JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8721
Hal 309-318
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Boosting Methods for Multi-label Data Cyberbullying

Fidya Farasalsabila[1,*], Mhd Adi Setiawan Aritonang[2], Faradiba Jabnabillah[1], Anip Moniva[3], Verra Budhi Lestari[4], Rizky Handayani[5]

[1]Sistem Informasi, Institut Teknologi Batam, Batam, Indonesia
[2]Teknik Komputer, Institut Teknologi Batam, Batam, Indonesia
[3]Kecerdasan Buatan dan Robotik, Politeknik AI Budi Mulia Dua, Yogyakarta, Indonesia
[4]Sistem Informasi, Universitas Media Nusantara Citra, Jakarta, Indonesia
[5]Bisnis Digital, Institut Teknologi Bisnis dan Kesehatan Bhakti Putra Bangsa Indonesia, Jawa Tengah, Indonesia
Email: [1*]fidya@iteba.ac.id, [2]adi@iteba.ac.id, [3]faradiba@iteba.ac.id, [4]anipmoniva@plai.ac.id,
[5]vera.budhi@mncu.ac.id, [6]rizkyhanda2000@gmail.com
Email Penulis Korespondensi: fidya@iteba.ac.id
Submitted **09-06-2025;** Accepted **28-06-2025;** Published **30-06-2025**

Abstract

Easy accessibility to the internet and social media allows individuals to communicate anonymously, providing opportunities for abusive and harmful behavior. The psychological impact of cyberbullying can be very detrimental, triggering stress, depression, and even causing more serious consequences such as suicide. This paper describes cyberbullying sentiment analysis with a focus on the use of four different boosting methods, namely Gradient Booster, Gradient Booster, XGBoost, AdaBoost, dan LightGBM on a multi-label public dataset covering 6 categories. The aim of this research is to compare and analyze the relative performance of these boosting methods in overcoming the challenges of multi-label sentiment analysis in the context of cyberbullying. Results reveal that XGBoost and LightGBM have a tendency to more effectively overcome the challenges of detecting cyberbullying in more complex categories, making a positive contribution to the development of superior detection systems in the context of multi-label sentiment analysis. This research contributes to the field by providing a comparative analysis of state-of-the-art boosting algorithms, highlighting their strengths in multi-label classification tasks, and offering practical insights for developing more accurate and reliable cyberbullying detection systems. The findings from this study are expected to serve as a reference for future development of machine learning-based tools that can help mitigate the psychological harm caused by online abuse, particularly in detecting subtle and complex forms of cyberbullying behavior.

Keywords: Cyberbullying; Multi-label Classification; Boosting Methods; Sentiment Analysis

1. INTRODUCTION

In the increasingly developing digital era, the problem of cyberbullying has become a major highlight in brave life. There are several research in the literature to prevent cyberbullying using intervention and preventative techniques. Such ideas are derived from the domains of psychology and education. These techniques, however, are uncommon on a worldwide scale. Furthermore, victims of cyberbullying frequently refuse to communicate with a parent [1], teacher , or other adult [2]. They spend a lot of time online [3], seek anonymous aid [4], and publish a need for information and assistance on the Internet [5]. However, the Internet is the most effective means of giving cyberbullying treatments[6]. Web-based methods can also be utilized anytime and whenever the patient wishes [7]. With the diversity of forms and contexts of messages that can be considered detrimental, sentiment analysis becomes an important instrument in detecting and categorizing cyberbullying behavior [8], [9]. Moreover, when we are faced with multi-label data, where a message can relate to more than one cyberbullying category, the challenge of sentiment analysis becomes even more complex [10], [11].

The latest research [11], [12], [13], [14], in this study shows progress in the application of model boosting methods. Boosting methods, such as AdaBoost, Gradient Boosting, However, its application to data involving more than one cyberbullying label still requires in-depth understanding. In an effort to fill this gap, the main objective of this paper is to analyze the extent to which the boosting method can be applied effectively to multi-label sentiment data related to cyberbullying. By presenting case studies of the different types of messages that can be found in online environments, we seek to understand the ability to improve methods for identifying and classifying cyberbullying messages related to more than one category. The key contribution of this research lies in providing a comprehensive comparative evaluation of multiple boosting algorithms specifically AdaBoost, Gradient Boosting, XGBoost, and LightGBM in the context of multi-label cyberbullying sentiment analysis, which remains underexplored in existing literature. The results are expected to offer practical insights for the development of more robust detection systems capable of handling the complexity and subtlety of real-world cyberbullying scenarios. This contribution is important because it not only highlights the strengths and limitations of each boosting method in handling multi-label classification, but also provides practical insights that can guide the development of more accurate, efficient, and scalable cyberbullying detection systems particularly those needed in real-world applications such as social media monitoring, online safety platforms, and digital mental health interventions.

2. RESEARCH METHODOLOGY

2.1 Research Method

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8721
Hal 309-318
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

The models proposed in this research are four boosting models, namely AdaBoost, Gradient Boosting, and XGBoost and LightGBM. There are several steps that need to be taken before applying this model in research to carry out multi-label classification. Figure 1 is the flow of the research that will be carried out.

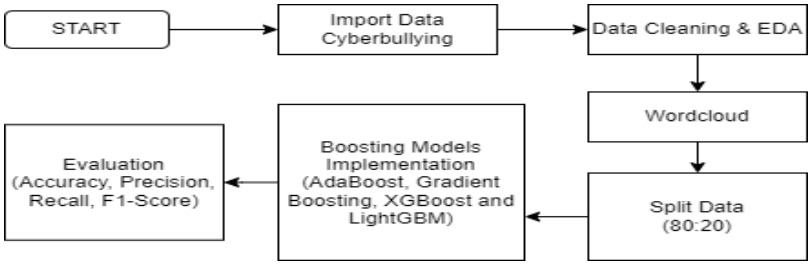

Figure 1. Research Flow Analysis of Boosting Methods for Multi-label Data Cyberbullying

Figure 1 depicts the flow of the proposed research. The first step, import and clean the cyberbullying dataset so that the data can be continued into the model implementation process. A wordcloud will be displayed before the process of implementing the data that has been preprocessed into the proposed model. Next, the data will be divided into several new subsets which will be implemented into the model. The final step is to evaluate the performance of the proposed 4 models on multi-label cyberbullying data.

2.2 Datasets

The raw data set in this research was obtained from [15] and stored in "comma separated values" (CSV) and can be accessed by the public via IEEE Dataport.

Table 1. Amount of raw data

No	Data Source	Data Label	Total
1	IEEE Dataport	religion (7997)	47.656
2		age (7992)	
3		ethnicity (7959)	
4		gender (7948)	
5		other_cyberbullying (7823)	
6		not_cyberbullying (7937)	

This dataset includes comments on Twitter written in English that use words that refer to cyberbullying, such as bad descriptions and offensive words in expressing opinions or describing people who have negative qualities, and some also use the word -words that refer to cyberbullying. as shown in table 1, the number of tweets was 47,692 data with 6 classes, namely religion, age, ethnicity, gender, not_cyberbullying and other_cyberbullying.

2.3 Data Cleaning

For the purposes of this research, the collected data underwent a series of pre-processing steps, designed to clean and prepare the text in preparation for analysis implemented into the model. The methods applied to the data used in this investigation are listed in Table 3.

Table 2. Data cleaning method

No	Function	Description
1	strip_emoji(text)	Removes emoji from text and uses regular expressions to replace all emoji characters with empty strings.
2	strip_all_entities(text)	Removes punctuation, stopwords, links, mentions, and newline characters from text. Apart from that, it also converts text to lowercase and removes stopwords (common words that often do not provide important information).
3	clean_hashtags(tweet)	Clean hashtags at the end of sentences and keep hashtags in the middle of sentences by only removing the # symbol.
4	filter_chars(text)	Filters special characters such as & and $ that are present in multiple words.
5	remove_mult_spaces(text)	Removes excess spaces (more than one space in a row) in text.
6	filter_non_english(text)	Checks if the text is in English and Returns an empty string if the text is not in English.
7	expand_contractions(text):	Replace word contractions with full forms. For example, changing "don't" to "do not".
8	remove_numbers(text)	Removes numbers from text
9	lemmatize(text):	Lemmatizes the words in the text and converts the words to their basic form (lemmat).

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8721
Hal 309-318
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2.4 Wordcloud

A wordcloud is a visual representation of frequently used words arranged in a cloud-like form. The terms that appear most frequently in the data text will have a large text size, and vice versa [16]. Wordcloud facilitates the identification of keywords commonly used in the field of cyberbullying. The size of terms in a wordcloud corresponds to the frequency with which they appear in the text. Terms that have a higher frequency will be displayed in larger sizes, which will visually highlight the most prominent or controversial terms. By examining words in a wordcloud, you can gain an initial understanding of the attitudes expressed in texts related to cyberbullying. For example, determining the dominant sentiment of a word as positive or negative or a certain class [17].

2.5 Boosting Methods

Boosting is a powerful ensemble learning technique in machine learning that aims to improve the predictive performance of weak learners by combining them into a strong composite model through a sequential training process[18]. A weak learner is defined as a model that performs only slightly better than random guessing, with decision stumps (shallow decision trees) commonly used in practice. The fundamental idea of boosting lies in its iterative correction of classification errors: at each stage, a new model is trained to emphasize the instances that were previously misclassified by the ensemble. This is typically achieved by adjusting the weights of the training samples, where higher weights are assigned to incorrectly predicted data points to draw greater focus in subsequent iterations. Over time, the ensemble learns from its own mistakes, leading to improved generalization and robustness[19], [20]. The final prediction is made by aggregating the outputs of all individual learners, often through a weighted majority vote or a sum of weighted scores, depending on the task (classification or regression). Several widely adopted variants of boosting include AdaBoost (Adaptive Boosting), which introduced the reweighting concept; Gradient Boosting, which frames the learning process as a gradient descent optimization on a loss function; and its modern extensions such as XGBoost and LightGBM, each offering enhancements in speed, regularization, and scalability[21]. These models have proven highly effective in domains with complex, imbalanced, or noisy data attributes common in cyberbullying detection tasks where textual inputs often contain subtle, context-dependent, and multi-label abusive content. Due to its iterative learning process, capacity to handle non-linear relationships, and ability to reduce both bias and variance, boosting remains one of the most competitive and widely used approaches in supervised learning today[19].

2.6 Confusion Matrix

In this research, we will use the confusion matrix as an evaluation matrix. Confusion Matrix is a matrix that shows how well a model predicts a particular situation. Confusion Matrix is one method to illuminate the performance of classification models. In its most basic form, a confusion matrix contains a 2x2 table to categorize a model with data as A or B [22].

Table 3. Confusion Matrix Formula

Actual Class	Predict Class	
	Positive	**Negative**
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Various evaluation methodologies are used to assess the effectiveness of classification on previously unseen test data. The most commonly used metrics for text classification are precision, recall, F-measure, and accuracy. The general goal is to maximize all metrics with values between 0 and 1. As a result, higher values indicate better categorization performance [23].

3. RESULT AND DISCUSSION

3.1 Data Visualization

The case study taken is a cyberbullying case study whose dataset was obtained through tweets from the Twitter platform [15]. A total of 47,692 raw tweet data with 6 different labels, namely religion (7997), age (7992), ethnicity (7959), gender (7948), not_cyberbullying (7937) and other_cyberbullying (7823). At the data preprocessing stage, a series of preprocessing techniques are used to clean and prepare text data before further analysis. To provide a better understanding of the dominant themes and vocabulary used within the cyberbullying dataset, a wordcloud visualization was generated. This visualization offers an intuitive overview of frequently occurring terms, which is crucial in identifying commonly used expressions that may indicate abusive or harmful language patterns.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8721
Hal 309-318
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

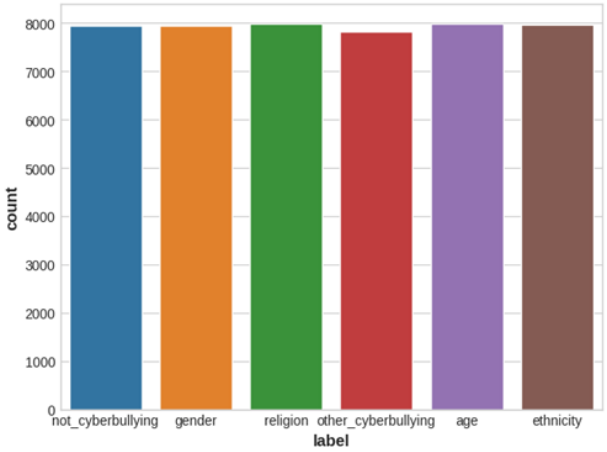

Figure 2. Wordcloud based on all words

Figure 2 illustrates a wordcloud created from the entire dataset, where word size reflects frequency. Prominently featured words represent terms that appear most often across all cyberbullying-related comments. These include words with strong emotional or discriminatory connotations, signaling their prevalence in online bullying discourse. The visualization supports the pre-analysis stage by highlighting potential keywords that are semantically or sentimentally relevant for classification, and helps justify the need for robust text cleaning and classification techniques used in the modeling phase. A sequence of functions is used to preprocess 47,692 text data before it is utilized in the boosting model. The first step is to remove any emoji, punctuation marks, stop words, links, tags, and new line characters from the text, as well as convert it to lowercase. Then, delete the '#' symbol from the hash mark in the midst of the sentence and clean the hash mark at the conclusion of the phrase. Aside from that, special characters like '&' and '$' that appear in a variety of terms are also filtered. To decrease noise in the data, this preprocessing procedure additionally removes double spaces from text. Then it tests to see if the text is in English and returns an empty string if it is not. At this level, the data is further cleaned by expanding contractions in the text, eliminating numeric digits from the text, and lemmatizing terms in the text. Continue by deleting short words (those that are shorter than the stated minimum length), replacing long words with their basic form, removing repetitive punctuation, and reducing extra spaces. The last sequence calls all of the cleaning routines in the correct order in order to preprocess the entire tweet. Overall, these pre-processing methods try to clean and unify text data, making it more acceptable for analysis or input into boosting models. At the completion of the preprocessing step, there were 41,409 data points with six classes: religion (7916), age (7818), ethnicity (7418), gender (7285), not_cyberbullying (6066), and other_cyberbullying (4906).

The most commonly repeated terms in a multiclass cyberbullying classification dataset were shown using a wordcloud. Figure 3 is a visualization of words grouped according to the 6 labels in the data used. To gain deeper insight into the linguistic patterns specific to each cyberbullying category, individual wordclouds were generated based on the labeled data. This approach allows for the identification of distinctive keywords associated with particular forms of abuse, providing more targeted understanding for model development.

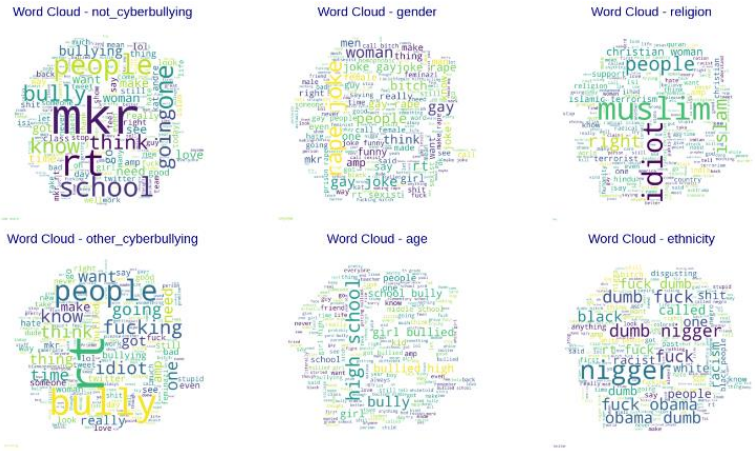

Figure 3. Wordcloud based on labels

Figure 3 displays a set of wordclouds grouped by the six labels in the dataset: religion, age, ethnicity, gender, not_cyberbullying, and other_cyberbullying. Each wordcloud reveals the dominant vocabulary used within its respective category. For instance, discriminatory or derogatory terms specific to religious or gender-based bullying are more prevalent in their corresponding labels. This helps highlight the unique semantic features of each class, which can be

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8721
Hal 309-318
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

leveraged to improve model sensitivity and accuracy in multi-label classification. In addition to category-specific analysis, a comprehensive wordcloud was created to visualize the combined frequency of all terms across the entire dataset. This serves as a holistic view of the language landscape, useful for guiding global feature extraction and preprocessing strategies.

Figure 4 shows the wordcloud based on all words found in the dataset. Words with higher frequency are displayed in larger fonts, indicating their prominence in cyberbullying discourse. The visualization captures key offensive, emotional, and discriminatory terms that frequently appear in user-generated content. This allows researchers to identify linguistically rich features that may contribute to model training and supports the argument for using advanced classification techniques to handle such complex language patterns.

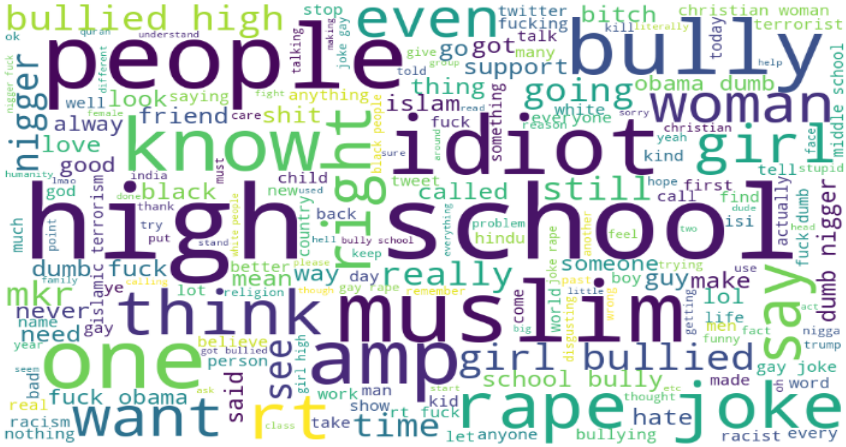

Figure 4. Wordcloud based on all words

Figure 4 presents a single consolidated wordcloud showing all terms in the dataset without label separation. Words with the highest frequency are displayed in larger fonts, indicating their widespread use across various categories. This visualization underscores the overall thematic concentration of the data, helping researchers identify common terms that may appear in multiple abuse types. It also assists in detecting generic versus context-specific words, which is valuable for optimizing tokenization and feature engineering processes.

In the next phase, the four boosting models are implemented for classification. The first step is to divide the collected data into 2 subsets, namely Train-Test Data and Train-Validation Data. Train-Test data is used to train the model and evaluate model performance. Meanwhile, Train-Validation Data is used for further tuning the model and monitoring model performance during training.

3.2 Evaluation

The Gradient Booster model demonstrates a notably strong performance in the classification of classes 0 and 1, exhibiting high levels of both precision and recall, which indicates its capability to correctly identify relevant instances while minimizing false positives. This suggests that for labels with sufficient representation and distinct linguistic patterns, the model is highly effective in capturing semantic features from the textual data. However, a noticeable decline in classification performance is observed for classes 2, 3, and 4, where the model struggles to maintain the same level of accuracy, likely due to increased complexity, overlapping features, or lower label frequency within these categories. Despite these disparities, the overall accuracy of the model remains relatively high at approximately 81%, suggesting a robust general capability across the dataset. Nevertheless, this aggregate performance may obscure the model's weaknesses in handling less dominant or more nuanced labels. Consequently, targeted efforts such as advanced resampling techniques or customized loss functions may be required to enhance the model's discriminative power for these challenging classes. The average values of precision, recall, and F1-score hover around 0.79, reflecting a relatively well-balanced model that does not excessively favor either the detection of true positives or the reduction of false alarms, thereby maintaining equilibrium between sensitivity and specificity across the classification task. Figure 5 presents the classification results obtained using the Gradient Booster model. This figure is intended to demonstrate how this model performs across different cyberbullying categories in terms of precision, recall, and F1-score.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8721
Hal 309-318
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Gradient Booster Report:

	precision	recall	f1-score	support
0	0.95	0.96	0.96	1558
1	0.96	0.96	0.96	1450
2	0.87	0.82	0.85	1455
3	0.54	0.55	0.55	1255
4	0.47	0.51	0.49	964
5	0.93	0.90	0.92	1602
accuracy			0.81	8284
macro avg	0.79	0.79	0.79	8284
weighted avg	0.82	0.81	0.82	8284

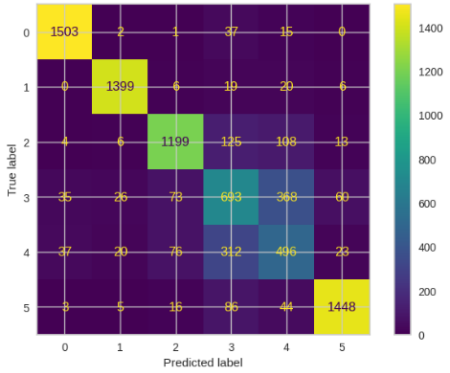

Figure 5. Gradient Booster Result

Figure 5 shows that the Gradient Booster performs well in classifying classes 0 and 1, which correspond to higher-frequency labels such as religion and age. However, the model's performance decreases notably in classes 2, 3, and 4, indicating difficulties in handling less frequent or more ambiguous categories. The overall accuracy of 81% suggests decent general performance, but with room for improvement in handling minority labels. These results emphasize the need for additional strategies to balance prediction performance across all labels.

The XGBoost model exhibits highly commendable performance in the classification of classes 0 and 1, achieving strong metrics in terms of accuracy, recall, and F1-score. These results indicate that the model is particularly well-suited for detecting more prevalent or semantically distinct labels, likely due to its ability to efficiently construct and optimize gradient-boosted decision trees with advanced regularization techniques. Compared to the Gradient Booster, XGBoost consistently outperforms across various evaluation metrics, demonstrating superior robustness and generalization capability. The model achieved an overall accuracy of 83%, which reflects a substantial improvement over its counterparts and highlights its potential as a primary classification engine in multi-label text-based tasks such as cyberbullying detection. Moreover, the model maintains a balanced trade-off between precision and recall at both the macro-average and weighted-average levels, suggesting its competence in handling label imbalances to a certain degree. Nonetheless, a noticeable drop in performance persists in the prediction of classes 2, 3, and 4, which may be attributed to the lower frequency or greater ambiguity of these labels. These deficiencies point to the need for future optimization strategies, such as label-specific feature augmentation, synthetic oversampling, or the incorporation of hierarchical label structures, to further enhance the model's classification effectiveness in underperforming categories. Figure 6 shows the classification results produced by the XGBoost model, one of the most advanced and widely used gradient boosting techniques.

XGBoost Report:

	precision	recall	f1-score	support
0	0.96	0.97	0.97	1558
1	0.97	0.98	0.97	1450
2	0.87	0.84	0.85	1455
3	0.57	0.57	0.57	1255
4	0.49	0.52	0.51	964
5	0.95	0.93	0.94	1602
accuracy			0.83	8284
macro avg	0.80	0.80	0.80	8284
weighted avg	0.83	0.83	0.83	8284

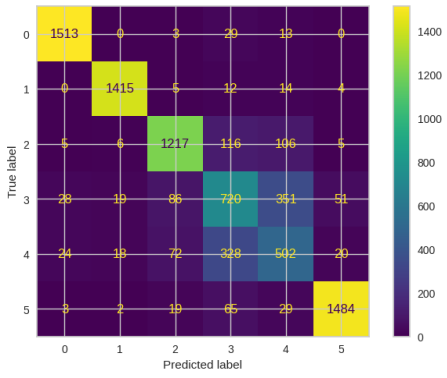

Figure 6. XGBoost Result

Figure 6 indicates that XGBoost achieves strong and balanced performance across all major evaluation metrics. With an overall accuracy of 83%, it outperforms the Gradient Booster and handles class imbalance more effectively. While it performs best in classes 0 and 1, it also shows better resilience than other models in more complex labels such as class 4. This makes XGBoost a reliable candidate for real-world cyberbullying detection systems, especially where overlapping and ambiguous abuse types are common.

The AdaBoost model demonstrates moderate classification performance, particularly in classes 0 and 1, where it shows reasonable levels of precision and recall. This suggests that the model can effectively capture patterns in the data when label distributions are more balanced or linguistically distinguishable. However, as the classification task extends to more complex or underrepresented labels namely classes 2, 3, and especially 4 a noticeable decline in performance is observed. Interestingly, despite this general trend, class 4 registers unexpectedly high values in precision, recall, and F1-score, indicating that AdaBoost may still capture certain high-impact patterns specific to that class. The overall accuracy of the model stands at approximately 78%, which, while lower than that of XGBoost and LightGBM, still reflects a

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8721
Hal 309-318
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

reasonable level of classification capability. Nevertheless, the varying performance across classes reveals a susceptibility of AdaBoost to label imbalance and feature sparsity, both of which are common in multi-label textual data. Given the algorithm's iterative nature and reliance on simple base learners, enhancements such as introducing more complex weak learners or adapting the algorithm through cost-sensitive boosting may be warranted to achieve more stable and consistent classification performance across all categories. Figure 7 illustrates the performance of the AdaBoost model, which represents a classic boosting technique often used for its simplicity and interpretability.

AdaBoost Report:

	precision	recall	f1-score	support
0	0.91	0.96	0.93	1558
1	0.90	0.94	0.92	1450
2	0.82	0.80	0.81	1455
3	0.51	0.57	0.54	1255
4	0.44	0.30	0.35	964
5	0.88	0.90	0.89	1602
accuracy			0.78	8284
macro avg	0.74	0.75	0.74	8284
weighted avg	0.77	0.78	0.77	8284

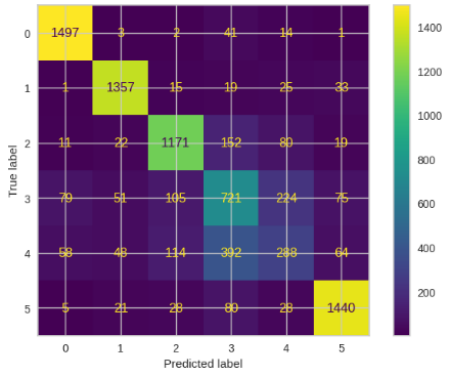

Figure 7. AdaBoost Result

Figure 7 shows that AdaBoost achieves moderate performance, with an overall accuracy of around 78%. It handles dominant labels such as religion and age relatively well but struggles significantly with classes like ethnicity and other_cyberbullying. Interestingly, the model shows an unexpected peak in class 4 performance, possibly due to distinctive patterns captured by its weak learners. However, the results confirm that AdaBoost is more vulnerable to label imbalance and lacks the sophistication of newer boosting frameworks.

The LightGBM model delivers highly competitive results, closely mirroring the performance of XGBoost while exhibiting strengths in efficiency and scalability. It performs particularly well in classes 0 and 1, with high precision, recall, and F1-score values that underscore its effectiveness in capturing dominant patterns from the preprocessed textual inputs. For the more challenging classes 2, 3, and 4 the model still maintains relatively strong classification metrics, surpassing the performance of Gradient Booster and AdaBoost in most cases. With an overall accuracy of 83%, LightGBM proves itself to be a highly suitable candidate for multi-label classification tasks involving large-scale and imbalanced datasets, such as those encountered in cyberbullying detection on social media platforms. The model's histogram-based learning and leaf-wise tree growth mechanism contribute to its faster training time and improved generalization, especially in scenarios with high-dimensional features. While its performance is consistent across macro and weighted average evaluation metrics, the model could benefit further from integrating semantic-rich embeddings or feature selection mechanisms tailored to multi-label structures. As such, LightGBM presents itself not only as a powerful standalone model but also as a promising component in future hybrid architectures combining deep contextual learning with efficient boosting techniques. Figure 8 presents the classification results of the LightGBM model, a modern and efficient boosting algorithm optimized for high-speed training and large-scale data.

LightGBM Report:

	precision	recall	f1-score	support
0	0.97	0.97	0.97	1558
1	0.97	0.97	0.97	1450
2	0.87	0.84	0.85	1455
3	0.57	0.57	0.57	1255
4	0.50	0.55	0.52	964
5	0.95	0.92	0.93	1602
accuracy			0.83	8284
macro avg	0.80	0.80	0.80	8284
weighted avg	0.83	0.83	0.83	8284

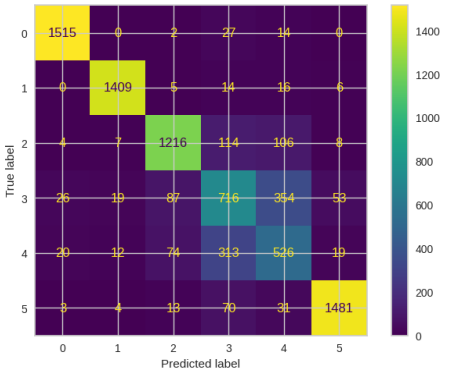

Figure 8. LightGBM Result

Figure 8 shows that LightGBM delivers results comparable to XGBoost, with an accuracy of 83% and strong macro-average performance. It performs consistently across dominant and minority classes and handles class imbalance better than Gradient Booster and AdaBoost. LightGBM's histogram-based learning and leaf-wise growth mechanism enable it to capture complex patterns efficiently. These characteristics position LightGBM as a robust and scalable solution for cyberbullying detection in real-world applications.

Figure 9 provides a detailed comparative analysis of the classification accuracy achieved by four prominent boosting algorithms: AdaBoost, Gradient Booster, XGBoost, and LightGBM. Each bar in the chart visually represents

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 3, Juni 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i3.8721
Hal 309-318
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

the overall accuracy score of a given model, with the x-axis labeling the respective algorithms and the y-axis quantifying the corresponding accuracy percentage. The visualization clearly highlights that XGBoost and LightGBM outperform the other models, each attaining a commendable accuracy level of approximately 83%. This high accuracy not only reflects the superior ability of these models to correctly classify multi-label text data but also underscores their effectiveness in managing the inherent complexities of cyberbullying content, such as overlapping label semantics, context-dependent abuse, and class imbalance.

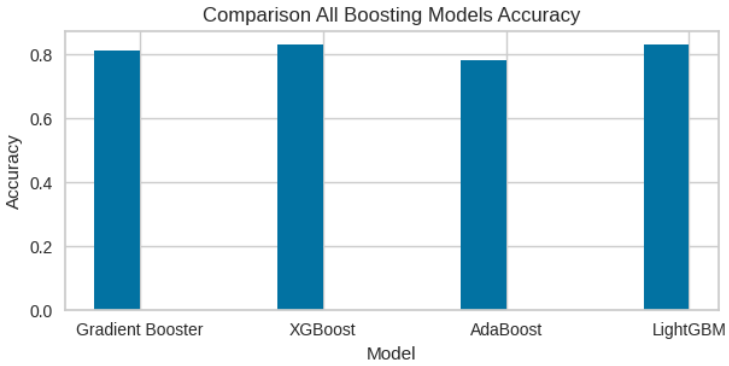

Figure 9. Comparison Boosting Model Result

The strength of XGBoost and LightGBM can be attributed to their sophisticated learning mechanisms. XGBoost leverages regularized gradient boosting with optimized tree structures and shrinkage, allowing it to reduce both variance and bias. LightGBM, on the other hand, employs a leaf-wise tree growth strategy with histogram-based decision rules, enabling faster training and better scalability for large datasets. These advanced techniques provide these models with a strategic advantage in identifying subtle patterns within high-dimensional text data a crucial requirement for accurate cyberbullying detection across multiple abuse categories. In comparison, the Gradient Booster model yields a slightly lower yet still respectable accuracy of approximately 81%. While this performance level is adequate for general-purpose classification tasks, it suggests that Gradient Booster may lack some of the structural efficiencies and regularization benefits offered by its more modern counterparts. Nevertheless, it remains a viable option for multi-label classification where computational resources or simplicity of implementation are priorities. Meanwhile, AdaBoost demonstrates the lowest performance among the four, with an accuracy of around 78%. This relatively modest result may stem from its reliance on sequentially applied weak learners (typically decision stumps) and its vulnerability to noisy or imbalanced data distributions common characteristics in real-world social media text datasets.

In conclusion, the comparative accuracy results affirm that XGBoost and LightGBM are the most suitable choices among the evaluated models for the task of multi-label cyberbullying detection. Their combination of high predictive accuracy, resilience to noise, and computational efficiency positions them as powerful tools for real-world deployment in automated moderation systems. Future work may explore hybrid strategies that combine these boosting techniques with deep contextual language models (e.g., BERT or RoBERTa) to further enhance semantic understanding and predictive robustness, particularly for underrepresented abuse categories.

3.3 Discussion

The findings of this study demonstrate that both XGBoost and LightGBM outperform the Gradient Booster and AdaBoost models, with accuracy scores reaching 83%. This superiority suggests that modern boosting algorithms, designed with computational efficiency and advanced tree pruning strategies, are more effective in handling the complexity and variability of cyberbullying text features. Moreover, both models maintain a strong balance between precision and recall, particularly in high-frequency labels such as "religion" and "age." These results indicate that XGBoost and LightGBM are not only capable of accurately detecting dominant labels but also exhibit resilience against label imbalance an inherent challenge in multi-label classification tasks. Although the overall performance is promising, the models demonstrate relatively lower recall scores for minority labels such as "other_cyberbullying" and "not_cyberbullying." This suggests a model bias toward majority labels, likely due to their higher representation in the dataset. Nevertheless, LightGBM shows improved performance in certain minority classes compared to other models, indicating its potential for handling imbalanced label distributions. Future enhancements may include advanced resampling techniques, synthetic data generation methods such as SMOTE, or the implementation of cost-sensitive loss functions to better capture the characteristics of underrepresented labels in multi-label classification.

Compared to previous studies, the findings of this research demonstrate both alignment and advancement in the field of cyberbullying detection using machine learning. Previous works [3][4],have highlighted the challenges in identifying cyberbullying content, especially due to imbalanced labels and the subtlety of offensive language in social media text. While ensemble-based models were recognized for their potential, their application in multi-label classification settings was limited. This research addresses that limitation by implementing and comparing multiple boosting algorithms in a multi-label context, demonstrating that XGBoost and LightGBM outperform other models in both accuracy and label balance. Earlier studies [4], [7] primarily focused on binary classification or employed basic

machine learning models, which may not effectively capture the complexity of cyberbullying behavior that spans multiple overlapping categories. In contrast, this study provides a more comprehensive evaluation tailored to the realities of multi-label cyberbullying detection in dynamic social media environments.

Other comparative studies [10], [11], [13], [14] have explored multi-label classification or boosting methods independently but did not combine both aspects in a cyberbullying-specific context. While some research has introduced customized boosting frameworks or incorporated classifier chains to enhance multi-label learning, the results presented here show that standard implementations of XGBoost and LightGBM can already deliver strong performance when applied with appropriate preprocessing and evaluation strategies. Furthermore, studies on imbalanced datasets in other domains have shown that modern boosting models consistently achieve superior results compared to traditional approaches, which aligns with the findings of this work. By applying and evaluating these models specifically for cyberbullying detection on social media, this study provides empirical evidence of their effectiveness and offers practical insights for developing automated moderation systems capable of handling real-world, multi-label abusive content.

The use of four distinct boosting models in this research provides a broad perspective on classifier selection for multi-label cyberbullying detection. The methodological strength lies not only in the variety of models evaluated but also in the comprehensive preprocessing pipeline implemented. Techniques such as normalization, lemmatization, and special character filtering significantly reduce noise, ensuring cleaner and more informative model inputs. The division of data into train-test and train-validation subsets also supports a robust model evaluation framework. These methodological advantages highlight a replicable and structured approach applicable to other multi-label text classification domains, particularly in socially sensitive areas like digital abuse detection. This study paves the way for several future research trajectories. First, integrating boosting models with transformer-based deep learning architectures such as BERT or RoBERTa in a hybrid ensemble format could enhance semantic understanding and classification accuracy. Second, the incorporation of explainable AI (XAI) frameworks like SHAP or LIME would contribute to interpretability, especially in content moderation scenarios. Third, adopting adaptive boosting frameworks built on classifier chains could exploit label dependencies more effectively, thus improving prediction in multi-label settings. Lastly, extending this work to other platforms such as Instagram or TikTok could broaden model generalizability and address emerging forms of cyberbullying. Consequently, this research represents a foundational effort toward building robust, adaptive, and ethically sound cyberbullying detection systems.

4. CONCLUSION

Cyberbullying has been identified as abusive behavior in the digital world, involving the use of degrading, aggressive, and threatening communications. The focus of this research is to build and improve a cyberbullying detection system, aiming to analyze and overcome online bullying through social media. The results of this study indicate that the XGBoost and LightGBM models achieve higher accuracy compared to Gradient Booster and AdaBoost. XGBoost and LightGBM dominate in terms of predictive performance, particularly in handling multi-label classification challenges. Meanwhile, Gradient Booster and AdaBoost models show slightly lower performance. However, model selection should not rely solely on accuracy. Other factors such as interpretability, training speed, scalability, and the specific requirements of the application should also be considered when choosing the most suitable algorithm. The limitations of this study particularly the lack of focus on model explainability open opportunities for further exploration. This research contributes by providing a comparative evaluation of state-of-the-art boosting methods for multi-label cyberbullying sentiment classification, an area that is still underexplored. These insights can support researchers and developers in selecting more effective machine learning approaches tailored to real-world online abuse detection challenges. Thus, through continuous efforts to develop and refine cyberbullying detection systems, this research is expected to contribute significantly to the prevention of online harassment. Ultimately, these systems can support the creation of a safer and more positive digital environment for social media users.

REFERENCES

[1] J. Li, G. Huang, C. Fan, Z. Sun, And H. Zhu, “Key Word Extraction For Short Text Via Word2vec, Doc2vec, And Textrank,” Turkish Journal Of Electrical Engineering And Computer Sciences, Vol. 27, No. 3, Pp. 1794–1805, 2019, Doi: 10.3906/Elk-1806-38.

[2] W. Medhat, A. Hassan, And H. Korashy, “Sentiment Analysis Algorithms And Applications: A Survey,” Ain Shams Engineering Journal, Vol. 5, No. 4, Pp. 1093–1113, Dec. 2014, Doi: 10.1016/J.Asej.2014.04.011.

[3] M. A. Al-Garadi Et Al., “Predicting Cyberbullying On Social Media In The Big Data Era Using Machine Learning Algorithms: Review Of Literature And Open Challenges,” Ieee Access, Vol. 7, Pp. 70701–70718, 2019, Doi: 10.1109/Access.2019.2918354.

[4] A. Muneer And S. M. Fati, “A Comparative Analysis Of Machine Learning Techniques For Cyberbullying Detection On Twitter,” Future Internet, Vol. 12, No. 11, Pp. 1–21, Nov. 2020, Doi: 10.3390/Fi12110187.

[5] D. W. Hosmer, Stanley. Lemeshow, And R. X. Sturdivant, Applied Logistic Regression.

[6] F. Farasalsabila, E. Utami, And M. Hanafi, “Analysis Of Public Opinion On Indonesian Television Shows Using Support Vector Machine,” Jurteksi (Jurnal Teknologi Dan Sistem Informasi), Vol. 10, No. 2, Pp. 239–246, Mar. 2024, Doi: 10.33330/Jurteksi.V10i2.2935.

[7] V. S. Chavan And S. S. S, Machine Learning Approach For Detection Of Cyber-Aggressive Comments By Peers On Social Media Network. 2015.

[8] F. Farasalsabila, E. Utami, And H. Hanafi, "Deteksi Cyberbullying Menggunakan Bert Dan Bi-Lstm," J Teknol, Vol. 17, No. 1, May 2024, Doi: 10.34151/Jurtek.V17i1.4636.
[9] F. Farasalsabila, E. Utami, And S. Raharjo, "Multi-Label Classification Using Bert For Cyberbullying Detection."
[10] J. Bogatinovski, L. Todorovski, S. Džeroski, And D. Kocev, "Comprehensive Comparative Study Of Multi-Label Classification Methods," Expert Syst Appl, Vol. 203, Oct. 2022, Doi: 10.1016/J.Eswa.2022.117215.
[11] J. Li, X. Zhu, And J. Wang, "Adaboost.C2: Boosting Classifiers Chains For Multi-Label Classification," 2023. [Online]. Available: Www.Aaai.Org
[12] G. Ke Et Al., "Lightgbm: A Highly Efficient Gradient Boosting Decision Tree." [Online]. Available: Https://Github.Com/Microsoft/Lightgbm.
[13] S. Rahman, M. Irfan, M. Raza, K. M. Ghori, S. Yaqoob, And M. Awais, "Performance Analysis Of Boosting Classifiers In Recognizing Activities Of Daily Living," Int J Environ Res Public Health, Vol. 17, No. 3, Feb. 2020, Doi: 10.3390/Ijerph17031082.
[14] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, And M. Asadpour, "Boosting Methods For Multi-Class Imbalanced Data Classification: An Experimental Review," J Big Data, Vol. 7, No. 1, Dec. 2020, Doi: 10.1186/S40537-020-00349-Y.
[15] Jason Wang, Kaiqun Fu, And Chang-Tien Lu, "Fine-Grained Balanced Cyberbullying Dataset," 2020.
[16] A. Rafid Rizqullah, A. Wedhasmara, R. Izwan Heroza, A. Putra, And P. Putra, "Analisis Masalah Pada Data Review Aplikasi Terhadap Layanan E-Commerce Menggunakan Metode Text Classification," 2023.
[17] D. D. Nur Cahyo Et Al., "Sentiment Analysis For Imdb Movie Review Using Support Vector Machine (Svm) Method," Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi, Vol. 8, No. 2, Pp. 90–95, Mar. 2023, Doi: 10.25139/Inform.V8i2.5700.
[18] P. Florek And A. Zagdański, "Benchmarking State-Of-The-Art Gradient Boosting Algorithms For Classification," May 2023, [Online]. Available: Http://Arxiv.Org/Abs/2305.17094
[19] H. Mulyo And A. Khanif Zyen, "Bulletin Of Computer Science Research Pengaruh Hyperparameter Tuning Gradient Boosting Terhadap Prediksi Pemilihan Program Studi Mahasiswa Baru," Media Online), Vol. 5, No. 2, Pp. 131–137, 2025, Doi: 10.47065/Bulletincsr.V5i2.454.
[20] M. Rama Hadi Suryanto And D. Wahyu Utomo, "Pembelajaran Ensemble Untuk Klasifikasi Ulasan Pelanggan E-Commerce Menggunakan Teknik Boosting," Vol. 15, No. 02, 2024, Doi: 10.35970/Infotekmesin.V15i2.2314.
[21] A. Mayr, H. Binder, O. Gefeller, And M. Schmid, "The Evolution Of Boosting Algorithms: From Machine Learning To Statistical Modelling," Methods Inf Med, Vol. 53, No. 6, Pp. 419–427, 2014, Doi: 10.3414/Me13-01-0122.
[22] N. Ritha Et Al., "Sentiment Analysis Of Health Protocol Policy Using K-Nearest Neighbor And Cosine Similarity," In Icsedti 2022, European Alliance For Innovation N.O., Jan. 2023. Doi: 10.4108/Eai.11-10-2022.2326274.
[23] E. Beauxis-Aussalet And L. Hardman, Simplifying The Visualization Of Confusion Matrix. 2014.