# Bully Comments Classification on TikTok Using Support Vector Machine and Chi-Square Feature Selection

Amelia Putri [a,1*], Abdiansah [b,2], Alvi Syahrini Utami [b,3]

[a,b] Informatics Engineering, Faculty of Computer Science, Universitas Sriwijaya, Indonesia
[1] amellptrr@gmail.com*; [2] abdiansah@unsri.ac.id; [3] alvisyahriniutani@ilkom.unsri.ac.id
* corresponding author

ARTICLE INFO

ABSTRACT

TikTok has been named the world's most popular social media platform. The high level of TikTok use makes it easier for an irresponsible user to do unethical things such as spreading hateful comments on someone's account. TikTok developers can prevent bullying by using policies such as word detection and filtering features that indicate comments fall under the category of bullying or non-bullying comments. Therefore, we conducted this study to classify bullying comments using Machine Learning methods for convenience purposes on TikTok usage, a method that we used in this research is the SVM method to classify the data and Chi-Square as the feature selection. Tests were carried out using the Linear, Polynomial, and RBF kernel functions with the C parameter, namely 0,1, 1, and 10 for each kernel. The results of this research show that the Support Vector Machine method with Chi-Square Feature Selection has a better performance. This was proven by the increased accuracy in RBF kernel C=0,1 which was 0,20

## 1. Introduction

Social Media has become an important part of the day-to-day life of most people. Specifically, it has become an essential tool to interact with each other, for information, and entertainment. TikTok has been named the world's most popular social media platform. TikTok allows its users to create videos with a duration of no more than 60 seconds that can bring many features, such as adding music, changing voices, and adding effects on the face. Based on a survey conducted by Statista, TikTok has more than 1 billion monthly active users as of the last quarter of 2022, with a total of 10 million active users in Indonesia. [1] The high number of users who are not accompanied by an understanding of the ethics of socializing through social media leads irresponsible people to cyberbully [2]

Several methods have been conducted to classify bully comments, two of them which are Naïve Bayes (NBC) and Support Vector Machine (SVM). Previous research from [3] compared classification methods to classify bully behavior on Twitter, and it was found that the SVM algorithm produced better accuracy of 99.6% while the Naïve Bayes Classifier produced an accuracy of 97.99%.

The second research from [4], feature selection with Chi-Square for news classification reveals that Chi-Square is a feature selection that can eliminate many irrelevant features without reducing the level of accuracy to optimize the performance of the classification method. Therefore, this research will use the SVM method with three different kernels, including Linear, Polynomial, and RBF, and the C parameter, namely 0,1, 1, and 10 for each kernel. Besides, we also use Chi-Square as a feature selection.

This study aims to classify bully comments on TikTok using the SVM method with Chi-Square so the results can be used as a reference for related research and for TikTok developers in minimizing the appearance of bully comments in TikTok.

## 2. Literature Study

### A. Text Preprocessing

Text Preprocessing is one of the fundamental processes in the field of NLP to prepare the data so that the system can read it correctly and achieve optimal classification results. [5] The goals steps are as follows:

1. Cleaning aims to clean documents of unnecessary characters
2. Case Folding aims to replace uppercase letters with lowercase letters
3. Tokenizing aims to break sentences into words
4. Normalizations, aim to replace slang words with standard words
5. Filtering aims to remove punctuation and blank characters
6. Stopword Removal aims to eliminate meaningless frequently used words
7. Stemming aims to replace the base word of each formed word

### B. Support Vector Machine

Support Vector Machine is an algorithm classification that can classify data non-linearly. SVM works by defining the limit between classes through the creation of a hyperplane with a maximum margin. [6] Margin is the distance between the hyperplane and the nearest data of each class. The general hyperplane function is shown in Equation 1.

$$w_i . x_i + b = 0 \tag{1}$$

If $w_i . x_i + b = +1$ is a supporting hyperplane of class +1 (positive) and the equation $w_i . x_i + b = -1$ is a supporting hyperplane of class -1 (negative) then the margin can be calculated by finding the distance of the two hyperplanes supporting the two classes. The margin value is defined as follows.

$$\frac{2}{\|w\|} \tag{2}$$

To determine the best hyperplane, the Quadratic Programming (QP) problem formula is used, which is a form of optimization using the inverse of the following equation.

$$\min \frac{1}{2} \|w\|^2 \tag{3}$$

Where $(y_i(wx_i) + b - 1 \geq 0$ and the limitations functions can be written as follows.

$$\sum_{i=1}^{n} a_i [y_i (wx_i) + b - 1] \tag{4}$$

Optimization of Quadratic Programming problems can be solved with the Lagrange Multiplier method defined through the following function.

$$L(\overrightarrow{w}, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{n} a_i (y_i(\overrightarrow{x_i}.\overrightarrow{w} + b) - 1)) \tag{5}$$

With $a_i$ is the value of Lagrange multipliers with a value of zero or positive ($a_i \geq 0$). The optimum value of Lagrange multipliers can be calculated by minimizing the variables $\overrightarrow{w}$ and b and maximizing the $a_i$ variables. Thus the equations can be written as follows.

$$\sum_{i=1}^{n} (a_i y_i x_i) = 0 \tag{6}$$

The Lagrange function in (5) is converted into (7) by substituting (6) into (5)

$$L(a) = \sum_{i=1}^{n} = a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j \overrightarrow{x_i}.\overrightarrow{x_j} \tag{7}$$

The determination of the best separation of fields is written as follows.

$$\max (L(a)) = \sum_{i=1}^{n} = a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j \overrightarrow{x_i}.\overrightarrow{x_j} \tag{8}$$

Based on the function of the obstacle, equation (8) becomes,

$$w = \sum_{i=1}^{n}(a_i\, y_i x_i) \qquad (9)$$

So the value of b is obtained as follows,

$$b = -\frac{1}{2}\,\widehat{w}.\,[x_r + x_s] \quad (10)$$

Where $x_r$ and $x_s$ are support vectors that comply with each class. The decision function obtained based on (9) and (10) is written as follows.

$$f(x) = sgn\,(\widehat{w}.\,x_i + b) \qquad (11)$$

## C. Chi-Square

Chi-Square is used to measure the value of relationships or relationships between one variable and another. This technique measures the distribution of dependency values between features and classes. The calculation for the Chi-Square is shown in the following equation.

$$X^2(t,c) = \frac{N\,(\,AxD - BxC)^2}{(A+B)x(C+D)x(A+C)x(B+D)} \qquad (12)$$

Where t is the term searched for in class c, N is the number of training data, A is the number of documents in class c that contain term t, B is the number of documents, not class c but contains term t, C is the number of documents in class c that does not contain term t, and D is the number of documents that are not class c and do not contain term t.

## D. Evaluation

Evaluation is a process to measure the performance of the method being tested. In measuring the performance of the classification process, four standard calculations can be used, namely accuracy, precision, recall, and f-measure.

1. Accuracy is a comparison between the data that is clarified correctly and the overall data

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision is the ratio of the amount of data of the correctly classified positive category to the total data classified as positive

$$Precision = \frac{TP}{TP + FP}$$

3. Recall, the success rate of the software in separating similar documents into the same classification

$$Recall = \frac{TP}{TP + FN}$$

4. F-measure, the average ratio of Precision and Recall values

$$F - measure = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall}$$

The parameters in these calculations are obtained from the Confusion Matrix table. A confusion Matrix is a method used to review classification results obtained by calculating the frequency of truth.

**Table 1.** Confusion Matrix

| Actual Class | | Prediction | |
|---|---|---|---|
| | | **Negative** | **Positive** |
| | Negative | TN | FP |
| | Positive | FN | TP |

Where TN (*True Negatives*) shows the value obtained when the model can classify negative data into negative, TP (*True Positives*) shows the value obtained when positive data is classified as positive, FP (*False Positives*) shows the value obtained when the model classifies negative data into positive, and FN (*False Negatives)* shows the value that obtained when positive data is classified to negative.

## 3. Research Method

### A. Data Collection

The type of data used in this study is primary data in the form of Indonesian comment data collected from several TikTok accounts with as many as 1,000 data comments. The data is divided into 800 training data and 200 test data.

The raw data that has been collected will then go through a labeling process carried out manually by experts. This classification consists of 2 classes with 500 bully comments and 500 non-bully comments.

### B. Framework

Based on the research steps that have been described, to simplify the explanation of the process carried out, the process will be described in a framework as shown in Figure 1.
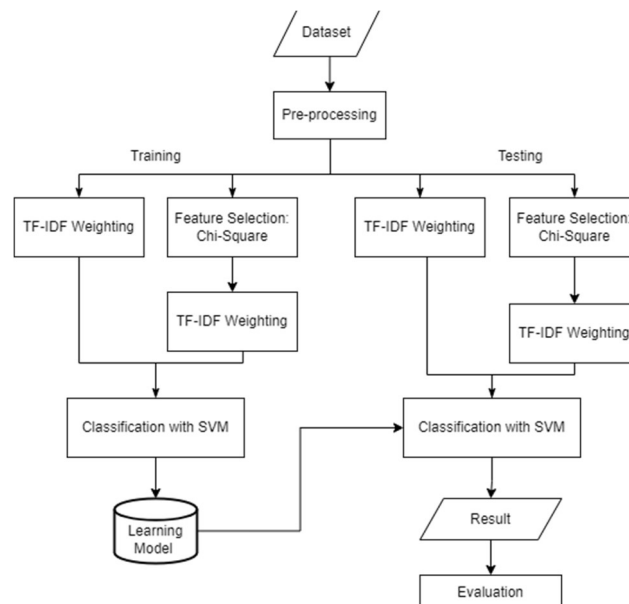


**Fig. 1.** Framework

Based on Figure 1, starting from text pre-processing, which consists of cleaning which removes noisy data, case folding which includes changing capital letters to lowercase, and stopword removal which includes removing unnecessary words that occur commonly across all the documents in the

stopword list. and filtering which includes removing punctuation and empty characters. Then splitting the data which divides into training data and testing data with a split of 80% and 20%.

After performing the preprocessing, the TF-IDF weighting procedure is performed, which is useful for converting words to numbers. Next, the TF-IDF weighted data is classified using SVM, thereby creating a training model. After that, the final results are evaluated in the form of a confusion matrix including accuracy, precision, recall, and f1 score.

## 4. Result and Discussion

In this research, we conducted two scenarios. In the first scenario, we compared several SVM kernels, including linear, polynomial, and RBF, using no Chi-Square feature selection. The second scenario is the same as the first scenario, however, implemented Chi-square as feature selection. Based on these two experiments, the highest accuracy value and another confusion matrix such as recall, precision, and f-measure will be obtained.

A. The result of testing

The result of testing based on the process that has been done, obtained a confusion matrix as shown in Table 2.

**Table 2.** Result of Confusion Matrix

| Algorithm | Kernel | C Parameters | TP | FP | TN | FN |
|---|---|---|---|---|---|---|
| SVM | Linear | 0,1 | 45 | 8 | 42 | 5 |
| | Polynomial | 1 | 42 | 4 | 46 | 8 |
| | RBF | 10 | 42 | 4 | 46 | 8 |
| SVM+CS | Linear | 0,1 | 44 | 3 | 47 | 6 |
| | Polynomial | 1 | 44 | 3 | 47 | 6 |
| | RBF | 10 | 44 | 3 | 47 | 6 |

Table 2 displays the results of the Confusion Matrix calculations for the best model of the SVM and SVM+CS classification procedures for each kernel and its varying parameters. Furthermore, the Confusion Matrix results from the best fold will be used to calculate the evaluation value. The results are compared to the comparison table on each kernel.

**Table 3.** Evaluation Data from Linear Kernel

| Algorithm | C Parameters | Kernel = Linear | | | | |
|---|---|---|---|---|---|---|
| | | Accuracy | Precision | Recall | F-1 | Time |
| SVM | 0,1 | 0,89 | 0,90 | 0,87 | 0,88 | 0,64 detik |
| | 1 | 0,88 | 0,90 | 0,83 | 0,86 | 0,55 detik |
| | 10 | 0,88 | 0,90 | 0,83 | 086 | 0,53 detik |
| SVM+CS | 0,1 | 0,91 | 0,94 | 0,88 | 0,91 | 0,14 detik |
| | 1 | 0,91 | 0,92 | 0,88 | 0,90 | 0,13 detik |
| | 10 | 0,91 | 0,93 | 0,88 | 0,90 | 0,16 detik |

**Table 4.** Evaluation Data from Polynomial Kernel

| Algorithm | C Parameters | Kernel = Polynomial | | | | |
|---|---|---|---|---|---|---|
| | | Accuracy | Precision | Recall | F-1 | Time |
| SVM | 0,1 | 0,91 | 0,93 | 0,89 | 0,91 | 1,20 detik |
| | 1 | 0,91 | 0,91 | 0,89 | 0,90 | 1,21 detik |
| | 10 | 0,89 | 0,91 | 0,85 | 0,88 | 1,02 detik |
| SVM+CS | 0,1 | 0,93 | 0,95 | 0,90 | 0,92 | 0,22 detik |
| | 1 | 0,91 | 0,94 | 0,88 | 0,90 | 0,16 detik |
| | 10 | 0,90 | 0,91 | 0,88 | 0,89 | 0,14 detik |

**Table 5.** Evaluation Data from RBF Kernel

| Algorithm | C Parameters | Kernel = RBF | | | | |
|---|---|---|---|---|---|---|
| | | Accuracy | Precision | Recall | F-1 | Time |
| SVM | 0,1 | 0,61 | 0,68 | 0,78 | 0,61 | 1,56 detik |
| | 1 | 0,92 | 0,94 | 0,88 | 0,91 | 1,44 detik |
| | 10 | 0,90 | 0,91 | 0,88 | 0,90 | 1,45 detik |
| SVM+CS | 0,1 | 0,81 | 0,74 | 0,94 | 0,82 | 0,37 detik |
| | 1 | 0,93 | 0,94 | 0,92 | 0,93 | 0,27 detik |
| | 10 | 0,92 | 0,92 | 0,90 | 0,91 | 0,33 detik |

The table above is a comparison of the results of the evaluation of the SVM method classification which is divided into 3 kernels, namely linear, polynomial, and Rbf. The graph is a visualization to explain the influence of the parameters used.
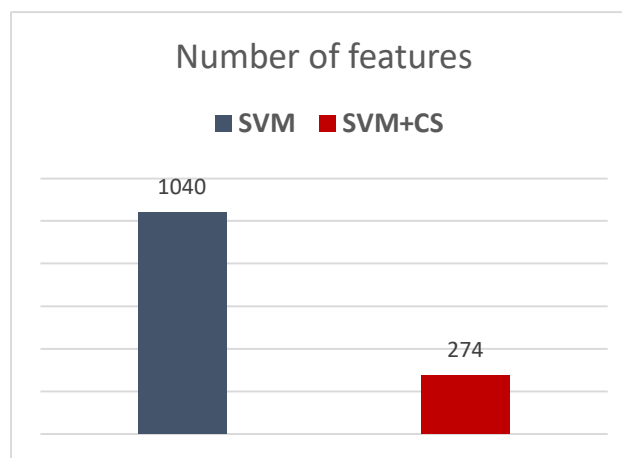


**Fig 2.** Number of Features

Figure 2 shows that in the absence of a selection process, the number of features in the data contains 1040 features. Data using the feature selection process consists of only 274 data features. This implies that chi-square feature selection can eliminate features that are judged to be unrelated to data, improve performance, and decrease the processing time of Support Vector Machine methods.

## 5.  Conclusion

This research successfully proved that the classification of bully comments on the TikTok application using the SVM method with Chi-Square Feature Selection results in better performance compared to using the SVM method. The Chi-Square Selection feature can reduce the performance of the SVM method. The Chi-Square approach can minimize the number of irrelevant features in the

dataset, improving accuracy, precision, recall, and f-measure values while reducing the computational time of the SVM classification process. Testing with RBF kernels with C=0.1 yielding a value of = 0.20 was the most substantial improvement in accuracy seen.

The results show that the use of parameter C affects the performance of each method in the SVM kernel. For linear and polynomial kernels, using C=0.1 gives the best performance, but using RBF kernels with C=0.1 gives the worst performance. This implies that the large size of the C value affects kernel performance. When a large C value is used, a small C value results in a minimum marginal value large enough to allow the resulting hyperplane to ignore outliers in the data.

## References

[1]     Clement, J. *Most popular categories on TikTok worldwide 2020, by hashtag views*. Retrieved from https://www.statista.com/statistics/1130988/most-popular-categories-tiktok-worldwide-hashtagviews. 2020.

[2]     Hidayah, A., Marcelawati, Y., Saputra, H. "*Cyber Harassment*: Fenomena Hate Comment Di Era Pandemi Covid-19 Pada Akun Tik-Tok." doi:10.31629/jmm.v5i1.3419 (2021). 10-17.

[3]     Nur Chamidah & Reiza Sahawaly. "Comparison Support Vector Machine and Naive Bayes Methods for Classifying Cyberbullying in Twitter". *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, (2021). 4-6.

[4]     Rahmad, A. N. "Pemilihan Feature Dengan Chi Square Dalam Algoritma Naive Bayes Untuk Klasifikasi Berita.". 2015.

[5]     Nadhia, S. A., Danang, T., dan Kemas, M.L. "Toxic Comment Classification on Social Media Using *Support Vector Machine* and *Chi-Square Feature Selection*". *Intl. Journal on ICT*. 2021.

[6]     Fakhri T, Said Al Faraby, & Mahendra, D. P. "Klasifikasi Teks Multi Label pada Hadis Terjemahan Bahasa Indonesia Menggunakan *Chi-Square* dan SVM". *Jurnal Tugas Akhir Fakultas Informatika*, (2021). 1-3.

[7]     Ade Clinton, S., Wanayumini, & Zakarias, S. "Analisis Kinerja *Support Vector Machine* dalam Mengidentifikasi Komentar Perundungan pada Jejaring Sosial." *Jurnal Media Informatika BudiDarma*, (2021). 475-478.

[8]     M. Fortunatusa, P. A. "Combining textual features to detect cyber*bully*ing in social media posts. *24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Procedia Computer Science", 176*, (2020). 612-6 [9]Mohkammad, R. T. "Analisis Sentimen Review Transportasi Menggunakan Algoritma *Support Vector Machine* Berbasis *Chi-Square"*. *Smart Comp Vol.9*, (2020). 36-38.

[10]    Rohmawati, U., Sihwi, S., & Cahyani, D. "SEMAR: An Interface for Indonesian Hate Speech Detection Using Machine Learning. *International Seminar On Research Of Information Technology And Intelligent (ISRITI)*. "doi:10.1109/isriti.2018.8864484. 2018.

[11]    Ramadlani, M. F. "SELEKSI FITUR MENGGUNAKAN ALGORITMA *CHI-SQUARE* UNTUK PREDIKSI CACAT PERANGKAT LUNAK". 2019.

[12]    Luque, A., Carrasco, A., Martin, A. & de las Heras, A. "*The impact of class imbalance in classification performance metrics based on the binary Confusion Matrix", 91*, (2019). 216-231.

[13]    M. U. Hasan, S. Ullah, M. J. Khan, & K. Khurshid. "Comparative Analysis of SVM, ANN and CNN for Classifying Vegetation Species Using Hyperspectral Thermal Infrared Data" . *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W13*, 1861-1868. doi:https://doi.org/10.5194/isprs-archives-XLII-2-W13-1861-2019.