Sentiment Analysis of Arabic Jordanian Dialect Tweets

Jalal Omer Atoum[1], Mais Nouman[2]

Computer Science Department

Princess Sumaya University for Technology, Amman, Jordan

Abstract—Sentiment Analysis (SA) of social media contents has become one of the growing areas of research in data mining. SA provides the ability of text mining the public opinions of a subjective manner in real time. This paper proposes a SA model of Arabic Jordanian dialect tweets. Tweets are annotated on three different classes; positive, negative, and neutral. Support Vector Machines (SVM) and Naïve Bayes (NB) are used as supervised machine learning classification tools. Preprocessing of such tweets for SA is done via; cleaning noisy tweets, normalization, tokenization, namely, Entity Recognition, removing stop words, and stemming. The results of the experiments conducted on this model showed encouraging outcomes when Arabic light stemmer/segment is applied on Arabic Jordanian dialect tweets. Also, the results showed that SVM has better performance than NB on such tweets' classifications.

Keywords—Sentiment analysis; Arabic Jordanian dialect; tweets; machine learning; text mining

I. INTRODUCTION

Sentiment Analysis (SA) or opinion mining is defined as the task of finding authors' opinion with respect to a topic or issue. Also, it is focused on analyzing sentences or classifying texts into either positive, negative, or neutral opinions. Furthermore, SA has gain high popularity in recent years to analyze and benefit from the available data that exit on online social media such as blogs, wikis, and tweeter [1]. Such analysis could be based on knowledge or statistics [2]. Finally, SA requires dealing with many natural language processing issues such as conceptual primitives [3], sarcasm [4], aspects-based [5], and subjectivity detection [6].

Arabic sentiment analysis of tweets may include the opinion of the public in regard to a specific topic [7], [8]. The performance of Arabic SA tools have become deeply engaged with the compatibility of the social media availability. Researchers have been working on sentiment analysis and opinion mining using different tools to define people's views and comments from negativity, positivity and neutrality opinions. Our research considers the sentiment analysis of how Jordanians, using their own dialect of local idioms and words, react to trends and news over Twitter.

Our target is to establish a SA model for classification of Arabic Jordanian dialect tweets into either negative, positive, or neutral, by recognizing words; named entities, stop words, and stemmers. To accomplish this task, we have collected tweets according to their locations, then we filtered these tweets to collect different types of terminologies in order to identify Jordanian Arabic dialect efficiently.

The rest of the paper is organized as follows: Section two lists some related work of sentiment analysis that emphasizes on Arabic SA. Section three discusses some background concepts needed for this research such as machine learning classification techniques. Section four presents the proposed Arabic Jordanian dialect tweets SA model. Section five provides the evaluation measures, the experimental results and the evaluation of this model. Finally, Section six presents the conclusions and some future work.

II. RELATED WORK

There are many researches that have considered Arabic sentiment analysis. In [9], the authors proposed a hybrid approach which combines SVM and semantic orientation on Egyptian dialect corpus of tweets. In [10], the authors presented a model for sentiment analysis of Saudi Arabic tweets to extract feedback from Mubasher products. In [11], the authors developed Corpus for Arabic Sentiment Analysis of Saudi Tweets. In [12], the authors explained how mining social networks can be done on Arabic Slang comments by proposing a SVM based classifier that applies sentiment analysis to classify youth news comments on Facebook.

The authors in [13], studied the effect of social media (Libyan tweets) during the Arab Spring based on two sources of information; the language of leaders in public speeches, and the language of the public in social media. Some researches such as in [14], have focused on Arabic opinion mining using a combined approach of lexicon based method and k-nearest method for classifying documents.

This research is similar to the work of [15] that designed a framework for data collection, statistical analysis, sentiment analysis, and language model comparison to understand the interests of Twitter users towards news headlines. However, we differ by not using the behavior of the English language entities. Instead, we use the behavior of Arabic Jordanian dialect. Also, this research provides a formulation of the opinion mining problem, identifies the key pieces of information that should be mined, and describes how a structured opinion summary can be produced from unstructured tweets. This research is also different from other related researches by focusing specifically on Arabic Jordanian dialect tweets written by Twitter users who comment and writes special local idioms and words that are

mainly used in Jordan. Hence, our research takes such locality of words and idioms into consideration during SA.

III. Background

Machine Learning (ML) has two main approaches, supervised learning and unsupervised learning. The problem with unsupervised machine learning is that they may overlap and learn to localize tweets with minimal unsupervised algorithms. Therefore, we have used two supervised ML approaches for the classifications of Arabic Jordanian dialect tweets, namely, Naïve Bayes (NB) and Support Vector Machine (SVM).

Naïve Bayes (NB) is a learning method in which it introduces the multinomial model, or a probabilistic learning method. NB often relies on the bag of words presentation of a document, where it collects the most used words neglecting other infrequent words. Bag of words depends on the feature extraction method to provide the classification of some data [16]. Furthermore, NB has a language modeling that divides each text as a representation of unigram, bigram, or trigram and tests the probability of the query corresponding with a specific document.

Support Vector Machine (SVM) makes non-probabilistic binary vectors as a learning algorithm to be applied for classification. The most important models for SVM text classifications are Linear and Radial Basis functions. Linear classification tends to train the data-set then builds a model that assigns classes or categories [17]. It represents the features as points in space predicted to one of the assigned classes. SVM provides good classification performance in several fields; but mostly applied for image recognition and text classification.

Many researchers used supervised learning approaches on data related to publically released corpuses for Arabic SA [11]. Such researches use the Modern Standard Arabic (MSA). MSA can handle neutral indications in which they are written in questions' forms or as unknown purposes such as the phrase (الحمدالله Praised God), in which it could mean something good or bad depending on the mode of the writer.

Furthermore, various methods of Arabic text mining have been discussed and researched. A hybrid approach of sentiment analysis by [9] combines SVM and Semantic Orientation (SO). The challenges that they faced for a given word are in terms of its root or its spelling or its different meanings.

Other classifications for SA are based on predicted classes and polarity, and/or on the level of classification (sentence or document). Lexicon based SA text extraction is annotated with semantic orientation polarity and strength. SA proved that light stemming comes in handy for the accuracy and for the performance of classification [18].

Finally, an automatic classifier of Arabic text documents based NB and SVM algorithms was presented in [19], the results indicated that the SVM algorithm handled the text documents classification better than the NB algorithm.

IV. Proposed Arabic Jordanian Dialect Tweets SA Model

In this section, we present our proposed model that analyzes, mines, and classifies Arabic Jordanian dialect tweets as illustrated in Fig. 1.

This model consists of the following phases:

A. Collecting Tweets

A connection to Twitter is created in order to collect a corpus of Jordanian dialect tweets. A read only application is built to collect written tweets from Twitter. The collected tweets are based on the following parameters; users timeline, home timeline, trends, and searching for queries. For each parameter, a corpus of 1000 tweets is collected.

B. Tweets Extraction

Tweets extraction helps in extracting the important content of a tweet (the essence). Hence, what is needed from a tweet is written after the hash-tags, and subsequently extracting the feature words, words that carry a message for the user whether it is a positive, negative, or neutral tweet. Also, tweets extraction is needed to facilitate analyzing the features vector and selection process (unigrams, bigrams and trigrams), and to facilitate the classification of both training and testing sets of tweets.

C. Cleaning and Tweets Annotations

All 'http/https' shortening, and special symbols such as (*, &, $, %,-,_,:,!,><) are removed from the collected tweets. Then each special character is replaced with a space character.

D. Tweets Preprocessing

Several preprocessing stages have to be done on the collected tweets in order for the SA process to be more effective. These stages are as follows:
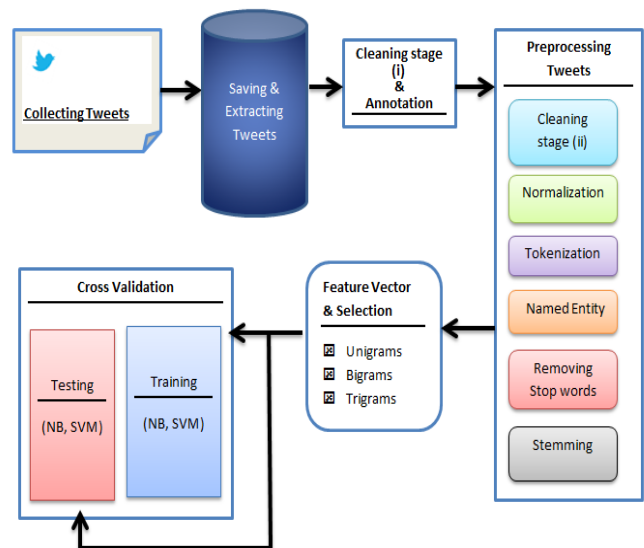

Fig. 1. Arabic Jordanian Dialect Tweets SA Model.

1) Cleaning stage: Dealing with Arabic characters and letters needs further cleaning process, so we have to continue cleaning the tweets for each line that contains special symbols, and various characters such as emoticons. Those symbols and characters may lead us to a different classification from what the user is intended in the tweet. Hence, we took the emoticons and emotion characters into consideration during the sentiment classification. Table 1 presents some special symbols that we have used along with their meanings and sentiments. Finally, the cleaned extracted tweets are stored into a database in a comma separated values format for further manipulation.

The annotation process of the collected tweets is done by an Arab Jordanian student. They were fully aware of the Jordanian dialect meanings and domains. As a result of this annotation, they labeled each tweet with either positive, negative, or neutral.

Positive tweets may include good words indicators that express happy feelings, and/or love, excitement, etc. For example, the following tweets are positive tweets:

- جمعة مباركة بالخير للجميع (Blessed Friday for everyone)

- افتخر ببلدي, اللهم زده من نعم (Proud of my country, God raise it with blesses)

Negative tweets may include bad words, negativity indicators that express bad feelings, sad, depressed, and/or anger. For example, the following tweets are negative tweets:

- زحلق من هون ما بدي شوفك (Move from here I don't want to see you)

- انطم واسكت ما بتعرف تركب كلمتين على بعض (Shut up, you don't know how to put two words together)

Neutral tweets may not include positive nor negative words, they may include questions about something, uncertainty, etc. Also, a tweet is considered neutral if it does not bear any opinion. For example, the following tweets are neutral tweets:

- اترك من ايدك, انا بكمل (Leave your hands out of it, I will continue)

- الجو معتدل هاليومين (The weather is fine these two days)

2) Normalization: Normalization of tweets is needed since there is no single convention of spelling of some Arabic letters, for example either one of the letters (ة Taa) or (ه haa) could be used as the last letter of a word. Hence, if we have the word (**"نهفة"** Funny), then its normalized form is **"نهفه"**. This normalization stage starts by removing all extra spaces, then the occurrence of any un-normalized letter is replaced with its normalized form as shown in Table 2.

All non-standard words that have numbers and/or dates are identified. Such words would be mapped into special built in vocabularies. This results in smaller number of Arabic tweet vocabularies and improves the accuracy of the classification task.

TABLE I. SOME SYMBOLS AND THEIR MEANINGS

Character/Symbol	Meaning	Sentiment
♥	Heart or Love	Positive
☺	Smile	Positive
❀	Snow	Neutral, Positive, Negative
?	Question	Neutral
✈	Bird or Airplane	Neutral

TABLE II. NORMALIZATION OF SOME ARABIC LETTERS

UnNormalized Arabic Letters	Normalized Letters
ؤ، ىء Hamza	ء
ا، أ Alef	ا
لا، لأ Lamalef	لا
ى Yah	ى
ة Hah	ه
ؤ Wow_hamzah	و

Some Arabic words have several Tatweels (stressing a literal by repeating it several times). The occurrences of such a letter in a word are replaced by only one occurrence. For example, the word (ألمــــــــا) (name of a person) would be normalized into (ألمَا).

Moreover, Tashkeel (added vocalizations or diacritization on Arabic alphabets) could affect the performance of the stemmers, and the performance of the classification process. Hence, words with Tashkeels will be replaced with their corresponding words without Tashkeels. Table 3 presents some examples of Arabic words in their un-normalized vocalized Tashkeel forms and their corresponding unvocalized normalized forms.

3) Tokenization: Tokenization is an important step in SA since it reduces the typographical variation of words. Feature extraction process and the bag of words require tokenization. Dealing with Arabic language requires a high level component that uses a dictionary of features that transforms these words into feature vectors, or feature indices; such that the index of the feature (word) in the vocabulary is linked to its frequency in the whole training corpus.

4) Named Entity Recognition (NER): NER is a significant tool in Natural Language Processing (NLP); it allows the identification of proper nouns in an unstructured text. NER has three categories of name entities; ENAMEX (person, organization, and country), TIMEX (date and time), and NUMEX (percentages and numbers). For example the word "أردن" means a country name (Jordan), and the word "الأحد" means the name of day in the week in Arabic (Sunday).

TABLE III.   SOME NORMALIZED TASHKEEL WORDS

Tashkeel Symbol	UnNormalized Example	Normalized Example
◌ُ Dama	ضُمه	ضمه
◌ّ Shada	شدّه	شده
◌ْ Skoon	سكوْن	سكون

5) Some of the Arabic language names start with " ال The" which helps in defining each word that starts with these two letters as a named word. Furthermore, NER helps in making the classification faster than processing the whole sentence. Hence, we have built a function to define Arabic named entities, so that in a future work we would apply SA only on those Arabic names. We estimated the frequency of each name in the tweets to see the most frequent names being used, and we measured the probability of having these names in positive tweets, in negative tweets, and/or in neutral tweets.

6) Removing stop words: Some stop words can help in attaining the full meaning of a tweet and some of them are just extra characters that need to be removed. Some examples of the Jordanian dialect stop words that don't affect the tweets meaning and can be removed from tweets are:
هداك/هضاك this/that، امتى/امتن when، بس enough، شوي little،
كمان more }

7) Stemming:- The last stage in the preprocessing of Arabic Jordanian dialect tweets is stemming. It is done by removing any attached suffixes, prefixes, and/or infixes from words in tweets. A stemmed word represents a broader concept to the original word, also it may lead to save storage [11]. The goal of stemming tweets is to reduce the derived or inflected words into their stems, base or root form in order to improve SA. Furthermore, stemming helps in putting all the variation of a word into one bucket, effectively decreasing our entropy and gives better concepts to the data.

Tashaphyne is an Arabic light stemmer/segment tool developed by [20] for exploring the sentiment analysis of Arabic roots. This tool has demonstrated the potential of mapping Arabic words into their basic roots for the process of SA task, showing noteworthy improvements to baseline performance [21]. Hence, we have used this tool as a useful light stemmer for our Arabic Jordanian dialect tweets. For instance, in Arabic Jordanian dialect the word "نشمي" which means a man with good Jordanian manners would be stemmed into "نشم".

Moreover, N-gram is a traditional method that takes into consideration the occurrences of N-words in a tweet and has the ability to identify formal expressions [22]. Hence, we have used N-gram in our SA.

Finally, in this research, we have implemented the term frequency using *weka* [23]. Term frequency assigns weights for each term in a document in which it depends on the number of occurrences of the term in a document, and it gives more weight to those terms that appear more frequent in tweets because these terms represent words and language patterns that are more used by the Arabic tweeters.

V. EXPERMINAL RESULTS AND EVALUATION

Several experiments have been conducted to compare the performance of Naïve Bayes (NB) and Support Vector Machines (SVM) classifiers of Arabic Jordanian dialect tweets. The classifications are conducted on three balanced and unbalanced classes namely; positive, negative, and neutral tweets. We have used a total of 3550 Jordanian dialect tweets as follows; 616 positive tweets, 1313 negative tweets, and 1621 neutral tweets. The first experiment used un-stemmed tweets with unigram feature, the second experiment used stemmed tweets, the third experiment used rooted tweets, the fourth experiment used stemmed and rooted bigram tweets, the fifth experiment used stemmed rooted tri-gram tweets, and the final sixth experiment used stemmed and rooted n-grams tweets.

Three measures of sentiment classification performance are used, namely; Accuracy, Precision, and Recall. In addition the Receiver Operating Characteristics (ROC) introduced in [24] is also used to measure the performance of the classifiers. The ROC graphs are used to visualize, organize, and select classifications based on the performance. The difference between the ROC and accuracy is that the ROC is helpful in managing unbalanced instances of classes, whereas, the accuracy is a single number to sum up the performance. Finally, we have used, also, the F-measure to evaluate the effectiveness of our proposed sentiment model of Arabic Jordanian dialect tweets.

Furthermore, in order to evaluate the performance of sentiment analysis model, cross validation is used in which 10-fold equal sized sets are produced. Each set is divided into two groups, training and testing, the testing set is taken by 10-fold from the training tweets.

The results obtained from conducting these experiments are shown in Table 4. From this table, it is shown that the SVM classifier performs better than the NB classifier in all measures of every experiment. Both classifiers have better performance on all measures when the set of tweets were balanced.

The ROC performance reached an average of 0.71 on NB and an average of 0.77 on SVM on all experiments, which are considered to be good values taking into account the instances used and the prediction of data since ROC compares the true positive and false positive rates, which is the fraction of sensitivity or recall in machine learning.

It is also noticed that the classification results using the unigrams experiments are better than using the bigrams experiments in SVM and vice versa in NB. This is due to the fact that NB classifies according to the probabilities, in addition to the fact that using bigrams would increase the probability of estimation.

TABLE IV. EXPERIMENTAL RESULTS

Measure		1st Experiment W/O Stems & With Unigrams		2nd Experiment With Stems		3rd Experiment With Roots		4th Experiment With Stems & Roots Bigrams		5th Experiment With Stems & Roots Trigrams		6th Experiment With Stems & Roots N-grams	
		Balance	UnBal	Balance	UnBal	Balance	UnBal	Balance	UnBal	Balance	UnBal	Balance	UnBal
Precision	NB	0.53	0.5	0.55	0.5	0.49	0.46	0.64	0.68	0.68	0.71	0.51	0.52
	SVM	85	0.67	0.83	0.65	0.75	0.54	0.77	0.8	0.82	0.79	0.73	0.65
Recall	NB	0.53	0.5	0.54	0.5	0.47	0.43	0.55	0.58	0.55	0.57	0.5	0.49
	SVM	0.84	0.67	0.83	0.65	0.75	0.55	0.74	0.75	0.76	0.7	0.73	0.64
F-Measure	NB	0.53	0.5	0.54	0.5	0.47	0.43	0.5	0.52	0.5	0.5	0.5	0.5
	SVM	0.84	0.66	0.83	0.65	0.75	0.54	0.74	0.74	0.75	0.67	0.72	0.64
Roc-Area	NB	0.7	0.65	0.7	0.65	0.64	0.61	0.74	0.79	0.86	0.83	0.67	0.67
	SVM	0.87	0.73	0.82	0.73	0.8	0.64	0.78	0.79	0.79	0.74	0.78	0.72
Accuracy	NB	0.53	0.5	0.54	0.5	0.47	0.43	0.55	0.58	0.55	0.57	0.5	0.5
	SVM	84	0.66	0.82	0.66	0.75	0.55	0.74	0.75	0.76	0.69	0.73	0.64

For the evaluation of bigram and trigram experiments, it is noticed that the measures obtained from using stemmed trigrams' features experiment are lower than those of using bigrams experiment, but in the rooted experiment, the performance of trigram is higher with accuracy of 55% in NB and of 76% in SVM. Hence, we can conclude that rooted balanced trigrams perform better than unigrams and bigrams.

A final experiment was conducted to determine the optimum threshold (ROC curve) for each stemmed unigram features using SVM. We have tried different term frequency thresholds as features until we got the best results of the ROC area with different values for each positive, negative, and neutral. Fig. 2 shows the optimum threshold for positive, negative, and neutral stemmed unigram features.

In this final experiment, the thresholds curve is applied for each class with its corresponding Area Under Curve (AUC) plot, positive tweets have AUC value of 0.8904, negative tweets have AUC value of 0.8666, and neutral tweets have AUC value of 0.8666. Hence, positive tweets have performed more accurately than the other sentiments; and the amount of data that positive tweets holds are more accurate to predict the correct positive sentiments.
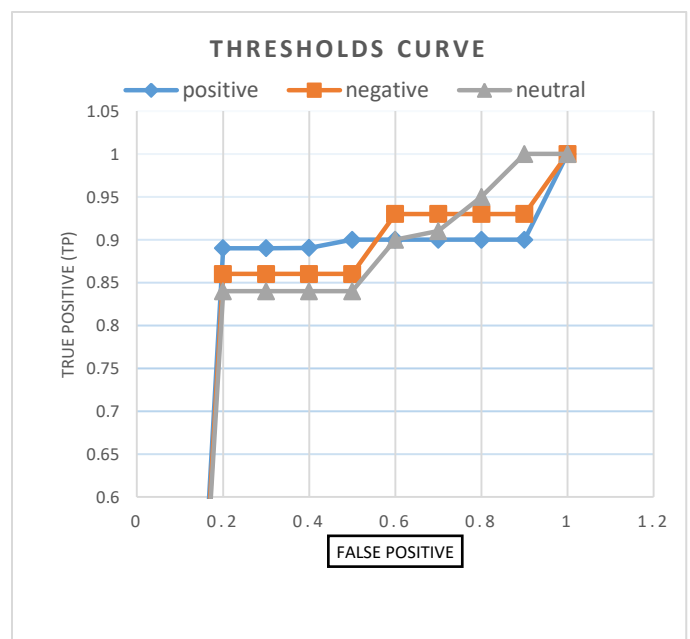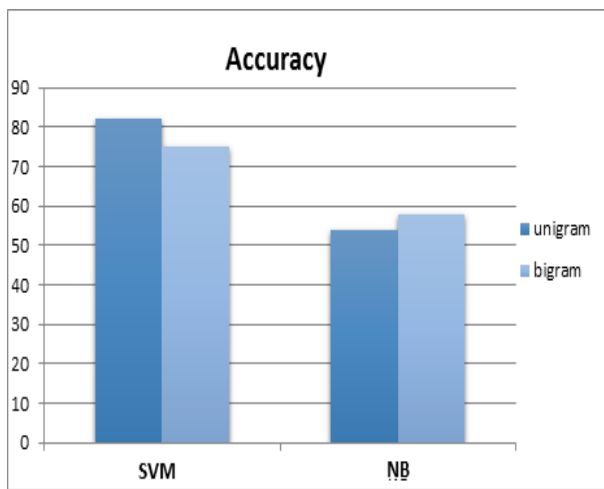

Fig. 2. Optimum Thresholds Curve.

Fig. 3. SVM and NB Accuracy of Stemmed Tweets.

Finally, Fig. 3 illustrates the accuracy differences between SVM and NB classifiers, and proves that classification with SVM provides us with higher accuracies in both cases of stemmed unigrams and bigrams.

VI. CONCLUSION AND FUTURE WORK

Sentiment analysis or opinion mining has increasingly evolved since the growth of social media networks; it is the process of evaluating the person's feelings to a specific subject.

The sentiment analysis model we have proposed in this research is based on three classes/labels; positive, negative, and neutral. Our model of sentiment analysis started with collecting and extracting of Arabic Jordanian dialect tweets followed by cleaning and annotating of such tweets. Then, these tweets were gone through various steps of preprocessing that includes; normalization, tokenization, name entity recognition, removing of stop word, and stemming.

Several experiments were conducted on the proposed model using supervised ML. We conclude that classifications using SVM on Arabic light stemming always yield better results than using NB. Furthermore, imposing a balance between the three classes, and reducing the number of instances to the most used instances improved the accuracy. The outcomes were very promising as SVM achieved an accuracy of 82.1%.

Despite the fact that rooted experiments have difficulties in correctly classifying Arabic Jordanian dialect tweets, we took the annotation and preprocessing steps very seriously.

Then we demonstrated the positive effect of classification using light stemming mechanism on the un-stemmed tweets. We have used all the polarity of classes on both (stemmed and un-stemmed tweets). After that we tested the same mechanism on the rooted tweets. Our experiments' results proved that stemming affects the accuracy of tweets' classifications.

Furthermore, the conducted experiments showed many aspects and beneficial insights about Arabic Jordanian dialect users. For example; we discovered that the amount of negative emotions the users obtain is more than positive emotions.

Finally, one direction to extend this research would be the improvement of light stemming, and rooting mechanism, by monitoring the performance of each rule on Jordanian dialects to test the improvement of the overall performance and the classification process. Another direction for future research is to apply the semantic orientation approach and building a list of positive, negative, and neutral tweets for better understanding of the Jordanian Arabic dialect tweets.

REFERENCES

[1] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, "Affective computing and sentiment analysis," in A Practical Guide to Sentiment Analysis, vol. 5 of Socio-Affective Computing, pp. 1–10, Springer, Cham, Switzerland, 2017.

[2] E. Cambria, and A Hussain, "Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis," Vol. 1, Springer, Cham, Switzerland, 2015. ISBN: 978-3-319-23654-4.

[3] E. Cambria, S. Poria, R. Bajpai, and B. Schuller, "Senticet 4: a semantic resource for sentiment analysis based on conceptual primitives," In proceeding of the COLING, 2016, The 26th International Conference on Computational Linguistics: Technical papers, pp. 2666–2677.

[4] S. Poria, E. Cambria, D. Hazarika, and P. Vij, "A deeper look into sarcastic tweets using deep convolutional neural networks," In proceeding of the COLING, 2016, The 26th International Conference on Computational Linguistics: pp. 1601–1612.

[5] Y. Ma, H. Peng, and E. Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM," in Proceedings of the 32nd (AAAI) conference on Artificial Intelligence, (2018). Pp. 5876-5883.

[6] I. Chaturvedi, E. Cambria, R. Welsch, and F. Herrera, "Distinguishing between facts and opinions for sentiment analysis: Survey and challenges," Information Fusion 44 (2018) 65–77.

[7] A. AlOwisheq, S. AlHumoud, N. AlTwairesh, and T. AlBuhairi, "Arabic Sentiment Analysis Resources: A Survey," In: Meiselwitz G. (eds) Social Computing and Social Media. SCSM. pp. 267–278. Springer, 2016, Toronto Canada.

[8] I. Awajan, and M. Mohamad, "A Review on Sentiment Analysis in Arabic Using Document Level," International Journal of Engineering and Technology. Vol. 7, No. 3.13 (2018): Special Issue, pp. 128-132.

[9] A. Shoukry, and A. Rafea, "A Hybrid Approach for Sentiment Classification of Egyptian Dialect Tweets," In: First International Conference on Arabic Computational Linguistics (ACLing). pp. 78–85. , Cairo, Egypt (2015).

[10] H. Al-Rubaiee, R. Qiu, and D. Li, "Identifying Mubasher Software Products Through Sentiment Analysis of Arabic Tweets," Presented at the 2016 International Conference on Industrial Informatics and Computer Systems (CIICS), 2016, pp. 1–6.

[11] N. Al-Twairesh, H. Al-Khalifa, A. Al-Salman, andY. Al-Ohali, "AraSenTi-Tweet: A Corpus for Arabic Sentiment Analysis of Saudi Tweets," In Proceeding of the 3rd International Conference on Arabic Computational Linguistics, ACLing 2017, 5-6 November 2017, Dubai, United Arab Emirates. pp. 63–72.

[12] T. Hassan, A. Soliman, and A. Ali, "Mining Social Networks' Arabic Slang Comments," In Proceedings of IADIS European Conference on Data Mining 2013 (ECDM'13), pp.22–24, (2013)

[13] C. Brown, J. Frazee, D. Beaver, X. Liu, F. Hoyt, and J. Hancock, "Evolution of Sentiment in the Libyan Revolution," (2011), White Paper at https://webspace.utexas.edu/dib97/libya-report-10-30-11.pdf

[14] A. El-Halees, "Arabic Opinion Mining Using Combined Classification Approach," In the Proceedings of the International Arab Conference on Information Technology (ACIT'2011), Naif Arab University for Security Science (NAUSS), Riyadh, Saudi Arabia December 11-14, pp. 264-271, 2011.

[15] L. Larkey, L. Ballesteros, and M. Connell, "Improving Stemming for Arabic Information Retrieval: Light Stemming and Co-occurrence Analysis," In Proceedings of the 25th annual international conference on research and development in information retrieval (SIGIR 2002), Tampere, Finland, August 11–15, 2002, pp. 275–282.

[16] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian Network Classifiers", Machine learning, Vol. 29, No. 2–3, pp. 131-163, 1997.

[17] C. Cortes, and V. Vapnik. "Support-vector networks". Machine Learning,. Vol. 20, No. 3, pp. 273–297,1995 doi:10.1007/BF00994018.

[18] N. Abdulla, N. Ahmed, M. Shehab, and M. Al-Ayyoub, "Arabic Sentiment Analysis: Lexicon-Based and Corpus-Based," In Proceedings of the 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies AEECT pp. 1-6.

[19] S. Alsaleem, "Automated Arabic Text Categorization Using SVM and NB," International Arab Journal of e-Technology, Vol. 2, No. 2, pp. 124-128, 2011.

[20] T. Zerrouki, (2010). "Tashaphyne, Arabic Light Stemmer/Segment". http://tashaphyne.sourceforge.net.

[21] S. Oraby, Y. El-Sonbaty, M. El-Nasr, "Exploring the Effects of Word Roots for Arabic Sentiment Analysis," In Proceedings of the Sixth International Joint Conference on Natural Language Processing, 2013, pp. 471-479.

[22] L. Chen, W. Wang, M. Nagaraja, S. Wang, and A. Sheth, "Beyond Positive/Negative Classification: Automatic Extraction of Sentiment Clues from Microblogs,", Kno.e.sis Center, Technical Report, 2011.

[23] G. Holmes, A. Donkin, and I. Witten, "WEKA: A Machine Learning Workbench," In Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information Systems, Brisbane, 29 November-2 December 1994, 357-361.Khalifa, K., and Omar, N., "A Hybrid Method Using Lexicon-Based Approach and Naïve Bayes Classifier for Arabic Opinion Question Answering," Journal of Computer Science 10 (11): 1961-1968, 2014 ISSN: 1549-3636.

[24] D. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," Journal of Machine Learning Technologies. 2 (1): 37–63, 2011.