
The Detection of Bullying Against Indonesian National Team Players Using Support Vector Machine

Sunggito Oyama^{1)*}, Desty Nur Kumalasari²⁾

^{1)*,2)}Informatics Study Program, Faculty of Science and Technology, Universitas PGRI Yogyakarta, Indonesia

^{1)*}oyama@upy.ac.id, ²⁾destynurk@gmail.com

ABSTRACT

Detection is a process to check or conduct an examination of something using certain methods and techniques. Detection can be used for various problems, for example in detection bullying, especially on social media, is a significant problem with negative impacts on mental health, especially for public figures such as Indonesian National Team players. This study aims to detect bullying comments on the Instagram platform using the Support Vector Machine (SVM) algorithm. The research dataset consists of 3,100 comments collected from the official Indonesian National Team account, which are classified into bullying and non-bullying categories. The data preprocessing stages include case folding, tokenizing, normalization, removing stopwords, and stemming. The processed data was analyzed using the Term Frequency-Inverse Document Frequency (TF-IDF) method for feature weighting before being classified using SVM with a linear kernel and Naïve Bayes. The results showed that SVM performed better with an accuracy of 89%, a bullying category precision reaching 93%, and a recall of 83%. Meanwhile, the Naïve Bayes method produced an accuracy of 79%, with a bullying category precision of 76% and a recall of 86%. The non-bullying category in Naïve Bayes has higher precision (84%) but lower recall (72%). Thus, SVM is proven to be more effective in detecting negative comments due to a better balance between precision and recall. However, challenges such as informal language variations and data imbalance remain obstacles in the development of this model. This study contributes to the development of cyberbullying detection technology and supports the creation of a healthier social media environment.

Keywords: Detection; Bullying; Instagram; Indonesian National Team Players; Support Vector Machine (SVM).

1. INTRODUCTION

Bullying, both physical and psychological, has become a serious issue in society, especially among young people and athletes. In Indonesia, bullying often occurs in schools and on social media, including against national team players who are often in the public spotlight. With the increasing use of social media, such as Twitter and Instagram, a form of bullying known as cyberbullying is increasingly prevalent. Cyberbullying can have significant negative impacts on the mental and physical health of victims, making it important to identify and detect this act effectively.

Support Vector Machine (SVM) is an effective method for data classification. This method works by finding the best hyperplane that separates two classes in the data, so it can be used to classify comments or tweets that contain bullying and those that do not. Previous research has shown that SVM has high accuracy in detecting negative sentiment in text, including in the context of bullying.

For example, research by Farras Rachmanisa Noor and Dwi Agustin Nuriani Sirodj found that SVM can achieve an accuracy of up to 84.5% in analyzing tweets related to bullying issues (Noor et al., 2024). In addition, other studies also show that SVM is superior to other algorithms such as Naïve Bayes in terms of classification accuracy (Soraya, 2023).

Other research was also conducted by "Comparison of Support Vector Machine and Decision Tree Methods for Sentiment Analysis of Comment Reviews on Online Transportation .

Applications". In this study, the Support Vector Machine has a higher level of accuracy in classifying text data. In the tested study, the Support Vector Machine produced an accuracy of up to 90.20%, higher than the Decision Tree

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

method.

In another study, three classification methods were compared, namely Neural Network, Support Vector Machine, and Naive Bayes in detecting breast cancer using the Wisconsin Breast Cancer dataset. The results showed that using a Support Vector Machine (SVM) had a higher classification accuracy rate than the other two algorithms. The accuracy achieved by SVM reached 98.9%. Therefore, this study concluded that SVM is the best model for breast cancer prediction when compared to Neural Network and Naive Bayes (Wahyuningsih, 2023). Other studies also state that the advantage of SVM lies in its ability to handle non-linear and high-dimensional data. This is very relevant in diabetes analysis because clinical data is often complex and involves many variables. SVM can provide reliable and reliable results in identifying patterns and trends that may be difficult to recognize by conventional methods (Septiana et al., 2024).

Other studies also state that the advantage of SVM lies in its ability to handle non-linear and high-dimensional data. This is very relevant in diabetes analysis because clinical data is often complex and involves many variables. SVM can provide reliable and reliable results in identifying patterns and trends that may be difficult to recognize by conventional methods.

Seeing the gaps in previous studies, an approach is needed that focuses on comments on Instagram directed at Indonesian national team players. Instagram is one of the most popular platforms, with a large number of comments and a variety of informal language.

2. LITERATURE REVIEW

A. Detection

Detection is a process to check or conduct an examination of something using certain methods and techniques (Pambudi et al., 2016). Detection can be used for various problems, for example in detecting cyberbullying Instagram comments, where the system identifies comments that contain positive or negative elements. The purpose of detection is to solve a problem in various ways depending on the method applied so as to produce a solution.

B. The Concept of Bullying

The concept of bullying refers to a concept adopted from English. Bullying itself comes from the term bully which relates to the actions of someone who suppresses or disturbs a weaker individual. Several phrases in Indonesian that are often used by the public to describe the phenomenon of bullying involve acts of oppression, bullying, hazing, extortion, exclusion, or intimidation (Yuliani, 2019). Bullying is also known as the behavior of using power to hurt individuals or groups, either through words, physical actions, or psychological aspects, which impacts the victim by making them feel depressed, traumatized, and unable to respond or fight back (Siswanto et al., 2022). Bullying in educational settings is a globally recognized public health issue, with a significant impact on the mental well-being of students and teachers, as well as their physical health (Dietrich et al., 2023).

C. Indonesian National Team Players

The Indonesian National Team (Timnas) players are individuals selected to represent Indonesia in international football competitions at various levels. They are part of the national team managed by PSSI (Persatuan Sepakbola Seluruh Indonesia), the governing body of football in Indonesia. These players are selected through a rigorous selection process based on their performance, ability, and contribution to their respective clubs, domestically and internationally.

D. Instagram

Instagram is a photo-sharing application that allows users to upload photos directly from their mobile phones (Ikbal et al., 2019). Instagram also has a post-comment feature which is an advantage of Instagram. Instagram comments are limited to 2,200 characters, while Twitter is only 280 characters (Muhaddisi et al., 2021). As many as 36.4% of Instagram users in Indonesia are between 18 and 24 years old (Ashilah et al., 2021). These figures show that Instagram has an important role in Indonesia, causing an increasing number of government agencies interested in using Instagram to communicate with the public (Rodríguez, Velastequí, 2019). This is because of the widespread appeal of social media, its relatively simple use, its ability to reach a wide audience, low cost. Instagram is also used to search for information, socialize, and convey messages or comments based on their views (Graham et al., 2013).

E. Support Vector Machine

Support Vector Machine is a technique that is useful for making predictions, in classification and regression. The basic principle of SVM is to perform linear classification, this means that the classification cases can be separated in a

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

linear way. However, SVM was developed to handle non-linear problems by utilizing this basic concept in a higher-dimensional workspace (Pratama et al., 2020).

F. Python

Python is one of the best programming languages to tackle data visualization because it has the support of a very active community and has many supporting libraries, especially in the context of intelligent computing. Python is also capable of managing large amounts of data with optimal performance (Galhartlambang & Khotiah, 2021).

3. METHOD

A. Research Methodology

This research begins with the process of collecting data in the form of comments taken from Instagram social media, the data will later be labeled according to its class. Before being processed, the data will be prepared first through preprocessing with several stages, after which the data enters the weighting process using TF-IDF and classification using a Support Vector Machine. Furthermore, data imbalance or Handle Imbalance is overcome by the SMOTE technique and then divided into test data and training data for the next model classification stage using the Support Vector Machine (SVM) algorithm. The last stage is model evaluation using accuracy, precision, recall, and F1-score metrics.

B. Research Object

The object of this research is in the form of text data comments on Instagram social media that are specifically related to the Indonesian National Team players. Comments were taken from the official Indonesian National Team account which discussed the performance of the players after the Indonesian National Team match against Australia which was posted on September 10, 2024. The main focus of the research is on comments that have the potential to contain elements of bullying, such as insults, hate speech, or non-constructive criticism.

These comments were analyzed to identify patterns and classifications based on two categories: bullying and non-bullying. With this approach, the study not only aims to understand the phenomenon of bullying in general but also how Indonesian National Team players as public figures become targets of this behavior on social media.

C. Data collection

Data collection was carried out using the scraping technique to retrieve comments from the official @timnasindonesia post account discussing the performance of the players after the Indonesian National Team match against Australia in the post dated September 10, 2024. Data was collected using the IG Comments Export Tool, an extension available on the Google Chrome browser. This extension allows users to automatically export comments from Instagram posts. The results are saved in the CSV (Comma Separated Values) file format. The total data collected was 3,100 comments.

D. Labeling

The labeling process is done by grouping comments based on their content manually and automatically in labeling the data. Manual labeling is done by annotators based on predetermined bullying and non-bullying labeling guidelines, while automatic labeling is done using a lexicon-based method that refers to manual results.

Table 1. Sample bullying and non-bullying data

No	Instagram Comments	Category
1	Please Marselino do not play Formerly before the national team safely and pass the world cup.	Bullying
2	No effect, play or not play	Bullying
3	Kick marselino	Bullying
4	A lot of chance sadly	Bullying
5	Number 7 please don't do a lot of blunder, king of blunder	Bullying
6	Number 7 just like kawa-kawa, make me dizzy	Bullying
7	I agree yeah same statement Kochi if Rafael Struijk is Not yet worthy of being a senior player, he is Better to play in U23 first.	Bullying

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

8	Marcelino Cool than you 🤔	Non-bullying
9	@abellrazta Nathan match yesterday was Good, maybe because of the field too	Non-bullying
10	the defender 🔥🔥🔥🔥 very cool, the quality players	Non-bullying

E. Pre-processing

In this study, data preprocessing was carried out to change the dataset that was initially messy or unstructured into a more organized one. By processing the data through several steps such as case folding, stemming, normalization, stopwords removal, and tokenizing. This sequence of stages is expected to improve data quality to support more optimal analysis results.

Table 2. Data Preprocessing

Stages	Results
Before Preprocessing	@abellrazta Nathan match yesterday was Good, maybe because of the field too
Folding Case	Nathan's match yesterday was Good how come Possible influenced the field too?
Tokenization	[" Nathan ", "match", " yesterday ", " good ", " why ", " maybe ", " influence ", " field ", " also"]
Normalization	" why " was changed to " only ".
Stopword Removal	Results: [" Nathan ", "match", " yesterday ", " good ", " just ", " maybe ", " influence ", " field ", " also"]
Stemming	[" Nathan ", "match", " good ", " influence ", " field "]

F. TF-IDF

In this study, TF-IDF is applied after the preprocessing stage to give weight to each word based on how often the word appears in the data. So that the analysis becomes more accurate.

G. Testing

The results of the accuracy test with the SVM method are displayed in the form of a Confusion Matrix. The following is a Confusion Matrix table for detecting bullying against Indonesian national team players on Instagram data, which consists of two categories, namely bullying and non-bullying.

H. System Design

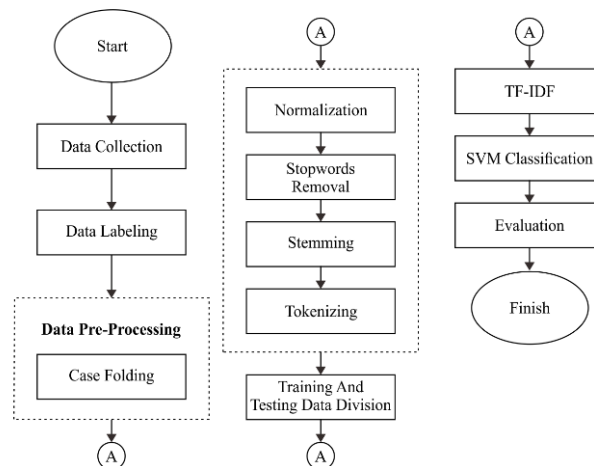


Fig. 1 Design System

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

The first stage begins by collecting data such as text from social media such as Instagram or other documents. After the data is collected, data labeling is carried out, namely giving bullying and non-bullying categories to form a dataset that will be used in the classification process. Furthermore, the data goes through a preprocessing stage to make it cleaner and ready to be analyzed. This stage includes case folding to change all letters to lowercase for consistency. Then stopword removal is carried out to remove unimportant common words, normalization to equate non-standard words to standard, tokenizing to break text into words, and stemming to return words to their basic form.

After preprocessing is complete, the dataset is divided into training data and testing data, where training data is used to train the model and testing data is used to test its performance. Furthermore, the TF-IDF method is used to assign weights to each word based on how often the word appears in the document compared to the entire dataset. These TF-IDF weights are used as input features in the classification stage using SVM (Support Vector Machine), which is tasked with predicting data categories. After the model has been trained, an evaluation is carried out to measure the model's performance using metrics such as accuracy, precision, or recall. This stage aims to ensure that the model works well and provides accurate results. This process is then ended after the evaluation is complete and the analysis results are obtained.

4. RESULT

A. Splitting Data

In the process of splitting divide the data into training data (80%) and test data (20%) using the function `train_test_split`. The "stemming" column is used as features and the "label" column as the target. The result is four variables: `x_train`, `x_test`, `y_train`, and `y_test` are readily used for model training and testing.

B. TF-IDF

In TF-IDF we use `TfidfVectorizer` from `sklearn` to convert text into numeric vectors, to measure the importance of a word in a document and how rare it is across the corpus. The `tvec` object can be used to process text into features ready for machine learning models.

C. Support Vector Machine

This process builds an SVM model with a linear kernel using Pipeline, which combines `TfidfVectorizer` to convert text to vectors and `SVC` for classification. The model is trained with training data `x_train` and `y_train` using the `fit()` method.

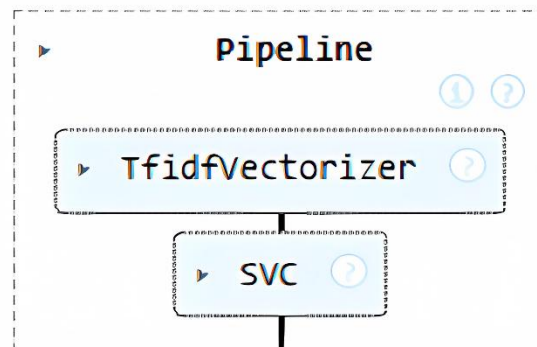


Fig. 2 Support Vector Machine Model

D. Classification Report

Testing using the confusion matrix is done to predict labels on the test data `x_test` using the trained model, then evaluate the results with `classification_report` to produce metrics such as accuracy, precision, recall, and F1-score, which are then printed.

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

```
Classification report (SVC):
              precision    recall  f1-score   support

   bullying      0.93      0.83      0.88       283
 non-bullying    0.86      0.94      0.90       325

 accuracy              0.89       608
 macro avg           0.90      0.89      0.89       608
 weighted avg       0.89      0.89      0.89       608
```

Fig. 3 Classification Report Results

E. Confusion Matrix

Process confusion matrix used To evaluate the model, comparing predictions with the original labels. The matrix is then visualized as a heatmap using **Seaborn**, showing the number of correct and incorrect predictions with clear labels.

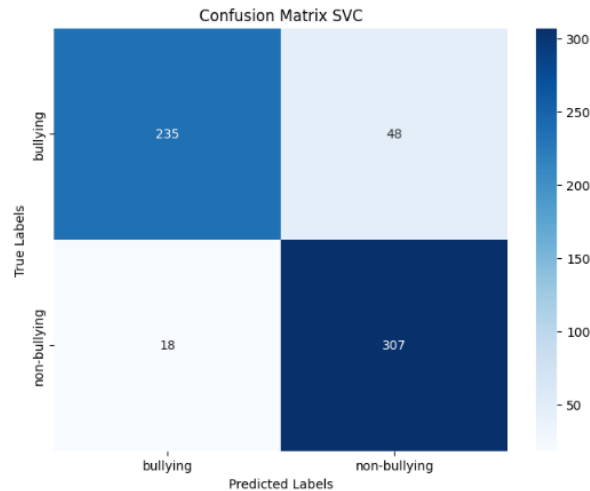


Fig. 4 Confusion Matrix Results

F. Testing Teks Input

This process trains an SVM model with training data x_{train} and y_{train} , then creates a `classify_text` function to classify the input text using the model. This function predicts the category of the given text and stores the results in a dictionary. Two sample texts are tested: "Marcelino is ugly" and "Semangat timnas", and the classification results are printed for each model used.

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

```
✓ [35] #@title **Testing**
0s input_text = "marcelino played badly"

results = classify_text(input_text)

print("Word Input:", input_text)

print("\nClassification Results:")

for model, prediction in results.items():
    print(f"{model}: {prediction}")
```

Word Input: marcelino played badly

Classification Results:
Support Vector Machine: bullying

Fig. 5 Testing Results 1

The program receives the input "marcelino played badly" and analyzes it to determine whether it contains elements of bullying or not. Based on the classification results, the text is identified as "bullying," which means that the program detects elements of bullying in the given sentence. This shows that the SVM model used has been trained to recognize language patterns that contain negative or detrimental elements.

```
✓ [35] #@title **Testing**
0s input_text = "national team spirit"

results = classify_text(input_text)

print("Word Input:", input_text)

print("\nClassification Results:")

for model, prediction in results.items():
    print(f"{model}: {prediction}")
```

Word Input: national team spirit

Classification Results:
Support Vector Machine: non-bullying

Fig. 6 Testing Results 2

The program receives the input "national team spirit" and analyzes it to determine whether it contains elements of bullying or not. The classification results show that the text falls into the category of "non-bullying," which means that the program detects that the sentence is positive or neutral, with no indication of bullying. This shows that the SVM model used is able to distinguish between texts that contain negative elements and texts that are supportive or positive.

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

5. DISCUSSIONS

The test results in this study indicate that the Support Vector Machine (SVM) method delivers high performance in detecting bullying comments directed at Indonesian national football players on Instagram. This is evidenced by an accuracy of 89%, a precision of 93% for the bullying category, and a recall of 83%. When compared to the Naïve Bayes algorithm also used in this study, SVM demonstrated superior performance, as Naïve Bayes only achieved an accuracy of 79%. These findings suggest that SVM-based modeling is more suitable for analyzing Instagram comment data, which often features informal and unstructured language.

In terms of methodology, preprocessing steps such as case folding, normalization, stopword removal, stemming, and tokenizing successfully prepared the data for feature weighting using TF-IDF. These processes significantly contributed to the improvement of classification accuracy. Compared to other models, SVM has the advantage of handling imbalanced data effectively, especially after the application of the SMOTE technique. As a result, this modeling approach provides an effective solution for addressing real-world challenges, such as detecting negative comments automatically in large volumes, which would be impractical to monitor manually.

Furthermore, the study found that bullying comments often contain complex language variations, ranging from sarcasm and offensive words to subtle hate speech. This highlights the need for more advanced approaches in future work, such as integrating SVM with more sophisticated natural language processing techniques. Overall, these findings contribute significantly to the development of cyberbullying detection technology in Indonesia and promote a healthier digital environment, especially for public figures like players of the Indonesian national football team.

6. CONCLUSION

Based on the results of this study, the Support Vector Machine (SVM) algorithm is proven to be able to detect bullying and non-bullying comments with good performance, achieving an accuracy of 89%. Data preprocessing processes, such as case folding, tokenizing, normalization, stopword removal, and weighting using TF-IDF, play an important role in improving the quality of the analysis.

The model showed adequate precision and recall for both categories, bullying, and non-bullying, with F1 scores of 0.88 and 0.90, respectively. The use of the SVM algorithm with a linear kernel and TF-IDF weighting proved effective in classifying Indonesian text comments, despite challenges such as data imbalance and informal language variations. This study provides a significant contribution to supporting the automatic detection of bullying comments on social media, especially on the Instagram platform. Meanwhile, the Naïve Bayes method produces 79% accuracy, with a bullying category precision of 76% and a recall of 86%. The non-bullying category in Naïve Bayes has a higher precision (84%) but a lower recall (72%).

7. REFERENCES

- Ashilah, K., Digital, M., Sosial, M., & Media, S. (2021). *INSTAGRAM SEBAGAI MEDIA PEMASARAN DIGITAL WASABI SUSHI JAPANESE FOOD JEMBER* Khittah Ashilah Universitas Islam Jember. 1(September), 149–160.
- Dietrich, L., Jurkowski, S., Schwarzer, N., & Zimmermann, D. (2023). The role of teachers in the bullying involvement of students with emotional and behavioral difficulties. *Teaching and Teacher Education*, 135(August), 104311. <https://doi.org/10.1016/j.tate.2023.104311>
- Galahartlambang, Y., & Khotiah, T. (2021). *Visualisasi Data Dari Dataset COVID-19 Menggunakan Pemrograman Python*. 4902(x).
- Graham, M., Avery, E. J., & Ph, D. (2013). *Government Public Relations and Social Media : An Analysis of the Perceptions and Trends of Social Media Use at the Local Government Level*. 7(4), 1–21.
- Ikbal, M., Ghofur, M. A., Studi, P., Komunikasi, I., Tribhuwana, U., & Malang, T. (2019). *EFEKTIVITAS UNGGAHAN FOTO DI AKUN INSTAGRAM*. 8(1), 52–64.
- Muhaddisi, A., Prastowo, B. N., Utami, D., & Putri, K. (2021). *Sentiment Analysis With Sarcasm Detection On Politician 's Instagram*. 15(4), 349–358.
- Noor, F. R., Agustin, D., & Sirodj, N. (2024). *Penerapan Support Vector Machine untuk Klasifikasi Opini Masyarakat Terhadap Isu Bullying*. 57–66.
- Pambudi, A. R., Informatika, T., Komputer, F. I., Karawang, U. S., Segmentasi, M., Based, R., Contour, A., Tepi, D.,

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

- Metode, M., & Dengan, C. (2016). *Deteksi keaslian uang kertas berdasarkan watermark dengan pengolahan citra digital*. 69–74.
- Pratama, F., Nasir, M., & Sauda, S. (2020). *Implementasi Metode Klasifikasi Dengan Algoritma Support Vector Machine Untuk Menentukan Stok Persediaan Barang Pada Koperasi Karyawan Pangan Utama*. 1(2), 71–81.
- Purwokerto, U. A., & Kunci, K. (2021). *PERBANDINGAN METODE SUPPORT VECTOR MACHINE DAN DECISION TREE UNTUK ANALISIS SENTIMEN REVIEW KOMENTAR PADA APLIKASI TRANSPORTASI ONLINE* *Abstraksi Keywords : Pendahuluan Tinjauan Pustaka*. 2(2).
- Rodríguez, Velastequí, M. (2019). *PERANAN NEW MEDIA DALAM TRANSFORMASI FUNGSI KOMUNIKATOR DAN FUNGSI KONSTRUKSI (DRAMATURGY AKUN INSTAGRAM WALIKOTA DENPASAR RAI MANTRA TAHUN 2019)*. 1–23.
- Septiana, T., Muda, M. A., Budiyanto, D., Pratama, M., & Putra, W. (2024). *Analisis Penggunaan Support Vector Machine pada Deteksi Dini Penyakit Diabetes Melitus*. 4(3), 1631–1640.
- Siswanto, S., Mar'ah, Z., Sabir, A. S. D., Hidayat, T., Adhel, F. A., & Amni, W. S. (2022). The Sentiment Analysis Using Naïve Bayes with Lexicon-Based Feature on TikTok Application. *Jurnal Varian*, 6(1), 89–96. <https://doi.org/10.30812/varian.v6i1.2205>
- Soraya, A. (2023). *Klasifikasi sentimen pada opinion cyberbullying di twitter menggunakan metode support vector machine dan naïve bayes*.
- Wahyuningsih, N. (2023). *Perbandingan metode klasifikasi dalam analisis sentimen masyarakat terhadap identitas kependudukan digital (ikd)*. 8(4), 1218–1227.
- Yuliani, N. (2019). *Fenomena kasus bullying di sekolah*.

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0
International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)