THE SCIENTIFIC TEMPER (2025) Vol. 16 (12): 5146-5156

E-ISSN: 2231-6396, ISSN: 0976-8653

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.12.01

https://scientifictemper.com/

RESEARCH ARTICLE

Cyberbullying Detection Using Continuous Based Bag of Words with Machine Learning by Text Classification

P. Vivekananth[1*], Navneet Sharma[2]

Abstract

The breakneck advancement in internet for Social Media (SM) have generated enormous text data that became a challenging as well as valuable task in identifying an adequate measure to analyze text data using machine. Natural Language Processing (NLP) technique is one of the text classification methods that applicable for several applications sectors such as e-commerce and customer service. Bulling over SM for individuals have resulted with calumny, chastise and threatening. This kind of cyberbullying generates increase in serious mental health issues particularly for young generation which resulted to lessened self-esteem as well as increase of suicidal reflection. A generation of young adults will be affected by mental health and self-esteem problems, if action is not taken to stop cyberbullying. However, cyberbullying has become the ultimate challenge for Artificial Intelligence (AI) studies as well as more beneficial in the real-life applications. Therefore, initial step of the machine is for understanding the text by text representation whereas the most preferable method is Bag-of-Words (BoW). This paper has proposes a Continuous Based BoW (CBBoW) method assist to perform better significance for minimizing the training time requirement and even accomplish the training accuracy rate. The results determine that suggested method accomplishes performance with best accuracy in detection of cyberbullying words. The suggested techniques are tested using conventional BoW and Word2Vec approaches on open-source datasets with predetermined data partitions provided accessible through an open digital repository to promote replication.

Keywords: Machine Learning; Social Media; Natural Language Processing; cyberbullying; Bag of Words.

Introduction

SM has permitted users in communicating and sharing with enormous individuals right away, at any time, and with anyone. The global population of SM users exceeds 3 billion. National Crime Security Council (NCPC) define cyberbullying in any online behaviour in which someone purposefully hurts or embarrasses another person through the use of a mobile device, a video game app, or for various communications. The cause of cyberbullying in internet may happen at any time and has been accessed by anyone as well as anywhere in the world. Cyberbullying may be documented in an anonymous way using text, images, or videos. The source of the post may be hard to find at times even untraceable.

Furthermore, it is difficult in deleting those messages over future. Online bullying occurs most frequently on SM sites including Wikipedia, Snapchat, Facebook, YouTube, Instagram, Twitter, and Skype. A few SM platforms, includes Facebook, as well as advice provision on bullying prevention in which the dedicated section assist in reporting cyberbullying and keep the user from being blocked. Users have the option to monitor or block users who upload images and videos on Instagram that they find uncomfortable. One of the most concerning issues that have emerged as an effect of the widespread use and potency of these platforms is cyberbullying. Cyberbullying is a type of abuse or harassment that people with psychotic symptoms inflict on frequent users of these networks. A few forms of harassment include hate speech, verbal abuse,

[1]PhD Research Scholar, Department of Computer Science, IIS Deemed to be University, Gurukul Marg, SFS, Mansarovar, Jaipur, India.

[2]Associate Professor, Department of Computer Science, IIS Deemed to be University, Gurukul Marg, SFS, Mansarovar, Jaipur, India.

***Corresponding Author:** P. Vivekananth, PhD Research Scholar, Department of Computer Science, IIS Deemed to be University, Gurukul Marg, SFS, Mansarovar, Jaipur, India, E-Mail: vivekanandhan2024@outlook.in

How to cite this article: Vivekananth, P., Sharma, N. (2025). Cyberbullying Detection Using Continuous Based Bag of Words with Machine Learning by Text Classification. The Scientific Temper, **16**(12):5146-5156.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.12.01

Source of support: Nil

Conflict of interest: None.

Received: 02/11/2025 **Accepted:** 18/11/2025 **Published:** 25/12/2025

and blackmail. Although cybercrime and bullying have long been prevalent issues in society, abusers have more ways to hide their identities behind virtuality, as seen by the rising incidence of these crimes (Raj, C., *et al.,* 2021).

This inverse trend unmistakably indicates a trend toward an increase in cyberbullying incidents as the number of research subjects' declines. Children and those under the age of eighteen made up more than thirty percent of the dead. Strict procedures are needed to evaluate the safety of these young people on SM platforms, since their lack of understanding *often* leads them to engage in such harmful online behaviours. Consequently, it becomes imperative to utilize a certain automated method that is capable of identifying such heinous incidents without depending on or subjecting to human review. These models seem to work really well with such small datasets. They are unable to handle extremely noise data as well as produced with various accuracy level values while applying it to real time datasets, like the ones employed in this work (Kumar, A., & Sachdeva, N. 2022). Cyberbullying (Ahmed, S., & Rajput, A. E. 2023; is the deliberate, repetitive, aggressive, and abusive behaviour directed against an individual or group of individuals through the use of digital media to post offensive information or participate in other types of social violence (Rosa, H., *et al.,* 2019). Teenagers are highly susceptible to cyberbullying, either as the victim, the offender, or a witness. The main challenge addressed by this research is evaluating the effects of modern intelligent methods on extremely complex as well as unstructured real time datasets that have been obtained from worldwide top researchers. This has been employed with two different kinds of datasets namely Twitter and Instagram (Kumari, K., & Singh, J. P. 2021) for describing the general methodology of proposed model, and assist with text-based classification challenges for other platforms also.

Text Classification

The process of giving text units, such as documents, SM posts, or articles from the news, categorized topics is known as text categorization. Text categorization research is a tremendously dynamic topic, as seen by the sheer number of innovative techniques reported in recent surveys (Zhou, X., *et al.,* 2020). One of the key ideas in Machine Learning (ML) is text classification (Ramalingam, G., *et al*., 2024). Many applications, such as spam filtering, emotion analysis, intention mining, etc., are related to text classification. This study includes sentiment analysis too whereas the BoW model is one of the most prevalent techniques for sentiment analysis. One of the better models for representing text data is the BoW model. Text is the format used for all online content, including Wikipedia, Gmail, and other resources. The BoW model categorizes this text and extracts certain features. Text preprocessing, feature selecting, extracting features, calculating similar, and determining the classifier are all common steps in text classification research (Deng, X., *et al.,* 2019). Although it is normal for individuals to determine if a document directly pertains to a certain topic by reading and interpreting it due to the benefit of knowing human language, a machine cannot do this procedure. Thus, text representation is the first step in a computer's text classification process as it converts text data towards a format is suitable to computer processing.

This work aims to conduct an extensive experimental investigation on integrating alternative for the sentimental classification to the text of Brazilian Portuguese that considered from general and conventional selection of models, in five freely available databases to guarantee generalization as well as replication of the results. The task is annoyed by lack of experimental research involving most current NLP frameworks and the increasing significance of user's assumptions from huge data. While several publications stress the value of neutrality in sentiment analysis, this work assume a binary sentiment analysis job in this case by concentrating just on positive and negative evaluations and ignoring the neutral ones (Valdivia, A., *et al.,* 2018).

From past decade, researchers have been exploring ML as well as DL techniques for text classifications for categorizing SM content as bullying or non-bullying. In general, researchers initially focus on character-level representations such as BoW used supervised ML techniques in conjunction with NLP approaches.

Chia, Z. L., *et al.,* (2021) have investigated the DL domain for the intricacy and efficacy on huge datasets, with more opportunity for development. Goal of the author was to detect ironic and sarcastic language in a dataset and it was taken from Twitter. Techniques from feature engineering and ML were applied to detect these data. Many techniques were employed to get an optimal results, beginning with the Naïve Bayes (NB) main classifier and moving on to Support Vector Machine (SVM), Convolutional Neural Network (CNN), and k-Nearest Neighbour (k-NN) to reduce errors in identifying cyberbullying texts. Yuvaraj, N., *et al.,* (2021) have created unique automated classification approach that did not need the data to be fitted into a sizable dimensional space. In order to minimize overfitting, the innovative deep decision tree classifier's limitation of tree depth was addressed by combining an upgraded decision tree classifier with DNN in the classification model. The authors were motivated to improve optimization by investigating DL models and assessing their dependability on large amounts of real-world data, even though this method has demonstrated a high degree of accuracy in detecting cyberbullying. Raj *et al*. have advanced the model that identifies cyberbullying in tweets and other SM messages through the Deep Neural Networks (DNNs) usage. When compared to traditional methods, DNNs are successful.

Balakrishnan, V., *et al.,* (2020) have attained this accuracy by utilizing occurrence based training and Decision Tree (DT) which enhances the cyberbullying detection. In this paper, the author includes personalities and sentiment as the features. It was also possible to identify cyberbullying using a number of DL-based algorithms. The DNN-based model is used to identify cyberbullying using real-world data. Al-Ajlan, M. A., & Ykhlef, M. (2018) employed transfer learning to accomplish the detection job after first conducting a systematic evaluation of cyberbullying. To identify cyberbullying, a model based on CNNs has been presented. However, this author used word embedding as words with similar meanings consists of same embeddings. Hence, the issue gets addressed by suggested XBully, a novel cyberbullying detection method is initiated by reformulating multi-modal SM data as a diverse network as well as attempted to learn node for embedding representations.

Roy, P. K., *et al.,* (2020) have discussed about Deep CNN (DCNN) benefits in creating request for detecting hate kind of speech on Twitter. ML algorithms utilized in identifying the tweets associated with hate speech in Twitter. Features are removed by one of the NLP technique is TF-IDF method. These algorithms include SVM, NB, Logistic Regression (LR), DT, Gradient Boosting (GB), Random Forest (RF), and KNN. The identified best ML model is SVM, although in a 75%:25% dataset has been utilized for test the train which was capable in predicting 53% of tweets with hate speech. Data with inconsistent was the cause for low prediction scale. Thus, the model is predicated on the identification of hateful tweets. The same results can be obtained from advanced learning methods based on CNN Long-Term Memory (LSTM) as well as their Contextual LSTM (CLSTM) combination as from independent distributed databases. By combining the proposed DCNN model with 10-fold cross validation with high recall rate was achieved.

The complex network based BoW technique is utilized to text representation. Due to its consideration of word correlations represented in the text network, the Attribute of Network Extended BoW (AEBoW) technique improves upon the BoW technique. When every connection between words that make up an edge are taken into account, the organization of a text network changes. Yan, D., *et al.,* (2020) have presented the concepts of the static and dynamic networks. Additionally, a hybrid network have incorporates relationships from both the static and dynamic networks is suggested. The effectiveness of AEBoW in text categorization has been compared against seven text representation techniques. According to experimental findings, the suggested AEBoW could achieve the highest efficiency and optimal performance. The shortest-path-based text network attribute known as eccentricity was the finest aspect of AEBoW. Soumya, S., & Pramod, K. V. (2020) have presented a sentiment analysis experiment using tweets written in Malayalam. Several ML classifiers, including RF, SVM, and NB were used by the researchers to categories the tweets. Furthermore, the researchers introduced two word embedding techniques: BoW and Term Frequency-Inverse Document Frequency (TF-IDF). The lack of sentiment-tagged Malayalam corpus presents an obstacle for this researcher. SM users have lately recognized that cyberbullying became severe issues of public health as well as developed an efficient detection algorithm with substantial scientific value. Al-Garadi, M. A., *et al.,* (2016) have presented a set of particular features that are obtained from Twitter, such as user behaviour and tweet content. An approach to the identification of cyberbullying on Twitter-based networks has been developed: supervised ML.

Kumar, Y. J. N., *et al.,* (2024) focused on identifying cyberbullying on SM, and taken a thorough approach that incorporates ML, text analysis, and group tactics. The goal is to build a resilient system that can adjust to the ever-changing environment of online interactions rather than just provide a one-time fix. This strategy, which is based on ethical principles, gives ethical and appropriate cyberbullying detection as top priority. Essentially, this effort is a commitment to creating a more secure and inclusive digital world rather than merely a technical feat. The primary objective is to make a significant contribution to the continuous efforts to prevent cyberbullying and encourage beneficial online interactions by merging technology innovation with ethical concerns and a commitment to continuous development. This study is evidence of the effectiveness of multidisciplinary methods in tackling complex societal problems in the context of digital technology. However, DNNs basically outperform conventional methods. A systematic review (Balakrishnan, V., & Kaity, M. 2023) highlights determined challenges in leveraging ML to cyberbullying detection includes multi-language platforms as well as unknown areas in unsupervised learning. The undeveloped and complex cyberbullying nature continues to challenge yet advanced ML algorithms but the frequent question is whether traditional ML, DL like Transformers-based models, or pre-trained LLMs offer the best balance among reliability as well as simplicity. Utilizing conventional feature extraction and leveraging word vectors to keep up tweet semantics, cyberbullying detection can be improved. For best results, a CNN's parameters were adjusted through an optimization technique. A multi-layer CNN was used in an additional investigation. Further, current neural architectures have been investigated for the identification of cyberbullying across various languages (Obaid, M. H., *et al.,* 2023).

Need for Study

However, previous research has worked on text classification using NLP technique and hyperparameter tuning the ML methods but individually in detecting and classifying

cyberbullying texts. These traditional methods have failed to provide better accuracy in their techniques. Hence, to accomplish higher accuracy and precision in detecting cyberbullying texts a modified version of BoW is proposed for considering hypertuning concept within this model through Continuous based technique which cannot be found in the previous literature. The research aims to compare the achieved accuracy using the modified methods with the traditional BOW and word2vec techniques. This research work identified to bridge the gap by achieving higher accuracy using modified techniques for classifying cyberbullying texts and ensure earlier prediction. Thus, the research concentrated as follows

To improve cyberbullying detection by text classification using NLP libraries and modifying the probability condition of neighbor words in the BOW technique is named as CBBOW.

To determine the proposed CBBOW model through evaluation metrics with various ML method for generating better prediction of cyberbullying classification detection.

Materials and Methods

Generally, this research focuses on adopting the text classification pipeline for implementing the different kind of embedding to improve the better detection of bulling words and its working procedure is discussed. Initially, the tweets are considered as the collected data from user reviews and data cleansing is done by removing the punctuation. Once it gets completed, the text data is further split in term of textual sub-units is said to be token. The importance of lower semantic tokens has been extracted based on the method under analysis. Once, the single feature vector has been generated for feeding the model classification. This research concentrated in accomplishing significant features to text classification, whereas the experiments have provided the similar classifier considered through LR for accessing the performance of embedding to obtain an alternate analysis. In the case of large size corpus, previous researcher has faced major issues and more time consumption during training. The suggested CBBoW method has introduced to sort out the above mentioned issue. This research implements ensemble method with voting classifier as the specific reflection in improving the cyberbullying detection. There are two other NLP methods like BoW and Word2Vec is considered for representing the word as vectors. However, the model principle and its extensions have been illustrated through comparison based model as well as its extension hybrid with various methods. Hence, the paper concentrated in developing automated cyberbullying detection in detection of "Cyberbullying" or "Non cyberbullying" classes for users on Twitter (social media). This may perform in misleading comments, once it gets detected and classified through CBBOW with ensemble model.

Dataset collection and data preprocessing

As SM becoming increasingly ubiquitous for day to day communication across all ages, the majority of citizens rely on it. The goal of data collection is to highlight the problems associated with cyberbullying. On April 15, 2020, UNICEF issue a warning about the increasing cyberbullying risk during the COVID-19 pandemic because of prevalent school closures, which increases screen period and reduces in-person social interaction. Statistics showing that 36.5% of schoolchildren report having been the victim of cyberbullying, and 87% of middle as well as high school students report the victim has experienced depressive symptoms, suicidal thoughts, and a decrease in academic performance resulted with cyberbullying raise alarms about the problem of cyberbullying. This study used 47692 tweets with various cyberbullying classes as substitutes for the form of cyberbullying illustrated in Figure 1.

The real-time datasets used in this study contain enormous quantities of noise that are dynamic nature results with different SM platforms. The dataset dependability and quality have been guaranteed thorough data pretreatment step that was carried out prior to the application of multivariable analytic techniques. To do this, the social media content had to be cleaned and filtered to get rid of clutter, pointless information, and repetitive posts. To normalize the textual data, we also performed text normalization, stop word removal, and stemming. The approach supports text based classification for identifying cyberbullying incident, as well as NLP is commonly employed to support the preprocessing stage of data. Initially, the dataset get transformed into data frame, which made editing them simpler.

This research concentrated on detecting the cyberbullying tweets from the overall tweets of the twitter SM. The experimental study initially removes the punctuation, numbers, Http links and the emoji from the tweet text. In order to improve the bulling words, the frequent words selection is considered through CBBoW as feature extraction method which is done after word tokenization. This CBBoW technique improves the frequent word usage present in the tweet texts that assist in building hyperparameter (θ) in BoW method. Once feature extracted is done, the NLP model is evaluated through sentiment status and its confident score and evaluated with various ML method for determining classification class detection through confusion matrix metric. The several ML methods are ensemble through voting classifier by its various meta-prediction methods and the proposed hybrid combination of NLP with ML has determined better detection of cyberbullying in the SM. The workflow of the proposed hybrid model is illustrated in the Figure 2.

CBBoW working principle

CBBoW approach has predicated based on the concept of words that occur near together in the text should have

tweet_text	cyberbullying_type
In other words #katandandre, your food was crapilicious! #mkr	not_cyberbullying
Why is #aussietv so white? #MKR #theblock #ImACelebrityAU #today #sunrise #studio10 #Neighbours #WonderlandTen #etc	not_cyberbullying
Ian Shit Ain't It The Truth . I Hate You 2 !	other_cyberbullying
@XochitlSuckkks a classy whore? Or more red velvet cupcakes?	not_cyberbullying
@Jason_Gio meh. :P thanks for the heads up, but not too concerned about another angry dude on twitter.	not_cyberbullying
I want to just slap that smug look off Kats face. #annoying #mkr @mykitchenrules	gender
@RudhoeEnglish This is an ISIS account pretending to be a Kurdish account. Like Islam, it is all lies.	not_cyberbullying
@Raja5aab @Quickieleaks Yes, the test of god is that good or bad or indifferent or weird or whatever, it all proves gods existence.	not_cyberbullying
@PressTV Like every Muslim country, they find money for weapons, but they don't have the money for schools.	religion
Itu sekolah ya bukan tempat bully! Ga jauh kaya neraka	not_cyberbullying
Karma. I hope it bites Kat on the butt. She is just nasty. #mkr	not_cyberbullying
@stockputout everything but mostly my priest	not_cyberbullying
Rebecca Black Drops Out of School Due to Bullying:	not_cyberbullying
@Jord_Is_Dead http://t.co/UsQInYW5Gn	not_cyberbullying
The Bully flushes on KD http://twitvid.com/A2TNP	not_cyberbullying
Ughhhh #MKR	not_cyberbullying
RT @Kurdsnews: Turkish state has killed 241 children in last 11 years http://t.co/JlvkE1epws #news ##GoogleÃ‡eviriciTopluluÄŸuKÃ¼rtÃ§eyideEâ€¦	not_cyberbullying
Love that the best response to the hotcakes they managed to film was a non-committal "meh" from some adolescent. #MKR	not_cyberbullying
@yasmimcaci @Bferrarii PAREM DE FAZER BULLYING COMIGO =(UHAHUAH BANDO DE PRETO	not_cyberbullying
@sarinhacoral @Victor_Maggi tadinhu de mim , sofrendo bulling viu MIMI'	not_cyberbullying
@0xabad1dea @kelseytheodore2 twitter is basically the angry letters of our generation.	not_cyberbullying
Best pick up line? Hi, you're cute... ?: I love how people call James Potter is a bully. - mypatronusisyou: http://tumblr.com/xol3xl14zy	not_cyberbullying
Now I gotta walk to classss?! I officially hate the stupid bus system! -_-	not_cyberbullying

Figure 1: Dataset of tweets and its cyberbullying classes

very similar meanings, but words that appear far apart typically have different meanings. Consequently, given the circumstances surrounding the formation of neighboring words, a center word is anticipated to develop. In general, the CBoW model initial setting of its own known parameters and the sentences mentioned using one-hot word vectors. The one-hot vectors as an input is represented as $x^{(s)}$ and output as $y^{(s)}$. Since, this model consists of only one output, the one hot vector of known center word is represented as y. According to CBBoW approach, the unknown can be defined by creating two matrices,

$$A \in \mathbb{R}^{n \times |V|} \text{ and } B \in \mathbb{R}^{|V| \times n}$$

Where,
n = arbitrary size that defines the embedding space size
A = Input word matrix
B = Output word matrix
V = vector for input words and output words
The i^{th} column of A with n-dimensional embedded vector for words (w_i) and the j^{th} row of B is an n-dimensional embedded vector for words (w_j). For each word, two vectors are learned that is defined as input word vector as A_i and output word vector B_i.

Algorithm of CBBoW method

Step 1: Generate one hot word vectors $\left(x^{(s)},\ldots,x^{(s-1)},x^{(s+1)},\ldots,x^{(s+m)}\right)$ for the input context size as m.
Step 2: The word vector gets embedded for the context

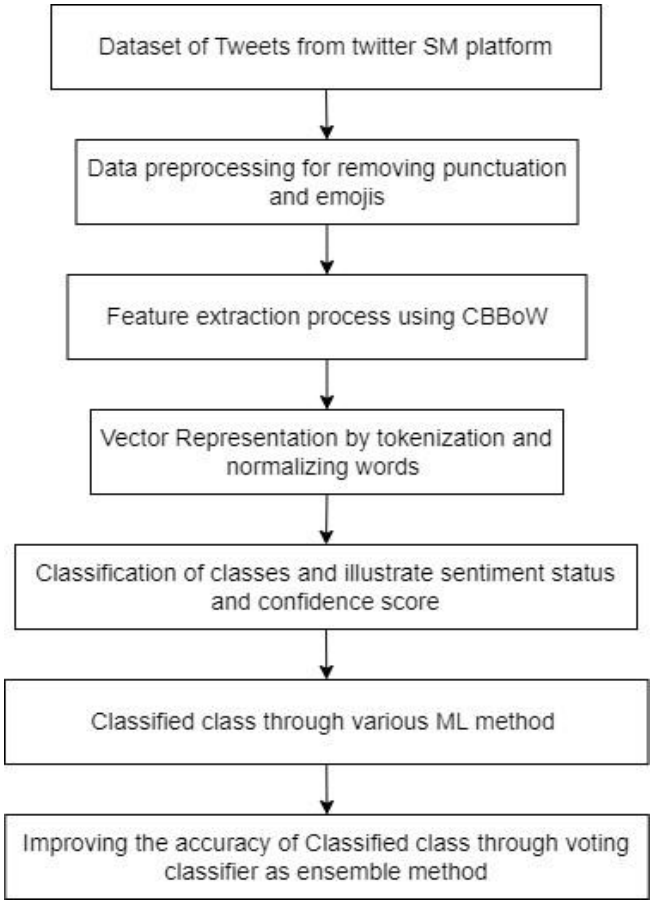

Figure 2: proposed flow diagram of hybrid model for detecting cyberbullying in SM

$$\left(v_{s-m}=Ax^{(s-m)},v_{s-m+1}=Ax^{(s-m+1)},\ldots,v_{s+m}=Ax^{(s+m)}\right)$$

Step 3: Average the vectors $v=\left(v_{s-m}+v_{s-m+1}+\ldots+v_{s+m}\right)\div\left(2m\right)$

Step 4: Generating the score vector as $\ddot{u}=$ ˆ

Step 5: The scores are measured and converted into probabilities $\hat{y}=Softmax\left(z\right)$

Step 6: Finally the generated probabilities $\hat{y}$ is to match the true probability y which is considered to be actual word one hot vector.

Based on the above concept, the probability can be learned from certain true probability whereas the information theory may provide information theory for calculating the distance among two distributions. The best choice may be calculated through cross entropy H(y, $\hat{y}$). The novel approach is determined by considering the cross entropy usage for discrete cases and can be determined through loss functions shown in equation 1.

$$\mathrm{H}\left(\mathrm{y},\hat{y}\right)=-\sum_{\mathrm{j}=1}^{|\mathrm{v}|}\mathrm{y}_{\mathrm{j}}\log\left(\hat{\mathrm{y}}_{\mathrm{j}}\right) \quad (1)$$

However, the word embedding involved in this approach with 'y' as one-hot vector. Hence, the loss can be simplified is shown in equation 2.

$$\mathrm{H}\left(\mathrm{y},\hat{y}\right)=-\mathrm{y}_{\mathrm{j}}\log\left(\hat{\mathrm{y}}_{\mathrm{j}}\right) \quad (2)$$

Moreover, the formulation has considered the index as's' for correct word's one hot vector is 1. Therefore, the prediction will be perfect and $\hat{\mathrm{y}}_{\mathrm{s}}$ =1 and it can be calculated as $\mathrm{H}\left(\mathrm{y},\hat{y}\right)=-1\log(1)$ = 0. Thus, the perfect prediction does not buzzed with any loss. In contrast, prediction is considered to be bad and $\hat{\mathrm{y}}_{\mathrm{s}}$ =0.01. The loss may be calculated as $\mathrm{H}\left(\mathrm{y},\hat{y}\right)=-1\log(0.01)$ ~4.605. The probability distribution with cross entropy provides a good distance measures. The optimization objective is formulated in equation 3.

$$Minimize\,J=-\log P\left(w_s|w_{s-m},\ldots,w_{s-1},w_{s+1},\ldots,w_{s+m}\right)$$

$$=-\log P\left(B_s|\hat{v}\right)$$

$$=-\log\frac{expB_s^T\hat{v}}{\sum_{j=1}^{|A|}expB_s^T\hat{v}}$$

$$=-B_s^T\hat{v}+\log\sum_{j=1}^{|A|}exp(B_s^T\hat{v}) \quad (3)$$

Thus, the use of gradient descent to update all relevant word vectors B_{s} and A_{j}.

Working of LGBM classifier

One of the gradient boosting frameworks is LGBM that completely relies on Decision Tree (DT) for improving the model efficiency and minimize the usage of memory. This method includes two methods namely Exclusive Feature Bundling (EFB) as well as Gradient based One Side Sampling (GOSS) in which mean of GOSS and EFB with Gradient Boosted Decision Tree (GBDT) in which GDBT is expressed in equation 3.

$$F\left(x,w\right)=\sum_{t=0}^{T}\alpha_t h_t\left(x,w\right) \quad (4)$$

Where,

F( ) = GBDT predictive value

$\mathrm{h_t}$( ) = $\mathrm{t^{th}}$ DT method function

w = Parameters for DT

x = Samples for input

α = each tree weight

If the loss function minimization L( ) for mapping the space x and y. The optimal model get resolved is illustrated in equation 4.

$$\hat{F}=argarg\,min\;F\sum_{i=0}^{N}L\left(y,F\left(x,w\right)\right) \quad (5)$$

The large gradient was kept in the sample of little gradients that were chosen using a constant weight, while the GOSS sampling algorithm is applied as LGBM. The order of presentation in raw data remains unchanged, with the GOSS primarily concentrating on an inadequately trained sample. Equation 5 expresses the splitting of instances in variation gain for instance over subsets M and N.

$$\hat{V}_j=\frac{1}{n}\left(L_1+L_2\right) \quad (6)$$

Where,

L_1 and L_2 = Variable for subset M and N.

Furthermore, M and N's representation of the sample with a large gradient has determined the size of selected arbitrarily. The negative gradient loss function with respect to the GBDT method performance has utilized for all iteration boosting of gradient. The premise of LGBM classifier has employed not only GOSS method for optimizing the train sample but also included EFB to extract features in order to speed up network training. Due to highly dimensional data for mutually contrary features, whereas the sparse features have attached together through feature based EFB and the data is significantly sparse. This is used for creating recent features and reconstructed based on new features of histogram equations and coding are shown in Figure 3 & 4.

The hyperparameter search space setup for each of the parameters you mentioned in detail, based on what they control in LightGBM classifier. The boosting_type has specified the boosting algorithm using gradient boosting

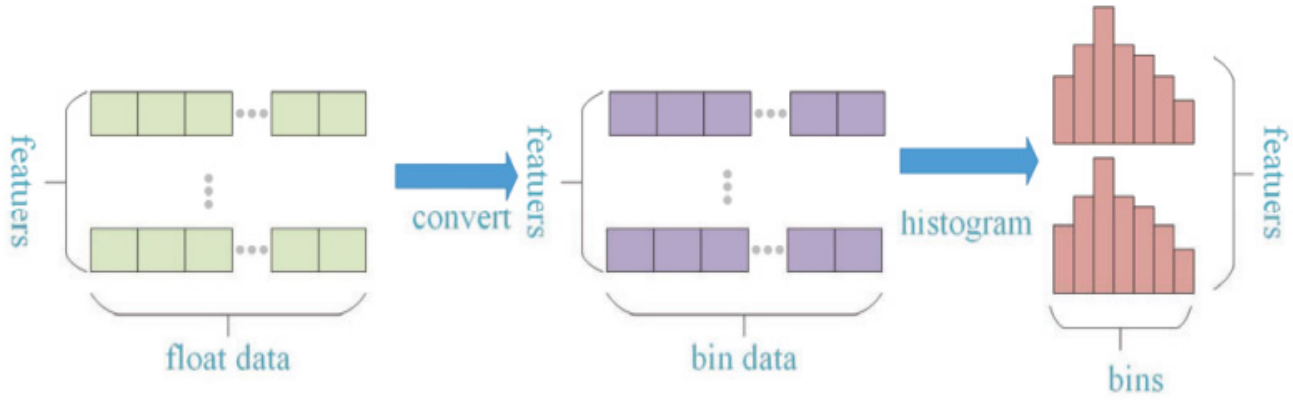

Figure 3: LGBM with histogram based reconstructed features

decision tree to sequential boosting that minimize the bias with high accuracy. The purpose of learning rate parameter assist to shrinks the contribution of each tree whereas the lower the learning rate may require more trees to improve the generalization. Similarly, the n_estimators define the number of boosting iterations in which the larger value generate better accuracy wile paring with low learning rate but the training time consumption is comparatively more. The parameter num_leaves illustrates the maximum number of leaves per tree wherein the large num_leaves to determine the more complex model. The parameter max_depth has illustrates the maximum depth of each tree that assist to avoid overfitting by selecting the shallow depth. The parameter reg_lambda illustrate L2 penalty on leaf weights in which larger values with stronger regulations minimize overfittings.

LGBM employs the leaf-wise method of growth. In order to prevent fruitless node splitting and conserve computer resources, it can be understood as choosing the most advantageous leaf nodes for growth at each divisive node. In addition, the tree's growth is constrained by the maximum depth, which helps to manage the network's complexity and prevent over-fitting. The generalization capacity of the LGBM model is also ensured by increasing the network's training speed. Therefore, the CBBoW with LGBM has generated high accuracy through better training of data modeling in tweet dataset and it can be evaluated through confusion matrix metrics measure and compared with various ML classifiers.

Hard voting classifier for the proposed ensemble:

A meta-classifier is used for combining ML models that are conceptually similar or conceptually dissimilar for prediction using vote majority. This voting classifier technique consists of two types are

Hard voting technique

Soft voting technique

The prediction of classes occur more frequent between the base that receives major votes is said to be hard voting and this type of category prediction is considered to be final prediction is shown in Figure 5. One of the best performances in ensemble model can be determined through voting classifier. Measures such as accuracy and

```
from skopt.callbacks import DeadlineStopper, DeltaYStopper
from skopt.space import Real, Categorical, Integer
# Setting the search space
search_spaces = {
    'boosting_type':(['gbdt', 'rf']),
    'importance_type':(['split', 'gain']),
    'reg_sqrt': Categorical([True, False]),
    'learning_rate': Real(0.001, 1.0, 'log-uniform'),     # Boosting learning rate
    'n_estimators': Integer(30, 5000),                    # Number of boosted trees to fit
    'num_leaves': Integer(2, 512),                        # Maximum tree leaves for base learners
    'max_depth': Integer(-1, 256),                        # Maximum tree depth for base learners, <=0 means no limit
    'subsample': Real(0.01, 1.0, 'uniform'),              # Subsample ratio of the training instance
    'subsample_freq': Integer(1, 10),                     # Frequency of subsample, <=0 means no enable
    'colsample_bytree': Real(0.01, 1.0, 'uniform'),       # Subsample ratio of columns when constructing each tree
    'reg_lambda': Real(1e-9, 100.0, 'log-uniform'),       # L2 regularization
    'reg_alpha': Real(1e-9, 100.0, 'log-uniform'),        # L1 regularization
  }
```

Figure 4: Hyperparameter involved for LGBM classifier

confusion matrix measurements are used to assess the hard Voting Classifier's performance on a different testing dataset. These metrics demonstrate how accurately the ensemble forecasts the eventual identification of cyberbullying related text present in the SM.

Moreover, the proposed hard Voting Classifier plays a robust ensemble technique that predicts the cyberbullying context present in the tweets. Hence, the ensemble method has improves accuracy and permanence by aggregating forecasts by majority voting, building it a helpful tool for exact prediction of cyberbullying content as well as decision-making to remove the content and identify the individual to block their account in the SM.

Algorithm:

Step 1: Dataset has to be loaded, preprocess the data, and divide it into test and training sets.
(Testing phase: 30%; training phase: 70%)
Step 2: Define the base classifiers (in this case, DT, LR, and SVM or models).
Step 3: Create the Voting Classifier using the list of tuples and specify voting='hard'.
voting_classifier = VotingClassifier(estimators=classifiers, voting='hard')
Step 4: Training the Voting Classifier using train data.
Step 5: Use the trained Voting Classifier in predicting the test set.
Step 6: Calculating the ensemble classifier accuracy and print accuracy = accuracy_score(y_test, predictions)
Step 7: Model evaluation using accuracy, F1, recall, sensitivity and Specificity

This research promptly focuses on minimizing the bias in the NLP method and improves the accuracy in detecting the cyberbullying text from the SM. Moreover, it assists the management of the respective SM platform to identify the individuals who made this kind of illegal work.

Result and Discussion

According to the experimental studies, redeemed literary texts with 38,706 features have been divided into six categories of classification as literary texts using experimental data including 47,692 samples. The training and testing datasets, which contain 33,384 and 14,308 texts, respectively have been included.

A standard PC with a 1 TB hard drives, a 16 GB of RAM, 2.60 GHz CPU, and loaded with 64-bit Windows 10 Enterprise Edition is considered in this evaluation model metrics using python library. While developing, optimizing, and putting models through training are crucial phases in the analytics lifecycle, it is more crucial to comprehend the model's performance. The proposed CBBoW is compared with traditional BoW and skip-gram as Word2vec. These NLP methods have been compared with testing tweet datasets with 1000 tweets as a sample. The CBBoW sentiment status and its confidence score is shown in table 1.

Table 1 has determined the confidence score for the sentiment status predicted by the NLP method as CBBoW method.

This tweet sentence is individually checked for all the tweets and listed with sentiment status and its confident score for the offensive (cyberbulling) or non offensive (Non cyberbulling) status. Moreover, this study focuses on determining the classes of the cyberbulling status from the tweet are determined through various ML methods. Hence, the different ML classifier is considered for the evaluation of CBBoW method of text classification.

The performance of classification method is generally endured by the model prediction results using LGBM classifier, RF Classifier (RFC), Extra Tree Classifier (ETC), Extreme Gradient Boosting (XGB) classifier, DT Classifier (DTC), Bagging Classifier (BC), and Bernoulli Navies Bayes (BNB). Hence, the CBBoW technique with the given ML model through evaluation metrics is considered to be finest when comparing with given ML models. The detection of cyberbullying prediction in SM has been evaluated by metrics of confusion matrix like accuracy. Figure 6 illustrates the heatmap of confusion matrix for CBBoW with LGBM classifier in which the True positive has increased due to modification of diagonal values for all classes from 0 to 5.

The accuracy of various ML classifiers with CBBoW along with ensemble model as voting classifier with CBBoW is shown in Table 2. The accuracy is high in voting classifier with CBBoW is 83.73% tailed by 83.26%, 83.54%, and 82.07% from the CBBoW with LGBM classifier, CBBoW with XGB classifier and CBBoW with ETC respectively. The CBBoW with DTC has very less accuracy has determined that very poor prediction in detection of cyberbullying in SM. In contrast, ensemble using voting classifier with CBBoW has high accuracy that

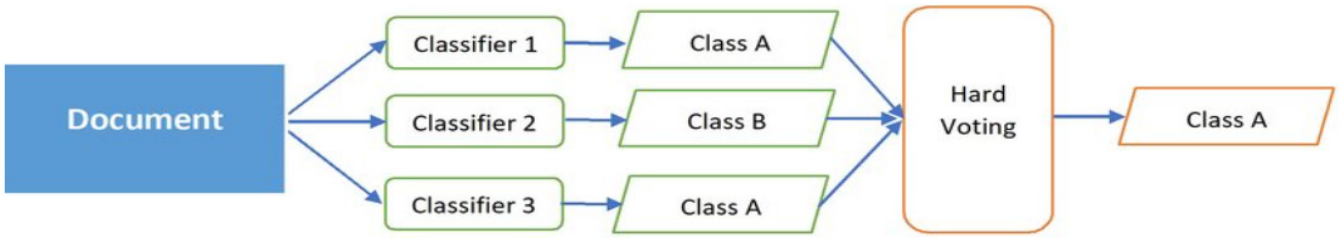

Figure 5: Model ensemble hard voting classifier

Table 1: Confidence score and sentiment status of cyberbullying type for the tweet

tweet_text	*Cyber bullying_ type*	*clean_text*	*Sentiment Status*	*Confidence Score*
FYI the phrase is «out your RABID mind» dumb fucks smh niggers these days *walks off*	ethnicity	fyi phrase «out rabid mind» dumb fucks smh niggers days *walks off*	NEGATIVE	0.9994
U are Still scared to call out Jihad Still scared to call out Pakistan Still scared to call out Islamic terrorism Only the platitudes - violence.. Innocents & bla bla . U r brushing the role of abv by calling it a Violence as if its some street brawl & both sides r indulging	religion	u still scared call jihad still scared call pakistan still scared call islamic terrorism platitudes - violence innocents &amp bla bla u r brushing role abv calling violence street brawl &amp sides r indulging	NEGATIVE	0.9916
You can read Gibbon on Rome, Thucydides on Athens & Sparta, Ibn Khaldun on the dynastic Muslim empires, or the Brook series on Imperial China - many explanations of rise, fall and change but one factorhaunts them all - at the time of collapse there's some'effing idiot in charge!	religion	read gibbon rome thucydides athens &amp sparta ibn khaldun dynastic muslim empires brook series imperial china - many explanations rise fall change one factorhaunts - time collapse effing idiot charge	NEGATIVE	0.8459
I was never bullied. But in gr11 the popular guy I had a crush on from day 1 of high school asked me to be his girlfriend. He took me to the volleyball wrap up party. One of the popular girls whipped a scotch glass at my head bc I wasnâ€™t pretty/ popular/good enough to be w/ him	age	never bullied gr popular guy crush day high school asked girlfriend took volleyball wrap party one popular girls whipped scotch glass head bc wasnt pretty/popular/ good enough w/	POSITIVE	0.8215
My dream is for one of the girls who bullied me in jr high to follow me because they donâ€™t know itâ€™s me and then Iâ€™ll tweet â€œone of you is a girl who bullied me in jr high and donâ€™t know itâ€™s me. Iâ€™m funny now huh?â€ And then all the mean girls will sweat thinking itâ€™s them.	age	dream one girls bullied jr high follow dont know ill tweet one girl bullied jr high dont know im funny huh mean girls sweat thinking	POSITIVE	0.7452

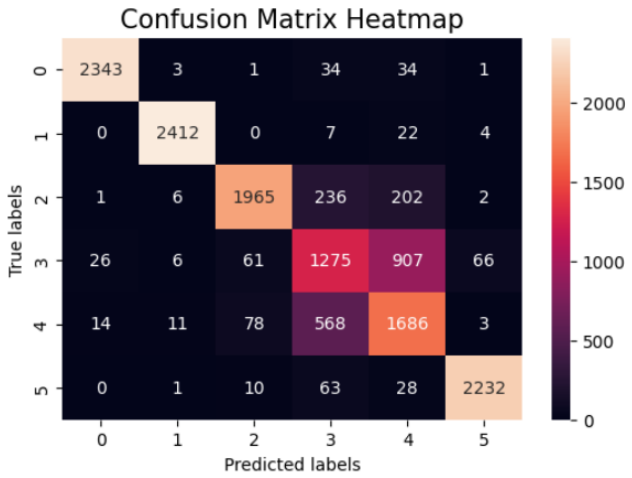

Figure 6: Heatmap of confusion matrix for CBBoW with LGBM Classifier

assists high prediction in detection of cyberbullying in SM.

Moreover, the comparative traditional NLP methods such as BoW and skip-gram as Word2Vec with various ML classification methods are resulted with model prediction such as LGBM classifier, XGB classifier, ETC, RFC, BC, DTC, and BNB. Hence, the evaluation metrics of BoW and word2vec with LGBM method are considered with high accuracy when

Table 2: Accuracy and balanced accuracy for CBBoW with various ML classifiers

CBBoW with ML classifier	*Balanced Accuracy (%)*	*Accuracy (%)*
LGBM Classifier	83.11	83.26
ETC	81.88	82.07
XGB Classifier	83.38	83.54
RFC	81.85	82.02
DTC	79.14	79.34
BC	80.49	80.68
BNB	81.41	81.52
Ensemble with voting classifier	83.57	83.73

comparing with other classifier models. The detection of cyberbullying prediction in SM is evaluated by metrics of confusion matrix as accuracy for BoW and word2vec with LGBM classifier is shown in table 3.

Table 4 and Figure 7 illustrates the ensemble voting classifier with CBBoW has better cyberbullying detection with 83.62% which is comparatively higher than CBBoW with LGBM classifier and traditional Bow with LGBM classifier and Word2Vec with LGBM classifier are 83.26%, 79.70%, and 79.81% respectively. The result of CBBoW with voting classifier performance overcomes the detection

Table 3: Accuracy and balanced accuracy of various classification methods with BoW and Word2vec

Classification Methods	*Bow*		*Skip-gram as Word2Vec*	
	Accuracy (%)	*Balanced Accuracy (%)*	*Accuracy (%)*	*Balanced Accuracy (%)*
LGBM Classifier	79.70	79.53	79.81	79.63
XGB Classifier	77.23	77.02	77.50	77.29
ETC	77.22	77.12	77.49	77.17
RFC	77.56	77.37	78.04	77.85
DTC	70.60	70.38	69.26	69.05
BC	75.54	75.26	75.70	75.49
BNB	69.61	69.47	67.54	67.45
Ensemble with voting classifier	79.86	79.47	79.92	79.86

Table 4: Performance accuracy of cyberbullying detection classifier

Cyberbullying detection classifier	*Accuracy (%)*
CBBoW with LGBM classifier	83.62
CBBoW with ensemble votting classifier	83.26
BoW with LGBM classifier	79.70
Word2 Vec with LGBM classifier	79.81

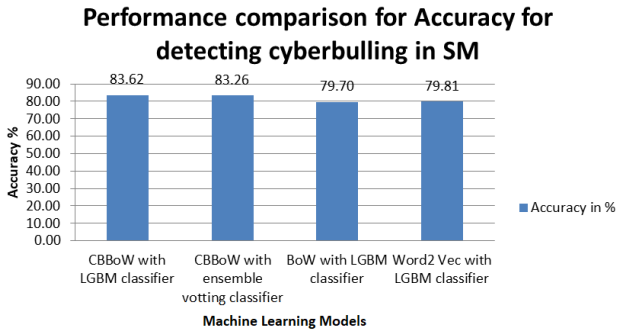

Figure 7: Performance accuracy of cyberbullying detection classifier

cyberbulling accuracy than BoW with LGBM classifier and Word2Vec in social media.

Conclusion

This paper proposes the CBBoW technique to the feature words issues, which arises in various disciplines with different meanings and frequencies across numerous disciplines, to be used for classifying literary texts. For the purpose of distinguishing qualities, the present discipline classification provides guidance. The goal of word meanings across domains is to boost confidence in the BoW algorithm's results. In comparison, the experiments on detecting cyberbullying by selecting features using text classification by ensemble through voting classifier as proposed hybrid model. The hybrid model improves text classification performance through accuracy as 83.62% which is higher compared to BoW with LGBM classifier and Word2Vec with LGBM classifier in SM which has the capability in generalization and reliability for the algorithm. In future work, the text classification is focused on the hybrid DL method in improving cyberbullying detection of Twitter tweets.

Acknowledgement

We would like to thank IIS Deemed to be University, Gurukul Marg, Jaipur, for the helpful support of conducting research in an effective manner.

References

Ahmed, S., & Rajput, A. E. (2023). Denial, acceptance and intervention in society regarding female workplace bullying: A mental health study on social media. *The Scientific Temper, 14*(4), 1544–1556. https://doi.org/10.58414/SCIENTIFICTEMPER.2023.14.4.70

Al-Ajlan, M. A., & Ykhlef, M. (2018). Deep learning algorithm for cyberbullying detection. *International Journal of Advanced Computer Science and Applications, 9*(9).

Al-Garadi, M. A., Varathan, K. D., & Ravana, S. D. (2016). Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network. *Computers in Human Behavior, 63*, 433–443.

Balakrishnan, V., & Kaity, M. (2023). Cyberbullying detection and machine learning: A systematic literature review. *Artificial Intelligence Review, 56*(1), 1375–1416.

Balakrishnan, V., Khan, S., & Arabnia, H. R. (2020). Improving cyberbullying detection using Twitter users' psychological features and machine learning. *Computers & Security, 90*, 101710.

Chia, Z. L., Ptaszynski, M., Masui, F., Leliwa, G., & Wroczynski, M. (2021). Machine learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection. *Information Processing & Management, 58*, 102600. https://doi.org/10.1016/j.ipm.2021.102600

Deng, X., Li, Y., Weng, J., & Zhang, J. (2019). Feature selection for text classification: A review. *Multimedia Tools and Applications, 78*(3), 3797–3816.

Kumar, A., & Sachdeva, N. (2022). A Bi-GRU with attention and CapsNet hybrid model for cyberbullying detection on social media. *World Wide Web, 25*, 1537–1550. https://doi.

org/10.1007/s11280-021-00920-4

Kumar, Y. J. N., Vanapatla, R. R., & Pinamon, V. K. (2024). Detecting cyberbullying in social media using text analysis and ensemble techniques. *E3S Web of Conferences, ICFTEST-2024*, 01069. https://doi.org/10.1051/e3sconf/202450701069

Kumari, K., & Singh, J. P. (2021). Identification of cyberbullying on multimodal social media posts using genetic algorithm. *Transactions on Emerging Telecommunications Technologies, 32*(1), e3907. https://doi.org/10.1002/ett.3907

Obaid, M. H., Guirguis, S. K., & Elkaffas, S. M. (2023). Cyberbullying detection and severity determination model. *IEEE Access*.

Raj, C., Agarwal, A., Bharathy, G., Narayan, B., & Prasad, M. (2021). Cyberbullying detection: Hybrid models based on machine learning and natural language processing techniques. *Electronics, 10*(22), 2810. https://doi.org/10.3390/electronics10222810

Ramalingam, G., Logeswari, S., Kumar, M. D., Prabakaran, M., Nishant, N., & Ahmed, S. A. (2024). Machine learning classifiers to predict the quality of semantic web queries. *The Scientific Temper, 15*(1), 1777–1783. https://doi.org/10.58414/SCIENTIFICTEMPER.2024.15.1.28

Rosa, H., *et al*. (2019). Automatic cyberbullying detection: A systematic review. *Computers in Human Behavior, 93*, 333–345. https://doi.org/10.1016/j.chb.2018.12.021

Roy, P. K., Tripathy, A. K., Das, T. K., & Gao, X.-Z. (2020). A framework for hate speech detection using deep convolutional neural network. *IEEE Access, 8*, 204951–204962. https://doi.org/10.1109/ACCESS.2020.3037073

Soumya, S., & Pramod, K. V. (2020). Sentiment analysis of Malayalam tweets using machine learning techniques. *ICT Express, 6*(4), 300–305. https://doi.org/10.1016/j.icte.2020.04.003

Valdivia, A., Luzón, M. V., & Cambria, E. (2018). Consensus vote models for detecting and filtering neutrality in sentiment analysis. *Information Fusion*. https://doi.org/10.1016/j.inffus.2018.03.007

Yan, D., Li, K., Gu, S., & Yang, L. (2020). Network-based bag-of-words model for text classification. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.2991074

Yuvaraj, N., Chang, V., Gobinathan, B., Pinagapani, A., Kannan, S., Dhiman, G., *et al*. (2021). Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification. *Computers & Electrical Engineering, 92*, 107186. https://doi.org/10.1016/j.compeleceng.2021.107186

Zhou, X., Gururajan, R., Li, Y., Venkataraman, R., Tao, X., Bargshady, G., Barua, P. D., & Kondalsamy-Chennakesavan, S. (2020). A survey on text classification and its applications. *Web Intelligence, 18*(3), 205–216.