Cyberbullying Detection by Including Emotion Model using Stacking Ensemble Method

Natasia[*], Sani Muhamad Isa

Computer Science Department, BINUS Graduate Program – Master of Computer Science Bina Nusantara University, Jakarta, Indonesia 11480

A B S T R A C T

Cyberbullying is a serious problem and caused an immense impact to the victim. To prevent the cyberbullying, the solution is to develop an automatic detection system. In this research, we propose a combined model for cyberbullying detection and emotion detection by using stacking method. The experiment is to create a better model for cyberbullying detection using SVM (support-vector machine), KNN (K-Nearest Neighbors), and Naive Bayes method, then combine the best model with the emotion model. The result conducted that using SVM classifier give the best accuracy for both emotion and cyberbullying detection. Emotion detection yields an accuracy of 96.7%. Cyberbullying detection using SVM classifier yields an accuracy of 72.73%. Then, both model are combined using the stacking ensemble method and yield an average accuracy of 77.8%. It concluded that including the emotion model would improve the accuracy of detection.

1   Introduction

Technologies have grown rapidly and give a big impact on our social life. It makes communication easier and become a big role in life, especially in communication and relationship among people [1]. However, the more people use these platforms, the more crime and offense occurs. Bullying in social media become common and afford a great impact to many people, especially a teenager. In 2016, showed that 41% - 50% of the teenager had experienced cyberbullying [2].

Cyberbullying could make a great impact on victims, especially young adults. Research shows that the ability to regulate emotional responses are influenced by age [3]. However, the use of the internet for social networks are around 41.7% - 46.7% for young adult [4]. This would create a long-lasting impact on mental health. Some who experienced cyberbullying shows emotional distress and a likelihood of acting out. Some also experiencing mental health problems, such as drugs, abuse, and suicide [5].

Cyberbullying is a serious problem, which could harm the mental of social media users. The increase of social media users and how social media is easy to use made it more difficult to validate the posts and comments. Moreover, many users left to be unknown are showing a bad habit in commenting. To reduce cyberbullying and the bad manner in comments, the solution is to automatically detect and validate the content.

To prevent cyberbullying, the solution is to develop an automatic detection system to categorize which content is a bully content and reporting if the system detects the kind of bullying [6]. By using natural language processing and text mining, the cyberbullying feature are extracted and classified [1], [6].

The previous study has conducted the cyberbullying detection with a psychological feature by using big five personality traits and dark triad features. The used features are number of followers, following, popularity, user favorite count, status count, extraversion, agreeableness, neuroticsm, psychopathy, and sentiment. The experiment resulting 90.1 – 91.7% accuracy, where the baseline + psychopathy have the highest accuracy [7].

In this paper, we approach to determine how user got bullied or not by including emotion features. The emotion features are extracted into 4 classes and resulting a model of cyberbullying detection. Then, the emotion model and cyberbullying model will be used as a feature to create a new model for better detection.

The study is organized as follows. In section 2, we introduce some of related works. Then, we propose our methodology for the emotion detection, cyberbullying detection, and the ensemble method. In section 4, the experiment and evaluation method are detailed. In section 5, experimental results are detailed and analyzed. Finally, concluding study results are provided in section 6.

[*]Corresponding Author: Natasia, Bina Nusantara University, Jakarta, Indonesia 11480, & natasia001@binus.ac.id

2   Background and Related Work

Considering the influence of cyberbullying on victim's mental health, several studies have been conducted. With the improvement of technology, the automatically cyberbullying detection could be applied. The main goal of researching this is to maintain the utilization of social media and reduce bullying activity in social media.

Some previous works use a heterogeneous technique to perform the detection using text mining principal. Using text mining and natural language processing concept, the step to do are preprocessing, feature generation (Bag of Words), feature selection, classification, and analyzing results [1], [6].

There also a detection by using Big Five and Dark Triad features [7], [8]. The effectiveness of cyberbullying detection is compared to the baseline model among the Big Five and Triads models. In the experiment, higher accuracy is produced for three personalities in Big Five and one personality in Dark Triad, which are extraversion, agreeableness, neuroticism, and psychopathy [7].

2.1   Text Processing

Text processing or text mining is used to extract knowledge, information, or pattern and convert the unstructured information into structured data to solve the problem [9], [10]. Preprocessing in text mining could be done in several ways. The most commonly used for preprocessing are tokenization, stemming and lemmatization, stop word and n-gram [10], [11].

Tokenization is used to split or segmented text into tokens, such as words and punctuation by separating each text from whitespace and punctuation. Stemming and lemmatization are used to derivationally the word into the base form. The stemming works by cutting off the end of the word into a list of common prefixes or suffixes. Lemmatization work by analyzing the word and link the text with its basic form.

Stop word is used to filter out the word that does not use for further process. For high-frequency words, such as "the", "I", "we", etc, the word is filtered to improve the accuracy in data. N-gram is used to determine the connection with each word with a contiguous sequence of n-words. This is used to calculating the probabilities and improving the predictions of a word that is connected.

After the pre-processing, the next step is information extraction or feature extraction. Feature extraction is a process for raw data to be converted into a manageable resource. The feature extraction contains clustering, where the extracted data will be clustered based on their feature [9].

This process will convert the text into base-features and effectively reducing the amount of data to be processed while still accurately describing the original data. Feature extraction could be using several methods. In the previous work, it used POS-Tag, BoW, EboW [1], Word2Vec [11], or TF-IDF [12], [11].

TF-IDF (term frequency-inverse document frequency) is a measurement that determined the usefulness of word collection in a document [13]. With given D as a document, w as word, f as frequency, the importance of a word is defined in equation 1 [14].

$$wd = f(w, d) * log(|D|/f(w, d)) \tag{1}$$

The next step is classification. Classification is a process to categorize the set of data into classes. In text mining, there are many algorithms to classified data, for example, using SVM, KNN, Decision Tree, Naïve Bayes, and many else [15], [16], and [11].

SVM uses kernel trick to map the data into a higher-dimensional space. The simplest SVM formula is linear SVM, where the hyperplane are placed on the space of the input data. The objective of using SVM algorithm is to find a boundary / hyperplane that classifies the data features (Figure 1) [17], [18].
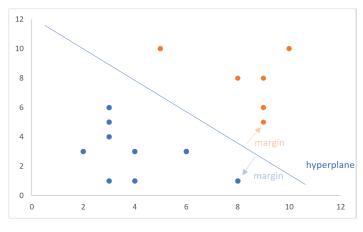

Figure 1: SVM Example

When SVM uses the hyperplane, KNN relies on the labeled data and builds a learning method to obtain the prediction of given unlabeled data. Before inputting the data into KNN classification, the data must be extracted into weight feature, by using TF-IDF, frequency, or a score. KNN will create a mapping for each class, with a border value to the other class.

2.2   Emotion Detection

Social media are frequently used to shared emotional feelings and opinions in form of a text message. Through the text message, the user show their emotion [19]. To detect the emotion automatically, the text message must be processed with several methods.

In [19] research, emotions are divided into four major classes, which are Happy-Active, Happy-Inactive, Unhappy-Active, Unhappy-Inactive [20]. The detection is developed using a supervised machine learning approach and including two tasks, an offline training task and an online classification task. The first task is called Emotex and the second task is called EmotexStream.

The dataset is obtained from Twitter tweet data. The collected data are selected when the comment would indicates an emotion. The preprocessing is to mitigate the common text which not indicate an emotion feature.

To convert the feature, the researcher use unigram, emoticon, and negation features. Unigram features are used to capture emotion in text. Emoticon features are used to portrait emotion by using an emoticon. For example, using ":)" to illustrated "happy" emotion.

Negation features are converted by using a single-phase by selecting the list of phases in LIWC dictionary. For example, "not happy" as the negation feature will be converted into "sad". Lastly,

by using Naïve Bayes as a probabilistic classifier, SVM as a decision boundary, and KNN classifier, the emotions are classified.

2.3 Cyberbullying Detection

Many researchers have researched on cyberbullying detection. For instance, using bullying features [1], create a distinction between cyberbullying and cyber aggression [21], investigate cyberbullying behavior [21], using a socio-linguistic model [22], using preprocessing of natural language processing [23] and even using pronunciation based CNN (Convolutional Neural Network) [24]. Summarizing the previous work, we attain that the best solution is using SVM as a classifier for the learning method [1] and [21].

A various dataset is taken from social media comment and post, such as Twitter and Instagram [1], [23], [24]. The TF-IDF is applied for feature extraction and used as input to classifiers. Bullying features will be pre-defined based on linguistic resources, which are named as insulting seeds [1].

After that, feed the learned feature into a linear SVM pattern classifier. The features are classified into bullying features and non-bullying features. Bullying features are separated into a threat, insult, defense, sexual talk, curse / exclusion, defamation, and encouragement of harasser [23].

For testing, the feature extraction is using text mining preprocessing and TF-IDF, which contains unigram to label every text which contains the bullying features. Then, we use SVM classifier to classify and label the data into non-bullying trace and bullying trace [1].

2.4 Ensemble Method

Several researchers have proven that the ensemble method could improve the performance in classification and regression problems [25], [26], [27], and [28]. This method is introduced by combining different learning models and create a new prediction. The most used ensemble methods are bagging, boosting, and stacking.

In the stacking method, the idea is to learn from several weak learners and combine them by training a meta-model. The output from base models is used as input for the meta-model. Unlike bagging, the models in the stacking method are typically different and fit on the same dataset.

3 Methodology

The proposed method in this research are defined in Figure 2. The stages are creating emotion detection model, cyberbullying detection model and ensemble the model.

The data are collected from Formspring and Twitter posts, which have been labeled as bully content or not. Afterward, implementation of the model will be done in 2 steps: emotion detection and bullying detection. The emotion model is divided into 4 classes: Happy-Active, Happy-Inactive, Unhappy-Active, and Unhappy-Inactive. When the victims show a sign of unhappiness, the probability of certain content is highly bully content.

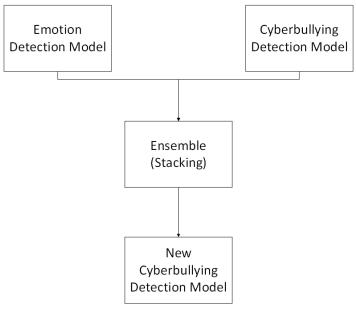In this research, we proposed the cyberbullying detection by using emotion detection through text. This method is performed by combining the model of cyberbullying detection and emotion detection. By ensemble the result of emotion detection and cyberbullying detection, we assume to have a better result. The fusing method is using stacking ensemble method.

Figure 2: Proposed Model

3.1 Emotion Detection

We perform three steps, which are preprocessing, feature selection, and classifier [29]. By using WordNet synsets, the emotions are expanded. Then from Twitter's tweet, the data are collected with the following criteria:

1. Contain one or more hashtags that defined emotion hashtag.

2. Ignore the retweet (which begins with "RT" keyword).

After collecting the data, the labeled tweets are processed by using the following rules:

1. User ID and URL are separated.

2. Text normalization.

3. Remove conflicting hashtag. For tweet which contains different emotion class will be removed.

4. The hashtag at the end of tweets is considered an emotion label.

5. Change negation word into the antonym word.

For capturing emotion, we performed three parts to convert the features, which are unigram, emoticon, and negation feature. In unigram feature selection, each word is classified into emotion categories in LIWC (Linguistic Inquiry and Word Count) lexical, which contains a dictionary of thousand words with specified emotion indicative.

The emoticon features are used to express the emotion, such as using :) as happy, :(as sad, :/ as angry, :0 as afraid/surprised ,

etc., obtained from [29] and combined with nltk sentiment documentation. The negation list are obtained from the nltk sentiment documentation.

Table 1: Emoticon List

Emotion	Emoticon
Happy	:) ;) =) :] :P :-P ;P :D ;D :> :3 :-) ;-) :^) :o) :~) :^) ;o) :') :-D :-> :-) :o) :] :3 :c) :> =] 8) :} :^) :-D 8-D 8D x-D xD X-D XD =-D =D =-3 =3 :-)) :'-) :* :^* >:P :-P :P X-P x-p xp XP :-p :p =p :-b :b >:) >;) >:-) <3
Sad	:(=(:-(:^(:o(:-(:'(:-< :L :-/ >:/ :S >:[:@ :-(:[=L :< :-[:-< =\\ =/ >:(:'-(:'(:\\ :-c :c :{ >:\\ ;(: −‖
Angry	>:S >:{ >: x-@ :@ :-@ :-/ :-\ :/
Afraid	:-o :-O o_O O_o :$
Sleepy	-_- ˜_˜

Table 2: Negation List

Negation List	never, no, nothing, nowhere, noone, none, not, havent, hasnt, hadnt, cant, couldnt, shouldnt, wont, wouldnt, dont, doesnt, didnt, isnt, arent, aint

There are possibilities that tweets have a word that conflicting with others' word. One word could be classified as happy emotion, while others are sad. For example, "Finally the last exam! Must study hard to get a better score #depressed #excited". This example shows that the post has exited feeling of upcoming exam and at the same time are depressed. To reduce these ambiguities, the negation features are separated and converted into the antonym word.

We use TF-IDF for the feature extraction [14]. Then, SVM and KNN Classifier are chosen as its level of accuracy are highest compared to Naïve Bayes and Decision Tree [19].

Table 3: Preprocessing for Emotion Detection

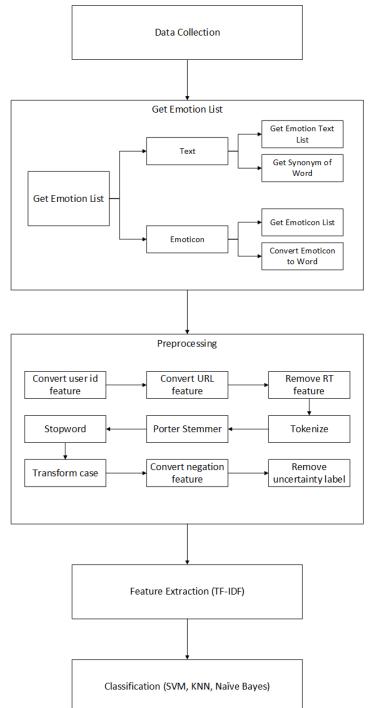| Feature	Process
'@' feature	A symbol that represented USERID
url links	Replaced into URL
Tweets contain more than one '#' and represented as two different classes	Remove the content to reduce ambiguities
Tweets contain more than one emoticon and represented as two different classes	Remove the content to reduce ambiguities
Tweets contain conflict between emoticon and '#'	Remove the content to reduce ambiguities
'#' in the end of content	Stripped the tags

Figure 3: Emotion Detection

3.2 Cyberbullying Detection

The data for cyberbullying are obtained from Kaggle.com, which provide Formspring data for cyberbullying detection. The data contains a question-and-answer field. Every content is reviewed by 3 annotators, with "yes" if the content is detected as a bully and "no" if not. Also, the annotator indicated the content severity with a 0-10 level.

The preprocessing is defined in 7 steps, which are data cleansing and data balancing, tokenization, transform case, stop word removal,

filter token, stemming and generate n-gram [12].

We use 2 and 4 classes for data cleansing and data balancing, which shown better accuracy [12]. The next step is tokenization, where the words are sliced into separated for each word (separated by a special character). Next, transform case will transform all the words into lower case.

Stop word removal will delete the unnecessary words based on nltk library. Filter token is used to remove the word that contains below 3 or more than 25 characters. Using Porter Stemmer, the words are converted into basic words. Lastly, we use n-gram to form a set of the word.

To produce an input that can be used as an input classifier, feature extraction is required. The preprocessing output is transformed into a vector space model and represented with a vector, with the calculation of weight. We use TD-IDF to extract the feature. The classifier of cyberbullying is using SVM method with the kernel poly. The classification is defined in 2 classes: non-bully and bully and 4 classes: class no, class low, class middle, and class high.
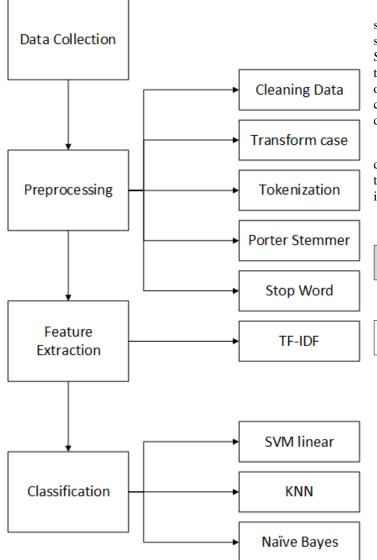
3.3 Stacking Ensemble Method

The idea of stacking is to learn several weak learners and combines them by training a meta-model to create a new prediction based on the new model (the combined model). The stacking architecture involves two or more base-models (level-1) and meta-model. Meta-models are trained using the predictions made by the base model (the output of the basic model is used as input in the model meta). The training dataset are prepared using k-fold cross-validation of the base-models.

The stacking method has three steps. The first step is creating the base model classifiers. Then, construct a new dataset for the meta-model. Lastly, construct the meta-model classifier using the output from the base model prediction.

In this research, we use the best model from cyberbullying and emotion detection. After both models are constructed, the models are combined using the stacking ensemble method.

This stacking method is done in three steps: learning basic classifiers, loading a new dataset and creating a meta-model. The first step is to create a base classifier, including emotion detection using SVM and cyberbullying prediction using SVM. In the emotion detection, the unhappy-active will be labeled as cyberbullying and the others will be labeled as non-bullies. Unhappy-active emotions are classified as bully content since these emotions are constituted by displeasure and aggressive form.

The next step is to load a new dataset for the meta-model. The data will be pre-processed like in cyberbullying detection. Lastly, the meta-model is constructed from the base classifier and resulting in a new prediction.
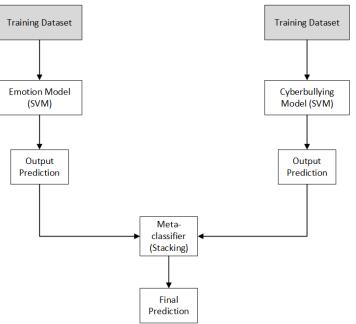

Figure 4: Cyberbullying Detection

Figure 5: Stacking Ensemble Method for Cyberbullying Detection

---

Algorithm 1: Stacking Ensemble Method

  Result: Meta-model Prediction
  Step 1: Learn base classifier ;
  Initialize models ;
  Build Emotion Classifier (SVM) ;
  Build Cyberbullying Classifier (SVM) ;
  Step 2: Load dataset ;
  Load and pre-processing dataset ;
  Step 3: Build meta-model ;
  Add all classifier as input for stacking classifier ;
  Return new prediction ;

---

4   Experiment and Evaluation Method

The experiment for cyberbullying and emotion detection is using SVM, KNN, and Naïve Bayes classifier. The dataset will be divided into 3 parts, training data, validation data, and testing data. 60% will be used for training data, 20% for validation, and 20% for testing. The validation is using K-Fold Cross-validation. This validation is used to create a less biased model. The K-Fold Cross Validation will randomly divide a dataset into k disjoint folds, then each fold will be used to test the model [30].

The dataset for emotion is obtained from Twitter data and the cyberbullying dataset is obtained from Formspring data. The Formspring data are labeled with severity level for each comment. The amount of data after cleansing are 10,890, with 10,504 non-bullying data, 146 labeled as low, 151 labeled as middle, and 89 labeled as high severity.

Figure 6: Total Data for Each Severity Level

To create an unbiased model, the processed data are separated with the following details:

1. 2 Class

   - Class No: 330 data with severity 0
   - Class Yes: 330 data with severity 1-10

2. 4 Class

   - Class No: 80 data with severity 0
   - Class Low: 80 data with severity 1-3
   - Class Middle: 80 data with severity 4-7
   - Class High: 80 data with severity 8-10

For the emotion detection, the data is obtained from Twitter tweets with 316,341 data after cleansing. The data are separated into 4 labels: happy-active 131,646 data, happy-inactive 34,036 data, unhappy-active 124,278, and unhappy-inactive 26,381 data. For creating a balanced model, we decide to use only 25,000 data for each class.

5   Result and Discussion

The result of the classification for cyberbullying is detailed in Table 4 and Table 5. The classification is using SVM linear, KNN with 3 neighbors, KNN with 5 neighbors, and Naïve Bayes. For the cyberbullying with four classes, the classification is using SVM linear. In the result, it could be seen that using two classes and SVM linear classification have the highest accuracy (72.73%) with an F1 score is about 72.7% and a recall score is 73.7%.

Table 4: Result for 2 Class Cyberbullying Detection

Classifier	Accuracy
SVM linear	**72.73%**
KNN 3 neigh	54.96%
KNN 5 neigh	50.94%
Naïve Bayes	66.66%

Table 5: Result for 4 Class Cyberbullying Detection

Classifier	Accuracy
SVM linear	52.5%

We use a confusion matrix to evaluate the accuracy of the model. Accuracy rates are calculated by using equation 2, where true positive (tp) and true negative (tn) is counted as true prediction, false positive (fp) and false negative (fn) is counted as a false prediction.

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{2}$$

For the emotion detection, the result is detailed in Table 6. The experiment is using two methods of classification, which are SVM linear and KNN (with 3 and 5 neighbors). The result shown that using SVM linear have the most accuracy (96.6 %) with an F1 score is about 94.81% and a recall score is 94.89%.

Table 6: Result for 4 Class Emotion Detection

Classifier	Accuracy
SVM linear	**96.6%**
KNN 3 neigh	62.38%
KNN 5 neigh	71.46%

Based on the results obtained, it appears that the best classification for cyberbullying and emotion detection is using SVM

classification. From several studies, it is also shown that SVM is better when used for text classification [1] and [12] . Besides, there are not many training data available, so when using the KNN, the results obtained are not optimal.

After constructing both models from emotion and cyberbullying, we create an ensemble model by using the stacking method. The result and comparison for the ensemble model are detailed in Table 7.

Table 7: Result and Comparison for Ensemble Model

Model	Accuracy
Emotion	96.6%
Cyberbullying	72.73%
Ensemble (Emotion + Cyberbullying)	**77.8%**

The results showed that by using the ensemble method for cyberbully and emotion models, the accuracy of cyberbullying detection could increase by about 5%. Detection on cyberbullying increased when combined with emotion detection using a method of stacking. This method is done by making each base model and making predictions, then the prediction results will be used as a feature for the meta-model. From this meta-model, new predictions will be generated.

By using this ensemble method, prediction errors from cyberbullying can be covered by predictions from emotion detection, thereby increasing the accuracy of the predictions. Accuracy can be increased due to the features of emotion. For words that have unhappy-active elements, such as angry, annoyed, frustrated will be the feature that determines whether the tweet is a bully content. This is because in most cases the bully's words constituted by displeasure and shown aggressive.

Also, the result shows that emotion detection has the highest accuracy when compared to cyberbullying detection model and an ensemble model. However, cyberbullying detection may not simply use the emotion model, as there are tweets that have no emotion feature but could be labeled as a bully tweet. Hence, the best solution is to use the stacking ensemble model from emotion and cyberbullying model to predict the content.

6 Conclusion

By including emotion feature in cyberbullying detection and using ensemble method, we ought to have a better accuracy. For cyberbullying, the most optimal classification is by using 2 class (bully and non-bully) and use SVM classifier with linear kernel, with an average accuracy is 72,73%. For emotion detection, the most optimal classification is by using 4 class (happy-active, happy-inactive, unhappy-active, unhappy-inactive) and use SVM classifier with linear kernel, with an average accuracy is 96,6%. Then, by ensemble both model using stacking method, the accuracy for cyberbullying improve to 77,8%.

This research has proven that including emotion in cyberbullying detection could improve the accuracy of the detection. This is because emotions that are indicated as unhappy-active generally show words that are less wearing and emit emotions such as anger, displeasure, distress, and other negative emotions. By including

these emotional features, accuracy, and precision of detection of cyberbullying increases.

For future works, the challenge is to try to make a model of cyberbullying detection with more data or in a different language. It is also challenging to consider sarcasm in tweet content for the possibility of bullying content.

References

[1] R. Zhao, A. Zhou, K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features," ACM International Conference Proceeding Series, **04-07-Jan**(January), 2016, doi:10.1145/2833312.2849567.

[2] A. Sinring, F. Aryani, N. H. AR, "Development of Guidance and Counseling Videos to Reduce Cyber Bullying among Vocational Students," **227**(Icamr 2018), 324–328, 2019, doi:10.2991/icamr-18.2019.80.

[3] V. Orgeta, "Specificity of age differences in emotion regulation," Aging and Mental Health, **13**(6), 818–826, 2009, doi:10.1080/13607860902989661.

[4] A. Weinstein, D. Dorani, R. Elhadif, Y. Bukovza, A. Yarmulnik, P. Dannon, "Internet addiction is associated with social anxiety in young adults," Annals of Clinical Psychiatry, **27**(1), 4–9, 2015.

[5] T. Safaria, "Prevalence and impact of cyberbullying in a sample of indonesian junior high school students," Turkish Online Journal of Educational Technology, **15**(1), 82–91, 2016.

[6] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, S. Mishra, "Analyzing labeled cyberbullying incidents on the instagram social network," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **9471**, 49–66, 2015, doi: 10.1007/978-3-319-27433-1_4.

[7] V. Balakrishnan, S. Khan, T. Fernandez, H. R. Arabnia, "Cyberbullying detection on twitter using Big Five and Dark Triad features," Personality and Individual Differences, **141**(September 2018), 252–257, 2019, doi: 10.1016/j.paid.2019.01.024.

[8] V. Balakrishnan, S. Kan, H. R. Arabnia, "Improving Cyberbullying Detection using Twitter Users' Psychological Features and Machine Learning," 2019.

[9] S. A. Salloum, M. Al-emran, A. A. Monem, "Using Text Mining Techniques for Extracting Information from Research Using Text Mining Techniques for Extracting Information from Research Articles," (November), 2018, doi: 10.1007/978-3-319-67056-0.

[10] S. Vijayarani, R. Janani, "TEXT MINING: OPEN SOURCE TOKENIZATION TOOLS – AN ANALYSIS," Advanced Computational Intelligence: An International Journal (ACII), **3**(1), 37–47, 2016, doi:10.5121/acii.2016.3104.

[11] A. Muneer, S. M. Fati, "A comparative analysis of machine learning techniques for cyberbullying detection on twitter," Future Internet, **12**(11), 1–21, 2020, doi:10.3390/fi12110187.

[12] Noviantho, S. M. Isa, L. Ashianti, "Cyberbullying classification using text mining," Proceedings - 2017 1st International Conference on Informatics and Computational Sciences, ICICoS 2017, **2018-January**, 241–245, 2018, doi: 10.1109/ICICOS.2017.8276369.

[13] A. Rajaraman, J. D. Ullman, "Data Mining," Mining of Massive Datasets, 1–17, 2011, doi:doi:10.1017/CBO9781139058452.002.

[14] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," Proceedings of the first instructional conference on machine learning, **242**, 133–142, 2003.

[15] M. Ikonomakis, S. Kotsiantis, V. Tampakas, "Text classification using machine learning techniques," WSEAS Transactions on Computers, **4**(8), 966–974, 2005, doi:10.11499/sicejl1962.38.456.

[16] H. Rosa, N. Pereira, R. Ribeiro, P. C. Ferreira, J. P. Carvalho, S. Oliveira, L. Coheur, P. Paulino, A. M. Veiga Simão, I. Trancoso, "Automatic cyberbullying detection: A systematic review," Computers in Human Behavior, **93**(October 2018), 333–345, 2019, doi:10.1016/j.chb.2018.12.021.

[17] T. Evgeniou, M. Pontil, "Workshop on Support Vector Machines : Theory and Applications," Machine Learning and Its Applications: Advanced Lectures, (January 2001), 249–257, 2001, doi:10.1007/3-540-44673-7.

[18] M. Awad, R. Khanna, "Efficient learning machines: Theories, concepts, and applications for engineers and system designers," Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, (July 2018), 1–248, 2015, doi:10.1007/978-1-4302-5990-9.

[19] M. Hasan, E. Rundensteiner, E. Agu, "Automatic emotion detection in text streams by analyzing Twitter data," International Journal of Data Science and Analytics, 2018, doi:10.1007/s41060-018-0096-z.

[20] J. A. Russell, "A Circumplex Model of Affect," Journal of Personality and Social Psychology, **39**(6), 1161–1178, 1980, doi:http://dx.doi.org/10.1037/h0077714.

[21] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, S. Mishra, "Detection of Cyberbullying Incidents on the Instagram Social Network," 2014.

[22] S. Tomkins, L. Getoor, Y. Chen, Y. Zhang, "A socio-linguistic model for cyberbullying detection," Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018, 53–60, 2018, doi:10.1109/ASONAM.2018.8508294.

[23] C. Van Hee, G. Jacobs, C. Emmery, B. DeSmet, E. Lefever, B. Verhoeven, G. De Pauw, W. Daelemans, V. Hoste, "Automatic detection of cyberbullying in social media text," PLoS ONE, **13**(10), 1–22, 2018, doi:10.1371/journal.pone.0203794.

[24] X. Zhang, J. Tong, N. Vishwamitra, E. Whittaker, J. P. Mazer, R. Kowalski, H. Hu, F. Luo, J. Macbeth, E. Dillon, "Cyberbullying Detection with a Pronunciation Based Convolutional Neural Network," 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), 740–745, 2017, doi:10.1109/icmla.2016.0132.

[25] T. G. Dietterich, "Ensemble Method in Machine Learning," Springer, 1–15, 2000.

[26] A. Onan, S. Korukoğlu, H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," Expert Systems with Applications, **57**, 232–247, 2016, doi:10.1016/j.eswa.2016.03.045.

[27] I. Perikos, I. Hatzilygeroudis, "Recognizing emotions in text using ensemble of classifiers," Engineering Applications of Artificial Intelligence, **51**, 191–201, 2016, doi:10.1016/j.engappai.2016.01.012.

[28] Yan-Shi Dong, Ke-Song Han, "A comparison of several ensemble methods for text categorization," 419–422, 2004, doi:10.1109/scc.2004.1358033.

[29] M. Hasan, E. Rundensteiner, E. Agu, "EMOTEX: Detecting Emotions in Twitter Messages," SocialCom Conference, 27–31, 2014.

[30] T. T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," Pattern Recognition, **48**(9), 2839–2846, 2015, doi:10.1016/j.patcog.2015.03.009.