Revue d'Intelligence Artificielle
Vol. 37, No. 1, February, 2023, pp. 179-184
Journal homepage: http://iieta.org/journals/ria

An Ensemble Approach for Cyber Bullying: Text Messages and Images

Zarapala Sunitha Bai[1*], Sreelatha Malempati[2]

[1] Department of Computer Science and Engineering, Y.S.R University College of Engineering & Technology, Acharya Nagarjuna University, Guntur 522510, Andhra Pradesh, India
[2] Department of Computer Science and Engineering, R.V.R&J.C College of Engineering, Chowdavaram, Guntur 522019, India

Corresponding Author Email: zsunithabai@gmail.com

https://doi.org/10.18280/ria.370122

Received: 9 November 2022
Accepted: 10 February 2023

Keywords:
convolutional neural networks (CNN), text mining (TM), Term Frequency (TF)-Inverse document frequency (IDF), Deep Neural Network (DNN)

ABSTRACT

Text mining (TM) is a domain used to find valuable patterns from various text documents. Cyberbullying is the term used to abuse a person online or offline platform. Nowadays, cyberbullying has become more dangerous to people who are using social networking sites (SNS). Cyberbullying is of many types, such as text messaging, morphed images, videos, Etc. It is a challenging task to prevent this type of abuse of the person in online SNS. Finding accurate text mining patterns gives better results in detecting cyberbullying on any platform. Cyberbullying is developed with the online SNS to send defamatory statements or orally bully other persons, or by using the online forum to abuse in front of SNS users. Deep Learning (DL) is one of the significant domains used to extract and learn the quality features dynamically from the low-level text inclusions. In this scenario, Convolution neural network (CNN) are DL models used to train text data, images, and videos. CNN is a compelling approach to preparing these data types and achieving better text classification. This paper describes the Ensemble model with the integration of Term Frequency (TF)-Inverse document frequency (IDF) and Deep Neural Network (DNN) with advanced feature-extracting techniques to classify the bullying text, images, and videos. Feature extraction technique extracts the features of cyber-bullying patterns from the text and images. A limited number of datasets are used to classify the data. The proposed approach also focused on reducing the training time and memory usage, which helps the classification improvement.

1. INTRODUCTION

The usage of social networking sites (SNS) is increasing rapidly every day. SNS is a platform that gives enormous opportunities and communication to people in several fields. People may discuss various issues that are more popular using this platform. In the SNS platform, cyber-bullying is one of the significant issues in the present situation. Cyber-bullying is increasing daily through several types of messages and images. In 2021, 77.96% of SNS users felt wrong about cyber-bullying [1]. 95% of people accepted that they witnessed some cyberbullying occurring online. So, this is the time to stop cyber-bullying [2]. Cyber-bullying is of many types, such as abusing the person using an SNS platform with comments, personal messages, morphed images, Etc. Cyber-bullying has become a more complicated issue and creates many problems in my personal life. Many SNS providers try to solve this issue by blocking users based on their behavior. Still, this is an unsolved issue in SNS. Text classification is a domain that belongs to various fields used to solve the various misclassification issues present in this domain.

In text classification, the features extracted by removing the noise from the given text inputs are called words, sentences, phrases, etc. [3]. It is essential to find the patterns that belong to a specific language, such as English. Various feature extraction methods are used to classify the different types of text messages.

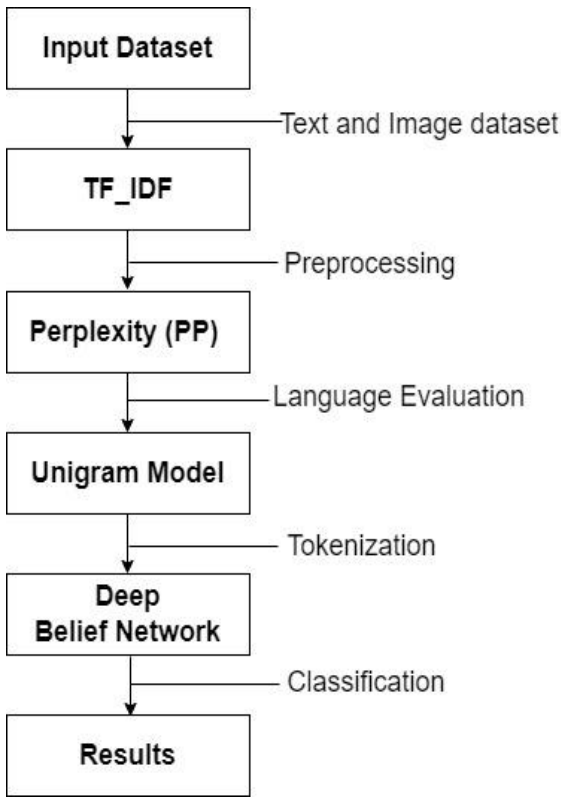


Figure 1. Architecture diagram

Figure 1 gives a detailed analysis of cyber-bullying detection by using various methods and techniques. The dataset is provided as an input, and TF_IDF is used for the preprocessing. Perplexity is utilized to evaluate the English language. The unigram model is used for tokenization which will split the sentences into words to understand the meaning of the words. Finally, DBN gives the classification of the dataset.

Organizing the bullying messages has a significant impact on using these feature extraction methods. This paper mainly focused on finding the bullying content from text messages and images in the SNS. Sentiment analysis (SA) is one of the significant tasks in finding the sentiments from the user messages or tweets in online SNS [4]. Various feature extraction models are used to extract the text and image features to analyze the feelings. These techniques improve the classification of sentiments present in the dataset. Online SNS are platforms for attackers to attack the victim with message bullying and image bullying.

Machine Learning (ML) is most widely used to detect language and Images automatically and prevents these attacks [5]. Many researchers are trying to develop an automated cyber-bullying model to detect and prevent this message [6]. Parts of Speech (POS) are most widely used to find the features that belong to polarity [7]. In the study [8], the author developed cyber-bully detection (CBD) using SA and emojis. Emoji is an expression-based image used to express a person's emotions.

In this paper, the deep neural networks (DNN) model Deep Belief Networks (DBN) is used to analyze the tweets data and image data for the classification of cyber-bullying. The proposed model integrates pre-processing techniques, tokenization, feature extraction techniques, and DBN. Two real-time datasets to analyze the performance of DBN. Figure 1 shows the overall system architecture.

2. LITERATURE SURVEY

Krizhevsky et al. [9] introduced a DCNN approach to find the objects in images. This approach extracts better features from object detection. Anand and Eswari [10] proposed the LSTM with and without integration of word GloVe embeddings to find the abused comments, store the websites circulating these types of messages, and prevent these websites and improve the safety discussions in online platforms. This paper uses the Kaggle dataset to find the various kinds of toxic comments.

Li et al. [11] proposed the text mining approach used to classify text messages using the term-based technique. In the existing methods, various issues are identified, such as polysemy and synonymy. For many years, pattern-based procedures have performed better than term-based methods. These methods cannot work on large datasets, which remains a massive text mining issue.

Nobata et al. [12] developed an ML-based approach to detect the hate speech collected from user comments present online by using two domains. A dynamic corpus consists of user comments annotated for abusive language. Kim [13] proposed that the CNN model trains the vectors to classify sentence-level tasks. The author combines several DL models to give extreme outputs on multiple datasets. Ibrohim et al. [14] proposed the integrated model using the word embedding (word2vec) feature. The proposed approach, combined with part of speech and emoji, was used to identify hate speech and abusive language on Twitter in the Indonesian language. The classification algorithms used in this study were SVM, RF, DT, and LR. Combining unigram features, part of speech, and emoji obtained the highest accuracy value of 79.85% with an F-Measure of 87.51%. Waseem and Hovy [15] introduced the method of finding hate speech from the publicly available corpus of 16k tweets. To improve hate speech detection, the extra-linguistic features with the integration of character n-grams detect accurate hate speech. Vigna et al. [16] proposed the alert-based approach (ABA), which sees hate speech in SNS. ABA focused on finding personal, caste, and religious abuse based on the text. ABA combined with the SVM and LSTM to classify the hate speech words and also by speech recognition. The two classification approaches give accurate hate speech recognition. Yenala et al. [17] proposed the novel DL approach that automatically finds irrelevant language. The novel approach solves several issues in finding irrelevant language. Irrelevant language means spelling mistakes and variations present in the language. The proposed approach is called Convolutional Bi-Directional LSTM (C-BiLSTM), combined with CNN and BLSTM. BLSTM is used to filter the irrelevant language, and CNN is used to extract the significant features present in the given dataset. Thus the C-BiLSTM obtained better accuracy compared with the existing models.

Islam et al. [18] developed a practical approach to detect bullying messages online. This approach merged with NLP and ML approaches. This combination of BoW and TF-IDF achieved better accuracy than existing ML algorithms. Shekhar and Venkatesan [19] proposed a novel technique used to detect cyberbullying with the help of the Bag-of-Phonetic-Codes model. The wrong-spelled and abused words are to be removed based on the pronunciation. The proposed approach used the BoW model to extract the textual features. The Soundex algorithm focused on creating phonetic code to increase the performance of the proposed system. Experiments show that the novel technique obtained the accurate detection of cyber-bullying detection. Sharma et al. [20] described a new model to find the accurate meaning of the text. A new model is also used to reduce the spreading hurtful messages over the internet. Adopting the features of NLP with ML gives better performance. Wadhwani et al. [21] developed a new model that solves various issues, such as mismatched bullying and irrelevant content detection. This paper mainly focused on detecting the Injurious comments that trouble online users in SNS. The proposed DNN model finds the patterns of the input message and analyses the type of the messages based on the metrics such as toxic, hate, serious harmful, threat, etc. Wu and Bhandary [22] introduced the classification based on the videos that are normal or hate videos. The dataset videos were collected from online sources using the online crawler.

Murshed et al. [23] proposed the DEA_RNN model that detects cyberbullying messages in online SNS. The approach applied to 10k tweets data to analyze the cyber-bullying text. The proposed method DEA_RNN obtained best results over existing models such as Bi-LSTM, RNN, SVM and, MND, RF. The accuracy is up to 91.54% and 90.67% precision, 89.89% recall, 90.21% F1-score, and 91.84% specificity. Bai and Malempati [24] proposed the ETMA approach to detect cyberbullying messages in real-time applications like Twitter. ETMA classifies the text into bullying and non-bullying notes, and ETMA merged with TF-IDF and CNN model. ETMA gives accurate results for the classification of cyberbullying messages.

3. DATA PRE-PROCESSING WITH TF-IDF

TF-IDF is the feature extraction technique used to extract the features present in the dataset. TF-IDF is the statistical technique in NLP and information retrieval (IF). TF-IDF is used to remove the bullying words present in the dataset. TF-IDF calculates the importance of bullying terms within every review and dataset. Text vectorization is the process utilized to transform the words in the given study. Of the many vectorization techniques, TF-IDF is one of the practical approaches.

Term Frequency (TF): The raw count of the bullying words from the document or review extracts by using several steps. Here document means dataset. Thus the frequency is adjusted based on the dataset instances or the frequency of words in the dataset.

IDF: Measures a particular word equal to overall messages in the dataset, divided by reviews of specific work.

The formula is given as,

$$\mathrm{TF} - \mathrm{IDF} = \mathrm{TF} * \mathrm{IDF} \tag{1}$$

$$\mathrm{TF}(\mathrm{t}, \mathrm{d}) = \log\big(1 + \mathrm{freq}(\mathrm{t}, \mathrm{dt})\big) \tag{2}$$

$$\mathrm{IDF}(\mathrm{t}, \mathrm{D}) = \log\left(\frac{\mathrm{N}}{\mathrm{count}\ (\mathrm{d} \in \mathrm{D}: \mathrm{t} \in \mathrm{d})}\right) \tag{3}$$

Perplexity (PP): In this paper, to analyze the given text dataset, PP is used to evaluate the language model perplexity. This approach provides the inverse probability of the test set, and the no of words is normalized.

$$\mathrm{PP}(\mathrm{Words}) = \mathrm{P}(\mathrm{w}_1, \mathrm{w}_2 \ldots \ldots \mathrm{w}_\mathrm{N})^{-\frac{1}{\mathrm{N}}} = \sqrt[N]{\frac{1}{P(w_1, w_2 \ldots . w_N)}} \tag{4}$$

3.1 Unigram model

This model uses tokenization to process the raw text into small words. The input sentences break the text into terms that are called tokens. These tokens help the scenario and meaning of the sentences. In tokenization, the unigram model is used to consider every token. The probability of token X is given as the past scenario is the probability of token X. If the unigram model generates the text, this will always predict the available tokens.

$$\mathrm{P}(\mathrm{a}) = \prod_{\mathrm{x}=1}^{\mathrm{M}} \mathrm{p}(\mathrm{a}_\mathrm{x}) \tag{5}$$

$$\forall \mathrm{x}\ \mathrm{a}_\mathrm{x} \epsilon\, \mathcal{V}, \sum_{\mathrm{a}\epsilon \mathrm{v}} \mathrm{p}(\mathrm{a}) = 1 \tag{6}$$

a: Sentence
a: sub-word forming sentence
V: Vocabulary

3.2 Deep Neural Network (DNN)

This paper uses the DNN model to find the accurate cyber-bullying in online SNS and other online images. Two benchmark datasets, the Twitter dataset and an online synthetic dataset, consist of online bullying images. In DNN, various connected components are called nodes. In DNN, nodes are tiny and act as the neurons in the human brain. The neuron starts the process of the signal received by the neuron. The sign is transferred from one neuron to another based on the input received—the complex network created from this feedback. Here, the information is a text message (reviews or tweets) or bullying images, and nodes will process these data. In DNN, this is the general process for every dataset. Deep Belief Networks (DBN) is the other algorithm used in this paper for processing complex datasets such as Twitter and image data. DBN is one of the best algorithms in DNN used to process large and complex datasets.

3.3 Deep Belief Networks (DBN)

DBN is an innovative algorithm consisting of stacked RBMs. DBN follows the hierarchical representation of the input text dataset and image dataset. Bengio, et al. [25] introduced the DBN algorithm that trains a single layer simultaneously. Hinton [26] every layer processes the input text data and image data. In this, x is the visible component and ℓ is the hidden layer with joint distribution.

$$\mathrm{p}(\mathrm{x}, \mathrm{h}^1, \ldots . \mathrm{h}^\ell) = \mathrm{p}\big(\mathrm{h}^{\ell-1}, \mathrm{h}^\ell\big)\Big(\prod_{\mathrm{k}=1}^{\ell-2} \mathrm{p}\big(\mathrm{h}^\mathrm{k} \big| \mathrm{h}^{\mathrm{k}+1}\big)\mathrm{p}(\mathrm{x}|\mathrm{h}^1) \tag{7}$$

Hence, every layer of DBN creates the RBM; DBN training is the same as RBM.

The classification for the given dataset used the DBN training. The following steps are used for training 1) learning of stacked RBM in layer based manner, 2) deep tuning classifier for supervised learning An optimization issue is solved at every stage. Training the dataset $\mathrm{D} = \{\big(\mathrm{a}^{(1)}, \mathrm{b}^{(1)}\big), \ldots \ldots, (\mathrm{a}^{(|\mathrm{D}|)}, \mathrm{b}^{(|\mathrm{D}|)})\}$ with a as input and b as label, the optimization issue is solved in pre-trained phase at every layer k,

$$\min_{\theta_\mathrm{k}} \frac{1}{|\mathrm{D}|} \sum_{\mathrm{i}=1}^{\mathrm{D}} \big[-\mathrm{logp}\big(\mathrm{a}_\mathrm{k}^{(\mathrm{i})}; \theta_\mathrm{k}\big)\big] \tag{8}$$

The parameters in RBM models represent the following metrics such as $\theta_k = (W_k, b_k, c_k)$. a_k^i is the visible layer k which is input $x^{(i)}$. The layers are updated in step by step wise and solve the ℓ issues from last to first hidden layer. For better filtering the optimization issue is solved by using:

$$\min_{\phi} \frac{1}{|\mathrm{D}|} \sum_{\mathrm{i}=1}^{\mathrm{D}} [\mathcal{L}(\emptyset; \mathrm{y}^{(\mathrm{i})}, \mathrm{h}\big(\mathrm{x}^{(\mathrm{i})}\big)\big)\ ] \tag{9}$$

where $\mathcal{L}()$ represents loss function, at layer ℓ the hidden features are represented, $\emptyset$ represents the metrics of the classifier. This is written as $\mathrm{h}(\mathrm{x}^{(\mathrm{i})}) = \mathrm{h}(\mathrm{x}_\mathrm{l}^{(\mathrm{i})})$. Thus this can be used to classify the text data and image data.

4. DATASET DESCRIPTION

The proposed approach performance is analyzed using two

benchmark datasets: Twitter and online bullying Image datasets. The Twitter dataset contains 17k reviews. From this, 6135 are bullying, 7235 are non-bullying and 2630 regular messages. Several types of bullying messages are present in this dataset. The second online bully image dataset contains 1500 training and 1500 testing images gathered from various online sources. These photos belong to bully images of several users on social media.

Table 1. Twitter dataset description

Message Type	Training	Testing
Bullying	10k	7k
Non-Bullying	4k	6k
Normal Messages	2k	2k

Table 2. Image dataset

Message Type	Training	Testing
Bullying	1500	1500
Non-Bullying	500	500
Normal Messages	1k	1k

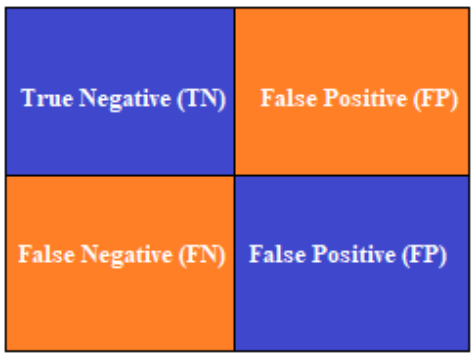
Table 1 and Table 2 show the total no of data belongs to training and testing.

5. EXPERIMENTAL EVALUTION

The implementation of proposed algorithms is done by using python language. To process the bullying datasets the system system requires 16 GB RAM with 256 SSD harddrive to solve the load balancing issues to process the large datasets. The system configuration also reduces the computation time and memory management.

6. PERFORMANCE METRICS

The performance of the model is analyzed by using the confusion matrix. This will specify the performance of classification models for given test data. This will specify the values for test data that are known. This matrix is divided into two attributes such as predicted values and original values along with an overall number of predictions.



True Negative (TN): *The estimated weight is zero, and original weight is also zero.*
True Positive (TP): *The estimated weight is one, and original weight is zero.*
False Positive (FP): *The estimated weight is one, and original weight is zero.*
False Negative (FN): *The estimated weight is zero, and original weight is one.*
Precision: The overall accurate outputs achieved by proposed model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

F1 Score: F1-score is the parameter which combines the recall and precision.

$$\text{F1} - \text{Score} = 2 \times \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{11}$$

Accuracy: Accuracy initializes the overall correctly classified data over total data records.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{12}$$

Recall: The total no of FN are to be predicted.

$$\text{Recall} = \frac{\text{TP}}{\text{No. of TP} + \text{No. of FN}} \tag{13}$$

Table 3. Performance of existing and proposed algorithms applied on twitter dataset

	SVM	CNN	TF-IDF+CNN	TF-IDF+DNN
Precision	79.23%	83.76%	90.45%	96.12%
F1-Measure	79.56%	85.9%	91.98%	96.56%
Accuracy	82.67%	86.8%	91.65%	96.12%
Recall	81.69%	86.72%	95.78%	96.67%

Table 3 compares several ML and DL algorithms to find better cyber-bullying messages and take better prevention methods. These algorithms are applied on twitter dataset for performance analysis.

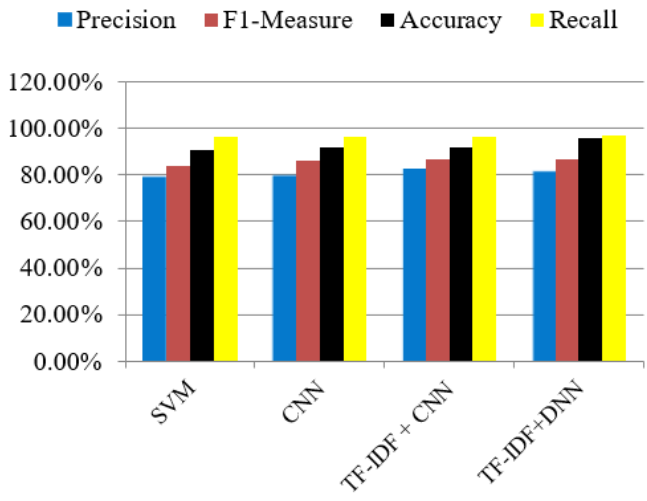


Figure 2. Comparison graph between existing and proposed algorithms

In Figure, 2 There is comparison between existing and proposed algorithms. In this comparison different algorithms

are applied on twitter dataset, so, we can improve the precision, recall, accuracy and F1-Measure.

Table 4, Figure 3 compares several ML and DL algorithms to find better cyber-bullying messages and take better prevention methods. These algorithms are applied on Images dataset for performance analysis.

Table 4. Performance of existing and proposed algorithms applied to image dataset

	SVM	CNN	TF-IDF+CNN	TF-IDF+DNN
Precision	77.12%	83.23%	90.12%	96.87%
F1-Measure	81.52%	85.67%	91.23%	97.45%
Accuracy	82.53%	86.12%	91.32%	98.56%
Recall	83.44%	87.12%	94.34%	98.23%

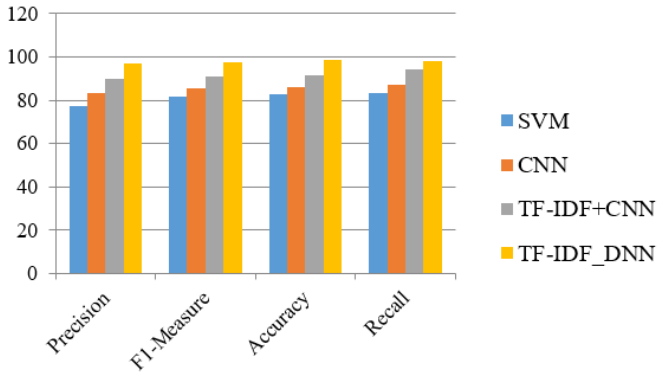


Figure 3. Performanceof previous and present algorithms

7. CONCLUSION

This paper's proposed DNN model accurately calculates bullying words and images. The proposed approach TF-IDF+DNN works better on online SNS to detect and find cyber-bullying words and Images. These types of Images and text create many issues for the victim. An efficient training model, pre-processing model, and word embedding methods make the system novel. The system proves to help analyze the cyber-bullying scores on different social media platforms so that preventive steps should take to decrease the cyber-bullying rate. The performance of the proposed approach TF-IDF+DNN achieved a precision of 96.87%, an F1-measure of 97.45%, an accuracy of 98.56%, and a recall of 98.23% and also the proposed approach focused on reducing the overall computation time.

REFERENCES

[1] Prusa, J.D., Khoshgoftaar, T.M. (2017). Improving deep neural network design with new text data representations. Journal of Big Data, 4(1). https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0065-8

[2] Ozcan, S., Homayounfard, A., Simms, C., Wasim, J. (2022). Technology Roadmapping Using Text Mining: A Foresight Study for the Retail Industry. IEEE Transactions on Engineering Management, 69(1): 228-244. https://doi.org/10.1109/TEM.2021.3068310

[3] Zhang, X., LeCun, Y. (2016). Text Understanding from Scratch. arXiv:1502.01710 [cs], https://doi.org/10.48550/arXiv.1502.01710

[4] Prusa, J.D., Khoshgoftaar, T.M., Dittman, D.J. (2015). Impact of feature selection techniques for tweet sentiment classification. Proceedings of the 28th International FLAIRS Conference, 2015: 299-304.

[5] Hani, J., Nashaat, M., Ahmed, M., Emad, Z., Amer, E., Mohammed, A. (2019). Social Media Cyber-bullying Detection using Machine Learning. International Journal of Advanced Computer Science and Applications, 10(5): https://doi.org/10.14569/ijacsa.2019.0100587

[6] Alotaibi, M., Alotaibi, B., Razaque, A. (2021). A multichannel deep learning framework for cyberbullying detection on social media. Electronics, 10(21): 2664. https://doi.org/10.3390/electronics10212664

[7] Agarwal, A, Xie, B, Vovsha, I., Rambow, O., Passonneau, R. (2011). Sentiment analysis of twitter data. In: Proceedings of the workshop on languages in social media. Association for Computational Linguistics, 2011: 30-38.

[8] Maity, K., Saha, S., Bhattacharyya, P. (2022). Emoji, Sentiment and Emotion Aided Cyberbullying Detection in Hinglish. IEEE Transactions on Computational Social Systems, pp. 1-10. https://doi.org/10.1109/tcss.2022.3183046

[9] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. Communications of the ACM. https://doi.org/10.1145/3065386

[10] Anand, M., Eswari, R. (2019). Classification of abusive comments in social media using deep learning. 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 974-977. https://doi.org/10.1109/ICCMC.2019.8819734

[11] Li, Y., Algarni, A., Albathan, M., Shen, Y., Bijaksana, M.A. (2015). Relevance Feature Discovery for Text Mining. IEEE Transactions on Knowledge and Data Engineering, 27(6): 1656-1669. https://doi.org/10.1109/TKDE.2014.2373357

[12] Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., Chang, Y. (2016). Abusive Language Detection in Online User Content. Proceedings of the 25th International Conference on World Wide Web - WWW '16, pp. 145-149. https://doi.org/10.1145/2872427.2883062

[13] Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv.org. https://doi.org/10.48550/arXiv.1408.5882

[14] Ibrohim, M.O., Setiadi, M.A., Budi, I. (2019). Identification of hate speech and abusive language on indonesian Twitter using the Word2vec, part of speech and emoji features. Proceedings of the International Conference on Advanced Information Science and System. https://doi.org/10.1145/3373477.3373495

[15] Waseem, Z., Hovy, D. (2016). Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. Proceedings of the NAACL Student Research Workshop, pp. 88-92. https://doi.org/10.18653/v1/n16-2013

[16] Vigna, F., Cimino, A., Dell'orletta, F., Petrocchi, M., Tesconi, M. (2022). Hate me, hate me not: Hate speech detection on Facebook. https://ceur-ws.org/Vol-1816/paper-09.pdf., accessed on Dec. 2, 2022.

[17] Yenala, H., Jhanwar, A., Chinnakotla, M.K., Goyal, J. (2017). Deep learning for detecting inappropriate content in text. International Journal of Data Science and Analytics, 6(4): 273-286.https://doi.org/10.1007/s41060-017-0088-4

[18] Islam, M.M., Uddin, M.A., Islam, L. Akter, A. Sharmin, S., Acharjee, U.K. (2020). Cyberbullying detection on social networks using machine learning approaches. 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast. https://doi.org/10.1109/csde50874.2020.9411601

[19] Shekhar A., Venkatesan, M. (2018). A Bag-of-Phonetic-Codes Modelfor Cyber-Bullying Detection in Twitter. 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, pp. 1-7. https://doi.org/10.1109/ICCTCT.2018.8550938

[20] Sharma, R., Ramakrishnan, A., Pendse, P., Chimurkar, Talele, K.T. (2021). Cyber-bullying detection via text mining and machine learning. 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, pp. 1-6. https://doi.org/10.1109/ICCCNT51525.2021.9579625

[21] Wadhwani, A., Jain, P., Sahu, S. (2021). Injurious Comment Detection and Removal utilizing Neural Network. 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), Noida, pp.165-168.https://doi.org/10.1109/ICIPTM52218.2021.9388331

[22] Wu, C.S., Bhandary, U. (2020). Detection of Hate Speech in Videos Using Machine Learning. International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, pp. 585-590. https://doi.org/10.1109/CSCI51800.2020.00104

[23] Murshed, A.H., Abawajy, J., Mallappa, S., Saif, M.A.N., Al-Ariki, H.D.E. (2022). DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform. IEEE Access, 10: 25857-25871.https://doi.org/10.1109/ACCESS.2022.3153675

[24] BaiZ.S., Malempati, S. (2022). An Enhanced Text Mining Approach using Ensemble Algorithm for Detecting Cyber Bullying. International Journal of Engineering Trends and Technology, 70(9): 393-399.

[25] Bengio, Y. (2009). Learning deep architectures for AI. Foundations and trendin Machine Learning, 2(1): 1–127. http://dx.doi.org/10.1561/2200000006

[26] Hinton, G.E., Osindero, S., Teh, Y.W. (2016). A fast learning algorithm for deep belief nets. Neural Computation, 18(7): 1527-54.