

Research Article

Manal Mostafa Ali*

Arabic sentiment analysis about online learning to mitigate covid-19

<https://doi.org/10.1515/jisys-2020-0115>

Received Nov 16, 2020; accepted Jan 15, 2021

Abstract: The Covid-19 pandemic is forcing organizations to innovate and change their strategies for a new reality. This study collects online learning related tweets in Arabic language to perform a comprehensive emotion mining and sentiment analysis (SA) during the pandemic. The present study exploits Natural Language Processing (NLP) and Machine Learning (ML) algorithms to extract subjective information, determine polarity and detect the feeling. We begin with pulling out the tweets using Twitter APIs and then preparing for intensive preprocessing. Second, the National Research Council Canada (NRC) Word-Emotion Lexicon was examined to calculate the presence of the eight emotions at their emotional weight. Third, Information Gain (IG) is used as a filtering technique. Fourth, the latent reasons behind the negative sentiments were recognized and analyzed. Finally, different classification algorithms including Naïve Bayes (NB), Multinomial Naïve Bayes (MNB), K Nearest Neighbor (KNN), Logistic Regression (LR), and Support Vector Machine (SVM) were examined. The experiments reveal that the proposed model performs well in analyzing the perception of people about coronavirus with a maximum accuracy of about 89.6% using SVM classifier.

From a practical perspective, the method could be generalized to other topical domains, such as public health monitoring and crisis management. It would help public health officials identify the progression and peaks of concerns for a disease in space and time, which enables the implementation of appropriate preventive actions to mitigate these diseases.

Keywords: Sentiment analysis, Machine learning, Arabic, Covid-19, Online education, Emotions

1 Introduction

Covid-19 has affected everyone's daily lives. It becomes one of the trending topics on Twitter since January 2020 and has continued to be discussed to date. The contingency assessments of public health officials and other agencies such as the World Health Organization (WHO) advised social distancing as a primary precautionary measure to mitigate the pandemic [1]. Under such a disruption, educational sectors over the globe have been hardly hit by the outbreak of the pandemic, which has led to profound changes in the educational deliveries. As a response to this, Internet and online courses become the best solution [2, 3]. But shifting from physical classrooms to online ones has not been without problems. Some challenges faced by online teaching platforms proposed by [4]. Thus, there is a critical need for practice-ready studies about distance learning so that authorities can make data-driven decision making from the insights of social media sentiment mining in real-time.

Sentiment Analysis (SA) is a classification process that identifies the opinions and emotions of users through the written contents [5, 6]. Researchers have mainly studied SA at three levels; 1) Document-level SA

Article note: Paper included in the Special Issue entitled: Intelligent Systems and Computational Methods in Medical and Healthcare Solutions

***Corresponding Author: Manal Mostafa Ali:** Al-azhar University, Faculty of Engineering, Computer and System Department; Email: manalmustafa10@yahoo.com

aims to classify a textual review, which is given on a single topic, as expressing positive or negative sentiment. 2) Sentence-level SA reflects the sentiment polarity of a single sentence. 3) Aspect-level SA is designed to solve the problem when multiple aspects show up in a complex sentence [7]. It needs to discover all aspects involved in the text and then perform SA for each aspect [8].

SA approaches can be classified into 1) Machine Learning (ML) approach, 2) Lexicon-Based approach and 3) Hybrid approach [9]. The first can be categorized into supervised learning that requires labeled data, unsupervised learning which reduces the need for using expounded trained data and finally semi-supervised technique that combines the two precedent approaches and includes supervision data, but, not for all the examples [10]. The second approach uses a sentiment dictionary with opinion words and matches them with the data to determine polarity [11]. The third approach collectively exhibits the accuracy of a ML approach and the speed of the lexical approach [12].

Classification performance can be improved with Feature Selection (FS) methods that select the most informative and relevant features. FS techniques can be classified into filters, wrappers and embedded methods [6, 13]. Filter methods exploit the new subset based on a performance measure and is independent of any ML algorithm. Wrapper methods elect the new subsets by evaluating the quality of performance on a modeling technique, which is considered as a black box evaluator. Hybrid based methods perform FS during the modeling algorithm's execution to select the optimal parameters. Eventually, SA comprises a multi-step process namely data extraction, text preprocessing, data analysis, and identification of useful knowledge [7]. Furthermore, most of the current studies related to this topic focus mainly on English texts with very limited resources available for other languages such as Arabic [14, 15]. Arabic SA is one of the more complicated SA tools of social media due to the informal noisy contents and the rich morphology of Arabic language [16, 17]. Lack of resources [15] and multi-dialectical forms add more challenges and ambiguities to analyze the Arabic sentiments [18].

Therefore, this paper identifies public Arabic sentiments about online learning during the pandemic. It applies different Natural Language Processing (NLP) techniques and ML algorithms to identify public Arabic sentiments. The framework begins with collecting Arabic sentiments about online learning and ends with analyzing these sentiments with common ML algorithms. Different stages of analyses were performed. Two types of lexicons were constructed and a method for negation handling was presented. Emotions are identified and analyzed using National Research Council Canada (NRC) emotion lexicon. Several experiments were conducted before and after applying Information Gain (IG) as a filtering method. The negative sentiments were detected and analyzed to recognize the potential reasons behind these negative sentiments.

This approach offers governments and decision maker with a solution to monitor and measure public satisfaction with respect to online learning during the occurrence of Covid-19 from people's posts on Twitter. The method could be generalized to other topical domains, such as public health monitoring and crisis management. In addition, the system could help public health officials identify the progression and peaks of concerns for a disease in space and time, which enables the implementation of appropriate preventive actions to mitigate these diseases. It also helps authorities in advocating effective personal hygiene and promoting social responsibility in spreading awareness to the public. It can provide a rapid and effective monitoring mechanism to manage future crises scenarios on a large scale at a low cost.

The rest of this paper is organized as follows. Discussion of Covid-19 related issues and some similar works of analyzing Arabic sentiments are provided in section 2. Proposed methodology and system architecture are explained in Section 3. Experimental setup and results analysis were discussed in Section 4. Conclusion and possible directions for future work are given in Section 5.

2 Related Work

2.1 Sentiment Analysis during Covid-19

Public SA during the Covid-19 outbreak provides insightful information in making appropriate responses. Over the last few months, many studies are working on SA during Covid-19. The authors of [19] designed a model that can effectively predict the sentiment expressed by people on social media platforms amidst this pandemic. Different classification techniques revealed that both Support Vector Machine (SVM) and Decision Tree (DT) have performed extremely well, but SVM classifier was more robust and consistent throughout all the experiments. A further investigation was introduced by [20] who used textual analytics methodologies to analyze public sentiment about Covid-19. They introduce public sentiment scenarios (PSS) framework that can manage future crises scenarios. The study examines two potential divergent scenarios - an early opening and a delayed opening, and the consequences of each. A similar study is conducted by the researchers in [21] who identified the reaction of citizens and people's sentiment about subsequent actions taken by countries during coronavirus. Deep long short-term memory (LSTM) models are used for estimating the sentiment polarity. They provide interesting insights into collective reactions on coronavirus outbreak on social media. The work presented by [22] was interested in recognizing characteristics of negative sentiment about Covid-19 related comments. Therefore, they analyzed public concerns by selecting coronavirus related Weibo posts to identify characteristics of negative sentiment. The results proved that people concern four aspects regarding Covid-19; virus origin, symptom, production activity and public health control.

Exploratory SA during coronavirus pandemic is examined in the study conducted by Samuel *et al.* [23]. Two groups of data containing different lengths of tweets are used for testing. The first group comprises shorter tweets with less than 77 characters, and the second one contains longer tweets with less than 120 characters. NB achieved an accuracy of 91.43% for shorter tweets and 57.14% for longer tweets, whereas, a worse performance is obtained by Logistic Regression (LR), with an accuracy of 74.29% for shorter tweets and 52% for longer ones respectively. [24] presented a model that analyzes the student's sentiment in the learning process during the pandemic using Word2vec and ML techniques.

The emotional classification methods mainly include dictionary-based methods, rule-based methods, ML methods, composite methods, and multi-label methods. [25] developed a lexicon-based approach for emotion analysis of Arabic text on Facebook and Twitter datasets. They showed that the lexicon-based approach is effective with an accuracy of 89.7%. Another work focuses on emotional reactions during the Covid-19 outbreak by exploring the tweets investigated by [26]. A random sample of about 18,000 tweets had examined for classification along with the basic eight emotions. The findings showed that there exists an almost equal number of positive and negative sentiments. The fear among the people was the number one emotion that dominated the tweets, followed by the trust of the authorities. Also, emotions such as sadness and anger of people were prevalent. Thus, the key findings of the literature review show that the research trends of SA using social media data during the pandemic and natural disasters are still evolving. Moreover, the aforementioned researches concerning the Covid-19 pandemic are in English language. Accordingly, emotional mining and SA in Arabic language need further attention.

2.2 Tracking Arabic Sentiment

Comparing Arabic with the other languages reveal that only few studies have investigated Arabic SA [7]. [11] showed that out of the 1458 SA-related papers that were published in 4 different databases, *i.e.*, Association for Computing Machinery (ACM), ScienceDirect (SD), IEEE Xplore (IEEE), and Web of Science (WoS), till May 2017, only 48 were related to Arabic SA.

Different surveys have discussed the characteristics of Arabic language such as [7, 15, 27]. Guellil *et al.* [7] surveyed the most recent resources and advances that have been done for Arabic SA. The authors found that the significant problems related to the treatment of Arabic and its dialects had been the lack of resources.

Therefore, they focus on the construction of sentiment lexicon and corpus. [15] introduced an exhaustive review of different approaches to Arabic SA. They review Arabic SA in-depth and outline the limitations of the current resources. But they proved that most SA approaches fail in Arabic social media space due to dialects and suggest shifting from word-level to concept-based SA. The work reported by [27] discussed existing work on Arabic SA. They surveyed a large number of studies, methods and the available Arabic sentiment resources.

Other work conducted by [11] focused on the various characteristics, state-of-the-art, and the level of SA along with NLP applied in the Arabic SA. They gave particular attention to dialectal Arabic and found it's difficult to handle the diverse slang due to its linguistic complexity. The finding also showed that the accuracy of the SA method is based on the existence of large annotated corpora, which is a limited resource for Arabic language. Alomari *et al.* [28] introduced a new Arabic Jordanian annotated twitter corpus and investigated several ML techniques for evaluating their performance. They also exploit different preprocessing strategies, N-grams and different weighting schemes. The best performance achieved by combining SVM classifier and term frequency-inverse document frequency (TF-IDF) weighting scheme with stemming through Bigrams. A comprehensive study of the different tools for Arabic text preprocessing, feature reduction and classification is presented by [29]. Their experiments proved the superiority of SVM followed by DT and NB.

AlSalman [16] proposed a corpus-based approach for Arabic SA of tweets. Their method use Discriminative Multinomial Naïve Bayes (DMNB) algorithm with N-grams, stemming, and TF-IDF techniques. Their results improved accuracy by 0.3%. The authors of [14] addressed the issue of the multi-way SA problem for Arabic reviews. That work examined a dataset of more than 63,000 book reviews based on a 5-star rating system. The evaluation accuracy showed that Multinomial Naïve Bayes (MNB) had the highest classification accuracy for both balanced and unbalanced datasets with average accuracy reached 46.4%.

[30] developed a system called "SAMAR" that jointly classifies subjectivity of a text as well as its sentiments. They showed how complex morphology characteristics of Arabic can be handled in the context of subjectivity and SA. Their system is based on Modern Standard Arabic (MSA) and Egyptian dialect. In that work, the authors used SVM classifier and achieved accuracy up to 84.65%. A similar finding was reported by [31] who analyzed sentiment embedded in blogs that are written either in Arabic or English on web forums. The authors represented each review using a set of syntactic and stylistic features. Although the previous papers [30, 31] used varieties of feature sets, they avoided semantic features because they are language dependent and need lexicon resources. Furthermore, few types of research used Arabic WordNet (AWN) as a semantic resource for improving classification results such as [32].

From the literature it can be inferred that besides the general challenges of SA, there are other challenges related to Arabic varieties and morphology. Availability of annotated datasets and lack of lexicons are common challenges of analyzing Arabic sentiments. Most of the existing Arabic SA approaches are semantically weak. Words are considered as independent features and the semantic associations have been ignored. As a result, synonymous words are represented as different independent features. There is also a lack for considering negation that can invert the Arabic text completely. Thus, Arabic SA needs further investigation.

However, this study differs from the previous papers by providing a newly exhaustive model that analyzes Arabic sentiments using NLP and ML techniques. Specifically, the method begins with collecting and preparing corpora about online learning during Covid-19 to explore the contexts and trends associated with this pandemic. Intensive preprocessing including morphological and semantic analysis had been performed. Objective and adjective lexicons were constructed. A new method for negation detection was developed. NRC emotion lexicon was employed to classify the tweets into one of the basic eight categories such as fear, sadness, anger and disgust. The research is also interested in analyzing the latent reasons behind the public sentiment variations regarding online education.

3 Proposed Methodology

Accordingly, the proposed approach consists of the following steps: fetching the Arabic tweets about online learning in the context of Covid-19, intensive preprocessing of the tweets, construction of Bag of Words (BoW),

morphological analysis, semantic analysis, analyzing emotions, employing IG as a filtering technique and finally, comparison of different ML approaches for classification of domain-specific tweets on two different datasets. System architecture is given in Figure 1 and will be discussed in the following subsections.

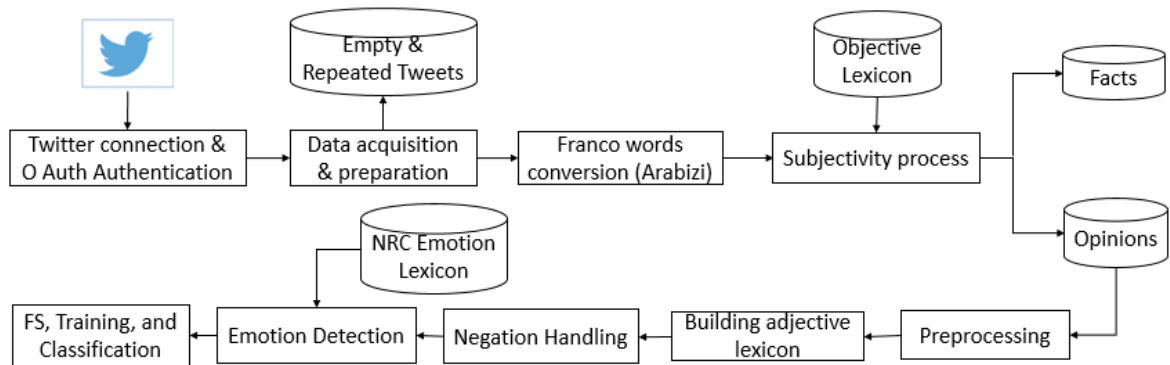


Figure 1: System Architecture

3.1 Data acquisition and preparation

Twitter is used as the primary data source in order to gather tweets specific to online teaching in the context of coronavirus. However, Twitter does not provide developers with an API to download historical data. Standard Search API of Twitter allows developers to collect tweets published in the past 7 days [8]. Tweets were extracted using the hashtags expressed in Table 1. The current data collection is limited to “Arabic” tweets only.

Table 1: Keywords for dataset collection

| Arabic word | English Equivalent |
|---------------------|-------------------------|
| #كوفيد19 | #Covid-19 |
| #فيروس كورونا | #Coronavirus |
| #كورونا | #Corona |
| #التحول الرقمي | #Digital transformation |
| #التعليم عن بعد | #Online learning |
| #التعليم الإلكتروني | #Online learning |
| #العالم الرقمي | #Digital world |

Moreover, the most frequently used words regarding Covid-19 were studied and drawn in word clouds where the size of every word shows how important it is in the text as depicted in Figure 2.

3.2 Franco words conversion

The crawled datasets contain a lot of empty and repeated tweets that need to be efficiently excluded. Besides, Arabic tweets may contain “Arabizi”, where Arabic words are written using Latin characters [33]. Therefore, franco words were converted to its Arabic equivalent by examining google’s API.

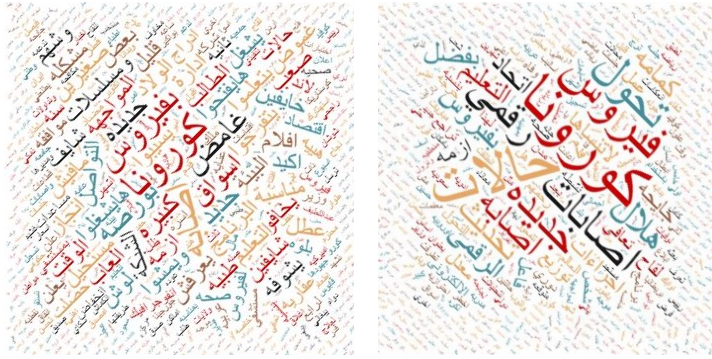


Figure 2: Instances of word cloud in twitter data

3.3 Subjectivity process

The textual datasets have two primary configurations; either facts or opinions [30, 34]. Facts are known as objective information about elements, objects, occasion and their properties while opinions are generally subjective expressions that illustrate an individual's sentiments. Some samples of facts and opinions are presented in Table 2. Many methods have been identified for subjectivity analysis including using some patterns of word usage, detecting of certain kinds of adjectives, presence of emojis, and occurrences of certain discourse connectives [35].

Table 2: Some instances for facts and opinions

| Tweet No | Tweet Text | Subjectivity |
|----------------|--|-----------------------|
| T ₁ | #عاجل وزير التعليم خلال كلمته عن بُعد في لقاء "العلوم والتكنولوجيا" بطوكيو: الدولة تحولت سريعاً إلى العالم الرقمي مع بدء جائحة #كورونا | Fact |
| T ₂ | التعليم ليس استعداداً للحياة، إنه الحياة ذاتها! استثمر في طفلك مع #المعهد_العالمي_للغات 🌐 #التعليم عن بعد. للتواصل: #####... | Fact |
| T ₃ | #التعليم الإلكتروني ومع عودة جميع الكادر التدريسي وتكديسهم في المدارس في هذه الفترة ستجد الزيادة في حجم الإصابات بفيروس كورونا كبير | Opinion (Negative) |
| T ₄ | من أجمل الحقايب الممتعة والمثمرة التي حضرتها حقيقي تحول رقمي هائل و رائع | Opinion (Positive) |

This study extracts facts by defining special dictionaries called the objective lexicon, containing synonyms for objective words. The main flaw of this approach is the necessity of manual selection of terms for the filters that can extract out facts, news and advertisements. For instance, T₁ in Table 2 is labeled as a fact because it has word “عاجل” which often used by channels and news sites to express some news. Similarly, T₂ in Table 1 is considered as objective as it contains the word “للتواصل” to announce some advertisements. These tweets are early excluded without any processing and categorized as objective information.

3.4 Data preprocessing

Before proceeding emotion mining and SA, the tweets are preprocessed to produce clean texts for NLP tasks. The most common techniques are removing hashtags, removing URLs, identifying emoticons, and removing user mentions and extra spaces as established in Figure 3. Punctuation is replaced with a single space and

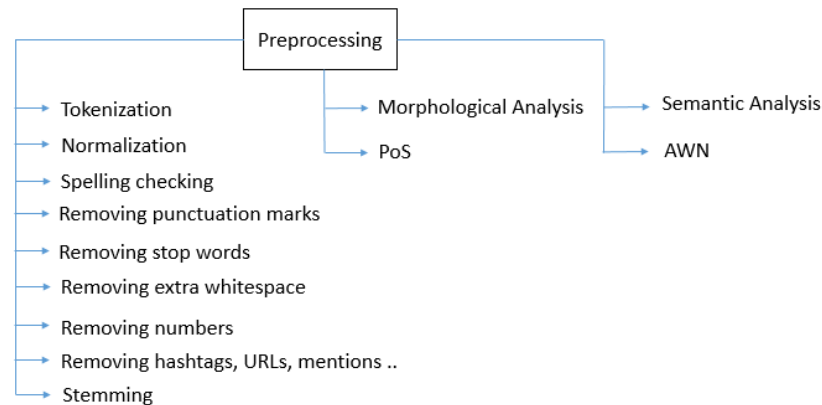


Figure 3: Preprocessing.

spelling correction was conducted to prepare the dataset ready for stemming. Besides, before classifying the sentiment of tweets, it is important to preprocess the text such that specific letters are normalized [36]. In particular, to reduce noise and sparsity in Arabic text orthographic normalization of certain Arabic letters had been performed. Orthographic normalization is the process of unifying the shape of some Arabic letters that have different shapes [17, 37].

Stop words¹ are not very helpful as we would expect them to be evenly distributed across the different texts and can be effectively removed [6]. In many cases, some of the words (e.g., *يتعلموا*, *علوم*, *تعليمهم*, *التعليم*) have multiple morphemes but still mean similar thing. A single representative word would be sufficient. Instead of using all of these forms. Lemmatization, an improved version of word stemming, uses morphological analysis of words to remove inflectional endings. Part of Speech (PoS) tagging is also performed for marking words, in a text, based on their nature and their relationships with adjacent and related words in that text. Every word has been associated with a relevant tag showing its role in the sentence. The entire list of tags along with their typical meaning is based on the ICA Tagset². PoS for every term in T_4 is given in Table 3.

Table 3: PoS for T_3 .

| Word | PoS |
|---------|---------|
| من | PPN |
| اجمل | NOU |
| الحقائب | NOU |
| الممتعه | ADJ |
| والثمره | ADJ |
| اللي | Unknown |
| حضرته | VER |
| حقيقى | NOU |
| تحول | NOU |
| رقمي | ADJ |
| رائع | ADJ |
| وهائل | ADJ |

¹ <https://sites.google.com/site/kevinbouge/stopwords-lists>

² <http://www.unlweb.net/wiki/index.php/Tagset>

It should be considering that the original form of a term is returned if it is unknown to the PoS tagger. Moreover, T_4 in Table 2 contains (هائل - رائع) as different adjectives but have a similar meaning. Classifiers cannot deal with such adjectives as correlated words that provide similar semantic interpretations. So,AWN [32] semantic relations are used to group those phrases into one synset.

3.5 Adjective lexicon

Adjectives reflect most of subjective information in a given text. Adjective lexicon “AdjLex” development includes many steps; first, a list of words, called seed words had been created. We start with some adjectives collected manually as seeds from different datasets in Arabic language. Second, the common lexicon techniques require human annotation. Third, the initial lexicon is expanded by collecting synonyms, morphemes and antonyms of seed words. Finally, this lexicon was extended through google translate to get more acronyms for the Arabic adjectives.

3.6 Negation handling

The presence of negation words may force the sentence into opposite polarity. This work considers 3 steps for automatic negation handling that can be summarized as follows. 1) Recognizing the negation word such as “ما - ليس” that means “not”, 2) Identifying the scope of negation (which words are affected by the negation word) and 3) Finally capturing appropriately the impact of the negation. Traditionally, the negation word is determined from small hand-crafted words. So, we define a list of negation terms in Modern Standard Arabic (MAS) and Egyptian dialect which can change the sentiment polarity (around 35 words in Arabic). The scope of negation is considered as the first word follows the negation particle. Negation words can be detected as follows.

```

for each  $T_k$  in the dataset { //  $T_k$  refers to tweet
  if ( $T_k$  contain negation article){
    flip the first word after the negation word
    classify
  }
}
```

All features are then grouped for Bag of Words (BoW) construction which will be used for training and classifications.

3.7 Emotion detection

The tweets are full of emojis and emoticons that are widely used by people to express their feelings. All emojis and emoticons provided by Twitter are kept and considered as a part of the texts. There is a list of emojis with their weights listed in the National Research Council Canada (NRC) emotion lexicon³. With all emojis assigned weights, we can proceed to annotate each tweet based on the emoji it contains. The different emotion annotations for a target term were consolidated by selecting the emoji that has the highest weight as elaborated in Figure 4. The NRC emotion lexicon was examined to calculate the presence of the basic eight emotions

³ <http://sentiment.nrc.ca/lexicons-for-research/>

(“anger”, “fear”, “anticipation”, “trust”, “surprise”, “sadness”, “joy”, and “disgust”) and their corresponding valence in coronavirus datasets. The procedure can be summarized in the following steps.

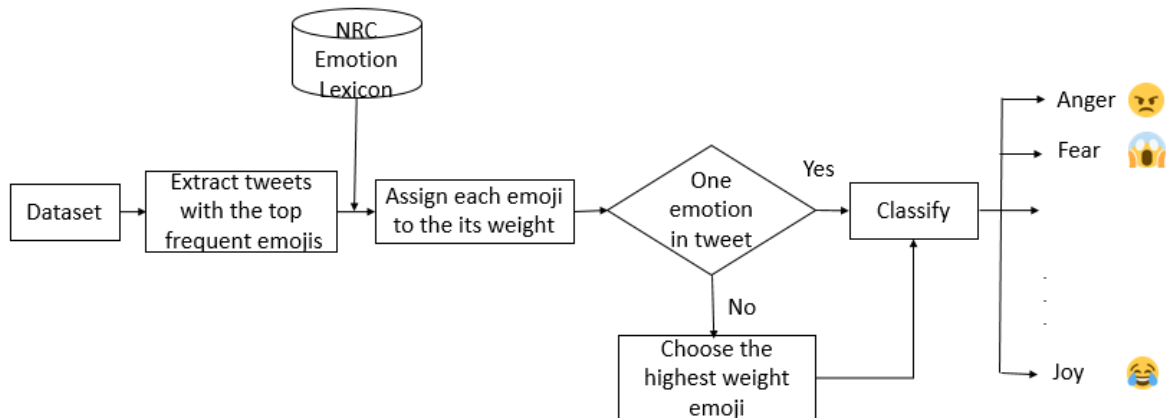


Figure 4: Emotion detection process.

1. Different datasets about coronavirus was scanned in order to estimate the top frequent emojis.
2. Every emotion is replaced with its typical weight by using NRC lexicon.
3. Emoticons are then converted to their corresponding words by using a specialized mapping table that maps common emoticons to their respective words.
4. Tweets with one emoji had been directly classified into one of the basic eight emotions.
5. The highest weight emotion is considered for tweets that contain more emojis and then categorized into one of the eight categories of emotions. Table 4 exhibits different samples of emojis distributed among the different tweets.

Table 4: Instances of tweets having emotions.

| Tweet No. | Tweet Text | Emotion Type | Weight | Category | Opinion |
|----------------|---|----------------|----------------|----------|----------|
| T ₁ | من اجمل الحقائق الممتعه والمثمره اللي حضرتها حقيقي تحول رقمي هائل و رائع 😊 | Smiling | 0.812 | Joy | Positive |
| T ₂ | مقبلون على تحول رقمي راح يتسبب في تقليص عدد كبير من الوظائف 😡😡 | Dizzy Angry | 0.562 0.824 | Angry | Negative |

T₁ in Table 4 contains only one emotion, so the emotion was directly replaced with the corresponding weight by employing NRC emotion lexicon. The second tweet T₂ includes two different emotions; dizzy and angry faces. Therefore, their weights are assigned and the highest weight which was angry was selected. After assigning the emojis into their categories, we label each tweet according to the emoji it contains.

4 Experiments and Results Analysis

4.1 Experimental setup

Two different datasets from Covid-19 related comments about online learning are used for the experiment. The datasets were collected between September 20, 2020 and October 15, 2020. Table 5 shows the target distribution of the two different datasets. During the classification process, we employed 10-fold cross-validation for all experiments by dividing the dataset into 10-folds, whereby one fold is used for testing purposes and the remaining 9 folds are used for training purposes. This process is repeated 10 times. Lastly, the average accuracy, average number of selected features, and average fitness across 10 runs are reported. All experiments were performed with the presence vectors as it reveals an interesting difference [38]. In each segment vector, the value of each dimension is binary regardless of how many times a feature occurs.

Table 5: Statistics of Arabic datasets.

| Dataset | D ₁ | D ₂ |
|--|----------------|----------------|
| Total number of tweets | 4689 | 5798 |
| Total number of positive tweets | 2344 | 2899 |
| Total number of negative tweets | 2345 | 2899 |
| Total number of words | 51814 | 59720 |
| Total number of unique words | 22563 | 25124 |
| Average number of words in each tweet | 11.05 | 10.3 |
| Average number of characters in each tweet | 52.6 | 61.03 |

Table 6: Classification algorithms and setting parameter.

| Classification Model | Description and parameter setting |
|-------------------------------|---|
| Naïve Bayes (NB) | Probabilistic classifier based on the Bayes' theorem Effective with real-world data, efficient and can deal with dimensionality. Over-simplified assumptions and limited by data scarcity |
| Multinomial Naïve Bayes (MNB) | Based on word appearance only. Can account for multiple repetitions of a word. Faster than plain NB. |
| K Nearest Neighbor (KNN) | Computes classification based on weights of the nearest neighbors, instance based. Easy to implement, efficient with small data, applicable for multi-class problems and sensitive to data quality. Noisy features degrade the performance. Distance measure: Euclidean distance and linear search |
| Logistic Regression (LR) | Probability of an outcome is based on a logistic (s-shaped) function. Regularized to avoid over-fitting. Expensive training phase. Distance measure: Euclidean distance. |
| Support Vector Machine (SVM) | Non-probabilistic binary classification model. Finds a decision boundary with a maximum distance between two classes. Kernel: RBF. Exponent = 1.0, Complexity (c) = 10.0. |

Naïve Bayes (NB), Multinomial Naïve Bayes (MNB), K Nearest Neighbor (KNN), Logistic Regression (LR) and Support Vector Machine (SVM) were used as classification techniques. Table 6 summarizes the description of the traditional ML models with suitable parameter settings.

4.2 Experimental results

The first experiment is represented by evaluating the performance of the proposed method using either NB, MNB, KNN, LR or SVM classifier. In this experiment, each classifier is examined only on the whole features space without applying any filtering technique. Figure 5 reports the sentiment prediction for D_1 in terms of precision, recall and F-Measure.

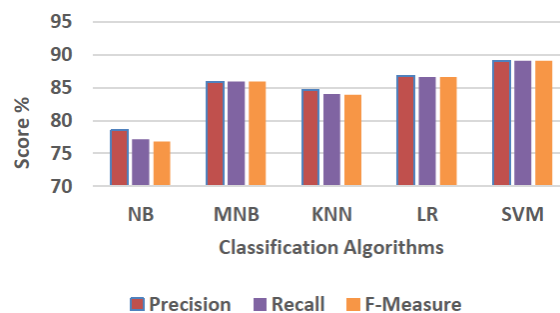


Figure 5: Sentiment prediction for D_1 .

For D_1 , The best classification accuracy is obtained by using SVM classifier which was 89.1%. On average, there is no significant difference in the performance of LR, and MNB which respectively recorded 86.6% and 85.9% followed by KNN which gives 83.9%. The lowest accuracy was given by NB at 76.8%. In general, every algorithm achieves approximately equal values of precision and recall.

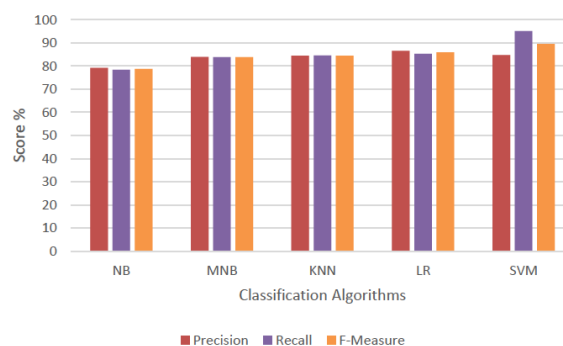


Figure 6: Sentiment prediction for D_2 .

Experimental results for D_2 are indicated in Figure 6. Based on these results, it's noticed that SVM still has the best accuracy which was 89.6% but returns recall values higher than those of precision. Higher recall values mean that, the algorithm returns most of the relevant results. LR followed by KNN and MNB have an almost similar accuracy and relatively equal values of precision and recall for each one. However, NB records the lowest accuracy at 78.8% and nearly equal values of precision and recall. The classification results of NB show that accuracy result is comparatively less than all other individual classifiers. Poor classification accuracy of NB for the two different datasets may be attributed to the fact that all features are independent.

Based on these results, it is confirmed that the SVM classifier outperforms other ML classifiers including KNN, NB, MNB and LR. Also, these results are consistent with the results of the previous works [28–30, 39, 40] that reveal the superiority of SVM in comparison with other classifiers over Arabic classification. Comparison of these studies in terms of datasets, preprocessing methods, features, classification algorithms and accuracies are tabulated in Table 7.

Thus, experimental results concluded that the proposed model can analyze public Arabic sentiment about online learning during Covid-19 using ML methods with a good accuracy result.

Table 7: A performance comparison of Arabic SA.

| Study | Dataset | Preprocessing | Features | Algorithm | F-Measure |
|-------|--|---|---|-----------|---|
| [28] | MSA and Jordanian Twitter | word stemming | TF-IDF or (TF) n-grams | SVM | 88.7% |
| | | | | NB | 83.6% |
| [29] | Saudi press agency Saudi newspapers Websites Writers Forums Islamic topics Arabic poem | Normalization and stop words removal | relative frequency, entropy, LTC, TFC, TFIDF, frequency and Boolean | SVM | The best accuracy ranges from 60.63 to 96.72% using seven corpora, with an average of 85.06%. |
| | | | | DT | |
| | | | | NB | |
| | | | | MLPs | |
| | | | | KNN | |
| [30] | MSA and Dialects Twitter, chat, forum | Tokenization and Lemmatization | POS Tags, standard, dialectal, and genre features | SVM | 84.65% |
| | | | | | |
| [39] | MSA and Moroccan Facebook comments | Cleaning noise, normalizing, tokenizing, and word stemming. | TF-IDF and n-grams | NB | 81.83% |
| | | | | SVM | 78.94% |
| [40] | The corpus contains 5070 documents of different lengths. These documents consist of six classes. | Normalization Stemming Stop word removal. | vector of terms (word stem) | DT | Best average 75.43% for DT and 84.29% for SVM. |
| | | | | SVM | |

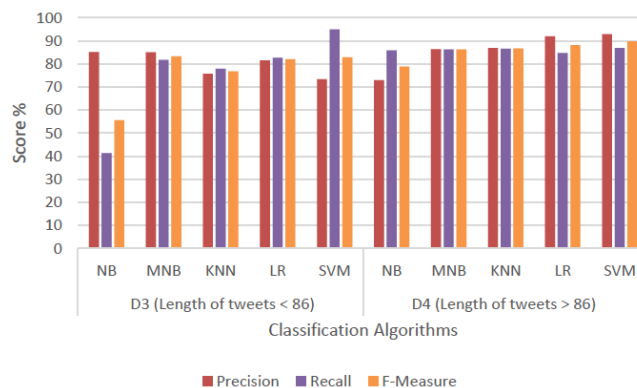


Figure 7: Sentiment prediction based on the length of tweets.

Furthermore, subsets of data were created based on the length of tweets to examine classification accuracy based on the length of tweets. Two groups were constructed where the first group consisted of Covid-19 related tweets which were less than 86 characters in length while the second group consisted of coronavirus tweets that have a length greater than 86 characters. These groups were then examined to ensure that the number of positive and negative tweets were balanced when being classified. The results have been drawn in Figure 7.

It is found sufficient directional support for classification algorithms with long length tweets, but the accuracy degrades with decreasing the length of tweets.

4.3 Filtering with IG

Even though words are good features for classification, we shouldn't employ all of those words. Many of these features may degrade the classifier performance and increase computational cost [41]. Hence, in the second experiment, IG [6] is involved with SVM and applied to rank all extracted features. The method can be expressed as follows. First, the IG is computed for each feature. Next, the IG scores of all the features are sorted from high to low and the top $k\%$ features are used in conjunction with the SVM. The k percentage may either be determined using validation data or it can be manually set. In this experiment, the top features are filtered according to their IG weights using 6%, 12%, and 18% ratios. Figure 8 shows the results after applying IG feature with SVM.

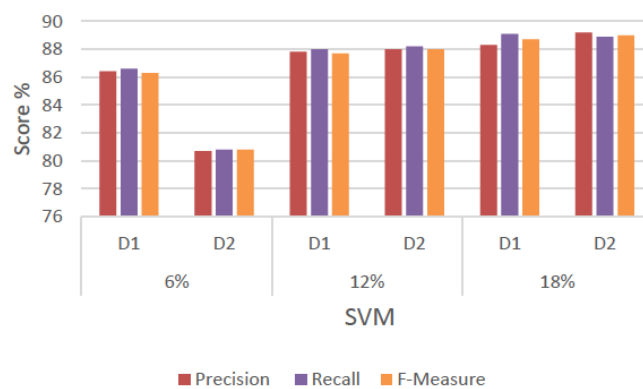


Figure 8: Sentiment prediction using IG and SVM with different k values.

The top 6% of ranked features selected by IG register uneven results when involved with SVM. For instance, the performance of SVM on D_1 has dropped from 89.1% (Figure 5) to 86.3% and falls from 89.6% (Figure 6) to 80.8% for D_2 . The ratio of 12% keeps reasonable results for both D_1 and D_2 . The best accuracy was achieved by selecting the top 18% of the ranked features where the classification accuracies were 88.7% and 89% for D_1 and D_2 respectively. These results prove that a suitable choice of ranker and feature subset size significantly impact the classifier performance.

4.4 Finding the latent reasons behind negative sentiment

To further enhance the readability of the mined reasons, we studied the most frequent words with the negative sentiments regarding coronavirus and online learning. Then tweets contain these words are determined and analyzed. It has been noted that people mainly concern with five aspects regarding negative sentiments about online learning in the context of Covid-19 that can be summarized as follows:

1. Online learning suffers from lack of supervision. This problem is mainly obvious for compulsory education. Most of the students at this age have poor self-management constraints and self-driving.

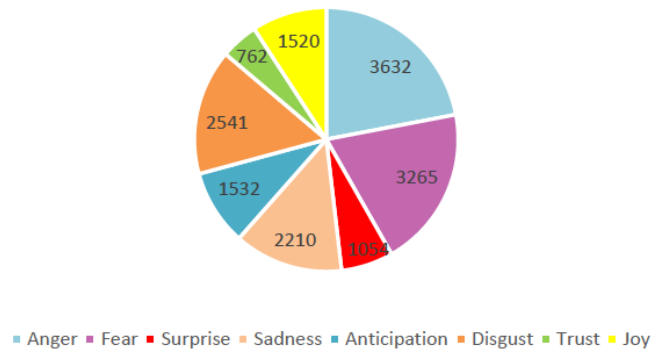


Figure 10: Statistical distribution of emojis on Covid-19 related tweets.

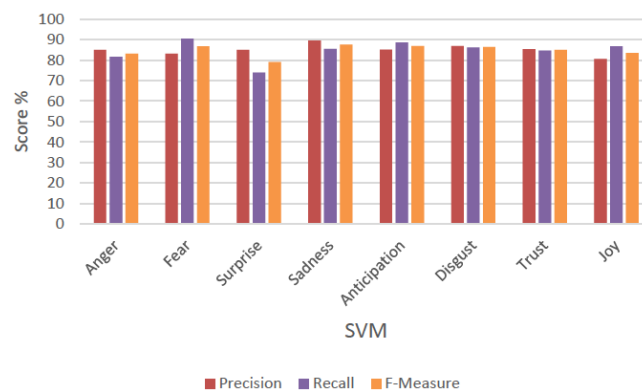


Figure 11: Sentiment prediction for emotions.

5 Conclusion

This paper has analyzed Arabic sentiments about online learning during the ongoing worldwide Covid-19 pandemic. Different aspects of analyzing Arabic text and intensive preprocessing were effectively performed. Various ML techniques are applied over two generated datasets and a comparison was done over their performance with well-known classification techniques. SVM-based models offer very high accuracy and consistency over other classifiers, including LR, MNB, KNN and NB. The best accuracy achieved by SVM was 89.6% for D_2 followed by 89.1% for D_1 . LR and MNB recorded 86.6%, 85.9% for D_1 and 85.9%, 83.8% for D_2 respectively. No clear difference between the performance of KNN on both datasets as it records 83.9% for D_1 and 84.4% for D_2 . The worst accuracy was given by NB that gives 76.8% and 78.8% for D_1 and D_2 respectively. It also observed that longer texts were more useful to identify sentiment than shorter ones. SVM registers 89.8% for longer tweets and 82.8% for shorter ones. Emotion analysis was also considered and the anger among the people was the top emotion that dominated the tweets, followed by the fear from the first attempt to try out distance learning. To further enhance the readability of the mined reasons, we select the most representative negative tweets to define the latent reasons behind the negative results of online learning. Absence of face to face communication, network system breakdown, ambiguity and games were the most significant reasons behind the negative sentiments.

In the future, open research challenges can be investigated, with a focus on the shortage of lexicon availability; use of Dialect Arabic (DA); lack of corpora; compound phrases and idioms.

References

- [1] Alam, A. S.; Lau, E.; Oh, C.; Chai, K. K. "An Alternative Laboratory Assessment Approach for Multimedia Modules in a Transnational Education (TNE) Programme during COVID-19", 2020 Transnational Engineering Education using Technology (TREET), Glasgow, United Kingdom, IEEE Xplore. Sep 2020.
- [2] Ping, Z.; Fudong, L.; Zheng, S.; "Thinking and Practice of Online Teaching under COVID-19 Epidemic", 2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI). Xinxiang, China, China, IEEE Xplore, Sep 2020.
- [3] Li, J.; Li, C.; "Exploration and Practice of the Teaching Pattern of Skill-oriented Courses in the Context of Online Home Schooling", 2020 15th International Conference on Computer Science & Education (ICCSE). Delft, Netherlands. IEEE Xplore, Sep 2020.
- [4] Feng, X., L.; Hu, X., C.; Fan, K., Y.; Yu, T.; "A Brief Discussion About the Impact of Coronavirus Disease 2019 on Teaching in Colleges and Universities of China", 2020 International Conference on E-Commerce and Internet Technology (ECIT), Zhangjiajie, China, IEEE Xplore. July 2020.
- [5] Hassonah, M., A.; Al-Sayyed, R.; Rodan, A.; Al-Zoubi, A., M.; Aljarah, I.; Faris, H.; "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter", Knowledge-Based Systems, Elsevier Vol. 192, March 2020.
- [6] Tubishat, M.; Abushariah, M., A., M.; Idris, N.; Aljarah, I.; "Improved whale optimization algorithm for feature selection in Arabic sentiment analysis", Springer, Nov 2018.
- [7] Guellil, I.; Azouaou, F.; Mendoza, M.; "Arabic sentiment analysis: studies, resources, and tools", Social Network Analysis and Mining, Springer, Vol. 9, No. 56, Sep 2019.
- [8] Long, Z.; Alharthi, R.; El Saddik, A.; "NeedFull – a Tweet Analysis Platform to Study Human Needs During the COVID-19 Pandemic in New York State". IEEE Access, Vol. 8, July 2020.
- [9] Prakash, T., N.; Aloysius, A.; "A Comparative study of Lexicon based and Machine learning based classifications in Sentiment analysis", International Journal of Data Mining Techniques and Applications Vol. 8, No. 1, pp. 43-47, June 2019.
- [10] Ahmad, M.; Aftab, Sh.; Muhammad, S., S.; Ahmad, S.; "Machine Learning Techniques for Sentiment Analysis: A Review", International Journal of Multidisciplinary Science and Engineering, Vol. 8, No. 3, pp. 27-32, April 2017.
- [11] 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, Vol. 7, Nov. 2019.
- [12] Gupta, I.; Joshi, N.; "Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic", Journal of Intelligent Systems, Vol. 29, No. 1, pp. 1611-1625, Sep 2019.
- [13] Lamirel, J., Ch.; Cuxac, P.; Hajlaoui, K.; Chivukula, A., S.; "A new feature selection and feature contrasting approach based on quality metric: application to efficient classification of complex textual data", PAKDD 2013 International Workshops on Trends and Applications in Knowledge Discovery and Data Mining. Pacific-Asia, pp. 367–378, 2013.
- [14] Al Shboul, B.; Al-Ayyoub, M.; Jararweh, Y.; "Multi-Way Sentiment Classification of Arabic Reviews" 2015 6th International Conference on Information and Communication Systems (ICICS). IEEE Xplore, Amman, Jordan. May 2015.
- [15] Oueslati, O.; Cambria, E.; Ben Haj Hmida, M.; Ounelli, H.; "A review of sentiment analysis research in Arabic language". Future Generation Computer Systems, Elsevier. Volume 112, pp. 408-430, Nov 2020.
- [16] AlSalman, H.; "An Improved Approach for Sentiment Analysis of Arabic Tweets in Twitter Social Media", 3rd International Conference on Computer Applications & Information Security (ICCAIS) IEEE Xplore, Riyadh, Saudi Arabia. May 2020.
- [17] Al-Twairish, N.; Al-Negheimish, H.; "Surface and Deep Features Ensemble for Sentiment Analysis of Arabic Tweets". IEEE Access, Vol. 7, June 2019.
- [18] Al-Azani, S.; El-Alfy, E., M.; "Enhanced Video Analytics for Sentiment Analysis Based on Fusing Textual, Auditory and Visual Information", IEEE Access, Vol. 8, July 2020.
- [19] Sethi, M.; Pandey, S.; Trar, P.; Soni, P.; "Sentiment Identification in COVID-19 Specific Tweets", 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE Xplore, Coimbatore, India, Aug 2020.
- [20] Samuel, I.; Rahman, Md., M.; Ali, G. G. Md. Nawaz; Samuel, Y.; Pelaez, A.; Chong, P., H., J.; Yakubov, M.; "Feeling Positive About Reopening? New Normal Scenarios From COVID-19 US Reopen Sentiment Analytics", IEEE Access, Vol. 8, Aug 2020.
- [21] Imran, A., Sh.; Daudpota, S., M.; Kastrati, Z.; Bhatra, R.; "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets", IEEE Access, Sep 2020.
- [22] Wang, T.; Ke Lu; Chow, K., P.; Zhu, Q.; "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model", IEEE Access, Vol 8, July 2020.
- [23] J. Samuel, M. N. Ali, M. M. Rahman, E. Esawi; Y. Samuel, "Covid-19 public sentiment insights and machine learning for tweets classification", Information, Vol. 11, No. 6, pp. 1–21, 2020.
- [24] Mostafa, L.; "Egyptian Student Sentiment Analysis Using Word2vec During the Coronavirus (Covid-19) Pandemic", Proceedings of the International Conference on Advanced Intelligent Systems and Informatics, Springer, pp. 195-203, Mar 2020.
- [25] Al-A'bed, M.; Al-Ayyoub M.; "A Lexicon-Based Approach for Emotion Analysis of Arabic Social Media Content", Proceedings of the International Computer Sciences and Informatics Conference (ICSIC 2016), Amman, Jordan, June 2016.

- [26] Kaila, R. P.; Prasad, A. K.; “Informational flow on twitter – corona virus outbreak – topic modelling approach,” *International Journal of Advanced Research in Engineering and Technology (IJARET)*, Vol. 11, No. 3, pp. 128–134, 2020.
- [27] Al-Ayyoub, M.; Khamaiseh, A., A.; Jararweh, Y.; Al-Kabib, M., N.; “A comprehensive survey of Arabic sentiment analysis”, Elsevier, Sep 2018.
- [28] Alomari, Kh., M.; Kh., M.; ElSherif, H., M.; Shaalan, Kh.; “Arabic tweets sentimental analysis using machine learning”, *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, Springer, pp. 602–610, 2017.
- [29] Khorsheed, M., S.; Abdulmohsen O. Al-Thubaity, “Comparative evaluation of text classification techniques using a large diverse Arabic dataset,” *Language Resources and Evaluation*, Springer, Vol. 47, No. 2, pp. 513–538, March 2013.
- [30] Abdul-Mageed, M.; Diab, M.; Kübler, S.; “SAMAR: Subjectivity and sentiment analysis for Arabic social media” *Computer Speech & Language*. ACM, Vol. 28, No. 1, pp. 20–37, Jan 2014.
- [31] Abbasi, A.; Chen, H.; Salem, A.; “Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums” *ACM Trans. Information Systems*, Vol. 26, No. 3, Article 12, June 2008.
- [32] Sforza, V., C.; Saddiki, H.; Bouzoubaa, K.; Abouenour, L.; “Bootstrapping a WordNet for an Arabic Dialect from Other WordNets and Dictionary Resources” *10th ACS/IEEE Int. Conf. On Computer Systems and Applications (AICCSA 2013)*, Fes/Ifrane, Morocco, May 2013.
- [33] Baly, R.; El-Khoury, G.; Moukalled, R.; Aoun, R.; Hajj, H.; Shaban, K/, B.; El-Hajj, W.; “Comparative Evaluation of Sentiment Analysis Methods Across Arabic Dialects”, *3rd International Conference on Arabic Computational Linguistics, ACLing. in Arabic Computational Linguistics*. Elsevier, Vol. 117, Pp. 266-273, Nov 2017.
- [34] Medhat, W.; Hassan, A.; Korashi, H.;. “Sentiment analysis algorithms and applications: A survey” *Ain Shams Engineering Journal*, Elsevier, Vol. 5, No. 4, pp. 1093–1113, Dec 2014.
- [35] Mohammad, S., M ; “Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text”, *National Research Council Canada* 2014.
- [36] Tan, S.; Li, Y.; Sun, H.; Guan, Z.; Yan, X.; Bu, J.; Chen, Ch.; He, X.; “Interpreting the Public Sentiment Variations on Twitter” *IEEE Trans. Knowledge and Data Engineering*, Vol. 6, No. 1, pp. 1–14, Sep 2012.
- [37] Abu-Farha, I.; ; Magdy, W. ; “Mazajak: An Online Arabic Sentiment Analyser”, *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, Florence, Italy, Association for Computational Linguistics. Pp. 192–198, Aug 2019.
- [38] Cambria, E.; Schuller, B.; Xia, Y.; Havasi, C.; “New Avenues in Opinion Mining and Sentiment Analysis” *IEEE Intelligent System*, Vol. 28, No. 2, pp. 15–21, 2013.
- [39] Maghfour, M.; Elouardighi, A.; “Standard and dialectal arabic text classification for sentiment analysis”, *International Conference on Model and Data Engineering*, pp. 282–291. Springer, Sep 2018.
- [40] Bahassine, S.; Madani, A.; Al-Sarem, M.; Kissi, M.; “Feature selection using an improved chi-square for Arabic text classification”, *Journal of King Saud University-Computer and Information Sciences*, Vo 32, No 2, pp. 225–231, Feb 2020.
- [41] Joseph, D., P.; Taghi, M., Kh.; David, J., D.; “Impact of Feature Selection Techniques for Tweet Sentiment Classification” *Proceeding of the 28th International Florida Artificial Intelligence Research Society Conference*. pp. 299–304, 2015.