Arabic CyberBullying Detection Using Arabic Sentiment Analysis

Samar Almutiry*[1], Mohamed Abdel Fattah**[2]

*College of Computer Science and Engineering, Taibah University
Saudi Arabia- Almadinah Almunawarah
[1]samar888abdullah@gmail.com

**Department of Electronics Technology, FIE, Helwan University, Cairo, Egypt
[2]maiahmed@taibahu.edu.sa

Abstract: ***The Sentiment Analysis is used for the text analysing, detecting opinion, and classification of the text attitude. It becomes quite challenging when it is applied to the Arabic language due to the structural and morphological complexity, known as "Arabic Sentiment Analysis (ASA)." For the implementation of ASA, we are using the computing advancement in the form of Machine Learning (ML) and Support Vector Machine (SVM) algorithm to train a dataset which is collected automatically through ArabiTools and Twitter API. The dataset contents are labelled by both means, automatic and manual, in order to maintain the efficiency of the detection of CyberBullying tweets. Use internet technology to bully a person by using aggressive and offensive words is known as CyberBullying. The dataset is automatically labelled with respect to the nature of the tweet. If a tweet contains one or more CyberBullying words, it is labelled as CyberBullying, while if there is not any word with aggressive meaning found, it is marked as the NonCyberBullying. After the data collection, there are several pre-processing techniques utilized, including the Normalization, Tokenization, Light Stemmer, ArabicStemmerKhoja, and Term Frequency-Inverse Document Frequency (TF-IDF)" term weighting schema." After the preliminary steps, (SVM), a standard "supervised algorithm," is used with WEKA and Python. There are three experiments that take place one with the WEKA tool using the Light Stemmer, the other is again with WEKA using ArabicStemmerKhoja, and the final experiment was performed with Python. The results are showing the WEKA is more efficient in classifying the text correctly, while Python is more effective with time to build the model. WEKA using the Light Stemmer have the efficiency of 85.49% and taken* **352.51 seconds***, and the WEKA using ArabicStemmerKhoja have the efficiency of 85.38% and taken 212.12 seconds, while the Python have the efficiency of 84.03% and taken 142.68 seconds.***

Keywords: ***CyberBullying, Text Classification, Arabic Text, Machine Learning, Sentiment Analysis, Support Vector Machine.***

1 INTRODUCTION

The Internet has revolutionized the lifestyle; a distance of thousands of kilometers is now just a number; a person can remain in-contact with another person with the help of the Internet. Social media has given a great boom to the Internet; people can share their opinions regarding any topic on social media like Instagram [1], Twitter [2], Facebook [3], etc. On the other hand, there are so many risks involved [4]. According to a strategy consultant, Steve Tobak, the tweet either make you famous or fired [5]. We share our opinion on Twitter on a daily basis; it is reachable to everyone. Some people agree with our point of view, and some do not. In this regard, we receive a lot of aggressive and offensive comments on our tweets even sometimes these comments are not in the context of our opinion; this causes a risk of CyberBullying, which means to bully a person by using the Internet and technology [6]. In this regard, we decided to find out the solution to this modern problem. Moreover, we found that the use of Machine Learning by using the "simulated annealing" can be effective in finding out the CyberBullying tweets by analyzing, polarizing, and classifying text into sentimental classes [7]. In addition to this, on Twitter, the users share their opinion most of the time in their local language, so it is another challenge to use the technology to find out CyberBullying in any language other than English. In this research, we are going to work on the tweets written in the Arabic language [8].

In order to obtain the solution to detect the CyberBullying in tweets written in the Arabic language, there are several experiments made. We are going to apply the Sentiment Analysis (SA) algorithm by using Machine Learning, which has the ability to check out the positive and negative words that lead to fulfilling our aim of this research. This technique will find out, analyze, and evaluate the opinions, attitudes, emotions, and views of a person to a particular issue or topic, such as social events, Saudi championship, and international brands [8]. In addition to this, the use of the SA algorithm approach is quite challenging because of the Arabic morphology complexity and diverse dialects of the Arabic language. For this experiment, we are utilizing the SA with the Machine Learning techniques with the support of a standard classifier known as "Support Vector Machine" (SVM). We are using the "Waikato Environment for Knowledge Analysis" WEKA and the Python as a tool for data mining in order to build the models for our experimentation. The dataset was obtained from Twitter by using "Application Programming Interface" (API) and ArabiTools.

A. Our Objectives

- To categorize Arabic tweets into CyberBullying and Non-CyberBullying.

- To create the automatic dataset, mark the tweets manually and automatically, and keep the corpus public available in order to aid the research community.
- To create different "Arabic Dialects Stop Words" and "Modern Standard Arabic" (MSA) list from our gathered dataset.
- To create a list of "Arabic CyberBullying" words from our gathered dataset.
- To apply a standard Machine Learning classifier (SVM) on our dataset.
- To create our model by the use of Python and WEKA, and comparison of the obtained results.

B. Motivation

It is noticed when a person writes a tweet, and he/she receive random sarcasm and bullying comments with respect to the tweet, profile picture, or any comment, sometimes this bullying is even not related to the writer's opinion. In addition, when someone posts his/her state in its opinion, it spread out all over the world through the platform of Twitter than the tweet, and the sarcasm comments have a subsequent effect on our society; this type of bullying is termed as CyberBullying. This is our motivation to develop a method through with we can prevent the tweets from the acts of CyberBullying by applying the approach of SA. Furthermore, it is observed that there is very little research on the Arabic Language with respect to CyberBullying, so this also motivates us to work with Arabic Tweets.

C. Research Scope

- This research is only focused on the "Arabic Language" tweets. We focused only on Arabic language tweets.
- Our dataset is obtained by random collection and by using the query-oriented method.
- This approach is mainly focused on classification accuracy instead of classification speed because the detection of CyberBullying tweets is an essential factor.

D. Research Questions

- Q1: What are the challenges and how to resolve them for the processing of the Arabic language?
- Q2: Different preprocessing steps can affect the classification accuracy of the model?
- Q3: Which data mining tool provides better results?
- Q4: Which data mining tools are more efficient with time for model classification?

E. Research Significance

This research work gave an awareness regarding the social medial CyberBullying and tried to detect the CyberBullying on Twitter. For the government and organization, it is quite essential to detect the CyberBullying. So, our designed models to detect from tweets can be useful for the systems working for CyberBullying. Our models can also be useful for the organization related to adult affairs and children like education, sports organization.

F. Research Organization

The rest of the research is organized as follows:

- **Background Section:** All the essential theories related to the topic are discussed.
- **Literature Review:** Previously done research are discussed.
- **Methodology:** All the proposed methods that are used for this research are explained.
- **Result Section:** All the proposed experiments are applied by using the WEKA and Python tools with the SVM classifier, and the results are shown in the tables and graphs.
- **Discussion Section:** Discuss all the experiments performed by WEKA and Python in this research and their results.
- **Conclusion Section:** Conclusive Summary of the whole research work.

2 BACKGROUND

This research is mainly based on "Natural Language Processing" (NLP), which analyzes the written text to make it possible for computers to understand the human language [9]. It has several applications; one of them is Text Classification (TC), used in this research. TC has the ability to perform the classification of texts in single or multiple predefined categories; it is also termed as Text Categorization or Text Tagging [10]. There is a type of TC known as Sentiment Analysis (SA), which uses the concepts of computer sciences to extract the sentiment from the text. In this research, we are going to use the SA by using the Machine Learning approach [11]. However, Machine Learning (ML) is a process to learn machine regarding a particular topic; there are two kinds of ML one supervised, and the other is unsupervised. In the supervised ML, the data is learned by using labels that indicate the class of each sample of data; further, the classification of new data is based on the training set of data. On the other hand, unsupervised ML is a way in

which there are unknown class labels of training data [12]. There is a learning algorithm, "Support Vector Machine" SVM, used for the classification tasks with SA and TC. It transforms the given data into a "higher-dimensional feature space" and finds out an optimal hyperplane, which will separate the given dataset in such a way that the variable of one category lies on one side of hyperplane and the other variables lies on the other side of hyperplane [13].

The Arabic language is the fifth most spoken language all around the world; it is the first language of 422 million people and 250 million people as the second language [14]. The Arabic language has its own importance due to graphical and religious reasons. There are three different kinds of the Arabic language, one is Classic Arabic in which the Holy Quran is written, the second is "Modern Standard Arabic" (MSA), which used in the books, media, education, etc., and Colloquial Arabic spoken in Lebanon, Syria, Algeria, Iraq, etc. [15]. This research is related to Arabic language use on Twitter, so there are about five million Arabic users, and the "Kingdom of Saudi Arabia" (KSA) has 2.4 million active Twitter users, which is about 40% of the total Arabic active users [16]. There are several challenges with the Arabic language; one of the most challenging is the complex morphology of Arabic; it makes the normalization, tokenization, and stemming very difficult. The structure of Arabic words is quite complex, in which the steam combines with the clitics (it includes the conjunctions, preposition, etc.) and affixes (has inflectional markers for tens like a number and/or gender). There is another challenge; the Arabic words are derived from root words, so the extract of root words from a conventional word is quite a difficult task. The Arabic language uses short vowels known as diacritics, that are used as the pronunciation extractor, and the exact meaning. The Arabic writing on Twitter should be without these diacritics; in addition, there are several dialects with no writing standards. Arabic has a wide range of synonyms; during the classification, it is quite difficult to classify a particular word by using the exact keyword. As a result, it degrades the performance of the system, in addition to the previously mentioned challenges, the "linguistic code-switching" between the MSA and other dialects. Furthermore, there are not enough tools available for Arabic morphology analysis and contents. Along with all of these challenges, there is a research gap with the Sentimental Analysis of the Arabic Language. Twitter is a "Micro-Blogging Social Media Network" where the users can share their opinions instantaneously in public. The users can follow sports clubs, influencers, fashion brands, etc. There are a lot of opinions on Twitter because the users share before and after their life-events [2]. Due to the high social activities on Twitter, there are a lot of changes in CyberBullying. According to the "Saudi Ministry of Communication and Information Technology," it is worse than traditional bullying because, in this condition, the bully is an anonymous person [17]. The data extraction from Twitter can be performed by using Twitter API, provide the interactions among web services and computer programs.

3 LITERATURE REVIEW

In the recent past, the application of SA is made on the classification of text with the English language [18]–[23]. However, there is a research gap in the application of SA for the Arabic language; we have very rare research such as [24]–[27]. In the use of SA, the researchers worked on the movie review, forums, news articles, and data of social media [28]. In the early stage of SA, there was work on the basis of text mining, but it revolved rapidly, and now it can be useful for complex feature and symbol recognition. In the past, this technique can only classify two sentiments, negative and positive, but it does not quite able to classify multiple sentiments like sad, happy, angry, etc. In order to work with the human language, the SA was not sufficient in the past. Because human writing can have different meanings with the same sentence, in this scenario, only positive and negative sentiments distinguish is not sufficient. In order to overcome this problem, the SA is developed by using the Machine Learning, where it can detect the polarity of the sentiment and label it [29]. The ML is used with several algorithms, including SVM, K-nearest neighbor (KNN) [28], Naïve Bayes (NB), and Decision Tree (DT). "Arabic Sentiment Analysis" (ASA) is conducted by using the MSA and other Arabic dialects like Egyptian [24]. As far as the concern about the TC, it is based on the steps, including data collection, preprocessing, selection of features, data classification, and the evaluation of data. The data can be collected from web pages, while the process of data preprocessing has further sub-processes like data cleaning, data normalization, removal of stop word, stemmer, and tokenization. According to a study, the selection of features can enhance the performance of classifiers [28].

Osaimi et al. [31] studied the sentimental analysis for the Arabic language and introduced an automatic method that can predict the sentiments of Arabic tweets along with emotion icons. This predicted process is based on several sub-processes, including the collection of tweets, preprocessing, and the model building using the KNN, NB, and RapidMiner tools. This proposed approach remains unable to provide great accuracy; the results showed if the KNN classifier used the accuracy found about 59.04%, and if it uses NB as a classifier, then the accuracy was 63.79%[30]. While Nalini et al. worked on the detection of CyberBullying and developed a method that can identify the active victims and bullies, they used "Term Frequency-Inverse Document Frequency" (TF-IDF). They compared their proposed method with the baseline methods of TF-IDF and semantic [31]. In addition to this, Bouchlaghem et al. [27] developed an approach based on the semantic analysis in order to translate the Arabic tweet orientation that is used for terroristic acts. They have developed a data representation method by using the N-gram features, "sentence-level features," linguistic features, syntactic features, and "tweets specific features. Generally, there is a problem with duplicate tweets while collecting data,

but these researchers solve this problem by the application of similarity measurement. The results of their research are showing that the SVM is the best performing classifier as compared to the Random Forest, NB, KNN, and DT. On the other hand, Magdy et al. [32] introduced a tool for the multilingual classification of tweets. It performs the automatic label's collection for the "training dataset," so it can be used for the classification. This tool has the ability to classify the tweets written in five different languages. They collected the Twitter data by using the Twitter API, they used the SVM classifier, and their results are showing an accuracy of 84% . Another group of researchers, Nandakumar et al. [33]., have applied the "binary classification" for the purpose of CyberBullying detection in the tweets by using the NB algorithm. They collected the data using Twitter API; after this, they removed the noise from the collected data, applied the NB classifier, and select the feature. After all these efforts, the list of "bullied tweets" is obtained. Further on, Abdelaal et al. [34] worked on the classification of Arabic tweets into predefined categories, including general, culture, technology, sports, and politics. They also tried to improve the accuracy of ensemble classification methods. There were three sub-processes in their proposed process, collection, preprocessing, and classification of tweets. Their result is showing that the ensemble methods are quite better as compared to the individual classifier in order to enhance the classification accuracy. The accuracy of SVM was increased by 2.2%, the accuracy of NB was improved by 1.6%, and the accuracy of DT was improved by 3.2%.

Lexicon approaches and sentiment analysis are suggested by AlHarbi et al. [35] to use for automatic cyberbullying detection. Java programming language has been used in this experimental work and the dataset has been done in a study. Datasets were taken from YouTube, Microsoft-Flow, and Twitter API comments. After that, they were collected into one file, which had nearly 100,327 comments and tweets. The data was arranged to bullying and non-bullying after the data preprocessing and cleaning step. The data was arranged by three individuals and employ an odd number of individuals to be the last division after the main opinion. When the data was configured and completed for lexicon generation usage, Authors employed Entropy, Chi-square, and PMI. As a result, the PMI approach provides the most efficient performance in identifying cyberbullying when compared with Entropy and Chi-square approaches. L. Cheng et al. [36] suggested PI-Bully, which is a principled personalized cyberbullying detection framework, which approach these interdisciplinary findings to alter and better the cyberbullying behavior prediction. Existing detecting cyber harassment models have concentrated on creating specific classification methods for all people trying to recognize bullying content from normal one. On the other hand, these methods skip special characteristics, which are used in the user-generated content. Empirical psychology research shows the role of individual differences, which are seen in users' specific personality motives, attitudes, traits, etc. and affect like-minded users as online bullying predictors. A. Bellmorea et al.[37], authors constructed a frequency vector for every tweet and tested a text classifier to provide an answer to main questions about online bullying ("What, Who, Why, When, and Where") with the help of a dictionary that contains words in a Twitter corpus. As the social networking systems prevalence keeps rising, network-based features such as relational centrality, network embeddedness, and the number of friends, are employed to identify the behaviors of cyberbullying. Furthermore, cyber harassment has been researched on other social media platforms, including Instagram by H. L. Haoti Zhong et al. [38] focusing on the cyberbullying detection in image sharing networks, paying attention to the early warning mechanism development for detecting pictures with high vulnerability to attacks. When it comes to image-sharing, authors concentrated on features of the captions and photos themselves, concluding that captions can serve as a huge predictor of future cyber harassment for a given picture. This project is an important step toward creating software tools that will help to monitor cyberbullying on social media platforms. Table 1 shows comparison of our work and some of the related works discussed earlier.

TABLE 1
COMPARISON BETWEEN SOME OF RELATED WORKS AND OUR WORK

Paper	**Dataset**	**Classifier**	**Accuracy**
[35]	YouTube, Microsoft-Flow, and Twitter API comment	Entropy, Chi-square, and PMI	PMI=81% Entropy=39.14% Chi-Square=62.11%
[36]	Twitter	KNN	84%
[37]	Twitter	SVM	86%
[38]	Instagram	SVM	93.2%
Our Work	Twitter	SVM	WEKA using the Light Stemmer have the efficiency of 85.49% WEKA using ArabicStemmerKhoja have the efficiency of 85.38% Python have the efficiency of 84.03%

4 METHODOLOGY

A. Overview and Structure of Methodology

The methodology of this experimentation is based on four steps, in the first step the data is collected by using ArabiTools and Twitter API and annotated by using the python. In the second, the collected and annotated data pre-processed before the classification from the noise. In the third step, the data is classified by using SVM, while in the final step the performance of model evaluated.

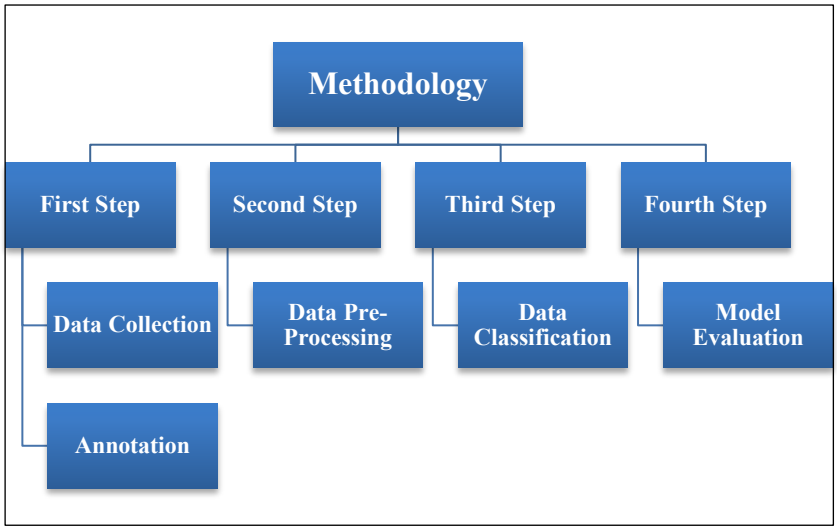

Figure 1 Structure of Methodology

1) Data Collection and Annotation

The data was collected by using the ArabiTools and Twitter API via two different methods, one is a random selection, where the words searched randomly, and the other is query-oriented, where the words are searched by using specific keywords, the words that mostly used to do Arabic CyberBullying such as Racist, عنصري. The dataset was obtained from the tweets written in MSA. There was total 17748 Arabic tweets collected, where CyberBullying tweets were 14178 while the Non-CyberBullying tweets were 3570, our dataset called AraBully-Tweets. Firstly, the dataset was stored in the "Comma Separated Value" (CSV), then it was converted to the "Attribute-Relation File Format" (ARFF) by using the WEKA. A sample of data in the ARFF is shown in Figure 1.

```
@relation CyberBullyingDetection
@attribute class {c, n}
@attribute tweet String
@data
نصصصيحه الصين بلد كئيب جدا و اكلهم يجيب المرض و نفوسهم شينه ,c
بع
و , n الطيبون في حياتنا رزق
```

Figure 2 ARFF Format Dataset Sample

After this, a list of "Arabic CyberBullying Words" is obtained from the two lexicons, the first one was SauDiSenti [39], and the Second was [40]. In addition to this, those words have an aggressive context are labeled as CyberBullying words. In our dataset, these words are gathered and counted by using the Python codes. These words include the bully and aggressive expressions, phrases, nouns, and adjectives. This obtained list of CyberBullying words is used to annotate the given dataset. Table 2 has some AraBully-words with their English meaning.

TABLE 2
SAMPLE OF ARABIC BULLYING WORDS

Words	English Translation
غبي	Stupid
أحمق	Foolish
متنمر	Bully

The annotation process (as shown in figure 2) is used with the aid of Python to annotate the dataset for AraBully-Words. The Python code compares every tweet with the list of Ara-Bully-Words; if the tweet has one or more Ara-Bully Words,

the tweet will be labeled as the CyberBullying; otherwise, it will be marked as the Non-CyberBullying. If a tweet contains a good word and a negative bullying word, the system will make the tweet as CyberBullying. After the Python-based automatic annotation, we applied a manual annotation process to check out the efficiency of Python-based Automatic Annotation. Three Arabic native speakers performed this manual annotation to check the automatic annotation.

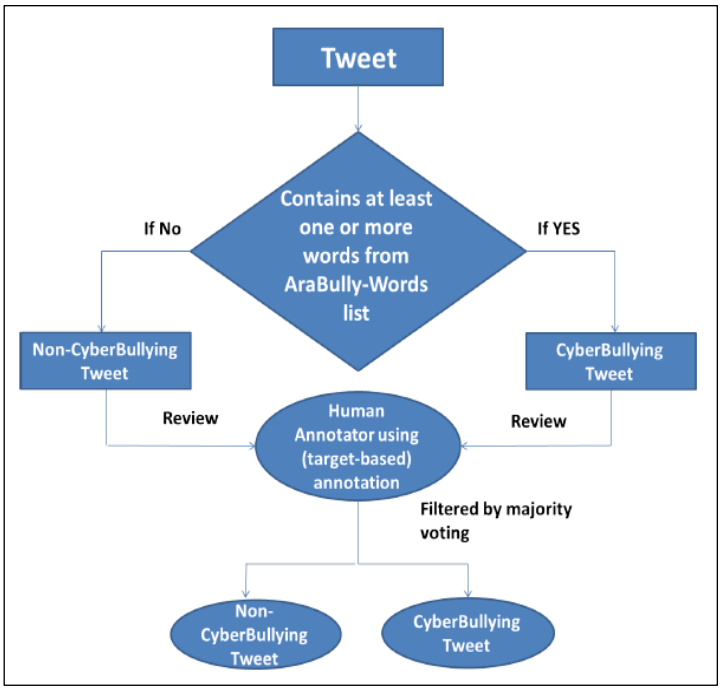

Figure 3 Process of Annotation

2) Data Preprocessing

Before the classification, the tweets should be cleaned and preprocesses from noise. In the noise, there are two features; one is the internal features, including the content, while the other is external features, which includes URL, hashtags, tweet size, etc. First, the dataset is cleaned by removing the items that cause noise in the dataset, a process known as Data Cleaning. The removing components are non-Arabic Letters, user mentions (@user), single Arabic characters, numbers special characters (%, &, +, /, %), duplicate tweets, re-tweets, and pictures of the tweets. This process is done by using the tools Microsoft ® Excel and Almoshatheb Alarabi [41]. After the Data Cleaning, the next step is Data Normalization where all different forms of Arabic words convert into a consistent shape like the ي converted to ى and the آ converted to ا. Furthermore, punctuation, diacritics, and the lengthening of the Arabic writing (Tatweel) are removed, as shown in Table 3.

TABLE 3
SAMPLE OF LENGTHENING REMOVING

Without Tatweel	With Tatweel
العربية	العـــــــــــربـيــة

There are several words in the writing that has no meaning, and they provide no information to the writing, known as Stop Words. These words are the prepositions, pronouns, conjunctions, etc. that are frequently utilized in sentences. These words are removed in this process; it will provide dimension reduction for the classification process. There are some stop words showed in table 4, with the MSA and Arabic Dialects.

TABLE 4
SAMPLES OF STOP WORDS

Arabic Dialects StopWords Examples	MSA StopWords Examples
عشان	متى
كذا	لاسيما
اللى	حيثما

Arabic is a highly derivative language; hundreds of words can be developed by using one root and adding the affixes. So, here is a step that is known as Stemming, where the affixes are removed and keep the word in root or stem forms. Such

as the words like كاتب, كتاب, مكتبه convert to the base root كتب. There are further two major approaches in the Stemming; one is Root-Based Stemmer where the word is reduced to its base root, it is also known as Khoja Stemmer and implement on the WEKA while the other is light stemming where the affixes (prefixes and suffixes) removed, this type of stemming is generally used in AraNLP. We have used both Light Stemming and ArabicStemmerKhoja. After this, the next step is Tokenization, where the unstructured text converts to the discrete token sequence by using the white spaces and punctuation marks. We have used single word tokenization; the output of this process will be used in the classification process. The sample of single word tokenization is shown in table 5.

TABLE 5
SAMPLE OF SINGLE WORD TOKENIZATION

Tweet	Tokenization
. ما في احد شكله غبي الا ياسر واتباعه	<'ما', 'في', 'احد', 'شكله', 'غبي', 'الا', 'ياسر', 'واتباعه'>

In order to apply the machine learning algorithms, the text should represent in the numerical form of a vector. There is another step known as Term Weighting. The vector length is equivalent to all "unique word's length" (t). Each term (t) in a document (j) is given a real-valued weight, W(tj). The documents are expressed as t-dimensional vectors:

$$document_j = (w_{1j}, w_{2j}, \dots, w_{tj})$$

For the better performance of TC, it is quite necessary to select the most appropriate "term weighting scheme." There are some "term weightings schemes" including "the Boolean model," "Term Frequency" (TF), "Inverse Document Frequency" (IDF), and TF-IDF. In this ongoing research, the (TF-IDF) "term weighting scheme" is utilized, it can be calculated by using the following formula.

$$W_{ij} = TF_{ij}\, x\, IDF_i = TF_{ij}\, x \log_2 \left(\frac{N}{DF_i}\right)$$

3) Data Classification

We have used the "supervised learning algorithms" (SVM). The Sentiment Analysis is performed by using the Machine Learning techniques. The Machine Learning algorithms are used to learn and build the "classifier model" by providing the training to dataset regarding the sentimental labeling on tweets as "CyberBullying and NonCyberBullying." This classifier analyzes the tweets and predicts the sentimental label for the tweets. We have performed three experiments by using the WEKA and Python using different preprocessing but there are some common preprocessing parts in the Python and WEKA including, normalization, data cleaning, removal of stop words, "term weighting schema" and tokenization. There are two Arabic stemmers used in WEKA, one is Light stemmer and the other is ArabicStemmerKhoja; while there is not any type of Stemmer used with the Python.

TABLE 6
EVALUATION MEASURES

Evaluation Matrices	Definition
A	"The number of instances or tweets that are correctly classified."
P	"The number of correctly classified positive tweets divided by the number of tweets labeled as positive by the system."
R	"The number of correctly classified positive tweets divided by the number of positive tweets in the dataset."
F	"It is the harmonic mean of P and R."

4) Model Evaluation

There is a way to measure the model performance of a test data based on the mixed number of incorrect and correct predictions, known as the Confusion Matrix. In the confusion matrix, two categories are made for the classification, namely Predicted Class and Actual Class, as shown in Table 6. Conclusively, this confusion matrix is used as the Evaluation Model for our research; in order to get the outcome of performance measures, there are several evaluation measures used like A, P, R, and F, as shown in Table 7. In the "matrix, the TP shows the number of those tweets that are

assigned correctly to the belong category while the FN shows the number of not correctly assigned. In addition, the FP shows the number of those tweets that are assigned incorrectly to the given category, and TN shows the number of not correctly" assigned.

TABLE 7
CONFUSION MATRIX FOR TWO-CLASS CLASSIFICATION PROBLEM

"Actual Class / Predicted Class"	**C1**	**-C1**
C1	"True Positive (TP)	"False Negative (FN)"
-C1	"False Positive (FP)"	"True Negative (TN)"

5 RESULTS

As earlier mentioned, the WEKA and Python are used for this experimentation; in the WEKA we used with Light Stemming and ArabicStemmerKhoja, SVM is used as the classifier in both of them. The encoding of "RunWeka.ini" was changed from the "Cp1252" to the "UTF-8" in order to make the Arabic characters visible in WEKA Explore. To handle the large size of data, the heap size is increased in "RunWeka.ini." First of all, the experiment is performed by using the WEKA with Light Stemmer and found there are 1525.63 tweets are correctly classified out of 17748, about 85.49%, while the P factor is 0.866, R is 0.859, and the F is 0.862, as shown in Table 8. The Threshold Curve is also generated by using the values of P (x-axis) and R (y-axis) factors to visualize the results, shown in figure 3. The time required to build this classification model is 352.51 seconds.

TABLE 8
RESULTS OF CLASSIFICATION BY USING WEKA WITH LIGHT STEMMER

Total Tweets	**A**	**P**	**R**	**F**
17748	15252.6312	0.866	0.859	0.862

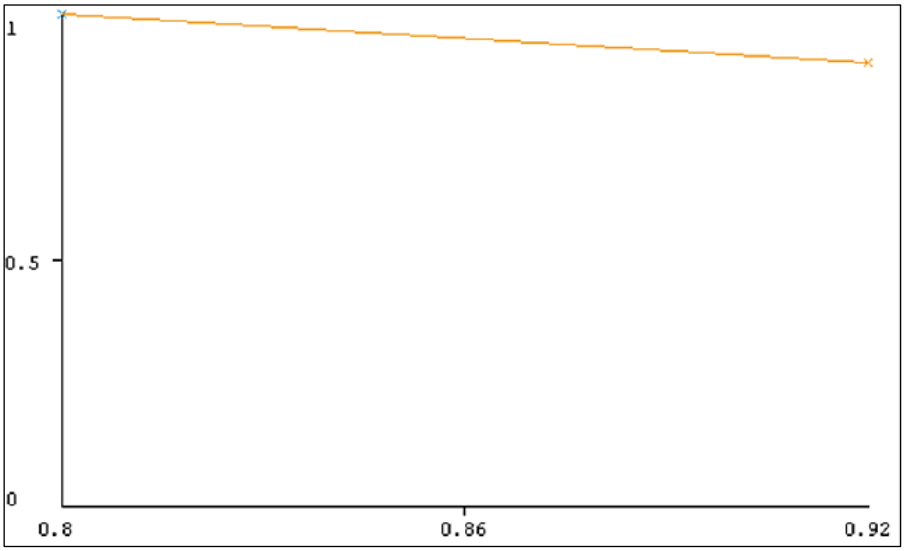

Figure 3 Threshold Curve for Classification by using WEKA with Light Stemmer.

The experiment is again performed by using the WEKA but this time with ArabicStemmerKhoja and found there is 15154 number of correctly classified tweets out of 17748, about 85.3843%. The value of P is 0.851, R is 0.854, and the F is 0.852. This result is shown in Table 9, and graphically represented with the help of Threshold Curve between P (x-axis) and R (y-axis) in figure 4. The time required to build this classification model is 212.12 seconds.

TABLE 9
RESULTS OF CLASSIFICATION BY USING WEKA WITH ARABICSTEMMERKHOJA

Total Tweets	A	P	R	F
17741	15154	0.851	0.854	0.852

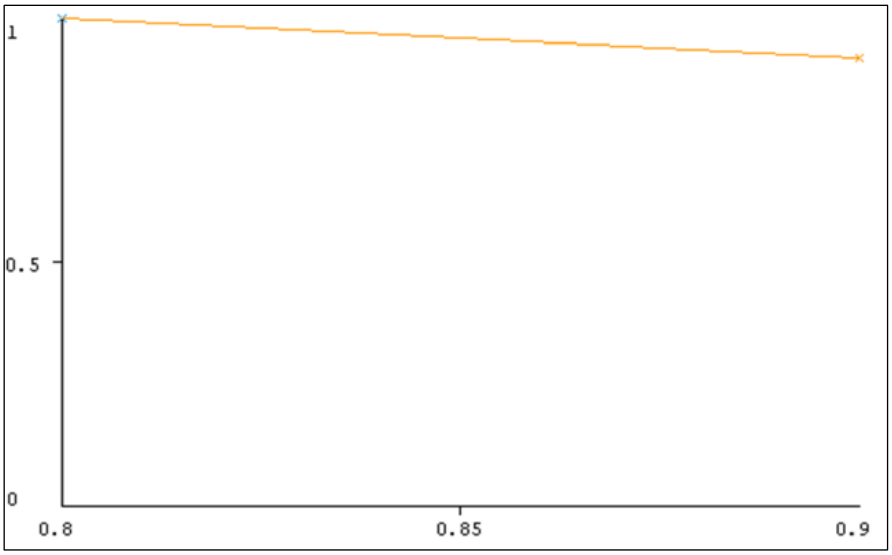

Figure 4 Threshold Curve for Classification by using WEKA with ArabicStemmerKhoja.

For the third turn, the experiment is performed with the Python supported with the SVM classifier and found that there are 14908.32 tweets corrects classified out of 17748 tweets, about 84.03%. The R is found about 0.84, P is 0.83, and the F is 0.835, as shown in table 10. The time required to build this classification model is 142.68 seconds.

TABLE 10
RESULTS OF CLASSIFICATION BY USING PYTHON

Total Tweets	A	P	R	F
17741	14908.32	0.83	0.84	0.835

6 DISCUSSION

In this research, we tried to detect the CyberBullying in tweets written in Arabic by using SA and ML. There are several challenges found with the Arabic languages, such as the morphology of the Arabic language is quite difficult, and it has a very complex structure due to the clitics and affixes. The Arabic words are based on the root words, so it is also very difficult to obtain the root word by removing the affixes and clitics. There are a lot of diacritics used for pronunciation. In Arabic, we have a lot of synonyms, which make the processing difficult, and the most important challenge with the SA of the Arabic language was the research gap. On the application of the detection model, we have made an annotation method where the tweet was checked by the automatic tools than it checked by the native Arabic humans for the testing purpose; the testing showed a great result with our approach. For the experimentation of this research, we have applied the WEKA and Python as data mining tools with the SVM classifier. For the WEKA, we have used two stemmers, one is Light Stemmer, and the other is ArabicStemmerKhoja. While we were using the python tool, we applied the normalization, stop word removal, tokenization; we used the "IF-IDF term weighting schema" and divided the main dataset into the testing dataset and the training dataset. While we were using the WEKA, we used the same method as we used in Python. In comparison, we found that the WEKA tool is better than the Python, the results showing that the WEKA correctly classified 15252.6312 (85.49%) tweets when used with Light Stemmer and correctly classified 15154 (85.3843%) tweets when used with the ArabicStemmerKhoja while Python correctly classified only 14908.32 tweets (84.03%). On the other hand, the Python tool is performed well with respect to the time for the building of classification models, it required only the time of 142.68 seconds, while the WEKA with Light Stemmer required 352.51 seconds and WEKA with ArabicStemmerKhoja required 212.12 seconds.

7 CONCLUSION

In this research, we have collected the data from Twitter by using tools like AraBully words. After the data collection, the preprocessing phase was performed where the data cleaning was done by using Microsoft® Excel and

Almoshatheb_Alarabi. After the cleaning, the normalizing method is applied to different words from a uniform version. It also removes the punctuations, diacritics, and removes the lengthening. After this tokenization is applied on the basis of white spaces, it is followed by the process of stop word removal from our dataset. Two data mining tools were utilized, WEKA and Python; for the WEKA, there are two stemmers used, one was Light Stemmer, and the other was ArabicStemmerKhoja. After the preprocessing, we worked on the classification step, where we used the SVM classifier tool, WEKA, and Python. There are different preprocessing steps for WEKA and Classifications, but they have some common too; the data cleaning, stop word removal, normalization, tokenization, and "(TF-IDF) term weighting schema" are the common preprocessing steps. The research is showing that the WEKA is better in classification the tweets correctly compared to Python; there is a slight difference in accuracy when using Light Stemmer and ArabicStemmerKhoja. But it is also observed that Python needs less time to build the classification model while the WEKA needs more time.

REFERENCES

[1] E. Moreau, "What Is Instagram and Why Should You Be Using It?," *Lifewire*, 2020. https://www.lifewire.com/what-is-instagram-3486316 (accessed Oct. 30, 2020).

[2] "What is Twitter and why should you use it? - Economic and Social Research Council." https://esrc.ukri.org/research/impact-toolkit/social-media/twitter/what-is-twitter/ (accessed Oct. 30, 2020).

[3] "Facebook: What is Facebook?," *GCFGlobal.org*. https://edu.gcfglobal.org/en/facebook101/what-is-facebook/1/ (accessed Oct. 30, 2020).

[4] A. Whiting and D. Williams, "Why people use social media: a uses and gratifications approach," *Qualitative Market Research: An International Journal*, vol. 16, no. 4, pp. 362–369, Jan. 2013, doi: 10.1108/QMR-06-2013-0041.

[5] "The Trouble With Twitter: It's a Double-Edged Sword - Steve Tobak," Jul. 16, 2018. https://stevetobak.com/2018/07/16/problem-with-twitter-double-edged-sword/ (accessed Oct. 30, 2020).

[6] C. Chelmis, D. Zois, and M. Yao, "Mining Patterns of Cyberbullying on Twitter," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 126–133, doi: 10.1109/ICDMW.2017.22.

[7] J. S. Sartakhti, H. Afrabandpey, and M. Saraee, "Simulated annealing least squares twin support vector machine (SA-LSTSVM) for pattern classification," *Soft Comput*, vol. 21, no. 15, pp. 4361–4373, Aug. 2017, doi: 10.1007/s00500-016-2067-4.

[8] D. Mouheb, R. Albarghash, M. F. Mowakeh, Z. A. Aghbari, and I. Kamel, "Detection of Arabic Cyberbullying on Social Networks using Machine Learning," in *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, Nov. 2019, pp. 1–5, doi: 10.1109/AICCSA47632.2019.9035276.

[9] S. Sun, C. Luo, and J. Chen, "A review of natural language processing techniques for opinion mining systems," *Information Fusion*, vol. 36, pp. 10–25, Jul. 2017, doi: 10.1016/j.inffus.2016.10.004.

[10] X. Zhang, J. Zhao, and Y. LeCun, "Character-level Convolutional Networks for Text Classification," in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 649–657.

[11] M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, Jul. 2013, pp. 1–5, doi: 10.1109/ICCCNT.2013.6726818.

[12] E. Alpaydin, *Introduction to machine learning*. MIT press, 2020.

[13] T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," in *Proceedings of the 2004 conference on empirical methods in natural language processing*, 2004, pp. 412–418.

[14] M. K. Saad and W. M. Ashour, "Arabic morphological tools for text mining," *Arabic morphological tools for text mining*, vol. 18, 2010.

[15] afreno, "The Three Major Types of Spoken Arabic Language," *Afreno*, Jul. 12, 2019. https://www.afreno.com/blog/the-three-major-types-of-spoken-arabic-language/ (accessed Oct. 30, 2020).

[16] "Twitter in the Arab Region," Mar. 2014. https://arabsocialmediareport.com/Twitter/LineChart.aspx (accessed Oct. 30, 2020).

[17] "The Growing Phenomenon Of Cyberbullying With The Increasing Use Of The Internet And Connected Devices," Jun. 22, 2015. https://www.mcit.gov.sa/ar/media-center/news/95841 (accessed Oct. 30, 2020).

[18] V. Hatzivassiloglou and K. McKeown, "Predicting the semantic orientation of adjectives," in *35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics*, 1997, pp. 174–181.

[19] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," *arXiv:cs/0212032*, Dec. 2002, Accessed: Oct. 30, 2020. [Online]. Available: http://arxiv.org/abs/cs/0212032.

[20] S. Bethard, H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky, "Automatic extraction of opinion propositions and their holders," in *2004 AAAI spring symposium on exploring attitude and affect in text*, 2004, vol. 2224.

[21] T. Wilson, J. Wiebe, and R. Hwa, "Just how mad are you? Finding strong and weak opinion clauses," in *aaai*, 2004, vol. 4, pp. 761–769.

[22] J. Wiebe and E. Riloff, "Finding Mutual Benefit between Subjectivity Analysis and Information Extraction," *IEEE Transactions on Affective Computing*, vol. 2, no. 4, pp. 175–191, Oct. 2011, doi: 10.1109/T-AFFC.2011.19.

[23] Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," *Expert Systems with Applications*, vol. 36, no. 3, Part 2, pp. 6527–6535, Apr. 2009, doi: 10.1016/j.eswa.2008.07.035.

[24] H. S. Ibrahim, S. M. Abdou, and M. Gheith, "Sentiment Analysis For Modern Standard Arabic And Colloquial," *IJNLC*, vol. 4, no. 2, pp. 95–109, Apr. 2015, doi: 10.5121/ijnlc.2015.4207.

[25] A. Mourad and K. Darwish, "Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs," in *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*, 2013, pp. 55–64.

[26] E. Refaee and V. Rieser, "An arabic twitter corpus for subjectivity and sentiment analysis.," in *LREC*, 2014, pp. 2268–2273.

[27] R. Bouchlaghem, A. Elkhelifi, and R. Faiz, "A Machine Learning Approach For Classifying Sentiments in Arabic tweets," in *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, New York, NY, USA, Jun. 2016, pp. 1–6, doi: 10.1145/2912845.2912874.

[28] N. Omar, M. Albared, T. Al-Moslmi, and A. Al-Shabi, "A Comparative Study of Feature Selection and Machine Learning Algorithms for Arabic Sentiment Classification," in *Information Retrieval Technology*, Cham, 2014, pp. 429–443, doi: 10.1007/978-3-319-12844-3_37.

[29] S. Alhumoud, T. Albuhairi, and M. Altuwaijri, "Arabic sentiment analysis using WEKA a hybrid learning approach," in *2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*, Nov. 2015, vol. 01, pp. 402–408.

[30] S. Al-Osaimi and K. M. Badruddin, "Role of Emotion icons in Sentiment classification of Arabic Tweets," in *Proceedings of the 6th International Conference on Management of Emergent Digital EcoSystems*, New York, NY, USA, Sep. 2014, pp. 167–171, doi: 10.1145/2668260.2668281.

[31] K. Nalini and L. J. Sheela, "Classification of Tweets Using Text Classifier to Detect Cyber Bullying," in *Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2*, vol. 338, S. C. Satapathy, A. Govardhan, K. S. Raju, and J. K. Mandal, Eds. Cham: Springer International Publishing, 2015, pp. 637–645.

[32] W. Magdy and M. Eldesouky, "ClassStrength: A Multilingual Tool for Tweets Classification," in *2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2017, pp. 593–596.

[33] Department of Information Technology, CUSAT Kerala, India and V. Nandakumar, "CYBERBULLYING REVELATION IN TWITTER DATA USING NAÏVE BAYES CLASSIFIER ALGORITHM," *ijarcs*, vol. 9, no. 1, pp. 510–513, Feb. 2018, doi: 10.26483/ijarcs.v9i1.5396.

[34] H. M. Abdelaal, A. N. Elmahdy, A. A. Halawa, and H. A. Youness, "Improve the automatic classification accuracy for Arabic tweets using ensemble methods," *Journal of Electrical Systems and Information Technology*, vol. 5, no. 3, pp. 363–370, Dec. 2018, doi: 10.1016/j.jesit.2018.03.001

[35] B. Y. AlHarbi, M. S. AlHarbi, N. J. AlZahrani, M. M. Alsheail, J. F. Alshobaili and D. M. Ibrahim, "Automatic Cyber Bullying Detection in Arabic Social Media," *International Journal of Engineering Research and Technology,* vol. 12, no. 19, 2019.

[36] L. Cheng, J. Li, Y. Silva, D. Hall and H. Liu, "PI-Bully: Personalized Cyberbullying Detection with Peer Influence," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*.

[37] A. Bellmorea, A. J.Calvina, J. M. Xu and b. Zhub, "The five W's of "bullying" on Twitter: Who, What, Why, Where, and When," *Computers in Human Behavior,* vol. 44, pp. 305-314, 2015.

[38] H. L. Haoti Zhong, A. Squicciarini, S. Rajtmajer, C. Griffin, D. Miller and C. Caragea, "Content-Driven Detection of Cyberbullying on the Instagram Social Network".

[39] "Automated Arabic Language Processing: Arabic MPQA subjective lexicon & Arabic opinion holder corpus," *Automated Arabic Language Processing*, May 23, 2012. http://nlp4arabic.blogspot.com/2012/05/arabic-mpqa-subjective-lexicon-arabic.html (accessed Oct. 26, 2020).

[40] S. Khoja, "Arabic stemmer." http://zeus.cs.pacificu.edu/shereen/research.htm (accessed Dec. 31, 2020).

[41] "المدونة العربية- معلومات إضافية," *The Arabic linguistic code for King Abdulaziz City for Science and Technology*. https://corpus.kacst.edu.sa/more_info.jsp (accessed Oct. 26, 2020)

BIOGRAPHY

Samar Almutiry got Master degree from Taibah University in 2019 and bachelor degree from Taibah University in 2015. She worked in Saudi Aramco in June 2013-Agust 2013. She is currently a Ph.D. candidate and Computer teacher for high schools.

Mohamed Abdel Fattah received the B.Sc. and M.Sc. degrees in Electronics from the Faculty of Engineering, Cairo University, Cairo, Egypt, in 1994 and 2003, respectively, and the Ph.D. degree in information science and intelligent systems from the University of Tokushima, Japan, in 2007. He was awarded a Japan Society of the Promotion of Science (JSPS) postdoctoral fellowship from 2007 to 2009 in Department of Information Science and Intelligent Systems, Tokushima University. He is currently a Professor with FIE, Helwan University, Cairo. His research interests include information retrieval, natural language processing, speech recognition and document processing.

TRANSLATED ABSTRACT

كشف التنمر الإلكتروني باللغة العربية باستخدام تحليل المشاعر العربية

سمر المطيرى*[1] ، محمد عبد الفتاح**[2]

*كلية هندسة و علوم الحاسب، جامعة طيبة، المدينة المنورة، المملكة العربية السعودي**

[1]samar888abdullah@gmail.com

*قسم تكنولوجيا الإلكترونيات، كلية الهندسة، جامعة حلوان، القاهرة، مصر***

[2]maiahmed@taibahu.edu.sa

ملخص

تحليل المشاعر يتم استخدامة لتحليل النص ، واكتشاف الرأي ، وتصنيف النص. "تحليل المشاعر العربية (ASA)" صعبًا للغاية بسبب التعقيد اللغوي للغة العربية. من أجل تحليل المشاعر باللغة العربية ، نستخدم التقدم التكنولوجي في شكل التعلم الآلي (ML) وخوارزميات مختلفة لتدريب النظام علي مجموعة من البيانات التي تم جمعها من خلال ArabiTools , واجهة البرمجة الخاصة بتطبيق تويتر Twitter API . يتم تصنيف البيانات بوسائل ، تلقائية ويدوية ، للحصول علي نظام كفئ في اكتشاف تغريدات التنمر الإلكتروني.

يُعرف التنمر لإلكتروني باستخدام كلمات عدوانية ومهينة عبر وسائل التواصل الاجتماعي . يتم تصنيف البيانات تلقائيًا وفقًا لطبيعة التغريدة و محتواها فإذا كانت التغريدة تحتوي على كلمة أو أكثر من كلمات التنمر ، فسيتم تصنيفها على أنها "تنمر عبر الإنترنت" ، بينما إذا لم يتم العثور على أي كلمة ذات معنى عدواني ، فيتم تصنيفها على أنها "ليست تنمر عبر الإنترنت". بعد جمع البيانات ، هناك العديد من تقنيات المعالجة المسبقة للنص ، بما في ذلك تطبيع النص (Normalization) و الترميز (Tokenization) و تصنيف المفردات إلى أقسامها المكونة لها (اسم، فعل، صفة) Light Stemmer و اداه TF-IDF .الخوارزميات التي تم استخداها في النظام هي خوارزمية SVM ، مع WEKA و لغة البرمجة بايثون. هناك ثلاث تجارب تم إجراؤها باستخدام WEKA مع Light Stemmer ، والأخرى باستخدام WEKA مع اداه ArabicStemmerKhoja ، اخيرا بايثون مع SVM. اظهرت النتائج أن WEKA أكثر كفاءة في تصنيف النص بشكل صحيح ، في حين أن بايثون أكثر فاعلية مع مرور الوقت لبناء النظام.

الكلمات المفتاحية

التنمر الإلكتروني ، تصنيف النص ، النص العربي ، التعلم الآلي ، تحليل المشاعر ، آلة المتجهات الداعمة.