

## Comparison of Machine Learning Approaches on Arabic Twitter Sentiment Analysis

Merfat M. Altawaier<sup>#</sup>, Sabrina Tiun<sup>#</sup>

<sup>#</sup>CAIT, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia  
E-mail: [mervatm1976@gmail.com](mailto:mervatm1976@gmail.com), [sabrinatiun@ukm.edu.my](mailto:sabrinatiun@ukm.edu.my)

**Abstract**— With the dramatic expansion of information over the internet, users around the world express their opinion daily on the social network such as Facebook and Twitter. Large corporations nowadays invest on analyzing these opinions in order to assess their products or services by knowing the people feedback toward such business. The process of knowing users' opinions toward particular product or services whether positive or negative is called sentiment analysis. Arabic is one of the common languages that have been addressed regarding sentiment analysis. In the literature, several approaches have been proposed for Arabic sentiment analysis and most of these approaches are using machine learning techniques. Machine learning techniques are various and have different performances. Therefore, in this study, we try to identify a simple, but workable approach for Arabic sentiment analysis on Twitter. Hence, this study aims to investigate the machine learning technique in terms of Arabic sentiment analysis on Twitter. Three techniques have been used including Naïve Bayes, Decision Tree (DT) and Support Vector Machine (SVM). In addition, two simple sub-tasks pre-processing have been also used; Term Frequency-Inverse Document Frequency (TF-IDF) and Arabic stemming to get the heaviest weight term as the feature for tweet classification. TF-IDF aims to identify the most frequent words, whereas stemming aims to retrieve the stem of the word by removing the inflectional derivations. The dataset that has been used is Modern Arabic Corpus which consists of Arabic tweets. The performance of classification has been evaluated based on the information retrieval metrics precision, recall, and f-measure. The experimental results have shown that DT has outperformed the other techniques by obtaining 78% of f-measure.

**Keywords**— Arabic sentiment analysis; opinion mining; twitter data

### I. INTRODUCTION

Recently, most of companies have an essential need in terms of verifying its services or products. Such demands depend on the consumers' perspectives toward those services or products. Therefore, knowing the consumers' opinions has become very challenging in order to enhance the quality of services or products by gathering customers' feedback and classifying it in classes such as, negative, positive or neutral. With the dramatic expansion of the World Wide Web, the processing of investigating people's opinions has become more accessible and straightforward. Usually, that information would be in a textual mode. Thus, the use of recent technologies such as, web mining and semantic web facilitate the text analyzing which leads to extract the knowledge. Such process called sentiment analysis [1].

Subjectivity is the way that emotions and opinion can be expressed in the language while objectivity refers to the factual phrases. The problem of identifying documents whether it is subjective (yields opinion) or objective (yields

fact) is called subjectivity classification. However, there are two kinds of sentiment analysis; the first one is binary classification which aims to identify the opinions into pre-defined class labels (positive and negative). And the other one is multiclass classification which aims to identify the opinions into several classes such as (strong positive, positive, medium, negative and strong negative) or ranking by numbers 1, 2, 3, 4, and 5 (1-2 negative and 4-5 positive).

Moreover, sentiment analysis has several components which are the object which is a product or service that has been opinionated and feature which is a specific item of an object that may be opinionated by the consumers, and opinion orientation or so-called opinion polarity which refers to the opinion itself such as, negative or positive, and finally the opinion holder which refers to the person that provides the opinion. Discovering those components is depending on the nature of the sentiment application, some applications need to discover all the components while others need specific components.

Sentiment Analysis has been addressed for complex languages such as Arabic by many researchers [2], [3], [4], [5], [6], [7]. In fact, identifying the polarity of sentiments in Arabic is a challenging task due to several reasons [8]. First, Arabic has many dialects that are being used in social media such as Facebook and Twitter by users who write posts in an informal language. This can hinder the process of identifying the meaning of words. Second, the lack of available lexicons such as WordNet for Arabic language increases the complexity of determining the polarity of adjectives and adverbs. Third, in Arabic, the single word could yield multiple meaning in which the adjectives that indicate a positive polarity may be formulated as the negative polarity. For example, the word 'زهيب' which means 'incredible' is used in Arabic to indicate both positive and negative polarities.

Furthermore, one of the significant impacts related to sentiment analysis is the selected machine learning technique. In fact, researchers have taken the advantages of machine learning techniques which lie on the model that is being built based on training [20]. Such model aims to train on the feature space in order to discriminate instances in the testing. Hence, the use of machine learning techniques could bring valuable outcomes for the process of identifying sentiment polarity. However, there are many machine learning techniques that could be used for sentiment analysis such as Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), K-nearest Neighbors (k-NN) and others.

#### A. Related Works

The Social networks play an essential role in our daily life activates not in terms of social life but also in terms of e-commerce, e-learning, and politics. With the rapid increase of the users in the Middle East, the Arabic language is used with different dialects. So that, the Arabic slang languages has become a challenging task to because it still suffers from new expressive words and idioms that are presented in an unstructured format. An efficient technique is required to extract the opinions from the Arabic slangs language.

In the works on Arabic twitter analysis using machine learning techniques can be seen in such as the work of [8] which Gaussian kernel SVM classifier for Arabic language was proposed in order to classify Arabic news and user comments that was used on Facebook, in which, a Slang Sentimental Words and Idioms Lexicon (SSWIL) was implemented and developed. The proposed system for sentiment analysis approach aimed to collect the unstructured and ungrammatical customers' comments which were based on Arabic language and mining these comments using Sentiment Words and Idioms Lexicon (SSWIL) and support vector machines SVM to classify the comments. The tested dataset was collected manually from microblogs. The experimental result showed that the proposed system for Arabic sentiment analysis achieved a good result of 86.86% accuracy rate with the precision of 88.63% and recall of 78%.

Shoukry and Rafea [7] had applied semantic sentiment analysis on Twitter social network using three different Twitter database for their experiments. They proposed Semitic features in Twitter sentiment classification and explored three different approaches for incorporating them

into the analysis with replacement, augmentation, and interpolation.

In the same context, [6] presented a newly collected data set of 8,868 gold-standard annotated Arabic twitter feeds. The corpus was manually labeled for subjectivity and sentiment analysis (SSA). In addition, the corpus was annotated with a variety of linguistically motivated feature-sets that had previously shown a positive impact on classification performance.

Opinion Corpus for Arabic (OCA) has been proposed by [9] for Arabic reviews which were extracted from web pages that were related to movies and films. This OCA corpus had been translated into English language in order to generate the EVOCA, the opinion corpus for English. The proposed system has been tested on many machine learning algorithms: The Support Vector Machines (SVM) and Naïve Bayes (NB), using 10-fold cross-validation. In the experiment, the results indicated that EVOCA were worse than OCA but EVOCA is still comparable with English methods.

A framework for Arabic sentiment analysis proposed by [4] was able to analyze twitter comments or tweets in order to classify them in three categories positive, negative and neutral sentiments. The novelty of the proposed framework was that it can handle the Arabic dialects, Arabizi, and emoticons. The collected dataset contained 350,000 tweets; for each tweet the label was assigned, in addition, the voting was used to decide the final label for every tweet. There were three classifiers which were used to evaluate the performance of the proposed framework: Naïve Bayes(NB), support vector machines (SVM) and k-nearest neighbor (KNN). The experimental result showed that the framework achieved good results.

Based on the previous works on Arabic Twitter Analysis, several machine learning techniques have been used [10], [4], [11], [12], [13], [7], [8]. The techniques that commonly used are Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Conditional Random Fields (CRF). Such proposed techniques showed different performances due to their various mechanism of classification. In addition, several features have been exploited such as unigram, bigram, trigram, TF-IDF, Adjectives and Adverbs. The variety of these techniques and features makes the process of determining the suitable approach for Arabic sentiment analysis a challenging task. Therefore, identifying an appropriate technique for Arabic sentiment analysis is a challenging task.

In this study, we aim to establish a comparison study of machine learning technique for Arabic sentiment analysis on twitter using simple features, which is the word weight term. In order to obtain the feature, pre-processing processes consist of Term Frequency-Inverse Document Frequency (TF-IDF) and word stemming are involved. Whereas, the machine learning techniques that will be investigated in this study consist of Support Vector Machine (SVM), Naïve Bayes (NB) and Decision Tree (DT). The results obtained from this study can be used as a simple benchmark for other future advanced study or as an initial work for any Arabic Twitter Analysis.

## II. MATERIAL AND METHOD

This research aims to investigate the performance of machine learning technique for classifying sentiment reviews that have been collected from Twitter. The framework of the proposed method as shown in Fig. 1 consists of six main phases, namely; Dataset, Transformation, Pre-processing, Word Vector Representation, Classification, and Evaluation. Dataset phase aims to clarify the dataset by determining the details of its content. Transformation aims to convert the data into an appropriate form that enables processing. Pre-processing phase aims to normalize the data by eliminating the irrelevant data. Word vector representation phase aims to turn the data into a vector space which facilitate the process of creating features. Such phase consists of two sub-tasks, which are Term Frequency-Inverse Document Frequency (TF-IDF) and word stemming. Classification phase aims to carry out three machine learning techniques, which are Support Vector Machine (SVM), Naïve Bayes (NB) and Decision Tree (DT). Finally, evaluation phase aims to evaluate the proposed method by establishing a comparison among the three classifiers (see Fig. 1).

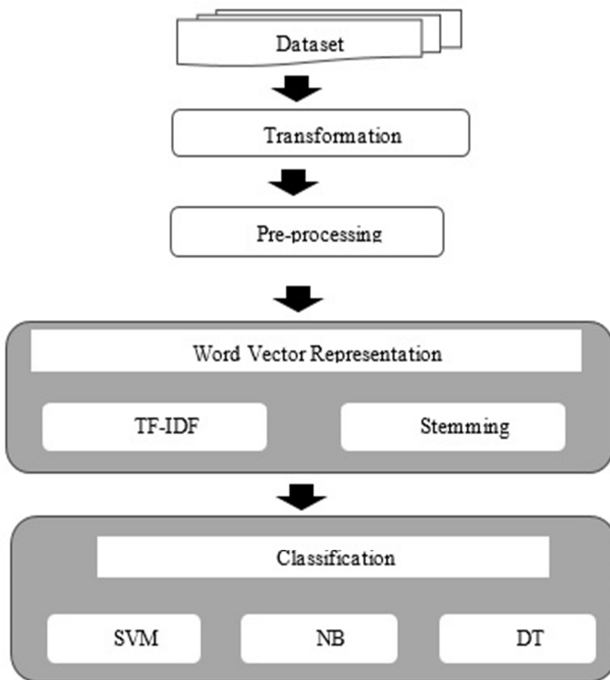


Fig.1 Research method framework

### B. Dataset

The dataset that has been used is a collection of Arabic tweets that were collected for the sentiment analysis purposes. Such dataset has been used by several researchers [2], [21], [16], [7]. This tweets dataset has been collected from UCI repository (<https://archive.ics.uci.edu/ml/datasets/Twitter>) which is a large repository for numerous datasets. The dataset contains 2000 labeled tweets (1000 positive tweets and 1000 negative tweets) related to two topics; 'politics' and 'arts'. The tweets in the datasets were written in Modern Standard Arabic (MSA). Table 1 depicts the details of such dataset.

TABLE I  
DATASET DETAILS

Details	Positive Class	Negative Class
Total tweets	1000	1000
Total words	7189	9769
Average words in each tweet	7.19	9.97
Average letters in each tweet	40.04	59.02

### C. Pre-processing

Due to the data is in Arabic language, an encoding task has to be performed. Encoding task aims to unify the encoding of letters to avoid the character appearance problem which is occurs when dealing with complicated language such as Arabic using ASCII Code encoding. Therefore, UTF-8 encoding has been used to do this purpose. UTF-8 is a character encoding capable of encoding all possible characters, or code points, in Unicode (Fig. 2).

1. # The MDI GUI
2. #mainclass = weka.ui.Main
3. # The GUIChooser
4. # The file encoding; use "utf-8" instead of "Cp1252" to display UTF-8
5. # GUI, e.g., the Explorer
6. FileEncoding = UTF-8

Fig. 2 Preferences adjusted of Weka for processing Unicode characters

In addition, a normalization task has been performed to normalize the data to remove the irrelevant data such as numbers, special characters, and stop-words. This data is unnecessary and does not affect the performance of classification thus; it has to be removed. For this purpose, the Arabic Morphological Analyzer (AraMorph) [22] tool has been used which is a tool for Arabic morphological analysis and the mentioned stop list has been embedded in Weka as in Fig. 3.

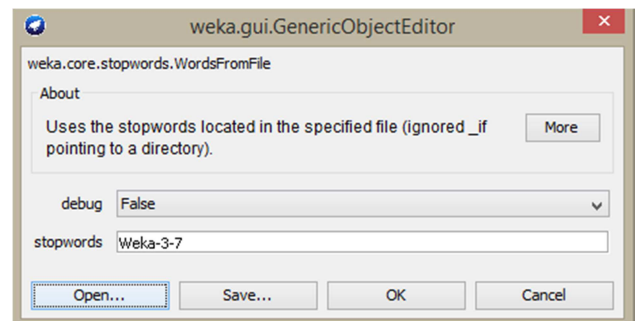


Fig. 3 Embedding Arabic stop-words list in Weka

This phase aims to represent the data in a suitable format for processing through the machine learning technique. In fact, this phase is crucial regarding to its significant impact on the performance of classification. Such representation is

performed by exploiting specific feature from the dataset. Such features are important to provide a numeric representation for the words in the data set. This phase consists of two sub-tasks which are illustrated in the following sub-sections.

1) *Term Frequency-Inverse Document Frequency (TF-IDF)*: Term Frequency Inverse Document Frequency is a useful feature in the text mining where the frequency of words is an indicator for important terms. In sentiment analysis, the frequency of terms plays a vital role in terms of identifying essential information [1]. For example, there are several words that can be frequently occurred such as ‘good’, ‘happy’, ‘mad’ and others. These words have a significant impact on identifying the polarity of opinion (i.e. whether positive or negative).

In fact, TFIDF contains the term frequency which is the number of the occurrences of the term in a given document [17]. It can be calculated as follow (see Eq. 1):

$$w_d(t) = TD(t, d) \quad (1)$$

Where TD is the frequency of term t in document d. In addition, TFIDF contains the inverse document frequency IDF which aims to provide high weight for rare conditions and low values for common conditions [18]. The formula as follows (see Eq. 2):

$$IDF_t = \log\left(\frac{N}{N_t}\right) \quad (2)$$

Where N is the number of documents in the documents and  $N_t$  is the number of documents that contain the term. Finally, TFIDF combines both the term frequency and inverse document frequency based on the following equation (Eq. 3):

$$w_t = TF(t, d).IDF_t \quad (3)$$

The TF-IDF is embedded already in Weka with other filters. Fig. 4 shows the application of TF-IDF in Weka in this study.

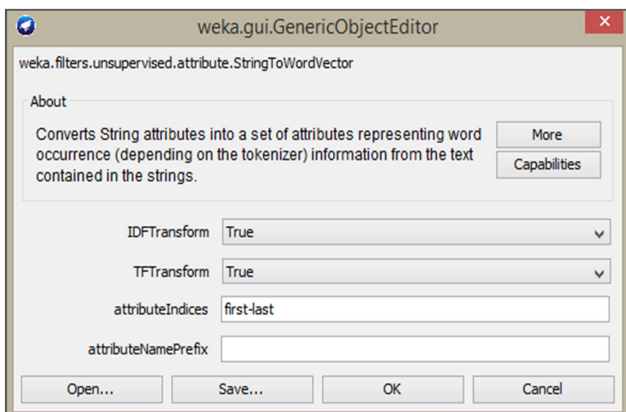


Fig. 4 Using TF-IDF in Weka

2) *Stemmed Words*: Stemmed word is the stem (root) of each word for instance, the word “reading” would be stemmed into “read”. For this purpose, Khoja stemmer [15] algorithm has been used in this study. The algorithm works by splitting a certain input word into prefix, stem, and suffix and then match these segments with the corresponding lexicon. presents an example of the stemming task and Fig. 5 depicts the successful insertion of Khoja stemmer in Weka.



Fig. 5 Insertion of Khoja Arabic stemming algorithm in Weka

Now, both of sub-tasks; TF-IDF calculation and Khoja stemming algorithm will be applied on each of words in the tweet. Such values will be used as weights for each word where the word with highest weights will be considered as a good feature for a tweet document. Fig. 6 shows a sample of results of this phase, in which the ‘weight sum’ value is used as the criteria to choose a word as a feature. In this example, the first four words show that these words have the highest weight sum, therefore they will be used as the feature for the particular tweet document.

وطن		
mean	0	0.0484
std. dev.	0.5729	0.5729
weight sum	999	995
precision	3.4372	3.4372
وفر		
mean	0	0.0339
std. dev.	0.6239	0.6239
weight sum	999	995
precision	3.7435	3.7435
وقد		
mean	0	0.0173
std. dev.	0.7176	0.7176
weight sum	999	995
precision	4.3056	4.3056
ولع		
mean	0	0.0096
std. dev.	0.7977	0.7977
weight sum	999	995
precision	4.786	4.786

Fig. 6 Sample results of TF-IDF and Khoja stemming sub-tasks processing

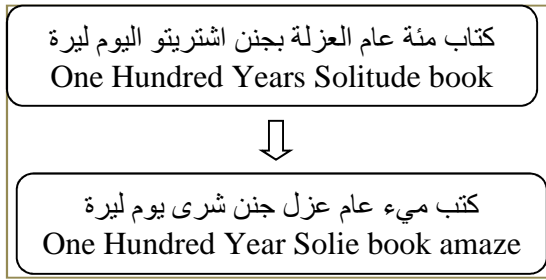


Fig. 7 Sample results of stemming processing

#### D. Classification

Supervised learning aims to train the data on certain pattern in order to identify it in the test part. This is a useful method in the field of sentiment analysis by train the data about a pattern that may indicate for whether the opinion is positive or negative. In this study, three classification techniques have been chosen including NB, SVM and DT. The experiments have been done using Weka [19], which is an open source application that enables user to carry out machine learning algorithms. The reason behind using such techniques lies on their effective ability to deal with text categorization where the number of features is huge [14].

### III. RESULT AND DISCUSSION

In this section, a comparative analysis has been performed in order to examine the performance of the three classification methods: NB, DT and SVM.

In the sentiment analysis field, most of the literature have used the common information retrieval metrics; precision, recall and f-measure in order to evaluate their methods [21]. This is due to evaluating the process of sentence classification is considered as an information retrieval task. However, the precision can be calculated as (Eq. 4):

$$Precision = \frac{TP}{(TP + FP)} \quad (4)$$

Where TP is the number of correctly classified sentences and FP is the number of incorrectly classified sentences. In addition, recall metric can be calculated as follows (Eq. 5):

$$Recall = \frac{TP}{(TP + FN)} \quad (5)$$

Where TP is the number of correctly classified sentences and FN is the number of sentences that were not classified at all. Now, it is possible to calculate the f-measure as follows (Eq. 6):

$$F - measure = \frac{Precision * Recall * 2}{(Precision + Recall)} \quad (6)$$

#### A. Naïve Bayes (NB) Classifier

NB classifier is applied with the proposed features (TF-IDF and Khoja stemmed words). In fact, the evaluation has been performed using the common information retrieval metrics which are precision, recall and f-measure. Table 2 shows such results.

TABLE II  
RESULTS OF NB

Training Portion	TP	FP	FN	Precision	Recall	F-measure
Positive	161	79	20	0.671	0.890	0.765
Negative	139	20	79	0.874	0.638	0.737
Average	300		99	0.782	0.752	0.750

As shown in Table 2, the precision of the positive sentences is (67.10%) less than the recall (89.00%), while the precision of negative sentences is (87.40%) more than the recall (63.80%). This is due the FN of positive sentences is equivalent to the FP of negative. In other meaning, because we have only two classes thus; when the classifier is incorrectly classified a sentence, it would classify it to the other class. Therefore, any incorrect match may increase the FP of class, leads to increase the FN of the other class similarly. Vice versa, any incorrect match may reduce the FP of class, leads to reduce the FN of the other class similarly. However, the average precision, recall and f-measure have achieved by NB classifier are 78%, 75% and 75% respectively.

#### B. J48 Decision Tree (DT) Classifier

J48-DT classifier is applied with the proposed features (TF-IDF and Khoja stemmed words) on the data. In fact, the evaluation has been performed using the common information retrieval metrics, which are precision, recall and f-measure. Table 3 shows the results.

TABLE III  
RESULTS OF DT

Training Portion	TP	FP	FN	Precision	Recall	F-measure
Positive	116	17	65	0.872	0.641	0.739
Negative	201	65	17	0.756	0.922	0.831
Average	317		82	0.809	0.794	0.789

As shown in Table 3, the precision of the positive sentences is (87.20%) more than the recall (64.10%), while the precision of negative sentences is (75.60%) less than the recall (92.20%). This is due the FN of positive sentences is equivalent to the FP of negative. In other meaning, because we have only two classes thus; when the classifier is incorrectly classify a sentence, it would classify it to the other class. Therefore, any incorrect match may increase the FP of class, leads to increase the FN of the other class similarly. Vice versa, any incorrect match may reduce the FP of class, leads to reduce the FN of the other class similarly. However, the average precision, recall and f-measure have achieved by DT classifier are 80%, 79% and 79% respectively.

### C. Support Vector Machine (SVM) Classifier

SVM classifier is applied with the proposed features (TF-IDF and Khoja stemmed words). In fact, the evaluation has been performed using the common information retrieval metrics which are precision, recall and f-measure. Table 4 shows such results. As shown in Table 4, the average precision, recall and f-measure have achieved by SVM classifier are 77%, 53% and 44% respectively.

TABLE IV  
RESULTS OF SVM

Training Portion	TP	FP	FN	Precision	Recall	F-measure
Positive	181	186	0	0.493	1	0.661
Negative	32	0	186	1	0.147	0.256
Average	210	32		0.777	0.534	0.440

### D. Comparative Analysis Between The Three Classifiers

A comparative analysis has been performed as well, in order to examine the performance of the three classification methods: NB, DT and SVM. Table 5 shows the results of this phase.

TABLE V  
RESULTS OF THREE CLASSIFIERS

Method	Precision	Recall	F-measure
NB	0.782	0.752	0.750
DT	<b>0.809</b>	<b>0.794</b>	<b>0.789</b>
SVM	0.777	0.534	0.440

As shown in Table 5, DT has outperformed the other classifiers by achieving 80% of precision, 79% of recall and 78% of f-measure. This was expected from the study of [23] where DT is outperforming NB when the number of classes is two. In addition, the study of [24] has concluded that SVM has a limitation when dealing with two classes. In fact, this is because SVM works by dividing the data into two classes. Therefore, the limitations lie on the margin of division or so-called hype-plane. Therefore, most of the SVM results refer to one class, which leads to increase precision and reduce the recall. Fig. 8 shows the comparison among the three classifiers.

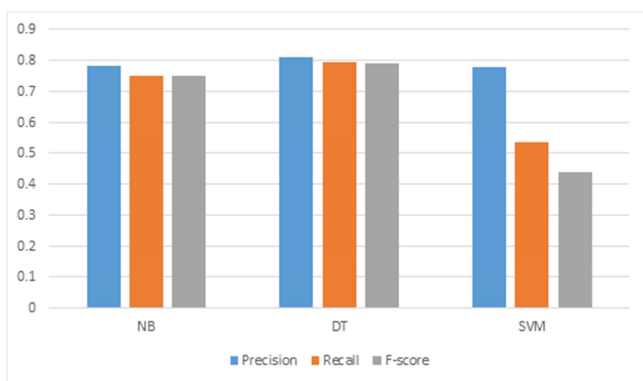


Fig. 8 Comparison among the three classifiers

## IV. CONCLUSION

This study investigates the performance of three machine learning (ML) techniques including Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT) when used on Arabic sentiment analysis based on a simple extracted feature. The dataset that has been used is a collection of Arabic tweets that were collected for the sentiment analysis purposes. It has been collected from UCI repository, which is a large-repository for numerous datasets. The dataset contains 2000 labelled tweets (1000 positive tweets and 1000 negative tweets) related to several topics including politics and arts. Results have shown that DT classifier has outperformed than the other ML by obtaining 78% of f-measure. Therefore, we can conclude that given a simple set of sub-tasks (TF-IDF and stemmed words) in extracting feature, Arabic sentiment analysis on two classes of opinions, will performed better if DT is used instead of SVM and NB. Thus, the study can be used as a base benchmark for more complex features and ML experiment of Arabic sentiment analysis. For future works, a lots of improvement can be done for this study, such as, by extending some factors such as follow: (i) Developing an Arabic lexicon corpus that covers several domains, and this could contributed toward more expanding and enhancing in the field of Arabic opinion question answering; (ii) exploiting more features such as emotional icons can be providing more accurate classification results by avoiding the incorrectly classified objects; (iii) using more machine learning techniques can contribute toward more efficient opinion question answering by reducing time and memory space; (iv) building a generic opinion question answering that can treat multi languages would be worth contribution. Another future work can be considered is the combination of multiple classifiers for sentiment analysis in Arabic in which DT, NB and SVM could be combined sequentially to produce better results.

## ACKNOWLEDGMENT

This research is partially funded by Malaysian Government under research grant: FRGS/1/2016/ICT02/UKM/02/14.

## REFERENCES

- [1] Pang, B. and Lee, L., "Opinion mining and sentiment analysis". *Foundations and trends in information retrieval* 2(1-2). 1-135, 2008.
- [2] Abdul-Mageed, M., Diab, M. T. and Korayem, M. 'Subjectivity and Sentiment Analysis of Modern Standard Arabic', *ACL (Short Papers)*, pp. 587-591, 2011.
- [3] Al-Subaihini, A. A., Al-Khalifa, H. S. and Al-Salman, A. S. "A proposed sentiment analysis tool for modern arabic using human-based computing". *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services*, 543-546, 2011.
- [4] Duwairi, R., Marji, R., Sha'ban, N. and Rushaidat, S., "Sentiment Analysis in Arabic tweets". *Information and Communication Systems (ICIS)*, 2014 5th International Conference on, 1-6, 2014.
- [5] El-Beltagy, S. R. and Ali, A., "Open issues in the sentiment analysis of arabic social media: A case study". *Innovations in Information Technology (IIT)*, 2013 9th International Conference on, 215-220, 2013.
- [6] Refaee, E. and Rieser, V., "An Arabic twitter corpus for subjectivity and sentiment analysis". *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*,

- Reykjavik, Iceland, may. European Language Resources Association (ELRA), 2014.
- [7] A. Shoukry and A. Rafea, "Preprocessing Egyptian Dialect Tweets for Sentiment Mining". *The Fourth Workshop on Computational Approaches to Arabic Script-based Languages*, pp. 47, 2012.
- [8] Soliman, T. H., Elmasry, M., Hedar, A. and Doss, M, "Sentiment Analysis of Arabic Slang Comments on Facebook". *International Journal Of Computers & Technology*, 12(5). 3470-3478, 2014.
- [9] Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña-López, L. A. and Perea-Ortega, J. M. "Bilingual experiments with an arabic-english corpus for opinion mining". 2011
- [10] Abbasi, A., Chen, H. and Salem, A, "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums". *ACM Transactions on Information Systems (TOIS)* 26(3). 12, 2008.
- [11] El-Halees, A., "Arabic opinion mining using combined classification approach". *International Arab Conference on Information Technology (ACIT2011)*, Riyadh, Saudi Arabia, 2011.
- [12] Elarnaoty, M., AbdelRahman, S. and Fahmy, A., "A Machine Learning Approach for Opinion Holder Extraction In Arabic Language". 2011.
- [13] Hamouda, A. E.-D. A. and El-taher, F. E., "Sentiment Analyzer for Arabic Comments System". *International Journal of Advanced Computer Science & Applications* 4(3), 2013.
- [14] Joachims, T. "Text categorization with support vector machines: Learning with many relevant features". *Machine Learning: ECML-98, Lecture Notes in Computer Science*, 1398, pp 137-142, 2005.
- [15] S. Khoja and R. Garside (1999) 'Stemming arabic text', Lancaster, UK, Computing Department, Lancaster University, (Access 1999).
- [16] M. McCord and M. Chuah, "Spam detection on twitter using traditional classifiers". *Autonomic and trusted computing, Springer*, pp. 175-186, 2011.
- [17] Yamamoto, M. and Church, K. W., "Using suffix arrays to compute term frequency and document frequency for all substrings in a corpus". *Computational Linguistics* 27(1). 1-30, 2001.
- [18] Tokunaga, T. and Makoto, I. "Text categorization based on weighted inverse document frequency". *Special Interest Groups and Information Process Society of Japan (SIG-IPSI)*, 1994.
- [19] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H., "The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* , 11(1). 10-18, 2009.
- [20] Cambria, E., Schuller, B., Xia, Y. and Havasi, C., "New avenues in opinion mining and sentiment analysis". *IEEE Intelligent Systems* (2). 15-21, 2013.
- [21] Gebremeskel, G., "Sentiment Analysis of Twitter Posts About news". *Sentiment Analysis. Feb.*, 2011.
- [22] Buckwalter, T., *Buckwalter {Arabic} Morphological Analyzer Version 1.0.*, 2002.
- [23] Amor, N. B., Benferhat, S. and Elouedi, Z. "Naive bayes vs decision trees in intrusion detection systems". *Proceedings of the 2004 ACM symposium on Applied computing*, 420-424, 2004.
- [24] Huang, J., Lu, J. and Ling, C. X. "Comparing naive Bayes, decision trees, and SVM with AUC and accuracy". *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, 553-556, 2003.