Original Research Article

LOGISTIC REGRESSION BASED CYBER HARASSMENT IDENTIFICATION

ABSTRACT

Increased online use and allowing users to engage with groups such as digital networking have contributed to the growth of hacking. Online abuse is a new type of harassment that has lately become more prevalent as online communities have grown in popularity. It tends to send messages which included defamatory claims or vocally harassing someone while in the internet group. Only if modern civilization recognizes harassment as it truly is, countless of hidden sufferers may continue to suffer. There have been several studies on cyber bullying, but none of them have been able to offer a solid remedy. By creating a model that can recognize and block bullying-related incoming and outgoing communications, we address this issue in our project. By employing supervised classification techniques on an open source dataset that has been carefully annotated, we hope to provide lexical baselines for this job. We have employed a logistic regression classifier for training and identifying instances of bullying behaviors.The dataset we used is a twitter dataset collected from kaggle.Our model classifies a message whether it's bullying or not.

Keywords: Machine Learning, cyber harassment, logistic regression, digital networking.

1.INTRODUCTION

A collection of Web 2.0-based programmes called SOCIAL Media make it possible to create and share user-generated content. These are all Internet-based applications. People may take use of social media to gain access to a wealth of knowledge, easy communication, etc.Cyber bullying is the term used to describe aggressive, deliberate acts committed by a person or group of individuals against a victim using digital communication channels like sending messages and leaving comments online.

Speech that is intended to stir up hatred for a specific group—a community, a religion, or a race—is referred to as hate speech. These assertions could or might not be true, but it is probable that they will result in violence. Worldwide increases in violence against minorities, such as lynchings, mass shootings, and ethnic cleansing, have been connected to hate speech on the Internet.

Simple word filters do not adequately address this issue, necessitating natural language processing that focuses on this symptom: What constitutes hate speech can be influenced by factors. The model is trained using the Tweeter dataset from Kaggle. We must initially use a single categorization algorithm to move further with these datasets. We utilized the 0-1 predictor to determine if the text contains cyber bullying material or not. This creates a binary space in which we can train our model and exclude out any grey possibilities. In order to properly classify data, it must first be cleaned of symbols, spacy tokenizer Addresses, mails, line breaks, spaces, digits, commas, separating, and individual characters. Together with an incisive analysis of some published research on methods for detecting cyber bullying, this study offers a thorough and organized overview of robotic incitement identification and examines a few of the existing methodologies.

2.LITERATURE REVIEW

- P. K. Roy, et al. [7] information on how to post a petition of bigotry on Facebook using an assistance from a deep neural convolutional network. With the help of machine learning algorithms, tweets containing hate speech have been foundUtilizing the tf procedure, functionalities on Facebook have now been removed. The best ml model is SVM, howeverIn a 4:1 sample used to evaluate trained predictions, it was capable of forecasting 53% of racial hatred messages.
- N. Tsapatsoulis, et al. [5] comprehensive analysis in newly implemented harassment on Facebook. Moreover, The significance of recognizing the numerous Facebook offenders is discussed. According to the Research report, there are a number of concrete measures that must be taken in order to construct a useful and successful software for detecting Online activity. I use characteristic types, ml models, and knowledge categorization and data logging.
- R. R. Dalvi, et al. [3] proposes the process forachieving them identify & stop digital abuse controlled ml techniques employing identified on Facebook. Inside this experiment, texts and sample sizes are compiled using the realtimeApi. The suggested model evaluates SVM and Bayesian Network on the gathered data sets. Use the TFIDF vectorizer to delete a feature. The findings demonstrate the accuracy of a model for internet abuse constructed using Vector

Assist. In comparison to Naïve Bayes classifier, the computer performs around 73.34% superior.

3.PROPOSED SYSTEM

Although social networking sites and online chat services give users a place to share their skills and information, they are seldom used to threaten other users with cyberharassment, which makes it difficult to use these services. In this research, we created a strategy for supervised learning to identify cyberharassment. The logistic regression approach It's utilized to test a algorithm for ml on the Random subset and improve it, which has been gathered with features and labels.

Online abuse/ bullying detection is a growing area of research that aims to automatically detect instances of cyberbullying in online interactions. These calculations are a widely used Stats model which can be applied to cyberbullying detection. The scope of cyberbullying detection using logistic regression is large and involves several stages of Preparing the data, choosing the features, building the model, testing it, and deploying it: In this work, a message can be detected whether it is hate speech or not.

This proposed scheme is an early version of a cyberbullying detection system that can be attached to social networking sites to clearly detect and keep track of cyberbullying.

Data collection: The program would gather information from websites and social networking sites like Youtube, Google, and Pinterest. Text, picture, and video data types are all possible. This collected data would be processed as follows

1. Pre-processing of data: The collected data will be cleaned, normalized, and pre-processed to remove irrelevant information and ensure that it is in a format suitable for analysis.

2. Feature extraction: relevant features are extracted from the preprocessed data. These features may include linguistic features such as the use of profanity, hate speech, and aggressive language, and behavioral features such as the frequency and timing of online interactions.

3. Feature selection: The most relevant features are selected for training the logistic regression

model. In this step, the features that have the greatest impact on the outcome variable, i.e., whether an online interaction is cyberbullying or not, is identified.

4. Model training: the logistic regression gathering all the necessary chosen by parameters is used to train the model. To discover the link between both the attribute values and the output vector, the computer must be trained.
5. Evaluation of this model: the performance of the logistic regression model is evaluated using parameters including highest accuracy, recollection, & accuracy. These metrics help determine the effectiveness of the model in correctly identifying cases of cyber bullying.

6. Deployment: once trained and evaluated, the logistic regression model can be deployed to identify cases of cyber bullying in real time. The model can be integrated into online platforms and social media networks to identify and flag cases of cyber bullying.

The proposed system for detecting cyber bullying using logistic regression would include collecting and preprocessing data, extracting and selecting relevant features, training and evaluating the model, and using the model to detect cyber bullying cases. The system can be a valuable tool for preventing and mitigating the harmful effects of cyber bullying.

4.METHODOLOGY

ML

Without being expressly designed, ml algorithms may acquire data and utilize it to learn on their own. So how exactly does the ml method operate? just by looking at the numbers.The three main categories of ml algorithms are: - Supervised machine learning - Task-oriented (classification and regression)

- ML without supervision - Driven by data (clustering)

- Strengthening computer learning - taking lessons from errors (reward or punishment)monitoring ml

In supervised learning, models are prepared on labeled data sets, where each input category's characteristics are taught to the algorithm. Two categories of supervised machine learning are recognized.

Logistic Regression

A machine learning technique for addressing categorization issues is logistical regression. It is a probability-based methodology of predictive analysis. The classification procedure of a binary number is used as the dependent variable in logistic regression.Establishing a connection between qualities and the similarities of a particular event is the goal of regression model. As example, based on the amount of time spent analyzing, the data set may be used to predict whether a child will succeed or not in a test.Many individuals are unclear about whether logistic regression falls under the classification or regression categories. Linear methods cannot perfectly depict it since it may have a value more than one or a little less than zero, which is unlikely depending on the regression analysis.

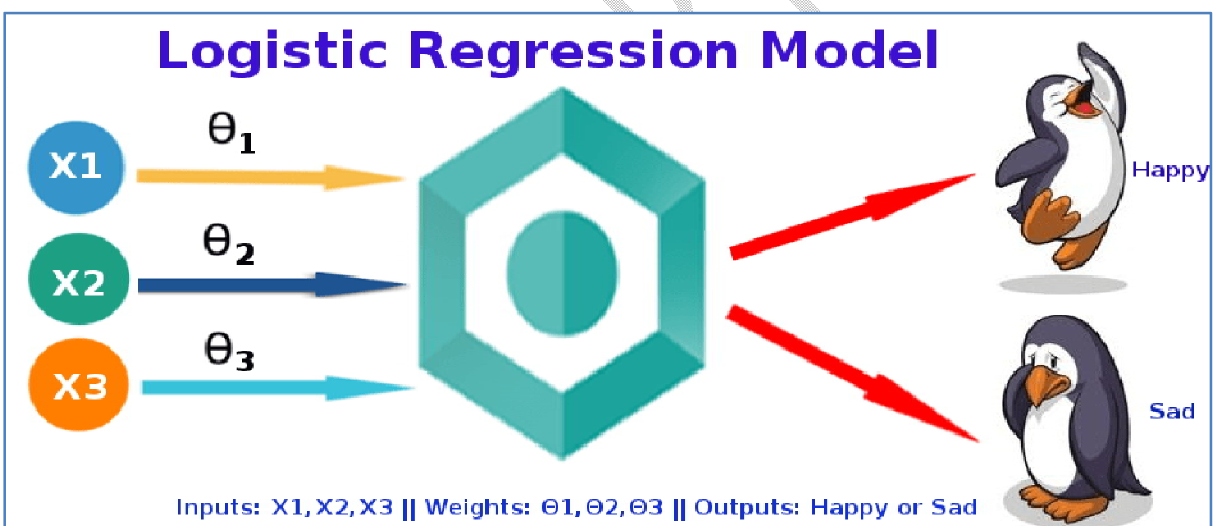

Figure 1: Logistic Regression Model

LR (LOGISTICAL REGRESSION) Equation:

- We understand that a perfect line's formula is:
  $y = b_0 + b_1x_1 + b_2x_2 + \dots\dots + b_nx_n$

- We split this preceding expression with (1-y) since the range of y in regression analysis is limited Zero to one:

$$\frac{y}{1-y};0\ for\ y\ =\ 0,\ and\ infinity\ for\ y\ =\ 1$$

- But since We need a band of length -[infinite] and +[nothingness], we need to take the equation's exponential, which is as follows:

$$\log\left[\frac{y}{1-y}\right] = b_0\ +\ b_1x_1\ +\ b_2x_2\ +\ \dots\dots + b_nx_n$$

The final the aforementioned expression is used in regression analysis.

Type of logistical regression:

Depending on the categories, there are three different forms of logistic regression:

o **Binomial Regression:**Just two different forms of the dependant can exist in binomial logistic regression,Variable, e.g., zero or one, etc.

o **Multinomial Regression:**Sometimes there may be 3or large distinct categories which are not in order in multinomial logistic regression., e.g., "cat," "dogs," and "sheep"

o **Ordinal Regression:**Sometimes there may be 3 or greater potential which are in a proper order categories belongs the dependant in ordinal logistical regression. Variable, e.g., "low," "medium,"/ "high."

Prerequisites for a functioning logistical regression

Any data sets may be used with this model, however if you want high performance, you still need to take some assumptions into account.

1. Binomial logistic regression requires a binary dependent variable.
2. It is best to include just the pertinent variables.
3. The considered variable must not be connected. That is, the model's co integration must be either non-existent or extremely low.

Decision Threshold - Logistic Regression

If the up to the optimum is less than 0.5, the pupil is considered to have passed; if not, they are recorded as failing. Clustering process come in both linearly & non-linear varieties. You can raise the cubic order to get a critical decision border.

4.IMPLEMENTATION

4.1 Database Setup

To complete the process of trend analysis, data sets must be accessible. The amount and quality of the dataset affect a classifier's performance. Online databases offer a number of datasets. Through Kaggle, we downloaded the Face book dataset. The dataset must be divided into training and testing datasets after it has been produced. It has three properties: the id, label (either 0 or 1), and tweet message.

id	label	tweet
1	0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run
2	0	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthanked
3	0	bihday your majesty
4	0	#model i love u take with u all the time in urÃ°ÂŸÂ“Â±!!! Ã°ÂŸÂ˜Â™Ã°ÂŸÂ˜ÂŽÃ°ÂŸÂ‘Â„Ã°ÂŸÂ‘Â…Ã°ÂŸÂ’Â¦Ã°ÂŸÂ’Â¦Ã°ÂŸÂ’Â¦
5	0	factsguide: society now #motivation
6	0	[2/2] huge fan fare and big talking before they leave. chaos and pay disputes when they get there. #allshowandnogo
7	0	@user camping tomorrow @user @user @user @user @user @user @user dannyÃ¢Â€Â¦
8	0	the next school year is the year for exams.Ã°ÂŸÂ˜Â¯ can't think about that Ã°ÂŸÂ˜Â #school #exams #hate #imagine #actorslife #revolutionschool #girl
9	0	we won!!! love the land!!! #allin #cavs #champions #cleveland #clevelandcavaliers Ã¢Â€Â¦
10	0	@user @user welcome here ! i'm it's so #gr8 !
11	0	Ã¢Â†Â #ireland consumer price index (mom) climbed from previous 0.2% to 0.5% in may #blog #silver #gold #forex
12	0	we are so selfish. #orlando #standwithorlando #pulseshooting #orlandoshooting #biggerproblems #selfish #heabreaking #values #love #
13	0	i get to see my daddy today!! #80days #gettingfed
14	1	@user #cnn calls #michigan middle school 'build the wall' chant '' #tcot
15	1	no comment! in #australia #opkillingbay #seashepherd #helpcovedolphins #thecove #helpcovedolphins
16	0	ouch...junior is angryÃ°ÂŸÂ˜Â#got7 #junior #yugyoem #omg
17	0	i am thankful for having a paner. #thankful #positive
18	1	retweet if you agree!
19	0	its #friday! Ã°ÂŸÂ˜Â€ smiles all around via ig user: @user #cookies make people
20	0	as we all know, essential oils are not made of chemicals.
21	0	#euro2016 people blaming ha for conceded goal was it fat rooney who gave away free kick knowing bale can hit them from there.
22	0	sad little dude.. #badday #coneofshame #cats #pissed #funny #laughs
23	0	product of the day: happy man #wine tool who's it's the #weekend? time to open up & drink up!
24	1	@user @user lumpy says i am a . prove it lumpy.
25	0	@user #tgif #ff to my #gamedev #indiedev #indiegamedev #squad! @user @user @user @user @user
26	0	beautiful sign by vendor 80 for $45.00!! #upsideofflorida #shopalyssas #love

Fig .2**Database Setup**

4.1 Implementation Steps

- Importing the required Libraries
- Loading Dataset
- Data Pre-processing
- Distribution of Sentiments
- Splitting of preprocessed Data

- ➢ Building Logistic Regression Model
- ➢ Evaluating Logistic Regression Model

Importing the required Libraries

We need to import numpy,pandas,matplotlib,nltk,sklearn to use them while building the model.

```
In [11]: import pandas as pd
         import numpy as np
         import re
         import nltk
         import seaborn as sns
         import matplotlib.pyplot as plt
         from matplotlib import style
         style.use('ggplot')
         from nltk.tokenize import word_tokenize
         from nltk.corpus import wordnet as wn
         from nltk.stem.wordnet import WordNetLemmatizer
         from nltk.stem import WordNetLemmatizer
         from nltk.corpus import stopwords
         stop_words = set(stopwords.words('english'))
         from wordcloud import WordCloud
         from sklearn.feature_extraction.text import TfidfVectorizer
         from sklearn.model_selection import train_test_split
         from sklearn.linear_model import LogisticRegression
         from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, ConfusionMatrixDisplay
```

Fig .3Importing the required Libraries

Loading Dataset

Twitter dataset is loaded in form of dataframe.

```
In [12]: tweet_df = pd.read_csv('cyberbullydetection_train.csv')

In [13]: tweet_df.head()

Out[13]:
```

	id	label	tweet
0	1	0	@user when a father is dysfunctional and is s...
1	2	0	@user @user thanks for #lyft credit i can't us...
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in ...
4	5	0	factsguide: society now #motivation

Fig.4 DatasetLoading

Data Preprocessing

```
In [16]: #creating a function to process the data
         def data_processing(tweet):
             tweet = tweet.lower()
             tweet = re.sub(r"https\S+|www\S+http\S+", '', tweet, flags = re.MULTILINE)
             tweet = re.sub(r'\@w+|\#','', tweet)
             tweet = re.sub(r'[^\w\s]','',tweet)
             tweet = re.sub(r'ð','',tweet)
             tweet_tokens = word_tokenize(tweet)
             filtered_tweets = [w for w in tweet_tokens if not w in stop_words]
             return " ".join(filtered_tweets)

In [17]: tweet_df.tweet = tweet_df['tweet'].apply(data_processing)

In [18]: tweet_df = tweet_df.drop_duplicates('tweet')

In [21]: lemmatizer = WordNetLemmatizer()
         def lemmatizing(data):
             tweet = [lemmatizer.lemmatize(word) for word in data]
             return data

In [23]: tweet_df['tweet'] = tweet_df['tweet'].apply(lambda x: lemmatizing(x))
```

Fig.5 Data Preprocessing

Distribution of Sentiments

```
In [26]: tweet_df['label'].value_counts()
Out[26]: 0    27352
         1     1993
         Name: label, dtype: int64

In [27]: fig = plt.figure(figsize=(5,5))
         sns.countplot(x='label', data = tweet_df)
Out[27]: <AxesSubplot:xlabel='label', ylabel='count'>
```

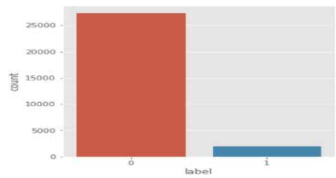

Fig.6 Distribution of Sentiments

```
In [28]: fig = plt.figure(figsize=(7,7))
         colors = ("red", "gold")
         wp = {'linewidth':2, 'edgecolor':"black"}
         tags = tweet_df['label'].value_counts()
         explode = (0.1, 0.1)
         tags.plot(kind='pie',autopct = '%1.1f%%', shadow=True, colors = colors, startangle =90,
                  wedgeprops = wp, explode = explode, label='')
         plt.title('Distribution of sentiments')
Out[28]: Text(0.5, 1.0, 'Distribution of sentiments')
```

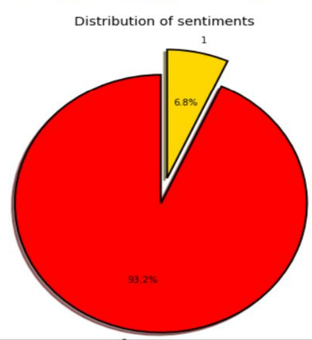

Fig.7. Twitter dataset

5. RESULTS AND ANALYSIS

Most frequent words in non hate speech:

```
In [29]: non_hate_tweets = tweet_df[tweet_df.label == 0]
         non_hate_tweets.head()
```

Out[29]:

	id	label	tweet
0	1	0	user father dysfunctional selfish drags kids d...
1	2	0	user user thanks lyft credit cant use cause do...
2	3	0	bihday majesty
3	4	0	model love u take u time ur
4	5	0	factsguide society motivation

```
In [30]: text = ' '.join([word for word in non_hate_tweets['tweet']])
         plt.figure(figsize=(20,15), facecolor='None')
         wordcloud = WordCloud(max_words=500, width=1600, height=800).generate(text)
         plt.imshow(wordcloud, interpolation='bilinear')
         plt.axis('off')
         plt.title('Most frequent words in non hate tweets', fontsize = 19)
         plt.show()
```

Fig.8 Tweet dataset 1

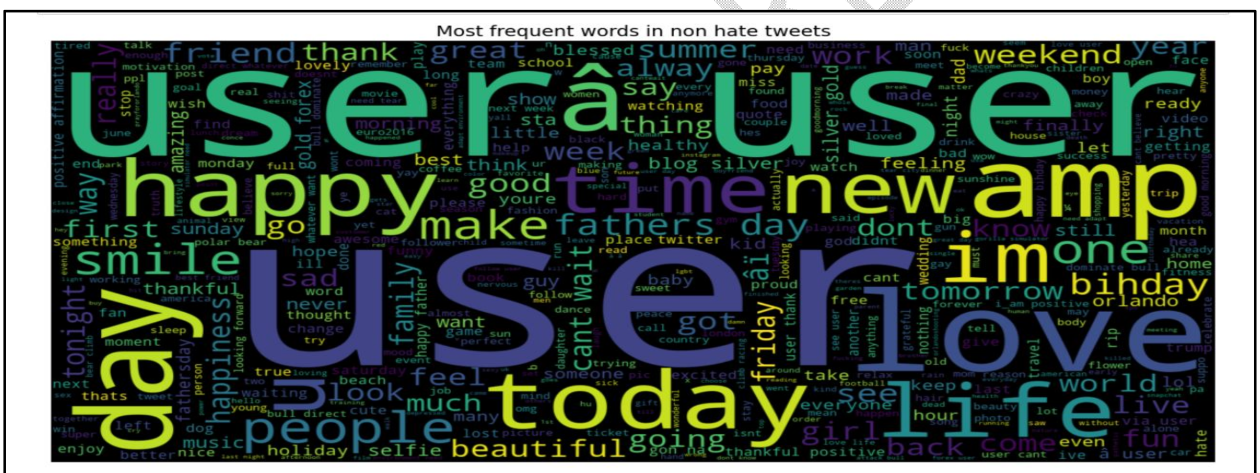

Fig.9 Most frequent words in non hate speech

Most frequent words in hate speech

```
In [31]: neg_tweets = tweet_df[tweet_df.label == 1]
         neg_tweets.head()
```

Out[31]:

	id	label	tweet
13	14	1	user cnn calls michigan middle school build wa...
14	15	1	comment australia opkillingbay seashepherd hel...
17	18	1	retweet agree
23	24	1	user user lumpy says prove lumpy
34	35	1	unbelievable 21st century wed need something l...

```
In [32]: text = ' '.join([word for word in neg_tweets['tweet']])
         plt.figure(figsize=(20,15), facecolor='None')
         wordcloud = WordCloud(max_words=500, width=1600, height=800).generate(text)
         plt.imshow(wordcloud, interpolation='bilinear')
         plt.axis('off')
         plt.title('Most frequent words in hate tweets', fontsize = 19)
         plt.show()
```

Fig.10 Tweet dataset 2

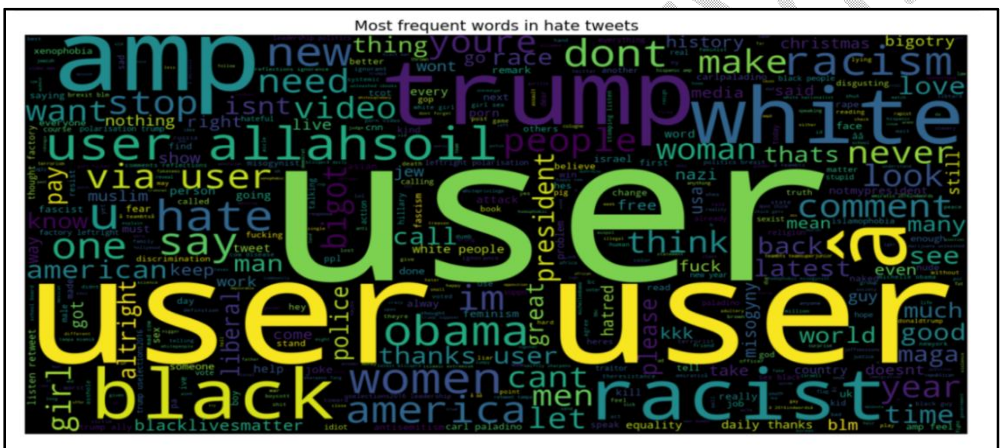

Fig.11 Most frequent words in hate speech

Splitting of preprocessed Data

```
In [37]: X = tweet_df['tweet']
         Y = tweet_df['label']
         X = vect.transform(X)

In [38]: x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=42)

In [39]: print("Size of x_train:", (x_train.shape))
         print("Size of y_train:", (y_train.shape))
         print("Size of x_test: ", (x_test.shape))
         print("Size of y_test: ", (y_test.shape))

         Size of x_train: (23476, 380305)
         Size of y_train: (23476,)
         Size of x_test:  (5869, 380305)
         Size of y_test:  (5869,)
```

Fig.12 Splitting of preprocessed Data

Building Logistic Regression Model

```
In [40]: logreg = LogisticRegression()
         logreg.fit(x_train, y_train)
         logreg_predict = logreg.predict(x_test)
         logreg_acc = accuracy_score(logreg_predict, y_test)
         print("Test accuarcy: {:.2f}%".format(logreg_acc*100))

         Test accuarcy: 93.17%
```

Fig.13 Building Logistic Regression Model

Evaluating Logistic Regression Model

```
In [41]: print(confusion_matrix(y_test, logreg_predict))
         print("\n")
         print(classification_report(y_test, logreg_predict))

         [[5458    0]
          [ 401   10]]


                       precision    recall  f1-score   support

                    0       0.93      1.00      0.96      5458
                    1       1.00      0.02      0.05       411

             accuracy                           0.93      5869
            macro avg       0.97      0.51      0.51      5869
         weighted avg       0.94      0.93      0.90      5869
```

Fig.14 Evaluating Logistic Regression Model

6.CONCLUSION

This research uses machine learning should address the issue of cyber abuse on the Social site. In the trials, ml methods that are controlled and uncontrolled had both been applied. Finding the appropriate set of keywords was found to be a crucial step for improving sentiment analysis outcomes. Findings show that our model performs quite well and might be employed to create practical monitoring applications to lessen the serious societal issue of cyber bullying.According to an experimental finding, the logistic regression-based approach performs better and achieves the best accuracy when user-specific data is added. Logistic Regression outperforms alternative classification algorithms for text classification because of the literature dataset's sequential reparability, scarce unimportant features, and large current environment.

7. REFERENCES

[1] J.W.PatchinandS.Hinduja,“BulliesmoveBeyondtheSchoolyard;aPreliminaryLookatCyberbullying,”YouthViolence and JuvenileJustice,vol. 4,no. 2, pp.148–169,2006.

[2] https://www.researchgate.net/publication/351131976_Cyberbullying_Detection_on_Social_Networks_Using_Machine_Learning_Approaches

[3] https://www.ijraset.com/best-journal/cyberbullying-detection-using-natural-language-processing

[4] N. E. Willard, Cyberbullying and Cyberthreats: Responding to the Challenge of Online Social Aggression, Threats, andDistress.ResearchPress, 2007.

[5] https://ieeexplore.ieee.org/document/8980379

[6] Dani Harsh, Jundong Li, and Huan Liu, “Sentiment Informed Cyberbullying Detection in Social Media” Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Spinger, Cham, 2017.

[7] Dinakar, Karthik, Roi Reichart, and Henry Lieberman. “Modeling the detection of Textual Cyberbullying.” The Social Mobile Web 11.02(2011):11-17

[8] K. Dinkar, R. Reichart and H. Liebernman, “Modeling the Detection of Textual Cyberbullying,” MIT. International.

[9] Huang, Qianjia, Vivek Kumar Singh, and Pradeep Kumar Atrey. ”Cyberbullying detection using social and textual analysis.” Proceedings of the 3rd International Workshop on Socially-Aware Multimedia. 2014.

[10] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, “Detection of harassment on web 2.0,” Proceedings of the Content Analysis in the WEB, vol. 2, pp. 1–7, 2009.

[11] T. Bin Abdur Rakib, L. K. Soon, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Springer Verlag, 2018), vol. 10751 LNAI, pp. 180–189.