*I.J. Intelligent Systems and Applications,* **2023, 2, 1-13**
Published Online on April 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2023.02.01

# Machine Learning in Cyberbullying Detection from Social-Media Image or Screenshot with Optical Character Recognition

**Tofayet Sultan***
Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
E-mail: tofayetsultanneon@gmail.com
ORCID iD: https://orcid.org/0000-0002-6259-0400
*Corresponding author

**Nusrat Jahan**
Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
E-mail: nusrat.jahan.jui23@gmail.com
ORCID iD: https://orcid.org/0000-0002-4958-2676

**Ritu Basak**
Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
E-mail: ritubsk@gmail.com
ORCID iD: https://orcid.org/0000-0003-0446-8607

**Mohammed Shaheen Alam Jony**
Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
E-mail: shaheenalamjony9@gmail.com
ORCID iD: https://orcid.org/0000-0003-3010-9130

**Rashidul Hasan Nabil**
Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
E-mail: rashidul@aiub.edu
ORCID iD: https://orcid.org/0000-0002-8414-6423

Received: 25 September 2022; Revised: 03 December 2022; Accepted: 04 January 2023; Published: 08 April 2023

**Abstract:** Along with the growth of the Internet, social media usage has drastically expanded. As people share their opinions and ideas more frequently on the Internet and through various social media platforms, there has been a notable rise in the number of consumer phrases that contain sentiment data. According to reports, cyberbullying frequently leads to severe emotional and physical suffering, especially in women and young children. In certain instances, it has even been reported that sufferers attempt suicide. The bully may occasionally attempt to destroy any proof they believe to be on their side. Even if the victim gets the evidence, it will still be a long time before they get justice at that point. This work used OCR, NLP, and machine learning to detect cyberbullying in photos in order to design and execute a practical method to recognize cyberbullying from images. Eight classifier techniques are used to compare the accuracy of these algorithms against the BoW Model and the TF-IDF, two key features. These classifiers are used to understand and recognize bullying behaviors. Based on testing the suggested method on the cyberbullying dataset, it was shown that linear SVC after OCR and logistic regression perform better and achieve the best accuracy of 96 percent. This study aid in providing a good outline that shapes the methods for detecting online bullying from a screenshot with design and implementation details.

**Index Terms:** Cyberbullying Detection, Data Mining, Machine Learning, NLP, OCR.

## 1. Introduction

Cyberbullying and rumors, particularly, have become a significant issue as the popularity of the Internet and social media grew. Cyberbullying is defined as using ICT infrastructure by anybody to shame or humiliate another individual or group of persons which may take several forms. [1]. It is now a severe national health issue, with sufferers having a much higher risk of suicidal thoughts. It was only a matter of time after the Internet's growth that bullies began to utilize this new and popular medium. Cyberbullying makes the sufferer feel as though he is being attacked from all sides because the internet world is merely a click away.

Even if the Internet is secure for users, its flexibility may lead to problems like cyberbullying, which has lately been identified as a public health concern [2]. As a result, it is crucial to explore the problem roots of cyberbullying. A system might distinguish between bullying and non-bullying messages and take appropriate action [3]. Because of the detrimental impact cyberbullying has on victims, detection and prevention are crucial. Many cyberbullying detection technologies have been developed to aid in the management and reduction of cyberbullying. Researchers working on cyberbullying detection have progressed, but many difficulties remain unresolved. Detecting cyberbullying in a photograph or screenshot is one of them. The phrase "Sentiment Analysis" (SA) refers to a technique for determining one's views in written language. People's frequent expressing of thoughts on social, economic, health, and product and brand concerns and the fact that social media is such an essential tool for people has paved the way for sentiment analysis [4]. Data mining and text mining are currently popular [5]. The most popular technique to categorize literature using SA is to seek the writer's point of view on a particular topic [6].

With the use of this study, we will be able to identify occurrences of cyberbullying from a snap across social media platforms. Our goal is to examine every potential sub-domain that is connected to the detection of cyberbullying so that we may develop a method that has a greater level of acceptance than any other now available. In order to do this, the first obstacle we face is extracting text from images, which requires us to find a system that has the least amount of complexity possible. The obtained text will be processed. Then, we will investigate every relevant sub-domain associated with the detection of cyberbullying. When the time finally comes, we will conduct an analysis and make any necessary adjustments to the procedure. Previous research in this area has had limited success, so our hopes are high for this one.

Seven sections make up the paper, which are arranged in the following order: Section 1 comprises the introduction; and Section 2 includes an introduction to several pertinent studies; In Section 3 the experimental and implementation setup is described with the dataset; We present the OCR tool processing to extract text from image and ML approach sentiment analysis algorithm in Section 4; after that the results are compared in Section 5; then the results are analyzed & evaluated in section 6. Finally, we bring the paper's observations in section 7.

## 2. Related Works

### *2.1. Background Study*

Cyberbullying has been studied from several contexts and viewpoints; this is why various definitions have been proposed. Because of these different viewpoints, we need to prevent this problem from different contexts as much as possible. Teenage to older, everyone is the victim of this harassment as it is a common problem on social media and other platforms. For these reasons, the detection of cyberbullying using machine learning or other algorithms has been done by many researchers and has been happening for years [7]. But eradicating online bullying is an ongoing activity, methods need to be routinely updated from different views to take into account the most recent developments [8].

### *2.2. Image to Text*

Deep learning has a variety of useful applications, one of which is optical character recognition. This work [9] describes the technique for partitioning text from character pictures, which may include visuals and computer-typed or handwritten words. The suggested character identification and extraction approach achieve encouraging results, demonstrating its resilience. A text extraction pipeline is described in this study [10] to address text extraction from varying quality photos obtained from social media. They collected datasets from 4 categories of images from social media. Several preprocessing techniques were included with Tesseract OCR to improve the accuracy of OCR. Using textual, visual, and infographic social data modalities, this paper [11] provides a deep neural network for cyberbullying detection. There infographic material separated from the image using Google Lens of the Google Photos App. There are many types of OCR: Keras-OCR, EasyOCR and Tesseract. In contrast to other libraries, the easyOCR library is extremely straightforward and lightweight to use [12]. It provides support for a variety of languages. It is also possible to improve its performance for particular use cases by adjusting the values of various hyper – parameters. When used to well-organized texts such as pdf files, receipts, and bills, it produces more accurate results though.

### *2.3. Text Processing & Classification*

According to a publication, they are extended a list of post-offensive phrases and assigned various strengths to create abuse components, which are then combined with Bag-of-Words and latent word meanings to produce the final form before passing them into a linear SVM Classification algorithm [13]. They proposed Encoding -enhanced Bag-of-Words, a novel representation learning approach for hate speech detection that is both simple and effective. In all evaluation metrics, their proposed method surpasses other compared methods. The suggested approach in this article [14] is used to identify harmful online interactions, such as abusive content spread through text and graphics. The proposed strategy uses Convolutional Neural Network and Bag of Words algorithms along with the current method to identify cyberbully images and text on the Instagram dataset. Islam et al. [15] reported that another model named TF-IDF outperforms bag of words feature when they analyzed it for four machine learning algorithms.

In a recent study, a suggestion for recognizing cyberbullying is made based on the fact that text message usage, settings, and language have changed over time. Sentiment analysis, in addition to conventional feature extraction techniques like TF-IDF, N-gram, and profanity, increases the system's accuracy. Information was given by over 20,000 students. The suggested method fails to identify the ironic component of cyberbullying. The suggested response has an F1 score of 74%, accuracy of 74.50%, precision of 74% and recall of 74% [16]. Alam et al. use two alternative feature extraction algorithms in conjunction with numerous n-gram analyses to demonstrate four machine learning classifiers and three ensemble models using Twitter data. Their suggested SLE and DLE models achieve 96% effectiveness when TFIDF feature extraction is combined with K-Fold cross-validation [17]. With an average accuracy of about 90.57%, the test findings in this study indicate that LR is better. SGD had the greatest precision (0.968), SVM had the highest recall (0.928), and logistic regression had the highest F1 score among the classifiers (1.00). The tests shown that LR outperforms other classifiers in terms of prediction time and improves with increasing data volume. As a consequence, SGD performs almost as well as LR, while the error is not as tiny [18]. This study's recommended approach achieved 90.3% accuracy using SVM with 4 grams and 92.8% accuracy using Neural Network with 3 grams while applying both TFIDF and sentiment classification. Their Neural Network greatly outperformed the SVM classifier, with a mean f-score of 91.9% compared to the SVM's 89.8% [19].

The authors of this article [20] aims to compare the predicting models for fundamental machine learning in cyberbullying detection and the suggested systems with participation techniques for feature selection, resampling, and utilizing two classifiers: Support Vector Machine (SVM) and Decision Tree for optimization. N-gram characteristics from the ASKfm corpus were used in word extraction and then applied to eight different experiment setups. The best performance is provided by Decision Tree, according to an analysis of performance metrics. Kumar et al. [21] used Sequential Machine Optimization, Random Forest, K-Nearest Neighbor and Naive Bayes. The data was provided from YouTube. They obtained data from 7962 comments on 60 YouTube videos, an average of 116 per video. The researchers discovered that when applied to Clips on YouTube for online bullying identification, K-Nearest Neighbor had the greatest accuracy of 83 percent, outperforming all other approaches. It has been shown in a published approach that Neural Networks outperform SVMs and achieve an efficiency of 92.8 percent, while SVMs achieve an efficiency of 90.3 percent [22]. A research paper by authors [23] described their works using deep learning for cyberbullying detection where they used several classifications with CNN-CB Architecture, and as a result, SVM gave 81% accuracy. Despite the methods and features to detect cyberbullying, researchers also used some other classifiers in different way.

From previous study's we have seen easyocr which can be used to get the efficient outcome for image to text extraction and for processing extracted text we are going to use two of the mentioned NLP technique which are Bag of words & TF-IDF [12, 15]. Besides, some of the followed effective text classification strategies we got which are Logistic Regression, Decision Tree, Random Forest classifier, SGD classifier and Linear SVC [17,18,23]. There are a variety of flaws in the research being done to identify cyberbullying using various machine learning techniques. Now it can be said that many platforms have tried to solve this issue automatically; then again, many countries have to take action manually, yet if the issue is significant, it takes time. That's why an expert system that has the ability to detect cyberbullying from snap with design and implementation details is proposing by us with other existing methods to give better precision and grouping. We should check each possible combination of every step to get the best combination for cyberbullying detection from snap for this reason we will employ not only the regular classifiers but also some other classifiers which may out performs the usual techniques. More specific about our suggested approach for detecting cyberbullying are found in the methodology section.

## 3. Dataset

We used a Kaggle dataset[1] on online bullying and toxicity that was gathered by the authors Fatma Elsafoury. Several datasets relevant to the automated detection of cyberbullying are collected in this info, which comes from a wide range of sources. A variety of social media sources were used to access the data, including Kaggle, Twitter, Wikipedia Talk pages, and YouTube. There is text in the dataset that has been classified as cyberbullying and text that has not been classified as bullying. In the statistics, there are numerous types of online bullying. From there, we utilized a dataset that had around 160000 instances that were associated with toxicity. There is text in the data that has been classified as bullying and text which has not been classified as bullying. We will implement the system for cyberbullying detection where we require the collected dataset to build a machine learning model for classification in Jupyter Notebook, an open-source software we planned to implement.

## 4. Methodology

We followed the diagram's stages here as research workflow in fig 1. We started by previously collected data from social media networks. After that, we preprocessed the dataset because it included some redundant components that could

[1] Dataset: https://www.kaggle.com/datasets/saurabhshahane/cyberbullying-dataset?select=toxicity_parsed_dataset.csv [24]

cause errors later. Then, we extracted the feature using the Bag of words model and the TF-IDF model. These features are used to train the text mining model. We used optical character recognition (OCR) software to extract text from snapchats and photos to detect cyberbullying. We use an OCR technology that's more accurate than others. After gathering text, it is simplified and made easier to understand before classification. Using a classification system, we'll decide if it's cyberbullying. In this scenario, we apply logistic regression, decision trees, gradient boosting, random forest, bagging, SGD, linear SVC, and adaboost classifiers. We got several results by using many NLP models and algorithms. Then, we compared the accuracy parameter to the others to see which is giving the best results. Here fig 2 illustrates the methodology or implementation details of this study.

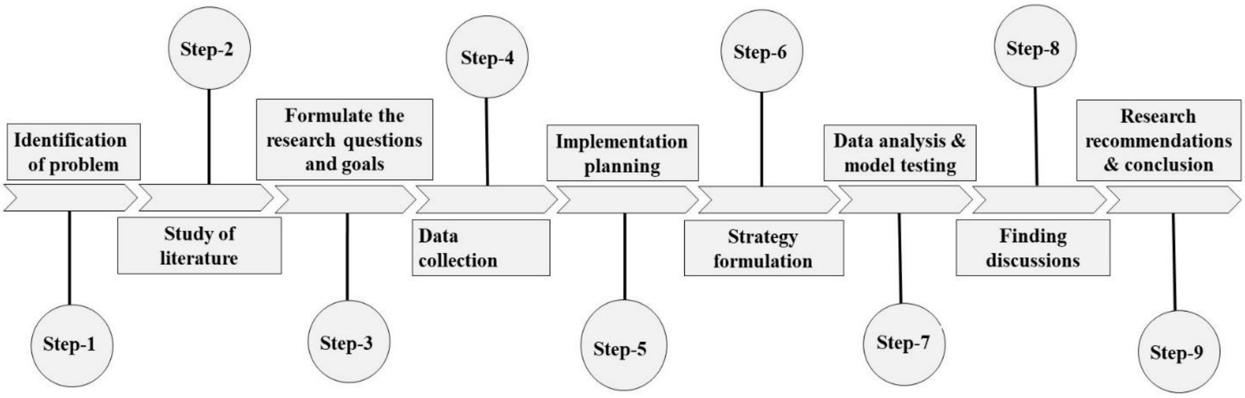


Fig.1. Research workflow.

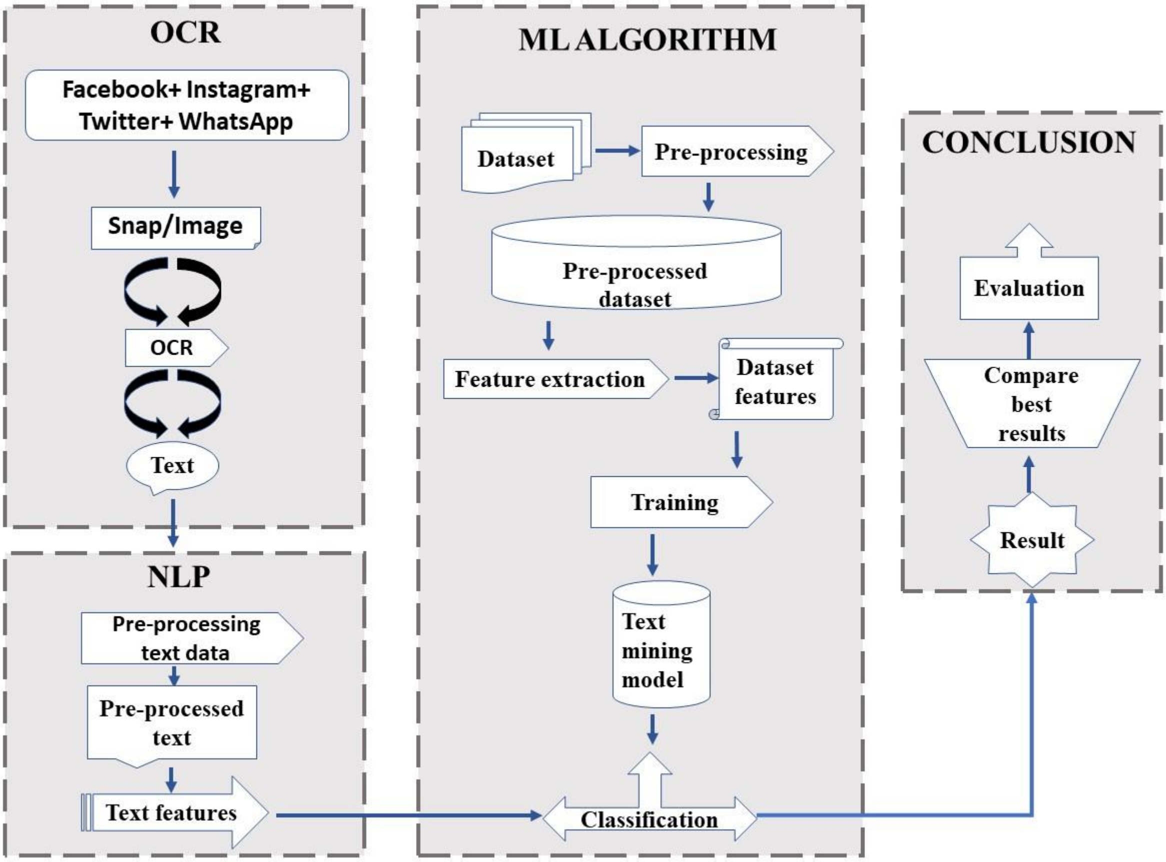


Fig.2. Methodology/ Implementation details.

## 4.1. *Text Extract from Image*

We use machine learning to detect cyberbullying from a snapshot. To categorize the text from the photos, we must extract it. This requires OCR. Optical Character Recognition (OCR) translates picture text to device text. Visuals can be printed sheets, symbols, or handwritten writing.

Literature review revealed easyOCR and tesseract-OCR. Easy OCR is more convenient and effective for extracting text from images. EasyOCR is a Python library for optical character recognition. It's the most user-friendly OCR approach because it supports over 70 languages, including English, Chinese, Japanese, Korean, and Hindi. It also offers functionality for many of those in the works. EasyOCR uses Python and the Pytorch; a graphics processing unit may speed up detection. It requires few lines of code to use and construct. It has acceptable performance for most of the images it has evaluated and may be employed in various dialects.

### *4.2. Natural Language Processing*

OCR analyzes the extracted phrase in milliseconds. Text data in its simplest form can't be machine-processed. We agree they want us to convert words to numbers the gadget can understand. Several symbols or words are unnecessary in real letters and writings. Digits or syntax may make bullying harder to see. We must clean and organize the postings before using data mining techniques on them. This process includes tokenization, stemming, and deleting stop-words and punctuation. The data include both bullying-related and unrelated terms. Two strategies translate text phrases into numeric vectors:

- **Bag-of-words:** The Bag of Words model seems to be the most basic numerical representation. Every single word in bag of words is given the exact same amount of weight. The data should be converted into vectors or integers. Before moving on to the next step, the processing data is stored in a "bag of words."
- **TF-IDF:** The Latent Semantic Document is used. A phrase's frequency in a corpus or collection is measured. So, frequent terms should be given more weight in TF-IDF.

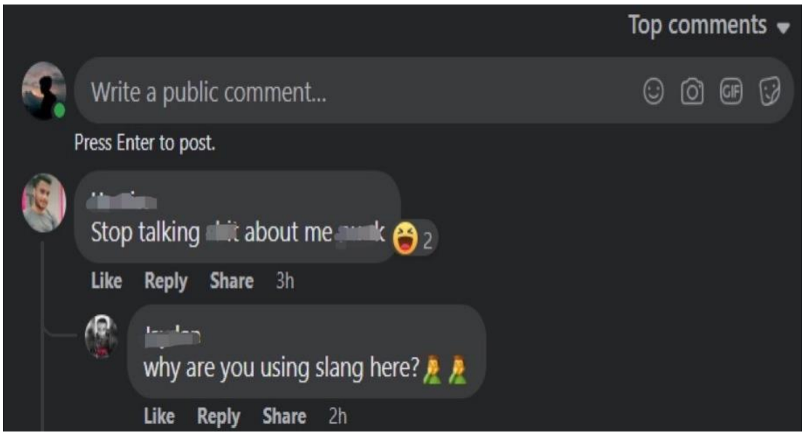
### *4.3. Machine Learning Algorithms*

Machine learning algorithms are increasingly becoming widespread in people's daily lives. Ready-to-use machine learning algorithms for speech recognition, language translation, text classifications, and other tasks are now being offered as cloud-based web services. Different machine learning algorithms fall into two categories. First, supervised learning aims to accurately predict an output variable based on input items. Unsupervised learning is the second-most-important machine learning category. In this type of learning, data are not labeled. Clustering involves finding a useful structure in incoming data and clustering it to solve these problems. We decided to use Logistic regression, Decision tree, Gradient boosting classifier, Random-forest classifier, Bagging classifier, SGD classifier, Linear SVC, and AdaBoost classifier.

### *4.4. Sample Input Output*

When implementing the system, we tried detecting cyberbullying from different images. Here we include some sample input and output where example sentences of bullying and non-bullying are included.

A sample image from social media that includes both bullying and non-bullying content is shown in Fig 3. We'll employ the Bag of Words approach for this.



[redacted]
Stop talking [redacted] about me [redacted] 2

[redacted]
why are you using slang here?

Fig.3. Sample input for the bag of words model from Facebook.

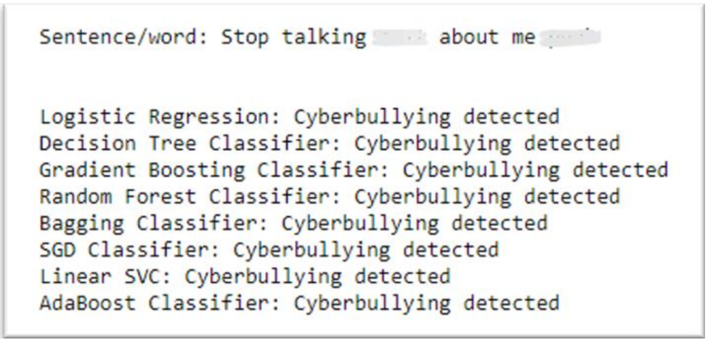
With the support of the Bag of Words processing technique, each classifier was able to identify the bullying-related portions of the text in Fig. 4.

Figure 5 shows a passage of text without any instances of bullying; using the Bag of Words processing technique, each classifier accurately identified this passage.

A sample Instagram photo with both bullying and non-bullying language is shown in Fig 6. We are employing the Bag of Words concept for this.

Fig 7 shows a passage of text without any instances of bullying, and each classifier used the Bag of Words approach to accurately classify it as such.

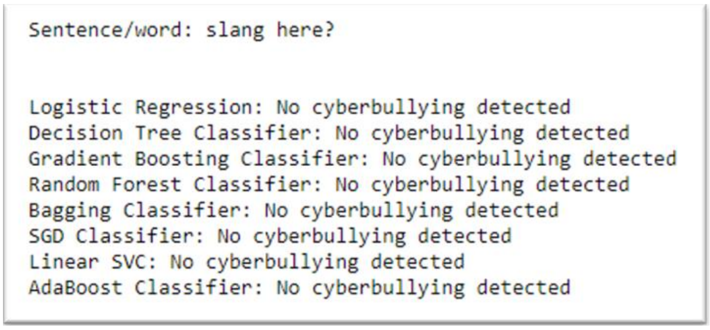
With the assistance of the Bag of Words processing technique, each classifier was able to identify the bullying-related portions of the text in Fig 8. However, the issue is that the profile name is not revised prior to classification. It shouldn't be a criterion to identify bullying because doing so could result in unintended consequences. In addition, the last two words of the entire phrase were classified independently due to a break, which may have swung the balance in favor of bullying or not.



```
Sentence/word: Stop talking [redacted] about me [redacted]


Logistic Regression: Cyberbullying detected
Decision Tree Classifier: Cyberbullying detected
Gradient Boosting Classifier: Cyberbullying detected
Random Forest Classifier: Cyberbullying detected
Bagging Classifier: Cyberbullying detected
SGD Classifier: Cyberbullying detected
Linear SVC: Cyberbullying detected
AdaBoost Classifier: Cyberbullying detected
```

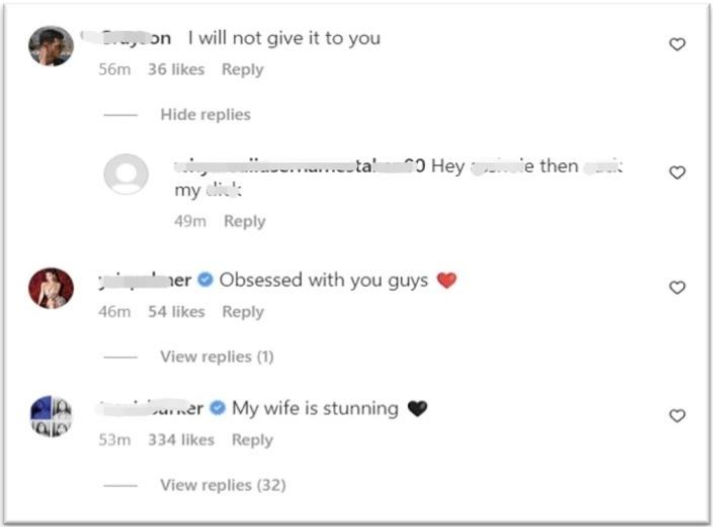
Fig.4. Partial output for each classifier on the bag of words model.



```
Sentence/word: slang here?


Logistic Regression: No cyberbullying detected
Decision Tree Classifier: No cyberbullying detected
Gradient Boosting Classifier: No cyberbullying detected
Random Forest Classifier: No cyberbullying detected
Bagging Classifier: No cyberbullying detected
SGD Classifier: No cyberbullying detected
Linear SVC: No cyberbullying detected
AdaBoost Classifier: No cyberbullying detected
```

Fig.5. Partial output for each classifier on the bag of words model.



[redacted] I will not give it to you
56m 36 likes Reply

Hide replies

[redacted] Hey [redacted] then [redacted] my [redacted]
49m Reply

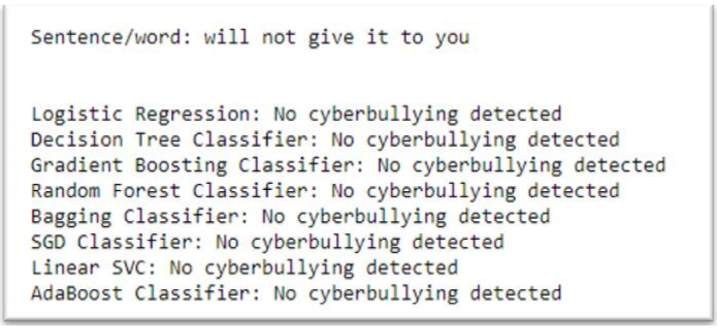
[redacted] Obsessed with you guys ❤
46m 54 likes Reply

View replies (1)

[redacted] My wife is stunning 🖤
53m 334 likes Reply

View replies (32)

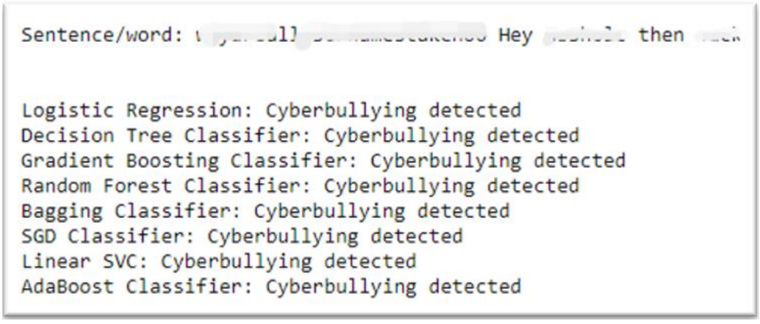
Fig.6. Sample input for the bag of words model from Instagram.



```
Sentence/word: will not give it to you


Logistic Regression: No cyberbullying detected
Decision Tree Classifier: No cyberbullying detected
Gradient Boosting Classifier: No cyberbullying detected
Random Forest Classifier: No cyberbullying detected
Bagging Classifier: No cyberbullying detected
SGD Classifier: No cyberbullying detected
Linear SVC: No cyberbullying detected
AdaBoost Classifier: No cyberbullying detected
```

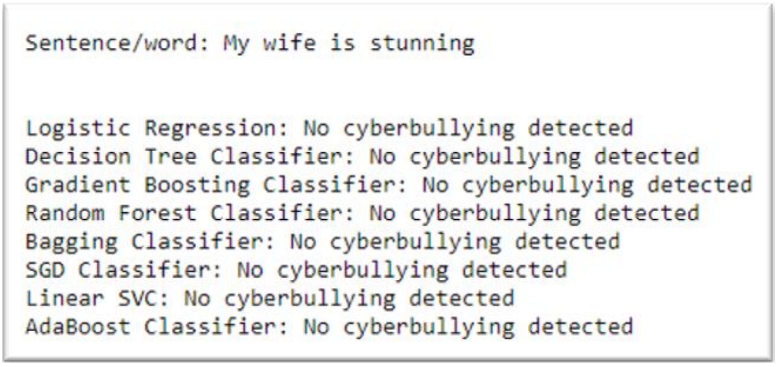
Fig.7. Partial output for each classifier on the bag of words model.



```
Sentence/word: [redacted] Hey [redacted] then [redacted]


Logistic Regression: Cyberbullying detected
Decision Tree Classifier: Cyberbullying detected
Gradient Boosting Classifier: Cyberbullying detected
Random Forest Classifier: Cyberbullying detected
Bagging Classifier: Cyberbullying detected
SGD Classifier: Cyberbullying detected
Linear SVC: Cyberbullying detected
AdaBoost Classifier: Cyberbullying detected
```

Fig.8. Partial output for each classifier on the bag of words model.

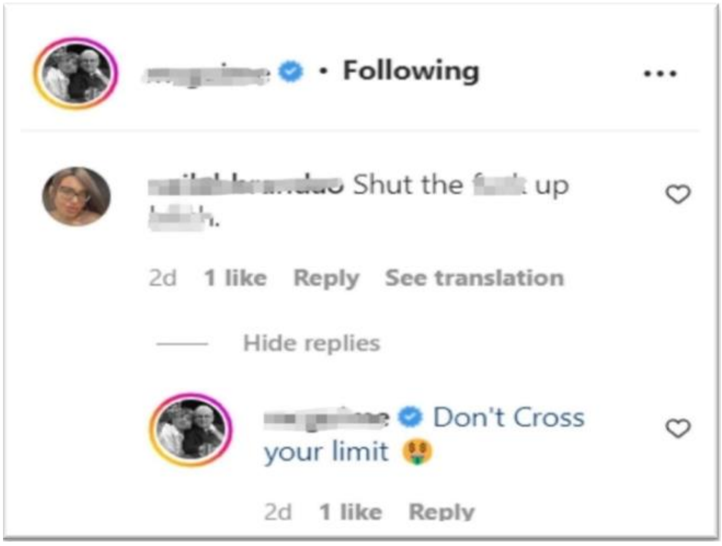


```
Sentence/word: My wife is stunning


Logistic Regression: No cyberbullying detected
Decision Tree Classifier: No cyberbullying detected
Gradient Boosting Classifier: No cyberbullying detected
Random Forest Classifier: No cyberbullying detected
Bagging Classifier: No cyberbullying detected
SGD Classifier: No cyberbullying detected
Linear SVC: No cyberbullying detected
AdaBoost Classifier: No cyberbullying detected
```

Fig.9. Partial output for each classifier on the bag of words model.

Fig 9 shows a passage of text without any instances of bullying, and each classifier used the bag of words processing technique to correctly identify it.



[redacted] • Following

[redacted] Shut the [redacted] up [redacted].

[redacted] Don't Cross your limit 🤑

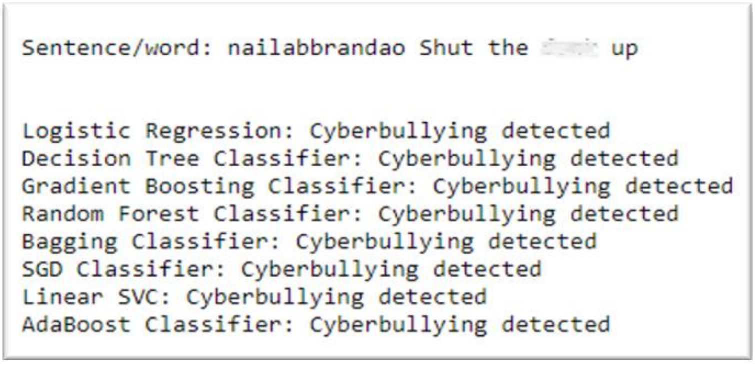
Fig.10. Sample input for the TF-IDF model from Instagram.

A sample social media screenshot with both bullying and non-bullying material is shown in Fig 10. We are using the TF-IDF model for this.

Using the TF-IDF processing technique, each classifier was able to identify the bullying-related text in Fig 11. However, it also uses the profile name to categorize the text, which increases the chance that the detection may be compromised.

Figure 12 shows a passage of text without any bullying-related language, which each classifier accurately marked using the TF-IDF processing technique.

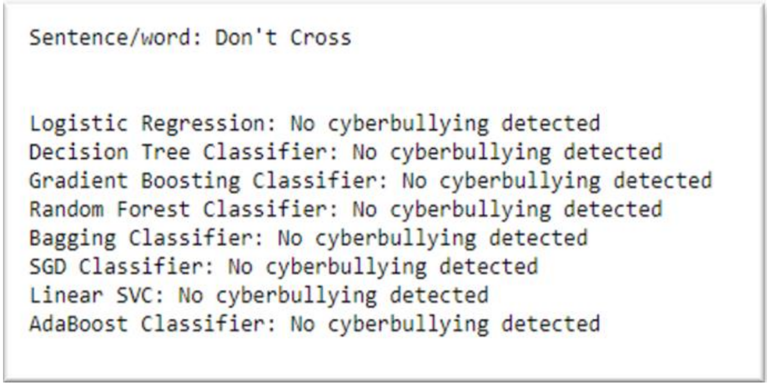
A sample social media screenshot with both bullying and non-bullying material is shown in Figure 13. We'll employ the TF-IDF model for this.



```
Sentence/word: nailabbrandao Shut the [redacted] up


Logistic Regression: Cyberbullying detected
Decision Tree Classifier: Cyberbullying detected
Gradient Boosting Classifier: Cyberbullying detected
Random Forest Classifier: Cyberbullying detected
Bagging Classifier: Cyberbullying detected
SGD Classifier: Cyberbullying detected
Linear SVC: Cyberbullying detected
AdaBoost Classifier: Cyberbullying detected
```

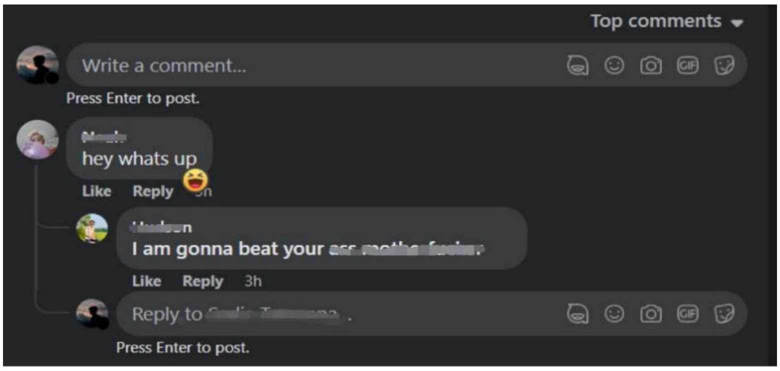
Fig.11. Partial output for each classifier on the TF-IDF model.



```
Sentence/word: Don't Cross


Logistic Regression: No cyberbullying detected
Decision Tree Classifier: No cyberbullying detected
Gradient Boosting Classifier: No cyberbullying detected
Random Forest Classifier: No cyberbullying detected
Bagging Classifier: No cyberbullying detected
SGD Classifier: No cyberbullying detected
Linear SVC: No cyberbullying detected
AdaBoost Classifier: No cyberbullying detected
```

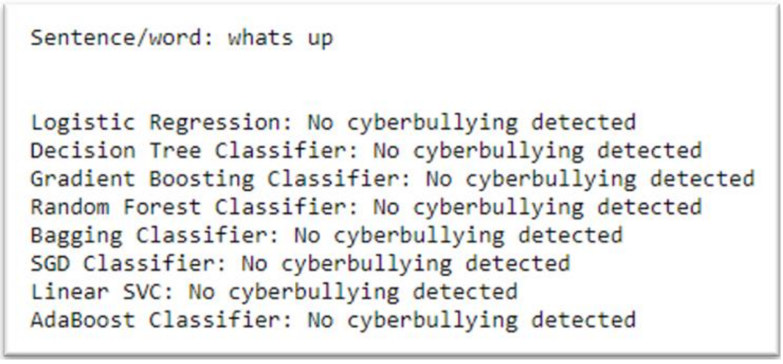
Fig.12. Partial output for each classifier on the TF-IDF model.



[redacted]
hey whats up

[redacted]
I am gonna beat your [redacted]

Fig.13. Sample input for the TF-IDF model from Facebook.

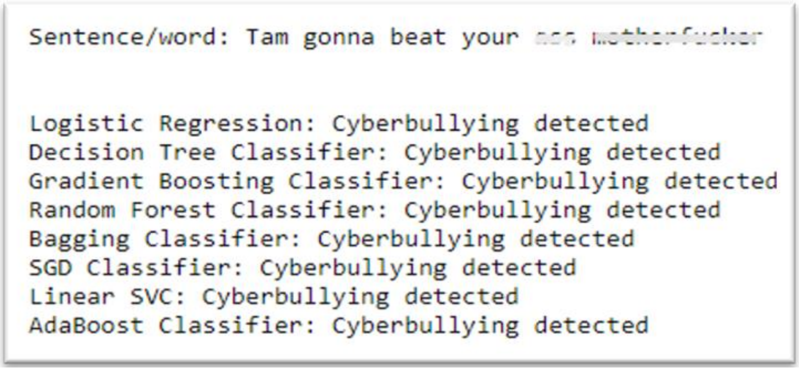


```
Sentence/word: whats up


Logistic Regression: No cyberbullying detected
Decision Tree Classifier: No cyberbullying detected
Gradient Boosting Classifier: No cyberbullying detected
Random Forest Classifier: No cyberbullying detected
Bagging Classifier: No cyberbullying detected
SGD Classifier: No cyberbullying detected
Linear SVC: No cyberbullying detected
AdaBoost Classifier: No cyberbullying detected
```

Fig.14. Partial output for each classifier on the TF-IDF model.

Figure 14 shows a passage of text with no instances of bullying, which each classifier accurately marked using the TF-IDF processing technique.



```
Sentence/word: Tam gonna beat your ass motherfucker


Logistic Regression: Cyberbullying detected
Decision Tree Classifier: Cyberbullying detected
Gradient Boosting Classifier: Cyberbullying detected
Random Forest Classifier: Cyberbullying detected
Bagging Classifier: Cyberbullying detected
SGD Classifier: Cyberbullying detected
Linear SVC: Cyberbullying detected
AdaBoost Classifier: Cyberbullying detected
```

Fig.15. Partial output for each classifier on the TF-IDF model.

With the aid of the TF-IDF processing technique, each classifier was able to identify the bullying component of the text in Fig 15, despite one word being incorrectly retrieved.

## 5. Results

### *5.1. Data Description and Accuracy Measures*

We used Logistic regression, Decision tree, Gradient boosting classifier, Random-forest classifier, bagging classifier, SGD Classifier, Linear SVC, and Ada boost classifier on both Bags of words and TFIDF feature extraction methods. Classification measures include accuracy, recall, precision and f-1 score. The below table describes the accuracy measures.

Table 1. Accuracy measures.

| Measures | Definition | Formula |
|---|---|---|
| Accuracy(A) | A simple ratio of correctly predicted to observed data though the success of our model will be determined by other criteria too. | $A = \frac{TN + TP}{TN + FN + TP + FP}$ |
| Precision(P) | Precision of a forecast is the proportion of accurately classified as positive measurements to the grand total. | $P = TP/(FP + TP)$ |
| Recall(R) | Proportion of positive readings correctly predicted to all positive findings in the actual class. | $R = \frac{TP}{FN + TP}$ |
| F-1 Score(F) | Recall and precision required for the F1 Score, which is weighted average. | $F = 2 * \frac{R * P}{P + R}$ |

### *5.2. Experimental Results*

The table below covers a range of performance values for all classification methods based on various criteria. The performance of each classifier in terms of Accuracy, Precision, F-1 score, and Recall values weighted average is listed concerning NLP approaches the BoW and the TF-IDF.

Table 2. Various measures for each classification algorithm.

| NLP | Bag of words | | | | TF-IDF | | | |
|---|---|---|---|---|---|---|---|---|
| Performance / Classifier | Accuracy | Precision | Recall | F-1 Score | Accuracy | Precision | Recall | F-1 Score |
| Logistic regression | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.95 |
| Decision tree | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| Gradient boosting classifier | 0.94 | 0.94 | 0.94 | 0.93 | 0.94 | 0.94 | 0.94 | 0.93 |
| Random forest classifier | 0.94 | 0.94 | 0.94 | 0.93 | 0.94 | 0.94 | 0.94 | 0.93 |
| Bagging classifier | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 |
| SGD classifier | 0.96 | 0.95 | 0.96 | 0.95 | 0.95 | 0.95 | 0.95 | 0.94 |
| Linear SVC | 0.95 | 0.95 | 0.95 | 0.95 | 0.96 | 0.96 | 0.96 | 0.96 |
| Ada boost classifier | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.95 | 0.95 | 0.94 |

The table 2 represents the classification result. This allows us to compare and assess the categorization method with the highest degree of accuracy. Performances of all classifiers based on various measures are plotted via graphs in below diagrams for both of the NLP model.

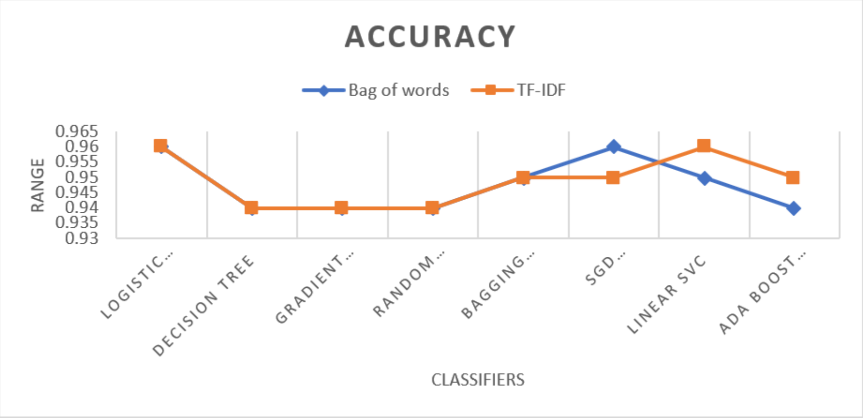


Fig.16. Accuracy for both Model.

According to the fig 16, three classifiers have the best overall accuracy value and that is 96%. Specifically, for the bag of words model, they are Logistic regression with SGD classifier and Logistic regression with Linear SVC for the TF-IDF model, respectively. Where Decision tree, Gradient boosting classifier, Random Forest classifier provides lowest accuracy out of the eight classifiers for both of the NLP model. Ada-boost is another classifier that performs at the lowest level of accuracy for both models. While logistic regression was the most accurate for both models.

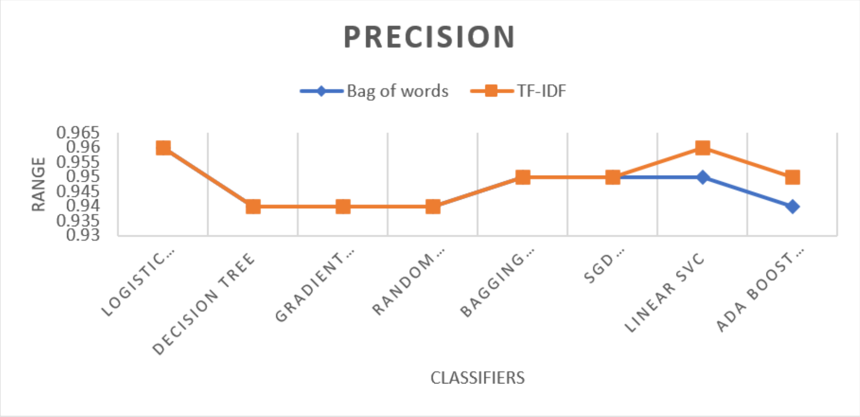


Fig.17. Precision for both Model.

Fig 17 illustrates Logistic regression is giving highest precision for both model which is 96% at the same time Linear SVC also provides equal precision rate on TF-IDF model. On the other hand, Decision tree, Gradient boosting classifier, Random Forest classifier provides and Ada boost classifier gives lowest precision out of them. Rest of the classifiers named Bagging and SGD stands same and close to highest precision.

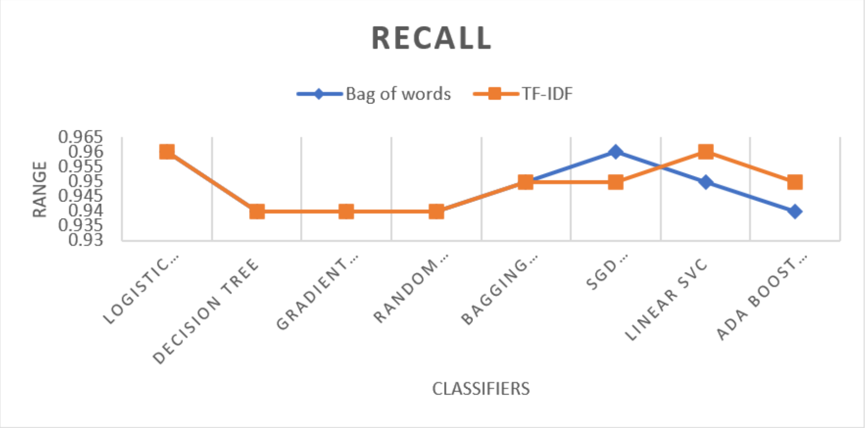


Fig.18. Recall for both Model.

Fig 18 demonstrates that for both of the model the recall value is high and same for Logistic Regression that is 96%. For TF-IDF model another highest recall rate comes from Linear SVC and SGD classifier gives another highest for Bag of words model. Like accuracy and precision, the recall value for Decision tree, Gradient boosting classifier and Random Forest classifier is the lowest. Another lowest recall rate given by Ada Boost classifier on Bag of words model.

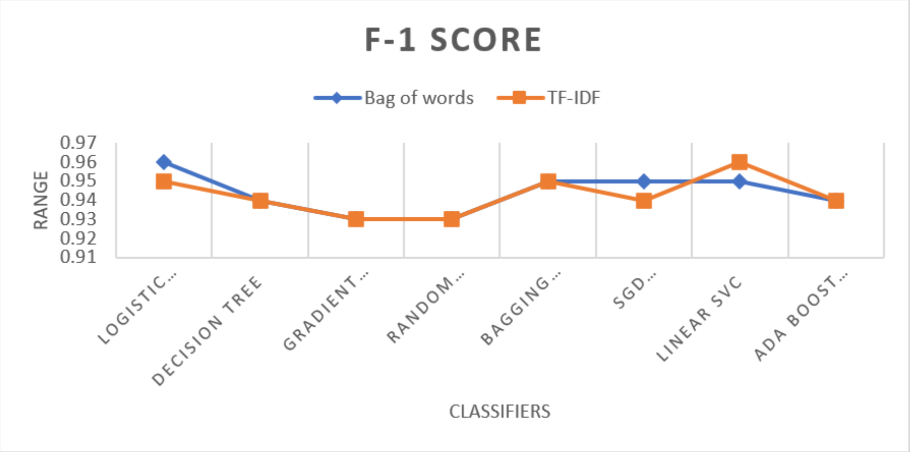


Fig.19. F-1 Score for both Model.

Fig 19 depicts that the F-1 score highest for Linear SVC on TF-IDF model at the same time on Bag of words model the highest F-1 score comes through Logistic regression. Rest of the classifiers f-1 score is considerable Except gradient boosting and random forest classifier as they showing same outcome like other accuracy parameters.

## 6. Discussion

This study aims to identify online bullying more quickly and accurately with improved precision so that people can benefit at any stage. Here, a gathered dataset is tested, and our solution is recommended in comparison to other existing methods. The idea of labeling or marking any images that contain cyberbullying is considered significantly. From previous study it was seen that 96% accuracy given by SLE and DLE models when TFIDF feature extraction is paired with K-Fold cross-validation [16] which means similar to the accuracy of our study but from fig 9 it can be said that similar kind of accuracy can be found from four different simpler combination which can be followed with less arrangements. In another study they got highest precision for SGD classifier which is 97% [17] where our study managed close to that rate for Logistic regression and linear SVC which is 96% with different model. Though the recall rate of Logistic regression, SGD classifier and Linear SVC outperformed previous study's which is 96%. Similarly, F-1 score of Logistic regression on Bag of words and Linear SVC on TF-IDF is not only similar but also higher than other study. Compared to the many criteria it has been discovered that Logistic regression on Bag of Words and Linear SVC on TF-IDF have the highest efficiency from different accuracy measures. As a result, we can say that these two classifiers, when used in conjunction with the appropriate approach, are capable of detecting cyberbullying with a great accuracy from the text extracted from the snap using easyocr. In methodology section we mentioned the sample outcome with our system where most of them are accurate compared to input. It is found that our approach offers better grouping and accuracy. This can support users while also safeguarding them from the negative effects of cyberbullying. However, numerous forms of pointless phrases, such as profile names, syntactic terms, and social network instructions, are automatically fed to the algorithm. Besides line break doesn't count by the system that's why a sentence with a break classified separately. These reasons may imbalance the detection of different text as they are taking part in classification. So, these gaps should be fixed in future though the objective of this paper matched closely with the outcome.

## 7. Conclusion & Future Works

In part as a result of the way today's internet has transformed the way people communicate, they are exposed to a variety of threats, including online bullying and other types of harassment, among other things. In recent years, online harassment, in particular, has grown in popularity as a result of the establishment of social networking sites and the increased use of social media by teenagers. So, these developments, online harassment, in particular, has begun to pose serious societal challenges that must be addressed. Though it is not employed enough by many platforms to use automated cyberbullying detection systems in order to mitigate the negative consequences of cyber harassment without delay. Making the necessary preparations ahead of time, regardless of whether the source is a snapshot, is critical to responding effectively.

It is discussed in this article how to detect cyberbullying in snap using simplest and effective techniques such as optical character recognition, natural language processing, and machine learning to identify the bully. Additionally, feature extraction is carried out utilizing it alone with both the Bag of Words model and the TFIDF. According to the findings, the accuracy of both the Bag of Words model and the TF IDF model was 96 percent when Logistic regression was used in both instances. Besides, when the Linear SVC's TF-IDF was employed in this investigation, we noticed that it had same accuracy rate, which was impressive. The findings indicate that when compared to prior similar studies, our technique surpassed other classifications in terms of accuracy, precision, recall, and the f-1 score. The suggested system can be used by the administration or any groups, institutions, monitors, policy makers, and enforcement bodies. Individuals' ability to communicate on social media in a safe atmosphere will undoubtedly be enhanced as a result of our research's high degree of precision in identifying cyberbullying. Since combating online bullying is a continuous process, strategies must be frequently revised to account for the most recent advancements. The goal is to develop a more complex system for detecting cyberbullying from images in the future, which will be able to distinguish between Bengali text messages and other languages, among other things.

## References

[1] A.Saravanaraj, J. I. Sheeba, S. Pradeep Devaneyan, 2016. Automatic Detection of Cyberbullying from twitter. IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN.

[2] M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana, “Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network,”Comput. Human Behav., vol. 63, 2016, pp. 433–443.

[3] Raisi, E. and Huang, B., 2017, July. Cyberbullying detection with weakly supervised machine learning. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 (pp. 409-416).

[4] Basarslan, M.S. and Kayaalp, F., 2020. Sentiment Analysis with Machine Learning Methods on Social Media.

[5] J. Han, M. Kamber and J. Pei, “Data Mining: Concepts and Techniques,” Elsevier, Morgan Kaufmann Series in Data Management Systems, vol. 3, 2011

[6] B. Liu, “Sentiment Analysis and Opinion Mining,” Synthesis Lectures on Human Language Technologies, Morgan & Claypool, vol. 5, no. 1, pp. 1-167, May 2012

[7] Hosseinmardi, H., Rafiq, R.I., Han, R., Lv, Q. and Mishra, S., 2016, August. Prediction of cyberbullying incidents in a media-based social network. In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 186-192). IEEE.

[8] Kargutkar, S.M. and Chitre, V., 2020, March. A study of cyberbullying detection using machine learning techniques. In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) (pp. 734-739). IEEE.

[9] Choudhary, A., Rishi, R. and Ahlawat, S., 2013. A new approach to detect and extract characters from off-line printed images and text. Procedia Computer Science, 17, pp.434-440.

[10] Akopyan, M.S., Belyaeva, O.V., Plechov, T.P. and Turdakov, D.Y., 2019, September. Text recognition on images from social media. In 2019 Ivannikov Memorial Workshop (IVMEM) (pp. 3-6). IEEE.

[11] Kumar, A. and Sachdeva, N., 2021. Multimodal cyberbullying detection using capsule network with dynamic routing and deep convolutional neural network. Multimedia Systems, pp.1-10.

[12] Ranjan S, Sanket S, Singh S, Tyagi S, Kaur M, Rakesh N, Nand P. OCR based Automated Number Plate Text Detection and Extraction. In2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) 2022 Mar 23 (pp. 621-627). IEEE.

[13] Zhao, R., Zhou, A. and Mao, K., 2016, January. Automatic detection of cyberbullying on social networks based on bullying features. In Proceedings of the 17th international conference on distributed computing and networking (pp. 1-6).

[14] Drishya, S.V., Saranya, S., Sheeba, J.I. and Devaneyan, S.P., 2019. Cyberbully image and text detection using convolutional neural networks. CiiT International Journal of Fuzzy Systems, 11(2), pp.25-30.

[15] Islam, M.M., Uddin, M.A., Islam, L., Akter, A., Sharmin, S. and Acharjee, U.K., 2020, December. Cyberbullying detection on social networks using machine learning approaches. In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (pp. 1-6). IEEE.

[16] Perera, A. and Fernando, P., 2021. Accurate cyberbullying detection and prevention on social media. Procedia Computer Science, 181, pp.605-611.

[17] Alam, K.S., Bhowmik, S. and Prosun, P.R.K., 2021, February. Cyberbullying detection: an ensemble based machine learning approach. In 2021 third international conference on intelligent communication technologies and virtual mobile networks (ICICV) (pp. 710-715). IEEE.

[18] Muneer, A. and Fati, S.M., 2020. A comparative analysis of machine learning techniques for cyberbullying detection on Twitter. Future Internet, 12(11), p.187.

[19] Hani, J., Mohamed, N., Ahmed, M., Emad, Z., Amer, E. and Ammar, M., 2019. Social media cyberbullying detection using machine learning. International Journal of Advanced Computer Science and Applications, 10(5).

[20] Wan Noor Hamiza Wan Ali, Masnizah Mohd, Fariza Fauzi, Centre for Cyber Security, Universiti Kebangsaan Malaysia Bangi, Selangor, 2020. Cyberbullying Predictive Model: Implementation of Machine Learning Approach.

[21] Kumar, A., Nayak, S. and Chandra, N., 2019. Empirical analysis of supervised machine learning techniques for Cyberbullying detection. In International Conference on Innovative Computing and Communications (pp. 223-230). Springer, Singapore.

[22] Hani, J., Nashaat, M., Ahmed, M., Emad, Z., Amer, E. and Mohammed, A., 2019. Social media cyberbullying detection using machine learning. Int. J. Adv. Comput. Sci. Appl, 10(5), pp.703-707.

[23] Monirah Abdullah Al-Ajlan, Mourad Ykhle, King Saud University, 2018. Deep Learning Algorithm for Cyberbullying Detection. (IJACSA) International Journal of Advanced Computer Science and Applications.

[24] Elsafoury, Fatma (2020), “Cyberbullying datasets”, Mendeley Data, V1, doi: 10.17632/jf4pzyvnpj.1

## Authors' Profiles

**Tofayet Sultan** was born in 1999. He received BSc Degree in Computer Science & Engineering from American International University – Bangladesh in 2022. His research focuses on Machine learning, Data science and Human Computer Interaction.

**Nusrat Jahan** was born in the year 2000. She attended the American International University - Bangladesh and earned a Bachelor of Science degree in Computer Science and Engineering in 2022. Her research focuses on areas such as Human-Computer Interaction and Machine Learning.

**Ritu Basak** was born in 2000. She is a current student at the American International University - Bangladesh, where she is majoring in Computer Science and Engineering. Deep Learning and Software Engineering are the primary subjects of her research.

**Md. Shaheen Alam Jony** was born in the year 1999. Currently, he is pursuing a degree in Computer Science and Engineering at the American International University - Bangladesh, where he is enrolled as a student. His research focuses mostly on Deep Learning and Software Engineering as his key areas of study.

**Rashidul Hasan Nabil** is currently working as a Lecturer in the Department of Computer Science under the Faculty of Science and Technology at American International University-Bangladesh (AIUB). Previously he has been working as a Lecturer in the Department of Computer Science and Engineering under the Faculty of Science and Engineering at City University, Bangladesh from 2017 to 2019. His research interest includes Human-Machine Interaction, Human-Computer Interaction, Machine Learning, and Deep Learning. He has published a number of his research works at different conferences.

**How to cite this paper:** Tofayet Sultan, Nusrat Jahan, Ritu Basak, Mohammed Shaheen Alam Jony, Rashidul Hasan Nabil, "Machine Learning in Cyberbullying Detection from Social-Media Image or Screenshot with Optical Character Recognition", International Journal of Intelligent Systems and Applications(IJISA), Vol.15, No.2, pp.1-13, 2023. DOI:10.5815/ijisa.2023.02.01