338

Comparison Support Vector Machine and Naive Bayes Methods for Classifying Cyberbullying in Twitter

Nur Chamidah [1], Reiza Sahawaly [2]

[1] Department of Mathematics, Airlangga University, Surabaya, Indonesia
[2] Statistics Study Program, Department of Mathematics, Airlangga University, Surabaya, Indonesia

ARTICLE INFO

ABSTRACT

Twitter users in Indonesia in 2019 were recorded at 6.43 million. The high level of Twitter users makes it allows for free opinion to anyone. It can cause cyberbullying. Victims of cyberbullying experienced higher levels of depression than other verbal acts of violence. The forms of cyberbullying that occur on Twitter are flaming, Denigration, and Body Shaming. The research contribution is able to make social media developers and users more aware of the type of cyberbullying that social media users sometimes do without realizing it. Social media developers can prevent cyberbullying by using policies such as word detection and filtering features that indicate cyberbullying more accurately by classifying it by type and using the most accurate method. To classify cyberbullying forms in twitter, in this study, we use the Naïve Bayes method and Support Vector Machine (SVM) and compare them based on classification accuracy. This research will also identify words that are characteristic of each category of cyberbullying so that each category is easy to identify by social media users and makes it easier to avoid cyberbullying. The results of this study are the classification accuracy of Naïve Bayes of 97.99% and the classification accuracy of SVM of 99.60%. It means that SVM is better than Naïve Bayes for classifying the forms of cyberbullying in Twitter.

Nur Chamidah,
Department of Mathematics, Faculty of Science and Technology, Airlangga University, Jl. Dr. Ir. H. Soekarno, Mulyorejo, Surabaya, Indonesia.
Email: nur-c@fst.unair.ac.id

1. INTRODUCTION

Social media is often used because it provides many conveniences for its users so that it allows communication with others via digital devices for daily interactions. The technological revolution is now changing the way people communicate and relate, which means not only communicating with people in certain environments but with other people in cyberspace [1][2][3]. One of the Social Media that is often used is Twitter. Twitter has a simple User Interface (UI) and is able to share information in the form of text, photos, and videos easily. Twitter users in Indonesia in 2019 were recorded as many as 6.43 million users [4]. The high level of Twitter users also allows for free opinions to anyone. It can cause cyberbullying. Cyberbullying is the harsh treatment carried out that has the intrinsic nature of bullying characters such as aggressive, intentional, and repeated actions over time by one or many people and has the aim of creating terror for the victims [5][6]. Cyberbullying, in a broader definition, is a crime based on three sources of risk, there are individual, society, and the environment [7]. Victims of bullying usually have been positioned as targets. If bullying occurs continuously, it will cause trauma, fear, anxiety, depression, and even death [8][9]. Cyberbullying can attack someone regardless of their status or religion. Pressure or intimidation, either physically or verbally, can cause depression, but the researchers found that victims of cyberbullying experienced higher levels of depression than other verbal acts of violence [10].

A crime can occur due to the factors that cause crime which is seen from two sides, factors that originate from inside the subject (internal factors) and factors that arise from outside the subject (external

factors). These factors in the case of cyberbullying can affect the bullies in using different words in their messages, thus forming cyberbullying categories that differ from each other [11][12]. Forms of cyberbullying that occur on Twitter are flaming, Denigration, and Body-shaming. Flaming is fighting online using electronic messages with abusive and vulgar terms, such as swearing, gossiping, or mocking. Denigration is sending or posting gossip or rumors about someone to ruin their reputation (defamation) [13][14]. Body Shaming is an act of criticizing or denouncing the shape, size, and physical appearance of others [15]. These actions can make the victim become insecure about their own body. The classification of cyberbullying categories on Twitter can use the text mining data processing method. Text mining is the process of extracting information in a large set of documents and can automatically identify patterns and special relationships of textual data [16].

Many methods have been developed to classify text, two of them which are Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) [17]. Naïve Bayes is a probability-based classification method. This method calculates a set of probabilities by sum the frequency and combination of values from a dataset. This method applies assumes that uses the Bayes Theorem and assumes all the attributes are independent or not interdependent determined by the value of the class variable [18]. SVM method is a non-probabilistic binary linear classification technique that represents each document in the form of a vector as a sample space. The SVM method is used to analyze vector data based on words with training models to determine hyperplane [19].

Research on cyberbullying about the classification of cyberbullying based on its form so far hasn't been done. Previous studies are only sentiment analyses of comments on social media to identify the document whether cyberbullying or not. These studies include sentiment analysis to detect cyberbullying on Facebook comments [20], but there is no classification based on the forms of cyberbullying. Other studies [21] identified cyberbullying tweets on Twitter social media. In that study, researchers identified comments that contain bullying and not bullying without any classification based on cyberbullying forms.

Based on the explanation above, this study will discuss the classifying of the tweets that indicated cyberbullying on Twitter using the Naïve Bayes Classifier method and Support Vector Machine. Indicated cyberbullying comments will be classified based on their forms, such as flaming, denigration, and body-shaming. This study aims to identify words that are characteristic of a cyberbullying category and to compare the best method to classify cyberbullying data between Naïve Bayes and SVM.

The research contribution is able to make social media developers and users more aware of the type of cyberbullying that social media users sometimes do without realizing it. Social media developers can prevent cyberbullying by using policies such as word detection and filtering features that indicate cyberbullying more accurately by categorizing it by type and using the most accurate method.

2. RESEARCH METHOD

The data used in this study are text data in the form of comments that indicated cyberbullying in Indonesia. The comments were taken from tweets of social media Twitter users. Retrieval of tweets data using website crawling techniques with the Application Program Interface (API) that has been provided by Twitter. The amount of data used is 1000 tweets, with three categories based on the types of cyberbullying, there are Flaming, Denigration, and Body Shaming. The tweets period taken is between December 2019 and February 2020. To find out the characteristics of the word cyberbullying category and classify it using the SVM and Naïve Bayes methods, the steps of research are website crawling using the API on Twitter, text processing, k-fold cross-validation, classifying data with SVM and Naïve Bayes methods, calculating the APPER value.

2.1. Website Crawling Using the API on Twitter

Website crawling using the API on Twitter is used to get cyberbullying tweets. The keywords used in taking tweets are words that indicate bullying. The steps are carried out with OSS-R software using the TwitterR package library.

2.2. Text Processing

Text processing is a process to prepare data so it can be read properly by the system and get optimal classification results. Some of the steps are as follows:
1. It is selecting data manually to retrieve tweets that are indicated cyberbullying and discard tweets that are not cyberbullying.
2. Label the classification of cyberbullying categories in the data manually using Microsoft Excel software.

3. Cleans documents from unnecessary characters such as emoticons, hashtags (#), numbers, symbols, punctuation marks, slangwords and stopwords.
4. They are transforming cases and stemming from data. At this stage, all the letters in the document will be changed to lowercase and basic words.
5. Form a word frequency matrix by using the Term Document Matrix function on OSS-R

2.3. K-Fold Cross-Validation

K-fold cross-validation is a technique for breaking data into k sections of data set of the same size. K-fold cross-validation is used to eliminate bias in the data. Training and testing are carried out k times. In the first try, the S_k subset is treated as test data, and the other subset is treated as training data. In the second trial, the subsets S_1, S_2, …, S_{k-2}, S_k become training data, and S_{k-1} become test data, and so on [22]. An example of applying the K-fold cross-validation method is illustrated in Fig. 1.
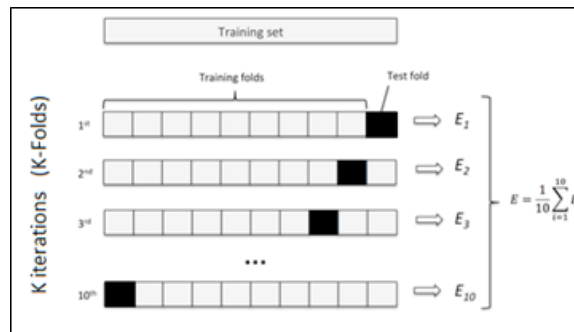

Fig. 1. Illustration of *K-fold cross-validation* with 10 iterations

2.4. Classifying Data with SVM and Naïve Bayes Methods
2.4.1 Naïve Bayes Classifier

Naïve Bayes has two stages in the classification of texts. There are training and testing stages. At the training stage, it will do an analysis process of the document sample in the form of vocabulary selection, which is the word that appears in the collection of sample documents that can be a representation of the document. Then the prior probability for each category is determined based on a sample document and determines the category value of a document based on the terms that appear in the classified document [23].

Assuming the collection of documents as $D= \{d_1, d_2, …, d_n\}$ and the category collection as $V= \{v_1, v_2, …, v_n\}$. The NBC classification begins by calculating the probability of $P(v_j|d_i)$, that is the probability of the v_j category if document d_i is known. The document d_i is an *n-tuple* of words in the document, that is $\{a_1, a_2, …, a_n\}$ whose frequency of occurrence is assumed to be a random variable with a Bernoulli probability distribution [24][25]. Furthermore, the classification of documents can be done by calculating the maximum posterior value based on the equation [26]:

$$V_{MAP} = \underset{v_j \in V}{arg\ max}\ P(v_j|a_1, a_2, …, a_n) \tag{1}$$

Based on (1), by applying the Bayes theorem. It can be written as follow

$$V_{MAP} = \underset{v_j \in V}{arg\ max}\ \frac{P(a_1, a_2, …, a_n|v_j)P(v_j)}{P(a_1, a_2, …, a_n)} \tag{2}$$

Since the value of $P(a_1, a_2, …, a_n)$ is constant, it can be ignored, so the equation (2) can be written as follow

$$V_{MAP} \propto \underset{v_j \in V}{arg\ max}\ P(a_1, a_2, …, a_n|v_j)P(v_j) \tag{3}$$

Next, in classification using the Naive Bayes method, it can be assumed that each word in $\{a_1, a_2, …, a_n\}$ is independent. Because of $P(a_1, a_2, …, a_n|v_j) = \prod_i P(a_i|v_j)$, the equation can be written as follow

$$V_{MAP} \propto \underset{v_j \in V}{arg\ max}\ P(v_j) \prod_i P(a_i|v_j) \tag{4}$$

The $P(v_j)$ value is determined in the training process, so the value is approximated by

$$P(v_j) = \frac{N_j}{N} \qquad (5)$$

Where N_j is the number of documents with category j in the training data, and N is the number of documents used for training data.

The term does not always appear in one of the categories during classification, so the value of $P(a_i|v_j)$ is zero [26]. To overcome this problem, *add-one smoothing* or *Laplace smoothing* is used by adding 1 of term frequency so that the equation (5) can be written as follow

$$P(a_i|v_j) = \frac{N_{ji} + 1}{N_j + B} \qquad (6)$$

2.4.2. Support Vector Machine

The algorithm of SVM basically works by defining the limit between two classes with the maximum distance from the closest data. To get this limit, we need the best hyperplane in the *input space* obtained by measuring the hyperplane's margin and looking for the maximum point. Margin is the distance between the hyperplane and the closest pattern of each class. The closest pattern is called a support vector [27]. The SVM method is illustrated in Fig. 2.
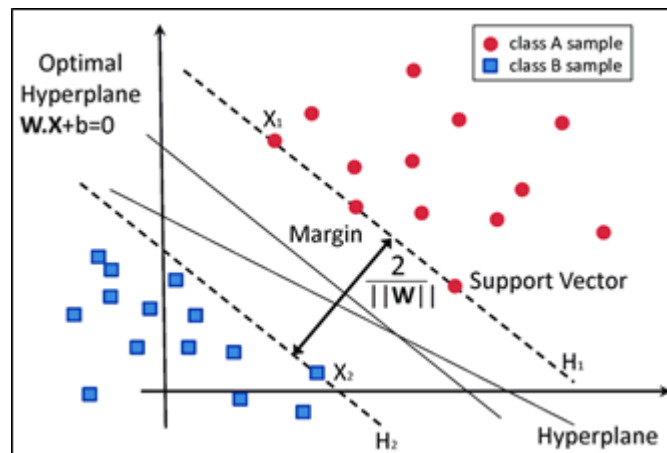

Fig. 2. Illustration of the SVM method

Mathematically, the problem of separating vectors that have two different groups in a number of m documents is formulated as follows.

$$(y_1, x_1), \ldots, (ym, xm), x \in R^n, y \in \{-1, +1\} \qquad (7)$$

and *hyperplane* is defined as follows [28].

$$(wx) + b = 0 \qquad (8)$$

Where $w = \{w_1, w_2, ..., w_n\}$ is a weighting vector, *n* is the number of attributes, x is the input data used as an attribute, and *b* is a scalar as an additional weight. The first delimiter field restricts the first class, while the second delimiter restricts the second class. The two bounding fields are stated in the equation as follows:

$$yi[(wx_i) + b] - 1 \geq 0 \qquad (9)$$

The margin value between the boundary planes based on the formula of the distance of the line to the center point is

$$\frac{2}{\|w\|} \qquad (10)$$

The margin value is maximized while still comply in (8). Maximizing $1/\|w\|$ is the same as minimizing $\|w\|^2$. So that the search for the best separator field with the largest margin value can be formulated into a constrained optimization problem, that is.

$$min \frac{1}{2}\|\boldsymbol{w}\|^2 \tag{11}$$

Where $y_i(\boldsymbol{wx_i}) + b - 1 \geq 0$ and the limitation functions can be written as follows.

$$\sum_{i=1}^{m} \alpha_i[y_i(\boldsymbol{wx_i} + b) - 1] \tag{12}$$

These problems can be solved using the *Lagrange multiplier* method. The equation can be transformed into the *Lagrange multiplier* function as follows.

$$L(\boldsymbol{w}, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i \{[(\boldsymbol{w}.\boldsymbol{x_i}) + b]y_i - 1\} \tag{13}$$

With $\alpha_i \geq 0$ is the Lagrange coefficient. Based on equation (8), the equation (13) can be written as follows.

$$L(\boldsymbol{w}, b, \alpha) = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{m} \alpha_i y_i(\boldsymbol{w}.\boldsymbol{x_i} + b) + \sum_{i=1}^{m} \alpha_i \tag{14}$$

To calculate the value of α_i (Lagrange coefficient), We must minimize L in (14) with respect to w and b. This value is obtained by calculating the partial derivative of L with respect to w and b, so the equation can be written as follows.

$$min \left(\frac{\partial}{\partial_b} L(\boldsymbol{w}, b, \alpha) = 0\right) \Rightarrow \sum_{i=1}^{m} \alpha_i y_i = 0$$

$$min \left(\frac{\partial}{\partial_w} L(\boldsymbol{w}, b, \alpha) = 0\right) \Rightarrow \boldsymbol{w} = \sum_{i=1}^{m} \alpha_i \boldsymbol{x_i} y_i \tag{15}$$

The langrage function in (14) is converted into (16) by substituting (15) into (14).

$$L(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=j=1}^{m} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i}.\boldsymbol{x_j}) \tag{16}$$

The determination of the best separation of fields is formulated as follows.

$$max(L(\alpha)) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=j=1}^{m} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i}.\boldsymbol{x_j}) \tag{17}$$

where the constraint function $\sum_{i=1}^{m} \alpha_i y_i = 0$ and $\alpha_i \geq 0$ for $i = 1, 2,..., m$. Based on the function of the obstacle, equation (17) becomes,

$$\widehat{\boldsymbol{w}} = \sum_{i=1}^{m} \widehat{\alpha_i} \boldsymbol{x_i} y_i \tag{18}$$

So the value of *b* is obtained as follows,

$$\hat{b} = -\frac{1}{2} \widehat{\boldsymbol{w}}.[\boldsymbol{x_r} + \boldsymbol{x_s}] \tag{19}$$

where $\boldsymbol{x_r}$ and $\boldsymbol{x_s}$ are support vectors that comply with each class, for $\widehat{\alpha_r}, \widehat{\alpha_s} > 0, y_r = 1, y_s = -1$. The decision function obtained based on (18) and (19) are as follows [29]

$$f(\boldsymbol{x}) = sign\left(\widehat{\boldsymbol{w}}.\boldsymbol{x_i} + \hat{b}\right) \tag{20}$$

2.5. Calculating the APPER Value

Apparent Error Rate (APPER) is a value that is used to see the probability of error in classifying objects. The calculation of classification accuracy is based on the results of the classification process that has been summarized into a classification table. The classification table with three categories is given in Table 1.

Table 1. Classification Table of Three Categories

Observation	Prediction			Sum
	Category 1	Category 2	Category 3	
Category 1	n_{11}	n_{12}	n_{13}	$n_{11} + n_{21} + n_{31}$
Category 2	n_{21}	n_{22}	n_{23}	$n_{12} + n_{22} + n_{32}$
Category 3	n_{31}	n_{32}	n_{33}	$n_{13} + n_{23} + n_{33}$
Total	$n_{11} + n_{21} + n_{31}$	$n_{12} + n_{22} + n_{32}$	$n_{13} + n_{23} + n_{33}$	$n_{11} + n_{21} + n_{31} + n_{12} + n_{22} + n_{32} + n_{13} + n_{23} + n_{33}$

Based on Table 1, errors in the classification of objects can be calculated using APPER [30], which is defined as.

$$APPER = \frac{n_{12} + n_{13} + n_{21} + n_{23} + n_{31} + n_{32}}{n_{11} + n_{21} + n_{31} + n_{12} + n_{22} + n_{32} + n_{13} + n_{23} + n_{33}} \times 100\% \qquad (21)$$

Meanwhile, to determine the accuracy of classification, we use the following equation:

$$Accuracy\ of\ Classification = \frac{n_{11} + n_{22} + n_{33}}{n_{11} + n_{21} + n_{31} + n_{12} + n_{22} + n_{32} + n_{13} + n_{23} + n_{33}} \times 100\% \qquad (22)$$

1.6 Flowchart

Based on the research steps that have been described, to simplify the explanation of the process carried out, the process will be described in a flowchart. Fig. 3 is a flowchart of the entire analysis process carried out.
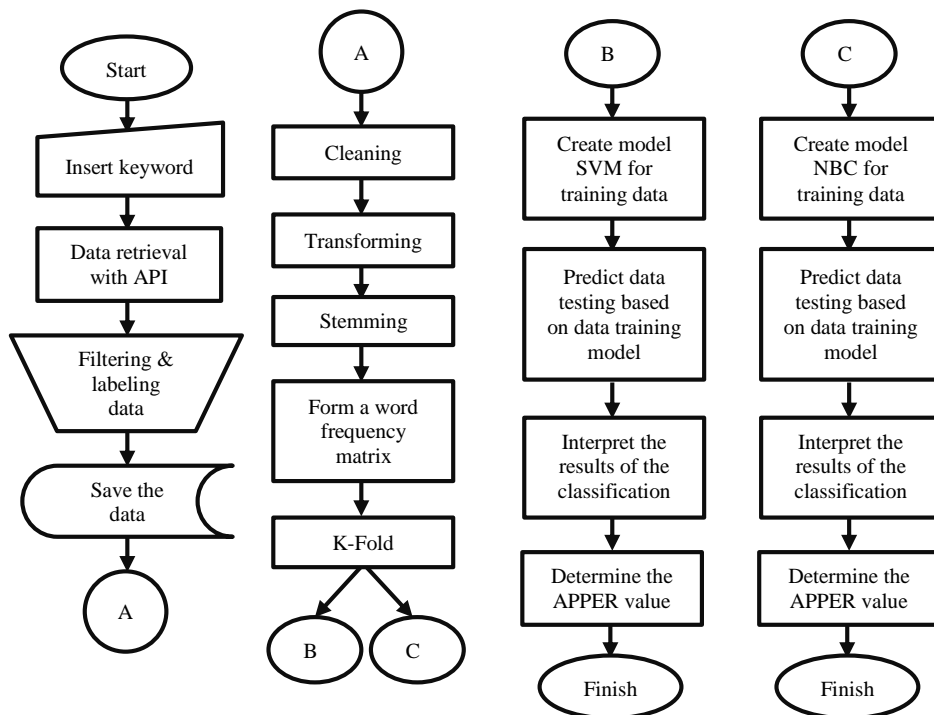
Fig. 3. Flowchart of the analytics method

3. RESULTS AND DISCUSSION

Table 2 gives the 10 words with the highest frequency of cyberbullying data in the body-shaming, denigration, and flaming categories.

Table 2. Ten Words with Highest Frequency in Three Cyberbullying Categories

Bodyshaming		Denigration		Flaming	
Word	**Frequency**	**Word**	**Frequency**	**Word**	**Frequency**
Cebol	58	Isu	180	Anjing	200
Jelek	52	Alih	159	Goblok	102
Botak	51	uang	67	bangsat	91
Gendut	50	becus	61	Orang	32
Item	50	plagiat	49	Anak	18
cungkring	39	haram	45	Bodoh	18
Dekil	38	kerja	39	Tahu	18
Sipit	32	orang	35	Tai	14
Buncit	27	kasus	35	Lihat	12
Orang	27	gelap	32	Baru	11

Based on Table 2, the words that characterize the body-shaming category are "cebol", "jelek", "botak", "gendut", "item", "cungkring", "dekil", "sipit", and "buncit" with "cebol" as a word that is widely used by bullier in bodyshaming. Words that characterize the denigration category are "isu", "alih", "uang", "becus", "plagiat", "haram", "kerja", "kasus", and "gelap" with the word "isu" as the word most often used by the bullier in denigration. Words that characterize the Flaming category are "anjing", "goblok", "bangsat", "bodoh", "tai". While, "anak", "tahu", "lihat", and "baru" are not considered as their characteristics because affixes word and small frequency, the word "anjing" is the most word widely used by bullier in flaming. The word "orang" is a word that is characteristic of cyberbullying in general because it is found in each category.

The classification accuracy of cyberbullying with 10-iterations by using Naïve Bayes and SVM methods obtained the following results in Table 3 and Table 4.

Table 3. Naïve Bayes Method analysis results

Iteration	Accuracy		Total Accuracy
	Training	**Testing**	
1	98%	93%	97.80%
2	98.33%	96%	98.10%
3	98.67%	97%	98.50%
4	97.56%	83%	97.30%
5	98.22%	95%	97.90%
6	98.44%	99%	98.50%
7	97.89%	91%	97.20%
8	98.56%	98%	98.50%
9	98.44%	95%	98.10%
10	98%	98%	98%
		Average	97.99%

Table 4. SVM Method analysis results

Iteration	Accuracy		Total Accuracy
	Training	**Testing**	
1	100%	97%	99.70%
2	100%	95%	99.50%
3	100%	94%	99.40%
4	99.89%	96%	99.50%
5	100%	97%	99.70%
6	100%	98%	99.80%
7	100%	95%	99.50%
8	100%	94%	99.40%
9	100%	95%	99.50%
10	100%	100%	100%
		Average	99.60%

Based on Table 3, the Naïve Bayes Classifier method can classify three forms of cyberbullying on Twitter with an average classification accuracy of 97.99%. Based on Table 4, the Support Vector Machine method can classify three forms of cyberbullying on Twitter with an average classification accuracy of 99.60%. It shows that the SVM method can classify cyberbullying forms better than the Naïve Bayes Classifier method.

4. CONCLUSION

Words that are widely used by bullier in the body-shaming category are *cebol*, *jelek*, *botak*, *gendut*, and item with almost the same frequency. Words that are often used by bullier in the denigration category are *alih* and *isu*. Words that are often used by bullier in the flaming category are *anjing*, *goblok* and *bangsat*. Based on classification accuracy of cyberbullying for body-shaming, denigration, and flaming categories using Naïve Bayes and SVM methods, the study result shows that the SVM method is very good to be used to identify forms of cyberbullying because it has better classification accuracy than the Naïve Bayes method.

Suggestions for further research in analytical methods are using other methods in machine learning to find out which is more accurate because the data will continue to be updated. Suggestions for the type of category, it is possible that other categories in cyberbullying may appear that can be added. Suggestions related to data, when more data is used in making the model, the model will be more accurate.

REFERENCES

[1] J. G. Green, "Using School Disciplinary Context to Understand Adolescent Health Behaviors," in *Journal of Adolesc Health*, vol. 62, no. 2, pp. 126-127, 2018. https://doi.org/10.1016/j.jadohealth.2017.11.287

[2] E. L. Meneses, E. V. Cano, M. D. G. Zamar, and E. A. Segura, "Socioeconomic Effects in Cyberbullying: Global Research Trends in the Educational Context," in *International Journal of Environmental Research and Public Health*, vol. 17, no. 12, p. 4369, 2020. https://doi.org/10.3390/ijerph17124369

[3] N. Kurniawati, E. H. Maolida and A. G. Anjaniputra, "The praxis of digital literacy in the EFL classroom: Digital immigrant vs digital-native teacher," in *Indonesian Journal of Applied Linguistics*, vol. 8, no. 1, pp. 28-37, 2018. https://doi.org/10.17509/ijal.v8i1.11459

[4] Wearesocial, Global Digital Report 2019, 2019. https://wearesocial.com/global-digital-report-2019

[5] K. Kircaburun, P. Jonason, M. D. Griffiths, E. Aslanargun, E. Emirtekin, S. B. Tosuntas, and J. Billieux, "Childhood Emotional Abuse and Cyberbullying Perpetration: The Role of Dark Personality Traits," in Journal of Interpersonal Violence, pp. 1–17, 2019. https://doi.org/10.1177/0886260519889930

[6] M. Zhong, X. Huang, E. S. Huebner, and L. Tian, "Association between bullying victimization and depressive symptoms in children: The mediating role of self-esteem," *Journal of Affective Disorders*, vol. 294, pp. 322-328, 2021. https://doi.org/10.1016/j.jad.2021.07.016

[7] S. A. Hemphill, J. A. Heerde, and K. E. Scholes-Balog, "Risk factors and risk-based protective factors for violent offending: A study of young Victorians," in *Journal of Criminal Justice*, vol. 45, pp. 94–100, 2016. https://doi.org/10.1016/j.jcrimjus.2016.02.012

[8] S.H. Bong, K. M. Kim, K. H. Seol, and J. W. Kim, "Bullying perpetration and victimization in elementary school students diagnosed with attention-deficit/hyperactivity disorder," *Asian Journal of Psychiatry*, vol. 62, p. 102729, 2021. https://doi.org/10.1016/j.ajp.2021.102729

[9] D. Yoon, S. L. Shipe, J. Park, and M. Yoon, "Bullying patterns and their associations with child maltreatment and adolescent psychosocial problems," *Children and Youth Services Review*, vol. 129, p. 106178, 2021. https://doi.org/10.1016/j.childyouth.2021.106178

[10] C. Liu, Z. Liu, and G. Yuan, "The longitudinal influence of cyberbullying victimization on depression and posttraumatic stress symptoms: The mediation role of rumination," Archives of Psychiatric Nursing, vol. 34, Issue 4, pp. 206-210, 2020. https://doi.org/10.1016/j.apnu.2020.05.002

[11] S. Yubero, R. Navarro, M. Elche, E. Larrañaga, and A. Ovejero, "Cyberbullying victimization in higher education: An exploratory analysis of its association with social and emotional factors among Spanish students," *Computers in Human Behavior*, vol. 75, pp. 439-449, 2017. https://doi.org/10.1016/j.chb.2017.05.037

[12] Y. Peled, "Cyberbullying and its influence on academic, social, and emotional development of undergraduate students," in *Heliyon*, vol. 5, no. 3, p. e01393, 2019. https://doi.org/10.1016/j.heliyon.2019.e01393

[13] S. Bauman, Cyberbullying: What Counselors Need to Know, John Wiley & Sons, 2014. https://doi.org/10.1002/9781119221685

[14] L. K. Watts, J. Wagner, B. Velasquez, and P. I. Behrens, "Cyberbullying in higher education: A literature review," *Computers in Human Behavior*, vol. 69, pp. 268-274, 2017. https://doi.org/10.1016/j.chb.2016.12.038

[15] J. P. Chaplin, Dictionary of Psychology, Rajawali Press, Jakarta, 2005.

[16] R. Feldman, and J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press, New York, 2007. https://doi.org/10.1017/CBO9780511546914

[17] A. Tripathy, A. Agrawal, and K. R. Santanu, "Classification of sentiment reviews using n-gram machine learning approach - Expert Systems with Applications," *Expert Systems with Applications*, vol. 57, pp. 117–126, 2016. https://doi.org/10.1016/j.eswa.2016.03.028

[18] S. Pramana, B. Yuniarto, S. Mariyah, I. Santoso, and R. Nooraeni, Data Mining dengan R, IN MEDIA, *Bogor, 2018.*

[19] H. T. Sueno, B. D. Gerardo, and R. P. Medina, "Multi-class Document Classification using Support Vector Machine (SVM) Based on Improved Naïve Bayes Vectorization Technique," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no 3, 2020. https://doi.org/10.30534/ijatcse/2020/216932020

[20] M. Fortunatusa, P. Anthonya, and S. Charters, "Combining textual features to detect cyberbullying in social media posts," 24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Procedia Computer Science vol. 176, pp. 612–621, 2020. https://doi.org/10.1016/j.procs.2020.08.063

[21] M. A. Al-garadi, K. D. Varathan, and S. D. Ravana, "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network," *Computers in Human Behavior*, vol. 63, pp. 433-443, 2016. https://doi.org/10.1016/j.chb.2016.05.051

[22] M. Bramer, Principles of Data Mining. Springer-Verlag, London, 2007. http://doi.org/10.1007/978-1-4471-7493-6

[23] Y. B. N. D. Artissal, I. Asror, and S. A. Faraby, "Personality Classification based on Facebook Status Text Using Multinomial Naïve Bayes Method," The 2nd International Conference on Data and Information Science, IOP Conf. Series: Journal of Physics: Conf. Series 1192, 012003, 2019. https://doi.org/10.1088/1742-6596/1192/1/012003

[24] A. A. Bimantara, A. Larasati, E. M. Risondang, M. Z. Naf'an, and N. A. S. Nugraha, "Sentiment Analysis of Cyberbullying on Instagram User Comments," in *Journal of Data Science and Its Applications*, vol. 2, no. 1, pp. 38-48, 2019. https://doi.org/10.21108/jdsa.2019.2.20

[25] J.-Q. Zhu, A. N. Sanborn, and N. Chater, "The Bayesian Sampler: Generic Bayesian Inference Causes Incoherence in Human Probability Judgments," *Psychological Review*, vol. 127, no. 5, pp. 719-748, 2020. http://dx.doi.org/10.1037/rev0000190

[26] T. M. Mitchell, Machine Learning, McGraw-Hill Science/Engineering/Math, 1997.

[27] L. Yang, and H. Dong, "Support Vector Machine with Truncated Pinball Loss and its Application in Pattern Recognition," in *Chemometrics and Intelligent Laboratory Systems*, vol. 177, pp. 89-99, 2018. https://doi.org/10.1016/j.chemolab.2018.04.003

[28] M. U. Hasan, S. Ullah, M. J. Khan, and K. Khurshid, "Comparative Analysis of SVM, ANN and CNN for Classifying Vegetation Species Using Hyperspectral Thermal Infrared Data," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W13, pp. 1861-1868, 2019. https://doi.org/10.5194/isprs-archives-XLII-2-W13-1861-2019

[29] S. R. Gunn, Support Vector Machine for Classification and Regression, Southampton: University of Southampton, 1998.

[30] L. S. Meyers, G. Glenn, and A. J. Guarino, Applied Multivariate Research, 3rd edition, *SAGE Publication*, United States of America, 2016. https://us.sagepub.com/en-us/nam/applied-multivariate-research/book246895

BIOGRAPHY OF AUTHORS

Nur Chamidah was born in Surabaya, East Java, Indonesia, on 2 June 1972. She received a bachelor (S.Si.) degree in Mathematics from Airlangga University, Surabaya, Indonesia, in 1997, Masters (M.Si.) degree in Statistics from Sepuluh Nopember Institute of Technology, Surabaya, Indonesia in 2002, and a Doctorate (Dr.) degree in Statistics from Sepuluh Nopember Institute of Technology, Surabaya, Indonesia in 2014. She works as a lecturer at the study program of Statistics, Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia, and a headmaster of Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia. Her research interests are Statistical Modeling in Life Science, Spatial Model, Data Mining and Nonparametric and Semiparametric regressions models. Email: nur-c@fst.unair.ac.id

Reiza Sahawaly is a graduate student of the Statistics Study Program, Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia. His research interests are computational statistics and machine learning focusing on text analytics and processing. Email: reiza.sahawaly-2016@fst.unair.ac.id