Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12629
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparison of NB and SVM in Sentiment Analysis of Cyberbullying using Feature Selection

Selamet Riadi[1)*], Ema Utami,[2)], Ainul Yaqin[3)]
[1,2,3)]Universitas AMIKOM Yogyakarta, seleman
[1)] selametriadi@students.amikom.ac.id, [2)]ema.u@amikom.ac.id, [3)]ainulyaqin@amikom.ac.id,

Submitted : Jul 3, 2023 | **Accepted** : Jul 19, 2023 | **Published** : Oct 1, 2023

Abstract: In the past few decades, the internet has become an inseparable part of human life. It provides ease of access and permeates almost every aspect of human existence. One of the internet platforms that is widely used by people around the world is social media. Apart from being spoiled with the convenience and efficiency offered by social media to support daily life, it has gained popularity among a wide audience. This has positive implications when utilized effectively, but it cannot be denied that there are negative consequences if not utilized properly. One such consequence is the prevalence of cyberbullying activities on social media. Cyberbullying has become a major concern for the public and social media users, prompting researchers to leverage information technology in developing technologies that can identify the elements of cyberbullying, particularly on social media platforms. Sentiment analysis has been employed by researchers to identify the components of cyberbullying in online platforms. Sentiment analysis involves the application of natural language processing techniques and text analysis to identify and extract subjective information from text. This study aims to compare the Naive Bayes algorithm and the Support Vector Machine algorithm, while utilizing feature selection, specifically chi-square, to enhance the accuracy of both algorithms in classifying Instagram comments. The experimental results indicate that the Multinomial Naive Bayes (MNB) algorithm outperforms the Support Vector Machine (SVM) algorithm, achieving an accuracy of 83.85% without feature selection and 90.77% with feature selection. Meanwhile, SVM achieves an accuracy of 82.31% without feature selection and 90% with feature selection. Evaluation through the confusion matrix and classification report reveals that MNB exhibits better precision and recall rates compared to SVM in identifying bullying and non-bullying classes. The use of feature selection enhances the performance of both algorithms in classifying Instagram comments related to cyberbullying.

Keywords: Naive Bayes; Support Vector Machine; Cyberbullying; feature selection; Chi-square.

INTRODUCTION

In recent decades, the internet has become an inseparable part of human life. It provides easy access and permeates almost every aspect of human existence. One of the internet platforms that is used by nearly all people worldwide is social media. Besides being indulged with the convenience and efficiency offered by social media to support life, it has gained increasing popularity among a wide audience. This has positive implications when utilized effectively, but it cannot be denied that there are adverse consequences when it is not used properly. One such consequence is the occurrence of

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12629
e-ISSN : 2541-2019
p-ISSN : 2541-044X

cyberbullying, which is no longer unfamiliar in the realm of social media. Cyberbullying incidents have been on the rise lately, and often, perpetrators are unaware that their expressions directed at someone constitute a form of bullying that can cause psychological issues for the victims. All statements that have already been sent through social media on the victim's personal account quickly spread and are easily readable by everyone connected through social media and maintaining friendships. (Aini & Apriana, 2019). Cyberbullying can have negative impacts on the mental health of teenagers, such as depressive stress, lowered self-esteem, feelings of despair, and even suicidal thoughts. (Chu et al., 2018).

Cyberbullying has become a primary concern for the public and social media users, with information technology playing a pivotal role in the development of technologies that can identify elements of cyberbullying, particularly in social media. Sentiment analysis has been utilized by researchers to identify elements of cyberbullying on the internet and social media. Sentiment analysis is the application of natural language processing techniques and text analysis to identify and extract subjective information from text. (Hussein, 2016).

The study conducted by (Naf'an et al., 2019) utilized sentiment analysis to identify cyberbullying, aiming to develop a system that can classify comments as containing cyberbullying elements or not. The classification results are utilized to detect comments related to cyberbullying. The Naïve Bayes Classifier algorithm was employed for classification, and each comment underwent preprocessing and feature extraction using the TF-IDF method. To evaluate and test the classification performance, K-Fold Cross Validation was employed, and the experiment was divided into two parts, with and without stemming. The study utilized a training dataset consisting of 455 instances. The experimental results showed the best performance with an accuracy rate of 83.53% from the experiment involving the stemming process.

In this study, the researchers aim to compare the Naive Bayes algorithm and Support Vector Machine (SVM) algorithm while utilizing feature selection, specifically chi-square, to enhance the accuracy of both algorithms in classifying Instagram comments.

.

LITERATURE REVIEW

The study conducted by (Ardianto et al., 2020) aims to evaluate opinions and differentiate between positive and negative sentiments related to e-sports education. This is done with the purpose of obtaining valuable information from social media platforms. In this research, two classification algorithms, Naive Bayes and Support Vector Machine, are used. After comparing the performance of both algorithms, the following prediction results were obtained: The Naive Bayes algorithm, using the SMOTE method, achieved an accuracy of 70.32% and an AUC (Area Under Curve) value of 0.954. On the other hand, the Support Vector Machine algorithm, using the SMOTE method, achieved an accuracy of 66.92% and an AUC value of 0.832.

(Azhar et al., 2020) A research study was conducted, focusing on customer sentiment analysis of the KlikIndomaret application. The study employed the Naive Bayes Classifier (NB) algorithm, which was compared with the Support Vector Machine (SVM) algorithm. Additionally, Feature Selection (FS) optimization was performed using the Particle Swarm Optimization method. The results of the study showed that without FS optimization, the NB algorithm achieved an accuracy of 69.74% and an Area Under Curve (AUC) value of 0.518. On the other hand, the SVM algorithm without FS optimization achieved an accuracy of 81.21% and an AUC value of 0.896. However, when cross-validation was performed with FS optimization, the NB algorithm achieved an accuracy of 75.21% and an AUC value of 0.598, while the SVM algorithm achieved an accuracy of 81.84% and an AUC value of 0.898.

A research study was also conducted by (Sharazita Dyah Anggita & Ikmah, 2020) comparing two algorithms, namely Support Vector Machine (SVM) and Naive Bayes (NB). In their research, before the classification process with the algorithms, the researchers utilized Particle Swarm Optimization (PSO) as a feature selection technique. The test results showed that the application of PSO to the Naive Bayes algorithm resulted in a 15.11% increase in accuracy. Meanwhile, for the SVM algorithm with the use of PSO, there was a 1.74% increase in accuracy specifically for the sigmoid kernel.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12629
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

Generally, in a research study, there is a methodology that outlines the steps or flow of the research. In this study, the author presents the method or model used in Figure 1. Figure 1 illustrates the steps or flow starting from data collection, followed by preprocessing, word weighting, and then feature selection and without feature selection stages. It then proceeds to the classification stage using predetermined algorithms, and finally, an evaluation is conducted.

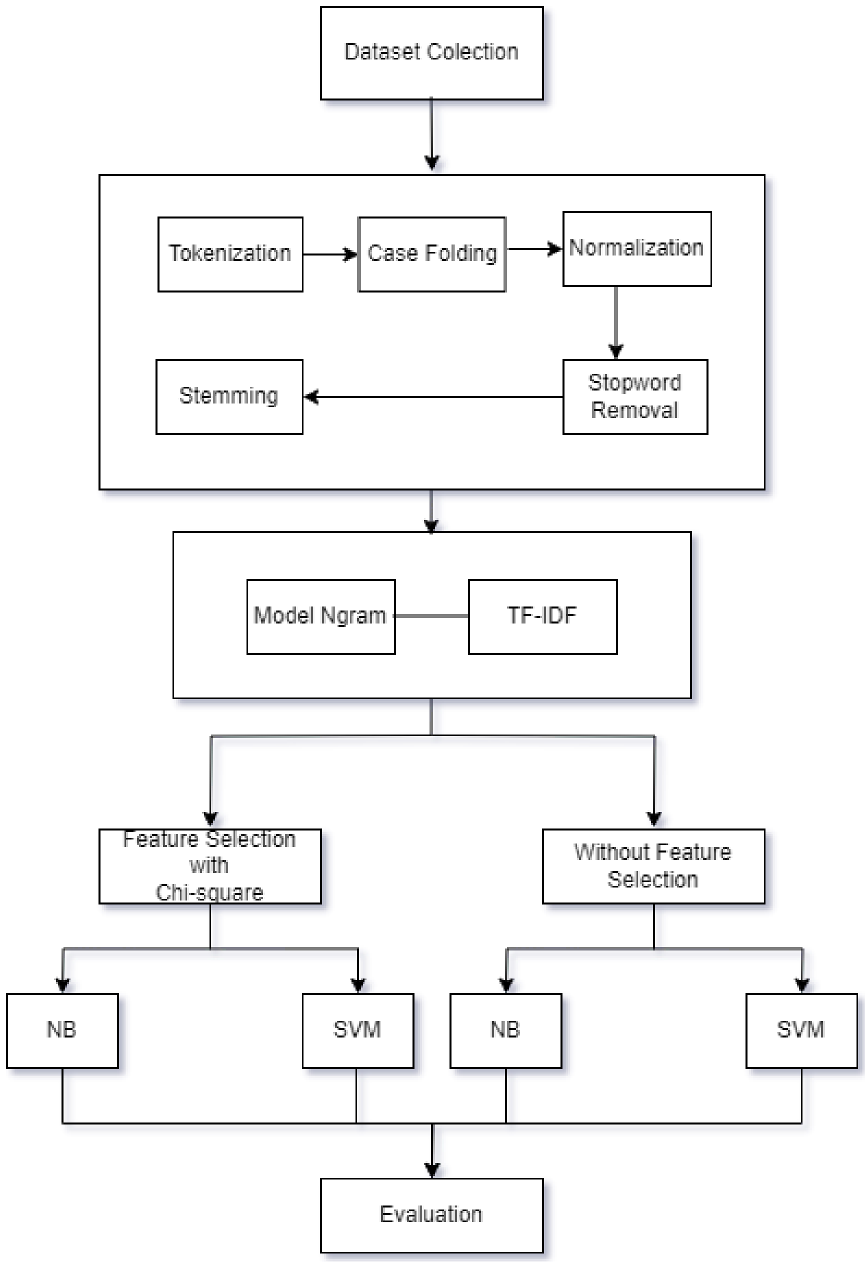


Figure 1. Research Flow

Table 1. Five Data Samples from the Dataset

No	Instagram Name	Comment	Category	Post Date	IG Account Name Artist/Celebrity
41	@maskersssssj	"Bra nya kedodoran kak make ukuran gede yaa"	Bullying	20/12/2020	@rosameldianti_
42	@yiunita	"Nah ini cantik seneng. Usahanya membuahkan hasil. Semangat ya	Non-bullying	20/12/2020	@rosameldianti_

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12629 p-ISSN : 2541-044X

		mel semoga menjadi pribadi yang lebih baik lagi"			
43	@anggrainikoenang	"mba bisa gk sih, kalau gk merem melek gitu matanya kaya orang cacingan"	Bullying	20/12/2020	@rosameldianti_
44	@nadhiya_za	"Semoga lancar terus usahanya mel"	Non-bullying	20/12/2020	@rosameldianti_
45	@riaros	"Cantik banget kalo meldi muka natural kek gini"	Non-bullying	20/12/2020	@rosameldianti_

Preprocessing

Preprocessing is the initial and most crucial stage in sentiment analysis when working with Twitter data, as it significantly affects the classification performance (Khairunnisa et al., 2021). The preprocessing process is carried out with the aim of obtaining higher-quality information from the text (Deolika et al., 2019). The preprocessing stages conducted in this study include the following:

a. Tokenization
Tokenization is the process of dividing the text document into sentences and words. This step also aims to remove numbers, punctuation marks, and spaces within the text (Prihatini, 2016).

b. Case Folding
Case folding is a data processing step that aims to convert or eliminate all capital letters in sentences within a document to lowercase letters (Merinda Lestandy et al., 2021).

c. Normalization
The normalization stage serves to transform non-standard words in social media datasets into standard base words or words that adhere to correct spelling. (Pramukti et al., 2022).

d. Stopword Removal
Stopword removal is a component of the text preprocessing step that aims to eliminate irrelevant words in a sentence based on a list of stopwords.

e. Stemming.
Stemming is the process of converting words to their base form. The goal of stemming is to produce consistent word forms (Ardiani et al., 2020). In simpler terms, stemming is used to remove word affixes in a dataset sentence.

Feature Extraction

Feature extraction is a stage of extracting or creating new features by applying functional mapping to the existing original features (Langgeni et al., 2010). In this study, feature extraction is performed using the Term Frequency Inverse Document Frequency (TF-IDF) method and the N-Gram approach with the aim of better understanding the structure and context of the text. The n-gram parameter used is the unigram parameter. Unigram is a type of token that consists of only one word (Marga et al., 2021). Next, the researcher applies the TF-IDF method to generate a feature matrix. Generally, the TF-IDF method is commonly used to determine the relationship between terms and documents or sentences. In this method, a weight or value is assigned to each word to indicate its level of relevance. (Arifin et al., 2021). The resulting feature matrix from the feature extraction process will then be utilized in the classification process with the predefined algorithm. The equation for TF-IDF is as follows:

$$tf = 0{,}5 + 0{,}5\; x \frac{tf}{max\;(tf)}$$

$$idf_t = log\left(\frac{D}{df_t}\right)$$

$$W_{d,t} = tf_{d,t}\; x\; IDF_{d,t}$$

Description :

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12629 p-ISSN : 2541-044X

tf : number of words searched in a document
$max(tf)$: the highest occurrence count of a term in the same document
Nilai *D* : document totals
df : the number of documents containing term t.
IDF : *inversed Document Frequency* (log2 (D/df))
d : the d-th document
t : the t-th word from the keyword.
W : the weight of document d towards word t.

Feature Selection

Feature selection is the process of removing less relevant features in the classification process (Maulana & Soebroto, 2019). The feature selection used in this study is Chi-square. Chi-square is one of the methods for feature classification in text classification (Luthfiana et al., 2020). Chi-square is categorized as a supervised feature method, which can remove or eliminate features without reducing the resulting accuracy level (Somantri & Apriliani, 2018). Here is the Chi-square equation :

$$X^2(t,c) = \frac{N(A \times D - B \times G)^2}{(A+B)X(C+D) \times (A+C) \times (B+D)}$$

Description:
t: term
n: total number of documents
A: number of documents in class c containing term t
B: number of documents not in class c but containing term t
C: number of documents in class c but not containing term t
D: number of documents not in class c and not containing term t

Naive Bayes

Naive Bayes is one of the popular and widely utilized classification methods. Naive Bayes is a simple method for probability-based classification that calculates a set of possibilities by summing the frequencies and combinations of values from the given dataset. (Saleh, 2015). Here is the Naive Bayes equation (Ridho Handoko & Neneng, 2021):

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Explanation:
P(Ci|X): the probability of hypothesis Ci given the fact or record X (posterior probability)
P(X|Ci): finding the parameter values that give the highest likelihood
P(Ci): prior probability of X (Prior Probability)
P(X): total probability of tuples occurring.

Support Vector Machine

Support Vector Machine (SVM) is an algorithm that performs nonlinear mapping on the original training data, allowing the data to be transformed into a higher-dimensional space. In this new dimension, SVM searches for a hyperplane that can separate the data linearly. By employing the appropriate nonlinear mapping to the higher-dimensional space, data from two classes can always be separated by this hyperplane. (Ritonga & Purwaningsih, 2018). Here is the equation for SVM (Rofiqoh et al., 2017):

$$f(x) = w.x + b$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12629
e-ISSN : 2541-2019
p-ISSN : 2541-044X

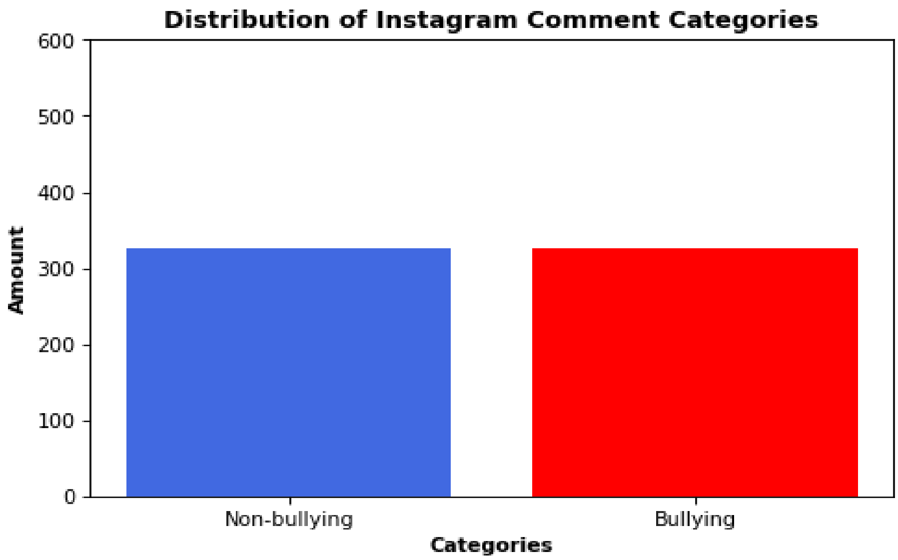
or

$$f(x) = \sum_{i=1}^{m} a_i y_i K(x_i x_i) + b$$

Description :
w : parameter of the hyperplane being sought (the line perpendicular to the hyperplane and the support vector point)
x : input data point for the support vector machine
a_i : weight value for each data point
$K(x_i x_i)$: kernel function
b : parameter of the hyperplane being sought (bias value)
In this study, a polynomial kernel is used..

RESULT

The dataset or Instagram comments used in this study consist of 650 labeled data, divided into Non-bullying and Bullying categories. The Non-bullying category contains 325 data, while the Bullying category contains 325 data. For more details, please refer to Figure 2.



Gambar 2. comment category.

Preprocessing

The preprocessing stage is performed with the aim of facilitating and optimizing the classification process or its results. For a clearer understanding, the example of the application of the entire preprocessing process in this study can be seen in Figure 3.

Tabel 7. Preprocessing process

Raw data	Preprocessing process	result
"Meldi udah dapat kontrak endorse skincare ya tapi kok mukanya tetap jelek si?"	Tokenization	Meldi udah dapat kontrak endorse skincare ya tapi kok mukanya tetap jelek si
	Case folding	meldi udah dapat kontrak endorse skincare ya tapi kok mukanya tetap jelek si

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12629
e-ISSN : 2541-2019
p-ISSN : 2541-044X

	Normalization	meldi sudah dapat kontrak endorse skincare ya tapi kok mukanya tetap jelek si
	Stopword removal	meldi kontrak endorse skincare ya mukanya jelek si
	Stemming	meldi kontrak endorse skincare ya muka jelek si

Feature extraction
Feature extraction or word weighting in this study is done using TF-IDF and N-gram. Unigram is elaborated, where the parameters in the N-gram are used to weight individual words within a sentence. Here is an example of the weighting result for one word. The 2 columns mentioned are taken from the 650 columns representing the TF-IDF results. For more details, please refer to Table 8.

Tabel 8. TF-IDF

word	Columns	Bobot
cantik	4	0.267155
	641	0.156491

Feature selection
After performing feature extraction, the next step is feature selection using Chi-square to select the best features. The scoring values are calculated for each feature computed earlier. Then, the features with the highest scores are selected. Here are the top 4 features based on the highest scores. For more details, please refer to Table 9.

Table 9. Features with the highest scores.

No	Fitur	nilai
1	anjing	19.302539
2	cantik	9.230268
3	muka	8.860719
4	babi	8.712018

Performa model Classification
The feature selection results obtained are then fed into the machine learning process for classification. In this study, the Naive Bayes (NB) and Support Vector Machine (SVM) algorithms are used. The accuracy results of the NB and SVM algorithms with and without chi-square can be seen in Table 10.

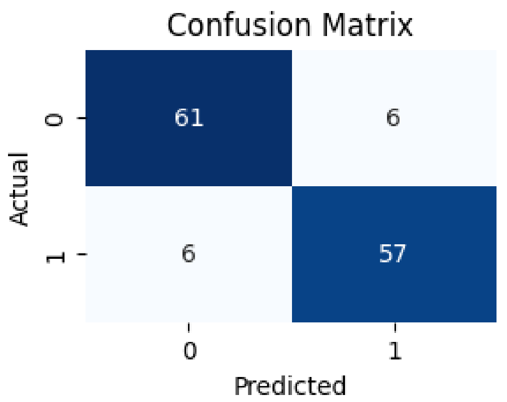
Table 10. NB Accuracy.

Algoritma	Tanpa chi-square	Dengan chi-square
	Akurasi	Akurasi
Naive Bayes (NB)	83.84615384615385 %	90.76923076923077 %
Support Vector Machine (SVM)	82.30769230769231 %	90.00000000000000 %

Next, to illustrate the performance level of each algorithm with and without using chi-square, it can be depicted using a Confusion Matrix (CM). Below is the Confusion Matrix for NB. For a clearer understanding, please refer to Figure 3 and Figure 4.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12629
e-ISSN : 2541-2019
p-ISSN : 2541-044X



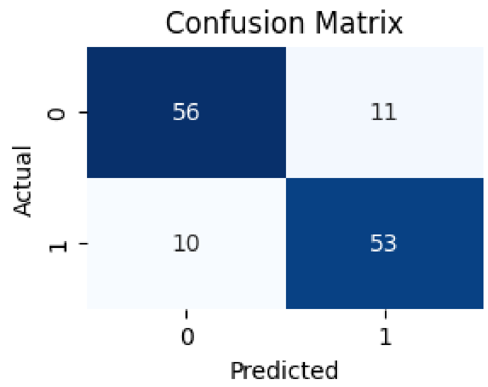
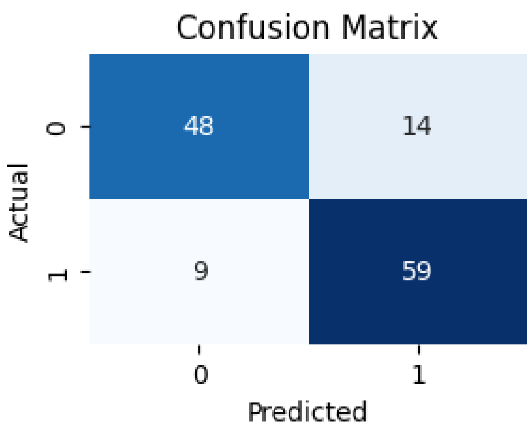


Figure 5. CM NB without chi-square

Figure 6. CM NB with chi-square

Next, the performance level with the Confusion Matrix for SVM, with or without using chi-square, can be seen in Figure 5 and Figure 6.



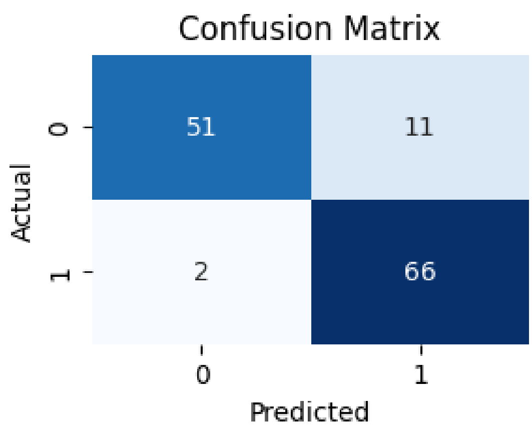


Figure 3. CM for SVM without chi-square.

Figure 4. CM for SVM with chi-square

To provide a more comprehensive understanding of the performance of the built classification models, the researcher utilized the evaluation model called classification report. Below are the results of the classification report for the algorithm models used, both with and without chi-square. Please refer to Table 11 for further details.

Tabel 11. Classification report

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12629
e-ISSN : 2541-2019
p-ISSN : 2541-044X

model	Feature Selection	Precision (Bullying)	Recall (Bullying)	F1-Score (Bullying)	Precision (Non-bullying)	Recall (Non-bullying)	F1-Score (Non-bullying)	Accuracy
NB	Without Chi-square	0.85	0.84	0.84	0.84	0.84	0.83	0.84
NB	With Chi-square	0.91	0.91	0.91	0.90	0.90	0.90	0.91
SVM	Without Chi-square	0.84	0.77	0.81	0.81	0.87	0.84	0.82
SVM	With Chi-square	0.96	0.82	0.89	0.86	0.97	0.91	0.90

DISCUSSIONS

In this study, we compared the Naive Bayes and Support Vector Machine algorithms, with and without using the feature selection technique. The specific type of Naive Bayes used was Multinomial Naive Bayes (MNB), and the Support Vector Machine utilized the "polynomial kernel" with a degree of 3. From the accuracy results, it can be observed that the MNB algorithm outperforms SVM both with and without the feature selection technique, using Chi-square. The MNB achieved an accuracy of 83.84615384615385% without feature selection and 90.76923076923077% with feature selection. On the other hand, SVM achieved an accuracy of 82.30769230769231% without feature selection and 90.00000000000000% with feature selection.

Next, we evaluated the performance of the NB and Support Vector Machine (SVM) algorithms without using feature selection. From the confusion matrix results, for the NB algorithm, we found 61 True Positives (TP), 6 False Positives (FP), 6 False Negatives (FN), and 57 True Negatives (TN). As for the SVM algorithm, there were 48 TP, 14 FP, 9 FN, and 59 TN. From these results, we can see that NB achieved a precision rate of 85%, recall rate of 84%, and F1-Score rate of 84 % for the bullying class, while for the non-bullying class, the precision was 84%, the recall was 84% and F1-Score was 83%. In SVM, the precision rate for the bullying class was 84% with a recall of 77% and F1-Score 81%, while for the non-bullying class, the precision was 81% with the recall was 87% and F1-Score was 84%.

Furthermore, we applied the feature selection technique to MNB and SVM. In NB with feature selection, we found 56 True Positives (TP), 11 False Positives (FP), 10 False Negatives (FN), and 53 True Negatives (TN). In SVM with feature selection, there were 51 TP, 11 FP, 2 FN, and 66 TN. From these results, we can observe the differences in the confusion matrix between NB and SVM with and without feature selection.

When evaluating using the classification report, we observed that NB without chi-square achieved a precision rate of 85% with recall rate of 84% and F1-Score rate of 84% for the bullying class, while for the non-bullying class, the precision was 84% with the recall was 84% and the F1-Score was 83%. When using chi-square, the precision rate of NB increased to 91% for the Bullying class and 90% for the Non-bullying class.

Meanwhile, SVM without chi-square had a precision rate of 84% with recall rate of 77% and F1-Score rate 81% for the bullying class, as well as a precision of 81% with recall of 87% and F1-Score of 84% for the non-bullying class. However, after using chi-square, the precision rate of SVM increased to 96% for the bullying class and 86% for the non-bullying class.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12629
e-ISSN : 2541-2019
p-ISSN : 2541-044X

CONCLUSION

Based on the presented results, it can be concluded that the use of feature selection techniques, such as chi-square, can improve the performance of NB and SVM classification models. Furthermore, SVM with chi-square demonstrates higher precision rates compared to NB in identifying the bullying class. However, it should be noted that a more comprehensive evaluation and proper metric selection are necessary to gain a better understanding of the performance of both algorithms.

REFERENCES

Aini, K., & Apriana, R. (2019). Dampak Cyberbullying Terhadap Depresi Pada Mahasiswa Prodi Ners. *Jurnal Keperawatan Jiwa*, *6*(2), 91. https://doi.org/10.26714/jkj.6.2.2018.91-97

Ardiani, L., Sujaini, H., & Tursina, T. (2020). Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, *8*(2), 183. https://doi.org/10.26418/justin.v8i2.36776

Ardianto, R., Rivanie, T., Alkhalifi, Y., Nugraha, F. S., & Gata, W. (2020). Sentiment Analysis on E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine. *Jurnal Ilmu Komputer Dan Informasi*, *13*(2), 109–122. https://doi.org/10.21609/jiki.v13i2.885

Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, *6*(2), 129. https://doi.org/10.30998/string.v6i2.10133

Azhar, M., Hafidz, N., Rudianto, B., & Gata, W. (2020). Marketplace Sentiment Analysis Using Naive Bayes And Support Vector Machine. *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic*, *8*(2), 91–100. https://doi.org/10.33558/piksel.v8i2.2272

Chu, X. W., Fan, C. Y., Liu, Q. Q., & Zhou, Z. K. (2018). Cyberbullying victimization and symptoms of depression and anxiety among Chinese adolescents: Examining hopelessness as a mediator and self-compassion as a moderator. *Computers in Human Behavior*, *86*, 377–386. https://doi.org/10.1016/j.chb.2018.04.039

Deolika, A., Kusrini, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*, *3*(2), 179. https://doi.org/10.36294/jurti.v3i2.1077

Hussein, D. M. E. D. M. (2016). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*, *30*(4), 330–338. https://doi.org/10.1016/j.jksues.2016.04.002

Khairunnisa, S., Adiwijaya, A., & Faraby, S. Al. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *Jurnal Media Informatika Budidarma*, *5*(2), 406. https://doi.org/10.30865/mib.v5i2.2835

Langgeni, D. P., Baizal, Z. K. A., & W, Y. F. A. (2010). Clustering Artikel Berita Berbahasa Indonesia Menggunakan Unsupervised Feature Selection. *Seminar Nasional Informatika 2010*, *2010*(semnasIF), 1–10.

Luthfiana, L., Young, J. C., & Rusli, A. (2020). Implementasi Algoritma Support Vector Machinedan Chi Squareuntuk Analisis Sentimen User FeedbackAplikasi. *Ultimatics*, *XII*(2), 125–128.

Marga, N. S., Isnain, A. R., & Alita, D. (2021). Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Bayes. *Jurnal Informatika Dan Rekayasa Perangkat Lunak (JATIKA)*, *2*(4), 453–463. http://jim.teknokrat.ac.id/index.php/informatika

Maulana, M. I., & Soebroto, A. A. (2019). *Klasifikasi Tingkat Stres Berdasarkan Tweet pada Akun Twitter menggunakan Metode Improved k-Nearest Neighbor dan Seleksi Fitur Chi- square*. *3*(7), 6662–6669.

Merinda Lestandy, Abdurrahim Abdurrahim, & Lailis Syafa'ah. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(4), 802–808. https://doi.org/10.29207/resti.v5i4.3308

Naf'an, M. Z., Bimantara, A. A., Larasati, A., Risondang, E. M., & Nugraha, N. A. S. (2019). Sentiment Analysis of Cyberbullying on Instagram User Comments. *Journal of Data Science and Its Applications*, *2*(1), 88–98. https://doi.org/10.21108/jdsa.2019.2.20

Pramukti, S. D., Nugroho, A., & Sunge, A. S. (2022). Analisis Sentimen Masyarakat Dengan Metode Naïve Bayes dan Particle Swarm Optimization. *Techno.Com*, *21*(1), 61–74.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12629
e-ISSN : 2541-2019
p-ISSN : 2541-044X

https://doi.org/10.33633/tc.v21i1.5332

Prihatini, P. M. (2016). Implementasi Ekstraksi Fitur Pada Pengolahan Dokumen Berbahasa Indonesia. *Jurnal Matrix*, *6*(3), 174–178.

Ridho Handoko, M., & Neneng. (2021). Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, *2*(1), 50–58. http://jim.teknokrat.ac.id/index.php/JTSI

Ritonga, A. S., & Purwaningsih, E. S. (2018). Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan Smaw (Shield Metal Arc Welding). *Ilmiah Edutic*, *5*(1), 17–25.

Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter dengan Metode Support Vector Machine dan Lexicon Based Features Twitter event detection View project Human Detection and Tracking View project. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1(12)*(October), 1725–1732. https://www.researchgate.net/publication/320234928

Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *CREATIVE INFORMATION TECHNOLOGY JOURNAL (CITEC JOURNAL)*, *2*. https://doi.org/10.20895/inista.v1i2.73

Sharazita Dyah Anggita, & Ikmah. (2020). Algorithm Comparation of Naive Bayes and Support Vector Machine based on Particle Swarm Optimization in Sentiment Analysis of Freight Forwarding Services. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *4*(2), 362–369. https://doi.org/10.29207/resti.v4i2.1840

Somantri, O., & Apriliani, D. (2018). Support Vector Machine Berbasis Feature Selection Untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *5*(5), 537. https://doi.org/10.25126/jtiik.201855867

*name of corresponding author