

Arabic Tweets Sentiment Analysis using Hybrid Approaches

Manal Essam

Department of Computer Science,
Faculty of Computers and
Information
Menoufia University, Egypt

Mohamed Elmenshawy

Department of Computer Science,
Faculty of Computers and
Information,
Menoufia University, Egypt

Hamdy M. Mousa

Department of Computer Science,
Faculty of Computers and
Information,
Menoufia University, Egypt

ABSTRACT

The unrestrained-access to social media makes people share their daily life; a Twitter platform allows its users to openly express and share their emotions about several issues in a predefined length. Thus, it becomes one of the most dominant networks in Arabic countries. Therefore, the Sentiment Analysis of Arabic tweets is a practical task of analyzing common sentiments and feelings. However, the existing resources regularly focus on the English language due to the shortage of Arabic Sentiment resources.

In this paper, a new sentence-based sentiment analysis system had developed for Arabic tweets. Initially, the main sentiment classification approaches had applied for the sentence-level to obtain the most suitable one. As a result, the steps towards the construction of a new dataset had evaluated. The experiments show that the supervised approach is the most accurate one, especially with the absence of the Arabic dialects' (informal) lexicons. Experiments comparisons achieve satisfactory results with high accuracy (78.08%) by supervised approach, Unsupervised gives acceptable accuracy (75%), and F-measure (74.1%) using a Hybrid classifier.

General Terms

Machine learning, Natural language processing, Data analysis

Keywords

Sentiment Analysis, Arabic social media, Twitter sentiment analysis, Arabic sentiment analysis, aspect-based sentiment analysis, machine learning.

1. INTRODUCTION

Twitter becomes habitual through Arabic countries, like other social media platforms. Thus, the number of Arabic Users Generated Content is expanding (UGC) [1] based on social media report [2], there are approximately 11 million active Arabic users of Twitter, and tweets exceed 849 million. Consequently, the tweets are commonly writing about daily life events, social issues, TV-based shows, and considering a key stimulus of most social actions [3]. For example, Arab spring tides have deemed a case in point showed the significant role of Twitter, where Arabic users used it to share approximately 40-45 tweets per minute [4] against the Egyptian government. Other examples include popular Arabic hashtags during this duration were #jan25, #Egypt, #Libya, and #protest [5]. Thus, Twitter considers a valuable source of opinions that coincides with the Sentiment Analysis (SA) process for classifying and predicting sentiment directions of texts. SA process is a multidisciplinary field of study, which deals with unstructured text to extract opinions and classify them as conveying either a negative or a positive sentiment

using computational methods [6]. Thus, SA is one of the most active research domains of Natural Language Processing (NLP) and Machine Learning (ML). However, Arabic Sentiment Analysis (ASA) process suffers from low scholarly research efforts and a lack of resources due to the Arabic language nature. Arabic is a complex morphology [7] where one word has multiple forms, and it has a diglossia, where formal language uses in writing thoroughly different from spoken [8]. Formal language knows as Modern Standard Arabic (MSA); it applies in academic books and media. On the other side, the spoken form knows as an informal language or Dialectal Arabic (DA); it has multiple dialects that differ from one country to another [9]. For that, the DA form is broadly used in social media [4] that sometimes affects the SA process performance. Besides, Arabic tweets have particular features such as colloquial expressions, abbreviations, and slang words that repeatedly find in tweets. These expressions change permanently, and new terms emerge, sometimes hold sentiment polarity, and may cause drawbacks classification approaches if not considered. Previous research efforts on ASA have focused on MSA, but recently efforts address the DA forms. Also, ASA worsened by the fact that most tweets share in an unstructured format due to the limited size of tweets characters [9]- [4].

This paper focused on classifying Arabic tweets because most of the existing sentiment analysis resources, whether participating in the workshop or, not are monolingual and usually concentrate on English texts [10]. For that, a new SA system had developed to deal with ASA challenges using main machine learning classification approaches (supervised and unsupervised).

2. LITERATURE REVIEW

The sentiment Analysis process usually designed for evaluating single form languages, especially English. Thus, the SA field needs data analysis for several languages, like Arabic. The Arabic language suffers from the absence of language standardization outside the academic and media contexts, diglossia phenomena, and the complicated morphology. Such challenges of the Arabic language led to hinder the SA task, prevent it from achieving mature, and lacks its resources. Also, Arabic social media data had covered well, so works conducted through the Arabic language had covered. According to SA levels (document, sentence, or aspect, classification approach used (supervised, unsupervised, or hybrid) and the language (MSA, or DA).

The authors in [11], collected 500 movie reviews from Arabic webpages, reviews annotated as a positive or a negative label. Thus, they have drawn on ML algorithms to develop their classifier. Also, they used the salient features on the cleaned

data to train the classifier. The classifier gives a promising result but didn't consider Arabic dialects and evaluated on document level, and their dataset was relatively limited.

In paper [3], the authors developed a hybrid approach for the SA process. They collected 1,103 Saudi tweets for classification. The tweets were cleaned from unwanted data. Then they used SentiWordNet for translating Arabic sentiment words to label tweets as either positive, neutral, or negative. The feature vector was used to train the SVM classifier. For the research limitation, sentiment classification had based on the accuracy of translated terms and only considered the MSA form.

The authors in [8], have developed a system called SAMAR for examining subjectivity and sentiment analysis (SSA) tasks. SSA task used for determining whether a text is objective or subjective, then classifying it into sentiment value. The system helps in collecting the data from various websites genres: Web Forums, Chat, Wikipedia Talk Pages, and Twitter. The system used the SVM light method for classification. Moreover, the authors focused on showing how it is complex to implement ASA. For this reason, the paper dealt with MSA and didn't give any accuracy for classification results.

In the paper [12], the authors have evaluated the unsupervised approach for ASA. Firstly, they built their lexicon list extracted from 380 seed words and then extended it. They used two algorithms to label these words. Then they merge two algorithms to calculate the overall SO of tweets retrieved from Twitter. The first algorithm sums the SO of each word in a tweet, and the other is the double sentiment polarity, where each word has negative and positive sentimental weight. This study used a small number of seed words for classification.

In [13], the authors collected 2400 comments from 220 Egyptian posts on Facebook to build their dataset. The dataset was manually annotated as attacking, supportive, and neutral according to the content of the text. Then comments had preprocessed by eliminating unwanted data like punctuation and stop words. Also, long comments that exceed 150 words were excluded. For feature extraction, they focused on the similarity between posts and comments addressed. For the classification of sentiment, they used three classifiers: SVM, NB, and DT. The SVM classifier achieved the best results concerning the extracted features. But, this paper highlights only the MSA form.

The authors in [14], developed an ASA system using human computation. The system was divided into two phases. The first is a game designed to label restaurant reviews collected from Qaym.com by game users. The users identify words of texts as positive, negative, or neutral for constructing a lexical list. Besides, the user classified the overall sentence. The second is a sentiment analyzer. The analyzer used two methods to determine the sentiments of reviews. The first one used the game extracted patterns to perform matching of sentence patterns to detect the polarity of the whole sentence. The second analyzes a text according to the lexical terms. Thus, such a study achieved low accuracy when depending on

human computation only for classification.

The authors in [15], provide a human-annotated Arabic dataset, called HAAD. The process of building HAAD is carried out in three steps: data collection, data annotation, and data annotation format. In the data collection step, texts were selected from Arabic reviews of the LABR dataset by seven groups of three graduate students. Then in the annotation, the students annotate the reviews for four subtasks (aspect extraction, aspect polarity, aspect category, and aspect category polarity). For the annotation format, the dataset puts in XML format for research. For the performance evaluation, the F-measure and accuracy methods had calculated. This research only focuses on MSA reviews and achieves low accuracy results.

In paper [5], the authors proposed a technique to identify the 'opinion target' of Arabic tweets. The system had executed in three stages. Primary, they collected tweets and manually annotated them. In the first phase: collected tweets had processed to remove unwanted data in three steps. Then, the system extracted the POS tags of words, named entities, and POS patterns. In the final phase: three ML classifiers trained using the extracted features to determine each tweet is either an opinion target or not. These classifiers are SVM, NB, and K-NN. The highest F-measure had achieved by using the K-NN method. This paper only classifies tweets to target or not without considering the second task of the ABSA process.

In paper [16], the authors present the Sem-Eval 2016 Aspect Based Sentiment Analysis (ABSA) shared task, which considered the extent task of the 2014 and 2015 versions. In the third version, it presented 19 training and 20 testing datasets for seven domains and eight languages. Besides, it introduces a standard evaluation procedure for system evaluating and testing. From the shared datasets, there were 25 for sentence-level and 14 for text-level ABSA and one dataset for the out-of-domain subtask at the start of evaluation duration. As a result, the task engaged 29 interested teams with 245 entries.

3. TWEETS SENTIMENT ANALYSIS

This research presents a new SA system for Arabic tweets that hold specific polarity (see Fig 1). Therefore, the classification approaches (supervised, unsupervised, and hybrid) had applied for investigating the efficient classifier predicts tweets and deal with the complex morphology nature of the Arabic language. In the supervised, the machine learning methods had based on the feature vector for classification. Thus, the system had utilized the N-gram, Term Frequency (TF), and term occurrence features for learning the supervised classifiers and predicting the sentiment outcome efficiently. In the unsupervised approach, the Arabic dialects lexicon list had applied to calculate the Semantic Orientation (SO) of tweets. Then, in the hybrid, the two previous methods had merged for analyzing the tweets that fail in the unsupervised, reducing the time and effort in labeling tweets and giving more accurate results with ML. Firstly, tweets had retrieved using Twitter API, clean and annotate tweets to build up the dataset due to a lack of Arabic resources [7] [1].

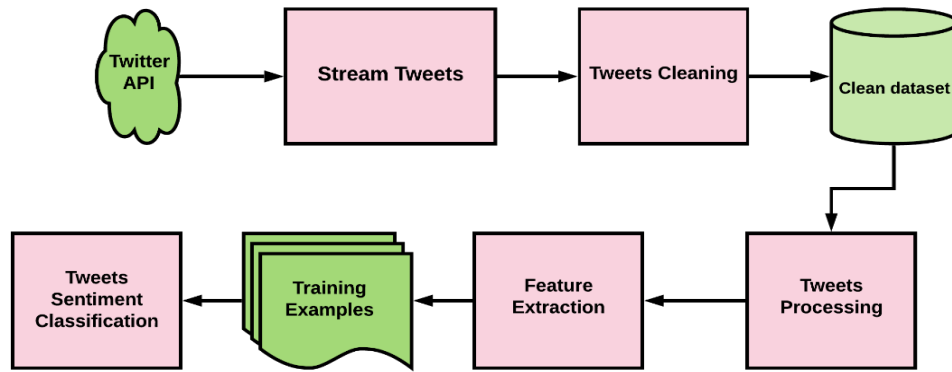


Fig1: Sentiment Analysis Workflow

3.1 Data Collecting

Twitter API service had used to crawl tweets for dataset generation (see Fig1), which allows retrieving a stream of tweets in a specific language. An R script had developed to gather tweets using language (“ar”) statements through the twitter-R library by creating a search query using top Arabic trends to enhance the chance of receiving tweets that hold sentiments from the trends Twitter hashtag like #Egypt. The tweets had retrieved from January 2020.

3.2 Data Cleaning and Annotating

The textual content (tweets) of Twitter usually comes in an unstructured format. Therefore, these data lead to low accuracy in classification results. As a result, tweets need to enhance data value by applying the cleaning functions that developed in R language to decrease the useless of non-sentiment data (see Table 1) using TM and arabicStemR R packages and ASAP utilities.

Table 1. Tweets Cleaning Functions

Function	REG
Strip URL	"(f htp(s?):/(.)*[a-z]+)"
Strip Retweets, Mentions	"(RT via)((?:\\b\\W*@[\\w+])+)")
Strip HTML	"&"
Strip Usernames	"@[\\w+]"
Strip Numbers	"[:digit:]"
Strip Latin Characters, Foreign words, Hashtags and Emotions. Remove Duplicated Tweets and Words.	TM and Arabic Stemmer Package, Functions

Then, tweets had annotated for SA tasks training and testing. According to previous SA studies [18], a sentiment polarity had defined as positive or negative. Each tweet was labeled with only one label using the dominant polarity concerning the writer’s perspective and linguistic guidelines by three Arabic native speakers. In cases where we disable to determine the polarity of tweets or don’t understand the dialect, they had excluded. Also, the ambiguity and sarcastic tweets had eliminated. Finally, we construct an Arabic tweet

dataset consisting of more than 2000 tweets (see Table 2) that shows the clean dataset examples. The output of this phase is a clean dataset.

Table 2. Clean Tweets Example

ORIGINAL TWEET	RT @NanaAnes1: مهاره عاليه عاليه من التنه في نيجريا
CLEAN TWEET	RT @NanaAnes1: مهاره عاليه عاليه من التنه في نيجريا
ANNOTATED TWEET	Negative: بيچندوا طلبه في ثانوي علشان يبلغوا عن صاحبهم حد سمع عن الموضوع ده
	Positive: كلنا اهلي وزمالك وراء المنتخب
	Uncertain: بتجري تسمع الاغنيه

3.3 Preprocessing

The preprocessing phase consists of NLP techniques to structure and prepare the dataset for classification experiments by using the clean dataset (see Fig1). First, the dataset represents with UTF-8 encoding. In the tokenization step, tweets had divided into individual words (tokens) by using white spaces. For normalization, tweets transformed into a consistent form by unifying some Arabic characters such as (أ، إ، ي، ه، ة). Then, the filtration step eliminated the Arabic stop words because these words don’t convey sentiment words like (من – علي – في – انت – انا). After that, the step had completed by the stemming process that performs by two methods: rooting (full stemming) and light stemming. The rooting converts the word into its root by eliminating prefixes, infixes, and suffixes. Thus, it reduces the data size by allocating several terms to the same stem. On the other side, the light stemming saves words meaning by eliminating the prefixes and suffixes. Therefore, it is a simple method. The Arabic language has two familiar stemmers: Khoja stemmer [18] for full stemming and light stemmer [19] for light. Thus, the previous stemming techniques had applied to evaluate their performance tweets Dialects nature. The RapidMiner and R packages had used for preparing data instances. Then, processed tweets fit into the extraction phase (see Table 3).

Table 3.Preprocessing Tweet Example

CLEANED TWEET	بيجندوا طلبه في ثانوي علشان ييلغوا عن صاحبهم حد سمع عن الموضوع ده
TOKENIZATION	"بيجندوا" "طلبه" "في" "ثانوي" "علشان" "ييلغوا" "عن" "صاحبهم" "حد" "سمع" "عن" "الموضوع" "ده"
FILTRATION	"بيجندوا" "طلبه" "ثانوي" "علشان" "ييلغوا" "صاحبهم" "الموضوع"
STEMMING	"جند" "طلب" "ثان" "علش" "بلغ" "صاحب" "موضوع"

3.4 Feature Extraction

This process is recognized as the first and fundamental task when one wants to present instances (training examples), the feature vectors, usually applied to make textual data suitable for computational (ML) methods [6]. Once features are selected, the data represented using the extracted features. For the work at hand, the features picked from the data itself to avoid Arabic data conflicting. These features usually include n-gram words (e.g., unigram/bigram and trigram), term frequency, term occurrences, and sentiment words and phrases. The syntactic features present textual data as a series of grams. For example, a gram of size 1 refers to a unigram; size 2 is a bigram; 3 is a trigram, and so on [20]. Besides, the term frequency weighting scheme (TF) assigns a weight to each word in document bases on the occurrences of the term in the document. Then, the unsupervised approach used sentiment words as a feature for the classification process.

3.5 Sentiment Classification approaches

This part addresses the machine learning classification approaches (supervised and unsupervised) that widely used for the SA process [20] to evaluate their performance, especially when dealing with the informal form of the Arabic language.

3.5.1 Supervised

Machine Learning (ML) is a process by which computers able to learn from a set of training examples, and ML methods can predict based on labeled data [20]. Thus, SA uses supervised methods to classify data into sentiment polarity. This approach requires a labeled dataset with sentiment labels (positive or negative) (see Table 2 and Table 3) fed into ML classifiers to construct a classification model that predicts new examples. In experiments, the extracted features had used for learning ML methods and selecting better features for each classifier. For testing and validation, the cross-validation method had applied, which is a statistical method for estimating and comparing training methods [19]. The standard form of cross-validation is the k-cross-validation, where k is the number of segments of data. Thus, it builds random subsets of testing data and measures the performance of each experiment by calculating the performance metrics. Then, K repetition of subsets had carried out so that, within every repetition, various segments caught for validation and the remaining utilized for learning. The ML attempts had carried out using the RapidMiner environment. RapidMiner software provides robust processing techniques for the Arabic language and provides an environment for evaluating ML algorithms.

3.5.2 Unsupervised

Various algorithms had developed to investigate the unsupervised approach for ASA to avoid time and effort consuming manual annotation. Thus, this approach had

generally based on a dictionary of opinion words (lexicons) to determine and extract the sentiment polarities (positive and negative) of the unlabeled dataset (Clean tweets) (see Table 2). For that, the lexicon-based tool had developed to analyze the sentiment polarity of tweets using the R language. The methodology for building our lexicon-based analyzer (see Fig 2). The classifier receives Arabic tweets and gives tweets with a sentiment polarity value. For analyzer evaluation, we based on two dialect lexicon lists that present positive and negative terms for the classification process. In the sentiment classification step, the analyzer corresponds between the word in the tweet and dictionary if it finds in the positive lexicon or negative. Then, the accumulative score computed using matching opinionated words in a tweet to predict its sentiment class (see Table 4) according to the following equation:

$$Score = \sum_{i=1}^n (Wp_i - Wn_i) \quad (1)$$

Wp is the number of positive words, and Wn is the number of negative ones. Finally, the classifier calculates the polarity of each tweet by summing the number of opinion words in each tweet. But some tweets don't label with the right polarity comparing to the original labeled dataset due to the limited size of the lexicon dictionary words. As a result, the tool fails to specify appropriate sentiment polarity for tweets that are free from any sentiment words belong to the dictionary.

INPUT: The Tweets text file (T), The sentiment words (Lexicons) L
OUTPUT: Sentiment orientation (SO) = Positive (P) or Negative (N)
INITIALIZATION: Sum = 0, where sum: calculates the sentiment weight for each tweet (t) in the dataset
Begin
 For each w in (T) do
 Calculate the score of positive and negative words in(T)
 Score = sum (positive words) – sum (negative words)
 If score > 0 then
 SO = positive
 Else
 SO = negative
 End if
 End for
End

Fig 2: Lexical Analyzer algorithm

Table 4 shows an example of the lexical analyzer input and output.

Table 4.Lexical Analysis Example

Tweet Dataset (Input)	Sentiment	Weight Score	Predictable Sentiment (Output)
انا متحمس وانا ما تابعته والله شكله مسلسل اسطوري	Positive	2	Positive

3.5.3 Hybrid

From previous approaches observation, the hybrid approach had investigated as a classifier to avoid some faults. The hybrid approach merges the lexical-analyzer with ML methods (see Fig 3). Therefore, it solves the supervised limitations like a manual annotation that requires time and

effort. On the other side, the unsupervised approach needs a big size dictionary to achieve good results. Thus, the lexicon list had expended with sentiment terms to improve the accuracy, but again it mislabels some tweets. Therefore, the hybrid approach had suggested to use the advantages and trying to avoid some limitations. The hybrid approach applied the lexicon-based classifier to label tweets alternatively to the manual annotation process. Therefore, it saves time and effort consuming. Then, the predicted dataset had utilized as training data for supervised to construct a sentiment classification model. Then, the classification model extracts the SO of unlabeled data that failed in the lexical analyzer.

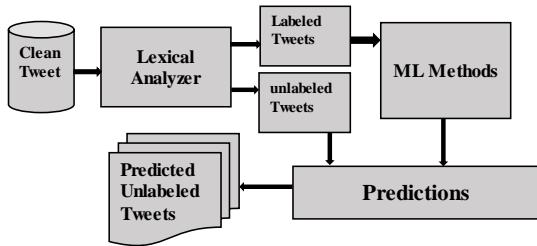


Fig 3: Hybrid Approach Workflow

4. EVALUATION CRITERIA

The evaluation conducted through the confusion matrix that contains four inputs: TP, TN, FP, and FN. TP is the correct predictions of the positive examples, TN is the negative samples, FP is the false predictions of positive samples, and FN is the wrong negative examples. Where: as Accuracy (Acc), Precision (Pre), Recall (Rec), and F -measure (F) calculated:

- Accuracy (ACC) is the percentage of the testing set of items that are classified correctly via the classifier. It calculated as follows:

$$Acc = \frac{TP + TN}{P + N} \quad (2)$$

- Precision (Pre) has defined as a measure of exactness (i.e., what the percentage of items labeled as positive are accurate). It can calculate as follows:

$$Pre = \frac{TP}{TP + FP} \quad (3)$$

- Recall (Rec) is a measure of completeness (i.e., what the percentage of negative items labeled as such). It can calculate as follows:

$$Rec = \frac{TP}{TP + FN} = \frac{TP}{P} \quad (4)$$

- F -measure (F) gives the equal weight to the precision and recall. It can calculate as follows:

$$F = \frac{2 \times Pre \times Rec}{Pre + Rec} \quad (5)$$

4.1 Supervised Results

In the supervised approach, tweets had classified using SVM and NB methods. In the attempts, the different features had extracted to construct word vectors, such as N-gram, TF, and term occurrences, with two stemming techniques (light and full) to achieve good results for the Arabic tweets' nature (see Table 5 and Table 6).

Table 5. SVM Method Results

	N-gram		Morphology Analysis		Schemes	Results			
	1	2	Light	Full		Acc	Pre	Rec	F
Method (SVM)					TF_IDF	78.08	77.87	79.59	78.66
					TF_IDF	73.14	67.46	91.02	77.49
					TF	77.26	74.99	83.13	78.78
					TF	72.36	65.24	89.82	75.60
					Term occurrences	74.49	70.04	87.33	77.67
					TF	76.98	74.74%	82.80	78.49
					Binary occurrence	75.69	72.08	85.46	78.13
					TF-IDF	67.85	62.69	90.49	74.05
					TF	72.26	68.93	82.45	75.08
					TF	67.89	62.65	90.78	74.14

From experiments results noticed that SVM advances the performance accuracy of NB. Besides, unigram features had gained better accuracy and precision within two methods due to its good data coverage without considering the word position. TF-IDF works well with SVM, and this gets similar to this paper [11] though it works on document formal language. Also, term occurrences worked well with two methods.

Table 6. NB Method Results

	N-GRAM		MORPHOLOGY ANALYSIS		SCHEMES	RESULTS			
	1	2	Light	Full		Acc	Pre	Re	F
NB					TF_IDF	73.48	70.82	81.20	75.63
					TF	74.80	70.93	85.46	77.48
					TF_IDF	67.4	64.1	83.04	72.32
					TF	72.81	68.47	86.40	76.33
					TF	71.59	67.07	84.7	74.90
					Binary term occurrences	76.34	76.64	76.99	76.74
					Term occurrences	76.18	76.26	77.18	76.65

The bigrams achieved better recall results within two methods because bigram features able to capture the Arabic expressions, for example (مش هينفع) this state by bigram considered direct negative opinions, so it gives high results with negative examples. Also, full stemming works well in our case with two methods because the stemming technique works to deal with the complex nature of Arabic Morphology by splitting words into small units. Thus, stemming less the sparsity in the dataset examples.

4.2 Unsupervised

For the unsupervised approach, our proposed lexical analyzer has applied. Table 7 and Table 8 illustrated that the classifier experiment1 missed predicting tweets correctly and gave low accuracy compared to supervised due to the limited size of the lexicon dictionary in experiment 1. Though, after we appended lexicons with more newly trendy opinionated words in experiment 2, the classifier obtained satisfying results of 75%.

Table 7. Lexical Analyzer prediction Experiments

Prediction	Pos. Ex1	Neg.Ex1	Pos. Ex2
<i>positive</i>	518	349	755
<i>Negative</i>	332	521	95

Table 8. Lexical Analyzer Experiments Results

	Acc.	Pre.	Rec.	F
<i>Experiment 1</i>	60.4	61.07	60.9	60.47
<i>Experiment 2</i>	75.1	79.0	80.9	77.4

The unsupervised experiments (see Table 8) shown that maximize the lexicon dictionary helps the classifier to perform well, especially for positive tweet directions, but Arabic social media slang terms change consequently and updated continuously. Thus, we try to classify tweets that get zero scores by using the hybrid approach.

4.3 Hybrid

This classifier merges two previous approaches to classify the unlabeled tweets in a lexical analyzer by using SVM and NB. From experiments (see Table 9 and Table 10), the Bayes theorem works well more than SVM with a small number of instances. Table 11 showed that the classifier predicts instances correctly with more than 66 probability. Also, positive tweets had classified correctly more than negative.

From classifiers results (see Fig4), supervised approach works well for classifying positive and negative tweets and able to deal with Arabic tweets sparsity nature, especially (SVM classifier through TF_IDF and full stemming).

Table 9. Lexical Analyzer and SVM results

	True Negative	True Positive	Class precision
Pred. Negative	128	59	66.67%
Pred. Positive	217	369	64.66%
Class recall	37.75%	87.03%	

Table 10. Lexical Analyzer with NB Results

	True Negative	True Positive	Class precision
Pred. Negative	219	155	58.56%
Pred. Positive	126	300	70.24%
Class recall	63.48%	65.39	

Table 11. Hybrid Average Results

	Average (ACC)	Average(F-measure)
<i>NB</i>	64.88	68.1
<i>SVM</i>	65.50	74.13

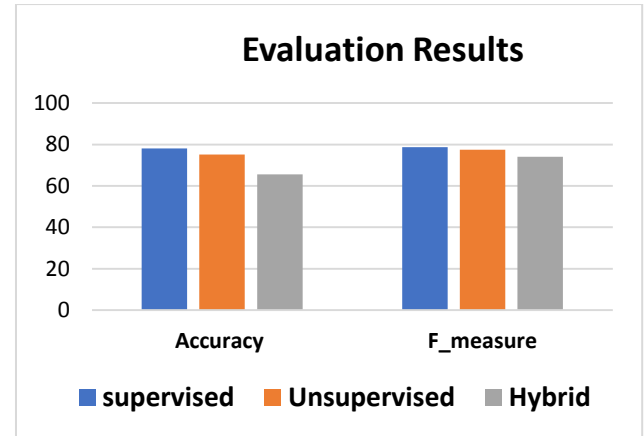


Fig4: Classifiers Evaluation Results

5. CONCLUSION

In this work, we have generated a dataset of Arabic tweets for sentiment analysis tasks. For that, steps for building up, cleaning, and annotating the dataset of SA tasks explicated. Also, we have addressed three sentiment classification approaches for predicting Arabic tweets. The experiments are carried out through different approaches to showcase and observe the developing degrees of classifiers' accuracy. It had noticed that SVM with full stemming and TF-IDF scheme gives the highest accuracy rate in the supervised approach reached 78%. For the unsupervised—after we appended the lexicon list—our analyzer achieves a reliable result reached 75% with the independent-domain dataset, but it fails to predict some tweets polarities values. In general, the supervised approach is better; it gives accurate results and able to predict most of the tweets correctly; because of the absence of dialect lexicon lists. As a result, we examined the hybrid approach as an alternative to avoid some defects of two classifiers and improve the classification rate. The hybrid achieves satisfying results where accuracy is 64.33%, and f-measure is 73.45%.

6. ACKNOWLEDGMENTS

Our My deep thanks are due to Dr. Hamdy M. Mousa and Dr. Mohamed El-Menshawy, my excellent supervisors, who supported me, answered my questions, and provided me with lots of facilities. I have reached with their supports this phase.

7. REFERENCES

- [1] M. Abdullah and M. Hadzikadic, "Sentiment analysis on arabic tweets: Challenges to dissecting the language," in International Conference on Social Computing and Social Media, pp. 191--202, 2017.
- [2] "Arab knowledge economy report 2016," 19 3 2016. [Online]. Available: <https://www.arabknowledgeeconomyreport2016.COM>. [Accessed 20 6 2018].
- [3] H. K. Aldayel and A. M. Azmi, "Arabic tweets sentiment analysis--a hybrid scheme," Journal of Information Science, vol. 42, pp. 782--797, 2016.
- [4] S. O. Alhumoud, M. I. Altuwaijri, T. M. Albuhaire and W. M. Alohaideb, "Survey on Arabic Sentiment Analysis in Twitter," World Academy of Science, Engineering and Technology, pp. 1--6, 2015.
- [5] M. Alhazmi and N. Salim, "Arabic opinion target extraction from tweets," ARPN Journal of Engineering and Applied Sciences, vol. 10, pp. 1023--1026, 2015.

- [6] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, New York, USA: Cambridge University Press, 2015.
- [7] N. Al-Twairish, H. Al-Khalifa and A. Al-Salman, "Subjectivity and sentiment analysis of Arabic: trends and challenges," in *11th International Conference on Computer Systems and Applications*. pp. 148--155, Doha, Qatar, IEEE Computer Society, 2014.
- [8] M. Abdul-Mageed, M. Diab and S. Kubler, "SAMAR: Subjectivity and sentiment analysis for Arabic social media," *Computer Speech & Language*, vol. 28, pp. 20--37, 2014.
- [9] M. Essam, M. ElMenshawyl and H. Mousal, "A survey on Sentiment Analysis for Arabic Social Media: Current Status and Future Work," *IJCI*, pp. 1--15, 2020.
- [10] J. Wagner, P. Arora, S. Cortes, U. Barman, D. Bogdanova, J. Foster and L. Tounsi, "Aspect Based Sentiment analysis SemEval-2014 Task 4," *Asian J. Comput. Sci. Inf. Technol.*, pp. 72--75, 2014.
- [11] M. Rushdi-Saleh, M. T. Martn-Valdivia, L. A. Urea-Lpez and J. Perea-Ortega, "OCA: Opinion corpus for Arabic," *Journal of the American Society for Information Science and Technology*, vol. 62, pp. 2045--2054, 2011.
- [12] S. R. El-Beltagy and A. Ali, "Open issues in the sentiment analysis of Arabic social media: A case study," in *International Conference on Innovations in Information Technology (IIT)*, pp. 215-220, 2013.
- [13] A. E.-D. A. Hamouda and F. E.-z. El-taher, "Sentiment Analyzer for Arabic Comments System," *International Journal of Advanced Computer Science and Applications (Int. J. Adv. Comput. Sci. Appl)*, vol. 4, no. 3, pp. 99--103, 2013.
- [14] A. A. Al-Subaihin, H. S. Al-Khalifa and A. S. Al-Salman, "A proposed sentiment analysis tool for modern arabic using human-based computing," in *Proceedings of the 13th international conference on information integration and web-based applications and services*, 2011, pp. 543--546.
- [15] M. Al-Smadi, O. Qawasmeh, B. Talafha and M. Quwaider, "Human annotated arabic dataset of book reviews for aspect based sentiment analysis," in *2015 3rd International Conference on Future Internet of Things and Cloud*, 2015, pp. 726--730.
- [16] Pontiki, M. Al-Ayyoub, Y. Zhao, B. Qin, D. Clercq and others, "Semeval-2016 task 5: Aspect based sentiment analysis," in *10th International Workshop on Semantic Evaluation (SemEval 2016)*, 2016.
- [17] M. A. Russell, *Mining the social web: data mining Facebook, Twitter, LinkedIn, Google+, GitHub, and more*, O'Reilly Media, Inc., 2013.
- [18] M. Korayem, D. Crandall and M. Abdul-Mageed, "Subjectivity and sentiment analysis of arabic: A survey," *International conference on advanced machine learning technologies and applications*, pp. 128--139, 2012.
- [19] D. K. Tayal and S. K. Yadav, "Sentiment analysis on social campaign "Swachh Bharat Abhiyan" using unigram method," *AI & SOCIETY*, vol. 32, pp. 633--645, 2017.
- [20] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool, 2012.
- [21] M. Al-Ayyoub, S. B. Essa and I. Alsmadi, "Lexicon-based sentiment analysis of Arabic tweets," *IJSNM*, vol. 2, pp. 101--114, 2015.
- [22] A. M. Alayba, V. Palade, M. England and R. Iqbal, "Improving sentiment analysis in Arabic using word representation," in *IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)*, IEEE, pp. 13--18, 2018.
- [23] A. Shoukry and A. Rafea, "Sentence level arabic sentiment analysis," in *Collaboration Technologies and Systems (CTS)*, 2012 International Conference, pp. 546--550, 2012.