Sentiment Analysis Data Annotation Across Arabic Dialects

Dr/Samar A. Assem
Lecturer & PhD Candidate
English Program Faculty of Alsun
King Salman International University (KSIU)Egypt

Abstract

Nowadays, with a vast amount of data being generated in the internet and social media platforms, there is much focus on sentiment analysis as it helps to extract data ultimately to analyze people's opinions. However, much of the research on this topic has been regarding the English language with less attention being paid to other common languages such as Arabic. That is why there is a noticeable gap in research especially when considering the multitude of dialects in the Arabic language. This multitude of dialects is considered to be a double-edged weapon as it offers a rich amount of data for researchers from one hand and it stands as a challenge in front of the annotators and the researchers from another. Accordingly, this article aims to carry out a critical review of the recent studies conducted regarding sentiment analysis and data annotation for Arabic dialects that have been published in the last 10 years. This review offers a taxonomy of data preprocessing and annotation methods. Moreover, it displays the challenges, motivations and recommendations in addition to an in-depth analysis of the current trends in the field of Arabic dialects and sentiment analysis. Accordingly, this literature review postulates new research gaps and future directions to drive more scholars into contributing to Arabic SA research. Finally, the contribution of the current study aims at making more successful multilingual sentiment analysis implementations in addition to providing some understanding of ASA in a plenty of settings.

Keywords: Sentiment Analysis – Arabic dialects – Dialectology

Introduction

The rise of social media has revolutionized communication, providing a platform for individuals to express their opinions and sentiments on a wide range of topics. Analyzing these sentiments has become crucial for understanding public opinion and gauging reactions to events, products, and policies. However, when it comes to Arabic dialects in social media, sentiment analysis presents unique challenges, particularly in the subtasks of pre-processing and annotation.

Sentiment analysis, also known as opinion mining, is a rapidly evolving field within natural language processing (NLP). Its primary goal is to determine the emotional tone or attitude expressed in a piece of text, providing valuable insights into public opinion, consumer behavior, and brand reputation management (Abbassi, Zribi, Belguith ,2015).

Sentiment analysis plays a crucial role in various fields due to its ability to provide valuable insights into public opinion and consumer behavior. It helps organizations understand how their products, services, and brand are perceived by customers, allowing them to make informed decisions about marketing strategies, product development, and customer service. For instance, sentiment analysis of social media posts can help companies identify trends, monitor reputation, and respond to customer feedback more effectively (Shahrour, & Meziane, 2020). Additionally, sentiment analysis can be used in the political sphere to gauge voter sentiment and track public opinions on key issues.

Abdul-Mageed, Diab, & Kübler (2014) elaborated that sentiment analysis plays a crucial role in understanding and managing public opinion on social media platforms. It helps businesses identify when and how to engage with their customers directly, allowing them to respond appropriately to positive or negative feedback. By analyzing the emotional content of customer conversations, sentiment analysis can provide valuable insights into consumer behavior and preferences, enabling companies to make informed decisions about branding, marketing strategies, and product development. Additionally, sentiment analysis is essential for reputation management, crisis management, and creating resonant messaging that fosters stronger connections with customers.

While sentiment analysis has gained considerable attention in recent years, its application to languages beyond English remains a challenging task. Among these languages, Arabic stands out as a particularly complex case due to the presence of multiple dialects and the intricacies of its linguistic structure.

Arabic dialects, with their diverse regional variations, colloquialisms, and lack of standardized orthography, pose significant hurdles for pre-processing. Social media further complicates this process with its informal language, creative spellings, and use of emojis and slang. Accurately identifying and interpreting these elements is essential for preparing the data for sentiment analysis.

Annotation, the process of manually labeling data for sentiment, also faces difficulties. The subjective nature of sentiment, coupled with the nuanced expressions and cultural contexts present in Arabic dialects, makes it challenging to establish consistent and accurate annotation guidelines. This is further compounded by the fast-evolving nature of social media language and the use of sarcasm and irony, which can easily be misinterpreted.

This paper aims to provide an overview of the current state of sentiment analysis in Arabic dialects, discussing both the challenges faced by researchers in this area and the opportunities for future advancements. By examining existing approaches and techniques employed in sentiment analysis for dialectical Arabic (DA), we hope to shed light on the complexities involved in capturing the emotional nuances of this rich and varied language.

Ultimately, our objective is to contribute to the ongoing efforts to improve sentiment analysis capabilities for Arabic dialects, enabling more accurate and comprehensive understanding of the opinions and attitudes expressed in these languages.

Sub- Tasks of Sentiment Analysis

Sentiment Analysis is a multi-task process that includes many steps inside to achieve the intended goals with high quality. It involves:

1- Text Preprocessing:

Raw text data is cleaned and prepared for further analysis. This involves removing special characters, punctuation, stop-words, and applying text normalization techniques like stemming or lemmatization.

2- Annotation:

Preprocessed text data is annotated or labeled with sentiment labels (e.g., positive, negative, neutral) to create a labeled dataset for training the sentiment analysis model (Zaghouani, Charfi & Boujelbane,2016).

3- Tokenization:

The preprocessed and annotated text is tokenized, breaking it down into individual words or tokens. Tokenization facilitates further analysis by converting the text into manageable units.

4- Feature Extraction:

Features or attributes are extracted from the tokenized text. Techniques such as bag-of-words (BoW), term frequency-inverse document frequency (TF-IDF), or word embeddings (e.g., Word2Vec, GloVe) are employed to represent the text data in a format suitable for modeling.

5- Sentiment Classification:

Using the annotated data, a machine learning or deep learning model is trained to classify the sentiment of text samples. Supervised learning algorithms, such as Support Vector Machines, Naive Bayes, or neural networks, learn to associate text patterns with sentiment labels based on the annotated examples.

6- Evaluation:

The trained sentiment analysis model is evaluated using metrics like accuracy, precision, recall, and F1-score to assess its performance. Evaluation helps gauge how well the model generalizes to unseen data and identifies areas for improvement.

7- Deployment and Monitoring:

Once the model has been trained and evaluated satisfactorily, it can be deployed in real-world applications. Continuous monitoring ensures that the model remains effective over time, with potential adjustments or retraining as needed to maintain performance

Pre-processing & Data Annotation

Pre-processing and annotation are two critical steps in sentiment analysis that can significantly affect the accuracy and reliability of results. Pre-processing refers to the initial cleaning and normalization of text data, including tasks such as tokenization, stop-word removal, stemming, and lemmatization. In the context of Arabic dialects, these processes can be particularly challenging due to the presence of non-standard words, abbreviations, and irregular spellings.

Annotation, on the other hand, involves manually labeling text data with sentiment categories, such as positive, negative, or neutral. This step is essential for training machine learning models and evaluating their performance. Moughrabi and Rambow (2018) argued that annotating Arabic dialects can be difficult due to the lack of standardized terminology and the need for domain expertise. Furthermore, the subjectivity inherent in sentiment analysis makes it challenging to achieve consensus among human annotators.

Arabic Dialects

Arabic is a Semitic language with over 300 million speakers worldwide. Although Modern Standard Arabic (MSA) serves as the common literary and administrative language across the Arab world, numerous dialects coexist, each with its distinct features and variations. Abdul-Mageed & Diab (2012) mentioned that These dialects pose significant challenges for sentiment analysis, as they often diverge from the standardized language and incorporate idiomatic expressions, colloquialisms, and non-standard grammar. Moreover, the use of social media platforms and informal communication channels further complicates the task of understanding the emotional content of texts written in Arabic dialects.

Arabic, as a language, is classified into three main varieties:

(1) Classical Arabic (CA), a form of Arabic language used in the Holy Quran and literary texts.

(2) Modern Standard Arabic (MSA), the form of Arabic used in everyday life

(3) Dialectal Arabic (DA), regional speech patterns of an informal register, common in spoken conversation but avoided in formal writing. Dialectal Arabic is specific to a region or social group who live in one nation or a group of nations with similar cultures. Dialectical Arabic is derived from CA, MSA, and foreign languages such as French, English, Spanish, Italian and Turkey

the Arab world's population prefers to use their dialects in daily conversation, which makes Arabic dialects increasingly used for communications in social media, radio programs, and TV shows

Questions of the Study

1- **Regarding collecting Data, what are the most common Arabic regions, sources of data and data sizes used in the reviewed studies?**
2- **What exactly the preprocessing techniques that are used in Arabic Dialect Sentiment Analysis?**
3- **What exactly the annotation techniques that are used in Arabic Dialect Sentiment Analysis?**
4- **What are the Approaches used in Arabic Dialect Sentiment Analysis?**

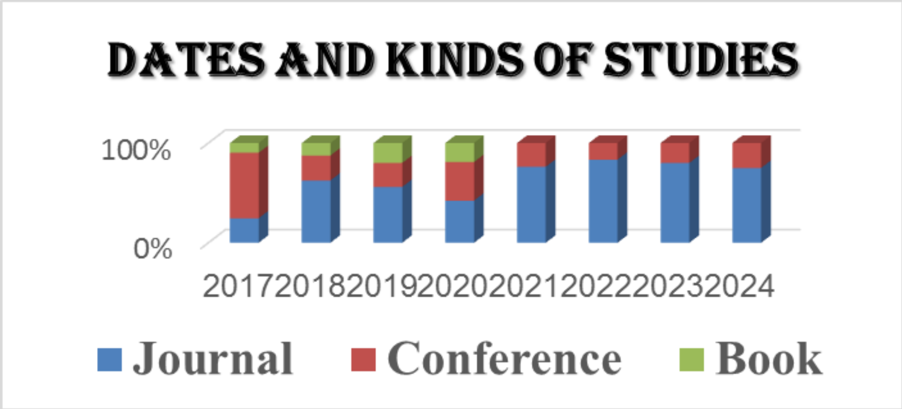
The Purpose of the study

In order to answer these questions, the study aimed at collecting the literature written in the last 7 years that tackles studying Arabic dialects in Sentiment Analysis.

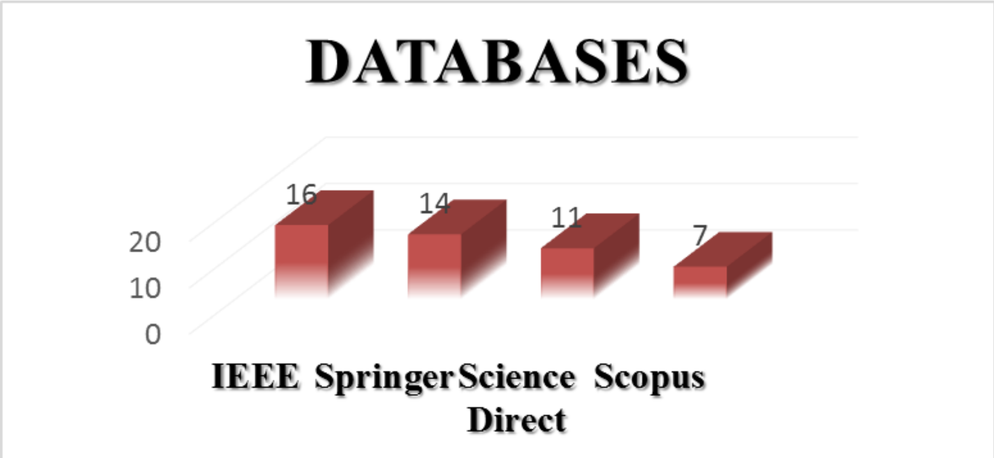
Methods

The current study adapted a plan of several steps in order to answer the targeted questions as follows:

First, the process initiated by examining the titles and abstracts of the located papers to assess their relevance. Accordingly, the studies which closely meet the required criteria were selected while the studies that were irrelevant got disregarded. The below chart shows the dates and types of the selected studies in addition to the databases used to retrieve the studies from



The below chart shows the most popular databases used to retrieve the studies



Conclusion

This section tackles answering the questions of the study as follows;

1- **Regarding collecting Data, what are the most common Arabic regions, sources of data and data sizes used in the reviewed studies?**

Upon reviewing the studies under investigation, it is found that the highest number of studies focused on multi-dialectal datasets include at least two different dialects. On the other hand, the least number of studies focused on Sudanese dialect dataset.

Regarding the mono-dialectal datasets, we found Maghrebi dialects (Moroccan, Algerian, and Tunisian) got the highest number of studies followed by the Gulf region followed by the Egyptian Dialect while the Levantine dialects (Lebanese, Syrian, Jordanian and Palestinian) come in the fourth rank. In general, most of studies included datasets of written content in Arabic dialect as well as MSA while a few numbers of datasets include exclusively a specific dialect.

Regarding the sources of the datasets under investigation, almost all the datasets were collected from social media platforms in addition to some customer review websites. Accordingly, most of the studies depend on twitter as a source for their datasets for two reasons; first, it contains a huge amount of people's opinions about a variety of controversial topics. Secondly, using the Twitter API makes it simple to scrape the public live tweets from Twitter. Facebook comes in the second rank as a source of dataset thanks to its role in the business field which helps in getting customers' reviews about products and services. Surprisingly, YouTube comes at the least rank which may reflect the least interest for the Arab audience.

When it comes to the size of dataset, studies range between those with less than 5000 comments to those with more than hundred thousand comments. What worth mentioning is that the most common number of comments among datasets ranges between ten to fifty thousand comments which indicates the sufficiency of this number of comments either for training or generalization. Accordingly, this indicates that that some machine learning techniques do not need a large number of data compared to deep learning techniques.

2- **What exactly the preprocessing techniques that are used in Arabic Dialect Sentiment Analysis?**

The complexity of preprocessing dialectical Arabic text on social media resort to many reasons, the presence of many dialect regions, common spelling mistakes, extra characters, diacritics. Accordingly, the steps of preprocessing vary according to the region of the dialect. Figure (3) below clarifies the Arabic Dialect Dataset Preprocessing Techniques.

The first step in preparing Arabic dialect text before annotation is cleaning it from all the noise and inconsistencies to make it ready in a computational format without changing the meaning. Therefore, multiple cleaning techniques have been used in most studies. The removal of non-Arabic characters is the most common technique, it has been used in 75% of the studies. Removing punctuation, repeated letters, usernames, diacritics, URLs and lengthening like "مبـروووووووووووووك" comes in the second ranking of the used techniques. Some cleaning techniques may be normalized instead of removed, such as emoticons, and hashtags. Furthermore, removing non-Arabic characters involves removing code switched, and Arabizi words such as “wa7shtony” “ alf Mabrouk” , which could result in the loss of important data.

Stop words refer to common words that are often filtered out or disregarded during NLP tasks, such as SA and text classification. These words are typically very common and do not carry significant meaning on their own in the context (Abdul-Mageed& Diab, 2014). Therefore, the majority of authors remove them by using a list of Dialectical Arabic words collected manually or using MSA stop words. Unfortunately, the majority of studies did not perform stop words removal as stop-word handling was not mentioned in 57% of the reviewed studies. on the other hand, the manual approach was the most used by 43% of the papers. Due to the structural diversity of linguistic resources in colloquial Arabic, automatic, and hybrid approaches have been less frequently used. This has led some authors, to generate stop words list using an automatic approach based merely on the frequency of letters. Examples of stop words from the Lebanese dialect: ... عمرك شمرة سألت نفسك “ ..... عم بيقوقلوا الحكومة".

Normalization is transforming a word or a letter into its standard form. It is applied to overcome spelling errors and the uniformity of Arabic characters. Many normalization techniques were used in the conducted studies such as converting hashtags to words, replacing emoticons by tags, and replacing numbers by words.

In some studies, a morphological analysis was conducted for normalization purposes such as that for Moroccan dialect. Although how important it is, 25% of the papers did not use normalization, in preprocessing stage. Almost all the normalization techniques still require manual efforts in terms of the construction of a standard lexicon. Some cleaning techniques may be avoided and treated through normalization, such as numbers, interrogative, and exclamatory punctuations, repeated letters, repeated words, and repeated negators, since they can contribute to the orientation of the sentiment.

Stemming is a computational procedure that produces the stem of a word by eliminating all of the suffixes, and prefixes from the word. However, removing prefixes and suffixes can typically change the meaning of words, especially in DA. For example, in Egyptian dialect, the root word of “متمشور" is شور which means (which could give the meaning of consulting). Light stemming, and root-based stemming are the main types of Arabic stemming procedures. The light-based stemming algorithms remove suffixes and prefixes to generate the original stems of words without the need for root extraction. While the root –based stemming employs linguistic procedures to extract the root of the word. Accordingly, the light stemming is the most preferable technique used by researchers to root stemming for two reasons; 1) it avoids semantic loss, 2) it is faster than root-based stemming method.

some studies have used a statistical stemmer based on MADA “ Morphological Analysis for Arabic Dialects," to carry out several analyses on multiple levels, such as complete lexemic, diacritic, morphological, and glossary information. MADA is a linguistic tool developed for the analysis of Arabic text, particularly focusing on dialectal variations of Arabic. It provides morphological analysis capabilities, which involve breaking down Arabic words into their constituent morphemes and providing information about their grammatical features such as roots, stems, prefixes, suffixes, and grammatical categories.

Translation &Transliteration

In several countries such as Morocco, Algeria, Tunisia and Egypt tend to use Arabizi. Many Arabic speaking generations tend to use foreign terms such as English, and French. As a result, using transliteration and translation to standardize Dialectical Arabic texts into MSA becomes an increasingly important. Reviewed Studies have used various techniques that handle Arabizi such as Buckwalter and rule-based approaches. Additionally, Google, Yamli, and Qalam APIs as well as the language model approach were rarely used. To avoid dealing directly with dialectal texts, some researchers have preferred to translate the text into MSA because of the lack of dialectical Arabic resources, lack of language models, and the lack of techniques capable of identifying the meaning of dialectal word. On the contrary, 82 % of the papers have ignored the translation step, and worked directly with DA in order to avoid the loss of relevant DA words.

Negation detection is considered a challenge in Dialectical Arabic, as it can flip the polarity of a sentiment. A few studies considered the negation issue, and included negation words as features in lexicon-based approaches. While, the machine learning based studies used the bi-gram as a feature extractor since it depends on the exitance of a prefix & a suffix in the same word as in the Moroccan dialect. for example: the positive word كنبغيك ,which means ‘‘I love you”, its negation is ‘مكنبغيكش‘ , which means ‘‘I do not love you”. With the use of the prefix ‘ما‘ , and the suffix ‘ش‘ , we can express the negation of the word ‘ بغيت ‘and the resulting word is ‘مابغيتش‘ , which means ”I do not want‘‘. In Egyptian also , مبحبكش could be the same idea of negation in Moroccan while in some urban areas they prefer putting مش in addition to the simple present of the verb as مش بحبك out of the studies reviewed, only 30 % dealt with the negation issue. Moreover, this problem has been processed only through a lexicon-based approach by using a set of rules linked to the dialect being worked upon.

3- **What exactly the annotation techniques that are used in Arabic Dialect Sentiment Analysis?**

Sentiment annotation is the labeling of an opinion / emotion inherent within a text with positive, negative or neutral polarity. As the most of datasets are extracted from social networks, it is necessary to associate a polarity to the content to conduct a sentiment analysis. It can be carried out manually, semi-automatically or automatically. Manual annotation is the most precise way since the annotation is done by linguistic experts, however, it is still a time consuming, and costly task. The automatic, and semiautomatic approaches aim to use Artificial intelligence methods. 93% of works have used manual annotation approaches. Manual annotation is sometimes subjective as it depends on the human interpretation of the sentences. In the absence of precise dialect rules, this task remains difficult to perform, that why automatic and semi-automatic annotation approaches were rarely used.

4- **What are the Approaches used in Arabic Dialect Sentiment Analysis?**

Mainly three techniques are used in the conducted studies. Machine learning based: algorithms that use pattern recognition techniques to automatically learn how to classify text according to its sentiment. They include a classical Machine Learning and

Deep Learning techniques. lexicon-based (LB) sentiment analysis: involves the use of specially crafted dictionaries that list words along with their corresponding sentiment scores. hybrid methods: attempt to combine the best features of both the LB and ML approaches. the Machine Learning based approach is the most utilized one with a total of 71.03%, while lexicon-based was the least used approach with 6.17% of the reviewed studies.

Recommendations

In conclusion, the study recommends using others ways that use linguistics tools and specifically the field of pragmatics for annotation. Moreover, it highly recommends solving the issues of Arabic dialects annotations as that will promote introducing more tools not only for the field of sentiment analysis but also to fields like text to speech and speech to text.

Bibliography

1. Abbassi, A., Zribi, I., & Belguith, L. H. (2015). Sentiment analysis in Arabic: A study of the impact of preprocessing. “International Journal on Natural Language Computing (IJNLC)”, 4(4), 33-44. doi:10.5121/ijnlc.2015.4403

2. Abdul-Mageed, M., & Diab, M. T. (2012). AWATIF: A multi-genre corpus for Modern Standard Arabic subjectivity and sentiment analysis. “Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)”, 1215-1220.

3. Abdul-Mageed, M., & Diab, M. T. (2014). SANA: A large scale multi-genre, multi-dialect lexicon for Arabic subjectivity and sentiment analysis. “Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)”, 3836-3841.

4. Abdul-Mageed, M., Diab, M. T., & Kübler, S. (2012). SAMAR: Subjectivity and sentiment analysis for Arabic social media. “Computer Speech & Language”, 28(1), 20-37. doi:10.1016/j.csl.2013.03.001.

5-Ahmed, A. A., Elmogy, A. M., & Sarhan, A. M. (2021). Sentiment analysis of Arabic tweets: Feature selection and ensemble classifiers. *Journal of Ambient Intelligence and Humanized Computing*, 12, 5931-5949. doi:10.1007/s12652-020-02605-1.

6-Alharbi, M., & Lee, M. (2020). An approach for sentiment analysis of Arabic tweets using hybrid classification. *Journal of Intelligent & Fuzzy Systems*, 38(5), 5453-5462. doi:10.3233/JIFS-179513

7-Al Sallab, A., Rafea, A., Shaalan, K., & Shoaib, M. (2015). Deep learning approach for sentiment analysis of Arabic tweets. “2015 IEEE International Conference on Data Mining Workshop (ICDMW)”, 387-392. doi:10.1109/ICDMW.2015.174

8- Al-Tamimi, J., & Hussein, R. F. (2015). Sociolinguistic variation and attitudes towards Jordanian Arabic dialects. “International Journal of Linguistics”, 7(3), 31-48. doi:10.5296/ijl.v7i3.7761

9- Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A., & Al-Ohali, Y. (2018). Sentiment analysis of Saudi dialect Arabic tweets: A hybrid approach. “Computers in Human Behavior”, 64, 210-218. doi:10.1016/j.chb.2016.11.059

10- Aliane, H., & Bouchiha, D. (2013). Sentiment analysis for Algerian dialect using a hybrid approach. “Journal of Information Technology and Software Engineering”, 3(1), 1-4. doi:10.4172/2165-7866.1000111

11-Amine, M. L., El Alaoui, I., & Sabri, A. Q. (2019). Arabic sentiment analysis using deep learning techniques. “International Journal of Electrical and Computer Engineering (IJECE)”, 9(6), 5050-5056. doi:10.11591/ijece.v9i6.pp5050-5056

12-Badaro, G., Baly, R., Hajj, H., Habash, N., & El-Hajj, W. (2014). A large scale Arabic sentiment lexicon for Arabic opinion mining. “Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)”, 165-173.
13-. Bouamor, H., Habash, N., & Oflazer, K. (2014). A multidialectal parallel corpus of Arabic. “Proceedings of the Ninth International

Conference on Language Resources and Evaluation (LREC'14)", 1240-1245.

14. Dabbagh, M., & Daud, A. (2016). A survey of sentiment analysis in Arabic: A comprehensive system perspective covering challenges and advancements. "IEEE Access", 4, 7299-7311. doi:10.1109/ACCESS.2016.2616449

15. Darwish, K., Hassan, H., & Mubarak, H. (2014). Using stemmers for Arabic information retrieval: Light, heavy, and root-based stemming. "Proceedings of the First Arabic Natural Language Processing Workshop (EMNLP)", 50-57.

16. El-Beltagy, S. R., & Ali, A. (2019). Open issues in the sentiment analysis of Arabic social media: A case study. *2013 IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)*, 72-75. doi:10.1109/AICCSA.2013.6616427

17. Elmadany, A., Azmi, A. M., Hussain, A., & Shamsuddin, S. M. (2018). Sentiment analysis and opinion mining: A survey. *Arabian Journal for Science and Engineering*, 43(8), 3745-3770. doi:10.1007/s13369-017-3034-3

18. Fahmy, A., & Darwish, K. (2014). LeSent: Lexicon-based sentiment analysis for informal Arabic. "Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)", 186-190.

19. Feldman, R. (2017). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82-89. doi:10.1145/2436256.2436274

20. Haddad, S., & Chouaib, H. (2016). Sentiment analysis of Algerian dialect: A preliminary study. "Proceedings of the Third International Conference on Social Networks Analysis, Management and Security (SNAMS)", 1-6. doi:10.1109/SNAMS.2016.7870577

21. Haddad, S., & Zitouni, M. (2021). A new annotated corpus for sentiment analysis of Algerian dialect. *Journal of King Saud University-Computer and Information Sciences*, 33(5), 561-569. doi:10.1016/j.jksuci.2018.05.004

22. Hitti, N., Jamous, R., & Daya, E. (2020). Sentiment analysis for modern standard Arabic and Lebanese dialect. *International Journal on Natural Language Computing (IJNLC)*, 9(4), 1-11. doi:10.5121/ijnlc.2020.9401

23. Hitti, N., Khouri, S., & Azar, D. (2021). Sentiment analysis for Arabic and its dialects: A survey. *Information Processing & Management*, 58(5), 102656. doi:10.1016/j.ipm.2021.102656

24. Hamouda, A., & El-Taher, M. (2015). Sentiment analyzer for Arabic comments system. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 4(3), 99-103. doi:10.14569/IJACSA.2013.040315

25. Karoui, M., Hadrich Belguith, L., & El Bahi, R. (2018). Sentiment analysis for Arabic and French tweets: An experimental study. “International Journal of Speech Technology”, 21(2), 251-263. doi:10.1007/s10772-018-9497-0

26. Medhat, W., Hassan, A., & Korashy, H. (2022). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113. doi:10.1016/j.asej.2014.04.011

27. Mulki, H., Haddad, H., Ali, A. B., & Aouad, L. (2019). Analyzing the sentiment of Arabic social media content: A survey. *Journal of Big Data*, 6(1), 1-22. doi:10.1186/s40537-019-0191-9

28. Moughrabi, I., & Rambow, O. (2018). Dialectal Arabic sentiment analysis using cross-dialectal and cross-domain data. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 187-192. doi:10.18653/v1/D18-1028

29. Shahrour, T., & Meziane, F. (2020). Sentiment analysis for Arabic dialects: A survey. “Journal of King Saud University-Computer and Information Sciences”. doi:10.1016/j.jksuci.2020.07.003

30. Zaghouani, W., Charfi, A., & Boujelbane, H. (2016). A web-based annotation tool for sentiment analysis of Arabic social media. *Procedia Computer Science*, 96, 1266-1274. doi:10.1016/j.procs.2016.08.167