Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 163 (2019) 257–265

Procedia Computer Science

www.elsevier.com/locate/procedia

16th International Learning & Technology Conference 2019

Detecting Arabic Depressed Users from Twitter Data

Salma Almouzini[a], Maher khemakhem[a,b] , Asem Alageel[c,*]

[a,b]*Department of Computer Science, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia*
[c]*Department of Psychiatry, Al Imam Mohammad Ibn Saud Islamic University, Riyadh, Kingdom of Saudi Arabia*

Abstract

Depression is one of the most common health issues impacting the world. People with severe depression symptoms are affected in their work, home, and social lives. Early diagnosis of mental illness is difficult, especially within the Arabic culture, because of the stigma of mental illness and lack of awareness in the field of psychiatry. Meanwhile, social media and its posts provide a new fertile source for mental health surveillance by people express their feelings, moods, and daily activity. Recently, the research field in detecting mental illness through social media has begun to be an exciting topic with the increase in popularity of social media platforms and the current studies in this area just covering English data. To our knowledge, this is the first study that has used Arabic data to explore depressive emotions in an online population. Our experiment, which is based on data collected from Twitter in the Gulf region, detects users who self-declared in their tweets as having been diagnosed with depression. Another set of tweets from non-depressed users was used as a standard group to construct a corpus with truth labels (depressed and non-depressed). We then built a predictive model based on supervised learning algorithms (Random Forest, Naïve Bayes, AdaBoostM1, and Liblinear) to predict whether a user's tweet was depressed or not. Our predictive model leveraged from an efficient features set which was extracted to cover not only the symptoms of clinical depression but also online depression-related behaviour on Twitter (e.g., interaction with trending hashtags and frequent emojis). We observed that optimal accuracy performance was with the Liblinear classifier at 87.5%.

Keywords: Depression; Arabic Sentiment Analysis; Mental Health; Twitter; Supervised Learning Approach

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: SalmaHoumod@gmail.com

10.1016/j.procs.2019.12.107

1. Introduction

Depression is considered one of the most devastating diseases in the world. According to the World Health Organization (WHO), depression affects 350 million people [1]. A primary problem with mental illness is that it is undiagnosed and untreated; this is especially true in Arab countries. Zolezzi et al. [2] point out the stigma associated with mental illness in Arab countries, which falls into one of the following three categories: beliefs towards a person with a mental illness (e.g., attributing the illness to religious and cultural beliefs), attitudes towards treatment for mental illness (e.g., fear of side effects and addiction to psychiatric drugs), and actions towards a person with a mental illness (e.g. such as social distancing in relationships or marriage). Previous studies provide strong evidence that it is possible to detect depression by observing aspects of a user's behaviour on social media. Social media posts provide more accurate and timely surveillance than traditional means, such as self-report surveys, as depression screening tools. For example, Choudhury et al. [3] identify social media as a tool to measure depression. Previous studies used English data to identify depression. To our knowledge, no study has used Arabic data to detect depression among Arabic users in an online population. Sentiment analysis is one of the most important research trends in healthcare that can obtain disease characteristics through textual analysis of social media. The overall goal of this study is to provide an example of a real population using Arabic social media to gain knowledge about a person's psychology. This can improve early identification and intervention so that treatment can begin early; potential outcomes can be improved, and the stigma associated with mental illness can be minimized.

The following are the main contribution of this study:

- We crawled a dataset of Arabic tweets by detecting users who self-declared, in their tweets, as having been diagnosed with depression and those who did not. Then, we created an annotated corpus with two truth labels (depressed, non-depressed) by identifying depression signs based on the Patient Health Questionnaire (PHQ-9) [4] and the Center for Epidemiologic Studies Depression Scale (CES-D), a screening test [5].
- We explored the depression-related behaviors of Arabic users that add value to a set of features that classify users' tweets to distinguish between the depressed class and the non-depressed class.

This paper is constructed as follows: section 2 presents an overview of depression, the main concepts related to sentiment analysis, and the related work. Sections 3 and 4 present the proposed methodology and experimental results, respectively. Section 5 consists of the conclusion and future work.

2. Background and Related Work

2.1. Depression Overview

Major Depressive Disorder (MDD) is a disorder that is diagnosed in patients who display at least five of the following symptoms of depression (where one of the symptoms must be a depressed mood or loss of interest) for at least two weeks [6]: depressed mood, loss of interest, change in appetite or weight, feeling guilty or worthless, sleep disorder, psychomotor agitation or impairment, fatigue or loss of energy, poor concentration and persistent thoughts of death or suicidal ideation [4].

The most famous of the psychometric self-report surveys for mental illnesses cited in the studies and used in this study are: the Patient Health Questionnaire (PHQ-9) and the Center for Epidemiologic Studies Depression Scale (CES-D). Both are self-report questionnaires that were developed to measure depressive symptoms in the general population. They are considered a screening instrument in primary care clinics and in research studies. The PHQ-9 is a 9-item questionnaire while the CES-D is a 20-item questionnaire. Both have a minimum score of zero. While the CES-D has a maximum score of 60, the PHQ-9 has a maximum score of 27. Generally, depression symptoms in these questionnaires are divided into levels based on the likelihood of having depression: mild, moderate, moderately severe, and severe [4][5].

2.2. Sentiment Analysis

Sentiment analysis (SA), also known as opinion mining, is a process that applies natural language processing (NLP) text analysis and machine learning to identify the sentiment of the text as positive, negative or neutral [7]. By analyzing people's opinions, SA uncovers people's sentiments towards different topics, products, services, events, issues, etc. [8]. An opinion is composed of five parts: an object (i.e., the target of the opinion), feature (i.e., the object's attribute), sentiment/opinion orientation (i.e., positive, negative, neutral), opinion holder (i.e., person/organization that expresses the opinion), and time (i.e., time opinion was expressed) [7].

There are two basic approaches currently being adopted in sentiment analysis research: the supervised approach and the unsupervised approach. The supervised approach is typically machine learning (ML) and the unsupervised approach is lexicon-based. The supervised approach utilizes corpus data with labels (positive /negative) to train a classifier or a set of classifiers. The classifiers that are used are machine learning algorithms such as the support vector machine (SVM), naïve Bayes (NB), artificial neural network (ANN), and k-nearest neighbor (KNN) [7][9]. On another hand, the unsupervised approach or the lexicon-based approach defines the orientation (polarity) of each word entry as a numerical value. Some of the dictionaries differentiate between positive and negative sentimental words using positive and negative values, with a value of zero to refer to neutral words. As opposed to the supervised approach, the unsupervised approach does not require labels [7][9].

2.3. Complexity in Arabic Sentiment Analysis

Arabic sentimental analysis followed the same general approach as sentimental analysis in English; however, there are differences in the underlying process based on the difference in language characteristics that creates extra processing challenges and required different processing mechanisms.
The Arabic language is the primary language in 27 countries and the secondary language in many others. Moreover, there are more than 422 million people who speak Arabic around the world [7]. The complexity of the Arabic language can be summarized as follows: Arabic orthography and Arabic morphology [7] [10]. According to Boudad et al. [7], orthographical complexities arise in most of the Modern Standard Arabic (MSA) texts are written without diacritical marks, which causes a lexical ambiguity problem. In contrast, morphological complexities arise on Arabic words that have multiple morphological aspects: derivation and inflection. Derivational morphology, words are created by combining derivational patterns with root words. The root is the word's original form that has no morphological pattern and from which different meanings can be formed by attaching different patterns. Inflectional morphology indicates variants of a word in different grammatical groups. An Arabic word can be a noun, verb (past, present, or future), or particle (all words that are not nouns or verbs, such as prepositions, adverbs, conjunctions).

2.4. Related Work

This section presents a review of prior studies related to depression detection on social media. The main characteristics of the studies are summarized in Table 1. The advantages and limitations of existing methods are also discussed.

Two strategies were reported in the literature to collect data and detect depression through social media: crowdsourcing and data collection from social media publically available. The crowdsourcing strategy is conducted in two stages [3] [15] In the first stage, responses from an online clinical depression survey, such as the CES-D survey, are gathered. In the second stage, contents are collected by accessing the Twitter data of the consented participants. Accordingly, predictive models with high-reliability data were built using the aforementioned strategy. The limitations of this strategy are the time-consumption and insecure responses to the survey. The second strategy [16], [12] is an alternative strategy for gathering data quickly and cheaply. As such, the data is collected directly from the social media that are publically available for participants with self-identified as a mental illness. The disadvantage of this strategy is its low-reliability.

Overall, various approaches were developed by extracting different set of features and using different predictive models. The features that have been extracted focused on the post's characteristics (emotion, linguistic, style and n-grams), while others focused on the author characteristics (post-time, retweets, reply). Several methods have been developed for the automatic detection of depression in social media content. For example, Linguistic Inquiry and Word Count (LIWC) sentiment tool is used to identify and distinguish between different emotions (i.e., non-depressed emotion, depressed emotion) [3], [11-12] [15-16].

Nadeem et al. [13] extracted n-grams and word frequency under the depressed class for an input text and represent each input as a bag of words. Aldarwish et al. [14] used the RapidMiner sentiment tool to generate word vectors of unigrams, yet a low performance was obtained as the created model was built based on low numbers input samples from the participants, which are potentially effecting the model accuracy.

Table 1. Summary of the previous studies.

Ref./Year	Source*	N(Users)/Period*	N(Posts)*	Basis of data collection*	Mental illness criteria	Features	Approach	Performance Measures
[11] (2012)	Twitter	69-participants /two-month	21,103 tweets	keyword: depression	Survey (CES-D)	LIWC Sentiment Tool	Multiple Regression based on LIWC	NA
[3] (2013)	Twitter	489-participants/NA	69,514 tweets	Twitter user's tweets	Survey (CES-D)	User Engagement Ego-Network n-grams Linguistic Style Emotion and Time	SVM Classifier	Accuracy=73% Precision =0.826
[12](2014)	Twitter	21,866 - participants/NA	NA	Twitter user's tweets	Self-declared	LIWC 1-gram (ULM) 5-gram (CLM)	Log-linear classifier	Precision Bipolar=0.48 Depression=0.64 PTSD=0.67 SAD=0.42
[13](2016)	Twitter	NA	2.5 million tweets	Twitter user's tweets	Self-declared	n-gram	Hunt's Algorithm SVM Classifier logistic Regression Naïve Bayes	Accuracy =81%
[14](2017)	Facebook Twitter LiveJournal	NA	6,773 posts	NA	Survey (BDI-II)	word vectors (unigrams)	SVM Classifier Naïve Bayes	Accuracy=63.3%

* Columns that are part of the Dataset, ULM (Unigram Language Model), CLM (Character Language Model), BDI (Beck Depression Inventory)

One of the most challenges facing depression detection is to identify the symptoms of mental illness in online health communities as there is symptom overlapping between multiple mental illnesses. As far as we know, no previous research has investigated the Arabic Twitter data for detecting whether a user's tweet will be depressed or not. To fill this literature gap, we detect signs relevant to depression using Arabic language tweets. To avoid the limitations related to data collection, we combine low cost and reliability data collection strategies. We are collecting data from public tweets and determine if such tweets indicated a genuine claim of MDD diagnosis.

3. The proposed methodology

The methodology was performed as follows: First the data was collected and manually labelled. Then, pre-processed data was cleaned, and sparse features were extracted from tweets as feature vectors. Finally, a supervised experiment was conducted using four popular classifiers: Random Forest, Naïve Bayes, AdaBoostM1, and Liblinear. To implement this experiment, we used *Weka*[1], a machine learning tool and *AffectiveTweets* [17] is a package to Weka involved multiple filters for converting tweets into feature vectors that can be fed into the machine learning classifiers executed in Weka. It is limited to the English language; therefore, some editing is required to this package's code in order to support the Arabic language.

3.1. Data Collection

We extracted more than seven thousand tweets from 97 users in the Gulf region, posted between 18th November 2016 and 23rd July 2018. Tweets in this study were a mixture of Modern Standard Arabic (MSA) and Arabian Gulf dialects. Three steps were taken to extract the tweets. First, we designed a Twitter crawler tool based on *Twitter4j*[2], a Java library for accessing the Twitter API. Second, to capture depressed users' tweets, we searched for users who self-declared in their tweets that they had depression, and their bio included explicit and implicit expressions of depression. We found 35 Twitter users with depression. Then, we searched for non-depressed users as a standard group based on their bio that included their real name, jobs, and interests, activities, and hobbies. We found 62 non-depressed users on Twitter. Third, we employed our crawler tool to retrieve the most recent tweets for each depressed user, up to 200 tweets in order to ensure tweets' contents were enough to evaluate whether the user was depressed or not. Next, we again crawled the most recent tweets for each non-depressed user to retrieve up to 50 tweets. Our crawler tool was designed to retrieve only Arabic tweets. URLs and retweets were also removed from the Twitter users' timeline.

3.2. Manually Labeling Process

The procedures for putting truth labels on the crawled dataset include two-phase. In the first phase, we screened depressed users; eight users with their tweets were eliminated as they were not diagnosed as depressed. This filtering left us with 6122 tweets from 89 users: 27 users classified as depressed, and 62 as non-depressed. In the second phase, we suggested that the manual annotation for the rest data, made up of 6122 tweets, must be based on the depression scale tools. We used two scales, CES-D and PHQ-9, to further confirm depression and eliminate noise.

We limited the diagnostic criteria for depression referred to in the above two scales exclusively into 12 specific categories. In addition, one new category was added called a health category. This indicates tweets from individuals containing very private information about themselves, such as receiving a diagnosis, taking medications, and meeting doctors. Figure 1 shows counts of total tweets for each diagnostic criterion for depression where 1359 tweets were classified as depressed. The depressed mood holds the highest number of tweets at 471 tweets. The second class is called non-depressed, which had 1363 tweets. This class included various words associated with positive emotions such as happiness, passion, ambition, family, friends, work, etc. 3400 tweets under a category called 'other topics' were eliminated in the labelling process as they are irrelevant to the symptoms of depression. Table 2 summaries the number of users and tweets for each class.

---

[1] https://www.cs.waikato.ac.nz/~ml/weka/
[2] http://twitter4j.org/en/index.html

Table 2. Number of tweets/users under each class.

Class	N(User)	N(Post)
Depressed	27	1359
Non-Depressed	62	1363
Total	89	2722

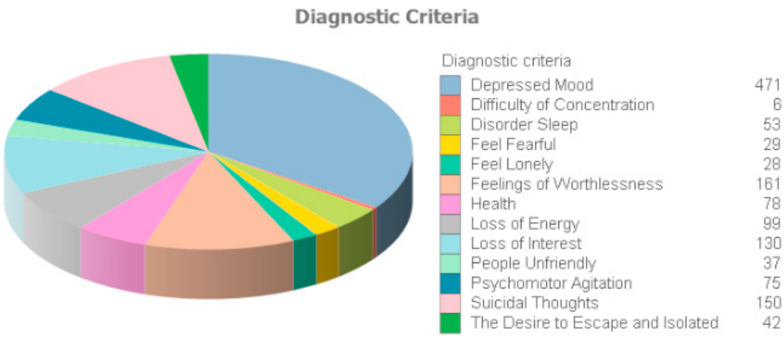

Fig. 1. number of tweets based on diagnostic criteria.

3.3. Cleaning and Pre-Processing Data

From the *AffectiveTweets* package, we selected the *TweetToSparseFeatureVector* filter that has many options for applying a cleaning process and pre-processing data. This package uses *TweetNLP*[3] library as a tokenizer. The cleaning data process involved deleting the following: English or Arabic numbers, English characters, punctuation marks, and @username. Then, this filter implemented pre-processing data that including four steps, as follows:

- Tokenization is used to split the text of a tweet into a sequence of words, phrases or symbols by removing punctuation marks. *TweetNLP*, after modifying the code, can extract the frequency of the Arabic unigrams based on Arabic punctuation.
- Stemming is the process of returning a word to its stem or root. we used the Arabic light stemmer [18] which removed the suffixes and prefixes from the word only without converting the word to the root.
- Stop words removal is the process of removing words that are not related to sentiment analysis words, such as “في” (in), “على” (on), “انت” (you), “من” (of), etc.
- Elimination of speech effect is the process of reducing the number of letters that are repeated more than two times in tokens, e.g., “مزااااجي” replaced “مزااجي”.

3.4. Features Extraction and Selection

We extracted depression symptoms and the online Arabic user’s behaviour from tweets’ texts as efficient features set to distinguish the depressed class from the non-depressed class. These features combined as bag-of-unigrams and negation handling. Where, the bag-of-unigrams considered one of either words, emojis, or hashtags. The words can be either positive emotional expressions, e.g., “سعادة” or negative emotional expressions of depressive symptoms, e.g.,

[3] http://www.cs.cmu.edu/~ark/TweetNLP/

“الموت” or “حزن” but frequent emojis or trending hashtags in the Gulf community were extracted from tweets, also known as Twitter-specific features. The negation words play an essential role in changing the sentiment orientation from affirmative to negative or vice versa (e.g., not happy, not sad, etc.). The *TweetToSparseFeatureVector* filter is adding the NEGTOKEN- tag to words occurring in negated contexts. This filter identifies the negation scope (start from the word following the negator and continues to the end of the sentence or next punctuation mark). We modified the filter’s code to become compatible with Arabic; then we created a list of 36 Arabic negation words, such as لست, لا, and لن, as negators. Finally, we were able to efficiently handle Arabic negation words in the tweets’ content.

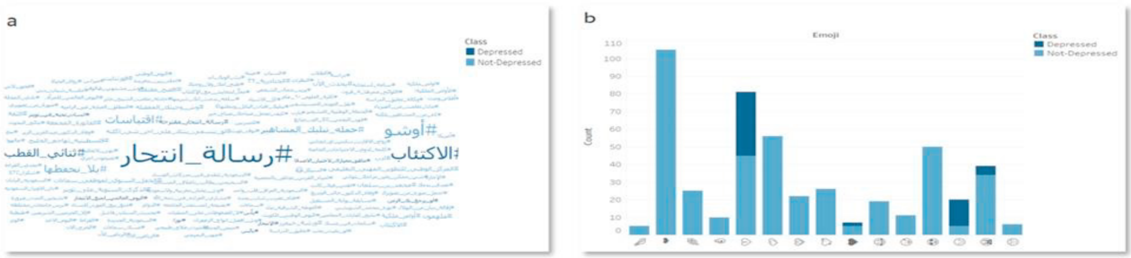

Fig. 2. User engagement with (a) Trending hashtags (b) Most popular emojis under each class.

Figure 2 illustrates the topmost frequent emojis used and interaction with trending hashtags per two classes (depressed/non-depressed). It is evident that the depressed class had less interaction with trending hashtags, which reflects the issues of the Gulf community, and tended to use emoji related to negative emotion more often than positive. This chart indicates that they have social isolation.

3.5. Classifiers Used and Performance Measures

In this study, the following four machine learning algorithms were used to determine the binary classification of the experimental analysis purpose:

- Random Forest: is an ensemble of decision trees that are trained using a bagging approach, which is one of the ensemble methods for multiple classifications that are based on dividing the dataset into multiple overlapped subsets. Accordingly, Random Forest is made by training multiple random decision trees with overlapped subsets of the original set. In the testing phase, the input test sample is classified using all the trained trees, and the final output is generated based on the majority voting of the output from all the trees [19].
- Naïve Bayes: is a probabilistic classifier that uses Bayes' theorem. It assumes that all features are statistically independent of each other and is suitable for high dimensional data [20].
- AdaBoostM1: is an ensemble classification that is based on a weighted sum of multiple classifiers. The idea of Adaboost is to overcome the limitation of each classifier using another classifier that is stronger in the underlying aspect [21].
- Liblinear: is an open-source package to support the linear SVM. It can perform efficiently in large-scale linear classification with better performance in training time and precision [22].

4. Supervised Experiment

The experiment is completely based on the supervised approach with class labels (depressed and non-depressed). We built a predictive model that uses extracted features from tweets, such as bag-of-unigrams and negation handling, based on four classifiers: Random Forest, Naïve Bayes, AdaBoostM1, and Liblinear. Each classifier takes labelled tweets and trains the model to generate predictions for new tweets.

We used 10-fold cross-validation to train the model. The performance metrics that were widely used in the literature [8] [23] (i.e., accuracy, precision, recall, and F-measure) were selected to evaluate the four classifiers. The results of the experiment are shown in Table 3.

Table 3. The performance evaluation of 4 classifiers

Algorithm	Accuracy (%)	Precision (%)			Recall (%)			F-Measure (%)		
		Dep	Non-Dep	Avg	Dep	Non-Dep	Avg	Dep	Non-Dep	Avg
AdaBoostM1	**55.2**	59.1	53.8	**56.4**	34.1	76.4	**55.3**	43.2	63.1	**53.2**
Random Forest	83	76	95.3	85.7	96.5	69.6	83.1	85.1	80.5	82.8
Naïve Bayes	75.6	73.7	77.9	75.8	79.5	71.8	75.6	76.5	74.7	75.6
Liblinear	**87.5**	86.4	88.7	**87.6**	89	86.1	**87.5**	87.7	87.4	**87.5**

4.1. Results

Table 3 shows a comparison of the performance measures of the four classifiers, Random Forest, AdaBoostM1, Liblinear, and Naïve Bayes. The Liblinear has better accuracy than other classifiers. The accuracy of Liblinear achieves 87.5, and recall is 87.5. On the other hand, the poorest accuracy is 55.2, and recall is 55.3 which were achieved by the AdaBoostM1 classifier. Figure 3 presents the accuracy and recall rate of the four classifiers in the 10-fold cross-validation.

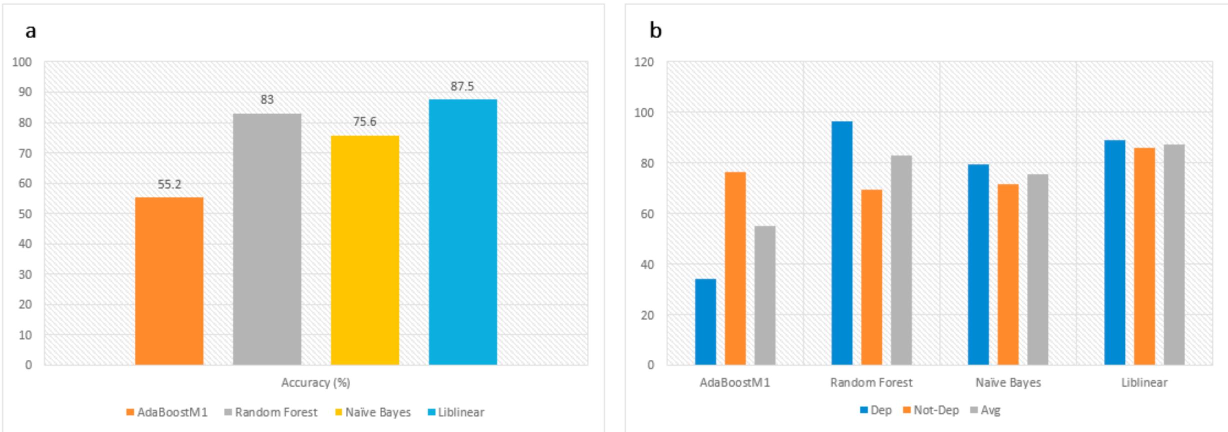

Fig. 3. (a) accuracy rate of 4 classifiers.; (b) recall rate of 4 classifiers.

5. Conclusion and Future work

In conclusion, we provide evidence as to whether Arabic users share their depressive feelings or admit their depression on widely used platforms such as Twitter. We built a predictive model to predict whether a user's tweet is depressed or not based on detecting depressed users using a supervised learning approach to Arabic sentiment analysis. We examined the performance of the four classifiers using a dataset collected from Twitter in the Gulf region based on manually constructed corpus with truth labels (depressed, non-depressed). We have observed that people suffering from depression are more socially isolated as evidenced by determining how they interacted with trending hashtags, popular emojis used in their tweets. We found the optimal accuracy with the Liblinear classifier at 87.5%. In future work, the performance of the model might be improved by including additional features of online Arabic user's behaviour (e.g., time of tweets or interaction with others).

References

[1] Marcus M, Taghi Yasami M, van Ommeren M, Chisholm D, Saxena S. (2012) "Depression: A global public health concern." WHO Department of Mental Health and Substance Abuse.

[2] Zolezzi, Monica, Maha Alamri, Shahd Shaar, and Daniel Rainkie. (2018) "Stigma Associated with Mental Illness and Its Treatment in the Arab Culture: A Systematic Review." International Journal of Social Psychiatry 64 (6): 597–609.

[3] Choudhury, Munmun De, Scott Counts, and Eric Horvitz. (2013) "Social Media as a Measurement Tool of Depression in Populations." in Proceedings of the 5th Annual ACM Web Science Conference, 47–56.

[4] Lyness, Jeffrey."Patient education: Depression in adults (Beyond the Basics)." UpToDate. https://www.uptodate.com/contents/depression-in-adults-beyond-the-basics

[5] Radloff, Lenore. (1977) "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." Applied Psychological Measurement 1 (3): 385.

[6] Simon, Gregory, and Ciechanowski,Paul. (2016) "Unipolar Major Depression in Adults: Choosing Initial Treatment." UpToDate. https://www.uptodate.com/contents/unipolar-major-depression-in-adults-choosing-initial-treatment

[7] Boudad, Naaima, Rdouan Faizi, Rachid Oulad Haj Thami, and Raddouane Chiheb. (2017) "Sentiment Analysis in Arabic: A Review of the Literature." Ain Shams Engineering Journal.

[8] Abdulla, Nawaf a, Nizar a Ahmed, Mohammed a Shehab, and Mahmoud Al-ayyoub. (2013) "Arabic Sentiment Analysis: Lexicon-Based and Corpus-Based." Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT΄13) 6 (12): 1–6.

[9] Shoukry, Amira, and Ahmed Rafea. (2012) "Sentence-Level Arabic Sentiment Analysis." In Proceedings of the 2012 International Conference on Collaboration Technologies and Systems, CTS 2012, 546–50.

[10] Elnagar, Ashraf. (2016) "Investigation on Sentiment Analysis for Arabic Reviews." In Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA.

[11] Park, M, C Cha, and M Cha. (2012) "Depressive Moods of Users Portrayed in Twitter." In Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD), 1–8.

[12] Coppersmith, Glen, Mark Dredze, and Craig Harman. (2014) "Quantifying Mental Health Signals in Twitter." In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 51–60.

[13] Nadeem, Moin, Mike Horn, Glen Coppersmith, and Sandip Sen. (2016) "Identifying Depression on Twitter."

[14] Aldarwish, Maryam Mohammed, and Hafiz Farooq Ahmad. (2017) "Predicting Depression Levels Using Social Media Posts." In Proceedings - 2017 IEEE 13th International Symposium on Autonomous Decentralized Systems, ISADS 2017, 277–80.

[15] Choudhury, Munmun De, Michael Gamon, Scott Counts, and Eric Horvitz. (2013) "Predicting Depression via Social Media." Icwsm 2: 128–37.

[16] Coppersmith, Glen a, Craig T Harman, and Mark H Dredze. (2014) "Measuring Post Traumatic Stress Disorder in Twitter." In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media Measuring, 579–82.

[17] Mohammad, Saif M., and Felipe Bravo-Marquez. (2017) "Emotion Intensities in Tweets," 65–77.

[18] Larkey, Leah S., Lisa Ballesteros, and Margaret E. Connell. (2007) "Light Stemming for Arabic Information Retrieval." Arabic Computational Morphology. Springer, Dordrecht. 221-243.

[19] Liaw, Andt, and Matthew Wiener. (2002) "Classification and Regression by RandomForest." R News 2, no. 3: 18-22.

[20] Shukla, Ashish, and Shweta Shukla. (2015) "A Survey on Sentiment Classification and Analysis Using Data Mining."International Journal of Advanced Research in Computer Science 6, no. 7.

[21] Wang, Gang, Jianshan Sun, Jian Ma, Kaiquan Xu, and Jibao Gu. (2014) "Sentiment Classification: The Contribution of Ensemble Learning." Decision Support Systems 57 (1): 77–93.

[22] Fan, Rong-En, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. (2008) "LIBLINEAR: A Library for Large Linear Classification." Journal of Machine Learning Research 9, no. Aug:p. 1871-1874.

[23] Awwad, Hunaida, and Adil Alpkocak. (2017) "Using Hybrid-Stemming Approach to Enhance Lexicon-Based Sentiment Analysis in Arabic." 2017 International Conference on New Trends in Computing Sciences (ICTCS), 229–235.