Jurnal Teknik Informatika (JUTIF)
Vol. 6, No. 1, February 2025, pp. 23-30

DOI: https://doi.org/10.52436/1.jutif.2025.6.1.4032
p-ISSN: 2723-3863
e-ISSN: 2723-3871

TEXT CLASSIFICATION OF BULLYING REPORTS USING NLP AND RANDOM FOREST.

Dasril Aldo*[1], Adanti Wido Paramadini[2], M. Yoka Fathoni[3]

[1]Informatics Engineering, Faculty of Informatics, Telkom University, Indonesia
[2]Biomedical Engineering, Faculty of Telecommunications and Electronics Engineering, Telkom University, Indonesia
[3]Information Technology, University Kuala Lumpur, Malaysia
Email: [1]dasrilaldo@telkomuniversity.ac.id, [2]adantip@telkomuniversity.ac.id, [3]fathoni.yoka@s.unikl.edu.my

(Article received: October 17, 2024; Revision: January 17, 2025; published: February 20, 2025)

Abstract

Bullying is a great concern that needs to be dealt with as early as possible, be it in the form of physical, verbal, social or cyber bullying. Using NLP algorithms, this paper intends to classify bullying report using Natural Language Processing in conjunction with Bag of Words. The study employs quantitative methodology. A total of 4671 reports of bullying are in essence categorized into physical, verbal, social, cyber and non-cyber bullying. We split the dataset into 80% training set (3737 reports) and 20% testing set (934 reports). The above model has achieved an accuracy of 94,76%, with good values of recall, precision and F1-score: 94,64%, 95,02% and 94,97% respectively. The dataset is then analyzed using Random Forest algorithm and Report of the Bullying Survey The model is to be effective in automatic Detection of Textual Bullying Reports Automated. While there has been no such effort in our institutions so far, automatic reporting of bullying will prove to be effective. This is because the system will allow a school or institution to have a precise constant monitoring of bullying reports. It will also allow an instantaneous action to be taken to protect the victim without letting the situation escalate.

Keywords: *Bag of Words, Bullying, Natural Language Processing, Random Forest, Text Classification*

1. INTRODUCTION

Bullying is an act of intimidation, oppression, or harassment that is deliberately carried out by one or more individuals against the victim, either physically, verbally, socially, or through cyberbullying [1]-[3]. This phenomenon is increasingly rampant, especially in the educational environment and social media, causing serious negative impacts, such as emotional, psychological, and even physical disturbances on victims [4], [5]. Therefore, early detection and classification of bullying reports are very important to reduce the negative impact caused.

Along with the rapid development of technology, data in the form of text, such as reports or complaints, has become an important source of information to detect bullying patterns [6]. Processing large amounts of text data requires special techniques such as Natural Language Processing, which allows computers to understand and analyze human language [7], [8]. One common approach used in Natural Language Processing is Bag of Words, which converts text into numerical representations so that it can be processed by machine learning algorithms [9]-[11].

Previous research has shown the effectiveness of Natural Language Processing techniques and machine learning models in detecting bullying. For example, the Natural Language Processing approach with Support Vector Machine (SVM) is able to detect bullying incidents on social media with quite high accuracy [12]. In addition, a text-based classification model for detecting cyberbullying utilizing Term Frequency-Inverse Document Frequency (TF-IDF) and Naive Bayes' algorithm also showed good performance in classifying text [13]. Other approaches such as Word2Vec and Long Short-Term Memory (LSTM) have succeeded in improving the accuracy of bullying classification on social media platforms by taking into account the context of the word [14]. The use of Natural Language Processing and Random Forest to detect types of verbal bullying in schools also showed significant performance [15]. Text-based bullying detection models with a combination of Bag of Words and Decision Tree have been used to detect various forms of bullying with quite good results [16].

Although these studies have made a significant contribution to text-based bullying detection, this study has some advantages compared to previous studies. First, the study covers different types of bullying, including physical, verbal, social, cyberbullying, and not bullying, which provides a more comprehensive understanding of the different forms of bullying. Most previous studies have focused on one type of bullying, such as cyberbullying or verbal bullying, so the scope of detection is limited.

Second, the approach used in this study involves a combination of Natural Language Processing, Bag of Words, and Random Forest, which are collectively able to handle more complex and diverse text data. Compared to the Support Vector Machine or Naive Bayes-based models in previous studies, the Random Forest algorithm has the advantage of handling unbalanced data and providing more accurate classification results by considering many decision trees.

Third, this study uses a larger and diversified dataset with 4,671 bullying reports categorized into five types. This data is divided into 80% for training and 20% for testing. The larger dataset size compared to previous studies allows the model to be better trained and produces more stable performance in detecting various forms of bullying.

The main contribution of this research is the development of a classification model that is able to detect various types of bullying through a combination of Natural Language Processing, Bag of Words, and Random Forest. The study not only identified one type of bullying, but was also able to differentiate between physical, verbal, social, and cyberbullying bullying, providing more comprehensive and detailed results compared to the more limited approaches in previous studies. In addition, this model can be implemented in various platforms, such as bullying detection systems in schools, social media, or other digital safety systems. Thus, this research is expected to contribute to the development of a more accurate and relevant bullying detection system for widespread use.

2. RESEARCH METHODS

This research focuses on developing a text classification model to detect various types of bullying, namely physical, verbal, social, and cyberbullying, using the Natural Language Processing, Bag of Words, and Random Forest algorithms. In this section, the formulation of the problem, the stages in data processing, and the classification method used in this study are explained.

2.1 Problem Formulation

The main problem in this study is how to classify bullying reports based on text into five categories: Physical, Verbal, Social, Cyberbullying, and Not bullying. Suppose there is a text report set T= {T1, T2,......,Tn}, where each Ti report consists of a string of words. The task is to determine the function of the classification f : Ti → C, where C={C1, C2, C3, C4, C_5} represents five categories of bullying (physical, verbal, social, cyberbullying, and not bullying). The purpose of this study is to maximize classification performance with evaluation metrics such as accuracy, recall, precision, and F1-score. The method used to calculate this probability is the Random Forest algorithm.

2.2 Stages of the Proposed Method

The research process is carried out in the following stages:

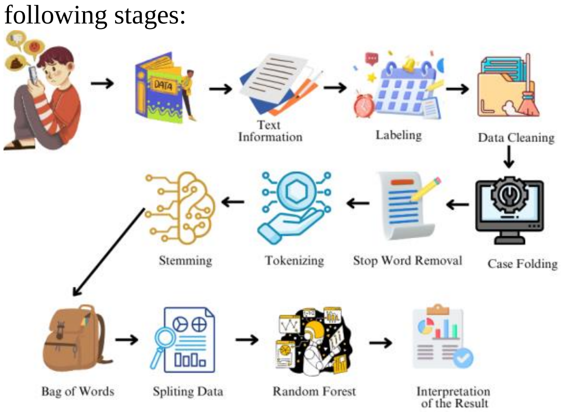

Figure 1. Research Methodology

The data used in this study consisted of 4,671 reports of bullying texts that had been categorized into five types: physical, verbal, social, cyberbullying, and not bullying. This data is divided into 80% for training data and 20% for test data.

The initial step of text data processing is carried out using Natural Language Processing (NLP) techniques to prepare text in numerical form that can be used by machine learning models. Stages in text processing include [17]:

a. Case Folding: All text is lowercased.
b. Removal of Stop Words: Words that do not have a significant meaningful contribution, such as "and", "which", and "or", are removed from the text.
c. Tokenization: Text is broken down into individual words.
d. Stemming: Each word is reduced to its basic form (example: "run" becomes "run").

After the text is processed, the representation of the text is done using the Bag of Words (BoW) technique. Each document is represented as a feature vector based on the frequency of occurrence of unique words throughout the corpus. IF V = {V1, V2, …, Vm} is a unique set of words corpus, then each report Ti, represented as a vector X_i = (Xi1, Xi2, …, Xim) where Xij is the number of occurrences of the word Vj in the report Ti.

For classification tasks, the Random Forest algorithm is used. Random Forest is an ensemble-based machine learning algorithm consisting of many decision trees [18]-[20]. This algorithm works by combining predictions from each decision tree to improve the overall accuracy of the prediction. On each tree, a random subset of features is selected to determine the separation at each tree node.

Mathematically, the classification in Random Forest is formulated as follows [21]:

$$\hat{y} = \text{Mode}\,(h1, (x), h2(x), \ldots, hn\,(x)) \quad (1)$$

Where $hi\ (x)$ is the prediction of the decision tree to-i is a prediction from the decision tree $\hat{y}$ is the final prediction which is the most value (mode) of the whole tree's prediction.

The data that has been represented is divided into two parts:

a. Training Sets: As many as 80% of the total data is used to train the model.
b. Test Sets: 20% of the total data is used to test the performance of the model after training.

Furthermore, the performance of the model is evaluated using several evaluation metrics, namely [22]:

a. Accuracy: Measures the percentage of correct predictions out of the total predictions made.

$$\text{Accuracy} = \frac{\textit{Number of correct predictions}}{\textit{Total Prediction}} \times 100\% \quad (2)$$

b. Precision, Recall, dan F1-score: It is used to measure the model's performance in detecting each bullying class. Mathematically, precision, recall, and F1-score are calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \quad (4)$$

$$\text{F1-score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (5)$$

TP is True Positives, FP is False Positives, and FN is False Negatives. With the proposed method, this research is expected to be able to provide an effective solution in detecting various forms of text-based bullying accurately and efficiently.

3. RESULTS AND DISCUSSION

3.1 Research Data

The data used in this study consisted of 4,671 bullying reports which were categorized into five types: physical bullying, verbal bullying, social bullying, cyberbullying, and not bullying. This data has been processed without typos to ensure optimal text input quality. The data used in this study consisted of 4,671 bullying reports which were categorized into five types: physical bullying, verbal bullying, social bullying, cyberbullying, and not bullying. This data has been processed without typos to ensure optimal text input quality.

Table 1. Sample Research Data

No	Report
1	Mereka bilang saya nggak pantas ada di sini, dan itu menyakitkan. Saya diejek karena cara saya berpakaian, katanya saya terlihat buruk. Setiap kali saya bicara, mereka langsung ketawa dan mengejek saya.
2	Saya diejek karena cara saya berpakaian, katanya saya terlihat buruk. Mereka selalu memanggilku dengan julukan yang menghina. Dia menghina penampilan saya di depan orang banyak. Mereka selalu mengataiku bodoh setiap kali saya berbicara.
3	Saya ditendang oleh dia setiap kali lewat di depannya. Dia pukul bahu saya sampai saya kesakitan. Dia memukul perut saya dan membuat saya sulit bernapas. Dia bilang saya jelek dan bodoh di depan teman-teman. Mereka selalu mengataiku bodoh setiap kali saya berbicara.
4	Saya di-bully online, mereka membuat akun palsu untuk menghina saya. Teman-teman selalu mengejek saya di Instagram, membuat saya merasa terhina. Mereka mengirim pesan kasar dan menghina saya di media sosial.
6	Kami biasanya berbicara tentang pelajaran dan bercanda bersama. Kadang saya berbicara dengan teman, tapi mereka baik-baik saja. Saya merasa diterima dengan baik oleh teman-teman. Mereka kadang bercanda, tapi tidak pernah melewati batas. Tidak ada yang salah, saya hanya merasa kurang nyaman sesekali. Saya tidak merasa diintimidasi, mereka memperlakukan saya dengan baik. Teman-teman saya ramah dan tidak pernah menghina saya.

3.2 Text Information

Text Information in the context of your bullying data consists of two main columns: "Report" and "Category". The "Report" column contains textual descriptions of various situations experienced by individuals, including verbal, physical, and cyber bullying experiences, as well as situations that are not bullyingg. The "Category" column classifies each report into specific categories such as Verbal Bullying, Physical Bullying, Cyberbullying, or Not bullyingg. This data is already in a structured text format, allowing for further analysis of the types of bullying reported and the characteristics of each category. This information can be used to understand bullying patterns, identify the most common types of bullying, and develop appropriate prevention strategies.

	Report	Category
0	Mereka bilang saya nggak pantas ada di sini, dan itu menyakitkan. Saya diej...	Verbal Bullying
1	Saya diejek karena cara saya berpakaian, katanya saya terlihat buruk. Merek...	Verbal Bullying
2	Saya ditendang oleh dia setiap kali lewat di depannya. Dia pukul bahu saya s...	Physical Bullying
3	Saya di-bully online, mereka membuat akun palsu untuk menghina saya. Te...	Cyberbullying
4	Kami biasanya berbicara tentang pelajaran dan bercanda bersama. Kadang ...	Not Bullying

Figure 2. Text Information Results

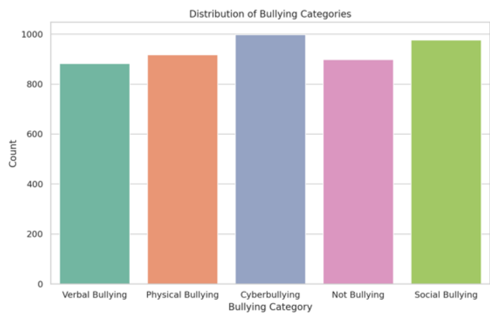

Figure 3. Chart Display by Category

3.3 Preprocesing Data

The preprocessing stage is an important step in ensuring that the text of the bullying report can be processed properly by machine learning algorithms.

In the initial stage, labeling is carried out where each report is given the appropriate category, namely physical bullying, verbal bullying, social bullying, cyberbullying, or not bullying. After that, data cleaning is carried out, where the text is cleaned of unnecessary characters such as punctuation, special symbols, or irrelevant words.

	Report	Category	Label
0	Mereka bilang saya nggak pantas ada di sini, dan itu menyakitkan. ...	Verbal Bullying	1
1	Saya diejek karena cara saya berpakaian, katanya saya terlihat buru...	Verbal Bullying	1
2	Saya ditendang oleh dia setiap kali lewat di depannya. Dia pukul ba...	Physical Bullying	2
3	Saya di-bully online, mereka membuat akun palsu untuk menghina ...	Cyberbullying	3
4	Kami biasanya berbicara tentang pelajaran dan bercanda bersama. ...	Not Bullying	0

Figure 4. Labeling

The next process is case folding, where all text is converted to lowercase letters to ensure consistency in text processing. This is done so that the same words with uppercase and lowercase letters are considered as one and the same entity.

	Report_Lowercase
0	mereka bilang saya nggak pantas ada di sini, dan itu menyakitkan. saya diejek karena cara saya berpak...
1	saya diejek karena cara saya berpakaian, katanya saya terlihat buruk. mereka selalu memanggilku deng...
2	saya ditendang oleh dia setiap kali lewat di depannya. dia pukul bahu saya sampai saya kesakitan. dia ...

Figure 5. Case Folding

After case folding, stop words removal is carried out, i.e. removing common words that do not provide much important information, such as "and", "which", or "or".

	Text_Without_Stopwords
0	bilang nggak , menyakitkan . diejek berpakaian , buruk . kali bicara , langsung ketawa mengejek .
1	diejek berpakaian , buruk . memanggilku julukan menghina . menghina penampilan orang . mengataiku ...
2	ditendang kali depannya . bahu kesakitan . memukul perut sulit bernapas . bilang jelek bodoh teman-te...

Figure 6. Stop Words Removal

Then, the text is broken down into individual words through a tokenization process to facilitate further processing.

	Tokens
0	["bilang","nggak",",","menyakitkan",".","diejek","berpakaian",",","buruk",".","kali","bicara",",","langsung","ke...
1	["diejek","berpakaian",",","buruk",".","memanggilku","julukan","menghina",".","menghina","penampilan","o...
2	["ditendang","kali","depannya",".","bahu","kesakitan",".","memukul","perut","sulit","bernapas",".","bilang","...

Figure 7. Tokenization

The final stage in preprocessing is stemming, which is the process of returning words to their basic or root form. This helps reduce the variation in the form of the word, so words like "run", "run", and "run" are considered to be one and the same word, which is "run".

	Stemmed_Tokens
0	["bilang","nggak","sakit","ejek","pakai","buruk","kali","bicara","langsung","ketawa","ejek"]
1	["ejek","pakai","buruk","panggil","juluk","hina","hina","tampil","orang","kata","bodoh","kali","bicara"]
2	["tendang","kali","depan","bahu","sakit","pukul","perut","sulit","napas","bilang","jelek","bodoh","kata","bo...

Figure 8. Stemming

3.4 Bag of Words (BoW)

Pada tahap ini, dilakukan representasi teks laporan bullying menggunakan model Bag of Words (BoW). Model ini berhasil memuat data dengan ukuran 4671 baris (dokumen) dan 88 kolom (kata unik). Setiap baris dalam BoW mewakili satu laporan bullying, dan setiap kolom merepresentasikan satu kata unik yang ditemukan dalam keseluruhan kumpulan data.

	abai	acara	ajak	ajar	aktivitas	akun	alami	anggap	asing	bahu
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0

Figure 9. Matrik BoW

Figure 9 shows the first 5 rows and the first 10 columns of the BoW model. Each cell contains a number that indicates the frequency of the occurrence of the word in the document. For example, in the 3rd row (index 2), the word "shoulder" appears once, while the other words in the column do not appear at all in the same document.

Additionally, here is a statistical summary for the first 7 words in the BoW model:

	abai	acara	ajak	ajar	aktivitas	akun	alami
count	4671	4671	4671	4671	4671	4671	4671
mean	0.2935131663	0.30229073	0.2956540355	0.1395846714	0.1466495397	0.1517876258	0.137871976
std	0.648271206	0.6646738488	0.6519123207	0.3465927126	0.3537940779	0.3588533289	0.344802475
min	0	0	0	0	0	0	0
25%	0	0	0	0	0	0	0
50%	0	0	0	0	0	0	0
75%	0	0	0	0	0	0	0
max	2	2	2	1	1	1	1

Figure 10. BoW Results Statistics

Figure 10 provides a statistical summary of the first 10 columns in the BoW model, which consists of:

a. Count: The number of documents (4671) is the same for all words.
b. Mean: The average occurrence of words throughout the document. For example, the word "ignore" appears an average of 0.29 times per document.
c. Std: The standard deviation of the occurrence of the word. Higher values indicate greater variation in the use of the word.
d. Min and Max: The minimum and maximum frequency of occurrences of words in a single document. For example, the word "aign" appears a maximum of 2 times in a single document.
e. Percentile (25%, 50%, 75%): The distribution of the frequency of words in the document. Most words have a median value (50%) equal to 0, which means they

don't appear in at least half of the documents.

From these results, we can conclude that the BoW model has successfully converted the text of bullying reports into numerical representations.

In the Bag of Words model applied to bullying reports, there were 20 words with the highest frequency of occurrence. These words often appear in reports and point to key themes relevant to bullying. Here is a list of words with the highest frequency:

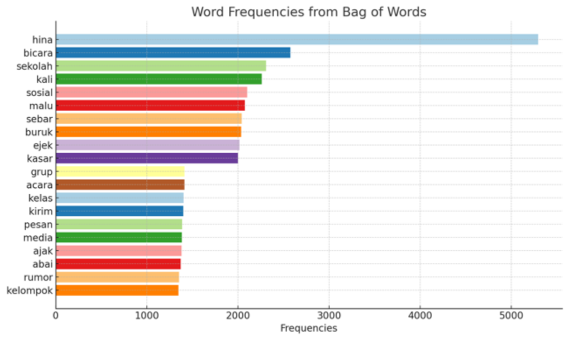

Figure 11. BoW Word Frequency

The word "insult" occupies the highest position with the number of occurrences as many as 5,295 times, followed by the word "talk" which appears 2,576 times, and "school" which appears 2,308 times. Words such as "embarrassed", "bad", and "rude" also appear with significant frequency, reflecting the emotional and verbal context of many bullying reports.

These words mostly figure out situations of bullying that occur verbally or socially, where the perpetrator often uses insults, ridicule, and rumors to demean the victim. Words such as "spread" and "group" indicate aspects of cyberbullying, where social media and group messages are the main means of spreading bullying. Meanwhile, the words "classroom" and "event" indicate places where many incidents of bullying occur, especially in school settings.

This distribution of words provides a clear figure on the dominant themes in bullying reports and can be used to better understand behavioral patterns in bullying cases. The BoW model effectively converts the report text into a numerical representation, which can later be used for further classification or other in-depth analysis.

This model captures information about how often certain words appear in a document, which can later be used as input for machine learning algorithms or further analysis.

3.5 Random Forest Algorithm

After going through a training process with 80% of the training data, the Random Forest model was applied to 20% of the test data to evaluate the accuracy of its classification. The classification results are shown through a confusion matrix that figures the model's ability to classify bullying reports into five categories: cyberbullying, not bullying, physical bullying, social bullying, and verbal bullying.

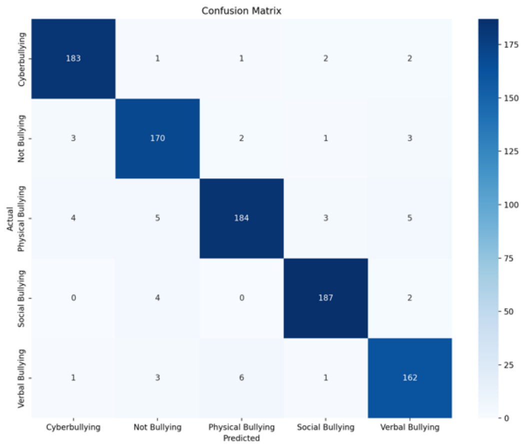

Figure 12. Confusion Matrix

From the confusion matrix in Figure 12, here are some key findings:

a. Cyberbullying: Of the 189 reports, 183 were correctly classified as cyberbullying, while there was 1 report that was incorrectly classified as not bullying and verbal bullying.
b. Not bullyingg: Of the 179 reports that should have been classified as not bullying, 170 of them were classified correctly, while there were few errors in other categories such as cyberbullying and social bullying.
c. Physical Bullying: Of the 201 reports, 184 of them were correctly classified as physical bullying, but there were errors in the classification of some reports into other categories.
d. Social Bullying: The model successfully classified 187 of the 193 reports as social bullying, with few misclassifications.
e. Verbal Bullying: Out of 173 reports, the model correctly classified 162, but some reports were incorrectly classified into other categories.

3.6 Interpretation of the Result

The Confusion Matrix Matrix shows the distribution of model predictions for each class. The main diagonal indicates the correct prediction, while the other cells indicate misclassification.

a. Precision: Shows the proportion of positive predictions that are truly positive. All classes have a precision above 0.91, with "Social Bullying" having the highest precision (0.9689).
b. Recall: Shows the proportion of positive cases that have been successfully identified. Recall scores were also high for all grades, with

"Social Bullying" having the highest recall (0.9689).

c. F1 Score: Is the harmonic average of precision and recall. A high F1 score shows a good balance between precision and recall. "Social Bullying" has the highest F1 score (0.9689).
d. Accuracy: The overall accuracy of the model is 0.9476 or 94.76%, which indicates excellent performance.

Based on these results, the Random Forest model shows good performance with a low classification error rate. Most of the reports are correctly classified into their respective categories, which indicates that the model managed to learn the patterns in the data quite well. However, some misclassification occurs among semantically similar categories, such as between verbal bullying and cyberbullying, which indicates an overlap of characteristics between these two categories.

4. DISCUSSION

The results of this study show that the Random Forest algorithm combined with Natural Language Processing (NLP) techniques based on Bag of Words (BoW) is able to classify bullying reports with a high level of accuracy, which is 94.76%. The model's performance in terms of precision, recall, and F1-score also showed excellent results, confirming that it was effective in distinguishing between different types of bullying such as physical, verbal, social, cyberbullying, and not bullyingg.

Previous research by Dedeepya et al. used the Support Vector Machine (SVM) algorithm to detect cyberbullying on Twitter, which achieved a fairly high level of accuracy [12]. However, this study only focuses on one form of bullying, which is cyberbullying, while our research includes five types of bullying, providing a more comprehensive view. In addition, the SVM model used in previous studies tended to have difficulty handling data imbalances, where the majority class dominated the prediction results, while Random Forest in this study showed better performance in handling unbalanced data.

Another study by Chingmuankim and Jindal used the Term Frequency-Inverse Document Frequency (TF-IDF) and Naive Bayes techniques for the classification of bullying texts, with adequate results in detecting cyberbullying [13]. However, Naive Bayes' algorithms are often not as efficient as Random Forest's handling complex variations in text data. Random Forest has the advantage of using multiple decision trees, which improves the accuracy of classification on a variety of more diverse bullying categories compared to probabilistic models such as Naive Bayes.

Research by Fati et al. uses a deep learning approach with an attention mechanism to detect cyberbullying on Twitter [14]. Deep learning approaches such as Long Short-Term Memory (LSTM) or attention mechanisms are indeed superior in capturing the deeper semantic context of the text, so they can detect the context of words better than simple models such as Bag of Words. However, the use of the Random Forest algorithm in this study showed very satisfactory results in terms of accuracy, with the main advantages being shorter training times and easier model interpretation compared to deep learning.

The advantage of this study lies not only in the combination of algorithms, but also in the large and diversified dataset size, which is 4,671 bullying reports. In comparison, the study by Esquivel et al. that focused on the emotional effects of bullying and cyberbullying used a smaller dataset, thus providing less stable results in terms of classification [5]. The larger dataset in this study allowed the model to be better trained and produce more consistent results in the classification of bullying reports.

Although the results achieved are very good, some classification errors still occur, especially between the categories of verbal bullying and cyberbullying. This is due to the semantic similarities between the two categories, especially in the case of verbal bullying that occurs online. Further research may consider the use of deep learning models such as Word2Vec or Transformer that are more effective in capturing word context and semantic relationships between words, to improve the accuracy of classification on overlapping categories.

5. CONCLUSION

This study shows that the combination of Natural Language Processing (NLP) techniques using Bag of Words (BoW) and the Random Forest algorithm succeeded in classifying bullying reports with high accuracy, reaching 94.76%. The model is capable of detecting various types of bullying, including physical, verbal, social, cyberbullying, and non-bullying, with satisfactory levels of precision, recall, and F1-score. The pre-processing process that includes case folding, stop word removal, tokenization, and stemming ensures that text data can be processed consistently, while the Random Forest algorithm has proven to be effective in handling unbalanced data, giving it an edge over other models such as the Support Vector Machine and Naive Bayes.

Nonetheless, the study faces challenges in misclassification of semantically similar categories, such as between verbal bullying and cyberbullying. To address this, more advanced NLP approaches such as Word2Vec or Transformer can be considered in advanced research. Overall, the results of this study contribute significantly to the development of text-based bullying detection systems, which can be applied across various platforms to help early detection and prevention of bullying more effectively.

ACKNOWLEDGMENTS

The author would like to thank the Institute for Research and Community Service (LPPM) of Telkom Purwokerto University for the support and facilities provided during this research process. The assistance and contribution provided by LPPM played a very important role in the smooth running and completion of this research. The author also thanks all parties who have contributed directly or indirectly to this research.

REFERENCES

[1] H. Torkos dan A. Anwer, "INVOLVEMENT OF TEENAGERS IN THE BEHAVIOR OF BULLYING & CYBER VIOLENCE," *J. PLUS Educ.*, vol. 31, no. 2/2022, hlm. 72–82, Nov 2022, doi: 10.24250/JPE/2/2022/HT/AA.

[2] M. F. H. Bornes, D. Terceiro, F. V. Peña, N. Burdisso, dan S. Terrasa, "Proposal for a short version of an international questionnaire to detect bullying: adaptation to Argentine Spanish language and exploratory factor analysis," *Arch. Argent. Pediatr.*, vol. 120, no. 1, Feb 2022, doi: 10.5546/aap.2022.eng.14.

[3] H. Korzhov dan M. Yenin, "Sociological dimensions of cyberbullying: essence, consequences, and coping strategies," *Sociol. Theory Methods Mark.*, no. 4, hlm. 103–120, 2022, doi: 10.15407/sociology2022.04.103.

[4] V. Veronica, "Bullying in School-Age Children," *Sci. Psychiatr.*, vol. 3, no. 2, hlm. 198–206, Apr 2022, doi: 10.37275/scipsy.v5i1.136.

[5] F. A. Esquivel, I. L. D. L. G. López, dan A. D. Benavides, "Emotional impact of bullying and cyber bullying: perceptions and effects on students," *Rev. Caribeña Cienc. Soc.*, vol. 12, no. 1, hlm. 367–383, Jun 2023, doi: 10.55905/rcssv12n1-022.

[6] W. Qian, S. Yu, Z. Nie, X. S. Lu, H. Liu, dan B. Huang, "Improved Hierarchical Attention Networks for Cyberbullying Detection via Social Media Data," dalam *2023 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, Marseille, France: IEEE, Okt 2023, hlm. 1–6. doi: 10.1109/ICNSC58704.2023.10319023.

[7] O. Kjell, S. Giorgi, dan H. A. Schwartz, "The text-package: An R-package for analyzing and visualizing human language using natural language processing and transformers.," *Psychol. Methods*, vol. 28, no. 6, hlm. 1478–1498, Des 2023, doi: 10.1037/met0000542.

[8] N. Alinda Rahmi and R. Wulan Dari, "IMPLEMENTATION OF NATURAL LANGUAGE PROCESSING (NLP) IN CONSUMER SENTIMENT ANALYSIS OF PRODUCT COMMENTS ON THE MARKETPLACE", *J. Tek. Inform. (JUTIF)*, vol. 5, no. 3, pp. 693-701, May 2024.

[9] M. Muffo, R. Tedesco, L. Sbattella, dan V. Scotti, "Static Fuzzy Bag-of-Words: a lightweight sentence embedding algorithm," 2023, doi: 10.48550/ARXIV.2304.03098.

[10] B. Probierz, A. Hrabia, dan J. Kozak, "A New Method for Graph-Based Representation of Text in Natural Language Processing," *Electronics*, vol. 12, no. 13, hlm. 2846, Jun 2023, doi: 10.3390/electronics12132846.

[11] D. H. Lubis, S. Sawaluddin, dan A. Candra, "Machine Learning Model for Language Classification: Bag-of-words and Multilayer Perceptron," *J. Inform. Telecommun. Eng.*, vol. 7, no. 1, hlm. 356–365, Jul 2023, doi: 10.31289/jite.v7i1.10114.

[12] P. Dedeepya, P. Sowmya, T. D. Saketh, P. Sruthi, P. Abhijit, dan S. P. Praveen, "Detecting Cyber Bullying on Twitter using Support Vector Machine," dalam *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India: IEEE, Feb 2023, hlm. 817–822. doi: 10.1109/ICAIS56108.2023.10073658.

[13] Chingmuankim dan R. Jindal, "Classification and Analysis of Textual data using Naive Bayes with TF-IDF," dalam *2022 4th International Conference on Electrical, Control and Instrumentation Engineering (ICECIE)*, KualaLumpur, Malaysia: IEEE, Nov 2022, hlm. 1–9. doi: 10.1109/ICECIE55199.2022.10000309.

[14] S. M. Fati, A. Muneer, A. Alwadain, dan A. O. Balogun, "Cyberbullying Detection on Twitter Using Deep Learning-Based Attention Mechanisms and Continuous Bag of Words Feature Extraction," *Mathematics*, vol. 11, no. 16, hlm. 3567, Agu 2023, doi: 10.3390/math11163567.

[15] A. Alabdulwahab, M. A. Haq, dan M. Alshehri, "Cyberbullying Detection using Machine Learning and Deep Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 10, 2023, doi: 10.14569/IJACSA.2023.0141045.

[16] P. Shiguihara dan L. Berton, "Exploring Deep Neural Networks and Decision Tree for Spanish Text Classification," dalam *2022 IEEE XXIX International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, Lima, Peru: IEEE, Agu 2022, hlm. 1–4. doi: 10.1109/INTERCON55795.2022.9870087.

[17] Mr Adepu Rajesh dan Dr Tryambak Hiwarkar, "Exploring Preprocessing Techniques for Natural LanguageText: A Comprehensive Study Using Python Code," *Int. J. Eng. Technol. Manag. Sci.*, vol. 7, no. 5, hlm. 390–

399, 2023, doi: 10.46647/ijetms.2023.v07i05.047.

[18] X. Fu, Y. Chen, J. Yan, Y. Chen, dan F. Xu, "BGRF: A broad granular random forest algorithm," *J. Intell. Fuzzy Syst.*, vol. 44, no. 5, hlm. 8103–8117, Mei 2023, doi: 10.3233/JIFS-223960.

[19] K. Dedja, F. K. Nakano, K. Pliakos, dan C. Vens, "BELLATREX: Building Explanations through a LocaLly AccuraTe Rule EXtractor," 2022, *arXiv*. doi: 10.48550/ARXIV.2203.15511.

[20] I. Tarchoune, A. Djebbar, dan H. F. Merouani, "Improving Random Forest with Pre-pruning technique for Binary classification," *Sci. Abstr.*, vol. 1, no. 2, hlm. 11, Jul 2023, doi: 10.59287/as-abstracts.1202.

[21] W. Gao, F. Xu, dan Z.-H. Zhou, "Towards convergence rate analysis of random forests for classification," *Artif. Intell.*, vol. 313, hlm. 103788, Des 2022, doi: 10.1016/j.artint.2022.103788.

[22] S. Riyanto, I. S. Sitanggang, T. Djatna, dan T. D. Atikah, "Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, 2023, doi: 10.14569/IJACSA.2023.01406116.