DETECTION OF CYBER-BULLYING IN SOCIAL-MEDIA USING CLASSIFICATION ALGORITHMS OF MACHINE LEARNING

Chandradeep Bhatt [1], Parul Goyal [2], Ghanshyam Prasad Dubey [3], Shalini Singh [4] and Vishal Kumar [5]

[1] Computer Science and Engineering, Graphic Era Hill University;
Adjunct Professor, Graphic Era Deemed to be University, Dehradun, India.
[2] Computer Science & Engineering, M. M. Engineering College,
Maharishi Markandeshwar Deemed to be University, Mullana, Ambala, Haryana, India.
[3] Associate Professor, Department of CSE, Amity School of Engineering & Technology,
Amity University Madhya Pradesh, Gwalior, India.
[4] Computer Science, ABES Engineering College, Ghaziabad, Up, India.
[5] Computer Science and Engineering, Roorkee Institute of Technology, Roorkee, India.
Email: [1]bhattchandradeep@gmail.com, [2]parul.goyal@mmumullana.org, [3]gpdubey@gwa.amity.edu,
[4]shalinisingh@abes.ac.in, [5]me.vishaldhiman221@gmail.com

DOI: 10.5281/zenodo.11203380

Abstract

Covid-19 has made everything available online. Consequently, parents are compelled to get a smartphone for their children. However, purchasing a smartphone implies that their children's social media troubles and other concerns can also worsen. The usage of social networks has skyrocketed since COVID-19. Thus, closely correlated with cyberbullying. The well-known Metaverse platforms are TikTok, Facebook, Instagram, Twitter, and WhatsApp. Every day, cyberbullying, which includes body shaming, appearance, behaviour, racism, sexual harassment, and other forms of online bullying, affects thousands of social media users. This project is intended only for young people who are not familiar with social media. The current research employs machine learning techniques to automatically identify the bullies' harsh word usage to stop this harassment. Following detection, if a child was bullied, it could use the child's dataset to determine the bullying and, in the unlikely event that the child was the victim of bullying, it would be determined from the dataset. If any abusive language is found, the developers will be alerted, the appropriate action will be taken, and an email alert will be sent if any abusive text is found in the conversation. Thus, the suggested approach is effective in identifying cyberbullying on social media and can be turned into a web application that requires users to link their social media profiles. Both parents who want to know what is happening with their children and the children themselves will find this project helpful. Since a lot of children develop social media profiles to stay up to date with their schooling, this idea will undoubtedly be helpful. Additionally, as everything is now done online, we can use this effort to stop bullying people online. My goal is to identify cyberbully words and their types and to send out an SMS alert if any are found. Consequently, we selected six machine learning techniques for classification and used count Vectorizer and time frequency - inverse document frequency to extract features as a bag of words. I employed Decision Trees, Naïve Bayes, Random Forest, Logistic Regression, K Nearest Neighbour, and Support Vector Classifiers. Using temporal frequency - inverse document frequency for feature extraction, support vector machines yield 91.98%.

Keywords: Cyberbullying, Machine Learning, Classification Algorithm, Social-Media.

I. INTRODUCTION

Social media is vital part of whole world’s life. It's useful for knowing the effects that are passing around the world. [1] After the coronavirus epidemic, there was a significant rise in the number of social media druggies. Nearly 10.5 increase was noted in the number of active social media accounts from 2019 to 2020. Instagram denoted a whopping 70 increase in observers. The increase in use of social media also rises concern in the adding of online cyber bullying. Even though social media has more benefits but it has few disadvantages. [2] By using social media some users’ acts are used to hurt someone’s feelings and their reputations in the society. In recent times cyberbullying is one of the main social-media issues in the world. [5] The Cyberbullying has physically and mentally impact on victims. The victims choose acts like suicide because of trauma of cyberbullying.

Social media gives us for communicator platform occasion and they further enlarge the susceptibility younger generations to forbidding circumstances online. Cyberbullying on a social media network is a world occurrence because of its extensive of active users [3]. The drift appears that the cyber bullying on social website is enlarging promptly day-to-day life. Recent studies shows that cyberbullying represent a enlarging problem among adolescent. Fortunate precaution depends on the reasonable spotting of potentially deleterious messages & the information overburden on the network need intelligent systems to identify potential risks adverbinevitably. In this paper, we focused on automatic cyberbullying detection.

Gather information about bullying language. Gathering bullying-related words from Twitter using the kaggle.com dataset [4] and processing the data using machine learning and natural language processing. Induce different machine learning algorithm model. In my project for cyberbullying detection,we have used classification algorithms which is under supervised learning in machine learning. The algorithms utilized for cyberbullying detection are Support vector machine, Random forest, Logistic regression, Decision tree classifier, Naïve Bayes and K nearest neighbour algorithms for the proposed system.[2] For feature extraction I used two extractions for cyberbullying and they are BOW (Bag of Words) and TF-IDF (Time FrequencyInverse Document Frequency).Support Vector machine using TF-IDF extraction gives better accuracy than BOW than any other machine learning algorithms used in this project. Get the messages from WhatsApp user account and preprocess it. Apply induced models on the text from WhatsApp and result whether the text is cyberbullying or non -cyberbullying word.

II. LITERATURE SURVEY

Kazi Saeed Alam et al. 2021, author define cyberbullying took several forms such as harassment, racism, and hateful comment. They took text data from twitter for detecting cyberbully word and for best performance. The data is classified into offensive and no offensive. In their work logistic regression and ensemble bagging model performs for detect cyberbully word. They gave best performance 96% with combination of TF-IDF and k fold validation[1]. Manowarul Islam etal.2020, according to the author, cyber-harassment is an online bully's computerised method. The purpose of research is to merge the natural processing language and machine learning algorithms. They took two set of datasets from Facebook and twitter. The results tell us TF-IDF gives better accuracy using support vector machine algorithm [2]. Roy Goldschmidt, and Yuval Elovici done research with social networks which describes statistics of online social networks, types, threats, improvementation of security, and solutions in protect social-media users [3]. Pradeep Kumar Roy, Asis Kumar Tripathy, Tapan Kumar Dasand Xiao-Zhi Gao article deals difficulty of hate-speech in tweets. Their hate-speech focused on gender, race, religion, and differently abled people. The authors developed automated system by Deep Convolutional Neural Network. They archive the precision as 0.97, recall as 0.88 and F1-score as 0.92 [4]. Kelly Reynolds, April Kontostathis and Lynne Edwards collected data from Formspring.me website. By using labelled dataset with machine learning techniques using Weka tool. The decision tree gives 78.5% accuracy [5]. Aaminah Ali and Adeel

M. Syed their study focus on accent before researchers and proposed a model to find cyberbully word and element of sarcasm. The took two datasets, one from formspring and another from tweets. The accuracy for Random Forest 91% Naïve Bayes 87% SVM 92% Logistic Regression 92% Ensemble 92% and this shows us SVM gives best accuracy [6]. Reem Bayari and Ameur Bensefia collected dataset from twitter using Arabic and Latin languages They focused on neural networks and machine learning algorithms, they used features ngram mainly bigram and trigram. They got best accuracy using CNN. SVM is the most common classifier for those languages [7]. Bandeh Ali TalpurID1 and Declan O'Sullivan collected data as tweets, and the categories are i)sexual, ii) racial, iii) appearancorelated, iv) intelligence, and v) political.

The authors applied Embedding, Sentiment, and Lexicon features with PMI-semantic orientation. Feature Extraction with Naïve Bayes, KNN, Decision Tree, Random Forest, and Support Vector Machine algorithms. They compared proposed and baseline features with algorithms. Random forest gives better accuracy 91.15% than others [8]. PatxiGal'an-Garc'a et al. (2016) applied supervised machine learning to a real-world instance of cyberbullying to identify troll profiles in the Twitter social network. The authors built a system to spot false profiles on twitter. The algorithms are KNN, Random Forest, Decision tree and Sequential Minimal Optimization. The SMO with poly-kernel gives better accuracy than others [9]. Tarek Kanan et al…2020 The author collected data from Facebook and tweets and Arabic language. They use different natural processing language tools. The results show random forest yields better FI-measure for Arabic and SVM gives better F1 measure for Facebook and tweets [10].

Sinchana C et al. 2020, used machine learning and NLP to detect the cyber-bully word by matching text. They have done detection in both image and textual. The algorithms are SVM, KNN, Naïve Bayes, Decision tree and neural networks. They detect cyberbully as three types, are none, sexism, and racism. Accuracy(94.37%), Precision(0.66), Recall(0.35), and F1-Score(0.45) [11]. Saloni Wade et al…2020 The author used real world's data as dataset from twitter. They classified as positive, negative, and neutral tweets. They used DEEP NEURAL NETWORK MODELS are Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). They detected cyberbully word and non-cyberbully word [12]. Harika Gummadavelly et al. 2021, goal is to use Naive Bayes and build a classification model with better accuracy and identifying cyber bully word. They have done web pages for chatting between clients. Naive Bayes' gives better accuracy as 97.11 [13]. John Hani et al…2019 The authors done project for detection and prevention of cyberbullying. They took dataset from Kaggle. They used Weka tools and feature extraction are TF-IDF. It shows Neural Network performs better accuracy of 92.8% than others [14]. Saloni Kargutkar and Vidya Chitre,the took dataset as tweets from twitter. They used DEEP NEURAL NETWORK MODELS are Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). They identified racism, non-bullying, and bullying [15].

Wahyu Adi Prabowo et al. 2019, done Pre-processing, Term Frequency, and Inverse Document Frequency as Feature extraction. The machine learning algorithm as Support Vector Machine and found Confusion Matrix, Application, Sentiment Analysis. They identified texts in the application as cyberbullying and non-cyberbullying [16].Omer Atoum2020, took dataset from twitter and feature extraction as n-gram and stemming and used two machine learning algorithms are SVM and Naïve Bayes. The SVM(91.64) algorithm gives better accuracy that naïve Bayes [17]. Karan Shah et al2022, authors collected real time tweets from twitter. They used machine learning algorithms and NLP for better performance. Their goal is to find the cyberbully text which from group chat. Count vectorizer and TF-IDF is used as feature extraction. The algorithms are SVM, Logistic regression, Bagging classifier, Decision tree, Ada boos, Random Forest, and Naive Byes. Decision tree using Tf-Idf gives better accuracy and fi score than others [18]. Nivethitha R et al. have collected dataset and pre-processed it and they have done data portioning, feature extraction and used machine learning models for detection of cyberbullying word. [19] Chahat Raj et al. 2021, collected real world cyberbullying dataset. They used feature extraction as TF-IDF and machine learning techniques Globle vectors and neural networks and BERT model. The neural networks give the better performance than others[20].

Proposed System Framework

This model can not only provide detailed analysis of past data, but can also identify abusive content from real time data. It can classify the chat dataset into 3 class's namely offensive language, hate speech and normal text classes. Apart from analysing the data, this model automatically sends an alert notification to the registered email id if any abusive content is found. The bully's number is also known through the email alert sent. If the text is normal, no alert notification will be sent. The advantage of proposed system is: Model can provide detailed sentiment analysis of real time data. Model has an accuracy of 90.98%. Model sends a SMS along with the bully's number, date, time, and its type if abusive content is found.The flowchart represents the cyber bullying detection in social media using classification algorithms that is reflected in figure 1.

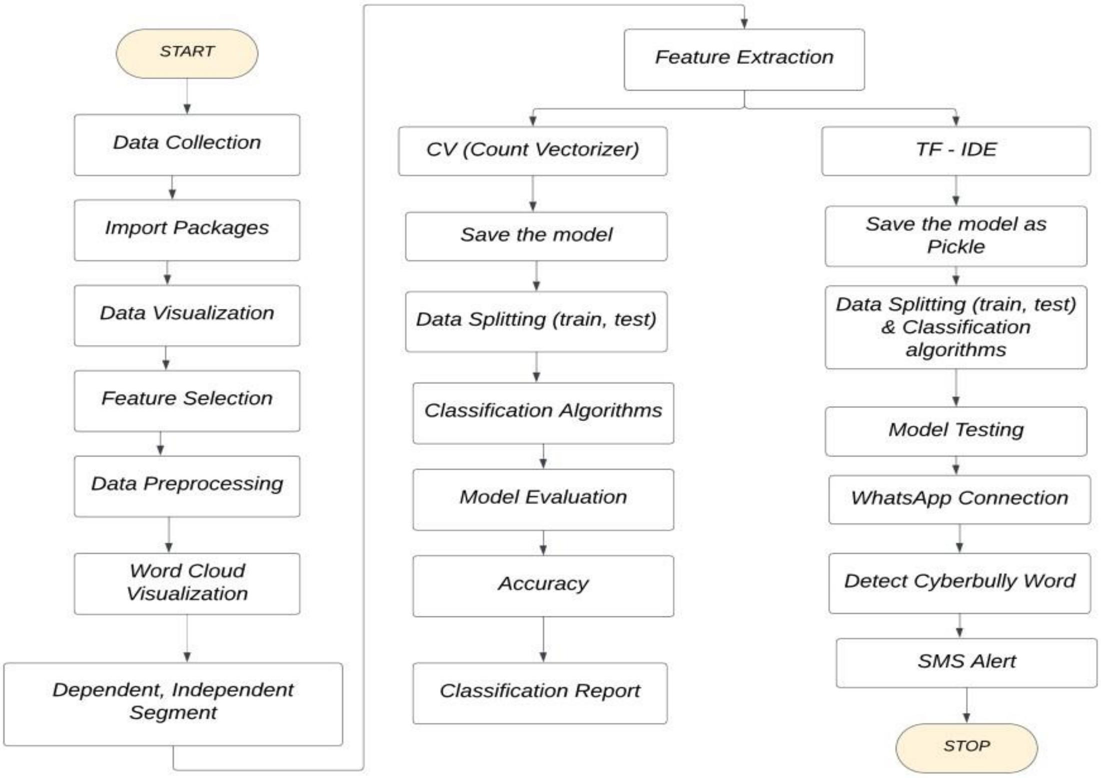


Figure 1: Flow chart

We developed my project with the help of programming language called python. Then after we will search and download the dataset that we need to classify from Kaggle website. After downloading the dataset, we will pre-process the data. And the next section is transfer to cv, Tf-Idf. Then we will generate codes for the machine learning algorithms are Naive Bayes, KNN, Logistic regression, Decision Tree, Random Forest, SVM using python.

So, we are using python as backend. The real-world text or post contain a greater number of unnecessary texts, symbols, and links. It not requires for cyber-bullying detection. So, I will remove all links and symbols. After that, the work is to apply the supervised machine learning algorithms for the spotted of bullying text.

In this section, the work is to remove unnecessary characters like symbols, emojis, numbers, links etc. And the next section is featuring extraction for classification, the two important features of the classification text is prepared: Bag of words: The machine learning algorithms are not work immediately with texts. So, we must convert them into some other form like numbers or vectors before applying machine learning algorithm to them.

In this way the data is converted by Bag-of-Words (BOW) so that it can be ready to use in next round. TF-IDF: One of the most important features to be considered is this. TF-IDF (Term Frequency-Inverse Document Frequency) is a quantitative measure. The words in text set are reconstruct into text vectorization task. The architecture for detecting cyberbullying in social media using classification algorithms is reflected in figure 2.

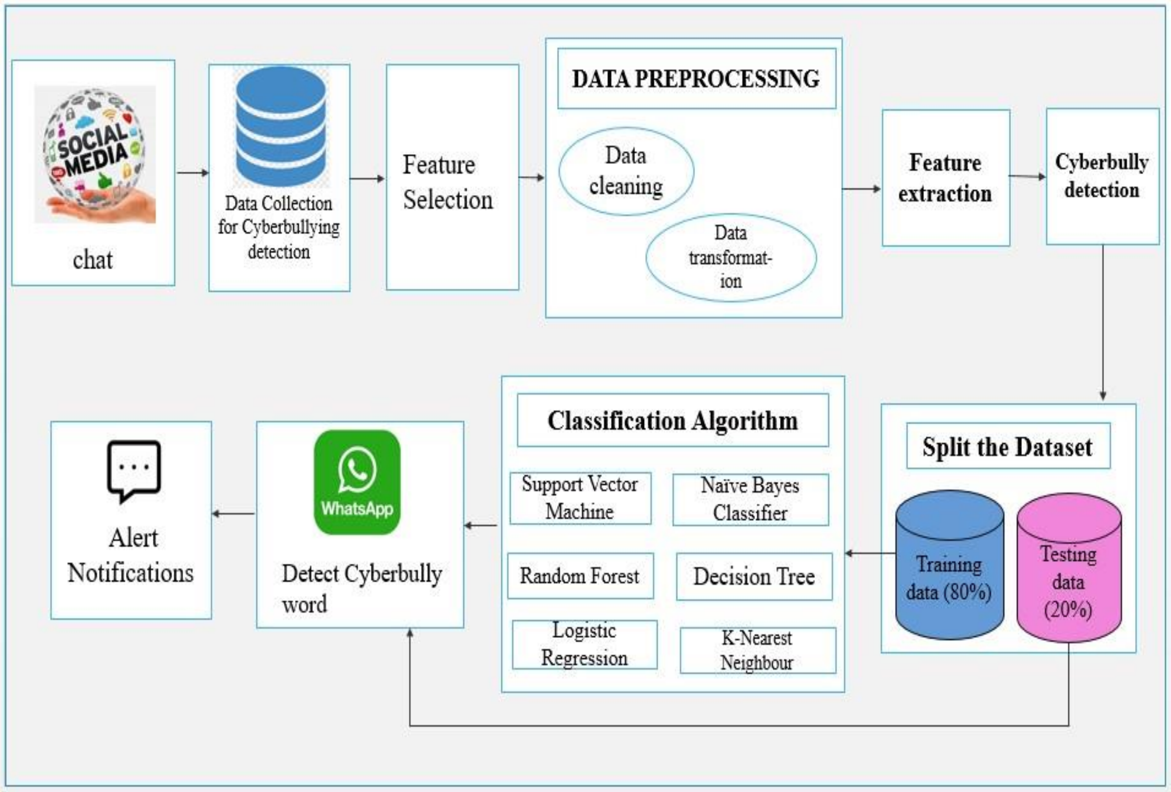
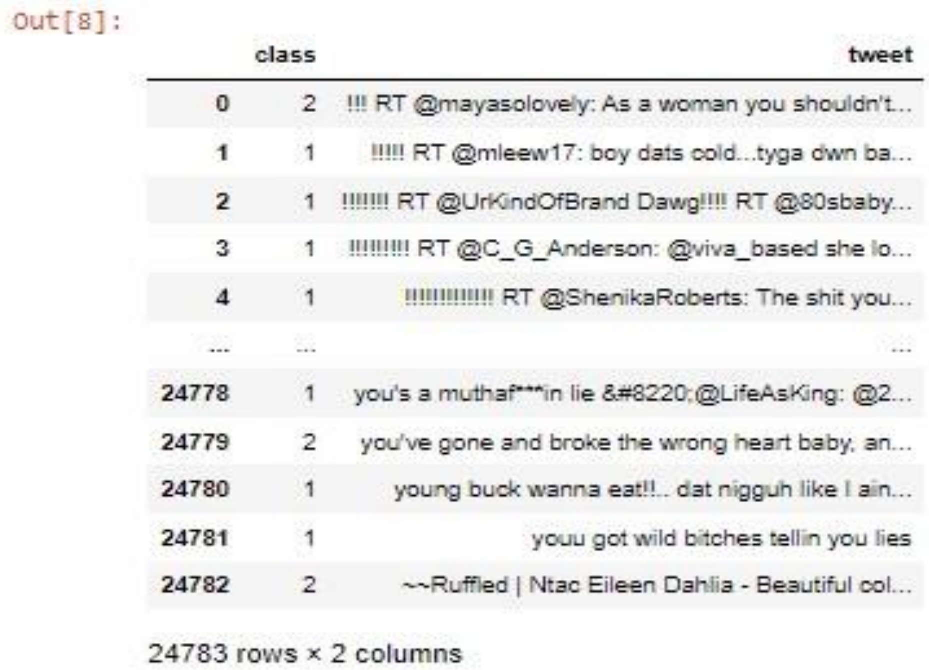


Figure 2: Proposed System Framework

Dataset Description

Cyberbullying also known as online bullying. I focused on detecting cyberbullying from twitter tweets. For this study the data set is took from website called kaggle.com. The downloaded dataset contains three classification that are hate speech, offensive language and normal.

The target is to find all the bullying text. The tweets are randomly taken 24783 tweets. These are stored as CSV file. HATE SPEECH: It will threaten some people mainly in religion and sexual.The next step is to select the required field for our project. In the dataset there are 7 columns and 24783 tweets. [4] For my project I need only two fields and they are class and tweets. The figure 3 represents after feature selection.



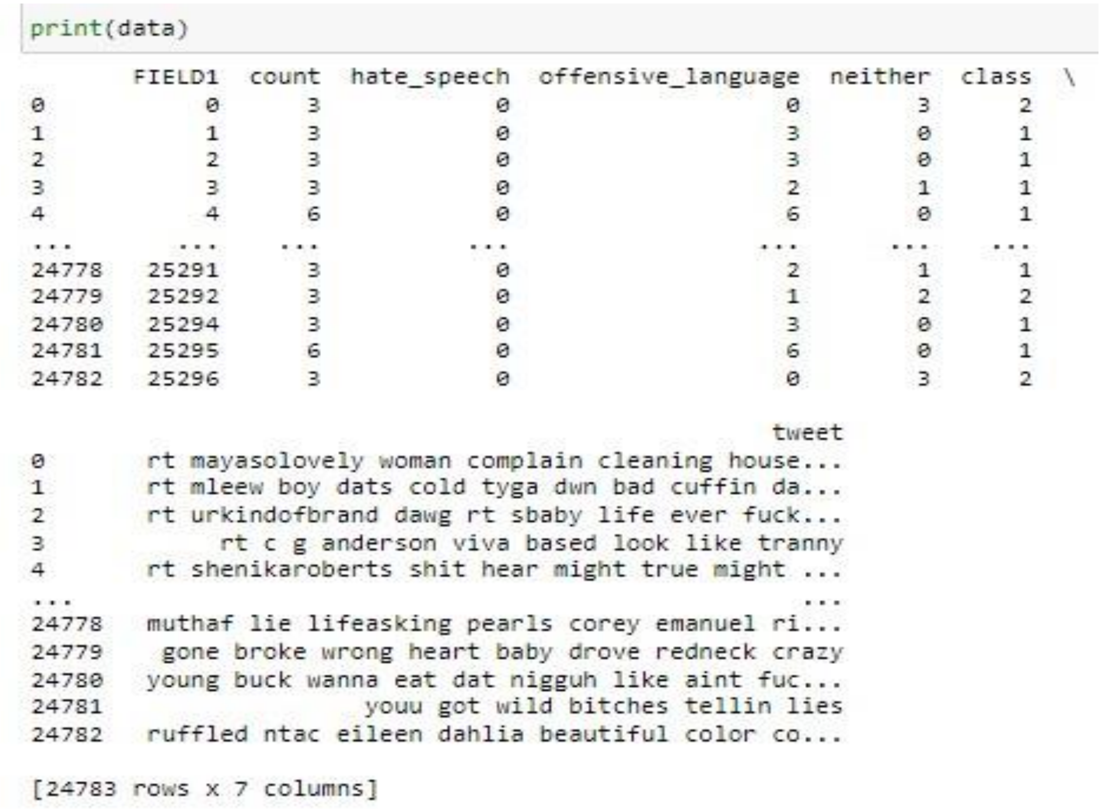
```
Out[8]:
       class                                              tweet
    0      2    !!! RT @mayasolovely: As a woman you shouldn't...
    1      1    !!!!! RT @mleew17: boy dats cold...tyga dwn ba...
    2      1   !!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby...
    3      1   !!!!!!!!! RT @C_G_Anderson: @viva_based she lo...
    4      1    !!!!!!!!!!!!! RT @ShenikaRoberts: The shit you...
  ...    ...                                                ...
24778      1  you's a muthaf***in lie &#8220;@LifeAsKing: @2...
24779      2    you've gone and broke the wrong heart baby, an...
24780      1    young buck wanna eat!!.. dat nigguh like I ain...
24781      1               youu got wild bitches tellin you lies
24782      2      ~~Ruffled | Ntac Eileen Dahlia - Beautiful col...

24783 rows × 2 columns
```

Figure 3: After Feature Selection

The next phase is to pre-process the dataset. As the tweets are collected from twitter was in not good format for the detection. It needs to be pre-processed for the implementation for detecting the cyberbullying. The processes are data cleaning, data transformation. In this process, it contains case conversion, removing special characters, removing short-hands, removing stop words, removing links, removing accents, and removing spaces are the following methods are done in data pre-processing. Case conversion is converting the whole dataset, all capital letters into small letters.

Removing special character's, theyare spaces, exclamation, double quote, hashtag, dollar sign, percent, ampersand, single quote, all parenthesis, asterisk, slash etc. Removing short-hands, for example ain't, aren't, cant't. Removing stop words for example a, is, the, are. Removing accents for example café',naive. Now it is suitable for the next Process. After all the process is completed, the data is pre-processed. Now it is structured form for execution. The figure 4 represents after pre-processed the dataset.



```
print(data)

       FIELD1  count  hate_speech  offensive_language  neither  class  \
0           0      3            0                   0        3      2
1           1      3            0                   3        0      1
2           2      3            0                   3        0      1
3           3      3            0                   2        1      1
4           4      6            0                   6        0      1
...       ...    ...          ...                 ...      ...    ...
24778   25291      3            0                   2        1      1
24779   25292      3            0                   1        2      2
24780   25294      3            0                   3        0      1
24781   25295      6            0                   6        0      1
24782   25296      3            0                   0        3      2

                                                 tweet
0      rt mayasolovely woman complain cleaning house...
1      rt mleew boy dats cold tyga dwn bad cuffin da...
2      rt urkindofbrand dawg rt sbaby life ever fuck...
3           rt c g anderson viva based look like tranny
4      rt shenikaroberts shit hear might true might ...
...                                                ...
24778  muthaf lie lifeasking pearls corey emanuel ri...
24779    gone broke wrong heart baby drove redneck crazy
24780  young buck wanna eat dat nigguh like aint fuc...
24781               youu got wild bitches tellin lies
24782  ruffled ntac eileen dahlia beautiful color co...

[24783 rows x 7 columns]
```

Figure 4: After Data Pre-processed

The next phase is to apply natural language processing (NLP). The feature extraction used for this cyberbullying detection are bag of words and time frequency-inverse document frequency. The machine learning classification algorithms are not going to work directly with texts. So, we must convert them into some other form like numbers or vectors before applying machine learning algorithm to them. In this way the data will converted by bow so that it can be ready to use in next round [8]. Bag of words is clear and accessible particularly for text data.

BOW is a text data rendering that describes word occurrences in the dataset. It will keep an eye on word counts, dismiss grammatical section and word categorize is known as BOW since any information regarding order of words in the data is abandon. The representation is just worried with known words, take place in the document. The drawback with text is jumbled and unshaped, but machine learning prefers only structured data.

By adapting these techniques, it will transform variable length toward fixed length vectors. The premier features to be considered is this. [9] Tf-idf is numerical measure to know the significance that a word brings in a document. The words in text set are reconstruct into text vectorization task. The easiest way to find Time frequency is rough calculation of word show itself in dataset. TF (e,f) = n(e,f)/ n(e,f). n(e,f) – number of times, word occur in documents.

n(e,f) – total number of words in dataset. The idf of text over set of data. This shows rare words in dataset. IDF = 1+log(N/dN). N – total no of documents in dataset. dN – total no of documents in which nth text occur. TF-IDF=TF*IDF.Then build the models by using distinct machine learning methods such as Decision Tree, Naïve Bayes, Random Forest, SVM, KNN, and Logistic regression.

RESULT AND DISCUSSION

The hardware requisite may supply for implementation of the system and there should be a fulfilled and dependable specifications of whole project system. PROCESSOR: INTEL CORE i7 (7th Gen). PROCESSOR SPEED: 3.00 GHZ. SSD: 240GB. RAM: 8GB. The software requirements are LANGUAGUE: PYTHON SOFTWARE: ANACONDA TOOL WITH JUPYTER NOTEBOOK. After applying six Classification algorithms from supervised machine learning, now it is time to find the accuracy, performance metrics are precision, recall, f1 score & support for six classification algorithms.

The data partition of training as 80% and testing as 20% using Time Frequency-Inverse Document Frequency gives better accuracy as 90.98% than others. Now, itexamines the SVM classifier's performance to determine how many inaccurate predictions it makes in comparison to the classifier for logistic regression. It must import confusion matrix function from package of Sklearntogenerate confusion matrix. Then, it will use a new variable called cm to call the function when it has been imported. Two parameters are required by the function, primarily y true (the actual numbers) and y pred. The figure 5 represents the Confusion matrix for SVM.

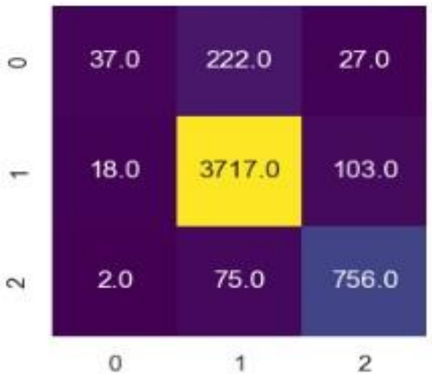
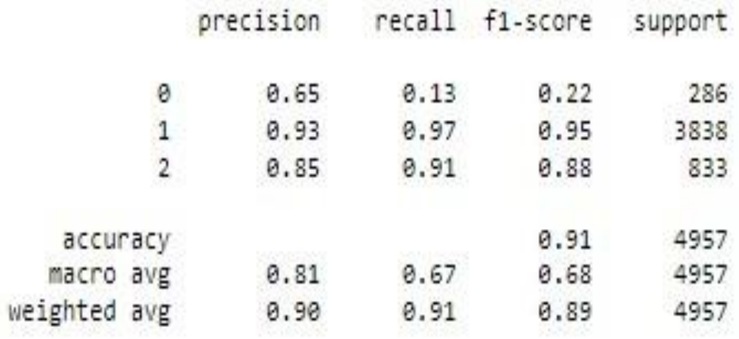


Figure 5: Confusion Matrix for SVM

Now calculating the classification report for implemented algorithms. The classification report contains precision, recall, f1 score & support by importing classification report library and implementing technique. The figure 6 represents the classification report for Support Vector Machine.



```
              precision    recall  f1-score   support

           0       0.65      0.13      0.22       286
           1       0.93      0.97      0.95      3838
           2       0.85      0.91      0.88       833

    accuracy                           0.91      4957
   macro avg       0.81      0.67      0.68      4957
weighted avg       0.90      0.91      0.89      4957
```

Figure 6: Classification Report for SVM

Table I: Performance Metrics Using Tf-Idf, Training (80%), Testing (20%)

Algorithms	Accuracy		Precision	Recall	F1 score	Support
SVM	90.98	0	0.65	0.13	0.22	286
		1	0.93	0.97	0.95	3838
		2	0.85	0.91	0.88	833
RF	89.28	0	0.55	0.11	0.18	286
		1	0.90	0.97	0.94	3838
		2	0.86	0.80	0.83	833
DT	79.40	0	0.32	0.22	0.26	286
		1	0.93	0.94	0.93	3838
		2	0.84	0.87	0.86	833
NB	88.82	0	0.00	0.00	0.00	286
		1	0.79	1.00	0.88	3838
		2	0.94	0.12	0.22	833
LR	90.03	0	0.73	0.12	0.20	286
		1	0.91	0.98	0.94	3838
		2	0.87	0.82	0.84	833
KNN	80.87	0	0.74	0.10	0.17	286
		1	0.81	0.99	0.89	3838
		2	0.88	0..21	0.34	833

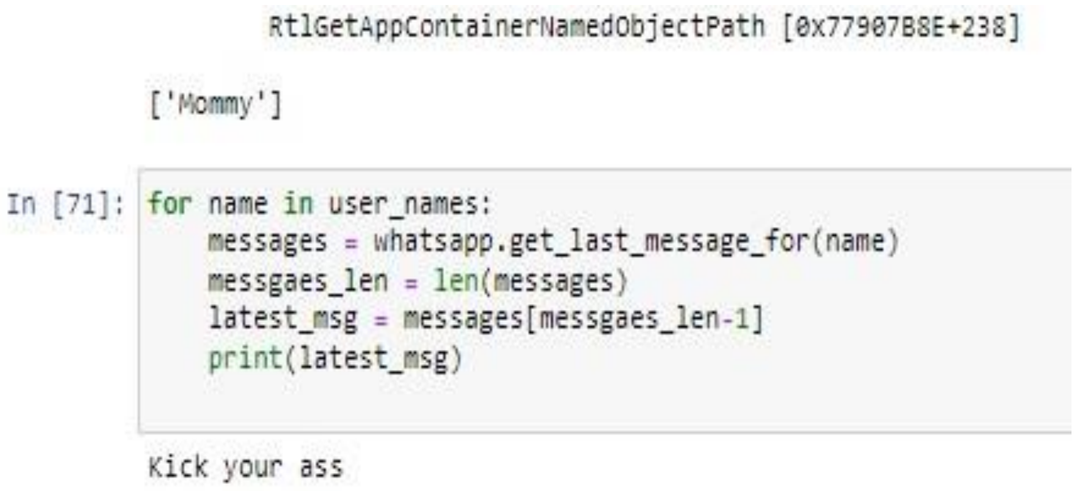
Table II: Parameters

Algorithm	Data Partition	Accuracy using TF-IDF	Accuracy using bag of words
SVM	70-30	90.92	89.87
	50-50	90.14	89.55
	67-33	90.74	89.92
	60-40	90.52	89.86
RF	70-30	89.17	88.74
	50-50	88.67	87.62
	67-33	88.60	88.05
	60-40	88.84	88.04
DT	70-30	88.72	88.97
	50-50	89.02	88.99
	67-33	88.91	89.03
	60-40	88.53	89.10
NB	70-30	79.20	87.10
	50-50	78.34	86.11
	67-33	79.11	87.05
	60-40	78.76	86.59
LR	70-30	89.75	90.65
	50-50	88.67	90.17
	67-33	89.76	90.52
	60-40	89.31	90.39
KNN	70-30	80.83	80.48
	50-50	80.38	79.77
	67-33	80.86	80.20
	60-40	80.72	79.80

The datais splitted into 5 partitions. They are Training and Testing as (1) 80 & 20%.(2) 70% & 30%. (3) 67% & 33%. (4) 60% & 40%. (5) 50% & 50% and used Feature extraction as bow and time frequency idf& founded accuracy for comparison is shown in table II.

Detecting Cyberbully Word

After applying six classification algorithms from supervised machine learning and predicting the accuracy, performance analysis and Testing. I connected my project to WhatsApp via web. It will send the message to the registered number. And it will go through the last message from any person, it will check whether there is a hate speech, offensive language. If its present means it will find the word and it will send SMS alert to the registered number with date, time, contact number, the abusive words, and its type of word (hate speech, offensive language).The Figure 7 represents Detection of cyberbully word.



```
RtlGetAppContainerNamedObjectPath [0x77907B8E+238]

['Mommy']

In [71]: for name in user_names:
             messages = whatsapp.get_last_message_for(name)
             messgaes_len = len(messages)
             latest_msg = messages[messgaes_len-1]
             print(latest_msg)

Kick your ass
```

Figure 7: Detect cyber bully word from Software

CONCLUSION

The current research is mainly utilized to explore the cyberbully detection using supervised machine learning techniques. Cyberbullying is not light matter. It should be considered as a serious problem because it has more victims.

In addition to it disturbs a people's mind. Most of the people are depressed after they got cyberbullied. This project successfully detected the cyberbully word from WhatsApp web and even sent the number of the bully to the registered number.

Parents not only parents, who take care of children can easily observe their child's movement using this system and therefore this system will be support for kids who are harassed in WhatsApp. Hence it is necessary to take the precautions to protect our kids from the bullies.

The current proposed system recognizes harassment texts from WhatsApp web alone. In future we can deploy it into an application which can read the chat automatically and detect if any abusive content is found in the chat. We can also develop our application to run on other social media networks like twitter, Facebook, Instagram etc.

This model will be really useful for kids and teenagers who need to be under the surveillance of their parents. The cyber bully detection system allows parents to monitor their child's online activity indefinitely.

References

1) Alam, K. S., Bhowmik, S., &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.

2) Islam, M. M., Uddin, M. A., Islam, L., Akter, A., Sharmin, S., &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.

3) Fire, M., Goldschmidt, R., &Elovici, Y. (2014). Online social networks: threats and solutions. IEEE Communications Surveys & Tutorials, 16(4), 2019-2036.

4) Reynolds, K., Kontostathis, A., & Edwards, L. (2011, December). Using machine learning to detect cyberbullying. In 2011 10th International Conference on Machine learning and applications and workshops (Vol. 2, pp. 241-244). IEEE.

5) Ali, A., & Syed, A. M. (2020). Cyberbullying detection using machine learning. Pakistan Journal of Engineering and Technology, 3(2), 45-50.

6) Ali, A., & Syed, A. M. (2020). Cyberbullying detection using machine learning. Pakistan Journal of Engineering and Technology, 3(2), 45-50.

7) Bayari, R., &Bensefia, A. (2021). Text mining techniques for cyberbullying detection: state of the art. Adv. Sci. Technol. Eng. Syst. J, 6, 783-790.

8) Talpur, B. A., & O'Sullivan, D. (2020). Cyberbullying severity detection: A machine learning approach. PloS one, 15(10), e0240924.

9) Galán-García, P., Puerta, J. G. D. L., Gómez, C. L., Santos, I., & Bringas, P. G. (2016). Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying. Logic Journal of the IGPL, 24(1), 42-53.

10) Kanan, T., Aldaaja, A., &Hawashin, B. (2020). Cyber-bullying and cyber-harassment detection using supervised machine learning techniques in Arabic social media contents. Journal of Internet Technology, 21(5), 1409-1421.

11) Kumar, R. (2021). Detection of Cyberbullying using Machine Learning. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(9), 656-661.

12) Wade, S., Parulekar, M., & Wasnik, K. (2020). Cyber Bullying Detection on Twitter Mining. https://www.jetir.org/papers/JETIR2108196.pdf

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

14) Kargutkar, S., & Chitre, V. Implementation of Cyberbullying Detection using Machine Learning Techniques.

15) Prabowo, W. A., & Azizah, F. (2020). Sentiment analysis for detecting cyberbullying using TF-IDF and SVM. Jurnal RESTI (RekayasaSistem dan TeknologiInformasi), 4(6), 1142-1148.

16) Atoum, J. O. (2020, December). Cyberbullying Detection Through Sentiment Analysis. In 2020 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 292-297). IEEE.

17) Shah, K., Phadtare, C., &Rajpara, K. Cyber-Bullying Detection in Hinglish Languages Using Machine Learning.https://eudl.eu/pdf/10.4108/eai.7-12-2021.2314577

18) Raj, C., Agarwal, A., Bharathy, G., Narayan, B., & Prasad, M. (2021). Cyberbullying Detection: Hybrid Models Based on Machine Learning and Natural Language Processing Techniques. Electronics 2021, 10, 2810.

19) Bhardwaj, A., Bhardwaj, H., Sakalle, A., Uddin, Z., Sakalle, M., & Ibrahim, W. (2022). Tree-based and machine learning algorithm analysis for breast cancer classification. Computational Intelligence and Neuroscience, 2022.

20) Islam, M. M., Uddin, M. A., Islam, L., Akter, A., Sharmin, S., &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.

21) https://www.javatpoint.com/machine-learning

22) https://www.stopbullying.gov/

23) Cuzcano, X. M., & Ayma, V. H. (2020). A comparison of classification models to detect cyberbullying in the Peruvian Spanish language on twitter. International Journal of Advanced Computer Science and Applications, 11(10).

24) Alakrot, A., Murray, L., & Nikolov, N. S. (2018). Towards accurate detection of offensive language in online communication in Arabic. Procedia computer science, 142, 315-320.

25) Kaur, S., Singh, S., & Kaushal, S. (2021). Abusive content detection in online usergenerated data: a survey. Procedia Computer Science, 189, 274-281.

26) Foody, M., Samara, M., &Carlbring, P. (2015). A review of cyberbullying and suggestions for online psychological therapy. Internet Interventions, 2(3), 235-242.

27) Mawardah, C. K. A., Normala, R., Azlini, C., Kamal, M. Y., & Lukman, Z. M. (2018). The Factors of Cyber Bullying and the Effects on Cyber Victims. Int. J. Res. Innov. Soc. Sci, 2, 59-61.

28) Elsafoury, F., Katsigiannis, S., Wilson, S. R., & Ramzan, N. (2021, July). Does BERT pay attention to cyberbullying?. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1900-1904).