Comparitive studies of cyberbullying and sarcasm detection on social media

John Hani, Mohamed Nashaat, Mostafa Ahmed, Zeyad Emad
Supervised by Dr. Eslam Amer and Eng. Menna Gamil

December 2, 2018

Abstract

In this comparative paper, we are going to discuss the different approaches and methods implemented by other papers to detect cyberbullying and sarcastic comments over social media. The reason we are did so, because we wanted to find the best technique to combat harassment on social media. Moreover, we focused on comparing between different classifiers that have been used in these papers like: (SVM)Support Vector Machine, Naive Bayes and Random Forest on advantages and drawbacks of each one of them related to the cyberbullying and sarcasm detection. We also mentioned the different preprocessing and the data-sets that have been used in these papers. Finally, we discussed and compared between the results of the papers in terms of accuracy and precision.

1 Introduction

As the increasing of social media nowadays there is an increasing in the cyber-crimes and we all know that everybody now is using social media in his daily life. Cyberbullying now is one of the bad effects of the social media according to bullying statistics.org over half of the youth have been cyberbullied so the main objective of this study is to carry out comparative studies on the various Cyberbullying and sarcasm methods and we want to address the problems and the drawbacks that was in this methods hence we want to make our cyberbullying and sarcasm detection system the agenda of this paper consists of 5 sections. Section 2 describes related work regarding cyberbullying and sarcasm detection system. Section 3 discusses the comparison of the methods deployed in cyberbullying and sarcasm detection addressing its problems, Section 4 contains results and discussions of the methods and Section 5 contains the conclusion.

2 Related Work

2.1 Information Retrieval Model

This section discusses numerous ways of dealing with cyberbullying detection as mainly the information retrieval approach depends on three fundamental modules; pre-processing phase, feature extraction phase and the classification methods.

2.1.1 Pre-Processing

Pre-processing transforms the raw data into a format that will be more easily and effectively processed for the purpose of the user and to prepare it for another processing procedure, There are a lot of Pre-processing methodologies for natural language processing and each has it's own unique contribution for such occasion to occur.

The removal of stopwords for instance showed a huge contribution in cleaning up the raw data as well as save much valuable processing time as parts of the sentence are being extracted and removed though they may not add much value to the the context and sentence [9].stop words are words which are filtered out before or after processing of natural language data[8]. Stop words also depend on a filter, for instance, some languages do have a filter of their own ex. English, German [11] while some other languages you have to create your own filter manually and add the words you see to add little to no value to the sentence. Such is the case when removing the unwanted words [4] while some datasets may contain a lot of test subjects, although not every subject is perfect it may be accompanied with encoded parts from where it originated from [17]that add too much unwanted noise .[24] These words have little or no meaningful content and therefore the topics that are discussed in the text that is analyzed is not taken into consideration [16]Some tools specifically avoid removing these stop words to support phrase search.

Stemming which is the process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words known as a lemma [22]. In social media, the posts made by users are often short, noisy and unstructured. Also, these posts are not necessarily about the same topic which causes the vocabulary size to be extremely large [9]. Hence, traditional text representation techniques such as n-grams and bag-of-words become extremely high dimensional. [10] .Some simple algorithms will simply strip recognized prefixes and suffixes, same is the case with the Porter Stemmer algo [11]. However, these simple algorithms are prone to error. For example, an error can reduce words like laziness to lazi instead of lazy. Such algorithms may also have difficulty with terms whose inflectional forms don't perfectly mirror the lemma such as with saw and see. In a nutshell the process of reducing inflected (or sometimes derived) words to their word stem is quite useful when you know what to look for exactly but however you may receive an undesired result [8] for example [19] stemming the word "computation" to its root "comput". This process facilitates the identification of similar tokens. Stemming is built upon the idea that words with the same stem are close in meaning [20]. So the words are stemmed to identify the words which are similar in meaning.

Tokenization is one of the main features of lexically analyzing the text [10]. Here the aggregation of the sequence of characters takes place. Decluttering the words of the sentence into tokens is very helpful in individual identification these instances of tokens are very useful. [19]In this operation, the extracted text is broken into a set of words or tokens, where a token is the most basic unit of a text,[20] A token could be a complete group of words or a whole paragraph, but most frequently, we use words as tokens. E.g. Using Hashtag-Tokenizer to separate the hashtags that contain connected words, e.g., splitting hashtag:Sarcastic irony into hashtag:sarcastic and hashtag:irony. In general [11], every word is identified or separated by other words by a space character, single quoting character ('), dot (.), semicolon (;), colon (:), so the tokenizing process uses nonletters mode to perform word separation. Apache OpenNLP [3] was used to perform different natural language processing tasks one of such is the tokenization and it showed promising results when combined with Gate Twitter part of speech tagger. Some used a special tokenizer which is specifically used for twitter [26] applied without any stemming or stopwords removal operations.

Word Correction tends to correct the misspelled words and also the words that has been written wrong to by pass the classification system like the word Idiot any one could write it !d!iot for the humans it remains readable but for the classifier it is a different word from the word it has trained on from the dataset. [11] Also consists of the word correction from informal word to formal word or to correct misspelled word and to convert the numeric characters used in the text into alphabetic character.[1] With the help of WordNet corpus and spell-correction algorithm2, correction of spelling and grammar mistakes in the raw sentences occurs by tasks such as deleting repeated letters in words, deleting meaningless symbols, splitting long words, transposing substituted letters,and replacing the incorrect and missing letters in words.[24]

2.1.2 Feature Extraction

Feature Extraction is phase where now that the data is now in a much more readable state and is formatted correctly with no spelling mistakes, it is time to extract some meaningful information from it, each feature extraction phase contributes in its own unique way and help us understand more and more from the posts that we see on the internet.

Bad words extraction and normalization has proven to be of worthy values when it comes to extracting features from plain text, these bad words may come from a bag of words(BoW) with their own weights, like when [8] used a ready made dictionary full of bad words and their weights from (http://www.noswearing.com/dictionary) it added more value to their work as it demonstrated great results

Pronouns also contribute in a considerable way, although you have to take into account that there are numerous pronoun types the most commonly used is the personal pronouns [22] and their iterations which are first and second and third person pronouns. Such is the case with [8] as they normalized these numbers where instances of personal pronouns were divided by how many words were in a paragraph/text.

Capital letters were also used to emphasize and stress in addition to pinpointing our area of interest by normalizing these [8] instances of capital letters throughout the text and finding their exact location and passing it to other feature extraction methods to make sure that these capital letters do mean something in that text. Sometimes add little to no value at all since the user may be bluffing in these capital letter and they mean nothing and thus making the whole point of looking for capital letters in a text meaningless.
TFIDF: used to calculate the importance of the words according to their number of occurrence in the document.but it have some problems.
(1)It computes document similarity directly in the word-count space, which may be slow for large vocabularies. (2) It assumes that the counts of different words provide independent evidence of similarity. (3)It makes no use of semantic similarities between words.and it is used in ([4], [9], [6], [16] and [11]).
LIWC word lists with computer program to extract basic counts / ratios. Contains dictionaries for English, German, Spanish, Dutch, and Italian. Extracts around 60 different word categories, including "positive emotions" and "negative emotions". The program can be purchased; their site also allows you to analyze texts one by one.

There are several word lists associated with the program. An English word list is available from James Pennebaker at University of Texas at Austin.
In related to emotion there are several word categories:[1]

Aggression: Swear words
Affective processes
Anxiety
Anger
Sadness
Positive emotion
Negative emotion
It was used in [12] and Liwc have problem with false postive.
Sentiment Analysis also known as Opinion Mining is a field within Natural Language Processing (NLP) that builds systems that try to identify and extract opinions within text. Usually, besides identifying the opinion, these systems extract attributes of the expression e.g.:

Polarity: if the speaker express a positive or negative opinion,
Subject: the thing that is being talked about, Opinion holder: the person, or entity that expresses the opinion.
So it is important to any text identifying system.and it is used in sarcasm papers mainly. It is used in [20], [19], [24], [17] and [2] Sentiment analysis is not enough. Because computer programs have problems recognizing things like sarcasm and irony, negations, jokes, and exaggerations - the sorts of things a

person would have little trouble identifying. And failing to recognize these can skew the results. 'Disappointed' may be classified as a negative word for the purposes of sentiment analysis, but within the phrase I wasn't disappointed", it should be classified as positive.Also Russell said, Like all opinions, sentiment is inherently subjective from person to person, and can even be outright irrational. Its critical to mine. So we have combined the three methods to make each one of them cover the weakness of the other one.

2.1.3 Classification Methods

Our main interest is the correct classification as to try and avoid the results which do not match the ground truth results and upset the overall credibility, in addition to being on of the crucial stages, using multiple classifiers will give different results as each classifier is unique in its method but ultimately the final results are of the same nature which is true or false. The diversity of the classes and what they have to offer allows the door of hybrid classifiers to open and introduce interleaved classifiers as well as cascading classifiers.

SVM (Support Vector machine) is a supervised learning algorithm, and is one of the most efficient and universal classification algorithms. Its goal is to find the optimal separating hyperplane which maximizes the margin of training data. Initially the classifier is trained with labelled data before being used to classify the data to test accuracy[1][11]. Before the data can be used to train our classifier, it is imperative to process it.However, the most serious problem with SVMs is the high algorithmic complexity and extensive memory requirements of the required quadratic programming in large-scale tasks. According to [13] [3] Their classification model consists of two parts the first part is the part that they classify the text in to three classes that it is positive or negative or neutral. Then the second part that they classify the positive text to find whether it is an opinion or neutral and if it is an opinion it is positive or negative. Another approach was using the linear kernel [4] with the equation to calculate and illustrate the hyperplane K(xi,xj)=xj xi where K(xi,xj) is the dot product of input data xi points mapped into large dimensional feature space xj by transformation function, while some papers proposed a method which is based on feature extraction results [7] 1)the ratio of capital letters in a comment 2)the number of emoticons 3)the occurrence of a second person pronoun followed by a profane word in profanity 4) the term frequencyinverse document frequency (Tf-Idf). Other methodologies [11] of SVM were used such as the Poly, RBF, and sigmoid kernels as to show how much potential the svm has with its different iterations.

The Naive Bayes family of classifiers are simple conditional probabilistic classifiers that work by applying Bayes theorem with naive independence assumptions between the different features [11] while assuming that the presence of a particular feature in a class is unrelated to the presence of any other features. Although [21][11] [1][20] the naive bayes feature extractor was assigned to eight different sub categories but it performed extremely well when applied with K-means as a method for clustering data.

Logistic Regression This algorithm provides probabilistic approach to data. The outcome are probabilities modeled as a function of predicted variables. However, logistic regression [4] cannot predict continuous outcomes. For example, logistic regression could not be used to determine how high an influenza patient's fever will rise, because the scale of measurement – temperature – is continuous. [4] [10] using a function that calculates the probability which is P = 1 / (1 + e*-t), where P is the probability of the observation is the first step of calculating the regression coefficient, regression algorithms deal with plotting the data coordinates alongside their yield value in space and then finding a function that can estimate the output value of a coordinate.

An expert machine evaluation system [7] presented by 12 experts on cyberbullying assigned weights and importance level to features such as the likelihood of a user belonging to a category on a four point scale, second is the importance also on a four point scale. Now the results being added to a machine learning classifier to create a cascading hybrid classifier that utilizes both an expert system and machine learning classifiers, alongside with the features categories used as a new set of features in the machine learning classifier.

As [11] Done in their model, they didn't just classify bullying to binary classification meaning it's bullying or not, but rather multiclass classification they made a severity intense in the bullying. Like they are dividing the bullying to several classes and take actions according to every class.[22]

3 Information Retrieval Approaches

This paper makes a comparative study on various methods to detect cyberbullying and sarcasm on social media. The problem in the previous work in cyberbullying detection is the low of accuracy and the number of false positive also they cannot detect sarcasm along with cyberbullying.

3.1 Datasets

The previous work in cyberbullying detection used different datasets. Walisa Romsaiyud ,etal [21] used from two different datasets the first one is the posted messages by members in Perverted-justice used as training datasets, and the second one is Twitter datasets from Stanford University as testing datasets. Sani Muhamad Isa, etal [11] and Vikas S Chavan etal [4] and [11] used a textual conversation taken from the Kaggle (www.kaggle.com) which provides 1,600 conversations in Formspring.me. Harsh Dani,etal [9] ,[1] used dataset from twitter and Myspace labeled as normal or bullying. Maral Dadvar,etal [7] collect his dataset from haresment comments other misbehavioursof YouTube videos Paras Dharwaletal [10] ,Edwin Lunando etal [13], MONDHER BOUAZIZI etal [3] and S.K.Bharti etal [2] in their papers of sarcasm detection they collected the dataset manually from Twitter.In [6] they have labled dataset collected from myspring.com with ground truth made by three students but the ground truth in this dataset is too small. In [17] and [16] they are available from the workshop

on Content Analysis for the Web 2.0, it contains data from Myspace and Formspring.me. 500 post from Formspring.me and 600 posts from Myspace.In [14] they collected the data using WOT(World Of Tanks) game api which reached 26,000 messages which 5,000 messages were manually classified to compare it to the automatic classifier.In [18] data is sampled from comments posted on Yahoo! Finance and News during the period between October 2012 and January 2014.In [19] they obtained datasets from the Myspace social networking website, which were manually marked as cyberbullying but this dataset is weak because it was marked mannualy. In [5] and [7] dataset retrieved from Youtube comments from postings in reaction to the top 18 videos.. Each comment includes a user id, a timestamp and text content. The user id identifies the author who posted the comment, the timestamp records when the comment was posted and the text content contained a users comments. The dataset contians comments from 2,175,474 users.In [24] they used two datasets Bad Words Dataset and Sensitive Words Dataset. In [23] they have collected dataset from four social media websites Facebook, Twitter, Ptt (https://www.ptt.cc) and CK101 https://ck101.com/). 100 as cyberbullying and 100 as not cyber bullying for each website but these data set is too smalls.In [26] they collected dataset from twitter and manually labeled them.In [12] they have collected from three social media website.
Instagram: 41K user ids were gathered with a snowball sampling method. About 25K public profiles, these 25K public user profiles are used as our normal Instagram users.
Ask.fm: starting from a seed node and again using snowball sample, 24K were used as the normal Ask.fm users.
they have collected 1M users information via snowball sampling. Only 4 percent of the users have mentioned their Instagram id in their profile information, thereby furnishing around 40K users ids for both social networks. Only 24K of these profiles were public. From this, we collected complete profile content for an 8K subset of these common users from each of the two social networks, forming our common Instagram users data and common Ask.fm users data.The dataset is [8] they collected 4626 comments from 3858 distinct users. The comments were manually labelled as bullying (9.7percent) and non-bullying (inter-annotator agreement 93 percent). this dataset is weak because it is collected manually.In [25] They used 1) 1313 messages from twitter
2) 13,000 messages from formspring.me.
In [15] they used utilized data provided by Fundacion Barcelona Media2 for the workshop on content analysis from the Web 2.0. The given data was collected from the three different SNs including Myspace, Kongregate, and Slashdot. In [20] they collected data from twitter and labelled it manually.

3.1.1 Differences Among the datasets

The main difference in the datasets is that datasets originating from social media don't have ground truth but the datasets from websites like kaggle .com have ground truth .

3.2 Table

Authors	Approach	Datasets	Performance
Vikas S Chavan and Shylaja S S	ML	They Collected thier datasets from Kaggle Website	Logistic Regression Accuracy = 73.76 And Support Vector Machine Accuracy = 77.65
to Edwin Lunando and Ayu Purwarianti	ML	They Collected datasets manually from Twitter	Naive Bayes Accuracy = 76.5 Maximum entropy Accuracy = 76.7 Support Vector Machine Accuracy = 77.3
S.K. Bharti n And B. Vachha And R.K. Pradhan And K.S. Babu And S.K. Jena	ML	They Got their dataset from Twitter	6 algorithms precision = 97
Harsh Dani And Jundong Li(B) And & Huan Liu	ML	They collected their Datasets from Twitter and MySpace	KNN F1 = 0.6105 AUC = 0.75 using SICD
Maral Dadvar And Dolf Trieschnig And Franciska de Jong	ML	Collected their datasets from Youtube	MCES discrimination capacity = 0.72 Naive Bayes with discrimination capacity Accuracy = 0.66
MONDHER BOUAZIZI AND TOMOAKI OTSUKI	ML	Collected their datasets from Twitter	SVM Accuracy = 83.1
Walisa Romsaiyud And Kodchakorn na Nakornphanom And ETAL	ML	They Collected datasets from Perverted Justice and Twitter	Naive Bayes Accuracy = 95.79
Noviantho And Sani Muhamad Isa And Livia Ashianti	ML	They Collected their dataset from Kaggle	Naive Bayes Accuracy = 92.81 SVM Accuracy = 97.11
Paras Dharwal And Tanupriya Choudhury And Rajat Mittal And Praveen Kumar	ML	They Collected their Dataset From Twitter	

Anukarsh G Prasad; Sanjana S Skanda M Bhat, B S Harish	ML	They got Twitter data set with manually labeling it.	Random Forest: 79.44 , 80.93 , 77.49 Naive Bayes: 74.56 , 73.9, 75.1
Santosh Kumar Bharti Reddy Naidu Korra Sathya Babu	ML	Collected Sarcastic Tweets from twitter manually	The system attains an accuracy of 75.12, 80.27, 80.67, 80.79, and 80.07 using NB, DT, SVM, RF, and AdaBoost respectively
Noviantho, Sani Muhamad Isa, Livia Ashianti	ML	Fromspring Dataset from Kaggle	Naive Bayes yields an average accuracy of 92.81, SVM with a poly kernel yields an average accuracy of 97.11.
Nobata, Chikashi and Tetreault	ML	Data is sampled from comments posted on Yahoo! Finance and News	All Features:F-Score of news dataset: 0.817 F-Score of Finance dataset: 0.795
Dadvar, Maral and De Jong, Franciska	ML	The data-set is provided by Fundacion Barcelona Media collected from MySpace its size is 381,000 posts, the ground truth data-set is 2,200 labeled by three students.	basline precision was 0.31 and gender-specific precision was 0.43
Nandhini, B Sri and Sheeba, JI	ML	it contains data from Myspace and Formspring.me. 500 post from Formspring.me and 600 posts from Myspace.	MySpace Acc. is 94.50 percent Formspring.me Acc.is 94.01 percent
Nandhini, B Sri and Sheeba, JI	ML	They are available from the workshop on Content Analysis for the Web 2.0, it contains data from Myspace and Formspring.me	MySpace Acc. is 87 percent Formspring.me Acc.is 86 percent
Murnion, Shane and Buchanan and William J	ML	They collected the data using World Of Tanks game api which reached 26,000 messages which 5,000 messages were manually classified to compare it to the automatic classifier	The simple naive automatic classification has reached 91.6 percent accuracy

4 Results and Discussions

5 Conclusion

As you could see from our paper there were many methods that had been introduced in the last couple of years and yet we are going to try to increase the accuracy and recoil of the detection of cyberbullying and also the detection of sarcasm. With this combination we are going to try to make the social media more secure.

6 References

References

[1] S. K. Bharti, R. Naidu, and K. S. Babu, "Hyperbolic feature-based sarcasm detection in tweets: A machine learning approach," in *2017 14th IEEE India Council International Conference (INDICON)*. IEEE, 2017, pp. 1–6.

[2] S. Bharti, B. Vachha, R. Pradhan, K. S. Babu, and S. Jena, "Sarcastic sentiment detection in tweets streamed in real time: a big data approach," *Digital Communications and Networks*, vol. 2, no. 3, pp. 108–121, 2016.

[3] M. Bouazizi and T. O. Ohtsuki, "A pattern-based approach for sarcasm detection on twitter," *IEEE Access*, vol. 4, pp. 5477–5488, 2016.

[4] V. S. Chavan and S. Shylaja, "Machine learning approach for detection of cyber-aggressive comments by peers on social media network," in *Advances in computing, communications and informatics (ICACCI), 2015 International Conference on*. IEEE, 2015, pp. 2354–2358.

[5] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, "Detecting offensive language in social media to protect adolescent online safety," in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. IEEE, 2012, pp. 71–80.

[6] M. Dadvar and F. De Jong, "Cyberbullying detection: a step toward a safer internet yard," in *Proceedings of the 21st International Conference on World Wide Web*. ACM, 2012, pp. 121–126.

[7] M. Dadvar, D. Trieschnigg, and F. de Jong, "Experts and machines against bullies: A hybrid approach to detect cyberbullies," in *Canadian Conference on Artificial Intelligence*. Springer, 2014, pp. 275–281.

[8] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, "Improving cyberbullying detection with user context," in *European Conference on Information Retrieval*. Springer, 2013, pp. 693–696.

[9] H. Dani, J. Li, and H. Liu, "Sentiment informed cyberbullying detection in social media," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2017, pp. 52–67.

[10] P. Dharwal, T. Choudhury, R. Mittal, and P. Kumar, "Automatic sarcasm detection using feature selection," in *2017 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*. IEEE, 2017, pp. 29–34.

[11] S. M. Isa, L. Ashianti *et al.*, "Cyberbullying classification using text mining," in *Informatics and Computational Sciences (ICICoS), 2017 1st International Conference on*. IEEE, 2017, pp. 241–246.

[12] H. H. S. Li, Z. Yang, Q. Lv, R. I. R. R. Han, and S. Mishra, "A comparison of common users across instagram and ask. fm to better understand cyberbullying," in *Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference on*. IEEE, 2014, pp. 355–362.

[13] E. Lunando and A. Purwarianti, "Indonesian social media sentiment analysis with sarcasm detection," in *Advanced Computer Science and Information Systems (ICACSIS), 2013 International Conference on*. IEEE, 2013, pp. 195–198.

[14] S. Murnion, W. J. Buchanan, A. Smales, and G. Russell, "Machine learning and semantic analysis of in-game chat for cyberbullying," *Computers & Security*, vol. 76, pp. 197–213, 2018.

[15] V. Nahar, S. Al-Maskari, X. Li, and C. Pang, "Semi-supervised learning for cyberbullying detection in social networks," in *Australasian Database Conference*. Springer, 2014, pp. 160–171.

[16] B. Nandhini and J. Sheeba, "Cyberbullying detection and classification using information retrieval algorithm," in *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015)*. ACM, 2015, p. 20.

[17] B. S. Nandhini and J. Sheeba, "Online social network bullying detection using intelligence techniques," *Procedia Computer Science*, vol. 45, pp. 485–492, 2015.

[18] C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang, "Abusive language detection in online user content," in *Proceedings of the 25th international conference on world wide web*. International World Wide Web Conferences Steering Committee, 2016, pp. 145–153.

[19] S. Parime and V. Suri, "Cyberbullying detection and prevention: Data mining and psychological perspective," in *Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on*. IEEE, 2014, pp. 1541–1547.

[20] A. G. Prasad, S. Sanjana, S. M. Bhat, and B. Harish, "Sentiment analysis for sarcasm detection on streaming short text data," in *Knowledge Engineering and Applications (ICKEA), 2017 2nd International Conference on*. IEEE, 2017, pp. 1–5.

[21] W. Romsaiyud, K. na Nakornphanom, P. Prasertsilp, P. Nurarak, and P. Konglerd, "Automated cyberbullying detection using clustering appearance patterns," in *Knowledge and Smart Technology (KST), 2017 9th International Conference on*. IEEE, 2017, pp. 242–247.

[22] G. Sarna and M. Bhatia, "Content based approach to find the credibility of user in social networks: an application of cyberbullying," *International Journal Of Machine Learning and Cybernetics*, vol. 8, no. 2, pp. 677–689, 2017.

[23] I.-H. Ting, W. S. Liou, D. Liberona, S.-L. Wang, and G. M. T. Bermudez, "Towards the detection of cyberbullying based on social network mining techniques," in *Behavioral, Economic, Socio-cultural Computing (BESC), 2017 International Conference on*. IEEE, 2017, pp. 1–2.

[24] A. Upadhyay, A. Chaudhari, S. Ghale, S. Pawar *et al.*, "Detection and prevention measures for cyberbullying and online grooming," in *Inventive Systems and Control (ICISC), 2017 International Conference on*. IEEE, 2017, pp. 1–4.

[25] X. Zhang, J. Tong, N. Vishwamitra, E. Whittaker, J. P. Mazer, R. Kowalski, H. Hu, F. Luo, J. Macbeth, and E. Dillon, "Cyberbullying detection with a pronunciation based convolutional neural network," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2016, pp. 740–745.

[26] R. Zhao, A. Zhou, and K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features," in *Proceedings of the 17th international conference on distributed computing and networking*. ACM, 2016, p. 43.