JOSEIT JOURNAL OF SYSTEMS ENGINEERING AND INFORMATION TECHNOLOGY
http://jurnal.iaii.or.id/index.php/JOSEIT

Classification of Bullying Comments on YouTube Streamer Comment Sections Using Naïve Bayes Classification

Ahlida Nikmatul H[1*], Didih Rizki C[2], Christian S.K. Aditya[3]
[123]Universitas Muhammadiyah Malang, Malang, Indonesia
Email: [1*]ahlida27@gmail.com, [2]didihkrizki@umm.ac.id, [3]christianskaditya@umm.ac.id

Abstract

One of the social media crimes that is rampant in the current era is cyberbullying. Cyberbullying is a form of intimidation by someone to harass other people using technological devices. this research uses a design for information decision making that aims to get the expected results. the data collection process is carried out manually with a time frame of 1 week by watching the live broadcast of the online game YouTube streamer then sorting out some bullying and non-bullying comments in the comment's column. Data labeling is done manually. The data obtained amounted to 1000 with 500 negative comments and 500 positive comments. The above test can be concluded that from the distribution of test data there are 90% - 10% have results that are superior to the results of other tests with an increase of 4% in the Naïve Bayes weighting Gain Ratio method. Based on the test data, the results of precision, recall, F1-score and accuracy of the Naïve Bayes classification method are obtained. The test analysis above can be concluded that from the distribution of test data, 90% - 10% have results that are superior to other test results with a 4% increase in the Naïve Bayes weighting Gain Ratio method. The existence of increased accuracy results is due to a randomized data processing process.

Keywords: Cyberbullying, Mobile Legends Bang-Bang, Naive Bayes, Gain Ratio

1. Introduction

One of the social media crimes that is rampant in the current era is cyberbullying. Cyberbullying is a form of intimidation by someone to harass another person using technological devices[1]. The perpetrators of cyberbullying attack the victim's mentality and mind, so it is not uncommon for victims to become depressed and lack confidence, even suicidal. In general, the perpetrators of cyberbullying actions are carried out on YouTube live streaming, it is not uncommon for these actions to attack someone's profession, the profession is an eSport athlete[2].

Bullying and harassment have become pervasive problems in the online world, and social media platforms such as YouTube are not immune to this phenomenon[3]. YouTube comment sections, in particular, have become a breeding ground for negative and abusive comments, which can have a significant impact on the mental health and well-being of streamers and their audiences. As such, there is a pressing need to develop effective tools for identifying and addressing bullying behavior in online communities[4].

This research paper focuses on the problem of classifying bullying comments on YouTube streamer comment sections using Naïve Bayes classification. Naïve Bayes is a simple but powerful classification algorithm that is widely used in text classification tasks[5], [6]. The algorithm works by assuming that the features (in this case, the words used in the comments) are independent of each other, and then calculating the probability that a comment belongs to a particular class (bullying or non-bullying) based on the occurrence of these features[7]–[9].

The goal of this research is to develop a Naïve Bayes classifier that can accurately detect and classify bullying comments on YouTube streamer comment sections[10]. To achieve this, we will collect a dataset of comments from various YouTube streamer comment sections, and manually annotate each comment as either bullying or non-bullying. We will then use this dataset to train and evaluate the Naïve Bayes classifier, using various metrics such as accuracy, precision, recall, and F1-score.

The results of this research will have important implications for addressing bullying behavior on YouTube and other online platforms. By developing an effective and accurate classifier for bullying comments, we can help streamers and their audiences to identify and address this harmful behavior, and create a more positive and supportive online community.

2. Method

At the method stage, this research uses a design for information decision making that aims to get the expected results. The stages of this research method design are contained in Figure 1 which includes data collection, preprocessing, TF-IDF, Gain Ratio, Naïve Bayes, and model evaluation.

Accepted by editor: 30-03-2023 | Final revision: 31-03-2023 | Online publication: 31-03-2023

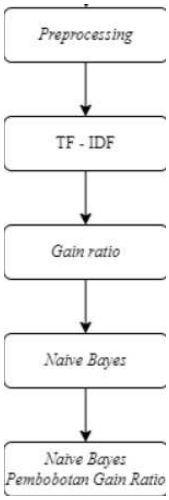

Figure 1. Research Design

2.1 Data Collection

In the data collection process, it is done manually with a time frame of 1 week by watching live broadcasts of online game YouTube streamers and then sorting out some bullying and non-bullying comments in the comment's column. Data labeling is done manually. The data obtained amounted to 1000 with 500 negative comments and 500 positive comments. The data sources obtained are some of the Mobile Legends YouTube channels: Bang Bang Official, Pascol Kintil, XINNN and several existing online game sites. In the process of retrieving comment data, 1000 comment data is obtained, data labeling is carried out using the Lexicon Based method on negative 500 comment data and positive 500 comments. The data obtained will be stored in the form of an excel file and converted into a CSV file.

2.2 Preprocessing

Preprocessing is the initial stage that prepares data before use with the aim of obtaining structured and clear data processes so as to produce information of good quality and ready to use.

a. Case Folding = At this stage, make changes to uppercase words equalized to lowercase letters.
b. Tokenizing = At this stage, remove punctuation marks and symbols so that weighting can be done. An example of tokenizing changes below.
c. Stopword Removal = At this stage, stopword removal is responsible for removing words that are not important in the document, such as conjunctions, pronouns and so on.
d. Steamming = At this stage, the process of converting affixed words into basic words is carried out.

2.3 TF-IDF

After doing the preprocessing stage, we can continue to do the TF-IDF stage to get the weight value for each document [6]. The following is the flow of explanation of the TF-IDF stage:

a. Prepare the data results that have been processed at the preprocessing stage.
b. TF (Term Frequency) process is carried out on each word contained in the document.
c. Performing the IDF (Inverse Document Frequency) stage to calculate the availability of terms in the document.
d. Then perform the calculation of the TF-IDF value.
e. Output data from the TF-IDF calculation.

The results of the TF-IDF stage process get data that already has a weight value for each word contained in the document. After doing the TF-IDF stage, the next process is the gain ratio.

2.4 Gain Ratio

After the preprocessing stage is carried out, then the Gain Ratio feature selection is carried out. Gain Ratio is an attribute weighting technique that is generally used in the Decision Trees method. This feature selection has a big effect on the prediction process and this feature also selects based on the Gain Ratio value ranking. So, the greater the value of the gain ratio can be a big influence on the prediction process.

The following is the flow of Gain Ratio weighting:

DOI: https://doi.org/10.29207/joseit.v2i1.5016

a. Calculate the Entropy value of each feature contained in each document.
b. Calculating the Information Gain value for each feature contained in each document.
c. Calculating the Split Information value of each feature contained in each document.
d. Calculating the Gain Ratio value of each feature contained in each document.
e. Sort the position of the value contained in the Gain Ratio and collect features that have an influence on the prediction results.

2.5 Naïve Bayes

After performing the preprocessing stage by weighting the Gain Ratio. Then the classification is carried out using the Naïve Bayes method [9]. The following is the flow of Naïve Bayes classification:

a. After the preprocessing stage process is carried out using the Gain Ratio weighting stage.
b. Calculate the probability value contained in the training data.
c. After calculating the Naïve Bayes probability, it is stored for testing on test data.
d. Next, calculate the probability value of the sentiment data to be tested.

3. Result and Discussion

This discussion explains the results of the application of the design that has been obtained in the previous chapter and the testing of the research that has been made using the method and phase according to the discussion in the previous chapter. In this test evaluation, we compare by splitting the data 3 times, namely 90% - 10%, 80% - 20% and 70% - 30%. From these 3 tests, we can find out which test is the best for classification.

Based on the test data, we get the results of precision, recall, F1-score and accuracy of the Naïve Bayes classification method. In the first test using 90% - 10% split data, the confusion matrix obtained positive precision results of 82% and negative precision of 78%, positive recall of 77% and negative recall of 83%, positive F1-score of 79% and negative F1-score of 81%, and accuracy of 80%. In the second test using split data 80% - 20%, the confusion matrix obtained positive precision results of 81% and negative precision of 77%, positive recall of 76% and negative recall of 82%, positive F1-score of 78% and negative F1-score of 80%, and accuracy of 79%.

In the third test using split data 70% - 30%, the confusion matrix obtained positive precision results of 77% and negative precision of 74%, positive recall of 72% and negative recall of 78%, positive F1-score of 74% and negative F1-score of 76%, and accuracy of 75%.

Table 1. Naïve Bayes Test Results

	Ratio	*Precision%*		*Recall %*		FI-score %		Accuracy %
		Positive	Negative	Positive	Negative	Positive	Negative	
Naïve Bayes	90%-10%	82%	78%	77%	83%	79%	81%	80%
	80%-20%	81%	77%	76%	82%	78%	80%	79%
	70%-30%	77%	74%	72%	78%	74%	76%	75%

Based on the test data, the results of precision, recall, F1-score and accuracy of the Naïve Bayes classification method are obtained. In the first test using 90% - 10% split data, the confusion matrix obtained positive precision results of 86% and negative precision of 82%, positive recall of 81% and negative recall of 87%, positive F1-score of 83% and negative F1-score of 84%, and accuracy of 84%.

In the second test using 80% - 20% split data, the confusion matrix obtained positive precision results of 83% and negative precision of 81%, positive recall of 80% and negative recall of 86%, positive F1-score of 82% and negative F1-score of 84%, and accuracy of 83%. In the third test using split data 70% - 30%, the confusion matrix obtained positive precision results of 79% and negative precision of 76%, positive recall of 74% and negative recall of 80%, positive F1-score of 76% and negative F1-score of 78%, and accuracy of 77%.

Table 2. Naïve Bayes Weighting Gain Ratio Test Results

	Ratio	*Precision %*		*Recall %*		FI-score %		Accuracy %
		Positive	Negative	Positive	Negative	Positive	Negative	
Naïve Bayes Weighting Gain Ratio	90%-10%	86%	82%	81%	87%	83%	84%	84%
	80%-20%	83%	81%	80%	86%	82%	84%	83%
	70%-30%	79%	76%	74%	80%	76%	78%	77%

From the test analysis above, it can be concluded that from the distribution of test data, 90% - 10% have results that are superior to other test results with a 4% increase in the Naïve Bayes weighting Gain Ratio method. The existence of increased accuracy results is due to a randomized data processing process.

DOI: https://doi.org/10.29207/joseit.v2i1.5016

4. Conclusion

In this study, which discusses the occurrence of bullying in the YouTube streamer column using the Naïve Bayes method with Gain Ratio weighting has tested 3 times. From these tests, using the Naïve Bayes method obtained an accuracy result of 80%, then the Naïve Bayes method with Gain Ratio weighting obtained a result of 84%, there was an increase in accuracy of 4%. With this research that Gain Ratio weighting is an effective method to be able to optimize accuracy results with a combination of the Naïve Bayes method.

References

[1] C. Zhu, S. Huang, R. Evans, and W. Zhang, "Cyberbullying Among Adolescents and Children: A Comprehensive Review of the Global Situation, Risk Factors, and Preventive Measures," *Frontiers in Public Health*, vol. 9. 2021. doi: 10.3389/fpubh.2021.634909.

[2] M. J. Wang, K. Yogeeswaran, N. P. Andrews, D. R. Hawi, and C. G. Sibley, "How Common Is Cyberbullying among Adults? Exploring Gender, Ethnic, and Age Differences in the Prevalence of Cyberbullying," *Cyberpsychol Behav Soc Netw*, vol. 22, no. 11, 2019, doi: 10.1089/cyber.2019.0146.

[3] P. Strickland and J. Dent, "Online harassment and cyber bullying," *House of Commons*, no. 07967, 2017.

[4] M. Weinstein, M. R. Jensen, and B. M. Tynes, "Victimized in many ways: Online and offline bullying/harassment and perceived racial discrimination in diverse racial–ethnic minority adolescents.," *Cultur Divers Ethnic Minor Psychol*, vol. 27, no. 3, 2021, doi: 10.1037/cdp0000436.

[5] Samsir *et al.*, "Naives Bayes Algorithm for Twitter Sentiment Analysis," *J Phys Conf Ser*, vol. 1933, no. 1, p. 012019, Jun. 2021, doi: 10.1088/1742-6596/1933/1/012019.

[6] D. Irmayani, F. Edi, J. M. Harahap, and ..., "Naives Bayes Algorithm for Twitter Sentiment Analysis," *Journal of Physics …*, 2021, [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1933/1/012019/meta

[7] C. A. P. Dita, P. Chairunisyah, and M. Mesran, "Penerapan Naive Bayesian Classifier Dalam Penyeleksian Beasiswa PPA," *Journal of Computer System and Informatics (JoSYC)*, vol. 2, no. 2, pp. 194–198, 2021.

[8] Y. Findawati, I. R. I. Astutik, A. S. Fitroni, I. Indrawati, and N. Yuniasih, "Comparative analysis of Naïve Bayes, K Nearest Neighbor and C.45 method in weather forecast," *J Phys Conf Ser*, vol. 1402, p. 066046, Dec. 2019, doi: 10.1088/1742-6596/1402/6/066046.

[9] K. U. Santoshi, S. S. Bhavya, Y. B. Sri, and B. Venkateswarlu, "Twitter Spam Detection Using Naïve Bayes Classifier," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, IEEE, Jan. 2021, pp. 773–777. doi: 10.1109/ICICT50816.2021.9358579.

[10] C. Zhang, G.-R. Xue, Y. Yu, and H. Zha, "Web-scale classification with naive bayes," *Proceedings of the 18th international conference on World wide web - WWW '09*, p. 1083, 2009, doi: 10.1145/1526709.1526867.