2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS), Yogyakarta, Indonesia

Detecting Online Child Grooming Conversation

Fergyanto E. Gunawan[1], Livia Ashianti[2], Sevenpri Candra[3], Benfano Soewito[4]
[1,2,4]Binus Graduate Programs, Bina Nusantara University, Jakarta, Indonesia 11480
[3]School of Business & Management, Bina Nusantara University, Jakarta, Indonesia 11480
Emails: [1]fgunawan@binus.edu, [2]lashianti@binus.edu, [3]scandra@binus.edu, [4]bsoewito@binus.edu

***Abstract*—Massive proliferation of social media has opened possibilities for perpetrator to conduct the crime of online child grooming. Because the pervasiveness of the problem scale, it may only be tamed effectively and efficiently by using an automatic grooming conversation detection system. Previously, Pranoto, Gunawan, and Soewito [1] had developed a logistic model for the purpose and the model was able to achieve 95% detection accuracy. The current study intends to address the issue by using Support Vector Machine and k-nearest neighbors classifiers. In addition, the study also proposes a low-computational cost classification method on the basis of the number of the existing grooming conversation characteristics. All proposed methods are evaluated using 150 conversation texts of which 105 texts are grooming and 45 texts are non-grooming. We identify that grooming conversations possess 17 features of grooming characteristics. The results suggest that the SVM and k-NN are able to identify grooming conversations at 98.6% and 97.8% of the level of accuracy. Meanwhile, the proposed simple method has 96.8% accuracy. The empirical study also suggests that two among the seventeen characteristics are insignificant for the classification.**

Keywords*—*Online Child Grooming, Support Vector Machine, k-nearest Neighbors, Grooming Classifier

I. Introduction

Online child grooming is defined as a process to approach, persuade, and engage a child, the victim, in sexual activity by using Internet as a medium. Perpetrator approach the victim to build not only sexual but also emotional relationship [2]. Massive profileration of social media has opened possibilities for perpetrators to conduct the crime of online child grooming in a larger scale [3], [4]. According to the Child Exploitation and Online Protection Agency, online child grooming is the most reported crime in UK in 2009–2010 [3]. It affects the victim life psychologically, physically, emotionally, behaviorally, and psycho-socially [5].

To reveal this type of crimes, investigator usually relies on the conversation texts where the grooming patterns are carefully analyzed [6]. With the vast amount of conversation text data, the process becomes extremely difficult and requires significant amount of time. The manual approach of investigating grooming pattern is also error prone [6]; besides, the grooming process usually takes about a month on the average [7].

For the reason described above, it is important to develop an automatic system to analyze a conversation text and to detect the possibility of the online child grooming conversation. During the last five years, a number of research works has been addressing the issue using various pattern detection schemes including using k-means clustering by Kontostathis, Edwards, and Leatherman [6], a ruled-based approach by McGhee et al. [8], Support Vector Machine (SVM) by Pandey, Khapaftis, and Manandhar [9]. Recently, Pranoto, Gunawan, and Soewito [1] developed a grooming detection system utilizing a logistic regression model. SVM method seems to work best for the text-based classification according to Ref. [10]. However, SVM has been also demonstrated for the image-based classification such as detection of corona artery disease [11] and breast cancer [12]. Reference [13] used SVM for developing an intrusion detection system.

This study intends to propose a simple method to detect an online child grooming conversation. In doing so, the study firstly identifies the main characteristics of the type of conversations. The proposed method is developed on the basis of the number of existing characteristics.

II. Supporting Theories

A. Characteristics of Online Child Grooming

Online child grooming conversation texts are complex as it varies in duration, type, and intensity depending on the perpetrator characteristics and behavior. However, in general, O'Connel [14] and Gupta [15] have identified the typical stages in an online child grooming process.

The first is the friendship forming stage. The perpetrator tries to get introduced to the child and then to establish a possibility of exchanging name, location, age, and etc. Furthermore, the perpetrator inquires other online informations related to the child, requesting photos in order to confirm that the child is indeed a child.

The second is the relationship forming stage. The perpetrator and the child talk about family, school, interest, and hobbies of the child so that he can exploit them by deceptively making the child believes that they are in a relationship.

The third is the risk assessment stage. The perpetrator tries to gage the level of threat and danger by talking to the child. He ensures that the child is alone and nobody else is reading their conversations.

The fourth is the exclusivity stage. The perpetrator tries to gain the complete trust of the child. Often, the concept of love and care are introduced by the perpetrator in this stage.

The fifth is the sexual stage. The perpetrator and the child talk about sexual activities and developing sex fantasy. Finally, the sixth is the conclusion stage. In this stage, the perpetrator approaches the child for meeting in person.

These stages of online child grooming may or may not occur in a sequence. The frequency, order, and extent of the occurrence of these stages may vary from chat to chat.

On the basis of the previous work [1], and Refs. [3], [15], we have identified 17 grooming characteristics, see Table I, and their relation to the grooming stages are presented in Table II. These characteristics would be used to classify the online conversation texts.

TABLE I. THE IDENTIFIED GROOMING CHARACTERISTICS.

No	Characteristics, Description, and Source.
1	Asking profile. Perpetrator and victim exchange information about personal info, such as, name, age, and location [15].
2	Other way contact. Perpetrator and victim talk about another way to communicate, such as, phone, email, and social media [15].
3	Asking picture. Perpetrator asks victim to send a his/her picture or vice versa [15].
4	Giving compliment. Perpetrator compliments the victim in order to make the victim happy and flattered [15].
5	Talking about activity, favourite hobby, and school. Perpetrator and victim talk about daily activities, favourite hobbies and victim's school activities [15].
6	Talking about friend and relationship. Perpetrator and victim talk about friendships or relationships, such as, asking about relationship with another person [15]. If the victim is not in a relationship with another person, it's easier for perpetrator to get closer.
7	Asking questions to know the risk of conversation. Perpetrator tries figure out the risk of their conversation, whether their conversation is known by victim's parents [1]. Usually, perpetrator will ask about anyone who uses victim's computer, location of the computer, and whether victim's parents know the password of the chat application.
8	Acknowledging wrong-doing. Perpetrator will inform to potential victim what they are doing is wrong, and have legal risks for perpetrator [1]. By telling this to victim, perpetrator has a purpose, which is perpetrator will be free from legal cases that will make him/her jailed in the future.
9	Asking if the child is alone or under adult or friend supervision. Perpetrator wants to make sure the victim whether is alone or under supervision [3].
10	Trying to build mutual trust. Perpetrator trying to build the mutual trust from victim, the next level relationships will be easier for perpetrator if perpetrator gain the trust from the victim [1], [15].
11	Using falling in love words. In conversation between perpetrator and the victim, they use words to express they are in love [3], [15].
12	Using word to express feeling. In a conversation between the perpetrator and victim, they use words to express their feelings [1].
13	Using word about biology, body, intimate parts, and sexual category. In a conversation between the perpetrator and the victim, they use words that contain sexual context [1].
14	Asking hot picture. Perpetrator asks victim for sexual theme photos or vice versa [1], [15]. These pictures can be used as fantasy or a tool to threaten victim to obey the perpetrator.
15	Introducing sexual stage. Conversation started with talking about sexual context, such as ask about sex experiences [1], [15].
16	Sexual stage. Conversation has entered the stage of sexual fantasies with words that show the interaction of activities and involve intimacy [15].
17	Arranging further contact and meetings. Perpetrator tries to get the victim address in order to have a meeting at the victim's house or to invite victim to meet somewhere [1], [15].

TABLE II. THE RELATION BETWEEN THE 17 GROOMING CHARACTERISTICS AND GROOMING STAGES.

No	Grooming Stage	Charateristic Name
1	Friendship forming	Asking profile
2		Other way contact
3		Asking picture
4		Giving compliment
5	Relationship forming	Talking about activity, favourite, hobby, school
6		Talking about friend and relationship
7	Risk assessment	Asking questions to know risk of conversation
8		Acknowledging wrong-doing
9		Asking if the child is alone or under adult or friend supervision
10	Exclusivity	Trying to build mutual trust
11		Using falling in love words
12		Using word to express feeling
13	Sexual	Using word about biology, body, intimate parts, and sexual category
14		Asking hot picture
15		Introducing sexual stage
16		Sexual stage
17	Conclusion	Arranging further contact and meetings

B. Support Vector Machine

In the present study, we only use the Support Vector Machine (SVM) for linearly separable data. The SVM is a numerical method to compute an hyperplane for separating a two-class dataset. It can easily be extended to multiple-class problem. The SVM establishes the hyperplane, governed by $(\mathbf{w}, b)$, by using the support vectors, which are the data points that are closest to the hyperplane. The following SVM formulation is derived from Refs. [16], [17]; readers are advised to the two sources for detail exposition.

We consider the point sets $\mathbf{x}_i \in \Re^d$, as the support vectors, with the categories $y_i \in [-1, +1]$. The hyperplane that separates $y_i = -1$ from those of $y_i = +1$ should satisfy

$$<\mathbf{w}, \mathbf{x}> +b = 0, \tag{1}$$

where $\mathbf{w} \in \Re^d$, $<\mathbf{w}, \mathbf{x}>$ denotes the inner dot product of $\mathbf{w}$ and $\mathbf{x}$, and b is a scalar constant. The hyperplane is obtained by solving:

$$\min_{\mathbf{w},b} L_p = \frac{1}{2} <\mathbf{w}, \mathbf{w}> - \sum_i \alpha_i \left[y_i \left(<\mathbf{w}, \mathbf{x}_i> +b \right) - 1 \right], \tag{2}$$

where $\alpha_i \geq 0$. For the case where the data are linearly not separable, the feature vector $\mathbf{x}_i$ would be transformed with a kernel function. Two types of the kernel functions would be evaluated: polynomial type where $K(\mathbf{x}, \mathbf{y}) = (1+ <\mathbf{x}, \mathbf{y}>)^d$ and Radial basis function (RBF) type where $K(\mathbf{x}, \mathbf{y}) = \exp(- < (\mathbf{x} - \mathbf{y}), (\mathbf{x} - \mathbf{y}) > /(2\sigma^2))$. The parameter d is an integer, and would be evaluated for $d = 1, 2, \text{and}, 3$, and σ has a positive value.

C. k-Nearest Neighbor

The k-Nearest Neighbor (k-NN) is an instant-based learning algorithm to classify data based on data from training dataset which is most similar to the data. In k-NN method, the method will retrieve k data from the most similar data from training dataset with the data. Similarity between data from training dataset and data are measure by calculating the distance. Before calculating distance, data and data from training dataset are represented into VSM. Usually Euclidean Distance is typically used in computing the distance between the vectors [18].

The training phase consists only of storing the feature vectors and class label of the training set. In the classification phase, using the testing data transformed into vector to calculate the distances with vectors from training dataset and k closest distance are selected. The annotated category of testing data are predicted based on the nearest point which has been assigned to a particular category, and then, assigned testing data to the class which contains most of the neighbors.

D. Accuracy Indicator

The classification accuracy is computed by:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{3}$$

where TP stands for True Positive, TN for True Negative, FP for False Positive, and FN for False Negative.

III. Research Methods

The research procedure is schematically shown in Fig. 1 and a few important steps are briefly explained in the following.

A. Dataset Preparation

Two types of conversation texts are required for this research: the first type is the conversations of actual online child grooming and the second type is the conversations of non-grooming conversations but has grooming characteristics.

The first type conversations are randomly selected from www.perverted-justice.com, a website that contains more than 500 texts of grooming conversations involving perpetrators and children, juvenile victims, or undercover law enforcements. Only 105 texts are selected. The source has also been used by the previous researchers [1], [6], [8], [9], [19].

The second type conversations are selected from www.literotika.com. The web contains conversation scripts of people expressing their sexual passion legally. Fourty five non-grooming conversation texts are randomly selected from the site.

B. Preprocessing

The text of online conversation contains many noises from the perspective of document classification. Those noises should be minimized or eliminated, if possible, prior the analysis to determine the grooming characteristics. The all texts in this research are subjected to the following processes. Tokenization: non-letter characters in the document would be removed and each document is partitioned into words Transformation: words in the document would be transformed into lowercase. Stopword elimination: words which frequently exists across document but not significantly useful would be erased. Stemming: words in the document would be reduced into their root using porter algorithm. Generating 3-gram: words in the document would be formed into 3-grams of 1 continuous sequence formed of 3 words from the document.

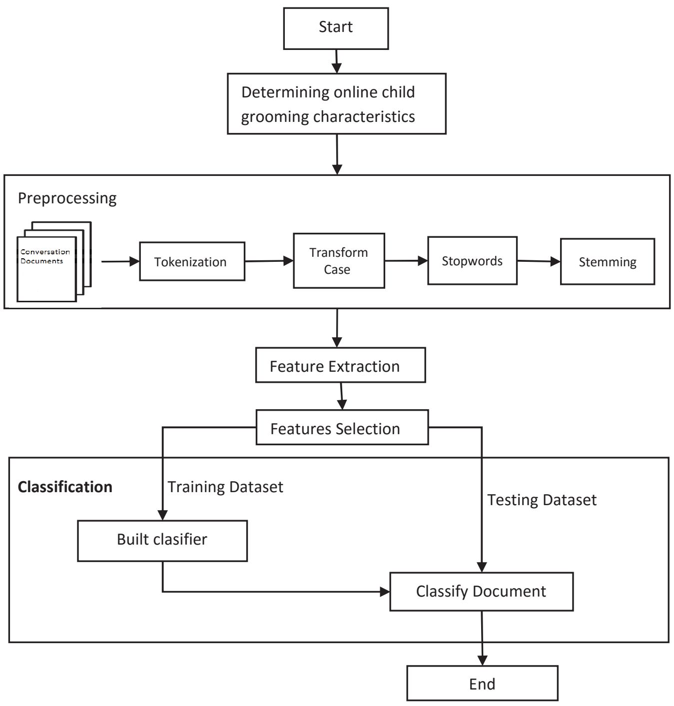

Fig. 1. The research procedure.

C. Feature Extraction

Texts that have been preprocessed would be transformed into a vector space model (VSM). The features are words or combinations of words that form the wordlist. The word list is denoted with $T_1, T_2, \ldots, T_t$. The feature extraction result from a document (d_m) is transformed into vector $d_m = \{w_{T_1,m}, w_{T_2,m}, \ldots, w_{T_t,m}\}$ where $m \in M$, M is number of documents and $w_{T_i,m}$ is weight calculation results from feature i using TF-IDF that represents how important features i in document m and all documents in the dataset.

D. Features Selection

Feature extraction results from each document in VSM would be used to create a grooming characteristic vector. Grooming characteristics used are 17 characteristics that have been determined in Table I. the vector is denoted $C_m = \{c_{m,1}, c_{m,2}, c_{m,3}, \ldots, c_{m,17}\}$ where $m \in M$ and $c_{m,j}$ is a value that indicates whether or not the characteristic j in the document m. If document m does not contain characteristic j then $c_{m,j} = 0$. If document m contain characteristic j then $c_{m,j} = 1$. To determine grooming characteristic j value in document m $c_{m,j}$, features from extraction will be selected in accordance with database which stores words or combinations of words that describe each grooming characteristic. The value of features that have been selected will be summed. If the result is 0 then $c_{m,j} = 0$ and if the result is greater than 0 then the characteristic value j in the document $c_{m,j} = 1$.

E. Classification

The classification would be performed using SVM (see Subsection II-B), k-NN (see Subsection II-C) and our proposed method, which is based on the number of grooming characteristics in the document.

IV. Results and Discussion

We have analyzed 150 conversation texts consisting of 105 grooming conversations, randomly taken from www.perverted-justic.com, and 45 non-grooming conversations, randomly taken from www.literotika.com. We have identified seventeen grooming characteristics by learning those grooming conversation and by considering previous works. Those characteristics are then represented in a vector space. These characteristics and their frequencies of occurrence in grooming and non-grooming conversations are presented in Table III.

What makes automatic classification difficult is that the grooming characteristics also appear on non-grooming conversations as shown by Table III. For example, the most prevalent characteristics, which is the 13th characteristics, “using word about biology, body, intimate parts, and sexual category” appears in 105 grooming text conversations and in 43 non-grooming text conversations.

Another insight shown by the table is that two characteristics, namely, the 14th characteristics, "asking hot picture", and the 7th characteristics, "asking question to know risk of conversation," apper rarely. We hypothesize: the two characteristics may not have significant contribution to the performance of the document automatic classification. This will be empirically evaluated.

In the following, we are going to discuss the results in term of the classification accuracy for various classification methods and with or without the 7th or 14th grooming characteristics. The training set consists of 70 grooming and 30 non-grooming conversations. The testing set consists of 35 grooming and 15 non-grooming conversations.

For the first research results, we compare the level of accuracy of the results by several SVM kernel functions, namely, RBF, quadratic, polynomial, and linear functions. The results, on the average accuracy, are shown in Table IV.

These results reveals some interesting phenomena, some are expected, some are unexpected. We expect that the highest level of accuracy would be achieved by using all grooming characteristics. This expectation is materialized for the three types of SVM kernels: polynomial, quadratic, and linear. Using the RBF kernel, the results are rather unexpected: the accuracies without the 7th and 14th characteristics are better than using the all characteristics. The expectation that the 7th and 14th grooming characteristics would only slightly affect the level of accuracy is only materialized for the three kernel: polynomial, quadratic, and linear. Using all grooming characteristics, the level of accuracy by means of SVM method is within the range of 83–98% depending on the selection of the kernel function. In comparison to the method utilizing the logistic regression model, see Ref. [1], the three kernels provide slightly better accuracies. The RBF kernel produces a lower accuracy than the logistic model.

For the second research results, we also compare the level of accuracy by using a different classifier, that is the k-NN method with the k values of 1, 3, and 5. The results, in average, are depicted in Table V.

These results completely agree with our expectation. The highest average level of accuracy is achieved by using all grooming characteristics. This result is materialized for all values of k. The expectation that the 7th and 14th grooming characteristics will only slightly affect the level of accuracy is materialized for all of k values. The average level of accuracy in classification without the 14th characteristics is the same with using all grooming characteristics. Using all grooming characteristics, the average level of accuracy by means of the k-NN method is within the range of 96.8–97.8% depending on the k value. However, it is not clear whether increasing the k value will increase or decrease the level of accuracy.

Finally, we propose a simple classification method, which requires very low computational cost and makes it suitable for implementation in the electronic mobile devices. The proposed method is to classify the conversation on the basis of the existing number of grooming characteristics. This method is proposed by observing the fact that the number of grooming characteristics are markedly different; see Table VI.

The table suggests that a conversation tends to be a grooming conversation if it contains the number of grooming characteristics within the range 8–17. Meanwhile, a conversa-

TABLE III. SEVENTEEN GROOMING CHARACTERISTICS AND THEIR FREQUENCIES OF OCCURRENCE ON GROOMING AND NON GROOMING CONVERSATION TEXTS.

No	Grooming Characteristics	Frequency	
		G	N
1	Asking profile	97	2
2	Other way contact	101	17
3	Asking Picture	102	10
4	Talk About friend and relationship	96	22
5	Giving Compliment	104	28
6	Talk About Activity, Favourite, Hobby, school	95	16
7	Asking Question To Know Risk Of Conversation	44	0
8	Acknowledging wrong doing	99	15
9	Asking if child is alone or adult supervision or friend	84	0
10	Trying building mutual trust	98	27
11	Using word in fallin in love	70	7
12	Using word in feel category	105	42
13	Using word in biology, body, intimate parts, and sexual category	105	43
14	Asking hot picture	13	0
15	Introduced sexual stage	101	34
16	Sexual Stage	97	44
17	Arrange further contact and meeting	100	5

G = Grooming, N = Non-grooming

TABLE IV. THE LEVEL OF CLASSIFICATION AVERAGE ACCURACY USING THE SVM METHOD WITH FOUR KERNEL FUNCTIONS AND WITH OR WITHOUT THE 7TH OR 14TH GROOMING CHARACTERISTICS.

Grooming Characteristics	The Type of SVM Kernel			
	RBF	Polynomial	Quadratic	Linear
All seventeen	83.8	97.6	98.6	98.6
Without the 14th	87.4	97.6	98.6	98.6
Without the 7th	89.4	96.6	97.4	97.8

TABLE V. THE LEVEL OF CLASSIFICATION AVERAGE ACCURACY USING THE k-NN METHOD WITH k VALUES OF 1, 3, AND 5, AND WITH OR WITHOUT THE 7TH OR 14TH GROOMING CHARACTERISTICS.

Grooming Characteristics	k Value		
	1	3	5
All characteristics	97.0	97.8	97.2
Without the 14th	97.0	97.8	97.2
Without the 7th	96.8	97.2	96.8

TABLE VI. THE NUMBER AND TYPE (GROOMING OR NON-GROOMING) OF DOCUMENTS AND THE NUMBER OF GROOMING CHARACTERISTICS CONTAINED IN THE DOCUMENT.

Number of the Grooming Characteristics	Number of Documents	
	G	N
1	0	0
2	0	1
3	0	1
4	0	4
5	0	5
6	0	10
7	0	4
8	1	8
9	1	7
10	2	4
11	5	1
12	8	0
13	10	0
14	19	0
15	23	0
16	30	0
17	6	0

G = Grooming, N = Non-grooming

tion tends to be a non-grooming conversation if it contains about 2–11 grooming characteristics. Thus, the number of grooming characteristics can simply be used as a classifier; despite the fact, there is an overlap in the number of grooming characteristics between the two categories. To evaluate a text conversation, we can set certain threshold, evaluate the number of grooming characteristics, and decide that the conversation is grooming if its number of grooming characteristics is equals or exceeds the threshold.

We empirically evaluate the method by varying the threshold value from 1 to 17. If the number of grooming characteristics in a document is less than the threshold, the conversation will be classified as a non-grooming conversation and vice versa. The results in the average accuracy are depicted in Table VII. These empirical data suggest that the highest average level of accuracy is achieved at the threshold value of 11. The best threshold provides an accuracy level of 96.8%. The expectation that the 7th and 4th grooming characteristics would only slightly affect the level of accuracy is materialized for all of the threshold values.

Finally, we compare the level of accuracy of the three classification methods: SVM, k-NN, and our proposal. For the SVM method, we only include the results of using the linear kernel as they are the best among the method. For the same reason, for the k-NN method, we include only the case of $k = 3$. The comparison is presented in Table VIII.

The three methods support the hypothesis that the accuracy would slightly drop when the 7th and 14th grooming characteristics are excluded. In addition, these results suggest that the SVM classifier is able to classify the best in term of the accuracy. The proposed method, despite of its simplicity, also performs rather well.

TABLE VII. THE AVERAGE ACCURACY OF THE CLASSIFICATION OF THE PROPOSED METHOD AS A FUNCTION OF THE THRESHOLD VALUES.

Threshold	The Level of Accuracy (%)		
	All	Without the 14th	Without the 7th
1	70.0	70.0	70.0
2	70.0	70.0	70.0
3	70.2	70.2	70.2
4	71.6	71.6	71.6
5	74.2	74.2	74.2
6	77.4	77.4	77.4
7	82.8	82.8	82.8
8	85.8	85.8	85.8
9	90.8	90.8	90.8
10	95.8	95.8	95.8
11	96.8	96.8	96.0
12	94.0	94.0	94.0
13	88.2	88.2	86.8
14	80.6	80.6	77.0
15	68.6	67.8	67.6
16	53.2	50.0	37.4
17	35.0	30.0	30.0

TABLE VIII. THE COMPARISON OF THE AVERAGE LEVEL OF THE CLASSIFICATION ACCURACY USING SVM, k-NN, AND CURRENT PROPOSED METHODS.

Grooming Characteristics	Classification Method		
	SVM	k-NN	Proposed Method
All	98.60	97.80	96.80
Without 14th	98.60	97.80	96.80
Without 7th	97.80	97.20	96.00

V. CONCLUSION

Automatic system to detect online child grooming has an important role in analyzing vast amount of conversation texts. For the reason, many studies have been performed using various pattern detection schemes. In the current work, seventeen characteristics of grooming conversation are identified and utilized for classification. Two traditional classification methods are used: SVM and k-NN. Moreover, this work proposes a simple classification method on the basis of the number of existing grooming characteristics in the conversation. The numerical analysis using empirical data suggests that the SVM method with the linear kernel is the best method among others with the average level of accuracy 98.6%. Our proposed method, despite of its simplicity, also performs well with average level of accuracy 96.8%. The empirical study also suggests that two among the seventeen characteristics are insignificant for the classification accuracy.

REFERENCES

[1] H. Pranoto, F. E. Gunawan, and B. Soewito, "Logistic models for classifying online grooming conversation," *Procedia Computer Science*, vol. 59, pp. 357–365, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050915020657

[2] L. N. Olson, J. L. Daggs, B. L. Ellevold, and T. K. Rogers, "Entrapping the innocent: Toward a theory of child sexual predators luring communication," *Communication Theory*, vol. 17, no. 3, pp. 231–251, 2007.

[3] H. Whittle, C. Hamilton-Giachritsis, A. Beech, and G. Collings, "A review of online grooming: Characteristics and concerns," *Aggression and violent behavior*, vol. 18, no. 1, pp. 62–70, 2013.

[4] B. Soewito and S. M. Isa, "Digital technology: the effect on connected world to computer ethic and family," *CommIT (Communication and Information Technology) Journal*, vol. 9, no. 1, pp. 23–28, 2015. [Online]. Available: http://journal.binus.ac.id/index.php/commit/article/view/880

[5] D. Michalopoulos and I. Mavridis, "Utilizing document classification for grooming attack recognition," in *Computers and Communications (ISCC), 2011 IEEE Symposium on*. IEEE, 2011, pp. 864–869.

[6] A. Kontostathis, L. Edwards, and A. Leatherman, *Text mining and cybercrime*, ser. Text Mining: Applications and Theory. Chichester, UK: John Wiley & Sons Ltd, 2010.

[7] P. Briggs, W. T. Simon, and S. Simonsen, "An exploratory study of internet-initiated sexual offenses and the chat room sex offender: Has the internet enabled a new typology of sex offender?" *Sexual Abuse: A Journal of Research and Treatment*, p. 1079063210384275, 2010.

[8] I. McGhee, J. Bayzick, A. Kontostathis, L. Edwards, A. McBride, and E. Jakubowski, "Learning to identify internet sexual predation," *International Journal of Electronic Commerce*, vol. 15, no. 3, pp. 103–122, 2011.

[9] S. J. Pandey, I. Klapaftis, and S. Manandhar, "Detecting predatory behaviour from online textual chats," in *Multimedia Communications, Services and Security*. Springer, 2012, pp. 270–281.

[10] B. Baharudin, L. H. Lee, and K. Khan, "A review of machine learning algorithms for text-documents classification," *Journal of advances in information technology*, vol. 1, no. 1, pp. 4–20, 2010.

[11] İ. Babaoglu, O. Findik, and E. Ülker, "A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3177–3183, 2010.

[12] U. R. Acharya, E. Ng, J.-H. Tan, and S. V. Sree, "Thermography based breast cancer detection using texture features and support vector machine," *Journal of medical systems*, vol. 36, no. 3, pp. 1503–1510, 2012.

[13] S.-J. Horng, M.-Y. Su, Y.-H. Chen, T.-W. Kao, R.-J. Chen, J.-L. Lai, and C. D. Perkasa, "A novel intrusion detection system based on hierarchical clustering and support vector machines," *Expert systems with Applications*, vol. 38, no. 1, pp. 306–313, 2011.

[14] R. A. O'Connel, "Typology of cybersex exploitation and online grooming process," Cyberspace Research Unit, University of Central Lancashire, the United Kingdom, Tech. Rep., 2014. [Online]. Available: http://netsafe.org.nz/Doc_Library/racheloconnell1.pdf

[15] V. Gupta and G. Lehal, "A survey of text summarization extractive techniques," *Journal of Emerging Technologies in Web Intelligence*, vol. 2, no. 3, pp. 258–268, 2010.

[16] N. Christianni and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, UK: Cambridge University Press, 2000.

[17] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York: Springer, 2008.

[18] B. Baharudin, L. H. Lee, and K. Khan, "A review of machine learning algorithms for text-documents classification," *Journal of advances in information technology*, vol. 1, no. 1, pp. 4–20, 2010.

[19] M. Wollis, "A linguistic analysis of online predator grooming," Ph.D. dissertation, College of Agriculture and Life Sciences, 2011.

[20] Hindarto and Sumarno, "Feature Extraction of Electroencephalography Signals using Fast Fourier Transform," *CommIT (Communication & Information Technology) Journal*, vol. 10, no. 1, pp. 49–52, 2016. http://journal.binus.ac.id/index.php/commit/article/view/1548/1421