I.J. Information Technology and Computer Science, **2021, 5, 41-50**
Published Online October 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2021.05.04

Data Mining for Cyberbullying and Harassment Detection in Arabic Texts

Eman Bashir[1]
[1]Collage of Computer Sciences and Information Technology, Sudan University of Science and Technology, Khartoum, Sudan
E-mail: eman_babiker@sustech.edu

Mohamed Bouguessa[2]
[2]Department of Computer Science, University of Quebec at Montreal, Montreal, QC, Canada
E-mail: bouguessa.mohamed@uqam.ca

Received: 31 July 2021; Revised: 07 August 2021; Accepted: 23 August 2021; Published: 08 October 2021

Abstract: Broadly cyberbullying is viewed as a severe social danger that influences many individuals around the globe, particularly young people and teenagers. The Arabic world has embraced technology and continues using it in different ways to communicate inside social media platforms. However, the Arabic text has drawbacks for its complexity, challenges, and scarcity of its resources. This paper investigates several questions related to the content of how to protect an Arabic text from cyberbullying/harassment through the information posted on Twitter. To answer this question, we collected the Arab corpus covering the topics with specific words, which will explain in detail. We devised experiments in which we investigated several learning approaches. Our results suggest that deep learning models like LSTM achieve better performance compared to other traditional cyberbullying classifiers with an accuracy of 72%.

Index Terms: Cyberbullying, Social network, Arabic text and Deep learning.

1. Introduction

1.1. Context

Online social networks are attracting more people where they communicate freely with each other and can share ideas, as well as comments on various events and issues; this information is beneficial to analysis. Data mining and machine learning techniques [1, 2, 3] provide tools needed to analyze complex, vast, and frequently changing social media data. Applying these techniques to social media has gained new perspectives on human behavior and human interaction in recent years. These techniques for social media can help deal with three main challenges. First, the dataset for social media is significant, considering there are millions of social media users without automated information processing for analysis; thus, the analyses will become unattainable for any reasonable amount of time. Second, social media can be noisy, which means much spam is available besides excessive trivial posts that are not desirable. Third, online social media datasets are dynamic, which change or frequently update over short periods [4].

Cyberbullying affects a lot of children all over the world, including Arab countries. Serious cyberbullying is on the rise worldwide; much cyberbullying research has been done in multiple languages in English, Chinese, Indian, the Dutch, but rarely in Arabic cyberbullying. Twitter is a popular social network where users can share short SMS-like messages called tweets. Arabic dataset contains data collected from Twitter stream API labeled automatically and manually for the study of cyberbullying traces in social media.

The Arabic language researches are scant for its complex underlying nature. There are three contrasts in the Arabic language, classical Arabic known as the Islamic manuscript language. Also, Modern Standard Arabic (MSA) is an official language known to all Arabs and used nowadays in news and textbooks. Finally, Arabic dialects are usually used informally among people [5].

1.2. Motivations

With our paper, we seek answers to the following issues, using the empirical results we have obtained as the following:

- Cyberbullying and harassment detection is a challenge that should use to protecting victims automatically in different social media platforms.

- A few successful efforts have protected offensive behaviors in online communities because no cyberbullying detection method has established at the time.
- The problem that we tackle is difficult because the Arabic dataset is not easier to collect and most of the related research used an English dataset.
- Some posts and comments in social media should analyze to detect hurtful words.
- Measure the LSTM algorithm effectiveness compared to traditional learning algorithms for cyberbullying and harassment detection.

Based on problem reviews, the question is what is an effective strategy for detecting and assessing cyberbullying/harassment? It may prevent people from exposing themselves to offensive textual content.

1.3. Objectives

One aim of this research is to investigate the problem of cyberbullying in Arabic languages. As seen in previous research [1, 2], there is some work done for cyberbullying in English, but none in the Arabic language. They considered the hypothesis that Arabic cyberbullying detection is a challenge. This paper focuses on developing an approach to detect cyberbullying and harassment in an Arabic text on Twitter automatically. We are working at capturing the semantic relationship between the words in sentiment analysis, using Word2Vec followed by supervised classification algorithms.

This paper has its main objectives, which are:

- Exhibit the accuracy of both the Continuous Bag of Words (CBOW) model and Skip-gram for the cyberbullying dataset.
- Demonstrate the concept of word embedding, where similar words have similar embedding using cosine distance improved classifier performance.
- Compare the accuracy of Random Forest, Naive Bayes Classifier for Multinomial Models, Linear Support Vector Machine, Logistic Regression, Ridge Classifier, Bernoulli Naive Bayes (BNB), and kNN Classifier.
- Demonstrate the superior accuracy of the LSTM algorithm compared to the traditional machine learning classifiers.

We organize the rest of this paper as follows: Section two describes the related work. Section three presents our approach in detail. Finally, Section four concludes and summarizes this paper.

2. Related Work

Among the main approaches to handle cyberbullying in social media is [6], where a system is made to join Arabic Twitter streaming API for gathering tweets and afterward to characterize Twitter clients as indicated by whether they have utilized any of these words in their tweets. Additionally, they added an enormous corpus of ordered user remarks that were erased from the Aljazeera Arabic news site for infringement of the site's principles. Their work presents an automated strategy to make and grow a rundown of profane words and an enormous corpus of explained user comments for vulgar and offensive language detection. They distinguished the contribution of pejoratives and obscenities in detecting offensive and introduce hand-authoring syntactic rules in identifying name-calling harassment by the Log Odds Ratio (LOR).

Another body of research [7] that introduced a model framework that was carried out in Arab nations is the use of Arabic swearing keyword lists to observe social network sites and identify bullying occurrences. Their features include tweet-based features, profile-based features, and social-graph features, which resulted in good predictions.

Further, the experiments approved their framework evaluated using different sets of tweets and features to determine the minimum sets of tweets and features that can achieve the highest classifier performance. They tested their approach using many learning algorithms: Naïve Bayes (NB), Support Vector Machine (SVM), and Decision Tree (J48) classifiers and recorded the best outcome with Naïve Bayes accomplishing an exactness of 90%.

Similarly, [5] is another study on cyberbullying detection done by authors who used traditional learning algorithms: Support Vector Machine classifiers (SVM) and Naïve Bayes (NB). They collected their data from Facebook and Twitter websites. When they applied both algorithms, the best result was accomplished by SVM with higher precision. The results obtained showed improvement in Precision, Recall, True Positive Rate, False Positive Rate, and F-Measure.

In [8], analysts used informational collections from MySpace consolidated client data like age and sex as highlights in their methodology. They applied SVM to detect cyberbullying and established that joining user-based substances improved the exactness of SVM. These studies indicate that many analysts believe that the use of gender highlighting can improve the separation limit of classifiers. However, the experiments were conducted on a limited size of dataset. Also, the authors must go to investigate other features which may differentiate writing styles of users such as age, profession, and educational level, not just gender.

The authors in [9] used FastText as their neural network model to identify an offensive text from YouTube,

MySpace, and Slashdot platforms. They compared the word embedding features with representations of traditional features such as n-gram features, and handcrafted features. They found that the pre-trained word embedding doesn't assure better performance in the FastText classifier because corpus characteristics used for pre-training the embeddings being different from their datasets.

In other later research found in [10], they tested their approach on a large-scale data set of user comments collected from the Yahoo Finance website. They were using Paragraph2Vec and BOW representation, using (TF) and (TFIDF) encoding. Once they had learned vector representations, they trained logistic regression classifiers and reported the classification performance of other methods after 5-fold cross-validation. Moreover, paragraph2vec achieved a higher AUC with 80% than BOW (TF) and BOW (TF-IDF).

Another research uses the Twitter dataset found in [11]. However, the approach automatically analyzes tweets into three classes: offensive, hateful, and clean. The authors considered n-grams as features and passing their term frequency-inverse document frequency (TFIDF) values to models. It recorded the best result with Logistic Regression achieving an accuracy of 95.6% because it is easier to implement, interpret, and very effective to train.

Cyberbullying detection was adopted in [12]; they improved features by taking emotions and words together in text messages. The study aims are to use social media messages written in the Turkish language to detect cyberbullying of Twitter and Instagram text. For most cases, the SVM accuracy is lower than Naive Bayes Multinomial (NBM) and kNN. However, SVM performance highly depends on parameter optimization. They used default parameters for SVM that caused the poor performance. Further, to increase the accuracy rate of classifiers, they applied chi-square (CHI2) feature selection and Information Gain (IG) methods, so the NBM classifier's accuracy had improved up to 84%.

In [13], the authors suggested rating tweets in French using a dataset of positive and negative emojis and training them to include Sentiment Specific Word Embeddings (SSWE) on top of an unsupervised Word2Vec model. It updated the embedding through deep learning with bidirectional LSTM on the auto-labeled data. They used French data to train restaurant reviews (Train) and another to test it (Test1) and used museum reviews as (Test2) and the manually labeled data set of tweets as (Test3). This sentiment-specific word embedding (SSWE) has performed better with an AUC improvement of 1.34%, 2.37%, and 0.79% on Test1, Test2, and Test3.

In extensive experiments, the authors of [14] have been utilizing three global datasets: Wikipedia, Formspring, and Twitter. They investigated Deep Neural Networks (Model1-CNN, Model2 LSTM, Model3-BLSTM, and Model4-BLSTM with attention) models which were using a variety of word representation approaches. They noticed that Model2 LSTM had weak performance than the Model3-BLSTM, Model4-BLSTM, and Model1-CNN. Moreover, they achieved better performance using three datasets for cyberbullying detection with no significant performance gap.

3. Proposed Approach

Our unit model consists of four main steps: data collection, pre-processing, sentiment classification, and evaluation.

3.1. Data Collection

For collecting data from Twitter, we just need to register our application to get the consumer secret, consumer key, the access secret, and access key which can be put in our Java code. Twitter provides an Application Program Interface (API) to collect tweets based on a list of specific cyberbullying words (table 1) dated from October 2018 to February 2019 and returns the tweets in JSON format and saves them as CSV files.

Table 1. List of Cyberbullying Keywords

Keywords		
سأقتلك (kill you)	****ابن ال	سأقتل نفسي (kill myself)
التحرّش الجنس (Sexual harassment)	***ن	بكرهك (hate you)
***سـأغ	***س	**** ابن
صور ع***	انتحر (Suicide)	**ن

Using specific keywords is very useful and can find that positive tweets and negative sentiment tweets can pool based on specific keywords to label the data automatically, thus removing the load of data described manually. When getting data from Twitter, the Twitter API contains comprehensive information such as user ID, user screen name, Tweet location, time and the date of tweets, and Tweet text (i.e., the main tweet text containing information about emotions, thoughts, behaviors, and other personally salient information).

We used this information to develop a set of features to classify data efficiently from Twitter and use them in different applications. All 36,056 tweets collected for this research related to Arabic swearing words; we use these words as searching seeds in the Twitter search engine.

3.2. Data Pre-Processing

We normalized tweets using two steps. The first step was cleaning up the tweet, and the second step was labeling.

A. Data Cleaning

A corpus is available in UTF-8 formats and has been cleaned and pre-processed. The unwanted data is removed, and the database is loaded into a single file, and we applied text filtering to keep only certain words. For example:

- Removed tweets that begin with RT.
- Removed @ from the text of the tweets.
- Removed hyperlinks and hashtags.
- Gets rid of punctuation and Arabic diacritic.
- Gets rid of English and Arabic numbers from tweets.
- Removed duplication.
- Removed all Latin chars.
- Removed all extra white spaces.
- Replacing the letter {ة} with {ه}
- Replacing the letter "ى" with "ي".
- Replacing the letter "آ ,أ ,إ" with "ا".

B. Data Labeling

One of the most dynamic areas of research in natural language processing (NLP) is sentiment analysis (SA), which can help analyze trending topics such as cyberbullying and harassment crises. SA can also help to predict a crisis before it occurs [15, 16].

They can accomplish sentiment analysis through both supervised and unsupervised learning techniques. They can accomplish sentiment analysis through both supervised and unsupervised learning techniques. Both techniques labeled data used to define the subjectivity, polarity, or features of the certain text. The polarity indicates whether certain content (tweet) is negative, positive, or neutral [15], as displayed in Table 2.

Table 2. Data Pre-processing

Clean data	Negative	Neutral	Positive
32,428	14516	17749	163

The presence of cyberbullying words can affect the meaning of the context, so they can have a significant effect on the text sentiment.

We construct a lexicon of all cyberbullying words and this feature will take on a value of -1 if the text contains one cyberbullying word and a value of 1 if the text contains one positive word, and a value of 0 otherwise.

4. Word Embedding

One of the oldest**,** and still one of the most commonly used is for text representation is vector space models (VSMs). Traditionally, the use of the vector space model is primarily to represent documents with the latest expansion of this model to the representation of the word or term. VSM depends on the distribution of the hypothesis, which expresses that words that occur in the same contexts have similar meanings. It bases two primary ways of building these representations on the methods of dealing with counting and predictive techniques.

Counting-based techniques measure repetition statistics between words and then map these statistics into a dense vector for each word. Predictive strategies attempt to predict a word from its neighbors in terms of a dense vector acquired for each word [16].

The term "word embedding" is used here to refer to create language models and feature learning techniques in Natural Language Processing (NLP), where words from the vocabulary are assigned to the vectors of real numbers [17]; Words with the same meaning have the same representation of the word embedding [16], so these word embeddings are considered numerical texts representations.

Various word combinations can be divided into two parts: frequency-based embedding and prediction-based embedding.

4.1. Frequency-based embedding

There are many applications of the frequency-based method like sentiment analysis and text classification. The frequency-based methods can easily classify the text with the machine learning algorithms because they extract positive and negative words from their text [18]. There are three kinds of vectors we experience under frequency-based strategies: TF-IDF vector, co-occurrence vector, and count vector.

4.2. Prediction based embedding

Both Continuous Bag of Words (CBOW) and skip-gram models estimate the effectiveness of word representations

in vector space executed as Word2Vec.

The Continuous Bag of Words (CBOW) model predicts The Continuous Bag of Words (CBOW) model predicts a given target word for context words as input [18]. Continuously distributed representation is used in the context that distinguishes it from the standard bag of the word model.

Let us understand this with an example: “The cat jumped over the puddle.” which treats {“The”; “cat”; “over”; “the”; “puddle”} as context, and from those words, someone can generate or predict the center word “jumped” [19].

There is a one-hot encoded layer of context words in the input layer. Also, the hidden layer is an N-dimensional vector picked to represent our word. The output layer is the output word, which is also one-hot encoded. We take the weight between the hidden layer and the output layer as a vector representation of the word [18, 19].

Skip-gram model (SG): it works as the opposite of CBOW. The aim of it is to use the target and context words by taking a word and predicting the context word from it. For example, the model will predict the words “The”, “cat”, “over”, “the”, “puddle” by giving the center word “jumped” [18, 19].

4.3. Building the model

Gensim is a solid open-source vector-space modeling and topic modeling toolkit implemented in Python. Gensim covers implementations of Word2Vec, Document2Vec algorithms [20]. A Gensim library provides an easy way to implement Word2Vec in Python by applying the following steps:

- Training our Word2Vec on Tweets Arabic corpus.
- Getting a word vector of a word.
- Printing the word similarity scores.
- Saving your model.

This research has built different Word2Vec models from the (CBOW) and (SG) models. We applied both models in different sizes (10, 50, 100, 150, and 200). We also used a small window size (the number of surrounding contexts or words) of 5 for Twitter because the highest length of a tweet is 140 characters. The minimum count = 3 is important to address the linguistic error. If we repeat the word in a dataset less than 3 times, we have not considered it.

Table 3. Sample of Models

Model	Sizes	وقح	منحط	Similarity value
Skip - gram	50	('خسيس', 0.9548759460449219) ('واسمه', 0.951399564743042) ('بيتز', 0.9505816102027893) ('لنيم', 0.9482298493385315) ('ردي', 0.9416403770446777) ('كلب', 0.9397455453872681) ('نعاج', 0.9377070665359497) ('ننن', 0.9341303110122681) ('مستبدة', 0.9297273755073547) ('فاسد', 0.924665093421936)	('قذر', 0.9311854839324951) ('حقير', 0.927175760269165) ('خبيث', 0.8939614295959473) ('تافه', 0.8831987380981445) ('وقح', 0.8830845355987549) ('خسيس', 0.8680808544158936) ('قبيح', 0.8676480650901794) ('خاين', 0.860995888710022) ('فاسد', 0.8588832020759583) ('نذل', 0.8523326516151428)	0.628459712532649
CBOW	150	('حقير', 0.9754908680915833) ('قذر', 0.9716367125511169) ('ساقط', 0.970047116279602) ('البعير', 0.9651697278022766) ('انفو', 0.9645867943763733) ('نجس', 0.9643831849098206) ('ياض', 0.9634526968002319) ('مختلط', 0.9583791494369507) ('خسيس', 0.9578557014465332) ('العوالم', 0.9577345848083496)	('ازرق', 0.9893141984939575) ('قذر', 0.9844918847084045) ('ساقط', 0.9832521677017212) ('زبالة', 0.9820957183837891) ('وقح', 0.9795246124267578) ('كوره', 0.9793844223022461) ('للهلال', 0.9757373929023743) ('حقير', 0.9756432771682739) ('لنيم', 0.9749078154563904) ('معلش', 0.9746178984642029)	0.881933588286418 1

4.4. Choosing a word2vec model

As we built models to implement as a piece of sentiment analysis, the principal approach to test the models is by checking the words similar to "منحط"(varmint) and "وقح"(wicked) as we use these words to express the negative sentiment. Another approach to test the models is by checking the highest similarity values. So, the best models are Skip-gram with a window size of 50 and CBOW with a size of 150.

A. Some Machine Learning Classifiers

In this experiment, the various machine learning classifiers will be applied with varying text feature selection methods to the dataset to improve accuracy. We get the text features using Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TFIDF). The TF is the frequency of each word in the dataset. TFIDF obtained by weighting each word in the dataset and combining the word frequency and inverse document frequency [21].

The classifiers that have been using are:

Multinomial Naive Bayes (MNB): It is a Naive Bayes classifier that used multinomial distribution for each of the features. The distribution is approximated by considering Naive Bayes' generative principle, which assumes that features are multinomial and distributed to count the probability of the document for each label and keep it to maximize its probability [22].

Linear Support Vector Classification (LSVC): it improves the traditional support vector machine (SVM) to longitudinal data by simultaneously rounding the well-known SVM separating hyperplane parameters with suggested temporal trend parameters [23].

Logistic Regression (LR): One of the standard algorithms used for binary classification and measures the relationship between the dependent variable (our label, what we want to predict) and one or more independent variables (the features); by rounding probabilities with uses the logistic function [24].

Ridge Classifier (RDG): one of the most fundamental regularization techniques that specialize in analyzing multiple regression multicollinearity data in nature. In this model, one predicted value in multiple regression models is predicted with others to achieve a certain level of accuracy [25].

Bernoulli Naive Bayes (BNB): It is a representation of a document as a vector of binary numbers. Each binary number represents the existence or absence of certain words. If the word is available in the document, it will take value 1 and otherwise is 0 [26].

Random Forest Classifier (RF): an ensemble model that uses an average to improve the model accuracy prediction and controls over-fitting. It grows many trees and classifies objects based on the "votes" of all the trees, which could ease the high-bias problem (over-fitting) [18].

kNN classifier: classified an object by a majority vote of the object-neighbors in the input space parameter. They assign the object to the class, which is most common among its k (an integer number), the nearest neighbor. kNN classifies objects based on feature similarity (feature = input variables) [18].

5. Long Short-Term Memory (LSTM)

Deep learning is a class of machine learning algorithms that has many models such as Long-Short Term Memory (LSTM) networks, Recurrent Neural Networks (RNN), and Convolutional Neural Network (CNN) networks [14, 27, 28].

Deep learning will be used in available data and combines it with automatically extracted hidden patterns within the text of the posts to detect many abusive behavioral norms that are highly interrelated.

In this article, we used (RNN) implementing (LSTM) model because they have shown success in understanding word chains and interpreting their meaning. This LSTM model can add or remove information to the state of cells through gate layers.

The LSTM is different from the typical neuron of RNN as the following points: LSTM has the control of deciding when to let the input enter the neuron, also the controlling of remembering what computed in the previous time step. LSTM also has the power in determining when to make the output pass on the next timestamp. Gates control all the input, output, and cell states in this way; the LSTM can decide to remember or forget the information in the recurrent layer. LSTM keeps the memory for a while in a memory cell and is controlled through three gates: inputs, output, and forget gate. The input gate activates the entry of information to the memory cell, and the forgetting gate selectively erases information in the memory cell and enables the storage to the next input; lastly, the output gate defines what information the memory cell should be output. That memory cell gives the model to remember insights derived from

the words through the comment and the capacity of learning the long-term dependencies, which is very helpful for our task [29].

We proposed the LSTM methodology to analyze texts and detecting cyberbullying and harassment as summarized in Fig.1.

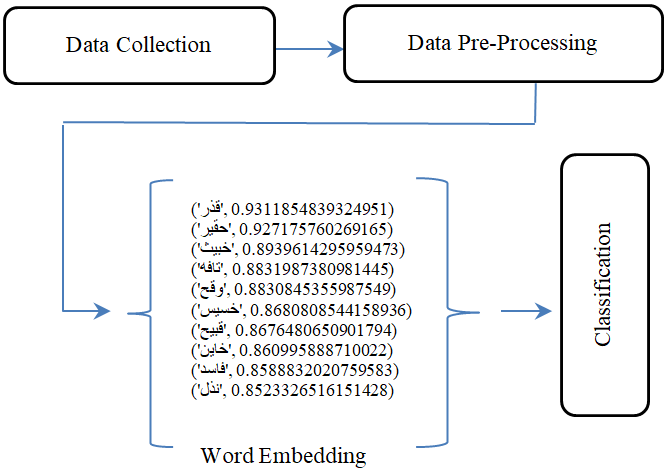

Fig.1. Flowchart Displays Research Methodology.

6. Result and Evaluation

In this section, the performance evaluation of the traditional approaches of cyberbullying detection algorithms depends on four measures: accuracy, precision, recall, and F-Measure [26, 30], as shown in table [3]. All these four metrics have the following equations that can be defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F1\_score = \frac{2*Precision*Recall}{Precision+Recall} \quad (4)$$

True - positive (TP) is used for correctly classified sentences, False - positive (FP) is used for wrongly classified sentences. False-negative (FN) is used for non-classified sentences predicted negative, True - negative (TN) is used for non-classified sentences predicted positive.

We used Anaconda3 software for classification. In table [4], we present results for the eight classifiers as illustrated. This experiment runs using the Spyder environment in Python, and LSTM was implemented using Keras.

Table 4. Classifiers Performance Results.

Algorithms	Accuracy	Precision	Recall	F1-Score
MNB	0.49	0.8	0.25	0.38
LSVC	0.44	0.79	0.22	0.35
LR	0.58	0.56	0.59	0.58
RDG	0.61	0.72	0.49	0.58
BNB	0.48	0.63	0.32	0.43
RF	0.68	0.91	0.54	0.68
KNN	0.67	0.88	0.56	0.61
LSTM	0.72	0.91	0.60	0.72

The Random Forest classifier (RF) shows the best result among the traditional classifiers with an accuracy of 68%, while the LSTM achieves the best overall accuracy of 72%.

In our experiment, the study aims to explore the performances of both the CBOW model and SK model which gave the COW model the best results as shown in Fig.2.

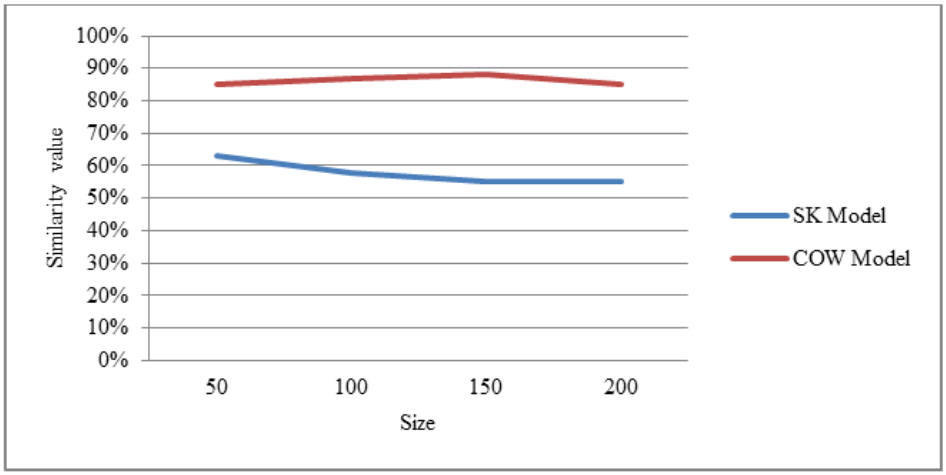

Fig.2. Continuous Bag of Words Model vs. Skip-gram Model

The results of this study showed the machine learning classifiers got an accuracy of less than 70%, while LSTM approaches got an accuracy of 72%, which was one of the main aims. Along with the performance of LSTM was raising its accuracy from 65% to 72% after the 10th epochs, as illustrated in Fig.3.

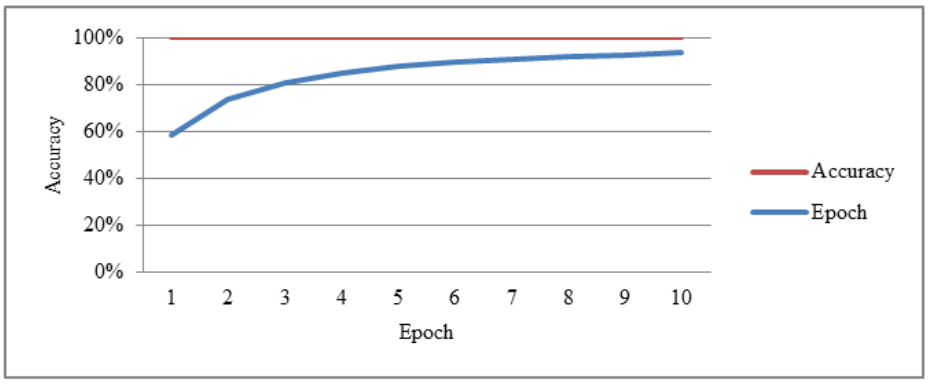

Fig.3. LSTM Model Accuracy

The LSTM makes cyberbullying detection automatically by skipping the two steps of feature extraction and feature collection. However, we not obtained satisfactory results to LSTM expectations, which achieved high accuracy rates with most huge datasets.

7. Conclusion

Most existing approaches proposed answers for recognizing cyberbullying is in the English language, yet none covered cyberbullying in the Arabic language. The present study was designed to determine the effect of machine learning approaches and LSTM that analyze tweets and comments to identify harassment and cyberbullying texts. To see if cyberbullying detection can be improved using the word embedding (Word2Vec) in the machine learning and LSTM algorithms. All processes applied to the same dataset with seven models of machine learning and LSTM. Furthermore, the performance was evaluated using four performance measures: Accuracy, Precision, Recall, and F1-Score. A limitation of this study is that the dataset size does not allow us to be more accurate, and it can only detect different cyberbullying and harassment based on specific swears words. Despite its limitations, the study certainly adds to our understanding of the LSTM is more accurate than machine learning algorithms which were listed in the result and evaluation.

Further, in future work, there will be an attempt to provide a more accurate assessment of cyberbullying/harassment detection in Arabic texts by extending the list of cyberbullying keywords and the proposed framework using new deep learning models. Our goal is to spare everyone, especially kids, from becoming victims of cybercrime by improving the accuracy of deep learning algorithms in all social media.

References

[1] Abdur Rahman, Mobashir Sadat, Saeed Siddik, "Sentiment Analysis on Twitter Data: Comparative Study on Different Approaches", International Journal of Intelligent Systems and Applications, Vol.13, No.4, pp.1-13, 2021.

[2] Marina Azer, Mohamed Taha, Hala H. Zayed, Mahmoud Gadallah, "Credibility Detection on Twitter News Using Machine Learning Approach", International Journal of Intelligent Systems and Applications, Vol.13, No.3, pp.1-10, 2021.

[3] Waheed G. Gadallah, Nagwa M. Omar, Hosny M. Ibrahim, "Machine Learning-based Distributed Denial of Service Attacks Detection Technique using New Features in Software-defined Networks", International Journal of Computer Network and Information Security, Vol.13, No.3, pp.15-27, 2021.

[4] Chen, Hsinchun. Dark web: Exploring and data mining the dark side of the web. Vol. 30. Springer Science & Business Media, 2011.

[5] Haidar, Batoul, Maroun Chamoun, and Ahmed Serhrouchni. "A multilingual System for Cyberbullying Detection: Arabic Content Detection Using Machine Learning." Advances in Science, Technology and Engineering Systems Journal 2.6 (2017): 275-284.

[6] Mubarak, Hamdy, Kareem Darwish, and Walid Magdy. "Abusive Language Detection on Arabic Social Media." Proceedings of the first workshop on abusive language online. 2017.

[7] Abozinadah, Ehab A., Alex V. Mbaziira, and J. Jones. "Detection of Abusive Accounts with Arabic Tweets." Int. J. Knowl. Eng.-IACSIT 1.2 (2015): 113-119.

[8] Dadvar, M., Trieschnigg, D., Ordelman, R. and de Jong, F. "Improving Cyberbullying Detection with User Context." European Conference on Information Retrieval. Springer, Berlin, Heidelberg, 2013.

[9] Chen, Hao, Susan McKeever, and Sarah Jane Delany. "Abusive Text Detection Using Neural Networks." AICS. 2017.

[10] Djuric, Nemanja, et al. "Hate Speech Detection with Comment Embeddings." Proceedings of the 24th international conference on world wide web. 2015.

[11] Gaydhani, A., Doma, V., Kendre, S. and Bhagwat, L."Detecting Hate Speech and Offensive Language on Twitter Using Machine Learning: An N-Gram and Tfidf Based Approach." arXiv preprint arXiv:1809.08651 (2018).

[12] Özel, S.A., Saraç, E., Akdemir, S. and Aksu, H. "Detection of Cyberbullying on Social Media Messages in Turkish." 2017 International Conference on Computer Science and Engineering (UBMK). IEEE, 2017.

[13] Saroufim, Carl, Akram Almatarky, and Mohammad Abdel Hady. "Language Independent Sentiment Analysis with Sentiment-Specific Word Embeddings." Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 2018.

[14] Agrawal, Sweta, and Amit Awekar. "Deep Learning for Detecting Cyberbullying across Multiple Social Media Platforms." European conference on information retrieval. Springer, Cham, 2018.

[15] Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A. and Al-Ohali, Y."Arasenti-Tweet: A Corpus for Arabic Sentiment Analysis of Saudi Tweets." Procedia Computer Science 117 (2017): 63-72.

[16] Al-Ayyoub, Mahmoud, et al. "A Comprehensive Survey of Arabic Sentiment Analysis." Information processing & management 56.2 (2019): 320-342.

[17] Goldberg, Yoav. "Neural Network Methods for Natural Language Processing." Synthesis lectures on human language technologies 10.1 (2017): 1-309.

[18] URL: https://medium.com/data-science-group-iitr/word-embedding-2d05d270b285. . Accessed Dec. 2019.

[19] Mikolov, T., Chen, K., Corrado, G. and Dean, J. "Efficient Estimation of Word Representations in Vector Space." arXiv preprint arXiv:1301.3781 (2013).

[20] URL: https://radimrehurek.com/about/. . Accessed Dec. 2020.

[21] Ghosal, Sambuddha, et al. "A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting." Plant Phenomics 2019 (2019).

[22] Lohar, P., Dutta Chowdhury, K., Afli, H., Hasanuzzaman, M. and Way, A."ADAPT at IJCNLP-2017 Task 4: A Multinomial Naive Bayes Classification Approach for Customer Feedback Analysis Task." (2017).

[23] Du, Wei, et al. "A Longitudinal Support Vector Regression for Prediction of ALS Score." 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2015.

[24] URL: https://www.experfy.com/blog/the-logistic-regression-algorithm. Accessed Jan. 2020.

[25] URL: https://mindmajix.com/ridge-regression. Accessed Jan. 2020.

[26] Diab, Diab M., and Khalil M. El Hindi. "Using Differential Evolution for Fine Tuning Naïve Bayesian Classifiers and Its Application for Text Classification." Applied Soft Computing 54 (2017): 183-199.

[27] Hossam Elzayady, Khaled M. Badran, Gouda I. Salama, "Arabic Opinion Mining Using Combined CNN - LSTM Models", International Journal of Intelligent Systems and Applications, Vol.12, No.4, pp.25-36, 2020.

[28] K Srinivasa Rao, G. Lavanya Devi, N. Ramesh, "Air Quality Prediction in Visakhapatnam with LSTM based Recurrent Neural Networks", International Journal of Intelligent Systems and Applications, Vol.11, No.2, pp.18-24, 2019.

[29] Munir Ahmad, Shabib Aftab, "Analyzing the Performance of SVM for Polarity Detection with Different Datasets", International Journal of Modern Education and Computer Science, Vol.9, No.10, pp. 29-36, 2017.

[30] Hilal Almarabeh,"Analysis of Students' Performance by Using Different Data Mining Classifiers", International Journal of Modern Education and Computer Science, Vol.9, No.8, pp.9-15, 2017.

Authors' Profiles

Eman Bashir received her Master of Computer Science and Technology from Collage of Computer Science and Technology at University of Al Gezira, Khartoum, Sudan. She received her B.Sc. Computer Science and Mathematics from Faculty of Computer Science and mathematics at University of Khartoum, Khartoum, Sudan. She is currently a Ph.D. student in the College of Computer Science and Information Technology, Sudan University of Science and Technology, Sudan.

Mohamed Bouguessa received the MSc and the PhD degrees, respectively, in 2005 and 2009 from the University of Sherbrooke, Quebec, Canada. He is currently an associate professor of computer science at the University of Quebec at Montreal (UQAM), Montreal, Quebec, Canada. His research covers a variety of data mining related topics. His current research projects include clustering, dynamic networks and graph neural networks.

How to cite this paper: Eman Bashir, Mohamed Bouguessa, "Data Mining for Cyberbullying and Harassment Detection in Arabic Texts", International Journal of Information Technology and Computer Science(IJITCS), Vol.13, No.5, pp.41-50, 2021. DOI: 10.5815/ijitcs.2021.05.04