

DOI: [https://dx.doi.org/10.21123/bsj.2020.17.1\(Suppl.\).0385](https://dx.doi.org/10.21123/bsj.2020.17.1(Suppl.).0385)

## A Hybrid Method of Linguistic and Statistical Features for Arabic Sentiment Analysis

Ahmed Sabah Ahmed AL-Jumaili<sup>1\*</sup>

Huda Kadhim Tayyeh<sup>2</sup>

Received 24/2/2019, Accepted 16/7/2019, Published 18/3/2020



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

### Abstract:

Sentiment analysis refers to the task of identifying polarity of positive and negative for particular text that yield an opinion. Arabic language has been expanded dramatically in the last decade especially with the emergence of social websites (e.g. Twitter, Facebook, etc.). Several studies addressed sentiment analysis for Arabic language using various techniques. The most efficient techniques according to the literature were the machine learning due to their capabilities to build a training model. Yet, there is still issues facing the Arabic sentiment analysis using machine learning techniques. Such issues are related to employing robust features that have the ability to discriminate the polarity of sentiments. This paper proposes a hybrid method of linguistic and statistical features along with classification methods for Arabic sentiment analysis. Linguistic features contains stemming and POS tagging, while statistical contains the TF-IDF. A benchmark dataset of Arabic tweets have been used in the experiments. In addition, three classifiers have been utilized including SVM, KNN and ME. Results showed that SVM has outperformed the other classifiers by obtaining an f-score of 72.15%. This indicates the usefulness of using SVM with the proposed hybrid features.

**Key words:** Arabic Sentiment Analysis, KNN, ME, POS Tagging, Sentiment Analysis, Stemming, SVM, TF-IDF.

### Introduction:

The exponential growth of textual information over web especially with the emergence of social media, an essential demand to analyze such information has emerged(1). Arabic language was one of the languages that caught great extent of such textual expansion due to approximately 100 million users who used such language(2). The recent research efforts have introduced Sentiment Analysis (SA) as a major task for analyzing the Arabic text over the social media. SA is the task of analyzing text and identifying the polarity of opinions within such text whether positive or negative(3). Apparently, not all text would have opinion where in many cases social network users are expressing their minds with facts rather than opinions.

Therefore, SA usually is subjected to a preliminary task which is the opinion identification in which the text is being examined in terms of holding opinion or not. Fortunately, the last decade has showed great efforts to provide a dataset of collected opinions which facilitate the process of SA for researchers without affording the identification of opinions (4).

Analyzing the opinions would require wide range of approaches such as linguistic, statistical, and semantic. Linguistic approaches refer to the grammatical annotation such adjectives and adverbs that have an important effect on determining sentiments (e.g. سعيد/*glad*, فرح/*happy*, جميل/*beautiful*, etc.)(5). Statistical approaches refer to the identification of frequent terms that imply the opinion (e.g. جيد / *good*, سيء / *awful*, رائع / *awesome*, مذهش / *amazing*, etc.)(6). Finally, semantic approaches refer to the utilization of the terms that have mutual meaning or sense (e.g. سيء / *bad*, قبيح / *ugly*, مستاء / *upset*, etc.)(7). However, machine learning techniques were the broader methods that have been extensively examined in terms of SA. This is due to its capabilities to include linguistic, statistical and semantic aspects. This can be represented by processing a data of sentiments and

<sup>1</sup>Department of Business Information Technology, College of Business Informatics, University of Information Technology and Communications, Baghdad, Iraq.

<sup>2</sup>Department of Informatics Systems Management, College of Business Informatics, University of Information Technology and Communications, Baghdad, Iraq

\*Corresponding author: [asabahj@uoitc.edu.iq](mailto:asabahj@uoitc.edu.iq)

\*ORCID ID: 0000-0002-3443-0725

generate the aforementioned aspects as features which can be used in the training. Such training paradigm refers to a model that is being built to guide the classification of new opinions into positive and negative(8).

Several studies have addressed the SA for Arabic language using different type of machine learning techniques and variety of linguistic and statistical features(9). Yet, there is still a demand for combining robust features from different aspects for the sake of enhancing classification accuracy. Therefore, this paper aims to propose a hybrid method of linguistic and statistical features using multiple machine learning techniques. The linguistic features contain stemming and "Part of Speech" tagging which also known as POS, while the statistical contain the Term Frequency (TF) and Inverse Document Frequency (IDF). Three classifiers are being used including Support Vector Machine (SVM), K-nearest Neighbor (KNN) and Maximum Entropy (ME).

## Related Work

Numerous studies have been presented for Arabic sentiment analysis for instance, Abbasi et al. (10) proposed a feature selection approach for the sake of determining best features for Arabic sentiment analysis. The authors have collected their data from web forums. Consequentially, the authors have applied genetic algorithm along with SVM classifier. Experimental results showed that the grammatical features such as adjectives and adverbs have been ranked as the most accurate features.

Abdulla et al. (8) proposed a semantic-based method for sentiment analysis in Arabic language using a scrapped data from Twitter. Such semantic approach is based on Arabic lexicon that contains numerous adjectives and adverbs with their corresponding polarity whether positive or negative. Eventually, the authors have applied different machine learning techniques including Decision Tree (DT), Naive Bayes (NB) and others.

Soliman et al. (11) proposed a semantic-based method along with SVM classifier for slang Arabic sentiment analysis. Since most of the social network contains informal Arabic words or slang idioms thus, the authors have used a specific lexicon for the slang Arabic terms. Using SVM classifier, the authors have classified sentiments collected from Twitter into positive and negative.

Abdul-Mageed et al. (12) proposed an approach for sentiment analysis in Arabic language based on linguistic features including lemmatization, POS tagging and other orthographical features. Such

linguistic approaches are intended to examine the root of words and the inflectional derivations added to some terms. Using a semantic lexicon, the authors have conducted a mapping between the processed terms and the words within the lexicon in order to classify Arabic sentiments.

Ahmed (9) has built an Arabic dataset of sentiments by gathering news articles from Arabic websites. Along with creating the dataset, the author has built a large-scale lexicon by translating an English lexicon of SentiWordNet. Finally, a classification for sentiments within the collected articles has been performed.

Al-Twairish et al. (13) have proposed a combination of feature engineering and lexicon-based approaches for Arabic sentiment analysis on Twitter. The authors have used the linguistic approaches such as morphological features for extracting significant sentiment characters such as question marks and emotions. In addition, a specific lexicon for Saudi dialect has been initiated in order to improve the semantic aspect. Eventually, an SVM classifier has been used to classify the sentiments into positive, negative and neutral.

Apart from the features, Oussous et al. (14) have focused on the performance of multiple classifiers regarding the Arabic tweets sentiment analysis (specifically for the Moroccan dialect). The authors have proposed an ensemble classification methods where three classifiers have been examined including NB, SVM and ME. The ensemble method has been represented by applying a voting combination for the three aforementioned classifiers.

## Proposed Method

The main architecture of the proposed method contains of three main parts which are data, feature extraction and classification as shown in Fig. 1. Basically, the first part is related to the Arabic corpus that will be used in the experiments which composed of sentiments collected from Twitter written in Arabic language. Second part is the feature extraction where two main categories of features are being used including linguistic and statistical. Linguistic contains stemming and POS tagging, while statistical contains the TF-IDF. Finally, the third part is related to the classification task in which the Arabic tweets will be categorized into positive and negative. Three classifiers have been used including SVM, KNN and ME. Following subsections will discuss each part in further details.

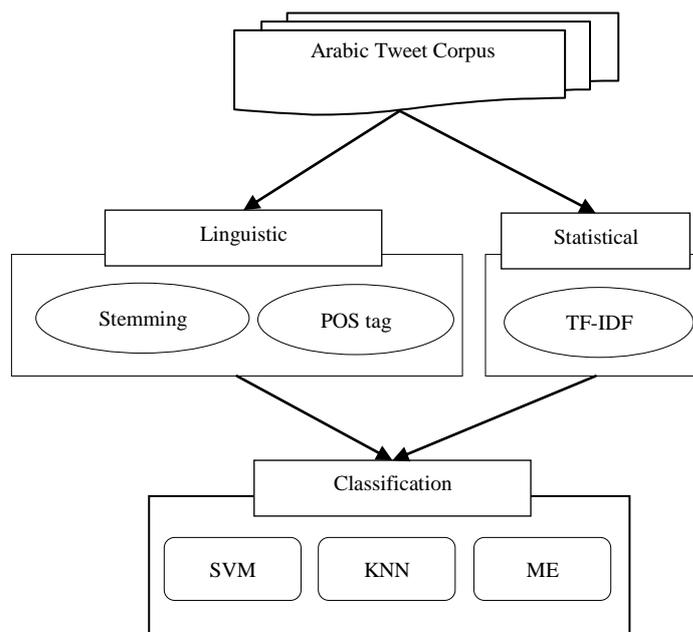


Figure1. Proposed method

### Arabic Tweet Corpus

Prior to perform sentiment analysis, a dataset that contains sentences with opinions should be identified. Therefore, a dataset of Arabic tweets has been used. Such dataset has been collected from Twitter where vast amount of Arabic written tweets have been gathered (15). The dataset contains around 10 thousands tweets with their corresponding class labels. Table 1 shows such classes with their explanation.

Table 1. Explanation of classes

Classes	Explanation
Positive	Tweets that have positive opinions
Negative	Tweets that have negative opinions
Objective	Tweets with no opinion
Neutral	Tweets that have negative and positive opinions at the same time

Table 2 depicts the distribution of tweet categories in accordance to the total number of tweets.

Table 2. Distribution of tweets' categories

Tweets	Quantity
Positive	799
Negative	1684
Objective	6691
Neutral	832
Total	10,006

### Feature Extraction (Linguistic Features)

In this section, the feature extraction task will be discussed especially the linguistic ones. Feature extraction is the process of identifying discriminative characteristic of the sentiments which may facilitate the classifier to predict the

class label of such tweet.

First of all, the stemming which is one of the linguistic features is used which aims to retrieve the roots of each term. For example, the word 'الجيد' / *the good ones*' will be stemmed into 'جيد' / *good*'. This will facilitate the process of identifying frequent adjectives and adverbs that are usually occurred with different derivational inflections such as 'جيدون', 'جيدون', and 'جيدات'. For this purpose, an Arabic stemmer called Khoja stemmer (16) has been used to accomplish this task.

In addition, POS tagging has been applied on each term in order to get the grammatical tags such as verb, noun, adverb and others. This will facilitate the process of identifying the adjectives and adverbs that have frequent occurrence with most of the opinions. For this purpose, an Arabic POS tagging tool introduced by (17) has been used to get the grammatical tags of the terms.

On the other hand, the statistical feature of TF-IDF will be applied by identifying the frequency of each term within the dataset. In fact, the frequency of terms is a significant mechanism that may imply the polarity of an opinion. This is because most of the opinions are expressed by terms that are usually used such as 'جيد', 'سيء', 'رائع', 'amazing', 'مزري' / *awful*' and others. The TF-IDF is calculated as follow:

$$W_t = TF(t, d).IDF_t \quad (1)$$

where  $TF(t, d)$  is the frequency of term  $t$  in accordance to a documented, and  $IDF_t$  is the inverse document frequency for the term  $t$ . In other words, the IDF is the quantity of document contains the term  $t$ .

To understand the TF-IDF mechanism when it

applied on the terms, assume a sample of tweets such as in Table 3.

**Table 3. Sample of tweets**

No	Tweets	Translation
1	سأبقى رائعاً معك إلى النهاية إن بقيتني رائعة معي إلى النهاية	I will be an amazing person for you as much as you will
2	الرائع سيظل رائعاً حتى نهاية العمر	The amazing person will remain impressing you until the end
3	كلامك جميل ورائع	Your words are amazing
4	بجد موضوع رائع وأنا أجد انه مناسب ايضاً لهذا الوقت	Seriously this topic is amazing and we need it these days
5	شعور سيء يخالجنني	I have a bad feeling

As shown in Table 3, five tweets have been depicted where four of them have contained the word 'رائع / amazing'. If we apply the term frequency  $TF(t, d)$  it will be 5 where it has occurred twice in the first tweet, once in the second, once in the third and once in the fourth tweet. However, if we want to apply the IDF the total number of tweets will be divided by the number of tweets that contain the word 'رائع / amazing'. In this regard, the frequency of terms will be addressed in total and in accordance to the tweets that contain them.

### Classification

After applying the linguistic and statistical features, three classifiers will be trained on the extracted features using 80% of the tweet dataset and then tested on 20% of the dataset. The testing refers to the ability of the classifier to predict the class label whether positive, negative, objective or neutral. For this purpose, three classifiers have been selected including SVM, KNN and ME. The reason behind such selected classifiers is that they have the most competitive performance in terms of classifying Arabic sentiments (14).

First classifier is SVM which is intended to classify the tweets into their classes based on a hyperplane. Such hyperplane is a margin that separate the tweets categorized under specific class label from the tweets that categorized under other class label(18). First of all, the data will be vectored in x-axis and y-axis, and based on the TF-IDF of each terms within the tweets, every tweet will be represented in a vector. The mechanism of identifying the hyperplane is depicted as:

$$f(\vec{x}) = \text{sgn}((\vec{x} \times \vec{w}) + b) \quad (2)$$

$$= \begin{cases} +1: & (\vec{x} \times \vec{w}) + b > 0 \\ -1: & \text{Otherwise} \end{cases}$$

where  $\vec{w}$  is the vector and  $b$  is the distance between nearest data point from one category and the hyperplane.

On the other hand, k-nearest neighbor or so called KNN is another classifier that utilize the similarity of the testing tweet with previous tweets from the training. In other words, after representing the tweets based on TF-IDF of their terms within the training, KNN will treat the testing tweet that needs to be classified by identifying the most similar tweet from the training(19). If the most similar tweet has been identified, its class label will be used for the testing one. KNN mechanism is very simple in which it relies on an assumption of similar tweets in meaning would definitely yield the same class label. The prediction of KNN will be performed based on the following equation:

$$f(x) = f(\text{Max sim}(x, x_t)) \quad (3)$$

In which  $x$  refers to testing instances, and  $x_t$  refers to training instances where the maximum similarity would lead to use the same class label.

Finally, the Maximum entropy (ME) is one of the statistical classifiers that intended to extend the regression task to include multiclass problems. Regression refers to the process of classifying a specific data into two classes thus, ME has extend the regression to include classifying data with multi-classes. The key characteristic behind ME lies on the assumption that data instances are case specific where every individual variable has a specific value for each case(20). In this regard, ME will address each individual variable by giving a score to generate the prediction model. Such score is calculated as follow:

$$\text{score}(X_i, k) = \mathbb{1}_k \cdot X_i \quad (4)$$

Where  $X_i$  is the vector of explanatory variables illustrating an observation  $i$ , and  $\mathbb{1}_k$  is the vector of weights regarding to a class label  $k$ .

### Results

Prior to the discussion of classification results, it is necessary to mention the evaluation method used in this paper. Since the evaluation will be held only on the testing tweets which correspond to 20% of the data thus, the evaluation will be conducted on whether the classifier has predicted a tweet correctly or not. This can be computed using precision which is depicted as follow:

$$\text{Precision} = \frac{|TP|}{|TP| + |FP|} \quad (5)$$

Where TP is the number of valid predicted tweets, and FP is the number of invalid predicted tweets. However, it is necessary also to consider how many tweets have been classified correctly in accordance to the total number of specific category.

Therefore, recall is also considered which can be depicted as follow:

$$Recall = \frac{|TP|}{|TP| + |FN|} \quad (6)$$

Where FN are the tweets that have been classified incorrectly into other class label. Finally, the f-score which is the harmony between precision and recall, will be considered. It can be depicted as follow:

$$F = \frac{2 \times |TP|}{(|TP| + |FN|) + (|TP| + |FP|)} \quad (7)$$

$$= 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

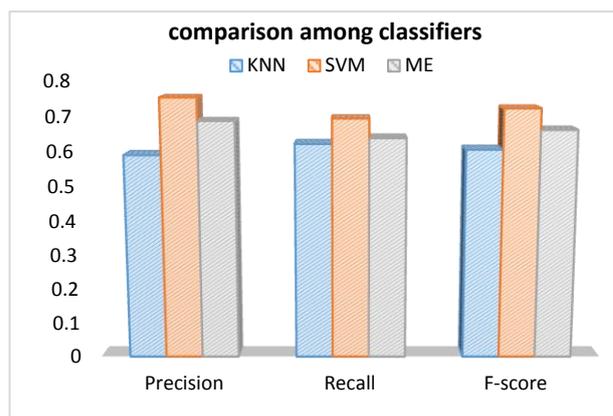
Based on the above mentioned measures, the results of classification for the three methods can be depicted in Table 4.

**Table 4. Results of classification**

Model	Precision	Recall	F-score
SVM	0.7521	0.6933	0.7215
KNN	0.5894	0.6213	0.6049
ME	0.6861	0.6377	0.6610

As shown in Table 4, SVM has achieved 75.21% precision, 69.33% recall and 72.15% f-score. Whereas, KNN has obtained 58.94% precision, 62.13% recall and 60.49% f-score. Finally, ME gained 68.61% precision, 63.77% recall and 66.10% f-score.

It is obvious that SVM has outperformed the other classifiers by obtaining the highest values of f-score. The reason behind such superiority lies on the capability of SVM to deal with large number of classes, as well as, the mechanism of vectorization which gives better representation in the text data (21). On the other hand, KNN showed the lowest values of f-score. The reason behind such poor performance lies on ineffective assumption of similar tweet in meaning would lead to similar class label. Sometimes a couple of tweet might yield similar context, meanwhile, they have different category of polarity. Figure 2 depicts the superiority of SVM in terms of precision, recall and f-score compared to the other classifiers.



**Figure 2. Performances of all classifiers**

By comparing the best results achieved by this study (i.e. 72.21%) against the state of the art such as (13) who achieved their best results at 69.9% of f-measure, it is clearly that the proposed method has the ability to produce competitive results of classifying sentiments.

### Conclusion:

This paper has proposed a combination of linguistic and statistical features along with classification methods for Arabic sentiment analysis. Linguistic has included the stemming and POS tagging, while statistical contained the TF-IDF. A benchmark dataset of Arabic tweets has been used within the experiments. Furthermore, three classifiers have been used including SVM, KNN and ME. Results showed that SVM has achieved the highest classification accuracy. For future work, utilizing more linguistic features such as parsing and word embedding would yield promising classification accuracy results.

### Conflicts of Interest: None.

### References:

1. Codella NC, Natsev A, Smith JR. Social media event detection and content-based retrieval. Google Patents. 2015.
2. Soliman AB, Eissa K, El-Beltagy SR. Aravec: A set of arabic word embedding models for use in arabic nlp. PCS. 2017;117:256-65.
3. Korayem M, Crandall D, Abdul-Mageed M, editors. Subjectivity and sentiment analysis of arabic: A survey. International conference on advanced machine learning technologies and applications. 2012: Springer.
4. Alhumoud SO, Altuwaijri MI, Albuhairei TM, Alohaideb WM. Survey on arabic sentiment analysis in twitter. ISI. 2015;9(1):364-8.
5. Mansour R, Hady M, Hosam E, Amr H, Ashour A. Feature Selection for Twitter Sentiment Analysis: An Experimental Study. In: Gelbukh A, editor. Computational Linguistics and Intelligent Text

- Processing. Lecture Notes in Computer Science. 9042: Springer International Publishing. 2015. p. 92-103.
6. Yazdani SF, Murad MAA, Sharef NM, Singh YP, Latiff ARA. Sentiment Classification of Financial News Using Statistical Features. IJPRAI. 2016;31(03):1750006.
  7. Nassirtoussi AK, Aghabozorgi S, Wah TY, Ngo DCL. Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. EXPERT SYST APPL.. 2015;42(1):306-24.
  8. Abdulla NA, Ahmed NA, Shehab MA, Al-Ayyoub M. Arabic sentiment analysis: Lexicon-based and corpus-based. 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT). 2013 3-5 Dec. 2013.
  9. Ahmed N. Large-Scale Arabic Sentiment Corpus And Lexicon Building For Concept-Based Sentiment Analysis Systems. 2018. PhD Thesis. Ankara Turkey: Hacettepe Universit.
  10. Abbasi A, Chen H, Salem A. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. TOIS. 2008;26(3):12.
  11. Soliman TH, Elmasry M, Hedar A, Doss M. Sentiment analysis of Arabic slang comments on facebook. IJCEIT 2014;12(5):3470-8.
  12. Abdul-Mageed M, Diab M, Kübler S. SAMAR: Subjectivity and sentiment analysis for Arabic social media. Comput Speech Lang. 2014;28(1):20-37.
  13. Al-Twairish N, Al-Khalifa H, Alsaman A, Al-Ohali Y. Sentiment Analysis of Arabic Tweets: Feature Engineering and A Hybrid Approach. arXiv e-prints [Internet]. 2018 May 01; Available from: <https://ui.adsabs.harvard.edu/abs/2018arXiv180508533A>.
  14. Oussous A, Lahcen AA, Belfkih S. Improving Sentiment Analysis of Moroccan Tweets Using Ensemble Learning. Big Data, Cloud and Applications. 2018; Cham: Springer International Publishing.
  15. Nabil M, Aly M, Atiya A. ASTD: Arabic Sentiment Tweets Dataset2015 sep; Lisbon, Portugal: Association for Computational Linguistics.
  16. Khoja S, Garside R. Stemming arabic text. Lancaster, UK, Computing Department, Lancaster University. 1999.
  17. AlGahtani S, Black W, McNaught J. Arabic part-of-speech tagging using transformation-based learning. Proceedings of the 2nd International Conference on Arabic Language Resources and Tools, Cairo. 2009.
  18. Yin C, Xiang J, Zhang H, Yin Z, Wang J. Short Text Classification Algorithm Based on Semi-Supervised Learning and SVM. IJMUE 2015;10(12):195-206.
  19. Lakshmi TRV, Sastry PN, Rajinikanth TV. Hybrid Approach for Telugu Handwritten Character Recognition Using k-NN and SVM Classifiers. IIRECOS. 2015;10(9):923-9.
  20. Xie X, Ge S, Hu F, Xie M, Jiang N. An improved algorithm for sentiment analysis based on maximum entropy. Soft comput 2017:1-13.
  21. Alsaleem S. Automated Arabic Text Categorization Using SVM and NB. Int Arab J e-Technol. 2011;2(2):124-8.

## دمج الآليات اللغوية والاحصائية لتحليل الرأي في اللغة العربية

هدى كاظم تايه<sup>2</sup>

احمد صباح احمد الجميلي<sup>1</sup>

<sup>1</sup>قسم تكنولوجيا معلومات الاعمال، كلية معلوماتية الاعمال، جامعة تكنولوجيا المعلومات والاتصالات، بغداد، العراق.  
<sup>2</sup>قسم ادارة انظمة المعلوماتية، كلية معلوماتية الاعمال، جامعة تكنولوجيا المعلومات والاتصالات، بغداد، العراق.

### الخلاصة:

تحليل الآراء هي عملية إيجاد تصنيف إيجابي أو سلبي لنص يحتمل احتوائه على آراء. اللغة العربية واحدة من اللغات التي تضخم محتواها بشكل كبير في العقد السابق وخصوصاً مع تصاعد وسائل الاتصال الاجتماعي مثل تويتر، فيسبوك وآخرين. دراسات كثيرة عاينت مهمة تحليل الآراء في اللغة العربية باستخدام تقنيات متعددة. أحد أكفأ الطرق المستخدمة في الدراسات السابقة كانت تعود لتقنيات تعلم الآلة وذلك لقدرتها على بناء قاعدة من التعلم من الحالات السابقة. مع ذلك هنالك قضايا كثيرة ممكن أن تواجه تقنيات تعلم الآلة في مهمة تحليل الرأي. واحدة من هذه القضايا هي كيفية إيجاد خصائص دقيقة في اللغة العربية التي بدورها ممكن أن تساعد على التفريق بين الآراء السلبية والإيجابية. هذه الدراسة تهدف الى اقتراح خليط من الأدوات اللغوية والاحصائية في سبيل الحصول على خصائص مميزة لتحليل الرأي في اللغة العربية. الأدوات اللغوية تحتوي على تقنيات إرجاع الكلمة لأصلها وتصنيف الكلمات بالنسبة لنوعها النحوي، بينما الأدوات الاحصائية تحتوي على تقنيات إيجاد أكثر الكلمات تردداً. تمت التجاربي باستخدام قاعدة بيانات آراء باللغة العربية. بالإضافة الى ذلك، تم استخدام ثلاث أنواع من تقنيات تعلم الآلة وهم (اس في ام)، (كي ان ان) و (ام اي). النتائج أظهرت بأن ال (اس في ام) تفوقت على الطرق الأخرى باستخدام الخصائص المقترحة وذلك بحصولها على دقة تساوي 72.15 بالمئة. تشير هذه النتائج الى فائدة استخدام ال (اس في ام) مع الخصائص المقترحة في تصنيف الآراء باللغة العربية.

**الكلمات المفتاحية:** تحليل الرأي باللغة العربية، طريقة العناصر الاقرب، الانترنت الأعلى، تجزير الكلمات، تحليل الرأي، نحويات، طرق تعليم الآلة، احصاء المصطلحات.