaches for Arabic NLP tasks.

In response to these challenges, early research efforts focused on the creation of sentiment lexicons for Arabic. Abdul-Mageed and Diab (2012) [11] made significant contributions in this area with their work on AWATIF, a multi-genre corpus for modern standard Arabic subjectivity and sentiment analysis. This lexicon-based approach laid the foundation for more advanced techniques in Arabic sentiment analysis.

As the field progressed, researchers began to explore machine learning techniques for Arabic sentiment classification. Support

Vector Machines (SVM) emerged as a popular choice due to their effectiveness in handling high-dimensional data. Rushdi-Saleh et al. (2011) [12] demonstrated the efficacy of SVM in their work on the Opinion Corpus for Arabic (OCA), which became a benchmark dataset for subsequent studies.

Comparative studies of various machine learning algorithms further enriched the field. Duwairi et al. (2014) [13] conducted a comprehensive analysis of different classifiers, including Naive Bayes and Decision Trees, for sentiment analysis of Arabic tweets. Their work provided valuable insights into the relative performance of these algorithms in the context of Arabic text.

The advent of deep learning techniques marked a significant turning point in Arabic sentiment analysis. Convolutional Neural Networks (CNNs) showed particularly promising results for classification tasks. Al-Sallab et al. (2017) [14] introduced AROMA, a recursive deep learning model that demonstrated superior performance in opinion mining for Arabic, especially in the context of limited resources.

Long Short-Term Memory (LSTM) networks also gained traction due to their ability to capture long-range dependencies in text. Altowayan and Tao (2016) [15] applied LSTMs in conjunction with word embeddings for Arabic sentiment analysis, achieving notable improvements over traditional machine learning approaches.

The most recent advancements in Arabic sentiment analysis have been characterized by three main trends.

1. Transfer Learning and Pre-trained Models: Researchers have begun to adapt large pre-trained models, such as BERT, for Arabic language understanding. Antoun et al. (2020) [16] introduced AraBERT, a transformer-based model specifically designed for Arabic NLP tasks, including sentiment analysis.
2. Multi-dialect Sentiment Analysis: Recognizing the importance of dialectal variations in Arabic, recent work has focused on developing systems capable of handling multiple Arabic dialects. Abu Farha and Magdy (2019) [17] presented Mazajak, an online Arabic sentiment analyzer that demonstrated robust performance across various dialects.
3. Cross-lingual Sentiment Analysis: Efforts to leverage resources from other languages to improve Arabic sentiment analysis have gained momentum. Mohammad et al. (2016) [18] investigated how translation affects sentiment, providing valuable insights for cross-lingual approaches in Arabic sentiment analysis.

The field of Arabic sentiment analysis has witnessed significant advancements over the past two decades. From addressing the fundamental challenges of Arabic NLP to the application of sophisticated deep learning techniques, researchers have made substantial progress in developing accurate and robust sentiment analysis systems for Arabic text. As the field continues to evolve, future research directions may include further improvements in handling dialectal variations, the development of more comprehensive Arabic language models, and the exploration of multimodal sentiment analysis techniques for Arabic content.

2.1. SA progress

The study of sentiment analysis was initiated in the year 2000 by Pang et al. [1]. The proliferation of social media platforms, online forums, and review platforms has resulted in a significant increase in the amount of digital data available, particularly in the form of opinionated texts such as statuses, comments, and arguments. As a result, the SA sector has experienced substantial growth in order to effectively handle this influx of data.

SA has emerged as a prominent and much discussed issue in the field of NLP since the year 2000. The majority of scholarly publications tend to concentrate on the English language, which is often preferred due to the availability of standardized datasets.

2.1.1. Process of SA

In general, the SA procedure is carried out in the five steps depicted in Fig. 2. Data are first gathered through the APIs of social media platforms or websites connected to a specific domain. This dataset can come from comments made on social media, reviews, polls, or any other text-based source. Unstructured data are transformed into structured form during the data preparation stage by transliteration or by deleting unimportant and distracting stuff that is not necessary for sentiment analysis. Computing activities are carried out in the sentiment detection step to locate and the textual dataset's emotion or opinion. The fourth phase, sentiment classification, puts subjective statements into categorization groups using lexicon-based, machine learning, deep learning, or hybrid algorithms. This stage's classification groups can be further separated into moods like joy, sadness, regret, melancholy, etc. The sentiment analysis results must be visualized and reported as the last stage. This can reveal information about the overall sentiment of the Arabic text dataset and can be done using graphs, charts, or tables.

2.1.2. Levels of SA

The analysis of sentiment has been explored across various levels, ranging from documents to aspects, phrases, and sentences. Fig. 3 illustrates the different levels of SA, which include document, sentence, phrase, and aspect levels.

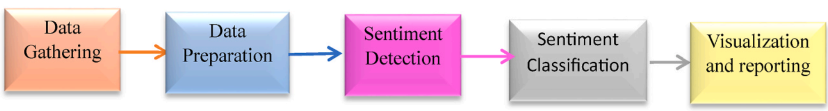

Fig. 2. General SA process.

In general, SA can be classified into four distinct levels.

1. **Document-level sentiment**: This entails assessing a text's overall tone. The goal is to classify the text as good, negative, or neutral. At the document level, just one entity is considered about. This level requires determining if it's favorable, negative, or neutral. For instance, the SA system assists in assessing whether a review of the most recent movie reflects a positive, negative, or neutral emotion about it.
2. **Sentence-level sentiment**: it is typically carried out in two subphases. The first phase involves classifying a given sentence as positive, negative, or neutral, while the second phase involves determining the strength or intensity of the sentiment expressed in the sentence. There are various types of sentences encountered in SA, such as opinionated, factual, mixed sentiment, and neutral sentences. Opinionated sentences contain an explicit or implicit opinion about a particular topic, and are generally easier to analyze for sentiment since the opinion is clearly stated. Factual sentences provide information about a specific topic but do not express an opinion, and are usually classified as neutral in sentiment. Mixed sentiment sentences, on the other hand, contain both positive and negative sentiment about a topic, making them more challenging to analyze. Neutral sentences do not express any strong sentiment, positive or negative, about a topic.
3. **Phrase-level sentiment:** it involves determining the sentiment expressed in a specific phrase or group of words within a larger text. It may be favorable, negative, or neutral. SA algorithms can be used to determine the sentiment of a phrase in a text. SA on opinion words at the phrase level classifies each phrase, which may have numerous or single aspects. This level of analysis is useful for product reviews, where a single aspect may be expressed in a phrase [19].
4. **Aspect-level sentiment:** Feature level is another name for aspect level. It conducts more precise SA analysis. This involves analyzing the sentiment of specific aspects or features within a document or text. For example, you might analyze the sentiment of a product review in relation to its price, quality, or customer service. Aspect level SA goals to detect and classify the corn of specific aspects or features within the text.

Each level of SA can provide different insights and be useful for different purposes. For example, document-level SA can be used to understand the overall sentiment of a large collection of texts, while aspect-level SA can be used to identify specific areas of satisfaction or dissatisfaction in customer feedback.

3. Methods

The following provides an outline of the procedures that were carried out in order to complete this analysis of ASA.

3.1. Review protocol development

The present systematic review conducted a comprehensive search across prominent electronic databases and prominent conferences in the field to identify pertinent research studies. Subsequently, research selection was conducted based on the use of inclusion and exclusion criteria. The study subjects were designed and subsequently finalized, leading to the selection of specific research projects. These studies were then subjected to a comprehensive analysis of their outcomes.

3.2. Querying the research

This research presents a rigorous review of ASA methodologies and techniques. ASA researchers' lexical and lexicon resources and tools are also included. To conduct the systematic review efficiently, Table 2 lists research questions.

3.3. Data sources

Before conducting the search, researchers chose appropriate electronic databases to locate pertinent research publications. Utilizing various academic search engines—specifically Google Scholar, Science Direct, Scopus, and IEEE Xplore—they sought out research papers. These search engines, accessible through their respective URLs, were employed to locate scholarly literature relevant to the study. The majority of the identified research papers were published in prestigious conferences centered on natural language processing, linguistics, and Arabic sentiment analysis. Notably, Google Scholar offered comprehensive coverage of these publications.

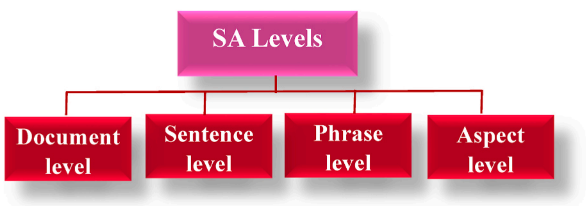

Fig. 3. Sa levels.

Table (2)
Research questions.

Question numbers	Research questions
Q1	Why Arabic Sentiment Analysis?
Q2	What is the existing standard dataset for ASA and annotation approaches?
Q3	Which sentiment analysis methods are the most frequently used in ASA?
Q4	what are limitations and challenges for ASA?
Q5	What are the most used Arabic data preprocessing and text representation approaches for ASA?
Q6	What are future directions & recommendations from this systematic review for ASA?

To ensure the inclusion of only relevant research publications, duplicate papers from Science Direct, Scopus, and IEEE Xplore were eliminated from consideration.

3.3.1. Inclusion/exclusion criteria

A systematical keyword-based advanced search has been followed to retrieve the significant research studies from the e-databases Various keyword terms were searched including ("Sentiment Analysis for Arabic"), ("Arabic opinion mining") and ("Arabic text mining"). In Fig. 3, We offer these articles that were found using keywords annually. We set out our study of articles over the past 10 years from 2012 to 2022. It is found that the social media data plays a huge role for analyzing Arabic text sentiments, especially in the Arab world, where every phenomenon exists; most people go to social media to express their feeling, moods, and attitudes. Due to disturbing things that were going on during the time in the Middle East of The Arab Spring and also due to the emergence and spread of the Covid-19 virus from.

2014–2020. However, several studies have been presented for ASA. In 2022, the highest rise in ASA research articles counts occurred. Fig. 4 demonstrates the number of research article last ten years. In light of this, a comprehensive investigation was conducted to integrate all relevant research findings within this field. The inclusion criteria encompassed conferences, journals, workshops, as well as master's and PhD theses, in accordance with the specified exclusion criteria. In order to do the electronic search, persons conducted searches on conferences and journals related to Natural Language Processing (NLP), ASA, and linguistics. The inquiry yielded 150 research studies, as delineated in Fig. 5, subsequently winnowed to 110 through title scrutiny, 100 through abstract assessment, and 58 through comprehensive evaluation of full-text content. Following this winnowing process, the 87 remaining research articles underwent thorough analysis to curate a definitive compilation of research studies.

3.3.2. Survey outcomes

The objective of this study is to catalogue the previous studies that have been conducted on ASA. This objective is presented in Table 1 in the form of research questions. The year-by-year status of publications on ASA, as well as the provenance of those articles, have been investigated, and the results are depicted in Fig. 6 below.

Examining articles published during the past decade, from 2012 to the 2022; a visual representation of the publishing status each year from 2012 to 2022 is provided in Fig. 3. The analysis of the graph shows that this area of study is expanding at a rapid rate. Analyses have shown that most research articles on Arabic sentiment analysis are published in a wide range of conference proceedings and journals. As can be seen in Fig. 6, around 58 % of research publications appear in journals, 26 % in conferences, 10 % in workshops, and the remaining 6 % are presented at either a thesis defense or the Congress of the Association for Computational Linguistics. Journals accounted for the bulk of the research publishing, while conferences came in second.

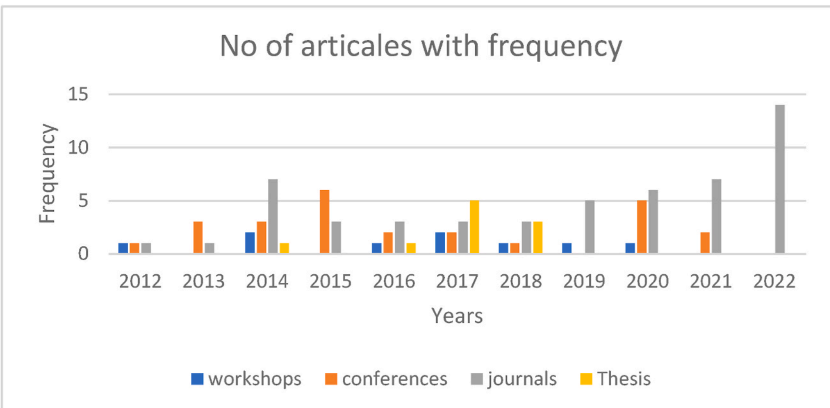

Fig. 4. Number of research articles for ASA.

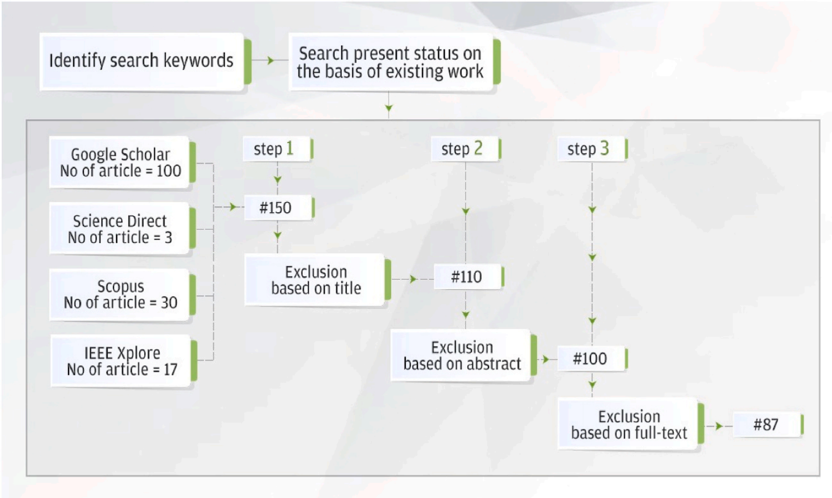

Fig. 5. Review method used.

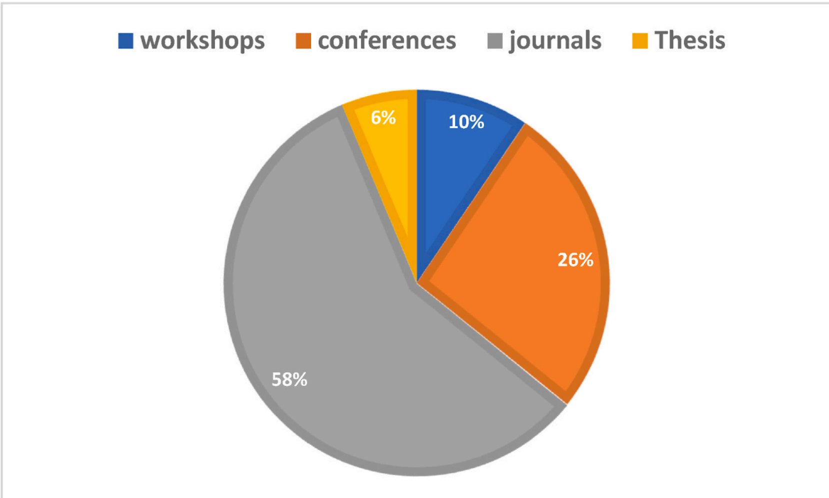

Fig. 6. Sources of publications.

3.4. Prefaces for ASA

3.4.1. Datasets

Dataset gathering is the initial step in SA. Twitter, blogs, discussion forums, and review sites for movies, products, and travel have been utilized for ASA. The information in Table 3 helps answer the research question RQ2. Several annotated datasets of tweets and reviews are available online. This table summarizes ASA's online annotated datasets.

Twitter sentiment dataset for Morocco (MSTD) [20]**:** dialectal sentences

With four-way sentiment classification that was extracted from tweets. The binary dataset piques their attention.

Arabic Sentiment Tweets Dataset (ASTD) [21]**:** Arabic tweets amount roughly 10K. Positive, negative, neutral, or mixed tweet annotations are possible. Includes nearly 10,000 Arabic emotion tweets divided into four categories: subjectively positive, subjectively negative, subjectively mixed, and objective. Each annotation was hand-annotated using Amazon Mechanical Turk (AMT).

(ArTwitter) Twitter dataset for Arabic Sentiment Analysis [22]**:** There are 2000 tweets on it, inscribed in equally Jordanian and MSA. Positive and negative annotations are added to tweets.

The BBN Blog Posts Emotion Corpus is a randomly selected subset of 1200 Levantine dialectal sentences in the BBN Arabic-Dialect-English Parallel Text [23]. Social media sentences translate. CrowdFlower, a crowdsourcing service, manually categorised sentiment into positive, negative, and neutral categories.

Saif et al. [24] introduce the Syria Tweets Sentiment Corpus (SYR), 2000 Syrian tweets (a country where Levantine dialectal Arabic is commonly spoken). Twitter API calls in May 2014 obtained the tweets. Portage Machine Translation translated the dataset into

Table 3
Datasets for Arabic Sentiment Analysis.

Authors	Dataset	Classified of Arabic language	Sentiment			Total
			Positive	Negative	Neutral	
[20]	**MSTD**	Moroccan dialectal sentences (DA)	866	2769	–	**3635**
[21]	**ASTD**	Egyptian multidialectal Arabic tweets	799	1684	6691	**9174**
[22]	**ArTwitter**	MSA and Jordanian dialect	1000	1000	–	**2000**
[23]	**BBN**	Levantine dialectal sentences	498	575	126	**1199**
[24]	**SYR (Syria)**	Syrian dialectal tweets	448	1350	202	**2000**
[25]	**MD**	multidialectal Arabic tweets	734	377	–	**1111**
[26]	**RR**	multidialectal Arabic tweets	1752	1941	3698	**7391**
[Twitter API]	**MASTD**	multidialectal Arabic tweets	415	777	658	**1850**
[27]	**ArSAS**	dialectal sentences (DA)	4643	7840	7279	**19,762**
[26]	**GS**	dialectal sentences (DA)	559	1232	2400	**4191**
[28]	**Main-AHS**	multidialectal Arabic	628	1398	–	**2026**
[29]	**NU**	Egyptian dialectal tweets	1046	976	724	**2746**
[29]	**NBI**	Saudi and MSA	2686	3225	3745	**9656**
[30]	**SR**	Egyptian dialectal tweets	1612	1604	1604	**4820**
[31]	**LABR**	multidialectal Arabic	8224	8224	–	**16,448**
[32]	**MPQA**	multidialectal Arabic	4597	5399	–	**9996**
[33]	**Multi-Domain Resource**	multidialectal Arabic	24948	6650	–	**31,598**
[34]	**AJGT**	Jordanian dialect	720	1080	–	**1800**
[35]	**ASTC**	multidialectal Arabic tweets	29849	28902	–	**58751**

English. The positive, negative, and neutral tree emotions are manually marked. This dataset underwent preprocessing to remove user IDs and non-Arabic words.

The 1111 tweets in the Arabic language that make up the **MD: Mourad Darwish** Dataset were annotated by hand by two native Arabic speakers. The tweets were collected by Mourad and Darwish [25] between January and December of 2012. There are 377 positive tweets and 734 bad ones. The SYR dataset underwent pre-processing to get rid of Latin characters, digits, and proper names (i. e., URLs, emails).

RR Dataset: Refaee and Rieser are the authors of this dataset [26]. There were manually tagged tweets that were 833 positive, 1848 negative, and 3685 neutral.

Dataset of Minimal Arabic Tweets with an Attitude (MASTD): Using the Twitter API, we were able to collect almost 6000 tweets, although we only ended up using about 1850 of them. Other people are ignored because of irony, disagreement, or repetition. Annotating tweets as neutral, bad, or positive is a laborious process.

ArSAS [27]**:** More than 21,000 tweets. Positive, negative, neutral, and mixed tweets are annotated.

Gold Standard ASA using Twitter Data [26] (GS): Tweets exceed 8000. Good, bad, neutral, or mixed tweets are categorised. The major difficulty with this dataset is that it just provides tweet ids and sentiment polarities, leading the acquired tweets to sometimes deviate. Only 4191 of the 8 K tweets were collected.

3.4.1.1. Arabic health services dataset. A Twitter-based Arabic analysis dataset was presented by the authors of [28]. Two classes (positive and negative). The dataset has 2026 tweets, with 1398 unfavorable and 628 positives. We call it “Main-AHS" after its authors.

The NU Talaat et al dataset [29]**:** This dataset was collected and annotated. 3436 tweets are largely in Egyptian. These tweets are used to create 2746 training and 683 test sets. Ten46 positive tweets, 976 negative tweets, and 724 tweets with no polarity make up the training set. Test dataset distribution: 192 neutral, 263 positive, 228 negative. The article’s author provides this dataset.

KSA CSS (NBI): Dr. Nasser Al-Biqami led this Saudi Arabian research facility’s collection of this dataset [29]. This dataset contains mostly Saudi and MSA tweets with a few Egyptian and other dialects. This dataset also had 9656 training tweets and 1414 test tweets. 2686 positive, 3225 negative, and 3745 neutral tweets are in the training set, whereas 403, 367, and 644 are in the test set.

Shoukry Rafea (SR) [30]**:** 1604 neutral, 1612 positive, and 1612 neutral tweets from Egypt make up this dataset. This dataset was pre-processed to remove mentions, hashtags, and URLs before we received it.

LABR: A Massive Arabic Book Review Dataset Arabic book reviews were presented in Ref. [31]. Almost 63,000 Arabic book reviews are in this collection. In March 2013, the website2 collected book reviews. The review ID, user ID, book ID, rating (1–5), and review text are included in each book review.

Multi-Perspective Question Answering News stories were displayed in Ref. [32]. The collection includes news stories from a range of sources manually tagged for opinions and other private states (i.e., beliefs, emotions, sentiments, speculations, etc.)

Huge Arabic Multi-domain Resources for Sentiment Analysis [33] suggested 33K automatically annotated Movie, Hotel, Restaurant, and Product Reviews. The datasets cover four areas.

1. **HTL, or Hotel Reviews,** as TripAdvisor3 reviews were removed.
2. **Reviews of Eating Establishments (RES):** Reviews for the restaurant were stolen from Qaym4 and TripAdvisor.
3. **The Movie Reviews (MOV)** dataset is part of the movies domain and was constructed by scraping revie
4. Ws from Elcinemas5 for a total of roughly 1000 films.

5. For the Products domain, a dataset of reviews was scraped from the Souq6website. Iv. **Product Reviews (PROD).** Customer feedback from Egypt, Saudi Arabia, and the UAE is included in the dataset.

There are 1800 tweets in the **Arabic Jordanian General Tweets (AJGT)** [34] dataset, and they've all been labeled as good or negative. In this case, we have split the data set in two: a training set those accounts for 80 % and a testing set those accounts for 20 %. The training set has 720 distinct positive and negative categories. For the purpose of the test, 180 classes are split evenly between the “positive" and “negative" categories.

3.4.1.2. ASTC: Arabic Sentiment Twitter Corpus dataset. Arabic Sentiment Twitter Corpus (ASTC) [35] was collected in April 2019 from Twitter.

In total, there are 58751 Arabic tweets with positive and negative labels. It is divided into

Two parts: a training set and a test set. There are 22,626 negative classes and 22,810

Positives classes in training set. There are 5703 positive and 5656 negative classes in the testing set.

3.4.2. Preprocessing linguistic resources

Pre-processing of linguistic resources in Arabic sentiment analysis typically involves several steps, incorporating tokenization, stop-word elimination, text cleaning and stemming.

1. Text cleaning involves removing any unwanted characters or symbols from the text, such as punctuation marks or numbers. This step is important to ensure that the text is in a consistent format.
2. The technique of tokenizing involves separating the text into single words or phrases. This step is important for the analysis of individual words and phrases, and their relationships.
3. Stop-word removal involves deleting meaningless words like “the", “is", and “and". Eliminating these words reduces text size and emphasizes more meaningful terms.
4. Stemming and lemmatization are techniques for reducing words to their base forms.

This step is useful for reducing variations of words to a common form, which can help to increase the accuracy of the analysis.

Once these pre-processing steps have been completed, the text can be used for further analysis, such as feature extraction, classification, and visualization.

Data processing cleans up the data set each word can be tokenize into a single one it makes difference when we say (x, y, z) and (xyz) cleans the data to provide the highest performance of the model this is the only difference when the data is excellently cleaned it has the text representation very clear for classify but if we delete the comma the model will deal as one word and this will be overfit.

Preprocessing Procedures:

Many steps applied to the raw Arabic text data, such as.

1. Tokenization and segmentation techniques used to handle Arabic's complex morphology.
2. Text cleaning steps like removing URLs, usernames, HTML tags, diacritics stripping, handling of special characters, or other language-specific considerations. Etc.
3. Normalization or standardization procedures (e.g., handling different Arabic dialects or colloquial varieties).
4. Encoding steps: unicode representation, handling Arabic orthographic variations
5. Techniques for data cleaning, noise removal, or handling missing values.

3.4.3. SWN: SA lexical resource

Researchers have manually or WordNet-built SA lexical resources. SWN is an SA-focused lexical resource. Polarity-based annotation of all WordNet synsets—positive, negative, and neutral—created SWN. SWN is a sentiment analysis lexicon. It assigns positive or negative sentiment polarities and scores to each word in a language. The resource adds sentiment to WordNet synsets, a lexical database of English words. SentiWordNet synsets have positive, negative, and objective scores. The objective score indicates neutrality, whereas the positive and negative values indicate the synset's sentiment [2]. guides readers through creating a manually annotated dataset and terminology.

Here are some example lexical resources that could be leveraged for Arabic sentiment analysis:

ASA is an essential field for understanding public opinions and emotions. Let's explore some resources and tools related to this topic.

1. Arabic Senti-Lexicon:

The Arabic Senti-Lexicon is a publicly available resource containing 3880 positive and negative synsets. These synsets are annotated with information such as part of speech, polarity scores, dialects, and inflected forms [36].

Additionally, there's a Multi-domain Arabic Sentiment Corpus (MASC) with 8860 positive and negative reviews from various domains [37].

2. SANA (Subjectivity and Sentiment Analysis of Arabic):

SANA is a large-scale cross-lingual lexical resource specifically designed for subjectivity and sentiment analysis of Arabic and its dialects. It covers communication contexts like chats, online news, YouTube comments, and tweets [36].

3. ArSenL (Arabic Sentiment Lexicon):

ArSenL was created by relying on several resources, including English WordNet (EWN), Arabic WordNet (AWN), English SentiWordNet (ESWN), and SAMA (Standard Arabic Morphological Analyzer). It's a valuable resource for sentiment analysis in Arabic [38].

4. Building Large Arabic Multi-domain Resources:

Researchers have generated large multi-domain datasets for sentiment analysis in Arabic. These datasets include 33,000 annotated reviews covering movies, hotels, restaurants, and products. Lexicons were also built from these datasets [39].

5. Deep Analysis of Arabic Sentiment Classification:

This study explores the effect of lexicon expansion on sentiment analysis. Researchers describe the steps followed to build lexicons and perform tasks related to enriching existing resources [40].

3.5. Feature extraction techniques

Text feature extraction converts raw text input into numerical features for machine learning applications including classification, clustering, and recommendation. Text feature extraction methods include.

1) Bag-of-words (BoW): BoW expresses text data as word counts. It ignores word order and solely considers document frequency.
2) TF-IDF weights words based on their document frequency and corpus inverse frequency.
3) Word embeddings: large text corpora are used to learn dense vector representations of words. Sentiment analysis, machine translation, and information retrieval use them to record semantic and syntactic links between words.
4) N-grams: N-grams are n-word sequences from a text. They capture document words' context and syntax.

Probabilistic topic models find hidden subjects in text. Each document is a blend of subjects with probability distributions over words.

3.6. Programming tools

These programming tools and libraries are used for ASA.

1 Arabic Preprocessing Libraries:
  - Tashaphyne - The Python package provides functionality for Arabic lightweight stemming, tokenization, and POS tagging.
  - QCRI Anlatic Analyzer - Tokenization, diacritization, stemming and other preprocessing tools.
  - MADAMIRA - Performing morphological analysis, tokenization, and POS tagging specifically for the Arabic language.
2 Arabic NLP Toolkits:
  - CAMeL Tools - Collection of tools designed for the manipulation and analysis of Arabic text data.
  - NLTK-ArabicSentimentAnalysis - Toolkit extending NLTK for ASA.
  - SAFAR (System for Arabic Fundamental and Applied Research) - Suite of linguistic analyzers.
3 Word Embeddings/Language Models:
  - AraBERT - BERT model pretrained on Arabic corpora from Google AI.
  - AraVec - Word embedding models for Arabic from Solvemate.com.
  - Arabic-ELECTRA - ELECTRA language model pretrained on Arabic data.
4 Deep Learning Frameworks:
  - PyTorch - With Arabic language support through packages like torcharabic.
  - TensorFlow/Keras - Frameworks supporting Arabic text processing.
  - AllenNLP - LibraryBuilt for Arabic among other languages.

Based on a wide range of significant and prominent publications in the field of ASA, it can be inferred that the NLTK, Gensim, and TextBlob libraries are the most valuable for Python ASA tasks. Regarding Java ASA libraries, it can be inferred that Weka and CoreNLP tools are widely utilized and have demonstrated excellent performance in this particular research field [41].

4. SA methods and evaluation metrics

4.1. SA methods

The three methods for SA that are most frequently used are lexicon-based approach, ML approach, and hybrid approach, according to the thorough survey. Additionally, researchers are

Always looking for more efficient methods to complete the task with higher accuracy and less computing expense. Overview of the many approaches utilized in sentiment analysis, as shown in Fig. 7 [42]. These methods comprehend the whole task scenario and method workflow.

4.1.1. Lexicon based approach

Lexicon-based sentiment analysis uses a pre-defined lexicon to classify a text's sentiment. Each word in the lexicon is assigned a sentiment score, usually positive or negative, and the text's overall sentiment is calculated by summing all the sentiment values.

A dictionary determines a word or sentence's semantic orientation (polarity) in the unsupervised lexicon-based technique. Each word in the dictionary has a simple polarity value (+1, −1, or 0 for positive, negative, or neutral, respectively) and a polarity "strength" value (for positive polarities, for example, the range (+1 to +5)), where words with polarities of +5 are significantly more positive than those with +1 [2].

Some professionals consider the lexicon-based approach unsupervised because it doesn't require training data [43].

The fundamental drawback of lexicon-based approaches is that they are heavily domain-focused, and terms from one domain cannot be used in another [44].

In sentiment analysis, words are often assigned a positive or negative sentiment score based on how they are used in a particular domain or context. Such as, a word like "love" may be considered positive in general, but if it is used in a negative context (such as "I love how frustrating this project is"), it may be assigned a negative sentiment score. Similarly, words like "hate" or "dislike" may be considered negative in general, but if they are used in a positive context (such as "I hate how much I love this movie"), they may be assigned a positive sentiment score. It's important to consider the context in which words are used when performing sentiment analysis. Lexicon-based approaches primarily employ two methods: Statistical and Corpus-Based Approach, the following points show Advantage and Disadvantage for it.

There are several benefits to using a lexicon-based approach to SA.

1. Simplicity: A lexicon-based approach is relatively simple, as it relies on a fixed list of words and corresponding sentiment scores.
2. Speed: A lexicon-based approach can be very fast, as it does not require any training or optimization.
3. Interoperability: Lexicons can be shared and used across different systems, making it easy to compare results and collaborate with others.
4. There are also some limitations to using a lexicon-based approach:

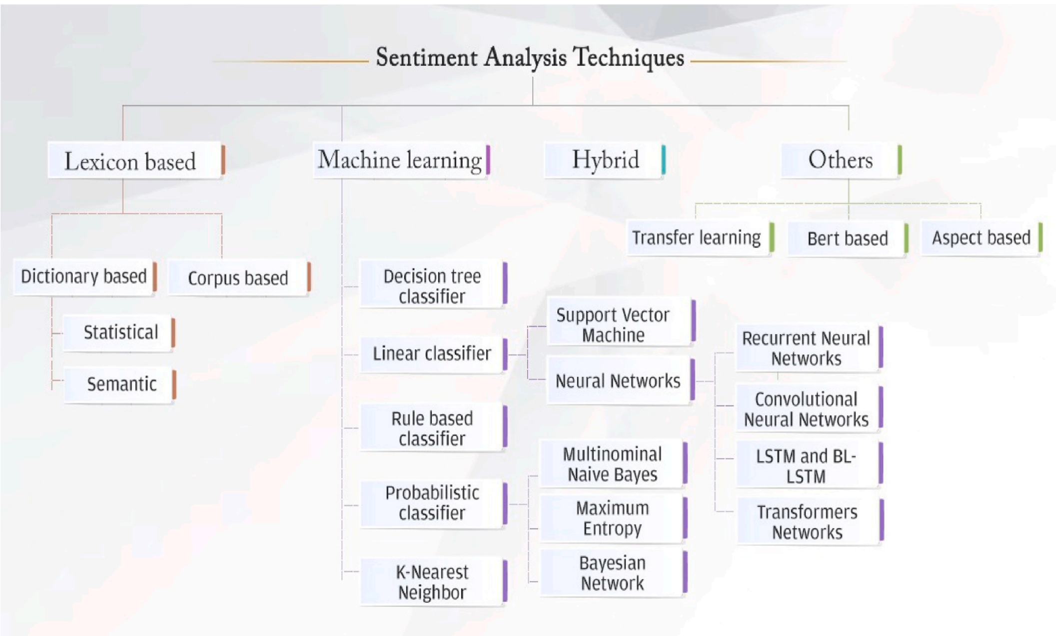

Fig. 7. Sentiment analysis approaches.

5. Limited coverage: A lexicon-based approach may not be able to accurately classify the sentiment of words or phrases that are not included in the lexicon.
6. Lack of context: A lexicon-based approach does not take into account the context in which words are used, which can lead to incorrect sentiment classifications.
7. Sensitivity to lexicon design: The quality and coverage of the lexicon can have a significant effect on the precision of the ASA.

4.1.2. ML approach

ML is a branch of artificial intelligence (AI) that involves teaching models to recognize patterns in data and make predictions or judgements without being specifically taught to do so. There are several different types of ML, each with their own strengths and weaknesses, including.

1. Supervised learning: In supervised learning, a model is trained on a labeled dataset, where the inputs (e.g., images, text, etc.) have already been labeled with their corresponding outputs (e.g., labels, classifications, etc.). The model then learns to predict the output for new, unseen input based on the patterns it has learned from the labeled dataset.
2. Unsupervised learning: In unsupervised learning, a model is trained on an unlabeled dataset. The model then tries to learn the underlying structure of the data, such as grouping similar inputs together or identifying patterns that are indicative of certain characteristics.
3. Semi-supervised learning: While learning semi-supervised, a model is trained on a partially labeled dataset, where only some inputs correlate to outputs.
4. A reinforcement learning agent aims to maximize a reward signal by acting in a certain way.
5. ML's deep learning branch that makes use of neural network topologies with numerous layers. It is frequently used in speech recognition, computer vision, and other fields where there is a wealth of data and computing power.

Because of its accuracy, supervised learning methods are becoming more popular. Before using real data, these algorithms must be taught on a training set. Text can yield characteristics. Syntactic and linguistic components are used in machine learning to classify sentiment in conventional text categorization. The categorization model assigns record attributes to class labels. The model predicts the class label for an unknown class instance. Categorizing instances with one label is tough. "Soft classification issues" involve assigning probabilistic labels to events. ML lets systems learn without programming. Sentiment analysis software can discern context, sarcasm, and incorrect word usage. Common supervised learning approaches include.

4.1.2.1. Naive bayes (NB). Naive Bayes technique is based on the Bayes theorem, which expresses the likelihood of an event occurring given that another event has already occurred, and assumes that the two events are independent of one another, is the foundation for the family of probabilistic algorithms known as naive bayes. Both categorization and training use the NB approach. They discovered that the NB classifiers produced superior outcomes than the DT classifiers in the study of [45]. Social network classification using machine learning. Along with an SVM model, they also suggested an NB model [46–48].

4.1.2.2. SVM-support vector machine. Assistive vector machines SVMs are a sort of supervised ML technique that can be applied to classification or regression applications. The goal of SVMs is to identify the optimal boundary (or "hyperplane") for classifying the data. In work of [49] used Support Vector Machines & NB to give tweets written in Arabizi an emotion or polarity designation. The study's findings demonstrate that SVM accuracy outperforms Naive Bayes accuracy. For the experiment [50], also used Twitter tweet data. They discovered that compiling the comments for microblogs was advantageous [51]. identified Facebook comments about Arabic news using a Gaussian kernel SVM classifier for Arabic slang.

4.1.2.3. LR- logistic regression. A supervised learning approach used for categorization issues is logistic regression. It is employed to forecast the likelihood that an instance will belong to a particular class. The objective of logistic regression is to train a mathematical function that can map the input characteristics to the output label given a dataset with input features (also known as independent variables) and a binary output label (also known as dependent variable). This function is represented by a set of parameters (often referred to as weights or coefficients), which are learned from the training data using optimization algorithms. In the work of [52], SVM is one of the most often used algorithms, BNB, MNB, LR, SGD, and KNN. Found that SVM and LR are the best two classifiers. According to the LR model [53], there is minimal to no multicollinearity between the predictive variables, and the dependent variable is binary.

4.1.2.4. XGBoost. A well-liked and effective open-source implementation of the gradient boosting algorithm is called XGBoost, for ML. XGBoost stands for "Extreme Gradient Boosting" and is designed to handle large-scale data and deliver high performance. The algorithm uses decision trees as the basic building blocks, and trains an ensemble of trees to optimize an objective function by minimizing the prediction error on the training data. XGBoost has proven to be effective on a wide range of problems, including regression, classification, and ranking, and has won many Kaggle competitions. It is available in many programming languages, including R, Python, and Julia, and provides a flexible and easy-to-use interface for tuning parameters and using custom loss functions.

4.1.2.5. DT- decision tree. DT is standing for Decision Tree, supervised learning method that uses the training example to construct a tree in order to categories the text's polarity. Data is recursively split into sections by DT using a criterion. In supervised learning, a decision tree is used to classify the data, and in unsupervised learning, it is used to perform clustering. The decision trees can be created by several algorithms like ID3, C4.5, C5.0 and CHAID. The most common algorithms used to generate decision trees are C4.5 and C5.0. They are also widely used in Random Forest and gradient boosting algorithms. In the work of [54] Study of Arabic Twitter Sentiment Using Different Machine Learning Techniques The outcomes demonstrated that the DT classifier outperformed the other ML Methods.

4.1.2.6. RF- Random Forest. A grouping technique for categorization is RF. and regression that is built by creating multiple DT and combining their predictions. The idea behind using multiple decision trees is that, if each tree is trained on a different subset of the data, and each tree makes predictions independently, then the ensemble of all the trees will make more accurate predictions than any individual tree.

4.1.2.7. K-nearest neighbors (KNN). KNN is a supervised learning technique that is employed in machine learning as a classifier. It searches for similarities between a given vector and another vector present in the collection. The k value and the distance metric are the two key parameters that must be set for KNN. The Euclidean function is applied to determine the distance, where $K = 3$. Based on a numerical analysis, k was set to 3, and the results are best at this value. The KNN matched fresh vectors to k of its nearest neighboring training instances. KNN is one of the widely utilized techniques for ASA, as seen in Refs. [55,56]. In the work of [57] 8000 customer reviews gathered from social media, Google Play, and the app store are part of a brand-new dataset used for this study. The dataset is subjected to the application of several different approaches, and the findings demonstrate that the (KNN) method produces the maximum accuracy when compared to other employed methods.

4.1.2.8. Maximum entropy (ME). The ME principle is often used in the construction of classifiers, particularly in the context of text classification and information retrieval. The ME classifier, also known as the logistic regression classifier or the exponential classifier, is a probabilistic model that makes predictions based on a probability distribution over the possible outcomes.

4.1.2.9. Semi-supervised learning. When the training dataset in this instance includes data with labels and without labels Semi-supervised education is a method of training machine learning models on a dataset that is partially labeled, meaning that only a portion of the examples have been labeled with their correct output or desired values. Aiming for semi-supervised learning is to use information in the labeled examples to learn the underlying patterns in the data that can be is utilized to foretell the labels of the samples that aren't labeled. In the work of [58] tapped into the synset relationships to apply semi-supervised learning to disseminate the scores in the Arabic WordNet. In Ref. [59] work is to employ semi-supervised online learning to produce a constant flow of consistently annotated Arabic twitter data.

4.1.3. Hybrid method

Hybrid sentiment analysis uses different algorithms to examine text data and assess its sentiment. Rule-based and machine learning sentiment analysis are hybrid methodologies. A rule-based technique employs predefined rules to recognize text sentiment, while a machine learning approach uses a trained model to determine sentiment based on dataset patterns. The rule-based technique, which is fast and straightforward to implement, and the machine learning approach, which can handle massive amounts of data and adapt to new patterns, could be combined in a hybrid approach. HILATSA uses lexicon-based and machine learning methods to identify tweet sentiment polarity [60]. In Ref. [61] proposed a semantic orientation-machine learning hybrid strategy. They demonstrated that lexical-based SVM can obtain the same outcomes as their hybrid technique. Their hybrid classifier has 84 F-measure and 84.01 % accuracy.

4.1.4. Other approaches

4.1.4.1. Sentiment analysis by aspect (ABSA). Understanding the aspects or entities being discussed in a text and the sentiment conveyed towards them is the goal of aspect-based sentiment analysis (ABSA), a natural language processing problem. As such, it goes beyond the scope of conventional sentiment analysis, which mainly entails labelling a text as generally good, negative, or neutral. ABSA can help you gain a more nuanced comprehension of the sentiments represented in a text by isolating the topics of conversation and assigning emotions to them. This has a number of potential uses, including market research and brand reputation management. Extracting the aspects of the text's primary entities and then determining the sentiment expressed for each aspect is the focus of aspect-based sentiment analysis (ABSA) [62]. ABSA studies of English text can be seen in places like movie reviews [63], reviews of electronic devices [64], and evaluations of local eateries [65,66]. The author of [67] is the first study of Arabic text in ABSA.

4.1.4.2. Transfer learning-based models. Transfer learning uses a model from one task to start a model for another. Use the architecture and weights of the pre-trained model to create a new model and train it on the new job. This is useful when the new work has minimal labeled data or is related to the old task.

Transfer learning for sentiment analysis might start with a pre-trained language model. For sentiment analysis, BERT can be fine-tuned on a new dataset of tagged tweets or reviews. The model can be trained on the new dataset using supervised learning with the pre-learned weights as a starting point. The model can classify new tweets and reviews by sentiment. Use a pre-trained model for feature extraction to apply transfer learning to sentiment analysis. By developing a model to classify text sentiment, you may use its characteristics to train a new model for the same purpose.

4.1.4.3. BERT based. BERT is Google's pre-trained neural network model. BERT may be fine-tuned for natural language interpretation and language translation using a huge corpus of text data. Bidirectional means it analyses word relationships in both directions to understand a sentence's meaning and context. BERT is one of the most common natural language processing models and performs well in many tasks.

4.1.4.4. Data augmentation. Various data augmentation techniques have been suggested for the purpose of Arabic text sentiment analysis. In Ref. [68] the authors proposed a system that utilizes Arabic morphology, synonymy lists, and grammatical rules to augment the dataset, leading to a 42 % improvement in accuracy. Beseiso [69] devised a subword attentive model for Arabic sentiment analysis, with a 95 % accuracy rate by the utilization of character-level convolutional neural networks and recurrent neural networks. In Sweidan's [70] study, autoregressive feature extraction was integrated with topic modeling to get exceptional results in aspect-based sentiment analysis. In Ref. [71] they presented a collection of labeled Arabic tweets and showcased the efficacy of data augmentation in enhancing the precision of deep learning models. The LSTM model outperformed the CNN and RCNN models.

4.1.4.5. Sentiment-aware machine translation. Several studies have studied sentiment analysis in Arabic, specifically targeting social media content. Bouchlaghem et al. [72] and Barhoumi et al. [73] conducted studies that demonstrated the effectiveness of machine translation in capturing sentiment in Arabic texts. Bouchlaghem's work specifically emphasized the significance of lexicon-based elements. in Ref. [74] the authors enhanced this field by creating extensive sentiment lexicons specifically for Arabic, resulting in enhanced precision of sentiment analysis algorithms. These results indicate that sentiment-aware machine translation has the potential to be a valuable tool for summarizing Arabic sentiment.

4.1.4.6. Sentiment summarization. The study in Ref. [75] presented a system for summarizing Arabic reviews based on their features, using natural language processing techniques and sentiment analysis. The algorithm produces succinct summaries of reviews by extracting data related to the domain and classifying sentiment. The inspection of the hotel domain data demonstrates an accuracy of 71.22 % for opinion mining, 73.23 % for positive summaries, and 72.46 % for negative summaries based on subjective assessments. These results indicate that the system is effective in extracting valuable information from Arabic reviews.

4.2. Evaluation measures

There are different evaluation metrics such Accuracy, recall, precision, and f1-score that can be used to evaluate various machine learning classifier for AS classification. The following is a brief overview of the various evaluation metrics.

1. **Accuracy:** This is the proportion of correctly classified instances out of the total number of instances. It's a simple and intuitive measure, but it may not be appropriate for imbalanced datasets.

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP})} \tag{1}$$

2. **Precision:** Precision measures the proportion of true positive predictions among all positive predictions. It focuses on the accuracy of positive predictions.

$$\mathrm{Precision}\ (\mathrm{P}) = \frac{TP}{(TP + FP)} \tag{2}$$

3. **Recall**: Recall measures the proportion of true positive predictions among all actual positive instances. It focuses on capturing all positive instances.

$$\mathrm{Recall}\ (\mathrm{R}) = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{3}$$

4. **F-measure:** The F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall.

$$\mathrm{F1} - \mathrm{measure} = \frac{(\mathrm{R})^*(\mathrm{P})}{(\mathrm{R}) + (\mathrm{P})}{}^*2 \tag{4}$$

5. Applications

The ASA involves understanding the attitudes, ideas, and emotions expressed in a text by analyzing the emotional tone of words. This technique is often used to gauge people's opinions about a particular topic, product, or service and can help identify patterns and trends in large amounts of data.

There are several potential applications for ASA, including.

1. Customer feedback: SA can be used to understand what consumers think of a product or service, and to identify areas for improvement.
2. Marketing: SA can be used to understand how people feel about a brand or campaign, and to inform marketing strategy.
3. Social media monitoring: ASA can be used to track the sentiment of social media conversations around a particular topic, to understand how people feel about it.
4. Politics: ASA can be used to track the sentiment of media coverage or social media conversations around a political candidate or issue, to understand how people feel about it.
5. Public health: ASA can be used to track public sentiment towards health issues, such as a vaccine or a particular disease. This can help public health agencies understand Arabic public opinion and take necessary actions to address public health concerns.

Marketing and Brand Management utilize sentiment analysis to gauge customer perceptions, drawing insights from social media and reviews. For instance, comparative analyses of Arabic and Turkish responses to Covid-19 and brand monitoring studies on Telsim and Turkcell demonstrate the strategic value of sentiment analysis tools like Brand24. These analyses underscore sentiment analysis' pivotal role in informing marketing strategies and brand positioning. Additionally, case studies showcase how AI-powered sentiment analysis aids in public relations efforts, contributing to effective brand reputation management and communication strategies [72].

Customer Service and Feedback Analysis utilizes ASA to monitor customer feedback on social media, enabling companies to track sentiment trends and promptly address concerns, thus enhancing satisfaction and loyalty. For instance, recent research highlights the need for nuanced understanding of customer sentiments across coffee products. However, analyzing Arabic text poses challenges due to its complex morphology. To tackle this, a new method has been developed to enhance the precision of Arabic sentiment analysis, particularly focusing on customer opinions about coffee products on platforms like Twitter [76].

Sentiment analysis offers significant potential for governmental and authoritative institutions, particularly in the domain of monitoring public sentiment as it manifests on online platforms. This capability proves particularly valuable for enhancing risk management and response strategies, especially during times of emergencies or major events. Consider the example of the COVID-19 pandemic, a global phenomenon that has deeply affected people's lives for more than two years. It has triggered extensive discussions on social media platforms, marked by a swift surge in posts, comments, and tweets related to the coronavirus. These online interactions act as a mirror, reflecting the thoughts and viewpoints of individuals concerning the pandemic [77,78]. A range of studies have explored sentiment analysis in Arabic e-commerce product reviews. Sghaier in Ref. [79] developed a tool for polarity detection in Arabic e-commerce reviews, despite challenges such as data scarcity and dialect complexity. Hakami [80] used sentiment analysis and topic modeling to identify factors influencing customer satisfaction in popular Saudi Arabian online shopping apps. Abbassy [77] proposed a rule-based emotion analysis system for Arabic customer reviews, while Harrage [78] focused on predicting product ratings based on sentiment in Arabic reviews. These studies collectively highlight the potential of sentiment analysis in understanding and improving customer experiences in Arabic e-commerce. Several studies have employed a diverse range of techniques to implement sentiment analysis, thereby uncovering public sentiments related to COVID-19 [79,80]. One significant application of sentiment analysis (SA) is its utility in assisting political parties or governments in gauging their likelihood of success in upcoming elections and assessing the level of popular satisfaction with their policies. It can also assist in forecasting the efficacy of a recently implemented social initiative, such as the 'Odd-even car regulation' in Delhi or the 'Swachh Bharat Abhiyaan'. The collection of public opinions can provide insight into the level of satisfaction with a new change, so enabling an assessment of the potential success of the new policy in the future. By adopting this approach, individuals can effectively conserve financial resources, time, and exertion, or implement essential measures to enhance the efficacy of their endeavors. This is particularly relevant as individuals predominantly articulate their viewpoints on social media platforms, and the volume of opinions escalates during periods of policy or regulatory modifications implemented by governmental bodies or organizations. It is possible to anticipate the financial performance of a film by utilizing the perspectives and evaluations of individuals. Furthermore, it has been observed that technology can play a significant role in enhancing the quality of teaching and facilitating student learning within the realm of education. Another significant use of sentiment analysis is to its utility for product companies in enhancing their revenues and client retention through the analysis of individuals' views regarding their items or services. Moreover, it is highly advantageous in forecasting client trends and facilitating the formulation of more captivating and influential marketing plans. Sentiment analysis (SA) is a commonly employed technique in the field of stock market forecasting. There exists a strong correlation between fluctuations in stock prices of a firm and the views conveyed regarding said company on social media platforms. Therefore, by considering the viewpoints shared on social media, one can make an informed assessment regarding the future profitability or loss of a firm, as well as the potential viability of investing in its stocks [1]. The range of uses of SA is extensive and limitless. Numerous topics have been investigated through the application of sentiment analysis, with ample potential for further exploration in this field.

Generally, ASA can be a valuable tool for businesses, organizations, and individuals looking to understand and track the emotions expressed in Arabic text data.

6. Challenges and limitations of ASA

Here, we discuss some of the difficulties that scientists today face. Sentiment analysis in Arabic is tough because of various obstacles. Reasonable difficulty in understanding the language: Arabic is a language with a complex grammar and a large vocabulary due to its extensive system of inflection. Especially when the literature in question makes use of figurative language or idioms, this can impede comprehension. Figurative language: Arabic speakers regularly utilize similes, metaphors, and other figurative devices to communicate meaning and emotion. When this happens, it can be challenging for a computer to grasp the meaning of the text. The sensitivity of emotions to setting: Rather than relying on overtly positive or negative language, context is often all that's needed to indicate the tone of a piece of writing. For example, a statement that appears neutral on its own may be interpreted as positive or negative depending on the context in which it is used. The lack of annotated data: ML based approaches to ASA require large amounts of annotated data for training. However, annotated data for sentiment analysis in Arabic is often difficult to find, which can make it challenging to develop accurate machine learning models. The influence of dialects and regional variations: Arabic is spoken in a wide variety of countries and regions, and there are important differences in the way the language is spoken and written in different areas. This can make it hard to progress a sentiment analysis structure that is effective across all regions.

The discipline of Arabic sentiment analysis is undergoing significant development, with an increasing emphasis on analyzing sentiment at the document level [81]. Nevertheless, further investigation is required in this domain, specifically regarding the utilization of rule-based methodologies and deep learning [82]. A comprehensive examination of the current body of literature on Arabic sentiment analysis has been conducted, which has identified the necessity for enhancements in preprocessing techniques, feature generation approaches, and methodologies for sentiment classification [83]. Moreover, there is a demand for the creation of semantic aspect-based sentiment analysis specifically for Arabic reviews, which has the potential to offer a more intricate comprehension of sentiment [84]. These observations indicate a potential future path for the discipline, emphasizing the enhancement of the precision and detail of sentiment analysis in Arabic literature.

6.1. *Use of arabizi*

Arabizi is a term used to refer to the informal use of the Latin alphabet to write Arabic, often using numbers and other non-alphabetic characters to represent certain sounds.

Arabic written using Latin characters is known as Arabizi, Arabish, or Romanized Arabic. MSA and Arabic dialects are frequently written in social media platforms. Studies that only focus on finding and translating Arabizi into Arabic have dealt with this particular literary style. Since texts are pre-processed to remove all Latin letters, Sentiment Analysis has not addressed this issue in the published publications. To the best of our knowledge, only [49] has dealt with this task in a published book.

6.2. *Arabic is intricate due to its distinctive characteristics*

Arabic is complicated due to its ambiguity and extensive morphological structure [85]. These traits matter most.

6.2.1. *Right to left*

Given the right-to-left nature of the Arabic script, it might be challenging to decipher emoticons within Arabic text. As a result, it's not uncommon for the tweet's emoticons to be mismatched, leading to a confusing portrayal of the author's emotions.

6.2.2. *Arabic words can take on several forms at the root level depending on the situation*

For instance, (كتب, كاتب, كتاب) in English means (He wrote, writer, book).

6.2.3. *The Arabic language has the unique trait of having words with the same spelling but different meanings*

As shown by the punctuation, for example (رجل) this means man and ((رجل this means leg, foot.

6.2.4. *When a sentence contains words like (لكن)*

Which means (But), it might simultaneously convey two opposing emotions.

The intricate word forms and morphology of Arabic provide challenges for SA. The presence of morphological modifications, such as inflectional forms (gender, number, case) and derivational forms (patterns, roots), poses a significant constraint. These modifications could potentially reduce the density of the feature space, hence posing challenges for SA models to identify patterns and make generalizations.

Utilizing advanced morphological analysis techniques such as stemming, lemmatization, and root extraction can potentially mitigate the impact of morphological variations. Standardizing dialectal text can enhance the performance of the model.

6.3. *Dialectal Arabic*

Every region or sub-national entity in a country has its own distinct variety or dialect of Arabic. This indicates that there are numerous variations of Arabic text that may be found online, each of which may convey a distinct meaning. As a consequence, conducting SA using several dialects introduces a high level of complexity.

The issue of Arabic dialectal variation is another concern. Egyptian, Levantine, and Gulf Arabic exhibit significant variations in vocabulary, grammar, and syntax compared to Modern Standard Arabic (MSA). Text sentiment can become unclear due to dialectal differences, particularly when the training data does not correctly represent the intended dialect. MSA-trained sentiment analysis models may have difficulties when processing dialectal Arabic text, hence limiting their applicability.

Applying transfer learning to refine pre-trained language models with Arabic sentiment data can effectively address the issue of limited data availability. Domain adaptation techniques can be employed to ensure effective performance of models across different domains or dialects.

6.4. *Lack of lexicon*

Insufficient as compared to English lexicons is the MSA lexicon. For spoken Arabic, there is no publicly available lexicon. A dictionary was created, however not every dialect was covered.

6.5. *Lack of corpora and datasets*

The most recent work that has been done on Arabic sentiment analysis was reported in Ref. [86], which was written by its authors. They discussed the most prevalent obstacles that can be found in each step of the analysis process before presenting the overarching procedure of Arabic sentiment analysis. For instance, they made some comments regarding the lack of corpora for the stage that was related to the gathering of data. In addition to that, they talked on the absence of Arabic sentiment lexicons.

6.6. *Named entity recognition*

Arabic names that are generated from Arabic adjectives may cause the classifier to become confused, hence it is crucial to identify Arabic names in sentences. For example, the word "فرح" corresponds to the adjective "فرح" which means "happiness". There is no means to identify Arabic names, in contrast to the English language, where it is common practice to start a name with a capital letter, the names of people in other languages are often written with lowercase letters.

6.7. *Sarcasm*

According to the findings of study [87], people will sometimes utilize sarcasm in their tweets combined with the opposing emoticon. In addition, choosing the incorrect emoticons to accompany a sentence might completely alter its significance.

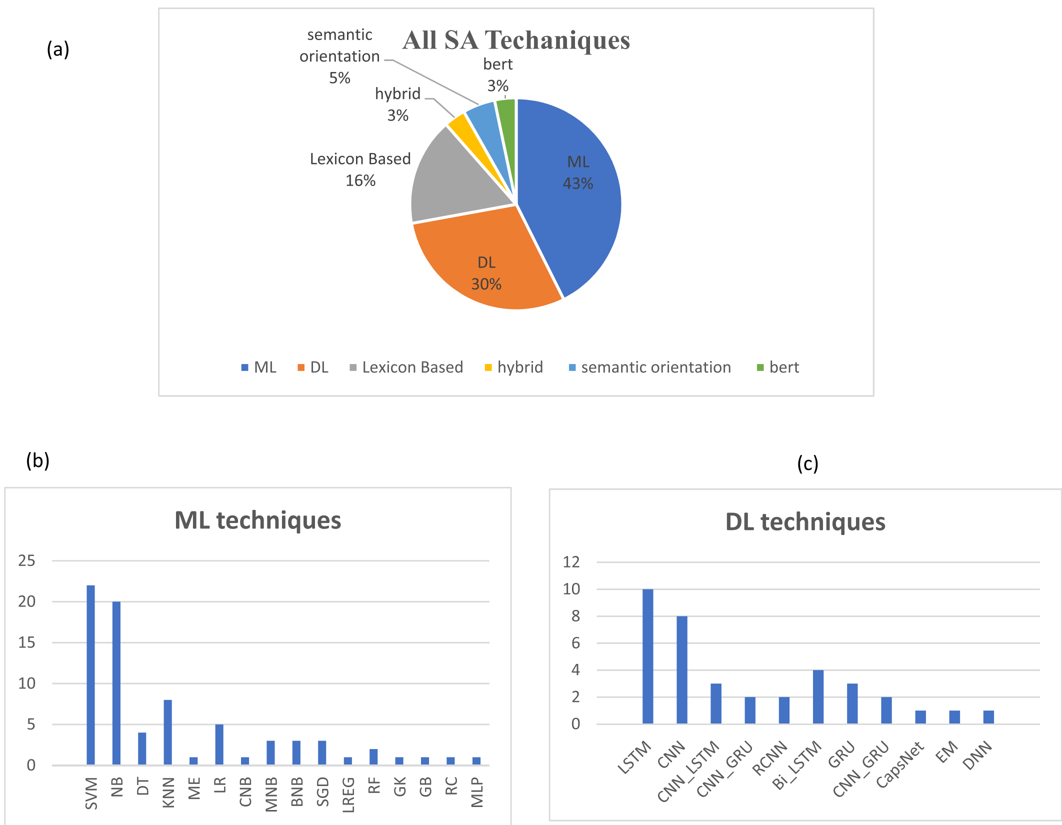

Fig. 8. Percentage of research work using **a** SA technique, **b** ML techniques, and **c** DL techniques.

6.8. Handling negation

Negation terms reverse the meaning of past or present tense verbs. e.g. "لم أعجب بهذا الكتاب". In English means "I disliked this book."

6.9. Phrases and idioms

Arabic speakers often communicate their thoughts through the usage of popular compound idioms and slogans. For instance, the expression (!الي شايخ) is used to indicate incredulity in what others are saying. Some words and phrases are employed for the purpose of expressing the sentiment, with usage playing a factor. There are a rising number of new phrases being coined every day [68].

6.10. Interdependence between domain and context

In various fields and situations, people may use the same words and phrases to signify something completely different. This technology is crazy … the patient is going crazy.

7. Findings of systematic survey

This section offers an overview of the findings obtained during the execution of this systematic survey, with a primary emphasis on addressing the research questions outlined in Table 2.

Arabic Sentiment Analysis is valuable in a wide range of applications, from business and marketing to social and political analysis, content moderation, education, and cultural understanding. It allows organizations and researchers to tap into the rich emotional and cultural landscape of the Arabic language, providing insights and opportunities for various sectors, as answered in RQ1. Most standard datasets for ASA were collected from Twitter, as summarized in Table 3, which answer to RQ2.

Fig. 8 offers an overview of the prevalent techniques employed in Sentiment Analysis (SA) within the context of the Arabic language, with a specific focus on addressing Research Question (RQ3). This visual representation encapsulates the landscape of SA research within the Arabic language domain, providing insights into how these techniques are distributed across various publications. The results indicate a preference for Machine Learning (ML) methods among the majority of researchers, followed by Deep Learning, lexicon-based approaches, hybrid methodologies, and other alternatives, all of which are detailed in Fig. 8a. Machine Learning has proven to be promising in reducing the manual effort traditionally associated with labor-intensive tasks, such as creating sentiment lexicons. Approximately 43 % of research studies have adopted ML techniques for conducting SA. Concurrently, there is a growing interest among researchers in exploring Deep Learning methods, primarily due to their improved accuracy, despite the potential for longer training times. About 30 % of research studies have embraced Deep Learning for SA, while 16 % have relied on lexicon-based strategies and other approaches, as illustrated in Fig. 8a.

In the realm of ML techniques, Support Vector Machines (SVM) have emerged as the preferred choice, representing roughly 29 % of research studies dedicated to developing SA systems, as highlighted in Fig. 8b. As for Deep Learning methods, Long Short-Term Memory (LSTM) networks have gained prominence, accounting for approximately 27 % of the research studies, as demonstrated in Fig. 8c. Section 8 lists the current challenges and limitations of ASA in prior studies, which answers RQ4. To answer RQ5 Indeed, established text representation methods excel at accommodating the Arabic language. These methodologies encompass a range of techniques for encoding and presenting textual data in a format comprehensible to machine learning algorithms. Several prevalent text representation techniques used for both Arabic and other languages as showing in Table 4.

8. Conclusions and recommendations

This comprehensive survey unveils the intricate landscape of Arabic sentiment analysis (ASA) research by systematically addressing the outlined inquiries. ASA demonstrates immense potential for deriving valuable insights across diverse domains, enabling organizations to harness Arabic's intricate linguistic and cultural nuances (RQ1). Twitter emerges as the predominant source for benchmark ASA datasets (RQ2, Table 3). The analysis unveils researchers' proclivity towards machine learning methodologies (43 %), followed by deep learning (30 %), lexicon-based (16 %), and hybrid techniques (RQ3, Fig. 8a). Support vector machines and long short-term memory networks emerge as the preeminent ML and deep learning approaches, respectively (Fig. 8b and c). Concurrently, challenges encompass grappling with dialectal variations, data paucity, linguistic complexities, and lack of robust evaluation frameworks (RQ4, Section 8). Established text representation paradigms have demonstrated adeptness in accommodating Arabic's idiosyncrasies, facilitating robust ASA modeling (RQ5, Table 4). This comprehensive exploration delineates the current state, preferred methodologies, and outstanding desiderata, fostering an ecosystem conducive to catalyzing future research endeavors towards unleashing Arabic NLP's latent potential.

As part of future work, advancing ASA necessitates exploring multiple promising research directions. Leveraging deep learning architectures like hybrid transformers is crucial for capturing Arabic's complex linguistics and dialectal variations. Multimodal approaches integrating textual and multimedia data are vital for Arabic social media sentiment analysis. Aspect-based techniques tailored to domains like e-commerce review analysis can enable identifying sentiment towards specific product attributes. Given Arabic's data scarcity, cross-lingual transfer learning from high-resource languages presents an opportunity to enhance models. Incorporating contextual cues such as user profiles, discourse context, and commonsense reasoning can facilitate more nuanced sentiment understanding. Developing interpretable and explainable models providing transparency into prediction rationales is also a

Table 4
Text representation techniques.

Text representation in ASA	Explain
Bag of Words (BoW)	Is a straightforward approach that portrays text as a collection of words, disregarding their order and structure. It assembles a vocabulary of unique words from the text and represents each document as a vector based on word frequencies.
Term Frequency-Inverse Document Frequency (TF-IDF)	TF-IDF is another method for text representation. It evaluates term frequencies within a document and adjusts them based on the prevalence of the term across the entire document corpus. This method helps discern the importance of terms within a document relative to the whole collection.
Word Embeddings	Such as Word2Vec, GloVe, and FastText, capture semantic relationships among words. They represent words as vectors in a continuous space where words with similar meanings are situated closely in the vector space.
Character Embeddings	Character-level embeddings represent words based on the characters they contain. This is especially advantageous for languages like Arabic, characterized by words sharing root letters and intricate morphological structures.
Subword Embeddings	Techniques like Byte Pair Encoding (BPE) and Sentence Piece capture subword information, proving valuable for languages with complex morphology, such as Arabic.
Contextual Word Embeddings	Models like ELMo, GPT (Generative Pre-trained Transformer), and BERT (Bidirectional Encoder Representations from Transformers) provide contextual embeddings by considering the words surrounding a given term in a sentence. This results in more precise representations.
Sentence Embeddings	These methods aim to encapsulate the overall meaning of a sentence or document, as exemplified by Doc2Vec and the Universal Sentence Encoder. They are useful for tasks like Arabic text summarization and classification.
Pre-trained Language Models	Models like BERT, including its Arabic variants, are pre-trained on extensive Arabic text corpora and can be fine-tuned for specific Natural Language Processing (NLP) tasks.
N-grams	Represent sequences of consecutive words or characters, capturing local context and relationships between terms.
Topic Models	Techniques like Latent Dirichlet Allocation (LDA) offer a means to represent documents by their distributions across various topics.

key direction. Exploring such computational methods and Arabic-specific solutions can drive performance gains and real-world deploy ability for ASA.

It is recommended to develop an open-source Arabic sentiment lexicon package would be beneficial that covering multi-class sentiment analysis, consisting of positive, negative, neutral, and mixed sentiments. Furthermore, developing code-switching models to map disparate Arabic dialects to Modern Standard Arabic, addressing negation, and applying state-of-the-art text representations with deep learning techniques should be pursued to enhance ASA performance, effectively addressing RQ6.

CRediT authorship contribution statement

Amani A. Aladeemy: Methodology, Investigation, Data curation, Conceptualization. **Ali Alzahrani:** Data curation, Conceptualization. **Mohammad H. Algarni:** Supervision, Software, Conceptualization. **Saleh Nagi Alsubari:** Resources, Project administration, Conceptualization. **Theyazn H.H. Aldhyani:** Methodology, Investigation, Data curation. **Sachin N. Deshmukh:** Supervision, Software, Conceptualization. **Osamah Ibrahim Khalaf:** Visualization, Validation, Software, Conceptualization. **Wing-Keung Wong:** Validation, Supervision, Software, Resources. **Sameer Aqburi:** Resources, Methodology, Investigation.

Data availability

There is no data the research work is survey study.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Declaration of competing interest

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

Acknowledgement

This work was supported through the Annual Funding track by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Research No. **GrantA No.364**).

References

[1] B. Pang, L. Lee, Opinion mining and sentiment analysis, Foundations and Trends® in information retrieval 2 (1–2) (2008) 1–135.
[2] N.A. Abdulla, N.A. Ahmed, M.A. Shehab, M. Al-Ayyoub, Arabic sentiment analysis: lexicon-based and corpus-based, in: 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), IEEE, 2013, December, pp. 1–6.

[3] N. Farra, E. Challita, R. Abou Assi, H. Hajj, Sentence-level and document-level sentiment mining for Arabic texts, in: 2010 IEEE International Conference on Data Mining Workshops, IEEE, 2010, December, pp. 1114–1119.
[4] G. Alwakid, T. Osman, T. Hughes-Roberts, Challenges in sentiment analysis for Arabic social networks, Procedia Computer Science 117 (2017) 89–100. .
[5] N. Jeong, J. Lee, An aspect-based review analysis using ChatGPT for the exploration of hotel service failures, Sustainability 16 (4) (2024) 1640. .
[6] N. Boudad, R. Faizi, R.O.H. Thami, R. Chiheb, Sentiment analysis in Arabic: a review of the literature, Ain Shams Eng. J. 9 (4) (2018) 2479–2490. .
[7] I. Guellil, A. Adeel, F. Azouaou, A. Hussain, Sentialg: automated corpus annotation for algerian sentiment analysis, in: International Conference on Brain Inspired Cognitive Systems, Springer, Cham.2018 ,, July, pp. 557–567.
[8] A.B. Sharaf, E. Atwell, QurAna: corpus of the Quran annotated with pronominal anaphora, in: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), 2012 ,, May, pp. 130–137.
[9] M. Abdul-Mageed, M. Diab, S. Kübler, SAMAR: subjectivity and sentiment analysis for Arabic social media, Comput. Speech Lang 28 (1) (2014) 20–37. .
[10] M. Maamouri, A. Bies, T. Buckwalter, W. Mekki, The penn Arabic treebank: building a large-scale annotated Arabic corpus, in: NEMLAR Conference on Arabic Language Resources and Tools, 27, 2004 ,, September, pp. 466–467.
[11] M. Abdul-Mageed, M.T. Diab, AWATIF: a multi-genre corpus for modern standard Arabic subjectivity and sentiment analysis, in: LREC, 515, 2012 ,, May, pp. 3907–3914.
[12] M. Rushdi-Saleh, M.T. Martín-Valdivia, L.A. Ureña-López, J.M. Perea-Ortega, OCA: opinion corpus for Arabic, J. Am. Soc. Inf. Sci. Technol. 62 (10) (2011) 2045–2054. .
[13] R.M. Duwairi, R. Marji, N. Sha'ban, S. Rushaidat, Sentiment analysis in Arabic tweets, in: 2014 5th International Conference on Information and Communication Systems (ICICS), IEEE, 2014, April, pp. 1–6.
[14] A. Al-Sallab, R. Baly, H. Hajj, K.B. Shaban, W. El-Hajj, G. Badaro, Aroma: a recursive deep learning model for opinion mining in Arabic as a low resource language, ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 16 (4) (2017) 1–20. .
[15] A.A. Altowayan, L. Tao, Word embeddings for Arabic sentiment analysis, in: 2016 IEEE International Conference on Big Data (big data), IEEE, 2016, December, pp. 3820–3825.
[16] W. Antoun, F. Baly, H. Hajj, Arabert: Transformer-Based Model for Arabic Language Understanding. arXiv Preprint arXiv:2003.00104, 2020 ,.
[17] I.A. Farha, W. Magdy, Mazajak: an online Arabic sentiment analyser, in: Proceedings of the Fourth Arabic Natural Language Processing Workshop, 2019 ,, August, pp. 192–198.
[18] S.M. Mohammad, M. Salameh, S. Kiritchenko, How translation alters sentiment, J. Artif. Intell. Res. 55 (2016) 95–130. .
[19] T.T. Thet, J.C. Na, C.S. Khoo, Aspect-based sentiment analysis of movie reviews on discussion boards, J. Inf. Sci. 36 (6) (2010) 823–848.
[20] S. Mihi, B. Ait, I. El, S. Arezki, N. Laachfoubi, MSTD: moroccan sentiment Twitter dataset, Int. J. Adv. Comput. Sci. Appl. 11 (10) (2020) 363–372, https://doi.org/10.14569/IJACSA.2020.0111045.
[21] Nabil M, Aly M, Atiya A. ASTD: Arabic sentiment tweets dataset. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. p. 2515–2519.
[22] N. Abdulla, N. Mahyoub, M. Shehab, M. Al-Ayyoub, Arabic sentiment analysis: corpus-based and lexicon-based, in: IEEE Conference on Applied Electrical Engineering and Computing Technologies (AEECT 2013), 2013.
[23] R. Zbib, et al., Machine translation of Arabic dialects, in: Presented at the Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2012, pp. 49–59.
[24] S.M. Mohammad, M. Salameh, S. Kiritchenko, How translation alters sentiment, J. Artif. Intell. Res. 55 (2016) 95–130.
[25] Mourad, K. Darwish, Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs, in: Presented at the Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2013, pp. 55–64.
[26] E. Refaee, V. Rieser, An Arabic twitter corpus for subjectivity and sentiment analysis, in: Presented at the LREC, 2014, pp. 2268–2273.
[27] A.A. Elmadany, W.M. Hamdy Mubarak, ArSAS: an Arabic speech-act and sentiment corpus of tweets, in: OSACT 3: the 3rd Workshop on Open-Source Arabic Corpora and Processing Tools, 2018, p. 20.
[28] A.M. Alayba, V. Palade, M. England, R. Iqbal, Arabic Language Sentiment Analysis on Health Services, 2017, pp. 114–118, https://doi.org/10.1109/asar.2017.8067771.
[29] T. Khalil, A. Halaby, M.H. Hammad, S.R. El-Beltagy, Which configuration works best? An experimental study on Supervised Arabic Twitter Sentiment Analysis, in: Proceedings of the First Conference on Arabic Computational Liguistics (ACLing 2015), Co-located with CICLing 2015, 2015, pp. 86–93. Cairo, Egypt.
[30] A. Shoukry, A. Rafea, A hybrid approach for sentiment classification of Egyptian dialect tweets, in: First International Conference on Arabic Computational Linguistics (ACLing), 2015, pp. 78–85. Cairo, Egypt.
[31] M. Aly, A. Atiya, Labr: a large scale Arabic book reviews dataset, in: Proceedings of the 51st Annual Meeting of theAssociation for Computational Linguistics 2, 2013, pp. 494–498. Short Papers.
[32] V. Stoyanov, C. Cardie, J. Wiebe, Multi-perspective question answering using the OpQA corpus, in: Proceedings of Theconference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005, pp. 923–930.
[33] H. ElSahar, S.R. El-Beltagy, Building large Arabic multi-domain resources for sentiment analysis, in: InternationalConference on Intelligent Text Processing and Computational Linguistics, 2015, pp. 23–34.
[34] K.M. Alomari, H.M. ElSherif, K. Shaalan, Arabic tweets sentimental analysis using machine learning, in: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Springer, Cham, 2017, June, pp. 602–610.
[35] Arabic sentiment twitter corpus, Available online: https://www.kaggle.com/mksaad/arabic-sentiment-twitter-corpus. (Accessed 2 April 2022).
[36] I. Guellil, F. Azouaou, M. Mendoza, Arabic sentiment analysis: studies, resources, and tools, Social Network Analysis and Mining 9 (2019) 1–17. .
[37] T. Al-Moslmi, M. Albared, A. Al-Shabi, N. Omar, S. Abdullah, Arabic senti-lexicon: constructing publicly available language resources for Arabic sentiment analysis, J. Inf. Sci. 44 (3) (2018) 345–362. .
[38] A. Alowisheq, S. Alhumoud, N. Altwairesh, T. Albuhairi, Arabic sentiment analysis resources: a survey, in: Social Computing and Social Media: 8th International Conference, SCSM 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17–22, 2016. Proceedings, 8, Springer International Publishing, , 2016, pp. 267–278.
[39] H. ElSahar, S.R. El-Beltagy, Building large Arabic multi-domain resources for sentiment analysis, in: International Conference on Intelligent Text Processing and Computational Linguistics, Springer International Publishing, Cham, 2015, April, pp. 23–34.
[40] I. Touahri, A. Mazroui, Deep analysis of an Arabic sentiment classification system based on lexical resource expansion and custom approaches building, Int. J. Speech Technol. 24 (2021) 109–126. .
[41] Y. Zahidi, Y. El Younoussi, Y. Al-Amrani, Different valuable tools for Arabic sentiment analysis: a comparative evaluation, Int. J. Electr. Comput. Eng. 11 (1) (2021). .
[42] M. Wankhade, A.C.S. Rao, C. Kulkarni, A survey on sentiment analysis methods, applications, and challenges, Artif. Intell. Rev. 55 (7) (2022) 5731–5780. .
[43] Z. Yan-Yan, Q. Bing, L. Ting, Integrating intra-and inter-document evidences for improving sentence sentiment classification, Acta Autom. Sin. 36 (10) (2010) 1417–1425. .
[44] A. Moreo, M. Romero, J. Castro, J.M. Zurita, Lexicon-based comments-oriented news sentiment analyzer system, Expert Syst. Appl. 39 (10) (2012) 9166–9180.
[45] L. Al-Horaibi, M.B. Khan, Sentiment analysis of Arabic tweets using text mining techniques, in: First International Workshop on Pattern Recognition, 10011, SPIE, 2016, July, pp. 288–292.
[46] N.A. Abdulla, M. Al-Ayyoub, M.N. Al-Kabi, An extended analytical study of Arabic sentiments, International Journal of Big Data Intelligence 1 1 (1–2) (2014) 103–113. .
[47] K.M. Alomari, H.M. ElSherif, K. Shaalan, Arabic tweets sentimental analysis using machine learning, in: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Springer, Cham, 2017, June, pp. 602–610.

[48] A. Shoukry, A. Rafea, Sentence-level Arabic sentiment analysis, in: 2012 International Conference on Collaboration Technologies and Systems (CTS), IEEE, 2012, May, pp. 546–550.
[49] R.M. Duwairi, M. Alfaqeh, M. Wardat, A. Alrabadi, Sentiment analysis for Arabizi text, in: 2016 7th International Conference on Information and Communication Systems (ICICS), IEEE, 2016, April, pp. 127–132.
[50] S.M. Ali, Z. Noorian, E. Bagheri, C. Ding, F. Al-Obeidat, Topic and sentiment aware microblog summarization for twitter, J. Intell. Inf. Syst. 54 (1) (2020) 129–156.
[51] T.H. Soliman, M.A. Elmasry, A. Hedar, M.M. Doss, Sentiment analysis of Arabic slang comments on facebook, International Journal of Computers & Technology 12 (5) (2014) 3470–3478. .
[52] M. Nabil, M. Aly, A. Atiya, Labr: a large scale Arabic sentiment analysis benchmark, arXiv preprint arXiv:1411.6718 (2014).
[53] H. Hamdan, P. Bellot, F. Bechet, Lsislif: crf and logistic regression for opinion target extraction and sentiment polarity analysis, in: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 2015, pp. 753–758.
[54] M.M. Altawaier, S. Tiun, Comparison of machine learning approaches on Arabic twitter sentiment analysis, Int. J. Adv. Sci. Eng. Inf. Technol. 6 (6) (2016) 1067–1073. .
[55] A.S.A. Al-Jumaili, H.K. Tayyeh, A hybrid method of linguistic and statistical features for Arabic sentiment analysis, Baghdad Science Journal 17 (1) (2020) 385–390.
[56] A.K. Al-Tamimi, A. Shatnawi, E. Bani-Issa, Arabic sentiment analysis ofYouTube comments, in: 2017IEEE Jordan Conf. On Applied Electrical Engineering and Computing Technologies, AEECT 2017, Aqaba,Jordan, IEEE, 2017, pp. 1–6.
[57] M. Hadwan, M. Al-Hagery, M. Al-Sarem, F. Saeed, Arabic sentiment analysis of users' opinions of governmental mobile applications, Comput. Mater. Continua (CMC) 72 (3) (2022) 4675–4689. .
[58] F.H. Mahyoub, M.A. Siddiqui, M.Y. Dahab, Building an Arabic sentiment lexicon using semi-supervised learning, Journal of King Saud University-Computer and Information Sciences 26 (4) (2014) 417–424. .
[59] E. Refaee, V. Rieser, An Arabic twitter corpus for subjectivity and sentiment analysis, in: LREC, 2014 ,, May, pp. 2268–2273.
[60] K. Elshakankery, M.F. Ahmed, HILATSA: a hybrid Incremental learning approach for Arabic tweets sentiment analysis, Egyptian Informatics Journal 20 (3) (2019) 163–171. .
[61] H.K. Aldayel, A.M. Azmi, Arabic tweets sentiment analysis–a hybrid scheme, J. Inf. Sci. 42 (6) (2016) 782–797. .
[62] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, G. Eryiğit, Semeval-2016 task 5: aspect based sentiment analysis, in: ProWorkshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, 2016 ,, pp. 19–30.
[63] T.T. Thet, J.C. Na, C.S. Khoo, Aspect-based sentiment analysis of movie reviews on discussion boards, J. Inf. Sci. 36 (6) (2010) 823–848. .
[64] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, August, pp. 168–177.
[65] G. Ganu, N. Elhadad, A. Marian, Beyond the stars: improving rating predictions using review text content, in: WebDB, 9, 2009, June, pp. 1–6.
[66] S. Brody, N. Elhadad, An unsupervised aspect-sentiment model for online reviews, in: Human Language Technologies: the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010 ,, June, pp. 804–812.
[67] M. Al-Smadi, O. Qawasmeh, B. Talafha, M. Quwaider, Human annotated Arabic dataset of book reviews for aspect based sentiment analysis, in: 2015 3rd International Conference on Future Internet of Things and Cloud, IEEE, 2015, August, pp. 726–730.
[68] R. Duwairi, F. Abushaqra, Syntactic-and morphology-based text augmentation framework for Arabic sentiment analysis, PeerJ Computer Science 7 (2021) e469.
[69] M. Beseiso, H. Elmousalami, Subword attentive model for Arabic sentiment analysis: a deep learning approach, ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 19 (2) (2020) 1–17.
[70] A.H. Sweidan, N. El-Bendary, E. Elhariri, Autoregressive feature extraction with topic modeling for aspect-based sentiment analysis of Arabic as a low-resource language, ACM Transactions on Asian and Low-Resource Language Information Processing 23 (2) (2024) 1–18.
[71] A. Mohammed, R. Kora, Deep learning approaches for Arabic sentiment analysis, Social Network Analysis and Mining 9 (1) (2019) 52.
[72] R. Bouchlaghem, A. Elkhelifi, R. Faiz, A machine learning approach for classifying sentiments in Arabic tweets, in: Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics, 2016, June, pp. 1–6.
[73] A. Barhoumi, C. Aloulou, N. Camelin, Y. Esteve, L. Belguith, Arabic Sentiment analysis: an empirical study of machine translation's impact, in: Language Processing and Knowledge Management International Conference (LPKM2018), 2018, October.
[74] S. Mohammad, M. Salameh, S. Kiritchenko, Sentiment lexicons for Arabic social media, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016, May, pp. 33–37.
[75] A.M. El-Halees, D. Salah, Feature-based opinion summarization for Arabic reviews, in: 2018 International Arab Conference on Information Technology (ACIT), IEEE, 2018, November, pp. 1–5.
[76] O. Alsemaree, A.S. Alam, S. Gill, S. Uhlig, Sentiment analysis of Arabic social media texts: a machine learning approach to deciphering customer perceptions, Heliyon (2024) e27863.
[77] M.M. Abbassy, A. Abo-Alnadr, Rule-based emotion AI in Arabic customer review, Int. J. Adv. Comput. Sci. Appl. 10 (9) (2019).
[78] F. Harrage, A. Alsalman, A. Alqahtani, Rating predictor: sentiment analysis of product reviews in Arabic, in: 2018 International Conference on Asian Language Processing (IALP), IEEE, 2018, November, pp. 44–49.
[79] S. Alhumoud, Arabic Sentiment Analysis Using Deep Learning for Covid-19 Twitter Data, International Journal of Computer Science and Network Security, , 2020, pp. 132–138.
[80] M. Alhajji, A. Al Khalifah, M. Aljubran, M. Alkhalifah, Sentiment Analysis of Tweets in Saudi Arabia Regarding Governmental Preventive Measures to Contain COVID-19, 2020 ,.
[81] I. Awajan, M. Mohamad, A review on sentiment analysis in Arabic using document level, Int. J. Eng. Technol. 7 (3.13) (2018) 128–132.
[82] H. Rahab, M. Djoudi, A. Zitouni, Sentiment Analysis of Arabic Documents: Main Challenges and Recent Advances. Natural Language Processing for Global and Local Business, 2021, pp. 307–331.
[83] A. Ghallab, A. Mohsen, Y. Ali, Arabic sentiment analysis: A systematic literature review. Applied Computational Intelligence and Soft Computing 2020 (2020) 1–21.
[84] S. Behdenna, F. Barigou, G. Belalem, Towards semantic aspect-based sentiment analysis for Arabic reviews, in: Research Anthology on Implementing Sentiment Analysis across Multiple Disciplines, IGI Global, 2022, pp. 1480–1493.
[85] S.R. El-Beltagy, A. Ali, Open issues in the sentiment analysis of Arabic social media: a case study, in: 2013 9th International Conference on Innovations in Information Technology (IIT), IEEE, 2013, March, pp. 215–220.
[86] M. El-Masri, N. Altrabsheh, H. Mansour, Successes and challenges of Arabic sentiment analysis research: a literature review. Social Network Analysis and Mining 7 (1) (2017) 1–22. .
[87] N. Altrabsheh, M. Cocea, S. Fallahkhair, Detecting sarcasm from students' feedback in Twitter, in: European Conference on Technology Enhanced Learning, Springer, Cham.2015 ,, September, pp. 551–555.