applied sciences

MDPI

Article

User Opinion Prediction for Arabic Hotel Reviews Using Lexicons and Artificial Intelligence Techniques

Rihab Fahd Al-Mutawa * and Arwa Yousef Al-Aama

Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; aalaama@kau.edu.sa
* Correspondence: rmohamedalmutawa@stu.kau.edu.sa

Abstract: Opinion mining refers to the process that helps to identify and to classify users' emotions and opinions from any source, such as an online review. Thus, opinion mining provides organizations with an insight into their reputation based on previous customers' opinions regarding their services or products. Automating opinion mining in different languages is still an important topic of interest for scientists, including those using the Arabic language, especially since potential customers mostly do not rate their opinion explicitly. This study proposes an ensemble-based deep learning approach using fastText embeddings and the proposed Arabic emoji and emoticon opinion lexicon to predict user opinion. For testing purposes, the study uses the publicly available Arabic HARD dataset, which includes hotel reviews associated with ratings, starting from one to five. Then, by employing multiple Arabic resources, it experiments with different generated features from the HARD dataset by combining shallow learning with the proposed approach. To the best of our knowledge, this study is the first to create a lexicon that considers emojis and emoticons for its user opinion prediction. Therefore, it is mainly a helpful contribution to the literature related to opinion mining and emojis and emoticons lexicons. Compared to other studies found in the literature related to the five-star rating prediction using the HARD dataset, the accuracy of the prediction using the proposed approach reached an increase of 3.21% using the balanced HARD dataset and an increase of 2.17% using the unbalanced HARD dataset. The proposed work can support a new direction for automating the unrated Arabic opinions in social media, based on five rating levels, to provide potential stakeholders with a precise idea about a service or product quality, instead of spending much time reading other opinions to learn that information.

Keywords: opinion mining; sentiment analysis; data mining; automation; artificial intelligence; rating prediction; machine learning; Arabic reviews; emoji and emoticon; Arabic lexicon

Citation: Al-Mutawa, R.F.; Al-Aama, A.Y. User Opinion Prediction for Arabic Hotel Reviews Using Lexicons and Artificial Intelligence Techniques. *Appl. Sci.* **2023**, *13*, 5985. https://doi.org/10.3390/app13105985

Academic Editor: Luigi Portinale

Received: 10 April 2023
Revised: 28 April 2023
Accepted: 8 May 2023
Published: 12 May 2023

1. Introduction

Opinion mining, also known as sentiment analysis, is a field of data mining concerned with analyzing people's opinions about entities, such as services or products [1]. It is an important topic of interest and an active field of study in all languages, especially for languages that lack sufficient scientific research and resources, such as datasets, lexicons, corpora, and tools [2]. One of these languages is the Arabic language, where the progress in Arabic opinion mining does not fit with the substantial Arab world population, contributing to generating enormous amounts of Arabic data on the web [3–5]. The other challenges of the Arabic language, including being inherently complex and using dialects of Arabic, are also slowing research advancement [6]. Taking advantage of the Arabic datasets that support the standard and dialectical Arabic and creating additional analytical methods for them can add considerable contributions to knowledge creation in various fields, including tourism and hospitality, which greatly enhance business value. This research utilized customer reviews of hotels, some of the most

effective online-generated content for users, to analyze expressed guest experience and to predict guest satisfaction for the stayed-in hotel. In the tourism and hospitality industries, there are two main concerning factors: hotel guest experience and guest satisfaction, as increasing them is crucial to customer loyalty, distributing positive word-of-mouth, and repeat visits, which are all critical factors in the hospitality industry [7]. Therefore, a hotel's reputation highly affects reservations, services, and product sales. The tourism and hospitality industries involve many service sectors that significantly boost the economies of many countries through destination spots and entertainment places and many services, including lodging services, such as hotels, guest houses, guest rooms, resorts, and motels. Thus, tourism and hospitality industries significantly increase gross domestic product growth and raise countries' economies [8]. Additionally, destinations and countries benefit significantly from it in terms of their innovation and social development [9].

Automating opinion mining can help to predict and to understand customers' ratings for a particular product or service after interpreting customer satisfaction levels based on their comments about it on social media. Thus, it saves time for both customers and stakeholders who do not want to read many comments to understand and obtain an insight into a specific product or service [10]. Tertiary sectors of cities and tourism can be among stakeholders of automated hotel rating predictors, as they can provide insight into customer satisfaction levels. Then, policymakers can improve hotel policies or upgrade hotel marketing strategies to contribute to promoting the tourism industry, thus helping a country maintain its economic sustainability. Automating opinion mining is particularly important when using social media platforms, such as Twitter. Customers or potential ones do not generally state their opinion explicitly through a rating on Twitter; instead, they talk about their experience with the object; this is due to the different nature of social media platforms when compared to review or trading applications and websites, where there is no particular mandatory place for rating. Social media is an attractive platform for spreading opinions about anything, rather than traditional methods, such as filling out a survey. Social media provide users with ease of use anytime and anywhere, thereby obtaining attention from others and expressing freedom of opinion. However, the spread of opinions can affect potential clients' decisions about specific products or services. Both parties of customers and organizations that provide a product or service can benefit from the predicted satisfaction level. Customers can find more appropriate products or services, and their providers can understand their needs and opinions and improve their outcomes. Moreover, classifying user opinions can be used to utilize recent user-generated data in developing large-scale datasets [10]. Alternatively, they can be used to recommend alternative or similar products or services for future customers based on the opinion analysis of their posts.

This paper predicts user opinion based on user ratings for Arabic hotel reviews using Arabic resources, different produced features, and artificial intelligence techniques. Furthermore, it seeks to bridge a gap in the opinion mining field, which is the lack of opinion lexicons consisting of both emojis and emoticons and assigning their score based on levels of ratings that comply with the data, since no lexicon includes a comprehensive list of emojis and emoticon symbols that suit any Arabic dataset consisting of those mixed symbols. Thus, this study generated a lexicon for its dataset to handle its emojis, in addition to both Eastern and Western emoticons. These were coupled with a new consideration for the five levels of satisfaction of users through data. This method was employed in calculating the emotion score. The list of emojis and emoticons can be increased to be more comprehensive with additional data in the future.

This paper provides the following main contributions:

- It proposes an Arabic emojis and emoticons opinion lexicon (ArEmo lexicon) for Arabic opinion mining application tasks. It contains emoticons and emojis with additional descriptions. Each emotion score was calculated based on five levels of ratings using the HARD dataset instead of the three levels in other studies.

- It suggests an emoticon disambiguation algorithm by applying regular expression and recursion techniques to prepare a text for calculating the total weight of emoticons and replacing emoticons with their meanings to reflect their emotional contents.
- It proposes a user opinion classification approach for five-star ratings of online Arabic hotel reviews using supervised learning, which combines shallow and deep learning methods and employs Arabic resources to obtain valuable features, in addition to the fastText word embeddings.

The remainder of this paper is structured as follows. Section 2 presents the relevant literature on Arabic users' opinion prediction and potentially valuable resources for this study. Section 3 details the proposed work, including feature production and emoticon disambiguation-related algorithms, the creation of an Arabic emoji and emoticon opinion lexicon, and the proposed modeling of users' opinion prediction. It is followed by showing and discussing results in Section 4. Finally, both the conclusion and suggested future work are presented in Section 5.

2. Related Work

Opinion mining is a study of extracting people's attitudes and emotions about a given object, such as a service or product, by employing computational techniques [11], [12]. Opinion mining for languages other than English is still in its infancy [13]. It is still challenging to analyze opinions written in the Arabic language because of the many challenges the Arabic language poses and the insufficient resources and tools [14]. Abo et al. [15] discussed the most common challenges that face Arabic opinion mining. These challenges include a lack of datasets, corpora, and lexicons. This section covers studies concerned with the used dataset and different opinion mining methods, available lexicons, and emojis and emoticons.

2.1. Arabic Dataset and Opinion Mining Methods

Ghallab et al. [16] provided information on studies that contribute to overcoming the lack of Arabic datasets related to opinion mining. Arabic opinion mining datasets extracted from the Twitter platform are the majority, followed by those extracted from reviews on websites and through applications. Accordingly, most Arabic opinion mining studies use available social media data with no more than three labels: positive, negative, and neutral, unlike the review data [14]. One study that overcame the lack of Arabic datasets is Elnagar et al.'s [17]. The authors in [17] presented the Hotel Arabic Reviews Dataset (HARD), the Arabic dataset's most extensive public hotel reviews for machine learning applications and subjective opinion mining. It consists of 409,562 unbalanced reviews about hotels collected from the Booking.com website. The balanced subset of the dataset is over a hundred thousand reviews. Each review has a rating on a scale of one to five stars. The review text is in modern standard and colloquial Arabic and includes emojis and emoticons. Emoji is the typical expression of emotion and concepts through digital Unicode graphic icons. This form of representation is the latest generation of emoticons that are most popular in social media through mobile applications [18].

The authors in [17] have published baseline results for opinion mining using the HARD dataset. They used lexicon-based methods and different machine learning methods. They achieved an accuracy of 76.1% as the highest testing result for rating classification using the logistic regression (LR) algorithm with term frequency-inverse document frequency (TF-IDF). This combination was used unigram and bigram on the unbalanced HARD dataset. TF-IDF is a mathematical measure that explicates a word's importance in a document in a series of documents [19]. The authors removed the neutral reviews with a rating of three from the balanced HARD dataset before the experiment. Thus, the balanced HARD dataset has four rating levels, while the unbalanced version has five. In [20], the authors developed different deep and machine learning models for

opinion mining in the reviews' domain. They used a subset of the HARD dataset, comprising 62,500 random reviews, to predict the five rating levels. The highest achieved accuracy for five labels of rating classification on the HARD dataset is 74.2%, while precision, recall, and f1-score results for each are 74.0% when using the random forest (RF) model. The following best models are based on convolutional neural networks (CNNs), decision trees (DTs), gated recurrent units (GRUs), and bi-directional recurrent neural networks (BiLSTMs). They achieved the following accuracy results, respectively: 67.4%, 66.4%, 65.1%, and 63.1%.

Different studies created baselines of opinion and emotion detection in other languages. For instance, Bashir et al. [21] proposed a deep neural network-based emotion detection model from the textual data for the Urdu language to classify emotions of their presented UNED corpus. Fei et al. [22] proposed a latent emotion memory network model that learns the latent emotion distribution in the data without external knowledge. The model receives an emotional bag-of-words as input, removes only stop-word tokens, and keeps the words presented in English or Chinese lexicons.

Aspect-based sentiment analysis is an active area of research, especially concerning review data. However, observations have shown that aspect-based sentiment analysis models trained on a dataset from one domain do not generalize well to another dataset belonging to another domain [23]. In [24], the authors enhanced the aspect-based sentiment analysis robustness by improving the model, data, and training. They tested data from different domains. Their methods can be applied to other artificial intelligence-based tasks. Other effective opinion mining methods include end-to-end neural-based methods, such as those in [25–28], as well as graph-based methods, such as those in [29–32].

2.2. Arabic Text-Based Lexicons

Two Arabic studies introduced large-scale public lexicons [33,34] to contribute more resources to Arabic opinion mining. Al-Twairesh et al. [33] presented a vast lexicon of Arabic tweets annotated for sentiment analysis, known as the AraSenTi lexicon. A sentiment lexicon is a dictionary of words or phrases reflecting positive or negative emotions. One of the popular lexicon usages is to calculate a text's sentiment score [35]. A lexicon is a significant resource for opinion mining to classify the words extracted from datasets as positive, negative, or neutral. There are two ways of constructing lexicons: automatically or manually. Usually, manual lexicons are more accurate than automatically constructed lexicons, but they have smaller sizes [22]. The AraSenTi lexicon reached 131,342 words, each associated with its sentiment score. It contains words of the Modern Standard Arabic and the Saudi dialect. The AraSenTi lexicon is suitable for various research directions, including the review of services and products, as a result of the lexicon's high coverage. There have been different usages for the AraSenTi lexicon in the Arabic literature. For instance, in [36], the authors collected, from the Twitter platform, an Arabic dataset related to COVID-19 to create a model for opinion mining that classifies Arabic tweets regarding the COVID-19 crisis. They labeled the tweets as positive or negative by using the AraSenTi lexicon. In [22], to support the study of Arabic domain-dependent opinion mining, they presented affect lexicons, where each word has a classification, either as positive or negative. The covered domains are social issues, technology, politics, and sports, where a word can be positive in one domain while negative in the other. The total number of positive and negative words in all four domains is 9461, less than those in the AraSenTi lexicon.

The lexicons for Arabic opinion mining in the literature are scarce, especially for public, large, and non-domain specific lexicons [16]. In [23], the authors built a fanatic lexicon with their methodology that automatically classifies Arabic texts in social media into fanatic and anti-fanatic emotions. The total number of unique phrases in the fanatic lexicon is 1766, distributed in twenty-one contexts. Sports fanaticism is a state of emotion that generates a blind hatred for the competitive teams paired with a blind love for the

favorite teams at the same time. Similarly, in [37], the authors produced twelve fanatic domain-specific lexicons that are helpful with Arabic social text for constructing a fanatic classification model to classify the texts into either fanatic or non-fanatic, building tools against fanatic attitude, or identifying and analyzing sports fanaticism.

2.3. Emojis and Emoticons Lexicons

In text-based opinion mining, emojis and emoticons can be valuable features. The emoji sentiment lexicon is a crucial element of the use of emojis in opinion mining [38]. In [39], the authors proposed an Arabic emoji sentiment lexicon (ArabESL) to conduct a comparative analysis for consistency of context-based emoji usage across Arabic and European cultures and languages. Despite their assumed consistent usage worldwide, there are indications that the meaning of an emoji may change across different cultures and languages. The Arab-ESL lexicon contains 1034 emojis extracted from public Arabic textual data from the Twitter platform, where these data consist of different dialects, in addition to Modern Standard Arabic. Each emoji has a name, label, and sentiment score based on sentiment labels that take one of three values: negative, neutral, or positive. The results indicate that some cultural-specific aspects, such as nature, affect the sentiment indications of some emojis. Notably, their study focused on emojis without emoticons. Likewise, in [18], the authors considered only emojis while proposing the first and most prominent European emoji sentiment lexicon, the emoji sentiment ranking. The lexicon consists of 751 emojis with their names and their sentiment scores with values between −1 and +1, which have been calculated based on the sentiment of the data in which they occur. The data are tweets from thirteen European languages, each classified into three positive, neutral, or negative sentiments. A unique visualization in the lexicon shows all emoji sentiments as a sentiment bar. The authors found that most emojis are positive, particularly the most frequently used ones. Besides, in comparison based on emoji rankings, there were no significant differences in emoji rankings between the thirteen languages. Therefore, the proposed emoji sentiment ranking is assumed to be an independent resource for European languages that supports automated opinion mining.

From another perspective, in [38], the authors created an Arabic emoji sentiment lexicon that is context-free, called CF-Arab-ESL. It is challenging to create an emoji sentiment lexicon that is context-free, as individuals interpret emojis according to their perspective, which is greatly affected by cultural background. Thus, the sentiment conveyed by each one is very subjective. In the presented lexicon, a total of 1069 emojis were labeled as positive, neutral, and negative by thirty-five native Arabians of different regions to discover how the sentiment of these emojis can be annotated without depending on a text-based context. The annotators agreed that only a subset of emojis represented particular sentiment. For example, there was an agreement on the positivity of all Arabian countries' flags as a sense of attachment to a nation, as well as an agreement that some animals' emojis, such as a horse, lion, and eagle, are positive, as they indicate positivity in Arabian culture, unlike lizards, pigs, and snakes.

3. Proposed Work

This section describes the proposed work for automating and predicting users' opinions of Arabic hotel reviews. As far as this research is aware, this is the first study that created a lexicon that incorporates emojis and emoticons to predict user opinion. The following subsections cover dataset preparation and preprocessing, the emoticon disambiguation method, Arabic emojis, and emoticon lexicon generation for opinion mining, feature production, and modeling.

3.1. Dataset Preparation

This study used the HARD dataset [17], the most significant standard and dialectical Arabic dataset containing emojis and emoticons. It was divided, similar to the authors'

approach, into 80% for training and 20% for testing. The study conducted experiments on two variations of the dataset. The unbalanced dataset consists of 409,562 reviews, and the balanced dataset consists of 62,500 reviews, where each rating has 12,500 reviews. The reviews' data needed preprocessing to extract the following features in order to be used with the machine learning models: the total weight of positive words, the total weight of negative words, the emotions score, and the number of the following—positive words, negative words, negation words, booster words, repeated characters, question marks, exclamation marks, periods, and commas. There are two preprocessing methods; the method selected depends on the required generated feature. Both applied preprocessing methods involve emoji removal, Arabic diacritics removal, URL removal, special characters removal, and tokenization, and they differ only in applying punctuation removal. An exception is calculating the emotion scores, which did not require any preprocessing, as they only need to detect emojis and emoticons in the text for the computation.

An example of applying data preprocessing steps with the case of removing punctuations from an Arabic review text is shown in Table 1; the other case of not removing punctuations is excluding the step that precedes the last one. The preprocessing steps involved using regular expressions, emojis, and NLTK libraries.

Table 1. An example of applying data preprocessing steps by removing punctuations in an Arabic text.

Preprocessing Step	Arabic Text
Original Arabic Text (Before Preprocessing)	مخيب للأمل. لولا موقع الفندق ولا كان الفندق ما يسوى ولا ريال. المويه انقطعت كذا مره + الكهرباء برضو انقطع تقريباً نص ساعة الفطور سيء جداً👎 Translation: "Disappointing. Had it not been for the hotel's location, the hotel would not have been worth anything. The water was cut off many times + the electricity also went out for about half an hour 👎 breakfast is very bad"
Emoji Removal	مخيب للأمل. لولا موقع الفندق ولا كان الفندق ما يسوى ولا ريال. المويه انقطعت كذا مره + الكهرباء برضو انقطع تقريباً نص ساعة الفطور سيء جداً Translation: "Disappointing. Had it not been for the hotel's location, the hotel would not have been worth anything. The water was cut off many times + the electricity also went out for about half an hour breakfast is very bad"
Arabic Diacritics Removal	مخيب للأمل. لولا موقع الفندق ولا كان الفندق ما يسوى ولا ريال. المويه انقطعت كذا مره + الكهرباء برضو انقطع تقريبا نص ساعة الفطور سيء جدا Translation: "Disappointing. Had it not been for the hotel's location, the hotel would not have been worth anything. The water was cut off many times + the electricity also went out for about half an hour breakfast is very bad"
URL Removal	مخيب للأمل. لولا موقع الفندق ولا كان الفندق ما يسوى ولا ريال. المويه انقطعت كذا مره + الكهرباء برضو انقطع تقريبا نص ساعة الفطور سيء جدا Translation: "Disappointing. Had it not been for the hotel's location, the hotel would not have been worth anything. The water was cut off many times + the electricity also went out for about half an hour breakfast is very bad"
Special Characters Removal	مخيب للأمل. لولا موقع الفندق ولا كان الفندق ما يسوى ولا ريال. المويه انقطعت كذا مره الكهرباء برضو انقطع تقريبا نص ساعة الفطور سيء جدا Translation: "Disappointing. Had it not been for the hotel's location, the hotel would not have been worth anything. The water was cut off many times the electricity also went out for about half an hour breakfast is very bad"
Punctuation Removal	مخيب للأمل لولا موقع الفندق ولا كان الفندق ما يسوى ولا ريال المويه انقطعت كذا مره الكهرباء برضو انقطع تقريبا نص ساعة الفطور سيء جدا Translation: "Disappointing Had it not been for the hotels location the hotel would not have been worth anything The water was cut off many times the electricity also went out for about half an hour breakfast is very bad"

Tokenization	['مخيب', 'للأمل', 'لولا', 'موقع', 'الفندق', 'ولا', 'كان', 'الفندق', 'ما', 'يسوى', 'ولا', 'ريال', 'الميه', 'انقطعت', 'كذا', 'مره', 'الكهرباء', 'برضو', 'انقطع', 'تقريبا', 'نص', 'ساعة', 'الفطور', 'سيء', 'جدا'] Translation: ['Disappointing', 'Had', 'it', 'not', 'been', 'for', 'the', 'hotels', 'location', 'the', 'hotel', 'would', 'not', 'have', 'been', 'worth', 'anything', 'The', 'water', 'was', 'cut', 'off', 'many', 'times', 'the', 'electricity', 'also', 'went', 'out', 'for', 'about', 'half', 'an', 'hour', 'breakfast', 'is', 'very', 'bad']

There is another variation of the HARD dataset made for the unbalanced and balanced data. The indicated meanings of emojis and emoticons are related to the experiment using unbalanced and balanced data via the deep learning-based method, which includes showing the results before and after replacing emojis and emoticons with their meaning, as presented in Figure 1.

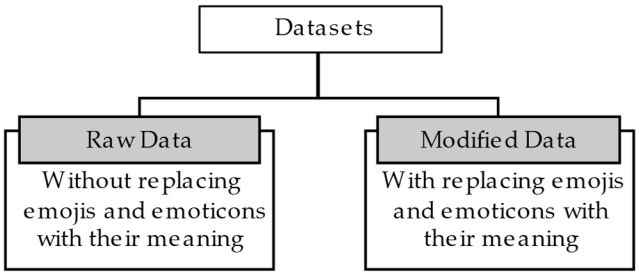

Figure 1. Two variations of input Arabic review data.

Figure 2 shows an overview of the review's replacement process. The upcoming subsections will cover the details of the algorithms and the ArEmo lexicon. It is worth mentioning that the emoticon disambiguation algorithm is required while preparing the review data and when calculating the emotion scores for the ArEmo lexicon.

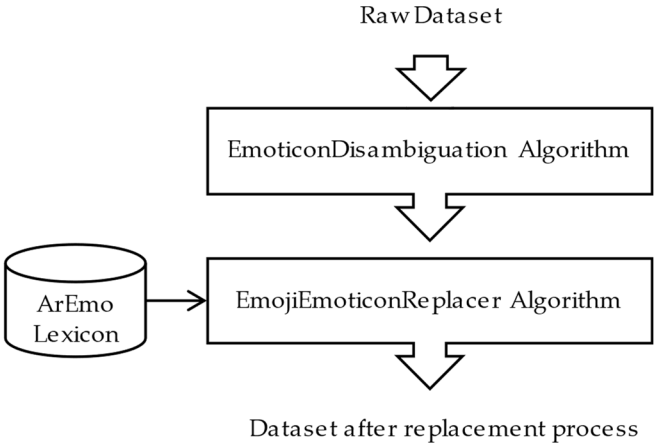

Figure 2. Extraction process for the modified input of Arabic review data.

Algorithm 1 represents the algorithm for replacing emojis and emoticons in the reviews. This algorithm considers adding a space next to the word when it does the replacement, as some emojis or emoticons are attached to the text.

Algorithm 1: Emoji and Emoticon Replacer Algorithm

Input: a review without uniform resource locators and enclosed brackets that do not belong to emoticons and the ArEmo Lexicon.

Output: an updated review after replacing its emojis and emoticons with their estimated equivalent meaning.

```
START
1.  def ReplaceEmojiEmoticonWithMeaning(Review, ArEmoLexicon):
2.  {
3.      ReviewWords = Review.split()
4.      for each key, value in ArEmoLexicon do
5.        for each word in ReviewWords do
6.          if key in word then
7.            UpdatedReview = Review.replace(key, f" {value} ")
8.      UpdatedReview = re.sub(r" +", " ", UpdatedReview).strip()
9.      return UpdatedReview
10. }
END
```

3.2. Emoticon Disambiguation

Preparing the review data for emotion score calculation and for replacing the emoticon in the reviews with their corresponding meaning, in addition to detecting the emoticon in the reviews during the computation process of emoticon score for ArEmo lexicon, the study suggests an emoticon disambiguation algorithm, presented in Algorithm 2, using regular expression and recursion techniques. This algorithm removes hyperlinks from reviews to avoid considering the following part between parentheses of the hyperlink: (:\) as an emoticon.

Algorithm 2: Emoticon Disambiguation Algorithm

Input: a review

Output: an updated review without uniform resource locators and enclosed brackets that do not belong to emoticons.

```
START
1.  URLRegEx='(?i)(http[17]?:\/\/\S+)'
2.  def DisambiguateEmoticon(Review):
3.  {
4.      UpdatedReview = re.sub(URLRegEx, '', Review)
5.      UpdatedReview = RemoveEnclosedBrackets(UpdatedReview)
6.      return UpdatedReview
7.  }
END
```

The emoticon disambiguation algorithm also calls the enclosed brackets removal algorithm in Algorithm 3 to remove all nested and non-nested enclosed brackets while keeping the brackets that precede or follow an emoticon, as part of it, such as (^_^) to resolve considering a colon next to a bracket whether from left or right as an emoticon, such as :(. Thus, it keeps only brackets that belong to emoticons, such as :), :], or :(. The enclosed bracket removal algorithm deals with four types of brackets, which are round brackets: (), angle brackets: <> or ⟨⟩, square brackets: [], and curly brackets: {}. They are also known as parentheses, chevrons, brackets, and braces. The case of equality at line 24 occurs with reversed brackets or with an emoticon surrounded by its brackets to stop the recursive function.

Algorithm 3: Enclosed Brackets Removal Algorithm

Input: a review.

Output: an updated review after recursively removing all enclosed and nested brackets while preventing removal of the emoticons surrounded by two brackets.

START

```
1.  BracketsRegEx = '\[(?! [\_\-\ ‿\^\ ˆ \$\ﷺ\*\نs])((.){0,}?)(?<![\_\-\ ‿\^\ ˆ \$\ﷺ\*\نs])\]'
2.  ParanthesesRegEx = '\((?![\_\-\ ‿\^\ ˆ \$\ﷺ\*\نs])((.){0,}?)(?<![\_\- \ ‿\^\ ˆ \$\ﷺ\*\نs])\)'
3.  SmallChevronsRegEx = '\<(?![\_\-\ ‿\^\ ˆ \$\ﷺ\*\نs])((.){0,}?)(?<![\_\- \ ‿\^\ ˆ \$\ﷺ\*\نs])\>'
4.  LargeChevronsRegEx = '\〈(?![\_\-\ ‿\^\ ˆ \$\ﷺ\*\نs])((.){0,}?)(?<![\_\- \ ‿\^\ ˆ \$\ﷺ\*\نs])\〉'
5.  BracesRegEx = '\{(?![\_\-\ ‿\^\ ˆ \$\ﷺ\*\نs])((.){0,}?)(?<![\_\- \ ‿\^\ ˆ \$\ﷺ\*\نs])\}'
6.  def RemoveEnclosedBrackets(Review):
7.  {
8.      if (Review.find('[') != −1 and Review.find(']') != −1) or (Review.find('(') != −1 and Review.find(')') != −1) or
(Review.find('<') != −1 and Review.find('>') != −1) or (Review.find('〈') != −1 and Review.find('〉') != −1) or
(Review.find('{') != −1 and Review.find('}') != −1) then
9.          if (Review.find('[') != −1 and Review.find(']') != −1) then
10.             Temp = Review
11.             Review = re.sub(BracketsRegEx, r'\1', Review)
12.         if (Review.find('(') != −1 and Review.find(')') != −1) then
13.             Temp = Review
14.             Review = re.sub(ParanthesesRegEx, r'\1', Review)
15.         if (review.find('<') != −1 and Review.find('>') != −1) then
16.             Temp = Review
17.             Review = re.sub(SmallChevronsRegEx, r'\1', Review)
18.         if (Review.find('〈') != −1 and Review.find('〉') != −1) then
19.             Temp = review
20.             Review = re.sub(LargeChevronsRegEx, r'\1', Review)
21.         if (Review.find('{') != −1 and Review.find('}') != −1) then
22.             Temp = Review
23.             Review = re.sub(BracesRegEx, r'\1', Review)
24.         if len(Review) == len(Temp) then
25.             return Review
26.         else return RemoveEnclosedBrackets(Review)
27.     else return Review
28. }
29.  //end of RemoveEnclosedBrackets
```

END

Table 2 shows some examples of Arabic reviews with ambiguous emoticons that should not be considered for the replacement with the equivalent meaning or for counting their emotion score.

Table 2. Examples of ambiguous emoticons in Arabic reviews.

Arabic Review	Translated Review into English	Ambiguous Emoticon
" مميز ?". - عندهم أفضل مركز صحي للمساج جربته بحياتي للآن .- أعجبني كذلك عبارتهم المميزة :(أشياء بسيطة تصنع الفارق)- يستحق التمييز ?	"Special ?". - They have the best health center for massages that I have tried in my life so far. - I also liked their distinctive phrase :(Simple things make a difference) - Worth the distinction?.	:(or :(

استثنائي. ممتاز جدا ويستحق الزيارة مره اخرى. عدم توفر بعض الخدمات الاخرى حول الفندق مثل : (المطاعم و المحلات ..)	Exceptional. Very excellent and worth visiting again. Unavailability of some other services around the hotel such as : (restaurants, shops...)	) : or : (Note: there is a space between the bracket and colon.
لا انصح فيه". اعجبني تواجد " الصراصير منظر غير لائق جدا ولا نصح فيه. النظافة معدومةصراصير الصورة هنا http://cdn.top4top.co/i_ba52fbf59d0.jpg	"I do not recommend it." I liked the presence of cockroaches a very inappropriate view, and I don't recommend it. Hygiene is lacking. Cockroaches photo here http://cdn.top4top.co/i_ba52fbf59d0.jpg	:/

3.3. Arabic Emoji and Emoticon Opinion Lexicon Creation

This section explains the method of generating the proposed Arabic emoji and emoticon lexicon for opinion mining (ArEmo lexicon) that uniquely considers both emojis and emoticons and some new emoticons not included in other lexicons. The ArEmo lexicon has 301 emojis and emoticons. Moreover, it computes the emotion score for each emoji or emoticon, concerning five classes of reviews' ratings, instead of the common consideration of three classes in the literature. The proposed ArEmo lexicon has two versions, both made using the train part of the HARD dataset, but they differ in the type of dataset that is either unbalanced or balanced. Tables 3 and 4 show the top 10 most frequently used emojis and emoticons extracted from the train part of the unbalanced and balanced HARD dataset. The classification column labels the emoticon as Western or Eastern, also known as vertical and horizontal emoticons, respectively.

Table 3. A sample from the ArEmo lexicon using unbalanced data.

Emo	Type	Emoji Unicode	Classification	Short English Naming	Short Arabic Naming	Emotion Score	Total Count
👍	Emoji	\U0001f44d	-	thumbs up	إبهام مرفوع لأعلى	0.797428	933
❤	Emoji	\u2764	-	red heart	قلب أحمر	0.885193	466
:)	Emoticon	-	Western	smiling face	وجه مبتسم	0.646119	438
:(	Emoticon	-	Western	sad face	وجه حزين	0.259574	235
👎	Emoji	\U0001f44e	-	thumbs down	إبهام متجه لأسفل	−0.22886	201
☆	Emoji	\u2b50	-	star	نجمة	0.763889	180
😍	Emoji	\U0001f60d	-	smiling face with heart-eyes	وجه مبتسم بعيون قلب	0.868571	175
👌	Emoji	\U0001f44c	-	ok hand	يد حسنا	0.849112	169
😊	Emoji	\U0001f60a	-	smiling face with smiling eyes	وجه مبتسم بعيون مبتسمة	0.732283	127
🌹	Emoji	\U0001f339	-	rose	ورد	0.756	125

Table 4. A sample from the ArEmo lexicon using balanced data.

Emo	Type	Emoji Unicode	Classification	Short English Naming	Short Arabic Naming	Emotion Score	Total Count
👍	Emoji	\U0001f44d	-	thumbs up	إبهام مرفوع لأعلى	0.708861	79
:)	Emoticon	-	Western	smiling face	وجه مبتسم	0.416667	60
👎	Emoji	\U0001f44e	-	thumbs down	إبهام متجه لأسفل	−0.5	56
❤	Emoji	\u2764	-	red heart	قلب أحمر	0.829268	41
:(	Emoticon	-	Western	sad face	وجه حزين	−0.21622	37
👌	Emoji	\U0001f44c	-	ok hand	يد حسنا	0.692308	26

😂	Emoji	\U0001f602	-	tearful laughing face	وجه يضحك بدموع	−0.11765	17
❌	Emoji	\u274c	-	error	خطأ	−0.61765	17
☆	Emoji	\u2b50	-	star	نجمة	0.59375	16
😡	Emoji	\U0001f621	-	pouting face	وجه عابس	−0.40625	16

The following two subsections cover emojis and emoticons' collection and scoring methods.

3.3.1. Emoji and Emoticon Collection

Extracting the emojis from the train part of the HARD dataset mandated utilizing an emoji pattern using regular expression. For extracting emoticons from the dataset, the following needed to be removed from the reviews' data: numbers, Arabic letters, Arabic diacritics, special Arabic letters, and English letters. Then, after obtaining each review in only symbols, the manual selection occurred for the symbols that look similar to an emoticon. Appropriately descriptive naming is given to naming the collected emojis and emoticons when no such one was found from multiple sources in [40–50].

3.3.2. Emotion Score Method

For calculating the emotion score for the ArEmo lexicon, this paper adapted and extended the solution model proposed by [18] to include five opinion labels that are −1, −0.5, 0, 0.5, and 1, instead of three labels. In this study, −1, −0.5, 0, 0.5, and 1, respectively, refer to the ratings 1, 2, 3, 4, and 5 in the HARD dataset, where the user opinion about the review can be very negative, negative, neutral, positive, or very positive. Only in this subsection the opinion labels 1, 2, 3, 4, and 5 are changed to different labels—−1, −0.5, 0, 0.5, and 1, indicating their equivalent meaning in order to avoid multiplying the positive discrete probability distributions with a larger number with disregard to the sign, than negative discrete probability distributions later, while computing the emotion score.

A discrete, 5-valued variable represents the opinion class for emojis or emoticons within the rated review, c:

$$c \in \{-1, -0.5, 0, +0.5, +1\}$$

The representations for the discrete distributions that reflect the opinion distribution are defined by the following equation:

$$N(c),\ \sum_{c=-1}^{+1} N(c) = N, \tag{1}$$

N indicates the number of all the occurrences of the emoji or emoticon in the reviews, and N(c) is the occurrences of the emoji or emoticon in reviews with the particular opinion label, c. Consequently, the discrete probability distribution based on each probability outcome is between 0 and 1, and the summation of all the outcomes from the discrete probability distribution must be one by definition, which is:

$$\sum_{c=-1}^{+1} P(c) = 1, \tag{2}$$

$$P(-1) + P(-0.5) + P(0) + P(+0.5) + P(+1) = 1, \tag{3}$$

The progression from the first probability, P(−1), until the last one, P(+1), consequently indicates high negativity, negativity, neutrality, positivity, and high positivity of a particular emoji or emoticon. Estimating probabilities typically depends on relative frequencies. Therefore, based on the emoji or emoticon occurrences, the discrete probability distribution, P(c), is:

$$P(c) = \frac{N(c)}{N}, \text{ when } N > 6, \tag{4}$$

However, when the sample is less than 6, it is more effective to apply the rule of succession for calculating the probability:

$$P(c) = \frac{N(c) + 1}{N + k}, \text{ when } N < 6, \quad (5)$$

where k is a constant denoting the cardinality of the class, and, accordingly, for the HARD dataset, k = |c| = 5. The discrete probability distribution can be used in computing the mean. The lower and higher limits of sigma are the smallest and largest values of c:

$$\bar{X} = \sum_{c=-1}^{+1} (c \cdot P(c)), \quad (6)$$

The emotion score, $\bar{S}$, is calculated as the mean of the discrete probability distributions. In contrast, each probability is multiplied by its opinion class as follows, and it has a value ranging from −1 to 1:

$$\bar{S} = (-1 \cdot P(-1)) + (-0.5 \cdot P(-0.5)) + (0 \cdot P(0)) + (0.5 \cdot P(+0.5)) + (1 \cdot P(+1)), \quad (7)$$

$$\bar{S} = 0.5\ P(0.5) + P(1) - (P(-1) + 0.5\ P(-0.5)), \quad (8)$$

3.4. Feature Production

Conducting experiments using a machine learning model involved generating new features from the dataset for all the reviews' data through employing multiple resources: the AraSenTi lexicon [33], the ArEmo Lexicon, a booster words list, and a negation words list, as demonstrated in Figure 3. The features include the total weight of positive words, negative words, and emotions score, as well as the number of following: positive words, negative words, negation words, booster words, elongated characters, question marks, exclamation marks, periods, and commas.

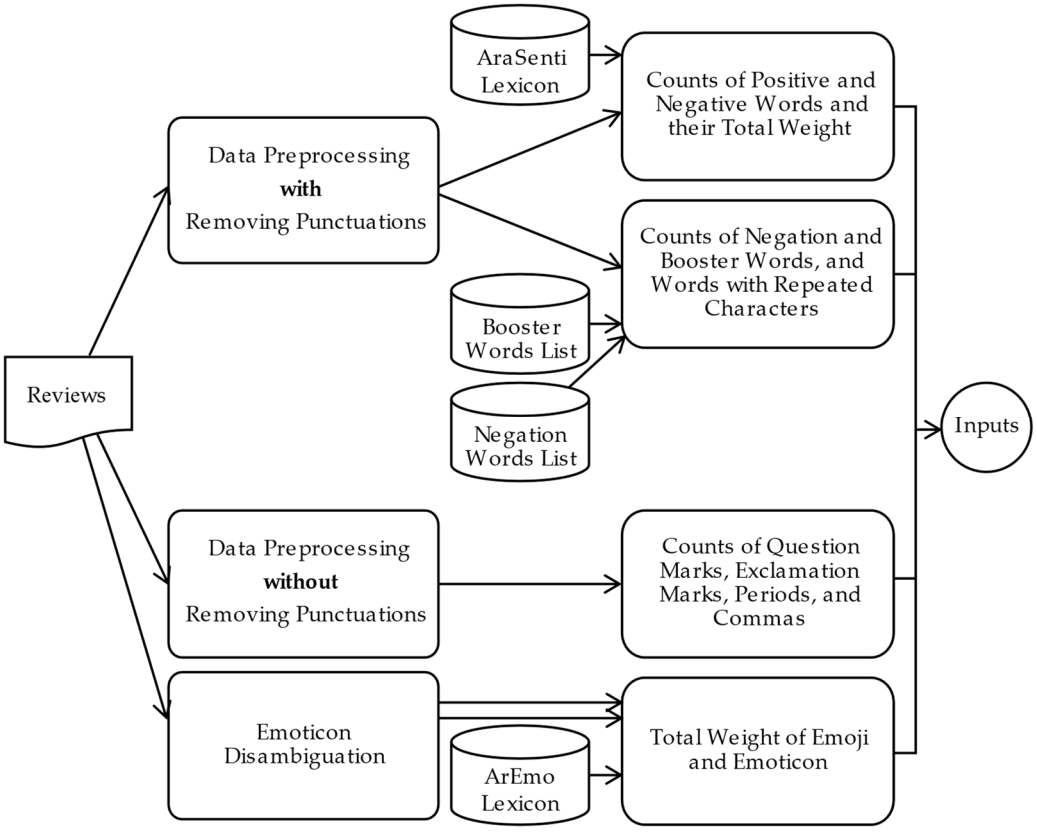

Figure 3. Produced features from the dataset for usage with a machine learning classifier.

After exploring different resources [51,52], the study formed booster and negation word lists. The booster and negation word lists consist of 33 words and 35 words, respectively. Tables 5 and 6 show a sample of 20 words from those lists. These tables classify each word, whether it is in Modern Standard Arabic or dialectal Arabic, according to their use, or whether it is used is a Modern Standard Arabic context or in a dialectal Arabic context.

Table 5. A sample Arabic booster word list.

Arabic Word	English Translation	Type of Arabic	Arabic Word	English Translation	Type of Arabic
جدا	Very	Standard	واجد	Very	Dialectal
جدن	Very	Dialectal	وايد	Very	Dialectal
قدا	Very	Dialectal	خالص	Very	Dialectal
قدن	Very	Dialectal	بزاف	A lot	Dialectal
مرة	Very	Dialectal	بالزاف	A lot	Dialectal
مره	Very	Dialectal	بقوة	Very	Dialectal
مرا	Very	Dialectal	بقوه	Very	Dialectal
كثير	A lot	Standard	مفرط	Excessive	Standard
كثيرا	A lot	Standard	بإفراط	Excessively	Standard
كتير	A lot	Dialectal	بافراط	Excessively	Dialectal

Table 6. A sample Arabic negation word list.

Arabic Word	English Translation	Type of Arabic	Arabic Word	English Translation	Type of Arabic
لا	No/Do not/Does not	Standard	لستم	You are not	Standard
لم	Did not	Standard	لسن	They are not	Standard
ما	No/Not	Standard	ليسوا	They are not	Standard
لن	Will not	Standard	مش	Not	Dialectal
لما	Not yet	Standard	مو	Not	Dialectal
إن	Not	Standard	مفي	There is no	Dialectal
لات	Not	Standard	مافي	There is no	Dialectal
غير	Not/Without	Standard	منو	Not	Dialectal
بدون	Without	Standard	مانو	Not	Dialectal
بلا	Without	Standard	ماكو	There is no	Dialectal

Different libraries were needed to extract the features using the Python programming language, including re for regular expression and collections for importing a counter.

3.5. Modeling

The study proposes two ensemble-based deep learning models utilizing a soft voting mechanism, one using unbalanced data, namely, EDLU, and the other using the balanced data, namely, EDLB. Figure 4 outlines the proposed ensemble-based deep learning model in both ways of using the unbalanced or balanced data.

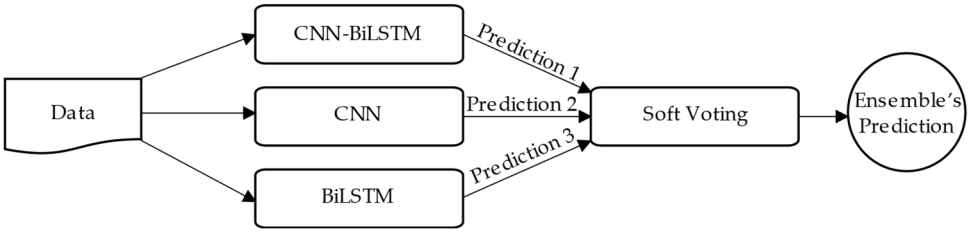

Figure 4. The suggested ensemble-based deep learning model.

Figure 5 represents the proposed final hybrid model's main steps in using the unbalanced or balanced data.

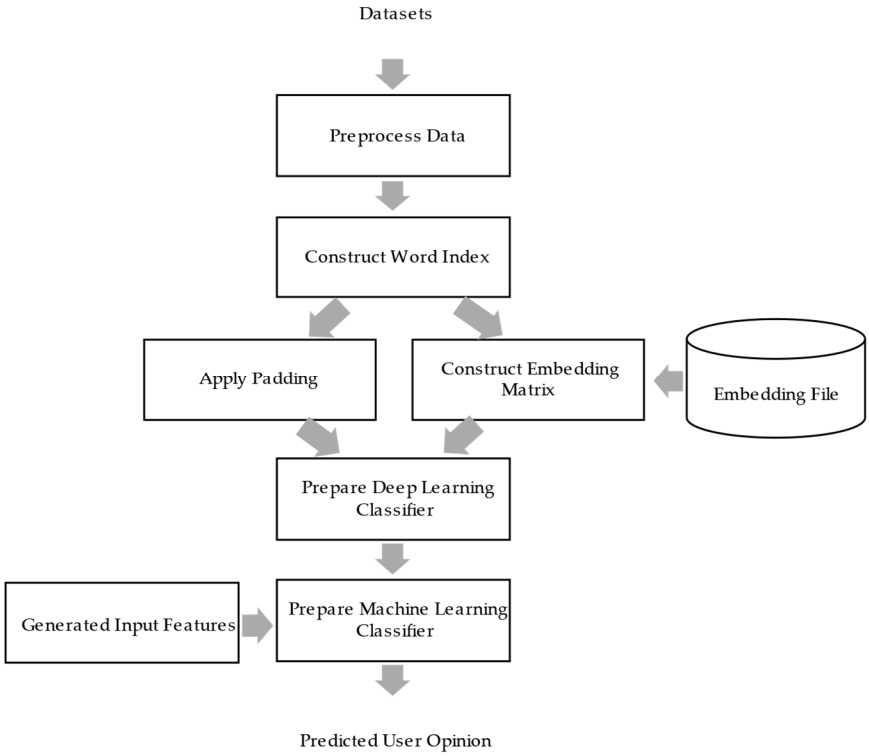

Figure 5. An overview of the steps involved in the proposed hybrid models.

The input datasets, consisting of train and test data, have experimented with the models in two variations: as raw data and as-modified after replacing each emoji and emoticon with their corresponding meaning from the proposed ArEmo lexicon, as already shown in Figure 1. A 'WE' is attached as a subscript to the models' names: EDLB$_{WE}$ and EDLU$_{WE}$ to indicate the case without emoji and emoticon replacements. The data preprocessing includes removing punctuation. After preprocessing and tokenizing the reviews, a word index is formed for the review data, followed by two processes: converting words in each review into sequences of integers that become padded to give these vectors the same length and creating the embedding matrix using the fastText pre-trained model of the Arabic language, which can represent rare words. Then, the base model is formed, depending on ensemble-based deep learning with the structure shown in Table A1 in Appendix A, consisting of three classifiers based on CNN-BiLSTM, CNN, and BiLSTM, respectively. The predicted user opinion from the proposed model is passed to the other proposed model using the MLP algorithm to be experimented with different generated features to obtain the final predicted user opinion, which labels the reviews with a five-star value—either one, two, three, four, or five.

4. Results and Discussion

As the selection for the emoticon was manual and not performed by using a previously prepared list of emoticons, it was found that some uncommon and unincluded emoticons occurred in this lexicon, but not in other lexicons. These include هـــنْ and (^__^). The found emoticons were Western and Eastern. Moreover, for the Western emoticon, the study found that Arabs do not always start writing the eyes from the left, but sometimes from the right; for example, the smiling face could be expressed as (: instead of ☺. The found studies related to those pictorial symbols in the literature did not cover these considerations. Therefore, this study recommends those considerations with

the newly generated lexicons for emojis and emoticons in the future to be more comprehensive.

When comparing Table 3 with Table 4 of the most frequently used emojis and emoticons, the study found that the most used emojis and emoticons can differ highly according to different variations of the same dataset for being balanced versus being unbalanced. Based on the emotion score, in the case of using the balanced dataset, it is interesting that there are equally five positive and five negative emojis and emoticons, while, in the other case, there are nine positive emojis and emoticons, and only one is negative. Similarly, comparing the top ten emojis and emoticons of the balanced version of the ArEmo lexicon in addition to its scores with the related emoji lexicons in literature in Table 7, it can be found that the frequently used emojis highly differ based on the differences in culture and used data and, accordingly, their scores differ, which leads to a different number of positive and negative emojis. Thus, this study used a custom-made context-based emoji and emoticon lexicon that depends on the used dataset instead of the available open-source lexicons.

Table 7. Comparison table of the proposed ArEmo lexicon with related emoji lexicons in the literature.

ArEmo Lexicon		Emoji Sentiment Ranking [18]		CF-Arab-ESL [38]		Arab-ESL [39]	
Emoji and Emoticon	**Score**	**Emoji**	**Score**	**Emoji**	**Score**	**Emoji**	**Score**
👍	0.709	😂	0.221	😂	0.839	😂	0.272
:)	0.417	❤	0.746	😍	0.946	💔	−0.934
👎	−0.5	♥	0.657	❤	0.911	❤	0.561
❤	0.829	😍	0.678	😢	−0.4	😭	−0.678
:(	−0.216	😭	−0.093	😭	−0.446	😍	0.87
👌	0.692	😘	0.701	♥	0.946	👌	0.766
😂	−0.118	😊	0.644	😘	0.946	☺	0.227
✖	−0.618	👌	0.563	🤣	0.786	💕	0.488
☆	0.594	💕	0.632	😁	0.839	😅	0.534
😠	−0.406	👏	0.52	😅	0.161	💗	0.616

For evaluating the performance of the proposed models, the study applied the statistical calculation of accuracy, precision, recall, and F1-score metrics using classification_report from the class of functions sklearn.metrics. It considered macro and weighted averages of precision, recall, and F1-score for the unbalanced dataset, and, as for the perfectly balanced dataset, both macro and weighted averages have exact results. The evaluation results of the proposed ensemble-based deep learning models EDLB and EDLU, when replacing the emojis and emoticons in the reviews with their meanings, are shown in Table 8. This was performed before experimenting with the different generated features, compared to the models before the replacement.

Table 8. Results of the proposed ensemble-based deep learning models EDLB and EDLU.

Model	Accuracy	Precision		Recall		F1-Score	
		Macro Avg	Weighted Avg	Macro Avg	Weighted Avg	Macro Avg	Weighted Avg
EDLB$_{WE}$	76.15%	75.90%	75.90%	76.15%	76.15%	75.74%	75.74%
EDLB	76.57%	76.38%	76.38%	76.57%	76.57%	76.22%	76.22%
EDLU$_{WE}$	77.59%	76.81%	77.45%	72.05%	77.59%	74.02%	77.41%
EDLU	77.75%	76.40%	77.64%	72.75%	77.75%	74.36%	77.62%

The results show an improvement after the replacement process using both model versions for balanced and unbalanced data. Moreover, the unbalanced dataset achieved higher accuracy than the balanced dataset. When a model trains on unbalanced data, it learns to obtain higher accuracy by consistently predicting the majority class, which is the class rating five in the HARD dataset [53]. Table 9 shows the detailed results for predicting each class using the EDLU model. Interestingly, the larger the review's train data in each class, the higher the f-score this class has. For instance, class 5, with the highest number of reviews, 115,422, has achieved the best f-score, followed by classes 4, 3, 2, and 1. These include the following number of reviews in each class: 105,757, 64,105, 30,888, and 11,478.

Table 9. Results of the EDLU model per class.

Class	Precision	Recall	F1-Score	Support
1	76.18%	60.37%	67.36%	2904
2	71.50%	68.39%	69.91%	7579
3	77.33%	72.49%	74.83%	16,221
4	74.70%	75.87%	75.28%	26,452
5	82.27%	86.67%	84.41%	28,756
Accuracy			77.75%	81,912
Macro Average	76.40%	72.75%	74.36%	81,912
Weighted Average	77.64%	77.75%	77.62%	81,912

Both ensemble models, EDLB and EDLU, have outperformed every single model included in their combined multiple models, as shown in Tables 10 and 11, which makes the ensemble models more accurate than single learners.

Table 10. Results of the ensemble model EDLB and its multiple internal models.

Model	Accuracy	Precision	Recall	F1-Score
CNN-BiLSTM	74.83%	74.75%	74.83%	74.47%
CNN	73.03%	72.72%	73.03%	72.55%
BiLSTM	75.06%	75.12%	75.06%	74.92%
EDLB	76.57%	76.38%	76.57%	76.22%

Table 11. Results of the ensemble model EDLU and its multiple internal models.

Model	Accuracy	Precision		Recall		F1-Score	
		Macro Avg	Weighted Avg	Macro Avg	Weighted Avg	Macro Avg	Weighted Avg
CNN-BiLSTM	76.79%	75.98%	76.79%	71.25%	76.79%	73.24%	76.69%
CNN	75.78%	75.06%	75.86%	70.11%	75.78%	72.04%	75.66%
BiLSTM	77.26%	75.12%	77.13%	72.97%	77.26%	73.91%	77.13%
EDLU	77.75%	76.40%	77.64%	72.75%	77.75%	74.36%	77.62%

The experimental results for combining the proposed models with the produced features from the HARD dataset after applying the MLP machine learning algorithm are presented in Tables 12 and 13. The features subsequently are the number of positive words, the number of negative words, the total weight of positive words, the total weight of negative words, the number of negation words, the number of booster words, the number of repeated characters, the number of question marks, number of exclamation marks, the number of periods, the number of commas, the total weight of emojis and emoticons, and all features together. The tables show that only some generated features impact improving the performance. More robust insights about the effect of the generated features on enhancing the performance can be driven using an additional different

dataset. The experiments achieved the highest improved results using the weight of negative words for the unbalanced data and the total weight of emojis and emoticons for the balanced data.

Table 12. Results of the proposed EDLB-MLP model after experimenting with different features.

Applied Feature	Accuracy	Precision	Recall	F1-Score
Num pos words	76.57%	76.38%	76.57%	76.22%
Num neg words	76.57%	76.38%	76.57%	76.22%
Weight pos words	76.56%	76.37%	76.56%	76.21%
Weight neg words	76.56%	76.37%	76.56%	76.21%
Num negation words	76.57%	76.38%	76.57%	76.22%
Num booster words	76.57%	76.38%	76.57%	76.22%
Num repeated characters	76.57%	76.38%	76.57%	76.22%
Num question marks	76.56%	76.37%	76.56%	76.21%
Num exclamation marks	76.57%	76.38%	76.57%	76.22%
Num periods	76.57%	76.38%	76.57%	76.22%
Num commas	76.57%	76.38%	76.57%	76.22%
Emo score	76.58%	76.39%	76.58%	76.23%
All features	76.54%	76.36%	76.54%	76.19%

Table 13. Results of the proposed EDLU-MLP model after experimenting with different features.

Applied Feature	Accuracy	Precision		Recall		F1-Score	
		Macro Avg	Weighted Avg	Macro Avg	Weighted Avg	Macro Avg	Weighted Avg
Num pos words	77.74%	76.39%	77.63%	72.75%	77.74%	74.35%	77.62%
Num neg words	77.74%	76.40%	77.63%	72.75%	77.74%	74.36%	77.62%
Weight pos words	77.75%	76.40%	77.63%	72.75%	77.75%	74.36%	77.62%
Weight neg words	77.75%	76.42%	77.64%	72.75%	77.75%	74.36%	77.62%
Num negation words	77.75%	76.40%	77.64%	72.75%	77.75%	74.36%	77.62%
Num booster words	77.75%	76.40%	77.64%	72.75%	77.75%	74.36%	77.62%
Num repeated characters	77.75%	76.40%	77.64%	72.75%	77.75%	74.36%	77.62%
Num question marks	77.75%	76.40%	77.63%	72.75%	77.75%	74.36%	77.62%
Num exclamation marks	77.74%	76.39%	77.63%	72.73%	77.74%	74.34%	77.61%
Num periods	77.74%	76.40%	77.63%	72.75%	77.74%	74.36%	77.62%
Num commas	77.75%	76.40%	77.64%	72.75%	77.75%	74.36%	77.62%
Emo score	77.75%	76.40%	77.64%	72.75%	77.75%	74.36%	77.62%
All features	76.33%	76.40%	77.61%	72.65%	77.72%	74.26%	77.59%

Table 14 compares the best-achieved results of balanced and unbalanced HARD datasets using the proposed models after emoji and emoticon replacement and using the features with the results of other studies in the literature that predicted user opinions based on five levels of ratings. For the balanced HARD data, using the EDLB-MLP model with the total weight of emojis and emoticons, the proposed classifier has achieved an increase of 3.21% in accuracy over Nassif et al.'s [20] top five models: random forest, convolutional neural network, decision tree, gated recurrent unit, and bi-directional recurrent neural network. The percentage increase formula is derived from the concept of percentage increase, as follows: [(New Accuracy − Old Accuracy)/Old Accuracy] × 100. The authors in [20] have used an initial embedding layer to learn embedding jointly with their deep learning-based models added to the tokenization process without using additional features. In contrast, this study used the fastText framework to learn the word embeddings from the HARD dataset in conjunction with the ensemble learning technique.

Employing this method with the MLP algorithm using the total weight of emojis and emoticons achieved a vast improvement of 237.27%, as compared to depending on the MLP algorithm without combining it with the proposed ensemble deep learning-based method. In the case of unbalanced HARD data, the proposed EDLU-MLP model, with the weight of negative words, had an increase of 2.17% in accuracy as compared with Elnagar et al. [17], where the authors used a logistic regression model using unigram and bigram features with TF-IDF. Similar to Elnagar et al. [17], the accuracy result of the proposed model, using an unbalanced HARD dataset, is higher than the proposed model using a balanced HARD dataset.

Table 14. Comparison of the experimental results for the proposed models using the balanced and unbalanced HARD datasets with models in the literature.

Model	Type of Data	Accuracy	Precision	Recall	F1-Score
Elnagar et al. [17] LR	Unbalanced	76.1%	-	-	-
Nassif et al. [20] RF	Balanced	74.2%	74.0%	74.0%	74.0%
Nassif et al. [20] CNN	Balanced	67.4%	77.3%	67.4%	72.0%
Nassif et al. [20] DT	Balanced	66.4%	66.0%	66.0%	66.0%
Nassif et al. [20] GRU	Balanced	65.1%	75.4%	65.3%	69.8%
Nassif et al. [20] BiLSTM	Balanced	63.1%	72.5%	63.2%	67.4%
Nassif et al. [20] MLP	Balanced	22%	20%	21%	14%
EDLU-MLP	Unbalanced	77.75%	76.42%	72.75%	74.36%
EDLB-MLP	Balanced	76.58%	76.39%	76.58%	76.23%

As accuracy is the common metric between the suggested models and the models in the literature, Figure 6 represents the performance comparison between all models using accuracy. The figure shows an improvement in the performance using both proposed EDLU-MLP and EDLB-MLP classifiers. UB indicates using unbalanced data with the model, while B indicates using balanced data.

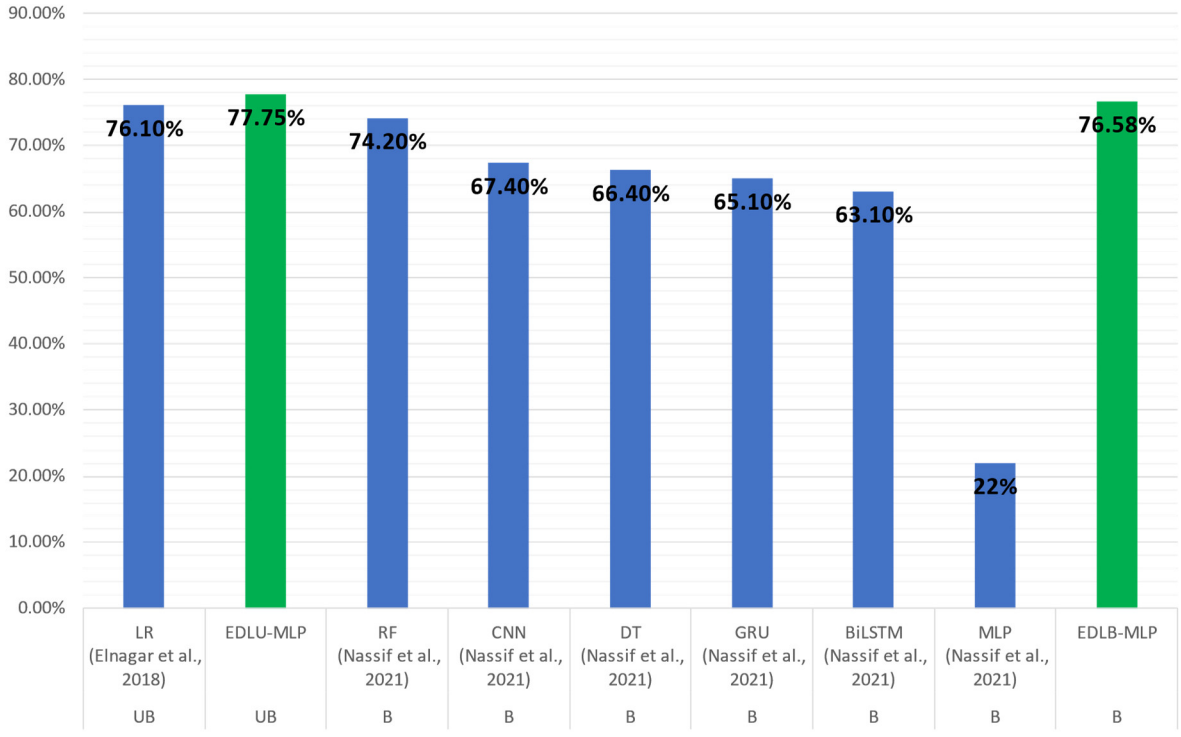

Figure 6. Accuracy results of user opinion prediction for models in the literature for the models LR [17], RF, CNN, DT, GRU, BiLSTM, and MLP [20] compared with the proposed models EDLU-MLP and EDLB-MLP.

5. Conclusions and Future Work

It is challenging to classify user opinion, especially with higher than two opposite classifications that either are positive or negative. For a dataset consisting of five levels of ratings, such as HARD, this is due to the relative levels of rating in the same category of good or bad, confusion of reviewers to determine the exact rating, mistakes to assign a rating, and could also be intended as spam or a misleading one from a competitor for instance. In order to contribute to the Arabic literature that needs more research and resources of opinion mining, this work has suggested an ensemble deep learning-based approach to predict the five-star rates of user opinions after substituting each emoji or emoticon with their equivalent meaning. The developed EDLB and EDLU models, based on balanced and unbalanced HARD datasets, added the advantage of transfer learning through pre-trained word embeddings, particularly the fastText technique to convert the review's words into numerical representations. Then, after experimenting by combining the result of the EDLB and EDLU classifiers with different features—generated from the HARD dataset—using MLP-based models that formed hybrid models, some results have shown an improvement. Comparing the results of this work with other research showed that the suggested method improved the accuracy by 3.21% using the balanced HARD dataset and improved the accuracy by 2.17% using the unbalanced HARD dataset.

In terms of future work, different avenues can be explored, including the following:

- Evaluate the use of context-based against context-free Arabic emoji and emoticon opinion lexicon using an Arabic dataset of emoji and emoticons.
- Improve non-Arabic emoji opinion lexicon for automated opinion mining by adding emoticons to it through suggested methods and algorithms.
- Work on more extensive extensible lists of emoji, emoticons, and negation words in both the standard Arabic language and dialectical Arabic. For instance, pictural symbols or negation words, combined with positive words, can be helpful to classify words while classifying text into negative when creating an Arabic corpus.
- Test the proposed approach using a dataset from a different domain while applying a robustness method and statistical significance analysis of the results.
- Apply an error detection and correction phase supporting the Arabic language as a preliminary step before applying the user opinion prediction approach to improve performance.

Author Contributions: Conceptualization, R.F.A.-M.; methodology, R.F.A.-M. and A.Y.A.-A.; software, R.F.A.-M.; validation, R.F.A.-M. and A.Y.A.-A.; formal analysis, R.F.A.-M. and A.Y.A.-A.; investigation, R.F.A.-M. and A.Y.A.-A.; resources, R.F.A.-M.; data curation, R.F.A.-M.; writing—original draft, R.F.A.-M.; writing—review and editing, R.F.A.-M. and A.Y.A.-A.; supervision, A.Y.A.-A.; project administration, R.F.A.-M.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The used HARD dataset and AraSenTi lexicon are publicly available from [17] and [33], respectively. The proposed ArEmo lexicons and the complete lists of Arabic booster and negation words are available at https://github.com/ralmutawa/files (accessed on 27 April 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

This appendix includes the layer and output shape structure of the proposed ensemble-based deep learning model.

Table A1. Structure of the proposed ensemble-based deep learning model.

Model 1		Model 2		Model 3	
Layer (Type)	**Output Shape**	**Layer (Type)**	**Output Shape**	**Layer (Type)**	**Output Shape**
embedding (Embedding)	(None, 615, 300)	embedding_1 (Embedding)	(None, 615, 300)	embedding_2 (Embedding)	(None, 615, 300)
conv1d (Conv1D)	(None, 613, 64)	conv1d_1 (Conv1D)	(None, 613, 64)	bidirectional_1 (Bidirectional)	(None, 615, 128)
max_pooling1d (MaxPooling1D)	(None, 306, 64)	max_pooling1d_1 (MaxPooling1D)	(None, 306, 64)	global_max_pooling1d_1 (GlobalMaxPooling1D)	(None, 128)
bidirectional (Bidirectional)	(None, 306, 128)	conv1d_2 (Conv1D)	(None, 304, 32)	dense_6 (Dense)	(None, 32)
global_max_pooling1d (GlobalMaxPooling1D)	(None, 128)	max_pooling1d_2 (MaxPooling1D)	(None, 152, 32)	dense_7 (Dense)	(None, 16)
dense (Dense)	(None, 16)	flatten (Flatten)	(None, 4864)	dense_8 (Dense)	(None, 8)
dense_1 (Dense)	(None, 8)	dense_3 (Dense)	(None, 16)	dense_9 (Dense)	(None, 5)
dense_2 (Dense)	(None, 5)	dense_4 (Dense)	(None, 8)		
		dense_5 (Dense)	(None, 5)		

References

1. Kuppusamy, S.; Thangavel, R. Deep Non-linear and Unbiased Deep Decisive Pooling Learning–Based Opinion Mining of Customer Review. *Cogn. Comput.* **2023**, *15*, 765–777. https://doi.org/10.1007/s12559-022-10089-1.
2. Farah, H.A.; Kakisim, A.G. Enhancing Lexicon Based Sentiment Analysis Using n-gram Approach. In *Smart Applications with Advanced Machine Learning and Human-Centred Problem Design*; Springer International Publishing: Cham, Switzerland, 2023; pp. 213–221. https://doi.org/10.1007/978-3-031-09753-9_17.
3. Chouikhi, H.; Alsuhaibani, M.; Jarray, F. BERT-Based Joint Model for Aspect Term Extraction and Aspect Polarity Detection in Arabic Text. *Electronics* **2023**, *12*, 515. https://doi.org/10.3390/electronics12030515.
4. El Khadrawy, A.S.A.I.; Abbas, S.; Omar, Y.K.; Jawad, N.H.A. Extracting Semantic Relationship Between Fatiha Chapter (Sura) and the Holy Quran. In Proceedings of the 8th International Conference on Advanced Intelligent Systems and Informatics, Cairo, Egypt, 20–22 November 2022.
5. Saeed, R.M.K.; Rady, S.; Gharib, T.F. Optimizing sentiment classification for Arabic opinion texts. *Cogn. Comput.* **2021**, *13*, 164–178. https://doi.org/10.1007/s12559-020-09771-z.
6. Alshahrani, A.; Dennehy, D.; Mäntymäki, M. An attention-based view of AI assimilation in public sector organizations: The case of Saudi Arabia. *Gov. Inf. Q.* **2022**, *39*, 101617. https://doi.org/10.1016/j.giq.2021.101617.
7. Zarezadeh, Z.Z.; Rastegar, R.; Xiang, Z. Big data analytics and hotel guest experience: A critical analysis of the literature. *Int. J. Contemp. Hosp. Manag.* **2022**, *34*, 2320–2336. https://doi.org/10.1108/IJCHM-10-2021-1293.
8. Zubair, F.; Shamsudin, M.F. Impact of covid-19 on tourism and hospitality industry of Malaysia. *J. Postgrad. Curr. Bus. Res.* **2021**, *6*, 6.
9. Darvishmotevali, M.; Altinay, L. Toward pro-environmental performance in the hospitality industry: Empirical evidence on the mediating and interaction analysis. *J. Hosp. Mark. Manag.* **2022**, *31*, 431–457. https://doi.org/10.1080/19368623.2022.2019650.
10. Ray, A.; Bala, P.K.; Jain, R. Utilizing emotion scores for improving classifier performance for predicting customer's intended ratings from social media posts. *Benchmarking Int. J.* **2021**, *28*, 438–464. https://doi.org/10.1108/BIJ-01-2020-0004.
11. Mammola, S.; Malumbres-Olarte, J.; Arabesky, V.; Barrales-Alcalá, D.A.; Barrion-Dupo, A.L.; Benamú, M.A.; Bird, T.L.; Bogomolova, M.; Cardoso, P.; Chatzaki, M.; et al. An expert-curated global database of online newspaper articles on spiders and spider bites. *Sci. Data* **2022**, *9*, 109. https://doi.org/10.1038/s41597-022-01197-6.
12. Antil, A.; Verma, H.V. Rahul Gandhi on Twitter: An analysis of brand building through Twitter by the leader of the main opposition party in India. *Glob. Bus. Rev.* **2021**, *22*, 1258–1275. https://doi.org/10.1177/0972150919833514.
13. Djatmiko, F.; Ferdiana, R.; Faris, M. A review of sentiment analysis for non-English language. In Proceedings of the 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), Yogyakarta, Indonesia, 13–15 March 2019; pp. 448–451. https://doi.org/10.1109/ICAIIT.2019.8834552.
14. Nassif, A.B.; Elnagar, A.; Shahin, I.; Henno, S. Deep learning for Arabic subjective sentiment analysis: Challenges and research opportunities. *Appl. Soft Comput.* **2021**, *98*, 106836 https://doi.org/10.1016/j.asoc.2020.106836.
15. Abo, M.E.M.; Raj, R.G.; Qazi, A. A review on Arabic sentiment analysis: State-of-the-art, taxonomy and open research challenges. *IEEE Access* **2019**, *7*, 162008–162024. https://doi.org/10.1109/ACCESS.2019.2951530.
16. Ghallab, A.; Mohsen, A.; Ali, Y. Arabic sentiment analysis: A systematic literature review. *Appl. Comput. Intell. Soft Comput.* **2020**, *2020*, 1–21. https://doi.org/10.1155/2020/7403128.

17. Elnagar, A.; Khalifa, Y.S.; Einea, A. Hotel Arabic-reviews dataset construction for sentiment analysis applications. *Intell. Nat. Lang. Process. Trends Appl.* **2018**, *740*, 35–52. https://doi.org/10.1007/978-3-319-67056-0_3.
18. Novak, P.K.; Smailović, J.; Sluban, B.; Mozetič, I. Sentiment of emojis. *PLoS ONE* **2015**, *10*, e0144296. https://doi.org/10.1371/journal.pone.0144296.
19. Nath, D.; Phani, S. Mood Analysis of Bengali Songs Using Deep Neural Networks. In *Information and Communication Technology for Competitive Strategies (ICTCS 2020) Intelligent Strategies for ICT*; Springer: Singapore, 2021; pp. 1103–1113. https://doi.org/10.1007/978-981-16-0882-7_100.
20. Nassif, A.B.; Darya, A.M.; Elnagar, A. Empirical evaluation of shallow and deep learning classifiers for Arabic sentiment analysis. *ACM Trans. Asian Low-Resource Lang. Inf. Process.* **2021**, *21*, 1–25. https://doi.org/10.1145/3466171.
21. Bashir, M.F.; Javed, A.R.; Arshad, M.U.; Gadekallu, T.R.; Shahzad, W.; Beg, M.O. Context aware emotion detection from low resource urdu language using deep neural network. *ACM Trans. Asian Low-Resource Lang. Inf. Process.* **2022**, *22*, 1–30. https://doi.org/10.1145/3528576.
22. Fei, H.; Zhang, Y.; Ren, Y.; Ji, D. Latent emotion memory for multi-label emotion classification. *Proc. AAAI Conf. Artif. Intell* **2020**, *34*, 7692–7699. https://doi.org/10.1609/aaai.v34i05.6271.
23. Chakraborty, S.; Goyal, P.; Mukherjee, A. Aspect-based sentiment analysis of scientific reviews. In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, Virtual, 1–5 August 2020; pp. 207–216. https://doi.org/10.1145/3383583.3398541.
24. Fei, H.; Chua, T.-S.; Li, C.; Ji, D.; Zhang, M.; Ren, Y. On the Robustness of Aspect-based Sentiment Analysis: Rethinking Model, Data, and Training. *ACM Trans. Inf. Syst.* **2022**, *41*, 1–32. https://doi.org/10.1145/3564281.
25. Li, X.; Bing, L.; Zhang, W.; Lam, W. Exploiting BERT for End-to-End Aspect-based Sentiment Analysis. In Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019), Hong Kong, China, 4 November 2019; pp. 34–41. http://dx.doi.org/10.18653/v1/D19-5505.
26. Fei, H.; Li, F.; Li, C.; Wu, S.; Li, J.; Ji, D. Inheriting the wisdom of predecessors: A multiplex cascade framework for unified aspect-based sentiment analysis. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, Vienna, Austria, 23–29 July 2022; pp. 4096–4103.
27. Shi, W.; Li, F.; Li, J.; Fei, H.; Ji, D. Effective Token Graph Modeling using a Novel Labeling Strategy for Structured Sentiment Analysis. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 22–27 May 2022; Volume 1, pp. 4232–4241. http://dx.doi.org/10.18653/v1/2022.acl-long.291.
28. Fei, H.; Shengqiong, W.; Jingye, L.; Bobo, L.; Fei, L.; Libo, Q.; Meishan, Z.; Min, Z.; Tat-Seng, C. LasUIE: Unifying information extraction with latent adaptive structure-aware generative language model. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 15460–15475.
29. Wu, S.; Fei, H.; Ren, Y.; Ji, D.; Li, J. Learn from Syntax: Improving Pair-wise Aspect and Opinion Terms Extraction with Rich Syntactic Knowledge. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21), Virtual, 19–26 August 2021.
30. Fei, H.; Shengqiong, W.; Yafeng, R.; Meishan, Z. Matching structure for dual learning. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 6373–6391.
31. Wu, S.; Fei, H.; Li, F.; Zhang, M.; Liu, Y.; Teng, C.; Ji, D. Mastering the explicit opinion-role interaction: Syntax-aided neural transition system for unified opinion role labeling. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–28 February 2022; Volume 36, no. 10, pp. 11513–11521. https://doi.org/10.1609/aaai.v36i10.21404.
32. Mo, X.; Tang, R.; Liu, H. A relation-aware heterogeneous graph convolutional network for relationship prediction. *Inf. Sci.* **2023**, *623*, 311–323. https://doi.org/10.1016/j.ins.2022.12.059.
33. Al-Twairesh, N.; Al-Khalifa, H.; AlSalman, A. Arasenti: Large-scale twitter-specific Arabic sentiment lexicons. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; Volume 1, pp. 697–705, http://dx.doi.org/10.18653/v1/P16-1066.
34. Alowisheq, A.; Al-Twairesh, N.; Altuwaijri, M.; Almoammar, A.; Alsuwailem, A.; Albuhairi, T.; Alahaideb, W.; Alhumoud, S. MARSA: Multi-domain Arabic resources for sentiment analysis. *IEEE Access* **2021**, *9*, 142718–142728. https://doi.org/10.1109/ACCESS.2021.3120746.
35. Alqmase, M.; Al-Muhtaseb, H.; Rabaan, H. Sports-fanaticism formalism for sentiment analysis in Arabic text. *Soc. Netw. Anal. Min.* **2021**, *11*, 52. https://doi.org/10.1007/s13278-021-00757-9.
36. Alhuri, L.A.; Aljohani, H.R.; Almutairi, R.M.; Haron, F. Sentiment analysis of COVID-19 on Saudi trending hashtags using recurrent neural network. In Proceedings of the 2020 13th International Conference on Developments in eSystems Engineering (DeSE), Virtual, 14–17 December 2020; pp. 299–304. https://doi.org/10.1109/DeSE51703.2020.9450746.
37. Alqmase, M.; Al-Muhtaseb, H. Sport-fanaticism lexicons for sentiment analysis in Arabic social text. *Soc. Netw. Anal. Min.* **2022**, *12*, 56. https://doi.org/10.1007/s13278-022-00871-2.
38. Hakami, S.A.A.; Robert, J.H.; Phillip, S. A Context-free Arabic Emoji Sentiment Lexicon (CF-Arab-ESL). In Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection, Marseille, France, 20 June 2022; pp. 51–59.
39. Hakami, S.A.A.; Robert, J.H.; Phillip, S. Arabic emoji sentiment lexicon (Arab-ESL): A comparison between Arabic and European emoji sentiment lexicons. In Proceedings of the Sixth Arabic Natural Language Processing Workshop, Virtual, 19 April 2021; pp. 60–71.

40. Hakami, S. Arabic_Emoji_Sentiment_Lexicon_Version_1.0.csv. GitHub, February 24, 2021. Available online: https://github.com/ShathaHakami/Arabic-Emoji-Sentiment-Lexicon-Version-1.0 (accessed on 18 August 2022).
41. UnicodePlus—Search for Unicode Characters. UnicodePlus—Search for Unicode Characters. 2021. Available online: https://unicodeplus.com/ (accessed on 27 September 2022).
42. Cyber Definitions. Available online: https://www.cyberdefinitions.com/ (accessed on 27 September 2022).
43. Full Emoji List, v15.0. 2023. Available online: https://unicode.org/emoji/charts/full-emoji-list.html (accessed on 21 March 2023).
44. Wikipedia, the free encyclopedia. "Wikipedia, the free encyclopedia," 2023. https://ar.wikipedia.org/.
45. Internet Slang Words—Internet Dictionary—InternetSlang.com. 2023. Available online: https://www.internetslang.com/ (accessed on 27 September 2022).
46. HiNative|A Question and Answer Community for Language Learners. 2023. Available online: https://hinative.com/en-US (accessed on 27 September 2022).
47. PC.net—Your Personal Computing Resource. 2023. Available online: https://pc.net/ (accessed on 27 September 2022).
48. Shaari, A.H. Accentuating illocutionary forces: Emoticons as speech act realization strategies in a multicultural online communication environment. *3L Southeast Asian J. Engl. Lang. Stud.* **2020**, *26*, 135–155. http://doi.org/10.17576/3L-2020-2601-10.
49. Trending—FastEmoji. n.d. Available online: https://www.fastemoji.com/ (accessed on 27 September 2022).
50. Amaghlobeli, N. Linguistic features of typographic emoticons in SMS discourse. *Theory Pract. Lang. Stud.* **2012**, *2*, 348.
51. Kaddoura, S.; Itani, M.; Roast, C. Analyzing the effect of negation in sentiment polarity of facebook dialectal arabic text. *Appl. Sci.* **2021**, *11*, 4768. https://doi.org/10.3390/app11114768.
52. Çoban, Ö.; Selma, A.Ö.; Ali, İ. Deep learning-based sentiment analysis of Facebook data: The case of Turkish users. *Comput. J.* **2021**, *64*, 473–499. https://doi.org/10.1093/comjnl/bxaa172.
53. Mohammed, A.; Arunachalam, N. Imbalanced machine learning based techniques for breast cancer detection. In Proceedings of the 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), Puducherry, India, 30–31 July 2021; pp. 1–4. https://doi.org/10.1109/ICSCAN53069.2021.9526422.