

## Hybrid approach for sentiment analysis of Arabic tweets based on deep learning model and features weighting



Altyeb Altaher\*

Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Jeddah, Saudi Arabia

### ARTICLE INFO

#### Article history:

Received 6 May 2017

Received in revised form

13 June 2017

Accepted 3 July 2017

#### Keywords:

Arabic text

Deep learning

Sentiment analysis

Features weighting

Classification algorithms

### ABSTRACT

The increasing adoption of social media networks as a platform for sharing opinions on different aspects emerged the sentiment analysis and opinion mining as an active research area. Recently, the sentiment analysis on Twitter has attracted considerable attention due to its many applications in various aspects of our lives. Many approaches have been presented for sentiment analysis based on English language, thus there is a need for efficient sentiment analysis approaches for Arabic language, since it has different structure when compared to other languages. This paper proposes a hybrid approach for sentiment analysis of Arabic tweets based on two stages. Firstly, the pre-processing methods like stop-word removal, tokenization and stemming are applied, and then two features weighting algorithms (information gain and chi square) are utilized to assign high weights to the most significant features of the Arabic tweets. Secondly, the deep learning technique is employed to effectively and accurately classify the Arabic tweets either as positive or negative tweets. The performance of the proposed approach has been compared with some of the classification methods such as Decision Tree (DT), Neural Networks (NN) and Support Vector Machine (SVM) using the dataset collected from Arabic tweets. The proposed approach outperforms the other approaches and achieved highest accuracy and precision of 90% and 93.7%, respectively.

© 2017 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

The fast growing of social media networks in the recently has made it as integral part of our daily lives. Millions of users utilize the social media networks in expressing their opinions and ideas, exchanging information and sharing news easily. Twitter is a social media platform that has been launched in 2006 and rapidly became a worldwide popular social media network, with more than 313 million monthly active users and more than 40 supported languages. The number of Twitter users in Arab World is increasing rapidly, and a huge number of Arabic tweets are generated daily. The extensive use of social media networks has emerged the opinion mining or sentiment analysis as a potential research field to evaluate the products and people's opinions (Thelwall et al., 2011).

Sentiment analysis is the process of using the natural language processing, computational linguistics and text analysis to find and mine

subjective (positive opinions or negative) or objective (neutral) information (Pang and Lee, 2008). The goal of sentiment analysis is to specify the behavior of the person based on the general contextual polarity of the text (negative, positive, or neutral). It has many useful applications in education, commerce, health, security and numerous others (Nakov et al., 2016; Imran et al., 2015). Since the sentiment is a language dependent, it is significantly important to consider the language and culture characteristics in analysing the sentiment. Several sentiment analysis studies have been carried out based on the English language (Nakov et al., 2016; Poria et al., 2016; Agarwal and Mittal, 2016; Cambria, 2016). The Arabic language is one of the top Semitic languages; approximately 422 million persons use the Arabic language in 27 countries (UNESCO, 2012). It is ranked as the fifth mostly used language among the top 100 languages in worldwide. There are three varieties of Arabic language: Traditional Arabic found in the Holy Quran, Modern Standard Arabic (MSA) practiced in official dealings and dialectical (Farghaly and Shaalan, 2009). Few sentiment analysis researches have been conducted in Arabic.

This paper primarily explores the efficiency of applying the deep learning and features weighting

\* Corresponding Author.

Email Address: [aaataha@kau.edu.sa](mailto:aaataha@kau.edu.sa)<https://doi.org/10.21833/ijaas.2017.08.007>

2313-626X/© 2017 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

techniques for sentiment analysis of Arabic tweets. The challenge is that accurate sentiment analysis needs more understanding than the traditional text classification, this is mainly because the sentiment usually written in a more subtle manner, while text classification depends on keywords. The contributions of this paper are as follows:

- Proposing an effective hybrid approach for sentiment analysis in Arabic tweets based on the deep learning model with features weighting.
- Finding the important features to discriminating between the positive and the negative sentiments in Arabic tweets.
- Investigating the potentials of deep learning techniques in the sentiment analysis of Arabic tweets.

The rest of the paper is organized as follows: Section 2 shows the related work. The proposed hybrid approach for the sentiment analysis of Arabic tweets in explained in section 3. Section 4 explains the experiments and the results discussion. Lastly, section 5 concludes this work.

## 2. Related works

Several researches have been conducted in the area of sentiment analysis. Firstly, the literature review of the sentiment analysis approaches for English will be covered, subsequently the Arabic language approaches.

There are many research efforts in the area of sentiment analysis for the English language, most of these research efforts are concentrating on the contents of social media. Pang et al. (2002) conducted experiments using different machine learning techniques, such as SVM, ME and NB to categorize the sentiments as positive or negative sentiments. They used the n-gram technique such as unigram, bigram for the sentiment classification. Their results show that NB algorithm achieved the lowest performance when compared to the other three algorithms, and SVM algorithm achieved the highest performance. Somasundaran et al. (2007) suggested a method for sentiment analysis of Tweets. The scheme is based on the interpretation of the emotion icons included in the tweets, the tweet was classified as positive sentiment if it includes the icon '( :)', or ':-)'; and as a negative sentiment if it includes the icon ':( '. Pak and Paroubek (2010) suggested a model for classifying the Tweets as positive, negative and objective. They collected tweets dataset including emoticons and developed a sentiment classifier using multinomial Naive Bayes and features like Ngram and POS-tags. Kim (2014) applied convolutional neural networks for English sentiment analysis using different datasets; the experimental results show that the convolutional neural networks achieved good performance in the sentiment analysis Niu et al. (2016) presented a Multi-View Sentiment Analysis (MVSA) dataset, and consists of images and texts collected from Twitter. Their approach of sentiment analysis achieved classification accuracy of 71.9%, and can be considered as lexicon based (set of opinion words

with pre-defined sentiment score) and statistic learning (several machine learning methods).

Aldayel and Azmi (2016) proposed a hybrid approach for Arabic tweets sentiment analysis that combines machine learning techniques and semantic orientation, in their method, the lexical-based classifier used to label the training data and the output is utilized to train the SVM machine learning classifier. Their results showed that the hybrid approach achieved sentiment analysis accuracy of 84.01%. In Rushdi-Saleh et al. (2011) collected An Arabic dataset consist of 500 movie reviews, they used NB and SVM as machine learning classifiers. In their study they considered several operations in the pre-processing stage. Their proposed approach achieved sentiment analysis accuracy of 89%. Shoukry and Rafea (2012) explored the sentences in Arabic text to perform sentiment analysis using 1000 tweets. They used machine learning techniques such as Naive Bayes (NB) and Support Vector Machines (SVM) for sentiment analysis. The sentiment analysis accuracy accomplished by the SVM Classifier is 72.60% while the NB classifier achieved accuracy of 65%. Salamah and Elkhilifi (2014) collected about 340,000 Arabic tweets related to the discussions in the Kuwait National Assembly. Random tree, SVM and decision tree were used for sentiment analysis. The average sentiment analysis precision of the used techniques was 76%. In Abdulla et al. (2013) 2000 Arabic tweets were collected for sentiment analysis. Corpus-based "Supervised Learning" and lexicon-based "Unsupervised Learning" were applied to the dataset. Four supervised machine learning techniques were applied, i.e., K-Nearest Neighbor, NB, SVM and Decision Tree. The SVM and NB achieved sentiment analysis accuracy of 80%. Using the lexicon-based "Unsupervised Learning" approach the achieved sentiment analysis accuracy was 58.6%. Al-Ayyoub et al. (2016) developed a lexicon-based approach for emotion analysis of Arabic text. To evaluate their approach, they collected a dataset of Arabic text from Twitter and Facebook; the experimental results show that the proposed lexicon-based approach achieved an accuracy of 89.7%.

## 3. The proposed hybrid approach

The proposed hybrid approach for sentiment analysis of Arabic tweets comprises two stages as shown in Fig. 1. The first stage involves pre-processing methods like stop-word removal, tokenization and stemming. Two features selection algorithms (information gain and chi square) were applied to assign high weights to the most significant features of the Arabic tweets. In the second stage, the deep learning technique is utilized to effectively and accurately classify the Arabic tweets as positive or negative. The proposed hybrid approach utilizes the combination of the advantages of deep learning and features weighting techniques.

### 3.1. Arabic tweets pre-processing

The pre-processing of Arabic tweets text is a significant phase in text mining; it is aim at

preparing the Arabic tweets for classification. The most difficult aspect of text mining is the huge size and dimensionality of textual data. Tokenization, stemming and filtering are examples for the operations that can be applied to decrease the dimensionality and the scope of the text. In this paper, the following steps for the pre-processing of Arabic tweets text were used.

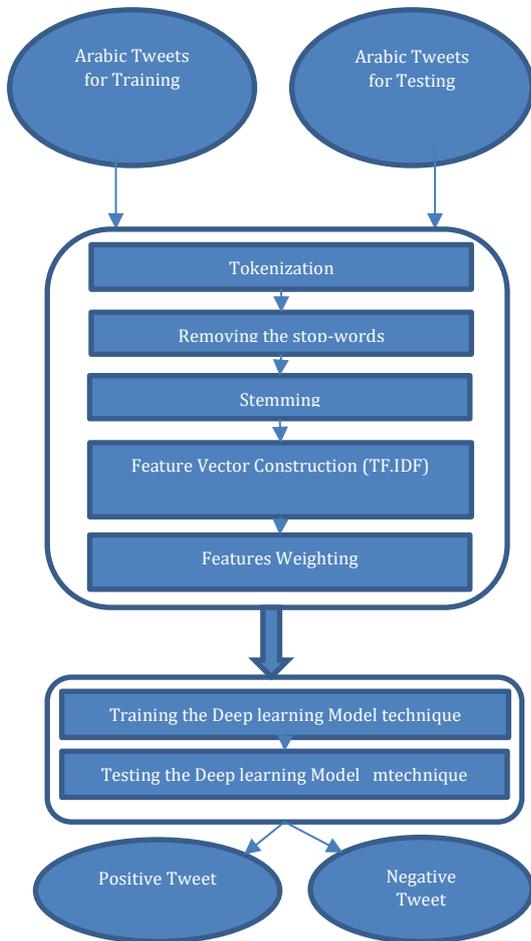


Fig. 1: The proposed approach for sentiment analysis of Arabic tweets

### 3.1.1. Tokenization, cleaning and normalization

Tokenization is the process of dividing a stream of text up into words or other significant components named tokens. Tokenization process aims to determine the potential keywords by exploring the words that compose the sentence (Verma et al., 2014).

Tokenization is important step in natural language processing and potentially affects the sentiment analysis of the texts used in the social media (Laboreiro et al., 2010; Bird et al., 2009). Morphologically, Arabic language is very rich (Duwairi et al., 2014), and hence the Arabic tweets text were cleaned by deleting the non-Arabic words characters and special Twitter characters from the sentence. In addition, this step includes the normalization of the alphabet that come in variant forms to the basic form. Examples of the normalization include the following:

- a. elimination of the diacritical markings, because

Arabic language has several diacritics that can change the meaning of the word based on it is presence or absence. For example, followers (mutābiʿwn) and following (mutābaʿwn), when the two words written in Arabic without diacritics they be read as same word.

- b. removal of the letter “Hamza”,
- c. unifying the letter “Alif”, so “ā” and “ʾā” are replaced by “a”,
- d. substituting “ī” with “h”,
- e. substituting “y” with “ʿy”.

### 3.1.2. Removing the stop-words

Stop-words used frequently and they are useless for text mining such as pronouns, articles and prepositions. Usually there are some words in the text that are not significant for the sentiment mining, and hence removing these words will meaningfully advances the performance of the system. Eliminating the stop-words from the Arabic tweets is potential in determining the most significant words. Examples for the words that have been eliminated include: “ʿly” (on), “mn” (of) and “fy” (in).

### 3.1.3. Stemming

Stemming is the procedure of eliminating the affixes in the words, and getting the roots of these words (Khoja and Garside, 1999). For example, Stemming would reduce the words “alḥasb” (in English the Computer) to the stem “ḥsb” (to Compute), while light stemmer reduces the word to “ḥasb” (Computer in English). Several stemming approaches have been proposed for Arabic language. There are two main categories for these stemmers. The first category is root extraction stemmer, and it reduces the words to their stems such as the stemmer introduced by Khoja and Garside (1999). The second category is light stemmers, and it deletes common affixes from words without reducing them to their roots such as the stemmer introduced in (Larkey et al., 2007).

### 3.1.4. Features selection

To obtain the most significant and meaningful features, the term frequency-inverse document frequency (TF-IDF) (Manning et al., 2008) has been utilized to select the terms that appear more frequently in the tweets.

### 3.1.5. Features weighting

The process of weighting the features is significantly important in determining the most effective feature in the system. In this paper two features weighting algorithms were used.

#### 3.1.5.1. Information gain

The information gain ratio (IGR) method (Mori, 2002) is used to calculate the weight of the terms in the Arabic tweets. The IGR finds the matches between sets of Arabic tweet terms, and then

compute and rank each term. The IGR is defined as in Eqs. 1 and 2:

$$gain_r(X, Y) = \frac{gain(X, Y)}{split\_info(Y)} \quad (1)$$

$$split\_info(Y) = \sum_i \left( \frac{|Y_i|}{|Y|} \right) \log \frac{|Y_i|}{|Y|} \quad (2)$$

where  $gain_r(X, Y)$  indicates the gain ratio of the term X existence in class Y.  $Y_i$  and  $|Y_i|$  indicates the existence of terms X in class  $Y_i$ , the subclass i of Y, and the total terms in  $Y_i$ .

### 3.1.5.2. Chi-square

The Chi-Square test of Independence is potential to decide if there is a significant correlation between two variables. For the sentiment analysis of Arabic tweets, the Chi-Square is used to assess the lack of independence between the term t and the class c, Chi-square is defined as follows (Eq. 3):

$$\chi^2_{(t,c)} = \frac{D \times (PN - MQ)^2}{(P+M) \times (Q+N) \times (P+Q) \times (M+N)} \quad (3)$$

where D represents the total number of A tweets. P is the number of tweets of class c having the term t. Q is the number of Arabic tweets containing t occurring without c. M is the number of tweets class c occurring without t. N is the number of tweets of other classes without t (Aggarwal et al., 2012).

### 3.2. The Deep learning technique

Deep learning is an effective approach for computational modelling (LeCun et al., 2015). The model composed of huge number of hidden layers and neurons to characterize the data with various abstractions. It works more efficiently with big datasets. The examples for deep learning approaches include Neural Networks that use several hidden layers, Recurrent Neural Networks and Convolutional Neural Networks (Schmidhuber, 2015). In this paper, the feed forward architecture used by H2O was employed; it is scalable, fast and open-source deep learning technique. H2O Deep Learning model utilizes a multi-layer feed-forward artificial neural network, which is trained using stochastic gradient descent based on back-propagation. It consists of big number of hidden layers containing neurons; Fig. 2 shows an example for the neuron. In the deep learning model, the weighted combination  $\alpha = \sum_i w_i x_i + b$  of the inputs is aggregated, then the output signal  $f(\alpha)$  is communicated by the connected neuron. The nonlinear activation function (f) is utilized in the network; the activation threshold of the neuron is calculated using the bias b. As shown in Fig. 3, the deep learning network consists of several layers of interconnected neuron elements. The first layer is the input layer to match the space of the feature and it is followed by layers of nonlinearity, and finally the last layer matches the space of the output. Fig. 2 shows example for the single neuron used in all the units of the model.

### 3.3. The used dataset

We collected a dataset consists of 500 Arabic tweets, and the tweets mainly discuss general topics about education. Using the stratified sampling approach the dataset was divided into two subsets: training and testing subsets to assess the efficiency of the proposed hybrid approach for sentiment analysis of Arabic tweets. 80 % of the entire dataset was used for training the deep learning model, while the remaining 20 % of the dataset was used for testing the proposed approach.

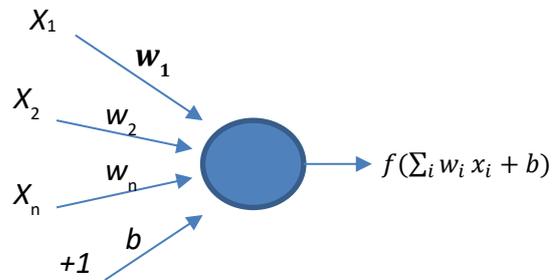


Fig. 2: Single neuron in deep learning model

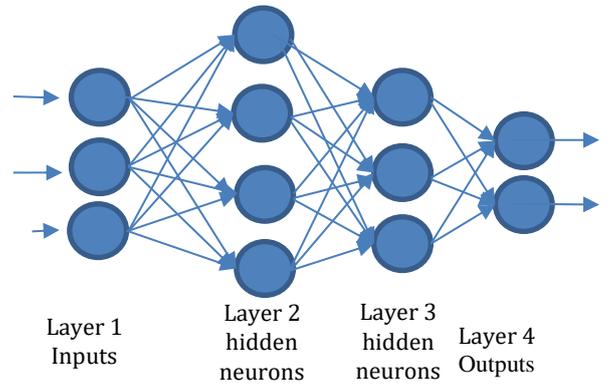


Fig. 3: The architecture of the deep learning network

### 3.4. Performance evaluation metrics

The following standard evaluation metrics have been used to assess the efficiency of the proposed approach:

**Accuracy:** It is a significant to evaluate the classification, and it is computed as the percentage of correctly classified samples to the total samples. Therefore, it can be represented mathematically as follow:

$$Accuracy = (TP + TN) / (TP + FP + TN + FN)$$

**Precision:** It evaluates the strictness of the classifier output. Precision is the percentage of samples classified as positive correctly to total number of samples classified as positive.

$$Precision = TP / (TP + FP)$$

**Recall:** the integrity of the classifier output is measured using the recall.

$$Recall = TP / (TP + FN)$$

Where

TP: the number of Arabic tweets that were correctly classified by the classifier as belonging to the correct class.

TN: the number of Arabic tweets that were correctly

classified by the classifier as not belonging to the correct class.

FP: the number of Arabic tweets that were incorrectly classified by the classifier as belonging to the correct class.

FN: the number of Arabic tweets was incorrectly classified by the classifier as not belonging to the correct class.

#### 4. Results and discussion

We focus on comparing the sentiment analysis accuracy of the proposed hybrid approach based on the deep learning with features weighting and other classifiers such as neural network, support vector machine and decision tree. Before approaching the experiments, all the models require parameter setting. According to the results of the deep learning model, three hidden layers with 100 neurons in each hidden layer provided the highest accuracy. For the neural network classifier, the traditional feed-forward network is used with a single hidden layer containing 10 neurons. The default parameters values of the support vector machine and decision tree were used. The performance of the proposed hybrid approach and other classifiers is evaluated in terms of accuracy, recall and precision.

In order to evaluate the efficiency of the proposed method, the pre-processing steps such as tokenization, stop-words, removing, stemming and the feature selection method (TF-IDF) are applied on the dataset. After that, two features weighting methods (Information Gain and Chai-square) were used in the experiments. The value of the Chi-square and Information Gain coefficient is used for ranking of all the feature in the datasets, the top 20 % features weighted by information gain and chai-square are selected as they gave higher classification performance, and used as inputs to the proposed deep learning model and three other classifiers (SVM, NN and DT) are used for sentiment classification.

##### 4.1. Hybrid approach based on deep learning with information gain

In this section, the results were presented by considering the information gain as features weighting method. Fig. 4 shows the deep learning with features weighted by the information gain algorithm compared with the SVM, Decision tree and neural network classifiers. It is clear from Fig. 4 that the deep learning and SVM classifiers achieved the same level of accuracy (87.5). However, the DT classifier achieved the lowest performance in terms of accuracy (67.5%) and recall (33.3%).

##### 4.2. Hybrid approach based on deep learning with chai-square

In this section, the results were presented by considering the chai-square as features weighting method. Fig. 5 demonstrates the performance of the proposed hybrid approach for Arabic sentiment analysis based on the deep learning model with chi-

square algorithm compared with the SVM, DT and NN classifiers. It is very clear that our proposed hybrid approach based on the deep learning model with chi-square algorithm achieved the highest performance in terms of accuracy (90%) and precision (93.7%) among all the classifiers, conferring the proposed hybrid approach based on features weighted by the chai square with a considerable advantage. Accuracy wise, we note that the proposed hybrid approach based on deep learning with chai-square produced achieved an improvement of 2.5% over the deep learning with information gain. We believe that precision is an important measure as it evaluates the classification correctness of the identified tweets. In terms of precision, the proposed hybrid approach based on deep learning with chai-square achieved a significant improvement of (9.7 %) over the deep learning with information gain. In terms of recall measure, our hybrid approach and SVM achieved the highest recall (83.33) attained an improvement of (5.6%). The DT classifier achieved the worst accuracy (67.5%) and recalls (33.3%).

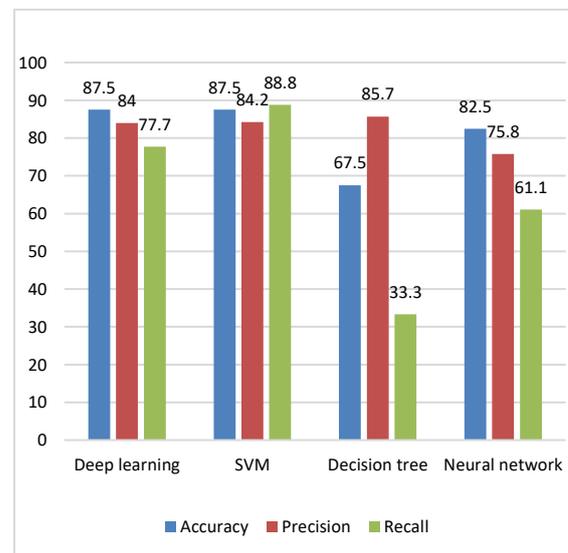


Fig. 4: Performance Comparison of the proposed hybrid approach and other classifiers based on information gain weighting algorithm

Although our proposed hybrid approach based on deep learning with chi-square achieved the high accuracy among all the tested classifiers, it has been also compared with other approaches. Consequently, the proposed hybrid approach for sentiment analysis of Arabic tweets has been compared with eight similar sentiment analysis approaches. The comparison results presented in Table 1. The accuracy of the classification of the approaches varied significantly. The proposed hybrid approach for sentiment analysis of Arabic tweets achieved the better performance with an accuracy of 90%, while the approach proposed in Shoukry and Rafea (2012) achieved the lowest performance and analyse the sentiment with accuracy of 72.60% among all classifiers. The authors in Shoukry and Rafea (2012), Aldayel and Azmi (2016), and Duwairi and Qarqaz

(2014) used machine learning classifiers and achieved low accuracy even when we compare it with our SVM and DT classifiers (Fig. 5). This may be attributed to their simple pre-processing approaches.

We believe that efficiency of our proposed hybrid approach is due to two main reasons: Firstly, we utilized the features weighting as an extra pre-processing step, by applying the chi-square for identifying the most significant features and reducing the input space. Secondly, the deep learning is applied in stage 2 of our proposed hybrid approach as a powerful technique for sentiment analysis.

### 5. Conclusion

Recently, there is a significant concentration in the research on knowledge extraction from social media networks. The huge volume of posted comments, and associated media is a potential source of collecting information. In this paper, we

investigated the problem of sentiment analysis in Arabic tweets.

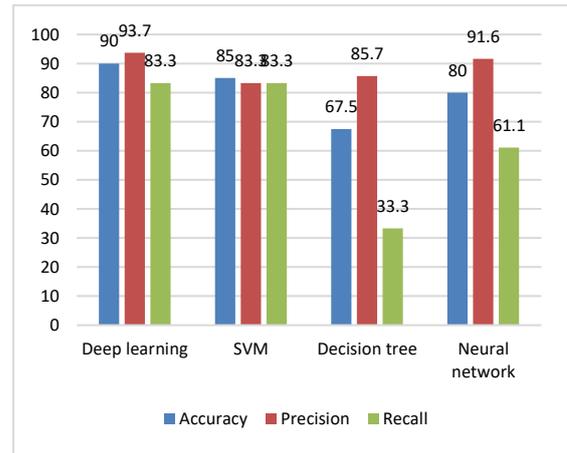


Fig. 5: Performance comparison of the proposed hybrid approach and other classifiers based on chi-square weighting algorithm

Table 1: Performance comparison of the proposed hybrid approach with other existing works. All the stated performance measures are as reported by the authors in the papers

Author	Pre-processing	Approach	Dataset used	Obtained accuracy
Shoukry and Rafea (2012)	Filtering technique	SVM	1000 Arabic tweets (500 positive, 500 negative)	72.60 %
Aldayel and Azmi (2016)	Filtering, normalization, stop-word removal, negation terms and stemming	Hybrid (lexical + SVM)	1103 Arabic tweets (576 positive, 527 negative)	84.01 %
Rushdi-Saleh et al. (2011)	spelling correction, tokenization, stop-words removal, and stemming,	NB and SVM	Arabic dataset consist of 500 movie reviews	89 %
Duwairi and Qarqaz (2014)	Tokenization, Stemming(Arabic), Filtering Stopwords, and Generating-n-Grams(Terms)	NB KNN and SVM	Tweet consists of 2591 (1518 negative, 1073 positive)	75.25 %
Khasawneh et al. (2015)	Filtering technique	ML Bagging technique	Dataset consist of 1500 Arabic tweets	85.95 %
Salamah and Elkhlifi. (2014)	Tokenization and segmentation of tweets	random tree, SVM and decision tree	Dataset consists of 340,000 Arabic tweets	76 %.
Abdulla et al. (2013)	spelling correction, tokenization, stop-words removal	K-Nearest Neighbour , NB ,SVM and D-Tree	Dataset consists of 2000 Arabic tweets (1000 positive tweets and 1000 negative ones)	80 %
Al-Ayyoub et al. (2016)	Text pre-processing include the emotion-bearing text excerpts of the dataset are manually explained by human experts.	lexicon-based approach	Dataset consists of Arabic tweets	89.7 %.
Our proposed hybrid approach	Text pre-processing include text cleaning, normalization, stop words removal and stemming	Features weighting and deep learning model.	Dataset consists of 500 Arabic tweets	90 %

We propose a hybrid approach for sentiment analysis of Arabic tweets based on deep learning with features weighting. The features weighting methods were used in the pre-processing stage to select the most significant features. The deep learning is employed as emerging powerful technique to analyze the sentiment of Arabic tweets based on the selected features.

The experimental results confirmed the efficiency of the proposed hybrid approach based on deep learning with chi-square method, and show that the

proposed hybrid approach outperforms the SVM, DT and NN classifiers achieved the highest performance in terms of accuracy and precision are 90% and 93.7% respectively. As for the future work, the semantic of the text in the Arabic tweets will be considered for more accurate analysis of sentiment.

### References

Abdulla NA, Ahmed NA, Shehab MA, and Al-Ayyoub M (2013). Arabic sentiment analysis: Lexicon-based and corpus-based. In the IEEE Jordan Conference on Applied Electrical

- Engineering and Computing Technologies (AEECT'13), IEEE, Amman, Jordan: 1-6. <https://doi.org/10.1109/AEECT.2013.6716448>
- Agarwal B and Mittal N (2016). Machine learning approach for sentiment analysis. In: Agarwal B and Mittal N (Eds.), *Prominent feature extraction for sentiment analysis*: 21-45. Springer International Publishing, Berlin, Germany.
- Aggarwal CC and Zhai C (2012). *Mining text data*. Springer Science and Business Media, Berlin, Germany
- Al-Ayyoub M, Nuseir A, Kanaan G, and Al-Shalabi R (2016) Hierarchical classifiers for multi-way sentiment analysis of arabic reviews. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 7(2):531-539.
- Aldayel HK and Azmi AM (2016). Arabic tweets sentiment analysis—a hybrid scheme. *Journal of Information Science*, 42(6):782-797.
- Bird S, Klein E, and Loper E (2009) *Natural language processing with Python: Analyzing text with the natural language toolkit*. O'Reilly Media Inc, Sebastopol, USA.
- Cambria E (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2): 102-107.
- Duwairi RM and Qarqaz I (2014) Arabic sentiment analysis using supervised classification. In the International Conference on Future Internet of Things and Cloud (FiCloud'14), IEEE, Barcelona, Spain: 579-583. <https://doi.org/10.1109/FiCloud.2014.100>
- Duwairi RM, Marji R, Sha'ban N, and Rushaidat S (2014). Sentiment analysis in Arabic tweets. In the 5th International Conference on Information and Communication Systems (ICICS'16), IEEE, Irbid, Jordan: 1-6. <https://doi.org/10.1109/IACS.2014.6841964>
- Farghaly A and Shaalan K (2009). Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing (TALIP)*, 8(4). <https://doi.org/10.1145/1644879.1644881>
- Imran M, Castillo C, Diaz F, and Vieweg S (2015). Processing social media messages in mass emergency: A survey. *ACM Computing Surveys (CSUR)*, 47(4). <https://doi.org/10.1145/2771588>
- Khasawneh RT, Wahsheh HA, Alsmadi IM, and Al-Kabi MN (2015). Arabic sentiment polarity identification using a hybrid approach. In the 6<sup>th</sup> International Conference on Information and Communication Systems (ICICS'15), IEEE, Amman, Jordan: 148-153. <https://doi.org/10.1109/IACS.2015.7103218>
- Khoja S and Garside R (1999). *Stemming Arabic text*. Computing Department, Lancaster University, Lancaster, UK.
- Kim Y (2014). Convolutional Neural Networks for Sentence Classification. In the Conference on Empirical Methods in Natural Language Processing (EMNLP'14), Association for Computational Linguistics, Doha, Qatar: 1746-1751. Available online at: <https://arxiv.org/abs/1408.5882>
- Laboreiro G, Sarmiento L, Teixeira J, and Oliveira E (2010). Tokenizing micro-blogging messages using a text classification approach. In the fourth workshop on Analytics for noisy unstructured text data, ACM, Toronto, Canada: 81-88. <https://doi.org/10.1145/1871840.1871853>
- Larkey LS, Ballesteros L, and Connell ME (2007). Light stemming for Arabic information retrieval. In: Soudi A, Neumann G, and Van den Bosch A (Eds.), *Arabic computational morphology*: 221-243. Springer Netherlands, Amsterdam, Netherlands.
- LeCun Y, Bengio Y, and Hinton G (2015). Deep learning. *Nature*, 521(7553):436-44.
- Manning CD, Raghavan P, and Schütze H (2008). *Introduction to information retrieval*. Cambridge University Press, Cambridge, United Kingdom.
- Mori T (2002). Information gain ratio as term weight: the case of summarization of ir results. In the 19<sup>th</sup> international conference on Computational linguistics, Association for Computational Linguistics, Taipei, Taiwan, 1: 1-7. <https://doi.org/10.3115/1072228.1072246>
- Nakov P, Ritter A, Rosenthal S, Sebastiani F, and Stoyanov V (2016). SemEval-2016 task 4: Sentiment analysis in Twitter. In the International Workshop on Semantic Evaluation (SemEval), Association for Computational Linguistics, San Diego, California, USA: 1-8. Available online at: <http://anthology.aclweb.org/S/S16/S16-1001.pdf>
- Niu T, Zhu S, Pang L, and El Saddik A (2016). Sentiment analysis on multi-view social data. In: Tian Q, Sebe N, Qi GJ, Huet B, Hong R, and Liu X (Eds.), *Multi Media Modeling: Lecture notes in computer science*: 15-27. Springer International Publishing, Berlin, Germany.
- Pak A and Paroubek P (2010). Twitter as a corpus for sentiment analysis and opinion mining. In the Louisiana Real Estate Commission (LREC'10), USA, 10: 1320-1326. Available online at: <http://crowdsourcing-class.org/assignments/downloads/pak-paroubek.pdf>
- Pang B and Lee L (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1-2): 1-35.
- Pang B, Lee L, and Vaithyanathan S (2002). Thumbs up?: sentiment classification using machine learning techniques. In the ACL-02 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Stroudsburg, PA, USA, 10: 79-86. <https://doi.org/10.3115/1118693.1118704>
- Poria S, Cambria E, Howard N, Huang GB, and Hussain A (2016). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*, 174: 50-59.
- Rushdi-Saleh M, Martín-Valdivia MT, Ureña-López LA, and Perea-Ortega JM (2011). OCA: Opinion corpus for Arabic. *Journal of the American Society for Information Science and Technology*, 62(10): 2045-2054.
- Salamah JB and Elkhlifi A (2014). Microblogging opinion mining approach for Kuwaiti dialect. In The International Conference on Computing Technology and Information Management (ICCTIM'14), Dubai, UAE: 388-396. Available online at: <https://journals.iauiip.de/index.php/icctim/article/view/225>
- Schmidhuber J (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61: 85-117.
- Shoukry A and Rafea A (2012). Sentence-level Arabic sentiment analysis. In the International Conference on Collaboration Technologies and Systems, IEEE, Denver, USA: 546-550. <https://doi.org/10.1109/CTS.2012.6261103>
- Somasundaran S, Wilson T, Wiebe J, and Stoyanov V (2007). QA with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news. In the International Conference on Web and Social Media (ICWSM'07), Boulder, Colorado, USA. Available online at: <http://icwsm.org/papers/2--Somasundaran-Wilson-Wiebe-Stoyanov.pdf>
- Thelwall M, Buckley K, and Paltoglou G (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2):406-418.
- UNESCO (2012). *World Arabic language day*. United Nations Educational, Scientific and Cultural Organization. Paris, France. Available online at: <http://www.en.unesco.org/>
- Verma T, Renu DG, and Gaur D (2014). Tokenization and filtering process in rapid miner. *International Journal of Applied Information Systems*, 7(2): 16-8.