

INDONESIAN LANGUAGE CLASSIFICATION OF CYBERBULLYING WORDS ON TWITTER USING ADABOOST AND NEURAL NETWORK METHODS

Kristiawan Nugroho

Computerized Accounting
AMIK Jakarta Teknologi Cipta
<http://www.amikjtc.com>
kristiawan@mhs.dinus.ac.id

Abstrak

Cyberbullying merupakan topik penelitian yang sangat menarik karena berkembangnya teknologi komunikasi terutama social media yang menyebabkan akibat negatif dimana orang bisa saling melakukan bullying sehingga menimbulkan korban bahkan sampai dengan bunuh diri. Fenomena deteksi Cyberbullying telah banyak diteliti dengan menggunakan berbagai pendekatan, Pada penelitian ini digunakan metode AdaBoost dan Neural Network yang merupakan metode machine learning dalam mengklasifikasikan kata Cyberbullying dari berbagai komentar yang diambil dari Twitter. Pengujian hasil klasifikasi dengan dua metode ini menghasilkan tingkat akurasi sebesar 99.5% dengan Adaboost dan 99.8% dengan menggunakan metode Neural Network, Sedangkan jika dibandingkan dengan metode lain diperoleh hasil akurasi sebesar 99.8% dengan SVM dan Decision Tree, 99.5% dengan Rendom Forest. Berdasarkan hasil penelitian metode Neural Network,SVM dan Decision Tree merupakan metode yang teruji dalam mendeteksi kata cyberbullying terbukti dengan pencapaian tingkat akurasi tertinggi dalam penelitian ini.

Kata kunci: Cyberbullying, Machine Learning, Adaboost, Neural Network

Abstract

Cyberbullying is a very interesting research topic because of the development of communication technology, especially social media, which causes negative consequences where people can bully each other, causing victims and even suicide. The phenomenon of Cyberbullying detection has been widely researched using various approaches. In this study, the AdaBoost and Neural Network methods were used, which are machine learning methods in classifying Cyberbullying words from various comments taken from Twitter. Testing the classification results with these two methods produces an accuracy rate of 99.5% with Adaboost and 99.8% using the Neural Network method. Meanwhile, when compared to other methods, the results obtained an accuracy of 99.8% with SVM and Decision Tree, 99.5% with Random Forest. Based on the research results of the Neural Network method, SVM and Decision Tree are tested methods in detecting the word cyberbullying proven by achieving the highest level of accuracy in this study.

Keywords: Cyberbullying, Machine Learning, Adaboost, Neural Network

INTRODUCTION

Cyberbullying is a phenomenon that continues to grow along with the growth of the social media industry around the world. Forms of crime in the form of slander, harassment, insults, and dirty words often exist in various communication media via the internet. According to Dewi Cyberbullying is a form of intimidation carried out through IT technology, especially the

internet and smartphones(Dewi, Suryani, & Sriyati, 2020), Cyberbullying is done because of a sense of boredom and a sense of personal pleasure that can be done in groups(Malihah & Alfiasari, 2018).

Bullying carried out through technology in cyberspace has resulted in many losses, ranging from minor impacts such as headaches, trouble sleeping, loss of appetite (Triyono & Rimadani, 2019) as well as serious disorders such as depression and even suicide(Utami, 2014). In research in the field of computer science, various

approaches with various methods have been widely used by researchers in detecting Cyberbullying. In general, research that has been done uses the machine learning approach (Talpur & O'Sullivan, 2020) which is one of the robust approaches in classification.

Research on cyberbullying has been carried out such as research conducted with the Support Vector Machine (Purnamasari, Fauzi, Indriarti, & Dewi, 2018) which achieves an accuracy rate of 75% and research conducted by (Luqyana, Cholissodin, & Perdana, 2018) which achieves the highest accuracy rate of 90%. Another method in machine learning is also used by researchers in classifying Cyberbullying words is Naive Bayes as in research conducted by (Maulana & Ernawati, 2020) which achieved an accuracy rate of 76% as well as research conducted by (Pardede, Miftahuddin, & Kahar, 2020) which achieved an accuracy performance rate of 80%. In other studies, the researchers also used the K-Nearest Neighbor method in classifying Cyberbullying words as in research conducted by (Candra & Rozana, 2020) which results in an accuracy rate of 77%, and research conducted by (Saputra & Rosiyadi, 2019) where this research also achieved a fairly high accuracy performance of 81.32%.

Research on Cyberbullying detection on Indonesian-language Twitter data uses the AdaBoost and Neural Network methods which are also robust machine learning approaches used in Cyberbullying detection. Research on Cyberbullying detection using the Neural Network algorithm has been conducted by (Hani et al., 2019) which obtained the best accuracy performance of 92.8% higher than other methods such as SVM which obtained 90.3% accuracy results. This study uses several stages of processing the Cyberbullying dataset with 2 preprocessing techniques, namely, Select Relevant Value and Impute Missing Values which are not used in other approaches such as SVM, Naive Bayes, and K-Nearest Neighbor. This preprocessing method is useful to select relevant features and data before the classification process is carried out using Neural Network and AdaBoost which aim to detect cyberbullying words originating from social media applications such as Twitter or other applications.

This study is presented in several sections which include Introduction which is a section that aims to explain the background of the research carried out based on the previous research developments. The next section is Research Methods, which is the part that explains the stages of research which include the formation of datasets, preprocessing, classification to

measuring evaluation results. The Results and Discussion stage contains the results of measuring cyberbullying word predictions using an orange application which is presented in the form of an accuracy and confusion matrix table. The last section contains Conclusions and Suggestions which contains conclusions and suggestions from the research that has been carried out.

RESEARCH METHODS

The Proposed Method

The research stage was carried out by providing source data from the Indonesian language Cyberbullying word dataset which was then carried out by the data preprocessing process followed by the classification and evaluation stages as shown in Figure 1 as follows :

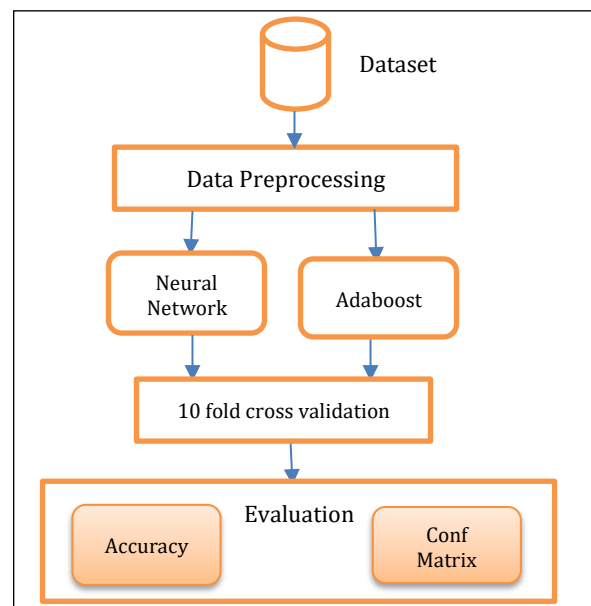


Figure 1. The Proposed Method

Dataset

This study uses a Cyberbullying dataset on social media Twitter which has been generated from research conducted by (Luqyana et al., 2018) in the form of a CSV file which is a dataset containing 400 lines of comments obtained from Twitter.

Data Preprocessing

The Cyberbullying dataset in Indonesian is then carried out by the preprocessing process by deleting rows containing empty values, selecting features to be used randomly, and normalizing each feature used in processing this dataset. This process can be seen in Figure 2 below:

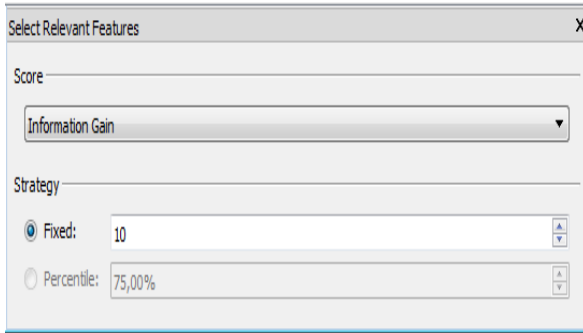


Figure 2. Selecting relevant features preprocessing

The "Select Relevant Value" stage is the process of selecting relevant features in the dataset. The scoring process uses Information Gain which is a technique that is often used in data preprocessing by using scoring to weight a feature by using maximum entropy (Maulida, Suyatno, Rahmania Hatta, & Mulwarman, 2016). The next stage of preprocessing data is Impute Missing Values as in Figure 3 as follows:

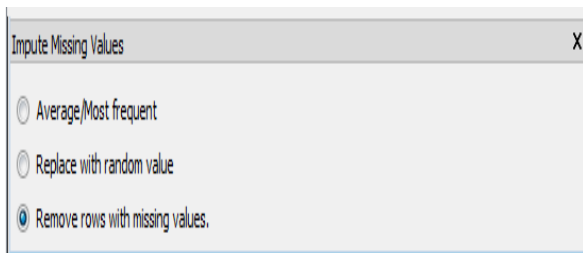


Figure 3. The preprocessing impute missing values

The data processing process in figure 3 above is a step by removing empty data so that it will simplify the next stage in processing data.

10 fold cross-validation

Cross-Validation (CV) is one of the statistical-based approaches used in evaluating a model in machine learning. 10 fold CV is a method that can be used by dividing data into 10 folds of the same size. From this process, the data will be tested by dividing 9 training data and 1 as testing data, and so on, then the performance results of the process are measured.

Classification

In this study, we used 2 machine learning methods which are known as robust methods in data mining classification, namely Neural Network and AdaBoost methods.

1. Neural Network

Neural Network or ANN is a machine learning method developed based on the way the human brain thinks. According to Muhamad Razak, ANN

consists of several parts which include the input unit and the output unit, which in each connection there is a weight that can be modified in getting the results of the predictions carried out (Razak & Riksakomara, 2017). The ANN method architecture can be seen as shown below:

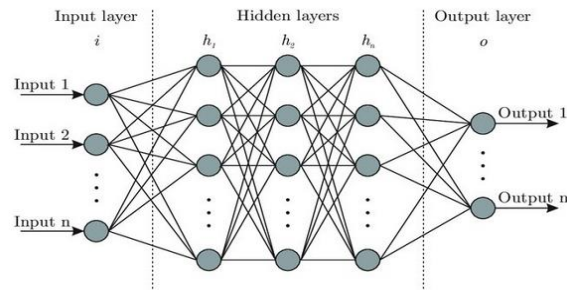


Figure 4. Neural Network Architecture

The Neural Network architecture in Figure 4 above (Bre, Gimenez, & Fachinotti, 2018) consists of several parts, namely the input layer, hidden layer, and output layer where the layers are connected in processing data in data mining.

2. Adaptive Booster (AdaBoost)

The Adaboost method or also known as Adaptive Booster includes ensemble learning which is often used in boosting techniques where the boosting process can be combined with other algorithm classifiers to improve performance in classification (Listiana & Muslim, 2017). The AdaBoost method was introduced by Yoav Freund and Robert Schapire in 1995 and is a machine learning method that is often used to improve the accuracy of a model.

Accuracy

An important step of the process carried out in data mining is the stage of evaluating the results of measuring the performance of the model. One way to measure it is by calculating accuracy. Accuracy is the total ratio of predictive calculations with true values that contain positive or negative values in a group of data. Accuracy calculations can be done using the following formula:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \dots\dots\dots (1)$$

- TP=True Positive
- TN=True Negative
- FP=False Positive
- FN=False Negative

Confusion Matrix (CM)



The CM approach is a technique that is often used in machine learning, CM is a technique for measuring the performance of a model that can produce an output in the form of two or more classes. The terms used in CM measurement are True Positive(TP), True Negative(TN), False Positive(FP), and False Negative(FN). The explanation of the confusion matrix can be seen in Figure 5 as follows:

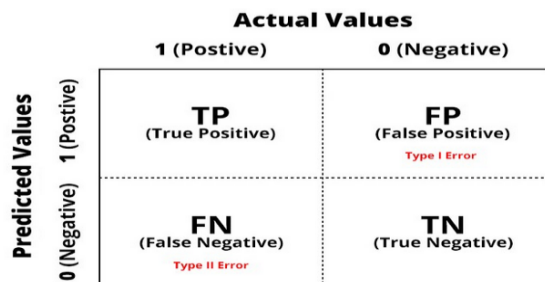


Figure 5. Confusion Matrix

RESULTS AND DISCUSSION

Orange Application

Orange is an application that can be used in analyzing data mining. This application was made by a scientist from Ljubljana University using the programming languages Python, Cython, C ++, and the C language. The orange application is designed to be user-friendly in the form of process visualization which is manifested in the form of an attractive image icon.

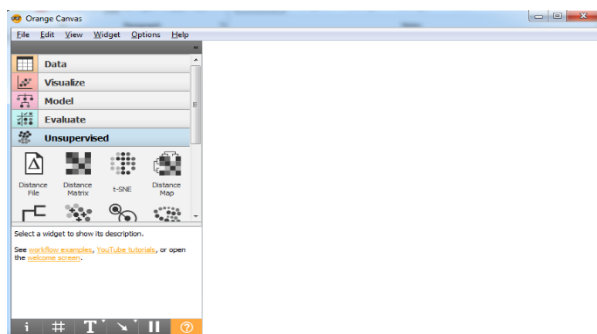


Figure 6. User Interface Orange App

The main part of this application consists of the Data menu which contains data sources that are used complete with visualization in table form. Also, this menu section provides preprocessing facilities where data can be processed with various techniques so that it will be easier to present in classification form. The next menu is Vizualize which contains a visualization of data presentation in various graphic forms. Models are part of the next menu which contains various machine learning models that can be used in processing

data. The next menu is Evaluate which contains the methods used in displaying the evaluation results of the models used. Also, the last menu presents Unsupervised which is the method used in the learning model without labels in machine learning.

Data Mining Analysis Design

The orange application is very helpful in analyzing data on Cyberbullying words that use Indonesian. In this study we used several stages starting in presenting data in files, visualizing data in tables, data preprocessing processes followed by data analysis using 2 machine learning approaches, namely Neural Networks and AdaBoost. The next stage is the evaluation process to display the results of the model performance and visualization of the analysis results in the form of a confusion matrix as shown in Figure 7 as follows :

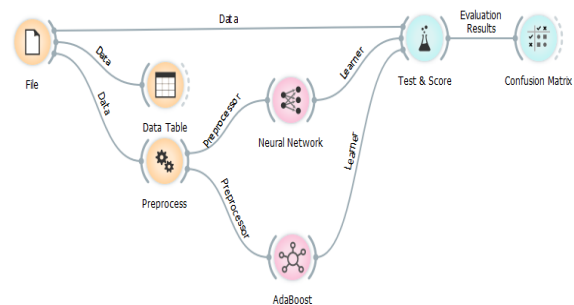


Figure 7. Data Analysis Design

File Configuration

The file arrangement to be taken as a data source is done by taking a CSV file with the name Cyberbullying-2.csv(Luqyana et al., 2018) into a load file menu, this file contains 400 words which will be given a positive class if they are 200 Cyberbullying utterances. and negative if it is not a Cyberbullying statement of 200 words.

Data Visualization

Speech data that has been loaded can then be visualized in table form so that it will look like Figure 8 below:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	id	Sentimen Instagram Comment Text																		
2	1	negative	<@SENAMAS> TOLU!! Gak ada hubungan nya keputoran dgn pake hijab sye!! yg lb bilang bany nya keparasan dikalem gak ada hubungan nya woyyyy!! Otak sama jempol lo gak sinkron																	
3	2	negative	Gedek! lo tata...mwoygt dibela2m balikan...hadew...ntar ditnggal lg nyalahin tuh cowo...padahal kitenya yg oon.																	
4	3	negative	kmm termenev2s4r lengket lg duwihh koi kaboi lgt sih mbak kya abege agih ato yg kmm cari sensasi biar top markotoppo etongk gk bertualitas																	
5	4	negative	Intinya kalau kesel dengan ATT nya gausah ke anaknya juga. Kesian buat perkembangan psikis anak kedepannya. Iu orang bener bener tolol, sng seandainya dia punya anak, terus anak yang sabayaa...mwoy Allah menjadi pembuka pintu syurga dan penghalang api neraka bagi kedua orang tuanya																	
6	200	positive	Lagu barunya mbak tata kan kisah nyatanya rumah tangga mbak tata... jadi rasanya mbak tata bener2 sayang get same suaminya... syukur kalau balikan... semoga avet...																	
7	203	positive	Yg komen kenapa si mbak ini ga seadil malah ketawa2... ya iyalah klo kita dah disakiti sm pasangan ompe berlaru2 nggakin juga sih kita seadil ampe deprei segitu... serahin sm Allah aja...																	

Figure 8. Data Vizualization

The figure above shows the Cyberbullying word data visualized in the form of a table consisting of 400 data which will then be processed in the preprocessing process so that the data is ready to be classified.

Data preprocessing is an important stage in data processing, this stage converts several data formats into a form that is ready for analysis. After the preprocessing process is carried out, the data structure will be displayed in Figure 9 as follows:

Data Preprocessing

	Sentiment	igram Comment	id
1	negative	<USERNAME> ...	1
2	negative	Geblek lo tata....	2
3	negative	Kmrn termewe...	3
4	negative	Intinya kalau ke...	4
5	negative	hadewwww p...	5
6	negative	Pantesan di tin...	6
7	negative	kebiasaan balaj...	7
8	negative	krn sebagian be...	8
9	negative	ayu janda bego...	9
10	negative	Anyiennnggg....	10
11	negative	Syarat sah nika...	11
12	negative	Cewek gatau di...	12
13	negative	<USERNAME> ...	13
14	negative	Jjrik liatnya suer...	14
15	negative	Emang bukan h...	15
16	negative	Yg jd masalah ...	16
17	negative	Anjing nya nga...	17
18	negative	<USERNAME> ...	18
19	negative	<USERNAME> ...	19
20	negative	Cuman jual go...	20
21	negative	Kasian ih nyari ...	21
22	negative	anjir tontonan ...	22
23	negative	Dlu inul ovan...	23

Figure 9. Data Preprocessing

Classification

The next stage in Cyberbullying data processing is the classification process. The results of the accuracy performance comparison can be seen in the accuracy measurement comparison table in table 1 as follows:

Table 1. Accuracy performance measurement results

Methods	Accuracy (%)
Neural Network	99.8
AdaptiveBooster (AdaBoost)	99.5

The results of measuring the performance of the model are compared using other methods, the results are also shown in Table 2 as follows:

Table 2. Accuracy measurement using other methods

Methods	Accuracy (%)
SVM	99.8
Random Forest	99.5
Decision Tree	99.8

The results of measuring the performance of the model in classifying Cyberbullying data

above indicate that the Neural Network method is an approach that produces a high level of accuracy, namely 99.8%, produces the same performance as the other two methods, namely, Support Vector Machine (SVM) and Decision Tree which has the same level of accuracy. Meanwhile, other methods such as AdaBoost and Random Forest also produce good performance by achieving an accuracy rate of 99.5%.

Confusion Matrix

The results of measuring the performance of the model using a confusion matrix can be seen in the following figures :

1. Neural Network

		Predicted		Σ
		negative	positive	
Actual	negative	200	0	200
	positive	1	199	200
Σ		201	199	400

Figure 10. Neural Network CM

Figure 10 shows that with the Neural Network method the negative and predicted negative (TN) data is 200, the negative data which is predicted to be positive (FP) is 0. Then the positive data which is predicted to be negative (FN) is 1 and the positive data which is predicted to be positive (TP) as many as 199 data.

2. Support Vector Machine(SVM)

		Predicted		Σ
		negative	positive	
Actual	negative	200	0	200
	positive	1	199	200
Σ		201	199	400

Figure 11. SVM CM

Figure 11 shows that with the SVM method the negative and predicted negative (TN) data is 200, the negative data which is predicted to be positive (FP) is 0. Then the positive data which is predicted to be negative (FN) is 1 and the positive data which is predicted to be positive (TP) as many as 199 data.

3. Decision Tree

		Predicted		Σ
		negative	positive	
Actual	negative	199	1	200
	positive	0	200	200
Σ		199	201	400

Figure 12. Decision Tree CM

The Decision Tree method in figure 12 shows that the negative and predicted negative (TN) data is 200, the negative data which is predicted to be positive (FP) is 1. Then the positive data which is predicted to be negative (FN) is 0 and the positive data which is predicted to be positive (TP) as many as 200 data.

4. Adaptive Booster (AdaBoost)

Figure 13 shows that with the AdaBoost method the negative and predicted negative (TN) data is 199, the negative data which is predicted to be positive (FP) is 1. Then the positive data which is predicted to be negative (FN) is 1 and the

positive data which is predicted to be positive (TP) as many as 199 data

		Predicted		Σ
		negative	positive	
Actual	negative	199	1	200
	positive	1	199	200
Σ		200	200	400

Figure 13. Decision Tree CM

5. Random Forest

		Predicted		Σ
		negative	positive	
Actual	negative	199	1	200
	positive	1	199	200
Σ		200	200	400

Figure 14. Random Forest CM

The Random Forest method in Figure 14 shows that the negative and predicted negative (TN) data is 199, the negative data which is predicted to be positive (FP) is 1. Then the positive data which is predicted to be negative (FN) is 1 and the positive data which is predicted to be positive (TP) as many as 199 data.

CONCLUSIONS AND SUGGESTIONS

Conclusion

Based on research on the classification of Cyberbullying words in Indonesian that has been done, The preprocessing process uses Select Relevant Value and Impute Missing Values combined with a machine learning approach make a good contribution in predicting the word Cyberbullying from the Twitter dataset. Neural Network, SVM, and Decision Tree methods produce the highest performance in the classification with an accuracy of 99.8% while the AdaBoost and Random Forest methods produce an accuracy of 99.5%.

Suggestion

In future research, the results of the Cyberbullying word classification performance need to be tested on a series of datasets that have a



large amount of data above 1000, the test results will be analyzed using various machine learning and deep learning methods. This test will provide information on the best approach that can be used in classifying Cyberbullying words with large amounts of data.

REFERENCES

- Bre, F., Gimenez, J. M., & Fachinotti, V. D. (2018). Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158(April), 1429–1441. <https://doi.org/10.1016/j.enbuild.2017.11.045>
- Candra, R. M., & Rozana, A. N. (2020). Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor. *IT Journal Research and Development*, 5(1), 45–52. Retrieved from <https://journal.uir.ac.id/index.php/ITJRD/article/view/4962>
- Dewi, H. A., Suryani, & Sriyati, A. (2020). Faktor-Faktor Yang Memengaruhi Cyberbullying Pada Remaja : A Systematic Review. *Journal of Nursing Care*, 3(2), 128–141. Retrieved from <http://jurnal.unpad.ac.id/jnc/article/view/24477>
- Hani, J., Nashaat, M., Ahmed, M., Emad, Z., Amer, E., & Mohammed, A. (2019). Social Media Cyberbullying Detection using Machine Learning. *International Journal of Advanced Computer Science and Applications(IJACSA)*, 10(5), 703–707. <https://doi.org/10.14569/IJACSA.2019.0100587>
- Listiana, E., & Muslim, M. A. (2017). Penerapan Adaboost Untuk Klasifikasi Support Vector Machine Guna Meningkatkan Akurasi Pada Diagnosa Chronic Kidney Disease. *Prosiding Seminar Nasional Teknologi Dan Informatika UMK 2017*, 875–881. Retrieved from <https://jurnal.umk.ac.id/index.php/SNA/article/view/1463>
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 4704–4713. Retrieved from <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3051>
- Malihah, Z., & Alfiasari, A. (2018). Perilaku Cyberbullying pada Remaja dan Kaitannya dengan Kontrol Diri dan Komunikasi Orang Tua. *Jurnal Ilmu Keluarga Dan Konsumen*, 11(2), 145–156. <https://doi.org/10.24156/jikk.2018.11.2.145>
- Maulana, F. A., & Ernawati, I. (2020). Analisa Sentimen Cyberbullying Di Jejaring Sosial Twitter Dengan Algoritma Naïve Bayes. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 529–538. Jakarta: Universitas Pembangunan Nasional Veteran Jakarta. Retrieved from <https://conference.upnvj.ac.id/index.php/senamika/article/view/619>
- Maulida, I., Suyatno, A., Rahmania Hatta, H., & Mulawarman, U. (2016). Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain. *Oktober 2016 Ijccs*, 17(2), 1–5.
- Pardede, J., Miftahuddin, Y., & Kahar, W. (2020). Deteksi Komentar Cyberbullying Pada Media Sosial Berbahasa Inggris Menggunakan Naïve Bayes Classification. 7(1), 46–54.
- Purnamasari, N. M. G. D., Fauzi, M. A., Indriarti, & Dewi, L. S. (2018). Identifikasi Tweet Cyberbullying pada Aplikasi Twitter menggunakan Metode Support Vector Machine (SVM) dan Information Gain (IG) sebagai Seleksi Fitur. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 5326–5332.
- Razak, A., & Riksakomara, E. (2017). Peramalan Jumlah Produksi Ikan dengan Menggunakan Backpropagation Neural Network (Studi Kasus: UPTD Pelabuhan Perikanan Banjarmasin. *Jurnal Teknik ITS*, 6(1), 142–148. <https://doi.org/10.12962/j23373539.v6i1.22129>
- Saputra, I., & Rosiyadi, D. (2019). Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp. *Faktor Exacta*, 12(2), 101–111. Retrieved from https://journal.lppmunindra.ac.id/index.php/Faktor_Exacta/article/view/4181
- Talpur, B. A., & O’Sullivan, D. (2020). Cyberbullying severity detection: A machine learning approach. *PLoS ONE*, 15(10 October), 1–19. <https://doi.org/10.1371/journal.pone.0240924>
- Triyono, T., & Rimadani, R. (2019). Dampak Cyberbullying Di Media Sosial Pada Remaja Dan Implikasinya Terhadap Pelayanan Bimbingan Dan Konseling (Studi Kasus pada Klien X di Padang Utara Kota Padang).

NeoKonseling, 1(1), 1-5. Retrieved from <http://neo.ppj.unp.ac.id/index.php/neo/article/view/96>

Utami, Y. C. (2014). Cyberbullying di Kalangan Remaja (Studi tentang Korban Cyberbullying di Kalangan Remaja di Surabaya). *Universitas Airlangga*, 3(3), 1-10. Retrieved from <http://journal.unair.ac.id/download-fullpapers-kmnts73d7a00d3dfull.pdf>

