2023 5th International Conference on Sustainable Technologies for Industry 5.0 (STI), 09-10 December, Dhaka

Bengali Cyberbullying Detection in Social Media Using Machine Learning Algorithms

Subrata Saha*, Md. Shamimul Islam†, Md. Mahbub Alam†, Md. Motinur Rahman*, Md. Ziaul Hasan Majumder*, Md. Shah Alam* and M. Khalid Hossain*
* Institute of Electronics, Bangladesh Atomic Energy Commission, Bangladesh
†Institute of Computer Science, Bangladesh Atomic Energy Commission, Bangladesh
Email: subratabaec@gmail.com, shamimul.jucse@gmail.com, mahbubcseju@gmail.com, motinurapece@gmail.com, majumderbd90@gmail.com, alam_ny_ni@yahoo.com, khalid.baec@gmail.com

2023 5th International Conference on Sustainable Technologies for Industry 5.0 (STI) | 979-8-3503-9431-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/STI59863.2023.10464740

Abstract— **Social media has become more prevalent and it is now fairly easy to communicate with people online. Social network users have many options to cooperate, interact positively, and exchange information. The same system might create a toxic environment that can create an unpleasant environment for online abuse and bullies. Young adults and celebrities are vulnerable to online abuse more often. That's why cyberbullying should be identified and eliminated from social media because it may significantly lead to psychological as well as emotional suffering. By utilizing Natural Language Processing (NLP), Machine Learning (ML), as well as Deep Learning Models based on Transformers like BERT, we can identify patterns in social media texts used by bullies and create an automated method that can detect abusive texts. In this study, we proposed a reliable machine-learning model for social media cyberbullying detection in the Bengali language. We applied text preprocessing, followed by feature extraction using the TF-IDF vectorizer. Then, we applied 4 ML algorithms and 1 transformer-based pre-trained BERT model and evaluated their performances by different performance metrics. Our study found that BERT worked best compared to other algorithms and achieved an accuracy of 90% and an AUC (Area under the ROC Curve) of 0.96.**

Keywords— Cyberbullying, Bangla bullying detection, Hate speech detection, NLP, Machine learning, BERT.

I. INTRODUCTION

Bullying is a type of aggression that inflicts victims with either immediate or long-term harm or distress. These aggressions include physical such as hitting, and tripping, verbal for example teasing, emotional, and social such as spreading false allegations. As a result, it may pose a severe public health risk [1]. Also, according to the Centers for Disease Control and Prevention (CDC), bullying is defined as any unwelcome hostile behavior(s) by another person or group of people that involves an imbalance of power, and is repeated or is very likely to repeat [2]. In recent times, as a result of advancements in information technology and easy access to digital devices to people, many forms of cyber security threats have surfaced and researchers are trying to mitigate the problems by incorporating different security measures [3]-[5]. Moreover, a new form of bullying has emerged in cyberspace dubbed as "cyberbullying." Cyberbullying can be defined as a sort of psychological harassment that takes place through the use of technology including smartphones, blogs, and Social networking sites such as Facebook, YouTube, and Twitter. This type of abuse can happen in different ways like sending threatening personal messages, commenting abusively on social media posts, and spreading rumors by manipulating social media posts [6],[7]. Some people consider cyberbullying as an extension of traditional harassment, however, cyberbullying differs from traditional bullying in several concerning ways. Unlink traditional bullying, a major part of cyberbullying is anonymous. In most of the cases, the perpetrator is unknown to the victim and also to the public. Furthermore, rumors can be spread to a larger audience with remarkable speed. This aspect of cyberbullying makes it incredibly challenging to regulate or deal with [8]-[9].

Over the past decades, Bangladesh has experienced a steady rise in internet usage. However, it is also thought to be contributing to an increase in harassment of women because of society's predominance of patriarchal views and customs as well as the lack of proper legal protection. In a poll conducted by ActionAid Bangladesh, 50% of the women who participated in the study reported experiencing internet harassment. More than 62% of the victims were under the age of 25. It's interesting to note that the victims identified Facebook as the main website where they experienced the most harassment. As a result of these circumstances, 76 percent of women had mental health issues like anxiety and sadness. About 30% of women were unaware of their options for filing complaints [10], [11].

Researchers have been trying to build automated systems to solve real-life problems [12], [13]. A lot of research has been done to incorporate the power of AI to develop an automated system that can classify texts to find offensive content on social media platforms in both English and other languages [14],[15]. Also, there are several research that focus on the Bengali language to identify the abusive language because the detection of cyberbullying in the Bengali language is of the utmost importance. Across a large portion of the world, Bengali is the most commonly used language for communication [16]-[18]. However, there is a critical need for an enhanced approach that can recognize different forms of offensive Bengali content.

To find and delete offensive Bengali content from social media networks, ML techniques can be very successful. In our study. We proposed an ML technique for effective cyberbullying classification in the Bengali language and to protect users from harassment on social media. Through focusing on the distinct Bengali language context, this study aims to progress the field of Bengali cyberbullying detection. In our study, we explored four machine learning methods, such as Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), XGBoost (Extreme Gradient Boosting), and 1 transformer- based BERT (Bidirectional Encoder

979-8-3503-9431-3/23/$31.00 ©2023 IEEE

Representations from Transformers) model "Bangla-bert-base". The word embedding technique that we used was Term frequency-Inverse document frequency (TF-IDF) and the findings showed that SVM performed better among 4 machine learning algorithms although Bangla-bert-base outperformed all ML algorithms explored in this work.

The remainder of the paper is arranged as follows. Section 2, examines the relevant articles for identifying cyberbullying and text classification. Our proposed methodology has been discussed in section 3. Section 4 focuses on the experimental setup and result analysis. Accuracy comparisons with previous studies have been discussed in section 5. And finally, the paper concludes by discussing some future prospects on how to improve our study.

II. RELATED WORKS

Multiple studies have been done on detecting cyberbullying using machine learning techniques. Some of the notable works in other languages are discussed here. The earliest work was done by Reynolds et al. [19]. They collected their data from Formspring.me which was a website focused on queries and answering with a significant number of comments related to bullying. Then labelled the dataset using Amazon's web services called Mechanical Turk. Later Weka tool was used to train and classify bullying texts. Finally, they successfully achieved 78.5% accuracy by applying the C4.5 algorithm and an instance-based learner. Another study done by Garadi et al. [20] worked on abusive behaviors on social media sites by reviewing cyberbullying classification techniques. They extensively studied algorithms related to feature selection and then studied ML models to identify abusive behaviors in social media. Hani et al. [21] explored a supervised ML approach to detect a pattern in cyberbullying texts. In their study, they showed that NN (Neural Network) performed better than SVM in abusive text classification with an accuracy of 92.8% compared to SVM (90.3% accuracy). Dalvi et al. [14] investigated a software-based approach to identify abusive tweets. They used SVM and NB algorithms and obtained about 71.25% and 52.70% accuracy respectively in identifying bullied texts. Saha et al. [22] investigated text sentiments from Twitter data using two different techniques named VADER and BERT. Later they surveyed 5 different ML algorithms for detection accuracy and achieved 92% accuracy using BERT.

Some important works also have been done on the Bengali language for classifying Bengali cyberbullying texts. In one study, methods based on deep learning (DL) as well as ML were utilized to identify abusive texts by Emon et al. [23]. They collected Bengali comments from several social media sites, online blogs, and newspapers and showed that the deep learning technique using Recurrent Neural Networks (RNN) performs better than other ML algorithms by achieving an accuracy score of 82.20%. They also proposed a new stemming technique for the Bengali language. Comments from more than 44000 users on the Facebook platform have been collected for classification by Ahmed et al. [24]. The dataset was categorized into 5 classes named non bully, sexual, threatening, troll, and religious. The researchers used neural networks and explored the effectiveness of both binary and multiclass classification models, ultimately achieved almost 88% and 85% accuracy respectively. The first Bengali corpus for bully identification in the Bengali language domain was created by Ishmam et al. [25]. They gathered and annotated

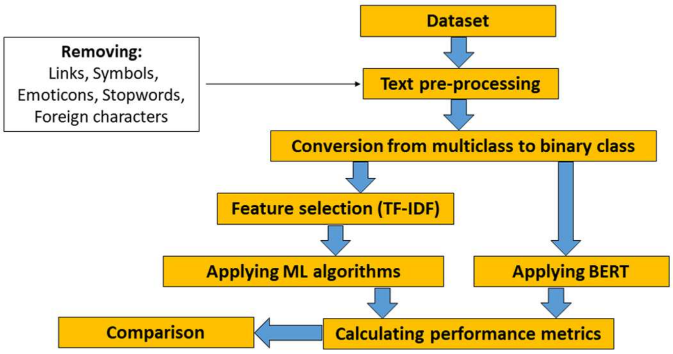


Fig. 1. Block diagram of the proposed model.

more than 5000 comments from Facebook and divided them into 6 categories. They applied the RF algorithm to achieve 52% accuracy, which was then increased by 18% after implementing the Gated Recurrent Unit (GRU). To create an automated system for cyberbullying text detection, both machine learning and natural language processing (NLP) techniques were used by Chakraborty et al. [26]. They regarded Unicode Bengali characters as well as Unicode emoticons as input for their suggested system and investigated 3 ML algorithms such as MNB, SVM, and Convolutional Neural Network (CNN). Among those, SVM showed 78% accuracy which was the maximum. Aurpa et al. [27] used the transformer-based pre-trained architectures BERT and ELECTRA to detect abusive emotions on data collected from Facebook. In their investigation, BERT exceeded ELECTRA in terms of performance. For the purpose of identifying sentence polarity, a brand-new rule-based system termed the "Bangla Text Sentiment Score" (BTSC) was created by Bhowmik et al. [28]. In their study, to determine a word's score and subsequently that of a sentence, this system takes into consideration words that are parts of speech tagger. Their model showed a maximum accuracy of 82.21% using SVM with the BiGram feature. Akhter et al. [29] used the TF-IDF vectorizer, balanced the dataset using Instance Hardness Threshold (IHT), and achieved 98.82% accuracy in multiclass classification. In order to detect fake news Sharifani et al. [30] used 3 ML classifiers on 2 datasets that were available publicly. Shakambhari et al. [31] applied an ensemble method to detect cyberbullying texts on the Reddit platform and proposed a model combining SVM, TF-IDF, and Bag of Words (BoW).

III. METHODOLOGY

The study consists of 5 parts. First dataset collection and preparation, followed by preprocessing, then feature extraction, and later application of ML algorithms, and finally evaluation of their performances. The whole proposed methodology is shown in Fig. 1 with a block diagram.

A. Dataset collection and preparation

The dataset used in this study includes people's opinions from the comments section in social media posts made by celebrities found on Facebook, including television and movie actors, models, sports figures, musicians, and politicians. In total, 44001 comments were collected for this dataset [32]. The dataset contains a total of 5 variables (columns) namely comment, Category, Gender, comment react number, and label, here label is the target variable. From the pie charts, there were about 32% of remarks were aimed at men, and 68.1% of comments were directed at women as shown in Fig 2(a).

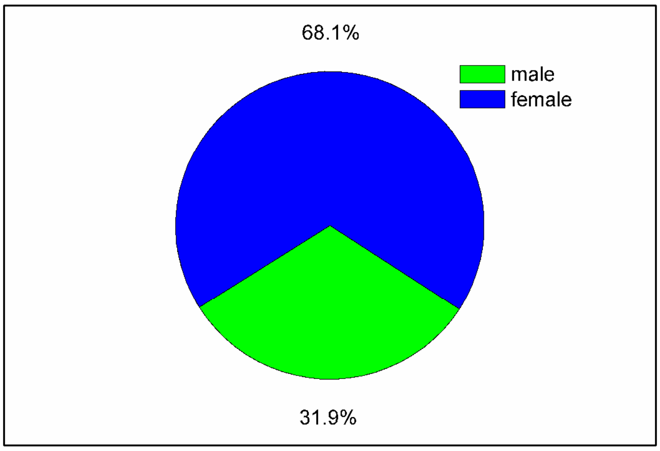
Moreover, Fig 2(b) demonstrates, that comments were categorized into 5 types such as threat, communal attacks, sexual, troll, and neutral. More than 36% of texts were neutral comments, 23.3% of texts were trolling, 19.5% were sexual attacks, 17.5% were religious and 3.3% of comments were threats.

B. *Text preprocessing*

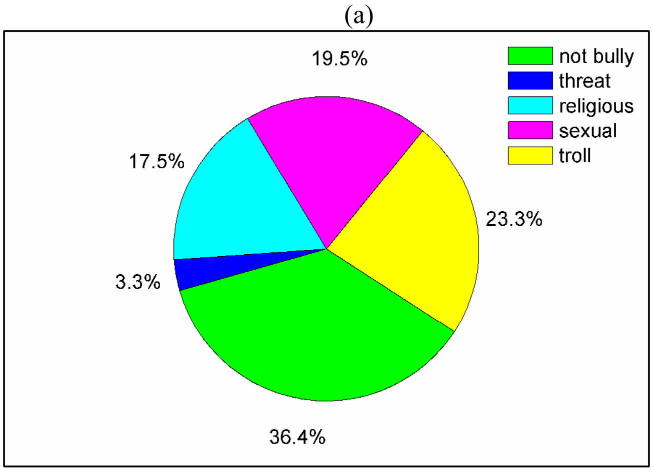
Making sure that the data can be understood by machines is a vital part of data analysis. That's why, before applying any kind of ML algorithm data needs to be processed through filtering and tokenization. Moreover, normal texts contain many kinds of symbols and words other than our desired language, so filtering is needed. In our research, by filtering techniques, we removed website links, punctuation, and symbols, signs, emoticons, any characters other than Bengali. Followed by the elimination of Bengali Stopwords. Tokenization is dividing each text into smaller words based on a delimiter (space). These words or smaller units are called tokens. Afterward, for our experiment, we converted our dataset from multiclass to binary class considering threat, communal attacks, sexual, and troll texts as bullying text, on the other hand, neutral comments were kept unchanged. For this purpose threat, communal attacks, sexual, and troll texts are represented as 1 (bully) and neutral as 0 (non-bully), at this point, the dataset contains 28661 bullying texts (1) and 15340 non-bullying texts (0). A sample from the dataset after filtering is depicted in Fig 3.

C. *Feature extraction*

Feature extraction is implemented because text data needs to be handled to train ML models. These methods are employed to represent the words numerically. There are several techniques used by researchers such as Bag of Words (BOW), One Hot Encoding, TF-IDF, and word2vec. In our study, we used TF-IDF as the text vectorizer.



(a)



(b)

Fig. 2. Pie charts of the (a) Gender and (b) label variables.

	comment	new_label	final_text
5	অন্যরকম .. ভালো লাগলো ..❤️	0	অন্যরকম ভালো লাগলো
100	কোনটা ছেড়ে কোনটা রাখি...। অসাধারণ ....অসা...	0	কোনটা ছেড়ে কোনটা রাখি অসাধারণ অসাধারণ তিশার তু...
116	এটা দিয়ে কি TikTok করা যাবে........	0	দিয়ে
372	গুড looks	0	গুড
11932	সাফা is not সাফা She is নাস্তিক আফা	1	সাফা সাফা নাস্তিক আফা
43911	আরেএএএএ!!!🖤🖤🖤	0	আরেএএএএ
43803	জাহান্নামি নারী......id er Name lihke rahkon.....	1	জাহান্নামি নারী নাস্তিক উপাদি অর্জন করছেননতুন ...

Fig. 3. Sample dataset after pre-processing.

Term frequency-Inverse document frequency vectorizer (TF-IDF) is a vectorization technique that converts text data into vectors. From raw texts, it creates a matrix of features.

Term Frequency (TF): TF counts the number of times a term appears in a document. TF is calculated using the equation 1.

$$TF = \frac{number\ of\ occurences\ of\ a\ term\ t\ in\ the\ document}{total\ words\ in\ the\ document} \quad (1)$$

Inverse Document Frequency (IDF): This is a weight that is an indicator of how frequently a term is used in a document. Its score decreases with increased usage across a document and scales up the less frequent words using equation 2.

$$IDF = log\frac{N}{DF_t} \quad (2)$$

Here, N is the total number of text documents and DF_t is the number of texts that use the term t.

TF-IDF: It is the product of IDF and TF as shown in equation 3.

$$TF - IDF(t) = TF(t) * IDF(t) \quad (3)$$

D. *BERT*

Bidirectional Encoder Representations from Transformers, which is termed as BERT, is a DL model that is based on Transformers. Every output element in a transformer is connected to every input, and attention-based dynamic weighting determines the relative importance of each element. The difference between BERT and previous language models is that BERT can simultaneously read texts in both directions contrary to other language models which could read text inputs in one direction only. BERT is a full language model that uses an embedding method as one of its constituent parts, not just an embedding technique [22], [33].

Here, bangla-bert-base is a pre-trained model for the bengali language utilizing mask language modeling which is detailed in BERT. Two primary sources were used to download Corpus which are Open Super-large Crawled Aggregated coRpus (OSCAR) and Bengali Wikipedia. Google BERT's code was used to train bangla-bert and the latest model contains 12 layers, 768 hidden layers, and 110 million parameters in its architecture [33].

We used the "simple transformer" NLP library to implement BERT and "ClassificationModel" as simple transformer model for binary classification. [34]

E. *Classification*

We tried 4 different ML classifiers (NB, RF, SVM, and XGB) and 1 BERT pre-trained model called bangla-bert-base [34] to train our dataset and classify Bengali cyberbullying texts.

Later we evaluated and compared their effectiveness considering different performance indicators.

F. *Performance metrics*

Accuracy: It is described as the proportion of all predictions made to the number of predictions that were accurate, as shown in equation 4.

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{4}$$

Precision: It is calculated by the ratio of the number of TP and the total number of positive predictions, shown in equation 5.

$$precision = \frac{TP}{TP+FP} \tag{5}$$

Recall: It is calculated by equation 6,

$$recall = \frac{TP}{TP+FN} \tag{6}$$

. F1 Score: It is determined as the harmonic average of recall and precision, shown in equation 7.

$$F\,1 = \frac{2}{precision^{-1}+recall^{-1}} \tag{7}$$

IV. Result analysis

After applying selected classifiers with hyper-parameter tuning using GridSearchCV and K fold cross-validation (where K=10) we got our performance indicators for all algorithms. A comparison of accuracy scores among all algorithms is shown in Fig 4. Table I shows TP, FP, FN, TN, TPR, FPR, and AUC scores. Moreover, Precision, recall, and F1 measures in detecting bully (1) and non-bully (0) texts are shown in Fig 5 and Fig 6 respectively.

From Fig 4, all the ML algorithms (except BERT) give almost similar accuracy scores ranging from 76% to 79%. However, BERT significantly improves the performance with an accuracy of 90%. The second-best accuracy is demonstrated by SVM (79% accuracy) and the lowest accuracy is achieved by NB (76% accuracy).

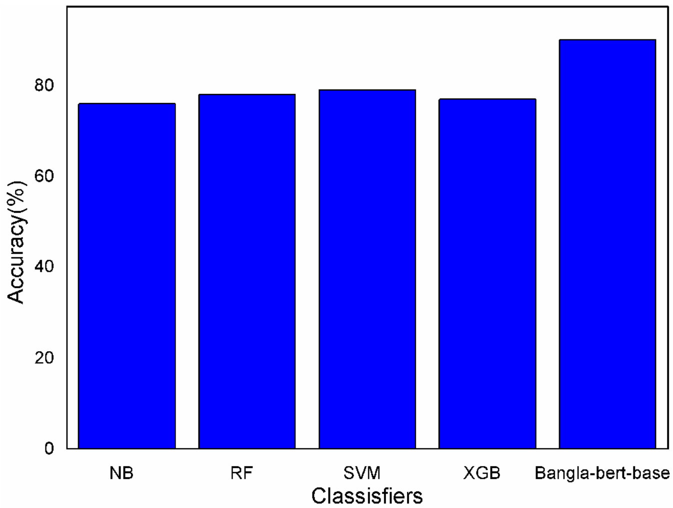


Fig. 4. Accuracy scores of ML algorithms.

TABLE I. TP, FP, FN, TN, AUC, TPR, FPR FOR ALL ALGORITHMS

	TP	FP	FN	TN	AUC ROC	TPR	FPR
NB	1562	2271	327	6841	0.83	0.82	0.24
RF	2061	1772	631	6537	0.84	0.76	0.21
SVM	2089	1744	544	6624	0.83	0.79	0.20
XGB	1799	2034	456	6721	0.83	0.79	0.23
BERT	6616	744	513	3128	0.96	0.93	0.19

As can be observed from Table I, among all the algorithms BERT successfully identifies 6616 bully texts and 3128 non-bully comments. Also, BERT achieves a maximum True Positive Rate (TPR) of 0.93 and the lowest False Positive Rate (FPR) of 0.19, of all the algorithms. This also indicates that

BERT performs better than other algorithms. Moreover, AUC ROC (Area under the curve of the Receiver Operating Characteristic) is a parameter that indicates how good a model is. The nearer the score towards 1, the better the model. In our study, BERT gives an AUC of 0.96 which is very close to 1. This demonstrates the model's efficiency in distinguishing Bangla bullying and non-bullying texts.

In our binary classification study, BERT performs well in contrast to other algorithms is also reflected in Fig 5 and Fig 6. Comparing Fig 5 and Fig 6, it can be seen that, in most cases, algorithms perform better in terms of precision, recall, and F1 score in predicting bullying texts compared to non-bullying texts. BERT gives precision, recall, and F1 scores as 0.92, 0.91, and 0.92 respectively for bullying texts (Fig 5) and 0.84, 0.84, and 0.84 for non-bullying texts (Fig 6). In both cases, BERT gives better precision, recall, and F1 scores compared to others (except in a single case for the NB algorithm).

Word clouds for Bangla bullying and non-bullying texts are shown in Fig 7 and Fig 8 respectively. In Fig 7 Some noticeable bully words are “নাস্তিক”, ”তুই”, ”তোর”, “সাফা ক□র”, ”পরকাল ”.These indicate significant number of comments are intended for religious attacks, calling out names, and towards a particular celebrity. From Fig 8 it can be seen some of the more frequent non-bullying or neutral words such

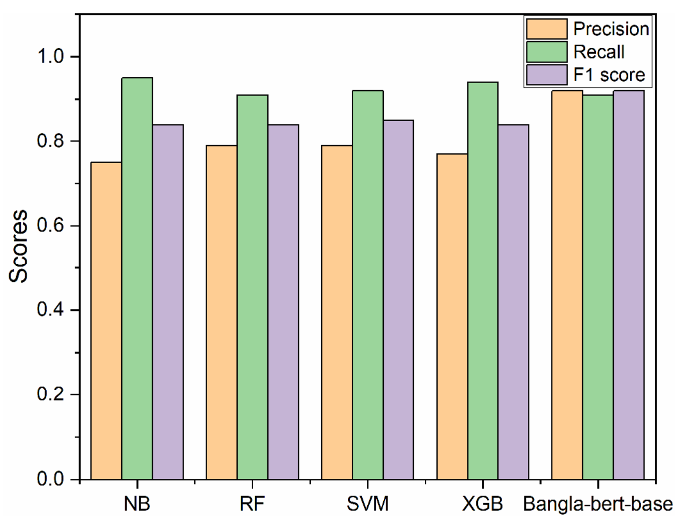


Fig. 5. Precision, Recall, and F1 score for bullying comments.

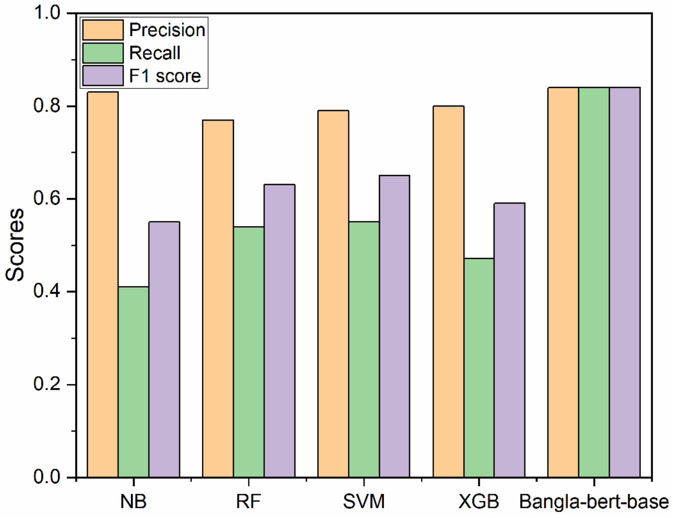


Fig. 6. Precision, Recall, and F1 score for non-bullying comments.

as, “ভালো”, ”সুন্দর”, ”হিরো আলম”, ”হেদায়েত দান”, ”মানুষ” which indicate most remarks contain good words and giving religious suggestions. Normally it is customary to address unknown people as “আপনি” or “তুমি” in Bengali language, here “তুই” , “তোর ” also indicate derogatory remarks shown in Fig 7.

V. Comparison with other studies

Our experimental results demonstrate that, compared to existing research findings for Bengali cyberbullying detection, our proposed model gets the highest accuracy rate for Binary classification. The relative effectiveness of various models is revealed by the comparative analysis in Table II.

VI. Conclusion and Future Works

Cyberbullying is more common than it is anticipated and might be difficult to effectively identify bullies on social media sites because, in most cases, culprits in cyberspace are not related to the targeted person. So it’s a necessity to identify and delete texts containing bullying sentiments. However, achieving this target for a variety of popular languages with a single model or technique is difficult. This is due to the diversity of languages. This article addresses recognizing text expressions for Bengali users on the Facebook platform and successfully identified bullying and non-bullying comments with an accuracy of 90% using BERT. We incorporated text preprocessing by removing links, symbols, and characters from other languages removing Bengali stop words, and then converting to binary class.

Fig. 7. Word cloud for frequent bullying words.

Fig. 8. Word cloud for frequent non-bullying words.

Following that, feature extraction using TF-IDF was utilized to convert the texts into numerical values. Afterward, we applied 2 techniques, at first, we implemented 4 ML algorithms (with hyperparameter tuning and 10-fold cross-validation), and all of them showed very close accuracy scores and almost similar AUC scores. The highest accuracy score was achieved by SVM (79%) although the maximum AUC was accomplished by RF which is 0.84. Then in the second phase, we tested 1 transformer-based pre-trained BERT model “bangla-bert-base”, which remarkably improved the performance by attaining 90% accuracy and AUC of 0.96, which showed the best result. The better performance of the transformer-based model is also demonstrated by TP, FP, FN, TN, and precision, recall, and F1 scores. Besides, Most used bullying words and neutral words were also studied and it was found that religious attacking words were prominent and females were mostly targeted.

Although showing potential in detecting Bengali cyberbullying texts with both ML and transformer-based approaches, our work has some issues and is subject to some challenges. Our dataset contains only Bangla text data, as a result, detecting cyberbullying texts in other languages is not possible with this model. Other ML, Deep Learning as well and pre-trained models need to be explored for Bengali cyberbullying detection whether there are any better alternatives or not. Moreover, the accuracy may be improved by collecting more text data on diverse fields. In our dataset, Non-bullying data is almost half of the bullying data, which makes it an imbalanced dataset, so better performance can be achieved with a balanced dataset.

The main theme of Industry 5.0 is the collaboration between people and robots as well as intelligent machinery to achieve sustainable development with the advent of newer technologies considering the values of the stockholders. It is all about utilizing innovative technologies such as Big Data, the Internet of Things (IoT), ML, Quantum Computing, and

TABLE II. Comparison with other studies

Author	Technique used	Best Accuracy
Emon et al. [23]	RNN	82.20%
Ahmed et al. [24]	Neural network	87.91%
Aurpa et al. [27]	BERT	85.00%
Ishmam et al. [25]	ML and GRU based Model	70.70%
Our study	ML and BERT-based model	90.00%

DL to enable robots to assist humans in doing tasks more efficiently and quickly. Our study focuses on detecting Bengali bullying texts in social media using different machine learning techniques and we believe, this can ensure a safe and sustainable hatred-free cyberspace for social media users.

REFERENCES

[1] A. M. Vivolo-Kantor, B. N. Martell, K. M. Holland, and R. Westby, "A systematic review and content analysis of bullying and cyber-bullying measurement strategies," Aggress. Violent Behav., vol. 19, no. 4, pp. 423–434, 2014, doi: 10.1016/j.avb.2014.06.008.

[2] R. M. Gladden et al., "Bullying surveillance among youths," Centers Dis. Control Prev. Atlanta, Georg., pp. 4–101, 2014.

[3] M. S. Islam, M. A. Uddin, D. M. S. Ahmed, and G. Moazzam, "Analysis and Evaluation of Network and Application Security Based on Next Generation Firewall," Int. J. Comput. Digit. Syst., vol. 13, no. 1, pp. 193–202, 2023, doi: 10.12785/ijcds/1smart116.

[4] M. S. Islam, N. K. Datta, M. I. H. Showrov, M. M. Alam, M. H. Ali, and M. D. Hossain, "Organizational Network Monitoring and Security Evaluation Using Next-Generation Firewall (NGFW)," 2023, pp. 133–147, doi: 10.1007/978-981-19-8032-9_10.

[5] K. Thakur, T. Hayajneh, and J. Tseng, "Cyber Security in Social Media: Challenges and the Way Forward," in IT Professional, Mar. 2019, vol. 21, no. 2, pp. 41–49, doi: 10.1109/MITP.2018.2881373.

[6] S. Shariff and R. Gouin, "Cyber-Dilemmas: Gendered Hierarchies, New Technologies and Cyber-Safety in Schools," Atl. Crit. Stud. Gender, Cult. Soc. Justice, vol. 31, no. 1, pp. 27–37, 2006.

[7] A. Maftei, A. C. Holman, and I. A. Merlici, "Using fake news as means of cyber-bullying: The link with compulsive internet use and online moral disengagement," Comput. Human Behav., vol. 127, p. 107032, Feb. 2022, doi: 10.1016/J.CHB.2021.107032.

[8] A. K. Das, A. Ashrafi, and M. Ahmmad, "Joint cognition of both human and machine for predicting criminal punishment in judicial system," 2019 IEEE 4th Int. Conf. Comput. Commun. Syst. ICCCS 2019, pp. 36–40, Feb. 2019, doi: 10.1109/CCOMS.2019.8821655.

[9] E. V. Altay and B. Alatas, "Detection of Cyberbullying in Social Networks Using Machine Learning Methods," in International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism, IBIGDELFT 2018 - Proceedings, Jan. 2019, pp. 87–91, doi: 10.1109/IBIGDELFT.2018.8625321.

[10] D. M. H. Kee et al., "Cyberbullying on social media under the influence of COVID-19," Glob. Bus. Organ. Excell., vol. 41, no. 6, pp. 11–22, Sep. 2022, doi: 10.1002/JOE.22175.

[11] "Combating cyber-bullying of women." https://thefinancialexpress.com.bd/views/opinions/combating-cyber-bullying-of-women-1650466631.

[12] M. M. Rahman, S. Saha, M. Z. H. Majumder, F. Akter, M. A. S. Haque, and M. Anzan-Uz-Zaman, "Design and Development of an IoT-based Smart System to Monitor and Control Environment of a Laboratory," 2022, doi: 10.1109/STI56238.2022.10103270.

[13] M. Motinur Rahman et al., "Energy Conservation of Smart Grid System Using Voltage Reduction Technique and Its Challenges," Evergreen, vol. 9, no. 4, pp. 924–938, Dec. 2022, doi: 10.5109/6622879.

[14] R. R. Dalvi et al., "Detecting A Twitter Cyberbullying Using Machine Learning," Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020, pp. 297–301, May 2020, doi: 10.1109/ICICCS48265.2020.9120893.

[15] A. Sharker, M. A. R. Farhab, T. A. Tamanna, U. Rumman, M. T. R. Shawon, and N. C. Mandal, "A Cross-Corpus Deep Learning Approach to Social Media Emotion Classification," 2022, doi: 10.1109/STI56238.2022.10103232.

[16] T. Mahmud, S. Das, M. Ptaszynski, M. S. Hossain, K. Andersson, and K. Barua, "Reason Based Machine Learning Approach to Detect Bangla Abusive Social Media Comments," Lect. Notes Networks Syst., vol. 569 LNNS, pp. 489–498, 2023, doi: 10.1007/978-3-031-19958-5_46/COVER.

[17] M. I. H. Emon, K. N. Iqbal, M. H. K. Mehedi, M. J. A. Mahbub, and A. A. Rasel, "Detection of Bangla Hate Comments and Cyberbullying in Social Media Using NLP and Transformer Models," Commun. Comput. Inf. Sci., vol. 1613 CCIS, pp. 86–96, 2022, doi: 10.1007/978-3-031-12638-3_8/COVER.

[18] T. Al Mahmud, S. Sultana, T. I. Chowdhury, and F. R. Anando, "A New Approach to Analysis of Public Sentiment on Padma Bridge in Bangla Text," 2022, doi: 10.1109/STI56238.2022.10103315.

[19] K. Reynolds, A. Kontostathis, and L. Edwards, "Using machine learning to detect cyberbullying," in Proceedings - 10th International Conference on Machine Learning and Applications, ICMLA 2011, 2011, vol. 2, pp. 241–244, doi: 10.1109/ICMLA.2011.152.

[20] M. A. Al-Garadi et al., "Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges," IEEE Access, vol. 7, pp. 70701–70718, 2019, doi: 10.1109/ACCESS.2019.2918354.

[21] J. Hani, M. Nashaat, M. Ahmed, Z. Emad, E. Amer, and A. Mohammed, "Social media cyberbullying detection using machine learning," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 5, pp. 703–707, 2019, doi: 10.14569/ijacsa.2019.0100587.

[22] S. Saha, M. I. H. Showrov, M. M. Rahman, and M. Z. H. Majumder, "VADER vs. BERT: A Comparative Performance Analysis for Sentiment on Coronavirus Outbreak," 2023, pp. 371–385, doi: 10.1007/978-3-031-34619-4_30/COVER.

[23] E. A. Emon, S. Rahman, J. Banarjee, A. K. Das, and T. Mittra, "A Deep Learning Approach to Detect Abusive Bengali Text," Jun. 2019, doi: 10.1109/ICSCC.2019.8843606.

[24] F. Ahmed et al., "Cyberbullying Detection Using Deep Neural Network from Social Media Comments in Bangla Language," Jun. 2021, Accessed: Aug. 18, 2023. [Online]. Available: https://arxiv.org/abs/2106.04506v1.

[25] A. M. Ishmam and S. Sharmin, "Hateful speech detection in public facebook pages for the bengali language," Proc. - 18th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2019, pp. 555–560, Dec. 2019, doi: 10.1109/ICMLA.2019.00104.

[26] P. Chakraborty and M. H. Seddiqui, "Threat and Abusive Language Detection on Social Media in Bengali Language," May 2019, doi: 10.1109/ICASERT.2019.8934609.

[27] T. T. Aurpa, R. Sadik, and M. S. Ahmed, "Abusive Bangla comments detection on Facebook using transformer-based deep learning models," Soc. Netw. Anal. Min., vol. 12, no. 1, pp. 1–14, Dec. 2022, doi: 10.1007/S13278-021-00852-X/METRICS.

[28] N. Ranjan Bhowmik, M. Arifuzzaman, M. Rubaiyat, H. Mondal, and A. History, "Bangla Text Sentiment Analysis Using Supervised Machine Learning with Extended Lexicon Dictionary," Nat. Lang. Process. Res., vol. 1, no. 3–4, pp. 34–45, Mar. 2021, doi: 10.2991/NLPR.D.210316.001.

[29] A. Akhter, U. K. Acharjee, M. A. Talukder, M. M. Islam, and M. A. Uddin, "A robust hybrid machine learning model for Bengali cyber bullying detection in social media," Nat. Lang. Process. J., vol. 4, p. 100027, Sep. 2023, doi: 10.1016/J.NLP.2023.100027.

[30] K. Sharifani, M. Amini, Y. Akbari, and J. Aghajanzadeh Godarzi, "Operating Machine Learning across Natural Language Processing Techniques for Improvement of Fabricated News Model." Oct. 01, 2022, Accessed: Oct. 24, 2023. [Online]. Available: https://papers.ssrn.com/abstract=4251017.

[31] Shakambhari, J. S. Raj, and S. Anantha Babu, "Smart Cyberbullying Detection with Machine Learning," Lect. Notes Electr. Eng., vol. 905, pp. 237–248, 2022, doi: 10.1007/978-981-19-2177-3_24/COVER.

[32] M. F. Ahmed, Z. Mahmud, Z. T. Biash, A. A. N. Ryen, A. Hossain, and F. Bin Ashraf, "Bangla Online Comments Dataset," 2021.

[33] "sagorsarker/bangla-bert-base Hugging Face." https://huggingface.co/sagorsarker/bangla-bert-base (accessed Aug. 21, 2023).

[34] "General Usage - Simple Transformers." https://simpletransformers.ai/docs/usage/ (accessed Oct. 24, 2023).