COMMENTS BULLYING DETECTION SOCIAL MEDIA WITH NEURAL NETWORK

Ladys Maopatni Sinaga 1, Ednawati Rainarli 2
[1, 2] Information Engineering - University Computer Indonesia
[1, 2] Jl. Dipatiukur 112-114 Bandung
e-mail:sinagaladys@gmail.com[1], ednawati.rainarli@email.unikom.ac.id[2]

ABSTRACT

Facebook social media is one of the social media are widely used. Indiscriminate users in the comments to be one reason for the content that is hurt (abuse) and result in bullying (harassment) in social media. Exploiting machine learning can be used to detect the comments containing bully. Prerocessing used include folding case, cleaning, convert negation, stopwords removal and tokenizing. TF-IDF is used in calculating the weighting of each term. Backpropagation Neural Network methods into methods used in this study. The test results obtained by changing the value of the epoch and learning rate got value - average Precision at 0:52, Recall with value - average of 0.40, and f-measures amounting to 0.5 where the lowest accuracy is 0:38 and the highest is 0.

Keywords : Machine learning, Facebook, bully, Neural Network Backpropagation, preprocessing, TF-IDF

1. PRELIMINARY

1.1. Background

Facebook is one of the social media are widely used in Indonesia. Social media is also often used because of the ease in which user interaction, the rapid spread of information, and freedom in response to the 'status' among users. Indiscriminate users in posts and comments be one reason for the content that is hurt (abuse) and result in bullying (harassment) in social media. To be able to detect bullying can be done by using machine learning. The use of machine learning in text processing had been done already by E Rainarli and KE Goddess in peringkasan text [1].

Research on the detection of bullying in social media has previously been done by Huascar Sanchez and Shreyas Kumar [2], In the study of social media is used twitter to detect the phrase that has the word "Gay", "Homo", "Dike", "Queer". Classification methods used are Naive Bayes and accuracy results in the study reached 70%.

Research related to text classification has been done before. Research conducted by Maulana Aziz Assuja regarding [3], indicated that based on the results of research, training, and testing done to get the value of accuracy of 78.34% accuracy and precision of 84.21%.

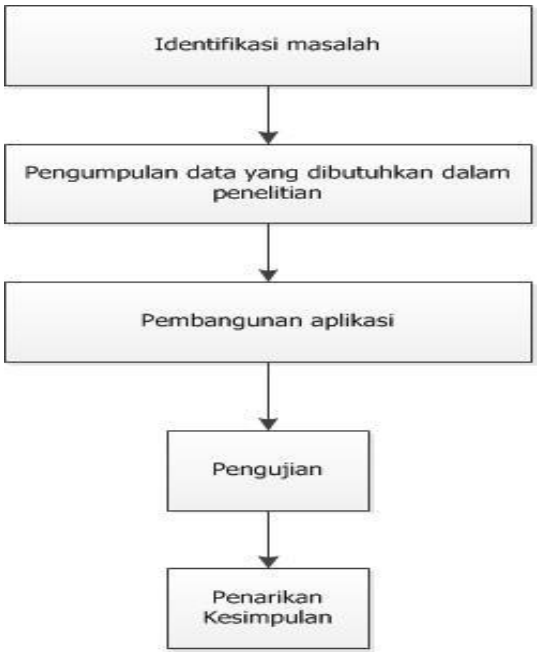
*feature*used by adding Contextual features after preprocessing process. Contextual features a contextual approach which includes a number of terms, the meaning of words, and proximity term (phrase or collocation) [4]. According to research conducted earlier by Jensen S Lee and Tony Martinez [3], suggesting that this could reduce errors as much as 33%, but to approach the level of classification of human beings, learning models should be abstracted training information and combine it with external knowledge that is not available in the data the training itself. A simple approach to represent the contextual relationship of the term is to count the number of times a term appears with another term in the contextual framework provided.

Based on what has been described above, this study will implement classification methods Neural Network Backpropagation with Contextual feature.

1.2. Research Methodology

The method used in this study is qualitative. Qualitative method is a method that is descriptive and tend to be analytical, departing from the data, utilizing the existing theory as a guide to focus on appropriate research.

The steps in this research can be seen in Figure 1



Picture1, Stages of research methodology

2. ISI RESEARCH

2.1. Harassment (Bullying)

Harassment (Bullying) is an act that is hurting someone physically or mentnal. In social media, harassment is one of the acts of cyber crime (cyber crime)

2.2. Social media

Social media is a means of digital communication that is connected online through the Internet network. Social media enable people to communicate with near and far and rapid dissemination of information. Some examples of social media is Facebook, Twitter, G-mail.

2.3. Facebook Automation

Facebook Automation is one of the tools to maximize online masrketing activities. The features found in tools that can add, delete, change or take (scraping) messages, status, comments, and information contained in the profile of individuals, community groups, events, pages like Facebook. In this study, Facebook Automation serve as a tool to extract data sets required. Results scraping stored in .CSV

Table 1 Examples of data that has been done scraping Train

Train Data	Category	Comment
P1	*Bully*	if already bad from the beginning so lost !!!! Ga bacod !!
P2	*Non Bully*	Lunox same opponents use gusion !! enough is enough !! harley
P3	*Bully*	Jirr Evos play noob udh ga weighted doing play njeeng!

2.4. process analysis

The process is done in this study included several stages that can be seen in Figure 2.

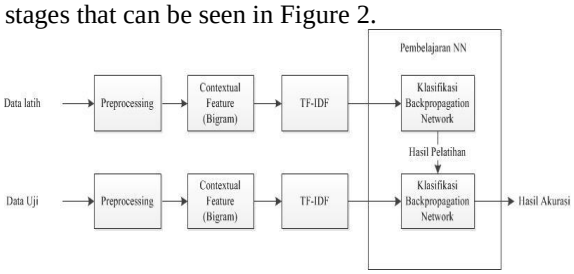


Figure 2. Stages Process Guide

The first stage of training, and then load the data set after the process is done preprocessing feature extraction using Contextual features and TF-IDF weighting calculation after Backpropahation.

2.5. *preprocessing*

*preprocessing*is to prepare the data processing unstructured data into structured data and prepared for the next process [5]. Stages - steps being taken in this research Case Folding, Cleaning, Convert negation, tokenizing, and Stopword Removal. Preprocessing block diagram is shown in Figure 3.

Figure 3. Block diagram of the preprocessing

a. Case folding

Case folding is one of the preprocessing process. This stage berjtujuan to homogenize each letter by changing all the letters in the document into lowercase (small)

Table 2 Results of Using Case Folding

Statement	Results Case Folding
if already bad from the beginning so lost !!!! Ga bacod !!	if already bad from the beginning so lost !!!! ga bacod !!
Lunox same opponents use gusion !! enough is enough !! harley	lunox same opponents use gusion !! enough is enough !! harley
Jirr Evos play noob udh ga weighted doing play njeeng!	jirr Evos play noob udh ga weighted doing play njeeng!

b. cleaning

cleaning are the steps to remove unnecessary as signs baca.Berikut cleaning results can be seen in Figure 3.

Table 3. Results Using Cleaning

Results Case Folding	Cleaning Results
if already bad from the	if already bad from the

beginning so lost !!!! ga bacod !!	beginning so lost ga bacod
lunox same opponents use gusion !! enough is enough !! harley	lunox same opponent harley use gusion enough is enough
jirr Evos play noob udh ga weighted doing play njeeng!	jirr Evos play noob udh ga weighted doing play njeeng

c. Convert negation

*Convert negation*is the stage to combine the negation word with the word thereafter. Here are the results Convert negation can be seen in Figure 4.

Table 4. Results Using Convert negation

Cleaning Results	**Results Convert negation**
if already bad from the beginning so lost ga bacod	if already bad from the beginning so lost gausah bacod
lunox same opponent harley use gusion enough is enough	lunox same opponent harley use gusion enough is enough
jirr Evos play noob udh ga weighted doing play njeeng!	jirr Evos play noob UDH gaberbobot doing play njeeng!

d. tokenizing

tokenizing is the stage to cut the string input by every word that constitute it. Tokenizing the following results can be seen in Figure 5.

Table 5. Results Using tokenize

Results Convert negation	**Results tokenize**
if already bad from the beginning so lost gausah bacod	reply already ugly from early yes lost No need bacod
lunox same opponent harley use gusion enough is enough	opponent lunox same gusion enough wear harley already enough
jirr Evos play noob UDH gaberbobot doing play njeeng	jirr Evos play noob uDH gaberbobot
	what are you doing play njeeng

e. stopword Removal

*stopword Removal*these steps to remove the words that are not descriptive. Removal Stopword following results can be seen in Figure 6.

Table 6. Results Using Stopword Removal

Results tokenizing	**Results Stopword Removal**
reply already ugly from early yes lost No need bacod	ugly early lost No need bacod
opponent lunox same gusion enough wear harley already enough	opponent lunox gusion harley
Jirr Evos play noob uDH gaberbobot what are you doing play njeeng	jirr Evos play noob gaberbobot play njeeng

2.6. Analysis of TF-IDF

TF-IDF is a weighting technique on any word or term, applying the weighted combination is multiplication local weight (tf) and global significance (idf). Idf value of a term (word) can be calculated using the

$$\text{IDF} = \log\left(\frac{D}{df}\right)$$

TF-IDF process flow is to calculate the per-*term* which can be seen in Figure 4.

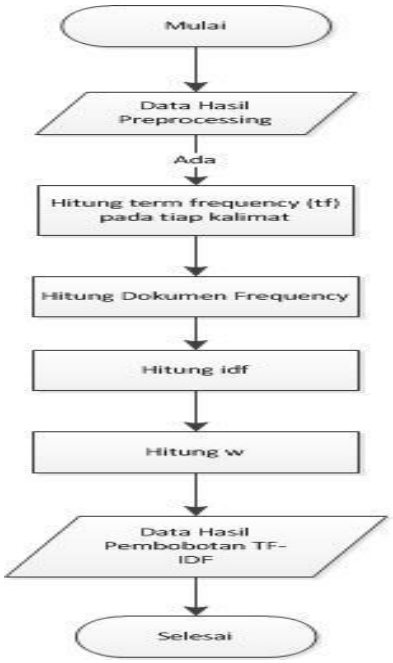


Figure 4. Flowchart of calculation TF-IDF

a. Calculating Term Frequency (tf)

b. Counting Document Frequency (df)

c. Calculating Inverse Document Frequency

IDF = log (= log () = $0.477\frac{D}{df}) \frac{3}{1}$

e. Mengiung weight / weight.

Calculating the weight (w) the word "ugly" in each statement can be calculated by the following:

WP1 = tfP1 * idf = 1 * 0.477 = 0.477

WP2 = tfp2 * idf = 1 * 0.477 = 0

WP3 = tfP3 * idf = 0 * 0.477 = 0

2.7. *Backrpopagation*

Back propagation neural network algorithms perform a two-stage calculation that advanced calculations to calculate the error between the actual output and the target, and the countdown for good propagates the error to correct the weights on all the existing neurons.

Backpropagation consists of layers including input layer (input layer), a layer consisting of either the input neurons of the first neuron to neuron network. Hidden layer (hidden layer), consist of at least one layer of hidden layer neuron from the first and so on, and the output layer (output).

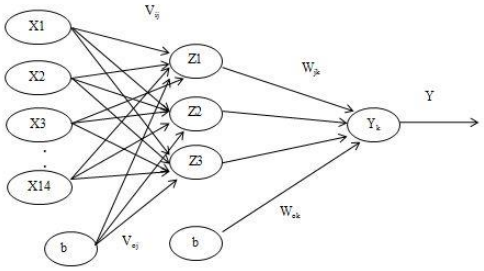


Figure 5. Architecture of Backpropagation (developed)

Information :

X : Rated Input (input layer)

Vij : The value of the connection weights on the input layer X
Unit to Z

Voj : Bias in the hidden layer unit X to Z

Z : The weight of the hidden layer (hidden layer)

Wjk :Value repair the connection to the lining
Hidden Z to Y unit

Wok: Exodus (*output*)

b : refraction

Y : Exodus (using one output)

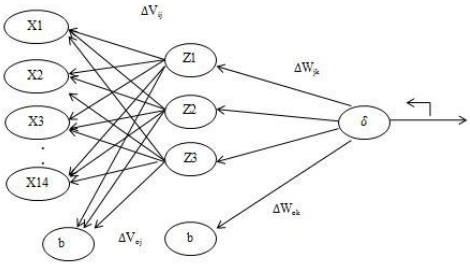


Figure 6. Architecture of Backpropagation (backward)

Information :

X : Rated Input (input layer)

ΔVij : The value of the connection weights on the input layer X
Unit to Z

ΔVoj : Bias connection on the input layer X to Unit
Z

Zi : The weight of the hidden layer (hidden *layer*)

ΔWjk: The value of the connection weight from Z to Y unit

ΔWok: Bias at the junction of Z to Y

b :Bias in the hidden layer and layer output

δ :Factors weighing the value of the connection settings
the output layer.

2.8 Backpropagation Training

At this stage there are variables coaching is determined as follows: the input layer, the number of neurons in the hidden layer, learning rate, limit the

epoch, and the iteration limit (error). Training variable in this study is based on research Maulana Aziz Assuja [2]. Here are the variables used training.

a. The number of terms are extracted with bigrams and TF-IDF (N) = 14
b. *input layer* N = 14 (+1 Bias Unit) = 15 *perceptron*
c. The number of neurons in the hidden layer N * 3 = 14 * 3 (+1 Bias Unit) = 103 *perceptron*
d. *Learning rate* = 0.1
e. Limit epoch = 100
f. Iteration limit (error) = 10%

Here is the result of improvements and the weight bias input layer to the hidden layer.

Table 7. Results Repair And Weight Bias input layer to the hidden layer.

	VI.1	VI.2	VI.3
X1	0.1 * 0.00003 * 1.87055 = 0.00019	0.1 * 0.00008 * 1.87055 = 0.000238	0.1 * 0.00007 * 1.87055 = 0.00018
X2	0.1 * 0.00003 * 1.97064 = 0.00020	0.1 * -0.00008 * 1.97064 = 0.000251	0.1 * 0.00007 * 1.97064 = 0.00018
X3	0.1 * 0.00003 * 1.82892 = 0.00019	0.1 * 0.00008 * 1.82892 = 0.000233	0.1 * 0.00007 * 1.82892 = 0.00017
refraction	0.1 * -0.00003 = 0.0001	0.1 * -0.00008 = 0.000127	0.1 * 0.00007 = 0.00009

The following is an improvement biased use and weight using the hidden layer and output layer to the input layer to the hidden layer can be seen in Table 8

Table 8 Results of corrective bias and weights hidden layer to the output layer.

	Vi.1baru	Vi.2baru	Vi.3baru
X1	0.86851 + (-0.00019) = 0.86831	0,18106+ (-0.000238) = 0.18082	0.71324 + (-0.00018) = 0.71307
X2	0.14882 + (-0.00020) = 0.14862	0.23276 + (-0.000251) = 0.23251	0.69606 + (-0.00018) = 0.69588
X3	0.49377 + (-0.00019) = 0.49358	0.65987 + (-0.000233) = 0.65963	0.46591 + (-0.00017) = 0.46574
Z	0,33131+0.00027 = 0.33194	0.70253 + (-0.000127) = 0.70320	0.56664 + (-0.00009) = 0.56726

Values above is the weight bias and bias values and new weights to be used as a further iteration

2.9 Testing Backpropagation

Testing is the phase propagation testing using methods backpropogation. At this stage only through the forward propagation process for identifying a predetermined pattern.

Figure 7. Block Diagram Testing Backpropagation

a. Results Pengujiann Model

Results of testing the model with training data and data testing 50 50. Tests carried out using the Epoch values of 100 and 600 and value of learning rate 0.1, 0:01 and 0001 test results can be seen in Table 9.

Table 9. Results of Testing Model

Epoch	*Learning rate*	accuracy	*recall*	*Precision*	*F-measure*
100	0.1	0.7	0.70	0.69	0.69
100	0:01	0.68	0.68	0.69	0.63
100	0001	0.74	0.74	0.75	0.72
600	0.1	0.72	0.72	0.73	0.69
600	0:01	0.74	0.74	0.78	0.71

b. The results of performance tests

The results of performance tests with the training data 100 and data testing scenarios 50. The results are shown in Table 10

Table 10. Performance Testing Results

Epoch	*Learning rate*	accuracy	*recall*	*Precision*	*F-measure*
100	0.1	0.62	0.62	0:38	0:47
100	0:01	0.72	0.72	0.76	0.68
100	0001	0.78	0.78	0.79	0.78
600	0.1	0.72	0.72	0.81	0.66
600	0:01	0.78	0.78	0.81	0.76

2.9 Evaluation Testing

Evaluation test on the detection of bullying text that has been done by comparing the text detection. Tests carried out calculations with equations 2.19 (recall), 2:20 equations (precision) and equations 2:21 (f-measure).

3. COVER

3.1. Conclusion

Based on the evaluation results of testing with the calculation of value - average precision, recall and F-measures in Table 4.7. Value - average gained namely Precision of 0.73 which shows the relevant word generated by the system. Recall the value - average of 0.73 which shows the words relevant to the detection of manual text so that it can be concluded value - average results text detection manually and results of text detection system with the average - average f-measures amounting to 0.70 where accuracy is the lowest was 0.62 and the highest was 0.82 ,

3.2. Suggestion

Suggestions for the development of bullying text detection in Indonesian language social media of this study are:

1. The data set used can be added to determine whether an increase in learning machine learning. Data sets can be replaced from other social media and language typo could be solved so as to minimize errors in the preprocessing stage.

2. Change the method used to compare the performance of the method in detecting bullying Indonesian text.

BIBLIOGRAPHY

[1] E Rainarli., TO Dewi Relevance Vector Machine for Summarization, IOP Conference Series, Bandung, 2018.

[2] Huascar Sanchez., Shreyas Kumar, "Twitter Bullying Detection", Journal ser. NSDI, Vol.12, UC Santa Cruz.

[3] Maulana Aziz Assuja., Saniati, "Sentiment Analysis Using Tweet *Backpropagation Neural Network*"TEKNOINFO Journal, Vol.10, No.2, ISSN: 1693-0010.

[4] Lee S. Jensen., Martinez, Tony., "Improving Text Classification by Using Conceptual and Contextual Features", Computer Science Department, Brigham Young University.

[5] Olweus, D., "Bullying at school: What we know and what we can do", New York: Blackwell, 1993.

[6] Siti Mujilahwati, "Pre-Processing Text Mining On Twitter Data", National Seminar on Information and Communication Technology, Yogyakarta, 2016.

[7] Irma Pujadayanti., Moch. Ali Fauzi., Yuita Arum Sari, "Prediction of the Auto Rating Beauty Product Reviews by Naïve Bayes Method and N-gram" Development Journal of Information Technology and Computer Science, Vol. 2, No. 11, pp 4421-4427.

[8] Dr. Suyanto, ST, M.Sc., "Data Mining for Classification and Data clustering", Information Technology, Bandung, 2017.