International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 411-418 www.ijaem.net ISSN: 2395-5252

Analysis of cyberbullying using machine learning algorithms

Kalyani.Chapa[1], A.Gowri Shankar[2], B.Surya Vara Prasad[3] , CH.Vinay Kiran[4], D.Phani Varma[5]

Department Of Computer Science and Engineering
RAGHU INSTITUTE OF TECHNOLOGY, Visakhapatnam, Andhra Pradesh, India.

Submitted: 01-06-2022 Revised: 05-06-2022 Accepted: 08-06-2022

ABSTRACT
The use of social media has fully grown exponentially over time with the expansion of the web and has become the foremost platform within the twenty first century. However, the improvement of social property usually creates negative impacts on society that contribute to one or two of dangerous phenomena like on-line abuse, harassment cyberbullying, crime and on- line troll. Cyberbullying often results in serious mental and physical distress, notably for ladies and kids, and even typically force them to aim suicide. On-line harassment attracts attention thanks to its sturdy negative social impact. Several incidents have recently occurred worldwide thanks to on-line harassment, like sharing non-public chats, rumours, and sexual remarks. Therefore, the identification of bullying text or message on social media has gained a growing quantity of attention among researchers. The aim of this analysis is to style and develop an efficient technique to discover on-line abusive and bullying messages by merging linguistic communication process and machine learning. There is distinct feature ie.. term frequency-inverse text frequency (TF- IDF), area unit accustomed analyse the accuracy level of 4 distinct machine learningalgorithms.

Index Terms—Cyberbullying, Machine learning, Naturallan- guage processing, Social media.

I. INTRODUCTION

Social media may be a platform that enables individuals to post any- factor like photos, videos, documents extensively and act with society . individuals connect with social media victimisation their computers or smartphones. the foremost fashionable social media includes Facebook, Twitter, Instagram, TikTok then on. Nowadays, social media is concerned in several sectors like education , business , and additionally for the noble cause Social media is additionally enhancing the world's economy through making several new job opportunities .

Although social media incorporates a heap of benefits, it additionally has some drawbacks. Using this media, malevolent users conduct unscrupulous and fallacious acts to harm others feelings and injury their name.

Recently, cyberbullying has been one amongst the most important social media problems. Cyberbullying or cyber-harassment refers to AN electronic methodology of bullying or harassment. Because the digital realm has grownup and technology has progressed, cyberbullying has become comparatively common, significantly amongst adolescents.

Approximately five hundredth of the teenagers in America expertise cyberbullying. This bullying incorporates a physical and mental impact on the victim . The victims select unsafe acts like suicide as a result of the trauma of cyberbullying that is difficult to be endured. Thus, the identification and hindrance of cyberbullying is very important to safeguard teenagers.

In this context, we advise a cyberbullying observation model supported machine learning that may detect whether or not a text relates to cyberbullying or not. we've investigated many machine learning algorithms, as well as Naive Bayes, Logistic regression, Decision Tree, and Random Forest within the projected cyberbullying detection model. we have a tendency to conduct experiments with one dataset from twitter posts. For performance analysis, we have a tendency to use TF-IDF.

The remainder of the paper is organized as follows literature survey, methodology, experiment and its results and finally concludes the paper and accentuate the future work.

II. LITERATURE SURVEY

In 2019, John Loloish et al. conferred a supervised learning approach to notice cyberbullying. As a section of the preprocessing step, information is cleansed by removing the noise and redundant text. this can be performed victimization tokenization, lowering text, stop words together with encoding cleaning and word correction. The second step is that the feature extraction step that is completed victimization TF military force and sentiment analysis technique as well as NGrams for considering completely different combos of the words like 2- Gram, 3-Gram, and 4- Gram. The cyberbullying dataset from Kaggle is split into ratios (0.8, 0.2)for train and take a look at. SVM and Neural networks square measure used as classifiers that run on a unique n-gram language model. Accuracy, recall and preciseness, and f-score square measure the performance measures. it's found that Neural Network performed higher than the SVM classifier. Neural Network achieved a median f-score of ninety one.9% and SVM achieveda median f-score of eighty nine.8%.

In 2018, Monirah Abdullah Al-Ajlanet and Mourad Ykhlef projected a unique formula CNN-CB that relies on a convolutional neural organization and adapts the concept of word embedding. The design includes four layers - Embedding, Convolution Layer, goop Pooling Layer, and Dense Layer. the primary layer, word embedding, creates a vector house of vocabulary that is that the input to the following layer, the convolutional layer,which compresses the input vector while not losing significantfeatures. The third layer, the goop pooling layer, takes the output of the second layer as its input and finds the utmost price ofthe chosen region to save lots of simply important highlights. The last layer, the Dense layer, will the classification. This gave a preciseness of ninety fifth.

In 2018, Monirah A. Al-Ajlan et al. projected optimized Twitter cyberbullying detection supported deep learning (OCDD) that doesn't extract options from tweets instead, it represents a tweet as a group of word vectors that square measure fed to a convolutional neural network (CNN)for classification.Hence the feature extraction and choice phases square measure eliminated during this approach. To represent the linguistics between words, word embedding is employed and is generated victimization (GloVe) technique. CNN uses loads of parameters and to optimize these values, a metaheuristic optimisation formula is employed to search out best or near-optimal values that may be used for classification. CNN showed nice results.

In 2017, Yee Jang Foong and Mourad Oussalah bestowed an automatic cyberbullying detection that uses language process techniques, text mining, and machine learning. For dataset facebook, a social media platform wherever users will anonymously or publically raise queries and think about a sample of a user's profile is employed. As a part of preprocessing he started to remove the different characters,emojis, incorrect wordings and additionally lexicons square measure replaced with equivalent matter expressions. a mixture of options has been used which has TF-IDF, uncommon capitalization count, LIWC, and Dependency computer programme. the information set is split into a seventieth coaching set and half-hour testing set. SVM was used as a classifier that was trained with a linear kernel on the coaching information.

In 2016, X. Zhang et al. planned a unique approach supported a pronunciation-based convolutional neural network (PCNN). Word-to-Pronunciation conversionis done to group A set of words spelled incorrectly, that have an equivalent that means and pronunciation, along with the corrected word. 2 separate CNN is employed to determine a baseline. For the primary baseline feature set, word-embedding supported Google's word-vector was used. For the creation of the feature set of the second baseline, CNN Random, willy-nilly generated vectorswere used.The phone codes were willy-nilly introduced into vectors for the feature set for PCNN. To handle category imbalance 3 techniques were implemented- threshold moving, value perform modify, and a hybrid resolution, out of that value perform adjusting is handiest.

In 2016, Michele Di Capua et al. bestowed Associate in Nursing unsupervised approach to observe cyberbullying employing a style model galvanized by Growing graded SOMs. Firstly, options ar divided into four groups: syntactical options, linguistics options, Sentiment options, Social options.GrowingHierarchical SelfOrganizing Map (GHSOM) network rule, that is like minded for an outsized assortment of documents that has got to be classified, is used. It uses a hierarchical data structure of multiple layers, wherever every layer consists of a range of freelance SOMs. one Kyrgyzstani monetary unit is used atthe rootlayer. for each unit, throughout this map, a Kyrgyzstani monetary unit can be extra to the following layer of the hierarchy. GHSOM

Network is trained and tested regarding a K-folded dataset, applying a K-fold partitioning of knowledge.

In 2014, Sourabh Parime Associate in Nursingd Vaibhav Suri bestowed an approach of mistreatment data processing and machine learning techniques to observe cyberbullying.Text mining is performed on unstructured knowledge mistreatment machine learning techniques to extract data from the text which has multiple stages like document cluster, knowledge pre-processing, attribute generation {for that|that} Associate in Nursing in-built classifier is employed to get labels from the options fed into it and occurrences ar counted and a weight is allotted to every label and digressive attributes ar removed which helps to estimate the character of the comments. Sentiment analysis is employed for determinative the tone of the given text. 2 categories of knowledge ar thought of one with positive emotions and therefore the different with negative emotions. These ar hold on into a vector and accustomed train a supervised learning rule SVM.

BUYLLYING DETECTION MODEL

In this section, we describe the cyberbullying detection framework which consists details of operations on dataset and methodologies.

A) Operations on Dataset:

We have collected twitter comments datset from kaggle.com .After data cleaning we performed data preprocessing and resampling on dataset.

1) **Data Preprocessing** :

In preprocessing process we appilied tokenization, removed digits and symbols from statements and appilied stemming. After all these processes we TF- IDF for extracting the features from the dataset and later we can use those feature in ml algorithms inorder to compare the accuracy between them.

TF-IDF: This is one of the features that we consider for our model. TF-IDF (Term Frequency-Inverse Document Frequency) is a statistical measure that can evaluate how relevant a word is to a document in a collection of documents. In bag of words, every word is given equal importance while in TF-IDFthe words that occur more frequently should be given more importance as they are more useful for classification.

2) Data Resampling:

As a data was skewed ,resampling had to be performed on the training data, Firstly the data was spilit into training and test in 80:20 ratio and resampling was performed on the training data. As we had ample data to work with ,we used oversampling of the minority class. This means that if the majority class had 1500 examples and the minority class had 122, this strategy would oversampling the minority class so that it has 1500 examples.

After resampling, the training data had 9750 Bullying and 9750 Non-Bullying instances.

B)Description of ML Algorithms

In this section, we discussed the basic mechanisms of sev- eral machine learning algorithms. We presented Decision Tree,Naive Bayes, Random Forest and logistic regression in each subsection.

1) Decision Tree : The decision tree classifier can be used inboth classification and regression . It can help represent thedecision as well as make a decision. The decision tree is a tree- like structure where each internal node represents a condition, and each leaf node represents a decision.

A classification tree where the target falls. A regression tree yields the predicted value for addressed input.

The classifier was implemented using sklearn.tree package.

2) Random Forest: Random Forest classifier is consists multiple decision tree classifiers . Each tree gives a class prediction individually. The maximum number of thepredictedclass is our final result. This classifier is a supervised learning model which provides accurate result because several decision trees are merged to make the outcome. Instead of relying on one decision tree, therandom forest takes the prediction from each generated tree and based on the majority votes of predictions, and itdecides the final output.

For example, ifwe have two classes namely A and B and the most of the decision tree predict the class label B of any instance, thenRF will decides the class label B as follows:

f(x)=majority of vote of all trees as b

The classifier was implemented using sklearn.ensemble package.

3) Naive Bayes: Naive Bayes is an efficient machine learn- ing algorithm based on Bayes theorem. The algorithm predicts depending on the probability of an object. The binary and multi-class classification problems can be quickly solved using this technique.

The classifier was implemented using sklearn.naive bayes package.

4) Logistic Regression: Regression analysis is a predictive modelling technique that analyzes the relation between the target or dependent variable and independent variable in a dataset. Regression analysis techniques get used when the target and independent variables show a linear or non-linear relationship between each other, and the target variable contains continuous values. Regression analysis involves determining the best fit line, which is a line that passes through all the data points in such a way that distance of the line from each data point is minimized.The classifier was implemented using sklearn.linear model package.

In figure 1 ,it describes the proposed framework for bullying detection.

III. EXPERIMENT AND RESULTS

For our analysis we particularly used , Decision Tree classifier (DT), Naive Bayes classifier(NB), logistic regression and Random Forest classifier (RF) and in our work its clear that naive bayes classifier showed poor accuracy rate when compared with other algorithms and random forest classifier gave the best results in term of every metrics.

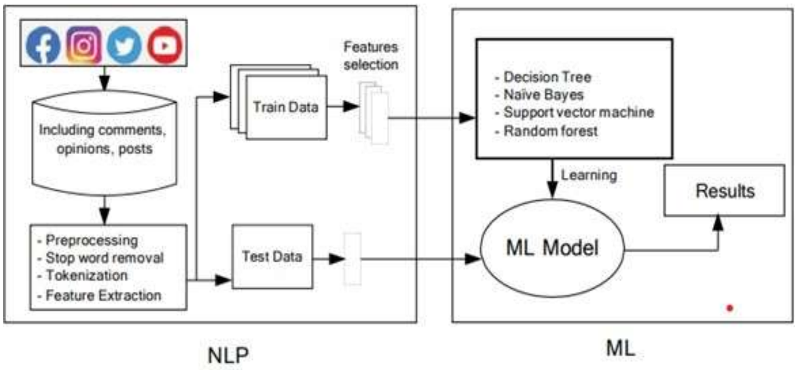

Fig 1: Proposed framework for bully detection

It wasn't surprising to see the Random Forest classifier performing the best. The Decision Tree classifier performed better than Naive Bayes classifier and Logistic Regression. The Random Forest Classifier came out on top in all the performance metrics, which was expected as it is an extension of the Decision Tree classifier, averaging out results of multiple recursions of the same.The bullying detection algorithms are are enforced victimization python machine learning packages. The performances are analyzed with relevancy the subsquent metrics.

- The classification results are listed in the confusionmatrix, additionally referred to as the contingency table. verity Positive higher left corner is that the range of people that were listed as true positive, whereas those were true. The False-positive lower right cell reflects the quantity of samples that, tho' false, were labeled as false negative. False- negative shows the quantity of people, whereas these were false, being counted as true. False-positive reflects, as these were true, the quantity of people that were listed as true.
- Based on confusion matrix, we will calculate metrics such as accuracy , recall and f-measure ROC curve and format of confusion matrix is shown in table 1.

	Condition Positive	Condition Negative
Predicted Condition Positive	True Positive	False Negative
Predicted Condition Negative	False Positive	True Negative

Table 1

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

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

$$Precision = \frac{TP}{TP + FP}$$

ROC Area, which denotes the area under thecurve formed by plotting TP rate. where,
TP = No. of True Positives
TN = No. of True Negatives FP = No. of False Positives FN= No. of False Negatives
Each individual algorithm which we have taken has showing the different precision , recall and f-measure results which is varying from one from another .

For considered dataset, Naïve bayes classifier resulted accuracy score is around 62% and its accuracy and ROC curves are represented in figure 2.

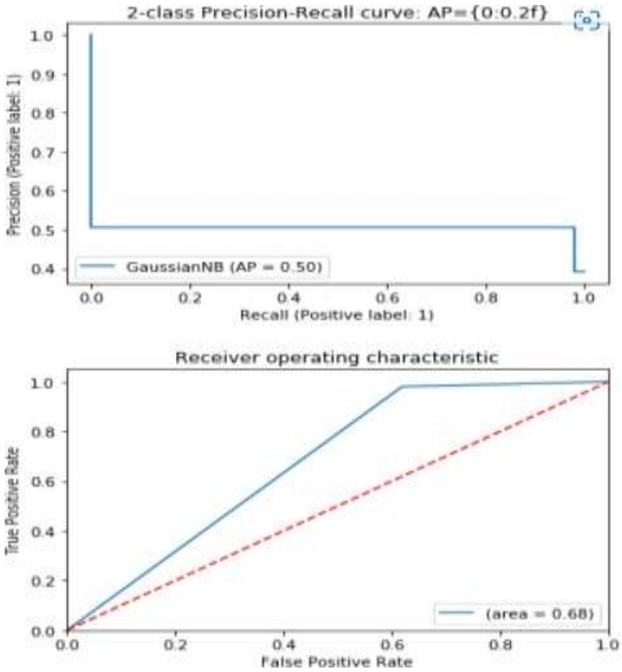

Figure 2

For considered dataset, logistic regression classifier resultedaccuracy score is around 80% and its accuracy and ROC curves are represented in figure 3.

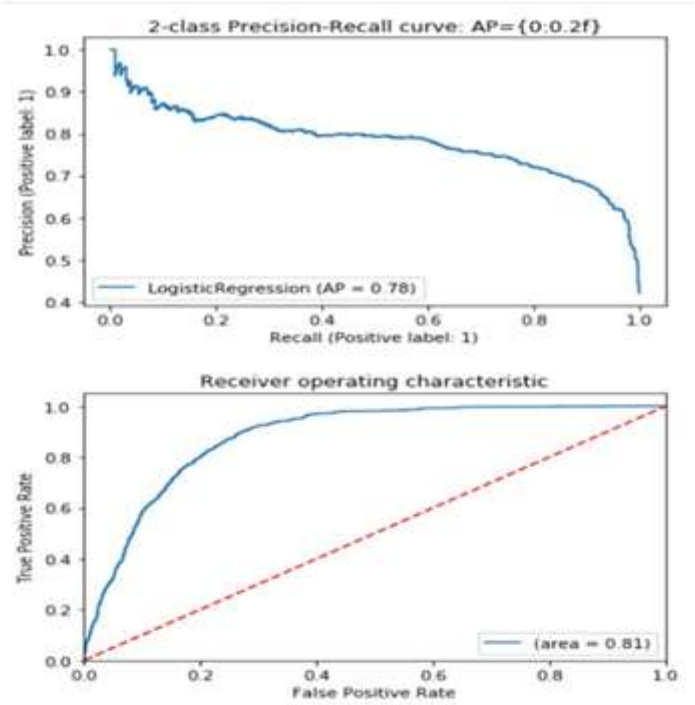

Figure 3

For considered dataset, Decision tree classifier resulted accuracy score is around 85% and its accuracy and ROC curves are represented in figure 4.

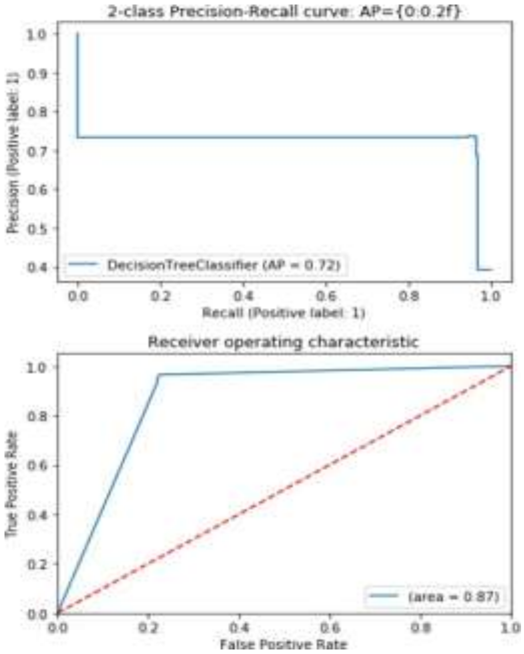

Figure 4

For considered dataset, Random forest classifier resulted accuracy score is around 92% and its accuracy and ROC curves are represented in figure 5.

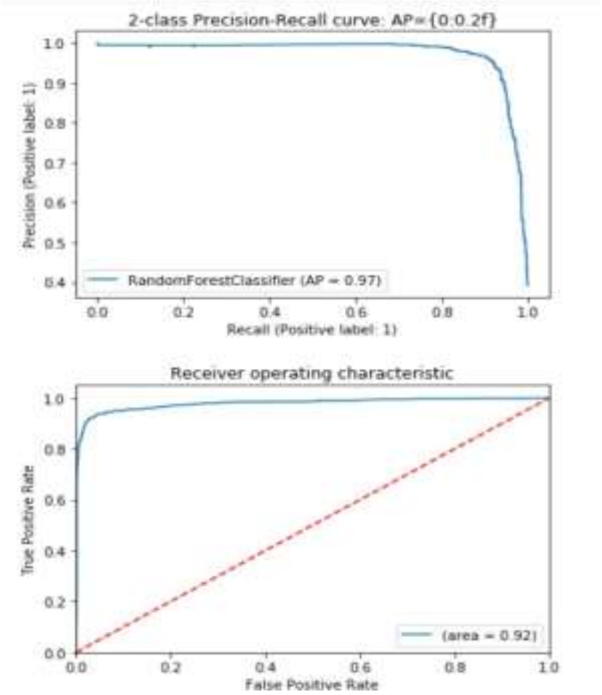

Figure 5

We represent the final result of each algorithm by using theprecision, F-measure, Recall and roc area in figure 6&7.

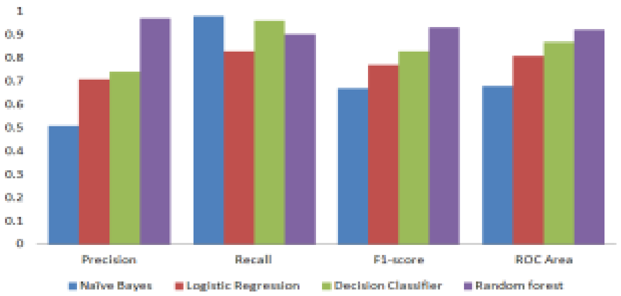

Figure 6

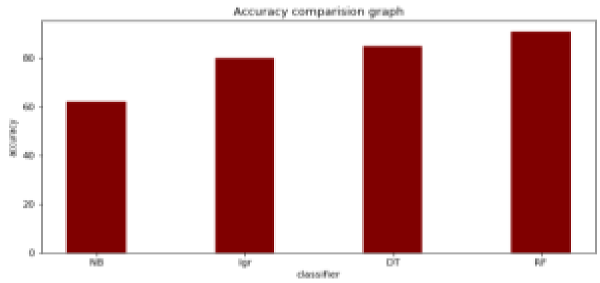

Figure 7

In our project we tried to implement random forest classifier in our framework inorder to predict the whether the input statements are Bullying or Non-Bullying.

IV. CONCLUSION

In specific , cyberbullying has become additional common thing in the internet and has begun to lift problems with the rising prevalence of social media sites and exaggeratedsocial media use by teenagers. There must style automatic cyberbullying detection technique to avoid unhealthy consequences of cyber harassment which happening through the social media. Considering the significance of cyberbullying detection, during this study , we tend to investigate the automatic of posts on social media associated with cyberbullying by considering a method from natural language processing technique(TF-IDF and Lemmatization). Four machine learning algorithms area unit accustomed determine bullying text and random forest classifier with TF-IDF provides the best accuracy to finding the results.

V. FUTURE SCOPE

In the future, we tend to area unit going to style a framework for automatic detection and classification and also we can further extend this analysis of cyberbullying project to classify the different language comments, tweets, messages are whether Bullying or Nonbullying statements. Probably , Deep Learning algorithms can give good results for this work. This could involve working in different social media official servers to identify the people who are publishing the bullying statements in their sites.

REFRENCES

[1]. N. Selwyn, "Social media in educational activity," The Galilean satellite world of learning, vol. 1, no. 3, pp. 1– 10, 2012.

[2]. H. Karjaluoto, P. Ulkuniemi, H. Keina¨nen, and O. Kuivalainen, "An- tecedents of social media b2b use in industrial selling context: customers' read," Journal of Business & Industrial selling, 2015.

[3]. W. Akram and R. Kumar, "A study on positive and negative effects of social media on society," International Journal of laptop Sciences and Engineering, vol. 5, no. 10, pp. 351–354, 2017.

[4]. D. Tapscott et al., The digital economy. McGraw-Hill Education,, 2015.

[5]. S. Bastiaensens, H. Vandebosch, K. Poels, K. Van Cleemput, A. Desmet, and I. Delaware Bourdeaudhuij, "Cyberbullying on social network sites. Associate in Nursing experimental study into bystanders' behavioral intentions to assist the victim or reinforce the bully," Computers in Human Behavior, vol. 31, pp. 259–271, 2018.

[6]. K. Reynolds, A. Kontostathis, and L. Edwards, "Using machine learning to discover cyberbullying," in 2011 tenth International Conference on Machine learning and applications and workshops, vol. 2. IEEE, 2019, pp. 241–244.

[7]. V. Balakrishnan, S. Khan, and H. R. Arabnia, "Improving cyberbully- ing detection victimisation twitter users' psychological options and machine learning," Computers & Security, vol. 90, p. 101710, 2020.

[8]. S. Agrawal and A. Awekar, "Deep learning for sleuthing cyberbullying across multiple social media platforms," in European Conference on data Retrieval. Springer, 2018, pp. 141–153.

[9]. P. Badjatiya, S. Gupta, M. Gupta, "Deep learning for hate speech detection in tweets," in, 2017, pp. 759–760.

[10]. M. A. Al-Ajlan and M. Ykhlef, "Deep learning algorithmic rule for cyberbullying detection," International Journal of Advanced applied science and Applications, vol. 9, no. 9, 2018.

[11]. L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, "Xbully: Cyberbullying detection among a multi-modal context," in Proceedings of the Twelfth ACM International Conference on internet Search and data processing, 2019, pp. 339–347.

[12]. K. Wang, Q. Xiong, C. Wu, M. Gao, "Multi-modal cy- berbullying detection on social networks," in 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2017, pp. 1–8.

[13]. T. A. Buan and R. Rama, "Automated cyberbullying detection in social media victimisation Associate in Nursing svm activated stacked convolution lstm network," in Proceedings of the 2020 the fourth International Conference on work out and information Analysis, 2020, pp. 170–174.

[14]. Slonje, R. and P.K. Smith, 2008. Cyberbullying: Another main type of bullying? Scand. J. Psychol., 49: 147-154. PMID: 18352984.