SEDAT: Sentiment and Emotion Detection in Arabic Text using CNN-LSTM Deep Learning

Malak Abdullah*, Mirsad Hadzikadic† and Samira Shaikh‡
*Department of Computer Science
Jordan University of Science and Technology
Irbid, Jordan
†College of Computing and Informatics
University of North Carolina at Charlotte
North Carolina, U.S
Email: *mabdullah@just.edu.jo, †mirsad@uncc.edu, ‡sshaikh2@uncc.edu

***Abstract*—Social media is growing as a communication medium where people can express online their feelings and opinions on a variety of topics in ways they rarely do in person. Detecting sentiments and emotions in text have gained considerable amount of attention in the last few years. The significant role of the Arab region in international politics and in the global economy have led to the investigation of sentiments and emotions in Arabic. This paper describes our system - SEDAT, to detect sentiments and emotions in Arabic tweets. We use word and document embeddings and a set of semantic features and apply CNN-LSTM and a fully connected neural network architectures to obtain performance results that show substantial improvements in Spearman correlation scores over the baseline models.**

***Keywords*-Sentiment Analysis; Emotion Detection; Machine learning; Deep Learning; LSTM; CNN; Twitter; Arabic;**

I. Introduction

Twitter is considered a rich data bank full of sentiments, emotions, and opinions. Sentiment analysis refers to classifying a subjective text as positive, neutral, or negative; emotion detection recognizes types of feelings through the expression of texts, such as anger, joy, fear, and sadness [1], [2]. In this article, we focus on the task of extracting sentiment and emotion from Arabic language tweets.

Although the Arabic language is considered one of the fastest growing languages on Twitter [3] and the official language for 22 countries, there has been relatively less work accomplished regarding analysis of Arabic language. This is partly due to the complexity and the variety of dialects in Arabic language, which make it hard to build a single system to detect sentiments and emotions for the Arabic language [4]. The language is classified into two main categories: Standard Arabic (SA) and Dialectical Arabic (DA). SA consists of two forms: Classical Arabic (CA) and Modern Standard Arabic (MSA) [5]. The current research paper studies the MSA and DA types of the Arabic language, which are used widely on Twitter. Although MSA is the primary language of Arab countries and is written in books and taught in schools, DA is spoken as a native language in people's informal daily communication and has a strong presence in texting and commenting on microblogging networks or in emails. There is only one MSA language for all Arabic speakers but several dialects with no formal written form [6]. This leads to a lack of lexicon resources for these dialects, and the official grammar rules do not work as efficiently with DA as with MSA.

Deep Neural Networks (DNN) have recently shown significant improvements over traditional Machine Learning (ML) based approaches on classification tasks [7]. In [7], [8], researchers show that Convolution Neural Networks (CNN) and Long-Short Term Memory Networks (LSTM) outperform the traditional machine learning approaches on text classifications, such as sentiment, emotion, and stance detections. Our system, SEDAT (Sentiment and Emotion Detection in Arabic Text), is **the first system** designed to detect and to predict the intensity of sentiments and emotions in Arabic Tweets using Deep Neural Networks (DNN). The data that is used as an input to our system is obtained from the public Twitter datasets of SemEval Task-1, Affect in Tweets [9]. The extracted features are mainly word embedding vectors [10] and semantic features acquired from the AffectiveTweets package [11], [12]. Our system applies these feature vectors to CNN-LSTM [13], [14] and a fully connected neural network architecture to classify sentiment, emotion as well as intensity of emotion in Arabic language tweets. The performance of our system shows substantial improvements in Spearman correlation scores over the baseline models, with 0.01-0.02 points difference between the state-of-the-art model and our proposed model.

The remainder of this paper is organized as follows. Section 2 gives a brief description of existing works in detecting sentiments and emotions. Section 3 provides detailed SEDAT system architecture to determine the presence and the intensity of sentiments and emotions in tweets. Section 4 presents the evaluations and the results. Section 5 describes how the system can handle different dialects and compares the results with different systems. Finally, section 6 concludes with future directions for this research.

II. Related work

Sentiment and Emotion Detection: There is great body of work from psychology that theorizes about emotions [2], [15]. Ekman [2] identified the six basic emotions as anger, disgust, fear, happiness, sadness, and surprise. In this work, we investigate how anger, joy, fear, and sadness are expressed in Arabic text. Very few corpora exist for emotion labeling. Prior research trained a classifier that automatically discovers the emotions in tweets. EmoTex [16] applied supervised learning methods to detect emotions. A group of researchers [11], [17] introduced shared tasks of detecting the intensity of emotion felt by the speaker of a tweet. State-of-the-art systems in these competitions [18], [19] used approaches of ensembling different models and applied feature vectors including word embeddings, semantic, and syntactic features to represent tweets.

Arabic Sentiment and Emotion Detection: To the best of our knowledge, sentiment and emotion detection for Arabic text is relatively new [4], [20]. The main challenges that most researchers face in analyzing sentiment and emotion in Arabic text can be classified under two main areas [21]: a lack of annotated resources and more complex morphology relative to other languages. Researchers in [22] had collected and annotated data and used a simplification of the SVM (known as SMO) and the NaiveBayes classifiers. Another two related works [23], [24] shared different tasks to identify the overall sentiments of the tweets or phrases taken from tweets in both English and Arabic. This current work investigates different neural network architectures with selecting best features to construct SEDAT system. Moreover, we are able to detect emotions in different dialects of Arabic language. To the best of our knowledge, SEDAT system is the first system to detect the intensity of emotions for both MSA and DA with using deep learning approaches.

III. SEDAT System

Our system, SEDAT, has the ability to determine the existence and the intensity of an emotion (Anger, Joy, Fear, or Sadness) in an Arabic tweet as a real-valued score between 0 (least intensity) and 1 (most intensity). It also classifies the intensity of emotion into one of four ordinal classes (0: no emotion, 1: low emotion, 2: moderate emotion, and 3: high emotion). Furthermore, the system is able to to determine the sentiment or valence in a tweet as a real-valued score between 0 (most negative) and 1 (most positive). It can also classify the sentiment intensity of a tweet into one of seven ordinal classes, corresponding to various levels of positive and negative sentiment intensity, starting with 3: very positive and ending with -3: very negative. TableI shows some examples of Arabic tweets with the English translations and the intensity of Anger and Joy emotion in the original Arabic tweets.

SEDAT system consists mainly of two sub-models. Figure 1 shows the structure of our system. More details about the system's components are provided in the following subsections: Section 3.1 describes the system's input and preprocessing step. Section 3.2 lists the feature vectors that are used in both sub-models, and Section 3.3 presents the different architectures of neural networks and deep learning that are used in both sub-models. Section 3.4 discusses the output results.

Arabic Tweets with Translations	**Task**	**Annotation**
وجود الاشخاص بغرفتي وانا مو فيها يزعجني Having people in my room when I am not there annoys me	Anger	0.484 1: low anger
ايه ده طب هنحضر خطوبتك امتى طيب انا عاوز ارقص What is this, so when are we going to attend your engagement, I want to dance.	Joy	0.672 2: moderate joy

Table I
Examples of annotated Arabic tweets with translations.

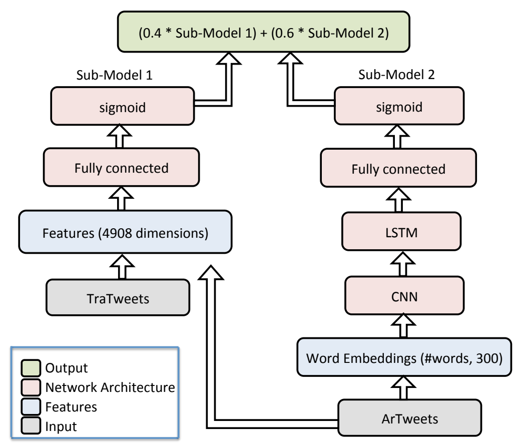

Figure 1. The structure of SEDAT system

A. Input and Preprocessing

The data inputs for our system have been attained from the public datasets of SemEval-2018 (Task 1: Affect in Tweets) [9]. The number of Arabic training and testing datasets for both sentiment and emotion tasks are illustrated in Table II. The Arabic tweets in our system have been used in two forms: as original raw Arabic tweets (ArTweets) or as translated into English (TraTweets) since English language has more preprocessing and feature extraction methods than Arabic. Several preprocessing steps have been applied to both ArTweets and TraTweets.

ArTweets: The original Arabic tweets in training and testing datasets have been tokenized, white spaces have been removed, and the punctuation marks have been treated as individual words (".,?!:;()[]#@'). It is worth mentioning that the preprocessing methods that were not included in our system are normalizing Arabic characters, removing

-	**Train Data**	**Test Data**
Anger	1027	373
Joy	952	448
Sadness	1030	370
Fear	1028	372
Sentiment	1070	730

Table II
NUMBER OF ARABIC TWEETS AS INPUT DATASET

diacritics, removing punctuations, and removing repeating characters. We have tried them but removed them after noticing that there is no improvement in the classification or regression results.

TraTweets: The Arabic tweets have been translated into English using a powerful translation tool written in Python (translate 3.5.0)[1]. The translated tweets then have been tokenized by converting the sentences into words, and all uppercase letters have been converted to lowercase. The preprocessing step also includes stemming the words and removing extraneous white spaces. Punctuation marks have been treated as individual words (".,?!:;()[]#@'), while contractions (wasn't, aren't) were left untreated. We need to highlight that applying preprocessing step for Arabic language before the translation was not appropriate for our system.

B. Feature Vectors

We explore different features to represent both ArTweets and TraTweets and select the best configurations that result in high performance. SEDAT system consists of two sub-models. The first sub-model uses ArTweets with a set of Arabic lexicons to produce ArabicFeature vector with 5 dimensions and TraTweets to produce other vectors with a total of 4903 dimensions (More details about the extracted features for the first sub-model are listed below). The second sub-model uses only ArTweets. Each word in a tweet in the second sub-model is represented as a 300 dimensional vector using the pretrained word embedding model AraVec (Twt-SG) [25] that is trained on Arabic tweets. Then, each tweet is represented as a vector with a fixed number of rows that equals the maximum length of dataset tweets and a standard 300 columns using padding of zero vectors. The following are the set of features for First sub-model:

AffectiveTweets-142: Each tweet in TraTweets is represented as a vector with 142 dimensions by concatenating three vectors obtained from the AffectiveTweets Weka-package [11], [12], 100 dimensional vector is obtained by vectorizing the tweets to embeddings attribute; two-dimensional vector using the Sentiment Strength feature; and finally 40 features have been extracted using the TweetToLexiconFeatureVector attribute that calculates attributes for a tweet using a variety of lexical resources. TweetToLexiconFeatureVector produces 43 dimensional vector, but after applying feature selection (LinearRegression and RandomForestRegressor) to these features we delete the least three significant features, this step shows 0.005 improvement in the results.

Doc2Vec-600: Each tweet in TraTweets is represented as a 600 dimensional vector using the document-level embeddings (doc2vec) [26], [27]. The 600 dimensions are acquired by concatenating two vectors of 300 dimensions each (dm and dbow). Averaging method has been applied to the vectors for each word in the tweet to attain 300 dimensions that best represent the tweet.

ArabicFeatures-5: This vector has been built using different features from [28], [29] and [30]. We start with 10 features then apply feature selection (LinearRegression and RandomForestRegressor) that helps to rank features and choose the configuration of 5 features, this step improves the performance of SEDAT model. The 5 best selected features are: Arabic Emoticon Lexicon, Arabic Hashtag Lexicon, Arabic Hashtag Lexicon (dialectal), Arabic translation of Bing Lius Lexicon, and one features that represents the emoji's in the tweet from [30].

DeepEmoji-64: Each TraTweet is represented as a 64 dimensional vector using deepMoji model [31], which is a model trained on 1.2 billion tweets with emojis to understand how language is used to express emotions. DeepMoji model predicts the sentiment of a tweet and produces different representation for a tweet. We extract the embeddings from the softmax layer with 64 dimensional vector.

UnsupervisedLearning-4096: Each word in TraTweet is represented as a 4096 dimensional vector that is extracted by using Unsupervised Sentiment Neuron [32], which learns an excellent representation of sentiment even though the model is trained only to predict the next character in the text.

EmojiFeature-1: Knowing that DeepMoji [31] model works only with 64 emoticons, we annotate them and assign values to each one of the 64 emoticon. We use DeepMoji model to produce different emoticons related to each tweet. Then by using our annotation, we build a one-dimensional vector to represent the tweet.

C. Network Architecture

Neural networks (NN) have recently become attractive to researchers for language modeling. A standard NN consists of simple connected processors called neurons, each producing a sequence of real-valued activations [33]. Deep learning model is composed of multiple processing layers to learn representations of data with multiple levels of abstraction. Although feed-forward networks can predict the next word of a sequence, the standard Recurrent Neural Networks (RNN) can take into account all of the predecessor words. RNNs are distinguished from feedforward networks by saying that RNNs have memory. A special kind of

[1] https://pypi.python.org/pypi/translate

RNNs are Long Short Term Memory networks (LSTM) [13], while Convolutional Neural Networks (CNN) [14] are feed-forward neural networks. In recent years, both networks have become the state-of-the-art models for a variety of machine learning problems. A common architecture for LSTM is composed of a memory cell, an input gate, an output gate and a forget gate. The cell stores a value (or state), for either long or short time periods. This is achieved by using activation function for the memory cell. CNN makes an efficient use of layers with convolving filters that are applied to local features [14]. The CNN LSTM architecture involves using CNN layers for feature extraction on input data combined with LSTM to support sequence prediction. This architecture was originally referred to as a Long-term Recurrent Convolutional Network or LRCN model [34]. The current research uses feed-forward, LSTM, and CNN to predict the sentiment and emotion in a tweet. The following is a full description of the network architectures for both sub-models in SEDAT system:

Sub-Model 1: The input 4908 dimensional vector feeds into a fully connected neural network with three dense hidden layers of 500, 200, and 80 neurons for each layer, respectively. The activation function for each layer is ReLU [35]. The output layer consists of one sigmoid neuron, which predicts the intensity of the emotion or the sentiment between 0 and 1. Two dropouts are used in this network (0.3, 0.2) after the first and second layers, respectively. For optimization, we use Stochastic Gradient Descent (SGD) optimizer (lr=0.01, decay=1×10^{-6}, and momentum=0.9)[2], augmenting for MSE loss function and ACCURACY metrics. Early stopping is also applied to obtain best results. Best weights for the output predictions are saved to predict the testing datasets. The fit function uses number of epoch=40, batch size=8, validation split=33%.

Sub-Model 2: This sub-model uses a CNN LSTM architectures by adding CNN layer on the front end followed by LSTM layer with Dense layers on the output. The input vector (300, maxLengthOfTweet) feeds a CNN layer with 64 filters, the kernel size is three, and the activation function is ReLU. A maxpooling with pool size=2 is added, then the vectors will be directed to an LSTM of 256 neurons. To avoid over fitting, we use dropout 0.3 after LSTM layer. We add two dense hidden layers with 200 and 80 neurons and ReLU activation function. The output layer consists of one sigmoid neuron, which predicts the intensity of the emotion or the sentiment between 0 and 1. For optimization, we use the same method as we use in Sub-Model 1. Also, best weights for the output predictions are saved to predict the testing datasets. The fit function uses the same epoch, patch size, and validation split parameters of Sub-Model 1.

[2]https://keras.io/optimizers/

Regression Task	sub-model 1	sub-model 2	Prediction
Anger Emotion	0.51	0.55	**0.595**
Joy Emotion	0.62	0.64	**0.747**
Fear Emotion	0.54	0.57	**0.622**
Sadness Emotion	0.459	0.649	**0.680**
Sentiment	0.78	0.74	**0.818**

Table III
THE SPEARMAN CORRELATION SCORES FOR SEDAT SYSTEM AND EACH SUB-MODEL IN THE SYSTEM

Output class	Angry	Joy	Fear	Sadness
0: no emotion	0-0.40	0-0.31	0-0.45	0-0.47
1: low emotion	0.40-0.55	0.31-0.51	0.45-0.56	0.47-0.54
2: moderate emotion	0.55-0.64	0.51-0.75	0.56-0.76	0.54-0.67
3: high emotion	0.64-1	0.75-1	0.76-1	0.67-1

Table IV
CLASSIFY THE OUTPUT TO ORDINAL CLASSES FOR ARABIC EL-OC.

Output class	Sentiment
-3: very negative emotional state	0-0.20
-2: moderately negative emotional state	0.20-0.37
-1: slightly negative emotional state	0.37-0.43
0: neutral or mixed emotional state	0.43-0.56
1: slightly positive emotional state	0.56-0.69
2: moderately positive emotional state	0.69-0.81
3: very positive emotional state	0.81-1

Table V
CLASSIFY THE OUTPUT TO ORDINAL CLASSES FOR ENGLISH AND ARABIC V-OC.

Task	Anger	Joy	Fear	Sadness	result
Emotion Reg	0.595	0.747	0.622	0.68	**0.661**
Emotion Clf	0.504	0.537	0.526	0.611	**0.569**
Sentiment Reg	-	-	-	-	**0.817**
Sentiment Clf	-	-	-	-	**0.786**

Table VI
THE SPEARMAN CORRELATION SCORES (REG=REGRESSION CLF=CLASSIFICATION

D. Output and Results

Each sub-model in SEDAT system produces a real-valued number between 0 and 1. It has been shown that the prediction of second sub-model gives higher Pearson correlations than first sub-model for emotion task, and vice versa for sentiment Task. Also, using averaging method for both predictions provides better results than each one separately. After trying different weights for both predictions, we find that taking 40% of Sub-model 1 and 60% from Sub-model 2 gives better results for all emotions, and vice versa for Sentiment. Table III produces comprehensive details on how each prediction of each sub-model is produced. Also, it shows the final prediction results with 40% of sub-model1 and 60% of sub-model2.

We classify the final results of the real-valued number to one of the ordinal classes. We determine the ranges of values for each ordinal class by studying the annotated datasets. Tables IV and V show the ranges of values to obtain the ordinal classes for emotions and sentiments, respectively. The final results for the regression and classification tasks for both emotion and sentiments are shown in Table VI.

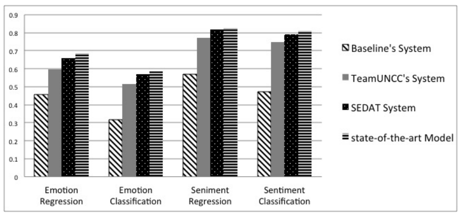

Figure 2. Comparing the Spearman correlation scores of SEDAT system with TeamUNCC and the baseline systems

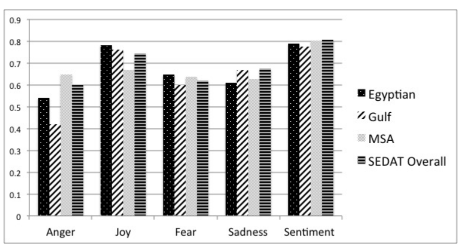

Figure 3. Analyzing the Spearman correlation scores of SEDAT system for each dialect

IV. Analysis and Evaluation

Table VI shows the performance of our system in regression and classification tasks for both sentiment and emotion. The performance of our system surpasses the SVM Unigrams Baseline model's performance, which is provided by the SemEval-Task 1's organizers. See Figure 2 to capture the difference between the two performances. SEDAT system also shows substantial improvements over TeamUNCC's system [36], which is the previous version of SEDAT system. SEDAT is only 0.01 to 0.02 points behind the first-ranked model in the challenge. It is worth mentioning that our results have been obtained using the task datasets without using any external data.

To gain insight on how SEDAT system performs in each language category, we manually split the testing dataset to MSA and two main dialects, Egyptian and Gulf dialects. We notice that the system performs best with MSA for Anger and Fear emotions, whereas Egyptian's dialect preforming the best with Joy emotion, and Gulf dialect with Sadness emotion (see Figure 3).

V. Conclusion

In this paper, we have presented our system SEDAT that uses deep learning architectures for detecting the intensity of emotions and sentiments in Arabic tweets. The performance of the system surpasses the performance of the baseline's model, indicating that our approach is promising. In this system, we uses word and document embedding models with feature vectors extracted from Arabic and translated tweets by using the AffectiveTweets package, Deepmoji, and Unsupervised Sentiment Neurons. These vectors feed different deep neural network architectures, feed-forward, CNN, and LSTM, to obtain the predictions. We uses the SemEval-2018 Task 1's datasets as input for our system, and shows that SEDAT is only 0.01-0.02 points behind the first-ranked model in the task's challenge. The system also proved a high proficiency in detecting sentiments and emotions in different dialectics and modern standard Arabic language.

There is no emotional lexicons that covers a wide variety of dialectical Arabic. Thus, much research can be carried out in the future that involve creating and annotating emotional lexicons for different dialects of Arabic language. Also, there is a need to annotate emotional tweets that can help training the model and improving the performance of prediction.

References

[1] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in *Proceedings of the workshop on languages in social media*. Association for Computational Linguistics, 2011, pp. 30–38.

[2] P. Ekman, "Facial expression and emotion." *American psychologist*, vol. 48, no. 4, p. 384, 1993.

[3] F. Harrag, "Estimating the sentiment of arabic social media contents: A survey," in *5th International Conference on Arabic Language Processing*, 2014.

[4] M. Abdullah and M. Hadzikadic, "Sentiment analysis on arabic tweets: Challenges to dissecting the language," in *International Conference on Social Computing and Social Media*. Springer, 2017, pp. 191–202.

[5] K. Shaalan, "Rule-based approach in arabic natural language processing," *The International Journal on Information and Communication Technologies (IJICT)*, vol. 3, no. 3, pp. 11–19, 2010.

[6] Z. Eviatar and R. Ibrahim, "Why is it hard to read arabic?" in *Handbook of Arabic literacy*. Springer, 2014, pp. 77–96.

[7] C. Guggilla, T. Miller, and I. Gurevych, "Cnn-and lstm-based claim classification in online user comments," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 2740–2751.

[8] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[9] S. M. Mohammad and S. Kiritchenko, "Understanding emotions: A dataset of tweets to study interactions between affect categories," in *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference*, Miyazaki, Japan, 2018.

[10] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[11] S. M. Mohammad and F. Bravo-Marquez, "Wassa-2017 shared task on emotion intensity," *arXiv preprint arXiv:1708.03700*, 2017.

[12] F. Bravo-Marquez, M. Mendoza, and B. Poblete, "Meta-level sentiment models for big social data analysis," *Knowledge-Based Systems*, vol. 69, pp. 86–99, 2014.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[15] R. Plutchik, "Emotions and psychotherapy: A psychoevolutionary perspective," in *Emotion, psychopathology, and psychotherapy*. Elsevier, 1990, pp. 3–41.

[16] M. Hasan, E. Rundensteiner, and E. Agu, "Emotex: Detecting emotions in twitter messages," 2014.

[17] S. M. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "Semeval-2018 Task 1: Affect in tweets," in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA, 2018.

[18] P. Goel, D. Kulshreshtha, P. Jain, and K. K. Shukla, "Prayas at emoint 2017: An ensemble of deep neural architectures for emotion intensity prediction in tweets," in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2017, pp. 58–65.

[19] V. Duppada, R. Jain, and S. Hiray, "Seernet at semeval-2018 task 1: Domain adaptation for affect in tweets," *arXiv preprint arXiv:1804.06137*, 2018.

[20] A. Assiri, A. Emam, and H. Al-Dossari, "Saudi twitter corpus for sentiment analysis," *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 10, no. 2, pp. 272–275, 2016.

[21] A. Assiri, A. Emam, and H. Aldossari, "Arabic sentiment analysis: a survey," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 12, pp. 75–85, 2015.

[22] O. Rabie and C. Sturm, "Feel the heat: Emotion detection in arabic social media content," in *The International Conference on Data Mining, Internet Computing, and Big Data (BigData2014)*. The Society of Digital Information and Wireless Communication, 2014, pp. 37–49.

[23] S. Kiritchenko, S. Mohammad, and M. Salameh, "Semeval-2016 task 7: Determining sentiment intensity of english and arabic phrases," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 42–51.

[24] S. Rosenthal, N. Farra, and P. Nakov, "Semeval-2017 task 4: Sentiment analysis in twitter," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 2017, pp. 502–518.

[25] A. B. Soliman, K. Eissa, and S. R. El-Beltagy, "Aravec: A set of arabic word embedding models for use in arabic nlp," *Procedia Computer Science*, vol. 117, pp. 256–265, 2017.

[26] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International Conference on Machine Learning*, 2014, pp. 1188–1196.

[27] J. H. Lau and T. Baldwin, "An empirical evaluation of doc2vec with practical insights into document embedding generation," *arXiv preprint arXiv:1607.05368*, 2016.

[28] M. S. Saif M. Mohammad and S. Kiritchenko, "Sentiment lexicons for arabic social media," in *Proceedings of 10th edition of the the Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia, 2016.

[29] S. Kiritchenko, S. M. Mohammad, and M. Salameh, "Semeval-2016 task 7: Determining sentiment intensity of english and arabic phrases," in *Proceedings of the International Workshop on Semantic Evaluation*, ser. SemEval '16, San Diego, California, June 2016.

[30] P. Kralj Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis," *PLoS ONE*, vol. 10, no. 12, p. e0144296, 2015. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0144296

[31] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, and S. Lehmann, "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm," *arXiv preprint arXiv:1708.00524*, 2017.

[32] A. Radford, R. Jozefowicz, and I. Sutskever, "Learning to generate reviews and discovering sentiment," *arXiv preprint arXiv:1704.01444*, 2017.

[33] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[34] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625–2634.

[35] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, no. 1, 2013, p. 3.

[36] M. Abdullah and S. Shaikh, "Teamuncc at semeval-2018 task 1: Emotion detection in english and arabic tweets using deep learning," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, 2018, pp. 350–357.