DOI: 10.15849/IJASCA.241130.06

Int. J. Advance Soft Compu. Appl, Vol. 16, No. 3, November2024
Print ISSN: 2710-1274, Online ISSN: 2074-8523

IRAQIDSAD: A Dataset for Benchmarking Sentiment Analysis Tasks on Iraqi Dialect based Texts

Hafedh Hameed Hussein and Amir Lakizadeh

Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran
e-mail: hafedh.h.hussein@gmail.com
Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran
e-mail:lakizadeh@qom.ac.ir

Abstract
Sentiment analysis is a widely used technique in psychology, politics, and marketing that makes use of computational linguistics and natural language processing. The main focus of this research is a thorough literature assessment on sentiment analysis that is especially used with dialectical Arabic. Different dialects have different syntax, morphology, and grammar, which makes it difficult to classify polarity in dialectical Arabic. Our Systematic literature review investigates several aspects of sentiment analysis for dialectical Arabic in order to address these problems and support scholars working on similar projects. We identify the following phases as crucial: preprocessing, feature extraction, text annotation, and the chosen methodologies. We also present a newly corpus of 14,141 Iraqi dialect Facebook comments for benchmarking all of the text sentiment analysis and polarity classification on Iraqi texts. This corpus can be considered as an invaluable tool for sentiment analysis research in Arabic language setting

Keywords: *Iraqi dialect; polarity classification; sentiment analysis; word Embedding.*

1 Overview

The most common Semitic language, Arabic, has between 80 and 400 million native speakers and is recognized as an official language in 28 nations. Its importance extends beyond its linguistic supremacy since, to the 1.6 billion Muslims worldwide, Arabic is the sacred language of the Quran. Three primary varieties of Arabic exist: Modern Standard Arabic, the current unified form taught in schools and used in media and news; Classical Arabic, which mimics the older form found in the Quran and includes several terms seldom used in modern speech. Conversely, dialectal Arabic includes a range of regional variants that are spoken informally [1].

The colloquial Arabic spoken in daily life, known as dialectal Arabic, varies greatly between and even within nations. It has different vocabulary and pronunciations, and its grammar is not standardized like Modern Standard Arabic's. Additionally, it includes

Received 5 July 2024; Accepted 7 November 2024

terms that are exclusive to a particular dialect or that are acquired from other languages. These differences also exist in textual form, which presents difficulties for automated processing [2], [3].

Arabic dialects vary somewhat in vocabulary and grammar, leading to different writing styles and pronunciations [4]. According to research by [5], it's surprising that certain towns in various nations may have more linguistic traits than towns inside the same region. Nonetheless, Arabic dialects are divided into six regional groupings according to a taxonomy by [6]: Levantine, Egyptian, Yemeni, Iraqi, Maghreb, and Gulf. Interestingly, countries in each area frequently talk the same dialects. In comparison to Modern Standard Arabic, Arabic dialects show less linguistic ties to classical Arabic, according to research by [7]. The authors also discovered that, in contrast to Modern Standard Arabic, Arabic dialects have less linguistic resources available to them. There are a few standard Dialectical Arabic resources, like the stop word list and a large Dialectical Arabic corpus, but more and more people are calling for the creation of novel natural language processing methods that can handle different Arabic dialects without being limited by their unique characteristics.

Sentiment analysis is the technique of identifying in different languages whether a text or voice conveys good, negative, or neutral attitudes. This technique involves several processes, including text annotation, preprocessing, extraction, and text classification using specialized models.

Sentiment analysis in dialectical Arabic has been the subject of some research. [8] investigated several categorization strategies and techniques used to sentiment analysis in Arabic dialect. Conversely, [9] and [10] concentrated on the shortcomings of sentiment analysis for Arabic dialect. Many researches have produced intriguing results, but a number of issues still exist, mainly because Arabic dialects differ greatly in their morphology and character. Several researchers have noted that sentiment expressions may differ between areas, enabling the development of area-independent sentiment models, including [11] and [12] . As [13] pointed out, feature extraction is still a difficult operation that can have a big influence on the model's performance, either favorably or adversely. [14] also stressed the need of language models and preprocessing methods in order to address the intricacy of Dialectical Arabic traits like Arabizi. These initiatives help to improve sentiment analysis's precision and usefulness for Arabic dialects.

Sentiment analysis for Iraqi dialects has a number of noteworthy obstacles and restrictions, chief among them being the dearth of publicly accessible, appropriately labeled datasets designed specifically for this purpose. To address this issue, this study introduces a newly compiled corpus of Facebook comments in Iraqi dialect, aimed at benchmarking text sentiment analysis and polarity classification for Iraqi texts. The Iraqi corpus was generated from comments on Facebook using Facepager software. This corpus is a valuable resource for sentiment analysis research in the context of the Arabic language. This work aims to provide a comprehensive overview of the literature concerning research methodology, preprocessing methods, lexicon-based approaches, and machine learning methods in sentiment analysis for Iraqi dialects. This systematic literature review analyzes different facets of sentiment analysis for dialectical Arabic, aiming to identify current challenges and support researchers in the field. The identified critical phases include preprocessing, feature extraction, text annotation, and the chosen methodologies. The article is organized as follows: Section 2 describes the methodology for the systematic review; Section 3 gives an overview of relevant works; Section 4 goes into depth about the findings; Section 5 talks about the results and future directions for research; and Section 6 presents the study's conclusion.

2 Methodical Review of the Literature

In order to fully analyze the body of research on sentiment analysis for the Arabic and Iraqi dialects, we performed a systematic literature review in this study. Both quantitative and qualitative research were combined to identify essential linkages, limits, and notable discoveries by critically reviewing and choosing pertinent papers written by Arabic speakers. This provided a deeper grasp of the state of the art in this field. As per the directives provided by [15], our study utilized exacting techniques for gathering and evaluating data. The study includes all necessary procedures, including a comprehensive search of the literature, data collection on Iraqi dialects, and the extraction of important information from the chosen literature. We investigate the areas, sources, and dataset sizes utilized in the reviewed works, as well as the numerous sentiment analysis procedures applied to Arabic dialect sentiment analysis, preprocessing strategies, and feature extraction methods.

3 Related Work

Growing interest in the sentiment analysis of diverse Arabic dialects has resulted in the introduction of several methods and instruments for the categorization and analysis of Modern Standard Arabic in recent years. Scholars have investigated a number of issues surrounding sentiment analysis for Arabic dialects, such as the intricacies associated with morphology, linguistic interdependence, Arabizi transliteration, removal of negation detection, stop words, and stemming. There are three primary methods that have been identified in the available literature: lexicon-based, machine learning, and hybrid techniques. These approaches are illustrated in Fig. 1.

Utilizing a pre-existing sentiment lexicon, the lexicon-based technique collects sentiment scores for every word in a text to assess the sentiment of the content as a whole. These lexicons were developed using two different methodologies: corpus-based lexicon creation and dictionary-based lexicon creation. It's important to remember that both kinds of lexicons are domain-specific, which means they are made for specific fields or subjects. The application of lexicon features (LF) obtained from part of speech tags (POS) and stylistic and syntactic aspects has been investigated in a number of papers. For example, [16] utilized emoticons, abbreviations, and interjections as lexicon features. They also used the SVM method with N-grams, and their accuracy was 75.31%. [17] carried out many tests on a dataset of 7698 comments pertaining to the Algerian dialect.

With a lexicon-based method and a common phrase similarity calculation module, they performed a variety of preprocessing approaches, such as transliteration, translation, and khoja stemming, and achieved an accuracy of 79.13%. [18] used WordNet in a different study to retrieve principles characteristics from a collection of data with 826 tweets. They examined many classifiers, including SVM and naïve Bayes (NB), and the F1-score measure showed that the SVM method performed best, scoring 95.63%. Furthermore, using a dataset of 3484 comments, a number of manually created characteristics were taken from various publications [19], [20], [21], [22], [23].

These characteristics allowed the polarity score of comments to be calculated. These elements included negation, POS, intensifiers, emoticons, and the semantic orientation of each word. Their astounding accuracy rates for positive and negative labels were thus 98.2% and 93.2%, respectively. Additionally, a lexicon-based strategy together with numerous lexicon expansion strategies was used by the Knowledge Media Institute, The Open University, UK et al. (2019), yielding an accuracy of 69%.

By training on a vast number of samples, machine learning (ML) systems have the benefit of automated improvement over time, unlike lexicon-based approaches that depend on human-labeled texts. The Arabic language's sentiment analysis has made considerable use of this. Interestingly, Word2vec [24] has been used as a prediction-based feature extractor, and it has produced good results, as shown in trials by [25] and [26], where Word2vec performed the top. The researchers trained aa group of data with 63,000 comments from Tunisia using the logistic regression (LR) technique in one experiment. The outcome was an F1-score of 81.88%. In a different experiment, LF combined with SVM produced an accuracy of 60.6% using a dataset of 1200 tweets from Egypt. However, the accuracy increased to 70% when Word2vec was used in conjunction with the Deep Learning (DL) algorithm as a prediction-based embedding (PBE) approach.

Additionally, [27] examined the efficacy of several preprocessing methods and machine learning techniques using a group of Jordanian data with 1000 comments that was gathered from Twitter. Transliteration negative handling, stop word removal, the techniques that were employed included translating the Jordanian dialect into Modern Standard Arabic and substituting emoticons. With an accuracy rate of 76.78%, the NB method surpassed the KNN and SVM algorithms. By a 93% accuracy rate, Doc2vec was used as a PBE approach in [28] to examine a dataset of 33,000 tweets. Another work by [29] used CNN architecture and term frequency with preprocessing to perform many trials on a Moroccan dataset of 2,000 comments, obtaining 96% accuracy. Applying preprocessing measures, however, only slightly increased performance by 3%. In another research conducted by Soumeur et al. in 2018, they obtained a classification accuracy of 60.11% by utilizing the Naive Bayes classifier and Bag-of-Words feature extractor on a dataset consisting of 25,475 comments from Algeria. It is worth noting that no preprocessing was performed on the dataset.

After that, adding preprocessing processes resulted in a noteworthy nearly 10% improvement. The BiLSTM architecture was trained with fast Text word embedding in [30], yielding an accuracy of 66.78%. In the meanwhile, 56.3% accuracy was obtained in 4-way classification studies utilizing LR and Term Frequency-Inverse Document Frequency (TF IDF) on the MSTD dataset [31]. An accuracy of 55.6% was obtained by combining the Support Vector Classifier (SVC) with the Bag of Words method (BOW). Utilizing linear TF-IDF and SVC with trigram, the greatest accuracy of 77.6% was obtained for 2-way classification (positive, negative). When compared to statistical methods, prediction-based embedding methods—especially sentiment analysis—have completely transformed natural language processing jobs, even if frequency-based feature extractors still produced satisfactory results.

Furthermore, word embedding approaches are frequently paired with recurrent neural networks and other deep learning (DL) techniques. For example, [32]experimented with Word2vec and many machine learning methods on a variety of datasets. Using the SVC algorithm, they were able to get the greatest performance of 81.46% accuracy on a group of Tunisian data with 16,448 comments.

The creation of language models such as AraBERT by [33] has resulted in significant progress in the field of natural language processing (NLP) in earlier works. This language model has proven to perform exceptionally well on a range of NLP tasks. Then, using the MSTD dataset and AraBERT, [34] carried out an experiment using Bidirectional long short-term memory (BiLSTM). Their 2-way classification accuracy was an astounding 83.24%. Furthermore, with a score of 80.82% in this trial, Support Vector Machine (SVM) was shown to be the most accurate machine learning method.

The relevance of lexicon features (LF) and bag of words was addressed by [35] who used SVM to reach an accuracy of 89.24%, whereas Word2vec scored 80.36%, emphasizing the need for ongoing attempts to enhance language models. In a similar vein, SVM outperformed the other ML algorithms in [36] evaluation of a dataset including 6750 observations, attaining an astounding 94% accuracy. Other studies by [37], [38], [39], [40], [41] have also consistently shown the superiority of SVM. In these studies, The combination of TF-IDF as a feature extractor with SVM consistently yielded the most optimal outcomes. Additionally, [42] discovered that SVM was the best method, obtaining an astounding 96.6% accuracy in their tests. Together, these results highlight the usefulness of SVM in the field of NLP and motivate more developments in language modeling methods. When used with different sentiment analysis feature extractors, the SVM classifier proved to be effective. For example, TF was used in [43], resulting in a 91% accuracy rate. Similar to this, [11] achieved 88% accuracy using BOW in conjunction with a smote-based augmentation data approach on a Jordanian dataset including 2790 comments. Different feature extractor and algorithm combinations were investigated in other experiments. Using Stochastic Gradient Descent (SGD) in conjunction with word2vec, [44], [45], [46] were able to generate F1-scores of 90.16%, 90.16%, and 88%, respectively. By contrast, [19] obtained an F1-score of 85% using the SVM technique and the identical word embeddings (WE). Numerous studies concentrated on sentiment analysis of particular dialects. For instance, studies on Tunisian dialects were carried out by [47], [48]. The latter used CNN architecture and mBert to gather 9196 comments from Facebook, achieving a 93.2% accuracy rate. Different algorithms and strategies were investigated in other investigations. TF-IDF with the KNN algorithm were used by [49], who attained 92% accuracy. [50] studied the Sudanese dialect, achieving accuracy scores of 83.5% and 60.85% for two-way and three-way classifications utilizing lexicon-based characteristics and DT, respectively. Studies such as [51] incorporated datasets utilizing TF-IDF and obtained 75.25% accuracy using the DT algorithm in the context of the Gulf dialect. On the other hand, with an accuracy of 89.9%, [52] discovered that LR was the most efficient of the other ML methods. A CNN-LSTM model was trained on 15,945 comments by [53], achieving 92.26% accuracy. Sentiment analysis has demonstrated significant potential for deep learning systems. For example, [54] coupled CNN and BiLSTM for an F1-score of 89.64%, whereas [55] reached 89% using LSTM. Additionally, Multiplicative LSTM (mLSTM) construction was utilized by [56], [57]. The latter achieved an outstanding 99.75% accuracy on a group of data with 5,615,943 comments. Furthermore, with accuracy ratings of 60.85% and 83.5% for two-way and three-way classification, respectively, the scientific investigation in [58] presented the superiority of the DT over NB method. Large datasets, plenty of features, and difficult tasks have been identified to be areas where neural networks excel. DL algorithms, including RNN and, CNN achieved better results than ML methods to some scope, according to experiments by [20], [34], [59]. According to particular studies, LSTM-GRU performed the top, scoring 94.32% on a group of data with 5288 comments [60]. But in contrast, even with the use of RNN architecture and lexicon-based features, [26] only managed to attain 58.5% accuracy.

Finally, a study conducted by [44] examined the performance of Doc2vec and BOW as feature extractors using several classifiers (DT, LR, SVM, NB, and RF). Findings showed that LR was the most accurate, with 78% accuracy using BOW and 59% accuracy using Doc2vec. The goal of a study by [61] was to examine user perceptions of herbal remedies for diabetes. Their dataset has 1013 positive and 3098 negative comments; they used a variety of preprocessing techniques to correct for this imbalance. They utilized the SVM and LR classifiers, and the remarkable accuracy rates they obtained with unigram were 94.94% and 92.58%, and with trigram, 95.27% to 92.85%. Comparably, in different research by [62], the researchers investigated frequency-based methods to classify 3063 comments from various websites, such as BOW, N-gram, TF, and TF-IDF. A number of preprocessing procedures were also used, including normalizing, the removal of diacritical marks, the removal of lengthening, and the removal of repetitive characters. They achieved the highest accuracy of 76.67% using the BOW and LR method by combining the SVM, NB, SGD, and LR algorithms. In a separate study, [61] used a variety of machine learning algorithms, including NB, SVM, KNN, SGD, RF, XgBoost, AdaBoost, DT, and LR, to assess the sentiment of 2732 YouTube comments. With an accuracy of 80.12%, the SVM algorithm surpassed the others, while the LR algorithm combined with the N-gram approach achieved 81% in the

F1-score. It is noteworthy that in the field of machine learning, deep learning, or DL, has become a promising method.

Numerous researches have investigated the hybrid technique, which combines the benefits of machine learning with lexicon-based approaches, with encouraging outcomes. For example, [63] used lexicon-based characteristics with the advantages of the SVM method to reach an F1-score of 82.9%. Similar to this, [64] used SVM with bagging to get an amazing 90% accuracy. Furthermore, the hybrid technique was used in additional studies by [21], [65] on a variety of datasets, including ArSAS gathered by [66]. For polarity labels, they obtained accuracies of 73.67% using RNN and81% and 84% using SVM technique, respectively. Alternatively, [67] used lexicon-based characteristics to apply machine learning algorithms to a dataset of 2500 tweets from Jordan, achieving an impressive 91.22% accuracy in the best trial. Additionally, [68] ran many experiments using a dataset of 22,761 Facebook comments from Algeria. They utilized maximum entropy (ME) for classification after extracting pertinent features using LF and n-grams, with a 78% classification accuracy. Furthermore, CNN architecture showed a 79% accuracy rate in their investigation. All things considered, the hybrid technique shows promise in combining the best aspects of many approaches, leading to impressive results on a range of natural language processing tasks.

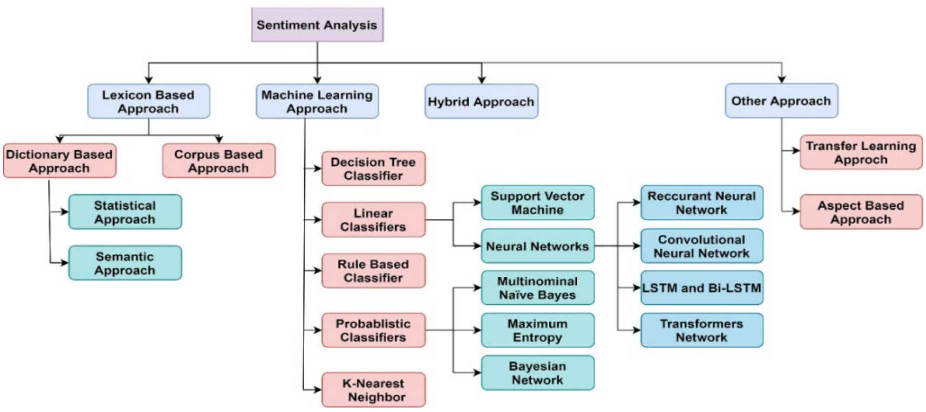

Figure 1: A summary of the popular sentiment categorization methods currently in use.

4 Review of Arabic and Iraqi Sentiment Analysis

Here we provide the main findings from our systematic literature review, with an emphasis on newly published works. Among the many publications we looked at between 2013 and 2023, these were picked because they included a variety of sentiment analysis studies in Arabic and Iraqi dialects. The combined experiment findings will be thoroughly investigated both quantitatively and qualitatively. This study will cover all phases of the lifetime of a sentiment analysis model.

4.1 Datasets Regions, Source, and Size

Because Arabic dialects differ within and across countries due to the effect of slang from adjacent countries, researchers gathered a variety of datasets from different places in order to conduct a thorough analysis. The link between Dialectical Arabic and Modern Standard Arabic allowed both datasets to be joined, containing both Arabic dialects and Modern Standard Arabic. Certain studies [44], [68], [69], [70], [71], [72] concentrated on datasets that only included Arabic dialects, for instance Gulf, Egyptian, Sudanese, Maghrebian, and Levantine dialects. Fig. 2 illustrates how many datasets from various geographic locations were gathered and used by the researchers.

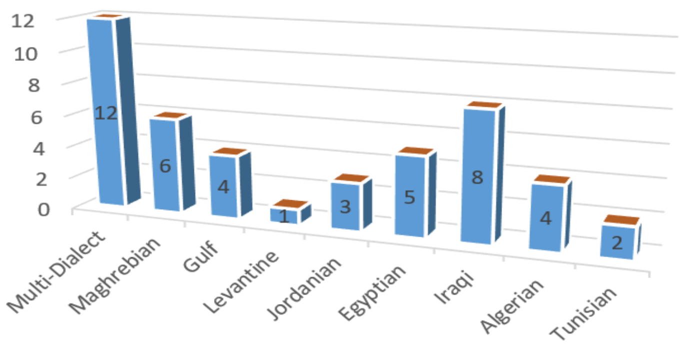

Figure 2: Quantity of datasets categorized by regions.

Information is gathered from a number of websites, including Facebook, Twitter, and others. The most popular social media network is Twitter, as seen in Fig. 3. The second-most popular is Facebook. However, just ten works have used websites, and lastly, movie reviews were the least popular data source.

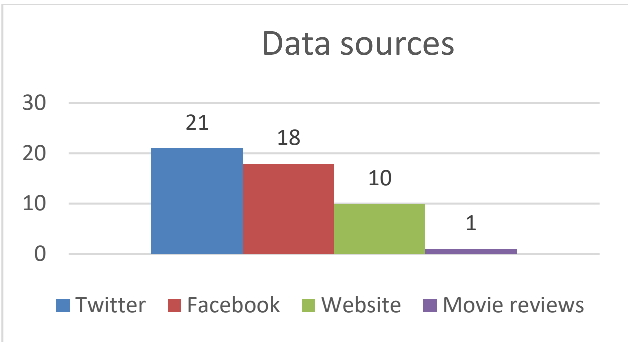

Figure 3: Quantity of datasets categorized by source.

Determining the polarity of text—that is, whether it is positive, negative, or neutral—is the main goal of sentiment analysis. Beyond that, though, it may be used to detect specific emotions like pleasure, fear, and rage in addition to the difficult work of recognizing intents. In order to better understand this topic, [59] carried out five studies in 2019 employing a variety of emotions, including fear, joy, sorrow, and rage. With an astounding peak performance of 99.82%, the average accuracy for all experiments was 48.56%. This achievement led to more research into the methods used, which showed that deep learning—in particular, the CNN architecture—performed better than machine learning (ML) techniques, which only obtained about 45% accuracy. Thus, it can be said that when data is supplied in a matrix format, the CNN architecture is more reliable. A graphic comparison of the diversity of classes among the grouped datasets is shown in Fig. 4. As a result, it aids in our understanding of the distribution and complexity of the data and allows us to determine which datasets have a greater or lower number of distinct classes.

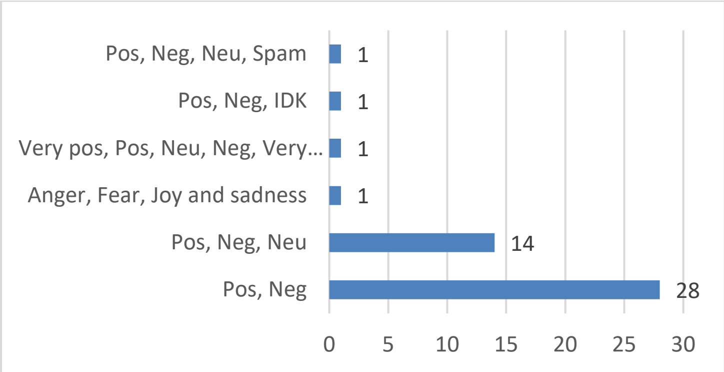

Figure 4: The type of classes of the datasets utilized.

While utilizing machine learning algorithms, the size of the dataset is a crucial factor. It is well established in the literature that a bigger dataset improves model learning and increases generalization phase accuracy. We conducted an examination of every dataset utilized in the chosen publications because the sizes of datasets might vary throughout research projects. datasets with fewer than 5000 comments at the beginning and datasets with more than 100,000 comments at the conclusion of each of the five periods created by grouping the dataset sizes as shown in Fig. 5.

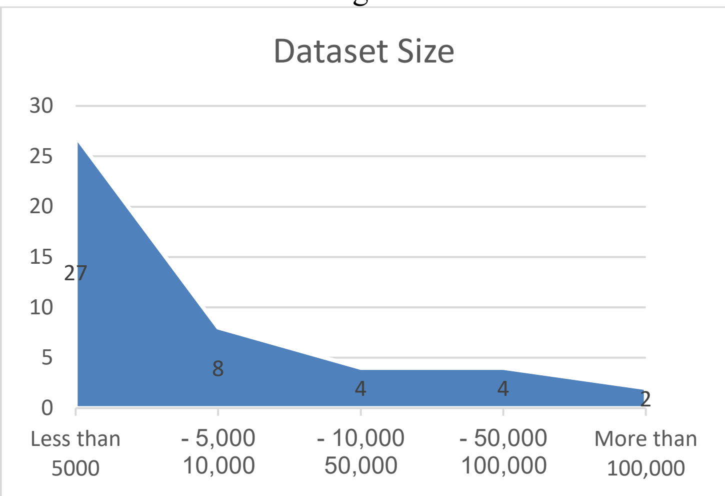

Figure 5: The quantity of the datasets utilized.

4.2 Arabic Dialectic Word Preprocessing

Analyzing and preparing dialectical Arabic writing from social media for further processing can be challenging for a number of reasons. These consist of the presence of several dialect areas, frequent misspellings, extra characters, diacritical marks, and elongations. The present study delineates the essential phases of preprocessing Arabic dialects, acknowledging that these stages may vary based on the particular area dialect diversity.

4.2.1 Cleaning

Several cleaning approaches are needed to prepare Arabic dialect text for computer analysis without changing its meaning or substance. These methods entail eliminating extraneous characters, punctuation, diacritical marks, elongations, and other irregularities. The majority of research employ various cleaning techniques to improve NLP understanding by removing superfluous information. Eliminating non-Arabic characters, punctuation, recurrent letters, lengthening, diacritical marks, usernames, and URLs are among the most often used strategies. But some cleaning methods—like handling hashtags and emoticons—are important for sentiment analysis and might be normalized rather than completely eliminated. Caution must be used when deleting non-Arabic letters because doing so might cause the loss of important information, such as code-switched and Arabizi words.

4.2.2 Stop-Words

irrelevant keywords, or Stop-words, are frequently removed in sentiment analysis in order to improve performance. For this reason, authors often use manual or Modern Standard Arabic stop-word lists; however, some have experimented with automated or hybrid techniques. NLP models can concentrate on important information by eliminating stop words, especially when it comes to sentiment analysis. Curiously, one research [73] defied the pattern shown in many earlier works that relied on manual procedures by achieving higher outcomes without stop-word removal. Notably, research by [74], [75] used an algorithmic method based on letter frequency to create lists of stop-words.

4.2.3 Normalization

The method of normalization, which involves putting words or letters into their standard forms, is mostly utilized by scholars to fix spelling mistakes and guarantee consistency in Arabic characters. Studies have used a variety of methods, such as word-for-word conversion of hashtags, emoticon tagging, and word replacement for numbers. Morphological analysis has also been used for normalization, as demonstrated by the work conducted by [76] for Tunisian dialect. The most popular method is to use ا و ه ي) (in place of characters like (ؤ إأآ ى ة) . Even with its importance in preprocessing, the majority of normalization methods still need human labor to create a standard vocabulary. Certain cleaning strategies, which might affect the sentiment orientation, whether positive or negative, can be handled by normalization. These techniques include handling numbers, repeating letters and phrases, negators, interrogative and exclamatory punctuation, and handling numbers. It's important to remember that distinct dialectical Arabic words may have different normalization terms. Regular expressions enable the automated replacement of certain letters and words.

4.2.4 Stemming

Stemming is a computer process that yields a word's core form by deleting its suffixes and prefixes. However, this process can drastically Alter the semantics of terms, particularly in dialectical Arabic. Arabic stemming techniques may be broadly classified into two categories: light-based stemming and root-based stemming. Light-based algorithms produce novel word stems without the need to extract roots; in contrast, the second method uses linguistic methodologies and heuristics to determine the word's root. Studies show that light stemming is the method of choice for academics since it maintains semantics better than root-based techniques and can be implemented more quickly.

Furthermore, a statistical stemmer based on MADA was used by [22] for thorough studies incorporating glossary, morphological, diacritic, and lexical data. However, because morphological modifications are necessary for Dialectical Arabic, the efficacy of stemming procedures may differ.

4.2.5. Translation and transliteration

Arabizi is a common writing style used in a number of countries, including Morocco, Algeria, and Tunisia. It involves writing Arabic texts using Latin letters. For example, the Moroccan Arabizi words "7awli" and (حَولي) both mean "a sheep" and are phonetically identical. Furthermore, a number of generations of Arabic speakers have a tendency to borrow vocabulary from other languages, such as French and English. As a result, the process of transliterating and translating Dialectical Arabic writings into Modern Standard Arabic has become increasingly important. Arabizi has been addressed using a variety of strategies, including as rule-based and Buckwalter approaches, as well as Qalam, Google, APIs, and Yamli, in addition to language model methodologies. Due to its casual writing style and disregard for traditional conventions, Arabizi presents substantial difficulties for sentiment analysis. Instead of tackling dialectal materials head-on, some academics have chosen to translate the information into Modern Standard Arabic. The dearth of language models, resources for dialectical Arabic, and methods for understanding dialectal word meanings are the main causes of this predilection.

4.2.6 Negation identification

Negation in dialectical Arabic can entirely invert the polarity of an idea or attitude, making it difficult to detect. A few earlier research, such those by S. The negation problem has been tackled by [1], [23], [77] by introducing words that express negativity as elements in lexicon-based techniques. On the other hand, bi-grams have been employed in machine learning research as feature extractors, taking advantage of certain prefixes and suffixes in words. Notwithstanding these endeavors, machine learning methodologies continue to face challenges in comprehending the intricacies of negation and its contextual use in Dialectical Arabic phrases.

4.2.7 Annotation techniques

Classifying a text's underlying opinion or feeling as good, negative, or neutral is known as sentiment annotation. Polarity association is required for sentiment analysis since the majority of datasets come from social networks. There are three types of annotation methods: automatic, semi-automated, and manual. Although manual annotation by language specialists produces the most accurate findings, it is expensive and time-consuming. Even while automated and semi-automatic approaches use AI to make annotating easier, they still have high mistake rates. Due to different human interpretations of sentences, manual annotation by non-experts may be subjective. Without specific dialect restrictions, this is still a difficult process to do.

4.3Arabic Dialects as Feature Extractors in Sentiment Analysis

Several feature extraction techniques, such as prediction-based embedding (PBE) and frequency-based embedding (FBE), have been used in sentiment analysis for Arabic dialect. It is still difficult to accurately capture semantic and contextual aspects of dialectical Arabic, nevertheless. Throughout time, FBEs became more and more common until PBEs appeared. Since then, PBEs have become more and more popular, mostly because of how well they are able to include semantic and contextual elements. Based on a word's frequency in the corpus, frequency-based embeddings portray every term/word as a vector according to how often it appears in each text. In the context of FBE, various techniques have been used to extract insights from the text, along with lexicon-based features that determine the sentiment orientation of words in the entire document. While these embeddings are computationally simple, they do not capture semantic or grammatical associations between words. Word frequency alone is not a reliable way to reliably characterize sentiment in Arabic dialects because of the vast variety of subtleties and meaning differences. The most often used technique among these approaches is TF-IDF, followed by lexicon-based features and Inverse Dense Frequency IDF. Because lexicon-based approaches take word meaning into consideration as well as negation and other language factors that affect sentiment, they may thus be more successful. Unfortunately, lexicon-based methods are not as good at managing the highly inflected character of dialectical Arabic and covering dialect-specific terminology, which results in less than ideal sentiment analysis outcomes. However, TF-IDF may lessen the effect of stop words by capturing word similarity and giving priority to important features in the corpus. The particular NLP job at hand and the properties of the corpus being studied influence the extraction strategy selection. Our data indicates that FBE techniques are frequently used in sentiment analysis in conjunction with ML algorithms. The number of recent research that employed frequency-based embedding approaches is shown in Fig. 6. The FBE method, reference, classifier, dataset size and source, dialect, and best efficiency are shown in Table1. This table's main goal is to present the study's conclusions on the best results, as determined by accuracy and F1-score, for each FBE approach. Interestingly, Table 1 shows that the SVM method is the majority commonly used classifier when combined with FBE approaches, consistently beating other ML classifiers.

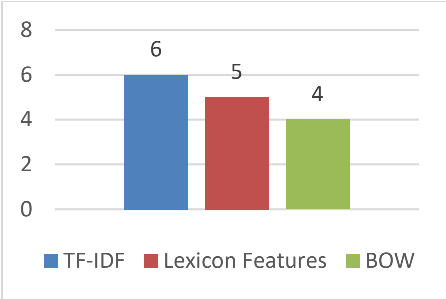

Figure 6: The quantity of recent studies used Frequency-Based Embedding techniques.

Table 1: A summary of the frequency-based approach.

Ref.	Technique	Classifier	Dialect type	Dataset size& source	F1-score & Accuracy
[57]	TF-IDF	RF	Several	5,615,943 - Twitter	98.04%Acc.
[71]	TF-IDF	SVM	Maghrebian	2569 - Facebook	90.67%F1
[1]	Lexicon features	SVM	Egyptian	2000 - Twitter	95.7%Acc.
[18]	Lexicon features	SVM	Several	826 - Twitter	95.63%F1
[78]	BOW	SVM	Levantine	1300 - Facebook	97.90%Acc.
[79]	BOW	NB	Levantine	22,550 - Twitter	87.60%F1
[43]	TF-IDF	SVM	Gulf	20,000 - Twitter	91%Acc.
[58]	TF-IDF	SVM	Maghrebian	6359 - Facebook	84.33%F1
[80]	TF-IDF	SVM	Maghrebian	147 - Website	83%F1
[81]	BOW	KNN	Iraqi	12000 -Facebook	80% Acc.
[82]	TF-IDF	NB	Iraqi	1080 -Facebook	81% Acc.
[83]	Lexicon features	SVM	Iraqi	1170-Twiter	78% Acc.
[84]	BOW	K-means	Iraqi	800-Twiter	72% Acc.
[85]	Lexicon features	RST	Iraqi	14200-Facebook	94% Acc
[86]	Lexicon features	SVM	Iraqi	1189-Movie reviews	92.8% Acc.

After training with a large group of data with 5,615,943 comments, the RF method with the TF-IDF approach achieved the maximum accuracy, showing an astonishing 98.04% accuracy. By comparison, the SVM classifier used the BOW approach with a Levantine dataset with 1300 comments, and it achieved an impressive accuracy of 97.90%. It's also important to keep in mind that deep learning methods, which mostly depend on word embeddings to capture syntactic and semantic features, do not frequently combine frequency-based approaches. Though, word embeddings for Arabic dialects might not have been advanced to the same extent as for Modern Standard Arabic. Consequently, deep learning models have difficulties in efficiently learning the nuances of Arabic dialects due to the deficiency of well-developed embeddings. The reviewed research indicates that because dialects differ significantly in language use, prediction-based algorithms have demonstrated better performance in sentiment analysis when compared to frequency-based techniques. Neural networks use continuous bag of words or skip-gram constructions to predict word occurrences in specific contexts, effectively capturing syntactic and semantic Word associations. Frequency-based methods rely on word or phrase occurrences in a corpus to identify sentiment, but this becomes difficult in dialects due to the diverse implications and meanings of words across regions. These methods provide word embeddings that may be used as feature extractors for different NLP applications. Notably, a number of words embedding methods have surfaced in the context of machine learning approaches, exhibiting remarkable results in several studies. These methods include Word2Vec, FastText, CBOW, Doc2Vec, AraVec, AraBert, mBert, and Glove. Word2Vec is the most popular feature extractor for prediction-based algorithms among them, with FastText following closely after. Deep learning architectures have been widely employed in conjunction with several prediction-based embedding (PBE) algorithms for sentiment analysis of Arabic dialects. A summary of the researches that employed the prediction-based strategy is shown in Fig. 7. Table 2 provides specific insights into the efficacy of each PBE approach.

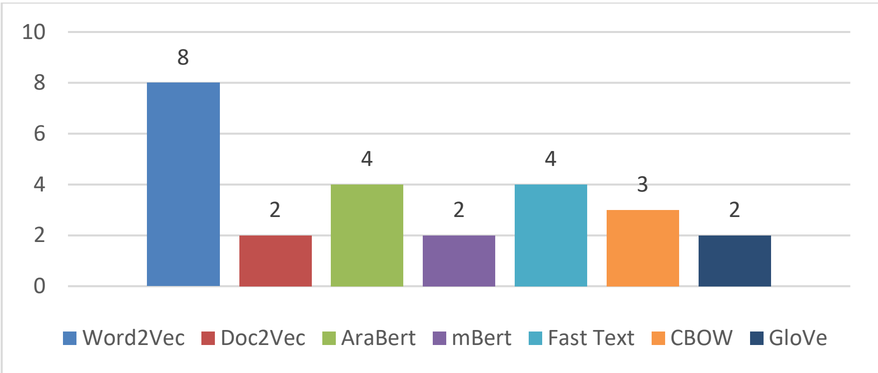

Figure 7: The quantity of recent studies used Prediction-Based Embedding techniques.

The greatest outcomes came from combining neural network-based deep learning architectures with PBE approaches. It is noteworthy, although, that the dataset's Maghrebian dialect yielded the least ratings, maybe as a result of the writing style that uses Arabici, a Latin script. Along with the experiments that implemented ML methods, [44] trained an LR classifier using Doc2vec on a group of data with 4000 embedded comments, achieving 59% accuracy. However, using a Twitter dataset of 5600 tweets, a CNN and Word2vec construction combo obtained an amazing 99.82% accuracy.

Table 2: Provides an overview of the prediction-based method.

Ref.	**Technique**	**Classifier**	**Dialect type**	**Dataset size& source**	**Accuracy & F1-score**
[57]	AraVec	mLSTM	Several	5,615,943 – Website	99.75% Acc.
[59]	Word2vec	CNN	Several	5600 – Twitter	99.82% Acc.
[45]	Word2vec	SGD	Maghrebian	1000 – Twitter	90.16% F1
[87]	AraBert	CNN BiLSTM	Several	91,000 – Website	94.20% Acc.
[75]	AraBert	GRU BiGRU	Gulf	56,674 – Twitter	90.21% F1
[60]	FastText	LSTM GRU	Several	5288 – Twitter	94.32% Acc.
[45]	FastText	MLP	Maghrebian	1000 – Facebook	90.20% F1
[48]	mBert	CNN	Maghrebian	9196 – Facebook	93.2% Acc.
[38]	GloVe	CNN	Several	54,000 – Website	90.02% Acc.
[44]	Doc2vec	LR	Maghrebian	4000 – Facebook	59% Acc.
[69]	Word2vec	LR	Maghrebian	17,541 – Facebook	81% F1
[88]	Doc2Vec	SVM	Iraqi	4000 -Facebook	82% Acc.
[89]	Word2vec	LSTM	Iraqi-Kurdish	14000-Facebook	71.35% Acc.

Recent advancements in natural language processing have led to the development of robust approaches for extracting features from textual data., particularly with regard to Arabic. Interestingly, the globally renowned open-source platform Hugging Face provides cutting edge pre-trained models created especially for Arabic language and its dialects.

4.4 Methods for Analyzing Sentiment in Arabic Dialects

The literature has offered three primary approaches for sentiment analysis: Machine learning-based, lexicon-based, and hybrid approaches. The lexicon-based (LB) method makes use of carefully chosen dictionaries that list words together with the associated emotion ratings. Table 3 summarizes the pertinent research and dataset attributes in order to assess the efficacy of this approach.

Table 3: Overview of the Lexicon-Based Approach.

Ref.	**Dataset size& source**	**Dialect type**	**Class type**	**Accuracy & F1-score**
[23]	3484 - Facebook	Egyptian	Pos	98.20% Acc.
			Neg	93.20% Acc.
[90]	4700 - Twitter	Gulf	Pos, Neg	85.40% Acc.

[17]	7698 - Facebook	Algerian	Pos, Neg, Neu	79.13% Acc.
[91]	1500 - Twitter	Gulf	Neg	77.4% F1
			Pos	59.1 % F1
			Neu	51.1% F1

Machine learning algorithms can automatically learn to categorize text according to its sentiment using pattern recognition techniques. The goal of hybrid approaches is to combine the benefits of machine learning (ML) with lexicon-based (LB) techniques. Table 4 summarizes the findings of several research that used the hybrid technique. According to the study, most of these experiments used SVM classifiers in conjunction with lexicon-based characteristics, with the first investigations showing the highest levels of accuracy. For example, by training the SVM classifier on a dataset with 2000 comments from Egyptian and Twitter websites, one research attained an astounding 95.70% accuracy. Comparably, 95.60% accuracy was obtained in another study that used the same classifier on a dataset of 2730 comments from Jordan. Nevertheless, another research containing many -dialect dataset consisting of 1200 tweets and labels produced a f1-score of 60.60%, which was significantly lower. An unbalanced and multiple classes dataset are the causes of this decreased performance.

ML-based methodologies, encompassing traditional ML and Deep Learning methods, serve as valuable tools for processing unstructured data, particularly text. The latter relies on multi-layered architectures to progressively extract higher-level features, achieved through transformers, a specific type of Deep Learning technique. Notably, starting from 2016, DL techniques have garnered interest in sentiment analysis for Arabic dialects, although their utilization remains comparatively lower than ML techniques. This disparity is attributable to DL's demand for significant computing power and large datasets to achieve accurate generalization. Conversely, ML models are simpler and impose fewer dataset size requirements. Within the realm of reviewed studies, a diverse range of ML algorithms have been employed, involving NB, SVM, and ensemble methods like bagging. The maximum popular procedures in ML techniques are shown in Fig. 8. It is evident that ensemble techniques are the least used, and SVM and NB are the most. In Table 5, the prevalence of various machine learning algorithms is depicted. Evidently, SVM and NB stand out as the most favored techniques, whereas ensemble methods are notably less prevalent. DT, LR, KNN, and RF were utilized at similar rates.

Table 4: Hybrid Approach Overview.

Ref.	**Source**	**Dataset size**	**Dialect type**	**Class type**	**Classifier**	**Accuracy & F1-score**
[1]	Websites &Twitter	2000	Egyptian	Neu, Neg, Pos	SVM	95.70% Acc.
[77]	Website	2730	Jordanian	Neg, Pos	SVM	95.60% Acc.

[73]	Twitter & Facebook	7366	Tunisian	Neg, Pos	SVM	94% Acc.
[67]	Website	2500	Jordanian	Neg, Pos	KNN	91% Acc.
[35]	Twitter	943	Multi-dialects	Neg, Pos	SVM	91% Acc.
[64]	Twitter	1111	Egyptian	Neg, Pos	Bagging with SVM	90% Acc.
[21]	Twitter	7800	Egyptian	Neu, Neg, Pos	SVM	84% Acc.
[63]	Websites	1350	Egyptian	Neg, Pos	SVM	82.97% F1
[26]	Twitter	1200	Several	Neu, Neg, Very Neg , Pos, Very Pos	SVM	60.60% F1
[88]	Facebook	4000	Iraqi	Neg, Pos	SVM	82% Acc.

In light of our analysis, it is obvious that ensemble learning approaches, which harness predictions from several models to boost performance, have not been fully exploited, alongside SGB and ME techniques. Moreover, the deep learning technique, built around neural networks, provides a diversity of designs. A quantitative analysis of the deep learning techniques employed in the evaluated papers is shown in Fig. 9.

Also, Table 6 presents a detailed statistical evaluation of the various DL designs implemented in the examined research. These comprise a varied variety of alternatives, including LSTM, CNN, GRU, BiLSTM, RNN, and BiGRU, employed either separately or in combination. Known for its ability in extracting local and position-invariant properties, as well as long-range semantic relationships, CNN and LSTM have earned considerable attention.

Table 5: Overview of Machine Learning Algorithms.

Ref.	**Dialect type**	**Source**	**Dataset size**	**Algorithm**	**Feature extractor**	**Class type**	**Accuracy & F1-score**
[78]	Levantine	Facebook	1300	SVM	BOW	Pos, Neg, Neu	97.90% Acc.
[18]	Several	Twitter	826	SVM	Lexicon features	Pos, Neg	95.63% F1
[42]	Several	Twitter	1500	NB	Lexicon features	Pos, Neg, Neu	96.60% Acc.
[18]	Several	Twitter	826	NB	Lexicon features	Pos, Neg	91.77% F1

[57]	Several	Twitter	5,615,943	LR	TF-IDF	Pos, Neg, Neu	97.72% Acc.
[45]	Maghrebian	Facebook	1000	LR	Word2vec	Pos, Neg	89.13% F1
[57]	Several	Twitter	5,615,943	RF	TF-IDF	Pos, Neg, Neu	98.04% Acc.
[46]	Maghrebian	Facebook	3048	RF	Word2vec	Pos, Neg	80% F1
[57]	Several	Twitter	5,615,943	DT	TF-IDF	Pos, Neg, Neu	97.71% Acc.
[55]	Several	Twitter	2000	DT	TF-IDF	Pos, Neg, Neu	75.4% F1
[78]	Levantine	Facebook	1300	KNN	BOW	Pos, Neg, Neu	96.80% Acc.
[32]	Several	Twitter Website	1951	SGD	Word2vec	Pos, Neg	79.52% Acc.
[92]	Maghrebian	Facebook	254,000	SGD	Word2vec	Pos, Neg	90.16% F1
[38]	Several	Website	54,000	XgBoost	Word2vec	Pos, Neg	88.71% Acc.
[87]	Several	Website	91,000	AdaBoost	AraBert	Pos, Neg	84.20% Acc.
[29]	Maghrebian	Multi-dialects	3355	ME	Word2vec	Pos, Neg	83.9% Acc.

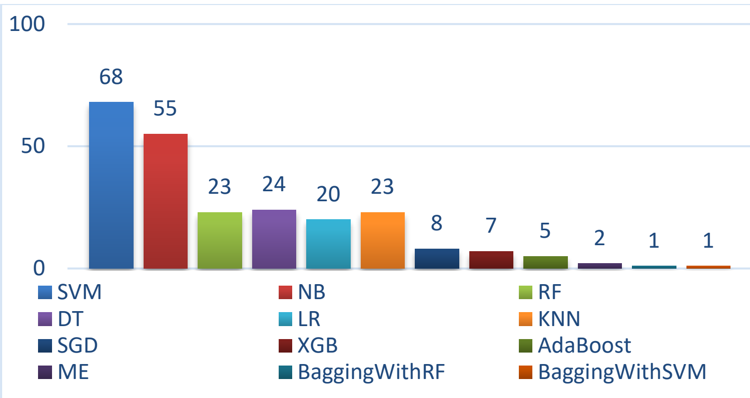

Figure 8: The quantity of recent studies used machine learning techniques.

Table 6: Overview of deep learning algorithms.

Ref.	Algorithm	Feature extractor	source	Dataset size	Dialect type	Class type	Accuracy & F1-score
[57]	mLSTM	AraVec	Twitter	5,615,943	Several	Neu, Neg, Pos	99.75% Acc.
[59]	CNN	Word2vec	Twitter	5600	Several	Sadness, Fear Joy, Anger	99.82% Acc.
[45]		FastText	Facebook	1000	Maghrebian	Neg, Pos	87.99% F1
[57]	GRU	AraVec	Twitter	5,615,943	Several	Neu, Neg, Pos	98.80% Acc.
[75]		AraBert	Twitter	56,674	Gulf	Neu, Neg, Pos	82.08% F1
[87]	BiLSTM	AraBert	Website	91,000	Several	Neg, Pos	93.70% Acc.
[69]		FastText	Facebook	1000	Maghrebian	Neg, Pos	88.19% F1
[29]	LSTM	Word2vec	Multi	2000	Maghrebian	Neg, Pos	97.20% Acc.
[55]		Word2vec	Twitter	2000	Several	Neu, Neg, Pos	87.5% F1
[60]	LSTM-GRU	FastText	Twitter Website	5288	Several	Neg, Pos	94.32% Acc.
[87]	CNN-BiLSTM	AraBert	Website	91,000	Several	Neg, Pos	94.2% Acc.
[87]	CNN-LSTM	AraBert	Website	91,000	Several	Neg, Pos	94% Acc.
[75]	BiGRU	AraBert	Twitter	56,674	Gulf	Neu, Neg, Pos	81.59% F1
[89]	LSTM	Word2vec	Facebook	14000	Iraqi-Kurdish	Neu, Neg, Pos	71.35% Acc.

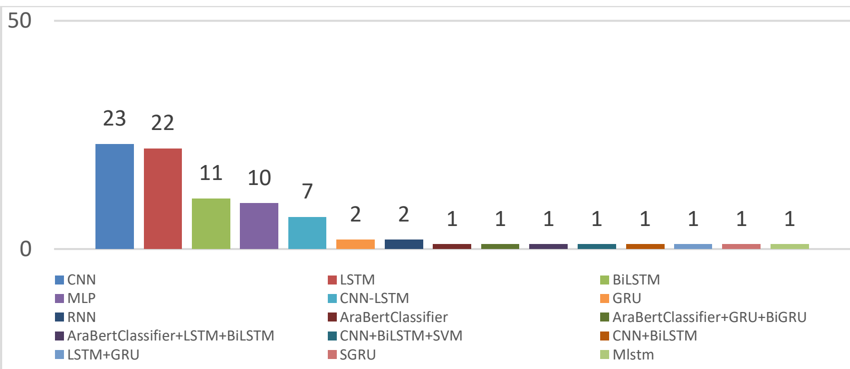

Figure 9: The quantity of recent studies used deep learning techniques.

5 IRAQISAT Corpus for Sentiment Analysis

Arabic is the official language among the many languages spoken by Iraqi people. Arabic dialects differ from standard Arabic in that they don't follow standardized dictionaries or standards and have different writing styles. Arabic dialect processing is more complicated than standard Arabic since social media users frequently express their thoughts in their dialects, which results in distinctive writing styles. So, the study's second objective is to create a corpus of Iraqi dialect from some Facebook pages.

Facepager program was used to harvest data from four Iraqi Facebook sites for this study. ("دليل مطاعم بغداد": Baghdad Restaurants Directory); ("برنامج ولايةبطيخ": Melon City show); ("ستيفن نبيل": Steven Nabil) and ("بغداد": Baghdad) page.Consequently, this program was used to get dozens of CSV files containing thousands of comments.Before processing, there were 18,656 sentences in the gathered comments.

Fig. 10 shows the steps involved in constructing a dataset. In order to address numerous concerns that might affect accuracy, a filtering sub-stage was built on the gathered comments following the data gathering step. Commenting with only one character or simple symbols, using profanity excessively, writing in Kurdi, English, or another language, having Facebook reactions (such as love, haha, wow, sad, or angry), having only tagged names, redundancies, and having links, mentions, or photo scraps are some examples of these problems. These kinds of remarks were eliminated from the dataset. The annotation step entailed manually classifying the remaining comments into four groups (0, 1, 2), which stood for positive, negative, and normal classifications, respectively, after the filtering sub-stage. Assuming that each remark represents an opinion, each one was carefully reviewed and given a label. Two distinct experts independently reviewed the annotations to confirm the legitimacy of the data annotations; their findings were 100% accurate and compatible with the annotations from the corpus. We ensured that all classes: positive, negative, and normal had the same comment size while developing corpus, as seen in Fig. 11. The created corpus is referred to as IRAQISAT, and it contains 14,141 annotated comments for sentiment analysis of Iraqi dialects.

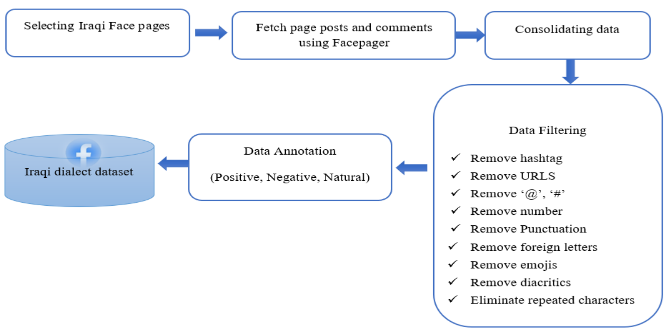

Figure 10: Iraqi corpus creation.

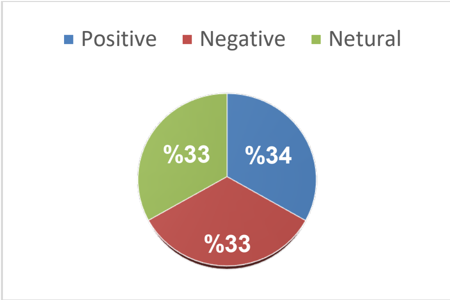

Figure 11: The data ratio for each class (positive, negative, and normal).

6 Discussion

This section focuses on the findings of our Systematic Literature Review and provides potential directions for future study. The objectives were twofold: first, to provide a comprehensive overview and analysis of the key stages involved in sentiment analysis for Arabic dialect; and second, to identify and address the limitations present in the reviewed studies. The study of sentiment analysis for Arabic dialect has mostly focused on three significant viewpoints. At first, the primary focus is on the dataset's features, such as its volume, source, and area, since these factors significantly impact the results of sentiment analysis methods for Arabic dialect. Subsequently, much emphasis is dedicated to the preprocessing approaches, which have a large influence on the effectiveness of sentiment analysis methods. Another key part is featuring extraction, which has a significant impact on constructing feature vectors and therefore strongly influences the sentiment analysis process. Finally, the classification technique comes into action, getting feature vectors from the extractor and establishing the sentiment characteristics. Given the variable dataset sizes among study articles, this topic warrants particular consideration in future investigations.

The enhancement of sentiment categorization may be obtained by cleaning inconsistent and noisy data, including non-Arabic fonts, repetitive letters, URLs, and punctuations. Our study illustrates the usefulness of data cleaning strategies at the word level in strengthening sentiment analysis systems. Inconsistent data may also be handled with normalization approaches, which minimize text variances and increase stemming efficiency. Removing stop words, which are frequent keywords with minimal information, is vital to minimize superfluous noise in NLP jobs. However, manual stop words removal typically leads to domain and region-specific difficulties. To address these constraints, a hybrid strategy to stop words removal is advised based on our study's findings. This methodology mitigates the drawbacks of both human and automatic procedures, assuring the preservation of the text's overall meaning. In the preprocessing step, feature extraction plays a vital role as it enables models to gain semantic and contextual information from feature representations. To examine this step's success, we did a quantitative comparison of different methodologies, indicating the prediction-based approach's small superiority in accuracy ratings, particularly when applying DL algorithms. Our study also demonstrated how accuracy may be impacted by numerous aspects including datasets, preprocessing approaches, and feature extractors. Given the ubiquitous usage of the ML technique, particularly Deep Learning (DL), for extracting highly effective features from enormous datasets, applying the DL model in Arabic dialect sentiment analysis becomes vital to address issues unsolvable by classic approaches like as SVM and NB. Notably, the SVM classifier displayed higher performance above the NB classifier. Interestingly, ensemble approaches, despite their demonstrated effectiveness, have not gotten significant attention. Consequently, to address this gap, we did a detailed comparison of outcomes achieved from ensemble and non-ensemble techniques.

Despite the gains of the ML technique, many difficulties such as negation remain unanswered, necessitating a lexicon-based approach. As a result, a hybrid technique incorporating both methodologies appear to give ideal outcomes. Several works, including [1], [23], [77], have effectively handled negation utilizing lexicon-based characteristics. Alternatively, [78] uses a mix of bag-of-words and bigram characteristics to collect negation words. Some additional investigations adopted a mixed strategy, with the SVM algorithm giving the most favorable results. Meanwhile, in DL-based techniques, multiplicative LSTM and CNN LSTM displayed greater performance.

7 Conclusion

In this study, we did a thorough literature assessment of significant publications on Arabic dialect sentiment analysis, while simultaneously concentrating on developing a sizable corpus to solve the issues in sentiment analysis for Iraqi dialects. The analyzed literature shows an increasing interest from the NLP research community towards sentiment analysis for Arabic dialects. We reviewed the most up-to-date Arabic research to investigate the various methodologies and methods utilized for sentiment analysis in this context. Preprocessing procedures, including stop word removal, stemming, negation, cleaning, translation, and transliteration, were carefully investigated. Moreover, we studied feature extraction methodologies covering both machine learning (ML) and deep learning (DL) algorithms. Additionally, we conducted an analysis of data representation with regards to its dimensions, extent, and sources., seeking to select the best acceptable datasets for sentiment analysis of Arabic dialects. Performance evaluation utilizing several ML approaches, such as probabilistic, non-parametric, and parametric algorithms, demonstrated the considerable influence of sentiment dataset quality on sentiment analysis findings. We discovered that combining deep learning approaches with lexicon-based features, such as negation features, offered the most promising outcomes.

8 Future Works

For the Iraqi language to have a future in technology, it still needs Iraqi computational resources, which means that researchers will need to work hard and patiently. Moreover, information about the Iraqi dialect is required for sentiment analysis. To improve the Iraqi Dialect corpus even further, labeled data is required. As future research, our work will compile the sentiment dataset from various sources, including Facebook, Twitter, YouTube, Instagram, and additional social media platforms. Furthermore, our work will focus on enhancing performance through the implementation of contemporary methodologies, including transformer-based strategies.

References

[1]. Ibrahim, H. S., Abdou, S. M., & Gheith, M. H. (2015). Sentiment analysis for modern standard Arabic and colloquial. *ArXiv*. Available online at https://api.semanticscholar.org/CorpusID:16258461

[2]. Khrisat, A. A., & AlYharthy, Z. A. (2015). Arabic dialects and classical Arabic language. *Advances in Social Sciences Research Journal, 2*(3), 254–260.

[3]. Mallek, F., Belainine, B., & Sadat, F. (2017). Arabic social media analysis and translation. *Procedia Computer Science, 117*, 298–303. https://doi.org/10.1016/j.procs.2017.10.121

[4]. Harrat, S., Meftouh, K., Abbas, M., & Saad, M. (2015). Cross-dialectal Arabic processing. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing* (pp. 620–632). Springer International Publishing.

[5]. Alsudais, A., Alotaibi, W., & Alomary, F. (2022). Similarities between Arabic dialects: Investigating geographical proximity. *Information Processing & Management, 59*(1).

[6]. Mubarak, H., & Darwish, K. (2014). Using Twitter to collect a multi-dialectal corpus of Arabic. In *ANLP@EMNLP*. Available online at https://api.semanticscholar.org/CorpusID:18393106

[7]. Alshutayri, A., & Atwell, E. (2018). A social media corpus of Arabic dialect text. *Linguistics, Computer Science.*

[8]. Al Shamsi, A. A., & Abdallah, S. (2021). Text mining techniques for sentiment analysis of Arabic dialects: Literature review. *Advances in Science, Technology and Engineering Systems Journal, 6*, 1012–1023. Available online at https://api.semanticscholar.org/CorpusID:233284449

[9]. Dashtipour, K., Mahdi, S., & Hussain, A. (2016). Multilingual sentiment analysis: State of the art and independent comparison of techniques. *Cognitive Computation, 8*(4), 757–771. https://doi.org/10.1007/s12559-016-9415-7

[10]. Boudad, N., Faizi, R., Oulad Haj Thami, R., & Chiheb, R. (2018). Sentiment analysis in Arabic: A review of the literature. *Ain Shams Engineering Journal, 9*(4), 2479–2490. https://doi.org/10.1016/j.asej.2017.04.007

[11]. Obiedat, R., Tarawneh, M., El-Reedy, A., & Ebidat, H. (2022). Sentiment analysis of customers' reviews using a hybrid evolutionary SVM-based approach in an imbalanced data distribution. *IEEE Access, 10*, 22260–22273. Available online at https://api.semanticscholar.org/CorpusID:246666702

[12]. Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: A tertiary study. *Artificial Intelligence Review, 54*(7), 4997–5053. https://doi.org/10.1007/s10462-021-09973-3

[13]. Hussein, S., Farouk, M., & Hemayed, E. S. (2019). Gender identification of Egyptian dialect in Twitter. *Egyptian Informatics Journal, 20*(2), 109–116. https://doi.org/10.1016/j.eij.2018.12.002

[14]. Ghallab, A., Mohsen, A., & Ali, Y. (2020). Arabic sentiment analysis: A systematic literature review. *Hindawi Limited.* https://doi.org/10.1155/2020/7403128

[15]. Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research. *Research Methods & Methodology in Accounting eJournal.* Available online at https://api.semanticscholar.org/CorpusID:18255235

[16]. Ben Hamouda, S., & Akaichi, J. (2013). Social networks' text mining for sentiment classification: The case of Facebook statuses updates in the 'Arabic Spring' era. Available online at https://api.semanticscholar.org/CorpusID:212560849

[17]. Mataoui, M., Zelmati, O., & Boumechache, M. (2016). A proposed lexicon-based sentiment analysis approach for the vernacular Algerian Arabic. *Research in Computing Science, 110*, 55–70. Available online at https://api.semanticscholar.org/CorpusID:31837949

[18]. Alowaidi, S., Saleh, M., & Abulnaja, O. (2017). Semantic sentiment analysis of Arabic texts. *International Journal of Advanced Computer Science and Applications.* Available online at www.ijacsa.thesai.org

[19]. Abdelli, A., Guerrouf, F., Tibermacine, O., & Abdelli, B. (2019). Sentiment analysis of Arabic Algerian dialect using a supervised method. In *2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS)* (pp. 1–6). https://doi.org/10.1109/ISACS48493.2019.9068897

[20]. Abdelhade, N., Soliman, T. H. A., & Ibrahim, H. M. (2018). Detecting Twitter users' opinions of Arabic comments during various time episodes via deep neural network. In *Advances in Intelligent Systems and Computing* (pp. 232–246). Springer. https://doi.org/10.1007/978-3-319-64861-3_22

[21]. El-Makky, N., et al. (2015). *The 3rd ASE International Conference on Social Informatics (Social Informatics 2014)*. In *Sentiment Analysis of Colloquial Arabic Tweets*. Harvard University, USA. Available online: https://www.researchgate.net/publication/282770064

[22]. Habash, N., Rambow, O., & Roth, R. (2009). MADA + TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming, and lemmatization. Available online: https://api.semanticscholar.org/CorpusID:10124099

[23]. Rouby, I., Badawy, M., Nour, M., & Hegazi, N. (2018). Performance evaluation of an adopted sentiment analysis model for Arabic comments from Facebook. Available online: https://api.semanticscholar.org/CorpusID:218899867

[24]. Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In E. P. Xing & T. Jebara (Eds.), *Proceedings of the 31st International Conference on Machine Learning*, *Proceedings of Machine Learning Research* (Vol. 32, pp. 1188–1196). PMLR. Available online: https://proceedings.mlr.press/v32/le14.html

[25]. Altowayan, A. A., & Tao, L. (2016). Word embeddings for Arabic sentiment analysis. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 3820–3825). https://doi.org/10.1109/BigData.2016.7841054

[26]. Baly, R., et al. (2017). Comparative evaluation of sentiment analysis methods across Arabic dialects. *Procedia Computer Science*, 117, 266–273. https://doi.org/10.1016/j.procs.2017.10.118

[27]. Duwairi, R. M., Marji, R., Sha'ban, N., & Rushaidat, S. (2014). Sentiment analysis for dialectical Arabic. In *Proceedings of the 6th International Conference on Information and Communication Systems (ICICS)* (pp. 166–170). IEEE. Available online: https://www.researchgate.net/publication/267993457

[28]. Alnawas, A., & Arici, N. (2018). Effect of word embedding variable parameters on Arabic sentiment analysis performance. In *5th International Conference on Computational and Experimental Science and Engineering (ICCESEN-2018)* (pp. 12–16).

[29]. Oussous, A., Benjelloun, F.-Z., Lahcen, A. A., & Belfkih, S. (2019). ASA: A framework for Arabic sentiment analysis. *Journal of Information Science*, 46, 544–559. Available online: https://api.semanticscholar.org/CorpusID:181819653

[30]. Adouane, W., Touileb, S., & Bernardy, J.-P. (2020). Identifying sentiments in Algerian code-switched user-generated comments. In *Proceedings of the 12th Edition of Language Resources and Evaluation Conference (LREC)* (pp. 2698–2705). European Language Resources Association (ELRA).
[31]. Mihi, S., Ait, B., El, I., Arezki, S., & Laachfoubi, N. (2020). MSTD: Moroccan sentiment Twitter dataset. *International Journal of Advanced Computer Science and Applications*, 11. Available online: https://api.semanticscholar.org/CorpusID:229214242
[32]. Elfaik Hanane, E. H., & Nfaoui, N. (2020). A comparative evaluation of classification algorithms for sentiment analysis using word embeddings. In M. Ezziyyani (Ed.), *Advanced Intelligent Systems for Sustainable Development (AI2SD'2019)* (pp. 1–11). Springer International Publishing.
[33]. Antoun, W., Baly, F., & Hajj, H. M. (2020). AraBERT: Transformer-based model for Arabic language understanding. *ArXiv*, vol. abs/2003.00104. Available online: https://api.semanticscholar.org/CorpusID:211678011
[34]. Matrane, Y., Benabbou, F., & Sael, N. (2021). Sentiment analysis through word embedding using AraBERT: Moroccan dialect use case. In *2021 International Conference on Digital Age & Technological Advances for Sustainable Development (ICDATA)* (pp. 80–87). https://doi.org/10.1109/ICDATA52997.2021.00024
[35]. Touahri, I., & Mazroui, A. (2021). Enhancement of a multi-dialectal sentiment analysis system by the detection of implied sarcastic features. *Knowledge-Based Systems*, 227, 107232. Available online: https://api.semanticscholar.org/CorpusID:236290539
[36]. Dahbi, R., Monir, M., & Saadane, S. (2020). Citizen sentiment analysis in social media: Moroccan dialect as case study. In R. Ben Ahmed, A. Boudhir, & K. Ismail (Eds.), *Innovations in Smart Cities Applications Edition 3* (pp. 16–29). Cham: Springer International Publishing.
[37]. Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A., & Al-Ohali, Y. (2017). AraSenTi-Tweet: A corpus for Arabic sentiment analysis of Saudi tweets. *Procedia Computer Science*, 63–72. https://doi.org/10.1016/j.procs.2017.10.094
[38]. Elnagar, A., Lulu, L., & Einea, O. (2018). An annotated huge dataset for standard and colloquial Arabic reviews for subjective sentiment analysis. *Procedia Computer Science*, 142, 182–189. https://doi.org/10.1016/j.procs.2018.10.474
[39]. Kwaik, K. A., Chatzikyriakidis, S., Dobnik, S., Saad, M., & Johansson, R. (2020). An Arabic tweets sentiment analysis dataset (ATSAD) using distant supervision and self-training. In *OSACT*. Available online: https://api.semanticscholar.org/CorpusID:219307254
[40]. Alowisheq, A., et al. (2021). MARSA: Multi-domain Arabic resources for sentiment analysis. *IEEE Access*, 9, 142718–142728. https://doi.org/10.1109/ACCESS.2021.3120746
[41]. Abuuznien, S., Abdelmohsin, Z., Abdu, E., & Amin, I. (2021). Sentiment analysis for Sudanese Arabic dialect using a comparative supervised learning approach. In *2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)* (pp. 1–6). https://doi.org/10.1109/ICCCEEE49695.2021.9429560
[42]. Ayadi, R., Shahin, O. R., Ghorbel, O., Alanazi, R., & Saidi, A. (2021). Sentiment analysis of COVID-19 tweets: Impact of preprocessing step. *IJCSNS International Journal of Computer Science and Network Security*, 21(3), 206. https://doi.org/10.22937/IJCSNS.2021.21.3.28
[43]. Almuqren, L., & Cristea, A. (2021). AraCust: A Saudi Telecom tweets corpus for sentiment analysis. *PeerJ Computer Science*, 7, 1–30. https://doi.org/10.7717/PEERJ-CS.510
[44]. Guellil, I., Adeel, A., Azouaou, F., & Hussain, A. (2018). SentiALG: Automated corpus annotation for Algerian sentiment analysis. *ArXiv*, abs/1808.05079. Available online: https://api.semanticscholar.org/CorpusID:52008929
[45]. Guellil, I., Azouaou, F., & Chiclana, F. (2020). ArAutoSenti: Automatic annotation and new tendencies for sentiment classification of Arabic messages. *Social Network Analysis and Mining*, 10. Available online: https://api.semanticscholar.org/CorpusID:221381261

[46]. Guellil, I., et al. (2021). A semi-supervised approach for sentiment analysis of Arabic messages: Application to the Algerian dialect. *SN Computer Science*, 2(2). https://doi.org/10.1007/s42979-021-00510-1

[47]. Medhaffar, S., Bougares, F., Estève, Y., & Belguith, L. H. (2017). Sentiment analysis of Tunisian dialects: Linguistic resources and experiments. In *WANLP@EACL*. Available online: https://api.semanticscholar.org/CorpusID:6825507

[48]. Messaoudi, A., Haddad, H., & Chawki, Y. (2021). Learning word representations for Tunisian sentiment analysis. In C. Djeddi, & M. Kessentini (Eds.), *Pattern Recognition and Artificial Intelligence* (pp. 329–340). Cham: Springer International Publishing.

[49]. Ismail, R., Omer, M., Mahadi, N., & Amin, I. (2018). Sentiment analysis for Arabic dialect using supervised learning. In *International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)* (pp. 1–6). IEEE.

[50]. Abo, M. E. M., Shah, N. A. K., Balakrishnan, V., Kamal, M., Abdelaziz, A., & Haruna, K. (2019). Subjectivity and sentiment analysis of Sudanese dialect Arabic. In *2019 International Conference on Computer and Information Sciences (ICCIS)* (pp. 1–5). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICCISci.2019.8716466

[51]. Alzyout, M., AL Bashabsheh, E., Najadat, H., & Alaiad, A. (2021). Sentiment analysis of Arabic tweets about violence against women using machine learning. In *12th International Conference on Information and Communication Systems (ICICS)* (p. 508). Valencia, Spain.

[52]. Aljabri, M., et al. (2021). Sentiment analysis of Arabic tweets regarding distance learning in Saudi Arabia during the COVID-19 pandemic. *Sensors*, 21(16). https://doi.org/10.3390/s21165431

[53]. Alwehaibi, A., Bikdash, M., Albogmi, M., & Roy, K. (2022). A study of the performance of embedding methods for Arabic short-text sentiment analysis using deep learning approaches. *Journal of King Saud University - Computer and Information Sciences*, 34(8), 6140–6149. https://doi.org/10.1016/j.jksuci.2021.07.011

[54]. Alharbi, O. (2021). A deep learning approach combining CNN and Bi-LSTM with SVM classifier for Arabic sentiment analysis. *International Journal of Advanced Computer Science and Applications*, 12(6). Available online: https://en.unesco.org/commemorations/worldarabiclanguageday

[55]. Abdellaoui, H., & Zrigui, M. (2018). Using tweets and emojis to build TEAD: An Arabic dataset for sentiment analysis. *Computacion y Sistemas*, 22(3), 777–786. https://doi.org/10.13053/CyS-22-3-3031

[56]. Al-Azani, E.-S. M., & El-Alfy, S. (2017). Hybrid deep learning for sentiment polarity determination of Arabic microblogs. In D. Liu & Y. Xie (Eds.), *Neural Information Processing* (pp. 491–500). Cham: Springer International Publishing.

[57]. Abdelminaam, D. S., Neggaz, N., Gomaa, I. A. E., Ismail, F. H., & Elsawy, A. A. (2021). Arabicdialects: An efficient framework for Arabic dialects opinion mining on Twitter using optimized deep neural networks. *IEEE Access*, 9, 97079–97099. https://doi.org/10.1109/ACCESS.2021.3094173

[58]. Maghfour, A., Mohcine, M., & Elouardighi, A. (2018). Standard and dialectal Arabic text classification for sentiment analysis. In H. Abdelwahed, M. G. Bellatreche (Eds.), *Model and Data Engineering* (pp. 282–291). Cham: Springer International Publishing.

[59]. Baali, M., & Ghneim, N. (2019). Emotion analysis of Arabic tweets using a deep learning approach. *Journal of Big Data*, 6(1). https://doi.org/10.1186/s40537-019-0252-x

[60]. Alharbi, A., Kalkatawi, M., & Taileb, M. (2021). Arabic sentiment analysis using deep learning and ensemble methods. *Arabian Journal for Science and Engineering*, 46(9), 8913–8923. https://doi.org/10.1007/s13369-021-05475-0

[61]. Yafooz, W. M. S., & Alsaeedi, A. (2021). Sentimental analysis on health-related information with improving model performance using machine learning. *Journal of Computer Science*. Retrieved from https://api.semanticscholar.org/CorpusID:234083781

[62]. Omar, A., Mahmoud, T. M., Abd-El-Hafeez, T., & Mahfouz, A. (2021). Multi-label Arabic text classification in online social networks. *Information Systems, 100*, 101785. https://doi.org/10.1016/j.is.2021.101785

[63]. Soliman, T., Elmasry, M., Hedar, A.-R., & Doss, M. (2013, May). Mining social networks' Arabic slang comments.

[64]. El-Naggar, N., El-Sonbaty, Y., & El-Nasr, M. A. (2017). Sentiment analysis of modern standard Arabic and Egyptian dialectal Arabic tweets. In *2017 Computing Conference* (pp. 880–887). https://doi.org/10.1109/SAI.2017.8252198

[65]. Elshakankery, K., & Ahmed, M. F. (2019). HILATSA: A hybrid incremental learning approach for Arabic tweets sentiment analysis. *Egyptian Informatics Journal, 20*(3), 163–171. https://doi.org/10.1016/j.eij.2019.03.002

[66]. Elmadany, A., Mubarak, H., & Magdy, W. (2018). ArSAS: An Arabic speech-act and sentiment corpus of tweets. Retrieved from https://api.semanticscholar.org/CorpusID:13693982

[67]. Al-Harbi, O. (2019). Classifying sentiment of dialectal Arabic reviews: A semi-supervised approach. *The International Arab Journal of Information Technology, 16*(6). Retrieved from http://dosweb.dos.gov.jo/ar/

[68]. Moudjari, L., & Akli-Astouati, K. (2020). An experimental study on sentiment classification of Algerian dialect texts. *Procedia Computer Science, 176*, 1151–1159. https://doi.org/10.1016/j.procs.2020.09.111

[69]. Guellil, I., Mendoza, M., & Azouaou, F. (2020). Arabic dialect sentiment analysis with zero effort: Case study on Algerian dialect. *Inteligencia Artificial, 23*(65), 124–135. https://doi.org/10.4114/intartif.vol23iss65pp124-135

[70]. Jerbi, H., Mohamed Amine, S., & Achour, E. (2019). Sentiment analysis of code-switched Tunisian dialect: Exploring RNN-based techniques. In K. Smaïli (Ed.), *Arabic Language Processing: From Theory to Practice* (pp. 122–131). Cham: Springer International Publishing.

[71]. Chader, A., Lanasri, D., Hamdad, L., Belkheir, M. C. E., & Hennoune, W. (2019). Sentiment analysis for Arabizi: Application to Algerian dialect. In *International Conference on Knowledge Discovery and Information Retrieval*. Retrieved from https://api.semanticscholar.org/CorpusID:204754582

[72]. Masmoudi, A., Hamdi, J., & Belguith, L. H. (2021). Deep learning for sentiment analysis of Tunisian dialect. *Computación y Sistemas, 25*(1), 129–148. https://doi.org/10.13053/CYS-25-1-3472

[73]. Mulki, H., Haddad, H., Ali, C. B., & Babaoglu, I. (2018). Tunisian dialect sentiment analysis: A natural language processing-based approach. *Computación y Sistemas, 22*. Retrieved from https://api.semanticscholar.org/CorpusID:59286097

[74]. Shoukry, A., & Rafea, A. (2012). Preprocessing Egyptian dialect tweets for sentiment mining. In *Conference of the Association for Machine Translation in the Americas*. Retrieved from https://api.semanticscholar.org/CorpusID:32734599

[75]. Al Wazrah, A., & Alhumoud, S. O. (2021). Sentiment analysis using stacked gated recurrent unit for Arabic tweets. *IEEE Access, 9*, 137176–137187. Retrieved from https://api.semanticscholar.org/CorpusID:238752721

[76]. Zribi, I., Boujelbane, R., Masmoudi, A., Ellouze, M., Belguith, L. H., & Habash, N. (2014). A conventional orthography for Tunisian Arabic. In *International Conference on Language Resources and Evaluation*. Retrieved from https://api.semanticscholar.org/CorpusID:9517956

[77]. Al-Harbi, O. (2017). Using objective words in the reviews to improve the colloquial Arabic sentiment analysis. *International Journal on Natural Language Computing, 6*(3), 1–14. https://doi.org/10.5121/ijnlc.2017.6301

[78]. Nahar, K. M. O., Jaradat, A., Atoum, M. S., & Ibrahim, F. (2020). Sentiment analysis and classification of Arab Jordanian Facebook comments for Jordanian telecom companies using lexicon-based approach and machine learning. *Jordanian Journal of Computers and*

Information Technology. Retrieved from https://api.semanticscholar.org/CorpusID:226432896

[79]. Duwairi, R. M. (2015). Sentiment analysis for dialectical Arabic. In *2015 6th International Conference on Information and Communication Systems (ICICS)* (pp. 166–170). https://doi.org/10.1109/IACS.2015.7103221

[80]. Rahab, H., Zitouni, A., & Djoudi, M. (2017). SIAAC: Sentiment polarity identification on Arabic Algerian newspaper comments. Retrieved from https://api.semanticscholar.org/CorpusID:194528725

[81]. Sabbar, B. M., Yousir, N. T., & Habeeb, L. A. (2018). Sentiment analysis for Iraqis dialect in social media using machine learning algorithms. *International Journal of Information and Communication Technology*.

[82]. Mohammed, N. T., Mohammed, E. A., & Hussein, H. H. (2023). Evaluating various classifiers for Iraqi dialectic sentiment analysis. In *Lecture Notes in Networks and Systems* (pp. 71–78). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-981-19-1412-6_6

[83]. Al-Jawad, M. M. H., Alharbi, H., Almukhtar, A. F., & Alnawas, A. A. (2022). Constructing Twitter corpus of Iraqi Arabic dialect (CIAD) for sentiment analysis. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 22*(2), 308–316. https://doi.org/10.17586/2226-1494-2022-22-2-308-316

[84]. Zaki, N. D., Hashim, N. Y., Mohialden, Y. M., Mohammed, M. A., Sutikno, T., & Ali, A. H. (2020). A real-time big data sentiment analysis for Iraqi tweets using Spark streaming. *Bulletin of Electrical Engineering and Informatics, 9*(4), 1411–1419. https://doi.org/10.11591/eei.v9i4.1897

[85]. Al-Bakri, N. F., Yonan, J. F., Sadiq, A. T., & Abid, A. S. (2022). Tourism companies assessment via social media using sentiment analysis. *Baghdad Science Journal, 19*(2), 422–429. https://doi.org/10.21123/BSJ.2022.19.2.0422

[86]. Nasrullah, H. A., Nasrullah, M. A., & Flayyih, W. N. (2023). Sentiment analysis in Arabic language using machine learning: Iraqi dialect case study. In *8th Engineering and 2nd International Conference for College of Engineering – University of Baghdad: COEC8-2021 Proceedings*. Retrieved from https://api.semanticscholar.org/CorpusID:257786424

[87]. Nassif, A. B., Darya, A. M., & Elnagar, A. (2021). Empirical evaluation of shallow and deep learning classifiers for Arabic sentiment analysis. *Transactions on Asian and Low-Resource Language Information Processing, 21*, 1–25. Retrieved from https://api.semanticscholar.org/CorpusID:244744807

[88]. Alnawas, A., & Arici, N. (2019). Sentiment analysis of Iraqi Arabic dialect on Facebook based on distributed representations of documents. *ACM Transactions on Asian and Low-Resource Language Information Processing, 18*(3). https://doi.org/10.1145/3278605

[89]. Awlla, K., & Veisi, H. (2022). Central Kurdish sentiment analysis using deep learning. *Journal of University of Anbar for Pure Science, 16*(2), 119–130. https://doi.org/10.37652/juaps.2022.176501

[90]. Assiri, A., Emam, A., & Al-Dossari, H. (2018). Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis. *Journal of Information Science, 44*(2), 184–202. https://doi.org/10.1177/0165551516688143

[91]. Al-Thubaity, A., Alqahtani, Q., & Aljandal, A. (2018). Sentiment lexicon for sentiment analysis of Saudi dialect tweets. In *Procedia Computer Science* (pp. 301–307). Elsevier B.V. https://doi.org/10.1016/j.procs.2018.10.494

[92]. Guellil, I., Azouaou, F., Benali, F., & Hachani, A.-E. (2021). Toward ONE model, ONE algorithm, ONE corpus dedicated to sentiment analysis of Arabic/Arabizi and its dialects. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 236–249). Association for Computational Linguistics. Retrieved from https://en.wikipedia.org/wiki/Arabic