Indonesian Journal of Electrical Engineering and Computer Science
Vol. 18, No. 3, June 2020, pp. 1494~1500
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v18.i3.pp1494-1500

Cyberbullying identification in twitter using support vector machine and information gain based feature selection

Ni Made Gita Dwi Purnamasari[1], M. Ali Fauzi[2], Indriati[3], Liana Shinta Dewi[4]
[1,2,3]Faculty of Computer Science, Brawijaya University, Indonesia
[4]Faculty of Cultural Studies, Brawijaya University, Indonesia

Article Info

Article history:

Received Apr 1, 2019
Revised Jul 7, 2019
Accepted Aug 21, 2019

Keywords:

Cyberbullying
Information gain
Support vector machine
Text classification

ABSTRACT

Cyberbullying is one of the actions that violate the ITE Law where the crime is committed on social media applications such as Twitter. This action is difficult to detect if no one is reporting the tweet. Cyberbullying tweet identification aims to classify tweets that contain bullying. Classification is done using Support Vector Machine method where this method aims to find the dividing hyperplane between negative and positive class. This study is a text classification where more data is used, the more features are produced, therefore this research also uses Information Gain as feature selection to select features that are not relevant to the classification. The process of the system starts from text preprocessing with tokenizing, filtering, stemming and term weighting. Then perform the information gain feature selection by calculating the entropy value of each term. After that perform the classification process based on the terms that have been selected, and the output of the system is identification whether the tweet is bullying or not. The result of using SVM method is accuracy 75%, precision 70.27%, recall 86.66% and f-measure 77.61% on experiment maximum iteration = 20, $\lambda = 0.5$, $\gamma = 0.001$, $\varepsilon = 0.000001$, and $C = 1$. The best threshold of information gain is 90%, with accuracy 76.66%, precision 72.22%, recall 86.66% and f-measure 78.78%.

Corresponding Author:

M. Ali Fauzi,
Faculty of Cultural Studies,
Brawijaya University, Malang, Indonesia.
Email: moch.ali.fauzi@ub.ac.id

1. INTRODUCTION

The advance of information and communication technology has been brought many benefits to the society. One of the brief example is social media in which people can find a lot of friends and extend their networks. However, the rise of this new technology tend to be double-edge knife as suggested by Segal [1] because it also bring some damaging effects such as cyberbullying [2]. Cyberbullying is one of the violent acts committed by a person against his victim on the internet, where the victim is humiliated, ridiculed, and intimidated [3]. Cyberbullying can have an impact on victim's mental, even there is many cases where the victims of bullying end up in suicide because they can not stand with much pressure [4-5]. Currently the Indonesian Ministry of Communication and Informatics (KOMINFO) is working with Google and Twitter in the prevention of negative content on the internet such as pornography, hoaxes, cyberbullying, and others. KOMINFO hopes that the prevention of the spread of negative content will be quickly resolved. Twitter has been providing a function where the report about negative content in a separate order to be responded more quickly. Therefore with the existence of an automated system for identification tweet cyberbullying can help in handling tweets hate rashers more quickly overcome and more efficient.

One of the research about cyberbullying is conducted by [6] entitled Cyberbullying Comment Classification on Indonesian Selebgram using Support Vector Machine (SVM) Method. This work focused

on comments in a selebgram (Instagram celebrity) account called Awkarin. The data used were taken from the comment section of one of the photos uploaded by Awkarin's account. The classification results obtained quite good performace with accuracy level of 79.412%.

The other research on cyberbullying [7] used SVM and Naïve Bayes (NB) method for the classification process. The data used were obtained from Kaggle. The dataset contained 1600 conversations Indonesian Language from Formspring.me website. The researchers also compared their results from previous studies by Reynolds [8] that used Decision Tree and K-Nearest Neighbour (K-NN). From the research, it was found that SVM method is better in cyberbullying classification than K-NN and Decision Tree method with accuracy level for SVM, Decision Tree, and KNN were 99.41%, 78.28%, and 89.01% respectively. Another work by Nahar et al. [9] on this topic also suggested that SVM could give a promising result for this task. However, SVM method has relatively high computational complexity [10]. Moreover, the most widely used feature in the cyberbullying classification task is Bag of Words (BoW) that has high dimensionality [11]. Using BoW as features, all unique words from the document collection are used as features so that a lot of features are used for the classification task. Therefore, feature selection is needed to select only relevant features in order to reduce the time complexity and increase the performace.

Several prior works suggested that the use of feature selection can make the text classification more effective and efficient. Information Gain (IG) is the widely used technique in the text classification task [12-13]. A study by Chandani, et al. [14] investigated the use of several feature selection methods including Information Gain (IG), Chi Square, Forward Selection and Backward Selection with several machine learning methods for review classification. The experiment results showed that SVM with IG feature selection method has the highest accuracy compared to other methods accuracy 81.50%.

In this study, we propose a cyberbullying classification from Twitter data using SVM method as the classifier and IG as the feature selection technique. We also investigate the effect of different SVM parameters and different IG selection threshold. The data used were tweets (twitter posts) that mentioned Vice Chairman of the Indonesian House of Representatives, Mr. Fadli Zon.

2. RESEARCH METHOD

The design of the system is depicted in Figure 1. The first step of this work is text preprocessing and followed by feature selection using IG. Then, the classification task is conducted using SVM method to get the cyberbullying identification result.

The preprocessing involves some processes such as tokenizing, filtering, stemming and then term weighting. The features used in this work are BoW with term frequency-inverse document frequency (TF-IDF) as the term weighting method. The features are then used as the input of SVM. Before the classification task using the SVM method conducted, the IG value of each features are calculated and the features with highest IG values are selected to represent the document. A certain threshold is used to select the features. After the feature has been selected, then the selected features are then used for classification using SVM. Finally, the results of cyberbullying identification are presented.

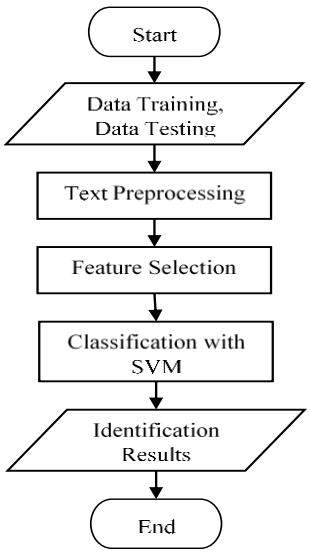

Figure 1. Main system flowchart

2.1. Text Preprocessing

There are 4 process of text preprocessing:

a) Tokenizing

Tokenizing is a process of separating the word on a sentence and deleting all punctuation and special characters. This process aims to separate each word to distinguish certain characters that are treated as word separators or not. The tokenizing process relies on the space character in the document as a word separator [15].

b) Filtering

The filtering stage is the stage of selecting important words from the token result. The most commonly used words are the unrelated words in Information Retrieval (IR) and text mining, these words are called Stopwords [16]. The dictionary of Indonesian stopwords used in this work is obtained from Tala [17].

c) Stemming

Stemming technique is used to find the root of a word. Stemming is done in addition to minimizing the number of different indexes of a document, also done to group words with similar words and basics but different shapes due to different affixes [18].

d) Term Weighting

The term weighting methos used if TF-IDF. TF-IDF is the product result of Normalized Term Frequency (WTF) and Inverse Document Frequency (IDF) to compute the impotance of a term in a document [19]. WTF is the normalized version of term occurances in a document. Using the WTF, the more the occurrence, the higher the weight of the term in the document. However, several insignificant terms often appear in many documents. Therefore, *IDF* is used to reduce the weight of terms that have occurances in many documents [20]. The TF-IDF formula can be seen in (1), (2), and (3).

$$tf - idf_{t,d} = wtf_{t,d} \cdot idf_t \tag{1}$$

$$wtf_{t,d} = \begin{cases} 1 + log_{10} tf_{t,d}, & if\ tf_{t,d} > 0 \\ 0, & if\ tf_{t,d} = 0 \end{cases} \tag{2}$$

Where *tf* is the term frequency that state how much the number of terms in a document.

$$idf_t = log_{10}(\frac{N}{df_t}) \tag{3}$$

Where *N* is the number of documents.

2.2. Feature Selection using Information Gain

Feature selection is one of the process of selecting relevant features in terms of target learning problems. The purpose of feature selection is to remove redundant and irrelevant features because well-chosen features can increase the classification performance [21]. IG is one of a method for feature selection that is widely used by researchers to define the boundaries of an attribute's importance [22]. The IG value is obtained from the entropy value before separation is reduced by the entropy value after the separation. This value is used for determining which attributes to be removed or used. Attributes that meet the weighting criteria will the be used for the classification process.

In the selection process of features with IG done with three following steps:

a) Calculates the entropy for each feature.
b) Specifies the threshold (limit). It is used to determine the number of attributes to be used within the threshold limit.
c) Fixed the dataset with selected attributes.

The IG (*t*) feature selection is formulated in (4)

$$IG(t) = -\sum_{i=1}^{|C|} P(C_i) \log P(C_i) + P(t) \sum_{i=1}^{|C|} P(C_i|t) \log P(C_i|t) + P(\bar{t}) \sum_{i=1}^{|C|} P(C_i|\bar{t}) \log P(C_i|\bar{t}) \tag{4}$$

where C_i is a data class, *P(Ci)* is an opportunity of the data class, *P(t)* and *P(t¯)* are probable term *t* occurring or not appearing in the document. In machine learning, information acquisition can be used to help determine feature ratings [23].

2.3. Classification using SVM

The concept of classification with SVM is to find the best hyperplane that servers as a separator class negative and positive class [24-25]. SVM is also capable of working on high dimensional datasets using

kernel trick. There are several functions of the SVM kernel, such as: Linear, Polynomial, Gaussian RBF, Sigmoid, Multi Quadratic Inverse, and Additive.

In this research kernel function used is SVM Polynomial. Linear SVM is used when data to be classified can be separated by a hyperplane, whereas a non-linear SVM is used when data can only be separated by curved lines. SVM Polynomial has a function definition with (5).

$$K(\overrightarrow{x_i}, \overrightarrow{x_j}) = (\overrightarrow{x_i} \cdot \overrightarrow{x_j} + 1)^d \quad (5)$$

where $K(\overrightarrow{x_i}, \overrightarrow{x_j})$ is a kernel function, *x* is a feature and *d* is an order.

Hyperplane in the optimal SVM is obtained by formulating it into the QP problem and solved using the library that is widely available in numerical analysis. But here is another alternative that is quite simple is the sequential method. This method was developed by Vijayakumar to find the value of α, which is described in step:

1. Initialization $\alpha_i = 0$

   Calculated Hessian Matrix value (6).

$$D_{ij} = y_i y_j (K(x_i . x_j) + \lambda^2) \quad (6)$$

Where is the class from the data i and j, $K(x_i . x_j)$ is a polynomial kernel function.

2. Calculate each level with three step i.e.:

$$\boldsymbol{E_i} = \sum_{j=1}^{n} \boldsymbol{\alpha_i D_{ij}} \quad (7)$$

$$\boldsymbol{\delta\alpha_i} = \boldsymbol{min}\{\boldsymbol{max}[\boldsymbol{\gamma}(\mathbf{1} - \boldsymbol{E_i}), -\boldsymbol{\alpha_i}], \boldsymbol{C} - \boldsymbol{\alpha_i}\} \quad (8)$$

$$\boldsymbol{\alpha_i} = \boldsymbol{\alpha_i} + \boldsymbol{\delta\alpha_i} \quad (9)$$

3. Repeating the step 2 until the α values reaches converging or until reaches maximum iteration.

ϒ is a parameter to control the speed of the learning process. Convergence can be defined from changes in the value of α.

3. RESULTS AND ANALYSIS

In the testing process, Accuracy, Precision, Recall, and F-measure were used. The amount of data used is 300 tweets, of which 150 tweet contain bullying and 150 tweet are not contain bullying. The data were manually labelled by an expert. In the experimtn, the data was splitted into 240 tweets as training data and 60 tweets as testing data.

3.1. Testing of Sequential Training SVM Parameters

There were six parameters tested on sequential training SVM with ten different experimental value i.e. lambda, gamma, epsilon, maximum iteration, and complexity (C) values. The sequential training SVM parameter values used in the test are λ = 0.5, ϒ = 0.001, ε = 0.0001, C = 1, and maximum iteration = 100. Figure 2 shows that the results obtained are constant for all tested lambda values, accuracy 70%, precision 68.75%, recall 73.33% and f-measure 70.96%. This happens because the lambda value is only used to perform hessian matrix calculations. And the hessian matrix is used when calculating the value of Ei for sequential training SVM. So the results obtained do not really affect the bias value. The value of lambda is 0.5.

Figure 3 shows that the best results are obtained at the gamma values of 0.001 and 0.01, then the value decreases as the gamma value is higher. This is because the gamma value is used to calculate the value of the alpha delta, where the value of the alpha delta is the value that determines whether the results are convergent or not. The value of gamma taken is 0.001 with accuracy 70%, precision 68.75%, recall 73.33% and f-measure 70.96%.

Figure 4 shows that the best results at epsilon value are less than 0.000001 and 0.0000001. This is because the epsilon value is used as the maximum limit for convergent results. As the epsilon value gets higher, the results will converge faster. Selection of epsilon value 0,000001 with accuracy 68.33%, precision 64.10%, recall 83.33% and f-measure 72.46%.

In Figure 5, the best results obtained are in the iteration of 20 with accuracy value of 75%, precision 70.27%, recall 86.66% and f-measure 77.61%. Then on the iteration of more than 1000 results are the same, this is because the calculation has entered convergent state at 1404 iteration. Figure 6 shows that the best result on test of C value is 1 with accuracy value of 75%, precision 70.27%, recall 86.66% and f-measure 77.61%. This is because the higher the value of C the polynomial kernel value will be higher too. When the polynomial kernel value is higher then the value of Hessian matrix is higher also, causing the calculation of the value of gamma constants whose results obtained smaller and then can cause iteration becomes faster convergent.

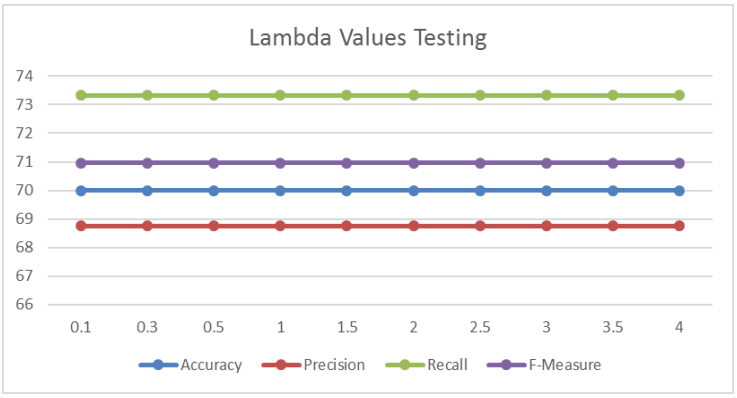

Figure 2. Graph of lambda testing result

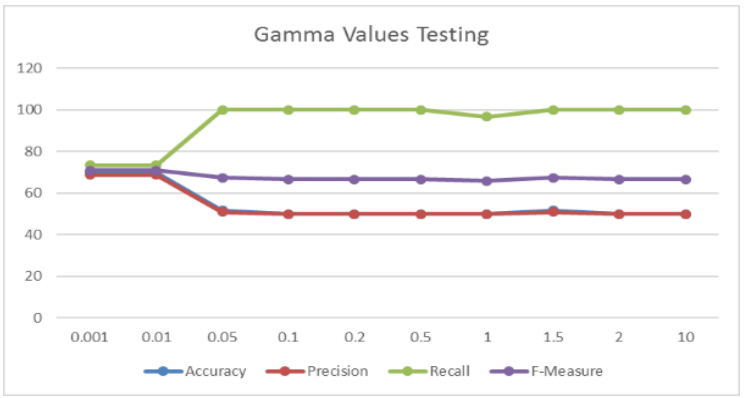

Figure 3. Graph of gamma testing result

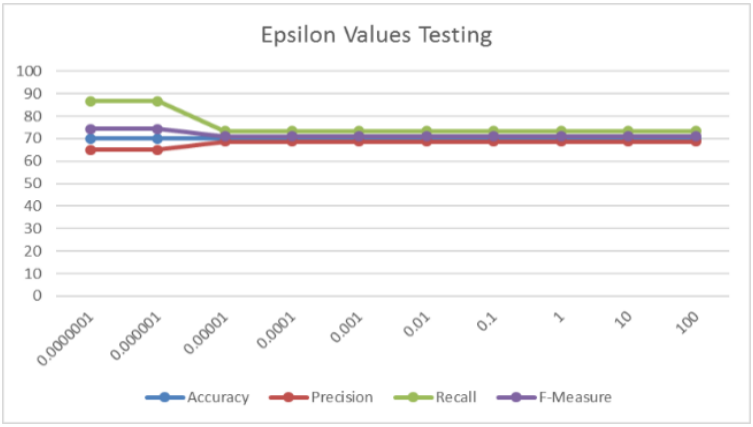

Figure 4. Graph of epsilon testing result

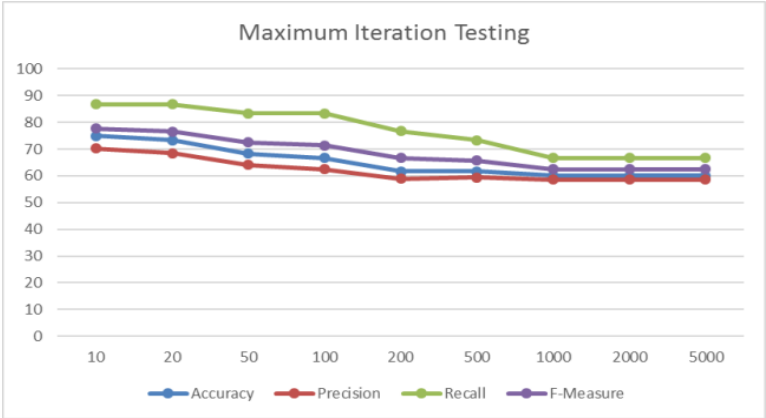

Figure 5. Graph of maximum iteration testing result

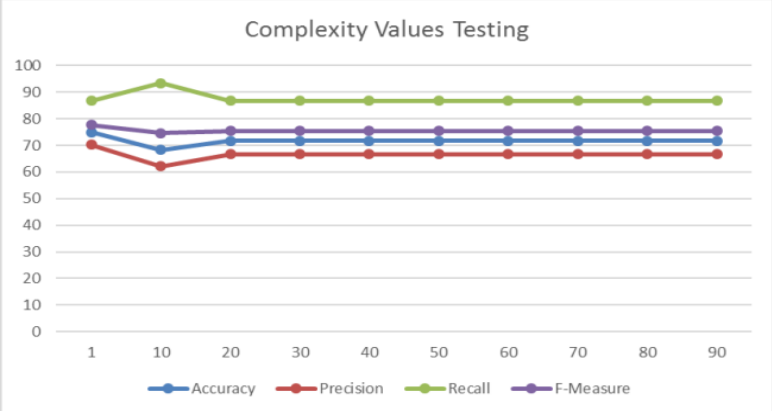

Figure 6. Graph of complexity testing result

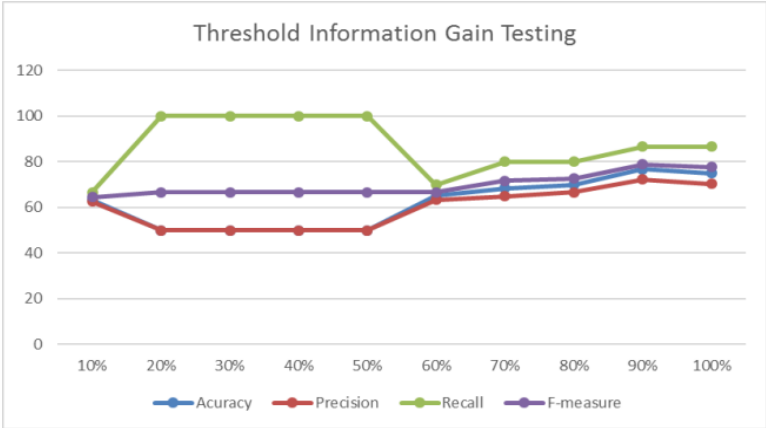

Figure 7. Graph of threshold information gain testing result

3.2. Testing of Information Gain Feature Selection Threshold

After performing the classifcation using SVM, the best SVM parameters obtained are maximum iteration = 20, λ = 0.5, γ = 0.001, ε = 0.000001, dan C = 1. Based on the graph displayed in Figure 7, the best result was obtained when 90% features were used with an f-measure of about 78.78%. This result is slightly higher than the results obtained using all of the features.

4. CONCLUSION

Based on the experiment result, it can be concluded that the cyberbullying tweet identification using SVM method and IG feature selection get a promising result. The most optimal SVM parameters obtained in this work are maximum iteration=20, λ=0.5, γ=0.001, ε=0.000001, and C=1. Meanwhile, the best threshold value of IG feature selection is 90% with accuracy of 76.66%, precision of 72.22%, recall of 86.66%, and f-measure of 78.78%. This is because the IG feature selection has a high value if it represents a particular class, and has a low value if it appears in all classes. So the results of cyberbullying tweet identification with feature selection get higher accuracy than using all the features. In the future works, several feature selections combination can also be used to improve the classification performance.

REFERENCES

[1] Nagle, M. (2008). "Casualties of bullying." *UMaine Today*, *8*(1), 2-9.

[2] Dadvar, M., Jong, F. D., Ordelman, R., & Trieschnigg, D. (2012). "*Improved cyberbullying detection using gender information.*" In Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012). University of Ghent.

[3] Rémond JJ, Kern L, Romo L. "A cyberbullying study: Analysis of cyberbullying, comorbidities and coping mechanisms". *L'Encephale*. 2015 Sep; 41(4):287-94.

[4] Nikolaou, D. (2017). "Does cyberbullying impact youth suicidal behaviors?." *Journal of health economics*, *56*, 30-46.

[5] Alavi, N., Reshetukha, T., & Prost, E. (2015). "Bullying including cyber bullying increases the risk of suicidal behaviour." *European Psychiatry*, *30*, 209.

[6] Andriansyah M, Akbar A, Ahwan A, Gilani NA, Nugraha AR, Sari RN, Senjaya R. *"Cyberbullying comment classification on Indonesian Selebgram using support vector machine method."* InInformatics and Computing (ICIC), 2017 Second International Conference on 2017 Nov 1 (pp. 1-5). IEEE.

[7] Isa SM, Ashianti L. "*Cyberbullying classification using text mining. InInformatics and Computational Sciences (ICICoS)*", 2017 1st International Conference on 2017 Nov 15 (pp. 241-246). IEEE.

[8] Reynolds, K., Kontostathis, A., & Edwards, L. (2011, December). "*Using machine learning to detect cyberbullying."* In 2011 10th International Conference on Machine learning and applications and workshops (Vol. 2, pp. 241-244).

[9] Nahar, V., Li, X., & Pang, C. (2013). "An effective approach for cyberbullying detection." *Communications in Information Science and Management Engineering*, *3*(5), 238.

[10] Domingos P. "A few useful things to know about machine learning". *Communications of the ACM*. 2012 Oct 1;55(10):78-87.

[11] Adeleke, A., Samsudin, N. A., Othman, Z. A., & Khalid, S. A. (2019). "A two-step feature selection method for quranic text classification." *Indonesian Journal of Electrical Engineering and Computer Science*, *16*(2), 730-736.

[12] Yang, Y., & Pedersen, J. O. (1997, July). "A comparative study on feature selection in text categorization." In *Icml* (Vol. 97, No. 412-420, p. 35).

[13] Uğuz, H. (2011). "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm." *Knowledge-Based Systems*, *24*(7), 1024-1032.

[14] Chandani V, Wahono RS. "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film". *Journal of Intelligent Systems*, 2015 Feb 18;1(1): 56-60.

[15] Fauzi MA, Arifin AZ, Yuniarti A. "Arabic Book Retrieval using Class and Book Index Based Term Weighting". *International Journal of Electrical and Computer Engineering (IJECE)*. 2017 Dec 1;7(6):3705-10.

[16] Fauzi MA, Utomo DC, Setiawan BD, Pramukantoro ES. "*Automatic Essay Scoring System Using N-Gram and Cosine Similarity for Gamification Based E-Learning*". In Proceedings of the International Conference on Advances in Image Processing 2017 Aug 25 (pp. 151-155). ACM.

[17] Tala FZ. "A study of stemming effects on information retrieval in Bahasa Indonesia". Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands. 2003 Jul.

[18] Fauzi MA. "Random Forest Approach for Sentiment Analysis in Indonesian Language". *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*. 2018 Oct 1;12(1).

[19] Saad, M. M., Jamil, N., & Hamzah, R. (2018). "Evaluation of Support Vector Machine and Decision Tree for Emotion Recognition of Malay Folklores." *Bulletin of Electrical Engineering and Informatics*, *7*(3), 479-486.

[20] Gaigole PC, Patil LH, Chaudhari PM. "*Preprocessing techniques in text categorization*". In National Conference on Innovative Paradigms in Engineering & Technology (NCIPET-2013) 2013 (pp. 1-3).

[21] Forman, G. (2003). "An extensive empirical study of feature selection metrics for text classification." *Journal of machine learning research*, *3*(Mar), 1289-1305.

[22] Deng H, Runger G. "*Feature selection via regularized trees*". *In Neural Networks (IJCNN)*, The 2012 International Joint Conference on 2012 Jun 10 (pp. 1-8). IEEE.

[23] Sui B. "Information gain feature selection based on feature interactions (Doctoral dissertation)."

[24] Meyer D, Wien FT. "Support vector machines". *R News*. 2001 Sep;1(3):23-6.

[25] Fauzi MA, Yuniarti A. "Ensemble Method for Indonesian Twitter Hate Speech Detection". *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*. 2018 Jul 1;11(1):294-9.